



Application of chemometric methods to the analysis of multimodal chemical images of biological tissues

Carmen Bedia¹ · Àngels Sierra² · Romà Tauler¹

Received: 30 December 2019 / Revised: 20 February 2020 / Accepted: 10 March 2020 / Published online: 30 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Current histology techniques, such as tissue staining or histochemistry protocols, provide very limited chemical information about the tissues. Chemical imaging technologies such as infrared, Raman, and mass spectrometry imaging, are powerful analytical techniques with a huge potential in describing the chemical composition of sample surfaces. In this work, three images of the same tissue slice using matrix-assisted laser desorption/ionization (MALDI) mass spectrometry, infrared microspectroscopy, and an RGB picture from a conventional hematoxylin/eosin (H/E) staining are simultaneously analyzed. These fused images were analyzed by multivariate curve resolution-alternating least squares (MCR-ALS), which provided, for each component, its distribution within the tissue surface, its IR spectrum fingerprint, its characteristic mass values, and the contribution of the RGB channels of the H/E staining. Compared with the individual analysis of each of the images alone, the fusion of the three images showed the relationship between the different types of chemical/biological information and enabled a better interpretation of the tissue under study. In addition, the least-squares projection of the MCR-ALS resolved spectra of components at low spatial resolution onto the IR and RGB images at high spatial resolution, provided a better delimitation of the sample constituents on the image, giving a more precise description of their distribution on the investigated tissue. The application of this procedure can be of interest in different research areas in which a good description of the spatial distribution of the chemical constituents of the samples is needed, such as in biomedicine, food, or environmental research.

Keywords Multimodal chemical imaging · Image data fusion · MCR-ALS · Mass spectrometry imaging · Infrared imaging · Histology

Introduction

The examination of histological sections or histology aims to identify and characterize the different parts of a tissue under study. In the context of diseases, histopathological studies refer to the microscopic analysis of a biopsy or surgical samples

to determine the presence and the nature of the disease [1]. In these studies, tissue slides, once frozen or processed by chemical fixation, are stained with different combinations of pigments to reveal the cellular content. One of the most used methods is hematoxylin/eosin staining (H/E), which stains the cellular nuclei in blue and the extracellular matrix and cytosol in pink [2]. In histochemistry, a branch of histology, the different parts of the tissue are revealed by chemical reactions between the reactant and specific cellular components. Examples of this are Perl's reaction to detect the presence of iron in the samples [3], or the Von Kossa technique, which uses a solution of silver nitrate to identify deposits of calcium [4]. Other staining variants include immunohistochemistry, which distinguishes particular chemical structures using specific antibodies that are further revealed using additional antibodies detected by chemiluminescence or by fluorescence [5]. Over the decades, the use of these histopathological techniques has contributed to improving our understanding of diseases, leading to the elaboration of more effective treatments.

Published in the topical collection *Euroanalysis XX* with guest editor Sibel A. Ozkan.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00216-020-02595-8>) contains supplementary material, which is available to authorized users.

✉ Carmen Bedia
carmen.bedia@idaea.csic.es

¹ Institute of Environmental Assessment and Water Research (IDAEA-CSIC), 08034 Barcelona, Spain

² Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Spain

Also, their inclusion in routine clinical protocols demonstrates their usefulness to make diagnosis decisions [1, 6].

The chemical information provided by these staining procedures remains limited to the detection of a restricted number of molecules of interest within the tissues. To fill the gap of chemical information, spectroscopic vibrational technologies have jumped into the histology scenario. These label-free and non-destructive techniques such as Raman and infrared (IR) microspectroscopy are able to provide vibrational spectral fingerprints of all the molecules present in each of the predefined pixels of the tissue [7]. Usually, their spatial resolution is very high, enabling the analysis at subcellular level with a high degree of chemical specificity. These technologies have become a powerful tool in biomedical research, enabling a better interpretation of microscope tissue images by pathologists [8]. Many reports evidenced the potential of IR and Raman spectroscopy for the detection of different types of cancer including breast, brain, and colon cancer [9]. In addition to biomedical purposes, these techniques have also been proven to be useful in microbiology [10], plant [11] and food [12] research areas.

Although the vibrational fingerprints provided by these technologies are unique and could be used to detect specific events in tissues, the molecular identity of the chemical compounds on the surface remains unknown. Mass spectrometry imaging (MSI) technologies have emerged to solve this issue. In general terms, in MSI, molecules are desorbed from the tissue surface, ionized, and sampled using a mass spectrometer while the spatial distribution of each ion is accurately recorded [13]. The most common MSI techniques employed are matrix-assisted laser desorption ionization (MALDI), secondary ion mass spectrometry (SIMS) and desorption electrospray ionization (DESI). The ability of MSI technologies to provide molecular information and molecule distribution within the tissues has opened a wide range of applications in different research areas. For instance, in biomedicine, MSI techniques have been applied to biomedical and preclinical studies as a novel approach to understanding the molecular mechanisms of disease [14]. In the pharmaceutical industry, MSI data can be used to study the pharmacokinetics and pharmacodynamics of drugs [15].

Both vibrational imaging and MSI data can be represented by a data cube structure, in which the two spatial dimensions of the sample surface are the x- and y-image pixels, and the third dimension contains the spectral information. If the interest of the study goes beyond the analysis of specific spectral bands or m/z values, the study of the whole image data cube requires the application of chemometric tools. The main data analysis tools to deal with hyperspectral imaging data are based on image segmentation, such as hierarchical or K-means clustering, or factor analysis based methods, such as principal component analysis (PCA) or multivariate curve resolution-alternating least squares (MCR-ALS). The MCR-

ALS method [16] has been successfully used to resolve the main components present in single and multiset hyperspectral imaging datasets, as well as their spatial distribution and spectral features [17–19].

Despite the importance of molecular identification in tissue analysis, MSI technologies focus on a subset of molecules at once (peptides, lipids, metabolites, etc.) detected in a specific ionization mode, whereas vibrational technologies capture chemical information about all types of molecules present in the sample. Therefore, the molecular pictures provided by both types of techniques are complementary in terms of molecular characterization of the tissue. In this context, multimodal imaging based on image data fusion and analysis of the same tissue slice using different types of technologies has been recently addressed. Piqueras et al. have fused mass spectrometry and Raman images of a green bean tissue and proposed the use of MCR-ALS for the analysis of incomplete image datasets [20]. Neumann et al. used a multimodal image fusion pan-sharpening procedure to combine MSI and IR microspectroscopy image data at different resolutions. Pan sharpening is based on the use of a Laplacian pyramid method, which used high spatial frequency components from the higher spatial resolution IR image to sharpen chemical images at lower resolution obtained by MSI [21].

In this work, we propose a multimodal imaging method based on the data fusion of three imaging modalities: MALDI imaging-mass spectrometry (from now MALDI-MS), IR microspectroscopy, and RGB pictures of H/E histological staining, and their joint analysis using the multivariate curve resolution-alternating least-squares (MCR-ALS) method. The combination of the chemical image data provided by MS and IR imaging techniques and the histological information obtained from a classical staining procedure can be extremely useful to better understand and interpret the tissue under study. Moreover, we propose a simple method to get high-resolution distribution maps of the chemical constituents on high-spatial-resolution IR and RGB images by projecting the IR and RGB spectra of these constituents obtained in the multimodal analysis of fused images at the lower spatial resolution. In this work, the tissue sample model used to implement and test the proposed chemometric analysis of the fused data from multimodal images is a mouse xenograft derived from a breast cancer patient.

