



HAL
open science

Data Processing

Landry Blanc

► **To cite this version:**

Landry Blanc. Data Processing. Master. Cours sur le traitement des données par les outils informatiques, y compris en IA, Frejus, France. 2024. <hal-05490726>

HAL Id: hal-05490726

<https://hal.science/hal-05490726v1>

Submitted on 3 Feb 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

DATA PROCESING

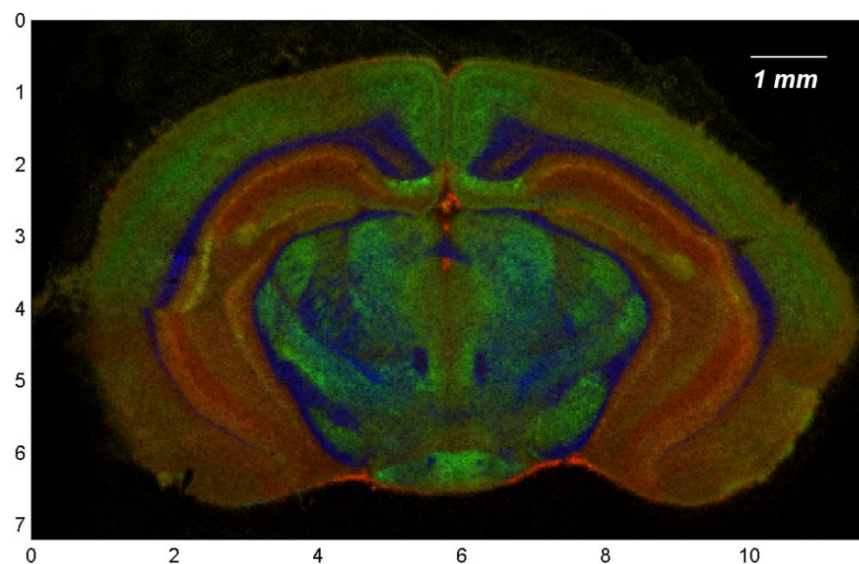
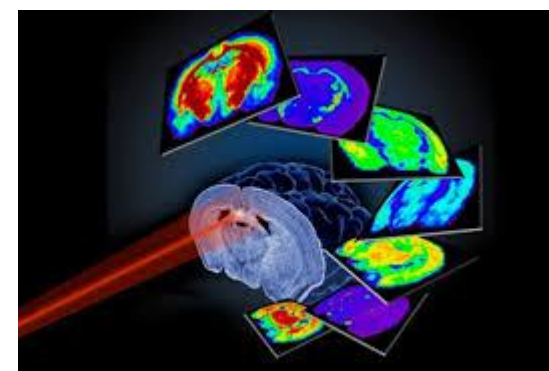
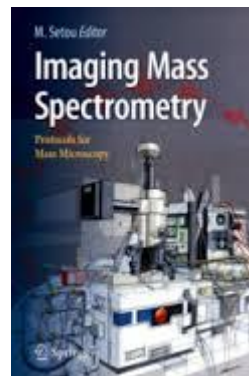
(IMPORT, LOGICIELS COMMERCIAUX ET LIBRES, INTERPRÉTATION DE SPECTRES,
ANALYSE STATISTIQUE DE GRANDS JEUX DE DONNÉES)

Ecole thématique GDR MSI – 2024

Landry Blanc



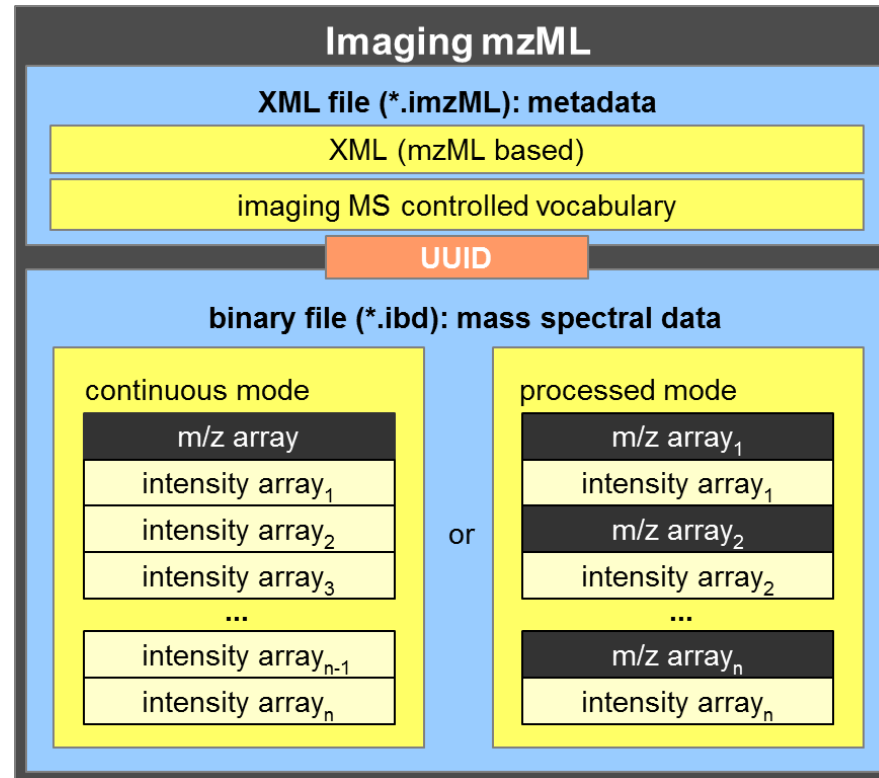
- **DATA PROCESSING: IMZML AND FREE / OPEN SOURCE SOFTWARE**
- **TOWARD MULTIMODAL IMAGING**



Existing MS software are often limited to proprietary data formats and make only limited use of MSI information. Several software tools have been developed to enable the flexible exchange and processing of MSI data, using a common format called **imzML** for “imaging mzML”.

