A metabolomic data fusion approach to support gliomas grading

A metabolomic data fusion approach to support gliomas grading

Valeria Righi[1], Nicola Cavallini[2], Antonella Valentini[3], Giampietro Pinna[3,4], Giacomo Pavesi[3,5], Maria Cecilia Rossi[6], Annette Puzzolante[3], Adele Mucci[2]*, Marina Cocchi[2]

[1]Dipartimento di Scienze per la Qualità della Vita, Università di Bologna, Campus Rimini, Corso D'Augusto 237, 49921 Rimini, Italy

[2]Dipartimento di Scienze Chimiche Geologiche, Università di Modena e Reggio Emilia, via G. Campi 103, 41125 Modena, Italy

[3]Dipartimento Integrato di Neuroscienze, Azienda Ospedaliero-Universitaria di Modena, Via Giardini 1355, Baggiovara, 41126 Modena Italy

[4]Current. Istituto di Neurochirurgia, Azienda Ospedaliera Universitaria Integrata Verona, Piazzale Aristide Stefani 1, 37126, Verona, Italy

[5]Dipartimento di Scienze Biomediche, Metaboliche e Neuroscienze, Università di Modena Reggio Emilia, via G. Campi 287, 41125 Modena, Italy

[6]Centro Interdipartimentale Grandi Strumenti, Università di Modena e Reggio Emilia, via G. Campi 213/A, 41125 Modena Italy

Corresponding author:

Prof. Adele Mucci, Dipartimento di Scienze Chimiche Geologiche Università di Modena e Reggio Emilia, via G. Campi 103, 41125 Modena, Italy

e-mail address: adele.mucci@unimore.it

tel. 00390592058636

Abstract

Magnetic resonance imaging (MRI) is the current gold standard for the diagnosis of brain tumors. However, despite the development of MRI techniques, differential diagnosis is often difficult in central nervous system (CNS) primitive pathologies such as lymphoma and glioblastoma or tumor-like brain lesions and glioma. MRI can be supported by *in vivo* magnetic resonance spectroscopy (MRS) to enhance its diagnostic power and multiproject-multicenter evaluations of classification of brain tumors have shown that an accuracy around 90% can be achieved for most of the pairwise discrimination problems. However, the survival rate is still low for patients affected by gliomas. The High-Resolution Magic-Angle-Spinning Nuclear Magnetic Resonance (HR-MAS NMR) metabolomics studies may be helpful for the discrimination of gliomas grades and the development of new strategies for clinical intervention. We propose to evaluate not only T_2-filtered spectra but also to try to extract as much information as possible from diffusion-filtered and conventional water-presaturated spectra fusing the data gathered by these different NMR experiments and applying a chemometric approach based on Multivariate Curve Resolution (MCR). Biomarkers important for glioma's discrimination were found. In particular, we focused our attention on cystathionine (Cyst) that seems an important biomarker for a better prognosis of glioma tumors.

Keywords: Gliomas, Brain tumors, Metabolomics**,** HR-MAS NMR, SIMCA, Multivariate Curve Resolution, Classification, Double cross-validation.

Abbreviations

World Health Organization (WHO), Multivariate Curve Resolution (MCR), Magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS), central nervous system (CNS), *N*-acetylaspartate (NAA), choline containing compounds (ChoCC), creatine (Cr), glutamate (Glu), glutamine (Gln), lactate (Lac), alanine (Ala), acetate (Ac), myo-inositol (Myo), glycine (Gly), taurine (Tau), glycerol in triacylglycerides ($Glyc_{tg}$), High-Resolution Magic-Angle-Spinning Nuclear Magnetic Resonance Spectroscopy (HR-MAS NMR), phosphocholine (PC), glycerophosphocholine (GPC), fatty acid (FA), phosphoethanolamine (PE), ß-glucose (Glc), arginine (Arg), isocitrate dehydrogenase 1 and 2 (*IDH1* and *IDH2*), Principal Component Analysis (PCA), Partial least squares-discriminant analysis (PLS-DA), Soft Independent Modelling of Class Analogies (SIMCA), cystathionine (Cyst), hypotaurine (HTau), 2-hydroxyglutarate (2OHG), cystathionine-β-synthase (CBS), cystathionine gamma-lyase (CTH).

1. INTRODUCTION

Primitive brain tumors are central nervous system tumors (CNS) and they are relatively rare (1-2% of human cancer). Every year in Europe about 8-10 new cases every 100,000 inhabitants are diagnosed, with a significant increase since 1970 in the industrialized Western countries. In the world, CNS tumors are ranked seventeenth for incidence. The probability of getting sick increases with age. When mortality is considered, brain tumors move up to the twelfth position.[1] Primitive brain tumors are mainly gliomas, with a number of histology subtypes: glioblastoma is the glioma of grade IV, which is the most invasive and frequent; anaplastic astrocytoma and anaplastic oligodendroglioma are gliomas of grade III and low-grade astrocytoma and low-grade oligodendroglioma are diffuse gliomas of grade II. Gliomas are classified into grades I to IV by the World Health Organization (WHO) using histopathological and clinical criteria. The new WHO 2016 classification prognosis is influenced by grading and biomolecular assessment in primary IDH mutation and MGMT methylation.[2,3] About 60-70% of primitive brain tumors are glioblastomas and their prognosis is still poor, with a median of 12 -14 months of overall survival despite the new surgical, oncological and radiotherapeutic technologies.

Magnetic resonance imaging (MRI) is the current gold standard for the diagnosis of brain tumors. However, despite the development of MRI techniques, differential diagnosis is often difficult in CNS primitive pathology such as lymphoma and glioblastoma or flogistic diseases and glioma or recurrent tumors and radionecrosis.[4] Further studies in neuro-oncology for diagnosis and care of brain tumors are needed. MRI can be flanked by *in vivo* magnetic resonance spectroscopy (MRS) to enhance its diagnostic power and multiproject-multicenter evaluations of classification of brain tumors have shown that an accuracy around 90% can be achieved for most of the pairwise discrimination problems.[5]

In vivo MRS provides, in principle, important metabolic information on tumors, detecting signals such as *N*-acetylaspartate (NAA), choline containing compounds (ChoCC) and creatine (Cr) and, in good quality spectra, also glutamate (Glu), glutamine (Gln), lactate (Lac), alanine (Ala) and other more overlapped signals from myo-inositol (Myo), glycine (Gly) and taurine (Tau). These metabolites are often aspecific and do not represent clear-cut markers for the different neuropathologies, even though recently Myo signals have been reported to be significantly higher in high-grade gliomas with respect to primary CNS lymphomas and lipid signals to be higher in gliomas with respect to other brain tumors.[6]

A complementary approach can be represented by *ex vivo* High-Resolution Magic-Angle-Spinning Nuclear Magnetic Resonance Spectroscopy (HR-MAS NMR) which enriches and strengthens the interpretation bases of *in vivo* spectra. The number of identifiable

metabolites found in *ex vivo* spectra using HR-MAS are about 10-fold those recognizable in *in vivo* spectra, due to higher resolution. For instance, HR-MAS NMR clearly shows, using two-dimensional techniques (COSY, TOCSY and HSQC), that the *in vivo* ChoCC peak receives contributions not only from choline (Cho), phosphocholine (PC) and glycerophosphocholine (GPC), but also from phosphoethanolamine (PE), (Myo, ß-glucose (Glc), Tau and arginine (Arg).[7,8]

The *ex vivo* HR-MAS NMR spectra most amenable for multivariate analysis are those T_2-filtered, obtained applying a Carr–Purcell–Meiboom–Gill (CPMG) sequence. They retain signals from small metabolites undergoing fast motion, and hence characterized by narrow linewidths, whereas signals from slowly tumbling species (usually lipids and macromolecules) that give broad spectral components, partially observed as baseline distortions, are filtered off. Detailed studies on different grade gliomas based on statistics applied to CPMG spectra have been reported.[9,10] Nevertheless, T_2-filtered signals represent only a part of the spectroscopic information, the part related to slowly diffusing molecules can be found in diffusion-filtered experiments and all the NMR visible metabolites contribute to the basic water-presaturated pulse-and-acquire spectra.

In this work, we evaluate not only T_2-filtered spectra but also try to extract as much information as possible from diffusion-filtered and conventional water-presaturated spectra fusing the data gathered by these different NMR experiments and applying the Multivariate Curve Resolution (MCR) approach on each experiment individually. MCR coupled with an aligning method is particularly important on tissues, since the chemical shifts of the signals of the same metabolite can slightly change in the different samples, for instance due to slight pH variations: these changes cannot be controlled by using buffers, as it is usually done in NMR on fluids, therefore the alignment step can be used for taking care of this potential issue.

The goals of this study are to investigate the different glioma grades and the biomarkers most important for their discrimination, in order to support the interpretation of *in vivo* MRS, thanks to the details that can be obtained by *ex vivo* HR-MAS NMR, to gain further insight into the biochemistry of the tumor subtypes, to help the detection of metabolic alterations in biofluids to be used in screening for early diagnosis and for therapy planning.

2. EXPERIMENTAL SECTION

2.1. Patient population

This study was approved by the local ethics committee (CE 131/10, approved in 07/09/2010). All patients received detailed information regarding the procedure and written

informed consent was obtained. Based on pre-surgical MRI parameters, image-guided tissue samples (n = 44) were collected from 35 enrolled subjects by biopsy or surgical resection, as summarized in Table 1.

