

Semi-supervised source extraction methodology for the nosological imaging of glioblastoma response to therapy

Sandra Ortega-Martorell
Department of Mathematics and Statistics
Liverpool John Moores University
Liverpool, United Kingdom
S.OrtegaMartorell@ljmu.ac.uk

Ivan Olier
Manchester Institute of Biotechnology
The University of Manchester
Manchester, United Kingdom
ivan.olier@manchester.ac.uk

Teresa Delgado-Goñi
The Institute of Cancer Research
London, United Kingdom
Teresa.DelgadoGoni@icr.ac.uk

Magdalena Ciezka
Departament de Bioquímica i Biologia Molecular
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Spain
magdalena.ciezka@uab.cat

Margarida Julià-Sapé
Networking Research Center on Bioengineering,
Biomaterials and Nanomedicine (CIBER-BBN)
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Spain
margarita.julia@uab.cat

Paulo Lisboa
Department of Mathematics and Statistics
Liverpool John Moores University
Liverpool, United Kingdom
P.J.Lisboa@ljmu.ac.uk

Carles Arús
Departament de Bioquímica i Biologia Molecular
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Spain
carles.arus@uab.es

Abstract—Glioblastomas are one the most aggressive brain tumors. Their usual bad prognosis is due to the heterogeneity of their response to treatment and the lack of early and robust biomarkers to decide whether the tumor is responding to therapy. In this work, we propose the use of a semi-supervised methodology for source extraction to identify the sources representing tumor response to therapy, untreated/unresponsive tumor, and normal brain; and create nosological images of the response to therapy based on those sources. Fourteen mice were used to calculate the sources, and an independent test set of eight mice was used to further evaluate the proposed approach. The preliminary results obtained indicate that was possible to discriminate response and untreated/unresponsive areas of the tumor, and that the color-coded images allowed convenient tracking of response, especially throughout the course of therapy.

Keywords—*non-negative matrix factorization; semi-supervised source extraction; nosological imaging; brain tumors; glioblastoma; response to therapy*

I. INTRODUCTION

Glioblastoma (GBM) are high-grade malignant brain tumors that evolve fast, killing patients in less than 12-15 months. None of the classical or experimental available therapies cures GBM in humans or in preclinical models. Nowadays, the standard clinical therapy procedure includes maximal safe tumor resection, radiotherapy and chemotherapy with temozolomide (TMZ) –an alkylating drug that interferes with DNA replication [1] and leads to apoptotic cell death. TMZ administered alone induces cell proliferation arrest in genetically engineered mice [2]. Some causes for the usual bad prognosis of GBM are the heterogeneity of their response to treatment and the lack of early and robust biomarkers to decide whether the tumor is responding to therapy.

Given that magnetic resonance (MR) signals (i.e. spectroscopy, MRS / spectroscopic imaging, MRSI) may represent a mixture of tissues, signal source extraction is a relevant approach for identifying them from the spectroscopic signal. In [3], the authors proposed a methodology to guide the source extraction in the direction of the class prototypes of brain tumors. Here, we propose to use this methodology in a

different context, which is guiding the source extraction in the direction of the glioblastoma response to therapy. We also propose to generate color-coded maps [4] displaying the spatial representation of these sources, which are nosological images of the response to therapy.

II. MATERIALS

A total of fourteen C57BL/6 female mice of 20-23g in weight, inoculated with GL261 mouse glioma cells as in [5], were used in the first stage of this study, which is the source extraction in a semi-supervised way. MR studies were carried out at 7T, with isoflurane anaesthesia (1.5-2%) at 37°C. Magnetic resonance images (MRI, T_{2w} , TR/TEeff 4200/36ms) were acquired for tumor volume measurement. Parameters for MRSI were as in [5]. Eight of the fourteen mice (C415, C418, C437, C525, C527, C575, C584, and C586) were treated following a previously developed therapy protocol for preclinical glioblastoma using GL261 tumors, consisting of three temozolomide (TMZ) cycles (given on the following days post-implantation: 11-15, 19-20 and 24-25), which induces response as tumor growth arrest, as detected by MRI [6]. The other six mice (C255, C288, C351, C520, C529, and C583) did not receive any treatment and were used as the control group. We validated the obtained results against histopathology.