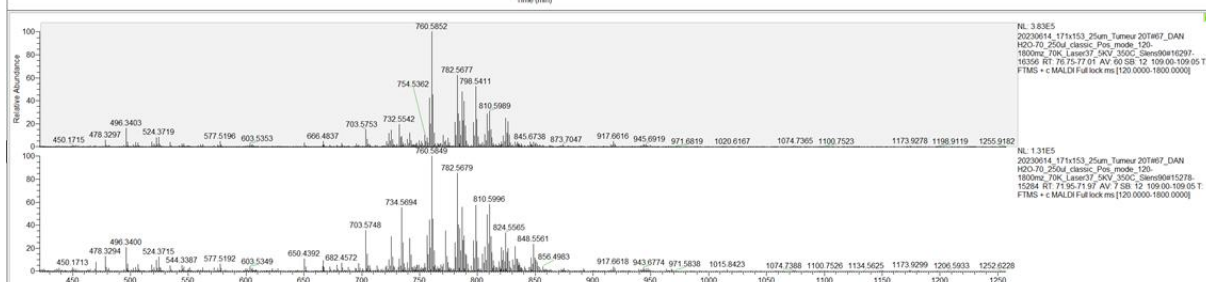
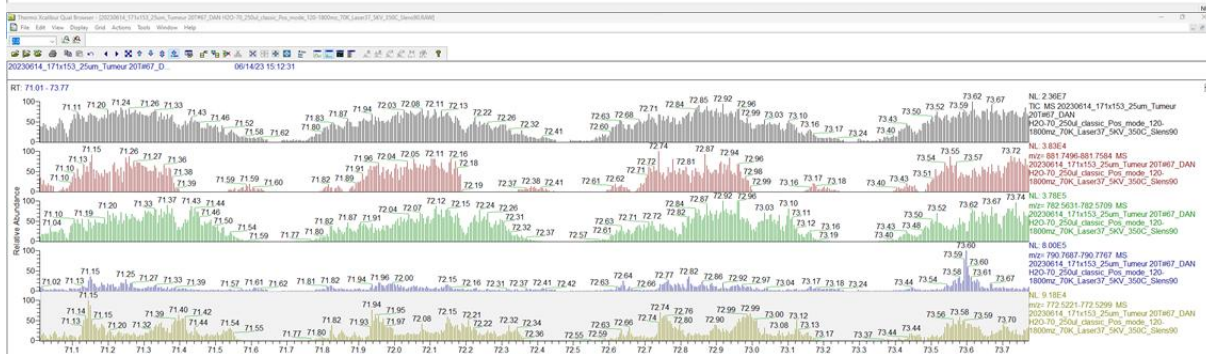
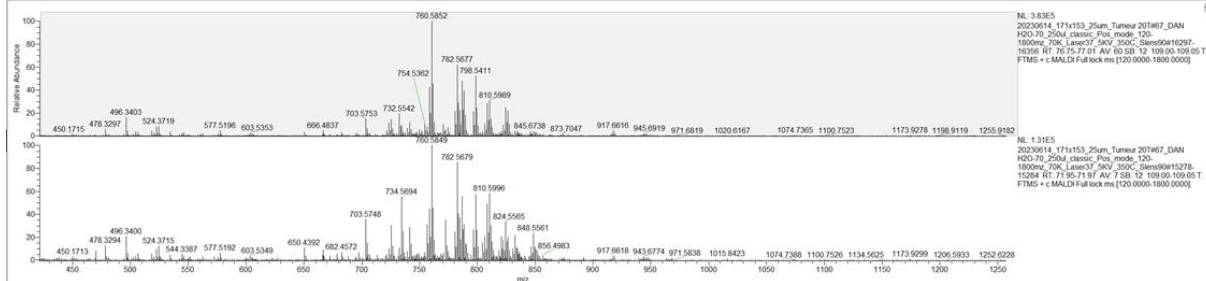
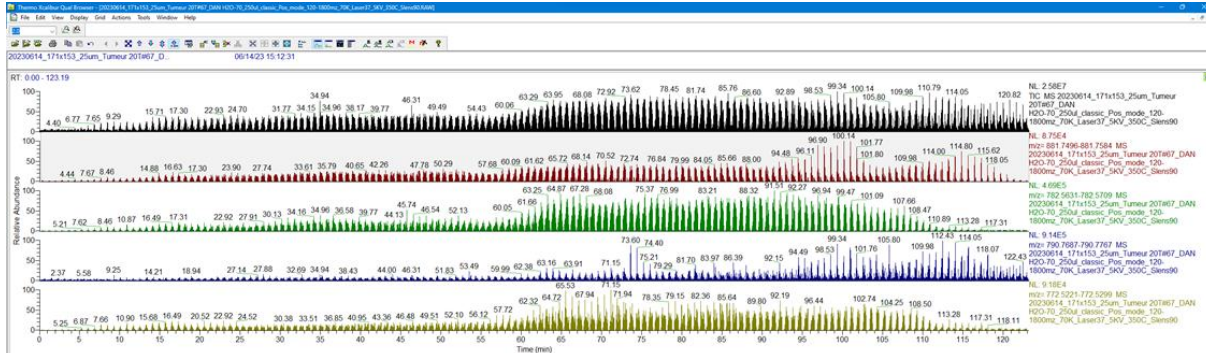
imzML = Flexibility + Exchangeability + Diversity

- **Workflow for MS imaging based on the imzML format:**
 - **MS imaging datasets obtained from different instrumental systems (ThermoFischer Scientific, Bruker Daltonics, AB Sciex, IonTOF, Shimadzu and Waters)**
 - **Converted to imzML**
 - **Data is displayed with identical parameters in the software packages**



T. Schramm et al, J. of Proteomics, 2012, 75, 5106-5110.

DATA PROCESSING: RAW TO IMZML CONVERTER



RAW+UDP to IMZML Converter V1.1r58

File Tools GUI Help

General Data

Selected input file:
20230614_171x153_25um_Tumour 20T#67_DAN H2O-70_250ul_classic_Pos_mode_120-1800mz

Found 26163 spectra in RAW file.

Imaging data/UDP data

Pixel count X	Pixel Count Y	Technical Info from RAW file	Skip # spectra at begin
171	153	Instrument SNO :Exactive Series slot #671	0
		Creation date :14/06/2023 15:12:31	
Pixel size X (um)	Pixel size Y (um)	Filter string : FTMS + c MALDI Full lock	Expect 26163 Spectra:
25	25		<input type="checkbox"/> omit spare spectra at end

Export as IMZML

File End

Contact information

Name
Your name here!

Address

E-Mail

Organisation

Reset to default Set as default...

TransMIT JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

- **RAW** (ThermoFischer): **“Raw to imzML” (JLU)**
- **GRD** (IonTOF), **Analyse** (AB Sciex, Shimadzu), **FID** (Bruker): **“ToimzML ModuleCea” (CEA)**
- **RAW** (Waters) : “High Definition Imaging” (Waters Corporation)
- **mzML**: **“imzML Converter” (Birmingham)** and **“mzml_to_imzml” (J.T. Prince)**
- **FID** (Bruker): “Bruker to imzML” (FlexImaging)
- **WIFF, FID, BAF, RAW....**: **“MSconverter” (ProteoWizard)**, *conversion in an intermediary format: mzML*
- **ASCII....**: **“ASCII to imzML” (Bayreuth)**

For more information: <https://www.ms-imaging.org/imzml/>

COMMERCIAL

- **MSIReader v2** : Statistical analysis, Classification, Quantification, etc. (David C Mudiman, Meredith C Muddiman, NC State University, MSI Software Solutions)



- **Quantinetix (Multimaging)**: Quantification, etc. (J. Stauber, ImaBiotech/Aliri)



- **MALDIVision**: ROI, Overlay of several masses up to 10, Co-registration, 3-D, etc. (PREMIER Biosoft)

- **SCiLS**: statistical analysis, overlay, co-registration, etc. (D. Trede, Bruker)



- **LipoSTAR** : Annotation, statistical analysis, overlay (Molecular discovery)



FREE

<https://www.ms-imaging.org/imzml/software-tools/>

- **SpectViewer: Comparison between images, Clustering, etc. (J.P. Both, CEA)**
- **DataCube Explorer: ROI (I. Klinkert, AMOLF)**
- **Mass Image Viewer: Read large data 100GB (K. Takahashi, AIST)**
- **mslQuant: Quantification, Co-registration, etc. (P. Källback, P. Andrén, Uppsala University)**
- **BioMap: Overlay of different masses, ROIs, etc. (M. Stoekli, Novartis)**
- **OpenMSI: Statistical analysis, etc. (B. Bowen, O. Rübel, Lawrence Berkeley National Lab)**
- **Mirion: Overlay of different masses, Fast images generation, Normalization of images, etc. (A. Hester, JLU)**

