Use of physiological information based on grayscale images to improve mass spectrometry imaging data analysis from biological tissues

Sílvia Mas Garcia, A. Torro, N. Bec, L. Fernández, G. Erschov, Céline
Gongora, C. Larroque, Pierre Martineau, A. de Juan, S. Marco

1 **Use of physiological information based on grayscale images to improve mass**

2 **spectrometry imaging data analysis from biological tissues.**

3 S. Mas[1,2*], A. Torro[3], N. Bec[3,4], L. Fernández[1,5], G. Erschov[3], C. Gongora[3], C.

4 Larroque[3,6], P. Martineau[3], A. de Juan[2] and S. Marco[1,4]

5 [1]Signal and Information Processing for Sensing Systems, IBEC, Baldiri Reixac 4-8,

6 08028 Barcelona, Catalonia, Spain

7 [2]Chemometrics Group. Department of Chemical Engineering and Analytical Chemistry.

8 UB. Av. Diagonal, 645. 08028 Barcelona, Catalonia, Spain

9 [3]Institut de Recherche en Cancérologie de Montpellier (IRCM), INSERM U1194,

10 Université de Montpellier, Institut régional du Cancer de Montpellier (ICM),

11 Montpellier, F-34298, France.

12 [4]Institute for Regenerative Medicine & Biotherapy (IRMB), INSERM U1183, CHRU of

13 Montpellier, 80 Rue Augustin Fiche, Montpellier, F-34295, France

14 [5]Department of Electronics and Biomedical Engineering, Universitat de Barcelona,

15 Marti i Franqués 1, Barcelona 08028, Spain.

16 [6]Supportive Care Unit, Institut du Cancer de Montpellier (ICM), 208 Rue des

17 Apothicaires, Montpellier, F-34298, France

18

19

20 ***Corresponding author:** Sílvia Mas Garcia, Signal and Information Processing for

21 Sensing Systems, Institute for Bioengineering of Catalonia (IBEC), Baldiri Reixac 4-8,

22 Barcelona 08028, Spain. Telf: +34 934 031118. smas@ibecbarcelona.eu

1   **Abstract**

2   The characterization of cancer tissues by matrix-assisted laser desorption ionization-

3   mass spectrometry images (MALDI-MSI) is of great interest because of the power of

4   MALDI-MS to understand the composition of biological samples and the imaging side

5   that allows for setting spatial boundaries among tissues of different nature based on

6   their compositional differences.

7   In tissue-based cancer research, information on the spatial location of necrotic/tumoral

8   cell populations can be approximately known from grayscale images of the scanned

9   tissue slices. This study proposes as a major novelty the introduction of this

10   physiologically-based information to help in the performance of unmixing methods,

11   oriented to extract the MS signatures and distribution maps of the different tissues

12   present in biological samples. Specifically, the information gathered from grayscale

13   images will be used as a local rank constraint in Multivariate Curve Resolution-

14   Alternating Least Squares (MCR-ALS) for the analysis of MALDI-MSI of cancer

15   tissues. The use of this constraint, setting absence of certain kind of tissues only in clear

16   zones of the image, will help to improve the performance of MCR-ALS and to provide

17   a more reliable definition of the chemical MS fingerprint and location of the tissues of

18   interest.

19   The general strategy to address the analysis of MALDI-MSI of cancer tissues will

20   involve the combination of MCR-ALS and K-means clustering. The resolution method

21   will provide the distribution maps and MS signatures of each tissue in the sample and

22   the resolved distribution maps for each biological component will be submitted as initial

23   information to K-means segmentation to obtain further information on the boundaries of

24   the different tissues studied. Such an approach is more powerful than the direct use of

25   K-means on the raw MSI spectra, since resolved non-biological signal contributions are

2

1   not used, and the weight given to the different biological components in the

2   segmentation process can be modulated by suitable preprocessing methods.

3

4   **Keywords**

5   MCR-ALS, K-means, local rank constraints, MALDI-MSI, grayscale images.

6

1. Introduction

Mass Spectrometry Imaging (MSI) is a powerful tool extensively applied in the field of biomedical research that provides exceptional advantages to analyze tissue specimens in detail. MSI incorporate information of a broad variety of analytes, ranging from small *(e.g., drugs and their metabolites)* to large molecules *(e.g. proteins, peptides)* and provide visualization of their spatial distribution. Mass spectrometry capabilities combined with microscopic information assists in the comprehension of molecular processes in specific cell types within a tissue.

Among MSI techniques, Matrix-Assisted Laser Desorption/Ionization (MALDI)-MSI is the most commonly employed in the field of tissue-based research [1–4]. The method MALDI-MSI, as developed by Caprioli et al. [5], permits the analysis of hundreds to thousands of molecules directly from tissue sections after matrix deposition and introduction of the sample into the ionization chamber. For each pixel, a complete mass spectrum is acquired. Since this spectrum may contain hundreds of distinct biomolecular ions, the information content of a MALDI-MSI data set is extremely complex. Therefore, the application of chemometric tools is necessary to successfully interpret MALDI images [6].

Two of the most important aspects to be investigated on sections of biological material are the identity and distribution of the biological components and the presence of sample regions with similar properties, often defined by tissue boundaries. The ideal approaches to achieve these two objectives are factor analysis and segmentation analysis, respectively. In the framework of MSI data, the most well-known factor analysis method is the principal component analysis (PCA) [7–10]. Other methods have also been applied, such as probabilistic latent semantic analysis [8,9], independent component analysis (ICA) [8,10] and non-negative matrix factorization (NMF) [10].

4

Recently, MCR-ALS has been proven to adapt particularly well to MALDI image resolution due to the ability to incorporate dedicated image constraints and to the chemical meaning of the resolved distribution maps and spectral information associated with the image constituents [10–12].

In the case of segmentation methods, unsupervised algorithms are commonly applied when insufficient or no prior knowledge is available, such as often occurs in tissue-based cancer research. Among them, K-means is widely used for large image data sets because it is computationally lighter than hierarchical approaches [13,14].

Recently, combination of MCR (factor analysis method) and K-means clustering (segmentation method), has been proposed as general strategy for the characterization of biological tissues in infrared and Raman images [15–17]. Piqueras *et al.* demonstrated that the use of MCR score as starting information in K-means allows a compound-wise selection and preprocessing of the input information to be submitted to the segmentation algorithm and, for this particular scenario, presents advantages in the interpretability of the class centroids fo[15]. MCR scores are chemically meaningful because they are concentrations of the pixel constituents in the images. This could help to explain that the meaning of centroids coming from MCR scores are easy to interpret. Moreover, the application of this strategy in the simultaneous analysis of several images coming from different samples provided a better differentiation among tissue subparts or among tissues in different conditions [16]. Therefore, the application of this proposed methodology improved the efficiency of the information obtained when K-means method was applied directly on the raw image spectra.

Nevertheless, resolution of hyperspectral images under constraints, like non-negativity, does not always guarantee unique solutions because of the rotation ambiguity inherent to the bilinear model decompositions of MCR [18,19]. The introduction of selectivity

and local rank constraints, i.e., information on the absence of one or more image constituents in certain pixels, are among the most powerful constraints to decrease the rotational ambiguity in resolution results. Some examples of introduction of these constraints to hyperspectral images have already been reported [20,21]. However, the presented work proposes for the first time the introduction of local rank constraints based on the information provided by biological grayscale images (scanned images of the tissues) in the MCR resolution. Thus, the spatial information from biological grayscale images will be used to force certain components to be absent in some pixels. To do so, grayscale images will be registered to their MALDI-MSI data acquisition. This multimodal registration technique will be used for anatomical labeling of the data and will be later coded under the form of local rank constraints.

Application of the combination of MCR followed by K-means for the simultaneous analysis of MALDI-MSI data and the benefits from setting this local rank constraint to obtain best-defined images and less ambiguous profiles will be shown in a real example corresponding to fifteen tissues of colorectal cancer images.

2. Experimental

Human colon adenocarcinoma cell lines sensitive HCT116 (S) and resistant HCT116-SN50 (R) to Irinotecan (chemotherapy drug) were used as a model of sensitive and resistant experimental tumors. Clonogenic tumor xenografts were generated by a subcutaneous injection of both unique cell lines in athymic mice whereas a model of heterogeneity was created injecting various mixtures of R and S clones. The tumors were then collected, sliced, and analyzed utilizing MALDI-MSI. The complete methodology is described in more detail in the following subsections.