Materials and methods

Tissue sample

Tissue samples of a patient-derived xenograft (PDX) from primary breast carcinoma implanted in the intramammary fat path of nude mice were obtained from a previous study [22]. Tumors were cryopreserved at $-80\text{ }^{\circ}\text{C}$ in a TissueTek

Cryomold Mold using Optimal Cutting Temperature (OTC, TissueTek). To obtain the sample slice used in this work, one of the tissues was mounted in a cutting chock using OTC on the base of the tissue. The tissue slice was made at 12 μm thickness using a cryostat (Leica CM 3050) and placed directly onto ITO glass slices (Bruker) to be analyzed.

Generation of imaging data

In the first place, the tissue was subjected to IR imaging using a Nicolet iN10 MX infrared microscope (Thermo Fischer) in the reflectance mode. IR spectra were acquired in the spectral range between 700 and 4000 cm^{-1} with 427 points spaced at 7.7 cm^{-1} and 8 accumulated scans. The number of pixels collected for this image was 55,685 (185 \times 301), with a pixel size of 25 μm . The second step was the acquisition of MALDI-MS images. The MALDI matrix 2,5-dihydroxybenzoic acid (DHB, Merck) was first applied by sublimation as described in previous work [18]. The MALDI-MS image was acquired in negative mode in the 400–1200 m/z range, in which the detected molecules were mostly lipids. The acquisition was performed using an Autoflex III MALDI-TOF/TOF instrument (Bruker) equipped with a Smartbeam laser operated at a 200-Hz laser repetition rate at the “large focus” setting. The size of the pixel was set to 150 μm , and the final size of the image was 1550 pixels (50 \times 31).

Once the MALDI-MS imaging data was acquired, the tissue was subjected to hematoxylin/eosin staining, a widespread method used in histology that displays a broad range of nuclear, cytoplasmic, and extracellular matrix features of tissues [2]. First, the DHB matrix was washed off in 70% EtOH in water (v/v) for one minute. After dip-washing in deionized water, the tissue was stained for 10 min in Mayer’s hematoxylin solution (Sigma). The excess of solution was then dip-washed in deionized water, and further washed for 1 more minute. The slide was then put into Eosin working solution (0.25%) for 1 min. Eosin stock solution 1% (w/v) was prepared in 75% EtOH in water (v/v). To prepare the eosin working solution, the stock solution was diluted with 80% EtOH in water (v/v). The slide was then washed in deionized water and 2 min each in 70% EtOH, 80% EtOH, 90% EtOH, 100% EtOH, and 100% EtOH again, all prepared in Histo-clear solution (v/v) (National diagnostics). Finally, the slide was washed 2 min with Histo-clear solution. A drop of mounting medium was put onto the slide and a coverslip was placed onto the tissue for microscope examination. Several \times 4 pictures were taken to cover the area of the stained tissue using a microscope (Nikon SMZ 1500 Stereo Microscope) fitted with a digital camera (Nikon DS-Ri1). Pictures were put together to have a high-resolution RGB image of the tissue. This picture provided the information of the red, green, and blue intensity of each of the pixels, on a scale from 0 to 255.

The workflow of the different steps followed in the analysis of the different image data sets is given in Fig. 1.

Data pre-processing

IR image IR spectra of the image were opened in the OMNIC TM Picta software (Thermo Fischer) and exported as *csv* extension files. These files were loaded into the MATLAB (The Mathworks Inc.) environment and converted to a data matrix of 55,685 rows and 427 columns, the former value corresponding to the number of pixels of the IR image and the latter to the number of IR wavelengths measured. Baseline correction was performed on this matrix using the asymmetric least-squares algorithm applying the smoothness parameter $\lambda = 1000$ and the asymmetry parameter $p = 0.001$, see reference [23].

MALDI-MS image. The raw data file obtained in MALDI-MS imaging was opened in SCiLS Lab Software (version 2014b, SCiLS GmbH) and exported to an *imzML* format file, the standard mass spectrometry data format. The file was imported to the MATLAB environment using the *imzML* converter tool. Then, the most relevant m/z values of data were selected using the Regions of Interest (ROI) procedure [24], a compression method that only selects the mass values whose signal intensities are above a predetermined threshold value, within a predefined mass error accuracy and are detected a minimum number of times. This procedure has been previously shown to be useful to select relevant m/z values from single and multiple MSI datasets [18, 25]. In the present work, the threshold value was set to 1 (0.5% of the maximum spectra intensity), the mass error was set to 0.55 Da, and the minimum number of times to be considered as a relevant signal was set to 15 (1% of the total image pixels). As a result and after one-by-one inspection of the ROIs selected, the number of m/z values was reduced to 102. The mean spectrum of raw MSI data and the representation of the selected ROI values are available in Electronic Supplementary Material (ESM) Fig. S1. MSI intensities of each pixel were normalized using the probabilistic quotient method (PQN) [26], using as a reference the median spectrum.

RGB picture A total of 16 pictures were combined to produce a 13,423,200-pixel image (4512 \times 2975) of the H/E stained tissue. This image was loaded into the MATLAB environment and converted to a data matrix of 13,423,200 pixels and 3 columns. Each pixel of the RGB picture is described by a combination of three variables (red, green, blue) with values that range from 0 to 255. In order to normalize the values to the lightness, which is the sum of R, G, and B values in one pixel, each of the RGB values was divided by the lightness value of the corresponding pixel.