A further set of mice was used in the second stage of this study, which involves the use of the semi-supervised sources to produce nosological images in cases not involved in the training stage from which the sources were obtained. This new test set is composed by seven well-known untreated mice from a previous study [4]; and one additional treated mouse (C819). The latter was handled (tumor induction and treatment) and scanned exactly as the treated mice from the training set, and with similar characteristics, but instead of acquiring one MRSI as in the training set, five MRSI were acquired at different time points (i.e. days 16, 18, 22, 36,45 post GL261 cells inoculation (pi) in C57BL/6 mice).

III. METHODS

In this study, we apply the semi-supervised methodology proposed in [3] for the extraction of meaningful source signals. These sources are used later on for the generation of nosological images of the response to therapy.

A. Non-negative Matrix Factorization

Non-negative Matrix Factorization (NMF) [7], [8] methods is a group of multivariate data analysis techniques aimed to estimate meaningful latent components, also known as sources, from non-negative data. Standard NMF methods decompose the data matrix X into two non-negative matrices S (the sources) and A (the mixing matrix). The differences between them is given by the different cost functions used for measuring the divergence between X and $S \cdot A$. In [9] the authors present a variant of NMF, namely Convex-NMF, in which the basis vectors of S are constrained to be convex combinations of the data points. Convex-NMF relaxes the NMF constraints to allow negative values both in the data matrix and the sources.

B. Semi-supervised extraction of sources signals

The semi-supervised methodology in [3] proposes to take benefit from the use of prior knowledge derived from class membership of the spectra to guide the source extraction. It involves three main stages:

- i. The definition of a Fisher Information (FI) metric [10] to model pairwise similarities and dissimilarities between data points, using a Multi-Layer Perceptron (MLP) classifier to estimate the conditional probabilities of class membership.
- ii. The approximation of the empirical data distribution in a Euclidean projective space in which NMF-based techniques can be applied. This is done in this study with Multidimensional Scaling methods, specifically with the iterative majorization algorithm [11], [12].
- iii. The application of Convex-NMF for the source decomposition of the data.

C. Nosological imaging of the response to therapy

For the generation of color-coded nosological images of each individual mouse we used the values of the mixing matrix, as they represent the proportions in which each source is being represented in each voxel. We produced nosological images throughout the course of therapy to track response changes.

D. Experimental settings

For the extraction of the sources with the semi-supervised methodology, we used a total of 508 spectral vectors from 14 mice, 8 of them treated and showing transient response to TMZ, and 6 of them from the control group. Animals treated with TMZ (two or three cycles) were expected to show a response pattern, based on tumor volume growth arrest, and increased survival in comparison to the control group [6]. The selected spectra corresponded to the subsets of voxels labeled as tumor (responding and untreated) and normal parenchyma, assuming that all the voxels inside each tumor belonged to the same category (responding or control). The robustness of this assumption was later on evaluated using the test set described before. Most of the voxels in the edges of the grid were discarded (except in case C586) due to low signal to noise ratio (SNR); and those located in the tumor boundaries with normal tissue were also discarded to avoid spectral pattern mixing or contamination from different tissue types essentially as described in [4], [13].

The 80% of those 508 spectra (randomly selected) were used to create the MLP model, which was assessed with the remaining 20% of them. This model was used to estimate the conditional probabilities of class membership for each case, which were then used to define the FI metric. Three sources were calculated then following the three steps of the semi-supervised methodology mentioned before, one of them to represent the responding tumor, another to represent the tumor without treatment, and a third for the normal tissue.

As the sources were calculated from a subset of voxel of different mice, the resulting mixing matrix is not useful for nosological imaging purposes. For the latter, we fixed the set

of extracted sources and calculated their corresponding mixing matrices. Equation (9) in [3] provided us with a mechanism to determine the extent to which a fixed set of sources are encoded in a new data set, facilitating the calculation of the new mixing matrices, and with them, the nosological images. We also produced an overall image compiling the results of the individual nosological maps of the sources, in which each voxel is colored with the color that represents the source with the maximum value of mixing matrix for that particular voxel. Figure 1 summarizes the general representation of the semi-supervised methodology and how we used it in this study to track response to therapy.