2.1 Materials and Methods

2.1.1 Chemicals and cell lines

Acetonitrile (ACN) and trifluoroacetic acid (TFA) were purchased from ThermoFisher Scientific (France). α-cyano-4-hydroxycinnamic acid was purchased from LaserBio Labs (France).

HCT116 (S) cell line was purchased from the American Type Culture Collection (ATCC, Manassas, Virginia). HCT116-SN50 (R) cell line was obtained from a clone of HCT116 as previously described in [22,23]. Cells were grown in RPMI1640 with L-glutamin supplemented with 10% fetal calf serum at 37°C under an atmosphere with 5% CO_2.

2.1.2 Tumor xenografts in nude mice

In vivo experiments were conducted by accredited researchers (Dr Adeline Ayrolles-Torro n°I-34UnivMontp-F1-12, Dr Celine Gongora n° 34-142) in compliance with the French regulation and ethical guidelines for experimental animal studies. Six weeks old female athymic mice were purchased from Harlan laboratories and were maintained in a specific pathogen-free facility in an accredited establishment (Agreement n° 34-172-27). Mice were xenografted subcutaneously in both flanks with 3.10^6 HCT116 or HCT116-SN50 alone or with mixture of these cells lines. Cells of the two lines of interest were mixed in 3 ratios 90% HCT116-10% HCT116-SN50; 50% HCT116-50% HCT116-SN50; 10%HCT116-90%HCT116-SN50. When tumors reached approximately $0.5cm^3$ in diameter, mice were euthanized and tumors were excised and frozen in liquid nitrogen.

2.1.3 MALDI-MSI analysis

Sample preparation

1 Frozen xenograft tumors were cut at 10μm thick slices with a HM 550 OVPD cryostat

2 (Fisher Scientific, Illkirch, France) operating at -20°C. Consecutive sections were

3 mounted on ITO coated conductive glass slides for IMS and on Superfrost Plus slides

4 (Microm) for immunochemistry experiments. Then, ITO slides were allowed to thaw

5 and were desiccated in a vacuum desiccator. The MALDI matrix application was as

6 follows: a 5 mg/ml solution of α-cyano-4-hydroxycinnamic acid dissolved in 50% ACN

7 and 0.1% TFA was sprayed on each tissue section (SunChrom SunCollect Maldi

8 spotter) with a distance in x of 0.5mm and in y of 2mm, Z offset at 1 and speed low7.

9 Twenty-two layers were applied at 10, 15, 20, 25 μl/min with a pause of 15 seconds

10 between each layer.

11 *Data acquisition*

12 *Scanned images*

13 The scanned images were acquired at 2400 dpi resolution with an Epson Perfection

14 4990 Photo Scanner (Seiko Epson, Nagano, Japan) into a tiff format as RGB images

15 stored in a x×y×3 matrix with data type uint8 (containing all whole numbers from 0 to

16 255). These scanned images were prior to MALDI-MSI acquisition.

17 *MALDI images*

18 All imaging mass spectrometry experiments were performed with a 4800 Plus MALDI

19 TOF/TOF[TM] Analyzer (AB Sciex) and the image acquisition was achieved using the

20 4800-imaging tool software (MSI imaging). Imaging of tumor sections was carried out

21 in a reflector positive mode, in the mass range of m/z 250-2000, with a resolution of

22 50000 points of m/z and 100μm in a 100μm x 1000 μm raster. The laser intensity was

23 set at 80% of full laser intensity as selected within the 4000 Series Explorer[TM]. At each

24 position of the tissue section, an averaged mass spectrum is generated from 1000

1   consecutive laser shots. The mass spectrum measured even for each pixel could not

2   have exactly the same size (number of measured m/z values) because only the non-zero

3   values were stored to reduce the size of the file acquired. The number of measured m/z

4   for each pixel could range between 50000 and 30000. The irregular MALDI images

5   (x×y×mz) were stored in Analyze7.5 format.

6   **2.1.4 Data description**

7   Different tumor samples and replicates of sections of the same tumor were analyzed in

8   order to obtain accurate results and identify differences of experimental origin. A total

9   of 15 images was analyzed: 3 slices of 100 % R cell line from the same tumor and

10   another slice from a different tumor (mice); 3 slices of 100 % S cell line from the same

11   tumor and another slice from a different tumor (mice); 2 slices of 50 % S and 50 % R, 3

12   slices of 90 % S and 2 slices of 90 % (R) of different tumors (mice). Description and

13   image labelling is summarized in table 1.

14   For each analysis, one scanned image and one MALDI image were provided. The size

15   (x×y number of pixels) of both scanned and MALDI images is also described in table 1

    3. Data analysis

16   **3.1 Data pretreatment**

17   In MALDI-MSI, the data generated can be arranged in a data cube in which the x- and

18   y- axis correspond to the pixel coordinates and the z-axis corresponds to the m/z values

19   registered in each pixel MS. Data preprocessing is needed to improve the signal quality

20   and to compress the raw data acquired into a list of useful peaks associated with

relevant m/z values for further analysis. In this work, almost all the data preprocessing was performed using the MALDIquant R package [24].

First, the raw data was imported into the R environment and, for each pixel, the spectrum was smoothed using a moving average filter and the baseline was corrected using the Statistics-sensitive Non-linear Iterative Peak-clipping algorithm (SNIP) [25]. Afterwards, the Median Absolute Deviation (MAD) was adopted as a noise estimation method, and all masses with signals above a signal-to-noise (S/N) ratio threshold of 2.5MAD, were identified as relevant peaks. This led to the identification of a list of masses and associated intensities for each pixel and helped to reduce the size of the data. Peaks of different m/z spectra of the same image or different images are associated with the same nominal m/z value by using a tolerance of 0.002 Da. Finally, the dataset including all 15 images was further reduced by keeping only the peaks present in at least 30 % of the spectra and, hence, avoiding artifacts and noise contributions The final dataset contained 758 different peaks between 250 and 1200 Da. For each pixel, the number of detected and preserved peaks ranged between 132 and 667 with a median value of 477. The most frequent peak (798.50 Da) was found in 98.9 % of the analyzed pixels, and 60.7 % of the peaks were present in at least half of the pixels.

For further analysis, this data set was imported to Matlab R2013b environment using a homemade script. A matrix $\mathbf{D}$ (n,m) of dimension n equal to $(x \times y)$ pixels by m m/z values was generated per each image. A multiset structure containing all 15 images was built, appending the 15 blocks of individual image pixel mass spectra one on top of each other.

3.2 Image registration

1   Spatially-related properties identified in the grayscale images from tissue slices could be

2   used to label the related parts in the MALDI images. Therefore, grayscale images were

3   registered to their related MALDI-MSI by using the MATLAB Image Processing

4   Toolbox 8.3 (R2013b) *"imregister"* function. This intensity-based automatic

5   registration uses an optimization algorithm to find the best transformation to register

6   two input images [26]. We used the "rigid" transformation (only involving translation

7   and rotation transforms) to register all grayscale images to MSI.

8   Before applying the registration, the pixels in the grayscale images were binned by a

9   factor of nine in *x* and *y* to have the same pixel size as the MALDI images. RGB

10   scanned images were converted to grayscale images with gray levels from 0 to 255 by

11   rbg2gray MATLAB function. This function converts RGB images to grayscale by

12   eliminating the hue and saturation information while retaining the luminance. MALDI

13   image cubes ($x \times y \times z$) were used to obtain 2D global intensity maps by adding all the

14   mass intensity values for each image pixel. Finally, the MALDI 2D global intensity

15   maps were converted to an image containing values in the range 0 (black) to 1.0 (white)

16   by mat2gray MATLAB function. Afterwards, we rescaled this intensity image

17   multiplying by 255 to be matched with the analogous grayscale images coming from the

18   scanner of tissue slices.

19   As an example, figure 1a, b and c show an original scanned grayscale image of 100R1a,

20   its related MALDI 2D global intensity map and its corresponding registered scanned

21   image, once binned to the MALDI pixel size, respectively. The scanned image in figure

22   1a reveals three different pixel types according to the gray intensity. Black color pixels

23   correspond to MALDI matrix, the dark gray pixels (external region) are attributed to the

24   region in which tumor cells have grown and light gray pixels (central region)

25   correspond to necrotic tissue.