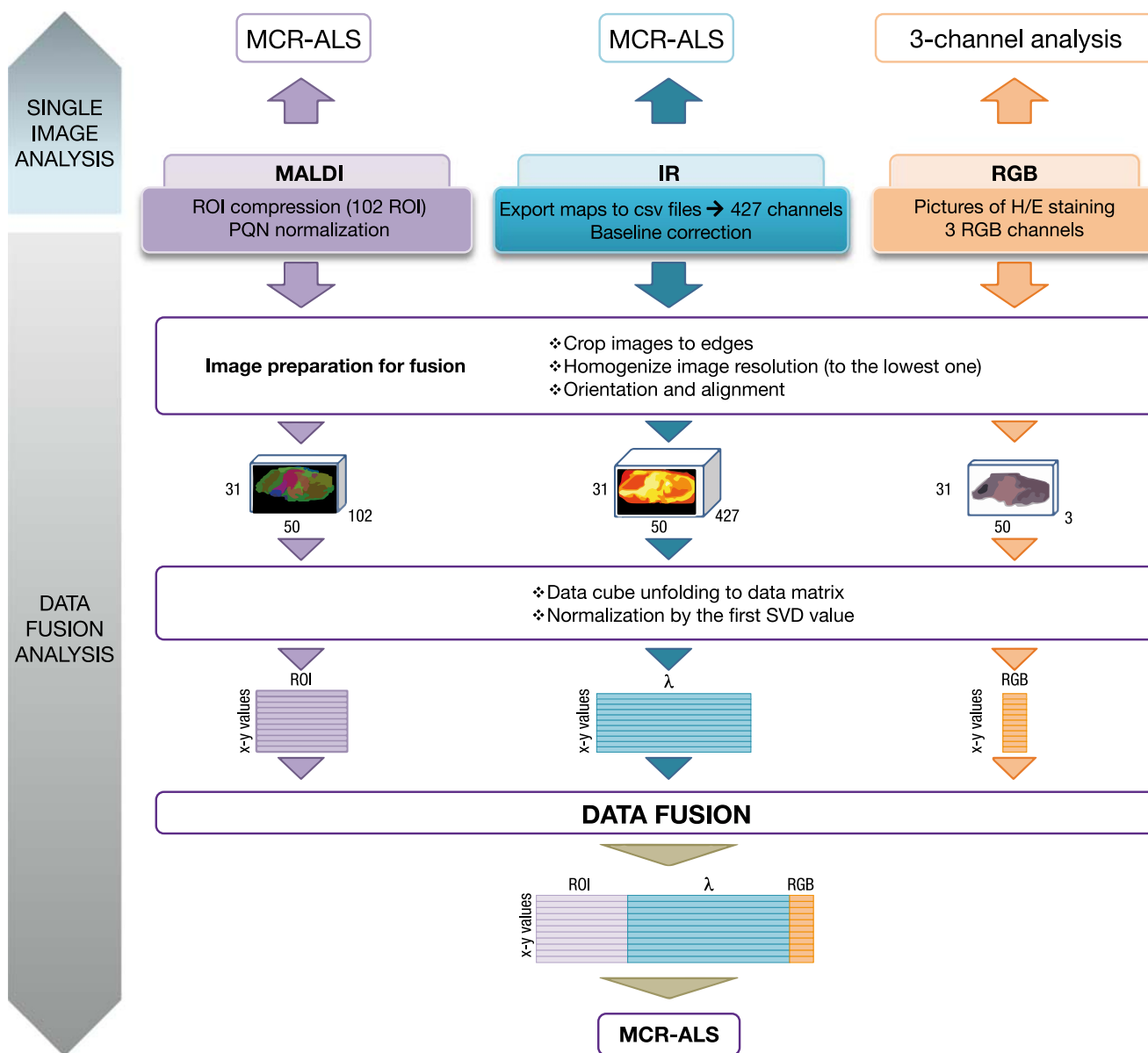


Fig. 1 Workflow of the single image and multimodal image data fusion analysis steps followed in this work

Image resizing and alignment

The experimental data from the images acquired in the analysis of the tissue using the three analytical methodologies were arranged in a data cube using the *reshape* function. Then, these data cubes were loaded into the multivariate image analysis software (MIA, PLS toolbox, Eigenvector Inc.) which provides many tools for multivariate image management. Using MIA, the x-y orientation of the images was initially corrected, and the images were cropped so that the three tissue images were adjusted to the borders. The modified images were then loaded into the MATLAB workspace and the IR and RGB matrices were resized to the same spatial dimensions of the MALDI-MS image (31×50), which had the lowest spatial resolution. To resize the IR and RGB images to the

same size as the MALDI-MS image, the *imresize* function from the Image Processing Toolbox of MATLAB was used. These images were then loaded again in the MIA image analysis interface to be further aligned. The alignment tool interface of MIA allows aligning two images using user-preselected points of both images helped by a guided user interface (GUI). Three variables of each image are initially chosen to guide the alignment process. Once the variables of each image are chosen, the GUI plots the two images side by side and represents the intensity of the variables selected within the tissue using different colors. This eases the next step, which is the selection of coincident points in the two images. Once two points are identified and selected in both images, the second image is shifted accordingly. Four pairs of points were selected to align first MALDI-MS and IR images, and

afterwards, the IR and the RGB images were aligned following the same procedure. The three new aligned images were then saved and loaded into the MATLAB workspace. The same procedure was followed to resize the RGB image to the higher spatial resolution of the IR image (301 × 185) and to align both images at this higher resolution.

Multimodal image data fusion

In order to perform the fusion of the imaging data, the three cube datasets were first unfolded to three augmented data matrices with the same number of rows (total number of image x-y pixels) and different number of columns depending on the spectroscopic technique (m/z ROI values, IR wavelengths, and RGB channels). Since the rows (x-y pixels) of these three augmented data matrices were already aligned and at the same spatial resolution, a new data matrix can be obtained by their row-wise matrix augmentation. In order to give the same importance to the three imaging data blocks (MS, IR, and RGB), prior to this fusion, each matrix was scaled by the first singular value of the corresponding data block. The row-wise augmented data matrix at low spatial resolution has a total number of 1450 rows (31 × 50 pixels) and 532 columns (102 m/z ROI + 427 wavelengths + 3 RGB channels). In the case of high-spatial-resolution IR and RGB matrices, their data fusion resulted in a matrix with 55,685 (301 × 185 pixels) rows and 430 columns (427 wavelengths + 3 RGB channels).

The dimensions of the image datasets used in the present work can be found in Table 1.

Multivariate curve resolution-alternating least squares

In this work, the MCR-ALS [16] method was applied to the analysis of the realigned and preprocessed image data matrices described above. MCR-ALS performs a bilinear decomposition of the image data matrices (**D**) into the product of two factor matrices (see Eq. 1), one factor matrix (**C**) providing the information about the relative concentration and spatial image distribution of the chemical constituents (distribution maps)

present in the analyzed sample tissue, and another factor matrix (**S^T**) related with the spectra of these sample constituents, respectively, according to the following equation:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where **E** accounts for the variance not explained by the bilinear model **C****S^T**, which should be related to the experimental error in the raw measurements. The **E** distribution maps of the MCR-ALS analyses performed in this study can be consulted in ESM Fig. S2. The MCR-ALS bilinear decomposition is performed for a number of components (number of columns of **C** and of rows of **S^T**) that explain optimally the data matrix **D** and diminished the error matrix **E**. MCR-ALS was applied using a different number of components. After results inspection, this number of components was selected as the one that explained most of the data variance and enabled their simplest interpretation. In MCR-ALS, the bilinear decomposition given in Eq. 1 is solved using an alternating least-squares optimization under constraints. In this work, the ALS optimization was performed under the constraints of non-negativity for the distribution maps (**C**) and spectra (**S^T**) of the components. The quality of the model MCR-ALS is measured with parameters, such as the percentage of variance explained (R^2) and the lack of fit (*lof*).

Explained variance:

$$R^2(\%) = 100 \times \left(1 - \frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2} \right) \quad (2)$$

Lack of fit (*lof*):

$$lof(\%) = 100 \times \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \quad (3)$$

where e_{ij} are the elements of the **E** matrix and d_{ij} are the elements of the raw dataset **D**. Subindexes i and j refer to the pixel and the wavenumber, respectively. Further details about the MCR-ALS method can be found in previous works [16, 27]. After the image data matrix bilinear decomposition

Table 1 Pixel dimensions of the data cubes and matrices of the hyperspectral images used in this study. *The first image data resizing is for the three-block data fusion at low resolution. The second image data resizing is for the fusion of IR and RGB images at a higher resolution

	MALDI	IR	RGB
x-pixel	50	301	4512
y-pixel	31	185	2975
total pixels	1550	55,685	13,423,200
hyperspectral dimension	102 ROI	427 wavelengths	3 channels
Data cube dimensions	50 × 31 × 102	301 × 185 × 427	4512 × 2975 × 3
Data matrix dimensions	1550 × 102	55,685 × 427	13,423,200 × 3
Resizing for fusion 1*	50 × 31 × 102	50 × 31 × 427	50 × 31 × 3
2*	–	301 × 185 × 427	301 × 185 × 3

by MCR-ALS, each of the C matrix columns are refolded according to the original image size (x - y pixels), which recovers the spatial information about the distribution of each of the resolved components on the image (distribution map) of the analyzed sample (tissue). Each of these components can be simultaneously identified from the counterpart row spectra in matrix S^T , containing information about MS, IR, or RGB contributions. MCR-ALS has already demonstrated its ability to deal with IR and MS imaging data, and also for multiset images, in which different image data matrices are column-wise augmented [18]. In this work, MCR-ALS has been applied to the analysis of the images of the same tissue acquired by different spectroscopic techniques, MS, IR, and RGB, in what is named multimodal imaging or image data fusion. MCR-ALS was applied first to the individual MALDI-MS and IR images separately, and then to the fused multimodal images. In the case of the 3-multimodal images fused dataset, the finally selected number of components was set to 6.