Table 1. Tissue samples and patient population

Diagnose	**Number of patients (tissues samples)**	**Weight (mg)**	**Age**
Glioma grade IV	19 (24)	9.7÷70.7	45-75
Glioma grade III	8 (12)	7.9÷47.3	42-61
Glioma grade II	3 (3)	7.6÷19.8	42-47
Lymphoma	5 (5)	4.2÷25.8	35-78

The pre-surgery MRI were acquired and used to plan the tissue sample locations prior to surgery. Samples were collected from a specific tumoral area identified by MR images (the same analyzed by *in vivo* MRS) and transferred to the neuronavigation system (BrainLAB Inc., USA). Each sample was split into two parts; one was signed and used for standard histology, the other for HR-MAS NMR measurements. The last one was snap-frozen in liquid nitrogen and stored at -80 °C. At the time of sample collection, the mutations in isocitrate dehydrogenase 1 and 2 (*IDH1* and *IDH2*) were not assessed. Mutations in *IDH1* and *IDH2* and complete codeletion of chromosome arms 1p and 19q have been recognized as favorable prognostic molecular markers. As these markers have represented a major breakthrough in the diagnosis of brain tumors, they have been integrated into the 2016 WHO classification of gliomas.[3,11] Only for some patients with recurrent cancer was performed the IDH mutation analysis.

2.2. HR-MAS Nuclear Magnetic Resonance Spectroscopy

HR-MAS NMR spectra were recorded with a Bruker Avance400 spectrometer (Bruker BioSpin, Karlsruhe, Germany) operating at a frequency of 400.13 MHz for ^{1}H and 100.61 for ^{13}C. The instrument was equipped with a ^{1}H,^{13}C HR-MAS probe whose temperature was controlled by a Bruker Cooling Unit. The samples, 4-70 mg, were placed into a MAS zirconia rotor (4 mm outer d), added with 5-20 µL of D_2O to provide deuterium for the lock system, and the volume was adjusted to 50 µL (or 12 µL in the case of very small samples) using the proper insert and then transferred into the probe cooled to 5 °C to prevent tissue degradation processes. The set up takes about 20 min. Experiments were performed at a temperature of 5 °C, spinning the samples at 4 kHz and three different

types of one-dimensional proton spectra were acquired[12] using the sequences implemented in the Bruker software: (i) a composite pulse sequence (zgcppr, hereafter zg) with 2 s water presaturation during the relaxation delay, 8 kHz spectral width, 32k data points, 32 scans, (ii) a water-suppressed spin-echo CPMG sequence (cpmgpr, hereafter cpmg) with 2.5 s water presaturation during the relaxation delay, 1 ms echo time (τ) and 360 ms total spin–spin relaxation delay ($2n\tau$), 8 kHz spectral width, 32k data points, 64 scans; and (iii) a sequence for diffusion measurements based on stimulated echo and bipolar-gradient pulses with longitudinal eddy current delay (ledbpgp2s1d, hereafter led) with big delta 200 ms, eddy current delay Te 5 ms, little delta 2*2 ms, sine-shaped gradient with 32 G/cm followed by a 200 µs delay for gradient recovery, 8 kHz spectral width, 8k data points, 256 scans. Assignments of ^{1}H signals were based on previous work,[7] checked with standard two-dimensional homonuclear and heteronuclear techniques (COSY, TOCSY and HSQC) and compared with data in databases, mainly HMDB (http://www.hmdb.ca/ version 4.0) and BRMB (http://www.bmrb.wisc.edu/).

2.3. Multivariate Data Analysis

Three data sets were considered, corresponding to the three types of NMR spectra which were acquired, namely: led (4096 data points), cpmg (15000 data points) and zg (7500 data points). The data analysis workflow consisted of the following steps: (i) preprocessing of each single data set; (ii) features extraction applying MCR; (iii) exploratory data analysis of each single data set (not reported for sake of brevity) and on the fused data set by Principal Component Analysis (PCA); classification (iv) of the different glioma grades by two different approaches, namely class-modeling applying Soft Independent Modelling of Class Analogies (SIMCA) and discriminant analysis applying Partial Least Squares-Discriminant Analysis (PLS-DA). Considering the limited number of samples, a double cross-validation scheme was applied for model validation. The salience of metabolites in differentiating the tumor grades were assessed using the discriminant power[13] in SIMCA and variables important in projection (VIPs),[14] regression vectors and selectivity ratio[15,16] in PLS-DA. Each step is detailed in the following sub-sections.

2.3.1. NMR data preprocessing

The three different groups of spectra, namely zg, cpmg and led, were processed using the Bruker software, applying 0.5, 0.5 and 5 Hz line broadening, respectively, prior to Fourier transformation and then phase-corrected. The zg and cpmg spectra were referenced to the chemical shifts of the CH_3 doublet of alanine at δ 1.48 ppm, whereas the led spectra

were referenced to the CH_3 peak of fatty acid (FA) chain at δ 0.89 ppm.

The zg NMR data points were reduced from 16k to 8k. The selected spectral width spanned from δ 9 ppm to δ 0.05 ppm and the final number of data points was 7500. For the cpmg spectra, the NMR data points were reduced from 32k to 16k. The selected spectral width spanned from δ 9 ppm to δ 0.05 ppm and the final number of data points was 15000. Finally, for the led spectra, the NMR data points were reduced from 16k to 8k. The selected spectral width spanned from δ 9 ppm to δ 0.05 ppm and the number of final data points was 4k. All the NMR data were exported in ASCII format and imported in Matlab using a routine written by us.

Preprocessing of NMR data was then performed on each spectral data set individually. The procedure consisted of three steps: denoising, baseline correction and peak alignment. Denoising was performed in wavelet (WT) domain[17] applying a global threshold criterion[18,19] to the detail coefficients obtained by WT decomposition of the NMR spectra using a Daubechies 4 wavelet filter and 5 as decomposition level.

Baseline correction was then performed on the reconstructed denoised spectra by means of weighted asymmetric least-squares[20] fitting a fifth order polynomial. Next, the denoised and baseline-corrected spectra were aligned, first globally (*i.e.* rigidly shifting the whole spectrum) and then operating on a set of carefully defined intervals, which were individually aligned.

To this aim, the icoshift algorithm[21] was used. The option 'average'[22] was selected both for the global and interval alignments: this setting performs the alignment procedure two times in a row, leading to better aligned peaks.

2.3.2. Features extraction and data fusion

MCR[23] was successfully applied to extract features from the NMR spectra, allowing to obtain a total of 95 resolved components from the three data blocks, of which 49 were resolved from the zg spectral dataset, 27 from cpmg and 19 from led. A brief description of MCR is given hereinafter, while a detailed description can be found in the literature.[24] MCR is a decomposition method based on Beer's Law which allows obtaining a pure spectral profile S_i and a vector of relative concentrations C_i (one concentration value for each sample) for each *i*th-extracted component. The pure spectral profiles carry the information about the signal's shape and position and can be easily used for matching signals with metabolites by comparison with literature, spectral library data and previous knowledge.

MCR can recover pure contributions from overlapping signals applying the decomposition equation reported in Figure 1 and solving it by alternating least squares, starting from an

initial guess of either matrix C (concentrations) or S (spectral profiles).[23]

Given the complexity of the NMR spectrum, it has been demonstrated that working on intervals instead of the whole spectral width can improve interpretability and model performances. Differences due to baseline noise, signal intensity and metabolites can be more easily handled, and meaningful chemical quantifications can be obtained from each region.[25]

An initial spectral profiles S matrix was estimated by means of SIMPLISMA algorithm,[26],[27] with 10% maximum noise allowed. The non-negativity constraint was applied to both the concentrations and the profiles, using in both cases the fast non-negative least squares algorithm.[28] Finally, the S matrix was normalized using the Euclidean norm. Maximum number of iterations was 500, and in those cases when convergence was not reached within this limit, the profiles' evolution was inspected, to make sure that the components of interest (those related to NMR signals) were stable enough at the last iteration.

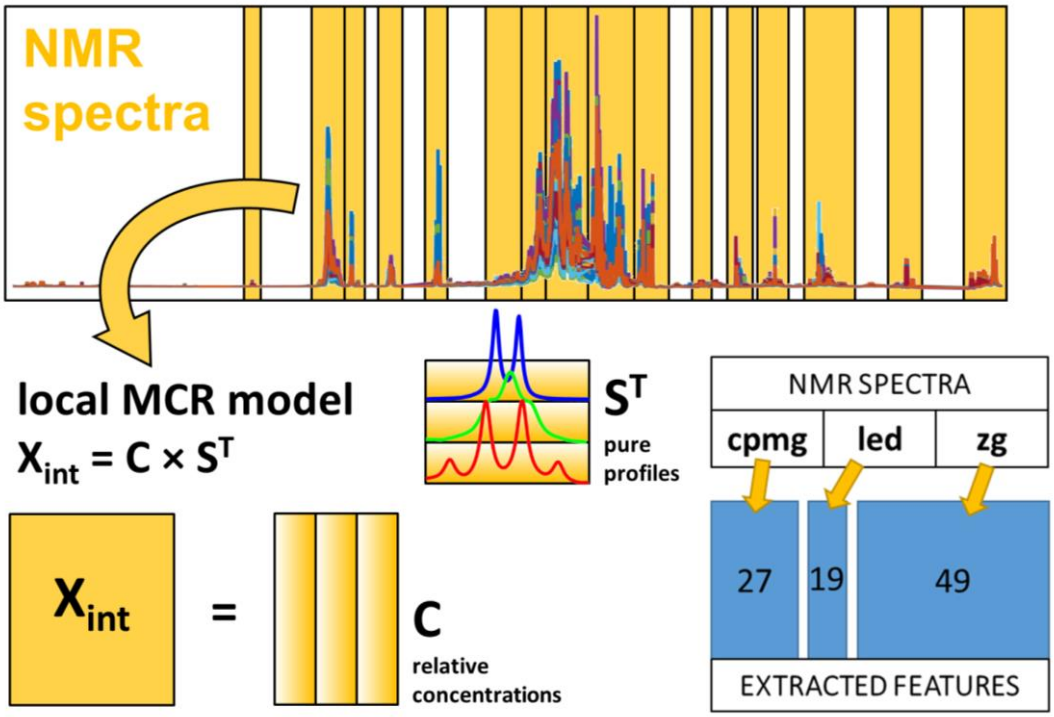

Figure 1. Extraction of chemical features from NMR spectra through interval-based MCR modeling.

