Metabolic phenotyping of human plasma by H-1-NMR at high and medium magnetic field strengths: a case study for lung cancer

Peer-reviewed author version

LOUIS, Evelyne; Cantrelle, Francois-Xavier; MESOTTEN, Liesbet; REEKMANS, Gunter; BERVOETS, Liene; VANHOVE, Karolien; THOMEER, Michiel; Lippens, Guy & ADRIAENSENS, Peter (2017) Metabolic phenotyping of human plasma by H-1-NMR at high and medium magnetic field strengths: a case study for lung cancer. In: MAGNETIC RESONANCE IN CHEMISTRY, 55(8), p. 706-713.

DOI: 10.1002/mrc.4577
Handle: http://hdl.handle.net/1942/24320

Metabolic phenotyping of human plasma by ^{1}H-NMR at high and medium magnetic field strengths: a case study for lung cancer

Running head: High-field (900 MHz) vs. medium-field (400 MHz) NMR metabolomics

Louis Evelyne[1], Cantrelle Francois-Xavier[2], Mesotten Liesbet[1,3], Reekmans Gunter[4], Bervoets Liene[1], Vanhove Karolien[1,5], Thomeer Michiel[1,6], Lippens Guy[2,7], Adriaensens Peter[4§]

[§]Corresponding author: Prof. dr. Peter Adriaensens, Applied and Analytical Chemistry, Institute for Materials Research, Hasselt University, Agoralaan Building D, 3590 Diepenbeek, Belgium. T: +3211268396. F: +3211268299. E: peter.adriaensens@uhasselt.be

[1]Faculty of Medicine and Life Sciences, Hasselt University, 3500 Hasselt, Belgium

[2]CNRS UMR 8576, Unité de Glycobiologie Structurale et Fonctionnelle, Université des Sciences et Technologies de Lille 1, 59655 Villeneuve d'Ascq Cedex, France

[3]Department of Nuclear Medicine, Ziekenhuis Oost-Limburg, 3600 Genk, Belgium

[4]Applied and Analytical Chemistry, Institute for Materials Research, 3590 Diepenbeek, Belgium

[5]Department of Respiratory Medicine, Algemeen Ziekenhuis Vesalius, 3700 Tongeren, Belgium

[6]Department of Respiratory Medicine, Ziekenhuis Oost-Limburg, 3600 Genk, Belgium

[7]Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INSA, University of Toulouse, CNRS, INRA, 31077 Toulouse, France

Abstract

Accurate identification and quantification of human plasma metabolites can be challenging in crowded regions of the NMR spectrum with severe signal overlap. Therefore, this study describes metabolite spiking experiments on the basis of which the NMR spectrum can be rationally segmented into well-defined integration regions, and this for spectrometers having magnetic field strengths corresponding to ^{1}H resonance frequencies of 400 MHz and 900 MHz. Subsequently, the integration data of a case-control dataset of 69 lung cancer patients and 74 controls were used to train a multivariate statistical classification model for both field strengths. In this way, the advantages/disadvantages of high versus medium magnetic field strength were evaluated. The discriminative power obtained from the data collected at the two magnetic field strengths is rather similar, i.e. a sensitivity and specificity of respectively 90% and 97% for the 400 MHz data versus 88% and 96% for the 900 MHz data. This shows that a medium-field NMR spectrometer (400-600 MHz) is already sufficient to perform clinical metabolomics. However, the improved spectral resolution (reduced signal overlap) and signal-to-noise ratio of 900 MHz spectra yield more integration regions that represent a single metabolite. This will simplify the unraveling and understanding of the related, disease disturbed, biochemical pathways.

Key words: nuclear magnetic resonance spectroscopy, ^{1}H, magnetic field strength, metabolic phenotype, plasma, lung cancer

Abbreviations: AUC: area under the curve; MS: mass spectrometry; OPLS-DA: orthogonal partial least squares discriminant analysis; ppm: parts per million; S/N: signal-to-noise ratio; VAR: variable; VIP: variable importance for the projection.

Introduction

Metabolomics is a powerful discipline in which small molecule metabolites are determined in biological samples such as plasma or urine [1-3]. The two major high-throughput analytical platforms used for metabolite analysis are mass spectrometry (MS) and ^{1}H-NMR spectroscopy [4]. Subsequently, the large amount of resulting data is analyzed by multivariate pattern recognition methods in an effort to reduce the complexity of the data and to extract diagnostic information regarding disease and to identify disturbed biochemical pathways [5-7]. As compared to mass spectrometry (with or without a liquid chromatography step in advance), NMR spectroscopy offers excellent stability and integration accuracy (linear response between metabolite concentration and signal intensity/integration value), allowing not only to identify but also to quantify the metabolites in biological samples. Moreover, since there is no need for sample extractions, the reproducibility is very high [4,8].

The application of ^{1}H-NMR-based metabolomics in the search for cancer biomarkers has increased enormously over the past decade [9-15]. Most of these studies used 400-600 MHz NMR spectrometers to study the disturbed metabolism of cancer cells [9,11,13-15]. However, recently, high-field NMR spectrometers with ^{1}H resonance frequencies up to 800 MHz have been employed [10,12]. Although NMR spectra obtained at these high field strengths have an improved spectral resolution and signal-to-noise ratio (S/N), the cost and housing facilities also raise strongly [16].

Recently, Louis et al. have performed metabolite spiking experiments on a 400 MHz spectrometer in order to rationally segment the human plasma ^{1}H-NMR spectrum into variable-sized integration regions [17]. Using these integral settings, the spectra of a large cohort of 233 lung cancer patients and 226 controls were analyzed and the resulting data used to train a statistical orthogonal partial least squares discriminant analysis (OPLS-DA) model (classifier). The resulting model was able to discriminate between the groups with a sensitivity of 78% and

a specificity of 92%. The validity of the model was demonstrated in an independent cohort of 98 lung cancer patients and 89 controls with a sensitivity of 71% and a specificity of 81%. Moreover, knowledge of aberrant metabolite concentrations can help to unravel the metabolic changes correlated with disease and responsible for the group distinction [18]. However, complete signal assignment is almost impossible on a 400 MHz spectrometer, especially in crowded regions with severe signal overlap [19,20].

In the present study, NMR spectra of the plasma of the same group of 69 lung cancer patients and 74 controls were acquired on a 400 MHz as well as a 900 MHz spectrometer. The signal assignment and setting of 105 well-defined integration regions in the high-field 900 MHz spectra is based on metabolite spiking. The signal assignment and setting of 74 well-defined integration regions in the medium-field 400 MHz spectra is based on metabolite spiking and the signal assignment of the 900 MHz spectra. The data results of both spectrometers were used to train a classifier in differentiating between the groups and to identify the metabolic alterations which are causing the group discrimination.

Materials and Methods

Subjects

Lung cancer patients (n=69) were included in the Limburg Positron Emission Tomography Center (Hasselt, Belgium) from March 2011 to January 2012. The diagnosis of lung cancer was confirmed by a pathological biopsy or a clinician specialized in interpreting radiological and clinical lung cancer data. Clinical staging of the tumors was performed according to the 7th edition of the tumor, node, metastasis classification of malignant tumors [21] and independently checked by two of the authors (EL and KV). Controls (n=74) were patients with non-cancer diseases who were included at Ziekenhuis Oost-Limburg (Genk, Belgium) between December 2011 and April 2012.