1

2   **FIGURE 1**

3   This difference among gray color intensity in the physiological grayscale image was

4   used as a criterion for tissue recognition. Histograms were computed from all the pixels

5   in the physiological grayscale images and the peaks and valleys in the histogram were

6   used to locate the boundaries among clusters related to MALDI matrix, tumor parts and

7   necrotic parts. The tissue assignment in the physiological images was then used to label

8   the related MALDI pixels. Figure 1d shows the histogram corresponding to the

9   registered grayscale image in Figure 1c. The black rectangle includes all the pixels with

10   information of the biological material that was considered for further analysis. The

11   region framed in blue represents the grey intensity range in which tumor parts are

12   clearly located and the red framed region includes the grayscale values associated with

13   necrotic pixels.   Pixels with grayscale intensities between both regions were not

14   assigned either to tumor or to necrotic parts to avoid incorrect labeling. Pixels outside

15   the black rectangle are related to background. This background would correspond to

16   MALDI matrix pixels, which give no information; therefore, they were not used for

17   further analysis. It is worth to mention that visually inspection of selected pixels was

18   made to ensure the correct labeling.

19   **3.3 Multivariate Curve Resolution—Alternating Least Squares (MCR-ALS)**

20   MCR-ALS decomposes the matrix of MS pixel spectra **D** into a bilinear model, where

21   all MS spectra in the image can be described by the sum of the MS fingerprints of the

22   different pure image constituents weighted according to their relative concentrations in

23   the different pixels. This model is expressed by the following equation in matrix form:

24   $\mathbf{D}(n,m) = \mathbf{C}(n,n_c)\mathbf{S}^{\mathbf{T}}(n_c,m) + \mathbf{E}(n,m)$ Equation 1

where $\mathbf{D}$ (n,m) is the matrix of MS pixel spectra obtained after the pretreatment described in section 3.1, $\mathbf{C}(n,n_c)$ is the matrix of the relative amounts or concentration of the n_c components in the n pixels and $\mathbf{S^T}$ is the pure MS spectra matrix associated with these n_c components. $\mathbf{E}$ is the matrix associated with noise or experimental error (variance not explained by the n_c resolved components). The indexes n and m refer to the number of pixels analyzed from the image and the number of m/z channels in each MS spectrum, respectively.

With the aim to recover the bilinear model expressed in Equation 1, an iterative alternating least squares optimization under constraints is used to obtain both $\mathbf{C}$ and $\mathbf{S^T}$ matrices with a minimum error in the reproduction of the original data set ($\mathbf{D}$). In order to initialize the iterative procedure, a previous determination of the number of pure signal contributions in the raw data set and the generation of $\mathbf{C}$ or $\mathbf{S^T}$ estimates is required.

During the alternating least-squares optimization, constraints are used to introduce information useful to provide chemically meaningful $\mathbf{C}$ and $\mathbf{S^T}$ profiles and to decrease the ambiguity in the final solutions obtained [19,27]. In image resolution, non-negativity in the concentration and in the spectral direction are the most commonly used constraints. Normalization of pure spectra in $\mathbf{S^T}$ is also a common constraint used to avoid scaling fluctuations in the profiles during the optimization. Moreover, the use of local rank constraints can be extremely helpful to improve the reliability of MCR-ALS solutions. Local rank information indicates the absence of one or more constituents in the different pixels. In tissue-based cancer research, local rank information about the presence/absence of different tissues (tumor or necrotic) in samples can be obtained from the grayscale images, as mentioned in section 3.2. In these cases, the use of this physiologically-based local rank information can help to decrease the inherent MCR

13

1   ambiguity and, consequently, to define better the location and chemical structure of

2   tissue type. The local rank information in terms of compound identification is translated

3   into a 'mask' matrix. This matrix, sized as $\mathbf{C}$ (n pixels $\times$ n_c of components), is used

4   to encode the local rank constraints, i.e., the information on the missing components in

5   the different pixels, setting them to have a concentration value equal to zero or smaller

6   than a very low predefined value [19,27].

7   A great advantage of MCR-ALS is the possibility to analyze simultaneously several

8   images in a single multiset structure to provide more reliable results, less affected by

9   ambiguity phenomena [19,28,29]. In a biomedical context, this means that resolved

10   spectral contributions will define much better general trends of the population of

11   samples analyzed together than if images were analyzed individually. In this study, the

12   multiset structure was obtained by appending the pixel MS blocks of the different

13   MALDI images one on top of each other to form a column-wise augmented matrix,

14   $\mathbf{D_{aug}}$. The decomposition $\mathbf{D_{aug}} = \mathbf{C_{aug}}\mathbf{S^T} + \mathbf{E}$ provides a single matrix $\mathbf{S^T}$ of pure spectra,

15   valid for all the images in the multiset, and a matrix $\mathbf{C_{aug}}$, formed by as many

16   submatrices $\mathbf{C_i}$ as images in the multiset. The profiles in each of these $\mathbf{C_i}$ submatrices

17   can be refolded conveniently to recover the related distribution maps of every

18   constituent in each image. A multiset structure also obeys the bilinear model seen in

19   Equation 1.

20   The percentage of explained variance is used to indicate the fit quality of the MCR-ALS

21   results. This parameter is calculated according to the expression:

$$r^2(\%) = 100 \times \frac{\sum\limits_{ij} d_{ij}^{2*}}{\sum\limits_{ij} d_{ij}^2}$$

22   Equation 2

14

where $d_{ij}*$ are the values of the data set reproduced by the bilinear model and d_{ij} the original values in the raw data set **D.** The number of components included in the MCR-ALS model should be compromise between this maximum variance explained by the model, model simplicity, and model interpretability

More details about the MCR-ALS method are given in [28,30,31] and a GUI to use the algorithm is freely available at http://mcrals.info.

3.4 K-means (segmentation analysis)

K-means clustering is an algorithm which works by partitioning the observations of a dataset into k clusters [32]. In image analysis, the pixels in the dataset are partitioned into groups according to their similarity, expressed by a similar spectral shape, composition, chemical and/or biological properties.

In this study, a single K-means analysis was performed on the 15 images simultaneously. This multiset segmentation analysis allowed finding clusters consistently present in all images and, thus, related to relevant biological structures of the samples analyzed. MCR scores coming from the augmented $\mathbf{C_{aug}}$ matrix obtained by MCR-ALS have been used as the pixel input information for K-means analysis because they are excellent compressed and noise-filtered representations of the pixel composition. Besides, each column in $\mathbf{C_{aug}}$ has constituent-specific information and this allows performing K-means on some or all the signal contributions resolved by MCR-ALS. When using $\mathbf{C_{aug}}$ for segmentation, the number of sought clusters should be equal or slightly higher than the number of MCR contributions used in the K-means analysis.

The K-means algorithm basically consists of three steps: 1. Initialization, where few pixels are randomly selected and used as cluster centroids. 2. Assignment of every pixel of the data set to its nearest centroid, and 3. Updating of the centroid position based on the groups formed. Steps 2 and 3 are iteratively performed until centroids do not shift any longer and clusters are stable or when a present number of iterations is exceeded. In this work, 100 replicates with random initial centroid positions were performed to achieve more reliable results and Euclidean distance was used as a similarity criterion for pixel assignment to clusters. Finally, suitable segmentation maps were obtained refolding the part of the segmentation vector that relates to each image. Segmentation maps can be used to be compared with physiological images.

Moreover, the use of MCR scores provides chemically meaningful spectra of the components and, hence, allows interpreting easily the centroid properties in each segmentation class. This is an advantage in comparison with the centroids coming from PCA scores, which are difficult to interpret because these composition-related values are abstract and lack chemical sense.

Figure 2 presents a schematic illustration of the combination of MCR-ALS with K-means for the simultaneous analysis of MALDI images previously described. First, multiset MCR-ALS is applied to the images of biological tissues to provide basic spectral signatures and distribution maps of the biological contributions. Finally, multiset K-means analysis is applied to obtain consistent tissue clusters (relevant biological parts) common to all samples. MCR scores from biological constituents in MSI images were used as input pixel information for segmentation analysis. Autoscaling was applied to each concentration profile in each submatrix C_i of the C_{aug} matrix, to balance the importance of all biological constituents present in each image in

16

1   the segmentation analysis. This strategy might be considered as generally applicable for

2   mass spectrometry imaging data analysis.