Lipid identification and IR band assignment

The peaks of the mean spectrum of MSI data were fragmented in the MALDI-TOF instrument (*lift* method) to obtain characteristic fragments of each compound and perform tentative identification. For each MCR-ALS resolved component, lipid compounds with mass intensities higher than the 25% of the maximum intensity were considered relevant and therefore identified. The fragment peaks of the phospholipid headgroups and acyl chains were obtained by negative ion collision-induced dissociation (CID), and compared with the fragments and mass values found in the literature [28] and public online databases such as LipidMaps [29]. The assignation of IR bands was done using public databases and available literature [30].

High-resolution distribution of MCR-ALS-resolved components

As said before, MSI data could only be obtained at lower spatial resolution than IR and RGB. Therefore, in order to correlate the information among the different spectroscopies and exploit the higher resolution of IR and RGB images, a new augmented image data matrix (D_H) was obtained using the IR and RGB images at their maximum IR resolution, as depicted in Fig. 2. Thus, the RGB image was resized to the resolution of the IR image (301×185), and both images were aligned and fused, as performed before for the three-low spatial resolution images above. Then, the IR-RGB information of S^T matrix obtained in the analysis of the fused low-resolution imaging data matrices (see below in the Results section) was projected by least-squares on to the high-spatial-resolution IR and RGB new fused data matrix, D_H , to calculate the new C_H matrix giving the sought information

about the distribution maps of the different components at higher spatial resolution.

K-means of resolved components

To improve the visualization and interpretation of the results, the C_H matrix, corresponding to the high-resolution distribution of the MCR-resolved components, was subjected to K -means segmentation, as implemented in the MIA Toolbox. The number of clusters was set to 6, the same number of components previously resolved using MCR-ALS.

Software and computer specifications

The calculations performed in the present work were carried out using MATLAB R2018a (The Mathworks Inc.), PLS ToolBox (Eigenvector Inc.), and MIA software (version 3.0.7, Eigenvector Inc.) running on a Fujitsu Celsius R940n workstation equipped with two Intel Xeon CPU E5-2620v3 processors and 128 Gb RAM using Microsoft Windows 7. MCR-ALS analysis was performed using the MCR-ALS toolbox freely available at <http://www.mcrals.info/>

Results

MCR-ALS analysis of single technology images

Prior to the multimodal analysis of three fused images, each of the individual images was first analyzed by MCR-ALS. In Fig. 1, a schematic view of the different image data analysis steps is given.

Before MCR-ALS analysis, each of the images was unfolded to the corresponding data matrix and preprocessed as described in the methodology section and in Fig. 1. The MCR-ALS distribution map and mass spectra of the MCR-ALS components in the analysis of each of these individual images are given in Fig. 3. The four components resolved in the analysis of the MALDI-MS image revealed different molecular distributions within the analyzed sample tissue. The total variance explained by the model was 98.2% with a *lof* of 13%. The components 1 and 3 (31% and 11% of explained variance, respectively) occupied the central part of the tissue with different distributions, whereas component 2 and 4 (23 and 59% of explained variance) were located in the lateral outer parts of the tissue with distinct patterns (Fig. 3a). Each of these components had a specific chemical identity. Component 1 was mostly composed by a phosphatidylcholine plasmalogen (PC(P-30:1)) and phosphatidic acid (PA(30:2)), whereas component 3, which shared some regions of the tissue surface with component 1, presented a very diverse composition including several species derived from the phosphatidic acid (PA), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and

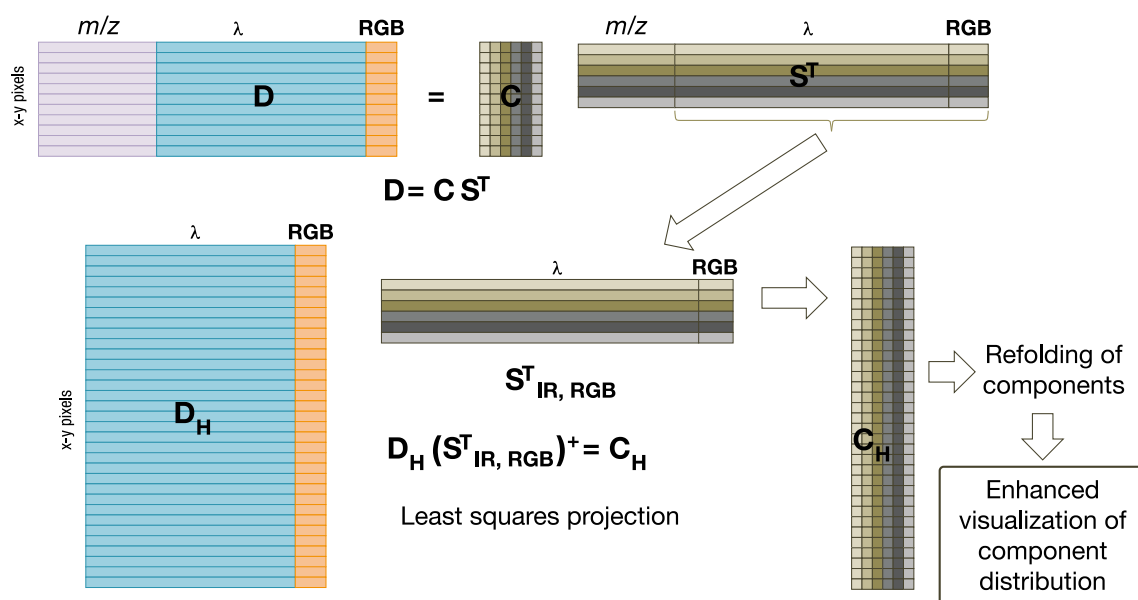


Fig. 2 Least-squares projection of S^T obtained by MCR-ALS of multimodal low spatial resolution images on high-spatial-resolution image data

phosphatidylinositol (PI). Components 2 and 4, located in the sides of the tissue image, had different compositions and shared strong signals of different PI species. Detailed information about the lipid m/z values and fragments is available in ESM Table S1. The sum of explained variances of the individual resolved components is 124%. This is due to the non-orthogonality of the overlapped MCR-resolved components (e.g., PI (38:3) appears in components 2, 3, and 4 in combination with other different lipids).