For each defined interval a distinct local MCR model was developed: the S profiles of the resolved components were inspected to select meaningful spectral profiles, while suspicious and baseline-like profiles were discarded. The identification of resolved components was based on literature assignments, the digital libraries of HMDB and BMRB and our knowledge about the shape and position of the signals.[7]

For each dataset, the concentration vectors corresponding to the identified components were organized into a new data matrix (MCR extracted features). The features matrices were then normalized by the sample's mass, to account for differences in signals' intensity that could affect the correct sample characterization.

Once each data set (zg, cpmg and led) was resolved, the extracted features, named after the identified metabolites, were combined in a new dataset. This approach of combining features extracted from different sources is referred to as mid-level data fusion[29,30] and all the results shown hereafter were obtained from the data-fused matrix (represented in blue in Figure 1).

2.3.3. Exploratory analysis

Exploratory data analysis was performed using PCA [31,32] on the autoscaled (*i.e.* the mean value was subtracted from each dataset's column, which was then divided by its standard deviation) mid-level fused data set containing glioma samples of grade IV and III. Results are presented in the Results and Discussion section.

2.3.4. Classification analysis

Taking into account both the clinical relevance and the limited number of samples, especially of grade III, two distinct models were derived: (i) a one-class model for assessing if a glioma is of grade IV using SIMCA and (ii) a PLS-DA model to discriminate between grade IV and grade III gliomas. The basics of the two methods are briefly recalled.

2.3.5. SIMCA

SIMCA[13,33] belongs to class-modelling methods in which each class is individually modeled, independently from the presence of other classes. The focus of class-modeling is on *intra*-class similarities, *i.e.* similarities among samples belonging to the same class, and the classification rule consists of establishing if a sample belongs to the modeled class or not. To capture characteristic class variability, SIMCA builds a distinct PCA model for each class being modeled (it is worth recalling that preprocessing is also applied distinctly to each class) and the number of PCs is established independently for each class and may differ for each of them.

The classification rule is then based on verification of whether a sample is accepted or rejected by the given class model in terms of distance of the sample to the class model. Two distances are defined: (i) the scores distance (T^2), which indicates how far a sample is from the training objects of the class in the PC space, and (ii) the orthogonal distance,

which measures the distance of a sample to the PC space of the class and is given by the sum of squared residuals (Q). The implementations of SIMCA[34] may differ with respect to how these two distances are combined in an overall distance measure and on how the statistical acceptance limit, for this measure, is defined. In this work we used the implementation of the PLS_Toolbox (see Software section) that considers distinct statistics for T^2 and Q and bases the classification rule on the reduced distance D_r:

$$D_r = \sqrt{\left(\frac{T^2}{T^2_{lim}}\right)^2 + \left(\frac{Q}{Q_{lim}}\right)^2}$$

Based on that, a sample is accepted if its D_r from a given class is smaller or equal to the square root of 2 or rejected if it is higher. The classification performance of the model was evaluated in terms of class sensitivity, *i.e.* percentage of samples belonging to the class, which are correctly accepted, and specificity, *i.e.* percentage of samples not belonging to the class, which are correctly rejected.

The class dimensionality was established according to the minimum classification error in leave-one-out cross-validation, which corresponds to a one-component model. Validation of the model was done using Double cross-validation[35] described below. Interpretation of the SIMCA model in terms of salient metabolites for discriminating the samples with respect to the modeled classes can be accomplished calculating the variable discriminatory power.[13] This parameter is calculated comparing the square residual standard deviations of a variable calculated for the objects when projected on the model of the class they do not belong to with respect to that of the class they belong to. The concept is that we expect the residuals of samples on a given variable being higher when fitted to a different class model with respect to their own, if this variable is important in discrimination. Therefore, this variable will have a higher value of discriminatory power. Since the discriminatory power is not upper bounded, we have used both the 95 percentile and the average of the calibration samples, as thresholds to assess the significance of each variable. Only to calculate the discriminatory power a SIMCA model for the grade III class was also evaluated.

2.3.6. PLS-DA

PLS-DA is a discriminant method[36] in which discriminant boundaries among classes are computed. As opposed to SIMCA, this method focuses on the differences between the classes, so the classes are not individually modeled and therefore at least two classes have to be defined. Moreover, samples will necessarily be assigned to one of the modeled classes. PLS-DA is based on PLS regression[14] where the dependent variables are dummy

variables (one for each modeled class) taking values of 0 if the sample does not belong to the class and 1 when it belongs to it. Implementation of PLS-DA may differ on the basis of how the classification rule is defined. Here we use the rule to assign a sample to the class for which the predicted y-value is the highest (*i.e.*, Y predicted values are continuous and not dummy as they were codified).

The number of PLS-DA components was established according to the minimum classification error in leave-one-out cross validation, which corresponds to a two-components model. The performance of the model was established in terms of non-error classification rate, *i.e.* percentage of samples correctly assigned to the respective classes.

2.3.7. Model validation: Double cross-validation

Double cross-validation (double CV) is an iterative procedure based on two cross-validation loops, one “external” and one “internal”. The internal loop is used to optimize the complexity of the model, while the external one is used to obtain prediction values for each one of the samples. The dataset is initially split into balanced segments, and it is processed according to these steps:

1. one segment *j* is excluded;
2. the remaining *j–1* segments are used to build a series of models using a leave-one-out CV scheme (internal loop), to compute the optimal complexity of each *j*th model;
3. once the optimal complexity is obtained, the *j*th model is built and the segment excluded in step 1 is predicted (external loop);
4. a set of prediction errors is stored.

Using this procedure, each sample is excluded and predicted one time, so one prediction value for each sample is obtained. These prediction errors are used to estimate the predictive capability of the model. A final calibration model is then built whose complexity (number of components) was established taking the median of the optimal complexity (established by the internal CV loop) of the *j* models.

Considering the presence of different number of replicates for each sample, here the internal and external segments have been customized, taking always replicates in the same segment. Moreover, in the case of PLS-DA representativeness of both classes in each segment was ensured.

Summarizing: the grade IV and III samples were divided into four balanced splits for the external loop, and for SIMCA modeling, only grade IV samples were considered in the modeling step.

2.3.8. Software

The whole data analysis process was carried out on MATLAB 2017b (Mathworks, MA, USA). PCA analysis was performed using the PLS_Toolbox 8.6 (Eigenvector Research Inc. WA, USA) and WT denoising using the Matlab Wavelet Toolbox[19]. NMR spectral alignment was operated using icoshift,[21] (http://www.models.life.ku.dk/icoshift, last access 30/05/2019). NMR interval-resolution was operated by means of the MCR-ALS GUI by Joaquim Jaumot, Anna de Juan and Romà Tauler,[37] (https://mcrals.wordpress.com/2014/11/17/mcr-als-gui-2-0-reference-and-videos/, last access 30/05/2019). SIMCA analysis was performed using a MATLAB routine kindly provided by Prof. Federico Marini. PLS-DA classification rule, SIMCA discriminant power and double-CV for SIMCA and PLS-DA were calculated using code implemented by us in MATLAB.

3. RESULTS AND DISCUSSION

35 subjects were evaluated, 3 patients with glioma II, 8 with glioma III, 19 affected by glioma IV and 5 patients with preliminary diagnosis of glioma and then reclassified after histological analysis as lymphomas. More than 70% of patients died between the first and second year after the surgery. Despite numerous treatment strategies, these data show that the survival rate is still low and the NMR metabolomics studies may be helpful for discrimination of gliomas grades and the development of new strategies for clinical intervention. Three one-dimensional experiments were recorded for each sample (Figure 2). The first one is a conventional ^{1}H HR-MAS NMR spectrum with water presaturation (zg), that permits to detect the whole NMR visible metabolome of intact tissues, formed by both rapidly and slowly tumbling molecules that give narrow and broad lines, respectively. This is the real and complete fingerprint of a tissue. The second one is a T_2-filtered spectrum, obtained through a cpmg sequence that filters off broad resonances, characterized by short spin-spin relaxation times (*i.e.* T_2) and highlighting signals from small metabolites (mainly monosaccharides and polyols, aminoacids, organic acids and osmolites). The last one is a diffusion-edited led spectrum that retains the broad signals coming from slowly moving species (macromolecules and lipids) at expense of the narrow ones due to rapidly diffusing metabolites.

The majority of HR-MAS NMR studies reported in the literature relies on the cpmg type of spectra for they show resolved signals and flat baselines. Despite the predictable difficulties that are encountered in the processing of water-presaturated zg and diffusion-edited led spectra, we believe that limiting the statistical approach to cpmg data could limit

the spectroscopic information that can be in principle gathered from tissues. Hence, we decided to try to derive metabolic and grading knowledge from the combined statistical analysis of the three types of spectra. In this perspective, zg, cpmg and led spectra were processed for each sample as described in the experimental section.