OPEN – SOURCE

<https://www.ms-imaging.org/imzml/software-tools/>

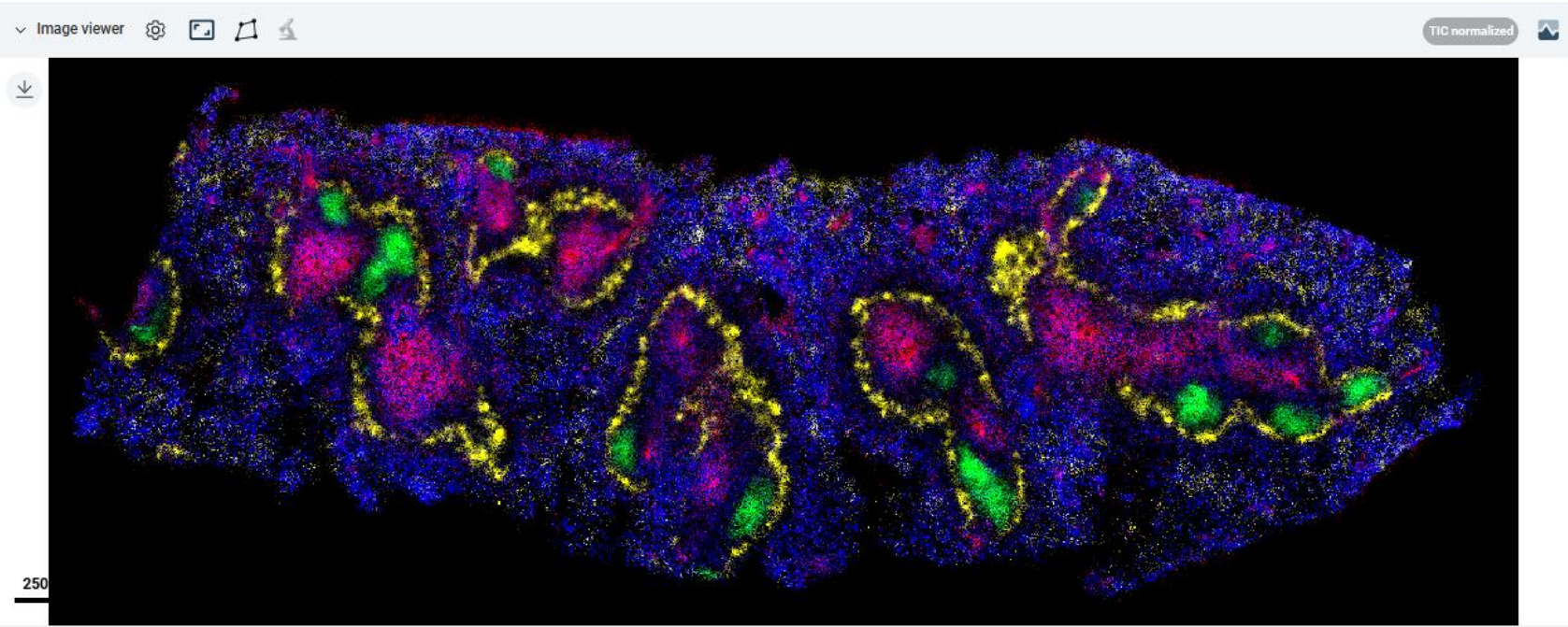
- **Cardinal: 3D images, Statistical tools, etc. (K. Bemis, O. Vitek, Purdue University) - R**
- **omniSpect: Multivariate analysis, etc. (M. Parry, Georgia Institute of Technonolgy) - Matlab**
- **MSI.R: Overlay of different masses, Co-registration, etc. (R. Winkler, CINVESTAV) -R**
- **METASPACE: Database annotation, Statistical tools, etc. (T. Alexandrov, EMBL Heidelberg) - python**
- **MSiReader v1: Overlay of different masses, Classification, Quantification, etc. (W.M. Keck, NC State University) - Matlab**
- **SpectralAnalysis: ROI, Statistical tools, etc. (A. Race, J. Bunch, National Physical Laboratory and University of Bayreuth) - Matlab**
- **Esmraldi: Co-registration, statistical tools, multimodal imaging, etc. (F. Grélard, CBMN) - Python**
- **Napari (MSI package) : imaging software (similar to ImageJ) with a MSI package for ,imzML - Python**

DATA PROCESSING: METASPACE

Add filter Database FDR Dataset

Annotation	m/z	Off-sample %	MSM	Max Intensity	Total Intensity	FDR
[C ₄₃ H ₈₁ O ₁₃ P - H] ⁻	835.5342	0 %	0.342	17762.9	531666144.0	5%
[C ₄₆ H ₈₀ NO ₁₀ P - H] ⁻	836.5447	0 %	0.035	7549.8	55740740.0	20%
[C ₄₃ H ₈₃ O ₁₃ P - H] ⁻	837.5499	0 %	0.485	26638.3	207607344.0	5%
[C ₄₃ H ₈₃ O ₁₃ P] ⁻	838.5577	0 %	0.590	12445.9	429847552.0	5%
[C ₄₆ H ₈₂ NO ₁₀ P - H] ⁻	838.5604	0 %	0.453	10419.0	273735488.0	5%
[C ₄₆ H ₈₈ NO ₁₀ P - H] ⁻	844.6073	78 %	0.058	1481.7	79692.2	20%
[C ₅₈ H ₈₈ O ₄ - H] ⁻	847.6610	95 %	0.039	1970.3	429936.0	20%
[C ₅₀ H ₉₄ NO ₇ P - H] ⁻	850.6695	100 %	0.278	8744.3	52157960.0	5%
[C ₄₈ H ₇₆ NO ₁₀ P - H] ⁻	856.5134	0 %	0.104	2013.7	2325032.5	10%
[C ₄₅ H ₇₉ O ₁₃ P - H] ⁻	857.5186	0 %	0.704	39880.9	1688511232.0	5%
[C ₅₀ H ₁₀₂ NO ₇ P - H] ⁻	858.7321	100 %	0.035	1945.1	478030.1	20%
[C ₄₅ H ₈₁ O ₁₃ P - H] ⁻	859.5342	0 %	0.215	17901.0	161370320.0	5%
[C ₄₅ H ₈₃ O ₁₃ P - H] ⁻	861.5499	0 %	0.747	29800.0	1265865088.0	5%
[C ₅₇ H ₈₆ O ₆ - H] ⁻	865.6352	0 %	0.143	4411.2	16798868.0	10%
[C ₅₀ H ₉₄ NO ₈ P - H] ⁻	866.6644	100 %	0.279	4293.6	17256770.0	5%
[C ₄₇ H ₈₁ O ₁₃ P - H] ⁻	883.5342	0 %	0.650	24333.7	1101808768.0	5%
[C ₄₇ H ₈₃ O ₁₃ P - H] ⁻	885.5499	0 %	0.930	347654.2	15826966528.0	5%
[C ₅₉ H ₉₈ O ₅ - H] ⁻	885.7341	100 %	0.038	2979.5	4201498.5	20%
[C ₄₇ H ₈₇ O ₁₃ P - H] ⁻	889.5812	0 %	0.033	6291.0	1930930.4	20%
[C ₄₈ H ₉₃ NO ₁₁ S - H] ⁻	890.6397	86 %	0.483	10479.3	90739.5	5%

[C₄₇H₈₃O₁₃P - H]⁻ 885.5499 m/z



> Molecules (18)