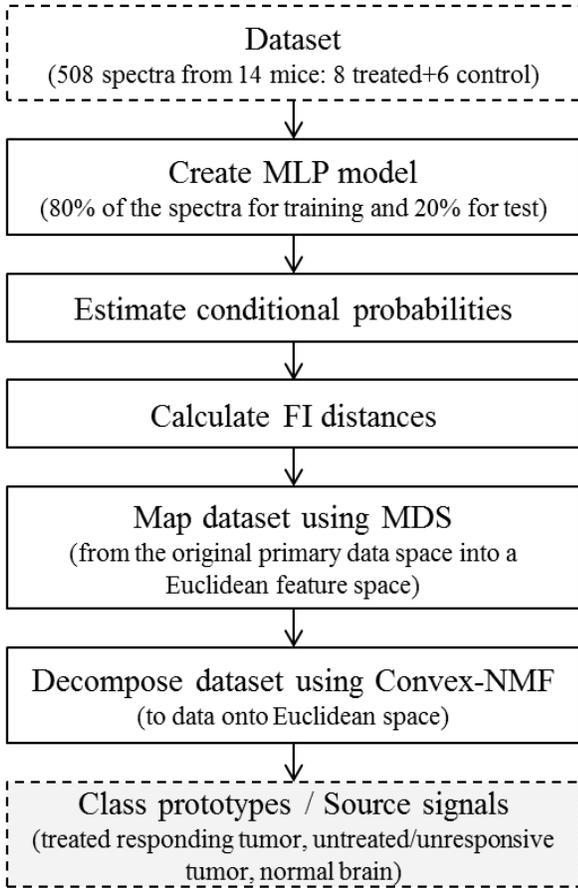


Fig. 1. General representation of the semi-supervised methodology used in this study to track response to therapy.

IV. PRELIMINARY RESULTS

A. Source signals calculated

Figure 2 illustrates the source signals calculated in a semi-supervised way, which represent the normal brain tissue and the treated (mainly in response to therapy) and untreated (actively proliferating) tumor regions.

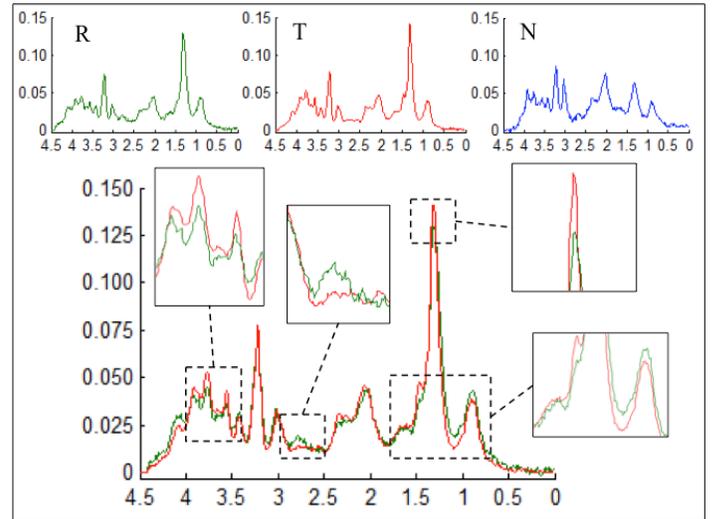


Fig. 2. The three sources calculated with the semi-supervised methodology (see top row): green represents treated responding tumor (R); red, untreated/unresponsive tumor (T); and blue, normal tissue (N). The main differences between sources R and T are also detailed (bottom). Horizontal axes indicate the frequencies measured in ppm.

B. Nosological imaging of the response to therapy

As explained before in the experimental settings section of Methods, the new mixing matrix calculated for each mouse fixing the sources are used in this study to generate color-coded nosological images. This section shows preliminary results on the use of these images to monitor response to therapy in a particular time point, and throughout the course of therapy.

For illustration purposes, in Figure 3 we show the results of three mice from the training set, two of them treated and one control. It shows the region of interest of each mouse with the original labeling (the tags given by the expert spectroscopist); the three nosological color-coded maps associated to the sources T, R and N; and the resulting nosologic image grouping the information of these maps, in which green represents response to therapy, red represents tumor (untreated or without response), and blue represents normal brain.

To test the validity of the methodology to produce nosological maps in new cases, we first use the seven control mice of the test set from [4] to check that they are correctly being represented by the tumor source. Figure 4 compiles the results.

