



**HAL**  
open science

# Système de spectrométrie de masse in vivo pour la chirurgie guidée des cancers œsogastriques

Léa Ledoux

► **To cite this version:**

Léa Ledoux. Système de spectrométrie de masse in vivo pour la chirurgie guidée des cancers œsogastriques. Cancer. Université de Lille, 2023. Français. NNT : 2023ULILS110 . tel-04609461

**HAL Id: tel-04609461**

**<https://theses.hal.science/tel-04609461>**

Submitted on 12 Jun 2024

**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.



École Doctorale Biologie Santé Lille

Université de Lille

## THESE DE DOCTORAT

En vue de l'obtention du grade de Docteur en Science de l'Université de Lille

Biochimie et biologie moléculaire

Présentée par

**LEA LEDOUX**

# Systeme de spectrométrie de masse *in vivo* pour la chirurgie guidée des cancers œsogastriques

Soutenue à Villeneuve d'Ascq, le 30 novembre 2023

**Directrice de Thèse** : Pr. Isabelle FOURNIER. PRISM, Inserm U1192. Université de Lille

**Co-encadrante** : Dr. Florence Renaud. Unité mixte SIRIC CURAMUS et 938, Inserm.

### Membres du jury

**Président du jury** : Pr. Jean Paul Borg, CRCM, Inserm, CNRS, Université Aix-Marseille.

**Rapporteur** : Pr. Vincent Carré, LCP-A2MC – EA4632, Université de Lorraine.

**Rapporteur** : Pr. Christine Enjalbal, IBMM, UMR 5247, CNRS, Université de Montpellier.

**Examineur** : Pr. Vincent Sobanski. INFINITE, Inserm U1286. Université de Lille.

**Examineur** : Pr. Guillaume Piessen, Département de chirurgie digestive et oncologique, CHU de Lille.

**Invité** : Pr. Michel Salzet. PRISM, Inserm U1192. Université de Lille.





« Saute, et le filet apparaîtra »

**John Burroughs**

« Pour ce qui est de l'avenir, il ne s'agit pas de le prévoir mais de le rendre possible »

**Antoine de Saint Exupéry**

« Peu importe que vous ayez du style, une réputation, ou de l'argent,  
si vous n'avez pas de bon cœur, vous ne valez rien »

**Louis de Funès**





## Remerciements

Je tiens à exprimer mes plus sincères remerciements à tous ceux qui ont contribué à la réalisation de cette thèse.

Je remercie ma directrice de thèse Pr. Isabelle Fournier pour son soutien, son expertise et ses conseils tout au long de cette aventure. Sans ton aide précieuse, ce travail n'aurait pas été possible. Avec le Pr. Michel Salzet, directeur du laboratoire PRISM, vous m'avez fait confiance et m'avez permis de grandir scientifiquement et humainement grâce à toutes les opportunités, que ce soit au niveau des collaborations ou des congrès, que vous m'avez confié. Je n'oublierai jamais que c'est grâce à vous que j'ai pu allier ma passion des voyages avec ma plus récente passion pour la recherche.

Je remercie ma co-encadrante Dr. Florence Renaud de m'avoir transmis son expertise sur le cancer œsogastrique. Je suis également reconnaissante envers tous les membres de mon comité de thèse d'avoir accepté d'en faire partie. Encore merci aux Pr. Vincent Carré, Pr. Christine Enjalbal, Pr. Vincent Sobanski, Pr. Jean Paul Borg et Pr. Guillaume Piessen pour leur temps accordé à l'évaluation de mon travail ainsi que leurs commentaires et suggestions constructives.

Je remercie également l'Initiative d'excellence de l'Université de Lille, la Région Haut de France (fonds Feder de l'UE) et OCR, tout particulièrement Mme Dominique Tierny, pour leur soutien financier.

Un remerciement tout particulier à la SpiderMass Team ; Nina, Lucas, Adrien, Atef, Paul et Philippe. Je suis très heureuse que nous ayons pu travailler tous ensemble.

Je tiens également à remercier Alice, Antonella, Tristan, Irène, Marie, Alexandre, Julien, Maxence, Diala, Christelle et Lucie pour nos discussions et collaborations durant ces années de recherche. Vos idées novatrices et votre savoir-faire m'ont inspiré et m'ont permis d'enrichir mes recherches. Je remercie évidemment toutes les personnes du laboratoire pour leur aide, leurs conseils et leur bonne humeur.

Mes plus sincères remerciements en particulier à Kamel, Diego et Lucas qui passent leur thèse en même temps que moi. Ces 3-4 années au laboratoire ont été très enrichissantes à vos côtés, malgré le stress de ces derniers mois. Kamel, merci pour ta joie de vivre et ton humour à toute épreuve. Lucas, merci pour ta perspicacité et tes conseils si précieux. Diego, thank you for being you and allowing me to improve my english skills. I hope one day you will have the time/desire to catch up on the Grey's Anatomy's seasons. Je vous souhaite la plus belle réussite dans vos projets futurs et espère que vous trouverez votre voie.

Lydia et Laurine, je garde en mémoire nos discussions, souvent autour d'un verre, que ce soit à la MDE, dans la magnifique maison de Laurine, mon petit appartement ou l'appartement de Lydia. Ah non, on attend toujours la crémaillère depuis 2 ans maintenant ! Vous avez toujours été là dans les moments forts mais aussi les plus difficiles. Laurine, ta thèse sera parfaite, je n'ai aucun doute là-dessus. Lydia, γur-m 4 n yiseggasen mi tuyaleđ d tamdakelt n tidet. U tanemmirt imi i yi-tqebled akken lliy. Am ass-a ad awed nuvam ula d kemmi ad nefrah yissem.

Je remercie mes amies, Sarah, Meryam, Célia et Amina, pour leur soutien moral, leur amitié et leur compréhension durant cette période. Leur soutien m'a permis de surmonter les défis et les difficultés inhérentes à la réalisation de cette thèse. Pendant ces 3 ans, nos vies ont beaucoup changé mais seule une distance géographique nous sépare.

Plus les remerciements avancent, plus il est difficile de mettre par écrit ce que je ressens. Papa, maman, comment exprimer en si peu de mots à quel point je suis reconnaissante d'être votre fille ? Merci d'avoir toujours cru en moi, d'avoir été présents à toutes les étapes de cette belle aventure et surtout j'espère vous rendre fière. Merci d'avoir donné autant de votre temps pour corriger toutes mes fautes d'orthographe. Grande sœur, tu as été l'un de mes plus grands soutiens tout au long de ce doctorat même si nos chemins de vie se sont éloignés. Un coup de téléphone et tout va mieux. Merci d'avoir mis sur ma route Lucas et surtout mon filleul Lyam. Petit frère, ton sourire et ta protection me rendent heureuse et sereine. Dire qu'il y a encore quelques jours, tu croyais que ma thèse avait pour sujet « la fusion » ... Mamie, merci d'être toujours là, les week-ends à Hazebrouck m'ont permis de me ressourcer afin d'être plus forte la semaine. Hâte de voir ton post Facebook pour mon obtention de doctorat ! Seb et Claudine, j'espère que ma soutenance sera à la hauteur de la scène dans l'Étudiante et qu'on pourra fêter mon titre de docteur comme il se doit.

Yanis. C'est comme si cette thèse était la nôtre tant ton apport en termes de recherche, d'idées et de suggestions a grandement amélioré la qualité de ce travail, et a permis d'approfondir les résultats obtenus. Ton soutien, ton inspiration et ton aide ont été des facteurs clés de la réussite de cette thèse. C'est à ton tour maintenant, c'est tellement mérité. J'ai découvert récemment cette citation « Je ne sais où va mon chemin mais je marche mieux quand ma main serre la tienne » d'Alfred de Musset et elle décrit vraiment ce que je ressens en pensant à toi. Toi et moi, c'est la plus belle chose qui me soit arrivée. Je suis fière d'être devenue la femme que je suis à tes côtés. On a vécu déjà tellement de choses ensemble et ce n'est que le début.

Ssaramey Ad tfarhed yis-i. Hēmlay-k.

## Curriculum Vitae

### Formations

Baccalauréat scientifique – mention assez-bien. Lycée Charles de Gaulle à Dijon. Obtenu en juillet 2015.

Licence de Biologie, parcours biochimie. Université de Lille, cité scientifique, Villeneuve d'Ascq. Septembre 2015 – Juillet 2018.

Master Chimie et Sciences du vivant, parcours Bio-analytique – mention assez-bien. Université de Lille, cité scientifique, Villeneuve d'Ascq. Septembre 2018 – Août 2020.

Doctorat biologie santé. Université de Lille, cité scientifique, Villeneuve d'Ascq. Octobre 2020 – Octobre 2023.

### Expériences

Stage L3 au laboratoire CMSB UGSF UMR 8576 CNRS. Université de Lille, cité scientifique. Villeneuve d'Ascq. Durée : 1 mois. Expression et purification de la lectine FimH et de l'anticorps Mannitou.

Stage M1 au laboratoire MSAP USR 3290 CNRS. Université de Lille, cité scientifique, Villeneuve d'Ascq. Durée : 2 mois. Optimisation de techniques d'enrichissement en phosphopeptides et en glycopeptides.

Stage au laboratoire PRISM Inserm U1192. Université de Lille, cité scientifique, Villeneuve d'Ascq. Durée : 6 mois. Développement d'une méthode de détection de stéroïdes dans les crins de chevaux pour le contrôle antidopage.

### Contributions scientifiques

#### Publications

- Acceptée

**Léa Ledoux**, Yanis Zirem, Florence Renaud, Ludovic Duponchel, Michel Salzet, Nina Ogrinc, Isabelle Fournier. Comparing MS Imaging of lipids by WALDI and MALDI: two technologies for evaluating a common ground truth in MS Imaging (acceptée le 15 septembre 2023 dans le journal Analyst). <https://doi.org/10.1039/D3AN01096A>

- En soumission

Yanis Zirem\*, **Léa Ledoux\***, Lucas Roussel, Claude Alain Maurage, Pierre Tirilly, Emilie le Rhun, Bertrand Meresse, Gargey Yagnik, Mark J. Lim, Kenneth J. Rothschild, Marie Duhamel, Michel Salzet, Isabelle Fournier. Real-time glioblastoma TME analyses by SpiderMass for surgeon decision making and patient management (soumis le 13 septembre 2023 au journal Cell reports medicine).

N. de Oliveira Alves, O. Boulard, A. Vaysse, G. Dalmaso, D. Nikitina, R. Ruez, **L. Ledoux**, T. Pedron, E. Bergsten, L. Autier, S. Allam, L. Motreff, P. Sauvanet, D. Letourneur, J. Gagnière, D. Pezet, C. Godfraind, I. Fournier, M. Salzet, I. Najjar, E. Lemichez, J. Iovanna, D. Mestivier, N. Barnich, C. Malabat, M. Monot, S. Kennedy, A. Mettouchi, R. Bonnet, I. Sobhani, M. Chamaillard. Spatially-resolved metabolomic identifies Colibactin-specific principles of reprogrammed lipid

metabolism promoting right-sided colon cancer progression (soumis le 15 septembre 2023 au journal Gut)

- En préparation

**Léa Ledoux**, Yanis Zirem, Nina Ogrinc, Florence Renaud, Guillaume Piessen, Charlotte Dufour, Michel Salzet, Isabelle Fournier. Artificial intelligence model for esogastric cancer diagnosis and prognosis based on bacterioscore

## Congrès

### *Bourses*

Young Investigator Educational Grant venant de MSACL, prise en charge de l'inscription au congrès. **MSACL 2023 – Monterey, CA, USA. 3-7 avril 2023.**

Aide financière pour une conférence international de 800 euros par la SFSM. **ASMS 2023 – Houston, TX, USA. 2-8 juin 2023.**

Aide financière pour l'hébergement de 500 euros par le comité du GDR-MSI. **ASMS 2023 – Houston, TX, USA. 2-8 juin 2023.**

### *Présentations orales*

**Léa Ledoux**, Nina Ogrinc, Florence Renaud, Guillaume Piessen, Michel Salzet, Isabelle Fournier. **Rencontres du club jeune de spectrométrie de masse - en ligne, 30-31 mars 2021.** SpiderMass, un nouvel outil diagnostique pour le cancer œsogastrique.

**Léa Ledoux**, Nina Ogrinc, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. **Rencontres du club jeune de spectrométrie de masse 2022 – Lyon, France. 7-11 mars 2022.** Révélation de profils lipidiques spécifiques du cancer œsogastrique par imagerie MALDI-MS et SpiderMass.

**Léa Ledoux**, Nina Ogrinc, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. **GDR-MSI – Bordeaux, France. 20-22 juin 2022.** Révélation de profils lipidiques spécifiques du cancer œsogastrique par imagerie MALDI-MS et SpiderMass.

**Léa Ledoux**, Nina Ogrinc, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. **Analytics 2022 – Nantes, France. 5-8 septembre 2022.** Révélation de profils lipidiques spécifiques du cancer œsogastrique par imagerie MALDI-MS et SpiderMass.

**Léa Ledoux**, Michel Salzet, Isabelle Fournier. **European Biotechnology congress 2022 – Prague, République Tchèque. 5-8 octobre 2022.** Toward in vivo intraoperative mass spectrometry technology.

**Léa Ledoux**, Michel Salzet, Isabelle Fournier. **Journée de travail « Couplages à la spectrométrie de masse » - Association AFSEP, Caen, France. 6 décembre 2022.** Imagerie par spectrométrie de masse.

**Léa Ledoux**, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. **Regional Mass Spectrometry Imaging Spring Workshop – Uppsala, Suède. 27-29 mars 2023.** Decreasing Relapse in Esogastric Cancer by Improved Diagnostic with SpiderMass Technology.

**Léa Ledoux, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. MSACL 2023 – Monterey, CA, USA. 3-7 avril 2023.** Decreasing Relapse in Esogastric Cancer by Improved Diagnostic with SpiderMass Technology.

**Léa Ledoux, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. ASMS 2023 – Houston, TX, USA. 2-8 juin 2023.** Decreasing Relapse in Esogastric Cancer by Improved Diagnostic with SpiderMass Technology.

*Posters*

**Léa Ledoux, Nina Ogrinc, Florence Renaud, Guillaume Piessen, Michel Salzet, Isabelle Fournier. Structure fédérative de Recherche Technologies de Santé et Médicament SFR-TSM – Lille, France. 14 décembre 2020.** SpiderMass, un nouvel outil diagnostique pour le cancer œsogastrique.

**Léa Ledoux, Nina Ogrinc, Florence Renaud, Guillaume Piessen, Michel Salzet, Isabelle Fournier. 9th International Singapore Lipid Symposium – en ligne, 1<sup>er</sup>-5 Mars 2021.** Towards Real-Time Diagnostic of Esogastric Cancer using SpiderMass Mass Spectrometry Based Device.

**Léa Ledoux, Nina Ogrinc, Florence Renaud, Guillaume Piessen, Michel Salzet, Isabelle Fournier. Ourcon 2021 – en ligne, 11-14 octobre 2021.** Revealing esogastric cancer specificity with lipid profiles by MALDI-MSI and SpiderMass technology.

**Léa Ledoux, Nina Ogrinc, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. IMSC 2022 - Maastricht, Pays-bas. 28 août- 2 septembre 2022.** Revealing esogastric cancer specificity with lipid profiles by MALDI-MSI and SpiderMass technology.

**Léa Ledoux, Nina Ogrinc, Yanis Zirem, Florence Renaud, Michel Salzet, Isabelle Fournier. OncoLille Days – Lille, France. 2-4 novembre 2022.** Révélation de profils lipidiques spécifiques du cancer oesogastrique par imagerie MALDI-MS et SpiderMass.

## Table des matières

Remerciements.....	4
Curriculum Vitae .....	6
Formations .....	6
Expériences .....	6
Contributions scientifiques.....	6
Publications.....	6
Congrès .....	7
Bourses .....	7
Présentations orales .....	7
Posters.....	8
Table des matières.....	9
Liste des figures .....	14
Liste des tableaux.....	19
Liste des abréviations.....	20
<b>Introduction</b> .....	26
<b>Chapitre 1 : État de l’art</b> .....	32
Spectrométrie de masse.....	32
Généralités .....	32
Les sources de spectrométrie de masse en condition ambiante.....	34
Basées sur l’utilisation d’un flux de solvant.....	36
Basées sur l’utilisation d’un faisceau de gaz ou de plasma .....	39
Basées sur l’utilisation d’un faisceau laser.....	41
Sources AIMS alternatives .....	43
La spectrométrie de masse en condition ambiante utilisée pour les analyses peropératoires ....	44
Généralités .....	44
iKnife .....	46
SpiderMass .....	47
MasSpec Pen.....	49
PIRL-MS.....	51
Analyse statistique des données moléculaires obtenues .....	52
Imagerie par spectrométrie de masse .....	54
Généralités .....	54
L’imagerie MALDI.....	54
L’imagerie WALDI .....	57
Cancer œsogastrique .....	58
Anatomie.....	58

Épidémiologie .....	59
Diagnostic .....	61
Cancer de l'estomac .....	61
Cancer de l'œsophage.....	63
Classification histologique.....	64
Traitement.....	66
Cancer de l'œsophage.....	66
Cancer de l'estomac .....	67
Lipidomique .....	68
Les lipides, une superfamille .....	68
Les acides gras.....	70
Les glycérolipides.....	70
Les glycérophospholipides .....	71
Les sphingolipides .....	73
Lipides dans le cancer .....	74
Les GPLs .....	75
Les céramides .....	78
<b>Chapitre 2 : Configuration et optimisations de la technologie SpiderMass .....</b>	<b>80</b>
Configuration de la technologie SpiderMass .....	80
Fonctionnement .....	80
Différentes configurations existantes.....	81
SpiderMass couplé à un spectromètre de masse Bruker.....	81
SpiderMass couplé à un spectromètre de masse ThermoFisher .....	82
SpiderMass couplé à un spectromètre de masse Waters .....	82
Optimisations de la technologie SpiderMass .....	83
Utilisation d'une infusion contenant un Lockmass .....	83
Optimisation du connecteur T.....	84
Changement de laser.....	84
<b>Chapitre 3 : Étude comparative de l'imagerie MALDI et de l'imagerie WALDI .....</b>	<b>87</b>
Introduction.....	87
Matériel et méthodes .....	89
Échantillons .....	89
Préparation des échantillons .....	89
MALDI-MSI.....	89
WALDI-MSI.....	90
Importation et traitement des données .....	90
Similarité des pics détectés et calculs de corrélation de Pearson des spectres de masse.....	91
Modèle CNN basé sur les données MALDI-MSI pour la prédiction des données WALDI-MSI.....	92



Analyse MS/MS.....	92
Résultats et discussion.....	93
Évaluation individuelle de la similarité des pics et des corrélations d'intensité et de variances des profils lipidomiques MALDI-MSI et WALDI-MSI.....	93
Similarité de la détection des pics et corrélation des intensités et variances entre MALDI et WALDI-MSI .....	94
Effet de la taille du spot laser et de l'épaisseur du tissu sur les profils moléculaires .....	97
Analyse comparative de tissus de cancer œsogastrique .....	98
Modèle CNN basé sur les données MALDI-MSI pour la prédiction des données WALDI-MSI.....	100
Conclusion et perspectives .....	102
<b>Chapitre 4 : Analyse en temps réel du microenvironnement tumoral du glioblastome grâce au SpiderMass : un outil pour guider les décisions chirurgicales et optimiser la prise en charge des patients</b> .....	106
Introduction.....	106
Cohorte Gliomic .....	109
Coloration histologique et annotations .....	109
Préparation des échantillons .....	110
Analyse SpiderMass.....	110
WALDI-MSI.....	111
Lignées cellulaires et analyse des cellules immunitaires .....	111
MALDI-MSI.....	112
Immunohistochimie couplée au MALDI-MSI (Tag-Mass) .....	112
Analyse des données SpiderMass .....	113
Prétraitement et importation des données.....	113
Modèle de classification optimal et prédiction à l'aveugle.....	114
Explication des prédictions.....	114
Biomarqueurs significatifs.....	114
Immunoscores .....	115
Quantification MALDI-IHC .....	116
Analyse MS/MS.....	116
Résultats et discussion.....	117
Évaluation du profil lipidomique des tissus de glioblastomes .....	117
Modèle de classification optimal et processus d'IA utilisé.....	117
Prédiction automatique aveugle de la classe histologique.....	119
Marqueurs lipidiques spécifiques à chaque classe moléculaire.....	120
Stratification moléculaire non supervisée et supervisée du glioblastome .....	123
Prédiction de la survie globale des patients atteints de glioblastome.....	125
Classification des patients en fonction de leur survie grâce au SpiderMass .....	125
Identification des marqueurs distinctifs liés à la survie des patients .....	126

Évaluation immunologique du glioblastome via l'identification de la signature SpiderMass des cellules immunitaires .....	128
Analyse des profils lipidiques des cellules immunitaires par le SpiderMass .....	128
Distribution et corrélation des scores immunitaires dans les images SpiderMass en lien avec le pronostic du patient .....	129
Conclusion et perspectives .....	133
<b>Chapitre 5 : Diminution de la rechute des patients de cancer œsogastrique grâce au diagnostic par l'analyse <i>in vivo</i> SpiderMass .....</b>	<b>140</b>
Introduction .....	140
Matériel et méthodes .....	144
Échantillons .....	144
Préparation des échantillons .....	144
Coloration histologique et annotations .....	144
Analyse MALDI-MSI .....	145
Analyse SpiderMass .....	145
Analyse WALDI-MSI .....	146
Analyse des souches bactériennes .....	146
Bactérioscore .....	146
Protéomique spatiale .....	147
Technique du backside grâce à la technologie SpiderMass .....	147
Analyse LC-MS et MS/MS .....	148
Traitement et analyse des données .....	149
Données MALDI-MSI .....	149
Données SpiderMass .....	149
Données protéomiques .....	150
Indentification des lipides .....	150
Résultats .....	151
Annotation moléculaire grâce à l'analyse MALDI-MSI .....	151
Évaluation du potentiel de diagnostic du cancer œsogastrique à l'aide de la technologie SpiderMass .....	152
Modèle de classification via analyse LDA .....	152
Optimisation du modèle de classification pour le diagnostic du cancer œsogastrique .....	154
Prédiction automatisée en aveugle de la catégorie histologique en utilisant la classification préalablement élaborée .....	155
Marqueurs lipidiques liés à la tumeur ou au tissu sain .....	155
Caractérisation des sous-types de cancer œsogastrique .....	158
Identification de la signature SpiderMass de différentes souches bactériennes : bactérioscore du cancer œsogastrique .....	163
Classification SpiderMass des profils lipidiques des souches bactériennes .....	164

Distribution des scores bactériens sur les images SpiderMass et corrélation avec le type de tissu .....	167
Analyse protéomique spatiale .....	170
Identification des protéines exclusives de chaque type de tissus .....	170
Identification des voies de signalisation spécifiques de chaque type de tissu .....	171
Identification des protéines alternatives .....	175
Conclusion et perspectives .....	177
<b>Chapitre 6 : Identification des spécificités de la Colibactine dans la reprogrammation du métabolisme des lipides favorisant la progression du cancer du côlon droit .....</b>	<b>182</b>
Introduction .....	182
Résultats préalables à l'analyse SpiderMass .....	184
CCD colonisé par CoPEC : pronostic moins favorable et immunosuppression liée aux bactéries tumorales .....	184
La métabolomique spatiale révèle une augmentation des glycérophospholipides dans les tumeurs associées à la Colibactine .....	185
CoPEC conduit à l'accumulation de gouttelettes lipidiques dans les cellules cancéreuses du côlon .....	185
Résultats de l'analyse SpiderMass .....	187
Méthodologie .....	187
Résultats .....	187
Résultats post-analyse SpiderMass .....	190
La mutation APC et la lectine de type c collaborent pour maintenir un glycérophospholipide bactérien dans le microenvironnement tumoral du cancer colorectal droit .....	190
La perte d'une lectine de type c chez la souris réduit la croissance tumorale après une infection par CoPEC .....	190
Discussion et conclusion .....	192
<b>Chapitre 7 : Conclusion générale et perspectives .....</b>	<b>196</b>
<b>Références .....</b>	<b>200</b>

## Liste des figures

Figure 1 : Schéma du prototype SpiderMass, utilisé dans ce manuscrit, constitué d'un laser émettant à la longueur d'onde IR de 2.94 $\mu\text{m}$ fibré, d'un tube de transfert, d'une interface non commerciale REIMS sur un spectromètre de masse Xevo qTOF (Waters). .....	28
Figure 2 : Différentes sources APMS. a) APCI <sup>31</sup> b) ESI <sup>31</sup> . c) AP-MALDI <sup>183</sup> .....	34
Figure 3 : Schéma non exhaustif des principales sources de spectrométrie de masse en condition ambiante. Celles-ci étant classées selon les 4 groupes traditionnels (bleu foncé), différenciées selon leur mode de fonctionnement (bleu ciel) et le nom de quelques exemples de sources (bleu clair) pour chaque groupe.....	35
Figure 4 : Différentes sources AIMS basées sur l'extraction par pulvérisation d'un liquide. a) DESI <sup>37</sup> . b) SESI. c) EASI <sup>36</sup> . d) EESI <sup>37</sup> . e) DICE <sup>42</sup> . f) AFADESI <sup>36</sup> . .....	36
Figure 5: Différentes sources AIMS basées sur l'extraction directe du liquide. a) LMJ-SSP <sup>36</sup> . b) NanoDESI <sup>36</sup> . c) LESA <sup>36</sup> .....	37
Figure 6 : Différentes sources AIMS basées sur la pulvérisation du substrat. a) PSI <sup>37</sup> . b) PESI... ..	38
Figure 7 : Différentes sources AIMS basées sur l'utilisation de plasma à haute température. a) DART <sup>37</sup> . b) DAPCI c) FAPA <sup>52</sup> . d) ASAP <sup>37</sup> .....	39
Figure 8 : Différentes sources AIMS basées sur l'utilisation de plasma à basse température. a) DBDI <sup>52</sup> . b) LTP <sup>37</sup> . c) PADI <sup>57</sup> .....	40
Figure 9 : Différentes sources AIMS utilisant un laser pour l'ablation couplé à une source d'ionisation secondaire. a) LAESI <sup>57</sup> . b) MALDESI <sup>37</sup> .....	41
Figure 10 : Exemples de sources AIMS alternatives. a) VSSI <sup>69</sup> . b) MAI <sup>70</sup> .....	44
Figure 11 : Représentation graphique des sondes de spectrométrie de masse portables développées pour une utilisation peropératoire <sup>74</sup> .....	45
Figure 12 : Installation de la technologie iKnife ex-vivo et in-vivo dans une salle d'opération. Le courant électrique est appliqué au tissu et les particules chargées qui en résultent sont extraites par la pièce à main et aspirées dans l'entrée de l'interface REIMS et analysées dans le spectromètre de masse pour produire des spectres de masse spécifiques aux tissus, qui sont ensuite soumis à une analyse statistique multivariée à l'aide d'ACP-LDA. En l'espace d'une à deux secondes, le diagnostic tissulaire en temps réel s'affiche sur un écran à l'intention du chirurgien <sup>78</sup> .....	47
Figure 13 : Photo du prototype actuel SpiderMass comprenant les lasers et leurs fibres, le bras robotisé, le tube de transfert, l'interface REIMS ainsi que le spectromètre de masse. Un modèle de classification obtenu par analyse statistique multivariée à l'aide d'ACP-LDA après analyse par SpiderMass d'une cohorte de tissus de cancer œsogastrique ainsi qu'un exemple de résultat obtenu par prédiction à l'aveugle d'un tissu de cancer PCC. ....	48
Figure 14 : Schéma de la plate-forme MasSpec Pen dans une salle d'opération ainsi que les spectres de masse obtenus in vivo à partir de la marge du col du pancréas et du corps du canal cholédoque d'un patient soumis à une intervention de Whipple. Les résultats correspondants de la prédiction Lasso (effectuée après la chirurgie) et les notes pathologiques finales sont illustrés <sup>100</sup> . ....	50
Figure 15 : Application du système PIRL-MS sur des xénogreffes de médulloblastome et de cancer du sein <sup>74</sup> . ....	52
Figure 16 : Méthodologie utilisée pour l'utilisation d'une technologie AIMS pour le diagnostic du cancer in vivo. ....	52
Figure 17 : Méthodologie utilisée pour l'imagerie MALDI afin d'obtenir une carte de la répartition de chaque ion présent dans le tissu. ....	55
Figure 18 : Anatomie de l'œsophage et de l'estomac avec un zoom à droite sur les différentes régions de l'estomac ainsi que de la paroi typique du tractus gastro-intestinal. ....	58

Figure 19 : a) Nombre de cas et de décès estimés pour les dix principaux cancers dans le monde sexe et âge confondus. Le cancer de l'estomac et de l'œsophage se classent respectivement 5 <sup>ème</sup> et 8 <sup>ème</sup> . b) Répartition des cas de cancer œsogastrique par région du monde en 2020 pour les deux sexes. c) Nombre de cas du cancer œsogastrique et de décès en 2020 selon le sexe et la région du monde. d) Estimation de nouveaux cas de cancer œsogastrique dans le mode d'ici 2040 <sup>1</sup> .....	60
Figure 20 : Différents types de tissus de cancer œsogastrique. a) PCC-SRC. b) PCC avec des cellules SRC (<10 %). c) PCC-NOS. d) ADK mucineux.....	64
Figure 21 : Traitements envisagés selon le stade du cancer de l'œsophage.....	66
Figure 22 : Exemple de céramides di- et tri- hydroxylés.....	73
Figure 23 : Différentes configurations possibles de la technologie SpiderMass.....	82
Figure 24 : Comparaison des spectres obtenus lors de l'analyse d'un tissu de cancer œsogastrique avec ou sans infusion d'IPA/LeuK en mode négatif.....	83
Figure 25 : Connecteur final permettant de relier l'infusion et le tube de transfert à la source REIMS.....	84
Figure 26 : Image du nouveau laser ainsi que la comparaison des spectres obtenus pour l'analyse d'un cerveau de rat avec l'ancien laser (en <b>haut</b> ) et le nouveau laser (en <b>bas</b> ) en mode négatif. .	85
Figure 27 : Comparaison des principales étapes des méthodologies utilisées pour l'imagerie MALDI et l'imagerie WALDI.....	88
Figure 28 : Comparaison des technologies d'imagerie MALDI et WALDI.....	88
Figure 29 : Explication de la similarité des pics détectés et calculs de la corrélation de Pearson du spectre de masse A vs les spectres B, C, D, E et inversement.....	92
Figure 30 : Réplicats MALDI-MSI et WALDI-MSI du cerveau de rat dans les deux modes d'ionisation. La matrice de corrélation d'intensité de Pearson et le tableau récapitulatif de la similarité des pics détectés, calculs de corrélation et coefficient de variance de chaque réplicat pour (a) MALDI-MSI en mode positif, (b) WALDI-MSI en mode positif, (c) MALDI-MSI en mode négatif et (d) WALDI-MSI en mode négatif.....	93
Figure 31 : WALDI-MSI et MALDI-MSI (matrice norharmane) du premier réplicat de cerveau rat en mode négatif. (a) Images d'ions sélectionnés et zoom spectral à m/z 790.6, 885.6 et 890.7. (b) Résultat de l'analyse pLSA. (c) La matrice de corrélation de Pearson pour les intensités et les variances.....	94
Figure 32 : Spectres WALDI-MSI (vert) et MALDI-MSI (matrice DAN et 9-AA) (violet) d'une section de cerveau de rat en mode négatif ainsi que les matrices de corrélation de Pearson correspondantes.....	95
Figure 33 : WALDI-MSI et MALDI-MSI (matrice norharmane) d'un tissu de cerveau de rat en mode positif. (a) Images d'ions sélectionnés à m/z 788.6 et 826.7. (b) Segmentation et dendrogramme de l'ensemble des données WALDI-MSI et MALDI-MSI. (c) Spectres WALDI-MSI (vert) et MALDI-MSI (violet). La boîte grise indique un groupe de pics présents uniquement dans les données WALDI-MSI. (d) Matrice de corrélation de Pearson pour les intensités et les variances des données WALDI-MSI et MALDI-MSI.....	96
Figure 34 : Spectres WALDI-MSI (vert) et MALDI-MSI (matrice DAN et DHB) (violet) d'une section de cerveau de rat en mode positif ainsi que les matrices de corrélation de Pearson correspondante.....	97
Figure 35 : WALDI-MSI et MALDI-MSI (matrice norharmane) pour différentes résolutions spatiales et épaisseurs de tissu en mode négatif. (a) Images d'ions sélectionnés à m/z 790.6 et m/z 890.7. (b) Spectres WALDI-MSI 250 µm (vert), MALDI-MSI 50 µm (violet) et MALDI-MSI 150 µm (rouge) d'une section de cerveau de rat analysée à différentes résolutions spatiales. (c) Spectres WALDI-MSI 20 µm (vert), MALDI-MSI 12 µm (violet) et WALDI-MSI 12 µm (rouge) d'une	

section de cerveau de rat de différentes épaisseurs. (d) Tableau récapitulatif de la similarité spectrale des pics détectés ainsi que des corrélations de variances et intensités. ....	97
Figure 36 : Données d'imagerie WALDI-MSI et MALDI-MSI (matrice norharmane) en mode négatif pour (a) la section de tissu gastrique n°1 et (b) la section de tissu gastrique n°2. ....	99
Figure 37 : Modèle CNN basé sur les données MALDI-MSI et les données WALDI-MSI pour la prédiction des tissus cancéreux et sains à partir de la section n°2 du tissu gastrique. ....	101
Figure 38 : Pipeline global pour le diagnostic moléculaire par spectrométrie de masse ambiante combiné à la prédiction par intelligence artificielle. Les flèches bleues sont obligatoires pour obtenir les résultats finaux (flèches vertes) qui sont le diagnostic à l'aveugle des tissus et l'identification des biomarqueurs confiants. Une étape facultative (flèche rouge) peut être ajoutée au pipeline comme la distribution spatiale des biomarqueurs confiants grâce à l'analyse par WALDI-MSI des tissus.....	108
Figure 39 : Le pipeline global pour voir la distribution des cellules cancéreuses et immunitaires (lymphocytes et macrophages de type M1 et M2) dans un tissu GBM analysé par SpiderMass-MSI. ....	115
Figure 40 : Identification du modèle de classification optimal après analyse de la cohorte rétrospective uniquement dans les deux modes d'ionisation. Tout d'abord, les tableaux des 24 classificateurs qui ont été entraînés et évalués, indiquant leur précision, leur précision équilibrée, leur score F1 et le temps nécessaire. De plus, les rapports de classification et les matrices de confusion de RidgeClassifier avec l'ensemble d'apprentissage et après une validation croisée 20 fois.....	117
Figure 41 : Matrice de confusion et rapport de classification en mode positif et négatif après l'ajout de la cohorte prospective à la cohorte rétrospective. ....	119
Figure 42 : Résultats de l'analyse supervisée et non supervisée pour la découverte de marqueurs en mode négatif et positif contenant le top 120 m/z qui contribuent positivement (vert) ou négativement (rouge) à la classification de chaque tissu dans le modèle ainsi que le heatmap des pics sélectionnés pour trois régions différentes. Le heatmap affiche l'intensité de chaque ion, le rouge indiquant une surexpression et le vert une sous-expression. ....	121
Figure 43 : Spectres MS/MS obtenus après fragmentation des ions m/z 788.6 et m/z 794.5, annotés ensuite en tant que PS (18:0_18:1) et PE (40:4) respectivement. ....	122
Figure 44 : Dix exemples de box plots d'ions biomarqueurs fiables des tissus bénins, tumoraux et nécrotiques .....	123
Figure 45 : Sous-stratification moléculaire non supervisée et supervisée. A) Segmentation globale basée sur les données lipidiques et protéomiques. B) et C) Rapport de classification et matrice de confusion du classificateur Ridge après une validation croisée, obtenus pour le modèle de classification qui distingue les 3 groupes de segmentation (rouge, jaune et bleue). D) Boîtes à moustaches de deux biomarqueurs lipidiques obtenus après la découverte non supervisée et supervisée de marqueurs pour chaque groupe de segmentation. ....	124
Figure 46 : Six biomarqueurs potentiels, leur boîte à moustaches correspondante, leur structure chimique, qui ont été trouvés spécifiques du pronostic <10 mois et >36 mois. A) en mode négatif. B) en mode positif.....	126
Figure 47 : Classification SpiderMass des profils lipidiques des cellules immunitaires. A) Modèle LGBM entraîné pour les cellules cancéreuses, les lymphocytes et les macrophages de type M1 et M2 qui ont été analysés précédemment par la technologie SpiderMass. B) Trois exemples de biomarqueurs et leur boîte à moustaches correspondante pour les cellules cancéreuses du GBM, les lymphocytes et les macrophages de type M1 et M2. ....	128
Figure 48 : Distribution des immunoscores sur six tissus frais congelés de GBM pour les cellules cancéreuses, les lymphocytes et les macrophages de type M1 et M2. Pour chaque cellule, le RPR	



en pourcentage est indiqué. Le pourcentage moyen de lymphocytes, de macrophages de type M1 et M2 et le ratio M1/M2 est représenté dans les tissus avec un pronostic <10 mois et >36 mois. ....	130
Figure 49 : Multiplex MALDI-IHC de 3 biomarqueurs sur deux coupes de tissus frais congelés de GBM. L'échelle d'affichage (unités arbitraires d'intensité) est la suivante (seuil d'intensité minimale/seuil d'intensité maximale) : 4/25 (Ki67), 2/6 (CD8 $\alpha$ ) et 4/25 (CD68).....	131
Figure 50 : Vue d'ensemble de tous les marqueurs lipidiques, trouvés à l'aide de la découverte supervisée et non supervisée, spécifiques de la méthylation de la tumeur, de l'histologie du tissu, de la sous-stratification de la tumeur GBM et de la durée de survie du patient, en mode positif et négatif. ....	136
Figure 51 : Aperçu général de l'ensemble des méthodologies utilisées pendant cette étude sur le cancer œsogastrique. ....	143
Figure 52 : Installation au laboratoire et schéma de la technique de l'irradiation par backside grâce au SpiderMass. ....	148
Figure 53 : Exemples de comparaison entre les annotations par le pathologiste et la segmentation spatiale obtenue dans les deux modes d'ionisation sur 3 tissus mixtes de cancer œsogastrique. ....	151
Figure 54 : Modèle de classification obtenu après analyse LDA afin de distinguer les tissus cancéreux des tissus sains en mode négatif et en mode positif. ....	153
Figure 55 : Matrice de confusion et rapport de classification en mode négatif et positif pour le modèle de classification permettant de distinguer les tissus cancéreux et les tissus sains. ....	154
Figure 56 : Spectre de fragmentation après analyse par spectrométrie de masse en tandem de l'ion m/z 701.5.....	156
Figure 57 : Quatre biomarqueurs fiables et leur boîte à moustache correspondante spécifiques des tissus cancéreux et des tissus sains. A) en mode négatif. B) en mode positif.....	157
Figure 58 : Distribution des ions sélectionnés dans un tissu mixte de cancer œsogastrique, constitué d'une zone cancéreuse au milieu ainsi que de deux zones saines de chaque côté. ....	157
Figure 59 : Matrice de confusion et rapport de classification en mode négatif et positif pour le modèle de classification permettant de distinguer les sous-types de cancer PCC et ADK.....	158
Figure 60 : Modèles de classification obtenus après analyse LDA afin de distinguer les sous-types de cancer en mode négatif et en mode positif. ....	159
Figure 61 : Quatre biomarqueurs fiables et leur boîte à moustaches correspondante spécifiques des sous-types de cancer ADK et PCC. A) en mode négatif. B) en mode positif.....	160
Figure 62 : Modèle de classification obtenu après analyse LDA afin de distinguer les différentes souches bactériennes en mode négatif et en mode positif. ....	164
Figure 63 : Exemple d'un biomarqueur fiable pour chaque souche bactérienne avec la boîte à moustaches correspondante en mode négatif (A) et mode positif (B). ....	167
Figure 64 : Box plot du rapport <i>S. bovis</i> et <i>L. gasseri</i> pour les différents types de tissus. ....	168
Figure 65 : Distribution des bactérioscopes sur 3 exemples de tissus de cancer oesogastrique pour les 6 souches bactériennes. Pour chaque souche bactérienne, le RPR en pourcentage est indiqué. Le pourcentage moyen des 6 souches bactériennes est représenté dans les tissus sains et cancéreux, ADK et PCC. ....	169
Figure 66 : Diagramme de Venn mettant en avant le nombre de protéines exclusives pour chaque type de tissus, sains, cancéreux de type PCC et cancéreux de type ADK. ....	170
Figure 67 : Analyse statistique ANOVA des protéines identifiées par irradiation backside correspondant aux différents types de tissus (sains, PCC et ADK). ....	171
Figure 68 : Voies de signalisation surexprimées dans le cluster 1, représentant les tissus sains. L'épaisseur des traits montre le degré de confiance pour le rapport entre les deux protéines.....	172

Figure 69 : Voies de signalisation surexprimées dans les clusters 2 et 3, représentant les tissus cancéreux du sous-type ADK. L'épaisseur des traits montre le degré de confiance pour le rapport entre les deux protéines.....	174
Figure 70 : Analyse statistique ANOVA des protéines alternatives identifiées par irradiation backside correspondant aux différents types de tissus (sains, PCC et ADK). .....	176
Figure 71 : Le modèle de classification LDA basé sur trois groupes : les cellules HCT116 infectées par la souche CoPEC (11G5), une souche mutante qui ne produit pas de CoPEC (11G5 $\Delta$ clbQ) et les cellules non infectées (contrôle négatif). .....	188
Figure 72 : Vue d'ensemble de l'analyse pan-cancer des marqueurs lipidiques trouvés à l'aide de la découverte supervisée et non supervisée, spécifiques de l'étude sur le glioblastome et sur le cancer œsogastrique, des cellules immunitaires et des bactéries, en mode positif et négatif.....	198



## Liste des tableaux

Tableau 1 : Tableau non exhaustif des principales matrices utilisées en imagerie MALDI ainsi que leur formule chimique, leur poids moléculaires, leur polarité d'utilisation, les molécules détectables et la première référence bibliographique les mentionnant. ....	56
Tableau 2 : Liste non exhaustive des classes et sous-classes existantes de lipides avec un exemple de structure chimique pour chaque classe.....	69
Tableau 3 : Liste exhaustive des 6 groupes de glycérophospholipides avec leur composition et un exemple de structure chimique pour chaque groupe.....	72
Tableau 4 : Liste des paramètres optimisés pour le dépôt des matrices norharmane, 2,5-DHB, 9-AA et DAN, avec le pulvérisateur HTX M5. ....	90
Tableau 5 : Calculs de la similarité des pics détectés et des corrélations pour tous les tissus, matrices et modes d'ionisation.....	103
Tableau 6 : Données démographiques et cliniques de la cohorte rétrospective.....	109
Tableau 7 : Résultats des prédictions à l'aveugle par le SpiderMass des 9 tissus (2 FFPE et 7 frais congelés (FF)) non utilisés dans l'entraînement du modèle de classification.....	120
Tableau 8 : Données démographiques et cliniques de la cohorte. ....	144
Tableau 9 : Évaluation des taux de bonne classification du modèle de classification cancer vs sain en mode négatif (orange) et en mode positif (bleu). ....	153
Tableau 10 : Liste exhaustive des biomarqueurs fiables des tissus sains et des tissus cancéreux ainsi que des sous-types de cancer PCC et ADK. ....	160
Tableau 11 : Liste exhaustive des biomarqueurs fiables des différentes souches bactériennes en mode positif et en mode négatif.....	165
Tableau 12 : Liste exhaustive des 23 biomarqueurs des cellules HCT116 infectées par la souche mutante 11G5clbQ, par la souche clinique de CoPEC 11G5 et les cellules témoins négatives. .	189

## Liste des abréviations

<b>9-AA</b> : 9-aminoacridine	56
<b>ACN</b> : acétonitrile	148
<b>ACP</b> : analyse en composantes principales	53
<b>ADK</b> : adénocarcinome	64
<b>AF</b> : acide formique	148
<b>AFADESI</b> : DESI assistée par flux d'air	36
<b>AG</b> : acide gras	70
<b>AGC</b> : contrôle automatique du gain	149
<b>AIMS</b> : sources d'ionisation en condition ambiante	27
<b>APC</b> : polypose adénomateuse coli	184
<b>APCI</b> : ionisation chimique à pression atmosphérique	34
<b>AP-MALDI</b> : ionisation par désorption laser assistée par matrice à pression atmosphérique	34
<b>APMS</b> : spectrométrie de masse à pression atmosphérique	33
<b>ASAP</b> : sonde d'analyse des solides atmosphériques	40
<b>AUC</b> : aire sous la courbe	91
<b>BSA</b> : sérum albumine-bovine	113
<b>CCD</b> : cancer du côlon droit	182
<b>CCG</b> : cancer du côlon gauche	182
<b>Cer</b> : céramides	73
<b>CI</b> : ionisation chimique	32
<b>CID</b> : dissociation induite par collision	92
<b>CIN</b> : instabilité chromosomique	65
<b>CoPEC</b> : <i>Escherichia Coli</i> productrices de Colibactine	182
<b>CMS</b> : sous-type moléculaire consensuel	184
<b>CNN</b> : réseaux neuronaux convolutionnels	53
<b>DAII</b> : ionisation d'entrée assistée par gouttelettes	44
<b>DAN</b> : 1,5-diaminonaphtalène	56
<b>DART</b> : analyse directe en temps réel	39
<b>DAPCI</b> : désorption chimique à pression atmosphérique	40

<b>DBDI</b> : ionisation par décharge à barrière diélectrique	40
<b>DDA</b> : acquisition dépendante des données	148
<b>DESI</b> : désorption - ionisation par électronébulisation	36
<b>DG</b> : diglycéride	71
<b>DHB</b> : acide 2,5-dihydrobenzoïque	56
<b>DICE</b> : désorption/ionisation par échange de charges	36
<b>EASI</b> : ionisation sonore – spray facile d'accès	36
<b>EAPPI</b> : photo-ionisation extractive à pression atmosphérique	43
<b>EBV</b> : virus d'Epstein-Barr	65
<b>EC</b> : cancer œsogastrique	26
<b>EDTA</b> : acide éthylènediaminetétraacétique	111
<b>EESI</b> : ionisation extractive par électronébulisation	36
<b>EI</b> : ionisation électronique	32
<b>ESI</b> : ionisation par électronébulisation	32
<b>FAB</b> : Fast Atom Bombardment	32
<b>FAPA</b> : spectroscopie de rémanence à pression	40
<b>FDR</b> : taux de fausses découvertes	150
<b>FI/FD</b> : ionisation et désorption de champ	32
<b>FISH</b> : imagerie par hybridation <i>in situ</i>	184
<b>FFPE</b> : fixés au formol et inclus dans la paraffine	28
<b>FREGAT</b> : recherche française des tumeurs de l'œsophage et de l'estomac	89
<b>FWHM</b> : largeur totale à moitié maximum	149
<b>GBM</b> : glioblastome	106
<b>GPL</b> : glycérophospholipides	70
<b>GS</b> : génomiquement stable	65
<b>HCCA</b> : acide $\alpha$ -cyano-4-hydroxycinnamique	56
<b>HCCA/ANI</b> : HCCA-aniline	56
<b>HPS</b> : coloration hémalum, phloxine et safran	89
<b>IA</b> : intelligence artificielle	104
<b>IFN</b> : interféron	111

<b>IHC</b> : immunohistochimie	116
<b>IPA</b> : isopropanol	83
<b>IR</b> : infrarouge	33
<b>IRM</b> : imagerie par résonance magnétique	107
<b>ITO</b> : indium tin oxide	89
<b>JOG</b> : jonction œsophago-gastrique	58
<b>keV</b> : kiloélectronvolts	33
<b>LAESI</b> : ionisation par ablation laser et électronébulisation	42
<b>LASSO</b> : opérateur de sélection et de moindre rétrécissement absolu	53
<b>LA – REIMS</b> : ablation laser couplée à la spectrométrie de masse à ionisation rapide par évaporation	53
<b>LCM</b> : microdissection laser	147
<b>LC-MS</b> : chromatographie liquide couplée à la spectrométrie de masse	33
<b>LDA</b> : analyse discriminante linéaire	53
<b>LDI</b> : ionisation/désorption par laser	33
<b>LDTD-APPI</b> : désorption thermique à diode laser couplée à la photo-ionisation à pression atmosphérique	42
<b>LESA</b> : analyse de surface par extraction liquide	37
<b>LeuK</b> : leucine enképhaline	83
<b>LFQ</b> : quantification sans étiquette	150
<b>LGBM</b> : light gradient boosting machine	115
<b>LIAD-APPI</b> : désorption acoustique induite par laser couplée à la photo-ionisation à pression atmosphérique	42
<b>LIME</b> : local interpretable model-agnostic explanations	114
<b>LMJ – SSP</b> : microjonction liquide – source d'échantillonnage de surface	37
<b>LPC</b> : lysophosphatidylcholine	185
<b>LPS</b> : lipopolysaccharide	111
<b>LSI</b> : ionisation par spray laser	41
<b>LTP</b> : ionisation par plasma à basse température	40
<b>MAI</b> : ionisation assistée par matrice	44
<b>MALDESI</b> : désorption laser assistée par matrice et ionisation par électronébulisation	42

<b>MALDI</b> : ionisation par désorption laser assistée par matrice	28
<b>MCSF</b> : macrophage colony-stimulating factor	111
<b>MeOH</b> : méthanol	90
<b>MG</b> : monoglycéride	71
<b>ML</b> : machine learning	29
<b>MS</b> : spectrométrie de masse	32
<b>MS/MS</b> : spectrométrie de masse en tandem	92
<b>MSI</b> : imagerie par spectrométrie de masse	54
<b>m/z</b> : rapport masse sur charge	32
<b>NaCl</b> : chlorure de sodium	113
<b>Nd:YAG</b> : neodymium-doped yttrium aluminium garnet	47
<b>OG</b> : Octyl $\beta$ -D-glucopyranoside	113
<b>OPO</b> : oscillateur paramétrique optique	47
<b>PA</b> : acide phosphatidique	71
<b>PADI</b> : désorption/ionisation assistée par plasma	41
<b>PBMC</b> : cellules mononuclées du sang périphérique	112
<b>PBS</b> : phosphate buffered saline	111
<b>PC</b> : phosphatidylcholine	72
<b>PCC</b> : adénocarcinome à cellules peu cohésives	26
<b>PCC – NOS</b> : adénocarcinome à cellules peu cohésives – non spécifié autrement	64
<b>PCC – SRC</b> : adénocarcinome à cellules peu cohésives en bague à chaton	64
<b>PD-L</b> : ligand de la mort programmée	65
<b>PDSM</b> : désorption par plasma	32
<b>PE</b> : phosphatidyléthanolamine	72
<b>PESI</b> : ionisation par électrospray de sonde	38
<b>PET</b> : tomographie par émissions de positons	64
<b>PFA</b> : paraformaldéhyde	112
<b>PFDC</b> : Pierre Fabre Dermocosmétique	81
<b>PG</b> : phosphatidylglycérol	71
<b>PI</b> : phosphatidylinositol	71

<b>PIRL</b> : laser infrarouge picoseconde	43
<b>PLC</b> : phospholipase C	77
<b>pLSA</b> : régression par les moindres carrés partiels	91
<b>PS</b> : phosphatidylsérine	72
<b>PSI</b> : ionisation par pulvérisation de papier	38
<b>PSM</b> : poste de sécurité microbiologique	111
<b>PTFE</b> : polytétrafluoroéthylène	49
<b>PTM</b> : modification post-traductionnelle	180
<b>REIMS</b> : spectrométrie de masse à ionisation rapide par évaporation	43
<b>RF</b> : random forest	53
<b>ROC</b> : receiver operating characteristic	91
<b>ROI</b> : région d'intérêt	150
<b>RMS</b> : root mean square	91
<b>RPMI</b> : Roswell Park Memorial Institute medium	111
<b>RPR</b> : ratio de présence relative	129
<b>SA</b> : acide sinapinique	56
<b>SAII</b> : ionisation d'entrée assistée par solvant	44
<b>SESI</b> : ionisation secondaire par électronébulisation	36
<b>SG</b> : survie globale	107
<b>SMA</b> : subminiature	48
<b>S/N</b> : rapport signal sur bruit	156
<b>TBS</b> : tampon tris salin	113
<b>TFA</b> : acide trifluoroacétique	90
<b>TIC</b> : nombre total d'ions	112
<b>TG</b> : triglycéride	71
<b>TP53</b> : protéine tumorale 53	66
<b>UV</b> : ultraviolet	33
<b>VSSI</b> : ionisation ambiante par pulvérisation vibrante à arêtes vives	43
<b>WALDI</b> : ionisation par désorption laser assistée par l'eau	28



## Introduction

---

Le cancer est la seconde cause de mortalité dans le monde avec presque 19.3 millions de nouveaux cas et 10 millions de décès<sup>1</sup> en 2020. Le cancer œsogastrique (EC) est constitué du cancer de l'estomac et de l'œsophage. Le cancer de l'estomac est le 5<sup>ème</sup> plus fréquent, touchant plus d'1 million de personnes par an dans le monde. Son taux de mortalité est très élevé, le mettant au 4<sup>ème</sup> rang des causes de mortalité par cancer<sup>1</sup>. En France, on estimait à 5013 le nombre de décès pour 7140 nouveaux cas de cancer gastrique en 2020, dont 65 % chez l'homme<sup>2</sup>. Quant au cancer de l'œsophage, le nombre de décès est estimé à 4274 pour 5313 nouveaux cas de cancer en 2020 dont 76 % chez l'homme<sup>2</sup>. Le pronostic des cancers œsogastriques n'est pas très élevé, il est inférieur à 15 %<sup>3</sup>.

L'adénocarcinome gastrique est la forme majoritaire (90-95 % des cancers œsogastriques). Plusieurs sous types ont été décrits dans la classification OMS en 2010. Parmi ceux-ci, l'adénocarcinome à cellules peu cohésives (PCC) correspond à des formes particulièrement agressives et apparaît chez des patients plus jeunes<sup>4</sup>.

La chirurgie reste le traitement habituel des cancers œsogastriques. Elle consiste en une gastrectomie totale avec curage ganglionnaire, dont la mortalité et la morbidité post-opératoire peuvent atteindre 3 % et 15 % respectivement<sup>5</sup>. Des normes existent pour maximiser le taux de résections complètes, unique garant d'une survie à long terme. Elles imposent d'optimiser les marges de la tumeur afin de diminuer le risque d'infiltration microscopique des marges<sup>6</sup>. Le respect de ces règles contraint parfois à réséquer des organes à proximité comme l'œsophage ou la rate, alourdissant alors le risque opératoire. Ces marges sont plus étendues en cas de tumeurs PCC (8 cm vs 5 cm), dont on connaît le potentiel infiltrant<sup>4</sup>.

Le diagnostic préopératoire d'un tissu de type PCC dans la tumeur est donc primordial alors même que l'hétérogénéité tumorale importante des adénocarcinomes gastriques peut gêner ce diagnostic. Le diagnostic peropératoire des PCC est quant à lui limité par la qualité et la durée de l'analyse extemporanée. Celle-ci est effectuée par l'anatomopathologiste en condition peropératoire sur les tissus excisés et souffre d'une faible fiabilité : le taux de faux négatif pouvant atteindre 30 % et correspond notamment à des cas de PCC<sup>7</sup>. L'analyse extemporanée, dont le résultat doit être rendu en moins d'une heure, se fait sur un tissu transporté rapidement au laboratoire puis congelé, coupé et coloré avec des techniques sources d'erreurs dont la qualité est assez basse. Le résultat est systématiquement rendu avec des réserves liées à la technique et il est souvent confirmé (ou infirmé) par une analyse définitive, plusieurs jours voire semaine après l'intervention. Ce dispositif



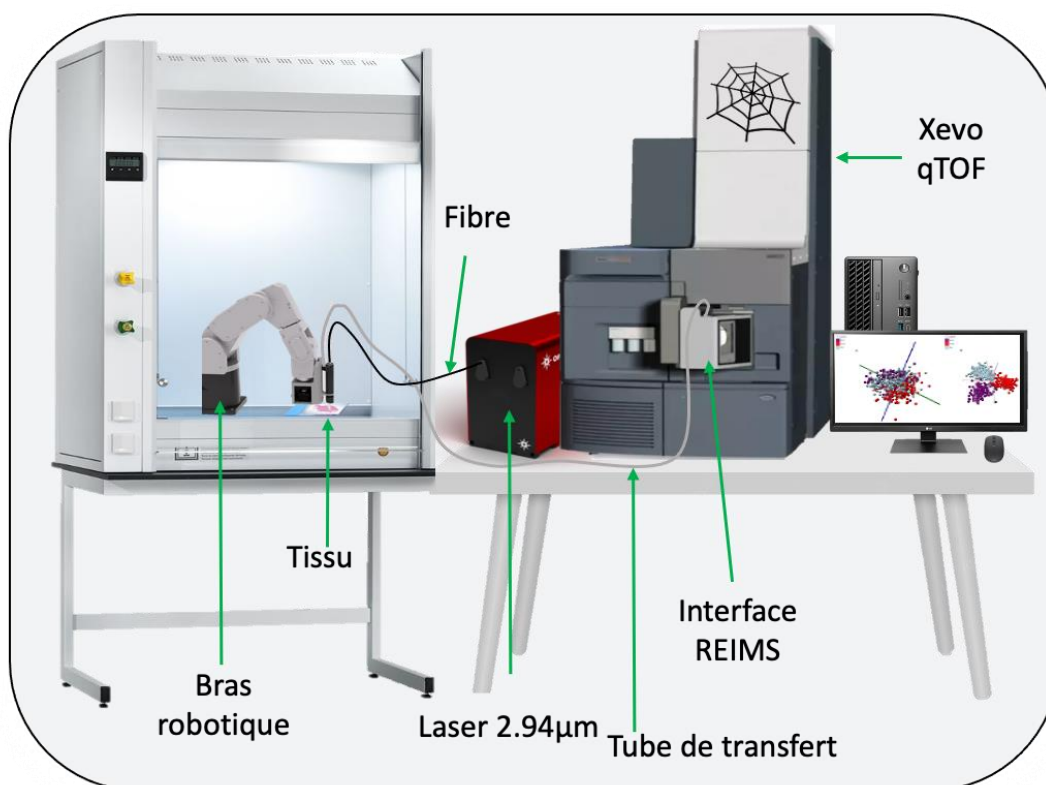
de diagnostic lourd, coûteux, peu fiable, chronophage et subjectif est néanmoins nécessaire en l'absence d'alternative.

Le développement d'un nouvel outil est urgent pour assurer une prise en charge ajustée à chaque patient. De plus, les tissus de type PCC, au pronostic plus sombre, nécessitant une réponse thérapeutique plus agressive, rendent leur diagnostic instantané fondamental alors que celui-ci est encore plus long et difficile. Il y a donc un besoin urgent d'un nouvel outil de diagnostic qui pourrait fonctionner dans la salle d'opération, fournissant une classification *in situ* en temps réel des différentes régions tissulaires sur la base de leur contenu moléculaire, sans nécessiter de marquage ou d'approches ciblées. Un tel outil permettrait d'effectuer un diagnostic ou une délimitation des marges, en peropératoire *ex vivo* dans le bloc pour une analyse rapide ou directement *in vivo* chez le patient et en temps réel. Cependant, ce nouvel outil doit obligatoirement être compatible avec les contraintes du bloc opératoire, c'est-à-dire collecter les informations moléculaires non ciblées en temps réel et de façon *in vivo*.

La stratégie de prise en charge des cancers œsogastriques manque d'un outil de diagnostic instantané, fiable et peu invasif pouvant guider le chirurgien en précisant le caractère tumoral ou non d'un tissu. Beaucoup d'enjeux dans la prise en charge de ce cancer sont actuellement l'objet de nombreuses recherches translationnelles et cliniques. En effet, nombreux essayent de caractériser moléculairement les tumeurs très hétérogènes qui constituent ce cancer, basés soit sur l'altération<sup>8</sup> ou sur l'expression<sup>9</sup> de certains gènes pouvant expliquer les mécanismes de cancérogenèse et permettre une prise en charge personnalisée des patients. Depuis moins de 20 ans, de nombreuses techniques non ciblées omiques (lipidomes, protéomes, transcriptomes, etc.) permettent d'identifier de nouveaux marqueurs tumoraux afin d'obtenir un diagnostic plus rapide et plus précis des cancers.

Par sa capacité à fournir des analyses rapides, robustes et sensibles à partir de mélanges complexes, la spectrométrie de masse est un outil d'intérêt pour ces applications. La spectrométrie de masse conventionnelle ne permet pas la réalisation d'analyses *in vivo* chez les patients. Cependant, le développement des sources d'ionisation ambiantes (AIMS) a contribué aux développements de technologies compatibles avec les analyses chez l'homme, ouvrant ainsi le champ de la médecine de précision. Aujourd'hui, plusieurs technologies basées sur la spectrométrie de masse ont été développées pour répondre au besoin d'un diagnostic en bloc opératoire *in vivo* et *ex vivo*<sup>10</sup>. C'est le cas notamment de l'iKnife, du SpiderMass, du MasSpec Pen et du PIRL-MS.

Parmi ces différentes techniques, le laboratoire PRISM Inserm U1192 développe depuis + de 10 ans un nouvel outil innovant appelé le SpiderMass (**Figure 1**).



**Figure 1** : Schéma du prototype SpiderMass, utilisé dans ce manuscrit, constitué d'un laser émettant à la longueur d'onde IR de 2.94  $\mu\text{m}$  fibré, d'un tube de transfert, d'une interface non commerciale REIMS sur un spectromètre de masse Xevo qTOF (Waters).

La technologie SpiderMass est composée d'une microsonde d'échantillonnage constituée d'un laser fibré émettant dans la gamme spectrale de l'infrarouge moyen, d'une ligne de transfert et d'un spectromètre de masse<sup>11</sup>. La microsonde laser a une longueur d'onde de 2.94  $\mu\text{m}$ , qui excite la bande d'absorption la plus intense des molécules d'eau (bande de stretching O-H) présentes dans les tissus. L'eau peut agir comme une matrice endogène dans un phénomène d'ionisation par désorption laser assistée par l'eau (WALDI) similaire au phénomène d'ionisation par désorption laser assistée par matrice (MALDI) conventionnelle<sup>11,12</sup>. Le matériel désorbé et ionisé est ensuite aspiré par un tube de transfert qui transporte l'aérosol vers le spectromètre de masse. Les profils moléculaires servent d'empreintes digitales et reflètent les phénotypes cellulaires présents dans les tissus.