3

4   **FIGURE 2**

5   **4. Results and discussion**

6   This section will include the results obtained from both image resolution and

7   segmentation approaches. Firstly, the information gathered from the comparison of the

8   analysis of the two types of tumor by MCR-ALS will be showed. Secondly, advantages

9   of the implementation of the new constraint in the simultaneous MCR-ALS analysis of

10   all 15 images will be presented. Finally, the results achieved from the multiset image

11   segmentation and a final interpretation of the chemical results will be offered. **4.1.**

12   **Resolution analysis (MCR-ALS)**

13   The first MCR-ALS analysis was oriented to identify significant contributions with a

14   specific "mass signature" for each tumor type (R or S). To do so, two multisets as

15   described in section 3.3. were built, each containing images related to a particular kind

16   of tumor and were analyzed separately. The first multiset was formed by the four

17   images corresponding to the tumors of 100% R, $\mathbf{D_{augR}}$= [$\mathbf{D_{100R1a}}$; $\mathbf{D_{100R1b}}$; $\mathbf{D_{100R1c}}$;

18     $\mathbf{D_{100R2}}$], and the second multiset by the four images corresponding to the tumors of

19   100% S, $\mathbf{D_S}$= [$\mathbf{D_{100S3a}}$; $\mathbf{D_{100S3b}}$; $\mathbf{D_{100S3c}}$; $\mathbf{D_{100S4}}$].

20   MCR-ALS was applied separately to each multiset structure ($\mathbf{D_R}$ and $\mathbf{D_S}$) following the

21   steps explained in section 3.3 and the following bilinear models were obtained:

22     $D_{augR} = C_{augR} * S^T_R$ Equation 2

23     $D_{augS} = C_{augS} * S^T_S$ Equation 3

1   and using non-negativity constraints in concentration and spectra profiles and spectra

2   normalization in matrix $\mathbf{S}^{\mathbf{T}}_{\mathbf{R \, or \, S}}$ (using the Euclidean norm).

3   Table 2 lists the number of resolved components and the explained variance obtained

4   from the MCR-ALS analyses of both multisets. Resolution of five contributions was

5   necessary in both cases. The inclusion of a different number of contributions gave

6   worse data fits or unreliable spectra or distribution maps (see Figure S1 in

7   Supplementary information for resolved spectra and distribution maps for these two

8   MCR-ALS analyses). Four contributions were related to biological regions of the

9   tissues; two associated with tumoral parts and two corresponding to necrotic parts,

10   according to the morphology seen in the related grayscale images. The need of two

11   biological contributions to describe the tumoral part of the sample and two additional

12   ones for the description of the necrotic regions of the samples is not surprising if we

13   take into consideration the heterogeneity associated with this kind of tissues. Moreover,

14   an additional non-biological contribution was needed to improve the resolution results

15   in both multiset structures. This extra contribution could be attributed to instrumental

16   background or could be originated to compensate for variations linked to slight

17   differences of sample preparation. Actually, the high noise of MSmass spectra and

18   possible experimental differences in the tissue preparation also justify that the explained

19   variance by the models does not exceed 90%.

20

21   **TABLE 2**

22

23   A correlation coefficient matrix generated by comparing by pairs the resolved spectra of

24   the components of both multisets ($\mathbf{S}^{\mathbf{T}}_{\mathbf{R}}$ and $\mathbf{S}^{\mathbf{T}}_{\mathbf{S}}$) by Pearson correlation shows a high

18

1   correlation between analogous biological components of the two tumor cell lines R and

2   S (see correlation matrix shown below).

3

$$
\begin{array}{c}
 & \text{Rc1} \quad \text{Rc2} \quad \text{Rn1} \quad \text{Rn2} \quad \text{Rb} \\
\begin{array}{c}
\text{Sc1} \\
\text{Sc2} \\
\text{Sn1} \\
\text{Sn2} \\
\text{Sb}
\end{array}
\left(
\begin{array}{ccccc}
\mathbf{0,9} & 0,0 & 0,4 & 0,1 & 0,0 \\
0,0 & \mathbf{1,0} & 0,1 & 0,1 & -0,1 \\
0,6 & 0,0 & \mathbf{0,8} & 0,4 & 0,4 \\
0,1 & 0,2 & 0,7 & \mathbf{0,9} & -0,1 \\
0,5 & 0,0 & 0,6 & 0,4 & 0,2
\end{array}
\right)
\end{array}
$$

4

5   Where Rc1/Rc2 and Sc1/Sc2 were the spectra corresponding to tumoral contribution 1/2

6   of $\mathbf{D_R}$ and $\mathbf{D_S}$, respectively. Rn1/Rn2 and Sn1/Sn2 were the spectra corresponding to

7   necrotic contribution 1/2 of $\mathbf{D_R}$ and $\mathbf{D_S}$, respectively. Finally, Rb and Sb were the

8   spectra corresponding to background contribution of $\mathbf{D_R}$ and $\mathbf{D_S}$, respectively.

9   No significant differences between the composition of tumors R and S were found.

10   Thus, resolved mass spectral signatures of the four biological contributions related to

11   tumor and necrotic parts were rather similar in both R and S tumoral tissues. Indeed, in

12   tumoral contributions, Rc and Sc, correlation coefficient values were equal or higher

13   than 0.9, whereas the lowest correlation coefficient between necrotic contributions, Rn

14   and Sn, was 0.8. Only the non-biological contributions, Rb and Sb, presented expected

15   differences in the spectral signatures, due to the noise random nature. Limits in the

16   selectivity of the MS detection system and the fact that both R and S clones are closely

17   related, since they derived from the same original HCT116 cell line, could be the

18   reasons to explain why a specific distinction and identification of "mass signatures" for

19   each tumor type was not possible. Owing to this limitation, the interpretation of the

20   differences among the R and S tumor compounds obtained was no longer the main goal

1 of the work. However, some general interpretation, valid for R and S tumors, can still be

2 done. Therefore, salient masses in the tumor signatures to masses around 750-800 could

3 be assigned to glycosyldiradylglycerols, glycerophosphocholines or

4 glycerophosphoglycerols and around 520-525 with oxidized glycerophospholipids or

5 fatty acyl glycosides. Relevant masses in the necrotic signatures to masses around 400-

6 430 could be assigned to sterols/fatty esters.

7 The spectral similarity between both tumors R and S suggested the possibility of

8 performing a simultaneous treatment of all 15 images corresponding to homogeneous

9 (100% R or S) and heterogenous (different percentages of R and S) tumor tissues in

10 order to obtain more reliable mass spectral signatures and distribution maps to integrally

11 describe the tumor and necrotic contributions common to both kind of R and S tumor

12 samples [19,28,29]. Therefore, MCR-ALS analysis of the multiset formed by the 15

13 images was performed. Figures 3a and b show the MCR-ALS resolved distribution

14 maps and pure spectra of the analyzed multiset. Results were clearly showed through

15 the example of two images (100S3a and 100S3b) with a higher magnification and the

16 full set of magnified distribution maps were provided as supplementary information (see

17 Figure S2). Related scanned grayscale images of all samples analyzed with MSI are also

18 shown (left plots in figure 3a). As mentioned in section 3.2, these physiological

19 grayscale pictures show the region in which tumoral cell populations have grown (dark

20 gray color) and the region where necrosis has been produced (light gray color).

21 Therefore, these grayscale images help to associate each resolved MCR contribution

22 with a particular kind of tissue types (tumoral or necrotic). A

23

24 **FIGURE 3**

20

1

Now, in contrast to the analysis of either $\mathbf{D_R}$ or $\mathbf{D_S}$ multiset structures, resolution of six components is proposed. As in the resolution of $\mathbf{D_R}$ and $\mathbf{D_S}$ multisets, the four components associated with relevant biological parts (tumor and necrotic) were observed. Looking at their related grayscale images, blue (component 1) and red (component 3) could be approximately associated with tumor contributions and cyan (component 4) and orange (component 6) with necrotic contributions. Moreover, two non-biological contributions (green and violet) were again required to perform the analysis and to cope with spectral differences due to instrumental and experimental variability. As expected, two non-biological spectral contributions were needed instead of one, in agreement with the lack of correlation between the Rb and Sb contribution resolved in the $\mathbf{D_R}$ and $\mathbf{D_S}$ multisets.

After the preliminary simultaneous MCR-ALS analysis of all 15 images, visual inspection of the distribution maps showed that the contributions 1 (blue) and 6 (orange) corresponding to a tumor contribution and a necrotic contribution, respectively, were blending with each other throughout the resolution. In some images corresponding to homogenous tumors (both 100% R and 100% S) the tumor contribution 1 was found in regions associated with necrotic tissue that should not be present and likewise for the necrotic contribution 6 that was present in parts of the tissues attributed to tumor tissue (see distribution maps of Figure 3).