Regarding the resolution of the IR image, six components were resolved (95.1% of explained variance, *lof* of 14%), as shown in Fig. 3b. Two of the six components resolved were not relevant to the tissue description since they were only located in the outer part of it. This is due to the data acquisition mode of the IR equipment, which records a squared area of the surface to analyze, in contrast to MALDI-MS acquisition, in which the borders of the tissue can be delimited before acquisition. IR signals from the slide surface and the OCT polymer present in the sample were resolved using other separate components (for more clarity, the results of these two components are available in ESM Fig. S3). Component 1 (24% of explained variance) was more present in the central part. Components 2 and 4 (12% and 10% of explained variance, respectively) were more abundant in the lateral parts of the tissue, and component 3 (41% of explained variance) had a wide distribution over the whole tissue image with a higher contribution in its right side (Fig. 3b). As previously mentioned for MALDI-MS image, the sum of the explained variance of each of the resolved components (126%) indicates that the information they provide is overlapped. The IR spectra resolved represent the sum of all the functional group signals of the compounds present in each of the components.

Component 1 was mainly constituted by protein and ester lipids, component 2 by alcohols and alkyl chains, component 3 by alkyl chains and proteins, and component 4 mainly by proteins and carboxylic acids. The RGB image of the same tissue sample confirmed the main differences observed in the MALDI-MS and IR images (Fig. 3c). The central part of the tissue was clearly redder than the lateral parts, which were prominently stained in a dark blue color. In the H/E staining, hematoxylin stains in dark blue the nuclei of the cells whereas eosin stains proteins, cytosol, and extracellular matrix in pink color. Hence, two main differentiated histological areas were present in the tissue analyzed, which were clearly recognized by the MS and IR imaging techniques, indicating the different chemical composition of each of these areas. However, the correlation between the MALDI-MS and IR chemical signals in these different areas cannot be assessed when images are analyzed separately. Therefore, in order to have a more exhaustive and reliable chemical description about the different areas described by H/E staining, the multimodal analysis of the fused data information provided by the three imaging modes was performed.

MCR-ALS analysis of multimodal fused imaging datasets

The next step leading to the multimodal image data fusion was to convert the IR and RGB image data blocks to the same spatial resolution of the MALDI-MS image. Then, images were aligned and data matrices were normalized by the first SVD value, as explained in the methodology section. Multimodal image fused data were then analyzed by MCR-ALS [31] to resolve the different components of the image

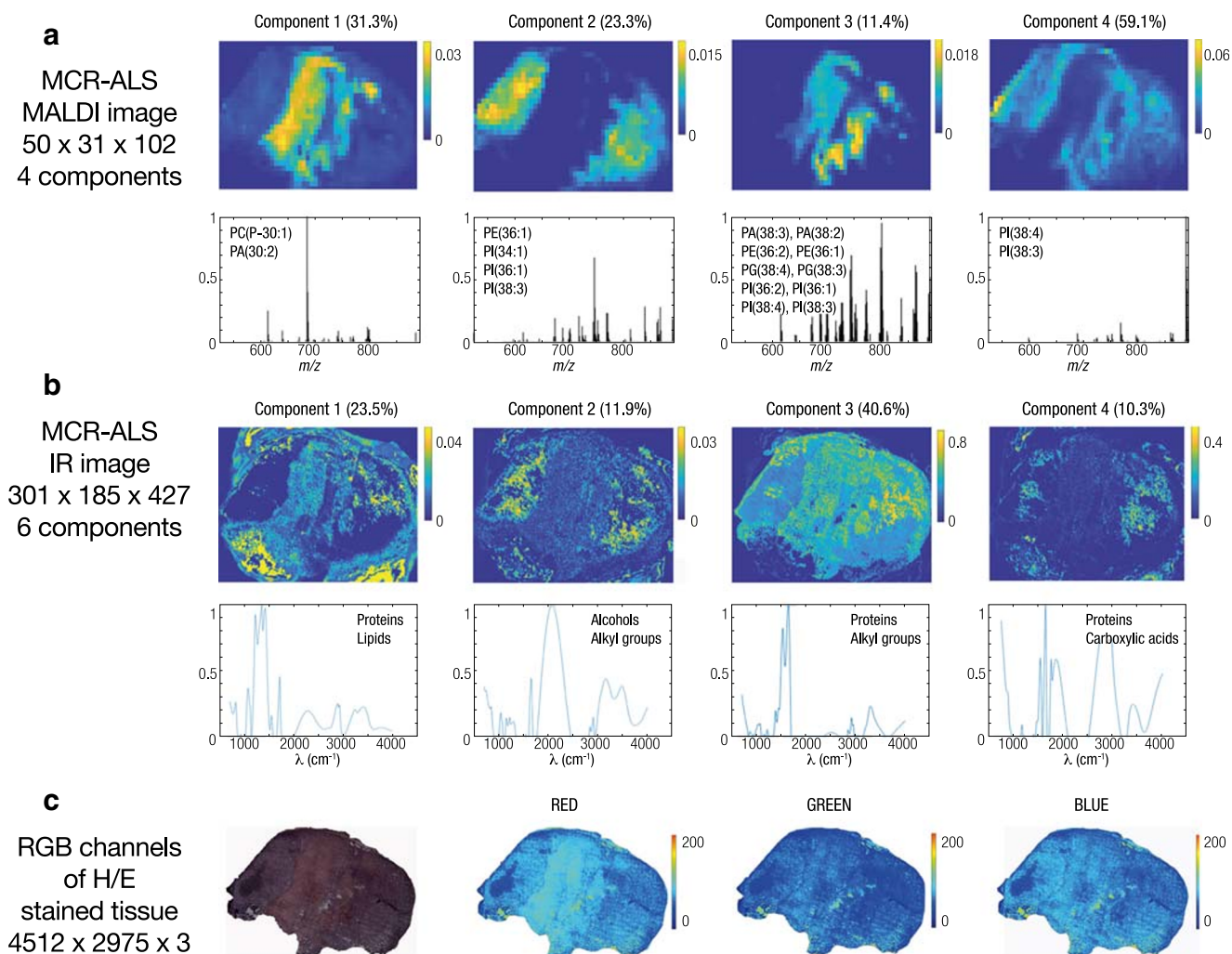


Fig. 3 Results of MCR-ALS of individual image datasets. Distribution maps and spectra of the different MCR-ALS resolved components from the **a** MALDI-MS image and **b** IR image. **c** Distribution of the red, green and blue channel intensities within the tissue. The mass spectrometry and IR MCR-ALS resolved spectra include the identification of the main

lipids and the main chemical structures and functional groups, respectively. PC: phosphatidylcholine; PA: phosphatidic acid; PE: phosphatidylethanolamine; PG: phosphatidylglycerol; PI: phosphatidylinositol. The numbers in brackets indicate the number of carbons and the total number of chain unsaturations

defined by their specific spatial distributions and spectra, giving the three types of chemical/biological information: MS lipid content, IR profile, and histological staining. The best resolution of the fused images was obtained using 6 MCR-ALS components (98.8% of explained variance, *lof* of 11%), 2 of which were discarded as they described mostly the surface outside the tissue, as it already happened previously in the individual resolution of the IR image. The homogeneous distribution observed in the **E** error matrix (see ESM Fig. S2) indicated that the difference between the calculated and the experimental data was not significant. The distribution maps of the informative four components resolved by MCR-ALS together with their chemical information and tissue staining patterns describing the analyzed tissue are given in Fig. 4. The complete information about the six components resolved is available in ESM Fig. S4.