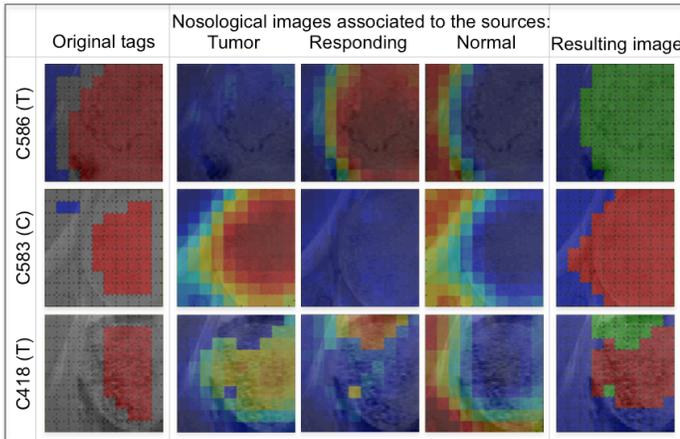


Fig. 3. Nosologic imaging for three mice from the training set. First column shows the map of original tags where red means (treated/untreated) tumor and blue means normal brain. The three color-coded maps from the middle are the nosological images associated to each source, where the scale of colors moves from red to blue, red indicates high values in the mixing matrix while blue indicates low values. Last column compiles the results from the previous three maps into one, in which the red color is for untreated tumor, green for responding tumor, and blue for normal brain. All the maps are superimposed over the corresponding T2w MR image. Please note the intermediate mixing matrix values for the tumour region in C418 (22 days pi).

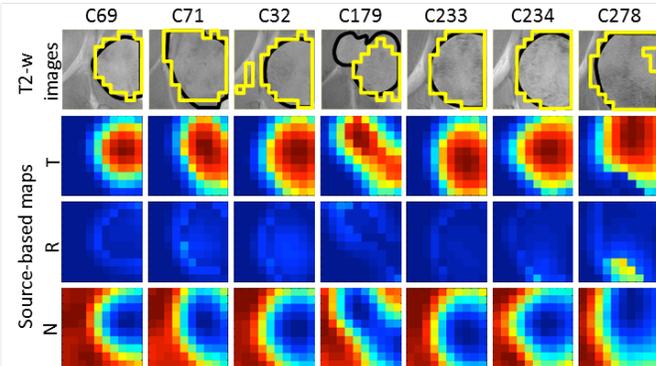


Fig. 4. Nosological maps (rows 2-4) associated to the tumor (T), responding (R) and normal (N) sources for the seven untreated mice of the independent test set. The top row shows the delineation of the tumor area as suggested by the MRI-abnormal region (black solid line) and the Proliferation Index (PI)>30% regions (yellow solid line), superimposed to the T2w images.

Secondly, we test the proposed approach in new, treated data, to show the usefulness of its use to track response to therapy. Very preliminary results with one mouse (C819) monitoring its response throughout the course of a therapy are shown in Figure 5. This figure shows the nosologic color-coded maps at selected time points (days 16, 18, 22, 36 and 45) during and after the three cycles of treatment. These days are highlighted (black triangles) in Figure 6, where we show how the tumor volume changes in both the control group of GL261 tumor bearing mice and a treated (C819) mouse.

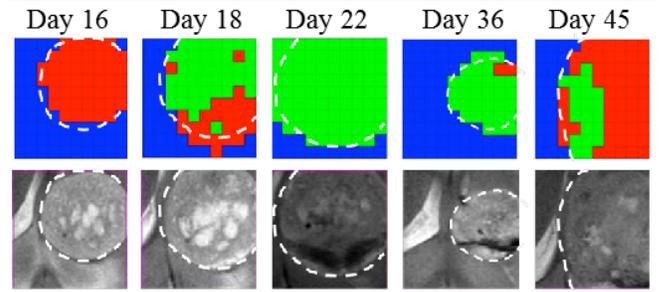


Fig. 5. Nosologic color-coded maps (top row) corresponding to mouse C819 at chosen time points. Blue voxels are assigned to normal brain parenchyma, red voxels to untreated/unresponsive tumor and green voxels to treated responding tumor. Tumor boundaries (T2w hyperintensity-derived, see bottom row) are marked with a white dotted line.