Ce système SpiderMass a déjà été appliqué dans différents domaines, par exemple, l'analyse du cancer de l'ovaire *ex vivo* et de la peau humaine *in vivo*<sup>11</sup> ainsi que des cellules cultivées sans préparation<sup>13</sup>. Le système a également révélé un taux de classification correcte de 97 % sur le typage et le classement des tumeurs pour l'analyse d'échantillons *ex vivo* de sarcomes canins<sup>14</sup> et a été suivi d'une analyse *in vivo* à la clinique Oncovet. Plus récemment, elle a démontré son applicabilité pour l'analyse d'échantillons fixés au formol et inclus dans la paraffine (FFPE) non traités<sup>15</sup>, de carcinome oral<sup>16</sup> et même *in vivo* sur des feuilles de cannabis<sup>17</sup>. Cette technique a aussi

récemment été associée à un bras robotisé avec une rotation sur 6 axes. Cette nouvelle technique WALDI-MSI fournit des images topographiques 2D et 3D<sup>18</sup> et des informations moléculaires corrélées.

L'objectif principal de ma thèse est de développer la technologie SpiderMass en tant que méthode de diagnostic des tumeurs œsogastriques en temps réel. Cela permettrait de délimiter les marges de résection chirurgicale chez un patient, permettant ainsi d'aider grandement la chirurgie oncologique. Cependant, j'ai tout d'abord réalisé différentes études et développements pour améliorer le traitement des données et les modèles de classification avant d'appliquer le SpiderMass à la problématique du cancer œsogastrique. En particulier, je me suis intéressée à la comparaison des spectres lipidiques issus du MALDI et du SpiderMass afin de déterminer s'il serait possible d'augmenter les données SpiderMass des données MALDI pour l'apprentissage. Par ailleurs, j'ai étudié l'utilisation de nouveaux algorithmes de machine learning (ML) pour la classification des données issues du SpiderMass ainsi qu'à la réalisation d'immunescores et de bactérioscopes pour l'amélioration du diagnostic et du pronostic des patients. Ainsi, le présent manuscrit s'articulera en six chapitres.

Le **chapitre 1**, en tant que chapitre bibliographique, permettra de procéder à un état de l'art de la spectrométrie de masse et des différentes sources existantes, de son utilisation en imagerie, du cancer œsogastrique et de la lipodominance.

Le **chapitre 2** sera dédié à une explication approfondie de la technologie utilisée tout au long de cette thèse, à savoir le SpiderMass, ainsi qu'aux optimisations réalisées en amont des travaux détaillés ultérieurement.

Le **chapitre 3** traitera de l'analyse comparative de deux technologies d'imagerie par spectrométrie de masse, MALDI et WALDI, afin de pouvoir utiliser ces dernières à l'unisson pour une meilleure classification des tissus sains, cancéreux ou mixtes.

Le **chapitre 4** abordera la problématique de l'analyse des données après l'obtention des résultats SpiderMass en mettant en avant une nouvelle méthodologie basée sur du ML développée sur une cohorte de tissus glioblastomes.

Dans le **chapitre 5**, seront présentés les résultats, obtenus après application des optimisations des chapitres précédents, sur la problématique de départ du cancer œsogastrique.

Un projet annexe, en collaboration avec le Dr. Chamaillard Matthias, OncoLille, sera présenté dans le **chapitre 6**. Celui-ci démontre l'utilisation de la technologie SpiderMass pour détecter des lipides spécifiques du changement métabolique dû à la présence de la colibactine favorisant le cancer du côlon droit.

Enfin, le **chapitre 7** permettra de conclure sur les optimisations effectuées tout au long de cette thèse afin de potentiellement permettre l'implantation du SpiderMass dans un centre hospitalier.



# Chapitre 1 : État de l'art

---

## Spectrométrie de masse

### Généralités

La spectrométrie de masse (MS) est une technique analytique permettant d'identifier différents composés allant d'atomes aux composés de poids moléculaire élevé comme les protéines, à partir d'échantillons d'une complexité variée (mélange simple ou complexe). Le principe de cette technique est de séparer les ions en phase gazeuse en fonction de leur rapport masse sur charge ( $m/z$ ) dans l'analyseur de masse<sup>19</sup>. Pour cela, des champs électriques ou magnétiques sont utilisés. Le résultat est un spectre de masse qui reporte le courant, après la conversion ion-électron, en fonction du rapport  $m/z$ . Ainsi, les spectromètres de masse se composent de trois parties principales : la source d'ionisation, l'analyseur de masse et le détecteur. Les ions en phase gazeuse, chargés positivement ou négativement, sont créés dans la source d'ionisation, ils sont ensuite dirigés et séparés dans l'analyseur de masse et finalement détectés par conversion du courant ionique en un signal électrique pouvant être traité de manière analogique ou numérique. La MS offre de nombreux avantages, tels que sa sensibilité, sa robustesse, sa reproductibilité et sa rapidité, et ce grâce aux améliorations constantes apportées à chacune des parties de l'instrument au cours des 50 dernières années.

Diverses avancées ont été nécessaires pour augmenter la diversité des échantillons analysables par MS. En effet, la création des ions en phase gazeuse via la source d'ionisation est une étape très importante de la mesure. Par exemple, les sources d'impact électronique (EI)<sup>20</sup> ou d'ionisation chimique (CI)<sup>21</sup> sont limitées à l'analyse de molécules de faible poids moléculaire et de polarité limitée. Quant à l'ionisation et la désorption de champ (FI/FD)<sup>22</sup>, elle ne convient pas aux échantillons thermiquement instables et ceux qui ne sont pas volatils. Des méthodes telles que la désorption par plasma (PDSM)<sup>23</sup>, le Fast Atom Bombardment (FAB)<sup>24</sup> ou l'ionisation par thermospray<sup>25</sup> ont contribué à surmonter ces limites, elles nécessitaient cependant une concentration importante de protéines et présentaient des limitations instrumentales (utilisation de fragments de fission du <sup>252</sup>Cf pour la PDMS, par exemple). Dans les années 1980, deux sources ont permis de révolutionner la spectrométrie de masse pour l'analyse d'échantillons d'intérêt biologique, la source MALDI et la source d'ionisation par électrobulbion (ESI). Celles-ci permettent d'analyser tout type de composés quel que soit son poids moléculaire ou sa polarité.

Le développement du MALDI par Karas et Hillenkamp<sup>26</sup> a résolu le problème de l'analyse de diverses molécules allant de petites molécules peu polaires aux grandes molécules polaires

comme les protéines par utilisation d'une matrice organique en ionisation/désorption laser (LDI). Simultanément, Koichi Tanaka (Prix Nobel de chimie 2002) introduit une méthode de désorption-ionisation douce pour l'analyse par spectrométrie de masse des macromolécules biologiques utilisant des matrices inorganiques<sup>27</sup>. La source MALDI utilise historiquement un laser ultraviolet (UV) pulsé et seulement quelques groupes de chercheurs ont démontré son utilisation grâce à un laser infrarouge (IR-MALDI)<sup>28,29</sup>. En MALDI, une matrice organique est introduite en grande quantité par rapport à l'échantillon. La matrice est choisie pour absorber à la longueur d'onde du laser et permet à la fois un transfert d'énergie efficace au système, tout en protégeant l'échantillon de recevoir une quantité d'énergie trop importante, contribuant à limiter la fragmentation des analytes. La matrice absorbe l'énergie du laser, permettant une éjection de matière (désorption) via un mécanisme d'explosion de phase. Elle contribue également à l'ionisation des analytes même si les mécanismes d'ionisation MALDI restent mal connus. Cette source d'ionisation douce permet la détection de tous types de famille de molécules ayant des propriétés physicochimiques variées. Elle facilite aussi l'analyse des échantillons solides, comme les surfaces biologiques. Cependant, l'utilisation d'un vide poussé et d'une matrice organique ne permet pas d'analyser directement les échantillons en condition native ou les échantillons liquides sans passage par un état solide.

La source ESI<sup>30</sup>, quant à elle développée par l'équipe de Fenn permet une création d'ions en phase gazeuse à pression atmosphérique. Cette source est basée sur le principe d'électronébulisation, consistant en la dispersion à pression atmosphérique d'un liquide sous forme de gouttelettes fortement chargées électriquement contenant des molécules d'analytes. Le liquide passe à travers un champ électrique de quelques kiloélectronvolts (keV), et le spray ainsi formé est évaporé durant le transfert vers le spectromètre de masse puis à l'intérieur de celui-ci. L'évaporation des gouttelettes provoque une explosion de ces dernières et la libération d'ions d'analytes multichargés qui sont transférés à l'analyseur de masse. L'avantage de l'ESI et qu'elle peut être couplée en amont à des techniques séparatives, comme la chromatographie liquide, ce qui permet d'analyser des mélanges d'échantillons très complexes par couplage chromatographie liquide couplée à la spectrométrie de masse (LC-MS).

L'ESI fait partie de la famille des sources à pression atmosphérique (APMS). En MS, il existe des sources fonctionnant sous vide poussé et d'autres à pression atmosphérique. Cependant, il existe également des sources fonctionnant en conditions ambiantes. Il est important de distinguer la MS en condition ambiante de l'ionisation à pression atmosphérique. En effet, la MS en condition ambiante nécessite peu ou pas de préparation d'échantillon et l'ionisation est effectuée en conditions ambiantes. Tandis que l'ionisation à pression atmosphérique peut nécessiter une préparation d'échantillon telle que la dissolution d'un échantillon solide pour des analyses par ESI

ou CI à pression atmosphérique (APCI)<sup>31</sup> (**Figure 2**). Celle-ci permettant surtout d'analyser des composés présents en faible quantité dans des mélanges complexes sans préparation d'échantillon.

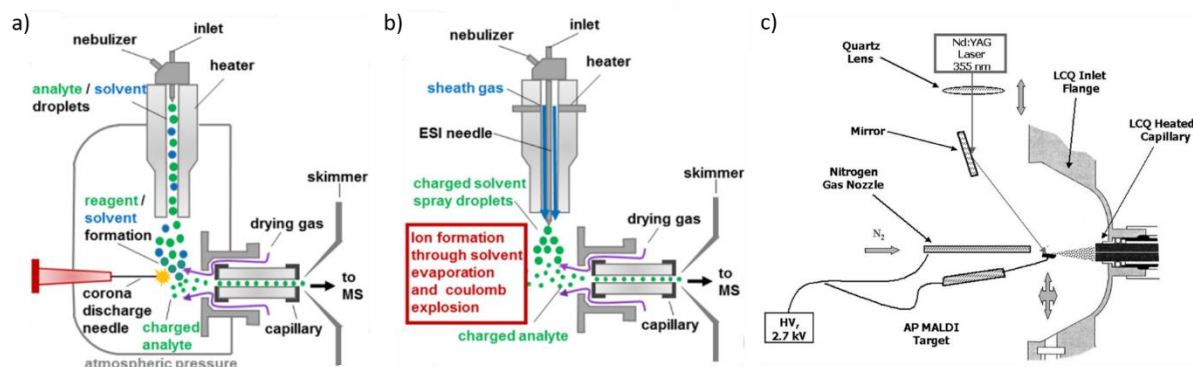


Figure 2 : Différentes sources APMS. a) APCI<sup>31</sup> b) ESI<sup>31</sup>. c) AP-MALDI<sup>183</sup>.

Une source à pression atmosphérique de la source MALDI existe (AP-MALDI). Les sources AP-MALDI ont été développées en 2000 et fonctionnent de manière similaire à la source MALDI sous vide poussé. La configuration des sources AP-MALDI est souvent orthogonale et la désorption laser se fait à un angle de 90 degrés par rapport à l'échantillon (**Figure 2**). Un flux d'azote, perpendiculaire au faisceau laser permet de favoriser le transfert du nuage d'ablation vers le spectromètre de masse. Bien que le processus d'ionisation soit plus doux que celui de la source MALDI, la formation de clusters est plus prononcée et nécessite l'ajout d'une matrice organique.

### Les sources de spectrométrie de masse en condition ambiante

L'analyse MS conventionnelle se fait souvent à basse pression, nécessitant l'introduction de l'échantillon dans un espace confiné sous vide poussé ce qui peut endommager l'échantillon ou l'instrument ainsi qu'une préparation de l'échantillon qui le modifie en grande partie. Depuis les années 2000, de nombreuses améliorations expérimentales ont permis de développer de nouvelles sources d'ionisation en conditions ambiantes. Ces développements visent à résoudre les inconvénients des sources à basse pression, notamment la nécessité de préparer les échantillons avant l'analyse<sup>32</sup>.

L'introduction des techniques AIMS a marqué le début d'une nouvelle ère dans ce domaine au début des années 2000. Depuis leur création, de nombreux chercheurs et laboratoires ont contribué à la mise au point de méthodes d'échantillonnage et d'ionisation dans des conditions ambiantes, ce qui a considérablement simplifié la complexité expérimentale et réduit le temps nécessaire pour les analyses de spectrométrie de masse. Ces 20 dernières années, les innovations dans le domaine de la spectrométrie de masse à ionisation ambiante ont connu une expansion



considérable, allant bien au-delà de leur concept initial et intégrant ces technologies de manière créative et stimulante dans la communauté scientifique.

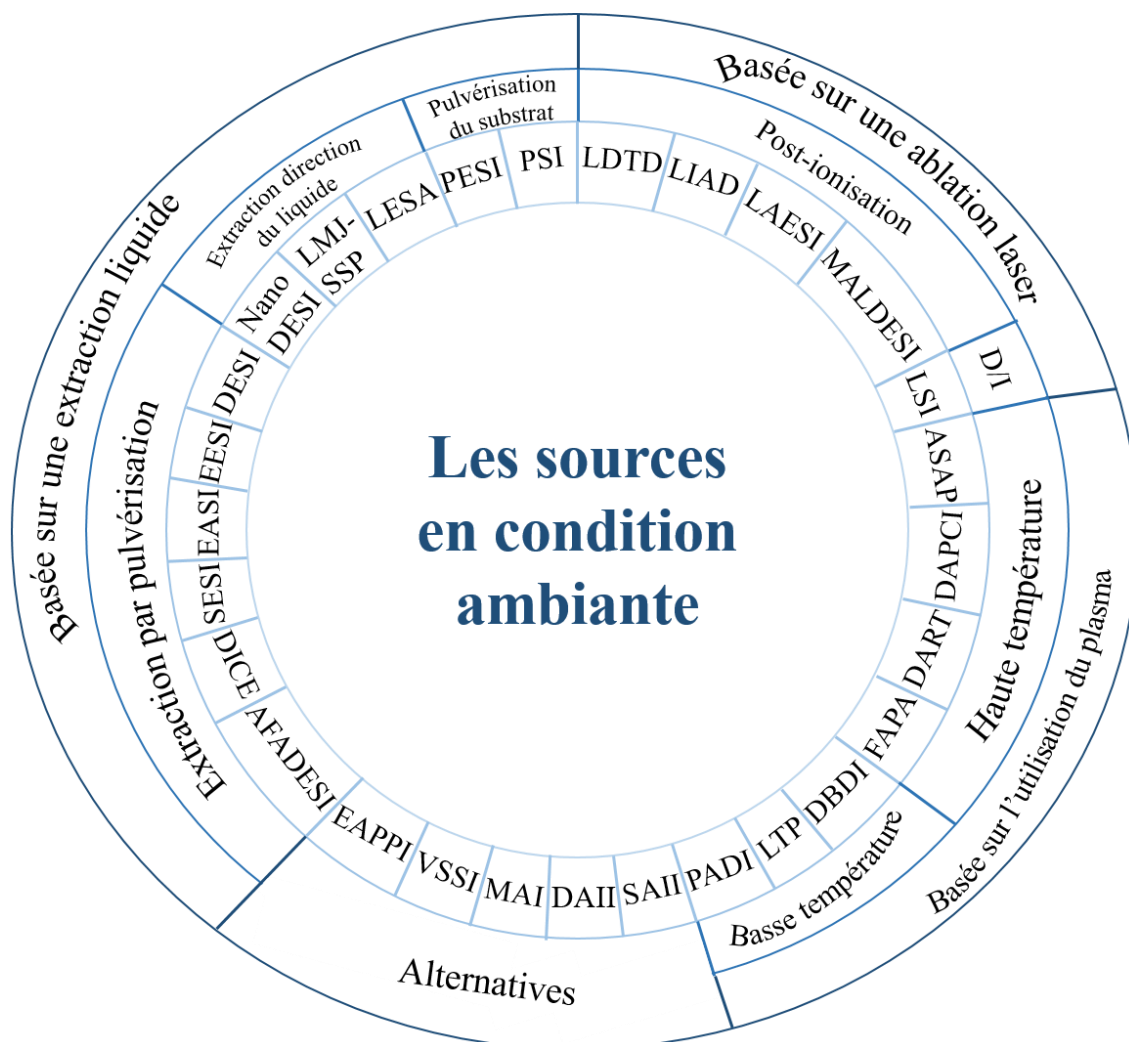
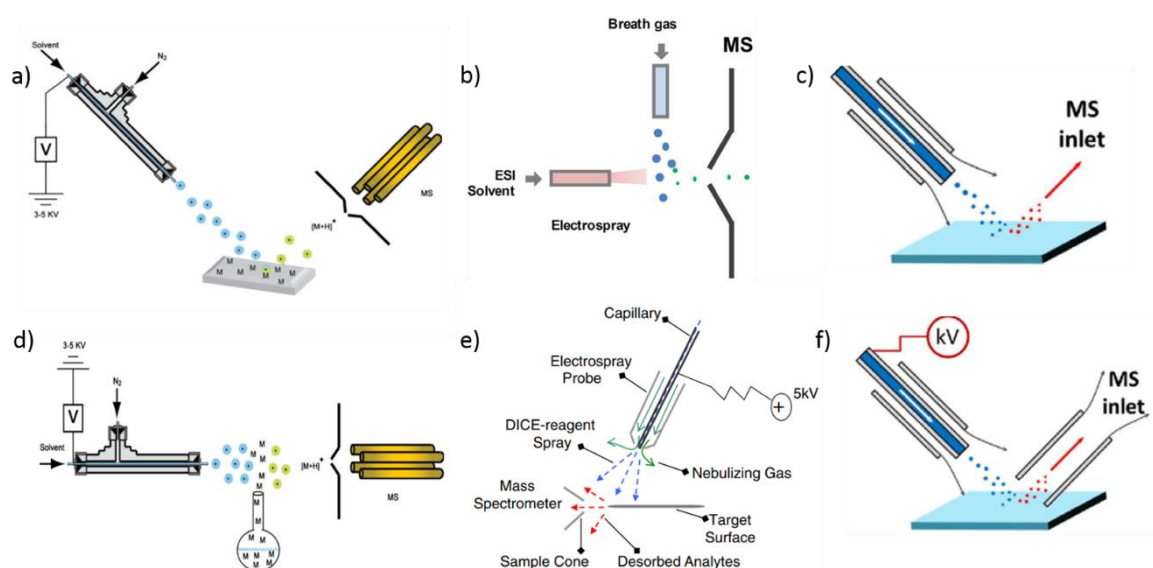


Figure 3 : Schéma non exhaustif des principales sources de spectrométrie de masse en condition ambiante. Celles-ci étant classées selon les 4 groupes traditionnels (bleu foncé), différenciées selon leur mode de fonctionnement (bleu ciel) et le nom de quelques exemples de sources (bleu clair) pour chaque groupe.

Les sources de spectrométrie de masse en condition ambiante sont capables d'analyser des échantillons de tissus biologiques vivants (analyse *ex vivo*) dans des conditions naturelles, ainsi que des organismes vivants tels que des bactéries, de petits mammifères et même des êtres humains (analyse *in vivo*). Elles sont très nombreuses et peuvent être catégorisées selon leur processus de désorption<sup>33</sup> (Figure 3). Le premier type implique l'utilisation d'un flux de solvant pour désorber les composés, tandis que le deuxième utilise un faisceau de gaz ou de plasma. La troisième catégorie consiste en une désorption induite par l'interaction avec un faisceau laser<sup>34</sup>.

### Basées sur l'utilisation d'un flux de solvant

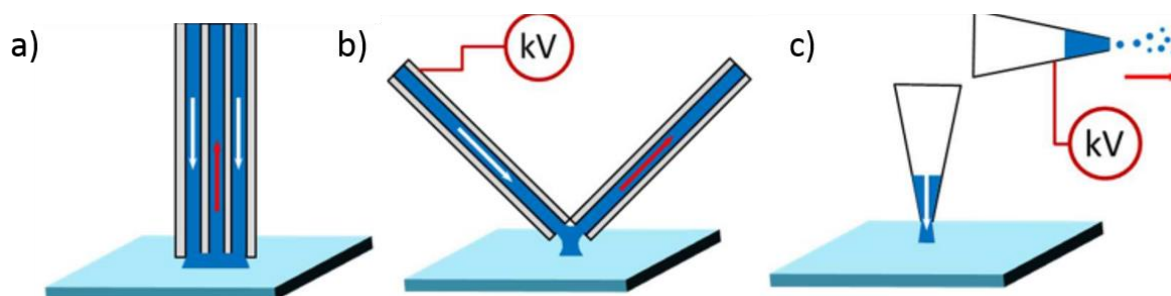
Les méthodes d'extraction liquide impliquent l'utilisation d'un solvant pour extraire ou désorber des molécules de la surface d'un échantillon. La majorité de ces techniques utilisent l'ionisation par électronébulisation ou des mécanismes similaires pour générer des ions. Ainsi, les techniques AIMS par extraction liquide sont principalement utilisées pour analyser des molécules polaires qui sont plus susceptibles d'être ionisées par des mécanismes basés sur l'ESI. Ces techniques peuvent être regroupées en trois catégories distinctes : l'extraction par pulvérisation, l'extraction directe de liquide et la pulvérisation de substrat<sup>35</sup>.



**Figure 4** : Différentes sources AIMS basées sur l'extraction par pulvérisation d'un liquide. a) DESI<sup>37</sup>. b) SESI. c) EASI<sup>36</sup>. d) EESI<sup>37</sup>. e) DICE<sup>42</sup>. f) AFADESI<sup>36</sup>.

Les techniques basées sur la pulvérisation impliquent l'utilisation d'un spray de solvant pour désorber les molécules de la surface de l'échantillon, qui sont ensuite analysées par le spectromètre de masse. Parmi ces techniques, on trouve l'ionisation par désorption/ionisation par électronébulisation (DESI), l'ionisation sonique – spray facile d'accès (EASI), l'ionisation extractive par électronébulisation (EESI), l'ionisation secondaire par électronébulisation (SESI), la désorption/ionisation par échange de charges (DICE) et du DESI assistée par flux d'air (AFADESI) (**Figure 4**). La DESI<sup>36,37</sup> et l'EASI sont similaires car elles utilisent toutes les deux un spray de solvant dirigé vers l'échantillon, en phase condensée, pour capturer les analytes, mais elles diffèrent par l'utilisation d'une haute tension appliquée pour le DESI entre le capillaire et la surface qui conduit à la création de gouttelettes de solvants chargées pour extraire et ioniser les analytes de la surface de l'échantillon, tandis que l'EASI n'utilise pas de tension et que le solvant est sprayé au travers d'un capillaire de silice fondue par un flux supersonique de gaz nébulisant coaxial au capillaire<sup>38</sup>. La simulation par dynamique des fluides computationnelle du processus DESI a montré que le spray initial forme un film fin de solvant à la surface de l'échantillon, à partir duquel

les analytes sont extraits et capturés par les gouttes suivantes qui rebondissent sur la surface pour repartir vers l'entrée du spectromètre où elles sont dirigées par un capillaire de transfert désigné à cet effet (initialement appelé « magic wand »). Une fois les gouttelettes entrées dans le spectromètre, elles subissent un processus de désolvatation similaire à l'ESI. A l'inverse du DESI et de l'EASI, l'EESI est majoritairement utilisée pour extraire des molécules d'échantillons liquides en utilisant deux panaches d'aérosols en collision, l'un contenant la solution de l'échantillon et l'autre le solvant extractif<sup>39</sup>. L'interaction entre ces panaches d'aérosols permet l'extraction et l'ionisation des analytes de l'échantillon, qui sont ensuite dirigés vers le spectromètre de masse pour être analysés. Cette technologie possédant une tolérance plus importante aux sels que la DESI. Le SESI est similaire à l'EESI, mais l'analyte est en phase gazeuse avant cette interaction<sup>40</sup>. Comme le DESI, la DICE débute par une pulvérisation primaire chargée. Au lieu d'utiliser des molécules de solvant protonées, les porteurs de charges sont des ions moléculaires de toluène à électrons impairs, formés par oxydation électrochimique dans le capillaire de pulvérisation<sup>41</sup>. Après l'extraction et la désorption, les ions d'analyte se forment par un processus de transfert de charge impliquant les ions moléculaires de toluène. Une méthode appelée AFADESI a été utilisée dans plusieurs études récentes pour faciliter l'extraction des analytes et le transport des gouttelettes contenant des analytes désorbés sur de plus longues distances en incorporant un flux d'air à haut débit<sup>42</sup>.

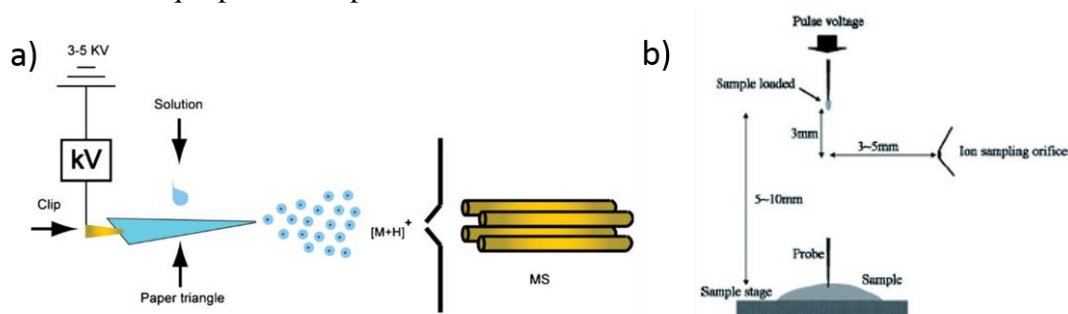


**Figure 5:** Différentes sources AIMS basées sur l'extraction directe du liquide. a) LMJ-SSP<sup>36</sup>. b) NanoDESI<sup>36</sup>. c) LESA<sup>36</sup>.

Les méthodes d'extraction liquide directe sont utilisées pour l'analyse des échantillons en permettant au solvant de réagir directement avec l'échantillon afin d'extraire les analytes qui sont ensuite transférés et ionisés pour l'analyse par MS. Différentes techniques existent, notamment la méthode de micro-jonction liquide - source d'échantillonnage de surface (LMJ-SSP), le nano-DESI et l'analyse de surface par extraction liquide (LESA) (**Figure 5**). La méthode LMJ-SSP utilise une sonde robotisée pour extraire les molécules de la surface de l'échantillon en générant un flux continu de solvant qui interagit avec la surface de l'échantillon<sup>43</sup>. La méthode nanoDESI utilise elle deux petits capillaires pour former une micro-jonction liquide sur la surface de l'échantillon, le premier étant soumis à une haute tension, et un gaz de nébulisation pour produire des gouttelettes chargées, permettant ainsi l'extraction localisée de matériaux liquides<sup>44</sup>. Tandis que le LESA,

introduit par van Berkel *et al*<sup>45</sup>, utilise une gouttelette de solvant fournie par une pointe de nano-ESI pour extraire les molécules de la surface de l'échantillon, pointe qui est ensuite basculée à 90° pour permettre la formation du spray de gouttelettes chargées comme en nano-ESI conventionnel<sup>46</sup>. Les techniques d'extraction liquide directe offrent souvent une plus grande sensibilité que les techniques basées sur la surface de l'échantillon pendant une période de temps plus longue et l'introduction complète de l'analyte dans le spectromètre de masse<sup>47</sup>.

D'autres techniques AIMS dérivés de l'ESI existent. C'est notamment le cas des techniques où le spray est réalisé directement à partir du substrat sur lequel se trouve l'échantillon (**Figure 6**). La technique la plus couramment utilisée est l'ionisation dite « paper spray » (PSI), très populaire en raison de son faible coût et de sa simplicité d'utilisation<sup>48</sup>. En PSI, l'échantillon liquide est déposé sur un papier filtre découpé en forme de triangle et laissé sécher. Le triangle de papier est ensuite utilisé comme tip d'ESI en appliquant un solvant et une tension qui permettent de réaliser le spray de gouttelettes chargées à partir de la pointe de papier. Cette méthode permet souvent de séparer les espèces interférentes des matrices complexes pour améliorer la sensibilité et la reproductibilité de l'analyse. La principale différence entre le PSI et les autres techniques d'AIMS basées sur l'électronébulisation est l'utilisation du papier comme filtre pour retenir certains composants de matrices biologiques complexes. Cependant, cette méthode a des difficultés pour l'analyse de composés hydrophiles et en mode ion négatif. C'est pour cela que des modifications physiques du substrat sont souvent utilisées pour améliorer la sensibilité pour les analytes ciblés. L'ionisation par électrospray de sonde (PESI) est une alternative à la PSI. Elle a été introduite pour la première fois par Hiraoka *et al.* au milieu des années 2000<sup>49</sup>. Celle-ci utilise une aiguille solide arrondie qui est en contact avec la surface de l'échantillon sous forme solide ou liquide, ce qui entraîne le transfert des analytes vers l'extrémité de l'aiguille. Ensuite, l'aiguille est positionnée devant l'entrée du spectromètre de masse, soit manuellement, soit par le biais d'un système automatisé, et l'application d'une haute tension provoque la formation d'un spray de gouttelettes chargées à partir de la pointe de l'aiguille, similairement au processus en ESI. Cette technique offre un moyen simple d'analyser des échantillons sans avoir besoin de pulvérisation de solvant et peut être achetée en tant que produit disponible dans le commerce.

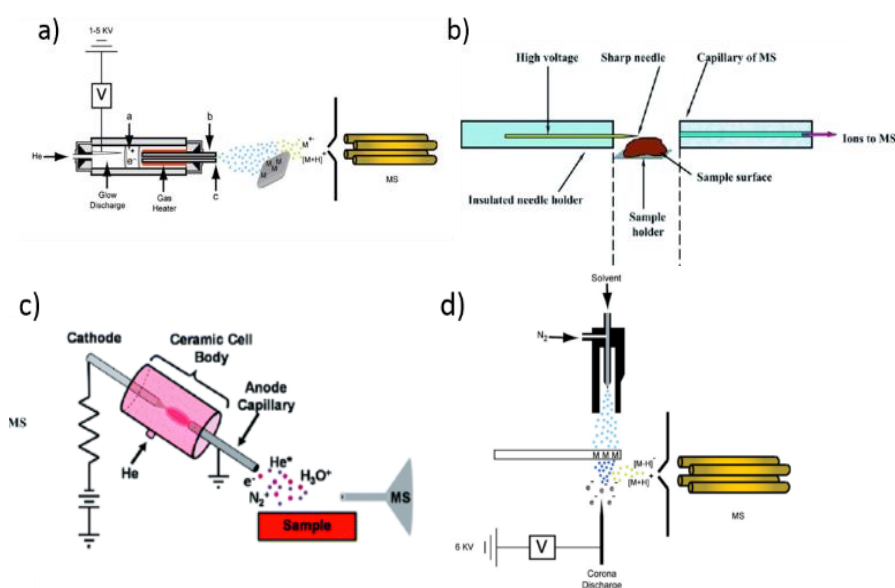


**Figure 6** : Différentes sources AIMS basées sur la pulvérisation du substrat. a) PSI<sup>37</sup>. b) PESI.

Une technologie basée sur l'extraction de liquide a été développée en 2017 par Livia Eberlin et ses collaborateurs pour faciliter l'échantillonnage de surfaces tridimensionnelles ou irrégulières, suivies d'une analyse par spectrométrie de masse<sup>50</sup>. Cette technologie, appelée MasSpec Pen, sera discutée en détail dans la suite du manuscrit en tant que potentielle technique de spectrométrie de masse utilisable directement *in vivo* au bloc opératoire afin de diagnostiquer le cancer.

### *Basées sur l'utilisation d'un faisceau de gaz ou de plasma*

Les méthodes d'extraction solide-liquide sont largement utilisées en AIMS mais les techniques basées sur un plasma sont également populaires<sup>51</sup> (**Figure 7**). Ces techniques utilisent des décharges électriques pour créer un plasma contenant des électrons, des espèces métastables et des radicaux qui induisent l'ionisation. Les méthodes d'ionisation ambiante par désorption au plasma découlent des mêmes principes chimiques que l'ionisation chimique à pression atmosphérique. Elles utilisent un processus similaire de décharge de plasma pour désorber et ioniser les molécules mais elles sont spécialement conçues pour l'analyse directe d'échantillons non modifiés sans couplage à des techniques de séparation.



**Figure 7** : Différentes sources AIMS basées sur l'utilisation de plasma à haute température. a) DART<sup>37</sup>. b) DAPCI c) FAPA<sup>52</sup>. d) ASAP<sup>37</sup>.

L'analyse directe en temps réel (DART) a été la première méthode d'ionisation ambiante basée sur le plasma et reste la plus répandue. Elle a été développée au début des années 2000 par Cody Durst et ses collègues<sup>52</sup>. En DART, un plasma est créé à partir d'une décharge électrique alors qu'un gaz inerte (He, Ar ou N<sub>2</sub>) circule dans le système. Le plasma contient des ions, des électrons et des molécules neutres excités électroniquement. Cependant, des lentilles électrostatiques sont placées pour ne permettre qu'aux molécules excitées de sortir du système. Ces espèces excitées vont réagir avec l'air ambiant pour former des clusters d'eau protonés. Ces clusters

vont réagir avec les analytes désorbés de la surface pour réaliser un transfert de proton du gaz réactif aux analytes. Tout comme en CI, le transfert de proton s'opèrera si l'affinité protéomique de l'analyte est supérieure à celle du gaz réactif. Le gaz réactif peut également être chauffé pour améliorer la désorption de l'analyte de la surface de l'échantillon et augmenter ainsi la sensibilité de la méthode. La DART est très polyvalente et peut être appliquée à des échantillons solides, liquides et gazeux, et elle est également commercialisée, ce qui en fait l'une des rares techniques AIMS disponibles sur le marché.

En sus de la technique DART, il y a plusieurs autres options pour les sources d'ions ambiantes à base de plasma qui peuvent être utilisées. Par exemple, la spectroscopie de rémanence à pression atmosphérique (FAPA) utilise la rémanence du plasma d'hélium ou d'argon pour générer des ions réactifs à transfert de charge qui peuvent ensuite interagir avec les analytes d'un échantillon et les ioniser<sup>53</sup>. Le FAPA utilise un courant plus élevé que le DART pour améliorer la désorption de l'échantillon sans l'ajout d'un gaz réactif chauffé. Ce courant plus élevé permet également d'obtenir une plus grande variété d'ions de gaz réactif. En DART comme en FAPA, les analytes présentent une énergie interne plus élevée que dans d'autres méthodes plus douces comme le DESI, et des fragmentations des analytes peuvent être observées.

Une autre méthode basée sur le plasma, la désorption chimique à pression atmosphérique (DAPCI), utilise également une décharge corona comme source d'ionisation, mais des solvants gazeux chauffés sont utilisés comme ions réactifs qui interagissent avec la surface de l'échantillon<sup>54</sup>. De plus, comme la DAPCI, la sonde d'analyse des solides atmosphériques (ASAP) est une construction modifiée d'une source standard APCI/ESI. Une caractéristique importante de l'ASAP étant de permettre d'ioniser des composés polaires et non polaires<sup>55</sup>.

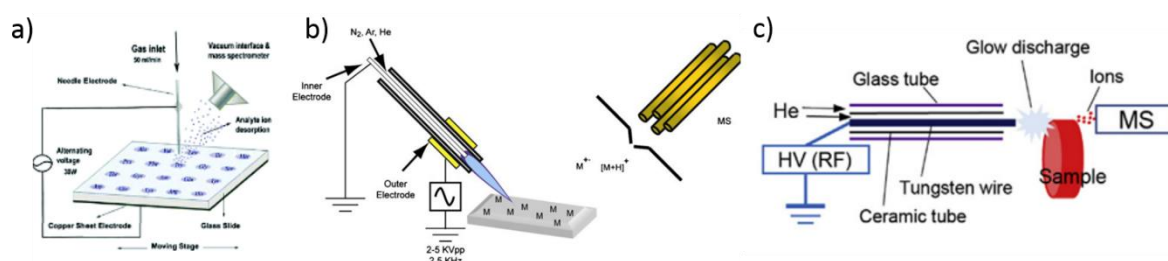


Figure 8 : Différentes sources AIMS basées sur l'utilisation de plasma à basse température. a) DBDI<sup>52</sup>. b) LTP<sup>37</sup>. c) PADI<sup>57</sup>.

Enfin, des techniques de spectrométrie de masse utilisant des plasmas à basse température ont été développées pour l'ionisation ambiante (**Figure 8**). L'ionisation par décharge à barrière diélectrique (DBDI) et l'ionisation par plasma à basse température (LTP) sont des exemples de techniques de spectrométrie de masse utilisant des plasmas à basse température pour l'analyse des échantillons en milieu ambiant. La LTP utilise une électrode portée à une haute tension et montée

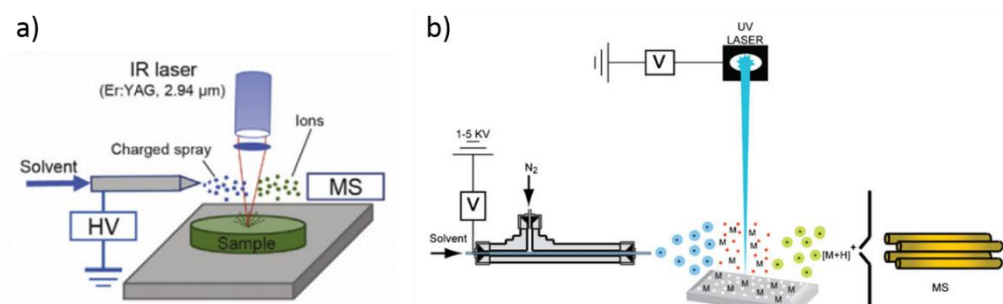


à l'extérieur de la barrière diélectrique alors que la contre-électrode en cuivre est placée dans le gaz de décharge. Cette configuration permet une analyse *in vivo* et est plus adaptée aux applications sur le terrain<sup>56</sup>. En effet, contrairement au DART, le LTP utilise un plasma froid qui est donc compatible avec une utilisation *in vivo*. Dans la méthode DBDI, le plasma est généré à l'extrémité d'une électrode à aiguille et mis en contact avec l'échantillon déposé sur une lame de verre. Une contre-électrode en cuivre est placée sous la lame de verre pour permettre la génération d'ions<sup>57</sup>. Comme la méthode DBDI, la désorption/ionisation assistée par plasma (PADI) est un exemple de sources directes de plasma. Dans les deux cas, une décharge de courant alternatif est créée directement entre une électrode et l'échantillon. Cependant, dans le cas du PADI, une décharge luminescente à radiofréquence est établie à l'extrémité d'une électrode à broche alimentée qui entre en contact direct avec l'échantillon<sup>58</sup>.

### *Basées sur l'utilisation d'un faisceau laser*

Les méthodes basées sur un processus d'électronébulisation et un plasma sont les plus couramment utilisées pour l'ionisation ambiante, mais il existe aussi d'autres techniques telles que l'ablation et/ou la désorption/ionisation laser, avec des sources laser UV ou IR. Ces méthodes présentent l'avantage d'une désorption efficace et précise grâce à une résolution spatiale plus élevée rendue possible par le contrôle de la focalisation du faisceau laser par des optiques appropriées contrairement aux techniques de désorption à base de solvants ou de plasma.

Une source ambiante d'ablation laser en une étape de désorption/ionisation existe, dérivée de l'AP-MALDI, la source d'ionisation par spray laser (LSI) est une source développée par l'équipe de Mc Ewen en 2010 où le faisceau laser irradie la face arrière du support où est déposé l'échantillon. La plume d'ablation est alors collectée par un capillaire de transfert chauffé connecté au spectromètre de masse.



**Figure 9** : Différentes sources AIMS utilisant un laser pour l'ablation couplé à une source d'ionisation secondaire. a) LAESI<sup>57</sup>. b) MALDESI<sup>37</sup>.

Cependant, ces approches ont une faible efficacité d'ionisation car la majorité des molécules désorbées par ablation laser restent neutres. Par conséquent, elles sont souvent couplées à une source d'ionisation secondaire, telle que l'ESI, la DART, l'APCI ou l'APPI pour améliorer la sensibilité. Parmi toutes ces techniques basées sur une ablation/désorption laser, deux sont

particulièrement populaires : l'ionisation par ablation laser et électronébulisation (LAESI) et la désorption laser assistée par matrice et ionisation par électronébulisation (MALDESI) (**Figure 9**).

Le LAESI est une méthode qui utilise un laser infrarouge pour ablater la surface de l'échantillon, produisant un nuage de molécules majoritairement neutres. La précision spatiale de cette méthode dépend du diamètre du faisceau laser focalisé, généralement positionné à une distance d'environ 200  $\mu\text{m}$  de la surface de l'échantillon<sup>59</sup>. Après l'ablation laser, les molécules sont capturées par un spray de gouttelettes de solvant chargées générées par ESI, placé perpendiculairement à la plume d'ablation. Le champ électrique créé par l'application d'une différence de tension entre le tip de l'ESI et la contre-électrode à l'entrée du spectromètre, permet aux gouttelettes chargées d'analyte d'être dirigée vers l'entrée de l'instrument où elles subissent, comme en ESI, un processus de désolvatation. Comme pour les autres méthodes basées sur l'ESI, la polarité de la molécule influe sur l'efficacité de l'ionisation. Bien que le LAESI ne nécessite pas de préparation de l'échantillon ou l'ajout de matrice, les échantillons doivent contenir une quantité suffisante d'eau pour absorber les rayons infrarouges et exciter correctement la cible. Récemment, la sensibilité de cette méthode a été améliorée par la recherche d'optimisation de la chambre d'ablation LAESI et de l'orientation du flux de gaz. Cette nouvelle conception a permis d'obtenir des performances similaires à celles du LAESI traditionnel sans les limites opérationnelles de la méthode initiale<sup>60</sup>. À l'origine, le MALDESI a été développé pour combiner le MALDI pour assurer la désorption avec une source ESI pour l'ionisation<sup>61</sup>. Pour cela, les échantillons sont co-cristallisés avec une matrice organique qui absorbe à la longueur d'onde du laser, puis ablatés avec le laser UV pulsé et ionisés avec une source ESI qui capture la plume de désorption comme en LAESI. Récemment, une fine couche de glace a été utilisée comme matrice avec un laser infrarouge pour l'ablation, ce qui élimine le besoin d'un dépôt de matrice organique et facilite son utilisation comme technique d'ionisation ambiante<sup>62</sup>.

Les dernières avancées dans les techniques de spectrométrie de masse basées sur l'ionisation ambiante par laser ont également porté sur l'intégration de l'ablation laser avec diverses méthodes d'ionisation autres que l'ESI. Ces méthodes incluent la désorption thermique à diode laser couplée à la photoionisation à pression atmosphérique (LDTD-APPI) et la désorption acoustique induite par laser couplée à la photoionisation à pression atmosphérique (LIAD-APPI). Ces méthodes, la LDTD-APPI<sup>63</sup> et la LIAD-APPI<sup>64</sup>, utilisent la photoionisation comme méthode d'ionisation, un processus où des photons UV à haute énergie interagissent avec des molécules neutres en phase gazeuse qui présentent de faibles potentiels d'ionisation et les ionisent, telles que les molécules de faible polarité.



Des avancées récentes dans le domaine de la spectrométrie de masse par ablation laser et ionisation ambiante incluent la technologie SpiderMass<sup>11</sup>, développée au laboratoire PRISM à Lille en France en 2010, ainsi que la technologie PIRL (PIRL-MS)<sup>65</sup>, dérivée du SpiderMass et proposée en 2016 par A. Zarrine Afsar et ses collaborateurs. Ces deux méthodes utilisent une source laser IR pour exciter la bande d'étirement O-H de l'eau dans des tissus biologiques principalement. Les ions générés sont ensuite acheminés vers le spectromètre de masse via un tube de transfert pour l'analyse MS. Il est important de noter que le PIRL utilise un laser picoseconde (0.3 ns exactement) tandis que le SpiderMass utilise un laser IR nanoseconde (6 ns). Ces deux technologies seront détaillées dans la suite du manuscrit en tant que potentiel outil de diagnostic du cancer en temps réel directement au bloc opératoire.

### *Sources AIMS alternatives*

Outre les méthodes d'extraction de liquides, de désorption par plasma et d'ablation par laser, des alternatives efficaces telles que la désorption thermique, vibratoire, acoustique et par évaporation ont été mises en place pour désorber et ioniser des molécules à partir d'échantillons complexes dans des conditions ambiantes.

Par exemple, la spectrométrie de masse à ionisation rapide par évaporation (REIMS) est une technique d'ionisation en milieu ambiant qui récupère les ions émis par un bistouri électrique pour l'analyse MS. En effet, il a été montré que la chaleur et l'énergie thermique produite par un bistouri électrique conduit à la production d'un aérosol qui contient des ions provenant de l'échantillon qui a été découpé<sup>66</sup>. Cette méthode a été largement utilisée pour l'analyse de divers échantillons complexes, notamment les cultures bactériennes, les produits alimentaires et les tissus cancéreux humains, avec un accent particulier sur l'analyse et le diagnostic des tissus cancéreux humains *in vivo* et *ex vivo*, qui sera plus amplement discuté dans la suite du manuscrit avec le développement de l'iKnife par l'équipe de Zoltan Takats.

Des techniques à ultrasons ou vibrations ont été employées pour transformer l'échantillon en aérosol avant l'analyse par spectrométrie de masse. Ces techniques incluent l'ionisation ambiante par pulvérisation vibrante à arêtes vives (VSSI) et la photoionisation extractive à pression atmosphérique (EAPPI) (**Figure 10**). Dans l'EAPPI, l'utilisation des ultrasons permet de générer un aérosol à partir d'un échantillon en solution, qui est ensuite ionisé pour l'analyse par spectrométrie de masse<sup>67</sup>. En ce qui concerne la VSSI, l'échantillon liquide est nébulisé en un spray à partir du bord tranchant d'une lame de microscope en verre par une vibration mécanique à haute fréquence produite par un transducteur piézoélectrique<sup>68</sup>. Ce spray est directement introduit dans l'entrée d'un spectromètre de masse, où l'ionisation des molécules se produit par évaporation du solvant et par des mécanismes similaires à ceux de l'ionisation à l'entrée. La VSSI est préférée aux

techniques similaires de pulvérisation de substrat, telles que PSI et PESI, car elle n'a pas besoin d'une haute tension, ce qui est avantageux pour les applications sur le terrain.

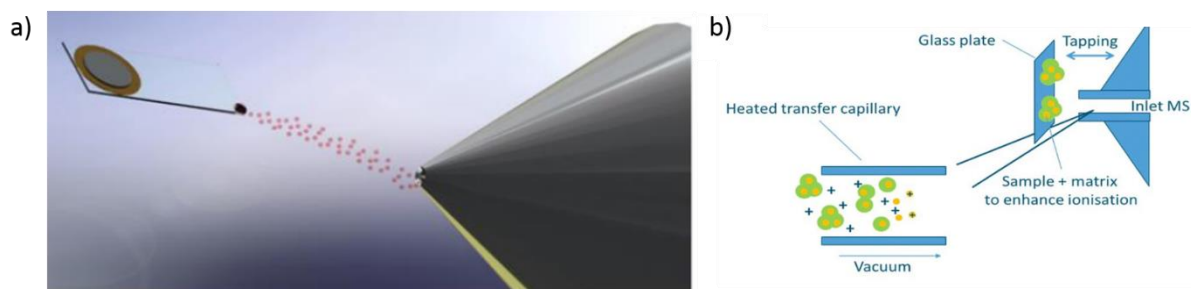


Figure 10 : Exemples de sources AIMS alternatives. a) VSSI<sup>69</sup>. b) MAI<sup>70</sup>.

La plupart des techniques AIMS utilisent une source d'ionisation externe pour produire des ions qui sont ensuite introduits dans le spectromètre de masse. Cependant, certains sous-ensembles fonctionnent en ionisant directement l'échantillon dans l'entrée du spectromètre de masse pour l'analyse, sans utiliser de source d'ionisation externe. Ces méthodes sont considérées comme des techniques d'ionisation ambiante si l'échantillonnage a lieu dans des conditions ambiantes et nécessite une préparation minimale de l'échantillon. Un exemple de cette méthode est l'ionisation assistée par matrice (MAI), où les échantillons solides sont co-cristallisés dans une matrice semi-volatile et placés près de l'entrée du spectromètre de masse<sup>69</sup>. La sublimation de l'échantillon libère des grappes d'analytes matriciels qui pénètrent dans le spectromètre de masse pour y être ionisés. Une variante de la MAI, appelée ionisation d'entrée assistée par solvant (SAII), utilise un mécanisme similaire pour ioniser les analytes dissous dans une matrice liquide<sup>70</sup>. Une variation séquentielle de la SAI, appelée ionisation d'entrée assistée par gouttelettes (DAII), a été développée plus récemment pour l'analyse des aérosols. Cette méthode utilise un condenseur pour générer des gouttelettes aqueuses à partir de particules d'aérosol contenant des analytes, qui sont ensuite directement introduites dans le spectromètre de masse<sup>71</sup>.

## La spectrométrie de masse en condition ambiante utilisée pour les analyses peropératoires

### Généralités

Le traitement le plus courant du cancer consiste à enlever chirurgicalement la tumeur dans son intégralité, accompagné de traitements complémentaires tels que la chimiothérapie et la radiothérapie. Pour une résection complète d'une tumeur solide, le chirurgien doit localiser précisément le tissu tumoral et évaluer son étendue en se basant sur des critères macroscopiques, des techniques d'imagerie et des connaissances sur le comportement spécifique du type de cancer. La principale préoccupation du chirurgien est de trouver le juste équilibre entre préserver le maximum de tissus sains et éliminer toutes les cellules cancéreuses. Une localisation précise de l'étendue de la tumeur permet au chirurgien de délimiter les marges d'excision de manière plus

précise. Même si le tissu environnant semble sain lors de l'examen histologique, il existe un risque accru de récurrence tumorale après la résection. Ces circonstances ont conduit à la nécessité d'améliorer les outils de diagnostic peropératoire des tumeurs.

L'avènement de l'AIMS a apporté progressivement plus de souplesse au domaine et élargi les applications de la spectrométrie de masse. Au fil des dernières années, le nombre de technologies AIMS a augmenté en raison de leur facilité d'utilisation, nécessitant peu ou voire aucune préparation d'échantillons. Nombreuses de ces techniques sont prometteuses pour des applications cliniques, notamment pour les essais *in vitro* et *ex vivo*, y compris l'analyse peropératoire<sup>72</sup>. Cependant, elles ne peuvent généralement pas être utilisées pour l'analyse MS *in vivo*. En réponse à ce besoin, des sondes AIMS en temps réel ont émergées, conçues pour des mesures rapides, sensibles, précises et exactes de façon *in vivo* pour la chirurgie guidée. Chaque technique favorise l'ionisation d'une manière spécifique, que ce soit par l'extraction de gouttelettes de solvant chargées, l'ablation ultrasonique des tissus, l'ablation/excitation laser et l'accumulation de fumée<sup>10</sup>. Pour s'adapter à l'environnement clinique, tous les systèmes sont constitués d'une sonde couplée à une ligne de transfert permettant de déporter le spectromètre de masse. Ces sondes fournissent une multitude d'informations chimiques à partir de l'analyse directe des tissus, offrant des profils moléculaires spécifiques de chaque tissu. Les codes-barres moléculaires, ainsi obtenus, peuvent être utilisés pour évaluer les tissus sains par rapport aux tissus cancéreux afin de diagnostiquer le cancer, détecter l'hétérogénéité intra tumorale ainsi que pour localiser les régions métastatiques. Ces sondes permettent préférentiellement une détection des petites molécules comme les lipides et les métabolites. Sachant qu'un métabolisme lipidique anormal est une caractéristique commune des cellules cancéreuses, même à des stades précoces du développement de la tumeur, il est alors possible par analyse moléculaire de délimiter les zones cancéreuses des zones saines chez un patient. Dans ce manuscrit, les quatre principales sondes existantes seront présentées en détail par ordre de création, l'iKnife, le SpiderMass, le MasSpec Pen et le PIRL (Figure 11).

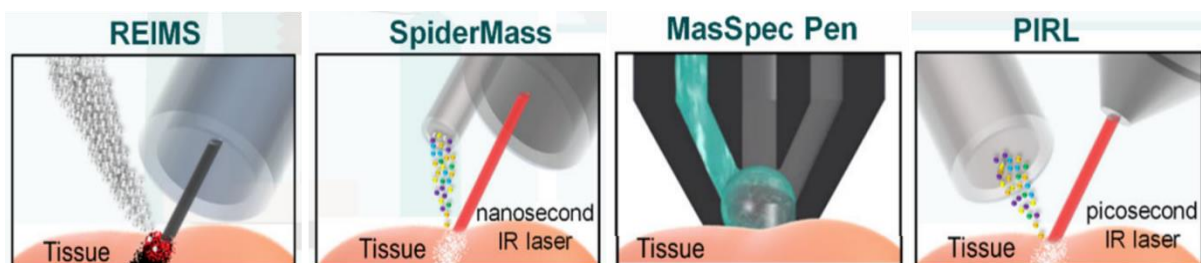


Figure 11 : Représentation graphique des sondes de spectrométrie de masse portables développées pour une utilisation peropératoire<sup>74</sup>.

## iKnife

L'iKnife, également connu sous le nom de couteau intelligent, est un concept développé en 2009 par le groupe de Zoltan Takats, pour la chirurgie guidée, qui combine la dissection et l'analyse. Cette technologie repose sur des techniques d'électrochirurgie qui utilisent un courant électrique à haute fréquence pour couper et cautériser les tissus<sup>66</sup>. Le courant électrique appliqué fait passer le tissu coupé de l'état liquide à l'état gazeux, provoquant une explosion de phase qui génère un panache de fumée contenant des traces du contenu intracellulaire du tissu. Cet aérosol est ensuite transmis au spectromètre de masse par un tube de transfert en polymère. L'iKnife est équipé d'une interface atmosphérique modifiée, comprenant une entrée capillaire chauffée et un dispositif de transfert d'aérosols orthogonal utilisant une pompe Venturi<sup>73</sup>. Les ions sont créés à l'entrée du spectromètre de masse en post-ionisation par l'interface REIMS grâce à une collision de l'aérosol sur une surface chauffée (un ruban de Khantal® porté à plusieurs centaines de °C)<sup>74,75</sup>. Chaque type de tissu produit une distribution unique de molécules lorsque les cellules sont vaporisées, des altérations distinctes se produisant lorsque le tissu devient cancéreux, ce qui établit une empreinte moléculaire unique qui peut être utilisée pour le diagnostic *in vivo*. Cette empreinte peut être utilisée pour distinguer un large éventail de pathologies cancéreuses en la comparant à une base de données stockée d'empreintes validées de cancers connus. Un logiciel propriétaire fait correspondre les empreintes digitales produites par l'aérosol et transmet au chirurgien l'identification histopathologique en temps réel pendant l'opération, guidant ainsi les marges de résection. L'iKnife combine les technologies de cautérisation électrique et de spectrométrie de masse pour faciliter le diagnostic intra-opératoire du tissu tumoral. L'un des avantages clés de cette technique est qu'il est basé sur des dispositifs chirurgicaux couramment utilisés. Cependant, il y a aussi certains inconvénients majeurs comme une faible résolution spatiale (de 0.5 à 2 mm), la destruction des échantillons pendant l'analyse et, par conséquent, l'impossibilité d'effectuer une validation croisée directe avec l'histologie.

Ce système a été testé *in vivo* et a permis d'identifier de manière certaine les tissus normaux, les marges et les tissus cancéreux provenant de patients atteints de cancer de l'ovaire<sup>76</sup>, du sein<sup>77,78</sup>, du col de l'utérus<sup>79</sup> et de l'endomètre<sup>80</sup>, ainsi que de classer de manière endoscopique la paroi intestinale, les cancers et les polypes<sup>81,82</sup> (**Figure 12**). La technologie REIMS a également été utilisée pour l'analyse des aliments<sup>83-85</sup> et pour l'identification des colonies bactériennes et la cartographie directe *in vivo* de la croissance bactérienne<sup>86</sup>. Ce système peut aussi être couplé à un laser CO<sub>2</sub> chirurgical plutôt qu'à un couteau diathermique. Cela permettant d'élargir son champ d'application car le laser génère des aérosols riches en lipides à partir de tissus durs tels que les os, la moelle osseuse et les tissus cartilagineux<sup>87</sup>. De plus, le laser produit moins de dommages aux tissus que le couteau diathermique (même si le laser coupe quand même les tissus) et une meilleure

reproductibilité du laser a été observée due à la position fixe du laser et à son temps d'ablation contrôlé.

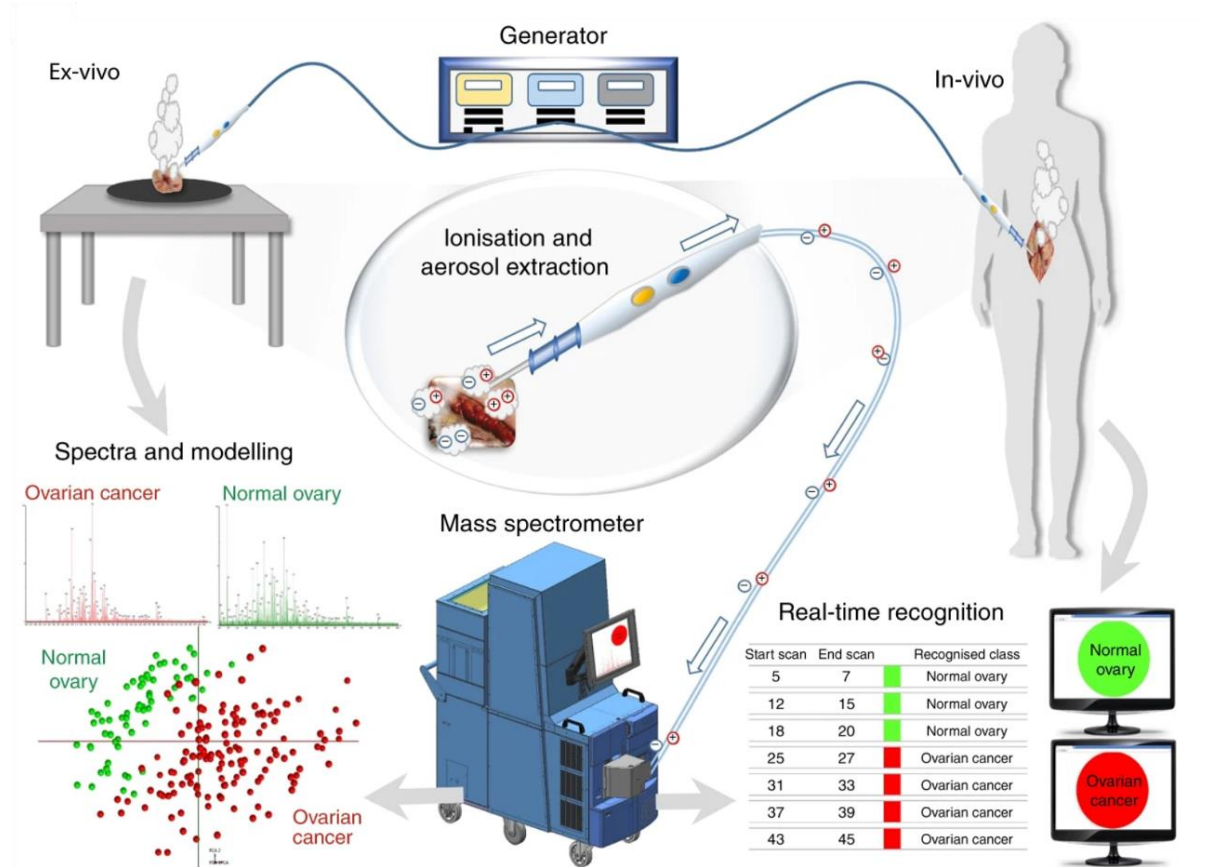


Figure 12 : Installation de la technologie iKnife *ex-vivo* et *in-vivo* dans une salle d'opération. Le courant électrique est appliqué au tissu et les particules chargées qui en résultent sont extraites par la pièce à main et aspirées dans l'entrée de l'interface REIMS et analysées dans le spectromètre de masse pour produire des spectres de masse spécifiques aux tissus, qui sont ensuite soumis à une analyse statistique multivariée à l'aide d'ACP-LDA. En l'espace d'une à deux secondes, le diagnostic tissulaire en temps réel s'affiche sur un écran à l'intention du chirurgien<sup>78</sup>.

### SpiderMass

Créée en 2010 au laboratoire PRISM à Lille puis brevetée en 2014, le SpiderMass est une technique qui permet des analyses *in vivo* en temps réel avec un minimum d'invasivité en étant basé sur l'utilisation d'un laser émettant dans l'IR (**Figure 13**). Le SpiderMass se compose d'une sonde de micro-échantillonnage, d'une ligne de transfert et de l'analyseur MS<sup>11</sup>. En effet, cette technologie utilise un laser IR comme sonde de micro-échantillonnage pour exciter la vibration la plus intense (mode d'étirement O-H) des molécules d'eau, ce qui entraîne un processus de désorption/ionisation. Un oscillateur paramétrique (OPO) pompé par un laser Nd:YAG à impulsion de 6 nanosecondes est utilisé afin d'ajuster la longueur d'onde du laser à 2.94  $\mu\text{m}$  pour une performance optimale. Ce processus de désorption et d'ionisation, appelé ionisation par désorption laser assistée par l'eau (WALDI), est similaire au MALDI, où l'eau joue le rôle de la matrice endogène pour produire des ions en phase gazeuse<sup>12</sup>. Pour l'analyse de tissus, le laser est focalisé



à la surface de l'échantillon via une fibre optique qui se termine par une pièce à main, permettant l'analyse sans contact de tout type de surface *ex vivo* ou *in vivo*. Un tube de transfert est placé au-dessus de la surface du tissu à analyser pour transférer la plume de désorption formée lors de l'ablation du tissu vers le spectromètre de masse. Ce système est peu invasif : lors d'une seule expérience, seulement 0.1 à 0.3 mm<sup>3</sup> d'échantillon sont prélevés (profondeur d'échantillonnage de quelques µm seulement) et les échantillons sont déshydratés de manière réversible. Les profils des spectres de masse obtenus avec le SpiderMass en modes d'ionisation positive et négative montrent la détection de divers acides gras et espèces lipidiques couramment présents dans les tissus.