The preliminary MCR-ALS analysis of the fifteen images was not fully satisfactory but provided better spectral signatures for the contributions to be modelled than the raw image and an approximate identification of the biological and non-biological contributions of the data set. The resolved spectral signatures were used as initial estimates in a second MCR-ALS analysis and the information provided by the

1 concentration profiles was used as an auxiliary source of information to set local rank

2 constraints, as described below.

3 Since spatial information of tumor and necrotic parts can be identified in the grayscale

4 images of tissue slices, this information was used to force the presence/absence of

5 tumor/necrotic contributions throughout the resolution. Once tumor and necrotic pixels

6 were identified using the image registration between grayscale and MS images (See

7 section 3), a masking local rank matrix was built to be used as a constraint (see section

8 3.2 for more detail in the implementation of this constraint). Taking as a basis the

9 preliminary results obtained by MCR-ALS analysis using only non-negativity

10 constraints, pixels related to necrotic parts were forced to be absent in component 1

11 (blue tumor contribution) and pixels corresponding to tumor cell populations were

12 forced to be absent in component 6 (orange necrotic contribution) in the images

13 corresponding to both 100% R and 100 % S. Please note that the use of this constraint

14 does not need either that all pixels in all images be constrained nor that all biological

15 contributions be subject to local rank conditions. The use of this constraint in small

16 areas clearly characterized as necrotic or tumor in simpler samples, such as 100% R or

17 100% S, is sufficient to improve significantly the results of the whole multiset.

18 Figure 4 shows the resolution results for the simultaneous analysis of all 15 images

19 using non-negativity, normalization and local rank information. The variance explained

20 obtained in the resolution using only non-negativity and normalization constraints was

21 86.8 % and now with the use of the local rank constraint is 85.5 %. The similarity

22 between the two values indicates that the new constraint introduced is correct and does

23 not perturb the natural behavior of the data set. Resolved distribution maps

24 corresponding to 100S3a and 100S3b were again zoomed for better visualization and

25 the full set of magnified distribution maps were provided as supplementary information

22

(see Figure S3). Moreover, pixel selection from the registered scanned images of necrotic (green color) and tumoral regions (red color) were also showed for these two examples (left side of the zoomed plot in Figure 4).

FIGURE 4

By examination of the MCR-ALS results, the positive effect of local rank constraints in the definition of MS signatures and in the morphology of the distribution maps of the different biological contributions can be clearly seen . Now, in contrast to the resolution only using non-negativity and normalization constraints, tumor contribution 1 and necrotic contribution 6 are no longer mixed with each other and are located in their correct physiological regions. Besides, the resolved profiles are more accurate and show less ambiguity due to the effect of the local rank constraint.

Once the MCR-ALS analysis on the all 15 images was performed, the MCR scores (C_{aug}, distribution maps of sample constituents) were used for segmentation analysis.

4.2 Multiset image segmentation. K-means cluster analysis.

Multiset image segmentation analysis was the next step in the strategy proposed (see figure 2), mainly performed to look for clusters present in all analyzed images. As mentioned before, K-means analysis was carried out using resolved MCR scores of the multiset structure (see Figure 4). The fact of using MCR scores allows selecting the information to be included in the segmentation process. In this case, only the four biological contributions related to tumor and necrotic parts (contributions 1, 3, 4 and 6) were selected and the background contributions (2 and 5) were left aside.

The individual profiles of the four biological contributions of each submatrix C_i of the C_{aug} matrix were autoscaled to balance the importance of each biological contribution present in each image of the multiset in the segmentation analysis. Doing it by block (each C_i separately), global intensity differences among MS images were also compensated. Autoscaled concentration profiles C_i were segmented simultaneously into four classes. In this case, a number of clusters equal to the number of biological contributions was selected. No reliable patterns of necrotic and cancer regions were achieved when the number of clusters in the segmentation procedure was increased. Figure 5 shows the segmentation results obtained by using the preprocessed MCR scores previously described. Zoom in an example (image 100S3a) was provided in Figure 5 for a better evaluation of the results and magnified results of the full set can be seen in supplementary information (see Figure S4).

FIGURE 5

The fact that K-means assign each pixel to only one cluster [32] helps to attribute that cluster to a single tissue type and, hence, be more easily comparable with the grayscale images. In figure 5 grayscale images were presented (left side of the pair images in Figure 5) for comparison with their related segmentation maps (right side of the pair images in Figure 5). It can be observed that the grayscale images match the location of tumor and necrotic parts with their related segmentation maps. Clusters 2 (green) and 3 (orange) correspond to the tumor parts, they have the same spatial distribution of the dark grey regions in the scanned image. Similarly, cluster 1 (blue) and 4 (brown) were associated with the necrotic parts. However, this information is richer than that provide

by the grayscale images, since heterogeneity within the same kind of tissue (tumor or necrotic) can be appreciated.

Centroid profiles using MCR scores in segmentation analysis are interpretable and, hence, it is possible to know which component or mixture of components is represented in each cluster. In our case, cluster 2 and 3 display dominance of contribution 1 and 3 of MCR contributions (tumor contributions), respectively, and low relative presence of other compounds, as expected. Likewise, cluster 1 and 4 show dominance of contribution 4 and 6 of MCR contributions (necrotic contributions). It must be reminded that segmentation clusters can be potentially formed by pure compound pixels or by pixel mixtures, whereas distribution maps from image resolution display the distribution of each pure image constituent. Since image constituents overlap (specially species 1 and 4), clusters are formed by some pixels with constituent mixtures of different composition. As mentioned before, an important asset is that MCR scores only include information related to biological contributions of the samples and discard information associated with background contributions. This benefit cannot be obtained if segmentation procedures are directly applied to raw MS image spectra.

The results from the combination of MCR with segmentation analysis could be potentially compared with conventionally stained sections to support determination of both anatomical and histopathological features and to provide a better chemical characterization of the compounds or mixtures of compounds involved in every sample region.

In future works, the incorporation of more selective MS detection systems and the use of conventionally stained sections is planned to allow for a more specific identification and location of biological regions to incorporate local rank constraints and for a better definition of the MS spectral signatures obtained to describe tumor heterogeneity.

1

5. Conclusions

The strategy of combining multiset resolution and segmentation has been shown to work for the analysis of biological tissues based on the use of hyperspectral images. For the first time, this strategy was applied to MALDI-MSI data and spatial information from grayscale images was used to set local rank constraints in the MCR-ALS resolution. This constraint improved significantly the morphology of the constituent distributions maps resolved by the MCR-ALS, helped in the suppression of the ambiguity in the results obtained and, therefore, contributed to a better definition of the related MS signatures for every biological contribution.

From the grayscale images, relevant features corresponding to anatomical areas of interest (tumoral or necrotic areas) could be identified. The use of this information to force certain components to be absent in particular pixels through MCR-ALS multiset analysis allowed providing reliably spectral signatures to define biological parts common in all the samples and to provide MCR scores for subsequent segmentation analysis. The results of image segmentation distinguished different tissue types (within tumor or necrotic areas) and their spatial distribution matched consistently with the grayscale images. Therefore, the segmentation maps and the chemical information linked to them could be eventually compared with the outcomes of standard histopathological staining protocols.

It is essential to stress the twofold benefit of using in a combined way physiological grayscale images and MSI measurements. On the one hand, the results obtained from MSI analysis by MCR-ALS are clearly improved by the easy setting of local rank constraints based on the visual location of different tissues in grayscale images. On the

other hand, the plain information in grayscale images, only marking the boundaries among tissue regions, is greatly enhanced by the MS signatures provided after MCR-ALS application associated with the different tissue regions. Such a detailed MSI analysis also provides more insight on the intrinsic heterogeneity of tissue regions, often necessarily described by more than one biological contribution.

The strategy proposed was shown to work for a case study including fifteen experimental tumors with different grade of heterogeneity but can be generally applicable in any study of biological material based on mass spectrometry imaging data analysis.

References

[1]   E.H. Seeley, R.M. Caprioli, MALDI imaging mass spectrometry of human tissue:

Method challenges and clinical perspectives, Trends Biotechnol. 29 (2011) 136–

143. doi:10.1016/j.tibtech.2010.12.002.

[2]   R.R. Drake, T.W. Powers, E.E. Jones, E. Bruner, A.S. Mehta, P.M. Angel,

MALDI Mass Spectrometry Imaging of N-Linked Glycans in Cancer Tissues, 1st

ed., Elsevier Inc., 2017. doi:10.1016/bs.acr.2016.11.009.

