Supplemental Digital Content 1

METHODS

Nuclear Magnetic Resonance data acquisition

Intact lung tissue samples (weighing approx. 10 mg) were examined using HR-MAS ^{1}H-NMR spectroscopy operating at 4 ºC to reduce metabolic degradation. ^{1}H-NMR spectroscopy was performed at 500.13 MHz using a Bruker AMX500 spectrometer 11.7 T (Bruker, Rivas-Vaciamadrid, Spain). The samples were placed into a 50 µl zirconium oxide rotor using a rinsed cylindrical insert, together with 15 µl of 0.1mM solution of Trimethylsilyl propanoic acid (TSP) in deuterium water (D_2O), and spun at 4000 Hz spinning rate to remove the effects of spinning side bands from the acquired spectra. Additionally, bronchoalveolar lavage (BAL) fluid lyophilized samples were diluted with 100 µl of D_2O and 100 µl of solution of TSP in D_2O and examined using a Bruker AV500 spectrometer operating at 4ºC.

Shimming and ^{1}H-NMR preparation time were reduced to a minimum, meanwhile the temperature for performing the ^{1}H-NMR analysis was chilled to 4 ºC to minimize metabolic changes. A number of bidimensional homonuclear experiments were performed to carry out the components assignments. Between consecutive 2D spectra, a control ^{1}H-NMR spectrum was always measured. No gross degradation was noted in the signals of multiple spectra acquired under the same conditions.

Standard solvent suppressed spectra were grouped into 16,000 data points for tissue samples analyses and into 32,000 data points for BAL analyses,

averaged over 256 acquisitions. The data acquisition lasted in total 13 min using a sequence based on the first increment of the Nuclear Overhauser Effect Spectroscopy (NOESY) pulse sequence to effect suppression of the water resonance and limit the effect of B_0 and B_1 inhomogeneities in the spectra (relaxation delay-90º-t1-90º-tm-90º-acquire Free Induction Decay signal) in which a secondary radio frequency irradiation field was applied at the water resonance frequency during the relaxation delay of 2 s and during the mixing period (tm = 150 ms), with t1 fixed at 3 s. Tissue samples acquisitions were performed using a spectral width of 8333.33 Hz and BAL fluid acquisitions were acquired using a spectral width of 7507.5 Hz. Prior to Fourier transformation, the Free Induction Decay signals were multiplied by an exponential weight function corresponding to a line broadening of 0.3 Hz. Spectra were referenced to the TSP singlet at 0 ppm chemical shift.

A standard gradient-enhanced Correlation Spectroscopy protocol was acquired under the following conditions: water presaturation during relaxation delay, spectral with of 5122.95 Hz in both dimensions, 2000 data points in f2 and 256 increments in f1. An unshifted sinusoidal window function was applied in both dimensions and zero filling in f1 dimension. ^{1}H-^{1}H Total Correlated Spectroscopy experiment was registered in the Time-Proportional Phase Incrementation phase sensitive mode, water pre-saturation during 1s relaxation delay, a spectral with of 5122.95 Hz in both dimensions, 60 ms mixing time, 2000 data points in f2 and 256 increments in f1. Zero filling in f1 and unshifted squared sinusoidal window function in both dimensions were applied before Fourier transformation.

Gradient-selected Heteronuclear Single Quantum Correlation experiments of serum and tissue samples were registered with the following parameters: 95 μs for Globally Optimized Alternating Phase Rectangular Pulse ^{13}C decoupling, 8333Hz and 21 kHz spectral widths in the ^{1}H and ^{13}C dimensions, respectively, 2000 data points in f_2 and 256 increments in f_1. Zero filling in f_1 and unshifted squared sinusoidal window function in both dimensions were applied before Fourier transformation.