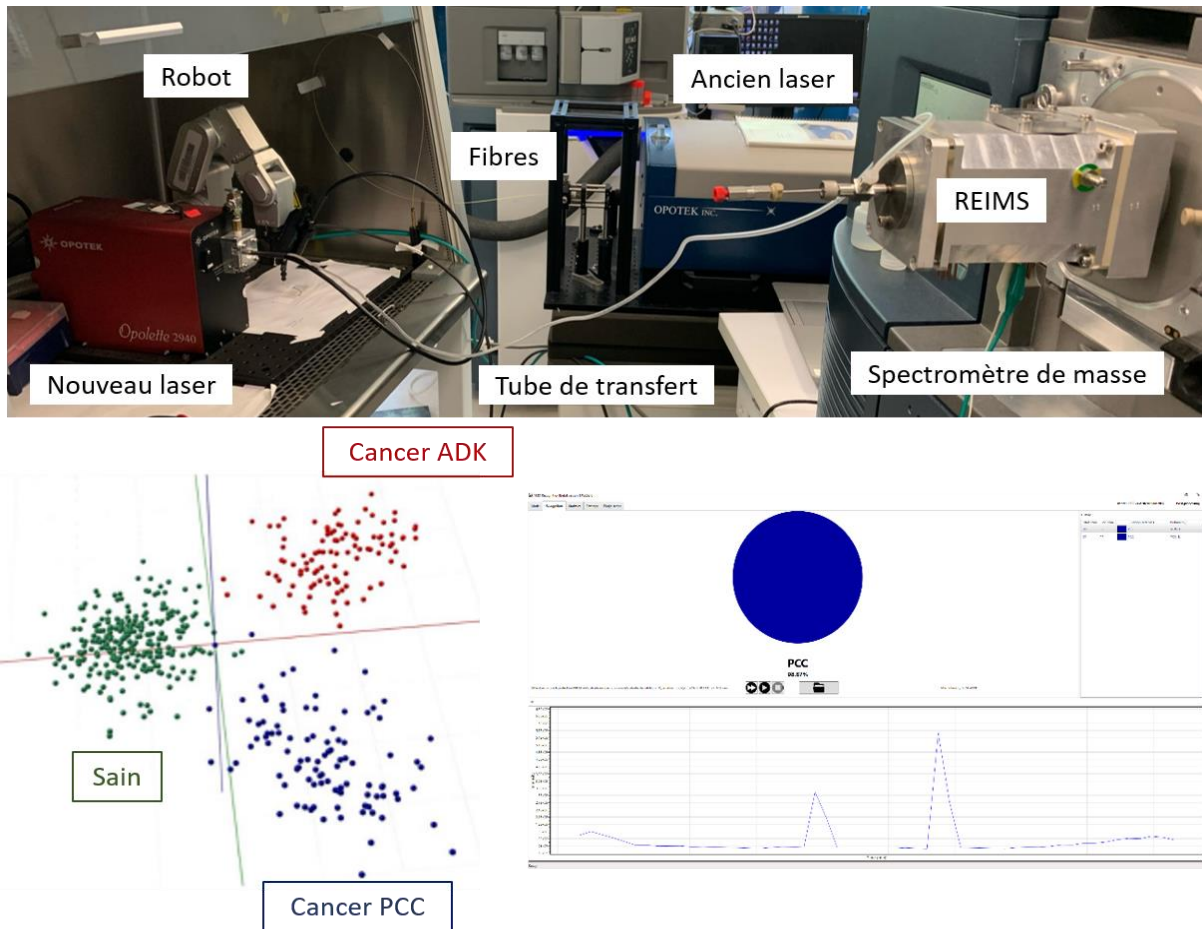


Figure 13 : Photo du prototype actuel SpiderMass comprenant les lasers et leurs fibres, le bras robotisé, le tube de transfert, l'interface REIMS ainsi que le spectromètre de masse. Un modèle de classification obtenu par analyse statistique multivariée à l'aide d'ACP-LDA après analyse par SpiderMass d'une cohorte de tissus de cancer œsogastrique ainsi qu'un exemple de résultat obtenu par prédiction à l'aveugle d'un tissu de cancer PCC.

Actuellement, le principal inconvénient réside dans la fragilité de la fibre nue utilisée dans l'installation, ce qui suscite des préoccupations en matière de santé publique en raison du risque de rupture de la fibre. La solution a été trouvée récemment grâce à une nouvelle optimisation du prototype SpiderMass en 2022. En effet, le laboratoire PRISM a pu faire l'acquisition d'un nouveau laser plus compact et équipé d'une fibre gainée présentant des connecteurs SMA à chaque extrémité permettant une connexion plug & play de la fibre au niveau du site d'injection et de la pièce à main.

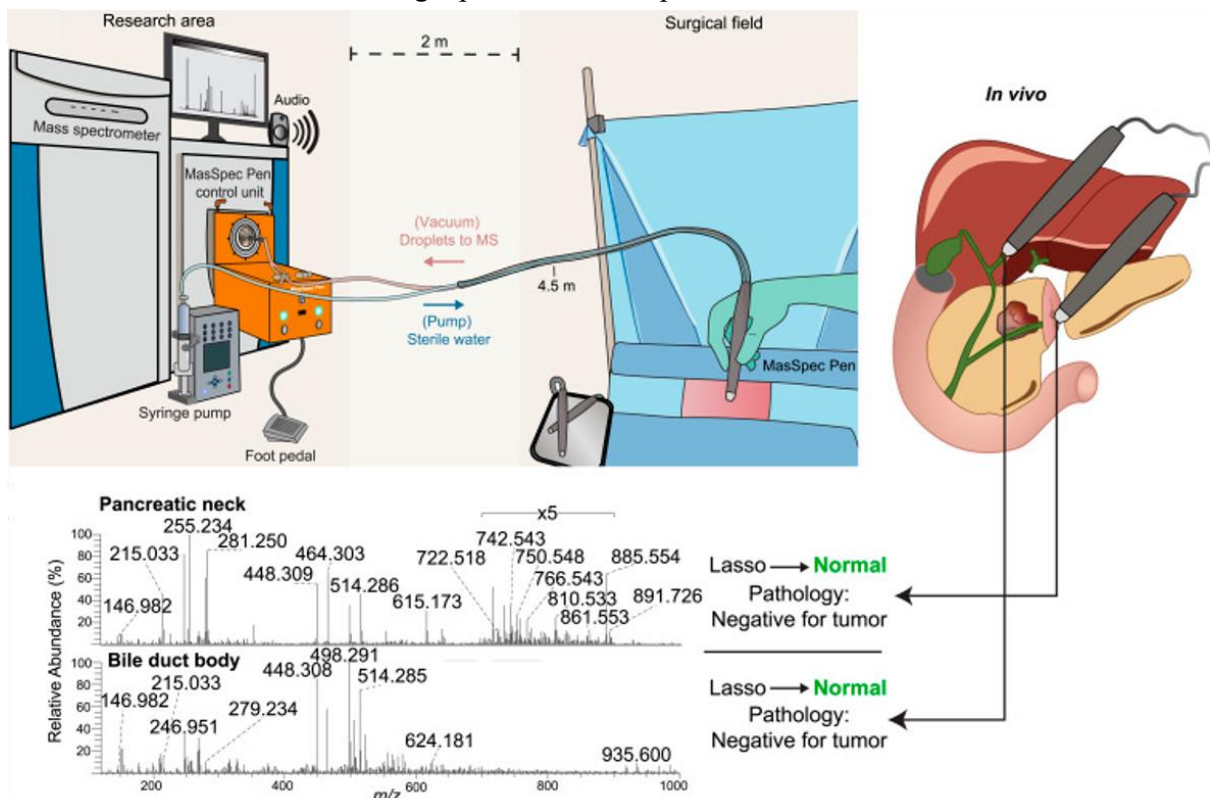
Le système SpiderMass a été utilisé pour analyser des biopsies de cancer de l'ovaire *ex vivo*, de la peau humaine *in vivo* ainsi que des cellules cultivées sans préparation<sup>11</sup>. De plus, il a également été utilisé pour analyser des échantillons de sarcome chez les chiens *ex vivo*, atteignant un taux de classification correcte de 97 % pour les types et les stades de tumeurs. Cette étude ayant été suivie par une analyse *in vivo* à la clinique Oncovet<sup>88,89</sup>. L'utilisation du SpiderMass a aussi permis d'analyser des tissus de carcinome épidermoïde buccal<sup>16</sup>, avec une bonne précision de classification des zones tumorales et non tumorales de 83 % avec une validation croisée 5 fois, ainsi que des tissus de glioblastomes<sup>90</sup> en tant que validation de données obtenues par imagerie MALDI.

Au cours des dernières années, le système a été utilisé pour analyser des lignées de cellules macrophages génétiquement modifiées ainsi que des protéines intactes<sup>91</sup>. Plus récemment, le SpiderMass a été utilisé dans de nouveaux domaines. En effet, cette technologie a aussi montré son applicabilité pour l'analyse d'échantillons de sarcomes FFPE non traités<sup>92</sup>. Il a aussi été performant pour l'analyse *in vivo* de plantes de cannabis<sup>17</sup> ainsi que pour l'analyse de feuilles de vignes. De plus, une thèse de doctorat a été soutenue en 2021 traitant de l'analyse métabolique et lipidomique *in vivo* de dermatoses par spectrométrie de masse en temps réel par la technologie SpiderMass en collaboration avec l'entreprise privée Pierre Fabre dermo-cosmétique<sup>93</sup>.

### MasSpec Pen

Le MasSpec Pen, développé en 2017 par le groupe de Livia Eberlin, repose sur des principes similaires à ceux de l'analyse de surface par extraction de liquide et de l'échantillonnage de surface par micro-jonction liquide. Il s'agit d'un dispositif portable automatisé, jetable et compatible avec le corps humain, qui permet l'échantillonnage direct et non destructif de tissus en extrayant des liquides. Le dispositif est composé de trois éléments : une pompe à seringue qui délivre des gouttelettes d'eau (4-10 µl) à la sonde de prélèvement, des tubes PTFE dotés de valves intégrées pour transporter l'eau vers le tissu et en provenance de celui-ci après un contact de 3 secondes, et un dispositif de la taille d'un stylo utilisé pour sonder les tissus biologiques<sup>94</sup>. L'utilisation de l'eau comme solvant permet d'extraire en douceur les composés hydrophiles et amphiphiles, mais nécessite un contact étroit entre la sonde et le tissu. En effet, le principal inconvénient du MasSpec Pen est son potentiel de contamination possible du système par des débris tissulaires, ce qui est dû au fait que l'appareil fonctionne en contact liquide avec le tissu<sup>10</sup>. Par ailleurs, le transport des gouttelettes du tissu vers le spectromètre de masse est plus long car les liquides circulent moins vite que les aérosols et peuvent être à l'origine de cross-contaminations entre les différents points analysés. Enfin, du fait du contact nécessaire avec le tissu, le MasSpec Pen n'est pas utilisable en mode imagerie.

Le MasSpec Pen a été utilisé pour analyser *ex vivo* des échantillons de biopsies humaines provenant de cancers du sein, de la thyroïde, de l'ovaire<sup>95</sup> et du poumon<sup>96</sup>, ainsi que des tissus sains, et ont pu diagnostiquer les tissus sains et les tissus cancéreux avec une grande précision, obtenant des taux élevés de sensibilité (96.4 %), de spécificité (96.2 %) et de précision globale (96.3 %). Ils ont également pu prédire avec précision les tumeurs bénignes et malignes de la thyroïde, ainsi que les différents sous-types histologiques du cancer du poumon. De manière notable, le MasSpec Pen a permis un diagnostic précis du cancer dans des zones tumorales présentant une composition histologique mixte et a également été utilisé avec succès pour diagnostiquer le cancer *in vivo* lors d'interventions chirurgicales menées sur des souris porteuses de tumeurs<sup>94</sup>. Les performances du MasSpec Pen ont été évaluées pour le diagnostic rapide du carcinome séreux de haut grade, démontrant une sensibilité clinique de 96.7 % et une spécificité de 95.7 %<sup>95</sup>. Récemment, le système a été utilisé *in vivo* par sept chirurgiens différents et sur des tissus excisés lors de 100 interventions chirurgicales<sup>97</sup>. Les modèles de classification basés sur les données moléculaires ont montré une concordance globale de 91.5 % avec la pathologie, une sensibilité de 95.5 % et une spécificité de 89.7 % pour distinguer les tissus pancréatiques normaux des cancers<sup>98</sup> (**Figure 14**). Plus récemment, le MasSpec Pen a pu être intégré au système chirurgical da Vinci pour l'analyse de tissus *in vivo* lors d'une chirurgie porcine assistée par robot<sup>99</sup>.



**Figure 14** : Schéma de la plate-forme MasSpec Pen dans une salle d'opération ainsi que les spectres de masse obtenus *in vivo* à partir de la marge du col du pancréas et du corps du canal cholédoque d'un patient soumis à une intervention de Whipple. Les résultats correspondants de la prédiction Lasso (effectuée après la chirurgie) et les notes pathologiques finales sont illustrés<sup>100</sup>.



Le MasSpec Pen peut être utilisés dans de nombreux domaines d'applications. En effet, il a aussi été utilisé pour l'analyse des aliments comme par exemple l'authentification de la provenance de la viande<sup>100</sup>. Cette technologie peut aussi avoir des applications dans le médico-légal afin de détecter des drogues comme la cocaïne ou l'oxycodone<sup>101</sup>. A l'aide de celui-ci, un dépistage rapide de COVID-19 directement à partir d'écouvillons nasopharyngés cliniques peut être effectué<sup>102</sup>. D'autres applications dans le domaine de la biologie ont également été décrites, par exemple avec la détection et le suivi des alcaloïdes de façon *in vivo* chez des grenouilles venimeuses, qui les utilisent en tant que défense chimique contre les prédateurs<sup>103</sup>.

### PIRL-MS

Le PIRL a été introduit comme un nouveau concept de chirurgie laser visant à minimiser la formation de cicatrices postopératoires<sup>104</sup>. Plus récemment en 2017, le groupe dirigé par Arash Zarrine-Afsar a développé une sonde portable basée sur un laser picoseconde directement lié à un spectromètre de masse, via une interface chauffée<sup>105</sup>. Ils ont utilisé un tube Tygon de 2 mètres de long avec un diamètre intérieur de 1.6 mm, qu'ils ont connecté au tube de transfert d'ions d'une source commerciale DESI de Waters. La température du capillaire de transfert d'ions et de la source MS était suffisamment élevée pour dissoudre les phospholipides et les acides gras libérés du tissu par la sonde PIRL manuelle. Le PIRL est une réadaptation de la technologie SpiderMass avec comme unique différence une impulsion laser plus courte, de l'ordre de la centaine de picosecondes. En effet, ce dispositif portable utilise des impulsions laser d'une durée de 300 picosecondes pour réduire les dommages thermiques aux tissus, permettant ainsi une ablation laser moins agressive<sup>106</sup>. Cependant, les lasers picosecondes restent coûteux et présentent des défis en termes de fibrage en raison de leur énergie de pointe requise pour être injectée dans les fibres. Par ailleurs, l'effet de la réduction de l'impulsion sur le rendement d'ionisation n'est pas connu. Enfin, avec une impulsion de 300 ns, les dommages réalisés au tissu en PIRL ne sont pas très différents du SpiderMass. Récemment, une interface double intégrant PIRL-MS et DESI-MS dans une seule plateforme a été décrite par le même groupe, permettant l'acquisition des informations de complémentarité fournies par ces techniques<sup>107</sup>.

Ce système a été appliqué à des tumeurs sous-cutanées de xénogreffes murines provenant de différentes lignées cellulaires de médulloblastome humain<sup>105</sup> (**Figure 15**). Les auteurs ont pu classer et différencier les sites cancéreux nécrotiques des sites cancéreux viables dans les tissus en se basant sur des profils moléculaires spécifiques, en utilisant des méthodes statistiques multivariées. Le système PIRL portable a été le plus largement exploré pour l'analyse des tissus du cancer du cerveau, en particulier pour le diagnostic et le sous-typage du médulloblastome, la tumeur cérébrale maligne la plus répandue chez l'enfant<sup>108</sup>. L'analyse PIRL-MS a obtenu un taux de réussite

de 98 % pour déterminer les sous-groupes observés sur 194 tissus tumoraux provenant de 19 tumeurs distinctes. Une étude a aussi montré que les aérosols tissulaires générés par un laser picoseconde permettaient de distinguer différents tissus cérébraux selon leurs profils lipidiques<sup>109</sup>. Plus récemment en 2022, une étude de faisabilité a démontré que le PIRL permettait de diagnostiquer en 10 secondes certains types de cancer de la peau. En effet, celui-ci a permis de détecter et de classer les mélanomes, les carcinomes épidermoïdes et les peaux normales avec des valeurs moyennes de sensibilité et de spécificité de 92 % et 95 %<sup>110</sup>.

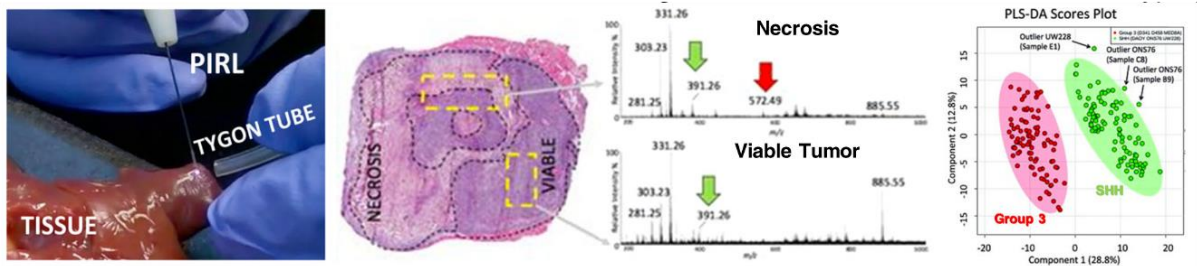


Figure 15 : Application du système PIRL-MS sur des xélogreffes de médulloblastome et de cancer du sein<sup>74</sup>.

### Analyse statistique des données moléculaires obtenues

Les sondes AIMS ont été créées pour permettre l'utilisation de la spectrométrie de masse directement dans la salle d'opération et entre les mains du personnel médical. La plupart de ces sondes sont conçues pour deux contextes cliniques différents : premièrement, une approche *in vivo* utilisée pendant les opérations de résection de tumeurs, et deuxièmement, une approche *ex vivo* utilisée en tant qu'outil de diagnostic par les cliniciens et les pathologistes<sup>111</sup> (Figure 16). Que ce soit dans le contexte *in vivo* ou *ex vivo*, l'instrument doit être entraîné à reconnaître les cellules, les types ou les stades du cancer avant d'être utilisé dans un contexte clinique. Comme cela, pendant l'opération, l'appareil utilise les données d'entraînement pour fournir un retour d'information instantané au chirurgien.

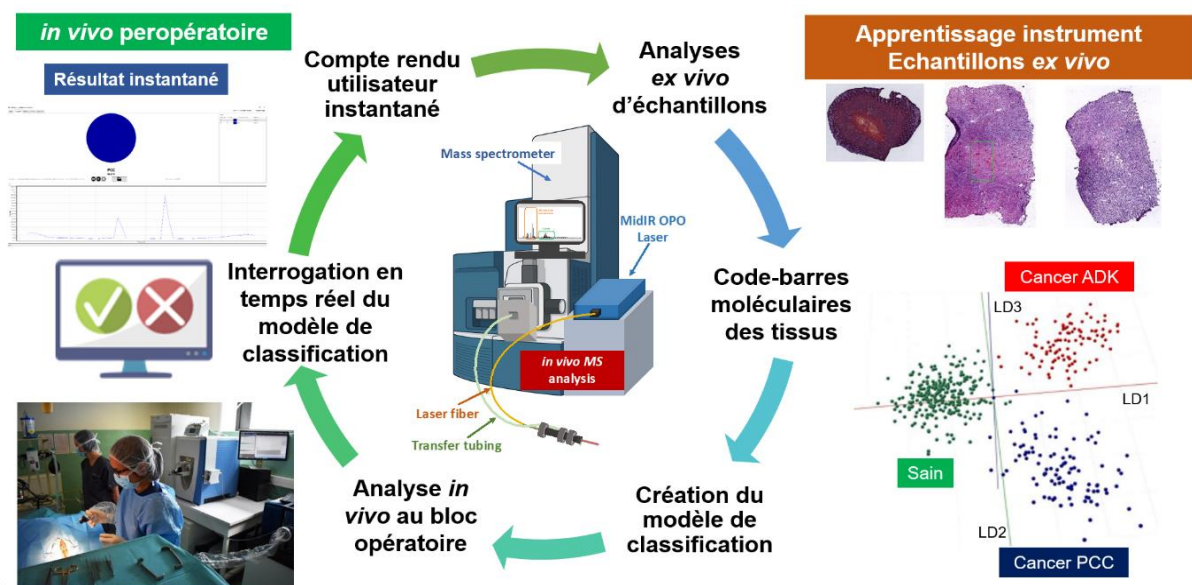


Figure 16 : Méthodologie utilisée pour l'utilisation d'une technologie AIMS pour le diagnostic du cancer *in vivo*.

La première étape, l'entraînement, est effectué en amont et peut être réalisé dans un laboratoire de recherche. Il implique la collecte de données à partir de coupes de tissus ou de tissus entiers excisés, frais ou congelés, dont le diagnostic est déjà connu, afin de constituer une base de données moléculaires basée sur des classes histopathologiques prédéfinies. Une fois que les spectres de masse sont collectés, les profils moléculaires peuvent être soumis à différents algorithmes d'apprentissage automatique. Par exemple, des méthodes multivariées telles que l'analyse en composantes principales (ACP) combinée à l'analyse discriminante linéaire (LDA) ont été utilisées pour créer des modèles de classification et classer les tissus normaux des tissus cancéreux<sup>73,77,82</sup>, ainsi que pour sous-typier et classer les sarcomes<sup>88</sup> et stratifier les patients atteints de gliomes<sup>105</sup>. La régression LASSO (opérateur de sélection et de moindre rétrécissement absolu) a été utilisée pour construire des modèles de classification basés sur les données MasSpec Pen, permettant de classer les cancers du poumon, les tumeurs thyroïdiennes<sup>94</sup> et les cancers ovariens<sup>95</sup> par rapport aux tissus normaux. Tandis que l'algorithme random forest (RF) qui a été utilisé pour la classification des données du LA-REIMS<sup>112</sup>, nécessite un prétraitement approfondi des données spectrales de masse<sup>113</sup>.

Récemment, de nouveaux algorithmes d'apprentissage profond, tels que les réseaux neuronaux convolutionnels (CNN), ont remis en question les approches d'apprentissage automatique traditionnelles en permettant l'extraction des caractéristiques et la classification en une seule étape. Dans le contexte peropératoire, des approches d'apprentissage par transfert ont été étudiées pour améliorer les taux de classification de plusieurs échantillons biologiques, y compris le sous-typage des sarcomes analysés par le SpiderMass<sup>114</sup>.

Une fois que les modèles de classification sont construits, ils peuvent être testés à l'aide de différentes techniques de validation croisées et sur des échantillons inconnus. Ainsi, pendant l'intervention chirurgicale, les modèles de classification sont utilisés en temps réel. Des logiciels spécifiques ont été développés pour corréliser les profils moléculaires en temps réel avec les modèles de classification préétablis, permettant ainsi une classification *in vivo* pendant l'opération chirurgicale et une lecture simplifiée pour le chirurgien, par exemple, à l'aide d'une représentation en code couleur.

Bien que le but premier de ces technologies soit de pouvoir différencier les tissus sains des tissus cancéreux *in vivo* en temps réel au bloc opératoire, il est aussi intéressant de pouvoir identifier des potentielles molécules biomarqueurs spécifiques des différents cancers analysés. Cela afin de donner un sens biologique et potentiellement découvrir de nouveaux traitements.

## Imagerie par spectrométrie de masse

### Généralités

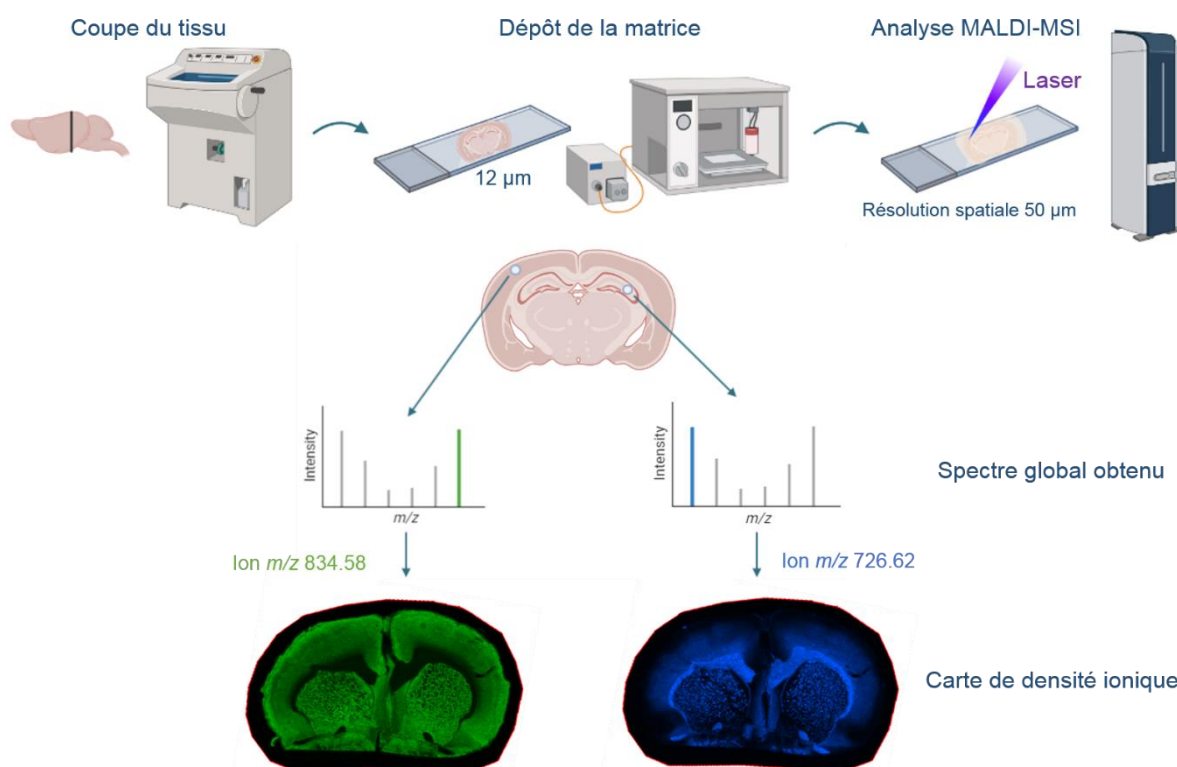
La technique d'imagerie par spectrométrie de masse (MSI) permet d'analyser la répartition spatiale des molécules sur des surfaces complexes. Cette méthode permet de combiner l'identification moléculaire et l'analyse de la localisation, créant ainsi un microscope chimique qui peut être utilisé pour caractériser directement la surface des tissus histologiques<sup>115</sup>. L'imagerie basée sur la spectrométrie de masse permet la détection et la cartographie sans étiquette d'une large gamme de composés biologiques de façon non ciblée. C'est donc une technique d'imagerie moléculaire puissante qui permet de détecter et de localiser des molécules endogènes et exogènes directement à partir de sections histologiques de quelques microns d'épaisseur<sup>116</sup>. Dans le cas d'une analyse en mode imagerie, l'ensemble de l'échantillon est analysé en mode pixel par pixel. A chaque position, un spectre de masse est enregistré avec les coordonnées x et y correspondantes. A la fin de l'acquisition, il est alors possible de mettre en évidence la distribution d'un composé sur la base de la distribution de l'intensité de l'ion correspondant. Les sources d'ionisation permettant de réaliser des expériences d'imagerie MS reposent sur leur capacité à pouvoir désorber des molécules à partir d'une surface solide. De nos jours, cette technique est largement utilisée en clinique<sup>117</sup> afin de mettre en évidence des biomarqueurs, pour le diagnostic, le pronostic et la mise en évidence de la réponse à un traitement ainsi que pour réaliser de l'histologie moléculaire dans le cas de la classification et de la stratification de patient<sup>90</sup>. Contrairement à d'autres techniques d'imagerie, telles que l'immunohistochimie, l'autoradiographie et la fluorescence, la MSI a l'avantage d'être sans marquage, d'avoir une large plage de détection et une sensibilité élevée.

De nombreuses sources de spectrométrie de masse sont utilisables en imagerie, que ce soit des sources à basse pression ou en condition ambiante. Cependant, seules les techniques d'imagerie MALDI et WALDI seront discutées dans ce manuscrit. En effet, ce sont les seules techniques qui ont été utilisées lors de ce travail de thèse.

### L'imagerie MALDI

La première méthode d'imagerie par spectrométrie de masse a été développée par le groupe de Caprioli en 1997<sup>118</sup>. Celle-ci utilise une source MALDI et permet de détecter ainsi que de localiser les molécules contenues dans un tissu de façon non ciblée. Comme discuté dans les paragraphes précédents, cette source utilise un laser UV ou IR et une matrice organique en large excès par rapport à l'échantillon pour créer un processus de désorption/ionisation. La matrice absorbe alors l'énergie du laser pour ioniser les composés, permettant ainsi la désorption et l'ionisation des molécules du tissu<sup>119</sup>.

Si la phase d'imagerie elle-même est l'élément central de la procédure, la préparation des échantillons pour l'analyse est une étape essentielle qui influence de manière significative la qualité des résultats obtenus lors de l'imagerie. Ainsi, l'obtention d'images MALDI nécessite une série d'étapes préalables à l'acquisition et au traitement des données. Les principales étapes incluent la préparation du tissu pour sa conservation, la réalisation des coupes, le dépôt de la matrice nécessaire à l'analyse MALDI, l'acquisition des données en mode imagerie, ainsi que le traitement des données pour les visualiser sous forme d'images (**Figure 17**). À chaque étape de cette séquence, des développements ont été réalisés au fil des années depuis l'introduction de la méthode, dans le but d'améliorer les performances de l'imagerie MALDI en termes de qualité d'analyse, de résolution spatiale des images, ainsi que de temps d'acquisition et de traitement des données<sup>120</sup>.



**Figure 17** : Méthodologie utilisée pour l'imagerie MALDI afin d'obtenir une carte de la répartition de chaque ion présent dans le tissu.

Chaque étape de préparation des échantillons est essentielle pour obtenir des résultats d'imagerie de haute qualité. L'une des étapes cruciales de la préparation d'échantillon est le dépôt de la matrice, tant dans le choix de la matrice que dans la procédure de dépôt de celle-ci. En ce qui concerne les stratégies d'application, il existe plusieurs choix, allant des sprayeurs manuels aux instruments de dépôt automatique. Toutes les méthodes sont satisfaisantes à condition qu'elles permettent d'obtenir un revêtement homogène et des cristaux de petite taille.

Le choix de la matrice joue un rôle crucial dans la réalisation des expériences d'imagerie MALDI, car elle détermine la classe de biomolécules détectée et influence la qualité du spectre obtenu (**Tableau 1**). Dans les approches de métabolomique et lipidomique, l'utilisation d'une



matrice avec un faible nombre de pics de matrice est bénéfique pour minimiser le risque de masquer des biomarqueurs potentiels. Il existe de nombreuses matrices disponibles, chacune ayant ses compatibilités spécifiques, ce qui signifie que la matrice doit être sélectionnée en fonction de chaque analyte. En général, pour l'analyse de petites molécules en mode d'ions positifs telles que les lipides, la matrice la plus couramment utilisée est la 2,5-dihydrobenzoïque (DHB)<sup>121</sup>, car elle présente peu de pics dans les basses masses. En ce qui concerne la détection des protéines, la matrice utilisée est la matrice acide sinapique (SA)<sup>122</sup> alors que pour la détection des peptides, la matrice acide  $\alpha$ -cyano-4-hydroxycinnamique (HCCA)<sup>123,124</sup> ou encore la matrice HCCA-Aniline (HCCA/ANI) développé au laboratoire PRISM sont privilégiées. En revanche, le mode négatif s'est avéré utile pour les métabolites et les lipides, pour lesquels le 1,5-diaminonaphtalène (DAN)<sup>125,126</sup> et la 9-aminoacridine (9-AA)<sup>127</sup> figurent parmi les matrices les plus couramment utilisées<sup>128</sup>. Bien que ces matrices restent les plus populaires en imagerie MALDI, de nouvelles matrices ont été développées ces dernières années pour améliorer la qualité des spectres et, par conséquent, des images. C'est le cas de la matrice Norharmane<sup>129</sup> qui peut être utilisée en mode d'ions négatifs et positifs afin d'analyser les lipides et métabolites<sup>130,131</sup>. De plus, il est aussi possible d'utiliser des matrices ioniques, cela comprend une matrice organique à laquelle a été ajouté un liquide ionique. Dans de nombreux cas, les performances des matrices ioniques solides sont supérieures à celles de leurs analogues classiques. C'est le cas lors de l'ajout d'Aniline à HCCA, on observe une amélioration significative du signal et de la sensibilité<sup>132</sup>.

**Tableau 1** : Tableau non exhaustif des principales matrices utilisées en imagerie MALDI ainsi que leur formule chimique, leur poids moléculaires, leur polarité d'utilisation, les molécules détectables et la première référence bibliographique les mentionnant.

Nom	Formule chimique	Polarité	Poids moléculaire	Molécules détectées	Référence bibliographique
9-AA	$C_{13}H_{10}N_2$	Négatif	194.23	Métabolites/Lipides	Chandler, C.J. <i>et al.</i> 1988
DAN	$C_{10}H_6(NH_2)_2$	Les deux	158.20	Métabolites/Lipides	Juhász, P. <i>et al.</i> 1992 <sup>133</sup>
2,5-DHB	$(HO)_2C_6H_3CO_2H$	Les deux	154.12	Lipides	Strupat, K. <i>et al.</i> 1991 <sup>134</sup>
HCCA	$HOC_6H_4CH=C(CN)CO_2H$	Positif	189.17	Peptides	Beavis, R.C. <i>et al.</i> 1992 <sup>135</sup>
HCCA/ANI	$HCCA - C_6H_5NH_2$	Positif	282.30	Peptides	Lemaire <i>et al.</i> 2006 <sup>132</sup>
Norharmane	$C_{11}H_8N_2$	Les deux	168.17	Métabolites/Lipides	Yamagaki, T. <i>et al.</i> 1999 <sup>136</sup>
SA	$C_{11}H_{12}O_5$	Positif	224.21	Protéines	Beavis, R.C. <i>et al.</i> 1989 <sup>137</sup>

L'imagerie par spectrométrie de masse MALDI permet également d'analyser des tissus FFPE. Cependant, leur analyse moléculaire est un défi car le formaldéhyde réagit avec des groupes libres amine, donnant lieu à des bases de Schiff et à une réticulation supplémentaire par la formation de ponts méthylènes. Différentes stratégies ont été proposées pour débloquent et imager les protéines

par MALDI-MSI, ce qui a donné lieu à des approches protéomiques ou glycaniques complètes et hautement reproductibles<sup>138,139</sup>. Quant à l'analyse des lipides et métabolites, ils sont détectables après déparaffinisation et réhydratation<sup>140</sup> mais une perte du nombre de molécules détectées a été confirmée après ces différentes étapes de préparation d'échantillon<sup>141</sup>.

### L'imagerie WALDI

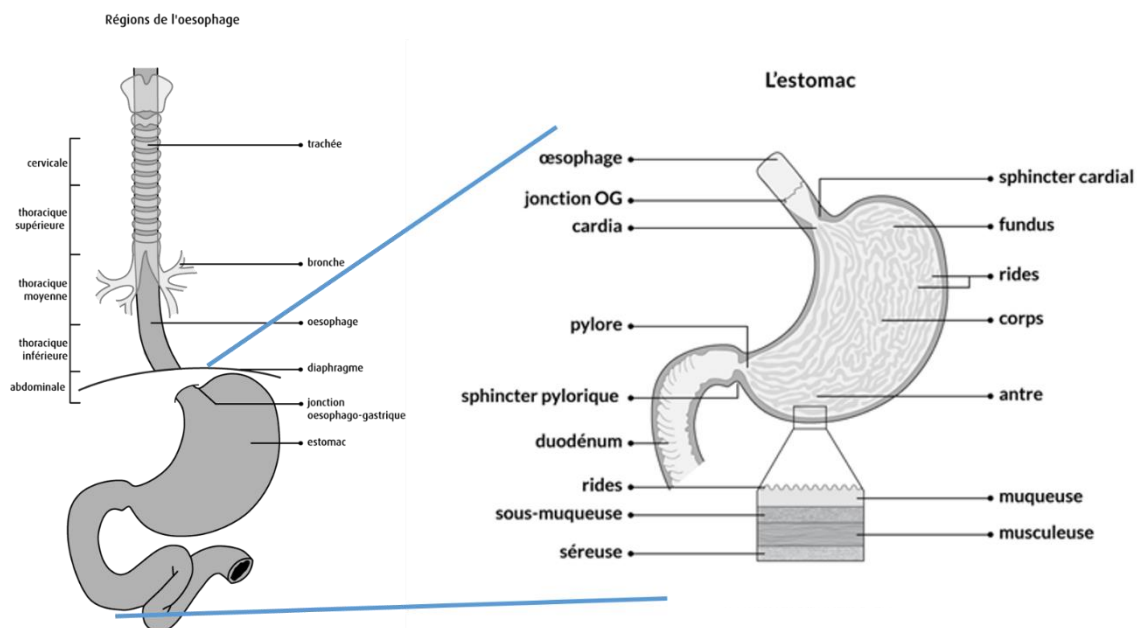
Récemment, la technologie SpiderMass s'est améliorée afin de permettre l'imagerie par spectrométrie de masse en condition ambiante de façon *in vivo*. En effet, la microsonde SpiderMass a été couplée à un bras robotisé à 6 axes de précision, avec une répétabilité de 5 µm et une précision de trajectoire inférieure à 100 µm dans les dimensions x, y et z. La pièce à main est fixée au bras robotique par un adaptateur "fait maison" imprimé grâce à une imprimante 3D<sup>18</sup>. Le bras robotisé et l'acquisition du spectromètre de masse sont contrôlés par une interface Matlab créée au laboratoire PRISM. Cela permettant de déclencher automatiquement et en même temps l'acquisition du spectromètre de masse ainsi que du laser. L'avantage ultime de l'imagerie WALDI en comparaison à l'imagerie MALDI est le fait qu'il n'y a aucune préparation d'échantillon. En effet, il n'y a pas d'ajout de matrice car l'eau endogène au tissu joue le rôle de la matrice, ce qui enlève de nombreux biais possibles lors de l'analyse 2D de tissus. Le seul désavantage de cette technique est la résolution spatiale qui n'est pour l'instant que de 250 µm par over-sampling. Cela étant dû aux capacités inhérentes au laser utilisé. Le but étant de pouvoir descendre jusqu'à 100 µm en améliorant la focalisation du laser dans un futur proche.

De plus, l'imagerie WALDI, grâce à l'ajout dans la pièce à main d'un capteur de distance, comprenant une caméra et une LED pour l'étalonnage de la hauteur z, permet aujourd'hui de faire aussi de l'imagerie en 3D *in vivo* par spectrométrie de masse. En effet, en plus des données moléculaires, les données topographiques du tissu analysé sont enregistrées et celles-ci sont co-enregistrées sur la matrice de l'image topographique, pixel par pixel. L'imagerie WALDI 3D a été testée sur un corps entier d'une souris fraîchement sacrifiée. Les images topographiques permettent de distinguer clairement différents organes tels que l'intestin, l'estomac, le foie, les poumons et le cœur. Pour la corrélation moléculaire, plusieurs zones plus petites ont été imagées, notamment la glande mammaire, le cœur, les poumons et le cerveau et permettent de localiser des ions spécifiques des différents organes. De plus, il a aussi été possible d'imager un doigt et une moule de façon *in vivo*<sup>18</sup>. Cela permet d'envisager de rendre ce système compatible avec l'utilisation dans le corps humain. Le système est envisagé pour une utilisation clinique future en chirurgie autonome. L'intégration du système d'imagerie développé avec une solution d'apprentissage profond permettrait de faciliter le processus de prise de décision pendant l'intervention chirurgicale.

## Cancer œsogastrique

### Anatomie

Le système digestif fait référence à l'ensemble des organes chargés de transformer les aliments, de les assimiler et de les absorber dans le corps. Après avoir été mâché dans la bouche, les aliments traversent différents organes par le biais du tube digestif. Les organes composant le système digestif humain comprennent la bouche, l'œsophage, l'estomac et les intestins (l'intestin grêle et le gros intestin) et le rectum. A chaque étape, ces organes réduisent les aliments en molécules de plus en plus petites qui sont ensuite transportées dans la circulation sanguine. Les résidus, sous forme de matières fécales, sont éliminés par l'anus. En plus des organes principaux impliqués dans la digestion et l'absorption des aliments, le système digestif est également composé d'organes accessoires qui interviennent à différentes étapes du processus. Ces organes comprennent le foie, la vésicule biliaire, le pancréas et les glandes salivaires. Le travail de thèse se concentrant sur le cancer œsogastrique, seule l'anatomie de l'estomac et de l'œsophage seront détaillées (**Figure 18**).



**Figure 18** : Anatomie de l'œsophage et de l'estomac avec un zoom à droite sur les différentes régions de l'estomac ainsi que de la paroi typique du tractus gastro-intestinal.

L'œsophage est une partie du tube digestif qui relie le pharynx au cardia de l'estomac et permet le passage des aliments de la bouche à l'estomac. Le point de contact entre celui-ci et l'estomac est appelé jonction œsophago-gastrique (JOG). Cette jonction se trouve près du diaphragme, qui est le muscle mince séparant la cavité thoracique de l'abdomen. L'œsophage est composé de trois sections qui correspondent aux trois régions anatomiques qu'il traverse : l'œsophage cervical, l'œsophage thoracique et l'œsophage abdominal. Sa longueur varie de 25 à 33 cm et sa partie la plus étroite mesure moins de 2.5 cm de large. Dans la partie supérieure, il est



positionné derrière la trachée tandis que dans la cavité thoracique, il est situé devant la colonne vertébrale.

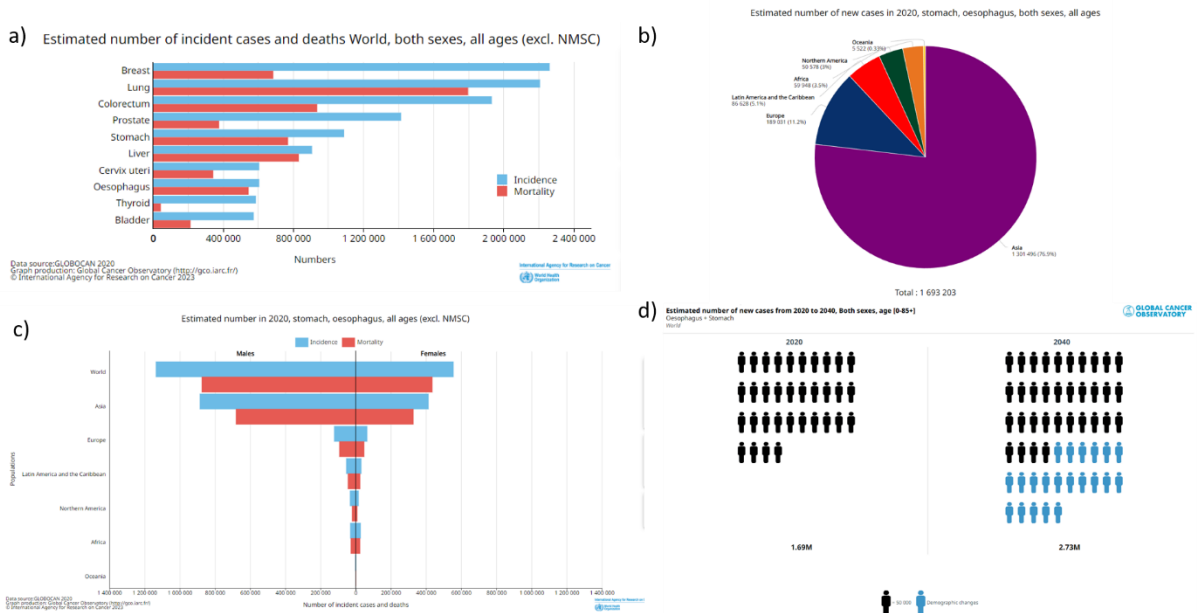
L'estomac est un organe en forme de réservoir qui se trouve dans la partie supérieure de l'abdomen, entre l'œsophage et le duodénum (la première partie de l'intestin grêle). Il se situe du côté gauche de la cavité abdominale, caché en partie par le foie et le diaphragme. La forme de celui-ci est dessinée par deux courbures principales ; la petite courbure sur la face médiale concave et la grande courbure qui correspond à la face latérale convexe. L'estomac est divisé en 5 régions. Tout d'abord, le cardia qui est situé dans la partie supérieure de l'estomac, près de l'œsophage. A cet endroit, se trouve le muscle sphincter cardia, qui empêche le reflux gastrique vers l'œsophage. La deuxième région est le fundus qui correspond à la région en forme de dôme située au-dessus du cardia, pouvant être temporairement un lieu de stock des aliments. Ensuite, le corps de l'estomac constitue la zone principale de l'organe et l'endroit où se décompose les aliments. La 4<sup>ème</sup> région est l'antrum pylorique qui se situe en-dessous du corps de l'estomac. Cette zone rétrécie sert de réservoir pour les aliments décomposés en attendant qu'ils soient déversés dans l'intestin. Enfin, la dernière région correspond au pylore qui permet la connexion entre l'estomac et l'intestin grêle. De forme conique, il est fermé par le muscle sphincter pylorique, qui contrôle la vidange du contenu de l'estomac vers l'intestin grêle.

La structure de la paroi de l'œsophage et de l'estomac suit le schéma typique retrouvé dans le tractus gastro-intestinal, comprenant une muqueuse, une sous-muqueuse, une couche musculaire (musculeuse) et une couche externe (adventice ou séreuse). La muqueuse constitue le revêtement interne des parois et est composée des éléments suivants : une fine couche d'épithélium à cellules pavimenteuses, une couche de tissu conjonctif appelée lamina propria et une mince couche de muscle connue sous le nom de musculaire muqueuse. La sous-muqueuse est une couche de tissu conjonctif qui entoure la muqueuse. Elle contient des glandes muqueuses, des vaisseaux sanguins, des nerfs et du tissu lymphatique. La musculature est la couche située immédiatement après la sous-muqueuse. Elle est composée d'un anneau interne de fibres musculaires circulaires et d'un anneau externe formé de longues fibres musculaires qui entourent la paroi de l'œsophage. L'adventice ou séreuse est une couche de tissu conjonctif qui soutient l'œsophage de manière lâche et qui le recouvre à l'extérieur dans la région du cou et du thorax.

### Épidémiologie

La seconde cause de mortalité dans le monde avec presque 19.3 millions de nouveaux cas et 10 millions de décès<sup>1</sup> en 2020 est le cancer. Sur presque 20 millions de nouveaux cas, la moitié se trouve en Asie, avec 59 % des décès, tandis que l'Europe et l'Amérique représentent seulement 23 % et 21 % des cas respectivement, dont 20 % et 14 % des décès mondiaux. Pour les deux sexes

confondus, les dix principaux types de cancer représentent plus de 60 % des cas de cancer nouvellement diagnostiqués et plus de 70 % des décès du cancer<sup>1</sup>. Le cancer du sein est le type de cancer le plus couramment diagnostiqué chez les femmes, représentant 11.7 % de tous les cas, suivi de près par le cancer du poumon (11.4 %), le cancer colorectal (10 %), le cancer de la prostate (7.3 %) et le cancer de l'estomac (5.6 %). Le cancer du poumon est la principale cause de décès liés au cancer, représentant 18,0 % du total des décès, suivi par le cancer colorectal (9.4 %), le cancer du foie (8.3 %), le cancer de l'estomac (7.7 %) et le cancer du sein chez les femmes (6.9 %). En Europe, le cancer représente plus de 4 millions de nouveaux cas en 2020 avec presque 2 millions de décès tandis qu'en France, cela représente environ 430000 cas par an pour plus de 180000 décès. Le cancer œsogastrique représente 1.7 millions de cas dans le monde avec 1.3 millions de décès et est présent dans le monde entier pour 77 % en Asie, 12 % en Europe, 8 % en Amérique, 4 % en Afrique et moins d'1 % en Océanie. Le cancer œsogastrique est constitué du cancer de l'estomac et de l'œsophage.



**Figure 19 :** a) Nombre de cas et de décès estimés pour les dix principaux cancers dans le monde sexe et âge confondus. Le cancer de l'estomac et de l'œsophage se classent respectivement 5<sup>ème</sup> et 8<sup>ème</sup>. b) Répartition des cas de cancer œsogastrique par région du monde en 2020 pour les deux sexes. c) Nombre de cas du cancer œsogastrique et de décès en 2020 selon le sexe et la région du monde. d) Estimation de nouveaux cas de cancer œsogastrique dans le monde d'ici 2040<sup>1</sup>.

Le cancer de l'estomac demeure une forme de cancer significative à l'échelle mondiale, en étant le 5<sup>ème</sup> plus fréquent, touchant plus d'1 million de personnes par an dans le monde. Son taux de mortalité est très élevé, environ 800000 décès ce qui représente un décès sur 13, le mettant au 4<sup>ème</sup> rang des causes de mortalité par cancer (**Figure 19**). Les taux de ce cancer sont deux fois plus élevés chez les hommes que chez les femmes. Il est le type de cancer le plus fréquemment diagnostiqué et la principale cause de décès par cancer chez les hommes dans plusieurs pays d'Asie centrale, tels que l'Iran, le Turkménistan et le Kirghizistan. Les taux d'incidence les plus élevés

sont observés en Asie de l'Est, notamment au Japon et en Mongolie qui présentent les taux les plus élevés chez les hommes et les femmes respectivement. On estime qu'en 2040, le cancer de l'estomac représentera plus de 1.77 millions de cas dont 1.27 millions de décès. En France, on estimait à 5013 le nombre de décès pour 7140 nouveaux cas de cancer gastrique en 2020, dont 65 % chez l'homme<sup>2</sup>.

Le cancer de l'œsophage, lui, occupe le 7<sup>ème</sup> rang en terme d'incidence avec plus de 600000 personnes par an avec un taux de mortalité de plus de 90 % (544076 décès en 2020) dont 418400 cas chez l'homme et 185800 chez la femme. Cela signifie qu'en 2020, environ 1 décès sur 18 serait dû au cancer de l'œsophage. Environ 70 % des cas surviennent chez les hommes, et les taux d'incidence et de mortalité sont deux à trois plus élevés chez les hommes que chez les femmes. Les pays en transition présentent des taux plus élevés que les pays développés chez les hommes mais les taux sont comparables pour les femmes. L'Asie de l'Est présente les taux d'incidence régionaux les plus élevés, en raison notamment de la charge importante de la Chine, suivie de l'Afrique du Sud, de l'Est et de l'Europe du Nord. Le Cap-Vert détient le taux d'incidence le plus élevé au monde chez les hommes, tandis que le Malawi détient ce titre chez les femmes. On estime une augmentation des cas jusqu'à presque 1 million dont 880000 décès d'ici 2040. En France, son nombre de décès est estimé à 4274 pour 5310 nouveaux cas de cancer en 2020 dont 76 % chez l'homme<sup>2</sup>.

Le pronostic de ce cancer est pessimiste en raison d'un diagnostic tardif et d'un état de santé souvent défavorable, avec dans 12 à 17 % des cas, un cancer des voies respiratoires associé. Selon les données de l'Institut National du Cancer basées sur des patients diagnostiqués entre 2010 et 2015<sup>142</sup>, la survie nette standardisée à 5 ans (la survie qui serait observée si le cancer était la seule cause de décès) est de 16 % chez les hommes et de 20 % chez les femmes. Bien que cette survie nette à 5 ans reste faible, elle a connu une amélioration, doublant entre 1990 et 2015, passant de 9 % à 18 %. Les taux de survie observés et la survie nette à 5 ans sont très similaires (respectivement 14 % et 16 %), ce qui suggère que les personnes atteintes d'un cancer de l'œsophage décèdent principalement de leur cancer dans les 5 ans suivant leur diagnostic.

## Diagnostic

### *Cancer de l'estomac*

Le diagnostic de nos jours du cancer de l'estomac se divise en 4 parties : le bilan diagnostic et pré-thérapeutique, le bilan d'extension, la recherche d'une prédisposition familiale et la recherche d'infection à *Helicobacter Pylori*<sup>143</sup>.

Le diagnostic repose sur l'utilisation d'une technique appelée gastroscopie, qui consiste en une endoscopie digestive haute avec biopsies, c'est l'insertion d'un tube (endoscope) muni d'une caméra à travers la bouche ou le nez pour examiner l'estomac. Cette procédure permet d'observer la lésion et d'évaluer son étendue. Des échantillons tissulaires, appelés biopsies, sont prélevés à partir des zones suspectes lors de la gastroscopie. L'analyse de ces biopsies par l'anatomo-pathologiste permet de confirmer si la lésion est cancéreuse ou non, ainsi que la classification histologique potentielle de la tumeur. Le bilan pré-thérapeutique est essentiel pour identifier des conditions telles qu'un état de santé général fragile, d'autres maladies ou des situations qui contre-indiquent la chirurgie ou d'autres traitements. Cette évaluation clinique et biologique est nécessaire avant toute prise en charge thérapeutique. Plusieurs parties composent ce bilan pré-thérapeutique comme un bilan nutritionnel, des analyses sanguines (globules rouges, globules blancs et plaquettes), un bilan biochimique rénal, un bilan hépatique, une évaluation cardiologique ainsi qu'un bilan pulmonaire afin d'évaluer la fonction respiratoire en cas de proposition de traitement chirurgical.

Une fois le diagnostic de cancer de l'estomac confirmé, il est nécessaire d'évaluer l'étendue de la propagation du cancer dans le reste du corps. Cela permet de déterminer le traitement le plus approprié pour le patient. Cette évaluation est appelée bilan d'extension et peut être faite sous différentes formes. Tout d'abord, un scanner thoraco-abdomino-pelvien est réalisé en administrant un agent de contraste afin d'évaluer la résecabilité de la tumeur et de recherche d'éventuelles métastases à distance. Il est possible aussi dans certains cas de faire une écho-endoscopie digestive haute afin d'évaluer l'extension des lésions sur l'œsophage, le pylore et le duodénum ou dans l'épaisseur de la paroi gastrique et au niveau des ganglions lymphatiques. Enfin, des analyses immunohistochimiques peuvent être effectuées afin d'obtenir les statuts MSS/MSI et HER2. Il est crucial d'identifier le statut MSI dès le diagnostic car les patients présentant une tumeur MSI ne tirent pas d'avantage de la chimiothérapie peropératoire.

La recherche d'une prédisposition familiale est un point fort du diagnostic du cancer gastrique, surtout chez les patients de moins de 40 ans. En effet, les deux principales formes héréditaires de cancers gastriques connues correspondent au syndrome de Lynch et aux formes associées aux mutations du gène CDH1. La reconnaissance de la forme génétique du cancer gastrique est essentielle car les modalités de prises en charge et de dépistage peuvent alors être adaptées. Une autre recherche est aussi très importante lors du diagnostic, c'est la recherche d'une infection à *Helicobacter Pylori* possible par sérologie ainsi que par un test respiratoire. En effet, la détection précoce suivie de l'élimination de la bactérie *Helicobacter Pylori* représente une stratégie

préventive efficace contre le cancer de l'estomac, en particulier lorsqu'elle est réalisée avant l'apparition de lésions gastriques précancéreuses.

Grâce aux différentes étapes du diagnostic, le stade et le grade de la tumeur peuvent être déterminés. En effet, les cancers gastriques, tout comme d'autres maladies tumorales, sont classés en fonction de leur stade d'avancement. Ce classement est réalisé selon le système TNM<sup>5</sup> qui décrit la progression locale de la tumeur (T), son atteinte aux ganglions lymphatiques environnants (N) et sa dissémination éventuelle sous forme de métastases (M). La tumeur est caractérisée par une valeur qui indique l'envahissement de la paroi de l'estomac. Cela peut varier de T1, lorsque seule la muqueuse de la paroi est envahie, à T4 lorsque la tumeur envahit la paroi jusqu'à la couche séreuse et atteint les tissus ou les organes adjacents. Le nombre de ganglions lymphatiques régionaux atteints est également pris en compte : le paramètre N varie de N0 (absence de ganglions atteints) à N3 lorsque plus de six ganglions sont touchés. Tandis que l'information sur les métastases est binaire avec M0 pour l'absence de métastases et M1 pour la présence de métastases. Ces critères permettent d'évaluer le stade de la maladie : les cancers de stade I sont les moins avancées (T1 et T, N0 ou N1, M0) et sont généralement plus facilement traitables. En revanche, le stade IV est le plus difficile à traiter car il implique la présence de métastases. Les stades II et III correspondent à des cancers de gravité intermédiaire. Le pronostic de la maladie dépend à la fois du stade et de sa localisation.

### *Cancer de l'œsophage*

Le cancer de l'œsophage peut se former dans différentes parties de cet organe et provient généralement des cellules de la muqueuse. Deux types principaux de cancer se distinguent : les carcinomes épidermoïdes et les adénocarcinomes. Les carcinomes épidermoïdes sont les plus fréquents et se développent principalement dans le tiers moyen et supérieur de l'œsophage à partir de cellules de l'épithélium. Les adénocarcinomes, quant à eux, se forment à partir des cellules glandulaires de la muqueuse, principalement situées dans la partie inférieure de l'œsophage<sup>144</sup>.

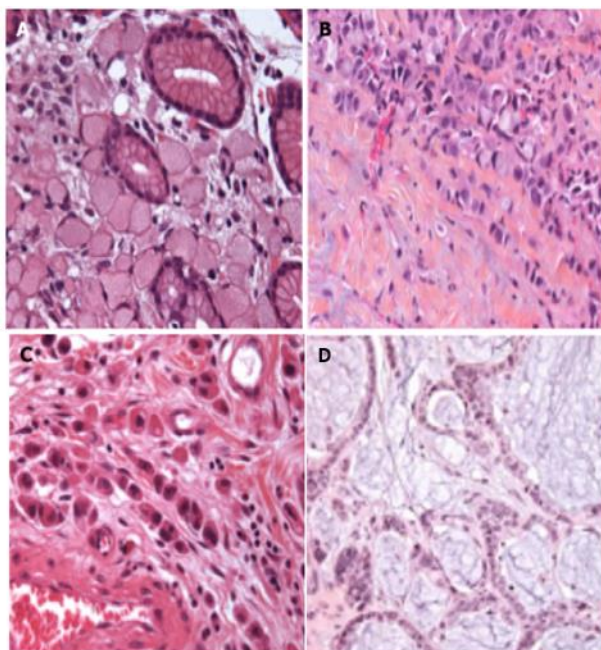
Le diagnostic du cancer de l'œsophage est très similaire à celui du cancer de l'estomac. En effet, la première étape est aussi de faire une endoscopie œsogastrique avec biopsies, cela afin de préciser la localisation tumorale, sa morphologie et son pourcentage de circonférence déjà atteint. Il est important d'évaluer le débord œsophagien et gastrique de la tumeur en centimètres afin de pouvoir guider le geste chirurgical. L'analyse des biopsies par l'anatomo-pathologiste permet de préciser l'invasivité de la lésion.

Lorsque la présence d'une tumeur cancéreuse est confirmée, il faut bien sûr faire un bilan d'extension pour obtenir de plus amples détails sur la tumeur afin de déterminer le traitement le

plus approprié. Cela peut être fait grâce à un scanner thoraco-abdominal ainsi que par une échographie œsophagienne si le cancer est non sténosant et non métastatique. Le scanner est effectué afin d'évaluer jusqu'où le cancer s'est propagé aux tissus voisins (trachée, bronches, aorte...) et donc de définir son stade. Le scanner peut parfois être associé à une tomographie par émissions de positons (PET) afin de compléter l'évaluation. Également connu sous le nom de PET-scan, cet examen indolore permet de générer des images en coupe de tout le corps après l'injection d'un traceur faiblement radioactif dans le sang sachant que ce traceur a la particularité de se fixer sur les cellules cancéreuses. L'écho-endoscopie œsophagienne est réalisée à l'aide d'un écho-endoscope radial et permet de préciser le stade de la tumeur selon son extension dans les différentes parties de la paroi de l'œsophage. Il est aussi important de faire une évaluation nutritionnelle et de la sévérité de la dénutrition afin de juger de la sévérité du cancer. Cette évaluation repose exclusivement sur des critères cliniques tels que l'indice de masse corporelle, la perte de poids par rapport au poids habituel avant le début de la maladie ainsi que la réduction de la prise alimentaire et des critères biologiques comme l'albuminémie et la pré-albuminémie. Le même système international TNM est utilisé pour déterminer le stade du cancer.

### Classification histologique

Les classifications les plus couramment utilisées pour le cancer œsogastrique sont celles de Lauren et de l'OMS 2019 (5<sup>ème</sup> édition). La classification de Lauren, assez ancienne, divise les cancers œsogastriques en 3 groupes : les tissus intestinaux, les tissus diffus (incluant les carcinomes à cellules isolées) et les tissus mixtes (approximativement 50 % de chaque type). La plus récente



**Figure 20** : Différents types de tissus de cancer œsogastrique. a) PCC-SRC. b) PCC avec des cellules SRC (<10 %). c) PCC-NOS. d) ADK mucineux.

étant celle de l'OMS selon laquelle le cancer œsogastrique ou adénocarcinome (ADK) peut se diviser en plusieurs groupes; l'ADK tubuleux, papillaire, mucineux, à cellules peu cohésives (PCC), mixte, micro papillaire et carcinome avec stroma lymphoïde (médullaire)<sup>145</sup>. L'adénocarcinome est un cancer dont la structure microscopique est de type glandulaire. L'ADK mixte doit contenir au moins 2 groupes distincts, l'un glandulaire et l'autre à cellules peu cohésives. Le type de cancer PCC peut être constitué de cellules présentant une forme de « chevalière » (PCC-SRC) ou de type PCC-NOS (non spécifié autrement).



La définition de la catégorie PCC selon l'OMS a évolué au fil des différentes éditions publiées de la classification de l'OMS. Dans la première édition publiée en 1977, la catégorie PCC-SRC était considérée comme un sous-type distinct d'EC et était définie comme « une tumeur contenant plus de 50 % de cellules malignes isolées ou de petits groupes de cellules malignes contenant de la mucine intracytoplasmique ». Cependant, dans la 4<sup>ème</sup> édition en 2010, la catégorie PCC-SRC a été redéfinie comme un sous-type de PCC, qui est composé de cellules néoplasiques isolées ou disposées en petits agrégats sans glandes bien formées<sup>146</sup>. Les SRC sont caractérisés par une gouttelette centrale de mucine cytoplasmique optiquement claire avec un noyau excentré. Les autres types cellulaires qui ne répondent pas à cette définition sont définis comme des PCC non spécifiés. En 2017, un groupe d'experts multidisciplinaire s'est réuni pour clarifier la définition pathologique des PCC. Ils ont proposé que seuls les PCC dont plus de 90 % des cellules présentent une morphologie SRC soient classés comme SRC, tandis que les autres catégories comprenant les PCC avec une composante SRC (moins de 90 % mais plus de 10 % de SRC) et les PCC-NOS (moins de 10 % de SRC)<sup>147</sup> (**Figure 20**). Cette nouvelle classification permet de prendre en compte l'expression du degré de différenciation du PCC en fonction de l'étendue du SRC dans la tumeur. Cette définition consensuelle est importante car elle permettra de standardiser les résultats des études futures et de faciliter la comparaison entre les études, évitant ainsi l'hétérogénéité qui a caractérisé les études sur les PCC-SRC au cours des dernières décennies. Il est aussi possible de déterminer les degrés de différenciation allant de non évaluable à bien différencié en passant par peu ou moyennement indifférencié. Concernant la JOG, la classification se fait selon sa localisation endoscopique, c'est la classification de Siewert. Son type est défini en fonction de la distance entre le centre de la tumeur et la ligne Z. En effet, le type Siewert I se situe entre -5 et -1 cm, le type Siewert II entre -1 et 2 cm et le type Siewert III entre 2 et 5 cm. Sachant que les tumeurs de la jonction œsogastrique sont classées comme des cancers de l'œsophage si leur point central se situe dans les 2 cm au-dessus du cardia (Siewert I et II). Si leur point central se trouve dans l'estomac à plus de 2 cm de la JOG ou s'ils sont localisés à moins de 5 cm de la JOG mais sans extension vers l'œsophage, ils sont considérés comme des tumeurs de l'estomac.