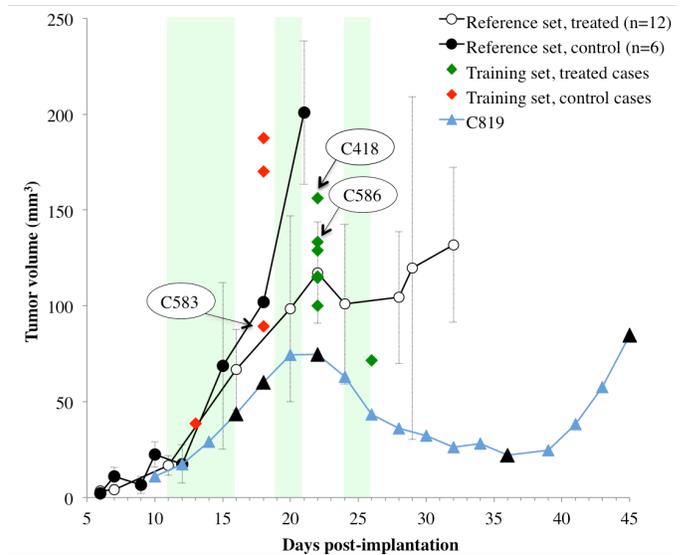


Fig. 6. Tumor volume change in treated and control mice. The circles represent the mean volume of a reference set [6]; the diamonds represent the tumor volume of the mice used for training (the day they were sacrificed); and the triangles represent the tumor volume of one treated (C819) mouse. Rectangles in light green highlight the 3 TMZ cycles. Black triangles highlight the selected days where nosologic maps were produced (see Figure 5).

V. DISCUSSION

A. Source signals calculated

Even when the treated and untreated sources have a very high correlation (0.97), preliminary results suggest that the subtle differences between them are able to discriminate these two regions in new/unseen mice data, based on their metabolome pattern changes recorded by MRSI.

The nature of the changes in the spectral pattern of treated tumors is mainly related to mobile lipid (ML) and polyunsaturated fatty acids (PUFA) resonances, which are very sensitive and accurate biomarkers for detecting therapy-related metabolic changes, as their intensities increase upon apoptosis

induced by treatment [14], [15]. Still, other contributions (e.g. lactate (1.3 and 4.1 ppm), glutamine, glutamate, alanine (≈ 3.8 ppm), and myo-inositol/glycine (≈ 3.5 ppm) (see figure 2) may be helping discrimination among tumor sources.

B. Nosological imaging of the response to therapy

Results shown in Figure 3 summarize the different types of response detected in the training set after applying the semi-supervised source extraction methodology. The control case (C583) shows high values for the mixing matrix of the untreated/unresponsive tumor source inside the abnormal mass assessed by T_{2w} imaging, as expected. On the other hand, tumor C586 is completely associated to the responding tumor source, in correlation with the volume measurement acquired at that time-point (the value is included in the average growth curve for treated mice represented in Figure 6). Finally, case C418 is partially associated with the unresponsive and responding sources, indicating a potential heterogeneity in tumor response to treatment. This partial response may be explained by the fact that the tumor volume registered on this day was slightly higher than the average growth curve for treated mice (see Figure 6), suggesting that it was not completely responding (yet) to TMZ therapy. These results suggest that the semi-supervised source extraction methodology is able to discriminate not only between responding and progressing tumors, but also between heterogeneous response regions in a particular case.

With respect to the control group, in the previous study [4] in which only two sources were extracted with Convex-NMF [9], the solid tumor region was validated by two indicators: proliferation index (PI) $> 30\%$ and abnormal area as suggested by the MRI (T_{2w} hyperintensity-derived). The results shown in Figure 4 indicate that, firstly, the tumor source (T) is the one that always represents the tumor masses in the seven mice (not the responding source). And secondly, the delineation of the tumor areas agrees with those suggested by the two indicators from [4]. Incidentally, in case C179, the nosologic image provides a better delineation of the abnormally proliferating area (Figure 4), including these areas in which proliferation is below 30%.

Nosologic images obtained for the treated mouse of the independent test set, namely C819, at representative time points during and after the course of therapy are presented in Figure 5. They allowed convenient tracking of response to treatment and differentiated the intratumoral heterogeneity of response, hinting the growth arrest and relapse (day 18 and 36, respectively), before changes in tumor volume were observed (Figure 6). Until the day 16 the tumor was recognized as fully unresponsive and the evidence of response started to be observed at day 18. This preceded the growth arrest (Figure 6) and the nosological image demonstrated full response on the day 22, followed by diminution of the mass. On the other hand, the image produced on the day 36 indicates signals of relapse, just before tumor started to regrow and finally on the day 45 was recognized as mostly fully growing again.

VI. CONCLUSIONS

In this communication we have applied the semi-supervised methodology proposed in [3] to assist the source extraction and obtain sources that better represent the therapy response in preclinical glioblastoma. The resulting sources were able to discriminate between response and untreated/unresponsive areas of the tumor, not only in the mice used for training but also in an independent test set of seven control and one treated (in different time points). We also provided a mechanism to produce nosologic images of the response to the therapy, based on the sources calculated in a semi-supervised way, which can serve as a valuable non-invasive tool in preclinical analysis for therapy response evaluation and monitoring.