The distribution patterns of these four components was clearly different, although two of them had a distribution in the side parts of the tissue (1 and 2) and the other two in the central part (3 and 4). A common feature in all the components is that their IR spectra share the same characteristic amide bands, which reflect that proteins are all over the tissue sample. Component 1 (19% of explained variance) and 2 (30%) had common lipids (PI(38:3) and PI(38:4) among other possible chemical constituents observed in the m/z values of the resolved mass spectra. In contrast, these two components had distinct IR spectra and different RGB contributions. The IR spectra of component 1 presented intense alcohol bands and an exclusive blue contribution in the RGB channels. Component 2 presented alkene bands and a more balanced contribution of red and blue colors on RGB. On the other hand, component 3 (17% of explained variance) and 4

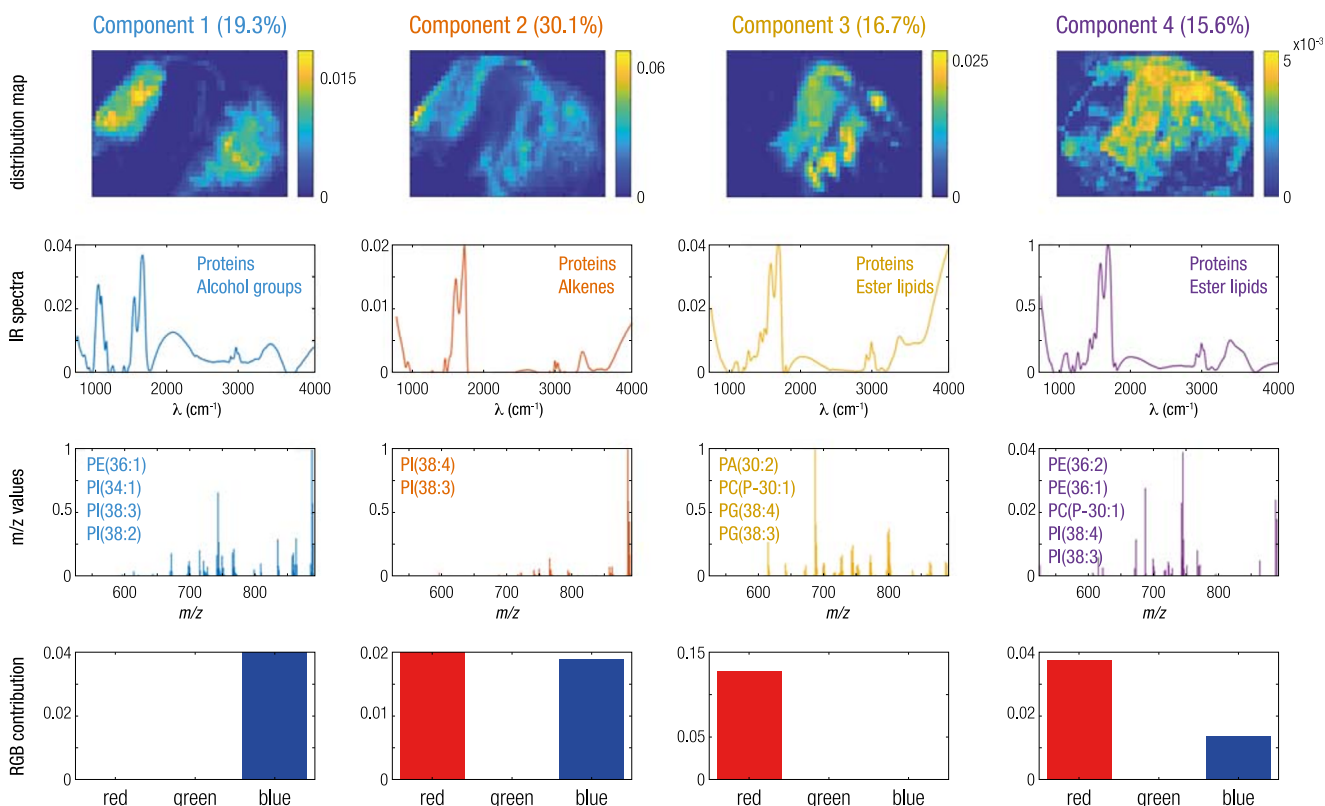


Fig. 4 Results of MCR-ALS analysis of the multimodal MS, IR, and RGB fused image data sets

(16%) presented a rather similar IR spectrum with the presence of protein and ester lipids bands. However, their lipid composition was very different (component 3 had PA, PC, and PGs in its composition, and component 4 presented PEs and PIs). RGB contribution of component 3 was exclusively represented by red, whereas component 4 had also some contribution of blue.

Finally, the spectra of the different components resolved by MCR-ALS in the analysis of the multimodal analysis of the images at low spatial resolution can be used to resolve the distribution maps of the same components at higher spatial resolution IR and RGB images. As said before, MSI data could only be obtained at lower spatial resolution than IR and RGB. However, since the spectral resolution (not the spatial) is the same in the images at low spatial resolution as in the images at high spatial resolution, the IR and RGB spectral information in the S^T matrix obtained in the multimodal analysis at low spatial resolution can be easily projected on the IR and RGB fused matrix at higher spatial resolution, D_H , to calculate the new C_H matrix at the higher spectral resolution (see Fig. 2). The resulting C_H matrix can be then refolded to give the high spatial resolution distribution maps of the MCR-resolved components. As a result, even if the chemical and biological information was the same in the images at low and high spatial resolution, the new C_H matrix, once refolded, provided higher resolution and more precise information about the localization of the resolved components (chemical

constituents) in the image or tissue sample, as shown in Fig. 5a.

To increase the interpretability of the resolved components, *K*-means segmentation was performed on the C_H matrix [32]. As a result, a new distribution map with the overlapped localization of the components was obtained, enabling the visualization of the distribution of the different chemical constituents in one single image (Fig. 5b).

Discussion

The aim of histology, either in biomedical or environmental toxicology studies, is to achieve the best characterization of the tissues under study. The characterization by histopathology techniques is often limited to general staining such as H/E or targeted staining performed using immunohistochemistry or immunofluorescence. Thus, the incursion of other imaging approaches more focused on tissue chemistry, such as vibrational and MSI techniques, represents a step forward in many research areas and clinical applications in which tissue interpretation is an essential issue.