D'un point de vue moléculaire, l'EC a été classé en quatre sous-types génomiques dans le cadre d'un projet historique de l'Atlas du génome du cancer<sup>8</sup>. Ces quatre sous-types comprennent: le sous-type du virus d'Epstein-Barr (EBV) (9 %) caractérisé par une hypermétrie extrême de l'ADN, des mutations PIK3CA récurrentes et d'une amplification de JAK2, du ligand 1 de la mort programmée (PD-L1) et de PD-L2 ; le sous-type d'instabilité microsatellite (21 %) contenant des mutations dans des gènes codant pour des protéines de signalisation oncogéniques et associé à une issue oncologique plus favorable ; le sous-type génomiquement stable (GS) (20 %) dans lequel la plupart des PCC sont classés ; et le sous-type d'instabilité chromosomique (CIN) (50 %) associé à

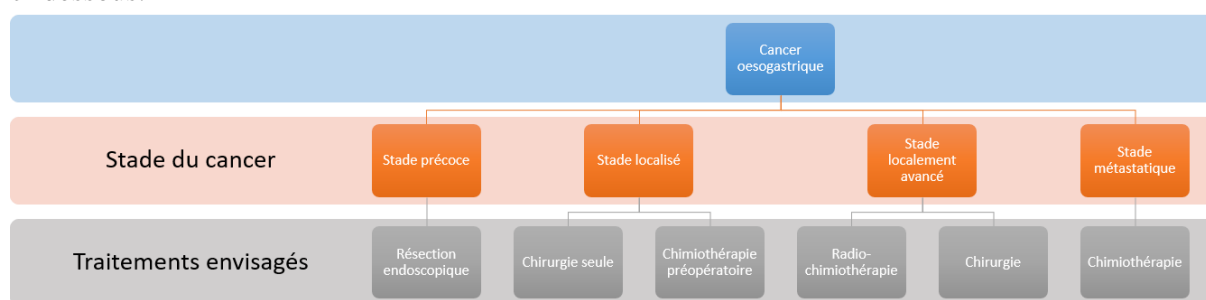
l'aneuploïdie et à l'amplification des gènes impliqués dans la signalisation des récepteurs tyrosine kinase/RAS/MAPK<sup>148</sup>. Plus récemment, une autre analyse moléculaire de l'EC a identifié quatre sous-groupes de tumeurs associés à des résultats cliniques distincts : un type mésenchymateux comprenant les tumeurs de type diffus et la plupart des tumeurs PCC ; un sous-type d'instabilité microsatellite caractérisé par de nombreuses mutations et un meilleur pronostic ; un sous-type actif de la protéine tumorale 53 (TP53) associé à des taux plus élevés d'infection par l'EBV et enfin un sous-type TP53-inactif similaire au sous-groupe CIN<sup>9</sup>.

Ces classifications moléculaires sont très importantes car elles fournissent une feuille de route pour la stratification des patients. Outre l'impact pronostique, il a été prouvé que ces sous-types génomiques sont associés à des caractéristiques distinctes en ce qui concerne la réponse tumorale. Cependant, il ne faut pas oublier que le cancer œsogastrique est constitué de tumeurs hétérogènes et que plusieurs composants histologiques et moléculaires peuvent être présents dans la même tumeur et peuvent être modifiés par les thérapies ciblées<sup>144</sup>. En outre, il n'existe pas de corrélation stricte entre les types histologiques et les sous-types moléculaires. Les PCC sont principalement des GS mais peuvent également être de type d'instabilité microsatellite ou de type EBV, avec des implications thérapeutiques potentielles puisque ces deux sous-types moléculaires sont associés à la réponse aux inhibiteurs du point de contrôle immunitaire<sup>149</sup>.

## Traitement

### *Cancer de l'œsophage*

Les traitements du cancer de l'œsophage diffèrent en fonction des spécificités de chaque tumeur, notamment de son type histologique ainsi que de son stade, comme le montre la **Figure 21** ci-dessous.



**Figure 21** : Traitements envisagés selon le stade du cancer de l'œsophage.

La chirurgie effectuée pour traiter le cancer de l'œsophage est appelée œsophagectomie. Selon la localisation et le stade de la tumeur, cette procédure consiste à enlever tout ou une partie de l'œsophage. Pendant l'opération, les ganglions lymphatiques proches de l'œsophage sont également retirés dans le cadre d'un curage ganglionnaire. De plus, la partie supérieure de l'estomac est généralement retirée simultanément. L'objectif de cette intervention chirurgicale est de retirer complètement la tumeur ainsi qu'une marge de tissu environnant. Il est essentiel que cette



marge de tissu ne soit pas atteinte par des cellules cancéreuses. Une reconstruction chirurgicale est réalisée en même temps afin de rétablir la continuité du tube digestif. Celle-ci est le plus souvent réalisée à l'aide de l'estomac, mais parfois également avec le côlon ou l'intestin grêle. Selon la localisation de la tumeur, une œsophagectomie totale ou subtotale sera réalisée.

La chimiothérapie conventionnelle est un traitement systémique qui utilise des médicaments anti-cancéreux administrés dans tout le corps. Cela permet d'atteindre les cellules cancéreuses, peu importe leur localisation, y compris celles qui pourraient être isolées et non détectées lors du diagnostic. Ces médicaments agissent en détruisant les cellules cancéreuses en ciblant leurs mécanismes de division. Les principaux médicaments de chimiothérapie classique utilisés dans le traitement du cancer de l'œsophage sont le 5-fluorouracile, le cisplatine et l'épirubicine. Dans certains cas d'adénocarcinomes métastatiques de la JOG présentant une surexpression de HER2, un traitement combinant la chimiothérapie et une thérapie ciblée peut être proposé. Celle-ci utilise le trastuzumab qui est un anticorps monoclonal. Cependant, la chimiothérapie présente de nombreux effets indésirables comme les nausées et vomissements, les diarrhées ainsi que la chute de cheveux.

La radiothérapie est un traitement localisé basé sur l'utilisation de rayonnements ionisants, générés par un accélérateur de particules, pour empêcher la multiplication des cellules cancéreuses. Son objectif est de cibler précisément la tumeur tout en préservant autant que possible les tissus sains et les organes voisins. Généralement, la radiothérapie est combinée avec la chimiothérapie, formant ainsi une approche de radio-chimiothérapie concomitante. Les médicaments de chimiothérapie sont utilisés pour accroître la sensibilité des cellules cancéreuses aux rayonnements. Ces médicaments, appelés radio-sensibilisants, renforcent l'efficacité respective de chaque traitement lorsqu'ils sont utilisés en association.

### *Cancer de l'estomac*

Tout comme le cancer de l'œsophage, plusieurs traitements peuvent être utilisés, seuls ou en combinaison, pour traiter un cancer de l'estomac ; la chirurgie, la chimiothérapie, la radiothérapie et plus rarement le traitement par endoscopie.

La chirurgie est le traitement standard pour les stades localisés du cancer et peut être faite aussi pour les stades localement avancés en la combinant à une chimiothérapie avant ou après l'opération. La chirurgie pratiquée pour traiter le cancer de l'estomac est appelée gastrectomie. Selon la localisation et le stade de la tumeur, le chirurgien peut effectuer une ablation totale de l'estomac (gastrectomie totale) ou une ablation partielle (gastrectomie partielle ou subtotale). Il est possible aussi de faire un curage ganglionnaire afin de retirer les ganglions lymphatiques à

proximité de l'estomac. L'objectif de l'intervention chirurgicale est de procéder à l'exérèse complète de la tumeur en prenant une marge de sécurité autour de celle-ci. Il est essentiel que celle-ci soit exempte de cellules cancéreuses (on parle de marges saines) afin d'éviter la rechute du patient. Les marges nécessaires varient en fonction du type de carcinome et de la localisation de la tumeur. Pour un adénocarcinome bien différencié, une marge de 5 cm est requise pour assurer une exérèse sans résidu cancéreux. En revanche, pour une tumeur de type PCC, les marges nécessaires sont de 8 à 10 cm. Il est recommandé de demander un examen extemporané en cas de doute.

Les principaux médicaments de chimiothérapie utilisés dans le traitement du cancer de l'estomac sont le 5-fluorouracile, le cisplatine, l'épirubicine, le docetaxel et l'oxaliplatine. La méthode couramment employée en radiothérapie externe pour le traitement du cancer de l'estomac est la radiothérapie conformationnelle en 3D. Cette technique vise à ajuster de manière précise le volume sur lequel les rayons seront dirigés afin qu'il corresponde au mieux à la zone à traiter.

De plus, il existe de nombreux essais cliniques dédiés au cancer œsogastrique qui ont pour but d'améliorer les soins et les traitements des patients, quel que soit le stade de la maladie. Ils offrent la possibilité d'évaluer l'efficacité et la tolérance de nouvelles thérapies avant de les rendre disponibles à tous les patients concernés. On compte parmi ces essais cliniques, l'étude IMHOTEP de phase 2, évaluant l'efficacité d'un traitement par immunothérapie, chez des patients ayant un carcinome non résecable non traité avec des MSI et une défaillance du système de réparation de l'ADN (MSI-dMMR) ou un cancer gastrique EBV+. Un autre essai clinique FIDES s'intéresse au dérazantinib seul ou en association avec le paclitaxel, le ramucirumab ou l'atezolizumab dans l'adénocarcinome gastrique. Ces deux essais cliniques prennent place au Centre Georges François Leclerc à Dijon.

## **Lipidomique**

La lipidomique est une étude à grande échelle des réseaux et des voies des lipides cellulaires dans les systèmes biologiques<sup>150</sup>. Le terme « lipidome » est utilisé pour décrire le profil lipidique complet d'une cellule, d'un tissu, d'un organisme ou d'un écosystème. La lipidomique est un domaine de recherche relativement récent qui utilise des technologies avancées telles que la MS, la spectroscopie de résonance magnétique nucléaire ainsi que la spectroscopie de fluorescence afin d'étudier le rôle des lipides dans de nombreuses maladies.

### **Les lipides, une superfamille**

Les lipides (du grec *lipos*, graisse) font partis des quatre principales biomolécules, avec les glucides, les protéines et les acides nucléiques. Ils sont une classe diversifiée et complexe de molécules, toutes étant caractérisées par une propriété physico-chimique, la solubilité<sup>151</sup>. En effet,

les lipides sont hydrophobes mais solubles dans des solvants non polaires. Ils sont présents naturellement dans la plupart des plantes, animaux et microorganismes. Les lipides se distinguent des autres métabolites cellulaires par leur nature grasseuse et sont la seule source de matière grasse dans les organismes vivants. Ce sont des petites molécules principalement composées de carbone, d'oxygène et d'hydrogène.

Bien que les lipides soient un groupe de molécules aux structures complexes et aux fonctions variées dans le monde du vivant, l'IUPAC a proposé une classification des lipides en 1978 basée sur la comparaison des squelettes hydrocarbonés des différentes molécules lipidiques<sup>152</sup>. Cette nomenclature a été mise à jour régulièrement en fonction des progrès technologiques dans l'analyse du lipidome. Classiquement, les lipides sont composés de deux grands groupes, lipides simples et lipides complexes. La différence étant que les lipides simples sont uniquement composés d'atomes de carbone, d'hydrogène et d'oxygène alors que les lipides complexes, en plus des trois éléments précédents, peuvent être constitués de phosphore, azote et/ou soufre<sup>153</sup>. A l'intérieur de ces deux groupes, un classement peut être fait en classes et sous-classes en fonction de critères de structure chimique tels que la structure de base de la molécule lipidique et les molécules annexes liées directement ou indirectement à cette structure<sup>154</sup> (**Tableau 2**).

**Tableau 2** : Liste non exhaustive des classes et sous-classes existantes de lipides avec un exemple de structure chimique pour chaque classe.

Classes	Sous-classes	Exemple
<b>Lipides simples</b>		
<b>Acides gras</b>	Acides gras saturés Acides gras insaturés Acides gras polyinsaturés	Acide arachidonique 
<b>Glycolipides</b>	Monoglycérides Diglycérides Triglycérides	DG 40:2 
<b>Stérides</b>	Cholestéryl esters	Cholestéryl linoléate 
<b>Cérides</b>		Palmitate de cétyle 
<b>Lipides complexes</b>		
<b>Glycérophospholipides</b>	Acides phosphatidiques Phosphatidylglycérols Phosphatidylcholines Phosphatidylinositols Phosphatidylsérines Phosphatidyléthanolamines	PI 36:3 
<b>Sphingolipides</b>	Céramides Phosphosphingolipides Glycosphingolipides	Sphinganine 
<b>Glycolipides</b>	Glycoglycérolipides Glycosphingolipides Glycophosphatidylinositols	Di-octadécadiénol-galactosyl-glycérol 

Les lipides simples sont donc composés des acides gras, des glycérolipides, des stérides ainsi que des cérides tandis que les lipides complexes sont divisés en 3 classes ; les glycérophospholipides (GPL), les sphingolipides et les glycolipides, tous eux-mêmes constitués de sous-classes<sup>153</sup>. Dans ce manuscrit, seuls les acides gras, les glycérolipides, les GPLs et les sphingolipides seront décrits. En effet, ces classes sont les principaux lipides observés grâce aux différentes technologies de spectrométrie de masse utilisées.

### *Les acides gras*

Les acides gras (AG) sont d'abord produits dans le cytoplasme des cellules à partir d'un précurseur appelé acétyl-coenzyme A, qui est libéré soit par la mitochondrie dans le cycle de Krebs soit par la voie de dégradation des acides gras et des acides aminés céto-gènes comme la lysine et la leucine. Pour produire les acides gras, la cellule a besoin de divers cofacteurs et enzymes pour carboxyler l'acétyl-coenzyme A en malonyl-coenzyme A. Les résidus de malonyl sont ajoutés successivement lors de cycles répétitifs, ce qui permet la synthèse cytosolique des acides gras jusqu'à l'acide palmitique qui contient 14 atomes de carbone. Ensuite l'extension de la chaîne se produit dans la mitochondrie de la cellule où l'acétyl-coenzyme A sert de brique élémentaire pour l'élongation.

Les AGs sont caractérisés par une répétition d'un groupe méthyle  $-CH_2-$ , dont la forme générique est  $CH_3-(C_xH_y)_n-COOH$ . Cette répétition forme une chaîne hydrocarbonée contenant généralement un nombre pair d'atomes de carbone (4 à 36 atomes), ce qui confère l'hydrophobicité aux acides gras, et se terminent par une fonction carboxylique ionisable et hydrophile. Leur hydrophobicité augmente avec la longueur de la chaîne carbonée. Les AGs peuvent être divisés en treize classes dont les AGs saturés, leurs dérivés insaturés, les eïcosanoïdes, les esters ou encore les amides d'acides gras. La nomenclature des AGs se fait grâce au radical alkyle correspondant, à la structure de la chaîne carbonée (nature des liaisons, nombre, position et configuration des doubles/triples liaisons) et de la nature de la fonction (acide dans notre cas). Par exemple, l'acide *5cis, 8cis, 11cis, 14cis* – eïcosatétraènoïque est un acide (terminaison « oïque ») à 20 atomes de carbone (eïcosa), à 4 (tétra) doubles liaisons (ène) positionnées sur les carbones 5, 8, 11 et 14 de configuration géométrique *cis*. Cet acide gras étant aussi appelé l'acide arachidonique dû à sa forte présence dans les arachides. Cette nomenclature triviale selon laquelle les noms des AGs sont donnés par rapport au produit où celui-ci a été découvert/produit ou celui où il est présent en grande quantité.

### *Les glycérolipides*

Les glycérolipides sont de grosses molécules lipidiques produites dans le réticulum endoplasmique lisse des cellules eucaryotes, puis transportées vers toutes les membranes de la

cellule. En effet, les AGs forment des liaisons ester avec les groupes hydroxyles de certaines molécules organiques pour créer des lipides simples, dont les plus courants sont les glycérolipides. Leur synthèse se produit par l'estérification d'un ou deux acides gras avec un radical phosphoglycérol, formant ainsi un acide phosphatidique qui se décompose ensuite par hydrolyse en mono (MGs) ou diglycérides (DGs). Pour former un triglycéride (TG), un troisième acide gras est estérifié sur la fonction alcool primaire du DG<sup>155</sup>. Les glycérolipides ont un rôle important dans la structure membranaire et dans la fluidité de la membrane plasmique, qui dépend de leur composition en acide gras. En plus de leur rôle structural, ils ont également un rôle énergétique ainsi qu'un rôle dans le métabolisme cellulaire.

Plus précisément, les MGs sont des composés produits lors de la dégradation partielle ou complète des TGs. En particulier, les 2-monoglycérides sont formés lorsque les lipases pancréatiques agissent sur les triglycérides dans le duodénum. Ils sont essentiels pour maintenir l'équilibre entre les lipides et l'eau et présentent une absorption intestinale maximale. De plus, dû à leurs propriétés tensio-actives, leur accumulation dans la cellule peut être dangereuse pour son intégrité membranaire. Quant aux diglycérides, bien qu'étant présents en minorité dans les cellules, leur fonction biologique est très importante. En effet, ils sont des intermédiaires dans le métabolisme cellulaire des TGs et de certains phospholipides<sup>155</sup>. Pour finir, le rôle principal des triglycérides est de servir de source d'énergie ou de se transformer en graisse de stockage (tissu adipeux) pour alimenter le métabolisme énergétique.

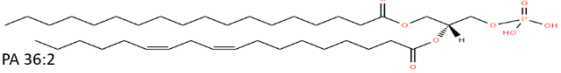
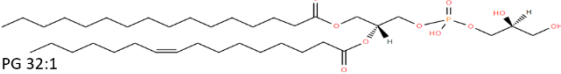
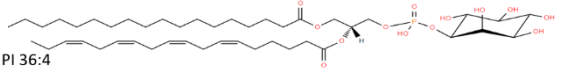
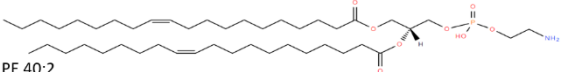
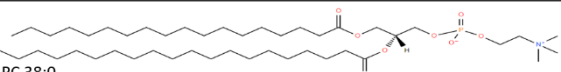
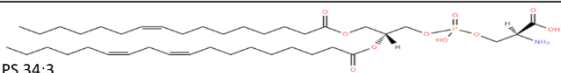
### *Les glycérophospholipides*

La synthèse des glycérophospholipides se fait dans le réticulum endoplasmique lisse. Les GPLs font partie des lipides complexes, ils sont composés de glycérol, d'acides gras et d'acide phosphorique. C'est une famille étendue de composés qui ont en commun une molécule de glycérol avec des acides gras estérifiés ainsi que de l'acide phosphorique. La structure de base, appelée phosphatidyle, donne lieu à plusieurs groupes de phospholipides complexes en raison de l'estérification possible de l'acide phosphorique avec différentes molécules contenant un groupe hydroxyle. Ce sont des molécules qui ont une structure bipolaire, avec un pôle hydrophile et un pôle hydrophobe, ce qui les rend amphiphiles. Cette amphiphilie est importante car elle permet aux phospholipides de se placer aux interfaces de milieux de polarités différentes, ce qui favorise la formation de structures orientées feuilletées ou micellaires. C'est pour cela que ce sont les lipides les plus représentés dans les systèmes membranaires et ils renferment dans leurs structures des composés ayant des propriétés de messagers biologiques secondaires.

Les phospholipides les plus fréquemment rencontrés sont décrits et classés en 6 groupes, les acides phosphatidiques (PA), les phosphatidylglycérols (PG), les phosphatidylinositols (PI),

les phosphatidyléthanolamines (PE), les phosphatidylcholines (PC) et les phosphatidylsérines (PS), dans le **Tableau 3** ci-dessous.

**Tableau 3** : Liste exhaustive des 6 groupes de glycérophospholipides avec leur composition et un exemple de structure chimique pour chaque groupe.

Phospholipides	Composition	Exemple
Acides phosphatidiques (PA)	Glycérol + AG + phosphate	 PA 36:2
Phosphatidylglycérols (PG)	Glycérol + AG + phosphate + <b>glycérol</b>	 PG 32:1
Phosphatidylinositols (PI)	Glycérol + AG + phosphate + <b>inositol</b>	 PI 36:4
Phosphatidyléthanolamines (PE)	Glycérol + AG + phosphate + <b>éthanolamine</b>	 PE 40:2
Phosphatidylcholines (PC)	Glycérol + AG + phosphate + <b>choline</b>	 PC 38:0
Phosphatidylsérines (PS)	Glycérol + AG + phosphate + <b>sérine</b>	 PS 34:3

Les acides phosphatidiques ont tous la même structure de base appelée phosphatidyle, mais varient par la nature des acides gras attachés à la molécule de glycérol. Ces molécules sont acides et ont une charge négative qui nécessite la présence d'un contre-ion.

Les phosphatidylglycérols, sont constitués du squelette phosphatidyle mais avec un résidu phosphate estérifié par un autre glycérol. Ces lipides portent deux charges négatives.

Le myoinositol est l'alcool qui estérifie le groupe carboxyle de l'acide phosphorique lors de la création d'un phosphatidylinositol. Ce composé est un isomère optiquement inactif de l'inositol, qui est le plus répandu dans la nature, et est présent en grande quantité dans le muscle, d'où son nom. Ces lipides sont rapidement dégradés dans la cellule pour produire des DGs et du phosphate d'inositol. Ces produits étant utilisés pour des processus de régulation. Par exemple un PI qui est impliqué dans le processus de signalisation cellulaire. Lorsqu'il est clivé par une protéine kinase C, il libère l'inositol triphosphate IP3.

Pour les phosphatidyléthanolamines, l'acide phosphorique est estérifié par l'éthanolamine, une molécule d'éthanol contenant un groupe amine. Ils sont parmi les plus abondants phospholipides et ont un grand rôle dans la constitution membranaire.

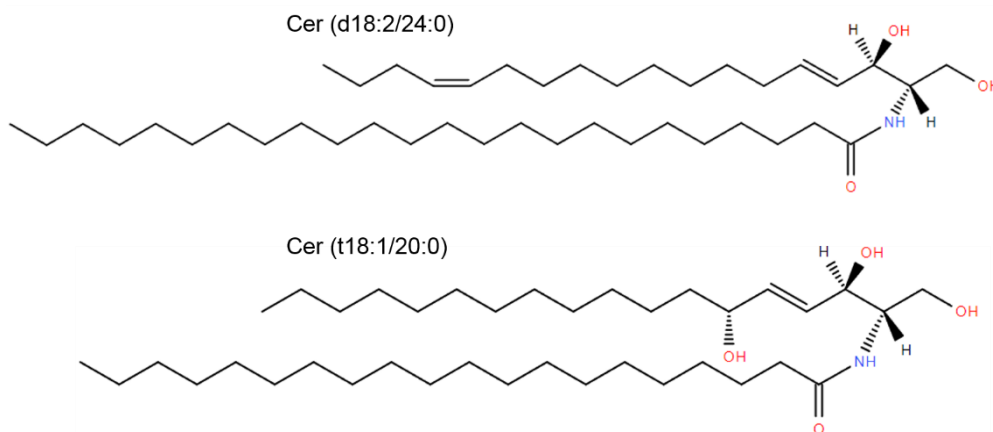
Les phosphatidylcholines sont des lipides dans lesquels l'acide phosphorique est estérifié par la choline, un dérivé triméthylé de l'éthanolamine. Ce sont les lipides les plus abondants des membranes cellulaires des tissus animaux.

Le groupe des phosphatidylsérines est constitué de molécules dont l'acide phosphorique est estérifié par l'hydroxyle d'un acide aminé aliphatique neutre appelé la serine. Ils sont faiblement acides et sont présents dans toutes les membranes cellulaires. Un exemple de phosphatidylsérine pouvant activer les mécanismes d'apoptose une fois transloquée vers la face extracellulaire de la membrane plasmique.

### *Les sphingolipides*

Dans cette catégorie de lipides complexes, le glycérol est substitué par un sphingoïde qui est une chaîne aliphatique qui comprend un aminodiol à l'une de ses extrémités. La plus couramment rencontrée étant la sphingosine. Afin de distinguer les sphingoïdes di- et tri-hydroxyles, la nomenclature 'd' et 't' est respectivement utilisée (**Figure 22**).

Les sphingolipides sont produits en ajoutant éventuellement un radical sur l'hydroxyle du carbone 1 de la sphingosine (ou autre sphingoïde) et en acylant l'amine en position 2. Le radical ajouté en position 1 est lié soit par une liaison ester, soit par une liaison glycosidique. Contrairement au glycérol, où les acides gras sont liés par une liaison ester, dans les sphingolipides, ils sont liés à la sphingosine par une liaison amide. Les sphingolipides peuvent être classés de manière générale en fonction de la nature du radical en position 1 du résidu sphingoïde. En effet, si le radical est un hydroxyle, un ester phosphorique ou un glucide, les sphingolipides obtenus feront partie de la classe des céramides, des phosphosphingolipides ou des glycosphingolipides respectivement.



**Figure 22** : Exemple de céramides di- et tri- hydroxylés.

Les céramides (Cer) représentent la forme la plus simple des sphingolipides. Ils se composent d'un sphingoïde auquel un acide gras est lié par une liaison amide à l'azote en position 2 de la chaîne carbonée. En raison de la longueur des chaînes carbonées qui les composent, les céramides sont plutôt hydrophobes et peu solubles dans les milieux aqueux, malgré la présence d'hydroxyles. Cependant, le groupe de tête des différents sphingoïdes, composé d'hydroxyles, joue un rôle déterminant dans certaines fonctions physiologiques telles que la barrière cutanée. Bien que les céramides soient principalement des intermédiaires dans la biosynthèse des sphingolipides



complexes, ils jouent un rôle dans plusieurs fonctions physiologiques telles que la différenciation, la transformation, la prolifération cellulaire et la régulation de l'apoptose<sup>156</sup>.

### Lipides dans le cancer

Avec plus de 17 millions de nouveaux cas chaque année à travers le monde et près de 10 millions de décès, le cancer demeure l'un des principaux défis de santé publique et représente un fardeau considérable pour la société. Selon les connaissances actuelles, le cancer est principalement causé par des altérations de l'ADN dans des gènes qui favorisent une croissance cellulaire excessive, la survie et la formation de métastases. Cela implique généralement une activation constante de récepteurs de facteurs de croissance et de processus de signalisation en aval, ainsi qu'une modification des processus métaboliques qui fournissent des nutriments et de l'énergie aux cellules cancéreuses pour leur croissance dans un environnement en constante évolution<sup>157</sup>. L'un des premiers changements métaboliques, découvert il y a près d'un siècle, est la façon dont les cellules cancéreuses modifient leur utilisation du glucose. En effet, depuis les années 1920, il est connu que, contrairement à la plupart des tissus normaux, les cellules cancéreuses ont une forte appétence pour le glucose, qu'elles transforment en lactate par un processus appelé glycolyse aérobie, indépendamment de la présence d'oxygène. Cette caractéristique, connue sous le nom d'"effet Warburg", est à la base de l'imagerie moderne du cancer par TEP-FDG. La glycolyse aérobie fournit non seulement de l'énergie aux cellules cancéreuses, mais également les éléments carbonés nécessaires à la synthèse des composants cellulaires essentiels, tels que les nucléotides et les lipides<sup>158</sup>.

Avec l'avènement d'outils efficaces pour étudier les lipides, notamment la lipidomique basée sur la spectrométrie de masse, les lipides apparaissent comme des acteurs centraux de la biologie du cancer. Ces dernières années, l'importance de l'analyse des lipides, notamment des glycérophospholipides, présents dans les tissus, a considérablement augmenté. Les GPLs sont maintenant reconnus comme des acteurs importants dans les mécanismes physiopathologiques<sup>159</sup>.

En effet, les lipides sont une famille de molécules très complexes en raison de la variété des acides gras et des autres composants qui les composent. Leurs propriétés et fonctions biologiques découlent de leur complexité structurelle. Les lipides sont des éléments constitutifs essentiels des membranes, servent de carburant pour les processus gourmands en énergie et jouent un rôle clé en tant que molécules de signalisation et régulateurs de nombreuses fonctions cellulaires. Cependant, un déséquilibre ou un mauvais fonctionnement du métabolisme lipidique peut causer des pathologies, notamment en cas de carence ou d'excès en acides gras. Il n'est donc pas surprenant que les cellules cancéreuses, ainsi que d'autres types de cellules dans le microenvironnement tumoral, exploitent divers moyens d'acquérir des lipides et réorganisent

considérablement leur métabolisme dans le cadre d'une reprogrammation métabolique plastique et dépendante du contexte, qui est pilotée par des signaux oncogéniques et environnementaux. Les changements qui en résultent dans le devenir et la composition des lipides aident les cellules cancéreuses à prospérer dans un microenvironnement changeant en soutenant les fonctions oncogéniques clés et les caractéristiques du cancer, y compris l'énergie cellulaire, la promotion de la signalisation oncogénique en amont, la résistance au stress oxydatif, la régulation de la communication intercellulaire et les réponses immunitaires<sup>160</sup>. De plus, les cellules cancéreuses possèdent de nombreuses adaptations pour résister à la réponse du système immunitaire et à la chimiothérapie. L'une des propriétés les plus significatives des cellules néoplasiques est l'altération du métabolisme des lipides et, par conséquent, la composition anormale de la membrane cellulaire. Ces changements peuvent entraîner la modulation de certaines enzymes, l'accumulation de matériel énergétique, qui pourrait être utilisé pour augmenter le taux de prolifération, une atténuation de la réponse du système immunitaire ainsi qu'une plus grande résistance aux thérapies existantes contre le cancer. Les changements sont si importants que certains lipides, pourraient même être considérés comme des biomarqueurs uniques du cancer, avec un potentiel diagnostique, pronostique et prédictif.

En effet, le profil lipidique de la membrane plasmique est une propriété cellulaire distinctive et peut être considéré comme caractéristique d'un type de cellule spécifique<sup>161</sup>. Les différences possibles comprennent la variabilité des groupes de tête polaires, ainsi que des chaînes d'hydrocarbures non polaires des acides gras<sup>162</sup>. Les groupes de tête dans le feuillet externe de la bicouche lipidique peuvent différer par des composés organiques simples ajoutés au groupe phosphate (choline, éthanolamine, sérine). Les groupements d'acides gras non polaires du feuillet interne varient en longueur de queue d'hydrocarbure et en nombre de doubles liaisons ou en position. La composition lipidique des cellules cancéreuses diffère du profil des cellules non malignes, mais elle varie également entre les types de tumeurs malignes.

### *Les GPLs*

#### *Les acides phosphatidiques*

Dans des conditions physiologiques, l'acide phosphatidique s'accumule du côté cytoplasmique de la membrane plasmique en raison de la charge négative présente dans son groupe de tête hydrophile. La proportion d'acide phosphatidique parmi les lipides membranaires ne dépasse généralement pas 1 % dans des conditions physiologiques. En raison de son groupe phosphate à charge négative, l'acide phosphatidique joue un rôle physiologique significatif.

Des niveaux élevés d'acide phosphatidique sont associés à une série de modifications dans le métabolisme des cellules cancéreuses. Tout d'abord, il active des kinases telles que la protéine

kinase activée par les mitogènes, la tyrosine kinase ABL 1 ou la protéine kinase-1 dépendante des phosphoinositides, qui sont impliquées dans les voies de signalisation du stress intracellulaire<sup>163</sup>. Ces enzymes susmentionnées sont également liées à l'initiation et à la progression du cancer. De plus, un niveau élevé d'acide phosphatidique renforce la transcription du facteur 1-alpha inductible par l'hypoxie, ce qui favorise l'angiogenèse et la prolifération des cellules cancéreuses<sup>164</sup>.

### Les phosphatidyléthanolamines

La phosphatidyléthanolamine est le deuxième phospholipide le plus couramment trouvé dans les membranes cellulaires (représentant de 15 à 25 %). Bien qu'elle soit répartie des deux côtés de la membrane plasmique, sa concentration la plus élevée se trouve généralement dans la couche cytosolique de la bicouche lipidique. Une observation intéressante est que, dans de nombreux types de cellules cancéreuses, la répartition de la phosphatidyléthanolamine est inversée, montrant une préférence pour la face externe de la membrane plasmique<sup>165</sup>.

La phosphatidyléthanolamine joue un rôle essentiel dans divers processus liés à la membrane cellulaire. Tout d'abord, elle agit comme un "chaperon lipidique" qui facilite le repliement des protéines membranaires. Par conséquent, une carence en phosphatidyléthanolamine peut entraîner une accumulation de protéines non repliées, contribuant ainsi à un changement métabolique vers un phénotype cancéreux. De ce fait, des altérations dans le profil lipidique peuvent affecter la fonction des protéines membranaires impliquées dans les voies de signalisation cellulaire, provoquant une perturbation du métabolisme et une réponse altérée aux signaux extracellulaires. De plus, la phosphatidyléthanolamine joue un rôle positif dans la régulation de l'autophagie, un processus cellulaire essentiel qui contribue à maintenir la qualité des cellules en les empêchant d'accumuler des organites endommagés ou anormaux<sup>166</sup>.

Enfin, la phosphatidyléthanolamine joue un rôle crucial en tant que composant de l'ancre glycosylphosphatidylinositol, responsable de l'attachement des protéines à la membrane. Dans ce mécanisme, la phosphatidyléthanolamine agit en fournissant le groupe phosphoéthanolamine, qui est lié soit à l'extrémité C-terminale des protéines, soit aux groupes hydroxyles du mannose situé dans le noyau glycanique<sup>167</sup>. Toute perturbation dans la biosynthèse de ces ancres GPI peut entraîner une instabilité des protéines au sein de la membrane. Par conséquent, cela peut avoir un impact sur les voies de signalisation intracellulaire, les processus de biosynthèse et de dégradation, ainsi que les mécanismes de communication cellulaire. Ces perturbations peuvent finalement conduire à des dysfonctionnements cellulaires et potentiellement à la formation de cancers.

### Les phosphatidylcholines

Bien qu'elle soit principalement considérée comme un lipide neutre, elle peut également être décrite comme un zwitterion, car elle est composée d'un groupe de tête choline, à charge positive, associé à un groupe phosphate substitutif, à charge négative. La phosphatidylcholine est le composé le plus abondant au sein de la bicouche lipidique. Par rapport à la phosphatidyléthanolamine, elle est principalement localisée dans la couche externe de la membrane plasmique.

Un large éventail de cancers montre des altérations dans le métabolisme de la phosphatidylcholine, principalement causées par une augmentation de l'activité de la choline kinase alpha, de la phospholipase C (PLC) ou de la phospholipase D. La PLC catalyse la dégradation de la phosphatidylcholine, libérant ainsi du DG, qui favorise les processus anaboliques cellulaires via l'activation de la protéine kinase A. Ces modifications métaboliques entraînent l'accumulation de molécules riches en énergie, fournissant ainsi des sources pour la prolifération cellulaire. Les carcinomes, tels que le cancer du sein, présentent une surexpression des enzymes mentionnées précédemment<sup>168</sup>.

### Les phosphatidylglycérols

Bien que le phosphatidylglycérol ne constitue qu'environ 1 % des phospholipides membranaires, il demeure un élément significatif de la face cytosolique de la membrane plasmique en raison de ses caractéristiques ioniques. Les groupements d'acides oléique et palmitique sont les substituants glycériques les plus fréquemment associés au phosphatidylglycérol. Des recherches ont montré que le phosphatidylglycérol joue un rôle dans l'activation de la protéine kinase C et qu'il participe à la régulation de la transcription virale<sup>169</sup>. Son importance lors des infections virales suggère que des niveaux anormaux de ce composé pourraient être observés dans les cancers associés aux virus, tels que le cancer du col de l'utérus. Ces fluctuations pourraient potentiellement favoriser une réplication virale plus efficace et la formation de néoplasies cellulaires. De plus, le phosphatidylglycérol est produit en tant que métabolite intermédiaire dans la voie de biosynthèse de la cardiolipine et est en mesure d'inhiber le transfert de phosphatidylcholine entre les membranes, ce qui peut entraîner des anomalies dans la structure membranaire des cellules cancéreuses.

### Les phosphatidylsérines

À pH physiologique, la phosphatidylsérine est classée comme un lipide à charge négative qui se trouve dans la couche interne de la membrane cellulaire, et sa quantité dans cette membrane est estimée à environ 6 %. Historiquement, la phosphatidylsérine était principalement associée à la signalisation de l'apoptose, cependant, des recherches récentes remettent en question son rôle en

tant qu'inducteur de la mort cellulaire par apoptose, suggérant plutôt une implication dans la nécroptose<sup>170</sup>. Ce phénomène survient lorsque la phosphatidylsérine migre vers la face externe de la membrane cellulaire grâce à l'action d'une enzyme dépendante du calcium appelée flippase. Normalement, lors de l'apoptose, la phosphatidylsérine est exposée à la surface de la cellule, mais chez les cellules cancéreuses, ce processus peut également se produire en réponse au stress oxydatif<sup>171</sup>. Par conséquent, une surabondance de phosphatidylsérine peut être considérée comme un indicateur de cancer.

La présence de phosphatidylsérine exposée à la surface des cellules cancéreuses a pour but d'éviter une réponse immunitaire auto-immune de l'organisme. Elle agit comme une barrière de protection dans le microenvironnement tumoral contre les cellules NK et d'autres cellules immunitaires cytotoxiques. Ainsi, les cellules cancéreuses échappent à la détection et à l'attaque du système immunitaire. Actuellement, les chercheurs s'efforcent de cibler les cellules présentant une exposition excessive de phosphatidylsérine dans le cadre des stratégies thérapeutiques contre le cancer<sup>171</sup>.

#### Les phosphatidylinositols

La famille des phosphatidylinositols est un groupe hétérogène de lipides membranaires qui diffèrent les uns des autres en ce qui concerne la composition des groupements d'acides gras et le nombre de groupes phosphates attachés au groupe de tête polaire du myoinositol. Dans des conditions physiologiques, les phosphatidylinositols sont présents dans le feuillet externe de la membrane<sup>172</sup>. Cependant, en raison de la grande diversité de la famille des phosphoinositols et de la variété de leurs fonctions métaboliques, il est impossible de comprendre pleinement la contribution des phosphoinositides au profil des membranes lipidiques.

#### Les céramides

Dans des conditions physiologiques, la quantité de céramides présente dans la membrane est minime, ne dépassant pas 0.5 %. Les céramides sont produites suite à une réaction catalysée par la sphingomyélinase, ce qui induit la libération du cytochrome c par les mitochondries et active ainsi la voie apoptotique. L'augmentation des niveaux de céramides dans la membrane entraîne l'inhibition de la prolifération des cellules cancéreuses et une régulation adéquate du cycle cellulaire<sup>173</sup>. En conséquence, l'augmentation de la biosynthèse des céramides semble être non seulement inoffensive, mais elle présente également un potentiel thérapeutique.

Des dysfonctionnements ou une surexpression des enzymes impliquées dans la dégradation des céramides peuvent perturber le cycle cellulaire et altérer le processus apoptotique, ce qui pourrait éventuellement contribuer à la cancérogenèse. L'augmentation de l'activité des enzymes

favorisant la formation de sphingolipides pro-malignes permet aux cellules de résister à l'apoptose et favorise la progression du cancer. Le transfert excessif de céramides du réticulum endoplasmique vers l'appareil de Golgi est considéré comme un facteur favorisant le cancer, car il éloigne le lipide anti-apoptotique de la membrane cellulaire<sup>165</sup>.

## Chapitre 2 : Configuration et optimisations de la technologie SpiderMass

---

### Configuration de la technologie SpiderMass

#### Fonctionnement

Le SpiderMass est une méthode d'analyse *in vivo* en temps réel, basée sur une source d'ionisation ambiante et caractérisée par une faible invasivité, et reposant sur l'utilisation d'un laser émettant dans le spectre infrarouge. Ce dispositif se compose d'une sonde de micro-échantillonnage, d'une ligne de transfert, et d'un analyseur MS. Notamment, cette technologie permet un micro-échantillonnage par excitation de la bande d'absorption la plus intense (mode d'étirement de la liaison O-H) des molécules d'eau à 2.94  $\mu\text{m}$  dans l'infrarouge, induisant ainsi un processus de désorption/ionisation. Pour atteindre une longueur d'onde de 2.94  $\mu\text{m}$  dans le moyen IR, le système laser utilisé repose sur un oscillateur paramétrique optique (OPO), alimenté par un laser de pompe de type Nd:YAG et accordable en longueur d'onde. La durée d'impulsion est de 6 nanosecondes afin de limiter les effets thermiques et assurer une faible invasivité tout en offrant un processus de désorption/ionisation efficace. Ce processus de désorption et d'ionisation, nommé désorption/ionisation laser assistée par l'eau (WALDI), est similaire à celui du MALDI, l'eau jouant le rôle de matrice endogène pour générer des ions en phase gazeuse. En ce qui concerne l'analyse des tissus, le laser est focalisé à la surface de l'échantillon via une fibre optique terminée par une pièce à main, qui permet un déplacement au-dessus de la surface, manuellement (si l'utilisateur prend le système en main) ou en automatique (si la pièce à main est couplée à un bras robotisé) et offrant ainsi une analyse par screening sans contact de tout type d'échantillon ou surface, qu'elles soient *ex vivo* ou *in vivo*. Un tube de transfert est positionné au-dessus de la surface du tissu à analyser pour aspirer la plume de désorption formée lors de l'ablation du tissu et la transférer vers le spectromètre de masse. Ce tube est un simple tubing en polymère. Plusieurs matériaux ont été testés avec succès (e.g. PTFE) néanmoins le tubing en Tygon biocompatible est privilégié en vue de l'application au bloc opératoire. Le tube de transfert permet non seulement de ramener l'aérosol généré au spectromètre de masse, mais il permet également de déporter ce dernier pour éviter de gêner les mouvements de l'utilisateur, encore une fois en considérant son utilisation dans l'environnement réel. Une longueur de tube jusqu'à 5m est utilisable et a été testée sans induire une perte de signal au niveau des spectres MS enregistrés<sup>174</sup>. Il est important de souligner que ce système est caractérisé par une faible invasivité, puisqu'au cours d'une seule expérience seulement 0.1 à 0.3  $\text{mm}^3$  d'échantillon sont prélevés (avec une profondeur d'échantillonnage de seulement quelques micromètres) et que les échantillons ne subissent qu'une déshydratation réversible. Les



profils des spectres de masse obtenus avec le SpiderMass en mode positif et négatif démontrent la détection de diverses molécules, telles que des métabolites, des lipides ainsi que des protéines à partir de standards. A partir de tissus, les métabolites et les lipides sont largement favorisés par leur forte abondance dans les tissus limitant, dans les conditions actuelles d'utilisation, l'observation des protéines à partir de tissus. La technologie SpiderMass peut être couplée à différents spectromètres de masse de toutes marques constructrices (**Figure 23**). Cependant, la source conventionnelle du spectromètre de masse doit être remplacée par une interface. En effet, la source n'est pas nécessaire puisque la formation des ions se produit en amont au niveau du site de désorption/ionisation. En revanche, une interface doit être ajoutée afin d'éviter un encrassement trop important de l'instrument requérant un nettoyage trop fréquent des optiques de transfert tel qu'observé dans les montages sans interfaces. Par ailleurs, il a été observé que l'interface permet d'assurer une désolvatation des agrégats présents dans l'aérosol et favorise la libération des charges, contribuant à une augmentation de l'intensité des signaux sur les spectres MS et donc de la sensibilité des analyses. Différentes interfaces ont été étudiées et utilisées. L'une d'elle, l'interface REIMS est commerciale et utilisée pour le couplage aux équipements de la société Waters. Par ailleurs, deux autres interfaces ont été développées au laboratoire et ont fait l'objet de brevets, l'une permettant un couplage avec les instruments de type Orbitrap et la seconde pour les configurations QToF commercialisées par la société Bruker.

### Différentes configurations existantes

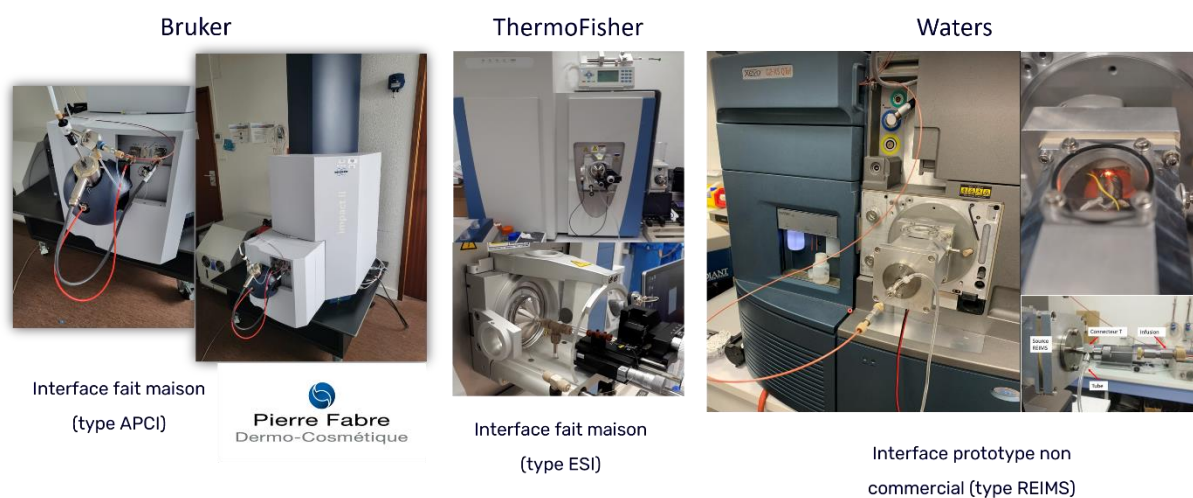
#### *SpiderMass couplé à un spectromètre de masse Bruker*

La configuration de la technologie SpiderMass couplée à un spectromètre de masse Bruker est disponible au sein de l'entreprise Pierre Fabre dermo-cosmétique (PFDC) dans le cadre d'une collaboration qui vise à évaluer les performances de la technologie SpiderMass en dermatologie pour les pathologies de la peau. Le spectromètre de masse utilisé est un QToF Impact II Bruker (Bruker Daltonics) sur lequel est monté une interface, développée et brevetée dans le cadre de la collaboration, dont le principe implique une décharge corona comme pour les sources de type APCI (**Figure 23**). Brièvement, les molécules à analyser sont nébulisées dans une chambre scellée et le spray est entièrement vaporisé à 400°C. Les molécules mélangées au solvant passent ensuite devant une aiguille à haute tension à l'extrémité de laquelle une décharge électrique est générée. Celle-ci ionise le solvant et transfère la charge aux analytes, ces derniers étant transférés au spectromètre de masse via le capillaire de transfert et le système d'ion-tunnels. Cette configuration a été utilisée dans le cadre d'une étude clinique (NCT04472546) pour étudier les variations métaboliques de la peau *in-vivo* dans le cadre de dermatose inflammatoires chroniques telles que l'acné, la dermatite atopique, la rosacée, la dermatite séborrhéique et le psoriasis. Les résultats obtenus avec cette

interface sont concluants, et il est particulièrement intéressant, de noter que cette interface offre une très bonne détection des petits métabolites, comme attendu d'une interface dont le principe reste proche d'une source APCI.

### *SpiderMass couplé à un spectromètre de masse ThermoFisher*

Une interface a également été développée pour le couplage avec les spectromètres de type Orbitrap de chez ThermoFisher, et en particulier, testé sur une configuration de type QExactive. Cette interface brevetée implique notamment un mécanisme d'électronébulisation en nanodébit (**Figure 23**). L'utilisation d'un processus de type nanoESI dans cette interface a montré la possibilité de détecter à la fois les lipides mais également les protéines directement à partir de tissus en condition *ex-vivo* comme *in-vivo*.



**Figure 23** : Différentes configurations possibles de la technologie SpiderMass.

### *SpiderMass couplé à un spectromètre de masse Waters*

La version du SpiderMass couplée à un spectromètre de masse de chez Waters, le Xevo G2-XS, est la version utilisée tout au long de ma thèse. Celle-ci utilise une interface prototype de l'interface REIMS commercialisée par la société Waters (**Figure 23**). Cette interface vient en lieu et place de la source ESI conventionnelle. Elle comporte un ruban chauffé de Kanthal dont la température peut être réglée en changeant l'intensité du courant qui traverse le ruban. Cependant, la température n'étant pas mesurable, celle-ci est estimée en fonction de la couleur du ruban chauffé. Les expériences réalisées ont montré que l'utilisation de l'interface permet une augmentation du courant ionique total et prévient le spectromètre de masse d'un encrassement trop important. L'hypothèse évoquée pour expliquer l'augmentation d'intensité du signal est l'obtention d'une meilleure désolvatation des agrégats de l'aérosol, sachant qu'un signal est observé même si le ruban n'est pas chauffé. C'est sur cette configuration que les optimisations de la technologie SpiderMass ont été effectuées.

## Optimisations de la technologie SpiderMass

Des optimisations du prototype SpiderMass ont été mises en place au début de ma thèse afin d'adapter au mieux le prototype de laboratoire pour une utilisation optimale future au bloc opératoire. Le spectromètre de masse de type Q-TOF, le Xevo G2-XS (Waters), a été choisi afin d'être compatible avec une utilisation lors d'opérations chirurgicales. De nombreux tests de routine ont été réalisés afin de déterminer les paramètres optimaux en mode d'ions positifs et négatifs afin d'augmenter la sensibilité de l'appareil, la qualité des spectres obtenus ainsi que la reproductibilité des expériences.

### Utilisation d'une infusion contenant un Lockmass

L'utilisation du spectromètre de masse Xevo G2-XS (Waters) avec son interface REIMS permet l'utilisation d'une infusion de solvant via l'interface pendant les analyses par SpiderMass. Différents tests ont été réalisés pour évaluer différents solvants à différents débits, et l'impact de ces changements sur l'intensité du signal et le profil des spectres. L'infusion contribue à une augmentation du signal total, en lien avec la solvataion des agrégats de l'aérosol et contribue à une augmentation du signal générale affectant certains lipides de façon plus importante que d'autres (**Figure 24**). Finalement, notre choix s'est porté sur l'utilisation d'une infusion d'isopropanol (IPA) dopée en leucine enképhaline (LeuK) à une concentration de 1 ng/μL pour un débit de 150 μL/min.

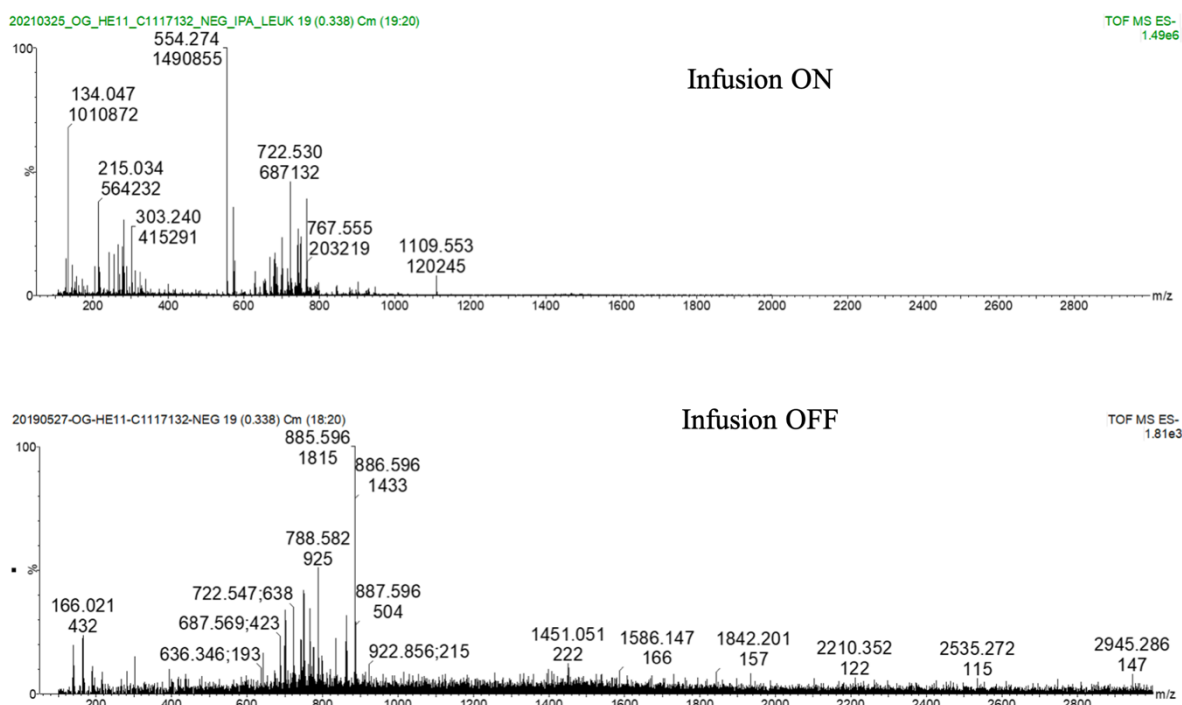


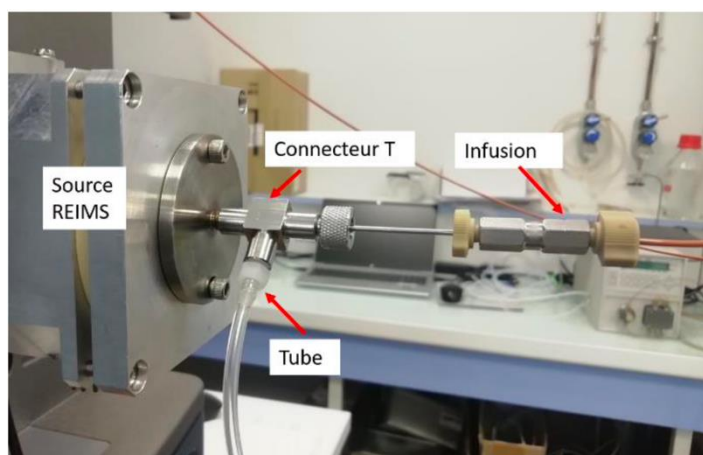
Figure 24 : Comparaison des spectres obtenus lors de l'analyse d'un tissu de cancer œsogastrique avec ou sans infusion d'IPA/LeuK en mode négatif.

La leucine enképhaline joue le rôle d'un lockmass, un pic de spectre correspondant à un composé de masse connue. Cela permet, pendant l'acquisition, de savoir si l'analyse est optimale ainsi qu'en,

post-acquisition, d'effectuer une correction de masse lors de l'analyse de données. La LeuK peut être utilisée autant en mode négatif ( $m/z$  554.3) qu'en mode positif ( $m/z$  556.3).

L'analyse d'un échantillon de cancer œsogastrique avec l'infusion IPA/LeuK permet d'obtenir des spectres dans les deux modes d'ionisation où l'on distingue très bien le pic caractéristique de la LeuK en plus des pics provenant du tissu analysé (**Figure 24**). Cette possible correction de masse en post-acquisition diminue grandement les problèmes de reproductibilité lors de l'analyse d'une cohorte de tissu sur plusieurs jours. De plus, l'infusion d'IPA permet de multiplier l'intensité par 1000 en obtenant des pics d'une intensité absolue dans la gamme des  $10^6$  au lieu de  $10^3$  sans infusion (**Figure 24**). De plus, le rapport signal sur bruit est augmenté et permet d'obtenir des spectres plus qualitatifs.

### Optimisation du connecteur T



**Figure 25** : Connecteur final permettant de relier l'infusion et le tube de transfert à la source REIMS.

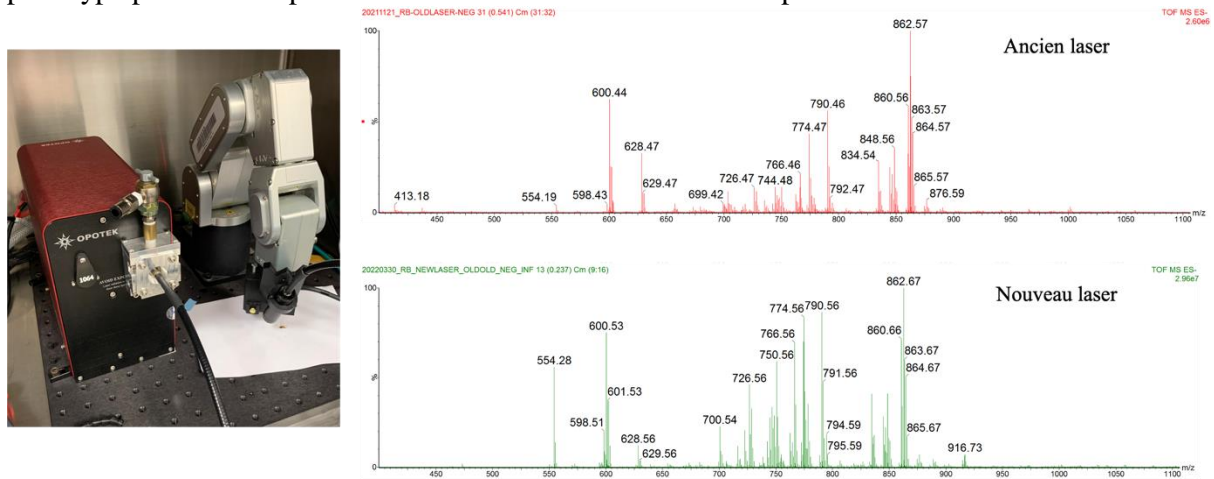
Afin de lier le tube de transfert au prototype d'interface REIMS ainsi que pour permettre l'infusion d'IPA/LeuK avant l'entrée dans le spectromètre de masse, un connecteur en forme de T a été développé. Sa forme et sa position ont été optimisées au maximum pour permettre une analyse optimale et reproductible. Le connecteur final en acier est composé de 3 entrées ; la première pour le capillaire relié au

système d'infusion, la deuxième permettant de relier le tube de transfert et la troisième pour le capillaire relié à la source REIMS (**Figure 25**). Ce connecteur permet d'obtenir des analyses quotidiennes optimales et reproductibles. Ces différentes optimisations permettent de disposer d'un système plus robuste et offrant une meilleure reproductibilité inter-analyses dans le temps.

### Changement de laser

Une optimisation du prototype SpiderMass a aussi été possible grâce au changement du laser utilisé pour le micro-échantillonnage des tissus. En effet, le laboratoire a pu faire l'acquisition d'un nouveau laser plus compact et équipé d'une fibre gainée présentant des connecteurs SMA à chaque extrémité permettant une connexion plug & play de la fibre au niveau du site d'injection et de la pièce à main. Par ailleurs, l'injection dans la fibre se fait sous  $N_2$  pour protéger cette dernière des poussières qui pourraient conduire à une dégradation rapide à l'entrée de la fibre, l'énergie en crête étant importante au vu des 9 mJ injectés sur des pulses d'une durée de seulement 6 ns dans la fibre. Les résultats obtenus sont très satisfaisants car ils sont très semblables aux résultats obtenus

avec l'ancien laser en termes de qualité de spectres avec même une meilleure intensité lors de l'analyse d'un cerveau de rat en mode négatif (**Figure 26**). Le laser est plus stable, une augmentation de la reproductibilité des spectres et un point d'ablation d'environ 300  $\mu\text{m}$  ont été observés. Par ailleurs, ce nouveau laser et ses fibres gainées permettent de limiter les fluctuations liées au changement de la fibre et son repositionnement mais aussi de permettre de rendre le prototype plus robuste pour une utilisation future en milieu hospitalier.



**Figure 26** : Image du nouveau laser ainsi que la comparaison des spectres obtenus pour l'analyse d'un cerveau de rat avec l'ancien laser (en **haut**) et le nouveau laser (en **bas**) en mode négatif.

La technologie est donc une nouvelle technologie basée sur la MS qui vise à aider la chirurgie des tumeurs solides en permettant au chirurgien de baser ses choix sur des données objectives provenant d'analyses moléculaires non ciblées, obtenues en temps-réel, et ne nécessitant par la réalisation de biopsies. Même si la technologie, reste encore en constante amélioration, elle est devenue assez robuste pour entamer des études cliniques. Chaque cancer présente ses particularités et ses questions, et les besoins pendant la chirurgie peuvent être différents. Dans le cadre de mon doctorat, nous nous sommes intéressés à la chirurgie des cancers œsogastriques étant donné l'importance de ces cancers dans la région des Hauts de France.

L'objectif principal de mon travail de recherche était donc d'évaluer la technologie SpiderMass pour le diagnostic et le pronostic en temps réel des tumeurs œsogastriques, contribuant ainsi notamment à préciser les marges de résection chirurgicale et l'extension loco-régionale de la tumeur. Avant d'appliquer le SpiderMass à la problématique du cancer œsogastrique, des études préliminaires ont été menées pour améliorer le traitement des données et les modèles de classification. Cela inclut la comparaison des spectres MALDI et SpiderMass dans la gamme des phospholipides pour déterminer si les données MALDI peuvent être utilisées pour enrichir les données d'apprentissage du SpiderMass. Par ailleurs, de nouveaux algorithmes de machine learning ont été explorés pour la classification des données SpiderMass, et des scores immunitaires et bactériens ont été développés pour améliorer le diagnostic et le pronostic.

Ainsi, le manuscrit est structuré en 4 parties qui reprennent les développements réalisés sur le traitement des données (chapitre 3 et 4) ainsi que l'application de ces développements au cancer œsogastrique (chapitre 5) et dans le cadre du cancer du côlon (chapitre 6).

Le chapitre 3 se concentrera sur la comparaison des technologies d'imagerie par spectrométrie de masse, MALDI et WALDI, dans le but d'une utilisation conjointe en vue d'une classification améliorée des tissus sains, cancéreux ou mixtes.

Le chapitre 4, quant à lui, abordera l'analyse des données SpiderMass en introduisant nouveau pipeline d'analyse des données basé sur du machine learning, et développé comme preuve de concept à partir d'une cohorte de tissus de glioblastomes.