Exclusion criteria were: 1) not fasted for at least 6 h; 2) fasting blood glucose concentration ≥ 200 mg/dl; 3) medication intake on the morning of blood sampling and 4) treatment or history of cancer in the past 5 years. The study was conducted in accordance with the ethical rules of the Helsinki Declaration and Good Clinical Practice and was approved by the ethical committees of Ziekenhuis Oost-Limburg and Hasselt University (Hasselt, Belgium). All study participants provided written informed consent.

Blood sampling and processing

Fasting venous blood samples were collected in 10 ml lithium-heparin tubes and stored at 4°C within 5 to 10 min. Within 8 h after blood collection, samples were centrifuged at 1600 g for 15 min and plasma aliquots of 500 µl were transferred into sterile cryovials and stored at -80°C until NMR analysis within six months.

NMR sample preparation

Plasma aliquots were centrifuged after thawing at 13000 g for 4 min at 4°C, followed by diluting 200 µl of the plasma with 600 µl deuterium oxide (D_2O, 99.9%) containing 0.3 µg/µl

trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP, 98%) as a chemical shift reference. All samples were placed on ice until ^{1}H-NMR analysis.

NMR measurements

^{1}H-NMR measurements were performed at 21.2°C on a 400 MHz spectrometer (9.4 Tesla; 54 mm bore-size; Varian Inova; Agilent Technologies Inc.; VnmrJ 3.2 RevisionA) and on a 900 MHz spectrometer (21.1 Tesla; 54 mm bore-size; Bruker Avance; Bruker Biospin). The 400 MHz spectrometer is equipped with an Agilent OneNMR 5mm probe, whereas the 900 MHz spectrometer has a triple resonance cryoprobe. Slightly T_2-weighted spectra were acquired using the Carr-Purcell-Meiboom-Gill pulse sequence (total spin-echo time of 32 ms), preceded by presaturation for water suppression. Other parameters were: a spectral width of 6000 Hz/14423 Hz (400 MHz/900 MHz), a 90° pulse length of 6.35/9.15 μs (400 MHz/900 MHz), an acquisition time of 1.2 s, a preparation delay of 3.5 s, and 96/64 scans (7min 44sec/5min 9sec on 400 MHz/900 MHz). Each free induction decay was zero-filled to 65 K points and multiplied by a 0.7 Hz exponential line-broadening function prior to Fourier transformation.

Metabolite spiking of reference plasma to assign the signals in the 400 MHz and 900 MHz NMR spectra

Fasting venous blood was collected from a healthy 44-year-old female. The plasma was obtained and processed as described above, and further referred to as reference plasma. For spiking, two stock solutions were prepared for each known metabolite by dissolving 0.01 mg or 1 mg in 100 µl of reference plasma. In a next step, 10 µl stock solution was added to a solution of 200 µl reference plasma and 600 µl D_2O containing TSP and subsequently analyzed on the 400 MHz and 900 MHz spectrometer as described above. When shifting of peaks was observed upon addition of the spiked metabolite, the concentration of the spike was decreased in order to assure correct signal assignments. This procedure was repeated for 37 different metabolites, i.e. alanine, arginine, asparagine, aspartate, cysteine, glutamine, glutamate,

glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, glucose, myo-inositol, acetate, acetoacetate, α-ketoglutarate, β-hydroxybutyrate, citrate, lactate, pyruvate, succinate, creatine, creatinine, aceton, betaine, choline, glycerol and methanol. **Table S1** presents the ^{1}H-NMR chemical shift values and J-coupling patterns of the spiked metabolites.

Spectral processing of 400 MHz and 900 MHz spectra

Based on the chemical shifts of the spiked metabolites, the 900 MHz ^{1}H-NMR spectra of lung cancer patients and controls were segmented into 105 variable-sized integration regions, excluding the water region (4.7-5.1 ppm) and TSP (-0.3-0.3 ppm). To check whether no positional changes of signals occur in the spectra between different plasma samples, two regions that contain well-resolved, non-overlapping signals (between 7.1-7.3 ppm and 3.45-3.9 ppm) were checked with respect to the location of the signals. The integration values were then normalized relatively to the total integrated area (except the area under the water and TSP signals), resulting in 105 normalized integration values, being the variables for multivariate statistics. By combining the information obtained from similar spiking experiments performed on a 400 MHz and 900 MHz spectrometer, the 400 MHz ^{1}H-NMR spectra of lung cancer patients and controls were segmented into 74 variable-sized integration regions, excluding the water region (4.7-5.0 ppm) and TSP (-0.1-0.3 ppm). Also for the 400 MHz ^{1}H-NMR spectra, two regions that contain well-resolved, non-overlapping signals (between 4.6-4.8 ppm and 3.7-3.85 ppm) were checked with respect to the location of the signals. The integration values were then normalized relatively to the total integrated area (except the area under the water and TSP signals), resulting in 74 normalized integration values, being the variables for multivariate statistics.

Statistical analysis

Multivariate statistics was performed using SIMCA-P+ (Version 14, Umetrics, Umea, Sweden). After mean-centering and Pareto scaling of the NMR data, supervised OPLS-DA was used to train a classification model in discriminating between lung cancer patients and controls [22]. The robustness of the classification models trained by means of the 400 and 900 MHz integration data, respectively, was further evaluated using Receiver Operating Characteristic Curve Explorer & Tester [23]. The most discriminating variables of the models and their corresponding variable importance for the projection (VIP) values were identified via an S-plot [24].

Results and Discussion

Figure 1 shows a zoom-in of ^{1}H-NMR spectra of a representative plasma sample acquired on a 400 MHz and a 900 MHz spectrometer. An improved spectral resolution as well as S/N can be observed for the 900 MHz spectrum as compared to the 400 MHz spectrum. To assign the resonance signals in both spectra, spiking experiments were performed with known metabolites. Louis et al. already reported the spiking of reference plasma with 37 different metabolites in relevant concentrations in order to assign the signals and to rationally divide the 400 MHz spectra into variable-sized, well-defined integration regions [17]. In this study, reference plasma was spiked with the same metabolites and analyzed on a 900 MHz spectrometer. The resulting information allows to rationally divide the 900 MHz spectra into 105 variable-sized, well-defined integration regions (**Table 1, left half**). Taking i) the chemical shifts of the spiked metabolite signals in the 400 MHz spectra and ii) the additional information obtained from the spiking experiments on the 900 MHz spectrometer into account, the 400 MHz spectra were divided into 74 variable-sized, well-defined integration regions (**Table 1, right half**).

Due to the improved resolution of a 900 MHz spectrum, an overlapping integration region in a 400 MHz spectrum can sometimes be divided into multiple integration regions that represent a single metabolite. An example is VAR34 in the 400 MHz spectrum, which can be divided into 6 integration regions (VAR50-55) in the 900 MHz spectrum, revealing that the relative plasma concentration of glucose (VAR54) is increased whereas the level of sphingomyelin and phosphatidylcholine (VAR55) is decreased for lung cancer patients. Remark that sphingomyelin and phosphatidylcholine can be discriminated from the other lipids on the basis of the strong singlet signal of the nine protons of the three methyl groups of the choline head group.