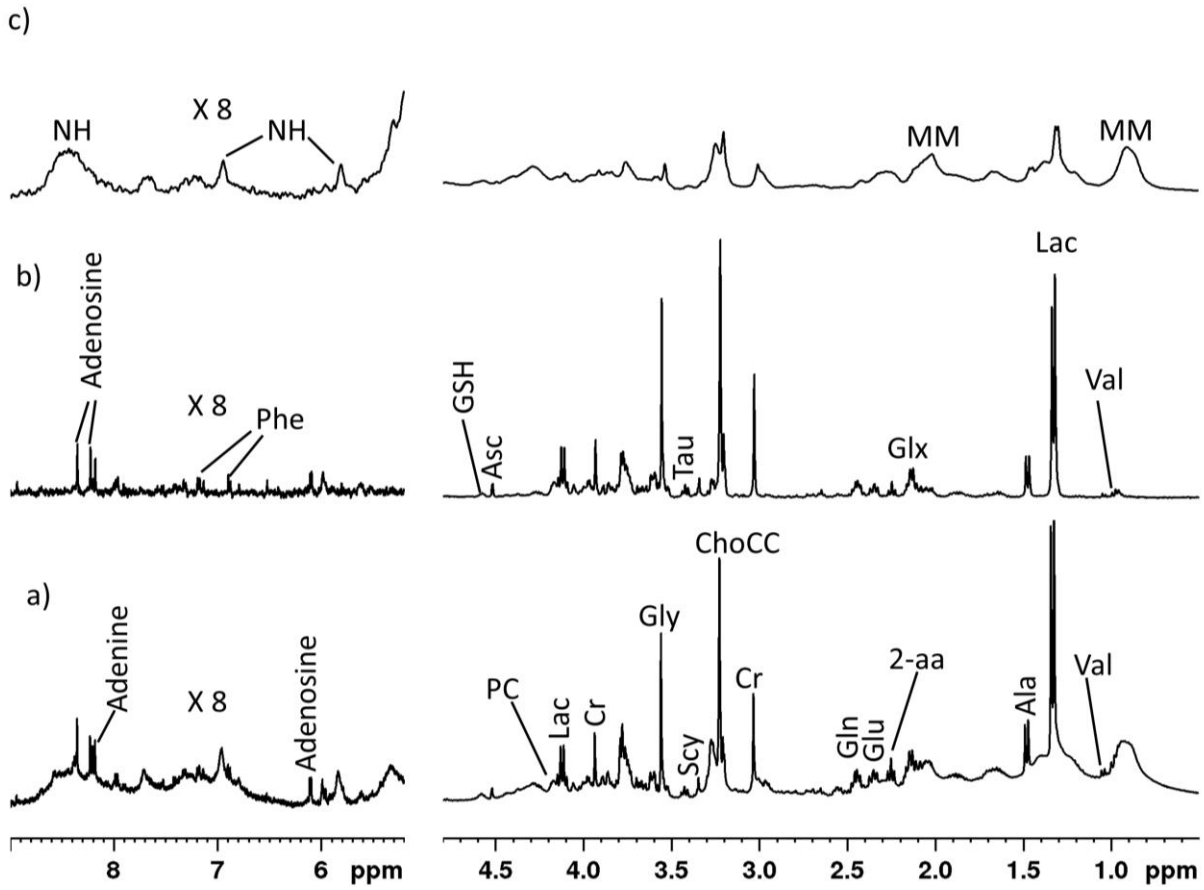

Figure 2. 1H NMR spectra of a grade IV sample obtained with a zg sequence (a), a cpmg sequence (b) and a led sequence (c). The first one displays both narrow and broad signals, the second one retains the narrow signals from fast tumbling molecules, whereas the third one only the broad resonances from slowly tumbling species. Major metabolites are labelled: alanine (Ala), 2-aminoadicip acid (2-aa), adenine, adenosine, ascorbate (Asc), choline containing compound (ChoCC), creatine (Cr), glutamine (Gln), glutamate (Glu), glutamine+glutamate (Glx), glutathione (GSH), glycine (Gly), lactate (Lac), macromolecules (MM), NH, phosphocholine (PC), phenylalanine (Phe), scylloinositol (Scy), taurine (Tau), valine (Val).

3.1. Exploratory analysis

A PCA model to explore the data was built prior to the classification step. Figure 3a shows a clear separation between grade III and IV samples, with some important exceptions. Three grade IV samples were found to be either extreme (04), extreme-borderline (28) or lying among the grade III samples (17). Two grade III samples also had interesting positions: sample 33-33_2 (replicates of the same specimen) was found in borderline

position, while sample 22 was lying among the grade IV samples. This can be better seen in Figure 3b.

The separation evidenced by the dash-dotted red curve (a guide for the eye) in Figure 3 correlates with neuroradiological characteristics of the two groups of tumors, in particular with the presence of necrosis in grade IV and with its absence and the reduced contrast enhancement in grade III. Samples 17 and 28 were not classified as grade III, but possess similar neuroradiological and prognostic characteristics.

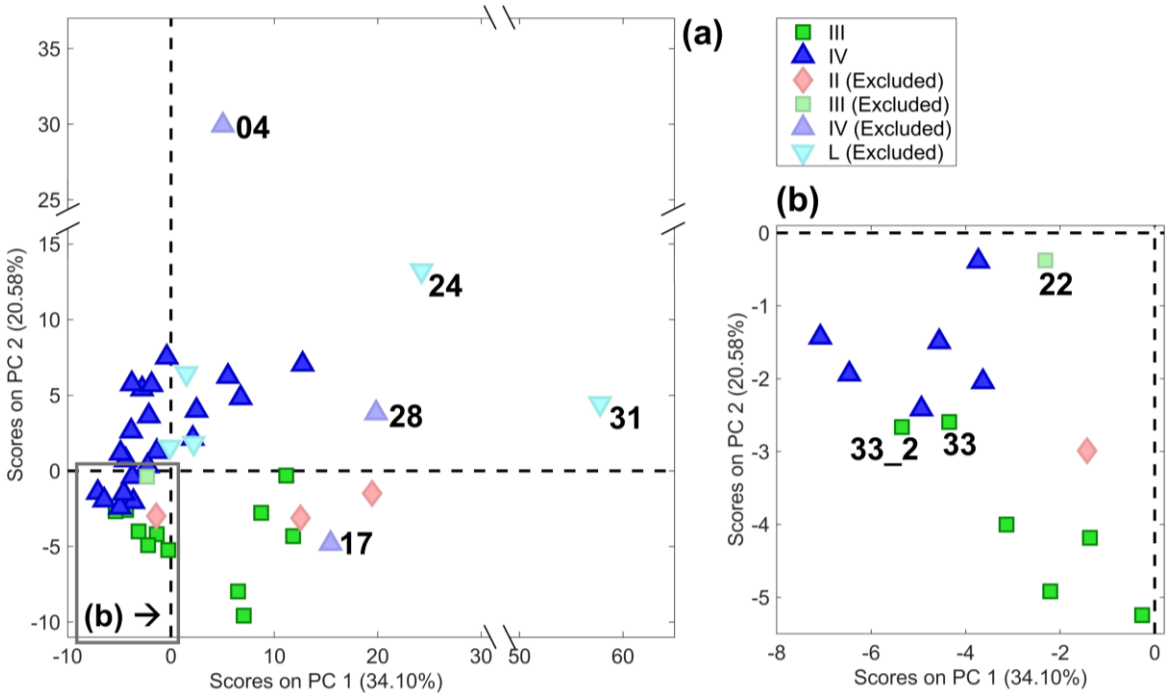

Figure 3. (a) PC1-2 score plot, also showing grade II, lymphoma and grade II and IV excluded samples; (b) zoom on the inter-class borderline zone, on samples 33 and 33_2. The dash-dotted red curve is a guide for the eye, see text.

Grade II gliomas and lymphomas were projected on this PCA model and they are also shown in Figure 3a. Lymphoma samples seem to be most similar to grade IV samples, while grade II gliomas are more similar to grade III gliomas.

It is evident that sample 04 is extreme, also compared to the other grade IV samples. Inspecting the corresponding T^2-contribution plot (Figure S1 reported in the Supplementary Materials), which shows the contribution of the original variables in determining the scores values of the samples, it emerges that its extreme position is mainly due to the significantly higher contributions from FAs. Sample 04 was obtained from a tumor lesion characterized by a cystic necrotic area, which could induce the high presence of FAs. Moreover, this sample, at the histological level, has features of atypical cells such as gemistocytic cells that define a different type of astrocytoma, that relapses

and has a poor prognosis.

Based on this analysis, three grade IV samples (04, 17, 28) and one grade III sample (22) were not considered in the calibration set for the classification step. A summary of spectra considered in classification models is presented on Table 2.

Table 2. Spectra used for classification analysis

Diagnose	**total number (44)**	**in SIMCA calibration set (21)**	**in PLS-DA calibration set (32)**	**excluded (12)**	excluded samples' labels
Glioma grade IV	24	21	21	3	*04, 17, 28*
Glioma grade III	12	0	11	1	*22*
Glioma grade II	3	0	0	3	
Lymphoma	5	0	0	5	

3.2. Classification analysis: SIMCA

Lymphomas and samples belonging to grade II were not considered in the classification modeling step, together with three samples of grade IV and one sample of grade III, as reported on Table 2. The total number of samples included for classification modelling was 23 (32 spectra), with 7 (11 spectra) grade III and 16 (21 spectra) grade IV. Due to the rather reduced and class-unbalanced numbers of samples, it was decided to build a single-class SIMCA model, focusing on grade IV. Moreover, instead of dividing the dataset into a training and a test set, the double CV approach was employed.

This allowed us to estimate sensitivity and specificity of the grade IV model in prediction using the classification assignments of the external loop. These results are reported in Table 3 and Figure 4.

The model wrongly rejected only three grade IV samples out of 21 (03_1, 16, 19). It is important to clarify that samples 03_1 and 16 lie within the acceptance limit in Figure 4, because this figure refers to their estimation when they are present in the calibration set. However, the above two samples were rejected during the external cross-validation assessment, which is the information to be used when referring to predictive capability for grade IV samples.

Table 3. SIMCA classification results in prediction

Diagnose	**Sensitivity[a]**	**Specificity[b]**	**labels of wrongly predicted samples**

Glioma grade IV	18/21 (86%)		*03_1, 16, 19*
Glioma grade III		10/11 (91%)	*33*
Glioma grade II		3/3	
Lymphoma		3/5	*23, 34*

[a] number of correctly accepted / total number of samples belonging to the class (in parentheses as percentage)

[b] number of correctly rejected / total number of samples of the given class (in parentheses as percentage)

The model wrongly accepted only one grade III sample (33) out of 11. It is interesting to notice that sample 33 was found in PCA (Figure 3b) very close to grade IV samples, together with its replicate 33_2, which was correctly rejected by the SIMCA model, but can be found very close to the acceptance limit in Figure 4. Samples 33 and 33_2 derive from the same cancer tissue, and their difference is probably due to the heterogeneity of the tissues.

The grade II glioma and lymphoma samples listed in Table 3 were also predicted by the SIMCA model, and the prediction results are reported in Table 4. All the grade II samples were correctly rejected, but two lymphoma samples (23, 34) were wrongly accepted.