Le chapitre 5 portera sur une étude du cancer œsogastriques par SpiderMass en faisant appel aux développement réalisés dans les chapitres précédents mais en introduisant également la notion de bastérioscore..

Enfin, le chapitre 6, présentera les résultats obtenus grâce à la technologie SpiderMass dans le cadre d'un projet de collaboration avec le Dr. Matthias Chamaillard et mettant en évidence l'utilisation de SpiderMass pour détecter des lipides spécifiques liés au changement métabolique induit par la colibactine, facteur favorisant le développement du cancer du côlon droit.



## Chapitre 3 : Étude comparative de l'imagerie MALDI et de l'imagerie WALDI

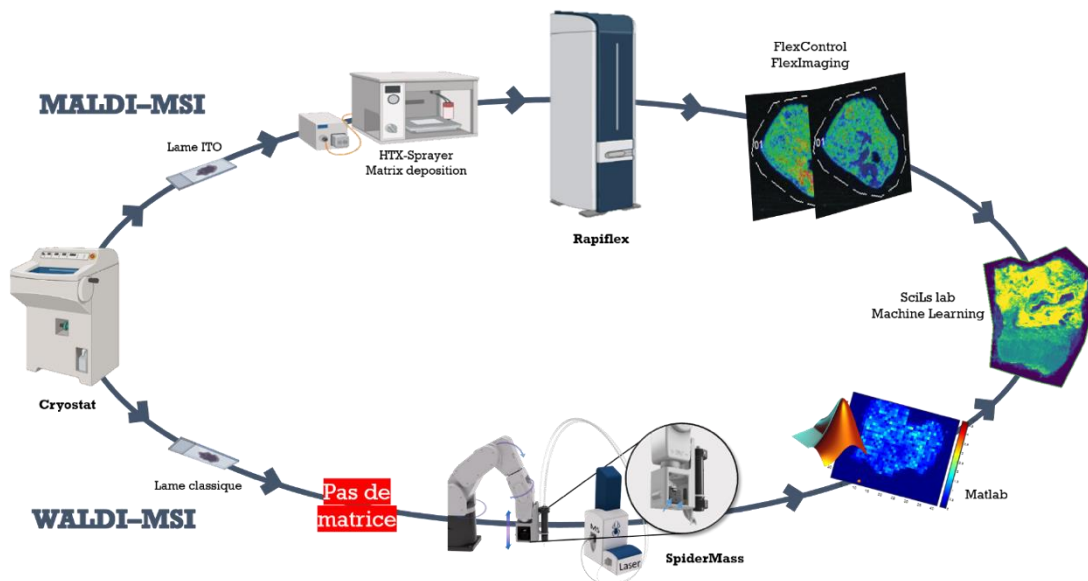
---

### Introduction

L'imagerie par spectrométrie de masse est une méthode moléculaire puissante et sans marquage qui permet d'analyser une large gamme de molécules endogènes et exogènes dans les tissus biologiques sans altérer l'intégrité ou la structure de l'échantillon<sup>116</sup>. En particulier, l'imagerie MALDI a démontré son immense potentiel pour des applications cliniques telles que la découverte de biomarqueurs, le diagnostic, le pronostic et la stratification des patients<sup>90</sup>. Cependant, le processus chronophage non trivial de préparation des échantillons dans le MALDI-MSI limite son utilisation *in-vivo*. La recherche sur les lipides tissulaires a considérablement gagné en importance au cours des dernières années en raison de leur rôle de médiateurs importants dans les mécanismes physiopathologiques<sup>159</sup>. Ainsi, l'imagerie MALDI-MSI est devenue un outil essentiel pour étudier la biochimie des lipides dans les tissus biologiques<sup>175,176</sup>. Elle a été utilisée pour examiner les variations de composition lipidique dans divers contextes, tels que l'oncologie<sup>177,178</sup> et les maladies neurodégénératives<sup>179,180</sup>, et elle est progressivement considérée comme une alternative à l'histopathologie, car elle permet d'étudier la distribution et l'abondance des métabolites et des lipides selon les phénotypes cellulaires<sup>181,182</sup>. Différentes technologies de spectrométrie de masse à ionisation ambiante, telles que le DESI<sup>37</sup>, l'AP-MALDI<sup>183,184</sup>, le LAESI<sup>59</sup> et le MALDESI<sup>61</sup>, ont été développées pour l'imagerie des échantillons dans des conditions proches de leur état naturel, nécessitant une préparation minimale contrairement aux approches traditionnelles en spectrométrie de masse sous vide. La technologie SpiderMass, récemment couplée à un bras robotique équipé d'un capteur de distance, permet d'obtenir des images topographiques en 2D et en 3D. Cette technologie permet également l'imagerie *ex vivo* de coupes de tissus ainsi que l'imagerie *in vivo* de tissus humains, sans ajout de matrice et à température ambiante, en utilisant l'eau endogène des tissus comme matrice pour l'ionisation WALDI (**Figure 27**).

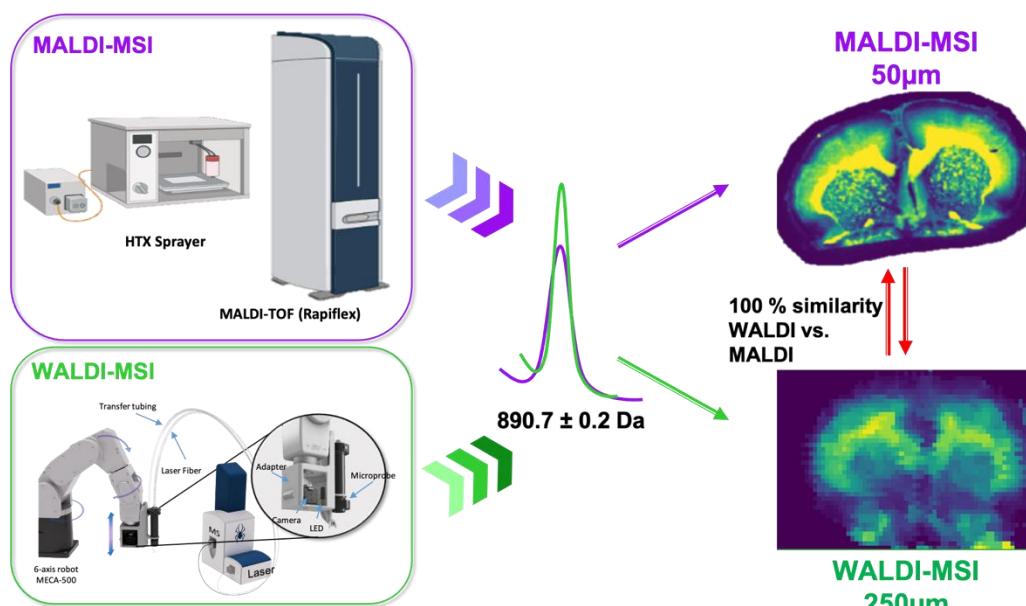
Les profils moléculaires obtenus par WALDI présentent des similitudes et des caractéristiques semblables à ceux obtenus par MALDI, notamment des effets collectifs, une faible fragmentation et une sensibilité équivalente en mode ions positifs et négatifs. En effet, un comportement collectif similaire à celui du MALDI a été observé précédemment dans le WALDI lors de l'accord de la longueur d'onde du laser pour obtenir l'excitation résonnante<sup>12</sup>. Dans ce contexte, la possibilité de comparer ces deux technologies que ce soit en mode profilage ou en mode imagerie grâce à une méthodologie précise et quantitative permettrait de mettre en lumière la possibilité d'utiliser ces deux technologies conjointement.





**Figure 27** : Comparaison des principales étapes des méthodologies utilisées pour l'imagerie MALDI et l'imagerie WALDI.

Dans cette étude, une comparaison directe approfondie entre les technologies MALDI et WALDI pour l'analyse des lipides a été menée, en utilisant le MALDI comme référence pour l'analyse moléculaire sans marquage dans la recherche biomédicale. Les comparaisons ont été effectuées en mode imagerie, en tenant compte des différentes matrices MALDI couramment utilisées pour l'analyse des lipides (**Figure 28**). Plus précisément, les profils lipidiques de plusieurs échantillons biologiques (cerveaux de rat et échantillons de cancer œsogastrique) ont été étudiés, et la similitude des pics détectés ainsi que les indices de corrélation de Pearson de l'intensité et de la variance du profil spectral entre les deux techniques ont été calculés. Enfin, l'étude a également cherché à déterminer si les données MALDI-MSI pouvaient être utilisées comme référence pour entraîner la technologie SpiderMass.



**Figure 28** : Comparaison des technologies d'imagerie MALDI et WALDI

## Matériel et méthodes

### Échantillons

Les trois cerveaux de rats frais congelés ont été obtenus auprès de l'animalerie de l'Université de Lille. Tandis que les échantillons de tissus provenant d'exérèses de cancer de l'estomac/œsophage ont été fournis par FREGAT (Recherche française Tumeurs de l'œsophage et de l'estomac) après accord de la tumorothèque du centre hospitalier universitaire de Lille. Tous les échantillons ont été congelés et entreposés à une température de -80°C jusqu'à leur analyse.

### Préparation des échantillons

Les cerveaux de rats congelés ainsi que les échantillons tissulaires cancéreux ont été découpés à l'aide d'un cryostat (Leica Biosystems CM1510S, Allemagne). En premier lieu, une section de 7 µm d'épaisseur a été montée sur une lame de verre ordinaire pour une coloration à l'hématoxyline phloxine safran (HPS). Par la suite, une seconde section d'une épaisseur de 12 µm a été déposée sur une lame de verre conductrice en oxyde d'indium et d'étain (ITO) (LaserBio Labs, Valbonne, France). Enfin, une troisième et une quatrième section, mesurant respectivement 20 µm et 12 µm d'épaisseur, ont été prélevées sur des lames de verre ordinaires en vue des analyses WALDI-MSI.

### MALDI-MSI

Plusieurs matrices avec des paramètres d'application optimisés ont été utilisées pour la réalisation de l'analyse MALDI-MSI. Les matrices norharmane et DAN ont été utilisés pour l'imagerie en mode polarité double, tandis que la matrice 9-AA a été employée exclusivement en mode négatif, et la matrice DHB en mode positif. Chaque matrice a été appliquée au moyen d'un pulvérisateur HTX TM-Sprayer (HTX Technologies, LLC) configuré selon les paramètres optimisés indiqués dans le **Tableau 4**.

Les analyses MALDI-MSI ont été exécutées sur un RapifleX MALDI TissueTyper™ (Bruker Daltonik GmbH) en mode positif et négatif, avec une calibration externe utilisant du phosphore rouge avant l'analyse. La résolution spatiale a été réglée à 150 µm ou 50 µm, et la gamme de masse a été définie entre  $m/z$  400 et 1500. L'énergie du laser a été paramétrée entre 70 % et 80 % de sa puissance maximale.

**Tableau 4** : Liste des paramètres optimisés pour le dépôt des matrices norharmane, 2,5-DHB, 9-AA et DAN, avec le pulvérisateur HTX M5.

<b>Paramètres</b>	<b>norharmane</b>	<b>2,5-DHB</b>	<b>9-AA</b>	<b>DAN</b>
Concentration (mg/mL)	7	20	10	10
Solvant	CHCl <sub>3</sub> :MeOH (2:1, v/v)	MeOH/H <sub>2</sub> O 0.1% TFA (7:3, v/v)	EtOH/H <sub>2</sub> O 0.1% AF (7:3, v/v)	ACN/H <sub>2</sub> O 0.1% AF (9:1, v/v)
Température (°C)	30	75	90	30
Nombre de passages	12	20	5	4
Motif	CC	CC	CC	CC
Débit (mL/min)	0.1	0.1	0.12	0.1
Vélocité de la buse (mm/min)	1200	1000	1200	1200
Espace entre les passages (mm)	2	3	3	3
Pression du gaz (psi)	10	10	10	10

#### WALDI-MSI

Le système SpiderMass est équipé d'un système laser à longueur d'onde réglable à distance à fibre optique (Radiant version 1.0.1, OPOTEK Inc., États-Unis) pompé par un laser Nd:YAG à commutation de 6 ns de largeur d'impulsion émettant un rayonnement de 1.064 µm (Quantel Laser, France). La microsonde laser infrarouge est réglée à 2.94 µm pour exciter la bande vibrationnelle la plus intense de l'eau (O-H). Une fibre laser biocompatible d'un mètre de long et de 450 µm de diamètre interne (HP fiber, Infrared Fiber Systems, États-Unis) est connectée à la sortie de l'OPO et focalisée pour obtenir un diamètre de faisceau de 400 à 500 µm. La microsonde Spider-Mass a été couplée à un bras robotique rigide décrit dans une publication antérieure<sup>18</sup>. La résolution spatiale a été réglée à 250 µm pour obtenir un suréchantillonnage et la gamme de masse a été fixée entre  $m/z$  100-1500. La séquence d'acquisition était composée de 3 tirs laser consécutifs et d'un intervalle de 3 secondes entre chaque étape. Les tirs laser et l'acquisition du spectromètre de masse ont été déclenchés automatiquement par l'interface MATLAB développée en interne pour la technique robotique WALDI-MSI<sup>18</sup>. Les données ont été acquises en mode positif et négatif sur l'analyseur de masse Xevo (G2-S, Q-TOF, Waters, Manchester, Royaume-Uni) à travers un tube Tygon biocompatible connecté à une interface prototype placée à l'entrée de l'instrument.

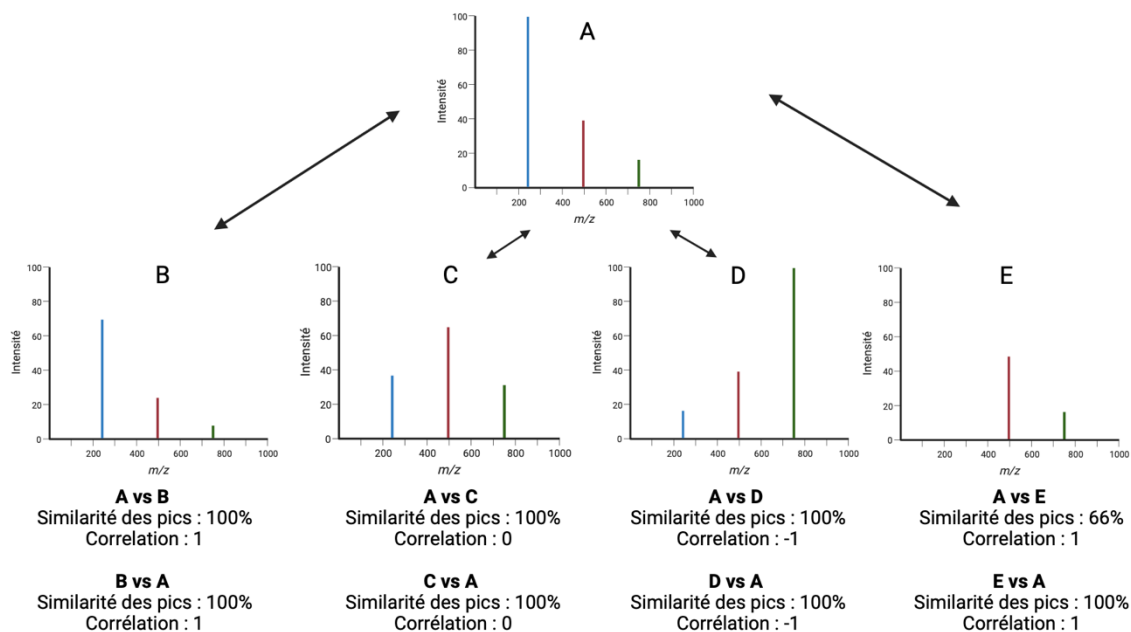
#### Importation et traitement des données

Un code Matlab a été développé pour extraire les scans MS de chaque pixel à partir du fichier brut Waters Raw. Les fichiers bruts SpiderMass ont été ensuite convertis en format mzML à l'aide de MSConvert (Proteowizard)<sup>185</sup>. Le convertisseur imzML a été employé pour reconstituer les fichiers d'imagerie<sup>186</sup>. Les données provenant de l'appareil Bruker ont été exportées au format

imzML en utilisant le logiciel FlexImaging 5.0. Toutes les données ont été importées dans le logiciel SCiLS (SCiLS™ Lab 2021, MVA) avec une correction top hat de la ligne de base. Les données ont été regroupées en bins de 0.3 Da. Premièrement, les données ont été soumises à une régression par les moindres carrés partiels (pLSA). La pLSA a été effectuée sur les spectres normalisés avec la normalisation Root Mean Square (RMS). La sélection automatique des pics a été réalisée en utilisant le pipeline de segmentation par défaut. Les intervalles  $m/z$  résultants ont été utilisés pour la pLSA. Sept composantes ont été calculées. Les composantes ainsi obtenues présentent des distributions distinctes dans le tissu, et chaque spectre est considéré comme une combinaison pondérée de ces composantes. Deuxièmement, la segmentation a été réalisée à l'aide de l'algorithme bisecting  $k$ -means en utilisant la distance corrélation. Les ions  $m/z$  qui permettent de différencier les données entre WALDI-MSI et MALDI-MSI ont été identifiés à l'aide d'une courbe ROC (receiver operating characteristic). Cette étape a été effectuée sur toute la plage de  $m/z$  analysée (300-1500). L'aire sous la courbe ROC (également appelée AUC) varie de 0 à 1 et reflète la capacité de discrimination du signal  $m/z$  en une seule valeur. Les valeurs discriminantes ont été considérées comme telles si l'aire sous la courbe ROC était supérieure au seuil 0.85. Les régions correspondant au tissu ont été exportées en tant que spectres globaux dans un fichier CSV, avec soustraction de la ligne de base et normalisation RMS. Ce fichier CSV a ensuite été utilisé pour les calculs de similarité des pics et de corrélation.

### Similarité des pics détectés et calculs de corrélation de Pearson des spectres de masse

Les analyses de similarité des pics détectés ont été effectuées à l'aide d'un algorithme Python dédié à la détection automatique des pics. Cet algorithme examine la présence ou l'absence de chaque pic dans une gamme de masse spécifique  $m/z$  600-1000 en mode positif et négatif. Le pourcentage a été calculé en se basant sur le nombre de pics présentant des différences. Un seuil a été appliqué pour comparer 90 % du signal des spectres, ce seuil étant choisi en fonction des pics de base. Ensuite, une matrice de corrélation de Pearson a été établie à l'aide du logiciel R pour évaluer la similarité des intensités relatives (abondances des ions) et des variances (écart-type de l'intensité) des mêmes pics détectés par les deux techniques. Le coefficient de corrélation de Pearson, noté  $r$ , est un indice reflétant une relation linéaire entre deux variables continues. Il varie de -1 à +1, où 0 indique une absence de corrélation, +1 une corrélation positive et -1 une corrélation négative (**Figure 29**).



**Figure 29** : Explication de la similarité des pics détectés et calculs de la corrélation de Pearson du spectre de masse A vs les spectres B, C, D, E et inversement.

## Modèle CNN basé sur les données MALDI-MSI pour la prédiction des données WALDI-MSI

La matrice de données a été élaborée en utilisant la segmentation MALDI-MSI précédente des deux groupes, à savoir la zone cancéreuse et la zone saine. Ces groupes ont été enregistrés au format imzML, et les données spectrales de chaque groupe ont été extraites en utilisant le module ImzMLParser de la bibliothèque pyimZML. La matrice finale comporte 37 560 caractéristiques qui correspondent aux plages  $m/z$  600-1000, et elle se divise en deux classes de cibles : la zone cancéreuse et la zone saine. Cette matrice a été séparée en un ensemble d'entraînement (80 %) et un ensemble de validation (20 %). Notre modèle CNN est construit comme une séquence de couches et a été mis en place en utilisant la bibliothèque Keras. La dernière couche est une couche de sortie dense avec 2 neurones et une fonction d'activation softmax, permettant de fournir les probabilités prédites pour chacune des 2 classes cibles. Le modèle est enregistré au format h5. Pour évaluer la performance du modèle CNN basé sur les données MALDI-MSI, 30 spectres aléatoires ont été prélevés à partir des données WALDI-MSI et utilisés pour interroger le modèle CNN, afin de déterminer leur classification correcte grâce à ce modèle.

### Analyse MS/MS

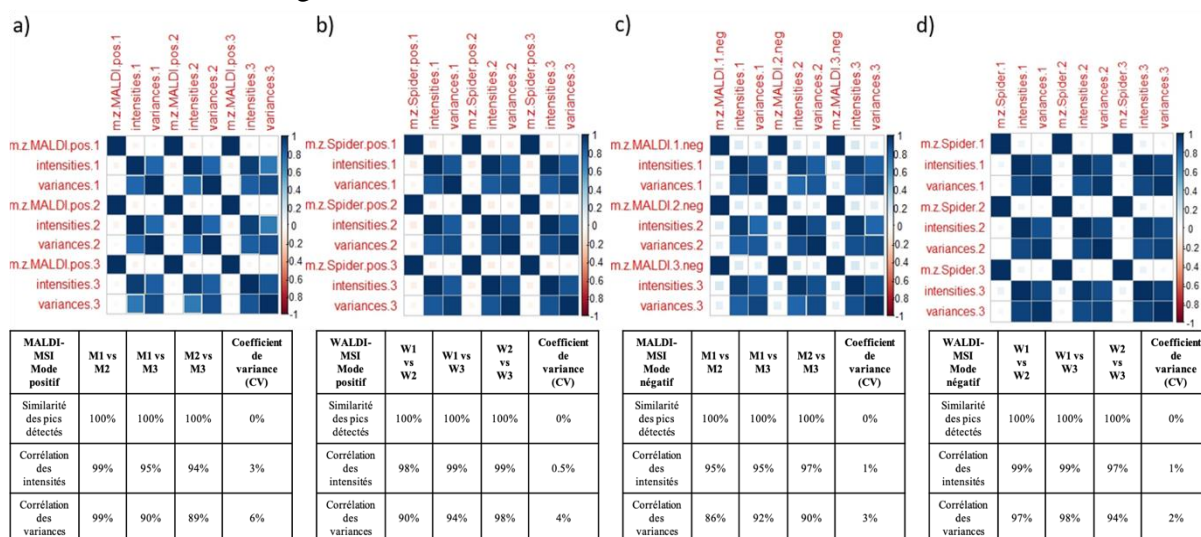
Chacun des ions lipidiques mentionnés a été soumis à une analyse de spectrométrie de masse en tandem (MS/MS) avec dissociation induite par collision (CID) à une énergie de collision normalisée située entre 30 et 40 eV. Les annotations concernant les lipides ont été établies de manière manuelle en se basant sur les informations des bases de données publiques Metlin, LipidMaps, Alex 123 ainsi que la littérature disponible<sup>187-189</sup>.

## Résultats et discussion

### Évaluation individuelle de la similarité des pics et des corrélations d'intensité et de variances des profils lipidomiques MALDI-MSI et WALDI-MSI

Pour commencer, nous avons évalué l'efficacité de notre nouvelle méthodologie de mesure de similarité et de corrélation en examinant des analyses en triplicat de tissus de cerveaux de rat effectuées en utilisant les modes positif et négatif pour les techniques d'imagerie MALDI et WALDI. Dans les deux modes d'ionisation, nous constatons que 100 % des pics ont été détectés entre chaque réplicat pour les analyses MALDI (M) et WALDI (W).

Les corrélations d'intensité sont hautement supérieures à 94 %, avec un coefficient de variance inférieur à 3 % pour les ensembles de données MALDI dans les deux modes. Tandis que les corrélations de variances sont toutes supérieures à 85 %, avec un coefficient de variance pouvant atteindre 6 % (**Figure 30A-C**). Pour les analyses en triplicat WALDI-MSI, une corrélation d'intensité légèrement plus élevée, dépassant les 97 %, est observée, avec un coefficient de variance inférieur à 1 %. Pour la corrélation des variances, elle est toujours supérieure à 90 %, avec un coefficient de variance n'excédant pas 4 % (**Figure 30B-D**). Indépendamment des différentes régions anatomiques des coupes de tissu de cerveau de rat, notre méthodologie a démontré une reproductibilité analytique et moléculaire élevée. Ces résultats mettent en évidence la robustesse et la précision de notre approche pour une évaluation approfondie de la corrélation et de la similarité entre les deux technologies.

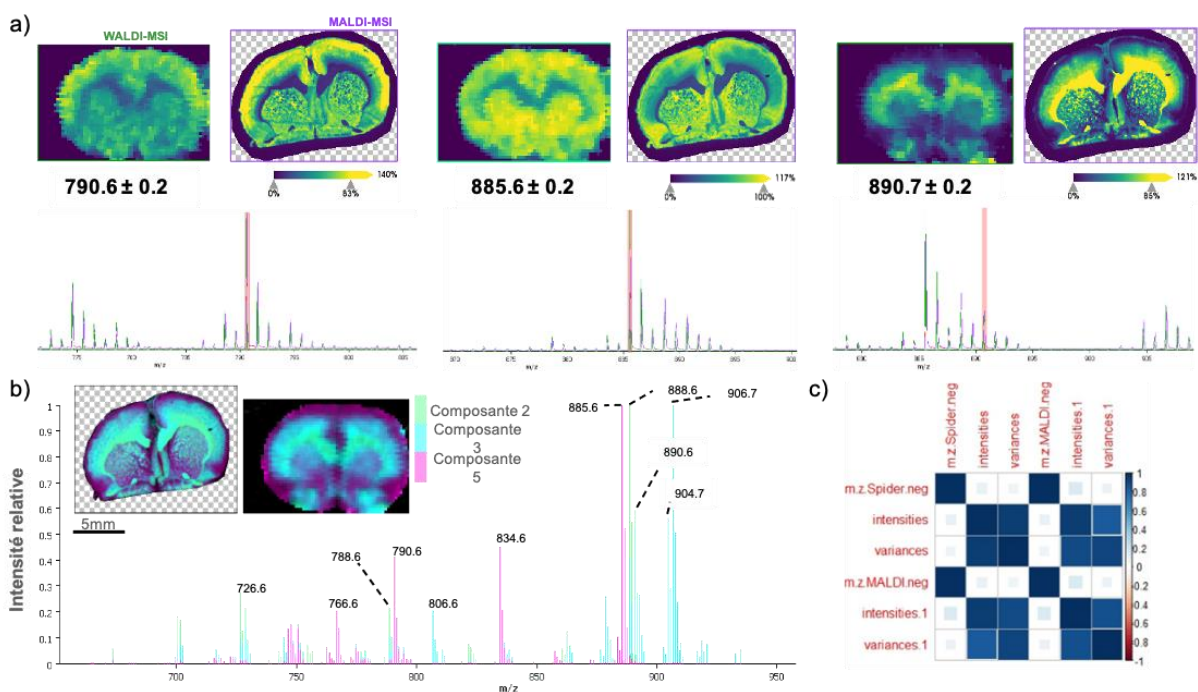


**Figure 30** : Réplicats MALDI-MSI et WALDI-MSI du cerveau de rat dans les deux modes d'ionisation. La matrice de corrélation d'intensité de Pearson et le tableau récapitulatif de la similarité des pics détectés, calculs de corrélation et coefficient de variance de chaque réplicat pour (a) MALDI-MSI en mode positif, (b) WALDI-MSI en mode positif, (c) MALDI-MSI en mode négatif et (d) WALDI-MSI en mode négatif.



## Similarité de la détection des pics et corrélation des intensités et variances entre MALDI et WALDI-MSI

Dans le contexte de la technologie SpiderMass, l'eau endogène remplit le rôle de matrice, tandis que dans le MALDI, il est bien établi que différentes familles ou sous-familles de molécules peuvent être détectées de manière plus optimale en modifiant la matrice utilisée. En conséquence, nous avons mené des expériences en utilisant les quatre matrices les plus couramment utilisées pour l'analyse MALDI-MSI des lipides, à savoir les matrices norharmane, 9-AA, 2,5-DHB et DAN. Les espèces de glycérophospholipides sont généralement observées dans la gamme de masse 600-1000. Les matrices DAN et 9-AA ont été utilisés dans les deux modes d'ionisation tandis que la matrice 2,5-DHB a été utilisée uniquement en mode positif et la matrice 9-AA uniquement en mode négatif, car il est connu que ces deux matrices sont plus spécifiques à l'un des deux modes<sup>190-192</sup>.

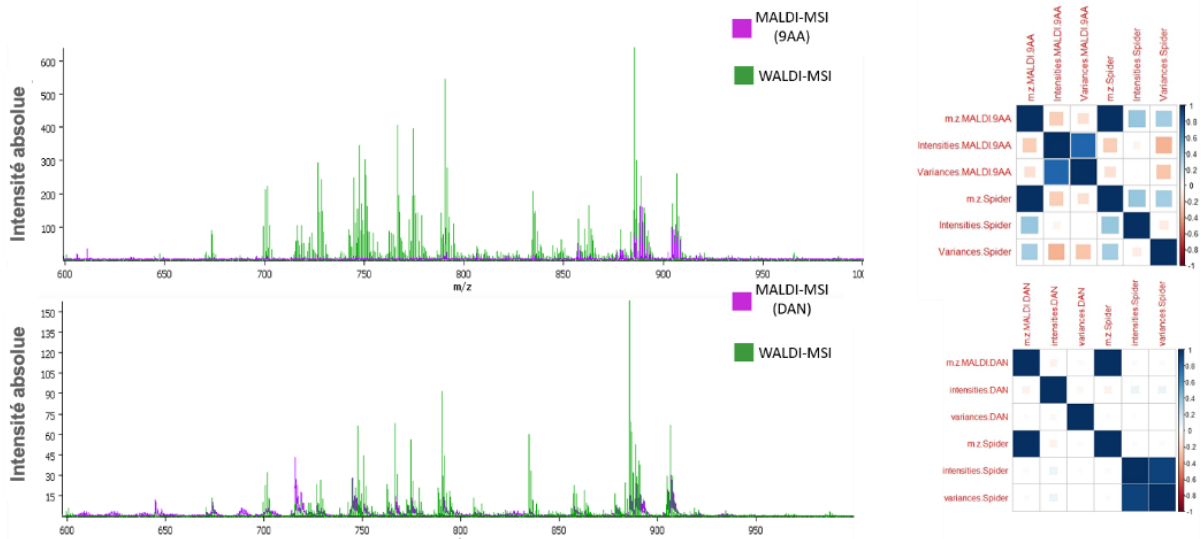


**Figure 31** : WALDI-MSI et MALDI-MSI (matrice norharmane) du premier réplicat de cerveau rat en mode négatif. (a) Images d'ions sélectionnés et zoom spectral à  $m/z$  790.6, 885.6 et 890.7. (b) Résultat de l'analyse pLSA. (c) La matrice de corrélation de Pearson pour les intensités et les variances.

Lors de l'analyse en mode négatif du premier cerveau de rat, on peut voir que le spectre MALDI-MSI obtenu avec la matrice norharmane présente un chevauchement significatif avec le spectre WALDI-MSI dans la gamme de masse des lipides entre 600-1000  $m/z$ . Cette observation est renforcée par les images et les spectres zoomés des ions sélectionnés  $m/z$  790.6 PE (18:0\_22:6),  $m/z$  885.6 PI (18:0\_20:4) et  $m/z$  890.7 SHexCer (d18:1\_24:0) (**Figure 31A**). Pour mieux comprendre les signaux spectraux résultant de caractéristiques spatiales particulières dans les deux sections du cerveau, nous avons utilisé la technique pLSA. Les composantes 2, 3 et 5 de la pLSA permettent de décrire les principaux éléments contribuant à la matière grise et à la matière blanche. Les ions  $m/z$  885.6,  $m/z$  888.6 SHexCer (d18:1\_24:1) et  $m/z$  906.7 SHexCer (42:1;3) présentent la



plus grande contribution à chaque composante (**Figure 31B**). Lorsque l'on regarde la similarité de détection de pics, aucune divergence des pics n'est observée. De plus, les corrélations de Pearson d'intensité et de variances atteignent 95 % et 92 % respectivement (**Figure 31C**). Après analyse de trois cerveaux de rat (triplicat biologique), la similarité de détection de pics atteint 100 % pour tous, tandis que la moyenne de corrélation de Pearson des intensités et des variances est de 93 % et 87 % (**Tableau 5**).

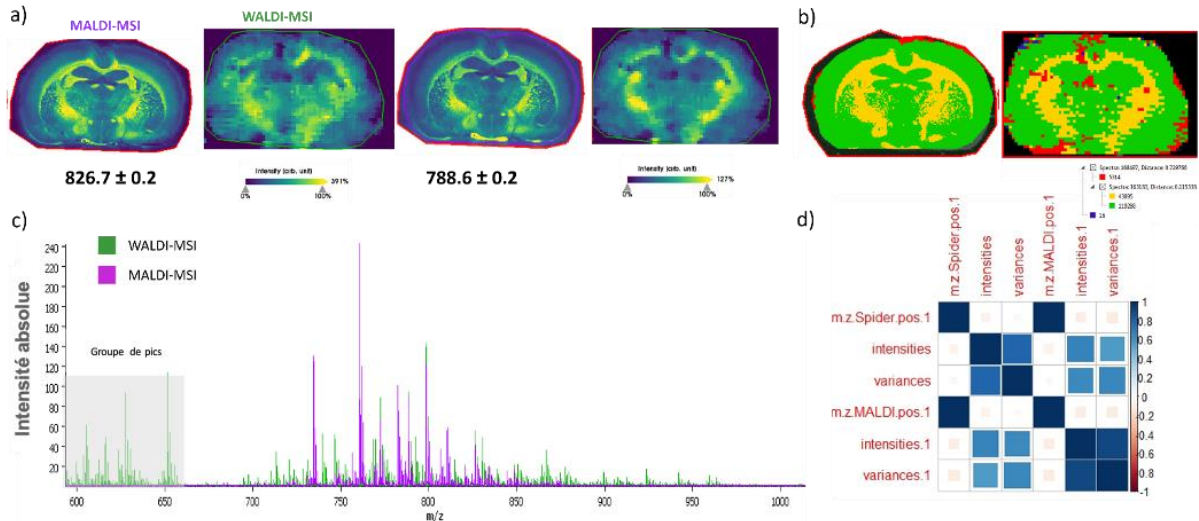


**Figure 32** : Spectres WALDI-MSI (vert) et MALDI-MSI (matrice DAN et 9-AA) (violet) d'une section de cerveau de rat en mode négatif ainsi que les matrices de corrélation de Pearson correspondantes.

Lors de l'utilisation de la matrice 9-AA, une similarité de seulement 84 % est observée entre les données MALDI et WALDI, tandis que la similarité des pics entre le WALDI et le MALDI est de seulement 48 % (**Figure 32**). Cela reflète l'utilisation traditionnelle de la matrice 9-AA pour l'imagerie des métabolites dans une gamme de masse plus basse. En ce qui concerne la matrice DAN, une similarité de 77 % des pics est observée, mais seulement 7 % de corrélation d'intensité (**Figure 32**).

En mode positif, le spectre WALDI-MSI se distingue visuellement de celui obtenu par MALDI-MSI par des différences marquées, où le spectre WALDI-MSI présente un profil riche, notamment dans la gamme de masse 600-700 (**Figure 33C**). Cette observation est corroborée par la similarité des pics détectés et par les calculs de corrélation de Pearson. En effet, la similarité des pics détectés atteint 100 % entre MALDI-MSI et WALDI-MSI, mais n'est que de 63.62 % entre WALDI-MSI et MALDI-MSI (**Figure 33D**). Cela semble être dû à la présence de plusieurs diglycérides dans les spectres SpiderMass dans la gamme de masse  $m/z$  600-700. L'analyse de corrélation montre une corrélation de seulement 52 % pour les intensités et 61 % pour les variances des mêmes pics détectés entre MALDI-MSI et WALDI-MSI. Malgré ce faible degré de corrélation, la distribution et la localisation des ions dans le tissu de cerveau de rat sont les mêmes pour les deux technologies. En particulier, l'ion  $m/z$  772.6 PC 32:0 est plus abondant dans la matière grise,

tandis que les ions  $m/z$  788.6 PC 36:1 et  $m/z$  826.7 PC 36:1 (adduit potassium) sont spécifiquement présents dans le corps calleux (**Figure 33A**). De plus, les résultats de segmentation obtenus pour l'analyse MALDI-MSI et WALDI-MSI sont similaires, le cluster jaune semble spécifique du corps calleux et le cluster vert spécifique de la matière blanche (**Figure 33B**). Après avoir analysé un triplicat biologique de cerveau de rat, la similarité de pics détectés MALDI-MSI vs WALDI-MSI atteint 83 % tandis que l'inverse est égal à 77 %. Pour la corrélation des intensités et des variances, on obtient une moyenne de 77 % et 65 % respectivement (**Tableau 5**).



**Figure 33** : WALDI-MSI et MALDI-MSI (matrice norharmane) d'un tissu de cerveau de rat en mode positif. (a) Images d'ions sélectionnés à  $m/z$  788.6 et 826.7. (b) Segmentation et dendrogramme de l'ensemble des données WALDI-MSI et MALDI-MSI. (c) Spectres WALDI-MSI (vert) et MALDI-MSI (violet). La boîte grise indique un groupe de pics présents uniquement dans les données WALDI-MSI. (d) Matrice de corrélation de Pearson pour les intensités et les variances des données WALDI-MSI et MALDI-MSI.

Lors de l'utilisation de la matrice DHB, la similarité entre MALDI-MSI et WALDI-MSI atteint 98 %. En revanche, la similarité entre WALDI-MSI et MALDI-MSI est plus basse, à 51 %. De plus, les calculs de corrélation de Pearson indiquent une corrélation plus faible de 34 % et 18 % avec la matrice DHB (**Figure 34**). En effet, on sait que la matrice DHB présente une couverture lipidique moindre par rapport à la matrice norharmane<sup>131</sup>, et donc une similarité plus basse dans les échantillons de cerveau de rat. La sensibilité et la couverture moléculaire réalisables par MALDI-MSI sont en fin de compte influencées par les efficacités d'ionisation (fraction d'ions par rapport aux neutres désorbés), qui sont relativement modestes et varient d'une matrice à l'autre. Les résultats sont encore plus faibles lors de l'utilisation de la matrice DAN. En effet, la similarité des pics détectés est de 75 % et la corrélation des intensités et des variances n'excèdent pas 10 % et 1 % respectivement (**Figure 34**).

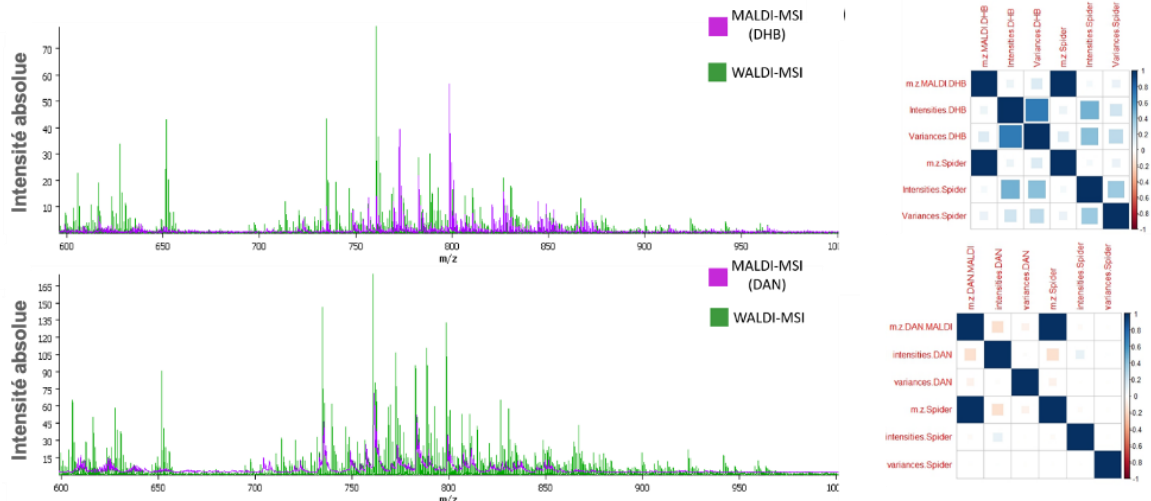


Figure 34 : Spectres WALDI-MSI (vert) et MALDI-MSI (matrice DAN et DHB) (violet) d'une section de cerveau de rat en mode positif ainsi que les matrices de corrélation de Pearson correspondante.

### Effet de la taille du spot laser et de l'épaisseur du tissu sur les profils moléculaires

Jusqu'à présent, les deux technologies, MALDI et WALDI, ne peuvent pas être utilisées avec la même taille de spot laser, car la résolution spatiale WALDI-MSI est limitée au diamètre du faisceau laser incident, dans cette étude autour de 400 à 500  $\mu\text{m}$ . Cela suggère que la comparaison entre MALDI et WALDI-MSI pourrait être affectée par des différences dans la quantité de matériau analysé.

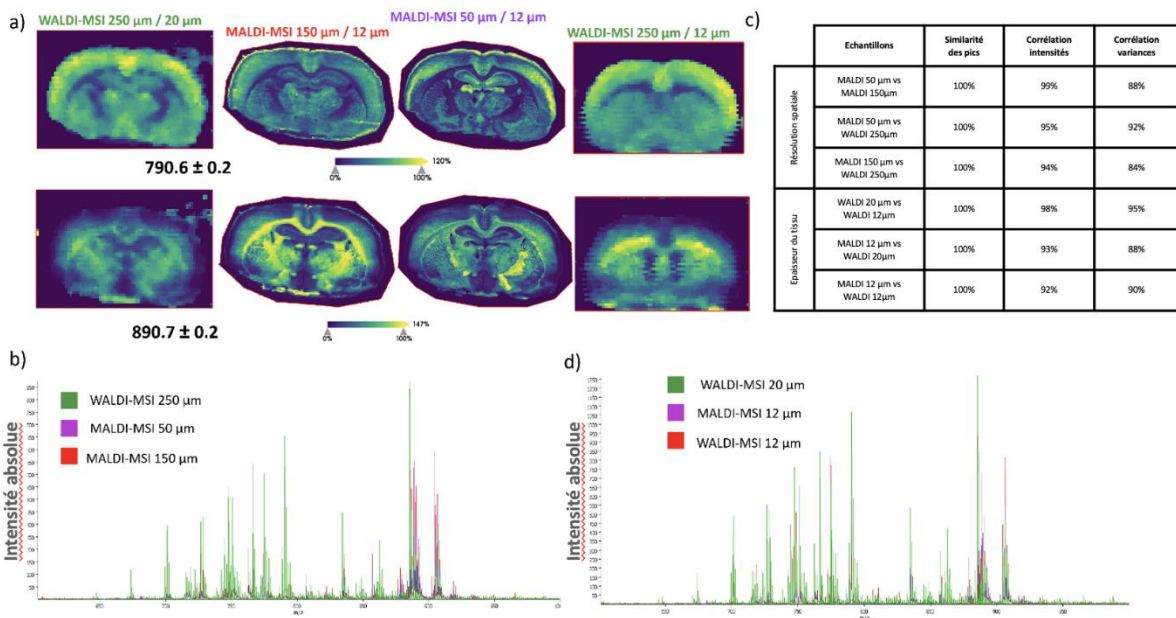


Figure 35 : WALDI-MSI et MALDI-MSI (matrice norharmane) pour différentes résolutions spatiales et épaisseurs de tissu en mode négatif. (a) Images d'ions sélectionnés à  $m/z$  790.6 et  $m/z$  890.7. (b) Spectres WALDI-MSI 250  $\mu\text{m}$  (vert), MALDI-MSI 50  $\mu\text{m}$  (violet) et MALDI-MSI 150  $\mu\text{m}$  (rouge) d'une section de cerveau de rat analysée à différentes résolutions spatiales. (c) Spectres WALDI-MSI 20  $\mu\text{m}$  (vert), MALDI-MSI 12  $\mu\text{m}$  (violet) et WALDI-MSI 12  $\mu\text{m}$  (rouge) d'une section de cerveau de rat de différentes épaisseurs. (d) Tableau récapitulatif de la similarité spectrale des pics détectés ainsi que des corrélations de variances et intensités.

Pour évaluer l'impact de la taille des spots laser sur les calculs de similarité et de corrélation, une analyse MALDI-MSI a été réalisée à une résolution spatiale réduite de 150  $\mu\text{m}$  en mode négatif. Comme le montrent les images des ions sélectionnés  $m/z$  790.6 et  $m/z$  890.7, la distribution de l'ion demeure constante, indépendamment de la résolution spatiale sélectionnée lors de l'imagerie MALDI (**Figure 35A**). Ceci est confirmé par la similarité des pics entre les deux méthodes qui est de 100 %, ainsi que ceux des calculs de corrélation de Pearson des intensités et variances. En effet, la corrélation d'intensité et de variance entre MALDI et WALDI-MSI diffère seulement d'1 % et de 8 %, respectivement, lorsque les données MALDI sont collectées à une résolution spatiale de 150  $\mu\text{m}$  ou de 50  $\mu\text{m}$  (**Figure 35C**). Des résultats similaires sont aussi observés lors de la comparaison de différentes épaisseurs de tissu. En effet, une analyse WALDI-MSI a été réalisée en mode négatif sur une coupe de tissu de 12  $\mu\text{m}$  au lieu de 20  $\mu\text{m}$ . La similarité des pics reste la même, et les corrélations d'intensité et de variances sont supérieures à 92 % et 89 % entre les épaisseurs de tissu de 12  $\mu\text{m}$  et 20  $\mu\text{m}$  (**Figure 35C**). Ces observations indiquent que la résolution spatiale et l'épaisseur du tissu n'ont qu'une influence limitée sur la similarité des pics ainsi que sur la corrélation d'intensité et de variance entre les deux modalités d'imagerie.

#### Analyse comparative de tissus de cancer œsogastrique

Bien que les coupes de tissus de cerveaux de rat aient servi de référence pour les analyses MALDI et WALDI-MSI, nous avons également cherché à confirmer notre hypothèse en utilisant d'autres tissus biologiques, tels que des échantillons de cancer œsogastrique humain. Dans cette optique, nous avons opté exclusivement pour l'utilisation de la matrice norharmane, car elle présente la plus grande similarité avec l'analyse WALDI-MSI. Nous avons utilisé deux sections de tissus hétérogènes provenant de carcinomes œsogastriques peu cohésifs pour notre analyse.

La première section de tissu contient une partie tumorale à hauteur de 10 % et une grande partie saine. La comparaison visuelle des images d'ions sélectionnés ainsi que des spectres révèlent une grande similitude entre les approches WALDI et MALDI-MSI (**Figure 36A**). Plus précisément, les ions  $m/z$  857.5 PI 36:4,  $m/z$  863.6 PI (18:1\_18:0) et  $m/z$  883.5 PI (18:1\_20:4) présentent non seulement la même distribution dans les deux zones, mais également des localisations spécifiques au sein du tissu. En effet, l'ion à  $m/z$  863.6 semble être caractéristique de l'épithélium, tandis que l'ion à  $m/z$  857.5 est prédominant dans le tissu muqueux. L'analyse des spectres révèle une similarité de détection des pics de 100 % ainsi qu'une corrélation d'intensité de 91 % et une corrélation de variance de 75 %.





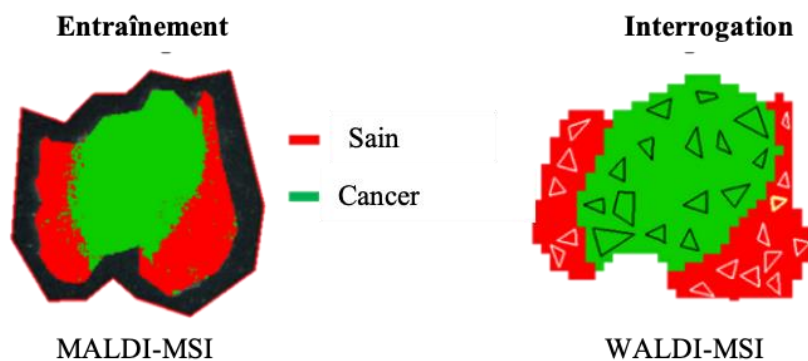
Une exploration plus approfondie des données d'imagerie en utilisant une analyse ROC (AUC supérieure à 0.85) a identifié uniquement 2 pics discriminatoires en MALDI-MSI et 5 pics discriminatoires en WALDI-MSI. Les pics  $m/z$  766.5 PE (18:0\_20:4) et  $m/z$  788.6 PS (18:1\_18:0) sont détectés dans les données WALDI, mais avec une intensité plus faible. En revanche, les pics  $m/z$  642.5 HexCer (30:1;2) et  $m/z$  687.6 PA (P-20:0\_16:0) ou (O-18:0\_18:1) ne sont pas observés dans les données MALDI-MSI. Cependant, les pics  $m/z$  701.5 PA (18:1\_18:0),  $m/z$  750.5 PE (16:0\_20:4) et  $m/z$  853.6 PI 36:6 y sont trouvés, mais avec une intensité réduite.

La deuxième section de tissu cancéreux analysée présente une importante zone tumorale ainsi que deux zones péri-tumorales de chaque côté de celle-ci. De la même manière que pour la première section de tissu, les distributions d'ions sélectionnés sont similaires entre les images WALDI et MALDI (**Figure 36B**). Les ions  $m/z$  857.5 et  $m/z$  863.6 sont spécifiques au tissu sain, tandis que l'ion  $m/z$  797.6 PG (18:1\_20:3) ou PG (18:0\_20:4) est présent uniquement dans le tissu cancéreux. Néanmoins, les spectres révèlent des différences d'intensité plus marquées. Bien que la similarité des pics détectés atteigne 100%, la corrélation des intensités n'est que de 77 %, et celle des variances de 67 %.

#### Modèle CNN basé sur les données MALDI-MSI pour la prédiction des données WALDI-MSI

Pour finir, nous avons réalisé une évaluation afin de déterminer si la forte similarité entre les données issues d'imagerie MALDI et WALDI permettait de mettre en place un entraînement avec les données MALDI pour ensuite interroger à l'aveugle les données WALDI.

À cette fin, nous avons construit un modèle CNN à deux classes en utilisant les données MALDI-MSI, dans le but de discriminer les tissus cancéreux et les tissus sains. Par la suite, ce modèle a été utilisé pour prédire le caractère cancéreux ou non de 30 échantillons aléatoires provenant des données WALDI-MSI. Nous avons accompli une classification précise à hauteur de 93 % pour la prédiction des tissus cancéreux, avec une précision moyenne de 81 %, et de 80 % pour la prédiction des tissus sains, affichant une précision moyenne de 80 %. Ces résultats consolident la notion que les données MALDI-MSI et WALDI-MSI reflètent une base d'information fondamentale similaire dans le contexte de l'imagerie par spectrométrie de masse, et attestent de la capacité à réussir l'entraînement du modèle WALDI à partir des données MALDI (**Figure 37**).



	Classe prédite	Précision	Classe prédite	Précision
Echantillons \ Classe	Cancer		Sain	
1	Cancer	69%	Sain	95%
2	Cancer	75%	Cancer	83%
3	Cancer	53%	Sain	79%
4	Cancer	78%	Sain	85%
5	Cancer	98%	Cancer	72%
6	Cancer	77%	Sain	57%
7	Cancer	92%	Sain	71%
8	Cancer	95%	Sain	63%
9	Cancer	89%	Sain	97%
10	Cancer	77%	Sain	66%
11	Cancer	82%	Sain	97%
12	Cancer	95%	Sain	75%
13	Cancer	90%	Sain	88%
14	Cancer	54%	Cancer	85%
15	Sain	55%	Sain	84%
<b>Total</b>	<b>14/15</b>	<b>81%</b>	<b>12/15</b>	<b>80%</b>

Figure 37 : Modèle CNN basé sur les données MALDI-MSI et les données WALDI-MSI pour la prédiction des tissus cancéreux et sains à partir de la section n°2 du tissu gastrique.



## Conclusion et perspectives

Tout d'abord, une nouvelle méthodologie a été développée afin de faire une analyse comparative quantitative et précise de deux technologies de spectrométrie de masse. Le principe étant de regarder les similarités visuelles des spectres obtenus ainsi que les images de la distribution d'ions sélectionnés entre les deux techniques. Ceci suivi ensuite du calcul de similarité de pics détectés et de la corrélation de Pearson uniquement des mêmes pics détectés. Grâce à cette nouvelle méthodologie, nous avons pu démontrer que les technologies MALDI et WALDI sont très proches dans les modes positif et négatif en utilisant la matrice norharmane comme matrice pour le MALDI. Dans ces conditions, non seulement les profils lipidiques présentent une similitude de 100 % des pics détectés, mais une corrélation d'intensité élevée entre les deux technologies est également obtenue, ce qui confirme la similitude des processus. Nous avons obtenu des résultats similaires à partir de différents échantillons de tissus, à savoir un triplicat de cerveaux de rats mais aussi deux échantillons mixtes de cancer œsogastrique. Les pourcentages obtenus pour l'analyse de similarité de pics détectés et de corrélation des intensités et des variances pour le triplicat biologique de cerveau de rat dans les deux modes d'ionisation ainsi que pour les deux tissus mixtes de cancer œsogastrique, sont disponibles dans le **Tableau 5**.

En résumé, notre étude a mis en évidence une forte similitude entre le nouveau WALDI-MSI et le MALDI-MSI pour l'analyse lipidique. Nous avons non seulement détecté les mêmes espèces moléculaires, mais nous avons également observé une corrélation significative entre les intensités et les variances des spectres obtenus par les deux techniques. Cette corrélation est particulièrement remarquable en mode ion négatif, et elle est favorisée par l'utilisation de la norharmane comme matrice MALDI. Malgré les différences de résolution entre les deux méthodes, nous avons réussi à identifier la même hétérogénéité tissulaire. Dans le cas du cancer gastrique, nous avons pu différencier les régions tumorales et péri-tumorales, ainsi que les cellules cancéreuses denses, l'épithélium et le tissu muqueux. Nos résultats démontrent clairement que la technique WALDI-MSI est une nouvelle méthode prometteuse d'imagerie par ionisation ambiante. Pour l'avenir, nous prévoyons de poursuivre le développement et l'amélioration de cette méthode. En particulier, nous souhaitons exploiter les données MALDI-MSI pour accélérer et améliorer la formation WALDI-MSI, en utilisant notamment des techniques d'apprentissage automatique supervisé adaptées à de grandes quantités de données.

**Tableau 5** : Calculs de la similarité des pics détectés et des corrélations pour tous les tissus, matrices et modes d'ionisation.

<b>Echantillon/matrice/ mode d'ionisation</b>	<b>Similarité MALDI vs. WALDI</b>	<b>Similarité WALDI vs. MALDI</b>	<b>Corrélation des intensités</b>	<b>Corrélation des variances</b>
RB1/Norharmane/Négatif	100 %	100 %	95 %	92 %
RB2/Norharmane/Négatif	100 %	100 %	91 %	82 %
RB3/Norharmane/Négatif	100 %	100 %	93 %	88 %
RB/9AA/Négatif	84 %	48 %	5 %	3 %
RB/DAN/Négatif	77 %	77 %	7 %	5 %
RB1/Norharmane/Positif	83 %	78 %	76 %	75 %
RB2/Norharmane/Positif	81 %	75 %	81 %	64 %
RB3/Norharmane/Positif	85 %	77 %	74 %	55 %
RB/DHB/Positif	98 %	51 %	34 %	18 %
RB/DAN/Positif	75 %	75 %	9 %	0.9 %
Tissu œsogastrique n°1 /Norharmane/Négatif	100 %	100 %	91 %	75 %
Tissu œsogastrique n°2 /Norharmane/Négatif	100 %	100 %	77 %	67 %

Cette étude a été d'un grand intérêt pour nous car il est ainsi légitime d'utiliser les données MALDI-MSI en plus des données WALDI pour entraîner l'outil de diagnostic SpiderMass afin de créer des modèles de classification beaucoup plus performants et précis. Sachant que les données MALDI-MSI peuvent être collectées en grandes quantités et permettent alors de palier à notre manque récurrent de nombres d'échantillons disponibles dans les cohortes de tissus de patient.

De plus, en plus d'améliorer la fiabilité des modèles de classification en fournissant des données complémentaires à celles obtenues par SpiderMass, l'utilisation de l'imagerie MALDI a plusieurs applications : elle peut servir à annoter les tissus avant d'effectuer une analyse SpiderMass ainsi qu'à valider les marqueurs découverts par SpiderMass,

Cette recherche a suscité un vif intérêt au sein de notre équipe, car elle a établi la légitimité d'incorporer des données MALDI-MSI en complément des données WALDI pour améliorer considérablement les performances et la précision de l'outil de diagnostic SpiderMass. Cette approche est particulièrement pertinente car les données MALDI-MSI peuvent être recueillies en grande quantité, ce qui permet de surmonter la limitation fréquente liée au nombre limité d'échantillons disponibles dans les cohortes de tissus de patients.

L'objectif suivant est de développer une façon d'analyser les données acquises par le SpiderMass afin d'établir des modèles de classification très robustes et fiables. En effet, dans le domaine de la spectrométrie de masse axée sur le diagnostic du cancer, les algorithmes d'apprentissage automatique supervisés jouent un rôle essentiel. Le choix et l'ajustement du modèle de classification approprié ont un impact significatif sur la précision des résultats. Il existe une variété d'algorithmes d'apprentissage automatique disponibles, allant de l'apprentissage traditionnel à l'apprentissage profond. Historiquement, la technologie SpiderMass pour le diagnostic moléculaire s'appuyait sur la modélisation de l'analyse discriminante linéaire avec l'utilisation de l'outil de conception de modèle AMX. Cependant, cette approche présente certaines limitations, notamment une dépendance exclusive à la classification LDA, l'absence d'évaluation et de validation croisée du modèle, un manque d'explication des prédictions et l'absence de tests statistiques pour évaluer les variations des caractéristiques.

Dans le but de surpasser ces restrictions, nous avons développé un pipeline d'intelligence artificielle (IA) amélioré et plus exhaustif. Notre objectif est de créer des modèles de classification qui ne seraient pas seulement précis et fiables, mais qui permettraient également une identification claire et indiscutable des biomarqueurs lipidiques.



## **Chapitre 4 : Analyse en temps réel du microenvironnement tumoral du glioblastome grâce au SpiderMass : un outil pour guider les décisions chirurgicales et optimiser la prise en charge des patients**

---

### **Introduction**

Les tumeurs cérébrales et celles du système nerveux central se classent parmi les formes de cancer les plus létales, avec un bilan de 308 102 nouveaux cas diagnostiqués et 251 329 décès répertoriés à travers le globe en 2020<sup>1</sup>. Parmi ces tumeurs, les gliomes, constituant 80 à 85 % des cas, demeurent la variété prédominante parmi les affections cérébrales chez les adultes<sup>193</sup>. Émanant de divers types cellulaires, les gliomes se présentent comme un défi de taille dans leur classification. Toutefois, la précision de cette classification revêt une importance cruciale pour anticiper le pronostic des patients, tout en permettant de concevoir des plans de traitement optimaux et des approches thérapeutiques personnalisées<sup>194</sup>. Depuis 2016, l'Organisation mondiale de la santé préconise l'intégration des caractéristiques histologiques et des altérations moléculaires dans la classification des tumeurs du système nerveux central<sup>195</sup>. Au sein des gliomes diffus chez les adultes, le glioblastome (GBM) se distingue comme la variante la plus répandue et la plus agressive, affichant une survie globale d'environ 15 mois, avec une médiane oscillante entre 14 et 24 mois (en contraste avec les gliomes de grade inférieur qui peuvent être gérés sur des périodes de plusieurs décennies).

Le glioblastome se révèle être une maladie d'une grande hétérogénéité, créant ainsi des défis pour la conception de traitements efficaces<sup>196,197</sup>. Jusqu'à présent, la gestion du glioblastome repose sur une approche multimodale. En première ligne, cela englobe une résection chirurgicale maximale sécuritaire, suivie d'une chimiothérapie adjuvante à long terme utilisant le témozolomide associé à une radiothérapie (connu sous le nom de protocole de Stupp<sup>198</sup>), un schéma qui perdure depuis de nombreuses années. En effet, l'évolution des tumeurs cérébrales perturbe l'équilibre du système nerveux central, c'est-à-dire que le système vasculaire, la matrice extracellulaire et la barrière hémato-encéphalique sont tous mis à contribution dans les processus évolutifs de la tumeur<sup>199</sup>. Il est reconnu que le microenvironnement tumoral possède des propriétés immunosuppressives, et que l'infiltration de diverses sous-populations de cellules immunitaires est directement liée à la réponse au traitement et à la survie globale des patients.

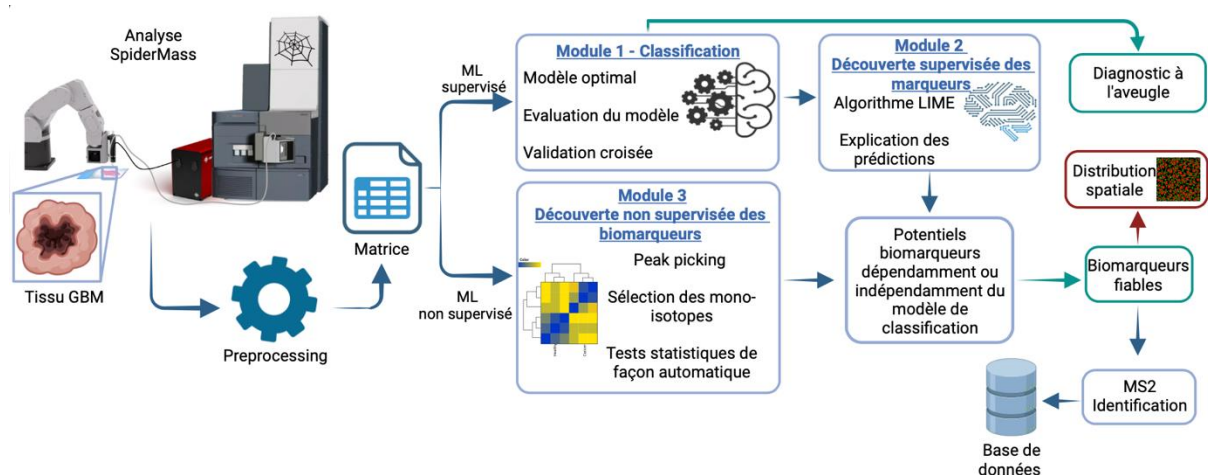
D'un autre côté, des preuves ont démontré que l'étendue de la résection est directement liée à la survie globale (SG), avec des résultats plus favorables associés à une résection totale brute par rapport à une résection subtotale ou à une biopsie<sup>200</sup>. Plus spécifiquement, il a été établi que l'augmentation de l'étendue de la résection à plus de 95 %, tout en préservant l'intégrité fonctionnelle des tissus, engendre des améliorations significatives dans les résultats chirurgicaux et la survie globale. Cependant, la réalisation d'une résection totale grossière reste complexe en raison de la nature diffuse du glioblastome, rendant ardue la délimitation précise des marges tumorales. C'est pourquoi, diverses méthodes ont été élaborées<sup>201</sup> pour faciliter la localisation des marges de résection, dont la navigation par imagerie par résonance magnétique (IRM), la chirurgie éveillée et l'utilisation de la spectroscopie Raman<sup>202-204</sup>. Par conséquent, la caractérisation moléculaire des tissus cérébraux pendant l'intervention chirurgicale, en vue de réaliser une résection sur mesure, ainsi qu'une meilleure appréhension du microenvironnement, pourraient devenir des éléments cruciaux dans la prise en charge neurochirurgicale des patients atteints de glioblastome<sup>205</sup>.