On the basis of the spiking experiments on the 900 MHz spectrometer, it can be concluded that the plasma levels of arginine, betaine, choline, cysteine, tryptophan and myo-inositol are below

the detection limit, even with the improved S/N of a 900 MHz spectrum. This explains why they are not shown in **Table 1**. The increased spectral resolution and S/N of the 900 MHz spectra further enable to define a larger number of integration regions that represent a single metabolite, thereby contributing to the identification of the discriminating metabolites and to the understanding of the underlying disturbed biochemical pathways of disease. More specifically, 58% of the variables in the 900 MHz spectrum (61 out of the 105) represent a single metabolite as opposed to 49% in the 400 MHz spectrum (36 out of 74).

Application to a lung cancer case-control dataset

In order to investigate the discriminative power of plasma metabolic phenotype data obtained at different magnetic field strengths, the plasma of a case-control dataset of 69 lung cancer patients and 74 controls was analyzed on a 400 MHz and a 900 MHz spectrometer. Subject characteristics are presented in **Table 2**. Supervised OPLS-DA analysis was conducted to train robust classification models in discriminating between lung cancer patients and controls. Using the 74 variables (integration values) obtained from the 400 MHz spectra, the trained model with 1 predictive component and 4 orthogonal components allows to classify 90% (62 out of 69) of the lung cancer patients and 97% (72 out of 74) of the controls correctly with an area under the curve (AUC) of 0.93 (**Figure 2a-b, Table 3**). For the 900 MHz dataset with 105 variables, the trained model with 1 predictive component and 6 orthogonal components enables to differentiate with a sensitivity of 88%, a specificity of 96% and an AUC of 0.90 (**Figure 2c-d, Table 3**). Independent permutation tests confirm that there is no overfitting for both classification models (**Figure S1a-b**).

Taking a closer look at the model characteristics (**Table 3**), it can be concluded that the trained 400 MHz and 900 MHz models both explain 86% of the intra-group variation ($R^2X(cum)$). In addition, the 900 MHz model explains 68% of the inter-group variation ($R^2Y(cum)$) compared to 66% for the 400 MHz model. Also, the predictive ability ($Q^2(cum)$) of both models, as

determined by sevenfold cross-validation, is quasi similar, i.e. 48% for the 900 MHz model versus 51% for the 400 MHz model. The-inter-variation for the models using multiple orthogonal components is shown in **Table 3**.

Taken all above in consideration, it can be concluded that the discriminative power and predictive ability of both models are comparable. Bertram et al. examined the impact of varying magnetic field strength, i.e. 250, 400, 500 and 800 MHz, on the urinary metabolic phenotype of 24 boys before and after a dietary intervention [16]. Although the study is rather limited in the number of subjects, indications were present that the power to discriminate between pre- and post-intervention samples improved upon increasing the magnetic field strength from 250 to 500 MHz, but that no further improvement was found for a further increase to 800 MHz. Our study on a larger cohort of 143 subjects confirms this finding.

In a next step, the variables discriminating most between lung cancer patients and controls (i.e. variables with a VIP value exceeding 0.5) were identified in an attempt to support in unraveling the disturbed biochemical pathways in lung cancer. It resulted in 35 and 42 variables from the 400 MHz and 900 MHz data, respectively. These variables (indicated with an asterisk in **Table 1**) are situated at the outer ends of the respective S-plots as demonstrated in **Figure 3a-b**. Evaluating these variables with respect to the controls, it is found that: i) in common for the 400 MHz and 900 MHz data, the plasma concentrations of glutamine, glucose, glycerol, isoleucine, leucine, N-acetylated glycoproteins, proline, threonine and valine are increased, whereas the concentrations of alanine, asparagine, citrate, lactate, non-cholinated lipids, phosphatidylcholine and sphingomyelin are decreased and, ii) according to the 900 MHz information, the plasma concentration of aspartate (VAR72, VIP: 0.77), β-hydroxybutyrate (VAR98, VIP: 0.91) and lysine (VAR94, VIP: 0.76) are increased. Since citrate and aspartate are overlapping in the 400 MHz spectrum (VAR47) and the concentration of citrate is decreased according to the 900 MHz (VAR74) and 400 MHz (VAR49) data, the increase of aspartate is

probably masked by the decrease of citrate in the 400 MHz spectra. Furthermore, knowing from the 900 MHz data that lysine (VAR94) and β-hydroxybutyrate (VAR98) are increased, information that cannot be retrieved from the 400 MHz data (VAR64, VAR68), it can be deduced that the increase of VAR42 in the 400 MHz spectrum and of VAR64 in the 900 MHz spectrum most probably arises from an increase of lysine rather than of α-ketoglutarate. Although somewhat beyond the goal of using this case-control dataset, the decreased levels of phospholipids confirm an enhanced membrane synthesis in lung cancer cells, while the other discriminating metabolites are pointing to a response on the Warburg effect, which is ongoing in the cancer cells [18].

Metabolomics and the NMR magnetic field strength

Taking all into account, it can be concluded that on the one hand medium-field NMR spectrometers can be satisfactory for clinical metabolomics studies in which group discrimination is the only aim. For the presented case-control dataset of 69 lung cancer patients and 74 controls, no significant improvement is observed if a higher field magnet is used (**Table 3**). High-field NMR spectra on the other hand have i) an increased spectral resolution as compared to 400 MHz spectra which results in a larger number of metabolites that can be detected via a unique, non-overlapping signal, an advantage if the analysis of the underlying disturbed biochemical pathways is an additional aim and ii) a higher S/N allowing a shorter total measuring time (and so higher throughput) as well as to detect smaller signals which might be important for other cancers or other diseases. An improved sensitivity might also become a consideration when the quantity of sample becomes limited (e.g. pediatric blood). However, one has to take into account that the NMR spectrometer costs also raise with the magnetic field strength, e.g. whereas the cost of a 400 MHz spectrometer is in the order of 300.000€ (500.000€ with a cryoprobe), this of a 900 MHz spectrometer is in the order of 2.750.000€ (3.000.000€ with a cryoprobe).

Acknowledgements

This study is part of the 'Limburg Clinical Research Program (LCRP) UHasselt-ZOL-Jessa', supported by the foundation Limburg Sterk Merk, province of Limburg, Flemish government, Hasselt University, Ziekenhuis Oost-Limburg and Jessa Hospital. We thank the Research Foundation Flanders for supporting via the MULTIMAR project. Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.