In this work, the simultaneous analysis of two completely different types of chemical imaging techniques such as MALDI-MS and IR, together with an RGB image using a classical H/E staining of the same tissue slice from PDX is presented for the first time. The results obtained here indicated

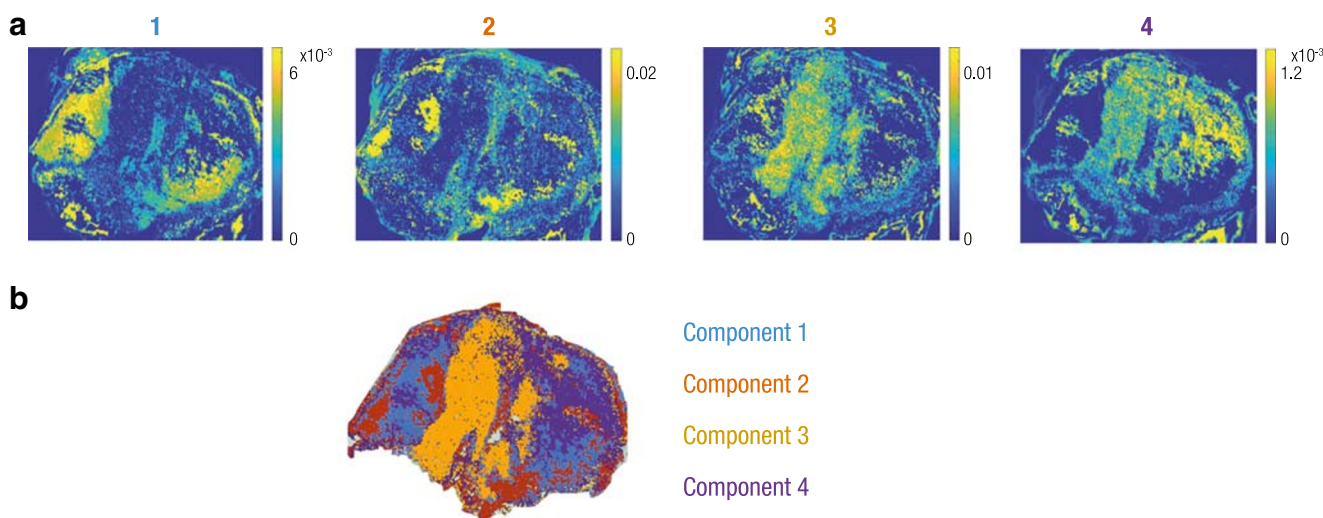


Fig. 5 Least-squares projection of S^T matrix on IR and RGB high-resolution fused data. **a** Distribution maps, C_H , of the four components resolved by MCR-ALS projected on the high-spatial-resolution image data. **b** Graphical representation of the K -means segmentation performed on C_H data

that, compared with the MCR-ALS analysis of the two MALDI-MS and IR images individually, the MCR-ALS applied to the multimodal fused images give more and better information about the characteristics of the analyzed tissue. The multimodal resolution of the three fused images revealed the different lipid compositions associated with specific IR fingerprints, their particular localizations, and their correlation with specific H/E staining color pattern. For instance, component 1, represented by 100% blue contribution in the RGB, is related to the presence of PI and PE lipid species. In addition, this external region of the image which is more stained in blue color, and therefore more associated with living cells in proliferation, is also related to an IR fingerprint with intense alcohol group absorption bands. This is in accordance with the probable presence of PI lipidic molecules, since they have five alcohol groups in their structure. In contrast, component 3, present in the central part of the imaged tissue, is mainly represented by the red RGB color, corresponding to the eosin staining of proteins, cytosol, and extracellular matrix. This fact, together with the low hematoxylin staining in this localization, indicated that the surface occupied by this 3rd component was a necrotic area. Therefore, the specific IR signature and the lipid profile found for this component should be associated with tissue necrosis. To illustrate the usefulness of applying the combination of multimodal chemical imaging with conventional histology, in a hypothetical study about the effectiveness of a cytotoxic compound, the IR results and the lipid profiles obtained by MSI could be useful to characterize and distinguish the cells that are still proliferative from the cells that undergo necrosis under a particular treatment.

Indeed, the use of high-resolution chemical imaging instrumentation is desirable to enhance the spatial correlation and overlapping of specific chemical signatures and discover characteristic distributions of the chemical constituents within the

tissue. Nowadays, MALDI-MS imaging instruments can already reach rather small pixel sizes ($10 \mu\text{m}$) and spectroscopic vibrational technologies can attain subcellular pixel sizes (below $1 \mu\text{m}$). Also, high-resolution mass spectrometers coupled to the imaging desorption and ionization sources (MALDI, DESI, SIMS, etc.) enable reliable identification of the chemical constituents of the analyzed sample, which is extremely useful to characterize the tissue and to ease the biological interpretation of the particular case under study. The combination of higher spatial resolution and higher spectral resolution instrumentations will provide more accurate information about the correlation between chemical and histological staining features. However, higher spatial resolution images require longer acquisition times and increased more expensive computer resources in order to handle and analyze the big data generated in multimodal hyperspectral imaging. Depending on the purpose of the study, a balance between resolution, time of acquisition, computer resources, and data management and analysis should be taken into consideration.

In this work, the limitation of the low spatial resolution of MALDI imaging instrumentation has been compensated using the higher resolution images obtained by the other two imaging techniques (IR and RGB). The spectra resolved during the analysis of the low spatial resolution images can be used in a second step to recover the distribution maps of the constituents at high spatial resolution by spectral projection. In this way, more detailed spatial information about the distribution of the chemical constituents within the analyzed tissue is achieved.

When the aim of the study is to compare and correlate the spectral features of the different tissues related to different treatments on healthy and diseased tissues, the analysis of the proposed multimodal data fusion strategy can be also used in the simultaneous analysis of images. In this way, the

specific spatial distribution and composition of the different chemical constituents and staining patterns can be correlated across the different tissue samples that are being compared. This methodology can be very useful to elucidate the relation between the chemical composition and staining signatures of diseased tissues and also to investigate the response of a tissue to specific treatments or environmental stressing conditions. The signatures and features resolved by the proposed multimodal image data fusion methodology can be extremely useful to understand the processes studied from the chemical and biological points of view, and also to identify the events occurring in different type of situations and experiments.

The methodology described in this work was applied using hyperspectral chemical imaging coupled to H/E staining, but this chemical information can also be coupled to other histological analysis techniques, such as immunohistochemistry, immunofluorescence, or other specific staining protocols of interest. Imaging data blocks containing tissue information can be added successively as long as their image pixels have been previously properly aligned. As more information is available in the new multimodal image, more precise and accurate will be the correlation between the chemical signatures to the histological stained features.

Conclusions

In this work, a new analytical procedure based on the multimodal fusion of images from different spectroscopy methods such as IR, MALDI-MS, and RGB is presented. On the one hand, the multimodal fusion of images at low resolution and their simultaneous MCR-ALS resolution provided information about the common localization of lipids, IR fingerprints, and H/E staining of each of the chemical constituents of the analyzed tissue sample. This type of multimodal fusion of images provides information about the correlation of specific lipids and chemical signatures with precise histological information defined by conventional and routine imaging procedures, such as H/E staining. On the other hand, the projection of the spectra of the sample constituents obtained in the analysis of multimodal images at a low spatial resolution on high-spatial-resolution IR and RGB images provide a more defined spatial delimitation of them on the investigated tissue. This procedure can be applied using different chemical imaging technologies, and to analyze one image or several images at once, as long as the pixels of the images of the same tissue sample obtained by the different hyperspectral imaging techniques are correctly aligned. The increase of knowledge about the tissue molecular composition and morphology provided by the proposed multimodal analysis can be important in many different research areas, such as in biomedicine, to better understand a disease and the response to treatments, in food research, or in environmental sciences, to analyze the

effects of environmental stressors on different organisms at molecular, cellular, and tissue levels.