De plus, le glioblastome a été le sujet de nombreuses études omiques utilisant des approches fondées sur la spectrométrie de masse, en particulier la protéomique shotgun et l'imagerie par spectrométrie de masse. Diverses modalités de MSI, dont MALDI et DESI, ont été employées pour tracer des marqueurs bien établis (comme les métabolites 2-hydroxyglutarate associés aux mutations IDH)<sup>206</sup> et identifier des biomarqueurs potentiels et des cibles thérapeutiques pour cette maladie<sup>90,207</sup>. Cependant, la spectrométrie de masse a récemment évolué grâce au développement de nouvelles techniques d'ionisation en milieu ambiant pour l'évaluation peropératoire<sup>10</sup>.

Néanmoins, pour que les données de spectrométrie de masse soient utiles au praticien, elles doivent être associées à l'apprentissage automatique<sup>208</sup>. Ainsi, dans le contexte de la spectrométrie de masse axée sur le diagnostic du cancer, les algorithmes d'apprentissage automatique supervisés jouent un rôle crucial en établissant des modèles de classification précis basés sur des ensembles de données annotées<sup>209</sup>. Le choix et l'ajustement du modèle de classification adéquat peuvent profondément influencer la précision des résultats. Il est également essentiel de prêter une attention minutieuse à la validation croisée à travers des méthodes d'apprentissage automatique non supervisées, ainsi qu'à la compréhensibilité des résultats. Un éventail d'algorithmes d'apprentissage automatique est à disposition, couvrant de l'apprentissage traditionnel jusqu'à l'apprentissage profond. Le choix du modèle de modélisation aura un impact direct sur la précision des modèles élaborés. Historiquement, le diagnostic moléculaire de la technologie SpiderMass reposait sur la modélisation de l'analyse discriminante linéaire, avec l'usage de l'outil de conception de modèle AMX. Cependant, cette approche présente certaines limites, notamment une dépendance exclusive

envers la classification LDA, l'absence d'évaluation et de validation croisée du modèle, le manque d'explication des prédictions ainsi que l'absence de tests statistiques pour évaluer les variations des caractéristiques. Par ailleurs, les modèles d'apprentissage automatique ont été critiqués pour leur manque de transparence et leur incapacité à expliquer le processus qui conduit à un résultat spécifique, ou encore la raison pour laquelle ce résultat est obtenu<sup>210,211</sup>.

Dans ce contexte, afin de surmonter ces limitations, nous avons élaboré un nouveau pipeline d'intelligence artificielle plus complet (**Figure 38**). L'objectif étant de fournir des modèles de classification non seulement précis et fiables, mais également de permettre une identification sans équivoque des biomarqueurs. Une cohorte rétrospective de 50 tissus FFPE et une cohorte prospective de 31 tissus frais de glioblastome ont été analysées par la technologie SpiderMass et utilisées pour le développement de cette nouvelle méthodologie d'IA d'analyse statistique des données MS. Cette approche vise à approfondir la compréhension du microenvironnement tumoral et à identifier des biomarqueurs fiables. La nouvelle méthodologie d'intelligence artificielle a démontré une précision dépassant les 90 % suite à des validations croisées. Cette précision concerne non seulement le diagnostic des tissus, mais également la sous-stratification et le pronostic des tissus affectés par le glioblastome. Ces résultats mettent en évidence le potentiel remarquable de l'association entre la technologie SpiderMass et l'IA. Cette synergie offre des perspectives prometteuses pour l'amélioration du diagnostic, du pronostic et des choix thérapeutiques destinés aux patients concernés.



**Figure 38** : Pipeline global pour le diagnostic moléculaire par spectrométrie de masse ambiante combiné à la prédiction par intelligence artificielle. Les flèches bleues sont obligatoires pour obtenir les résultats finaux (flèches vertes) qui sont le diagnostic à l'aveugle des tissus et l'identification des biomarqueurs confiants. Une étape facultative (flèche rouge) peut être ajoutée au pipeline comme la distribution spatiale des biomarqueurs confiants grâce à l'analyse par WALDI-MSI des tissus.



## Matériel et méthodes

### Cohorte Gliomic

Une cohorte rétrospective composée de 50 tissus de glioblastome FFPE a été obtenue auprès du service de pathologie de l'hôpital de Lille, dont les données démographiques et cliniques sont résumées dans le **Tableau 6**. De plus, une cohorte prospective de 31 tissus de glioblastome frais congelés a été incluse dans l'étude. Entre septembre 2014 et novembre 2018, 31 patients récemment diagnostiqués avec un glioblastome ont été recrutés de manière prospective au CHU de Lille, France (NCT02473484). Tous les patients ont fourni leur consentement éclairé par écrit avant leur inscription.

**Tableau 6** : Données démographiques et cliniques de la cohorte rétrospective.

		Cohorte rétrospective (n = 50)
	Age, médiane (min-max)	64.1 (45-82)
Sexe	Femme, <i>n</i>	20
	Homme, <i>n</i>	30
Location de la tumeur	Frontale, <i>n</i>	15
	Occipitale, <i>n</i>	3
	Pariétale, <i>n</i>	10
	Temporale, <i>n</i>	22
Statut de méthylation du promoteur MGMT	Non méthylé, <i>n</i>	36
	Méthylé, <i>n</i>	14
Mort	Oui, <i>n</i>	49
	Non, <i>n</i>	1
Survie globale	SG < 15 mois, <i>n</i>	24
	SG > 15 mois, <i>n</i>	26
	SG < 10 mois, <i>n</i>	14
	SG > 36 mois, <i>n</i>	9

### Coloration histologique et annotations

Les blocs de paraffine ainsi que les tissus récemment congelés ont été découpés en sections de 7 µm et déposés sur des lames préalablement enduites de poly-lysine. Cette section de tissu consécutive à celle analysée par SpiderMass a subi un traitement à l'hémalum pendant 1 minute, suivi d'un rinçage à l'eau courante. Pour la coloration, la section a été immergée dans une solution de phloxine à 0.1 % pendant 10 secondes, puis déshydratée dans des bains successifs d'éthanol à 70 % et à 100 %. Ensuite, la section a été nettoyée au xylène, deux fois à l'alcool, plongée dans une solution de safran pendant 5 secondes, et enfin, montée avec des lamelles en utilisant le milieu de montage pour lames EUKITT. En utilisant le scanner de lames Panoramic MIDI (3DHISTECH LTD. Budapest, Hongrie), la lame colorée a été numérisée pour obtenir une image numérique. Ces images ont été examinées et extraites à l'aide de QuPath 0.2.3. Les annotations par le pathologiste ont été faites sur les tissus ayant été colorés.

## Préparation des échantillons

Tous les blocs FFPE ont été découpés en tranches tissulaires de 8  $\mu\text{m}$  à l'aide d'un microtome (Leica Biosystems, Wetzlar, Allemagne) maintenu à température ambiante. Un processus de déparaffinage a été effectué en plongeant les coupes de tissu dans une solution de xylène à deux reprises pendant 5 minutes. Par la suite, chaque coupe de tissu déparaffiné de la cohorte rétrospective de glioblastome a été imprégnée manuellement d'une solution de glycérol/alcool isopropylique (IPA) (8:2, v/v) en deux passages consécutifs, en utilisant un pulvérisateur manuel (Agilent). Le débit du pousse-seringue (série 74900 Cole Parmer Instrument Company) a été ajusté à 500  $\mu\text{L}/\text{min}$ . Quant aux tissus frais congelés, un cryostat Leica CM1510S (Leica Microsystems, Nanterre, France) a été employé pour découper des sections de 20  $\mu\text{m}$ . L'analyse de ces sections avec le SpiderMass ne nécessite aucune préparation préalable de l'échantillon.

## Analyse SpiderMass

La configuration générale de l'appareil a été précédemment discutée<sup>212</sup>. En somme, le système se compose de trois éléments : le spectromètre de masse en lui-même, un système laser pour le micro-échantillonnage à distance des tissus, et une ligne de transfert permettant le déplacement du matériel micro-échantillonné. Le premier élément comprend un laser pulsé Nd:YAG (durée d'impulsion : 5 ns, Quantel, Les Ulis, France) qui alimente un OPO à longueur d'onde réglable (Radiant version 1.0.1, OPOTEK Inc., Carlsbad, CA, USA). Une pièce à main avec une lentille de focalisation de 4 cm est attachée à l'extrémité de la fibre laser biocompatible, qui est reliée à la sortie du système laser et possède un diamètre intérieur de 450 microns et une longueur de 1 m. Pour cette étude, l'intensité du laser a été réglée à 4 mJ/impulsion pour une durée d'irradiation fixe de 10 s, générant une fluence laser d'environ 3 J/cm<sup>2</sup>. Le deuxième élément du système consiste en une ligne de transfert de 2 m de tube Tygon ND 100-65 (Akron, Ohio, États-Unis). Cette ligne de transfert est directement connectée au spectromètre de masse (Xevo, Waters, Manchester, Royaume-Uni), dont la source conventionnelle électrospray a été retirée et remplacée par une interface REIMS d'un côté, et est attachée à la pièce à main laser de l'autre côté. Une infusion d'isopropanol à un débit de 200  $\mu\text{L}/\text{min}$  a été effectuée pendant chaque acquisition. Une concentration de 200  $\mu\text{g}/\text{mL}$  de leucine enképhaline a été ajoutée à l'infusion en tant que lockmass. L'acquisition consiste en une série de 10 tirs laser, chacun donnant lieu à un spectre individuel. Cette acquisition spectrale a été réalisée en mode ion positif et négatif, en mode sensibilité, avec un temps de balayage d'1 seconde. La gamme de masse analysée est de  $m/z$  50 à 2000.

## WALDI-MSI

La mise en place du système SpiderMass a été décrite dans la section précédente. Pour faire de l'analyse d'imagerie, la microsonde Spider-Mass a été connectée à un bras robotique rigide décrit dans une publication antérieure<sup>18</sup>. La résolution spatiale a été fixée à 250  $\mu\text{m}$ . Pour chacun d'entre eux, la résolution spatiale finale a été améliorée grâce à un algorithme d'interpolation. La gamme de masse a été réglée entre  $m/z$  100 et 1500. La séquence d'acquisition comprend 3 tirs laser consécutifs avec un intervalle de 3 secondes entre chaque étape. Les tirs laser ainsi que l'acquisition du spectromètre de masse ont été déclenchés automatiquement par une interface utilisateur MATLAB spécialement conçue<sup>18</sup>. Les données ont été acquises en mode ion négatif et en mode sensibilité.

### Lignées cellulaires et analyse des cellules immunitaires

La culture des lignées cellulaires et leur analyse a été effectuée par Lucas Roussel, doctorant au laboratoire PRISM. Les cellules humaines de gliome de stade IV, provenant de la lignée NCH82, ont été fournies par le Dr Regnier-Vigouroux. Ces cellules ont été cultivées dans un milieu de culture DMEM. Le milieu a été complété avec 10 % de sérum bovin fœtal et 100 U/mL de pénicilline-streptomycine, et les cultures ont été maintenues dans un incubateur à 37°C avec une atmosphère de 5 % de CO<sub>2</sub>. Lorsque les cellules ont atteint une confluence de 70 %, elles ont été lavées deux fois avec une solution de DPBS, puis séchées à température ambiante sous PSM pendant 10 minutes, avant d'être soumises à une analyse directe par SpiderMass dans la plaque cellulaire.

En ce qui concerne la génération de macrophages primaires, 50 mL de sang ont été obtenus de l'Établissement Français du Sang. Le sang a été dilué deux fois dans une solution de PBS-EDTA, et les leucocytes ont été isolés en effectuant une centrifugation à 2200 rpm pendant 25 minutes sur un gradient de densité de Ficoll. Après trois lavages des leucocytes avec une solution de PBS-EDTA, ils ont été resuspendus dans un milieu RPMI et incubés dans une plaque cellulaire pendant 1 heure et 30 minutes à 37°C. Les plaques cellulaires ont ensuite été lavées trois fois avec du PBS. Les macrophages obtenus ont été cultivés pendant 7 jours dans un milieu RPMI contenant 10 % de sérum bovin fœtal, 100 U/mL de pénicilline-streptomycine et du MCSF. Ensuite, les macrophages primaires ont été soumis à des stimulations dans deux conditions distinctes. Les macrophages de type M1 ont été stimulés avec 0.5 mg/mL de LPS et 20 ng/mL d'IFN- $\gamma$  pendant 48 heures, tandis que les macrophages de type M2 ont été stimulés avec 20 ng/mL d'Il<sub>4</sub> pendant la même durée. Tout comme les cellules NCH82, les macrophages ont été lavés deux fois avec une solution de DPBS, séchés sous PSM pendant 10 minutes à température ambiante, puis directement analysés par SpiderMass dans la plaque cellulaire.

De plus, une isolation de lymphocytes primaires a été effectuée. Pour ce faire, les cellules mononucléaires du sang périphérique (PBMC) ont été séparées à partir d'échantillons de sang total par une centrifugation avec un gradient de densité de Ficoll. Ensuite, les cellules ont été marquées avec des anticorps mixtes (Sony) : anti-CD3 PE (clone SK7), anti-CD4 FITC (clone A161A1), anti-CD8 APC (clone HIT8a) et anti-CD7 PE-Cy5 (clone CD7-6B7) pendant une période de 20 minutes à 4°C dans l'obscurité. Après le rinçage, les populations de cellules CD3<sup>+</sup>-CD4<sup>+</sup>, CD3<sup>+</sup>-CD8<sup>+</sup> et CD3<sup>+</sup>-CD7<sup>+</sup> ont été triées à l'aide du dispositif BD FACS ARIA II SORP. Un million de cellules de chaque population ont été transférées sur des lames de verre à l'aide d'une centrifugeuse Cytospin™ (Thermo Shandon Cytospin) et conservées à une température de -80°C. Les analyses SpiderMass ont ensuite été effectuées directement sur ces lames de verre.

### MALDI-MSI

Neuf tissus tumoraux ont été soumis à une analyse par MALDI-MSI. Pour ce faire, des sections de 12 micromètres ont été obtenues à l'aide d'un cryostat Leica CM1510S et ces sections ont ensuite été disposées sur des lames de verre ITO de LaserBioLabs (Valbonne, France). L'application de la matrice d'acide 2,5-DHB a été réalisée à l'aide d'un pulvérisateur manuel. Les analyses MALDI-MSI ont été exécutées sur un instrument MALDI-TOF/TOF Ultraflex II (Bruker) opérant en mode ions positifs. La résolution spatiale a été réglée à 70 micromètres et la gamme de masse a été définie de  $m/z$  60 à 1000. Les données MALDI-MSI ont ensuite été soumises à une segmentation sur le logiciel SciLS Lab (version 2023, SciLS GmbH). Une normalisation basée sur le nombre total d'ions (TIC) a été appliquée aux données. La segmentation des images a été réalisée à l'aide de l'algorithme bisecting  $k$ -means, facilitant ainsi une segmentation globale et individuelle des neuf images. Cette segmentation spatiale complète a permis de mettre en évidence des régions d'intérêt qui se sont avérées concorder avec celles obtenues à partir des données protéomiques.

### Immunohistochimie couplée au MALDI-MSI (Tag-Mass)

L'imagerie multiplex a été effectuée sur deux échantillons de tissus de glioblastome frais congelés, qui avaient déjà été analysés par SpiderMass-MSI précédemment. L'un de ces échantillons présentait une survie inférieure à 10 mois, tandis que l'autre affichait une survie supérieure à 36 mois. La préparation des échantillons et le protocole d'imagerie adopté étaient en accord avec les directives recommandées pour les sondes Miralys d'Ambergen. Dans un premier temps, les tissus ont été soumis à un processus de dessiccation sous vide durant 10 minutes, suivi d'une fixation avec une solution à 1 % de PFA pendant 30 minutes. Par la suite, les tissus ont été soumis à une série de bains : un bain dans une solution de PBS pendant 10 minutes, deux bains successifs dans de l'acétone pendant 3 minutes chacun, suivi d'un bain dans une solution de Carnoy pendant 3 minutes. Puis, les tissus ont été réhydratés au moyen de deux bains dans de l'éthanol à

100 % pendant 2 minutes chacun, suivis de trois bains consécutifs dans de l'éthanol à 95 %, de l'éthanol à 70 % et de l'éthanol à 50 %, chacun durant 3 minutes. Une étape de lavage avec une solution TBS pendant 10 minutes (50 mM Tris, pH 7.5, 200 mM NaCl) a précédé l'antigen retrieval, lequel a été réalisé dans un tampon Tris 20 mM à pH 9 pendant 30 minutes à 95°C. Après un rinçage de 10 minutes avec une solution TBS, les tissus ont été traités avec un tampon de blocage tissulaire (composé de 2 % de sérum normal de souris et lapin ainsi que de 5 % de BSA dans une solution de TBS-OG [TBS avec 0.05 % *w/v* Octyl  $\beta$ -D-glucopyranoside]) pendant 1 heure. Par la suite, les tissus ont été incubés à une température de 4°C pendant toute la nuit dans le même tampon de blocage, contenant les sondes CD68, CD8 $\alpha$ , Ki67, Vimentine et collagène, à une concentration de 2.5  $\mu$ g/mL.

Chaque lame a été soigneusement lavée à l'aide de trois bains TBS de 5 minutes et de trois bains de 2 minutes dans une solution à 50 mM de NH<sub>4</sub>HCO<sub>3</sub>, le tout réalisé dans des conditions d'obscurité. Les tissus ont ensuite été séchés sous vide pendant 1 heure et 30 minutes à température ambiante, suivi d'une exposition aux rayons UV à une longueur d'onde de 365 nm (grâce à l'utilisation d'une boîte UV Miralys™ Light Box de chez AmberGen, Inc., Billerica, MA) pendant 10 minutes, afin de cliver les sondes. Ensuite, la matrice DHB a été préparée à une concentration de 20 mg/mL dans une solution de MeOH:TFA à 0.1 % (70:30, *v/v*), et cette matrice a été déposée sur les tissus à l'aide du pulvérisateur HTX TM-Sprayer (HTX Technologies, LLC). Les deux lames ainsi préparées ont ensuite été soumises à une analyse MALDI-MSI en utilisant l'instrument RapifleX MALDI Tissuetyper™ (Bruker Daltonik GmbH). Cette analyse a été réalisée en mode réflecteur, en mode ion positif et avec une résolution spatiale de 20  $\mu$ m. Chaque pixel a été frappé par 500 tirs laser, et la normalisation TIC a été appliquée pour assurer un traitement multiplex efficace des images résultantes. Ces images ont ensuite été soumises à une analyse à l'aide du logiciel flexImaging (Bruker Daltonics, Billerica, MA).

### Analyse des données SpiderMass

Tout le développement informatique (écriture et amélioration du code python) a été effectué par Yanis Zirem, ingénieur en bio-informatique au laboratoire PRISM.

#### *Prétraitement et importation des données*

Avant d'obtenir les données matricielles pour chaque ensemble de données, plusieurs étapes de prétraitement ont été implémentées. En premier lieu, les données ont été regroupées par incréments de 0.1 Da afin de réduire la densité des points de données. Ensuite, une procédure d'alignement a été effectuée en se basant sur la lockmass ( $m/z$  554.3 et 556.3 en mode ion négatif et positif respectivement), et la gamme de masse a été restreinte entre 600 et 1100  $m/z$ . Enfin, la normalisation TIC a été appliquée pour éliminer toute variation d'intensité parmi les spectres.

Chaque ensemble de données final comprend alors 5000 caractéristiques  $m/z$ . Concernant l'importation des données, deux approches ont été utilisées pour extraire les données matricielles à partir des fichiers RAW provenant de Waters. La première méthode impliquait la conversion des fichiers bruts en fichiers au format mzML grâce à MSConvert (Proteowizard)<sup>185</sup>, suivi de l'importation des fichiers mzML dans Python à l'aide de la bibliothèque pyopenMS. La deuxième méthode, majoritairement utilisée, incluait l'importation directe des fichiers bruts dans "Abstract Model Builder" - AMX (version 1.0 1972.0, Waters, Hongrie) et l'extraction des données matricielles au format csv. Ensuite, la bibliothèque pandas a été utilisée pour importer ces fichiers csv dans Python.

### *Modèle de classification optimal et prédiction à l'aveugle*

Pour construire un modèle de classification optimal, nous avons exploité la bibliothèque Lazy Predict en développant plusieurs modèles à partir de la bibliothèque scikit-learn. Cette démarche a impliqué l'entraînement et le test de 24 classificateurs différents. Par la suite, nous avons reconstruit individuellement le modèle optimal en utilisant la bibliothèque scikit-learn. Cette approche nous a permis de procéder à un affinage minutieux des paramètres du modèle pour optimisation et évaluation de son exactitude. Afin d'obtenir une évaluation approfondie de la performance du modèle, nous avons effectué une validation croisée à 20 reprises en employant les fonctions KFold et cross\_val\_score. Pour mieux appréhender la qualité des prédictions, nous avons généré un rapport de classification grâce à la fonction classification\_report. Par ailleurs, pour visualiser la matrice de confusion, nous avons recouru à la fonction ConfusionMatrixDisplay de la bibliothèque matplotlib. Une fois le modèle optimal établi, nous l'avons sauvegardé et chargé pour réaliser des prédictions à l'aveugle. Cette étape a été réalisée en utilisant la bibliothèque joblib, à travers les fonctions joblib.dump et joblib.load.

### *Explication des prédictions*

Pour interpréter les prédictions du modèle de classification, nous avons fait appel à l'algorithme LIME (Local Interpretable Model-agnostic Explanations)<sup>213</sup>. Cet algorithme calcule les contributions des différents  $m/z$ , qu'elles soient positives ou négatives. Pour réaliser cette étape, nous avons utilisé la bibliothèque ELI5, qui nous a permis de générer un tableau LIME renfermant les poids attribués aux contributions de chaque  $m/z$ , en utilisant la fonction explain\_prediction.

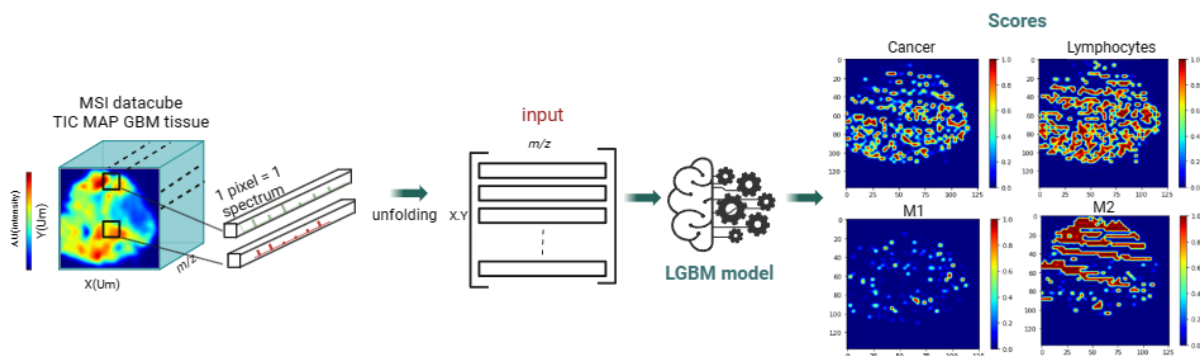
### *Biomarqueurs significatifs*

Dans le but de réduire le bruit provenant des instruments, nous avons mis en œuvre un algorithme de sélection des pics grâce à la fonction find\_peaks\_cwt de la bibliothèque scipy. Ensuite, un heatmap a été généré sur la base de la liste des pics. Celui-ci utilise le regroupement hiérarchique entre les variables  $m/z$  avec la distance euclidienne, affichant les ions qui sont plus ou

moins abondants dans chaque groupe. A cette fin, la fonction clustermap de la bibliothèque seaborn a été utilisée. Enfin, pour évaluer l'importance de chaque  $m/z$ , nous avons appliqué un test statistique non paramétrique Kruskal-Wallis avec correction de Bonferroni. Cela a été accompli en utilisant la fonction stats.kruskal de la bibliothèque Scipy. Seules les valeurs  $m/z$  qui se sont avérées significatives, c'est-à-dire avec une valeur p égale ou inférieure à 0.05, ont été retenues. Finalement, afin de ne garder que les pics mono-isotopiques, une étape de filtrage a été ajoutée. Les boîtes à moustaches correspondantes ont ensuite été générées à l'aide de la bibliothèque seaborn.

### Immunescores

Le modèle d'immunescore a été développé en utilisant la bibliothèque Python LGBM (Light Gradient Boosting Machine), développée par Microsoft. Ce modèle s'est appuyé sur les spectres cellulaires se situant dans la gamme de masse  $m/z$  de 600 à 1100 en mode négatif. Ces spectres ont été catégorisés en différentes classes de cellules, à savoir les macrophages (M1 et M2), la lignée cellulaire cancéreuse (NCH82) et les lymphocytes (NK, CD8, CD4) avec chacun 107 spectres. En raison du nombre restreint de spectres disponibles pour chaque type de lymphocyte et dans le but d'éviter la création d'un modèle déséquilibré, les différents types de lymphocytes ont été regroupés. Afin de prédire les types de cellules présents sur les images SpiderMass, nous avons utilisé la fonction predict\_proba du modèle LGBM (**Figure 39**).



**Figure 39** : Le pipeline global pour voir la distribution des cellules cancéreuses et immunitaires (lymphocytes et macrophages de type M1 et M2) dans un tissu GBM analysé par SpiderMass-MSI.

Cette fonction a fourni des estimations de probabilité pour chaque type de cellule, ce qui a permis d'obtenir une compréhension plus détaillée des probabilités relatives de la présence de chaque type de cellule dans le micro-environnement tumorale. De plus, nous avons calculé des scores de ratio pour évaluer la présence relative de chaque type de cellule sur l'ensemble de l'image. Ces scores ont été obtenus en additionnant les scores associés à chaque type de cellule, puis en divisant cette somme par le total des scores pour toutes les classes. Les ratios obtenus ont offert des indications sur la répartition des différents types de cellules dans l'ensemble de l'image, ce qui a permis d'évaluer de manière globale la composition cellulaire dans l'échantillon analysé.



## Quantification MALDI-IHC

La bibliothèque Python PIL a été utilisée pour convertir l'image MALDI - immunohistochimie (MALDI-IHC) au format PNG. Ensuite, toutes les images ont été converties en niveaux de gris et transformées en tableaux NumPy à l'aide de la bibliothèque NumPy. Un algorithme a été conçu et utilisé pour calculer le nombre de pixels noirs dans l'image à l'aide du tableau NumPy, où chaque canal va de 0 (représentant le noir) à 255 (représentant le blanc). La quantification de chaque marqueur IHC a été déterminée en soustrayant le nombre de pixels noirs du nombre total de pixels de l'image. La bibliothèque matplotlib.pyplot a été utilisée pour visualiser l'image et générer un diagramme à barres.

## Analyse MS/MS

Les spectres MS/MS ont été obtenus après l'isolement de l'ion précurseur, avec une fenêtre de 1 u, suivis d'une dissociation induite par collision dans la cellule de transfert. L'énergie de collision de transfert utilisée a varié entre 30 et 40 eV, en fonction de l'ion précurseur sélectionné. Les annotations des lipides ont été réalisées manuellement en se basant sur les spectres de fragmentation. Ensuite, ces annotations ont été comparées avec les informations disponibles dans la base de données publiques LipidMaps, Alex123, Metlin et les sources littéraires appropriées.

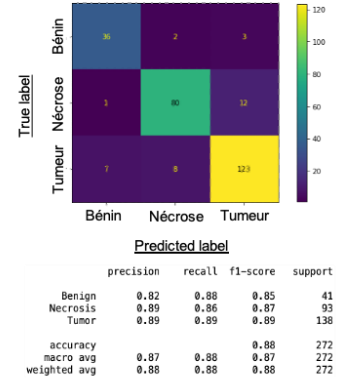
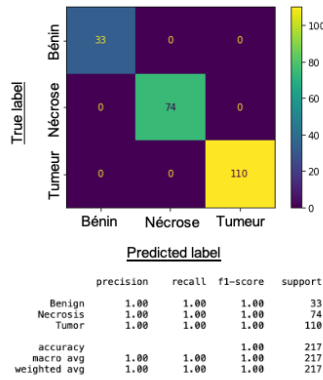
## Résultats et discussion

### Évaluation du profil lipidomique des tissus de glioblastomes

#### Modèle de classification optimal et processus d'IA utilisé.

#### Mode négatif

Model	Accuracy	Balanced Accuracy	F1 Score	Time Taken
RidgeClassifier	0.93	0.93	0.92	0.18
SVC	0.89	0.91	0.89	0.28
LogisticRegression	0.89	0.90	0.89	0.37
PassiveAggressiveClassifier	0.87	0.89	0.87	0.46
SGDClassifier	0.89	0.88	0.89	0.22
LinearDiscriminantAnalysis	0.85	0.87	0.85	0.34
RandomForestClassifier	0.85	0.87	0.85	0.55
ExtraTreesClassifier	0.85	0.87	0.85	0.32
LinearSVC	0.84	0.86	0.84	1.62
Perceptron	0.82	0.86	0.82	0.22
KNeighborsClassifier	0.82	0.81	0.81	0.14
LGBMClassifier	0.80	0.77	0.79	2.89
CalibratedClassifierCV	0.84	0.77	0.83	6.50
BaggingClassifier	0.78	0.76	0.78	1.82
AdaBoostClassifier	0.67	0.73	0.67	3.95
GaussianNB	0.62	0.72	0.61	0.17
DecisionTreeClassifier	0.71	0.71	0.71	0.43
BernoulliNB	0.47	0.59	0.44	0.17
NearestCentroid	0.49	0.59	0.48	0.14
ExtraTreeClassifier	0.62	0.54	0.61	0.14
QuadraticDiscriminantAnalysis	0.49	0.51	0.49	0.22
LabelSpreading	0.15	0.33	0.04	0.15
LabelPropagation	0.15	0.33	0.04	0.20
DummyClassifier	0.51	0.33	0.34	0.13



#### Mode positif

Model	Accuracy	Balanced Accuracy	F1 Score	Time Taken
Perceptron	0.88	0.88	0.88	0.21
RidgeClassifier	0.92	0.88	0.92	0.16
SGDClassifier	0.86	0.87	0.86	0.20
LogisticRegression	0.90	0.87	0.90	0.36
LinearSVC	0.86	0.86	0.86	0.55
PassiveAggressiveClassifier	0.86	0.86	0.86	0.33
LGBMClassifier	0.84	0.82	0.84	2.29
LinearDiscriminantAnalysis	0.84	0.79	0.84	0.28
SVC	0.86	0.78	0.85	0.28
RandomForestClassifier	0.84	0.77	0.84	0.53
ExtraTreesClassifier	0.84	0.76	0.83	0.32
CalibratedClassifierCV	0.86	0.75	0.84	1.85
BaggingClassifier	0.76	0.72	0.77	1.79
GaussianNB	0.69	0.70	0.69	0.15
DecisionTreeClassifier	0.67	0.63	0.68	0.51
KNeighborsClassifier	0.65	0.63	0.65	0.14
NearestCentroid	0.61	0.57	0.64	0.13
BernoulliNB	0.59	0.56	0.62	0.15
AdaBoostClassifier	0.67	0.55	0.63	3.73
ExtraTreeClassifier	0.57	0.45	0.54	0.15
QuadraticDiscriminantAnalysis	0.41	0.39	0.42	0.21
LabelPropagation	0.16	0.33	0.04	0.14
LabelSpreading	0.16	0.33	0.04	0.14
DummyClassifier	0.49	0.33	0.32	0.13

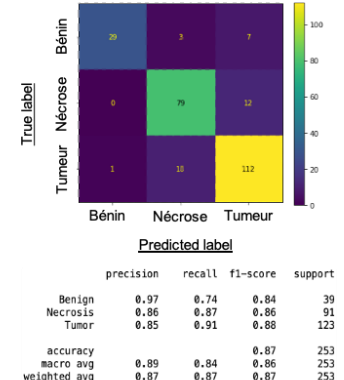
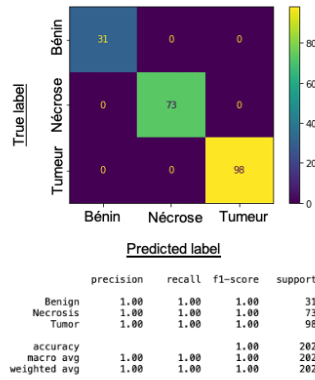


Figure 40 : Identification du modèle de classification optimal après analyse de la cohorte rétrospective uniquement dans les deux modes d'ionisation. Tout d'abord, les tableaux des 24 classificateurs qui ont été entraînés et évalués, indiquant leur précision, leur précision équilibrée, leur score F1 et le temps nécessaire. De plus, les rapports de classification et les matrices de confusion de RidgeClassifier avec l'ensemble d'apprentissage et après une validation croisée 20 fois.

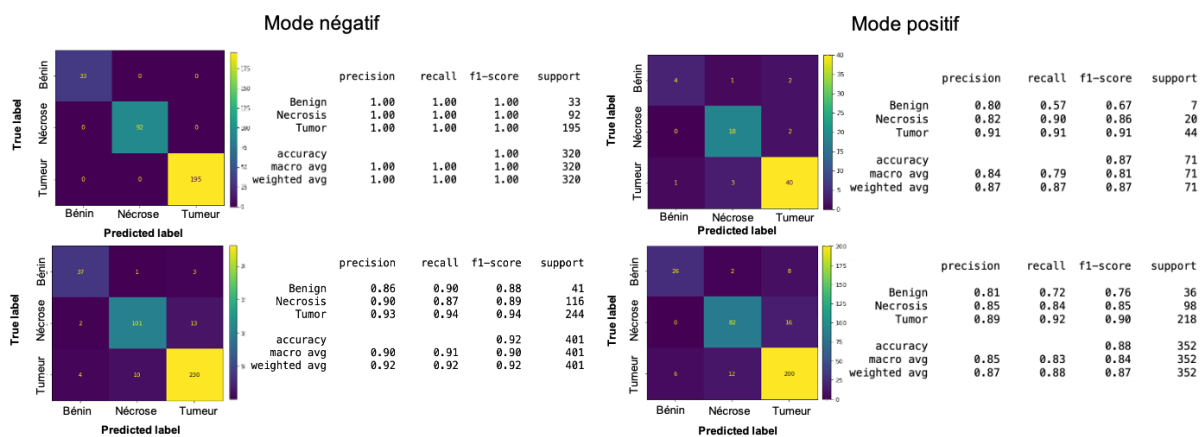
Pour être fonctionnel en bloc opératoire, la technologie SpiderMass doit être combinée à l'apprentissage automatique. Cependant, une variété d'algorithmes d'apprentissage automatique est disponible, couvrant des approches traditionnelles jusqu'aux techniques d'apprentissage profond, et le choix de l'algorithme de modélisation impacte directement la précision des modèles construits.

Historiquement, la méthode de diagnostic moléculaire par SpiderMass repose sur la modélisation d'LDA au moyen du logiciel AMX (Waters). Cependant, cette approche présente certaines limitations, notamment en se limitant à l'LDA pour la classification, en ne proposant pas d'évaluation ni de validation croisée des modèles, et en ne fournissant pas d'explications sur les prédictions ni de tests statistiques pour évaluer les variations des caractéristiques. Dans le but de surmonter ces limites, nous avons développé un nouveau processus d'IA plus complet visant à créer des modèles de classification précis et robustes. Ce processus vise également à identifier des biomarqueurs avec certitude, en se focalisant non seulement sur la classification en elle-même, mais aussi sur la validation des classifications par le biais d'une compréhension biologique.

Dans cette optique, nous avons évalué les performances de 24 classificateurs différents. Cette évaluation a été réalisée sur la base d'une cohorte rétrospective de tissus FFPE. Toutes les régions histologiques (comprenant 50 tumeurs, 42 nécroses et 18 tissus bénins) ont été soumises à l'analyse via SpiderMass en utilisant les deux modes d'ionisation. Des modèles de classification ont été élaborés pour différencier les tissus tumoraux, nécrotiques et bénins, en attribuant les régions à une classe dont l'appartenance ne fait pas l'objet de doute selon un pathologiste expert. Le modèle de classification optimal a été choisi en se basant sur la plus haute précision obtenue lors de la validation croisée de 20 %. Une fois le rapport de classification obtenu ainsi que la matrice de confusion associée au modèle de classification optimal, ses performances ont été évaluées au moyen des scores de précision et de F1. Ces scores, tenant compte à la fois de la justesse et du rappel, fournissent une mesure équilibrée des performances entre les différentes classes. En utilisant les modes négatif et positif, la classification optimale est obtenue en employant le classificateur RidgeClassifier, affichant une précision de 93 % et 92 % respectivement. Cependant, pour assurer la robustesse et la généralisation des données, une validation croisée a été effectuée à vingt reprises. Cette procédure révèle que la précision effective du modèle en mode négatif est finalement de 88 %, avec un écart-type de 0.03. Dans le cas du mode positif, le modèle obtient une précision effective de 87 %, accompagnée d'un écart-type de 0.02 (**Figure 40**).

Le nombre restreint d'échantillons dans la cohorte rétrospective peut expliquer la diminution de la précision finale après avoir été soumis à la validation croisée vingt fois. Afin de surmonter cet obstacle, nous avons ajouté un deuxième ensemble de données basée sur une cohorte prospective de 31 tissus frais congelés. Cette nouvelle cohorte se compose de 30 régions tumorales

et de 7 régions nécrotiques. Malgré certaines différences spectrales entre l'analyse des tissus frais et celle des tissus FFPE, l'utilisation d'un ensemble de données étendu a contribué à une amélioration de la précision du modèle RidgeClassifier. Il est particulièrement notable que les spectres MS extraits à partir de tissus frais congelés et de tissus FFPE ont affiché une similarité suffisante pour être utilisés conjointement, engendrant ainsi une construction et une amélioration efficaces de la classification. En effet, en mode négatif, la précision connaît une augmentation et se stabilise à 92 % à l'issue de vingt cycles de validation croisée. Par ailleurs, l'écart-type diminue de 0.03 à 0.02 (**Figure 41**). En mode positif, les performances montrent une légère amélioration. Il est crucial de souligner que la précision demeure constante même après la validation croisée, aboutissant à une précision effective de 88 %, accompagnée d'un écart-type minime de 0.02 (**Figure 41**).



**Figure 41** : Matrice de confusion et rapport de classification en mode positif et négatif après l'ajout de la cohorte prospective à la cohorte rétrospective.

### *Prédiction automatique aveugle de la classe histologique*

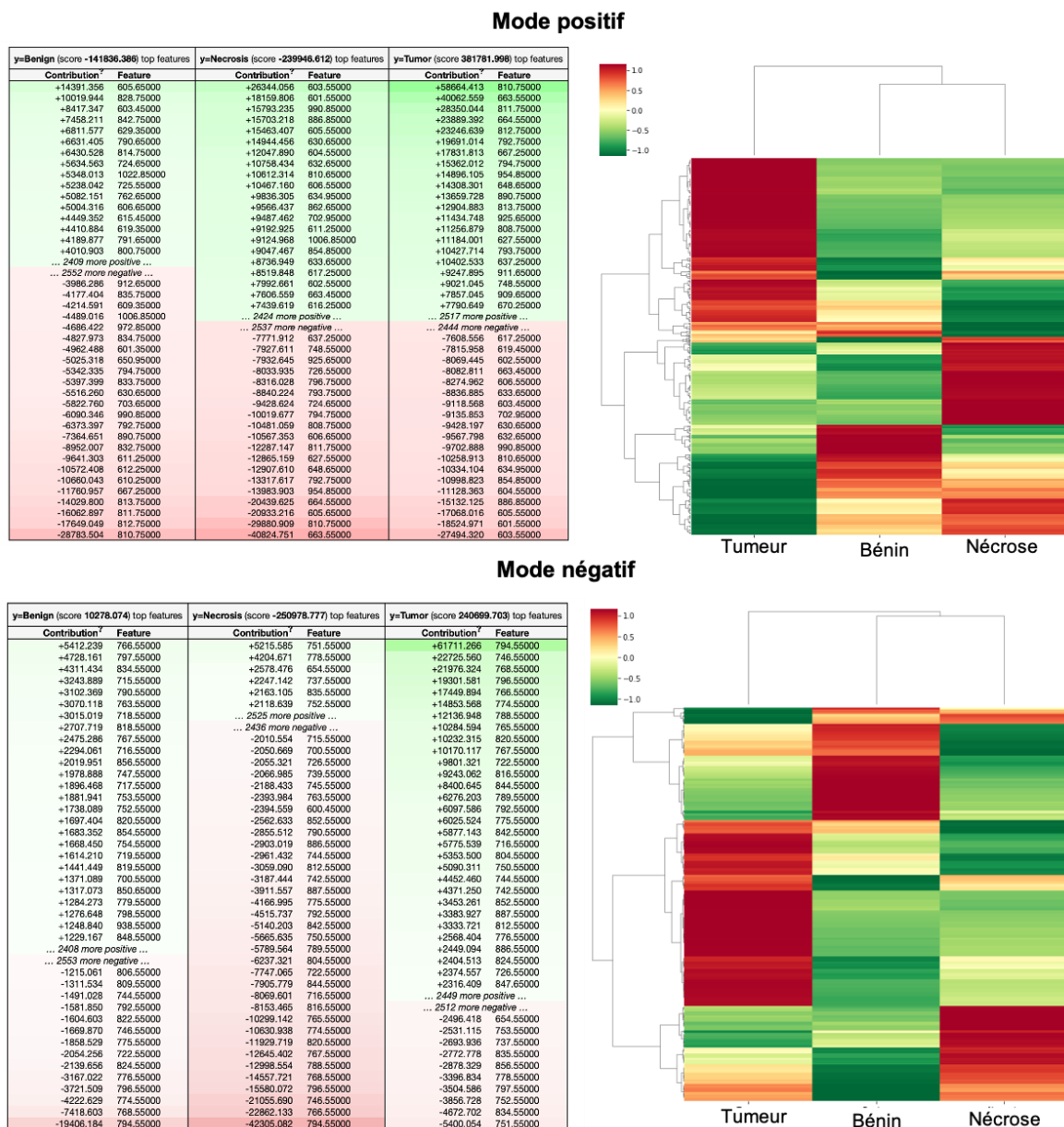
Afin d'évaluer davantage les performances du modèle de classification dans les deux modes d'ionisation, des prédictions à l'aveugle ont été réalisées sur une cohorte de validation indépendante composée de deux échantillons de tissus FFPE et de sept échantillons de tissus frais congelés. Ces neuf échantillons n'ont pas été utilisés dans la création et l'évaluation du modèle précédent, qui implique la catégorisation des régions en groupes bénins, nécrotiques et tumoraux. Les tissus ont été soumis à une analyse SpiderMass en aveugle et ensuite évalués par le modèle de classification après augmentation des données. La prédiction par le SpiderMass a ensuite été confrontée à celle obtenue par le pathologiste. En mode négatif, tous les tissus analysés sont correctement assignés. Tandis qu'en mode positif, deux régions sur vingt sont incorrectement classées, aboutissant à un taux de classification correcte de 90 % (**Tableau 7**). Toutes les régions nécrotiques sont prédites correctement, tandis que deux régions tumorales sont identifiées à tort comme nécrotiques. Ces deux erreurs de classification peuvent être attribuées au fait que le point d'analyse à l'aveugle se situe à la limite entre les zones tumorales et nécrotiques.

**Tableau 7** : Résultats des prédictions à l’aveugle par le SpiderMass des 9 tissus (2 FFPE et 7 frais congelés (FF)) non utilisés dans l’entraînement du modèle de classification.

Tissu	Annotation obtenue par le SpiderMass					Annotation obtenue par le pathologiste
	Mode négatif		Mode positif			
FF 32	Nécrose	Nécrose		Nécrose	Nécrose	Nécrose
FF 4	Tumeur	Tumeur		Tumeur		Tumeur
FF 9	Tumeur	Tumeur	Tumeur	Tumeur		Tumeur
FF 12-3	Tumeur	Tumeur		Tumeur	Tumeur	Tumeur
FF 12-4	Tumeur	Tumeur		Tumeur	Tumeur	Tumeur
FF 14	Tumeur	Tumeur		Nécrose		Tumeur
FF 15	Tumeur	Tumeur		Tumeur		Tumeur
FF 17	Tumeur	Tumeur		Tumeur	Tumeur	Tumeur
FFPE 8	Nécrose		Nécrose	Nécrose	Nécrose	Nécrose
FFPE 8	Tumeur		Nécrose			Tumeur
FFPE 4	Bénin	Bénin		Bénin	Bénin	Bénin
<b>Taux de bonne prédiction</b>	<b>100 %</b>			<b>90 %</b>		

*Marqueurs lipidiques spécifiques à chaque classe moléculaire.*

Enfin, nous nous sommes focalisés sur l'identification des lipides associés aux diverses classes histologiques, dans le but de mieux appréhender les voies biologiques sous-jacentes à la discrimination, et de résoudre la problématique commune de l'opacité inhérente à la plupart des modèles de classification. L'intégration de l'algorithme LIME dans le pipeline de traitement élaboré a permis de fournir des explications interprétables, incluant l'identification potentielle de biomarqueurs, en calculant les contributions de chaque ion  $m/z$  pour classifier chaque type de tissu. Les 120 ions  $m/z$  contribuant le plus positivement et le plus négativement et leur contribution sont indiqués dans la **Figure 42** pour le mode négatif et positif.

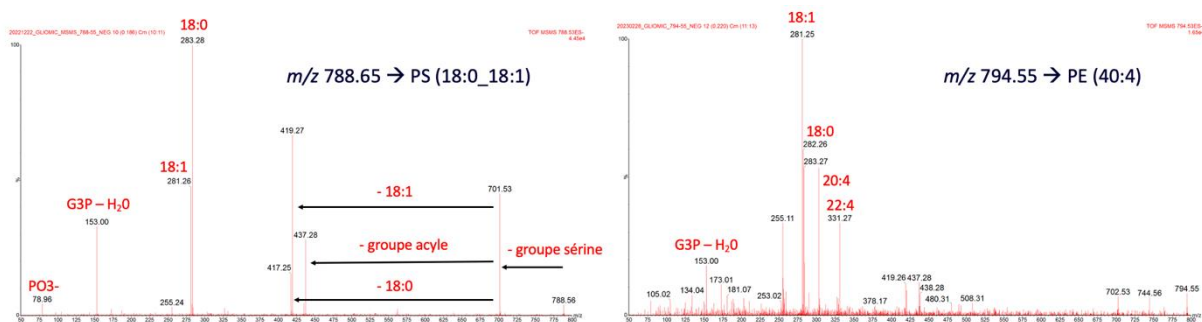


**Figure 42** : Résultats de l'analyse supervisée et non supervisée pour la découverte de marqueurs en mode négatif et positif contenant le top 120  $m/z$  qui contribuent positivement (vert) ou négativement (rouge) à la classification de chaque tissu dans le modèle ainsi que le heatmap des pics sélectionnés pour trois régions différentes. Le heatmap affiche l'intensité de chaque ion, le rouge indiquant une surexpression et le vert une sous-expression.

Pour identifier les potentiels biomarqueurs lipidiques indépendamment de la classification, les variations d'abondance des ions entre les différents types de tissu ont été explorées à travers une analyse statistique multivariée appliquée aux mêmes données. En mode négatif, le peak picking a abouti à 509 pics parmi les 5000 ions  $m/z$  grâce à un seuil permettant d'enlever le bruit présent dans les spectres de masse. Le regroupement hiérarchique a été utilisé pour construire un heatmap et nous a permis d'identifier trois groupes d'ions, chacun d'entre eux étant surexprimé dans un type de tissu spécifique (**Figure 42**). Après avoir visualisé le heatmap, un test de Kruskal-Wallis a été mené pour évaluer la signification statistique de chaque ion. Parmi ces 509 pics, 154 se sont avérés significatifs (avec une valeur  $p \leq 0.05$ ). Suite au filtrage des pics non résolus et des isotopes, 57

biomarqueurs potentiels ont été retenus. En mode positif, 636 pics ont été sélectionnés après le peak picking. Parmi eux, 100 ont été considérés comme significatifs, et un total de 37 ions  $m/z$  ont été conservés en tant que biomarqueurs potentiels après le filtrage des isotopes et des pics non résolus. Pour identifier les biomarqueurs lipidiques présentant un niveau de confiance élevé, nous avons croisé les données issues des analyses supervisées et non supervisées. Ainsi, afin qu'un ion  $m/z$  soit retenu comme un biomarqueur fiable, il doit être statistiquement significatif (soit une surexpression, soit une sous-expression) au sein d'un type de tissu spécifique et doit également afficher une contribution positive en cas de surexpression et une contribution négative en cas de sous-expression, toujours au sein du même type de tissu.

Au total, nous avons identifié 41 biomarqueurs fiables : 27 en mode négatif et 14 en mode positif, ces biomarqueurs étant associés aux tissus tumoraux, bénins et nécrotiques (**Figure 50**). En recourant à une analyse par spectrométrie de masse en tandem, ces 41 biomarqueurs ont pu être manuellement annotés et identifiés. Cette démarche permet d'explorer la prédominance de certaines catégories de lipides dans des types spécifiques de tissus (**Figure 43**).



**Figure 43** : Spectres MS/MS obtenus après fragmentation des ions  $m/z$  788.6 et  $m/z$  794.5, annotés ensuite en tant que PS (18:0\_18:1) et PE (40:4) respectivement.

En mode négatif, parmi ces 27 biomarqueurs fiables, neuf ions sont surexprimés dans les tissus tumoraux, principalement des phosphatidylsérines et des acides phosphatidiques. Sept ions sont spécifiques des tissus bénins, parmi lesquels six sont abondants et un DG 40:9 ( $m/z$  661.55) est peu présent. Dans les tissus bénins, une variété d'espèces lipidiques est présente, cependant, l'absence de diglycérides pourrait s'avérer être un marqueur de santé potentiel. Notamment, tous les ions très abondants identifiés dans les tissus bénins sont des phospholipides de grande taille, avec des acides gras comportant plus de trois insaturations. Au sein des cellules de gliome, la concentration élevée en acides gras polyinsaturés semble inhiber la division cellulaire et la formation de tumeurs<sup>214</sup>. Concernant les régions nécrotiques, trois ions sont surexprimés à des masses faibles tandis que huit ions sont sous-exprimés à des masses plus élevées. Ces tissus se distinguent également par l'absence de certaines molécules plutôt que par leur présence. Plus précisément, les phosphatidylinositols présentent une sous-expression dans ces tissus. En ce qui



concerne le mode positif, parmi les 14 biomarqueurs fiables, deux sont spécifiques des tissus bénins, à savoir  $m/z$  842.65 et  $m/z$  850.65. Dans les tissus nécrotiques, trois biomarqueurs fiables montrent une faible représentation, notamment les PIs et les PCs. Pour ce qui est des régions tumorales, huit biomarqueurs sont en surexpression, la majorité d'entre eux étant des PEs. Il est remarquable que les glycérophospholipides abondamment exprimés dans les tissus tumoraux sont principalement constitués de lipides saturés tels que PC 34:2 ( $m/z$  758.65) et PS 38:2 ( $m/z$  814.55) (Figure 44).

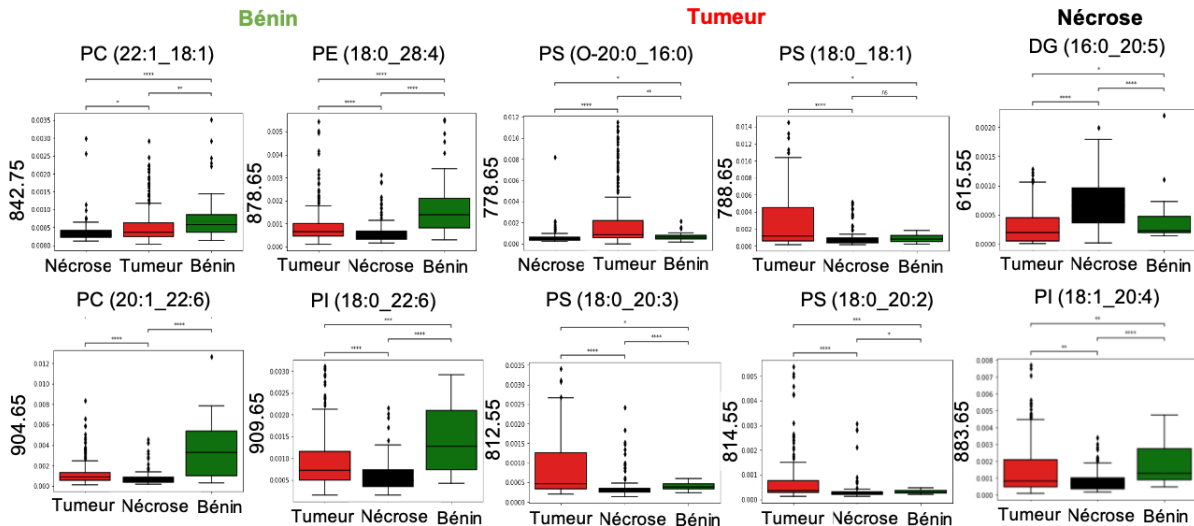


Figure 44 : Dix exemples de box plots d'ions biomarqueurs fiables des tissus bénins, tumoraux et nécrotiques

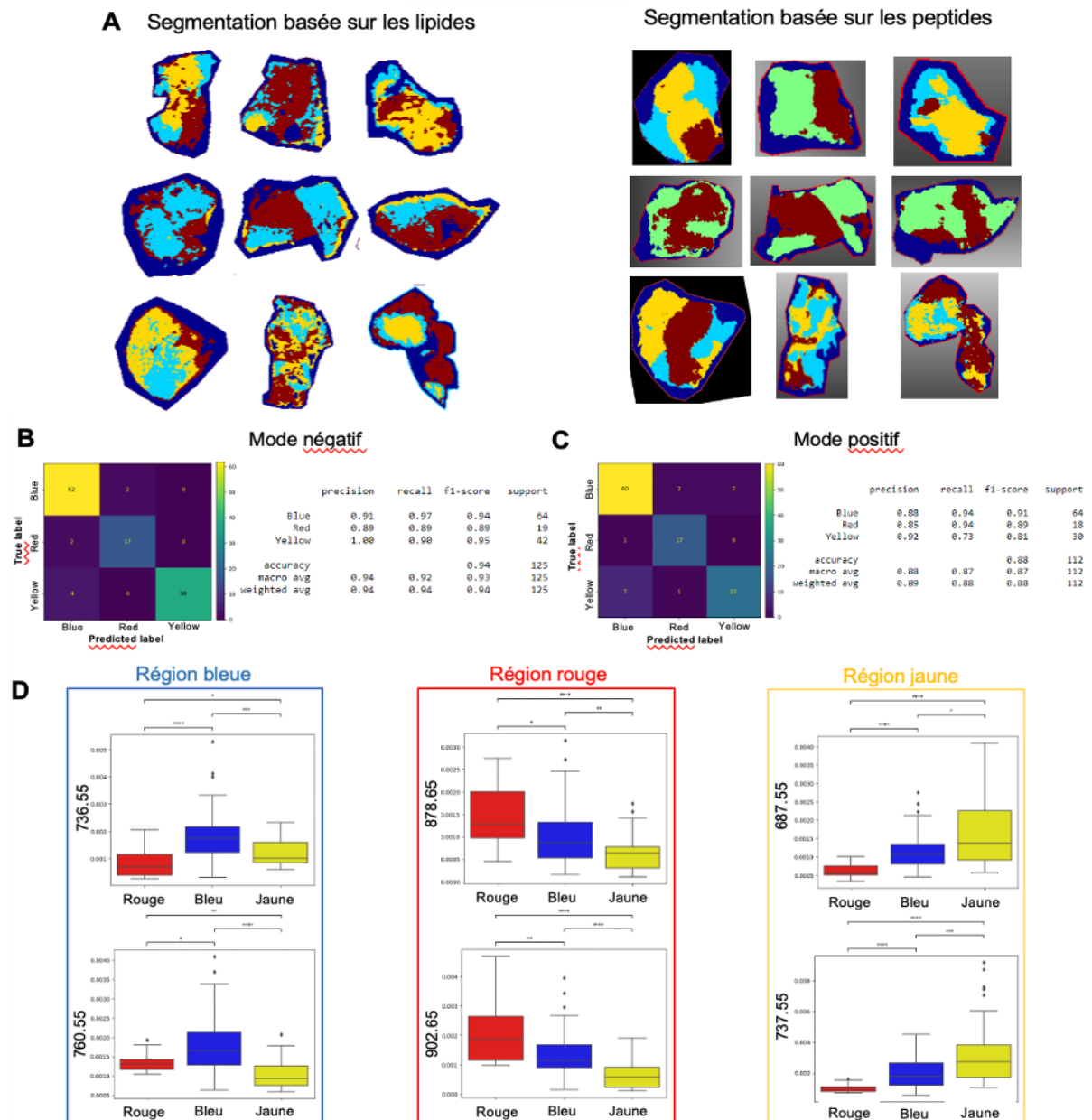
### *Stratification moléculaire non supervisée et supervisée du glioblastome*

Dans une étude antérieure qui a utilisé la protéomique guidée par l'imagerie MALDI-MS, il a été mis en évidence que, bien que le glioblastome de grade IV ne soit pas lié à un statut mutationnel spécifique, il pouvait être stratifié en trois sous-groupes moléculaires distincts. Il est remarquable que ces sous-groupes soient associés à différentes voies, à savoir i) les gènes liés au

développement neurologique, caractéristiques des lignées neuronales/gliales et des cellules progénitrices neurales (représentés en rouge), ii) le statut immunitaire avec une infiltration de macrophages (représentés en jaune), et iii) la réponse immunitaire antivirale, l'infection virale, ainsi que le traitement de l'ARN (représentés en bleu)<sup>90</sup>. En outre, une corrélation entre ces groupes et la durée de vie du patient a été observée : les patients pour lesquels des marqueurs d'inflammation et une infiltration de macrophages avaient été identifiés présentaient une durée de vie plus courte.

Par conséquent, dans cette partie, nous avons d'abord réalisé une analyse MALDI-MSI en mode positif à partir de 9 échantillons restants de la cohorte prospective. Nous avons ensuite effectué une segmentation individuelle pour chaque tissu ainsi qu'une segmentation globale en utilisant l'ensemble des tissus. Tout comme observé précédemment pour les peptides, nous avons identifié trois principaux groupes moléculaires présentant une forte corrélation avec la

segmentation globale obtenue à partir des peptides. Il est important de noter que les données lipidomiques et protéomiques ont permis de différencier les mêmes zones dans plusieurs tissus. Effectivement, la segmentation a conduit à la même sous-stratification des tissus GBM en se basant sur les données lipidomiques et protéomiques MALDI-MSI (**Figure 45A**).



**Figure 45** : Sous-stratification moléculaire non supervisée et supervisée. A) Segmentation globale basée sur les données lipidomiques et protéomiques. B) et C) Rapport de classification et matrice de confusion du classificateur Ridge après une validation croisée, obtenus pour le modèle de classification qui distingue les 3 groupes de segmentation (rouge, jaune et bleu). D) Boîtes à moustaches de deux biomarqueurs lipidiques obtenus après la découverte non supervisée et supervisée de marqueurs pour chaque groupe de segmentation.

De plus, l'obtention d'une telle stratification *in vivo* et en temps réel pendant une intervention chirurgicale pour adapter l'extension de la résection en fonction du pronostic du patient serait extrêmement intéressante. Bien que le MALDI-MSI ne puisse pas être utilisé *in vivo*, le

SpiderMass peut l'être. Il est à noter que les profils lipidiques obtenus par MALDI et SpiderMass sont très similaires, car la technologie SpiderMass est fondée sur la technique WALDI-MS, qui reproduit un processus similaire au MALDI en utilisant l'eau endogène comme matrice MALDI.

Cela a d'ailleurs été préalablement présenté dans le chapitre 2 de ce manuscrit de thèse. Grâce au nouveau pipeline de traitement IA développé, des modèles de classification ont été créés pour distinguer les trois sous-groupes moléculaires spécifiques, avec des taux de classification correcte de 94 % et 88 % en mode négatif et positif, respectivement, après validation croisée (**Figure 45B-C**). Notamment, la prédiction automatique en aveugle des différents sous-groupes a bien classé six des huit tissus inconnus. Ensuite, 18 biomarqueurs fiables ont été identifiés dans les deux modes, dont trois spécifiques aux zones bleues, neuf aux zones rouges et six aux zones jaunes (**Figure 45D**). Il est intéressant de remarquer que certains ions sont communs avec les marqueurs lipidiques des différentes zones histologiques mentionnées précédemment. Par exemple, le PS 34:1 ( $m/z$  760.55), spécifique de la zone tumorale, est également spécifique de la région bleue liée à la croissance tumorale en protéomique. De plus, deux ions  $m/z$  878.65 PE 46:4 et  $m/z$  902.65 PS 44:0 ont été identifiés comme spécifiques non seulement au tissu bénin, mais également à la région rouge associée aux gènes du développement neurologique. Cela souligne les multiples corrélations entre les données protéomiques et lipidiques, ouvrant ainsi la possibilité d'utiliser SpiderMass comme un outil *in vivo* pour la stratification du tissu du glioblastome. Cette stratification pourrait permettre d'adapter la réponse chirurgicale en fonction du temps de survie du patient.