> Colocalized annotations

> Diagnostics

> Metadata

1 ... 4 5 6 7 8 9
 164 matching records
 FDR levels: 5% 10% 20% 50%

Columns Export to CSV

MSiReader v1.00 [Matlab R2017a 64-bit standalone] 17:13:14 Wednesday, 2019.06.12

m/z Imaging Data

Spots per Line: 600 Spot Spacing: 5 um

Number of Lines: 1000 Line Spacing: 5 um

Clear Data Data loaded: 20190607 Tableau2_600x1000
5um_CHCA AcN
Xyl_Pos_200-2000mz_5kV
300C_laser33_70K_line

imzML File:

MSiReader 1.00

W.M. KECK
FTMS LABORATORY FOR HUMAN HEALTH RESEARCH

Post Processing

Noise Correction

BG File: none

Normalization: none

Reference peak: 519.14

Norm Cutoff: 1

Norm Scale: 1

Heatmap Mode

MSi File Data

Custom Data Load

MS Navigation

m/z: 6

m/z Tolerance ±: 15 ppm

Abundance: window sum

m/z Step: 0.01

Scale Override: 75360.414

min

↑

↓

0

max

↑

↓

33912.186

Appearance

Interpolation: linear

Order: 0

Edit Colormap

Log Color Scaling



Instructions

- 1- Load mzXML file, imzML file, ..., MAT session file.
- 2- Select m/z of interest, change interpolation and perform post processing if needed.
- 3- Change colormap and view MS using toolbar menu functions.

DATA PROCESSING: ESMERALDI

Esmraldi

File Process ROI Segmentation Registration Analyze Help

Current image: 20220413_1107x427_5um_  

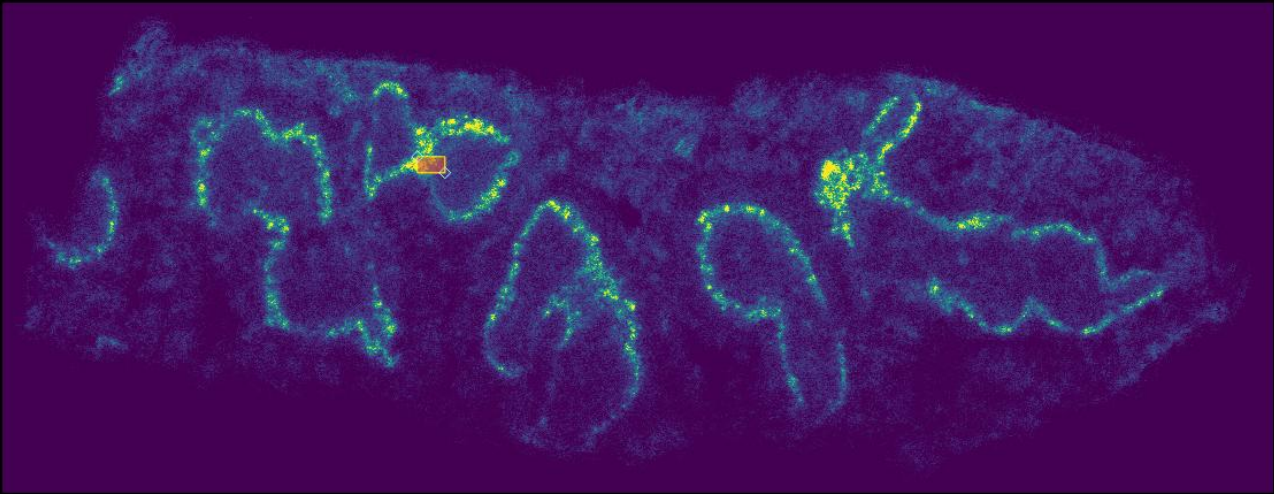
(0, 0): 0.000e+00
837.5490112304688

Peak picking

Prominence: 75 ?
Noise level: 0 ?

No peaks found yet.

OK Cancel

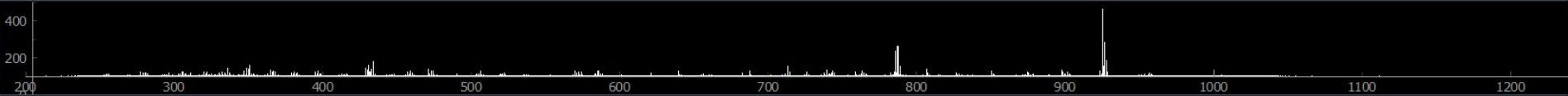


$\mu=5.677e+03$ $\sigma=3.511e+03$

ROI Norm Spectra

0 200 300 400 500 600 700 800 900 1000 1100 1200

Normalization
Type: Off TIC Ion (m/z) -1 Cut-off 0





400
200
200 300 400 500 600 700 800 900 1000 1100 1200

Type: Square Circle Polygon Image 20220413_1

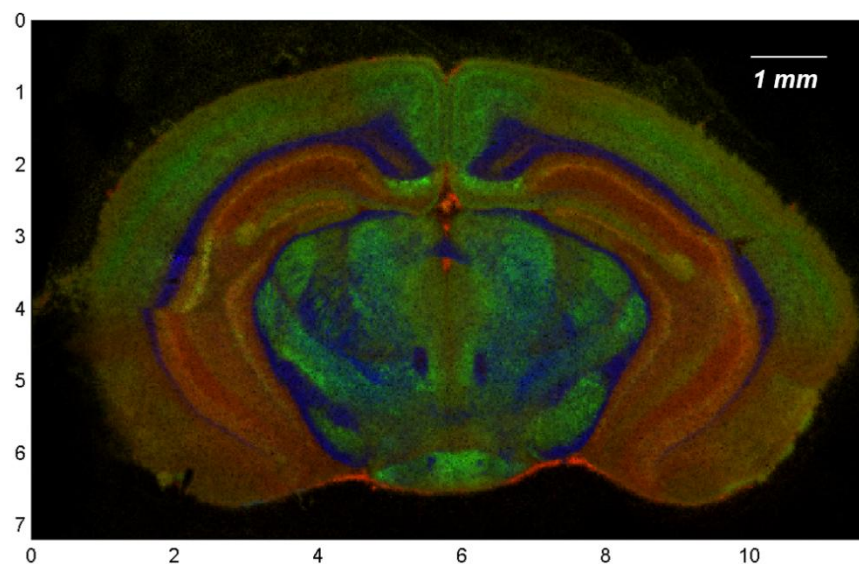
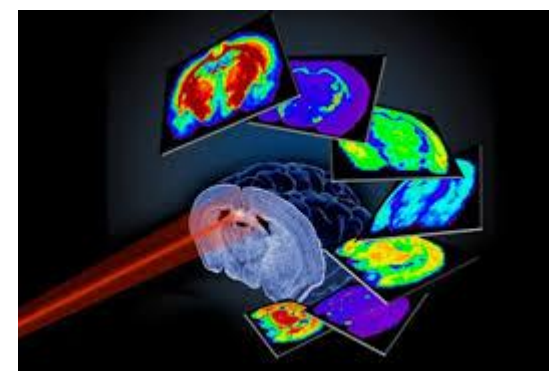
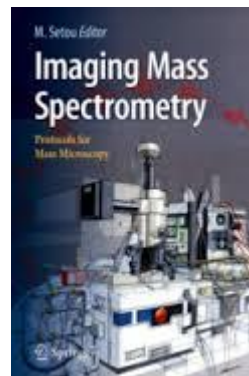
Range: 0 100 100