ACKNOWLEDGMENT

This work was funded by MINECO grant MARESCAN (SAF 2011-23870) and Centro de Investigación Biomédica en Red – Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN, [http://www.ciber-bbn.es/en]), an initiative of the Instituto de Salud Carlos III (Spain) co-funded by EU FEDER. M. Ciezka holds a FI-DGR grant from the Generalitat de Catalunya. S. Ortega-Martorell is funded by the European Union under the 7th FP, “Marie Curie” FP7-PEOPLE-2012-IEF.

REFERENCES

- [1] F. Stupp, R. Roila, “Malignant glioma: ESMO clinical recommendations for diagnosis, treatment and follow-up,” *Ann. Oncol.*, vol. 20, no. Suppl 4, pp. 126–128, 2009.
- [2] J. Chen, Y. Li, T.-S. Yu, R. M. McKay, D. K. Burns, S. G. Kernie, and L. F. Parada, “A restricted cell population propagates glioblastoma growth after chemotherapy,” *Nature*, vol. 488, pp. 522–526, 2012.
- [3] S. Ortega-Martorell, H. Ruiz, A. Vellido, I. Olier, E. Romero, M. Julià-Sapè, J. D. Martín, I. H. Jarman, C. Arús, and P. J. G. Lisboa, “A Novel Semi-Supervised Methodology for Extracting Tumor Type-Specific MRS Sources in Human Brain Data,” *PLoS One*, vol. 8, no. 12, p. e83773, 2013.
- [4] S. Ortega-Martorell, P. J. G. Lisboa, R. V. Simões, M. Pumarola, M. Julià-Sapè, and C. Arús, “Convex Non-Negative Matrix Factorization for brain tumor delimitation from MRSI data,” *PLoS One*, vol. 7, no. 10, p. e47824, 2012.
- [5] R. V. Simões, T. Delgado-Goñi, S. Lope-Piedrafita, and C. Arús, “¹H-MRSI pattern perturbation in a mouse glioma: the effects of acute hyperglycemia and moderate hypothermia,” *NMR Biomed.*, vol. 23, no. 1, pp. 23–33, 2010.
- [6] T. Delgado-Goñi, J. Martín-Sitjar, M. Acosta, S. Lope-Piedrafita, and C. Arús, “DMSO-based contrast for monitoring GBM response to therapy,” *Magn. Reson. Mater. Phys.*, vol. 24, no. S1, pp. 119–120, 2011.
- [7] P. Paatero and U. Tapper, “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values,” *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.
- [8] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [9] C. Ding, T. Li, and M. I. Jordan, “Convex and semi-nonnegative matrix factorizations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, 2010.
- [10] S. Amari, “Information geometry on hierarchy of probability distributions,” *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1701–1711, 2001.
- [11] W. J. Heiser, “Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis,” in *Recent*

Advances in descriptive Multivariate Analysis, W. J. Krzanowski, Ed. Oxford: Oxford University Press, 1995, pp. 157–189.

- [12] H. A. L. Kiers, “Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems,” *Comput. Stat. Data Anal.*, vol. 41, no. 1, pp. 157–170, 2002.
- [13] R. V. Simões, S. Ortega-Martorell, T. Delgado-Goñi, Y. Le Fur, M. Pumarola, A. P. Candiota, J. Martín, R. Stoyanova, P. J. Cozzone, M. Julià-Sapé, and C. Arús, “Improving the classification of brain tumors in mice with Perturbation Enhanced (PE)-MRSI,” *Integr. Biol.*, vol. 4, no. 2, pp. 183–191, 2012.
- [14] J. M. Hakumäki, H. Poptani, A. M. Sandmair, S. Ylä-Herttuala, and R. A. Kauppinen, “¹H MRS detects polyunsaturated fatty acid accumulation during gene therapy of glioma: implications for the in vivo detection of apoptosis,” *Nat. Med.*, vol. 5, no. 11, pp. 1323–1327, 1999.
- [15] L. Mirbahai, M. Wilson, C. S. Shaw, C. McConville, R. D. G. Malcomson, R. A. Kauppinen, and A. C. Peet, “Lipid biomarkers of glioma cell growth arrest and cell death detected by ¹H magic angle spinning MRS,” *NMR Biomed.*, vol. 25, no. 11, pp. 1253–1262, 2012.