References

1. K. Hollywood, D. R. Brison, R. Goodacre, *Proteomics*. **2006**, *6*(17), 4716-4723.
2. J. K. Nicholson, J. C. Lindon, E. Holmes, *Xenobiotica*. **1999**, *29*(11), 1181-1189.
3. A. C. Dona, B. Jimenez, H. Schafer, E. Humpfer, M. Spraul, M. R. Lewis, J. T. Pearce, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal Chem*. **2014**, *86*(19), 9887-9894.
4. J. C. Lindon, J. K. Nicholson, *Annu Rev Anal Chem*. **2008**, *1*, 45-69.
5. M. Eliasson, S. Rannar, J. Trygg, *Curr Pharm Biotechnol*. **2011**, *12*(7), 996-1004.
6. R. Madsen, T. Lundstedt, J. Trygg, *Anal Chim Acta*. **2010**, *659*(1-2), 23-33.
7. J. Trygg, E. Holmes, T. Lundstedt, *J Proteome Res*. **2007**, *6*(2), 469-479.
8. M. E. Dumas, E. C. Maibaum, C. Teague, H. Ueshima, B. Zhou, J. C. Lindon, J. K. Nicholson, J. Stamler, P. Elliott, Q. Chan, E. Holmes, *Anal Chem*. **2006**, *78*(7), 2199-2208.
9. E. Garcia, C. Andrews, J. Hua, H. L. Kim, D. K. Sukumaran, T. Szyperski, K. Odunsi, *J Proteome Res*. **2011**, *10*(4), 1765-1771.
10. E. Jobard, C. Pontoizeau, B. J. Blaise, T. Bachelot, B. Elena-Herrmann, O. Tredan, *Cancer Lett*. **2014**, *343*(1), 33-41.
11. E. E. Kline, E. G. Treat, T. A. Averna, M. S. Davis, A. Y. Smith, L. O. Sillerud, *The J Urol*. **2006**, *176*(5), 2274-2279.
12. D. Kumar, A. Gupta, A. Mandhani, S. N. Sankhwar, *J Proteome Res*. **2015**.
13. C. M. Rocha, J. Carrola, A. S. Barros, A. M. Gil, B. J. Goodfellow, I. M. Carreira, J. Bernardo, A. Gomes, V. Sousa, L. Carvalho, I. F. Duarte, *J Proteome Res*. **2011**, *10*(9), 4314-4324.
14. L. Zhang, H. Jin, X. Guo, Z. Yang, L. Zhao, S. Tang, P. Mo, K. Wu, Y. Nie, Y. Pan, D. Fan, *Clin Biochem*. **2012**, *45*(13-14), 1064-1069.

15. T. A. Fedele, A. C. Galdos-Riveros, H. Jose de Farias e Melo, A. Magalhaes, D. A. Maria, *Biomed Pharmacother*. **2013**, *67*(2), 146-156.

16. H. C. Bertram, A. Malmendal, B. O. Petersen, J. C. Madsen, H. Pedersen, N. C. Nielsen, C. Hoppe, C. Molgaard, K. F. Michaelsen, J. O. Duus, *Anal Chem*. **2007**, *79*(18), 7110-7115.

17. E. Louis, L. Bervoets, G. Reekmans, E. De Jonge, L. Mesotten, M. Thomeer, P. Adriaensens, *Metabolomics*. **2015**, *11*, 225-236.

18. E. Louis, P. Adriaensens, W. Guedens, T. Bigirumurame, K. Baeten, K. Vanhove, K. Vandeurzen, K. Darquennes, J. Vansteenkiste, C. Dooms, Z. Shkedy, L. Mesotten, M. Thomeer, *J Thorac Oncol*. **2016**, *11*(4), 516-523.

19. T. M. O'Connell, *Bioanalysis*. **2012**, *4*(4), 431-451.

20. C. Zheng, S. Zhang, S. Ragg, D. Raftery, O. Vitek. *Bioinformatics*. **2011**, *27*(12), 1637-1644.

21. P. Goldstraw, J. Crowley, K. Chansky, D. J. Giroux, P. A. Groome, R. Rami-Porta, M. P. Postmus, V. Rusch, L. Sobin. *J Thorac Oncol*. **2007**, *2*(8), 706-714.

22. J. Trygg, S. Wold, *J Chemometr*. **2002**, *16*(3), 119-128.

23. J. Xia, D. I. Broadhurst, M. Wilson, D. S. Wishart, *Metabolomics*. **2013**, *9*(2), 280-299.

24. S. Wiklund, E. Johansson, L. Sjostrom, E. J. Mellerowicz, U. Edlund, J. P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, *Anal Chem.* **2008**, *80*(1), 115-122.

25. M. Kriat, J. Vion-Dury, S. Confort-Gouny, R. Favre, P. Viout, M. Sciaky, H. Sari, P. J. Cozzone, *J Lipid Res*. **1993**, *34*(6), 1009-1019.

26. M. Oostendorp, U. F. Engelke, M. A. Willemsen, R. A. Wevers, *Clin Chem*. **2006**, *52*(7), 1395-1405.

27. A. Papathanasiou, C. Kostara, M. T. Cung, K. Seferiadis, M. Elisaf, E. Bairaktari, I. A. Goudevenos, *Hellenic J Cardiol*. **2008**, *49*(2), 72-78.

28. E. M. Lenz, J. Bright, I. D. Wilson, S. R. Morgan, A. F. Nash, *J Pharm Biomed Anal*. **2003**, *33*(5), 1103-1115.

Figure Legends

Figure 1. Zoom-in between 0.80-1.10 ppm of the ^{1}H-NMR spectrum (line-broadening of 0.7 Hz) of human plasma acquired at 900 (*top*) and 400 MHz (*bottom*). The frequency (Hz) labeled signals represent the doublets of the two methyl groups of the amino acid valine. Remark that the J-coupling (7.2 Hz) is independent of the field strength when expressed in Hz (see marked resonance frequencies), but not when expressed in ppm. Abbreviations: ppm: parts per million

Figure 2. (a) OPLS-DA score plot derived from the 400 MHz data, (b) receiver operating characteristic curve derived from the 400 MHz data, (c) OPLS-DA score plot derived from the 900 MHz data, (d) receiver operating characteristic curve derived from the 900 MHz data. Abbreviations: AUC: area under the curve, C: controls, LC: lung cancer patients, OPLS-DA: orthogonal partial least squares discriminant analysis

Figure 3. (a) S-plot of the OPLS-DA model derived from the 400 MHz data showing the variables contributing most to group discrimination. Variables situated at the right end are increased in the plasma of lung cancer patients, while those situated at the left end are increased for the controls, (b) S-plot of the OPLS-DA model derived from the 900 MHz data. Variables used to explain the disturbed biochemical pathways in lung cancer (VIP > 0.5) are marked (●). Abbreviations: Var: variable, VIP: variable importance for the projection

Tables

Table 1. Overview of the 105 variable-sized integration regions defined in the 900 MHz NMR spectra and their contributing metabolites (**left half**) versus the 74 variable-sized integration regions and their contributing metabolites in the 400 MHz NMR spectra (**right half)**.