Acknowledgments This work was supported by Generalitat de Catalunya (Suport a les activitats de Grups de Recerca, 2017 SGR 753), Spanish Ministry of Science and Innovation (Project CEX2018-000794-S), and FIS-PI14/00336 - FIS-PI18/00916 Grants, from the I+D+I National Plan, with the financial support from ISCIII-Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Underwood JCE. More than meets the eye: the changing face of histopathology. *Histopathology*. 2017;70:4–9. <https://doi.org/10.1111/his.13047>.
- Feldman AT, Wolfe D. Tissue processing and hematoxylin and eosin staining. *Methods Mol Biol*. 2014;1180:31–43. https://doi.org/10.1007/978-1-4939-1050-2_3.
- Meguro R, Asano Y, Odagiri S, Li C, Iwatsuki H, Shoumura K. Nonheme-iron histochemistry for light and electron microscopy: a historical, theoretical and technical review. *Arch Histol Cytol*. 2007;70:1–19. <https://doi.org/10.1679/aohc.70.1>.
- Meloan SN, Puchtler H. Chemical mechanisms of staining methods: Von Kossa's technique: what von Kossa really wrote and a modified reaction for selective demonstration of inorganic phosphates. *J Histotechnol*. 1985;8:11–3. <https://doi.org/10.1179/his.1985.8.1.11>.
- Ramos-Vara JA. Principles and methods of immunohistochemistry, in: *methods Mol. Biol.*, Humana Press Inc.; 2017, pp. 115–128. https://doi.org/10.1007/978-1-4939-7172-5_5.
- Beesley J. *Histopathology: advances in research and techniques*. Immunotherapy. 2011;3:825–8.
- Bunaciu AA, Hoang VD, Aboul-Enein HY. Vibrational microspectroscopy of human tissues analysis: review. *Crit Rev Anal Chem*. 2017;47:194–203. <https://doi.org/10.1080/10408347.2016.1253454>.
- Pahlow S, Weber K, Popp J, Wood BR, Kochan K, Rütger A, et al. Application of vibrational spectroscopy and imaging to point-of-care medicine: a review. *Appl Spectrosc*. 2018;72:52–84. <https://doi.org/10.1177/0003702818791939>.
- Lyng FM, Ramos IRM, Ibrahim O, Byrne HJ. Vibrational microspectroscopy for cancer screening. *Appl Sci*. 2015;5:23–35. <https://doi.org/10.3390/app5010023>.
- Harrison JP, Berry D. Vibrational spectroscopy for imaging single microbial cells in complex biological samples. *Front Microbiol*. 2017;8. <https://doi.org/10.3389/fmicb.2017.00675>.
- Türker-Kaya S, Huck CW. A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. *Molecules*. 2017;22. <https://doi.org/10.3390/molecules22010168>.
- Fu X, Ying Y. Food safety evaluation based on near infrared spectroscopy and imaging: a review. *Crit Rev Food Sci Nutr*. 2016;56:1913–24. <https://doi.org/10.1080/10408398.2013.807418>.
- Buchberger AR, DeLaney K, Johnson J, Li L. Mass spectrometry imaging: a review of emerging advancements and future insights.

- Anal Chem. 2018;90:240–65. <https://doi.org/10.1021/acs.analchem.7b04733>.
14. Norris JL, Caprioli RM. Imaging mass spectrometry: a new tool for pathology in a molecular age. *Proteomics Clin Appl*. 2013;7:733–8. <https://doi.org/10.1002/prca.201300055>.
 15. Nilsson A, Goodwin RJA, Shariatgorji M, Vallianatou T, Webborn PJH, Andr n PE. Mass spectrometry imaging in drug development. *Anal Chem*. 2015;87:1437–55. <https://doi.org/10.1021/ac504734s>.
 16. Tauler R. Multivariate curve resolution applied to second order data. *Chemom Intell Lab Syst*. 1995;30:133–46. [https://doi.org/10.1016/0169-7439\(95\)00047-X](https://doi.org/10.1016/0169-7439(95)00047-X).
 17. Bedia C, Tauler R, Jaumot J. Compression strategies for the chemometric analysis of mass spectrometry imaging data. *J Chemom*. 2016;30. <https://doi.org/10.1002/cem.2821>.
 18. Bedia C, Tauler R, Jaumot J. Analysis of multiple mass spectrometry images from different *Phaseolus vulgaris* samples by multivariate curve resolution. *Talanta*. 2017;175:557–65. <https://doi.org/10.1016/j.talanta.2017.07.087>.
 19. Olmos V, Ben tez L, Marro M, Loza-Alvarez P, Pi na B, Tauler R, et al. Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images. *TrAC - Trends Anal Chem*. 2017;94:130–40. <https://doi.org/10.1016/j.trac.2017.07.004>.
 20. Piqueras S, Bedia C, Beleites C, Krafft C, Popp J, Maeder M, et al. Handling different spatial resolutions in image fusion by multivariate curve resolution-alternating least squares for incomplete image multisets. *Anal Chem*. 2018. <https://doi.org/10.1021/acs.analchem.8b00630>.
 21. Neumann EK, Comi TJ, Spegazzini N, Mitchell JW, Rubakhin SS, Gillette MU, et al. Multimodal chemical analysis of the brain by high mass resolution mass spectrometry and infrared spectroscopic imaging. *Anal Chem*. 2018;90:11572–80. <https://doi.org/10.1021/acs.analchem.8b02913>.
 22. Mart nez-Aranda A, Hern ndez V, Guney E, Muix  L, Foj R, Baixeras N, et al. FN14 and GRP94 expression are prognostic/predictive biomarkers of brain metastasis outcome that open up new therapeutic strategies. *Oncotarget*. 2015;6:44254–73. <https://doi.org/10.18632/oncotarget.5471>.
 23. Eilers PHC. A perfect smoother. *Analytical Chemistry*. 2003;75:p3631–6. <https://doi.org/10.1021/ac034173t>.
 24. Gorrochategui E, Jaumot J, Tauler R. ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets. *BMC Bioinformatics*. 2019;20:256. <https://doi.org/10.1186/s12859-019-2848-8>.
 25. Bedia C, Tauler R, Jaumot J. Compression strategies for the chemometric analysis of mass spectrometry imaging data. *J Chemom*. 2016;30:575–88. <https://doi.org/10.1002/cem.2821>.
 26. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem*. 2006;78:4281–90. <https://doi.org/10.1021/ac051632c>.
 27. Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. *Chemom Intell Lab Syst*. 2015;140:1–12. <https://doi.org/10.1016/J.CHEMOLAB.2014.10.003>.
 28. Zemski Berry KA, Hankin JA, Barkley RM, Spraggins JM, Caprioli RM, Murphy RC. MALDI Imaging of Lipid Biochemistry in Tissues by Mass Spectrometry. n.d.. <https://doi.org/10.1021/cr200280p>.
 29. Lipidmaps. (n.d.) lipidmaps.org.
 30. Larkin PJ. Infrared and raman spectroscopy : principles and spectral interpretation. Elsevier; 2011.
 31. de Juan A, Tauler R. Data fusion by multivariate curve resolution, in: *data Handl. Sci Technol*, Elsevier Ltd. 2019;205–233. <https://doi.org/10.1016/B978-0-444-63984-4.00008-9>.
 32. Piqueras S, Duponchel L, Tauler R, De Juan A. Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares. *Anal Chim Acta*. 2011;705:182–92. <https://doi.org/10.1016/j.aca.2011.05.020>.
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.