[3]   T.J.A. Dekker, B.D. Balluff, E.A. Jones, C.D. Schöne, M. Schmitt, M. Aubele,

J.R. Kroep, V.T.H.B.M. Smit, R.A.E.M. Tollenaar, W.E. Mesker, A. Walch, L.A.

McDonnell, Multicentre matrix-assisted laser desorption/ionization mass

spectrometry imaging (MALDI MSI) identifies proteomic differences in breast

cancer-associated stroma., J. Proteome Res. (2014). doi:10.1021/pr500253j.

[4]   A. Buck, S. Halbritter, C. Späth, A. Feuchtinger, M. Aichler, H. Zitzelsberger,

K.P. Janssen, A. Walch, Distribution and quantification of irinotecan and its

active metabolite SN-38 in colon cancer murine model systems using MALDI

MSI, Anal. Bioanal. Chem. 407 (2015) 2107–2116. doi:10.1007/s00216-014-

8237-2.

[5]   * Richard M. Caprioli, and Terry B. Farmer, J. Gile, Molecular Imaging of

Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF

MS, 69 (1997) 4751–4760. doi:10.1021/AC970888I.

1   [6]   T. Alexandrov, MALDI imaging mass spectrometry: statistical data analysis and

2   current computational challenges., BMC Bioinformatics. 13 Suppl 1 (2012) S11.

3   doi:10.1186/1471-2105-13-S16-S11.

4   [7]   H.B. Mann, D.R. Whitney, On a Test of Whether one of Two Random Variables

5   is Stochastically Larger than the Other, Ann. Math. Stat. 18 (1947) 50–60.

6   doi:10.1214/aoms/1177730491.

7   [8]   M. Hanselmann, M. Kirchner, B.Y. Renard, E.R. Amstalden, K. Glunde, R.M.A.

8   Heeren, F.A. Hamprecht, Concise representation of mass spectrometry images by

9   probabilistic latent semantic analysis, Anal. Chem. 80 (2008) 9649–9658.

10   doi:10.1021/ac801303x.

11   [9]   H.C. Diehl, B. Beine, J. Elm, D. Trede, M. Ahrens, M. Eisenacher, K. Marcus,

12   H.E. Meyer, C. Henkel, The challenge of on-tissue digestion for MALDI MSI- a

13   comparison of different protocols to improve imaging experiments, Anal.

14   Bioanal. Chem. 407 (2015) 2223–2243. doi:10.1007/s00216-014-8345-z.

15   [10]   Y. Gut, M. Boiret, L. Bultel, T. Renaud, A. Chetouani, A. Hafiane, Y.M. Ginot,

16   R. Jennane, Application of chemometric algorithms to MALDI mass

17   spectrometry imaging of pharmaceutical tablets, J. Pharm. Biomed. Anal. 105

18   (2015). doi:10.1016/j.jpba.2014.11.047.

19   [11]   C. Bedia, R. Tauler, J. Jaumot, Analysis of multiple mass spectrometry images

20   from different Phaseolus vulgaris samples by multivariate curve resolution,

21   Talanta. 175 (2017) 557–565. doi:10.1016/j.talanta.2017.07.087.

22   [12]   J. Jaumot, R. Tauler, Potential use of multivariate curve resolution for the

23   analysis of mass spectrometry images, Analyst. 140 (2015).

24   doi:10.1039/c4an00801d.

1 [13] D. Trede, S. Schi, M. Becker, S. Wirtz, K. Steinhorst, J. Strehlow, M. Aichler,

2   J.H. Kobarg, J. Oetjen, A. Dyatlov, S. Heldmann, A. Walch, H. Thiele, P. Maass,

3   T. Alexandrov, Excluído-Exploring Three-Dimensional Matrix-Assisted Laser

4   Desorption/ Ionization Imaging Mass Spectrometry Data: Three-Dimensional

5   Spatial Segmentation of Mouse Kidney, Anal. Chem. 84 (2012) 6079–6087.

6 [14] T. Alexandrov, J.H. Kobarg, Efficient spatial segmentation of large imaging mass

7   spectrometry datasets with spatially aware clustering, Bioinformatics. 27 (2011)

8   i230–i238. doi:10.1093/bioinformatics/btr246.

9 [15] S. Piqueras, L. Duponchel, R. Tauler, A. De Juan, Resolution and segmentation

10   of hyperspectral biomedical images by Multivariate Curve Resolution-

11   Alternating Least Squares, Anal. Chim. Acta. 705 (2011) 182–192.

12   doi:10.1016/j.aca.2011.05.020.

13 [16] S. Piqueras, C. Krafft, C. Beleites, K. Egodage, F. von Eggeling, O. Guntinas-

14   Lichius, J. Popp, R. Tauler, A. de Juan, Combining multiset resolution and

15   segmentation for hyperspectral image analysis of biological tissues, Anal. Chim.

16   Acta. 881 (2015) 24–36. doi:10.1016/j.aca.2015.04.053.

17 [17] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. de Juan, A. Gorzsás, Vibrational

18   spectroscopic image analysis of biological material using multivariate curve

19   resolution–alternating least squares (MCR-ALS), Nat. Protoc. 10 (2015) 217–

20   240. doi:10.1038/nprot.2015.008.

21 [18] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the

22   evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemom.

23   Intell. Lab. Syst. 103 (2010) 96–107. doi:10.1016/j.chemolab.2010.05.020.

24 [19] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data

analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58. doi:10.1002/cem.1180090105.

[20] A. De Juan, M. Maeder, T. Hancewicz, R. Tauler, Use of local rank-based spatial information for resolution of spectroscopic images, J. Chemom. 22 (2008) 291–298. doi:10.1002/cem.1099.

[21] X. Zhang, A. De Juan, R. Tauler, Local rank-based spatial information for improvement of remote sensing hyperspectral imaging resolution, Talanta. 146 (2016) 1–9. doi:10.1016/j.talanta.2015.08.017.

[22] L. Candeil, I. Gourdier, D. Peyron, N. Vezzio, V. Copois, F. Bibeau, B. Orsetti, G.L. Scheffer, M. Ychou, Q.A. Khan, Y. Pommier, B. Pau, P. Martineau, M. Del Rio, ABCG2 overexpression in colon cancer cells resistant to SN38 and in irinotecan-treated metastases, Int. J. Cancer. 109 (2004) 848–854. doi:10.1002/ijc.20032.

[23] C. Gongora, N. Vezzio-Vie, S. Tuduri, V. Denis, A. Causse, C. Auzanneau, G. Collod-Beroud, A. Coquelle, P. Pasero, P. Pourquier, P. Martineau, M. Del Rio, New Topoisomerase I mutations are associated with resistance to camptothecin, Mol. Cancer. 10 (2011) 1–13. doi:10.1186/1476-4598-10-64.

[24] S. Gibb, K. Strimmer, Maldiquant: A versatile R package for the analysis of mass spectrometry data, Bioinformatics. 28 (2012) 2270–2271. doi:10.1093/bioinformatics/bts447.

[25] M.H.Ã. Zhu, L.G. Liu, Y.S. Cheng, T.K. Dong, Z. You, A.A. Xu, Nuclear Instruments and Methods in Physics Research A Iterative estimation of the background in noisy spectroscopic data, 602 (2009) 597–599. doi:10.1016/j.nima.2009.01.174.

1   [26]   B.G. Sharma, A. Thé, Automating Image Registration with MATLAB,

2   Mathworks.Com. (2013) 1–7.

3   [27]   a. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent

4   processes and mixtures, Anal. Chim. Acta. 500 (2003) 195–210.

5   doi:10.1016/S0003-2670(03)00724-4.

6   [28]   R. Tauler, Multivariate curve resolution applied to second order data, Chemom.

7   Intell. Lab. Syst. 30 (1995) 133–146. doi:10.1016/0169-7439(95)00047-X.

8   [29]   R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate

9   Curve Resolution, in: Compr. Chemom., Elsevier, 2009: pp. 473–505.

10   doi:10.1016/B978-044452701-1.00055-7.

11   [30]   R. de Juan, Anna; Rutan, S. and Tauler, Two-Way Data Analysis: Multivariate

12   Curve Resolution – Iterative Resolution Methods, in: B. Brown, S. D. ; Tauler, R.

13   and Walczak (Ed.), Compr. Chemom., Elsevier, 2010: pp. 325–344.

14   [31]   J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and

15   applications, Chemom. Intell. Lab. Syst. 140 (2015).

16   doi:10.1016/j.chemolab.2014.10.003.

17   [32]   T.N. Tran, R. Wehrens, L.M.C. Buydens, Clustering multispectral images: a

18   tutorial, Chemom. Intell. Lab. Syst. 77 (2005) 3–17.