900 MHz					400 MHz				
VAR	**VIP**	**Contributing metabolites**	**Start**	**End**	**VAR**	**VIP**	**Contributing metabolites**	**Start**	**End**
1		NI	8,4914	8,4796					
2		Formate	8,3702	8,3602					
3		NI	8,2601	8,2500					
4		NI	8,2300	8,2050					
5		NI	7,8561	7,8104	**1**		NI	7,9500	7,8104
6		His	7,7812	7,7544	**2**		His	7,7890	7,7480
7		Phe	7,4677	7,4380	**3**		Phe, NI	7,4840	7,3620
8		Phe, NI	7,4162	7,3755					
9		Phe	7,3675	7,3484	**4**		Phe	7,3620	7,33
10		NI	7,3484	7,3227	**5**		NI	7,33	7,2820
11		Tyr	7,2327	7,2046	**6**		Tyr	7,2550	7,2000
12		NI	7,1894	7,1591					
13		His	7,0792	7,0597	**7**		His	7,1070	7,0597
14		NI	7,0201	6,9652					
15		Tyr	6,9355	6,9056	**8**		Tyr	6,9430	6,9050
16		NI	6,7460	6,7004	**9**		NI	6,7460	6,7004
17*[d]	1.67	Lipids: -C**H**=C**H**- in FAC[#]	5,4422	5,2833	**10*[d]**	1.51	Lipids: -C**H**=C**H**- in FAC[#]	5,4422	5,2752
18*[i]	0.91	Glucose	5,2751	5,2542	**11*[i]**	0.62	Glucose	5,2752	5,2516
19		C_2H in glycerol backbone of PL and TG[#]	5,2542	5,2301	**12**		C_2H in glycerol backbone of PL and TG[#]	5,2516	5,2030
20		C_2H in glycerol backbone of PL and TG[#]	5,2186	5,2038					
21		NI	5,1525	5,1187					
22*[i]	0.92	Glucose	4,7088	4,6421	**13*[i]**	0.87	Glucose	4,7088	4,6421
23		C_1H and C_3H in glycerol backbone of TG[#]	4,3579	4,2902	**14**		C_1H and C_3H in glycerol backbone of TG[#]	4,4100	4,2902
24		O-C**H**$_2$-CH_2-$N^+(CH_3)_3$ of PC and SM[#], Thr	4,2852	4,2536	**15**		O-C**H**$_2$-CH_2-$N^+(CH_3)_3$ of PC and SM[#], Thr	4,2902	4,2332
25		β-hydroxybutyrate	4,2000	4,1607	**16**		β-hydroxybutyrate	4,2000	4,1607
26*[d]	1.44	C_1H and C_3H in glycerol backbone of PL and TG[#], lactate	4,1570	4,1276	**17*[d]**	1.09	C_1H and C_3H in glycerol backbone of PL and TG[#], lactate	4,1607	4,1260
27*[d]	0.54	NI	4,1276	4,0904	**18*[d]**	0.94	NI	4,1260	4,0904
28		Creatinine	4,0904	4,0780	**19**				

							Creatinine	4,0904	4,0700
29		C_3H_2 in glycerol backbone of PL[#], Asn, His, Phe, Ser	4,0400	3,9913	**20*[i]**	0.57	C_3H_2 in glycerol backbone of PL[#], Asn, His, Phe, Ser	4,0400	3,9900
30*[i]	0.73	Asn, His, Ser, Tyr	3,9903	3,9644	**21*[i]**	1.40	Creatine, Asn, His, Ser, Tyr	3,9900	3,9586
31		Creatine	3,9644	3,9586					
32		Tyr	3,9586	3,9527	**22*[i]**	3.09	Glucose, Tyr	3,9586	3,8306
33*[i]	2.08	Glucose	3,9527	3,9120					
34*[i]	1.15	Glucose	3,9120	3,8957					
35*[i]	1.86	Glucose	3,8881	3,8306					
36*[i]	0.56	Glucose, Ala, Gln, Glu, Ser	3,8286	3,8097	**23*[i]**	0.74	Glucose, Ala, Gln, Glu, Ser	3,8286	3,7956
37*[i]	1.26	Glucose, Ala, Gln	3,8097	3,7794	**24**		Glucose, Ala, Gln	3,7956	3,7794
38*[i]	1.92	Glucose	3,7776	3,7275	**25*[i]**	2.01	Glucose	3,7794	3,7141
39*[i]	2.32	Glycerol	3,7204	3,6453	**26*[i]**	1.69	Glycerol	3,7141	3,6376
40*[i]	0.66	Val	3,6453	3,6212	**27*[i]**	0.72	Val	3,6376	3,6212
41*[i]	1.15	Thr	3,6163	3,5861	**28*[i]**	0.88	Thr	3,6212	3,5861
42		Glycerol	3,5861	3,5771	**29*[i]**	1.05	Glucose, glycerol	3,5861	3,5649
43*[i]	1.04	Glucose	3,5771	3,5481	**30*[i]**	0.75	Glucose	3,5649	3,5360
44*[i]	1.80	Glucose	3,5355	3,4798	**31*[i]**	1.82	Glucose, Pro	3,5360	3,3980
45*[i]	0.54	Pro	3,4772	3,4576					
46*[i]	1.46	Glucose	3,4576	3,4093					
47		Methanol	3,3964	3,3924	**32**		Methanol, NI	3,3980	3,3746
48		NI	3,3924	3,3746					
49		Pro	3,3746	3,3465	**33**		Pro	3,3746	3,3256
50		Phe	3,3256	3,3132	**34*[d]**	2.09	O-CH_2-CH_2-N^+(CH_3)$_3$ of PC and SM[#], glucose, Phe, NI	3,3256	3,1972
51		Phe, NI	3,3132	3,3030					
52		NI	3,3030	3,2956					
53		NI	3,2956	3,2909					
54*[i]	1.12	Glucose	3,2909	3,2616					
55*[d]	2.28	O-CH_2-CH_2-N^+(CH_3)$_3$ of PC and SM[#]	3,2616	3,2085					
56		Tyr, NI	3,1972	3,1895	**35**		Tyr, NI	3,1972	3,1895
57		NI	3,1881	3,1821	**36**		NI	3,1881	3,1724
58		NI	3,1821	3,1724					
59		NI	3,1707	3,1571	**37**		NI	3,1707	3,1540
60		His, Phe	3,1541	3,1378	**38**		His, Phe	3,1540	3,1090
61		Tyr	3,0921	3,0769	**39*[d]**	0.70	Tyr	3,1090	3,0769
62		Creatinine	3,0769	3,0699	**40**		Creatinine	3,0769	3,0699
63		Creatine	3,0699	3,0635	**41**		Creatine	3,0699	3,0635
64*[i]	0.58	α-ketoglutrate, Lys	3,0635	3,0047	**42*[i]**	0.76	α-ketoglutarate, Lys	3,0635	2,9950
65		Lipids: =CH-C**H**$_2$-CH= in FAC[#]	3,0047	2,9655	**43*[d]**	1.13	Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn	2,9950	2,8885
66*[d]	0.58	Asn	2,9597	2,9201					
67		Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn, Asp	2,8874	2,8465	**44*[d]**	0.77	Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn, Asp	2,8885	2,8465