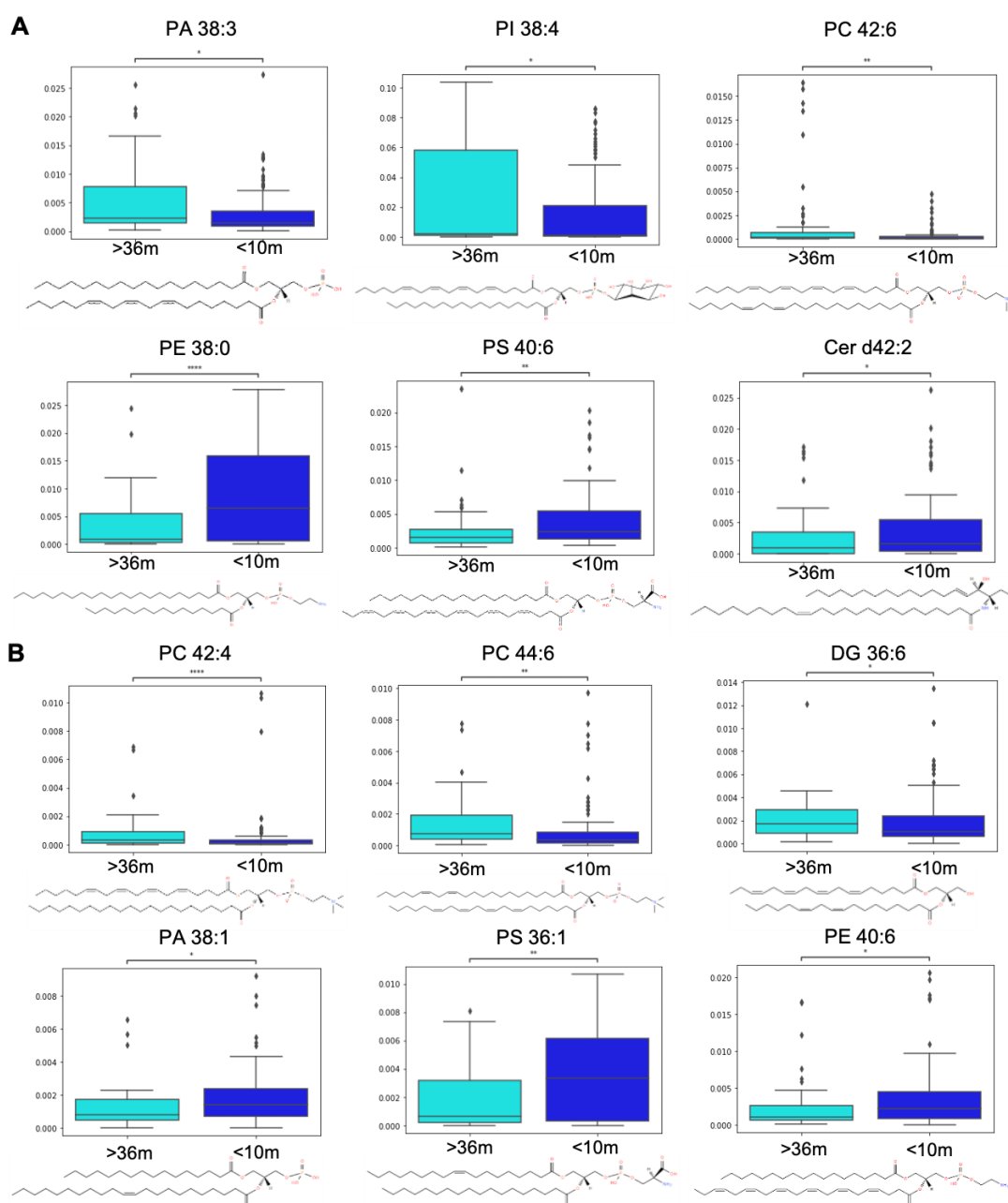
### Prédiction de la survie globale des patients atteints de glioblastome

#### *Classification des patients en fonction de leur survie grâce au SpiderMass*

En plus du diagnostic confirmé et de la stratification réalisée par SpiderMass, nous avons également entrepris d'évaluer son potentiel de prédiction du pronostic des patients. Utilisant une médiane de survie de 15 mois comme référence pour les patients atteints de glioblastome, nous avons mis en place une classification à l'aide du pipeline de traitement précédemment développé, divisant notre cohorte en deux groupes : ceux avec une survie inférieure à 15 mois et ceux avec une survie supérieure à 15 mois. Nous avons utilisé les spectres MS enregistrés à partir des tissus, indépendamment de leur localisation dans une région histologique. En utilisant le mode positif et négatif, nous avons pu élaborer des modèles de classification affichant un taux de classification correcte de 79 % et 87 % respectivement, après validation croisée. Pour renforcer la spécificité de la classification, nous avons ensuite focalisé notre attention sur les patients avec les survies les plus extrêmes, soit inférieures à 10 mois et supérieures à 36 mois. En limitant notre modèle aux cas les plus extrêmes, le taux de classification a augmenté jusqu'à 93 % et 87 % en mode négatif et positif, respectivement après la validation croisée.

### Identification des marqueurs distinctifs liés à la survie des patients

Lors de l'exploration des potentiels biomarqueurs associés à la survie des patients atteints de glioblastome, certains ions se sont avérés spécifiques à la courte ou longue survie. En effet, les deux modes d'ionisation ont abouti à l'identification de 48 biomarqueurs pronostiques statistiquement significatifs. En résumé, les phosphatidylcholines et les phosphatidylinositols sont plus abondants chez les patients ayant une survie plus longue, notamment les PC 42:4 et PC 44:6 en mode positif ainsi que les PI 38:4 et PC 42:6 en mode négatif. De plus, les ions  $m/z$  850.75 identifié comme PE (22:1\_22:4) et  $m/z$  920.75 comme PS 46:5, plus abondants dans les tissus bénins, sont également associés à une survie prolongée.



**Figure 46 :** Six biomarqueurs potentiels, leur boîte à moustaches correspondante, leur structure chimique, qui ont été trouvés spécifiques du pronostic <10 mois et >36 mois. A) en mode négatif. B) en mode positif.

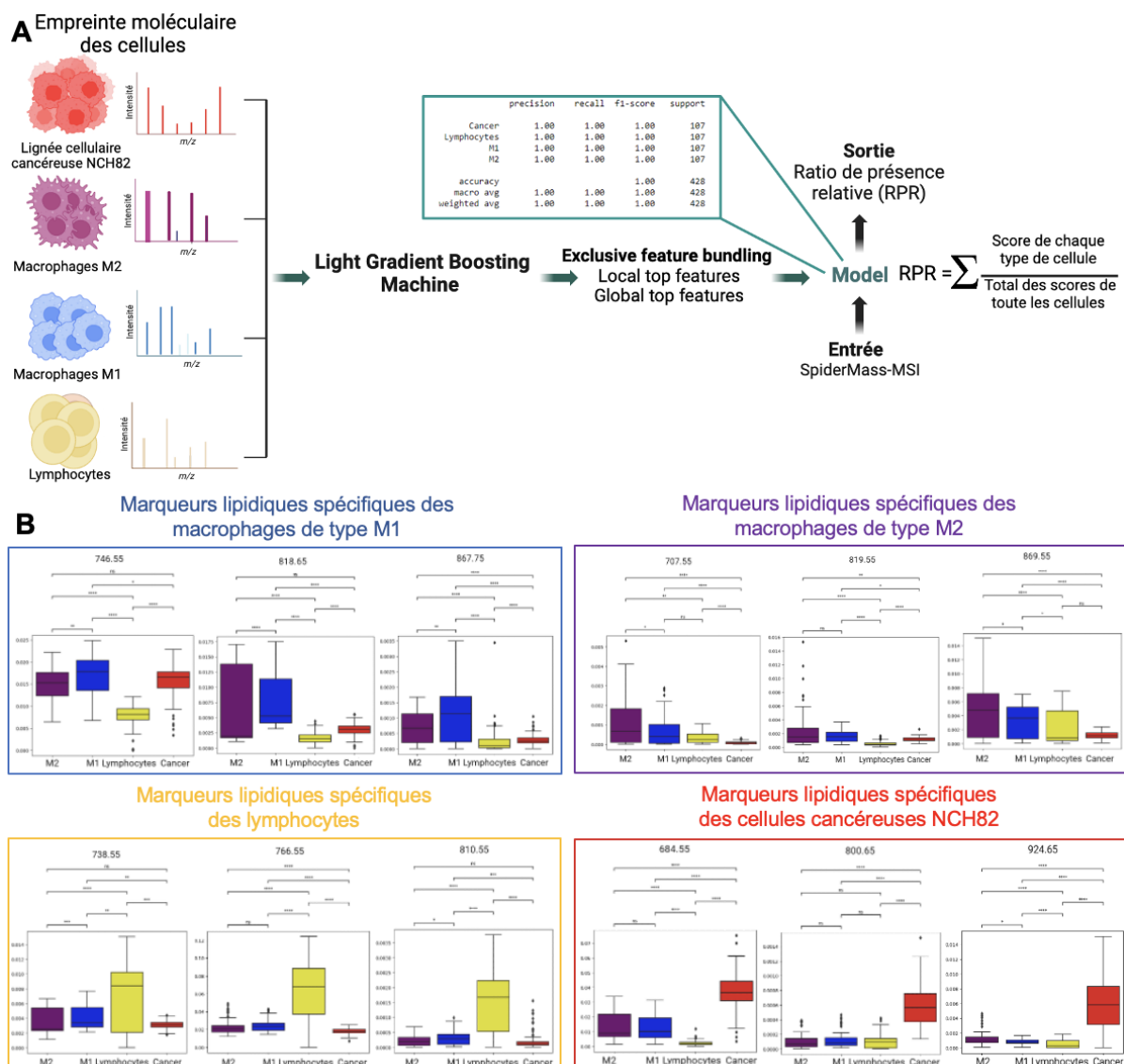
Il semblerait que les marqueurs moléculaires partagés avec les tissus bénins soient présents dans les tissus tumoraux des patients ayant une survie plus longue, ce qui pourrait contribuer à expliquer leur durée de vie prolongée. En revanche, les phosphatidylsérines, les acides phosphatidiques et les phosphatidyléthanolamines, tels que PA 38:1, PE 38:0, PE 40:6 et PS 36:1, sont davantage exprimés dans les tissus tumoraux associés à un pronostic moins favorable. Notamment, le PS 40:6 a été identifié comme lié à une survie plus courte dans les deux modes d'ionisation ( $m/z$  834.55 et 836.55) avec une signification statistique élevée et une contribution significative. De plus, les sphingolipides, en particulier les céramides, sont caractéristiques d'un pronostic défavorable dans notre étude, aucun n'ayant été détecté pour les survies plus longues. En effet, on démontre que la concentration de cinq céramides incluant le Cer d42:2 ( $m/z$  682.65) et le CerP 34:0 ( $m/z$  654.55), est considérablement plus élevée chez les patients avec une survie courte par rapport à ceux ayant une durée de vie prolongée, sans distinction de la longueur de la chaîne acyle ou du degré d'insaturation. De plus, le  $m/z$  844.65 identifié comme PS 40:1, spécifique à la zone histologique tumorale, est ici corrélé à une survie courte (**Figure 46 et Figure 50**).

Il est intéressant de noter que nos conclusions relatives aux lipides sont étroitement liées à celles obtenues au niveau de la protéomique<sup>90</sup>. En effet, les trois marqueurs de pronostic identifiés précédemment présentent une forte corrélation avec les marqueurs lipidiques de pronostic. Les tumeurs des patients avec une survie prolongée montrent une expression accrue des protéines RPS14 et PPP1R12A, tandis que celles des patients avec une survie limitée présentent une expression accrue d'ANXA11. ANXA11 est une protéine de liaison aux phospholipides qui dépend du calcium<sup>215</sup>, et les annexines sont des protéines solubles qui se lient aux phospholipides chargés négativement en présence d'ions calcium<sup>216</sup>. Les phospholipides identifiés comme marqueurs pour une survie courte, à savoir les PS, PA et PE, portent une charge négative, suggérant une relation spécifique entre les voies lipidiques et protéiques associées à un pronostic défavorable. Cibler cette voie à l'avenir pourrait potentiellement améliorer la survie des patients atteints de glioblastome. De plus, PPP1R12A (aussi appelé MYPT1) agit comme une sous-unité de la myosine phosphatase, régulant les voies de signalisation anti-tumorales telles que l'activation de l'expression de gènes suppresseurs de tumeurs comme Rb, Rap et c-Myc<sup>217</sup>. Par ailleurs, il existe une interaction établie entre les céramides et la protéine I2PP2A, conduisant à la dégradation de c-Myc<sup>218</sup>. Cela souligne un équilibre entre la présence de la protéine PPP1R12A, qui active c-Myc et améliore le pronostic, et les céramides, qui inhibent c-Myc et sont associés à une survie plus courte.

## Évaluation immunologique du glioblastome via l'identification de la signature SpiderMass des cellules immunitaires.

### Analyse des profils lipidiques des cellules immunitaires par le SpiderMass

L'infiltration de diverses populations de cellules immunitaires a été associée à l'évolution du cancer du côlon<sup>219–221</sup> et est désormais reconnue dans un éventail croissant de cancers<sup>222</sup>. Actuellement, l'immunescore est obtenu à partir de tissus excisés ou de biopsies par immunohistochimie, en utilisant des anticorps spécifiques pour différentes populations de cellules immunitaires. Cependant, l'immunescore est établi après l'intervention chirurgicale. Dans cette étude, nous avons exploré la possibilité de créer un immunescore basé sur les données de SpiderMass, qui pourrait éventuellement être obtenu *in vivo*.



**Figure 47 :** Classification SpiderMass des profils lipidiques des cellules immunitaires. A) Modèle LGBM entraîné pour les cellules cancéreuses, les lymphocytes et les macrophages de type M1 et M2 qui ont été analysés précédemment par la technologie SpiderMass. B) Trois exemples de biomarqueurs et leur boîte à moustaches correspondante pour les cellules cancéreuses du GBM, les lymphocytes et les macrophages de type M1 et M2.



Dans cette optique, nous avons analysé, avec la technologie SpiderMass, différentes populations de cellules immunitaires, telles que les macrophages (type M1 et M2), les lymphocytes T cytotoxiques CD8+, CD4+ et les cellules NK (regroupés sous le terme de lymphocytes), par rapport aux cellules cancéreuses (cellules de glioblastome NCH82). Les cellules immunitaires ont été directement analysées à partir de plaques de culture après tri cellulaire par cytométrie en flux, dans le but d'obtenir le profil moléculaire propre à chaque type cellulaire.

Étant donné que le SpiderMass n'est pas conçu pour analyser des cellules individuelles, des centaines de cellules sont analysées simultanément à partir d'un seul point d'analyse. Afin de compenser cette résolution spatiale plus restreinte par rapport à l'IHC, nous mettons en place une approche pour prédire le ratio d'un type de cellule à partir d'un spectre MS SpiderMass. Nous avons utilisé une bibliothèque Python appelée LGBM<sup>223</sup> pour développer un modèle de classification des cellules immunitaires. Celui-ci a été formé en utilisant 107 spectres MS pour chaque type de cellule, aboutissant à un taux de classification correcte de 100 % à la fois lors de l'entraînement et de la validation croisée (**Figure 47A**). En utilisant la nouvelle approche d'IA que nous avons mise en place, nous avons identifié 10 marqueurs lipidiques spécifiques des macrophages de type M1, 8 marqueurs lipidiques spécifiques des macrophages de type M2, 6 marqueurs lipidiques spécifiques des lymphocytes, ainsi que 3 marqueurs lipidiques spécifiques des cellules cancéreuses NCH82 (**Figure 47B**). Par exemple, les ions  $m/z$  818.65 (GlcCer d18:1\_22:0) et  $m/z$  819.55 (PG 18:1\_22:6) se sont avérés spécifiques des macrophages de type M1 et M2, respectivement. Concernant les lymphocytes, les phosphatidyléthanolamines sont identifiées comme marqueurs lipidiques, notamment PE (20:4\_16:0) ( $m/z$  738.55) et PE 38:3 ( $m/z$  768.55). Il est à noter que deux ions,  $m/z$  844.65 et  $m/z$  848.65, spécifiques des macrophages de type M1 et M2 respectivement, sont montrés comme spécifiques du tissu tumoral chez les patients ayant une survie inférieure à 10 mois.

#### *Distribution et corrélation des scores immunitaires dans les images SpiderMass en lien avec le pronostic du patient*

À partir d'une image SpiderMass, nous sommes maintenant en mesure d'obtenir la distribution des diverses populations de cellules immunitaires présentes dans le tissu. En effet, un second processus a été spécifiquement mis en place pour prédire la présence de chaque type de cellule à partir des images SpiderMass provenant des tissus de glioblastome frais congelés de la cohorte prospective. Les résultats obtenus génèrent des scores estimés pour chaque type de cellule, et ces scores sont ensuite utilisés pour calculer des ratios de présence relative (RPR) pour chaque type de cellule dans l'ensemble de l'image. Au total, six échantillons de tissus frais congelés ont été analysés, dont trois provenant de patients ayant survécu moins de 10 mois et trois provenant de patients ayant survécu plus de 36 mois (**Figure 48**). Ces ratios fournissent un aperçu de la



répartition des différentes populations cellulaires au sein de l'image, permettant ainsi une évaluation globale du microenvironnement cellulaire chez les patients avec des survies courtes ou longues. Il est pertinent de noter que le ratio prédit de cellules cancéreuses ne présente pas de différence significative en fonction de la survie.

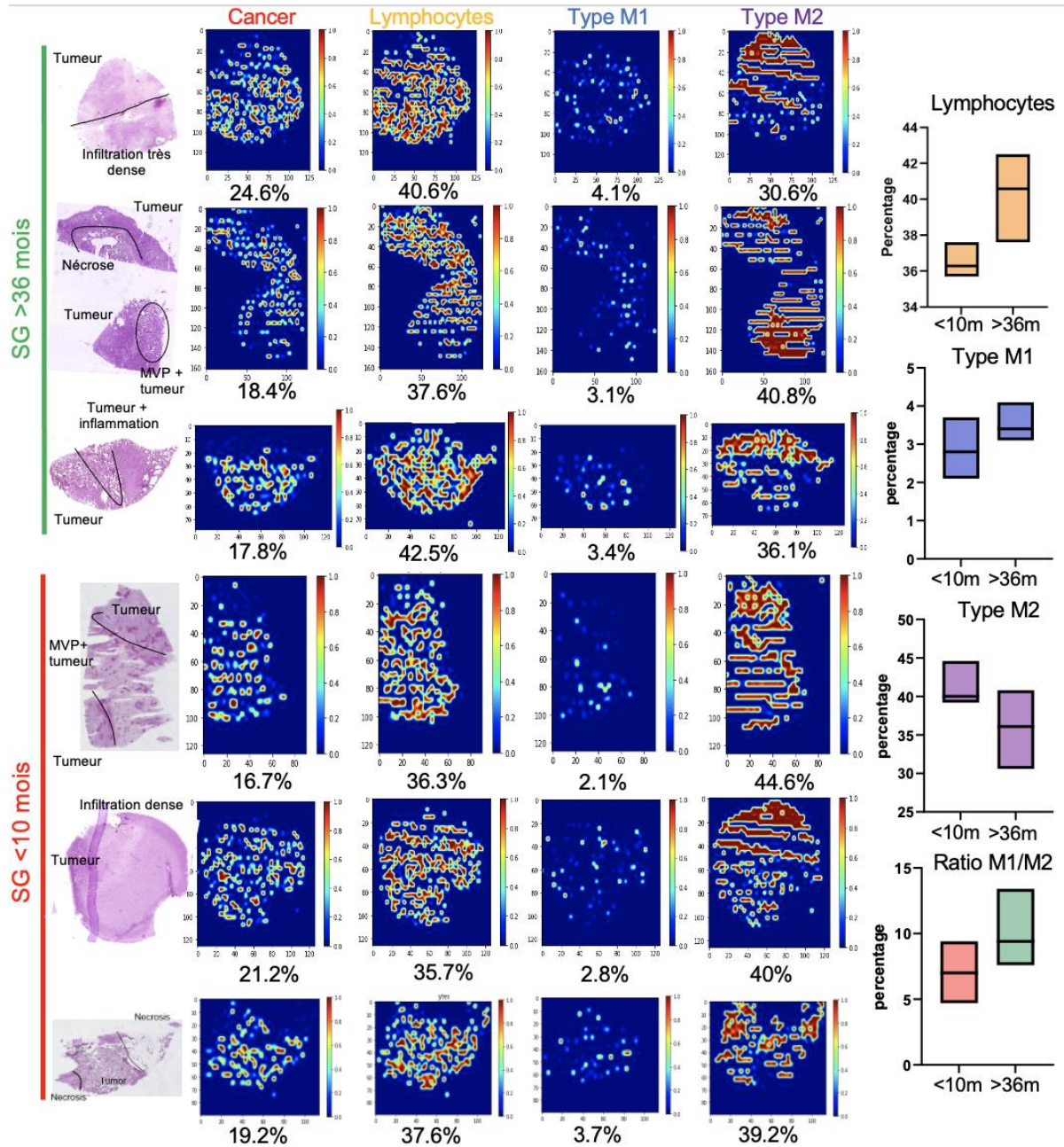
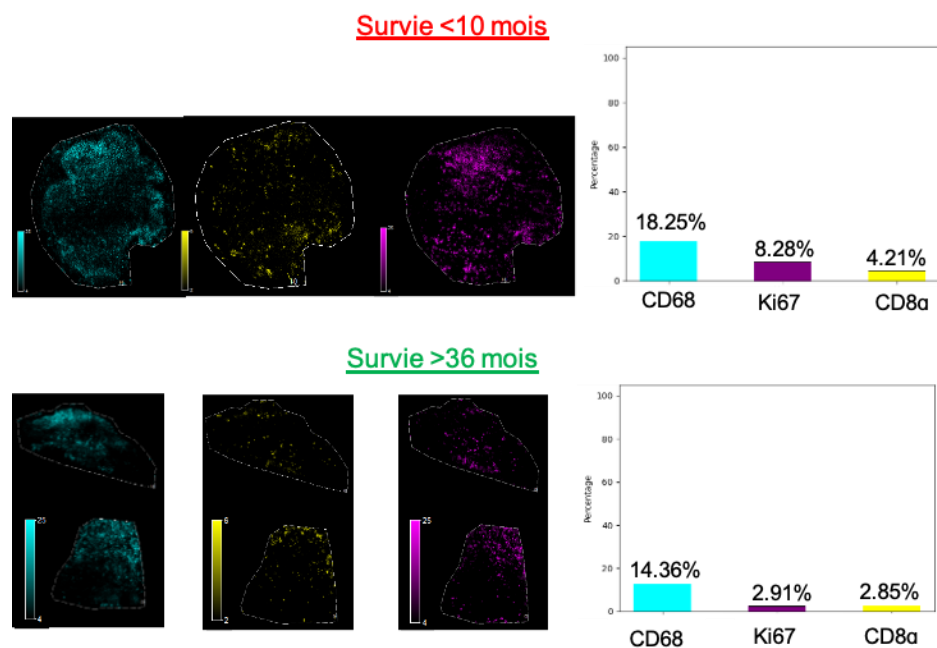


Figure 48 : Distribution des immunoscores sur six tissus frais congelés de GBM pour les cellules cancéreuses, les lymphocytes et les macrophages de type M1 et M2. Pour chaque cellule, le RPR en pourcentage est indiqué. Le pourcentage moyen de lymphocytes, de macrophages de type M1 et M2 et le ratio M1/M2 est représenté dans les tissus avec un pronostic <10 mois et >36 mois.

En revanche, les lymphocytes semblent plus abondants dans les tissus des patients ayant survécu plus de 36 mois. En effet, le RPR moyen des lymphocytes s'établit à plus de 40 % pour la survie prolongée, tandis qu'il s'est situé en dessous de 36 % pour la survie plus courte. De manière

similaire, le ratio prédit pour les macrophages met en lumière une relation intrigante entre les cellules immunitaires et la survie globale du patient. Plus spécifiquement, une proportion plus élevée de macrophages de type M1 (3.5 %) est détectée dans les tissus des patients ayant une survie longue que dans ceux des patients ayant une survie courte (2.8 %).

De façon contraire, les macrophages de type M2 manifeste un RPR moyen de 35.8 % pour les survies prolongées et de 41.3 % pour les survies courtes. Le rapport entre les macrophages de type M1 et les macrophages de type M2 pourrait devenir un indicateur de pronostic. En effet, lorsque la proportion de macrophages de type M1 par rapport aux macrophages de type M2 est inférieure à 7 %, cela est associé à une survie plus courte. À l'inverse, un tissu où les macrophages de type M1 représentent plus de 10 % est lié à une survie prolongée. De plus, il semble y avoir une corrélation entre le pronostic et l'équilibre entre la présence de lymphocytes et de macrophages. Plus précisément, en cas de pronostic défavorable, les macrophages ont tendance à prédominer par rapport aux lymphocytes (+22 %), tandis que l'inverse est observé chez les patients ayant un pronostic plus positif.



**Figure 49 :** Multiplex MALDI-IHC de 3 biomarqueurs sur deux coupes de tissus frais congelés de GBM. L'échelle d'affichage (unités arbitraires d'intensité) est la suivante (seuil d'intensité minimale/seuil d'intensité maximale) : 4/25 (Ki67), 2/6 (CD8α) et 4/25 (CD68).

Finalement, un 5-plex MALDI-IHC a été effectué pour corroborer les résultats précédents obtenus grâce à la nouvelle méthodologie d'immunoscoring. Pour ce faire, cinq biomarqueurs ont été sélectionnés, comprenant deux témoins : la vimentine et le collagène, servant de marqueurs matriciels cellulaires, ainsi qu'un marqueur de prolifération lié au cancer, le Ki67. De plus, des biomarqueurs pour les cellules immunitaires CD8α (cellules T cytotoxiques) et CD68 (macrophages) ont été inclus. La répartition de différentes populations cellulaires, notamment les

cellules immunitaires, obtenue par l'intégration de SpiderMass et de LGBM, est confirmée par MALDI-IHC (**Figure 49**). En effet, une concentration plus élevée de macrophages, de lymphocytes et de cancer est détectée dans le tissu d'un patient au pronostic défavorable.

Il est pertinent de souligner que l'approche SpiderMass-MSI, fondée sur l'immunescore, permet la distinction de diverses sous-populations immunitaires au sein du microenvironnement tumoral sans recourir à des méthodes sondées telles que l'IHC. En fin de compte, cette approche novatrice offre la possibilité de mieux appréhender la présence des cellules immunitaires dans le microenvironnement tumoral, tout en offrant la perspective d'une prédiction plus rapide de la survie des patients atteints de glioblastome.

## Conclusion et perspectives

Cette étude constitue une investigation approfondie visant à déterminer la faisabilité d'une classification histologique précise, ainsi que d'une sous-stratification et d'un pronostic pertinents pour le glioblastome. Pour atteindre cet objectif, la spectrométrie de masse à ionisation ambiante, grâce à la technologie SpiderMass, a été combinée avec un pipeline de prédiction basé sur l'intelligence artificielle. Nous avons exploré diverses méthodes de classification pour déterminer celle qui s'adaptait le mieux à notre ensemble de données moléculaires générées par la spectrométrie de masse. Étant donné l'absence de consensus au sein de la communauté scientifique et l'utilisation variée d'algorithmes, tels que LDA<sup>73</sup> et Lasso<sup>224</sup>, dans des contextes similaires avec des données très proches, cette démarche a été particulièrement pertinente. La combinaison de SpiderMass avec ce processus d'apprentissage automatique a permis de démontrer une perspective prometteuse pour avancer dans le domaine de la médecine de précision, tant pour le glioblastome que pour d'autres cancers caractérisés par leur hétérogénéité.

En effet, le modèle de classification optimal a été obtenu à l'aide du classificateur RidgeClassifier avec des taux de précision de 92 % et 88 % pour la classification des types histologiques de tissus, à savoir tumeur, nécrose et tissu bénin, respectivement, en utilisant les modes négatif et positif. Le RidgeClassifier<sup>225</sup> qui convient aux ensembles de données à haute dimension comme les nôtres avec 5000 dimensions  $m/z$ , est un classificateur linéaire qui utilise la régularisation L2 pour éviter l'ajustement excessif et qui a démontré de solides performances pour les données MS provenant de tissus<sup>226</sup>. Lors des prédictions en aveugle, où un ensemble de validation indépendant composé de dix tissus a été employé, aucune erreur de classification n'a été observée en mode ionique négatif, tandis qu'en mode positif, seulement deux spectres ont été incorrectement classés, cela est dû à leur localisation à la frontière entre la zone tumorale et la zone nécrotique. En complément, pour apporter une justification biologique à la classification basée sur les caractéristiques histologiques, nous avons intégré un processus de traitement des données secondaires reposant sur l'algorithme LIME. Cet algorithme, associé à des méthodes d'apprentissage automatique non supervisées et à des tests statistiques pour identifier des biomarqueurs potentiels indépendants de la classification, a été employé pour identifier des marqueurs lipidiques fiables. Grâce à cela, on a réussi à découvrir 41 biomarqueurs fiables, lesquels montrent des variations dans les classes de phospholipides en fonction du type de tissu. Les tissus bénins se démarquent par la présence d'acides gras polyinsaturés et l'absence de DG. Dans le cas des tissus nécrotiques, une absence de PI a été observée, tandis que les tissus tumoraux présentent une concentration significative de PS et de PA<sup>227,228</sup>.

De plus, cette méthodologie novatrice a réussi à clairement distinguer les trois sous-groupes moléculaires issus de la segmentation des lipides et des peptides par MALDI-MSI. La précision de la classification de ces sous-groupes a atteint 94 % et 88 % pour les deux modes d'ionisation respectifs. Il est important de noter qu'au sein des 18 biomarqueurs robustes identifiés grâce à une combinaison d'approches de découverte de marqueurs supervisées et non supervisées, trois biomarqueurs se sont déjà avérés aussi spécifiques à des types de tissus histologiques distincts. Un exemple significatif est un biomarqueur spécifique non seulement à la région tumorale, mais aussi à la zone bleue associée à la croissance tumorale. De même, deux autres biomarqueurs ont affiché une spécificité pour les régions bénignes et la zone rouge qui, à partir d'analyses protéomiques antérieures, est liée aux gènes du développement neurologique. Ces découvertes facilitent non seulement l'établissement de liens complexes entre les données protéomiques et lipidiques, mais elles ouvrent également la voie à l'utilisation de SpiderMass comme un outil *in vivo* pour la stratification des tissus atteints de glioblastome. Une telle stratification pourrait grandement contribuer à adapter les approches chirurgicales en fonction de la durée de survie projetée pour chaque patient.

De plus, ce nouveau processus d'apprentissage automatique ne se limite pas à une fonction de diagnostic, mais a également été évalué pour ses capacités pronostiques au sein de l'utilisation de SpiderMass. Le modèle de classification mis au point au cours de cette étude a réussi à différencier les tumeurs chez les patients dont la survie était inférieure ou supérieure à la médiane de 15 mois. Cependant, des conclusions plus précises ont été obtenues en examinant les patients avec des durées de survie nettement opposées, à savoir moins de 10 mois par rapport à plus de 36 mois. Dans les modes ioniques négatif et positif, des taux de précision de 93 % et 87 % respectivement ont été atteints pour cette distinction plus marquée. La présence de PCs et de PIs a été montrée en association à une survie plus longue, tandis que les PSs, PAs, PEs et Cer étaient plus fortement exprimés dans le tissu tumoral avec une survie plus courte. Nous démontrons également que des marqueurs lipidiques spécifiques sont liés à la classe histologique, au pronostic des patients et ainsi qu'à une stratification moléculaire identique entre lipides et protéines. Ces marqueurs lipidiques spécifiques étant, pour certains, liés à plusieurs groupes et permettant de lier, par exemple, la survie courte à la tumeur ainsi qu'au cluster bleu lié à la croissance tumorale. Cela se voit très facilement dans la **Figure 50** qui récapitule l'entièreté des biomarqueurs découverts tout au long de cette étude.

Le troisième aspect innovant de cette nouvelle stratégie réside dans l'exploitation de la technologie SpiderMass pour anticiper la durée de survie des patients atteints de glioblastome en fonction de l'infiltration des cellules immunitaires. Dans cette perspective, nous avons concentré



notre attention sur les macrophages et les lymphocytes comme première approche pour évaluer la faisabilité de prédire la présence d'une infiltration de cellules immunitaires en se basant sur les données générées par SpiderMass. Actuellement, l'immunescore repose sur l'immunohistochimie et est généralement effectué lors de l'examen pathologique après la chirurgie. La création d'un immunescore adapté à l'approche SpiderMass offrirait une manière de traduire l'immunescore en conditions *in vivo* et de prédire les résultats pendant l'intervention chirurgicale. Cette perspective permettrait au chirurgien d'adapter la résecabilité tumorale en fonction de l'évolution projetée et de personnaliser la thérapie, notamment en envisageant des traitements locaux peropératoires.

Pour ce faire, nous avons exploité les données des diverses cellules immunitaires (macrophages et lymphocytes) collectées par SpiderMass pour alimenter l'algorithme. Le phénotype M1 des macrophages correspond aux macrophages pro-inflammatoires et antitumoraux, tandis que le phénotype M2 correspond aux macrophages anti-inflammatoires et pro-tumoraux<sup>229</sup>. De manière frappante, nous avons démontré une association entre le type de macrophage (type M1 versus type M2) et le pronostic, révélant un rapport distinct entre les types M1 et les types M2 en fonction de la survie du patient. Les patients avec une survie réduite affichaient une prédominance de type M2 et une réduction de type M1, tandis que les patients avec une survie prolongée présentaient le schéma inverse de type M1 et de type M2. Ce phénomène est bien documenté dans le domaine oncologique, validant ainsi notre approche<sup>230-232</sup>. De plus, nous avons confirmé nos prédictions en utilisant une méthode de validation basée sur l'immunohistochimie par spectrométrie de masse, en employant des anticorps spécifiques ciblant les cellules immunitaires. Cette découverte de l'immunescore détient le potentiel de se transformer en un outil de prédiction du pronostic pour les patients atteints de glioblastome, éliminant le besoin de sondes spécifiques. Une telle avancée pourrait ouvrir la voie à des traitements personnalisés guidés par leur profil immunitaire. De plus, nous explorons le rôle crucial de l'infiltration des cellules immunitaires dans les tissus, mettant en évidence que le rapport entre les sous-populations de macrophages et des lymphocytes, peut servir à anticiper l'évolution du patient. La nouvelle méthodologie d'immunescores sur image SpiderMass permet de mettre en évidence l'utilisation multiple des données MS.

Ainsi, notre étude illustre que la technologie SpiderMass pourrait être employée pour réaliser une sous-stratification plus précise des patients atteints de glioblastome et fournir des informations pronostiques en temps réel. Cette approche permettrait d'optimiser à la fois la résection chirurgicale maximale et la gestion thérapeutique approfondie du patient.

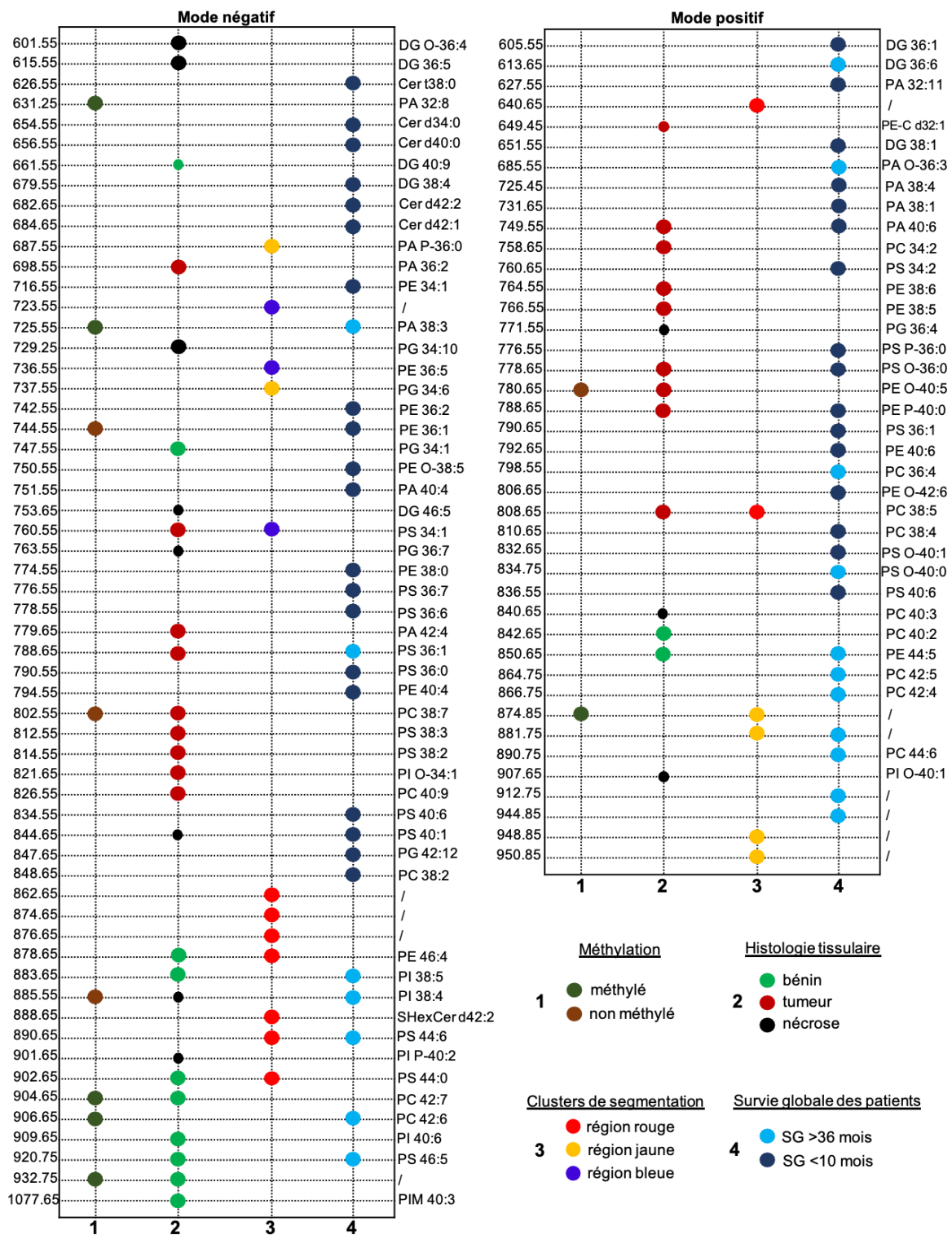


Figure 50 : Vue d'ensemble de tous les marqueurs lipidiques, trouvés à l'aide de la découverte supervisée et non supervisée, spécifiques de la méthylation de la tumeur, de l'histologie du tissu, de la sous-stratification de la tumeur GBM et de la durée de survie du patient, en mode positif et négatif.





Cette recherche marque un progrès important dans la compréhension du glioblastome. Elle démontre l'efficacité de la combinaison entre la spectrométrie de masse ambiante par SpiderMass et l'intelligence artificielle pour obtenir des classifications histologiques précises ainsi qu'une évaluation pronostique utile pour cette maladie complexe.

Ce troisième chapitre permet aussi de mettre en évidence la possibilité d'établir un immunescore spécifique pour l'approche SpiderMass. Cela ouvrirait la voie à une traduction de cet immunescore en conditions *in vivo*, permettant ainsi de prédire les résultats durant l'intervention chirurgicale. Cette possibilité offrirait au chirurgien la capacité d'ajuster la résection de la tumeur en fonction de l'évolution anticipée, ce qui permettrait une personnalisation plus précise du traitement, y compris l'envisagement de traitements locaux peropératoires.

Cette nouvelle approche méthodologique sera mise en œuvre et appliquée dans le quatrième chapitre de cette thèse, qui se consacre à l'analyse approfondie du cancer œsogastrique. Grâce à cette méthodologie novatrice, nous pourrions explorer en détail les différents aspects du cancer œsogastrique. Cette méthodologie apportera une contribution significative à notre compréhension globale de cette maladie, permettant ainsi des conclusions plus fiables dans le contexte de notre recherche.



## Chapitre 5 : Diminution de la rechute des patients de cancer œsogastrique grâce au diagnostic par l'analyse *in vivo* SpiderMass

---

### Introduction

La deuxième cause de décès la plus répandue au monde en 2020, avec près de 19.3 millions de nouveaux cas et 10 millions de décès, est le cancer<sup>1</sup>. Cette étude se focalise spécifiquement sur le cancer œsogastrique, qui englobe à la fois les cancers de l'estomac et de l'œsophage. Le cancer de l'estomac présente une incidence significative à l'échelle mondiale, se classant au cinquième rang en termes de fréquence, et touchant plus d'un million de personnes chaque année dans le monde. Son taux de mortalité est alarmant, avec environ 800 000 décès, ce qui en fait la quatrième cause de décès par cancer, avec un décès sur 13 lié au cancer. Le cancer de l'œsophage, quant à lui, se classe au septième rang en termes d'incidence, affectant plus de 600 000 personnes chaque année, avec un taux de mortalité dépassant les 90 % (544 076 décès en 2020), dont 418 400 cas chez les hommes et 185 800 chez les femmes. En d'autres termes, en 2020, environ un décès sur 18 était attribuable au cancer de l'œsophage. Le cancer œsogastrique est généralement diagnostiqué à un stade avancé en raison de symptômes peu spécifiques et latents, ce qui se traduit par un pronostic sombre. Des études ont montré qu'une détection précoce et précise de ce cancer pourrait augmenter le taux de survie à cinq ans à environ 90 %<sup>233</sup>.

L'endoscopie joue un rôle essentiel dans la détection du cancer œsogastrique car elle permet aux chirurgiens d'observer directement les sites cancéreux et leurs étendues. Il est très important d'établir un diagnostic précis à l'aide d'images endoscopiques afin d'améliorer le pronostic du patient. Cependant, des études récentes ont montré que la précision de détection de l'endoscopie conventionnelle n'atteignait que 79 %<sup>234</sup>. En raison de la lourde charge de travail que représente l'analyse des images médicales, des erreurs de diagnostic ont inévitablement lieu.

Bien que les thérapies adjuvantes telles que la chimioradiothérapie, la chimiothérapie et l'immunothérapie puissent améliorer dans une certaine mesure la survie des patients atteints de GC, la chirurgie radicale demeure la principale option pour traiter les cancers de cette région<sup>235</sup>. La résection curative avec des marges microscopiquement négatives (aucune cellule cancéreuse identifiée au niveau de la marge de résection lors de l'examen pathologique) a été acceptée comme le traitement le plus efficace sur la base de la philosophie chirurgicale, car même un nombre minime de cellules cancéreuses restantes peuvent engendrer des récurrences<sup>236</sup>. En effet, il est largement reconnu que l'ablation complète de la tumeur est associée à une meilleure survie globale et donc un meilleur pronostic pour la plupart des types de cancers solides. En revanche, des marges positives,

qui se définissent par la présence de cellules cancéreuses sur les bords de la zone résectionnée ou dans la zone d'excision, sont liées à une augmentation du risque de récurrence locale et à une réduction de la survie globale dans de nombreux types de cancers. Pour éviter de laisser des marges positives, une analyse extemporanée est généralement effectuée dans une salle proche de la salle d'opération, où les marges proximales et distales de la zone chirurgicale sont étudiées après congélation de la pièce opératoire, sa découpe et sa coloration puis son examen par un anatomopathologiste en peropératoire. Les résultats de cette évaluation pathologique peropératoire des marges sont ensuite communiqués au chirurgien et orientent ses décisions et son geste chirurgical. Par exemple, en cas de marges positives, le chirurgien peut décider de prolonger la résection en retirant davantage de tissu autour de la tumeur, ce qui nécessite souvent une réévaluation des marges par le pathologiste. Cependant, l'examen extemporané présente certaines limitations. En effet, le temps nécessaire à cette pratique (généralement au moins 20 minutes) ne permet l'examen que de quelques pièces, qui sont par ailleurs choisies par le chirurgien sur une base subjective. De plus, dans le laps de temps imparti, le pathologiste ne peut réaliser que des colorants histologiques et ne peut mettre en œuvre des techniques plus ciblées comme l'IHC. De plus, en raison des problèmes potentiels liés à la congélation des tissus et à la complexité de la morphologie des tissus, les résultats des analyses en temps réel des marges sont parfois subjectifs et peuvent manquer de fiabilité. Les anatomopathologistes eux-mêmes estiment que le taux d'erreur diagnostique lors de l'examen extemporané peut monter jusqu'à 30 % des patients subissant une résection pour un cancer œsogastrique. Ceci est particulièrement notable, dans des cas de diagnostic de certains sous-types de cancer œsogastrique comme le sous-type PCC, qui est plus rare, mais très agressif<sup>237</sup>.

Pour toutes ces raisons, le cancer œsogastrique a été le sujet de nombreuses études omiques, notamment par des approches non ciblées à grande échelle basées sur la spectrométrie de masse, comme la protéomique shotgun<sup>238,239</sup> ou l'imagerie par spectrométrie de masse<sup>240</sup>. Diverses modalités de MSI, dont le MALDI et le DESI, ont permis d'identifier des biomarqueurs potentiels et des cibles thérapeutiques pour cette maladie<sup>241,242</sup>. Cependant, si ces approches permettent la découverte de nouveaux marqueurs, elles ne répondent pas au besoin clinique de l'analyse peropératoire pour l'aide à la décision pendant la chirurgie. Mais la spectrométrie de masse a récemment évolué grâce au développement de nouvelles techniques d'ionisation en milieu ambiant (AIMS) et de nouvelles technologies adaptées à l'évaluation peropératoire qui sont apparues au cours de ces 10 dernières années<sup>10</sup>.

Dans un contexte d'utilisation de la MS au bloc opératoire pour la prise de décision du chirurgien, il est nécessaire de coupler les données générées par MS avec des solutions d'IA, afin de fournir une interprétation automatique des données collectées et les rendre lisibles par les

cliniciens<sup>208</sup>. Les techniques d'IA, qui simulent les processus cognitifs humains à l'aide d'un ordinateur, ont la capacité de traiter de grandes quantités de données, ce qui peut être extrêmement utile pour aider les chirurgiens gastro-entérologues à poser un diagnostic clinique et à prendre les meilleures décisions pour la prise en charge de leurs patients<sup>243</sup>. Jusqu'à présent, l'IA a été appliquée avec succès dans de nombreuses branches de l'imagerie médicale, telles que l'endoscopie, la pathologie et la tomodensitométrie. Dans le domaine du diagnostic assisté par l'IA en endoscopie, cela englobe la reconnaissance des caractéristiques visuelles dans les images<sup>244</sup>, la détection précoce du cancer gastrique<sup>245</sup>, la détection de conditions précancéreuses<sup>246</sup>, l'amélioration de l'endoscopie en utilisant l'imagerie à bande étroite<sup>247</sup> et l'application de l'endoscopie Raman<sup>248</sup>. L'IA intervient également dans l'assistance au diagnostic assisté en anatomopathologie, qui comprend l'identification automatique du cancer oesogastrique<sup>249</sup>, la détection basée sur l'imagerie de la lame entière<sup>250</sup>, la détection automatique des lymphocytes infiltrant la tumeur<sup>251</sup> et la segmentation des zones lésionnelles<sup>252</sup>. En ce qui concerne la tomodensitométrie assistée par l'IA, l'accent est mis sur l'identification des métastases péritonéales préopératoires<sup>253</sup> et la détection des ganglions lymphatiques métastatiques péri-gastriques<sup>254</sup>. Dans beaucoup de cas, les performances diagnostiques de ces modèles d'IA sont comparables, voire supérieures, à celles des experts humains.

Dans le domaine de la spectrométrie de masse appliquée au diagnostic du cancer, les algorithmes d'apprentissage automatique supervisés jouent un rôle essentiel en élaborant des modèles de classification précis à partir de données annotées<sup>209</sup>. C'est pourquoi dans cette étude, nous avons utilisé la nouvelle méthodologie d'IA développée dans le chapitre 2 sur le glioblastome en combinaison avec la technologie SpiderMass dans l'objectif de permettre un diagnostic peropératoire en temps-réel du cancer œsogastrique précis, permettant au chirurgien de retirer les tissus présentant des cellules cancéreuses en épargnant au maximum les tissus sains, limitant *de facto* la rechute des patients et augmentant leur survie globale. En plus de son utilisation pour le diagnostic du cancer œsogastrique ainsi que pour le sous-typage des cancers, la technologie SpiderMass a aussi été utilisée afin de prédire la probabilité de présence de différentes souches bactériennes dans les tissus gastriques. De plus, une analyse par protéomique spatiale des tissus de cancer œsogastrique a pu être réalisée via la méthodologie de l'irradiation backside par la technologie SpiderMass. Cette dernière étude permet de mieux appréhender les mécanismes physiopathologiques des cancers œsogastriques en montrant le lien entre les bactéries, les lipides et les protéines dans le cancer œsogastrique.

Figure 51 reprend la stratégie globale d'analyse multi-omique qui a été mise en œuvre autour du cancer œsogastrique grâce aux différentes utilisations de la technologie SpiderMass, de l'imagerie MALDI et de la protéomique shot-gun.

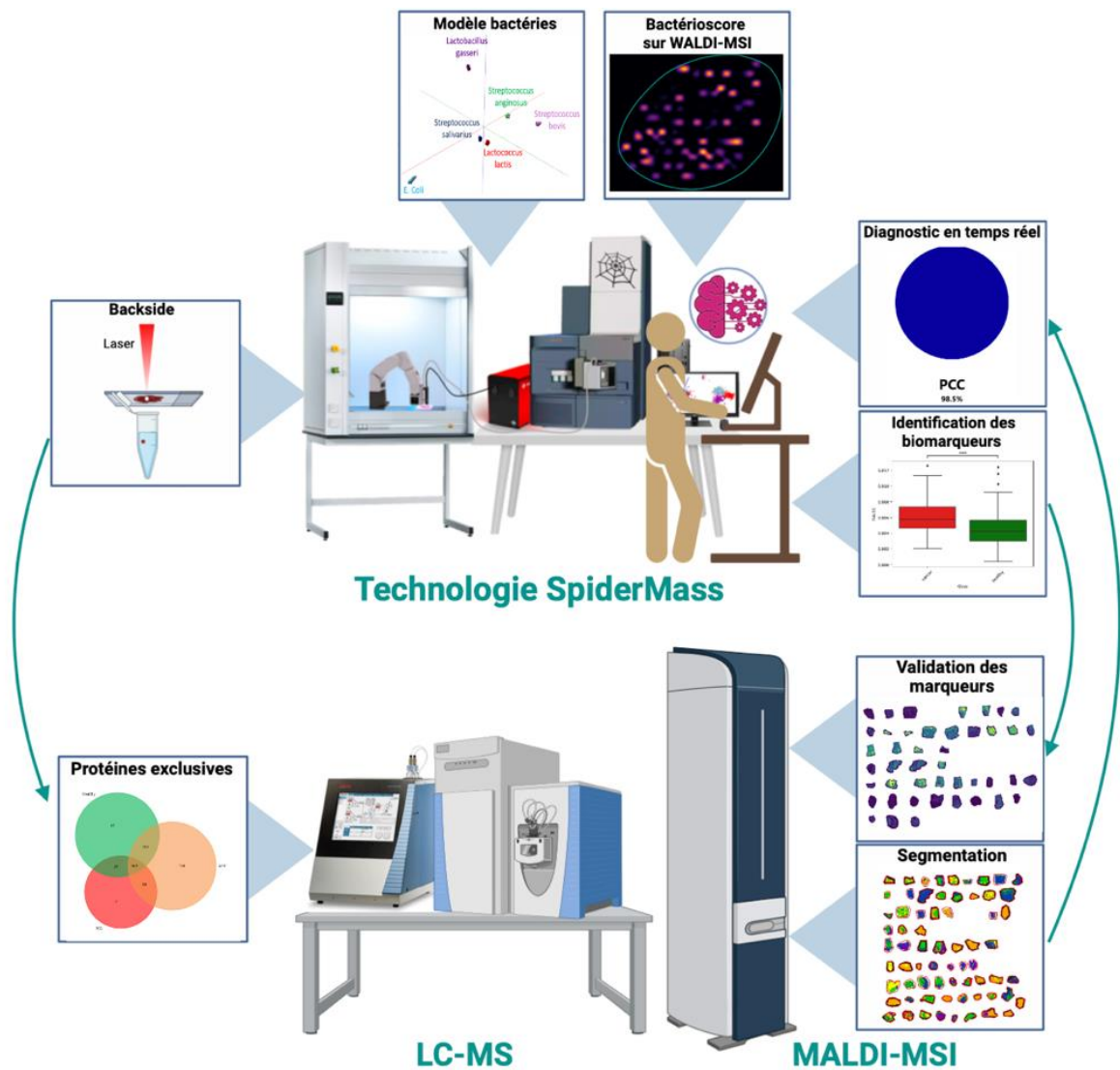


Figure 51 : Aperçu général de l'ensemble des méthodologies utilisées pendant cette étude sur le cancer œsogastrique.



## Matériel et méthodes

### Échantillons

Les tissus excisés lors de la chirurgie sur des patients atteints de cancer œsogastrique ont été obtenus au travers de la base de données clinico-biologiques nationale de l'INCA sur les tumeurs œsophagiennes et gastriques (FREGAT, <https://fregat-database.org/en/>), suite à l'approbation par le comité de la tumorotheque du CHU de Lille. Ces échantillons sont préservés par congélation et conservés à une température de  $-80^{\circ}\text{C}$ . Tous les patients ont donné leur accord éclairé avant d'être inclus dans l'étude. Cette cohorte prospective comprend 190 régions d'intérêt (ROI) venant de 176 échantillons de tissus frais congelés provenant de 114 patients, parmi lesquels 77 sont des tissus sains, 85 sont des tissus cancéreux et 13 sont des tissus mixtes, correspondant à 15 régions cancéreuses et 13 régions saines (**Tableau 8**). Les 100 régions d'intérêt cancéreuses sont divisées selon les sous-types de cancer œsogastrique, tel que 62 et 38 ROIs correspondent aux sous-types ADK et PCC respectivement.

**Tableau 8** : Données démographiques et cliniques de la cohorte.

		Cancer (100 ROI)		Sain (90 ROI)
		ADK (62 ROI)	PCC (38 ROI)	
Age, médiane (min-max)		67 (32-90)	58 (24-86)	61 (24-90)
Organe	Œsophage, <i>n</i>	27	6	71
	Estomac, <i>n</i>	35	32	19
Sexe	Femme, <i>n</i>	11	17	26
	Homme, <i>n</i>	51	21	64

### Préparation des échantillons

Tous les tissus ont été sectionnés sur un cryostat Leica CM1510S (Leica Biosystems, Nanterre, France). Trois coupes consécutives ont été réalisées par prélèvement : l'une ( $5\ \mu\text{m}$ ) pour la coloration HPS, la deuxième ( $20\ \mu\text{m}$ ) pour l'analyse SpiderMass, et la dernière ( $12\ \mu\text{m}$ ) pour l'analyse MALDI-MSI. Les différentes coupes ont été respectivement montées sur deux lames de verre standard et la dernière sur une lame de verre conductrice ITO.

### Coloration histologique et annotations

La coupe de tissu d'une épaisseur de  $5\ \mu\text{m}$  a été soumise à une séquence d'étapes. Premièrement, elle a été traitée avec une solution d'hémalum pendant une durée d'une minute, suivie d'un rinçage à l'eau courante. Ensuite, elle a été immergée dans des solutions d'éthanol à des concentrations de 70 % et de 100 % pour réaliser une déshydratation des tissus. Elle a préalablement été colorée pendant 10 secondes dans une solution de phloxine à 0.1 %. Une fois la coloration terminée, la coupe a été délipidée par un bain de xylène, puis deux bains d'alcool successifs. Enfin,

une immersion rapide dans du safran pendant 5 secondes a été effectuée avant de fixer la section en la recouvrant d'une lamelle de couvre-objet et du milieu de montage pour lames EUKITT. Afin d'obtenir une image, la lame colorée a été numérisée au moyen du scanner de diapositives Panoramic MIDI de 3DHISTECH LTD. (Budapest, Hongrie). Les images produites ont été ouvertes et exportées en utilisant QuPath 0.2.3. Les annotations par le pathologiste ont été faites sur les tissus ayant subis une coloration histologique.

### Analyse MALDI-MSI

La matrice norharmane à une concentration de 7 mg/mL dans un mélange  $\text{CHCl}_3$ :MeOH (2:1, v:v) a été employée comme matrice pour l'analyse MALDI-MSI dans les deux modes d'ionisation (positif et négatif). L'application de la matrice a été réalisée avec le sprayer HTX-M5 TM-Sprayer (HTX Technologies, LLC). Les paramètres utilisés pour le dépôt de matrice sont les suivants : une température de buse maintenue à 30°C, douze passages en suivant un motif en zigzag, un débit de 0.1 mL/min, une vitesse de déplacement de la buse à 1200 mm/min, un espacement de 2 mm entre les passages, et une pression du gaz à 10 psi. L'analyse MALDI-MSI a été exécutée dans les deux polarités sur un spectromètre de masse MALDI-TOF (Rapiflex, Bruker), avec l'utilisation du phosphore rouge comme calibrant externe. Les données d'imagerie par spectrométrie de masse ont été acquises en utilisant une résolution spatiale de 50  $\mu\text{m}$  et une zone de balayage du faisceau de  $20 \times 20 \mu\text{m}^2$  pour la modalité d'imagerie en double polarité. La gamme de masse a été fixée entre  $m/z$  400 et 1500.

### Analyse SpiderMass

Le système est constitué de trois composants principaux : le spectromètre de masse en lui-même, un dispositif laser conçu pour réaliser des prélèvements micro-échantillonnés à distance des tissus, ainsi qu'une ligne de transfert facilitant le déplacement du matériel micro-échantillonné. Le composant central consiste en un laser pulsé Nd:YAG (avec une durée d'impulsion de 5 ns). Il sert de source d'énergie pour un OPO à longueur d'onde ajustable (Radiant version 1.0.1, fourni par OPOTEK Inc. basée à Carlsbad, en Californie, aux États-Unis). Une fibre laser biocompatible d'une longueur de 1 mètre et d'un diamètre intérieur de 450 microns est connectée à la sortie du système laser. À l'extrémité de cette fibre, une pièce à main équipée d'une lentille de focalisation de 4 cm est fixée. L'intensité du laser a été réglée à 4 mJ par impulsion, avec une durée d'irradiation constante de 10 secondes, ce qui a permis d'atteindre une fluence laser d'environ 3 J/cm<sup>2</sup>. Le deuxième composant crucial du dispositif est une ligne de transfert de 2 mètres fabriquée à partir de tubes Tygon ND 100-65, ayant un diamètre intérieur de 2.4 mm et un diamètre extérieur de 4 mm. Cette ligne de transfert établit une liaison directe entre le spectromètre de masse (Xevo, Waters), dont la source d'électrospray classique a été substituée par une interface REIMS, et la

pièce à main laser positionnée à l'autre extrémité. Pendant chaque analyse, une infusion d'isopropanol à un débit de 200  $\mu\text{L}/\text{min}$  a été utilisée. Dans le but de servir de calibrant de masse, une concentration de 200  $\mu\text{g}/\text{mL}$  de leucine enképhaline a été incorporée à l'infusion. La localisation des tirs d'analyse laser a été déterminée en se basant sur les annotations histopathologiques. L'acquisition des données impliquait une succession de 10 tirs laser, chacun générant un spectre individuel. L'acquisition spectrale a été conduite en modes positif et négatif en configuration sensibilité. La gamme de masse analysée s'étendait de  $m/z$  50 à 2000.

### Analyse WALDI-MSI

La technologie SpiderMass a été détaillée dans la section précédente. Pour réaliser l'analyse d'imagerie, la microsonde SpiderMass a été connectée à un bras robotique rigide. Le faisceau est focalisé à 500  $\mu\text{m}$  mais les images ont été acquises en mode « oversampling » de façon à atteindre une résolution spatiale de 250  $\mu\text{m}$ . De plus, pour chaque image obtenue, la résolution spatiale finale a été améliorée par un facteur de deux à travers un processus d'interpolation. La gamme de masse analysée s'est étendue de  $m/z$  100 à 1500. Le processus d'acquisition consistait en une séquence de 3 tirs laser successifs, avec un intervalle de 3 secondes entre chaque tir. Les séquences de tirs laser ainsi que l'acquisition de données par le spectromètre ont été automatisées grâce à une interface utilisateur développée sous MATLAB spécifiquement pour le système robotique WALDI-MSI. Les données d'imagerie ont été acquises dans le mode négatif. Le Q-TOF était réglé sous le mode sensibilité.

### Analyse des souches bactériennes

Les différentes souches bactériennes qui ont été étudiées sont *Escherichia Coli* (ATCC 12435), *Lactococcus lactis* (ATCC 19435), *Lactobacillus gasseri* (ATCC 33323), *Streptococcus bovis* (ATCC 33317), *Streptococcus anginosus* (ATCC 700231) et *Streptococcus salivarius* (ATCC 7073). Toutes ces souches ont été obtenues auprès de l'American Type Culture Collection (ATCC). Concernant leur culture, *E. coli* a été cultivée sur des géloses Tryptone et NaCl, tandis que *Lactococcus lactis* et *Streptococcus salivarius* ont été ensemencées sur des géloses Brain Heart Infusion (BD 211065). Par ailleurs, *Streptococcus anginosus* et *S. bovis* ont été cultivées sur des géloses Tryptic Soy (BD 236950), et enfin *Lactobacillus gasseri* a été cultivée sur des géloses Lactobacilli MRS (BD 288210). Toutes les cultures ont été incubées à une température de 37°C pendant 24 heures. L'analyse par la méthode SpiderMass a été réalisée directement dans les boîtes de Pétri sans aucune préparation de l'échantillon.

### Bactérioscore

Le modèle d'analyse des souches bactériennes a été formé en utilisant la bibliothèque Python LGBM, développé par Microsoft. Ce modèle a été construit sur la base des spectres

lipidiques enregistrés dans la plage de  $m/z$  de 600 à 1100 en mode négatif. Les spectres ont été classifiés en différents types de souches bactériennes : *Escherichia Coli*, *Lactococcus lactis*, *Lactobacillus gasseri*, *Streptococcus anginosus*, *Streptococcus bovis* et *Streptococcus salivarius*, chacun comprenant 75 spectres. Pour prédire la souche bactérienne présente dans les images SpiderMass, la fonction `predict_proba` du modèle LGBM a été utilisée. Cette fonction génère des estimations de probabilité pour chaque variété bactérienne. En complément, des scores de ratio ont été calculés pour évaluer la proportion relative de chaque souche bactérienne à travers l'ensemble de l'image. Ces scores ont été obtenus en sommant les scores attribués à chaque souche bactérienne, et en divisant par la somme totale des scores associés à toutes les souches. Ces scores ont procuré des indications sur la distribution des diverses souches bactériennes à travers l'intégralité de l'image, fournissant ainsi une vue exhaustive du paysage bactérien dans l'échantillon analysé.

## Protéomique spatiale

### *Technique du backside grâce à la technologie SpiderMass*

La technique SpiderMass peut également être utilisée pour réaliser un micro-échantillonnage d'une région d'un tissu de la taille du faisceau laser au point focal. Ce micro-échantillonnage est très similaire à ce qui peut être obtenu après microdissection laser (LCM). L'avantage du SpiderMass eu égard à sa longueur d'onde et l'énergie qu'il permet d'atteindre est que le verre est transparent à la longueur d'onde utilisée et que l'irradiation peut être réalisée par l'arrière avec une ablation du tissu vers l'avant de la lame, ce qui facilite la collette du tissu ablaté. Le tissu ablaté collecté peut ensuite être traité pour réaliser tout type d'analyse omique. Dans le cas du cancer oesogastrique, nous nous sommes intéressés à une approche de protéomique à grande échelle non ciblée par shot-gun.

Ainsi, dix échantillons de tissus frais congelés, provenant respectivement des tissus sains et des tissus cancéreux de sous-type PCC et ADK, ont été soumis à un traitement d'irradiation par l'arrière avec le système SpiderMass. Cette approche novatrice implique le tir d'un laser d'énergie calibrée à 4.2 mJ depuis l'arrière de la lame sur laquelle le tissu est déposé. Comme le montre la **Figure 52**, le setup utilisé au laboratoire comprend la technologie SpiderMass à laquelle ont été rajoutés deux supports créés via une imprimante 3D. Le premier support a pour but de tenir la lame sur laquelle le tissu a été déposé et le deuxième support permet de tenir les tubes de récoltes juste en dessous de la lame.