19   doi:10.1016/j.chemolab.2004.07.011.

20

21

22

1

2

3

4

5

6 **Table 1**. Percentage of both R and S cell lines in xenografts, replicate number of either

7 the same or different tumors (mice), code, size of both scanned and MALDI imagesof

8 the image sections analyzed.

9

Percentage of cell lines	Replicate number of the same tumor	Replicate number of different tumors	Image code	Pixels of RGB scanned image	Pixels of MALDI image
100% R	1	1	100R1a	1541x1118	131x84
100% R	2	1	100R1b	1560x1008	132x75
100% R	3	1	100R1c	1536x948	132x73
100% R	-	2	100R2	792x654	72x56
100% S	1	1	100S3a	1272x888	105x63
100% S	2	1	100S3b	1188x912	104x72
100% S	3	1	100S3c	1224x963	105x70
100% S	-	2	100S4	1064x704	78x51
90% R	-	1	90R5	1040x784	134x79
90% R	-	2	90R6	1512x1232	83x64
90% S	-	1	90S7	1392x930	125x77
90% S	-	2	90S8	1614x900	150x83
90% S	-	3	90S10	1200x1194	105x97
50% S	-	1	50R11	1480x1080	136x84
50% S	-	2	50R12	1296x968	110x73

10

11

12 **Table 2.** Number of resolved components and variance explained by MCR-ALS
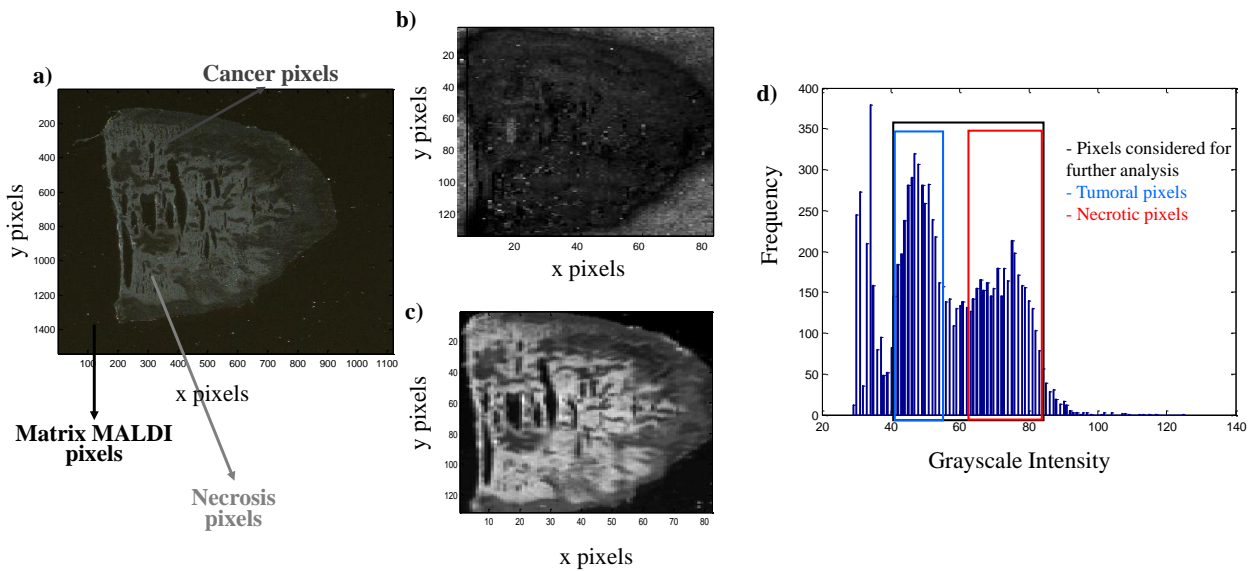13 analysis of $\mathbf{D_r}$ and $\mathbf{D_s}$ multiset structures.

1

| Multiset	Explained variance	Resolved components
$D_R = [D_{100R1a}; D_{100R1b}; D_{100R1c}; D_{100R2}]$	89.7%	5
$D_S = [D_{100S1a}; D_{100S1b}; D_{100S1c}; D_{100S2}]$	87.5%	5

2

3

8. Figures

5

Figure 1. a) Original RGB image of 100R1a sample, b) MALDI image of 100R1a after pretreatment of 3.1 section, c) registered scanned image of 100R1a, d) Histogram of registered image in 1c. Color interpretation in grayscale images in 1a and 1c is as follows: black pixels (MALDI matrix part), dark gray pixels (tumoral part) and light gray pixels (necrotic part). Grayscale intensity starts at zero for black color and goes until 255 for pure white. Values in between make up the different shades of gray.
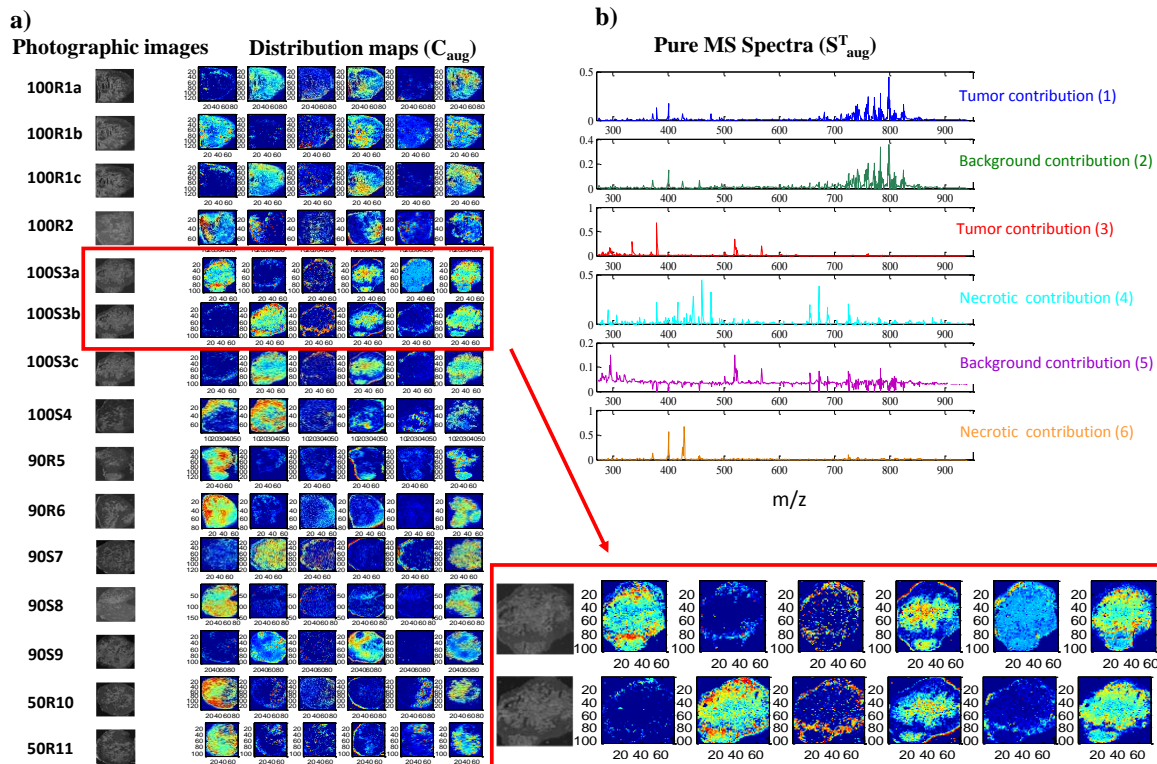
12
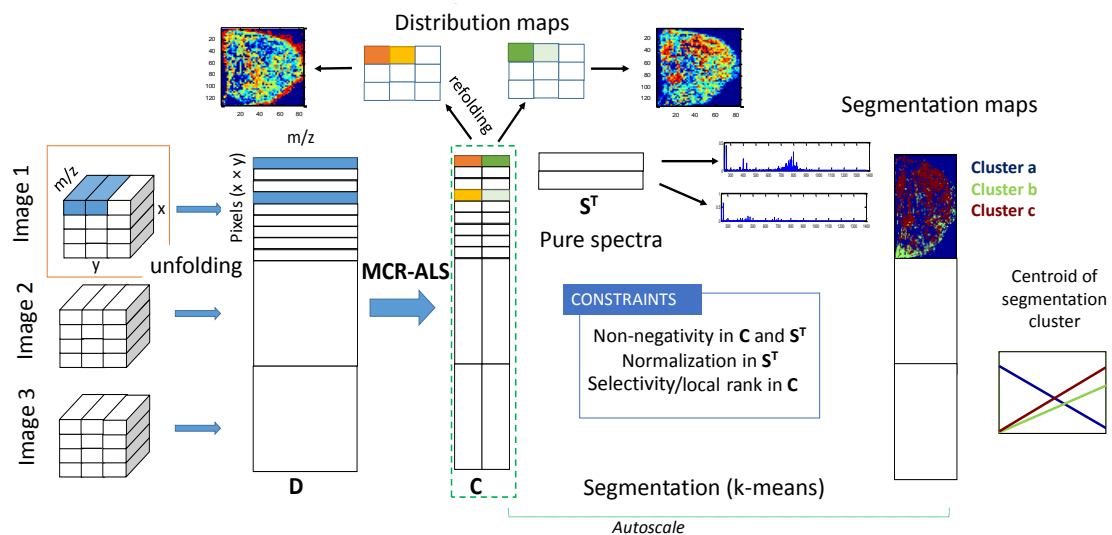
1