In order to remove the random effects of variation in the water resonance suppression, the chemical shift regions between 4.90 and 5.30 ppm of tissue spectra, 5.00-5.20 ppm of serum spectra and 4.85-5.00 of BAL spectra, were excluded from the analysis. Similarly, the chemical shift region from 0 to 0.04 ppm containing the internal reference (TSP) was also excluded. The baseline correction (method by Rocke and Xi) (**1**) was automatically performed with the baseline correction tool of the Metabonomic R package (**2**). The Metabonomic R package is a graphical environment for the metabolomic analysis developed and maintained by our group in the public domain R framework (rel. 2.11.1) (**3**). ^{1}H-NMR spectra were automatically data-reduced to integral segments or buckets of equal length (δ0.01 ppm for BAL spectra and δ0.04 ppm for tissue and serum spectra) in order to compensate for variations in resonance positions (**4**) and they were normalized to the total sum of the spectral regions. 2D spectral processing and editing was performed with MestRenova v. 6.03 (Mestrelab Research S.L., Santiago de Compostela, Spain).

NMR data treatment

Principal Components Analysis (PCA) is the fundamental method in chemometrics (**5**). In PCA the data collected on a set of samples is resolved

into principal components. The first principal component is defined by the spectral profile (loading) in the data which describes most of the variation, the second principal component, orthogonal to the first one, is the second best profile describing the variation, and so on. The principal components are composed of so-called scores and loadings. Loadings contain information about the variables (chemical shifts) in the data set and the scores hold information on samples (concentrations) in the data set. Prior to PCA, the data were centered and Pareto scaled.

A Partial Least Squares (PLS) analysis (**6**) is a commonly used multivariate method for analyzing high-dimensional data. PLS analysis was applied to these data to investigate the significant differences between groups. The potential biomarkers selected from PCA loading plots were confirmed from PLS correlation plots by Hotteling's T^2 tests (**7**). Using the selected metabolites or chemical shifts, a PLS-Discriminant Analysis (PLS-DA) was developed as classificatory model. We have used the algorithm proposed by Ding and Gentleman* (tolerance for convergence=1e-03, maximum number of iteration allowed = 100). The number of PLS components used was chosen by the percentage of variance explained, the R^2 and the Mean Squared Error of Cross Validation graphics. PLS-DA models were trained with a number of random testing subjects and used afterwards to classify the rest of subjects as an internal validation. This process was repeated 200 times with random

* Ding B, Gentleman R: The gpls package: Classification using generalized partial least squares, version 1.3.1. Available at: ftp://ftp.auckland.ac.nz/pub/software/CRAN/doc/packages/gpls.pdf. Accessed October 7, 2013.

permutations of the data to reduce type I error. The percentages of correct classification were calculated as a measure of the model performance. The statistical computing and spectral processing described above were also performed with the Metabonomic package (rel.3.3.1).

Mass Spectrometry (MS) serum sample preparation

Protein precipitation and metabolite extraction was performed by adding 1 part of serum to 3 parts of cold (-20 °C) mixture of methanol and ethanol (1:1). Samples were then vortex-mixed and stored at -20 °C for 5 min. The supernatant was collected by centrifuging at 16 000× g for 10 min at 4 °C, and then the supernatant was filtered through a 0.22 μm nylon filter.

Quality control (QC) samples were prepared by pooling equal volumes of serum from each of the 23 samples. Five samples were independently prepared from this pooled serum following the same procedure as for the rest of samples. QC samples were analyzed throughout the run to provide a measurement not only of the system's stability and performance (**8**) but also of the reproducibility of the sample treatment procedure.

MS data acquisition

The High Performance Liquid Cromatography system consisted of a degasser, two binary pumps, and autosampler (1200 series, Agilent, Santa Clara, CA); 10 μL of extracted serum sample was applied to a reversed-phase column (Discovery HS C18 15 cm × 2.1 mm, 3 μm; Supelco, St. Louis, MO) with a guard column (Discovery HS C18 2 cm × 2.1 mm, 3 μm; Supelco). The system was operated in positive ion mode at the flow rate 0.6 mL/ min with solvent A composed of water with 0.1% formic acid, and solvent B composed of

acetonitrile with 0.1% formic acid. The gradient started from 25% B to 95% B in 35 min, and returned to starting conditions in 1 min, keeping the re-equilibration at 25% B for 9 min. Data were collected in positive electrospray ionization mode in separate runs on a quadrupole time-of-flight (QTOF) (Agilent 6520) operated in full scan mode from 50 to 1000 m/z. The capillary voltage was 3000 V with a scan rate of 1.02 scan per second; the nebulizer gas flow rate was 10.5 L/min.