Plot spectra Reset ROI

m/z 837.5490
Tolerance 14

View:  

- DATA PROCESSING: IMZML AND FREE / OPEN SOURCE SOFTWARE
- **TOWARD MULTIMODAL IMAGING**



➤ *Spectral discrimination effects (ionization yields in complex mixture).*

➤ *Lack of (i) molecular information for some compound classes (proteins, polysaccharides, nucleic acids, etc.),*

↪ *Specific sample preparation*

➤ *(ii) correlation with anatomical structures or cellular types,*

↪ *Spatial resolution*

➤ *(iii) correlation with morphological nature of tissue,*

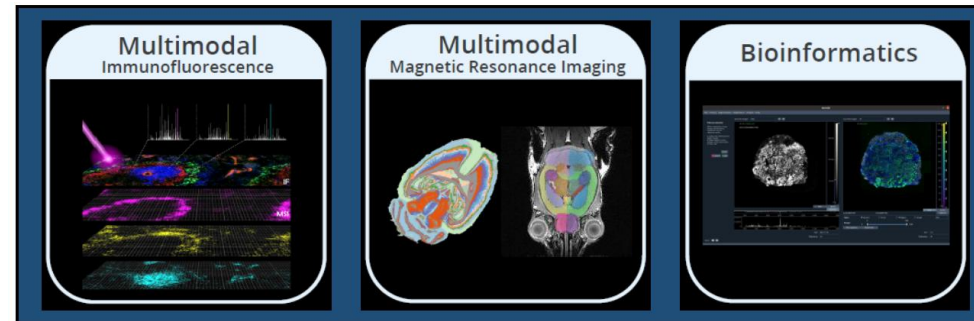
↪ *3-Dimension imaging*

➤ *(iv) computational strategy for multimodal imaging*

↪ *Dedicated software?*

Because no individual technique provides optimal answer to all question, it is highly attractive to combine information from two or more measurement platforms.

↪ **Multimodal Imaging**



- Multimodal imaging is becoming a grand challenge in chemical analysis.
 - To compensate for the limitation of individual imaging techniques,
 - To obtain more complementary chemical information,
- To elucidate the biochemistry (multi-omics investigation) of tissue samples,
 - For a better understanding of pathologies.



Which combinations

TABLE 1 | Collection of publications that combine MALDI-MSI with another modality. It attempts to classify each paper by its multimodal strategy and scientific goals. The list is separated *in vivo/ex vivo* and grouped by the spectral nature of its complementary modality, i.e., whether it is mono-, multi-channel, i.e., spectra with discrete bands, or hyperspectral, i.e., spectra with continuous bands.

Modality(ies)		Multimodal strategy			Scientific goal			References
		Acquisition(s) and registration method(s)	Data analysis		Result(s)	Omics	Biological background	
			Strategy(ies)	Method(s)				
<i>In-vivo</i>	MRI	Adjacent sections, & iconic: linear & non-linear	Class discovery	HSNE, & Pearson correlation	Colocalizing signal (3D)	Proteins	Murine kidney and pancreas, human colorectal cancer, microbial colonies, & human oral carcinoma	Abdelmoula et al. (2019)
	MRI	Adjacent sections, & iconic: linear	Class prediction, & class comparison	Automated anatomical interpretation & Wilcoxon rank test	Brain atlas, & region-specific ions	Proteins	Human brain	Verbeeck et al., (2017)
	MRI	Adjacent sections, & iconic: linear & non-linear	Class prediction	NMF & inverse NMF, & distance in reduced space	Colocalizing signal	Polysaccharides	Plant: wheat grain	Grélard et al., (2021)
	MRI	Adjacent sections, & iconic: linear & non-linear	Class discovery	Bisecting k-means, & Pearson correlation	Colocalizing signal (3D)	Proteins	Murine kidney	Oetjen et al., (2013)
	MRI	Adjacent sections Iconic: linear	Class comparison	Unpaired t-test	Region-specific ions (3D)	Proteins	Murine brain	Sinha et al. (2008)
	MRI	Adjacent sections, & iconic: linear	Overlay visualization	N/a	Integrated dataset (3D)	Proteins	Murine	Attia et al., (2012)

MULTIMODAL MALDI-MSI: CHOICE OF THE MODALITIES

Ex-vivo	H&E	Adjacent and same sections, & landmarks: linear	Class prediction	PLS	Pansharpening, & out-of-sample prediction	Lipids, proteins, metabolites, & drugs	Murine brain	Van De Plas et al. (2015)
	H&E	Adjacent sections, & registration not explained	Class comparison, & class prediction	Various multivariate analysis, ROC curve, & Kruskal-Wallis test	Biomarker discovery	Metabolites	Human Urachal Cancer	Neumann et al., (2021)
	H&E	Same section, & landmark	Class prediction	PLS regression	Out-of-sample prediction	Proteins	Murine brain, & kidney	Prentice et al., (2018)
	Autofluorescence Microscopy	Same and adjacent section, landmark, & iconic: Linear & nonlinear	Class comparison	Histogram of average intensities in multiple ROIs	Guided-acquisition. Region-specific ions	Lipids	Murine brain, kidney, spleen & <i>P. yoelii</i> -infected livers, human kidney	Patterson et al., (2018a)
	Autofluorescence Microscopy	Same & adjacent section, Landmark, & iconic: Linear and nonlinear	Class discovery	Weighted correlations	Colocalizing signal	Lipids, metabolites	Murine kidney, & brain	Patterson et al., (2018b)
	IHC	Same section, & registration not explained	Class comparison	ROC curves, & Mann Whitney U test	Biomarker discovery	Proteins	Human breast cancer & liver	Rujchanarong et al., (2021)
	IHC	Same section, & landmark	Class comparison	Overlay visualization, & 2D correlation	Integrated dataset	Lipids	Murine brain – Hunter's disease	Dufresne et al., (2017)
	IF H&E	Same section, & landmark	N/a	Overlay visualization	Region-specific ions	Lipids	Murine brain – Alzheimer's disease	Kaya et al., (2017b)
	IF	Same sections, & landmark	Class comparison, & class discovery	k-means clustering, & t-test	Integrated dataset	Lipids	Murine brain – Alzheimer's disease	Kaya et al., (2017a)
	IF	Same cells Registration not explained	Class discovery	Spatial correlation: Euclidean distance, Pearson correlation, & multivariate analysis: K-means clustering	Colocalizing signal	Lipids	Single cells	Nikitina et al., (2020)
IF	Same section, & landmark: linear	Class comparison, & class discovery	Correlation Network, Spearman rank order correlation, & Mann Whitney U test	Region-specific ions	Metabolites, lipids	Murine pancreas cancer	Prade et al., (2020)	

M. Tuck, L. et al, **Multimodal Imaging Based on Vibrational Spectroscopies and Mass Spectrometry Imaging Applied to Biological Tissue: A Multiscale and Multiomics Review.** *Analytical Chemistry* **2021** 93 (1), 445-477. DOI: 10.1021/acs.analchem.0c04595