2 **Figure 2.** Graphical representation of the combination of MCR-ALS with K-means for

3 simultaneous analysis of MALDI images for an example of two MCR contributions
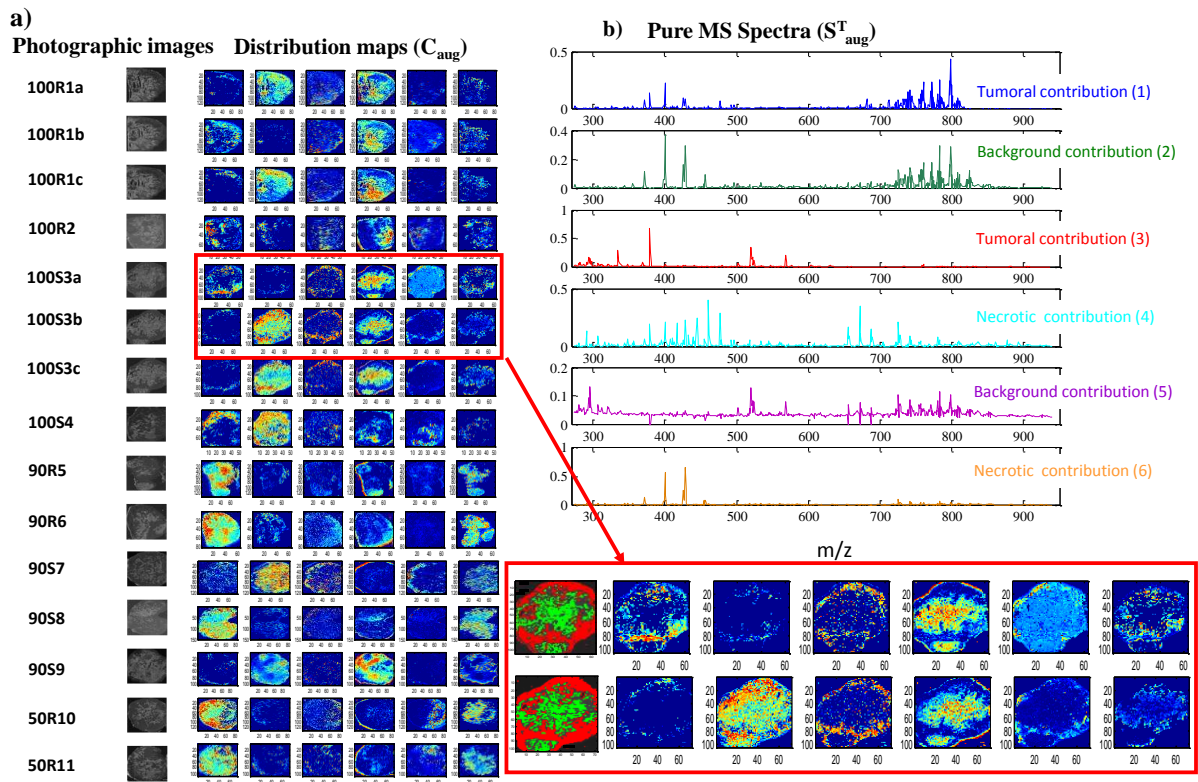
4 resolved.

5

6

7

8

Figure 3. MCR-ALS results for the multiset analysis of 15 tumor images using non-negativity and normalization constraints. Left plots: registered grayscale images. Middle plots: related MCR-ALS distribution maps. Right plots: resolved pure MS spectra. Color indices in grayscale images: dark gray (tumoral part) and light gray (necrotic part). Zoomed plot: gray scale images and distribution maps corresponding of images 100S3a and 100S3b

Figure 4. MCR-ALS results for the multiset analysis of 15 images using non-negativity, normalization and local rank constraints. Left plots: registered grayscale images. Middle plots: related MCR-ALS distribution maps Right plots: resolved pure MS spectra. Color indexes in grayscale images: black (MALDI matrix part), dark gray (tumoral part) and light gray (necrotic part). Zoomed plot: prior selection of necrotic and tumoral pixels and distribution maps corresponding of images 100S3a and 100S3b.
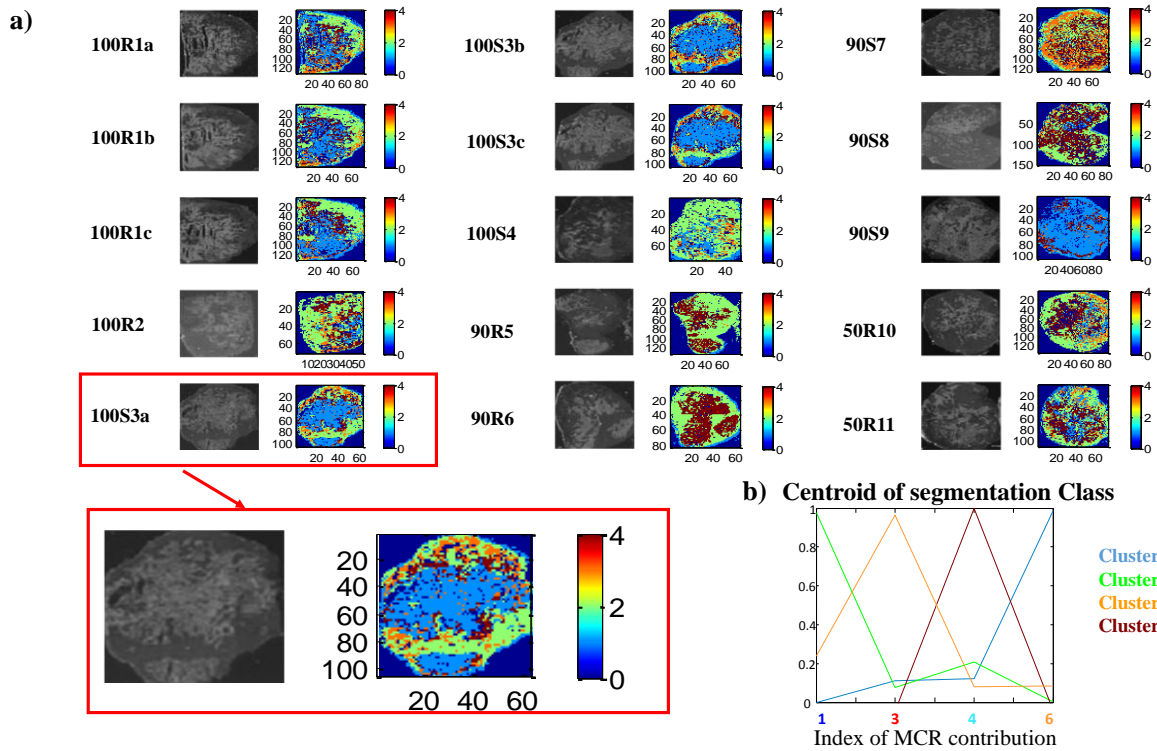
Figure 5. Segmentation results of for the multiset analysis of 15 images using MCR-ALS scores (distribution maps) coming from biological contributions in figure 4). a) Segmentation schemes and related registered grayscale images (left side) for all images. b) Centroid profiles. Color indexes in grayscale scanned image (c): black (MALDI matrix part), dark gray (tumoral part) and light gray (necrotic part). The same color is used for the class in the segmentation scheme and in the related centroid profile. Zoomed plot: gray scale images and segmentation map corresponding of images 100S3a.

1

2

3

4

5

6

7