The resulting data file was cleaned of extraneous background noise and unrelated ions by the molecular feature extraction tool in the MassHunter Qualitative Analysis Software U (Agilent). The molecular feature extraction then created a listing of all possible components as represented by the full QTOF mass spectral data. Exact mass databases quoted below were then searched for hits to identify the compounds.

MS data treatment

To avoid the risk of over-fitting for a PLS-DA model used for selection of statistically significant metabolites according to jack-knifed confidence intervals, the model was validated by use of a cross-validation tool **(9),** using a 1/3 out approach which has been described elsewhere **(10).** In brief, the dataset was divided into three parts, 1/3 of samples were excluded, and a model was built using the remaining 2/3 of samples. Excluded samples were then predicted by this new model, and the procedure was repeated until all samples had been predicted at least once. Each time the percentage of correctly classified samples was calculated.

QC of the MS methodology

QCs were checked along with the two other groups, VILI and control. For these three groups a PLS-DA model was built taking all variables (without any scaling) generated after molecular feature extraction in the mass spectrum (15016 variables in total). The robustness of the analytical procedure was evident by the tight clustering of QC samples obtained by mixing equal volumes of all samples. QCs were located in the center of the plot when sent to be classified by the model proving that separation between groups is not random, but due to real variability. The quality of the model built for three components was good (variance explained R2 = 0.907, and variance predicted Q2 = 0.383).

Compound identification

The nuclear magnetic resonance signals were identified according to the Human Metabolome Database (**11**), and characteristic cross-peaks from 2D spectra to help in the unequivocal assignation of these metabolites.

The identity of MS signals that were found to be significant in class separation was confirmed by liquid cromatography-MS/MS by using a QTOF (model 6520, Agilent). Experiments were repeated with identical chromatographic conditions as in the primary analysis. Ions were targeted for collision induced dissociation fragmentation on the fly based on the previously determined accurate mass and retention time. Comparison of the structure of the proposed compound with the fragments obtained can confirm the identity. Accurate mass data and isotopic distributions for the precursor and product ions can be studied and compared to spectral data of reference compounds, if available, obtained under identical conditions for final confirmation (HMDB, METLIN).

REFERENCES

1. Xi Y, Rocke DM: Baseline correction for nmr spectroscopic metabolomics data analysis. BMC Bioinformatics 2008;9:324

2. Izquierdo-Garcia JL, Rodriguez I, Kyriazis A, Villa P, Barreiro P, Desco M, Ruiz-Cabello J: A novel r-package graphic user interface for the analysis of metabonomic profiles. BMC Bioinformatics 2009;10:363

3. R_Development_Core_Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2010

4. Holmes E, Foxall PJD, Nicholson JK, Neild GH, Brown SM, Beddell CR, Sweatman BC, Rahr E, Lindon JC, Spraul M, Neidig P: Automatic data reduction and pattern recognition methods for analysis of 1h nuclear magnetic resonance spectra of human urine from normal and pathological states. Anal Biochemistry 1994;220:284-96

5. Hotelling H: Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24:417-41

6. Kramer R: Chemometric techniques for quantitative analysis. Marcel Dekker: New York, 1998; p. 203

7. Lin ZY, Xu PB, Yan SK, Meng HB, Yang GJ, Dai WX, Liu XR, Li JB, Deng XM, Zhang WD: A metabonomic approach to early prognostic evaluation of experimental sepsis by (1)H NMR and pattern recognition. NMR Biomed 2009;22:601-8

8. Gika HG, Macpherson E, Theodoridis GA, Wilson ID. Evaluation of the repeatability of ultra-performance liquid chromatography-tof-ms for global metabolic profiling of human urine samples. J Chromatogr B Analyt Technol Biomed Life Sci 2008;871:299-305

9. Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F. Assessment of PLSDA cross validation. Metabolomics. 2004;4:81–9

10. Ciborowski M, Teul J, Martin-Ventura JL, Egido J, Barbas C: Metabolomics with LC–QTOF-MS permits the prediction of disease stage in aortic abdominal aneurysm based on plasma metabolic fingerprint. PLoS One 2012;7:31982

11. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L: Hmdb: The human metabolome database. Nucleic Acids Res 2007;35:D521-6.