A comparison of the metabolites' relative content (peak areas obtained by MCR) of lymphoma samples 23 and 34 with respect to the correctly rejected ones reveals that the two rejected samples have significantly lower content of all metabolites. Moreover, sample 05 was correctly rejected but it can be found close to the acceptance limit (Figure 4), and its profile resulted more similar to the two wrongly accepted samples (23, 34) than to the other two rejected ones (24, 31). Recently, a paper has been published about lymphoma tumor.[38] This paper focuses on the discrimination between malignant lymphomas and gliomas, and the authors report that it is usually difficult to preoperatively distinguish between the two. The use of MRS can be useful for preoperative diagnoses, and quantitative analysis is considered to be a valuable method for distinguishing between gliomas and malignant lymphomas. The limitation of the cited[38] and of our study is the number of samples.

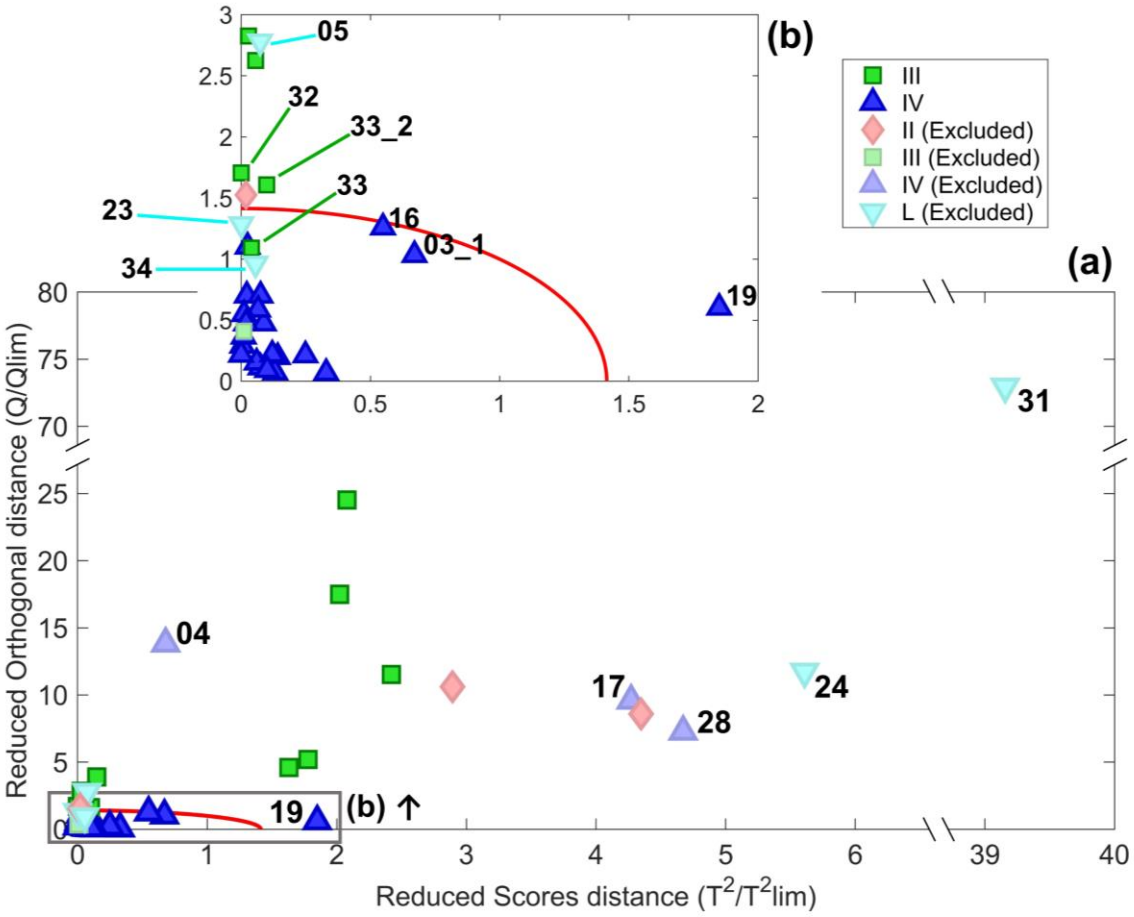

Figure 4. Reduced Orthogonal distance (Q/Qlim) vs. reduced Scores distance (T2/T2lim) of the SIMCA model built for the grade IV gliomas. The red circle highlights the acceptance area, which is delimited by the reduced distance threshold, computed using the formula reported in the paragraph about SIMCA, under the Experimental Section. Symbols are reported in the legend; the symbols with lighter colors represent the twelve projected samples, i.e. 3 glioma of grade II, 5 lymphoma, and the glioma grade III (1) and grade IV (3) listed in the last column of Table 3, which were not considered in modeling, as reported in PCA analysis.

As expected for their position in exploratory PCA model, the three glioma grade IV samples 04, 17, 28 (“excluded” in Table 2) were wrongly rejected and the glioma grade III sample 22 wrongly accepted. This last finding can be explained by the final diagnosis of 22 as xanthoastrocytoma.

3.3. Classification analysis: PLS-DA

Only samples belonging to grades III and IV were considered in this classification modeling step. As opposed to the SIMCA classification, it is not possible to predict samples belonging to classes different from grade III and IV using the PLS-DA model, as explained in the Multivariate data analysis section about PLS-DA. As with the SIMCA

classification, instead of dividing the dataset into a training and a test set, the double CV approach was used. The model performance results are reported in Table 4.

Table 4. PLS-DA classification results in prediction

Diagnose	**non-error rate[a]**	**labels of wrongly predicted samples**
Glioma grade IV	18/21 (86%)	*19 (IV)*
Glioma grade III	8/11 (73%)	*33_2, 21 (III)*

[a] number of correctly accepted / total number of samples belonging to the class (in parentheses as percentage)

[b] the Roman numbers in brackets are referred to the glioma grade to which the sample belongs to

As expected, the samples' distribution in the scores space of the PLS-DA model (Figure 5) resembles the one found in PCA (Figure 3), with a clear separation between the two classes. It is important to clarify the reason why the model wrongly predicted three samples, while Figure 5 shows a perfect separation between the two classes. As for the SIMCA model, the information obtained from the double CV is the one to be used when referring to the model's predictive capability. Figure 5 refers instead to the model built using all the samples altogether, which generally performs better than the single models built in cross-validation.

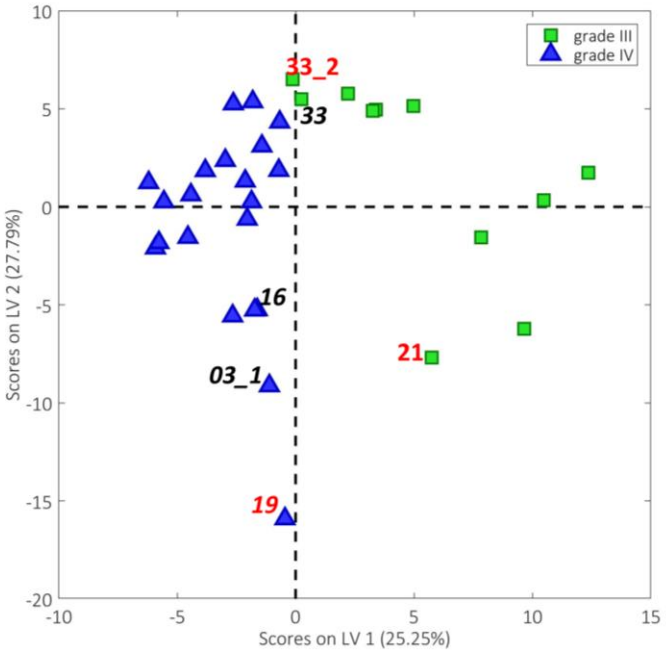

Figure 5. LV1-2 score plot showing grade III and grade IV samples. Labels are reported only for samples wrongly predicted using SIMCA (italics) or PLS-DA (red). Sample 19 is the only sample that was wrongly predicted by both methods.

It is interesting to inspect the position of the wrongly predicted samples. These samples result somehow extreme within their own class (Figure 5, labelled samples).

The PLS-DA model wrongly assigned sample 33_2 to grade IV, as opposed to the SIMCA model, which wrongly assigned sample 33 (a replicate of sample 33_2) to grade IV. These two replicates can be found close to each other in Figure 5, but also close to the grade IV samples. As already underlined, the difference found for these adjacent samples is probably due to the heterogeneity of the glioma sample. Furthermore, subject 33 was initially diagnosed with an oligodendroglioma that became an anaplastic tumor after a few years.

3.4. Significant metabolites

In this section the SIMCA and PLS-DA sets of discriminant metabolites are compared. The shared metabolites are reported on Figure 6.

Twenty-eight discriminant signals were recovered by SIMCA, thirty-six were recovered by PLS-DA and twenty-one were found in common between the two classification methods. Six signals related to FAs were found, and the main source for them were the led NMR spectra. Maleschlijski and co-workers[9] report that grade IV compared to grade III tumors have a higher content of FAs, and the same trend was found in our study, as it can be seen in Figure 6. In the same study, Cr and Myo are reported to be less abundant in grade IV compared to grade III tumors: this trend was also found in our study and it can be seen in Figure 6 as well.[9] Other metabolites such as NAA,[39] Cyst and acetate (Ac) follow the same descending trend (Figure 6).