68*[d]	0.56	Lipids: =CH-C**H**$_2$-CH= in FAC[#]	2,8465	2,7623	**45*[d]**	1.21	Lipids: =CH-C**H**$_2$-CH= in FAC[#]	2,8465	2,7571
69		Citrate	2,7571	2,7493	**46**		Citrate, NI	2,7571	2,7368
70		NI	2,7472	2,7390					
71		Citrate	2,7368	2,7251	**47**		Citrate, Asp	2,7368	2,6768
72*[i]	0.77	Asp	2,7237	2,6768					
73		Met	2,6768	2,6597	**48**		Met	2,6768	2,6597
74*[d]		Citrate	2,5865	2,5426	**49*[d]**	0.64	Citrate	2,5920	2,5340
75*[i]	0.68	Gln	2,5183	2,4428	**50*[i]**	0.53	Gln	2,5183	2,4428
76		β-hydroxybutyrate	2,4428	2,4280	**51**		β-hydroxybutyrate	2,4428	2,4148
77		Pyruvate	2,4060	2,3978	**52**		Pyruvate	2,4060	2,3978
78		Glu	2,3978	2,3648	**53**		Glu	2,3978	2,3640
79		β-hydroxybutyrate	2,3540	2,3194	**54**		β-hydroxybutyrate	2,3628	2,3180
80		Acetoacetate	2,3134	2,3067	**55**		Acetoacetate	2,3134	2,3067
81*[d]	0.63	Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#]	2,3067	2,2630	**56*[d]**	0.93	Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#], aceton	2,3067	2,2300
82		Aceton	2,2630	2,2563					
83		NI	2,1975	2,1814	**57*[i]**	0.72	Gln, Met, NI	2,2000	2,1270
84		Gln	2,1777	2,1670					
85		Met	2,1670	2,1619					
86*[i]	0.85	Gln	2,1619	2,1311					
87*[d]	1.76	Lipids: -C**H**$_2$-CH=CH- in FAC[#]	2,1289	2,0993	**58*[i]**	2.58	Lipids: -C**H**$_2$-CH=CH- in FAC[#], C**H**$_3$ of NAG[$]	2,1270	1,9680
88*[i]	2.74	Lipids: -C**H**$_2$-CH=CH- in FAC, C**H**$_3$ of NAG[$]	2,0993	1,9889					
89		Acetate	1,9547	1,9421	**59**		Acetate	1,9550	1,9421
90		Lys	1,9421	1,9028	**60**		Lys	1,9421	1,8800
91*[i]	1.05	Leu	1,8006	1,6758	**61*[i]**	0.85	Leu	1,8034	1,6758
92		Lipids: -C**H**$_2$-CH$_2$-C=O or -C**H**$_2$-CH$_2$-CH=CH- in FAC[#]	1,6530	1,5770	**62*[d]**	1.08	Lipids: -C**H**$_2$-CH$_2$-C=O or -C**H**$_2$-CH$_2$-CH=CH- in FAC[#]	1,6700	1,5500
93*[d]	0.52	Ala	1,5226	1,4919	**63*[d]**	0.56	Ala	1,5300	1,4900
94*[i]	0.76	Lys	1,4587	1,4201	**64**		Lys	1,4656	1,4200
95*[d]	1.53	Lactate	1,4169	1,3675	**65*[d]**	1.66	Lactate	1,4169	1,3620
96*[d]	2.77	Lactate	1,3675	1,3516	**66*[d]**	1.75	Lactate, Thr	1,3620	1,3450
97*[d]	4.58	Lipids: -CH$_3$-(C**H**$_2$)$_n$- in FAC[#]	1,3516	1,2366	**67*[d]**	2.99	Lipids: CH$_3$-(C**H**$_2$)$_n$- in FAC[#]	1,3450	1,2360
98*[i]	0.91	β-hydroxybutyrate	1,2366	1,2240	**68**		β-hydroxybutyrate	1,2360	1,2180
99	1.80	NI	1,2240	1,1766	**69**		NI	1,2180	1,1578
100		Val	1,0860	1,0592	**70**		Val	1,0900	1,0592
101		Ile	1,0513	1,0340	**71**		Ile	1,0569	1,0334
102*[i]	0.65	Val	1,0310	1,0106	**72**		Val	1,0334	1,0092
103*[i]	0.60	Leu	1,0083	0,9766	**73*[i]**	1.04	Ile, Leu	1,0092	0,9663
104*[i]	0.66	Ile	0,9766	0,9663					

105*^{d}	2.33	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC[#]	0,9663	0,7961	**74*^{d}**	1.78	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC[#]	0,9663	0,7961

[#]The assignment of lipid signals is based on literature (25-27). [$]The assignment of signals of N-acetylated glycoproteins is based on literature (28). Amino acids are presented by their 3-letter code. [*d]: Variables having a VIP value above 0.5 and which are decreased in the plasma of lung cancer patients, [*i]: Variables having a VIP value above 0.5 and which are increased in the plasma of lung cancer patients. Abbreviations: FAC: fatty acid chain, NAG: N-acetylated glycoproteins, NI: non-identified, PC: phosphatidylcholine, PL: phospholipids, ppm: parts per million, SM: sphingomyelin, TG: triglycerides, VAR: variable: VIP: variable importance for the projection

Table 2. Characteristics of the subjects included in the study. Data are presented as mean ± standard deviation and range, unless otherwise indicated.

	LC	**C**
Number of subjects, n	69	74
Gender, n (%)		
Male	46 (66.7)	44 (59.5)
Female	23 (33.3)	30 (40.5)
Age, yrs	68 ± 10	64 ± 13
(range)	(36 – 88)	(23 – 84)
BMI, kg/m^2	25.3 ± 4.6	26.3 ± 4.6
(range)	(17.5 – 38.5)	(16.5 – 39.0)
Smoking habits		
Smoker, n (%)	40 (58.0)	19 (25.7)
Ex-smoker, n (%)	26 (37.7)	28 (37.8)
Non-smoker, n (%)	3 (4.3)	27 (36.5)
Pack years	34 ± 21	18 ± 28
(range)	(0-125)	(0-175)
Laterality		
Left, n (%)	23 (33.3)	
Right, n (%)	39 (56.5)	
Bilateral, n (%)	5 (7.2)	
Unknown, n (%)	2 (2.9)	
Amount of tumors, n	74	
Histological subtype		
NSCLC-Adenocarcinoma, n (%)	27 (36.5)	
NSCLC-Spinocellular carcinoma, n (%)	18 (24.3)	
NSCLC-Adenosquamous carcinoma, n (%)	3 (4.1)	
NSCLC-Carcinoid, n (%)	1 (1.3)	
NSCLC-NOS, n (%)	3 (4.1)	
SCLC, n (%)	12 (16.2)	
Unknown	10 (13.5)	
Clinical stage according to 7th TNM edition		
IA, n (%)	18 (24.3)	
IB, n (%)	5 (6.7)	
IIA, n (%)	4 (5.4)	
IIB, n (%)	2 (2.7)	
IIIA, n (%)	15 (20.3)	
IIIB, n (%)	11 (14.9)	
IV, n (%)	19 (25.7)	

Abbreviations: BMI: body mass index, C: controls, LC: lung cancer patients, NOS: not otherwise specified, NSCLC: non-small cell lung cancer, SCLC: small cell lung cancer, TNM: tumor, node, metastasis

Table 3. Characteristics of the trained OPLS-DA classification models resulting from the 400 MHz and 900 MHz data.

	LV (P+O)	R^2X (cum)	R^2Y (cum)	Q^2 (cum)	Sens (%)	Spec (%)	AUC
LC vs. C (400 MHz)	2 (1+1)	0.70	0.50	0.44	75	95	-
	3 (1+2)	0.80	0.56	0.49	77	95	-
	4 (1+3)	0.84	0.61	0.50	84	96	-
	5 (1+4)	0.86	0.66	0.51	90	97	0.93
LC vs. C (900 MHz)	2 (1+1)	0.65	0.40	0.31	72	86	-
	3 (1+2)	0.74	0.45	0.34	77	89	-
	4 (1+3)	0.78	0.52	0.36	81	93	-
	5 (1+4)	0.81	0.58	0.38	86	95	-
	6 (1+5)	0.85	0.62	0.40	88	96	-
	7 (1+6)	0.86	0.68	0.48	88	96	0.90

Abbreviations: AUC: area under the curve, C: controls, LC: lung cancer patients, LV: latent variable, MHz: megahertz, O: number of orthogonal components, OPLS-DA: orthogonal partial least squares discriminant analysis, P: number of predictive components, R^2X(cum): total explained variation in X, R^2Y(cum): total explained variation in Y, Sens: sensitivity: Spec: specificity, Q^2(cum): predicted variation

Supplementary Material

Table S1. ^{1}H-NMR chemical shifts (δ in ppm) of low molecular weight plasma metabolites and their J-coupling constants (in Hz).