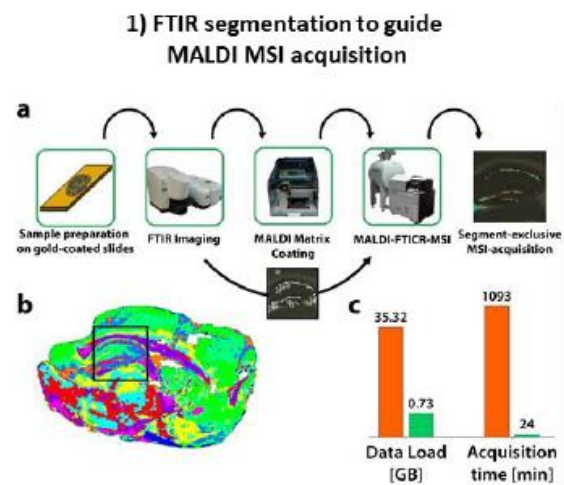
Ex-vivo	Imaging Mass Cytometry	Same section, & registration not explained	N/a	Overlay visualization	Integrated dataset	Lipids	Murine Brain, Human tonsil & breast cancer	Yagnik et al., (2021)
	FT-IR	Same section, & registration not explained	Class discovery	Random forest classifier	Integrated dataset, & colocalizing signal	Lipids, carbohydrates, & nucleic acids	<i>Eisenia fetida</i>	Ritschar et al., (2022)
	FT-IR	Same section, & Iconic: linear	Class comparison, & Class discovery	k-means clustering, & t-test	Integrated dataset, & guided acquisition	Metabolites, & lipids	Murine brain, & human gastrointestinal stroma tumors	Rabe et al. (2018)
	FT-IR	Same section, iconic	Class comparison, & class prediction	PCA Data Integration/ Laplacian Pyramid Sharpening, & ANOVA	Pansharpening, image fusion, & region-specific ions	Lipids, & peptides/ proteins	Murine brain	Neumann et al., (2018)

(Continued on following page)

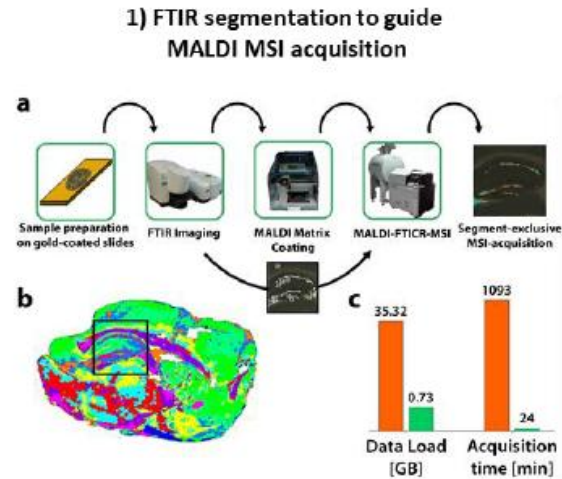
TABLE 1 | (Continued) Collection of publications that combine MALDI-MSI with another modality. It attempts to classify each paper by its multimodal strategy and scientific goals. The list is separated *in vivo/ex vivo* and grouped by the spectral nature of its complementary modality, i.e., whether it is mono-, multi-channel, i.e., spectra with discrete bands, or hyperspectral, i.e., spectra with continuous bands.

Modality(ies)	Multimodal strategy			Result(s)	Scientific goal		References
	Acquisition(s) and registration method(s)	Data analysis			Omics	Biological background	
		Strategy(ies)	Method(s)				
Raman Spectroscopy	Same section, landmark, & fiducial aided	Class discovery	PCA-PCA correlation	Integrated dataset, & colocalizing signal	Lipids, & peptides/ proteins	Cell Spheroids	Ahlf et al., (2014)
Raman Spectroscopy	Same section, & landmark: linear	Class prediction	PCA on combined dataset, & 2D correlation	Colocalizing signal	Lipids, & peptides/ proteins	Murine brain	Ryabchykov et al., (2018)
Raman Spectroscopy TOF-SIMS	Same section, & landmark: linear	Class prediction	NMF Data fusion	Pansharpening	Metabolites, & lipids	Murine brain	Race et al., (2020)
LA-ICP	Same & adjacent sections, & iconic: Linear and nonlinear	Class discovery, class comparison	Pearson correlation, k-means, Student's t-test	Colocalizing signal, region-specific ions	Metals, lipids, proteins	Murine spleen, & liver	Castellanos-Garcia et al., (2021)
TOF-SIMS	Same section, & landmark	Class discovery	Visual comparison	Guided acquisition, & colocalizing signal	Lipids	Human colon cancer	Desbenoit et al., (2018)
TOF-SIMS	Same section, & inherently registered	N/a	ROI selection	Guided acquisition	Metabolites	Biofilms	Lanni et al., (2014)
TOF-SIMS Microscopy	Same section, & landmark	Class discovery	Thresholding, granulometry, & visual assessment	Guided acquisition, & colocalizing signals	Lipids	Cells	Comi et al., (2017)
TOF-SIMS	Same sections, landmark: linear	Class prediction	CCA, & NMF	Pansharpening	Lipids	Murine brain	Borodinov et al., (2020)
DESI IMC H&E - IF	Same & adjacent sections, & landmark: linear	Class discovery	2D correlation	Colocalizing signal	Drugs	Murine Pancreatic cancer	Strittmatter et al., (2022)
DESI	Same section, Registration not explained	Class discovery	Visual comparison	Colocalizing signal	Lipids, & proteins	Murine brain, human glioma	Eberlin et al., (2011)
MALDI-MSI	Adjacent sections, & iconic: linear	Class prediction, class comparison	Multiblock OPLS	Pansharpening, region-specific ions	Lipids, & proteins	Murine hippocampus & Rat prostate	Wehri et al. (2020)
MALDI H&E	Same section, & inherently registered	Class comparison	Ion Fold change calculation	Integrated dataset, region-specific ions, & multi-omics	N-glycans, & peptides/proteins	human carcinomas, & tissue microarrays	Heijs et al., (2016)
LDI	Same section, & Landmark	N/a	Overlay visualization	Integrated dataset, & region-specific ions	Metabolites, Lipids	Murine brain, & lung	Fincher et al., (2020)