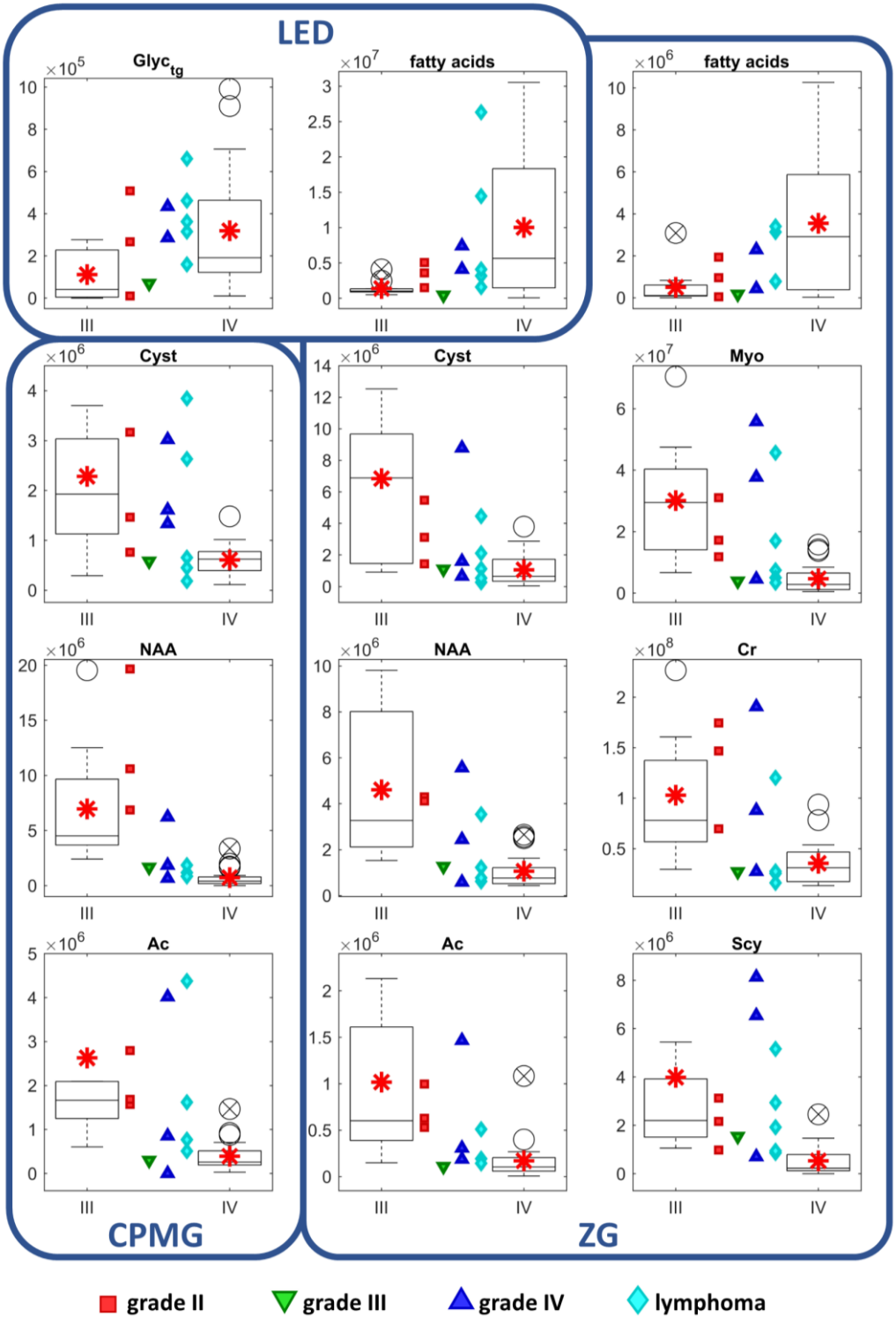

Figure 6. Boxplots of discriminant metabolites (the values reported correspond to the resolved peak areas), most in agreement between SIMCA and PLS-DA. The grade III and IV excluded from the classification, together with grade II and lymphoma samples, are plotted between each couple of box plots, for comparison.

These trends are also confirmed by Figure 7, where the discriminant metabolites important for grade IV (bars in blue) are mainly FAs, which are more abundant for this grade, while the other discriminant metabolites are important for grade III (bars in green) and are also

more abundant for this grade.

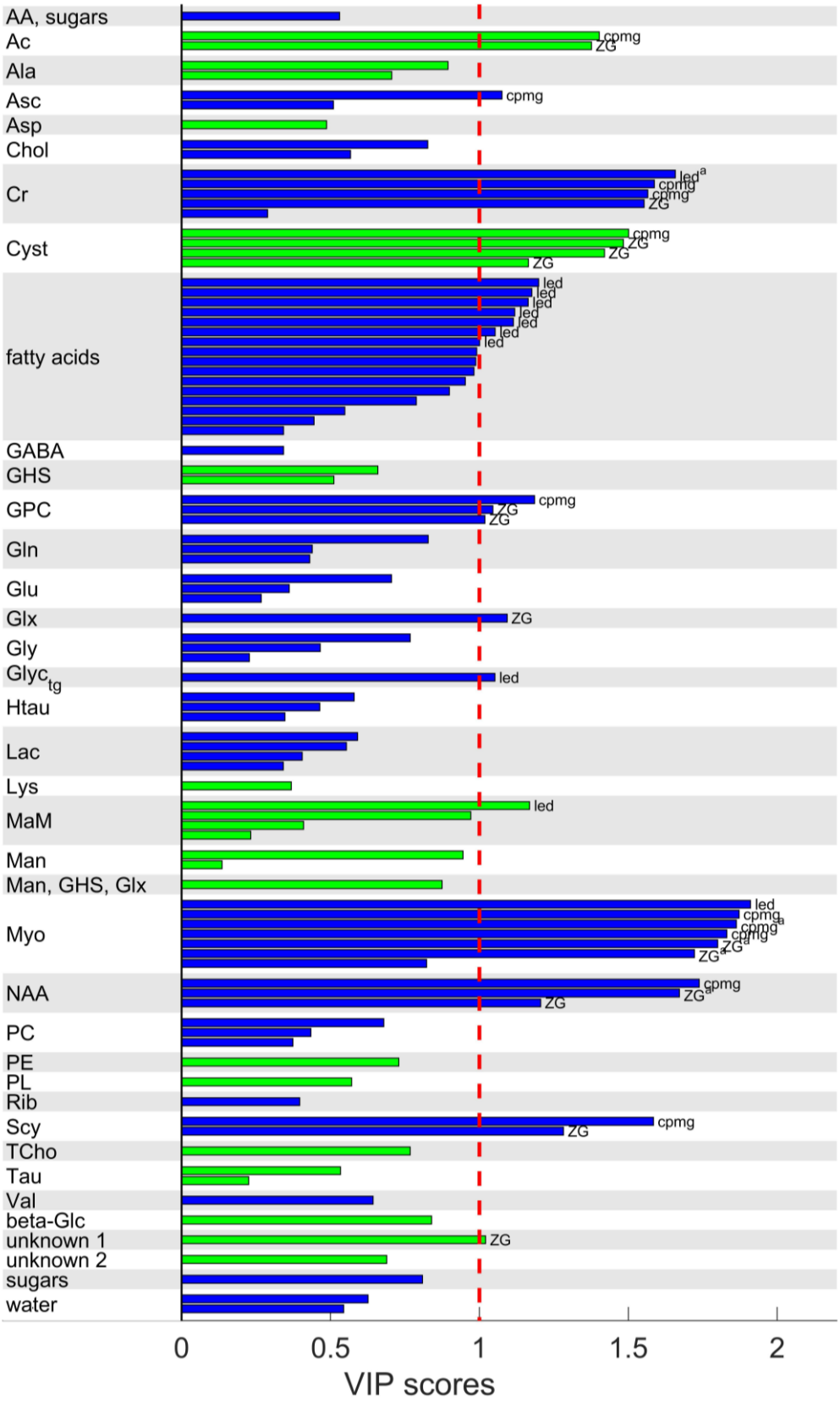

Figure 7. Variables important in prediction (VIPs) for class IV (blue) and class III (green), only metabolites exceeding the VIP threshold (=1) are labelled.

The role of FAs in the distinction between lymphomas and grade IV gliomas is still an open

question. It has been suggested that their signals can be useful to distinguish between malignant lymphomas and gliomas, selecting regions without necrosis, for malignant lymphomas and glioblastomas have different mechanisms for the generation of FAs.[38]
In the case of gliomas, a higher FA peak is due to cystic or necrotic components,[40] whereas, high lipids levels in malignant lymphomas seem to reflect the presence of activated or transformed lymphocytes and leukocytes that contain high levels of lipids.[41]

Most of the metabolites significant for the classification have already been reported, discussed and reviewed in the literature,[42] when dealing with *ex vivo* spectroscopic results: the enhancement of signals from lipids (FA and $Glyc_{tg}$, in this study) and the decrease of Cr, NAA and Myo. This study shed light on the importance of another metabolite, that is cystathionine (Cyst). Its presence in human brain is known[43] nevertheless its detection by HR-MAS NMR in human brain tumor has not been reported so far. We think that this could be due:

1. to the wide use in NMR metabolomics of ^{1}H cpmg spectra alone, where signals from Cyst [HMDB00099] are less intense than in zg spectra;
2. to the position of some of its signals, that are close to those of ethanolamine (at 3.13 ppm [HMDB00149]) and hypotaurine (HTau, at 2.66 ppm [HMDB00965]) and, if 2D experiments are not used to check them, these signals can be attributed to these two metabolites.

For this reason, we report in Figure 8 and 9 the fingerprints of Cyst and in Table S1 its chemical shifts (together with those of some selected metabolites). In details, the methylene protons of Cyst at 3.12 ppm correlate with a carbon at about 35 ppm, and not at 44 ppm as in ethanolamine, and the methylene protons of Cyst at 2.73 ppm correlate with a carbon at 30 ppm, and not at 58 ppm, as in the case of HTau. In some samples, both HTau and Cyst can be found and distinguished. In our samples instead, ethanolamine was present in small amounts and Cyst signals usually dominated ^{1}H spectra in the region around 3.1 ppm, as can be seen from Figure 8. The T_2 filter acts more on Cyst than on HTau, the signals of which become more evident in the cpmg spectrum.