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Amino acids					
Alanine (Ala)	$^{\alpha}CH$	3.790	q	7.2	α-β
(CHEBI:57972)	$^{\beta}CH_3$	1.509	d	7.2	β-α
Arginine (Arg)	$^{\alpha}CH$	3.690	t	6.1	α-β; α-β’
(CHEBI:32682)	$^{\beta}CH_2$	1.700	m	-	-
	$^{\gamma}CH_2$	1.902	m	-	-
	$^{\delta}CH_2$	3.266	t	6.9	δ-γ
Asparagine (Asn)	$^{\alpha}CH$	3.999	dd	7.8; 4.3	α-β; α-β'
(CHEBI:58048)	$^{\beta}CH_2$	2.845	dd	16.7; 4.3	β-β'; β-α
		2.962	dd	16.7; 7.8	β'-β; β'-α
Aspartate (Asp)	$^{\alpha}CH$	3.930	dd	8.9; 3.7	α-β; α-β'
(CHEBI:29991)	$^{\beta}CH_2$	2.702	dd	17.5; 3.7	β-β'; β-α
		2.850	dd	17.5; 8.9	β'-β; β'-α
Cysteine (Cys)	$^{\alpha}CH$	3.973	dd	5.7; 4.3	α-β; α-β'
(CHEBI:35235)	$^{\beta}CH_2$	3.052	dd	14.7; 4.3	β-β'; β-α
		3.112	dd	14.7; 5.7	β'-β; β'-α
Glutamine (Gln)	$^{\alpha}CH$	3.786	t	6.2	α-β; α-β’
(CHEBI:58359)	$^{\beta}CH_2$	2.160	m	-	-
	$^{\gamma}CH_2$	2.480	m	-	-

Table S1 continued.

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Glutamate (Glu)	$^{\alpha}CH$	3.788	dd	7.1; 4.9	α-β; α-β'
(CHEBI:29985)	$^{\beta}CH_2$	2.120	m	-	-
	$^{\gamma}CH_2$	2.388	m	-	-
Glycine (Gly)	$^{\alpha}CH_2$	3.586	s	-	-
(CHEBI:57305)					
Histidine (His)	$^{\alpha}CH$	4.012	dd	8.0; 4.9	α-β; α-β'
(CHEBI:57595)	$^{\beta}CH_2$	3.150	dd	15.5; 8.0	β-β'; β-α
		3.260	dd	15.5; 4.9	β'-β; β'-α
	$^{\gamma}CH$	7.780	s	-	-
	$^{\delta}CH$	7.086	s	-	-
Isoleucine (Ile)	$^{\alpha}CH$	3.673	d	4.0	α-β
(CHEBI:58045)	$^{\beta}CH$	1.990	m	-	-
	$^{\gamma}CH_3$	1.015	d	7.0	γ-β
	$^{\delta}CH_2$	1.476	m	-	-
	$^{\varepsilon}CH_3$	0.945	t	7.4	ε-δ
Leucine (Leu)	$^{\alpha}CH$	3.769	dd	7.0; 1.3	α-β; α-β'
(CHEBI:57427)	$^{\beta}CH_2$	1.742	m	-	-
	$^{\gamma}CH$	1.742	m	-	-
	$^{\delta}CH_3$	0.987	d	4.7	δ-γ
		1.003	d	4.7	δ'-γ

Table S1 continued.

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Lysine (Lys)	$^{\alpha}CH$	3.772	t	6.0	α-β; α-β’
(CHEBI:32551)	$^{\beta}CH_2$	1.928	m	-	-
	$^{\gamma}CH_2$	1.502	m	-	-
	$^{\delta}CH_2$	1.751	p	7.5	γ-δ; δ-ε
	$^{\varepsilon}CH_2$	3.060	t	7.5	ε-δ
Methionine (Met)	$^{\alpha}CH$	3.875	dd	7.0; 5.4	α-β; α-β'
(CHEBI:57844)	$^{\beta}CH_2$	2.195	m	-	-
	$^{\gamma}CH_2$	2.673	t	7.6	γ-β; γ-β'
	$^{\delta}CH_3$	2.167	s	-	-
Phenylalanine (Phe)	$^{\alpha}CH$	3.998	dd	7.7; 5.2	α-β; α-β'
(CHEBI:58095)	$^{\beta}CH_2$	3.140	dd	14.4; 5.2	β-β'; β-α
		3.310	dd	14.4; 7.7	β'-β; β’-α
	$^{\gamma}CH$	7.353	d	7.2	γ-δ
	$^{\delta}CH$	7.454	t	7.2	δ-γ; δ-ε
	$^{\varepsilon}CH$	7.414	t	7.2	ε-δ
Proline (Pro)	$^{\alpha}CH$	4.162	dd	8.9; 6.3	α-β; α-β'
(CHEBI:60039)	$^{\beta}CH_2$	2.382	m	-	-
	$^{\gamma}CH_2$	2.060	m	-	-
	$^{\delta}CH_2$	3.365	t	7.0	δ-γ
		3.441	t	7.0	δ'-γ

Table S1 continued.

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Serine (Ser)	$^{\alpha}CH$	3.845	dd	5.6; 4.0	α-β; α-β'
(CHEBI:33384)	$^{\beta}CH_2$	3.953	dd	12.2; 5.6	β-β'; β-α
		4.012	dd	12.2; 4.0	β'-β; β'-α
Threonine (Thr)	$^{\alpha}CH$	3.596	d	4.9	α-β
(CHEBI:57926)	$^{\beta}CH$	4.276	dq	6.6; 4.9	β-α; β-γ
	$^{\gamma}CH_3$	1.358	d	6.6	γ-β
Tryptophan (Trp)	$^{\alpha}CH$	4.086	dd	8.1; 5.2	α-β; α-β'
(CHEBI:57912)	$^{\beta}CH_2$	3.338	dd	15.3; 8.1	β-β'; β-α
		3.224	dd	15.3; 5.2	β'-β; β'-α
	$^{\gamma}CH$	7.351	s	-	-
	$^{\delta}CH$	7.770	d	7.8	δ-ε
	$^{\varepsilon}CH$	7.229	t	7.8	ε-δ; ε-ζ
	$^{\zeta}CH$	7.310	t	7.8	ζ-ε; ζ-η
	$^{\eta}CH$	7.570	d	7.8	η-ζ
Tyrosine (Tyr)	$^{\alpha}CH$	3.957	dd	7.8; 5.0	α-β; α-β'
(CHEBI:58315)	$^{\beta}CH_2$	3.076	dd	14.2; 7.8	β-β'; β-α
		3.227	dd	14.2; 5.0	β'-β; β'-α
	$^{\gamma}CH$	6.924	d	8.4	γ-δ
	$^{\delta}CH$	7.222	d	8.4	δ-γ
Valine (Val)	$^{\alpha}CH$	3.635	d	4.3	α-β
(CHEBI:57762)	$^{\beta}CH$	2.305	m	-	-
	$^{\gamma}CH_3$	1.021	d	7.2	γ-β
		1.074	d	7.2	γ'-β

Table S1 continued.