➤ *Multimodal imaging to guide acquisition*

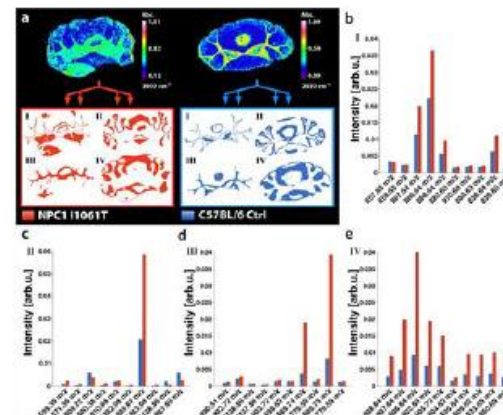


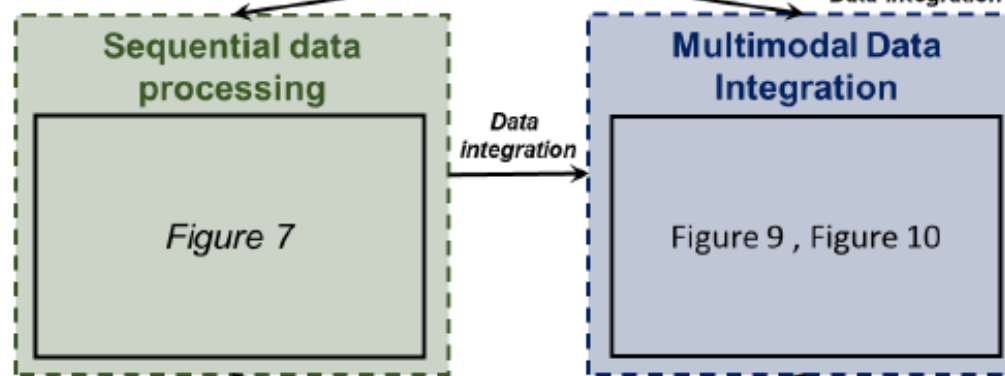
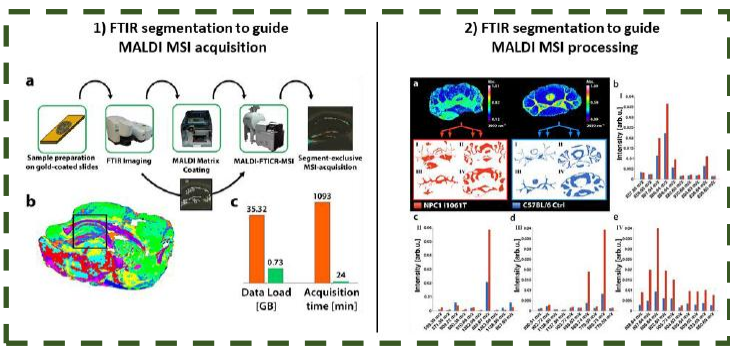
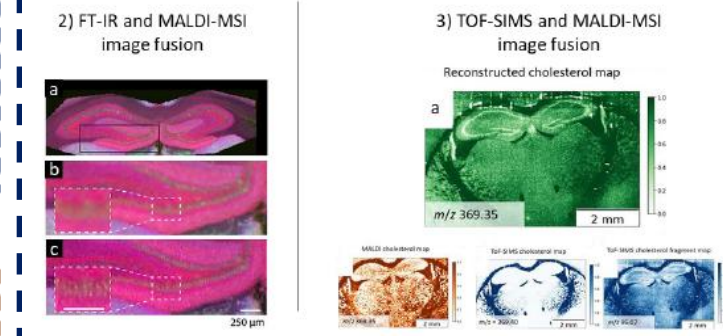
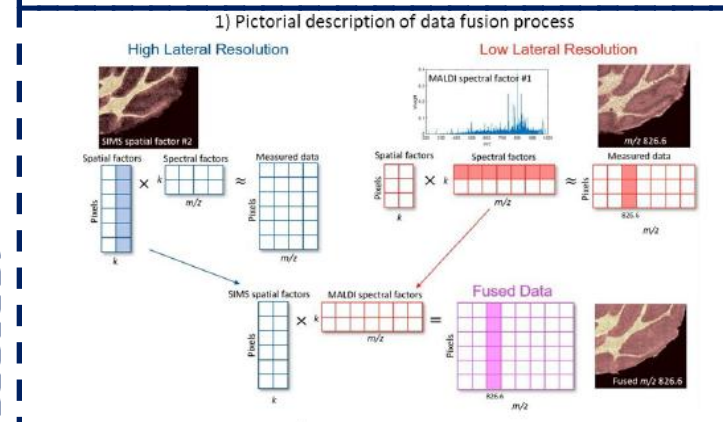
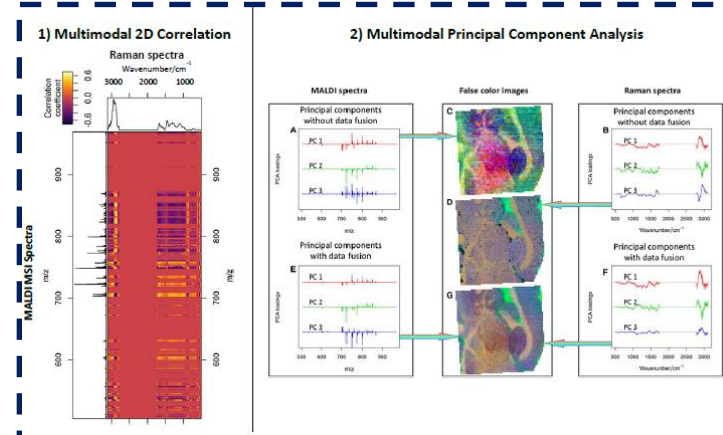
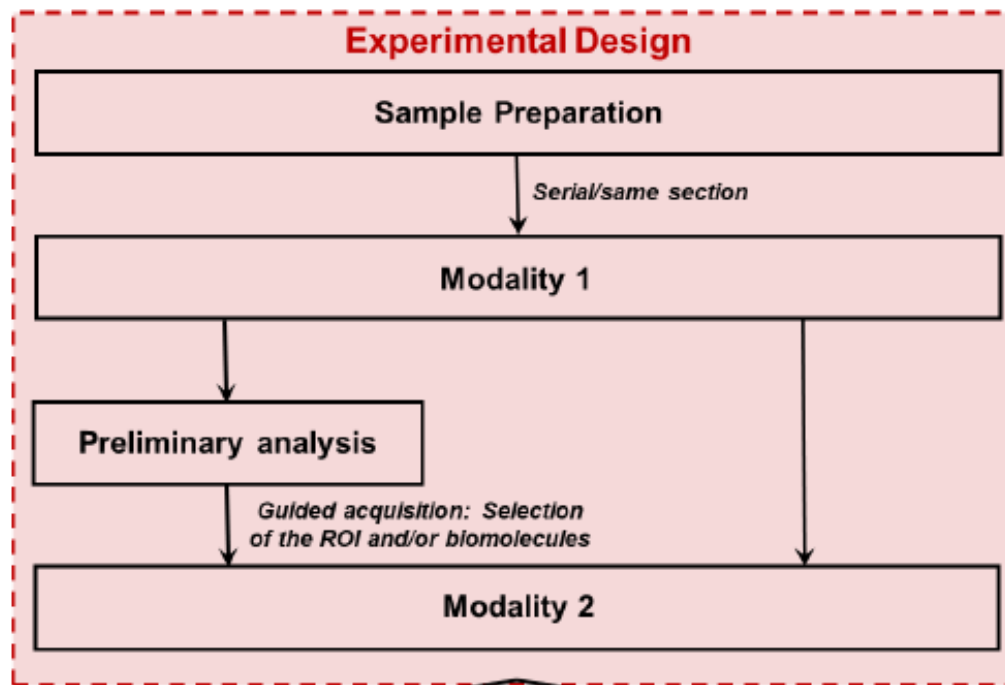
➤ Multimodal imaging to guide acquisition