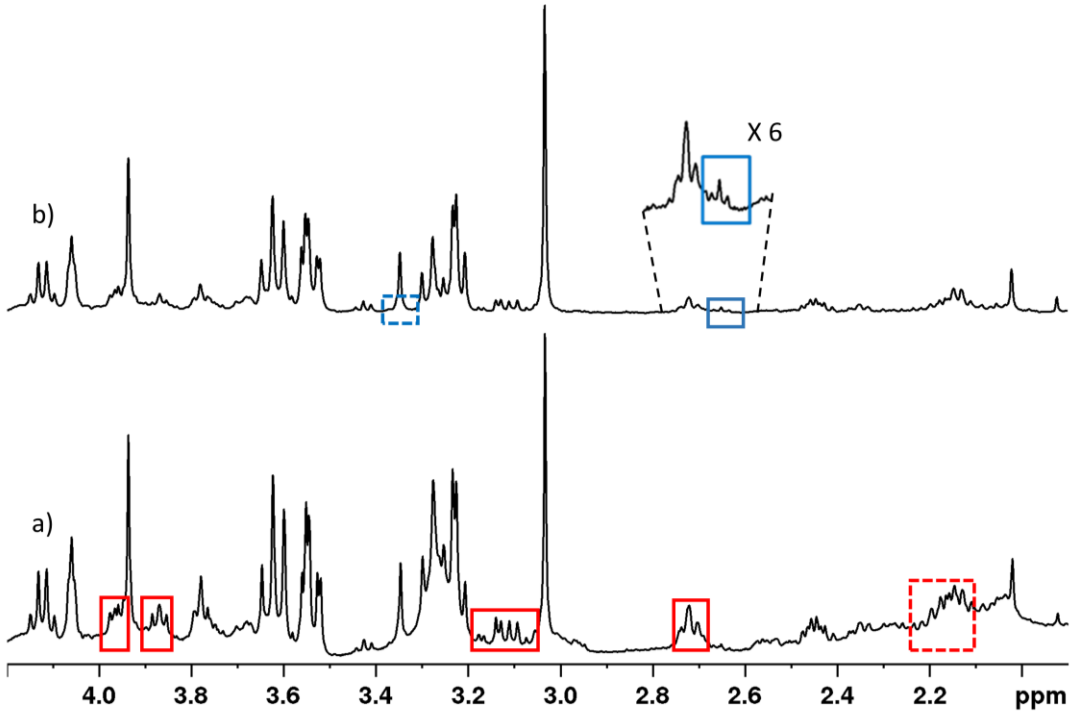

Figure 8. ^{1}H NMR water-presaturated spectra of a GIII sample obtained with a zg sequence (a) and a cpmg sequence (b). Signals from Cyst are marked by red squares. Signals from HTau are marked by blue squares. Those in broken line are highly overlapped to other signals.

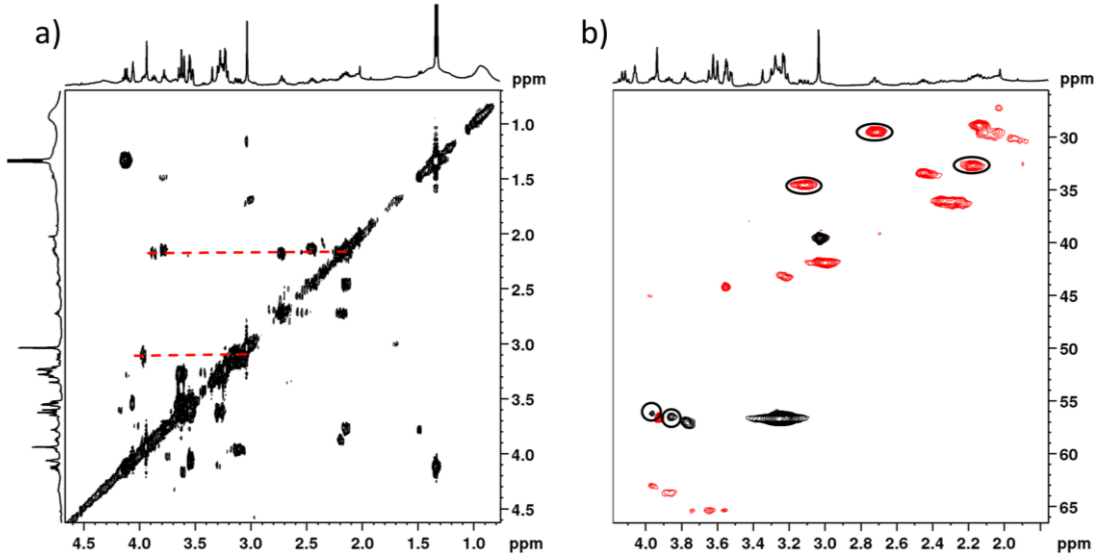

Figure 9. 2D NMR COSY (a) and HSQC (b) spectra of a GIII sample with Cyst correlations marked.

Only one recently published study has reported the first measurement of Cyst *in vivo* MRS.[44] The identification of Cyst was confirmed comparing *in vivo* spectra acquired gliomas (with *IDH* mutations) with the Cyst spectrum measured in a phantom. With the

present study we support these findings and the direct assignment of Cyst *in vivo* MRS.

Finally, we want to stress the importance of verifying 1D ^{1}H NMR assignments through selected 2D experiments, when possible. A signal (triplet) around 2.25 ppm can be attributed both to 2-aminoadipate and to 2-hydroxyglutarate (2OHG)[10] and the assignment can be done on the basis of TOCSY spectra (see Figure S2 and Table S1). TOCSY correlations are very helpful in this respect and allowed us to assign it to 2-aminoadipate in seven samples and to 2OHG only in one.

Cyst and 2OHG are considered important metabolites in cancer pathologies. Their role is related to IDH mutations. IDH mutations can be found in gliomas and are characterized by a specific cellular metabolism, causing the accumulation of 2OHG in tumor cells.[45,46]

Higher expression of cystathionine-β-synthase (CBS), the first enzyme of the transsulfuration pathway, has been associated with better prognosis in In *IDH-mutated* 1p/19q codeleted gliomas.[47]

Cyst derives from the condensation of homocysteine with serine catalyzed by CBS, and it is the initial and rate-limiting step in the transsulfuration pathway. Cyst is subsequently cleaved by the enzyme cystathionine gamma-lyase (CTH) to form cysteine, a precursor of GSH. Moreover, CBS participates in the desulfuration reactions that contribute to endogenous hydrogen sulfide production. Deregulation of CBS and the associated alterations in homocysteine and/or hydrogen sulfide levels leads to a wide range of pathological disturbances, and CBS activity also plays an important but complex role in cancer biology.[48]

Reduced serine biosynthesis may lead to increased reliance on the *CBS/CTH* pathway as a critical response to increased oxidative stress.[47,49] In particular, high *CBS* expression has been shown to confer better prognosis in *IDH*-mutated 1p/19q codeleted gliomas[47] in line with a previous study showing that decreased expression of *CBS* promotes glioma tumorigenesis in tumour xenografts.[50] Correlation between higher *CBS* expression and survival in *IDH*-mutated 1p/19q codeleted gliomas has been confirmed also from the POLA public dataset.[44] Fack *et al* have reported decreased Cyst in *IDH* mutant tumour xenografts compared with wild type, yet in the few samples reported in this last study, Cyst levels were roughly inversely correlated with codeletion status.[47]

4. CONCLUSIONS

This study highlights metabolic differences between anaplastic astrocytomas (grade III gliomas) and glioblastomas (grade IV gliomas), as identified at the time of the first diagnosis. The application of MCR allowed to efficiently extract important and meaningful

metabolic details from the brain tumour samples obtained using HR-MAS NMR. This approach is particularly important on tissues, since slight changes in the chemical shifts of the signals of the same metabolites are commonly observed in the different specimens, and they cannot be controlled using buffers, as usually done in NMR on fluid samples.

The three one-dimensional spectra datasets under examination (zg, cpmg and led) were analysed and modeled altogether, within the framework of data fusion.

Two classification models of different nature (SIMCA, class-modelling *vs* PLS-DA, discriminant analysis) were built, leading in both cases to good classification rates reported in Tables 3 and 4. Moreover, the two approaches provided valuable information about the metabolites important for the distinction between grade III and grade IV gliomas. Among these metabolites, Cyst surely is the most interesting result. Cyst and other related amino acids such as 2OHG identified in this study are relatively new in the MRS and HR-MAS NMR panorama and they seem to be good candidates as markers to monitor brain cancer prognosis and treatments.

Once again, we would like to underline that the use of 2D NMR spectra is mandatory for the correct assignment of metabolites with similar chemical shifts and signal shapes.

AKNOWLEDGEMENTS

Centro Interdipartimentale Grandi Strumenti of Università di Modena e Reggio Emilia is greatly acknowledged for the use of Bruker Avance400 spectrometer.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer.* 2014; 136:E359-E386.
2. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* 2007;114:97-109.
3. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*. 2016;131: 803-820.
4. Schlemmer HP, Bachert P, Henze M, Buslei R, Herfarth KK, Debus J, Von Kaick G. Differentiation of radiation necrosis from tumor progression using proton magnetic resonance spectroscopy. *Neuroradiology.* 2002;44:216-222.
5. García-Gómez JM, Luts J, Julià-Sapé M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-García E, Olier I, Postma G, Monleón D, Moreno-Torres À, Pujol J, Candiota A-P, Martínez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arús C, Robles M. Multiproject–multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magn Reson Mater Phys Biol*

Med. 2009;22:5-18.