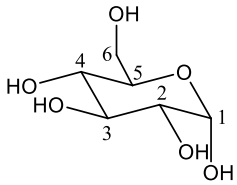
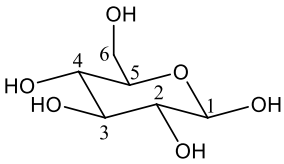
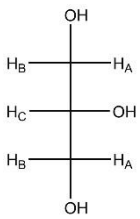
Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Carbohydrates					
D-glucose					
α-anomer	C1H	5.264	d	3.8	-
(CHEBI:17925)	C2H	3.563	dd	9.8; 3.8	-
	C3H	3.744	t	9.4	-
	C4H	3.439	t	9.4	-
	C5H	3.888	m	-	-
	C6H	3.858	dd	10; 2.2	-
	C6'H	3.792	dd	13.1; 6.3	-
β-anomer	C1H	4.678	d	7.8	-
(CHEBI:15903)	C2H	3.272	t	8.3	-
	C3H	3.518	t	9.2	-
	C4H	3.428	t	9.4	-
	C5H	3.492	m	-	-
	C6H	3.933	dd	12.2; 2.0	-
	C6'H	3.752	dd	12.2; 5.7	-
Glycerol	CH_C	3.814	m	-	-
(CHEBI:17754)	CH_2: H_A	3.591	dd	11.7; 6.6	-
	CH_2: H_B	3.682	dd	11.7; 4.4	-

Table S1 continued.

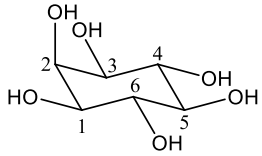
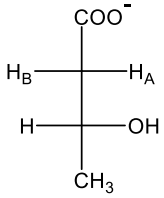
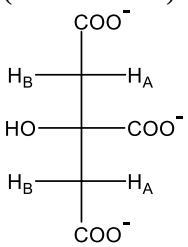
Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Myo-inositol	C5H	4.090	t	2.9	-
(CHEBI:17268)	C4H+C6H	3.562	dd	9.8; 2.9	-
	C1H+C3H	3.654	t	9.8	-
	C2H	3.306	t	9.3	-
Organic acids					
Acetate	CH_3	1.948	s	-	-
(CHEBI:30089)					
Acetoacetate	CH_2	2.319	s	-	-
(CHEBI:13705)	CH_3	3.480	s	-	-
α-ketoglutarate	CH_2-CO	3.040	t	6.9	-
(CHEBI:16810)	CH_2-COO^-	2.470	t	6.9	-
D-β-hydroxybutyrate	CH_A	2.400	dd	14.5; 7.3	-
(CHEBI:10983)	CH_B	2.300	dd	14.5; 7.3	-
	CH	4.184	m	-	-
	CH_3	1.231	d	6.3	-
Citrate	CH_A	2.717	d	15.8	-
(CHEBI:16947)	CH_B	2.566	d	15.8	-
					

Table S1 continued.

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
L-lactate	CH	4.138	q	6.9	-
(CHEBI:16651)	CH_3	1.354	d	6.9	-
Pyruvate	CH_3	2.402	s	-	-
(CHEBI:15361)					
Succinate	CH_2	2.439	s	-	-
(CHEBI:30031)					
Others					
Acetone	CH_3	2.264	s	-	-
(CHEBI:15347)					
Betaine	CH_3	3.300	s	-	-
Choline	CH_3	3.236	s	-	-
(CHEBI:133341)	CH_2	3.554	m	-	-
	CH_2OH	4.098	m	-	-
Creatine	CH_3	3.068	s	-	-
(CHEBI:57947)	CH_2	3.962	s	-	-
Creatinine	CH_3	3.075	s	-	-
(CHEBI:16737)	CH_2	4.087	s	-	-
Methanol	CH_3	3.396	s	-	-
(CHEBI:17790)					

Chemical shifts are expressed relatively to the singlet resonance of the trimethyl protons of TSP at δ 0.015 ppm and J-coupling patterns are described as: s, singlet; d, doublet; dd, double doublet; dq, double quadriplet; t, triplet; q, quadriplet; p, pentaplet; m, multiplet. Metabolite identifiers from the database of Chemical Entities of Biological Interest (ChEBI) are indicated. The atom numbering of the metabolites

follows the IUPAC-IUB nomenclature unless otherwise indicated in the structures included in **Table S1.**

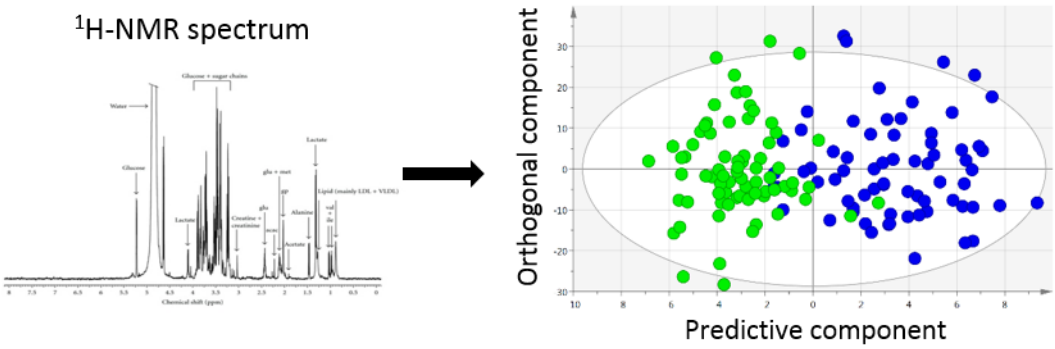
Figure S1. Permutation tests comparing the goodness of fit of the trained OPLS-DA models obtained from the 400 MHz data (A) and the 900 MHz data (B) with that of twenty permuted models. For both the 400 MHz and 900 MHz data, the R_2 and Q_2 values obtained for the permuted models (at the left) are lower than these of the original model (at the right), indicating that there is no overfitting. Abbreviations: R_2: explained variation, Q_2: predicted variation as determined by 7-fold cross-validation.

Figure S2. Superposition of zoom-ins (between 0.5 and 4.75 ppm) of five ^{1}H-NMR spectra obtained from five NMR samples prepared from the same plasma pool.

Graphical Abstract

Metabolic phenotyping of human plasma by ^{1}H-NMR at high and medium magnetic field strengths: a case study for lung cancer

Louis E., Cantrelle F-X., Mesotten L., Reekmans G., Vanhove K., Bervoets L., Thomeer M., Lippens G., Adriaensens P.*



This study describes metabolite spiking experiments on the basis of which the ^{1}H-NMR spectrum can be segmented into well-defined integration regions, and this for 400 and 900 MHz spectra. Subsequently, the integration data of a case-control dataset were used to train a multivariate classifier for both magnetic field strengths. The discriminative power of the resulting models is rather similar, i.e. a sensitivity and specificity of 90% and 97% for the 400 MHz data versus 88% and 96% for the 900 MHz data.