➤ Multimodal imaging to guide data processing

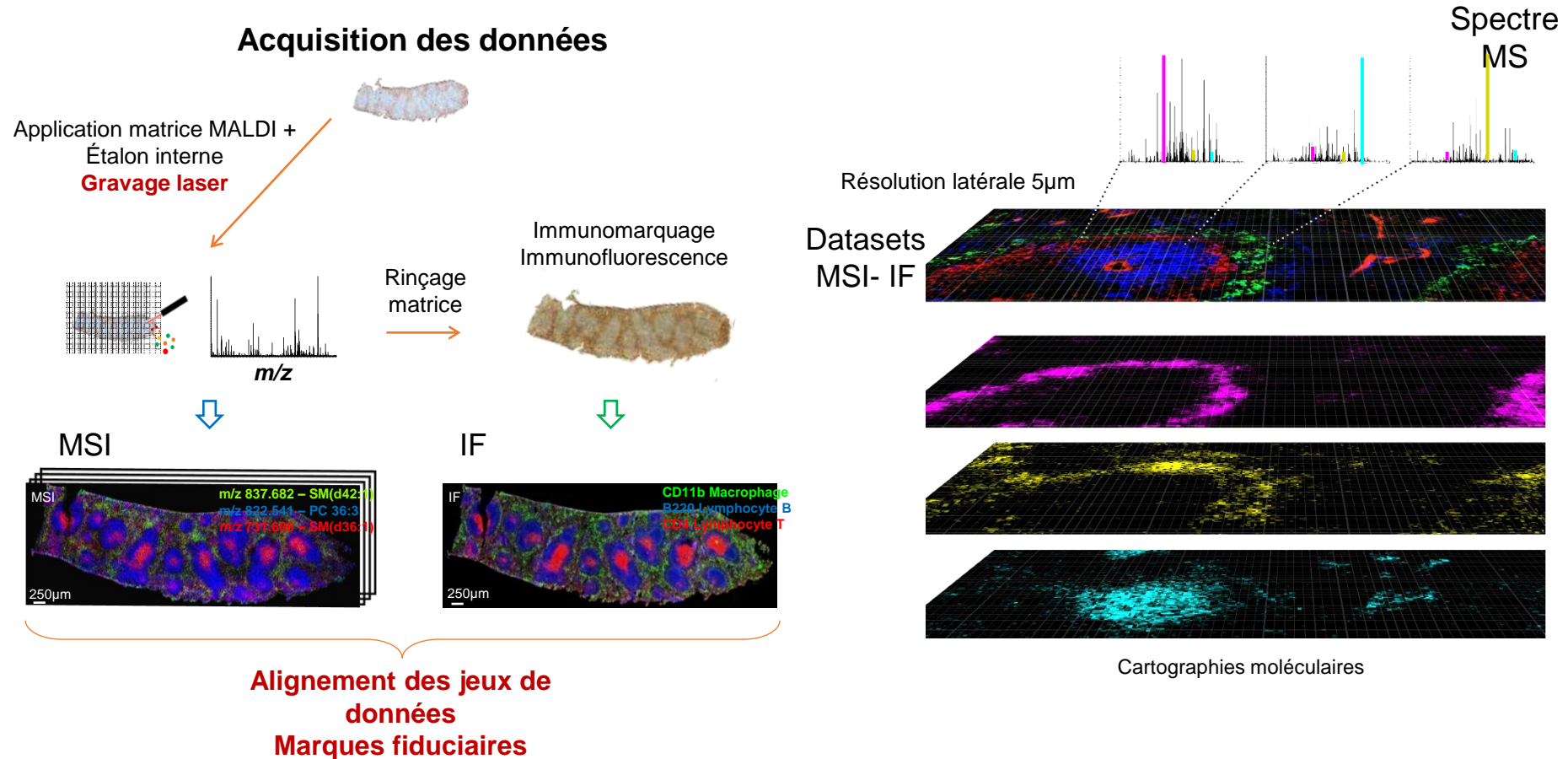
2) FTIR segmentation to guide MALDI MSI processing





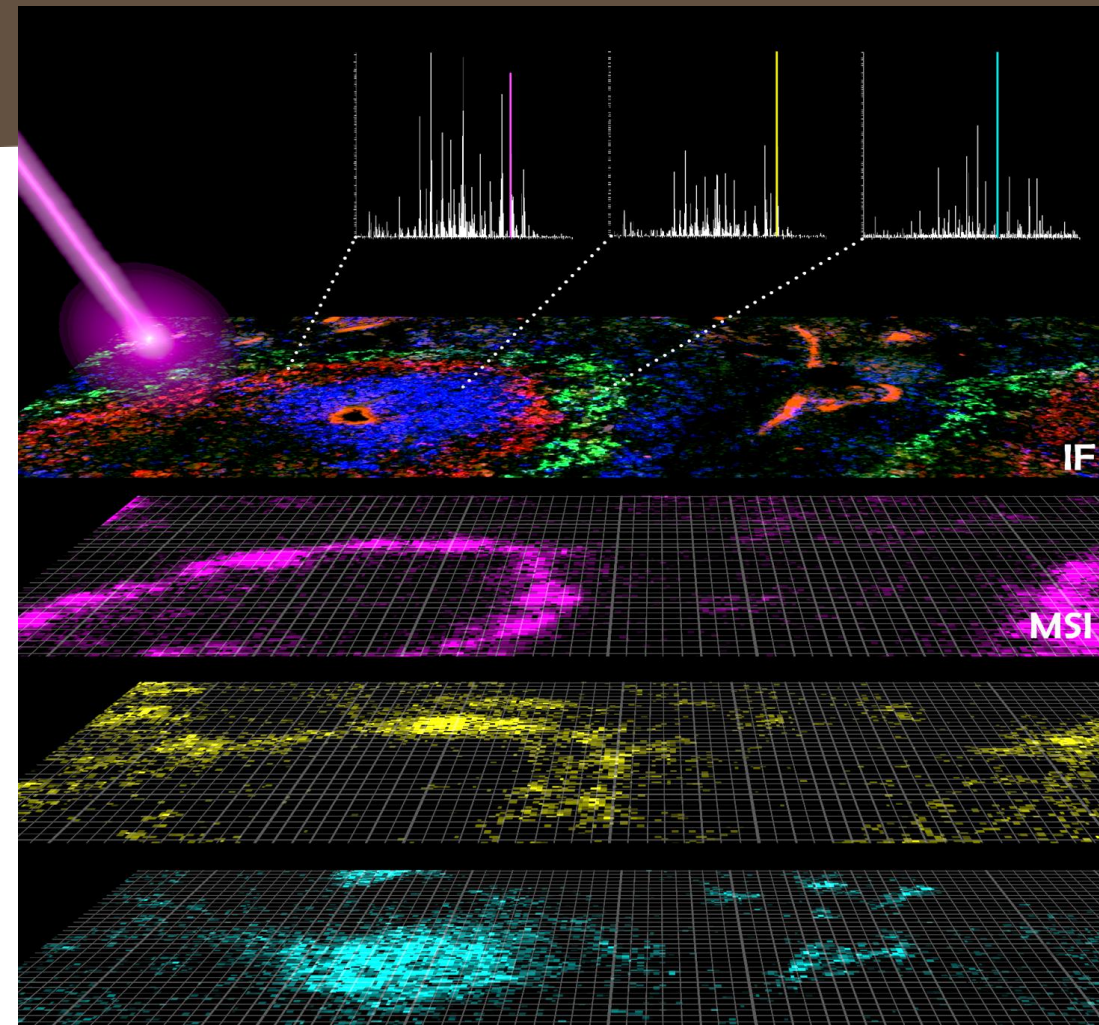
Interpretation / Discussion of results

- Objectif: Corréler images ioniques avec les types cellulaires
- Préserver l'intégrité du tissu entre les deux acquisitions (rendre MSI compatible avec IF)
 - Aligner précisément les jeux de données (marques fiduciaires gravées au laser)
 - Analyser les données fusionnées (stratégie computationnelle)



Changer de paradigme en MSI :
Ne plus raisonner en termes de pixels mais de cellules

Thank you for your attention



 **GDR** Groupement
de recherche
MSI Imagerie par Spectrométrie
de Masse