6. Nagashima H, Sasayama T, Tanaka K, Kyotani K, Sato N, Maeyama M, Kohta M, Sakata J, Yamamoto Y, Hosoda K, Itoh T, Sasaki R, Kohmura E. Myo-inositol concentration in MR spectroscopy for differentiating high grade glioma from primary central nervous system lymphoma. *J. Neuro-Oncol.* 2018;136:317-326.
7. Tugnoli V, Schenetti L, Mucci A, Parenti F, Cagnoli R, Righi V, Trinchero A, Nocetti L, Toraci C, Mavilla L, Trentini G, Zunarelli E, Tosi MR. Ex vivo HR-MAS MRS of human meningiomas: a comparison with in vivo 1H MR spectra. *Int. J. Mol. Med.* 2006;18:859-869.
8. Righi V, Roda JM, Paz J, Mucci A, Tugnoli V, Rodriguez-Tarduchy G, Barrios L, Schenetti L, Cerdán S, García-Martín ML. 1H HR-MAS and genomic analysis of human tumor biopsies discriminate between high and low grade astrocytomas. *NMR Biomed.* 2009;22:629-637.
9. Maleschlijski S, Autry A, Jalbert L, Olson MP, McKnight T, Luks T, Nelson, S. Strategy for automated metabolic profiling of glioma subtypes from ex-vivo HRMAS spectra. *Metabolomics*. 2017;13:149.
10. Elkhaled A, Jalbert L, Constantin A, Yoshihara HAI, Phillips JJ, Molinaro AM, Chang SM, Nelson SJ. Characterization of metabolites in infiltrating gliomas using ex vivo 1H high-resolution magic angle spinning spectroscopy. *NMR Biomed*. 2014; 27:578-593.
11. Leeper HE, Caron AA, Decker PA, Jenkins RB, Lachance DH, Giannini C. IDH mutation, 1p19q codeletion and ATRX loss in WHO grade II gliomas. *Oncotarget.* 2015;6:30295-30305.
12. Righi V, Durante C, Cocchi M, Calabrese C, Di Febo G, Lecce F, Pisi A, Tugnoli V, Mucci A, Schenetti L. Discrimination of Healthy and Neoplastic Human Colon tissues by ex vivo HR-MAS NMR Spectroscopy and Chemometric Analyses. *J. Proteome Res.* 2009;8:1859-1869.
13. Wold S, Sjostrom M. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy, in *Chemometrics: Theory and Application.* Kowalski BR. ACS Symposium Series; American Chemical Society: Washington, DC, 1977. 243-282 pp.
14. Wold S, Johansson E, Cocchi M. PLS: partial least squares projections to latent structures, in 3D QSAR *Drug Design: Theory, Methods and Applications*. Kubinyi H. Ed., ESCOM Science Publishers: Leiden, 1993. 523-550 pp.
15. Rajalahti T, Arneberg R, Berven FS, Myhr KM, Ulvik RJ, Kvalheim OM. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemom. Intell. Lab. Syst.* 2009; 95:35-48
16. Andersen CM, Bro R. Variable selection in regression-a tutorial. *J. Chemom.* 2010; 24:728-737
17. Nounou MN, Bakshi BR. Multiscale Methods for Denoising and Compression, in *Wavelets in chemistry*. Walczak B., Ed, Elsevier Science, Amsterdam, 2000.119p.
18. Birgé L, Massart P. From Model Selection to Adaptive Estimation, in *Festschrift for Lucien Le Cam*, L. Springer, New York, 1997. 55-87 pp.
19. Misiti M, Misiti Y, Oppenheim G, Poggi J-M. Wavelet ToolboxTM 4 User's Guide Product enhancement suggestions Wavelet ToolboxTM User's Guide, 1997. www.mathworks.com
20. Eilers PHC. Parametric Time Warping. *Anal. Chem.* 2004;76:404-411.
21. Savorani F, Tomasi G, Engelsen SB. Icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* 2010;202:190-202.

22. Savorani F, Tomasi G, Engelsen SB. Alignment of 1D NMR Data using the iCoshift Tool: A Tutorial, in *Magn. Reson. Food Sci. Food for Thought,* van Duynhoven J, Belton PS, Webb GA, van As H, Eds., Royal Society of Chemistry. 2013. 14–24 pp.
23. De Juan A, Tauler R. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Crit. Rev. Anal. Chem.* 2006;36:163-176.
24. De Juan A, Tauler R. Multivariate curve resolution-alternating least squares for spectroscopic data. In *Data Handl. Sci. Technol.* Ruckebush C. Ed., Elsevier 2016.5-51pp.
25. Savorani F, Rasmussen MA, Rinnan Å, Engelsen SB. Interval-Based Chemometric Methods in NMR Foodomics. In *Data Handl. Sci. Technol.* Ruckebush C. Ed., Elsevier 2013. 449-486 pp.
26. Windig W, Heckler CE, Agblevor FA, Evans RJ. Self-modeling mixture analysis of categorized pyrolysis mass spectral data with the SIMPLISMA approach *Chemom. Intell. Lab. Syst.* 1992;14:195-207.
27. Windig W. Two-Way Data Analysis: Detection of Purest Variables. In *Compr. Chemom*. 2009. 275-307 pp.
28. Bro R, De Jong S. A fast non-negativity-constrained least squares algorithm. *J. Chemom*. 1997;11:393-401.
29. Silvestri M, Bertacchini L, Durante C, Marchetti A, Salvatore E, Cocchi M. Application of data fusion techniques to direct geographical traceability indicators. *Anal. Chim. Acta*. 2013; 769:1-9.
30. Biancolillo A, Bucci R, Magrì AL, Magrì AD, Marini F. Data-fusion for multiplatform characterization of an italian craft beer aimed at its authentication. *Anal. Chim. Acta*. 2014;820:23-31.
31. Li Vigni M, Durante C, Cocchi M. Exploratory Data Analysis. In *Data Handl. Sci. Technol.* Ruckebush C. Ed., Elsevier. 2013. 55-126pp.
32. Bro R, Smilde AK. Principal component analysis. *Anal. Methods*. 2014;6:2812-2831.
33. Bevilacqua M, Bucci R, Magrı` AD, Magrı` AL, Nescatelli R, Marini F. Classification and Class-Modelling. In *Data Handl. Sci. Technol.* Marini F. Ed., Elsevier. 2013;28:215-224.
34. Durante C, Bro R, Cocchi M., A classification tool for N-way array based on SIMCA methodology. *Chemom. Intell. Lab. Syst*. 2011;106:73-85.
35. Varmuza K, Filzmoser P. Introduction to multivariate statistical analysis in chemometrics. *CRC Press*, United States, 2009.
36. Indahl UG, Martens H, Naes T. From dummy regression to prior probabilities in PLS-DA. *J Chemometr*. 2007;21;529-536.
37. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: New features and applications. *Chemom. Intell. Lab. Syst*. 2015;140:1-12.
38. Ohba S, Murayama K, Abe M, Hasegawa M, Hirose, Y. Magnetic Resonance Imaging and Proton Magnetic Resonance Spectroscopy for Differentiating Between Enhanced Gliomas and Malignant Lymphomas. *World Neurosurgery*. 2019; 127:e779-e787
39. Ghasemi K, Khanmohammadi M, Saligheh HR. Accurate grading of brain gliomas by soft independent modeling of class analogy based on non-negative matrix factorization of proton magnetic resonance spectra. *Magn. Reson. Chem*. 2016;54:119-125.
40. Yamasaki F, Takayasu T, Nosaka R, Amatya VJ, Doskaliyev A, Akiyama Y, Tominaga A, Takeshima Y,

Sugiyama K, Kurisu K. Magnetic resonance spectroscopy detection of high lipid levels in intraaxial tumors without central necrosis: a characteristic of malignant lymphoma. *J Neurosurg*. 2015;122:1370-1379.

41. Chawla S, Zhang Y, Wang S, Chaudhary S, Chou C, O'Rourke DM, Vossough A, Melhem ER, Poptani H. Proton Magnetic Resonance Spectroscopy in Differentiating Glioblastomas From Primary Cerebral Lymphomas and Brain Metastases. *J Comput Assist Tomogr.* 2010;34:836-841.
42. Dietz C, Ehret F, Palmas F, Vandergrift LA, Jiang Y, Schmitt V, Dufner V, Habbel P, Nowak J, Cheng LL. Applications of high-resolution magic angle spinning MRS in biomedical studies II—Human diseases. *NMR Biomed.* 2017;30: doi: 10.1002/nbm.3784.
43. Gjessing LR. Studies of Functional Neural Tumors: III. Cystathionine in the Tumor Tissue. *Scandinav. J. Clin. & Lab. Investigation.* 1963;15:479-482.
44. Branzoli F, Pontoizeau C, Tchara L, Di Stefano AL, Kamoun A, Deelchand DK, Valabrègue R, Lehéricy S, Sanson M, Ottolenghi C, Marjańska M. Cystathionine as a marker for 1p/19q codeleted gliomas by in vivo magnetic resonance spectroscopy*. Neuro-Oncology***.** 2019; 21:765-774.
45. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science.* 2008;321:1807-1812.
46. Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, Marks KM, Prins RM, Ward PS, Yen KE, Liau LM, Rabinowitz JD, Cantley LC, Thompson CB, Vander Heiden MG, Su SM. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462:739-744.
47. Fack F, Tardito S, Hochart G, Oudin A, Zheng L, Fritah S, Golebiewska A, Nazarov PV, Bernard A, Hau AC, Keunen O, Leenders W, Lund-Johansen M, Stauber J, Gottlieb E, Bjerkvig R, Niclou SP. Altered metabolic landscape in IDH-mutant gliomas affects phospholipid, energy, and oxidative stress pathways. *EMBO Molecular Medicine*. 2017;9:1681-1695.
48. Zhu H, Blake S, Chan KT, Pearson RB, Kang J. Cystathionine β-Synthase in Physiology and Cancer. *BioMed Res. Int.* 2018. doi: 10.1155/2018/3205125.
49. Branzoli F, Di Stefano AL, Capelle L, Ottolenghi C, Valabrègue R, Deelchand DK, Bielle F, Villa C, Baussart B, Lehéricy S, Sanson M, Marjanska M. Highly specific determination of IDH status using edited in vivo magnetic resonance spectroscopy. *Neuro Oncol.* 2018;20:907-916.
50. Takano N, Sarfraz Y, Gilkes DM, Chaturvedi P, Xiang L, Suematsu M, Zagzag D, Semenza GL. Decreased expression of cystathionine β-synthase promotes glioma tumorigenesis. *Mol Cancer Res.* 2014;12:1398-1406.