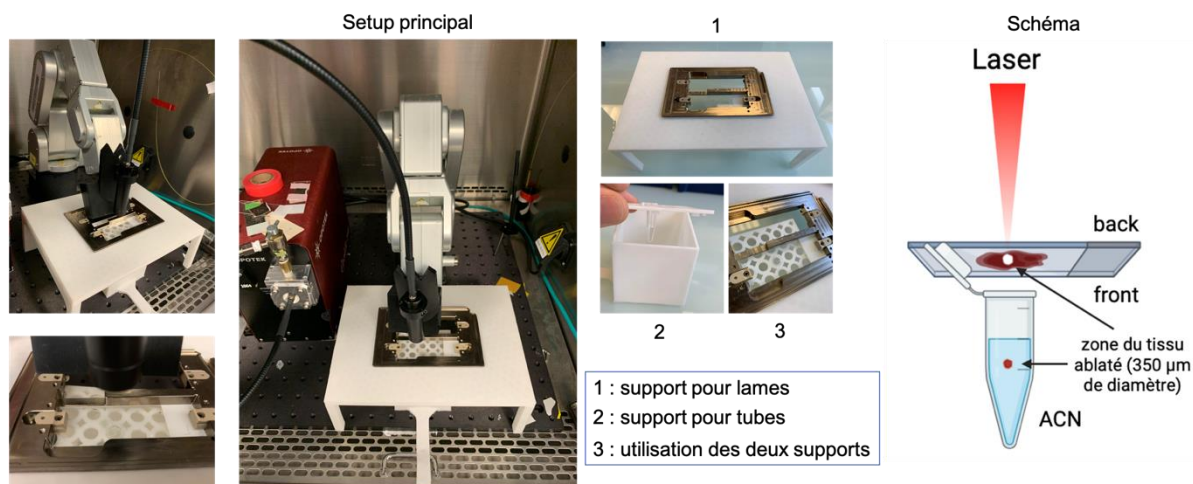


Figure 52 : Installation au laboratoire et schéma de la technique de l'irradiation par backside grâce au SpiderMass.

Ce procédé permet une forme de microdissection laser, qui extrait des portions de tissu d'un diamètre approximatif de 350  $\mu\text{m}$ . Les fragments de tissu extraits grâce à la micro-ablation SpiderMass ont été collectés dans un tube contenant 200  $\mu\text{l}$  d'acétonitrile (ACN), suivi d'une brève agitation et d'une sonication pendant 5 minutes. Ensuite, le matériel collecté a été séché à l'aide d'un dispositif SpeedVac. Par la suite, une solution de trypsine à une concentration de 40  $\mu\text{g}/\mu\text{l}$ , à raison de 50  $\mu\text{l}$ , a été ajoutée pour permettre une digestion liquide durant la nuit. Le jour suivant, afin d'ajuster l'acidité, 1  $\mu\text{l}$  de TFA à 5 % a été ajouté. Enfin, l'échantillon a de nouveau été séché à l'aide du SpeedVac avant d'être soumis à une analyse par LC-MS/MS.

#### *Analyse LC-MS et MS/MS*

Les échantillons séchés ont été réhydratés en utilisant 20  $\mu\text{L}$  d'une solution d'eau contenant 0.1 % de TFA. Ces échantillons réhydratés ont ensuite subi une étape de dessalage en employant des cônes ZipTip chargés avec de la résine C18, conformément aux indications du fabricant (Millipore, Saint-Quentin-en-Yvelines). Après le dessalage, les échantillons ont été séchés sous vide puis reconstitués dans un mélange de 20  $\mu\text{L}$  d'ACN et de 0.1 % d'acide formique (AF) dans un rapport de 98:2 (v/v).

Les échantillons ont ensuite été analysés par LC-MS/MS. Un système EASY-nLC 1000 (Thermo Scientific) a été couplé à un spectromètre de masse Q-Exactive Orbitrap (Thermo Scientific), relié par une source nanoélectrospray. Deux colonnes ont été utilisées : une colonne de préconcentration (Proxeon) de dimensions 75  $\mu\text{m}$  ID x 2 cm avec une taille de particules de 3  $\mu\text{m}$ , et une colonne analytique (Acclaim PepMap) de dimensions 75  $\mu\text{m}$  ID x 50 cm avec une taille de particules de 2  $\mu\text{m}$ , toutes deux fournies par Thermo Scientific. Un gradient linéaire sur 2 heures composé d'acétonitrile dans 0.1 % d'AF a été employé, avec une concentration d'acétonitrile variant de 2 % à 30 %. Le débit a été maintenu à 300 nL/min. L'acquisition a été réalisée en mode DDA, en travaillant en Top 10. La gamme de masse pour l'analyse par spectrométrie de masse a été fixée

entre  $m/z$  300 et 1600, avec une résolution de 70 000 FWHM (largeur totale à moitié de maximum), un contrôle automatique du gain (AGC) réglé à  $3 \times 10^6$  ions et un temps d'injection maximal de 120 ms. L'analyse MS/MS a été réalisée dans la gamme de masse  $m/z$  de 200 à 2000, avec un AGC de  $5 \times 10^4$  ions, un temps d'injection maximal de 60 ms et une résolution de 17 500 FWHM. Afin de réduire l'impact potentiel d'effets de batchs durant l'analyse, les échantillons à analyser a été randomisée, créant ainsi des séquences d'analyse impartiales.

## Traitement et analyse des données

### *Données MALDI-MSI*

Les données MALDI-MSI ont été analysées sous le logiciel SCiLS Lab MVS 2023b (Bruker Daltonik GmbH). Afin de garantir la cohérence, les données ont été normalisées au moyen de la normalisation RMS, et ce, en appliquant une configuration de débruitage modérée pour les deux polarités. Le pipeline de segmentation par défaut a été employé pour une sélection automatisée des pics. Pour chaque échantillon de tissu, une méthode de segmentation spatiale globale a été mise en œuvre dans le but de mettre en lumière les variations dans la composition moléculaire des tissus et mettre en évidence les différents phénotypes cellulaires. Ce processus de segmentation repose sur le regroupement des pixels ayant des spectres similaires, ce qui aboutit à la formation de segments visuellement distincts et étendus, chacun étant associé à une couleur différente. Les pixels faisant partie d'un même segment sont colorés en conséquence. Pour visualiser cette segmentation, un dendrogramme est généré, les branches étant établies en fonction des distances de corrélation entre les groupes. Les distances plus courtes traduisent une plus grande similitude dans les profils moléculaires entre les segments.

### *Données SpiderMass*

Toutes les données brutes générées par SpiderMass ont été traitées conformément aux explications fournies dans le chapitre précédent. En résumé, la bibliothèque Lazy Predict a généré plusieurs modèles en utilisant scikit-learn, lesquels ont été entraînés et évalués en utilisant 24 classificateurs. Le modèle optimal a ensuite été individuellement reconstruit dans l'environnement de scikit-learn, avec un affinement minutieux de ses paramètres. Pour évaluer la performance, une validation croisée à 20 reprises a été mise en œuvre, et un rapport de classification a été établi. Dans le but d'effectuer des prédictions à l'aveugle, le modèle a été enregistré en utilisant le module joblib. Pour mieux comprendre et interpréter le modèle, des explications locales interprétables (algorithme LIME) ont été employées. Ces explications ont permis d'attribuer des contributions positives/négatives à chaque ion  $m/z$  et de les visualiser sous forme de tableau grâce à l'outil ELI5. Pour réduire le bruit, un algorithme de sélection de pics provenant de la bibliothèque scipy a été utilisé. En outre, afin d'évaluer l'importance de chaque ion  $m/z$ , le test de Kruskal-Wallis avec



correction de Bonferroni a été appliqué de façon automatique. Seuls les ions  $m/z$  avec une valeur  $p \leq 0,05$  ont été retenus, et les pics correspondant à des ions mono-isotopiques ont été filtrés. Les boîtes à moustaches ont été générées en utilisant la bibliothèque seaborn. Pour l'identification de biomarqueurs lipidiques fiables, les approches supervisées et non supervisées ont été conjuguées, exigeant à la fois une signification statistique, une expression appropriée et une corrélation de contribution au sein d'une même région d'intérêt.

### *Données protéomiques*

Le logiciel MaxQuant (version 2.3.1.0)<sup>255</sup>, ainsi que le moteur de recherche Andromeda<sup>256</sup>, ont été utilisés pour l'identification à partir des données de LC-MS/MS. Cette recherche a été effectuée par rapport au protéome complet de l'Homo sapiens, en utilisant spécifiquement la version de juillet 2023 de la base de données UniProt. L'enzyme trypsine a été choisie, permettant deux clivages manqués, et les modifications variables comprenaient l'acétylation N-terminale et l'oxydation de la méthionine. Les paramètres de précision de masse pour les spectres MS et MS/MS étaient respectivement de 6 ppm et 20 ppm. Pour évaluer la crédibilité des correspondances de spectre peptidique (PSM) et des protéines, le taux de fausse découverte (FDR) a été fixé à 1 %. Pour qu'une protéine soit identifiée avec succès, il fallait au moins deux peptides, dont au moins un était unique. Pour la quantification des protéines sans marquage (LFQ), l'algorithme MaxLFQ a été utilisé. L'analyse ultérieure du fichier résultant a été effectuée à l'aide du logiciel Perseus (version 1.6.5.0). Les occurrences de la base de données inversée, les protéines contenant uniquement des peptides modifiés et les contaminants potentiels ont été supprimés. Finalement, un diagramme de Venn a été construit pour illustrer les protéines communes aux trois types de tissus (PCC, ADK et sain), ainsi que celles exclusives à chaque groupe. Ensuite, une analyse statistique a été réalisée en utilisant un test ANOVA avec une valeur seuil "valeur p" de 1 %. Les protéines significativement différentes ont été choisies et standardisées à l'aide d'un z-score. Les résultats ont été représentés sur heatmap où les échantillons sont regroupés en non supervisé en fonction des clusters de protéines communes.

### *Identification des lipides*

Les spectres MS/MS ont été obtenus après avoir isolé l'ion précurseur avec une fenêtre de 1 u, puis dissociation induite par collision. Une énergie de collision entre 30 et 40 eV, a été appliquée en fonction de l'ion précurseur choisi. Les annotations des lipides ont été exécutées manuellement avec l'aide des bases de données publiques comme LipidMaps, Alex123 et Metlin ainsi qu'en se rapportant à la littérature pertinente.

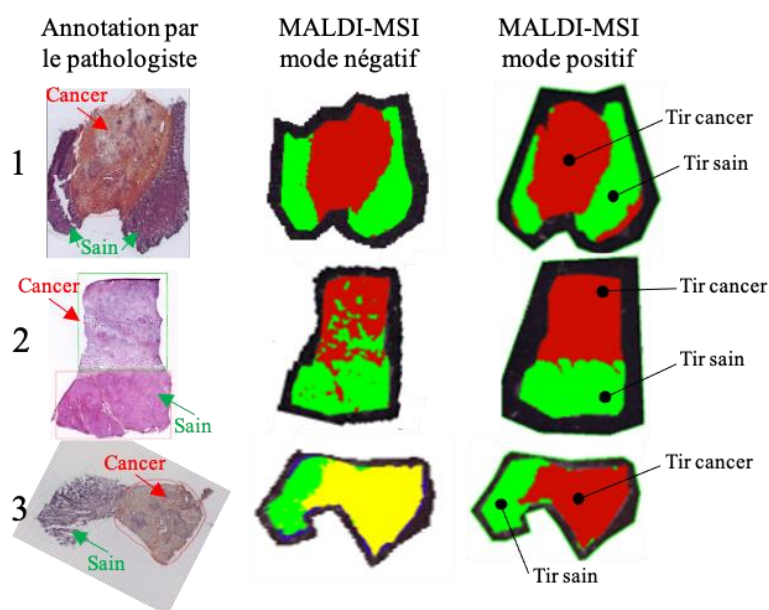


## Résultats

### Annotation moléculaire grâce à l'analyse MALDI-MSI

La solide similarité et la corrélation élevée entre les données d'imagerie MALDI et celles de l'imagerie WALDI (SpiderMass) ont été démontrées dans le premier chapitre de cette thèse. C'est pourquoi l'intégralité des échantillons des 176 patients atteint d'un cancer œsogastrique a été soumis à une analyse MALDI-MSI dans les deux modes d'ionisation. L'analyse par imagerie MALDI sert à plusieurs fins : elle permet d'utiliser l'analyse MALDI-MSI comme une annotation moléculaire préalable à l'analyse SpiderMass pour déterminer l'hétérogénéité moléculaire des tissus, comme une validation croisée des marqueurs découverts grâce au SpiderMass ainsi qu'en tant que données supplémentaires à celles obtenues avec le SpiderMass afin d'accroître la fiabilité des modèles de classification.

Tout d'abord, les résultats de l'analyse par MALDI-MSI ont été exportés sur le logiciel SCiLS. Les 13 échantillons mixtes ont été soumis à un algorithme de segmentation bisecting  $k$ -means en utilisant la distance corrélation. La segmentation spatiale regroupe tous les pixels présentant des spectres similaires en groupes colorés et extensibles appelés segments<sup>257</sup>. Tous les pixels correspondant à un même groupe sont alors représentés par la même couleur. La segmentation affiche généralement un profil moléculaire (plusieurs valeurs  $m/z$  et chaque  $m/z$  correspond à une molécule) associé à chaque grappe. La segmentation est représentée sous la forme d'un dendrogramme dont les branches sont calculées par les distances de corrélation entre chaque cluster. Plus la distance est petite, plus les profils moléculaires sont proches entre chaque segment.



**Figure 53** : Exemples de comparaison entre les annotations par le pathologiste et la segmentation spatiale obtenue dans les deux modes d'ionisation sur 3 tissus mixtes de cancer œsogastrique.

Comme présenté pour les 3 tissus de la **Figure 53**, la segmentation spatiale permet de distinguer les zones saines (cluster vert) des zones cancéreuses (cluster rouge/jaune) dans les tissus mixtes, que ce soit en mode positif ou en mode négatif. L'annotation moléculaire obtenue grâce à l'analyse MALDI-MSI est très similaire à l'annotation du pathologiste (voir tissu n°1) mais en étant encore plus précise, car basée sur une caractérisation moléculaire non ciblée. En effet, pour le tissu n°2, la marge entre la zone saine et la zone cancéreuse ne semble pas être aussi marquée que l'annotation de l'anatomopathologiste semble l'indiquer. Le cancer semble plutôt être infiltré dans un tissu toujours encore un peu sain, ainsi le tir par la technologie SpiderMass sera effectué en haut à droite pour être sûr d'obtenir un spectre caractéristique d'une zone cancéreuse et non d'une zone mixte. Tandis que pour le tissu n°3, la délimitation entre le tissu sain et tissu cancéreux semble être décalée à gauche.

Cela permet donc de mettre en évidence la possibilité d'utiliser l'analyse par imagerie MALDI en tant qu'annotation moléculaire de référence, plus précise que l'annotation histologique par le pathologiste. Celle-ci permet donc de nous guider pour l'analyse SpiderMass sur une lame consécutive dans l'objectif de l'établissement de modèle de classification pour lequel il est crucial de connaître avec certitude la classe à effectuer au spectre enregistré. En effet, la localisation des points de tirs est déterminée grâce à l'analyse par imagerie MALDI en validation des annotations subjectives du pathologiste.

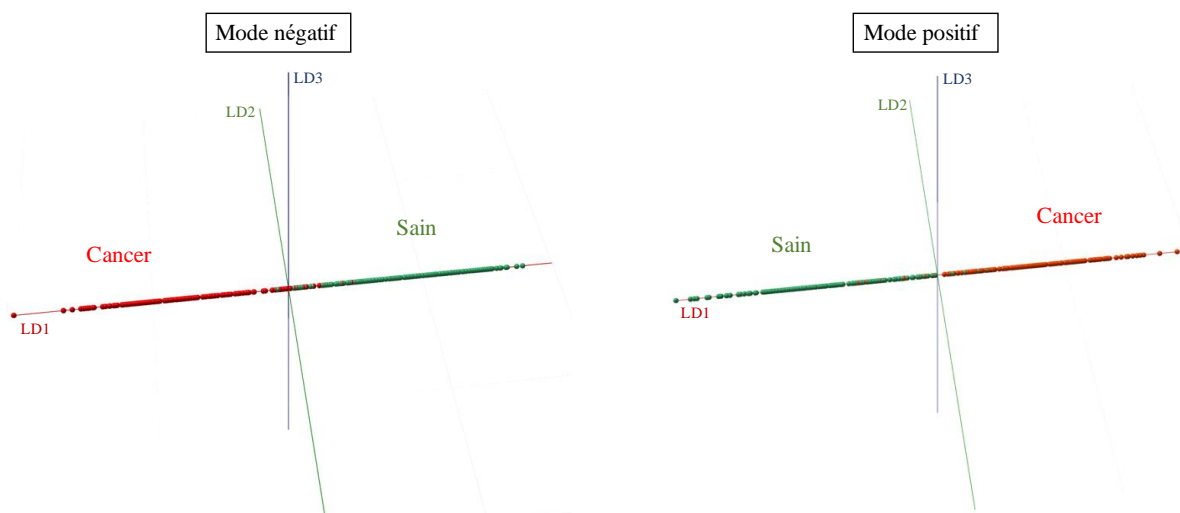
Évaluation du potentiel de diagnostic du cancer œsogastrique à l'aide de la technologie SpiderMass

#### *Modèle de classification via analyse LDA*

Concernant l'analyse SpiderMass, la cohorte de 176 échantillons de tissus, pour un total de 190 régions d'intérêt en raison de la présence de tissus mixtes, a été divisée en deux ensembles distincts : la cohorte d'entraînement et la cohorte de validation. La cohorte d'entraînement compte 151 régions d'intérêt. En parallèle, la cohorte de validation comprend, elle, 39 zones d'intérêt définies.

Les spectres générés, à partir de la cohorte d'entraînement, ont été soumis à différentes analyses statistiques multivariées, une analyse en composantes principales non supervisée puis à une analyse linéaire discriminante supervisée (LDA). Comme le montre la **Figure 54**, une séparation est visible, suivant la LD1, entre les tissus sains et les tissus cancéreux dans les deux modes d'ionisation. Cependant, il est à noter que la classification n'est pas nette et que certains spectres ne sont pas complètement séparés.

Pour vérifier la fiabilité du modèle de classification, le taux de bonne classification a été évalué via deux tests ; le test « 20 % out » où 20 % des spectres moyens sont exclus du modèle puis reclassés dans le modèle de classification créé sans ces données et le test « Full group out », similaire au test précédent mais ici chaque patient est exclu un à un. Ainsi, environ 81 % de ces spectres moyens sont correctement reclassés dans le bon groupe dans les deux modes d'ionisation pour le test « 20 % out » et le test « full group out » permet d'obtenir un taux de classement correct de plus de 70 % (**Tableau 9**). Il est d'ailleurs admis que le premier test surestime les performances du modèle alors que le deuxième a tendance à sous-estimer les performances lorsque certaines classes sont peu représentées.



**Figure 54** : Modèle de classification obtenu après analyse LDA afin de distinguer les tissus cancéreux des tissus sains en mode négatif et en mode positif.

**Tableau 9** : Évaluation des taux de bonne classification du modèle de classification cancer vs sain en mode négatif (orange) et en mode positif (bleu).

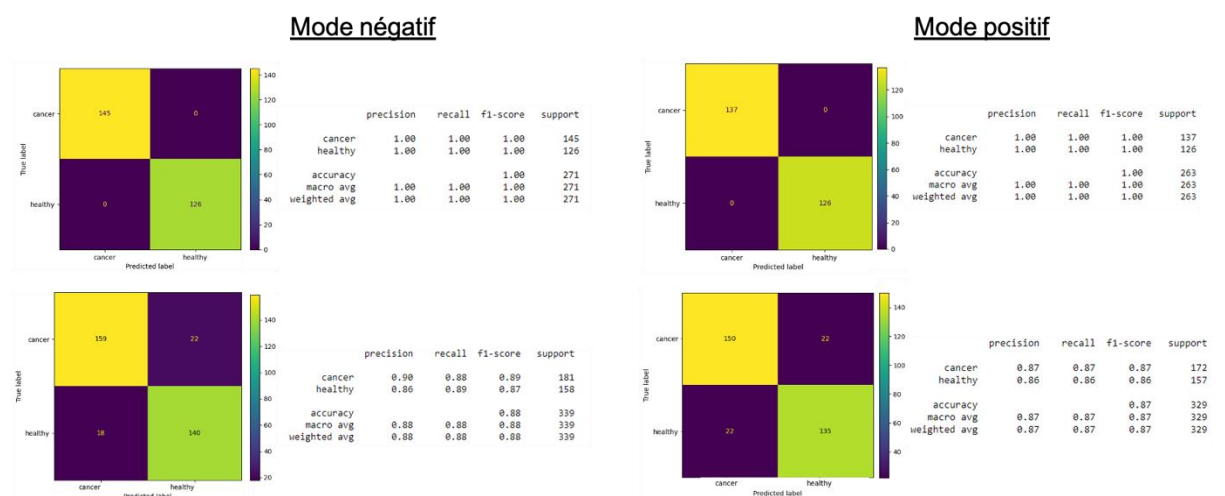
	Nombre de spectres	Nombre de passages	Nombre d'échecs	Nombre de valeurs aberrantes	Taux de classification correct
20 % out	338	279	58	1	82.79 %
Full group out	338	249	86	3	74.33 %
20 % out	329	263	65	1	80.18 %
Full group out	329	228	101	0	69.30 %

Historiquement, la méthode de diagnostic moléculaire par SpiderMass repose sur un modèle de type LDA au moyen du logiciel AMX (Waters) comme vu ci-dessus. Cependant, cette approche présente certaines limitations, notamment en se réduisant aux analyses statistiques

ACP/LDA pour la classification, en ne proposant pas d'évaluation ni de validation croisée des modèles, et en ne fournissant pas d'explications sur les prédictions ni de tests statistiques pour évaluer les variations des ions  $m/z$  selon le type de tissus. C'est pourquoi, comme montré dans le chapitre 2, nous avons développé un nouveau processus d'IA plus complet visant à créer des modèles de classification précis et robustes. Ce processus vise également à identifier des biomarqueurs avec certitude, en se focalisant non seulement sur la classification en elle-même, mais aussi sur la validation des classifications par le biais d'une compréhension biologique.

### *Optimisation du modèle de classification pour le diagnostic du cancer œsogastrique*

En utilisant la nouvelle méthodologie d'apprentissage, un modèle de classification a été élaboré en utilisant la cohorte d'entraînement, ceci visant à différencier de manière précise les échantillons de tissus cancéreux des échantillons sains dont l'attribution à leur classe respective a été obtenue par l'évaluation d'un pathologiste expert. L'algorithme RidgeClassifier a été utilisé pour obtenir le modèle de classification optimal. Les performances de ce modèle sur l'ensemble d'entraînement affichent une précision de 100 % pour les deux modes d'ionisation. Pour garantir la fiabilité et la généralisation des résultats, une procédure de validation croisée a été réalisée avec 20 itérations. Les résultats ont montré que la précision effective du modèle en utilisant le mode négatif atteint 88 %, avec un écart type de 0.09. Dans le cas du mode positif, le modèle présente une précision réelle d'environ 87 %, avec un écart-type de 0.07. En ce qui concerne les performances évaluées au moyen de la mesure F1, le modèle obtient des scores de 89 % et 88 % pour les échantillons sains et cancéreux respectivement, en utilisant le mode négatif. Pour le mode positif, les scores F1 sont respectivement de 86 % pour les échantillons sains et de 87 % pour les échantillons cancéreux (**Figure 55**).



**Figure 55** : Matrice de confusion et rapport de classification en mode négatif et positif pour le modèle de classification permettant de distinguer les tissus cancéreux et les tissus sains.

### *Prédiction automatisée en aveugle de la catégorie histologique en utilisant la classification préalablement élaborée*

Pour évaluer les performances du modèle de classification dans les deux modes d'ionisation, des prédictions à l'aveugle ont été réalisées indépendamment grâce à la cohorte de validation. Les 38 échantillons de tissus de la cohorte de validation sont totalement indépendants et n'ont pas été utilisés ni lors de la création ni lors de l'évaluation du modèle créé précédemment, qui a permis de différencier des sous-groupes de régions définies comme saines ou cancéreuses. Les régions d'intérêt ont été soumises à une analyse SpiderMass à l'aveugle puis annotées grâce au modèle de classification établi. Celle-ci a ensuite été comparée à l'annotation effectuée par le pathologiste.

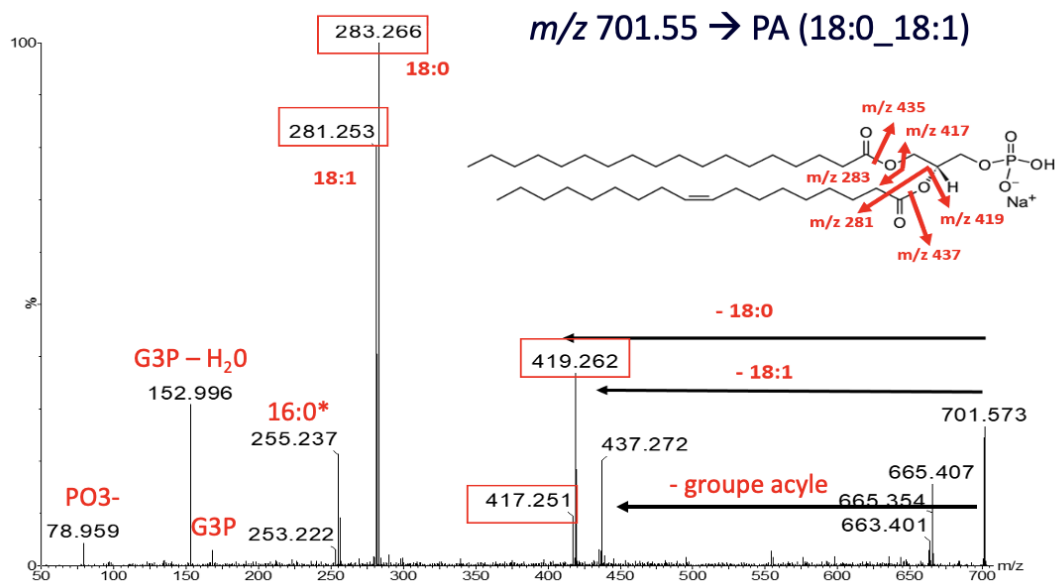
Lors de l'analyse en mode négatif, 7 régions sur 39 sont incorrectement classées, ce qui conduit à un taux de bonne classification de 82 %. En revanche, en utilisant le mode positif, seulement 6 échantillons sur 39 sont erronément classés, engendrant ainsi un taux de bonne classification de 85 %. Cependant, il est important de noter que dans certains cas, les erreurs de classification des échantillons peuvent être justifiées. À titre d'exemple, il y a une situation où un échantillon initialement identifié comme sain par l'hôpital a été ensuite annoté par le pathologiste en tant qu'échantillon cancéreux. La prédiction de SpiderMass, qui suggère que l'échantillon est sain, concorde avec les données de départ de l'hôpital, remettant ainsi en question l'annotation du pathologiste. D'autres situations concernent des échantillons mixtes, où certaines parties ont été identifiées comme saines et d'autres comme cancéreuses, comme l'a souligné le pathologiste. Dans de telles situations, les prédictions fournies par la technologie SpiderMass pourraient présenter des divergences en raison de la localisation de l'échantillonnage à proximité de la frontière entre ces deux zones distinctes. Ce genre de situation pourrait expliquer la variation entre les prédictions établies par SpiderMass et les annotations du pathologiste. Cela met en évidence la capacité de la technologie SpiderMass à être particulièrement performante dans la délimitation des marges de zones cancéreuses. En résumé, en ne tenant pas compte de ces cas ambigus, les prédictions en aveugle par SpiderMass permettent de correctement classer 32 échantillons sur 36 dans les deux modes d'ionisation, se traduisant par un remarquable taux de bonne classification de 90 %, ce qui est largement supérieur à ce que le pathologiste pourrait fournir en condition d'extemporanée.

### *Marqueurs lipidiques liés à la tumeur ou au tissu sain.*

De plus, nous nous sommes intéressés à l'identification des lipides associés au caractère cancéreux ou non d'un tissu œsogastrique, dans le but de mieux appréhender les voies biologiques sous-jacentes à leur discrimination. L'utilisation d'une méthodologie supervisée, tel que l'algorithme LIME, a permis de fournir une liste des contributions de chaque ion  $m/z$ , que ce soit

négativement ou positivement. En plus de cette méthodologie, indépendamment de la classification, les variations d'abondance des ions entre les différents types de tissu ont été explorées à travers une analyse statistique multivariée appliquée aux mêmes données. En mode négatif, sur les 179 ions  $m/z$ , présentant un rapport signal/bruit (S/N) > 10, 42 pics se sont révélés significatifs (avec une valeur  $p \leq 0.05$ ) tel que le met en évidence le test de Kruskal-Wallis. En mode positif, 34  $m/z$  ont été montrés significatifs sur les 273 signaux supérieurs au seuil S/N choisi.

Pour identifier les biomarqueurs lipidiques présentant un niveau de confiance élevé, nous avons croisé les données issues des analyses supervisées et non supervisées. Au total, nous avons identifié 60 biomarqueurs fiables : 32 en mode négatif et 28 en mode positif, ces biomarqueurs étant associés aux tissus tumoraux ou sains. Suite à l'annotation manuelle des biomarqueurs (**Figure 56**), il a été observé une prévalence marquée de certaines classes de lipides au sein de tissus particuliers.

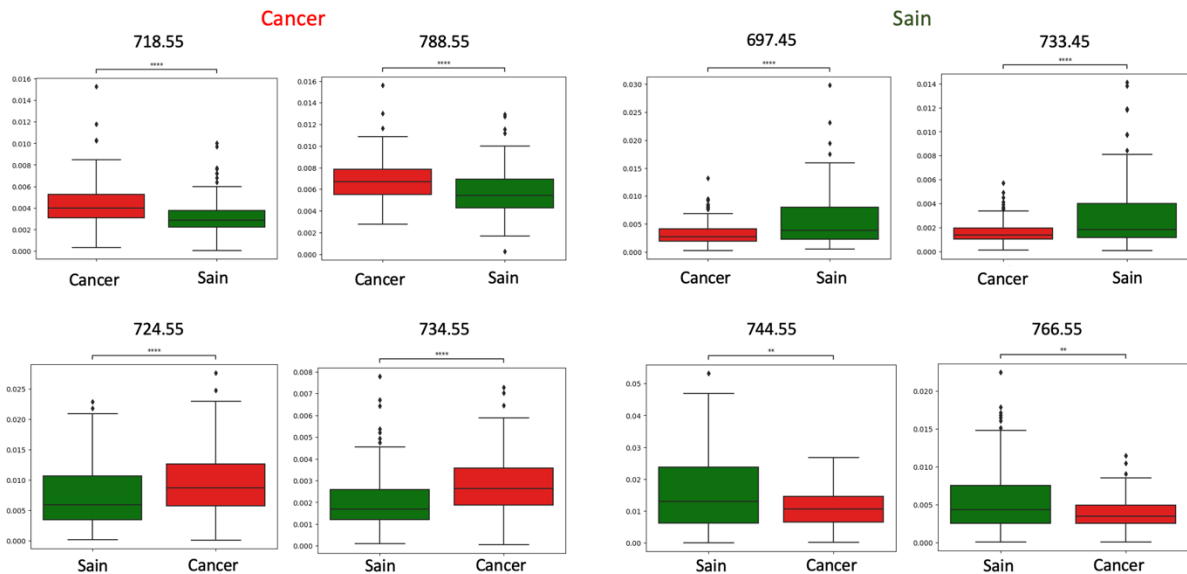


**Figure 56** : Spectre de fragmentation après analyse par spectrométrie de masse en tandem de l'ion  $m/z$  701.5.

En ce qui concerne le mode positif, parmi les 28 biomarqueurs fiables, 16 sont spécifiques des tissus cancéreux et 12 des tissus sains. Les PEs et les PCs sont surexprimés dans les tissus à caractère sain tandis que ce sont les DGs et les Cer qui sont en surexpression dans les tissus cancéreux. En effet, les ions  $m/z$  605.55 DG 36:1 et  $m/z$  712.65 Cer (t20:0\_24:0 (2OH)) sont spécifiques des tissus à caractère cancéreux. Au contraire des ions  $m/z$  724.55 PC 32:5,  $m/z$  792.55 PE 40:6 et  $m/z$  794.55 PE 40:5 qui sont surexprimés dans les tissus sains (**Figure 57B et Tableau 10**).

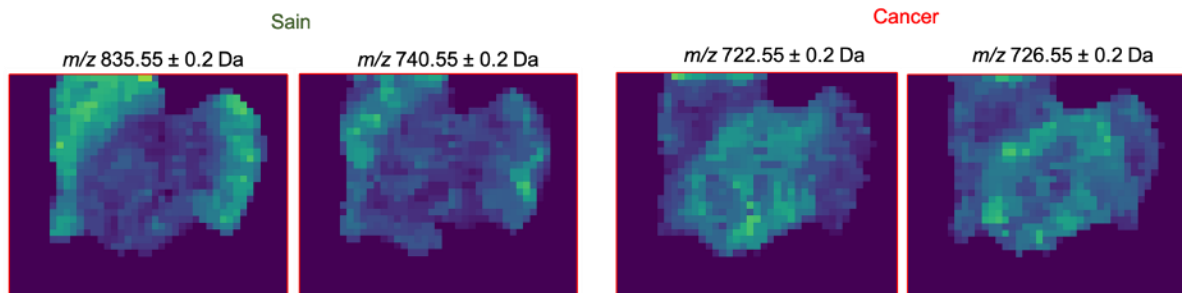
En mode négatif, parmi les 32 biomarqueurs fiables, 24 ions sont surexprimés dans les tissus tumoraux, principalement des phosphatidylsérines, des phosphatidyléthanolamines et des céramides. C'est notamment le cas des ions  $m/z$  698.55 Cer (d18:2\_24:0 (2OH)),  $m/z$  718.55 PE

34:0 et  $m/z$  788.55 PS (18:1\_18:0). De plus, les ions  $m/z$  819.55 PG (18:1\_22:6) et  $m/z$  846.65 HexCer (d18:1/24:0) spécifiques des tissus cancéreux ont été montrés, dans le chapitre précédent, comme des biomarqueurs fiables de la présence de macrophages de type M2. Pour les tissus sains, huit ions sont spécifiques, notamment certains phosphatidylinositols et acides phosphatidiques. Par exemple, l'ion  $m/z$  835.55 PI (18:1\_16:0) ainsi que le lipide PA (18:1\_18:2), celui-ci ayant été trouvé spécifique sous sa forme protonée négativement ( $m/z$  697.45) ainsi qu'avec un adduit chlorure ( $m/z$  733.45) (**Figure 57A et Tableau 10**).



**Figure 57** : Quatre biomarqueurs fiables et leur boîte à moustache correspondante spécifiques des tissus cancéreux et des tissus sains. A) en mode négatif. B) en mode positif.

De plus, il est aussi possible de confirmer ces biomarqueurs grâce à la localisation des ions dans un tissu mixte de cancer œsogastrique après imagerie par SpiderMass. En effet, comme le montre la **Figure 58**, la distribution des ions sélectionnés, de potentiels biomarqueurs des tissus sains et cancéreux, permet de mettre en évidence les zones saines et cancéreuses du tissu n°1 et ainsi confirmer la fiabilité des ions en tant que biomarqueurs.



**Figure 58** : Distribution des ions sélectionnés dans un tissu mixte de cancer œsogastrique, constitué d'une zone cancéreuse au milieu ainsi que de deux zones saines de chaque côté.



## Caractérisation des sous-types de cancer œsogastrique

L'objectif de notre étude ne se limitait pas uniquement à distinguer les tissus cancéreux des tissus sains, mais incluait également la capacité à différencier les différents sous-types de cancer œsogastriques, à savoir les sous-types PCC et ADK. Dans cette optique, nous avons utilisé deux approches : soit la création d'un modèle de classification basé sur 3 classes et permettant de discriminer les tissus sains, le tissu PCC et les tissus ADK, soit un modèle de classification à 2 classes visant uniquement à distinguer les tissus cancéreux entre les sous-types PCC et ADK.

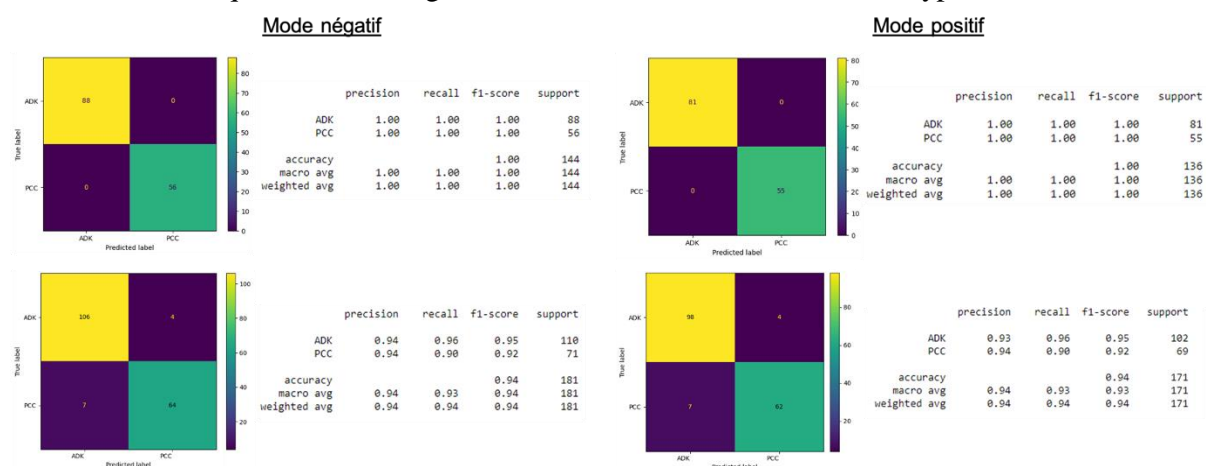
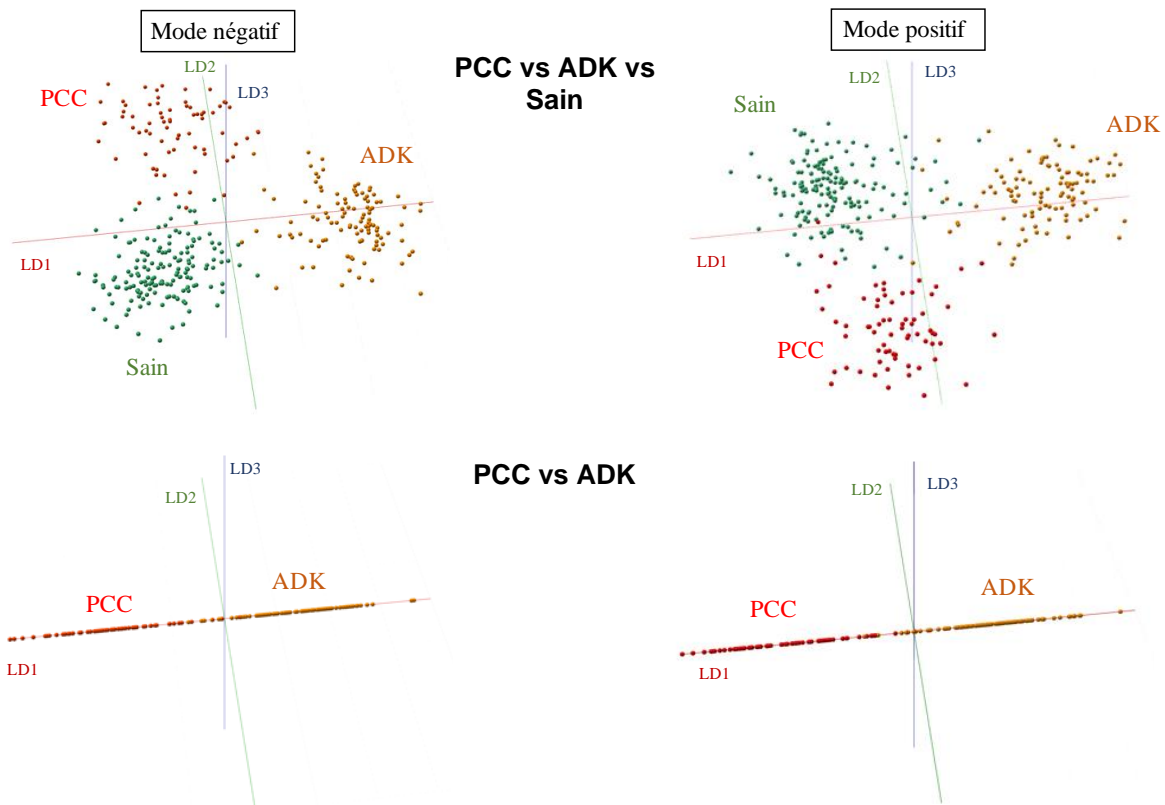


Figure 59 : Matrice de confusion et rapport de classification en mode négatif et positif pour le modèle de classification permettant de distinguer les sous-types de cancer PCC et ADK.

Le premier modèle atteint un taux de classification correcte de seulement 86 %, avec un écart-type de 0.07, et 83 %, avec un écart-type de 0.08, après une validation croisée à 20 reprises, respectivement en mode négatif et positif. Cependant, le deuxième modèle est beaucoup plus optimal, avec un taux de classification correcte de 94 %, et un écart-type de 0.08 dans les deux modes d'ionisation (Figure 59). Cela met en évidence une stratégie double qui semble être plus efficace, à savoir un modèle de classification en deux étapes qui détermine d'abord si le tissu est cancéreux ou sain avec une précision de 88 %, puis si un tissu est annoté cancéreux, le deuxième modèle identifie le sous-type avec une précision de 94 %. Ce résultat est intéressant car le sous-type PCC qui est très infiltrant est difficile à discriminer pour le pathologiste en conditions peropératoire, sur la simple base de coloration histologique et sans marquage IHC.

En comparaison, l'analyse par LDA, visant à distinguer les tissus sains, des tissus PCC et ADK, a uniquement permis d'obtenir 74.8 % et 78.5 % de bonne classification en mode positif et en mode négatif respectivement après le test « 20 % out » (Figure 60). Tandis que celui visant à distinguer uniquement les sous-types de cancer, PCC vs ADK, a uniquement permis d'obtenir plus de 87 % dans les deux modes d'ionisation (Figure 60).

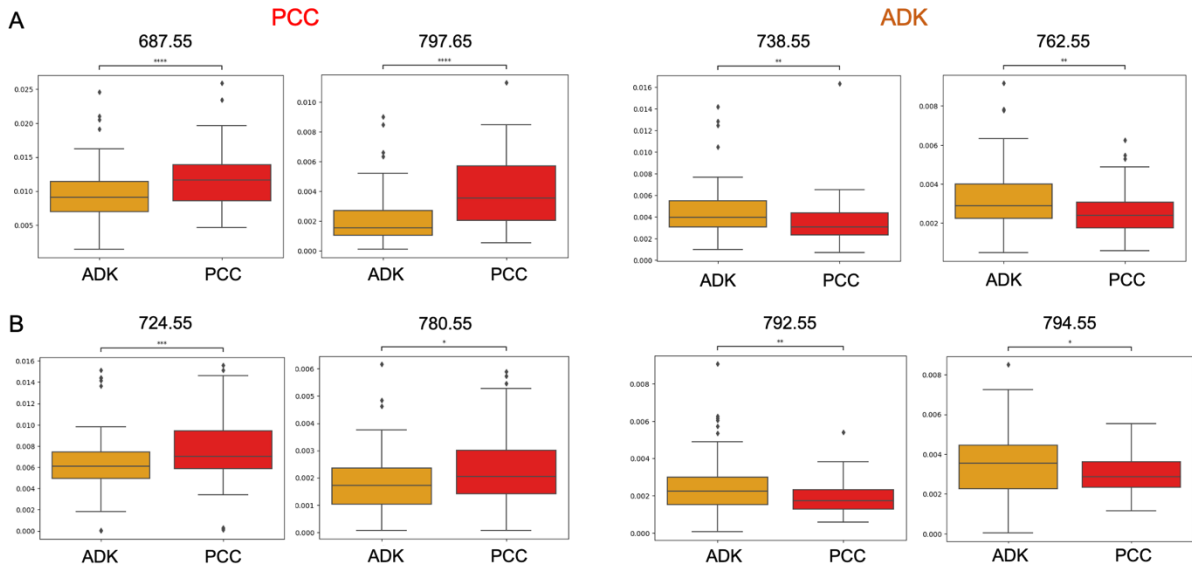


**Figure 60** : Modèles de classification obtenus après analyse LDA afin de distinguer les sous-types de cancer en mode négatif et en mode positif.

Ce qui est encore plus intéressant, c'est que le diagnostic à l'aveugle a permis de classer correctement 15 des 20 tissus inconnus, correspondant à 75 % de bonne classification. De plus, il a été possible d'identifier des marqueurs lipidiques fiables pour les deux sous-types cancéreux. En effet, en mode négatif, 19 biomarqueurs statistiquement significatifs ont été découverts ; 14 spécifiques des tissus ADK et 5 spécifiques des tissus PCC (**Figure 61A et Tableau 10**). Par exemple, les ions  $m/z$  671.45 PA 34:2,  $m/z$  673.45 PA 34:1 et  $m/z$  764.55 PE (18:1\_20:4) sont surexprimés dans les tissus cancéreux de sous-type ADK tandis que les ions  $m/z$  678.45 PE 32:6 et  $m/z$  687.55 PA O-36:1 sont plus présents dans les tissus PCC. De façon intéressante, certains biomarqueurs étaient déjà considérés comme spécifiques aux tissus sains dans le modèle initial. C'est le cas des ions  $m/z$  697.45 PA (18:1\_18:2),  $m/z$  764.55 PE (18:1\_20:4) et  $m/z$  835.55 PI (16:0\_18:1) préalablement trouvés spécifiques des tissus sains, mais ici sont découverts spécifiques du sous-type de cancer ADK. Il semblerait que les marqueurs moléculaires partagés avec les tissus sains soient présents dans les tissus du sous-type ADK, celui-ci étant un sous-type moins agressif et ayant une meilleure survie.

L'ion  $m/z$  819.55 spécifique des tissus cancéreux ainsi que des macrophages de type M2, est ici surexprimé dans les tissus de sous-type PCC. Ce résultat est particulièrement intéressant car les adénocarcinomes PCC correspondent à des formes particulièrement agressives qui apparaissent

particulièrement chez des patients plus jeunes et les macrophages de type M2 sont des macrophages anti-inflammatoires et pro-tumoraux contrairement aux macrophages de type M1. De plus, l'ion  $m/z$  738.55 PE (16:0\_20:4) spécifique des lymphocytes et l'ion  $m/z$  762.55 PE 38:6 spécifique des macrophages de type M1, sont trouvés ici surexprimés dans les tissus cancéreux de sous-type ADK.



**Figure 61** : Quatre biomarqueurs fiables et leur boîte à moustaches correspondante spécifiques des sous-types de cancer ADK et PCC. A) en mode négatif. B) en mode positif.

En mode positif, 18 biomarqueurs statistiquement fiables ont été décrits dont 10 sont spécifiques des tissus ADK et 8 spécifiques des tissus PCC. Par exemple, les ions, spécifiques du cancer,  $m/z$  724.55 PC 32:5 et  $m/z$  780.55 PC 36:5 sont surexprimés dans les tissus cancéreux de sous-type PCC. Le sous-type de cancer ADK semble caractérisé par la présence de céramides, tel que l'ion  $m/z$  650.65 Cer 42:1, et la présence de PEs, tels que les ions  $m/z$  702.55 PE O-34:2 et  $m/z$  810.65 PE O-42:4 (**Figure 61B** et **Tableau 10**).

**Tableau 10** : Liste exhaustive des biomarqueurs fiables des tissus sains et des tissus cancéreux ainsi que des sous-types de cancer PCC et ADK.

Mode négatif					
$m/z$	Cancer vs sain		PCC vs ADK		Possible annotation
	Cancer	Sain	PCC	ADK	
639.25					[PA (26:7)-H] <sup>-</sup>
671.45					[PA (34:2)-H] <sup>-</sup>
673.45					[PA (34:1)-H] <sup>-</sup>
678.45					[PE (32:6)-H] <sup>-</sup>
684.65					[Cer d42:1+Cl] <sup>-</sup>
687.55					[PA (O-36:1)-H] <sup>-</sup>
697.45					[PA (18:1_18:2)-H] <sup>-</sup>
698.55					[Cer (d18:2_24:0(2OH))+Cl] <sup>-</sup>

699.55					[PA (36:2)-H] <sup>-</sup>
700.55					[Cer (d18:1_24:0(2OH))+Cl] <sup>-</sup>
701.55					[PA (36:1)-H] <sup>-</sup>
714.55					[PE (16:0_18:2)-H] <sup>-</sup>
718.55					[PE (16:0_18:0)-H] <sup>-</sup>
720.45					[PE (32:2)+Cl] <sup>-</sup>
722.55					[PE (P-16:0_20:4)-H] <sup>-</sup>
726.55					[PE (P-18:0_18:2)-H] <sup>-</sup>
728.55					[PE (P-18:1_18:0)-H] <sup>-</sup>
733.45					[PA (36:3)+Cl] <sup>-</sup>
734.55					[PE (16:2_20:4)-H] <sup>-</sup>
736.55					[PE (36:5)-H] <sup>-</sup>
738.55					[PE (16:0_20:4)-H] <sup>-</sup>
740.55					[PE (36:3)-H] <sup>-</sup>
744.55					[PE (18:0_18:1)-H] <sup>-</sup>
748.55					[PS (O-34:0)-H] <sup>-</sup>
750.55					[PE (P-18:0_20:4)-H] <sup>-</sup>
762.55					[PE (38:6)-H] <sup>-</sup>
764.55					[PE (18:1_20:4)-H] <sup>-</sup>
772.55					[PE (18:1_20:0)-H] <sup>-</sup>
778.55					[PS (36:6)-H] <sup>-</sup>
788.55					[PS (18:0_18:1)-H] <sup>-</sup>
790.55					[PS (18:0_18:0)-H] <sup>-</sup>
794.55					[PE (40:4)-H] <sup>-</sup>
797.65					[TG (48:4)-H] <sup>-</sup>
819.55					[PG (18:1_22:6)-H] <sup>-</sup>
820.55					[PE (42:5)-H] <sup>-</sup>
835.55					[PI (18:1_16:0)-H] <sup>-</sup>
846.65					[HexCer (d18:1/24:0)-H] <sup>-</sup>
864.65					[PS (42:5)-H] <sup>-</sup>
883.55					[PI (18:1_20:4)-H] <sup>-</sup>
981.25					/
1006.65					[Hex2Cer (38:2)-H] <sup>-</sup>
<b>Mode positif</b>					
<i>m/z</i>	Cancer vs sain		PCC vs ADK		Possible annotation
	<b>Cancer</b>	<b>Sain</b>	<b>PCC</b>	<b>ADK</b>	
601.55					/
605.55					[DG (36:1)+H] <sup>+</sup>
627.55					[DG (O-36:8)+H] <sup>+</sup>

630.65					/
650.65					[Cer (42:1)+H] <sup>+</sup>
663.45					[PG (28:2)+H] <sup>+</sup>
666.45					[PS (O-38:0)+H] <sup>+</sup>
685.45					[PA (O-36:3)+H] <sup>+</sup>
689.55					[PA (36:8)+H] <sup>+</sup>
702.55					[PE (O-34:2)+H] <sup>+</sup>
703.55					[SM (d16:1_18:0)+H] <sup>+</sup>
710.65					[Cer (44:1)+H] <sup>+</sup>
712.65					[Cer (t20:0_24:0 (2OH))+H] <sup>+</sup>
716.55					[PE (18:1_16:1)+H] <sup>+</sup>
720.55					[PS (P-32:0)+H] <sup>+</sup>
723.45					[PA (38:5)+H] <sup>+</sup>
724.55					[PC (32:5)+H] <sup>+</sup>
728.55					[PE (O-36:3)+H] <sup>+</sup>
730.55					[PE (O-36:2)+H] <sup>+</sup>
734.55					[PE (36:7)+H] <sup>+</sup>
742.55					[PE (36:3)+H] <sup>+</sup>
744.55					[PE (36:2)+H] <sup>+</sup>
750.55					[PE (O-38:6)+H] <sup>+</sup>
758.55					[PE (34:2)+H] <sup>+</sup>
766.55					[PC (O-36:5)+H] <sup>+</sup>
774.55					[PS (O-36:2)+H] <sup>+</sup>
780.55					[PC (22:4_14:1)+H] <sup>+</sup>
784.55					[PC (36:3)+H] <sup>+</sup>
788.65					[PC (36:1)+H] <sup>+</sup>
792.55					[PE (40:6)+H] <sup>+</sup>
794.55					[PE (40:5)+H] <sup>+</sup>
805.55					[PA (44:6)+H] <sup>+</sup>
808.55					[PC (38:5)+H] <sup>+</sup>
810.65					[PE (O-42:4)+H] <sup>+</sup>
812.55					[PC (38:3)+H] <sup>+</sup>
814.55					[PE (42:9)+H] <sup>+</sup>
855.55					[PI (36:6)+H] <sup>+</sup>
875.55					[TG (54:8)+H] <sup>+</sup>
907.35					[TG (56:6)+H] <sup>+</sup>

## Identification de la signature SpiderMass de différentes souches bactériennes : bactérioscore du cancer œsogastrique

Le microbiote humain est composé d'une grande quantité de bactéries. Dans des conditions normales, il joue un rôle essentiel dans le maintien de la santé humaine, en contribuant au métabolisme de l'énergie, à l'absorption des nutriments, à la maturation du système immunitaire intestinal et à la protection contre les infections par des agents pathogènes<sup>258</sup>. Les perturbations du microbiote peuvent potentiellement être associées au développement du cancer. Par exemple, des types de bactéries susceptibles de favoriser le cancer, tels que *Fusobacteria* et *Escherichia coli*, sont plus abondants dans les cas de cancer colorectal<sup>259</sup>.

Cela a aussi été démontré dans le cas du cancer œsogastrique. En effet, de récentes avancées technologiques ont permis de caractériser le microbiote gastrique. Ces résultats indiquent que des bactéries autres que *H. pylori* jouent également un rôle important dans l'apparition et le développement du cancer gastrique. Avec le développement de la biologie moléculaire et de la métagénomique, les scientifiques ont une compréhension plus complète des microbes gastro-intestinaux, et il est maintenant admis que la dysbiose microbienne peut favoriser l'apparition du cancer gastrique par différents mécanismes<sup>260</sup>. De plus, des études ont montré le rôle majeur du microbiote non-*H. pylori* dans le développement du cancer gastrique chez les souris<sup>261</sup>. Chez les humains, l'infection chronique par *H. pylori* ou l'utilisation de médicaments comme l'oméprazole ont perturbé l'équilibre microbien de la muqueuse gastrique, favorisant la prolifération d'autres bactéries<sup>262</sup>. Cela a entraîné une augmentation des espèces nitrosantes, augmentant les niveaux de nitrite et de N-nitrosamine dans l'estomac, suggérant le rôle potentiel de microbiotes autres que *H. pylori* dans la carcinogénèse gastrique<sup>263</sup>. Cependant, notre compréhension de la flore gastrique chez l'homme reste limitée. Des études ont identifié des différences dans la composition et la fonction du microbiote gastrique entre les patients atteints de cancer gastrique et les groupes de contrôle, mais il n'y a pas de consensus sur les taxons microbiens spécifiques impliqués dans la pathogénèse du cancer gastrique. De plus, les modifications microbiennes dans le microenvironnement tumoral demeurent peu claires.

De plus, pour l'instant, ces différences de composition du microbiote gastrique, chez les patients atteints de cancer gastrique et chez les patients témoins, sont analysés principalement grâce à la génétique. En outre, le microbiote gastrique, qui est largement sous-étudié par rapport à son homologue intestinal, est examiné dans des cohortes mixtes de patients atteints de cancer gastrique sans tenir compte des différents sous-types<sup>264</sup>.

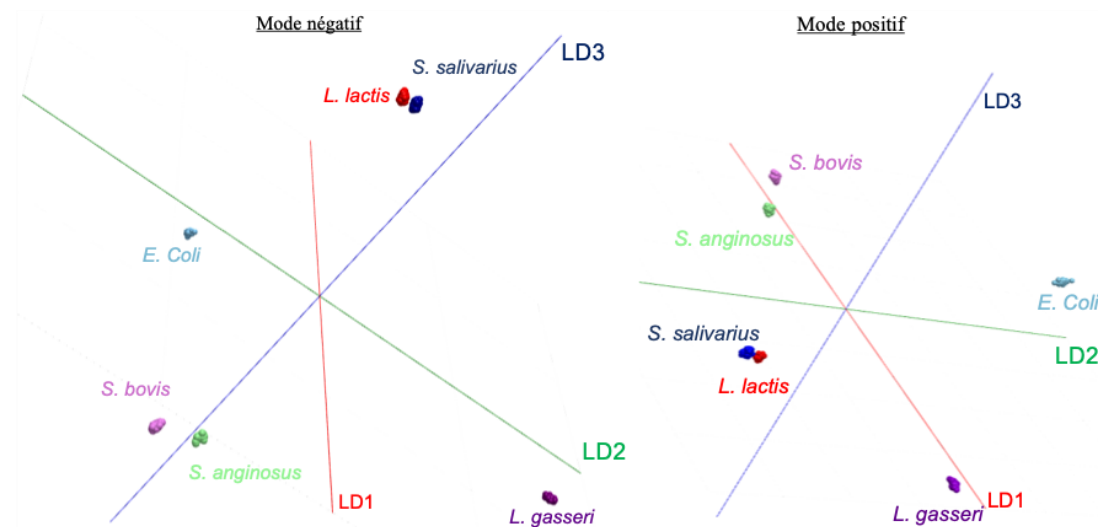
Dans le cas d'adénocarcinomes gastriques, les pathologistes sont exclusivement concentrés sur la détection de la présence de la bactérie *Helicobacter pylori*. Cette identification est

couramment effectuée par le biais de techniques telles que la coloration de Giemsa Lent ou l'immunohistochimie, utilisant des anticorps anti-*Helicobacter pylori*, à partir de sections de blocs FFPE uniquement. Néanmoins, cette donnée n'est disponible qu'après la chirurgie et ne prend pas en compte d'autres bactéries qui pourraient également jouer un rôle ou être fortement exprimées dans le cas du cancer œsogastrique.

### *Classification SpiderMass des profils lipidiques des souches bactériennes*

Tout d'abord, nous avons utilisé la technologie SpiderMass pour analyser diverses souches bactériennes séparément, incluant *Escherichia coli* (*E. coli*), *Lactococcus lactis* (*L. lactis*), *Lactobacillus gasseri* (*L. gasseri*), *Streptococcus anginosus* (*S. anginosus*), *Streptococcus salivarius* (*S. salivarius*) et *Streptococcus bovis* (*S. bovis*), afin d'identifier des lipides spécifiques à chacune de ces souches bactériennes.

Chaque souche a été directement analysée à partir de boîtes de Petri dans lesquelles elles ont été cultivées et après une journée d'incubation. Comme le montre la **Figure 62**, les souches bactériennes sont très bien différenciées grâce à un modèle de classification utilisant l'algorithme LDA en mode positif et en mode négatif.



**Figure 62** : Modèle de classification obtenu après analyse LDA afin de distinguer les différentes souches bactériennes en mode négatif et en mode positif.

Lors de l'utilisation du nouveau pipeline développé dans le chapitre 3, un modèle de classification a été élaboré pour différencier chacune des souches bactériennes, et celui-ci a atteint un taux de classification correcte de 100 % grâce à une validation croisée effectuée 20 fois dans les deux modes d'ionisation. Les méthodes de découverte de biomarqueurs fiables, qu'elles soient supervisées ou non supervisées, ont permis d'identifier 10 marqueurs lipidiques spécifiques de la souche *E. coli*, 5 marqueurs lipidiques spécifiques de la souche *S. anginosus*, 13 marqueurs lipidiques spécifiques de la souche *S. bovis*, 7 marqueurs lipidiques spécifiques de la souche *L.*



*lactis*, 8 marqueurs lipidiques spécifiques de la souche *S. salivarius* et 4 marqueurs lipidiques spécifiques de la souche *L. gasseri* (**Tableau 11**)

**Tableau 11** : Liste exhaustive des biomarqueurs fiables des différentes souches bactériennes en mode positif et en mode négatif.

Mode négatif							
<i>m/z</i>	<i>E. Coli</i>	<i>S. anginosus</i>	<i>S. bovis</i>	<i>L. lactis</i>	<i>S. salivarius</i>	<i>L. gasseri</i>	Annotation
660.45							[PE (30:1)-H] <sup>-</sup>
663.45							[PG (28:1)-H] <sup>-</sup>
665.45							[PE (28:0)-H] <sup>-</sup>
665.65							/
674.45							[PE (31:1)-H] <sup>-</sup>
681.45							[PA (35:4)-H] <sup>-</sup>
688.45							[PE (32:1)-H] <sup>-</sup>
702.55							[PE (33:1)-H] <sup>-</sup>
705.45							[PG (31:1)-H] <sup>-</sup>
707.45							[PA (37:5)-H] <sup>-</sup>
716.55							[PE (34:1)-H] <sup>-</sup>
733.55							[PG (33:1)-H] <sup>-</sup>
735.45							[PA (39:5)-H] <sup>-</sup>
737.55							[PG (34:6)-H] <sup>-</sup>
745.55							[PG (34:2)-H] <sup>-</sup>
747.55							[PG (34:1)-H] <sup>-</sup>
760.55							[PS (34:1)-H] <sup>-</sup>
763.55							[PG (35:0)-H] <sup>-</sup>
773.55							[PG (36:2)-H] <sup>-</sup>
789.55							[PG (37:1)-H] <sup>-</sup>
801.55							[PG (38:2)-H] <sup>-</sup>
843.45							[PI (35:4)-H] <sup>-</sup>
845.65							[PI (35:3)-H] <sup>-</sup>
869.55							[PI (37:5)-H] <sup>-</sup>
871.55							[PI (37:4)-H] <sup>-</sup>
899.55							[PI (39:4)-H] <sup>-</sup>
953.65							[PI (43:5)-H] <sup>-</sup>
996.55							/
1007.65							/
Mode positif							
<i>m/z</i>	<i>E. Coli</i>	<i>S. anginosus</i>	<i>S. bovis</i>	<i>L. lactis</i>	<i>S. salivarius</i>	<i>L. gasseri</i>	Annotation
615.45							/

639.45							[PG (36:0)+H] <sup>+</sup>
671.55							/
672.55							/
676.45							[PE (31:1)+H] <sup>+</sup>
697.45							[PA (36:4)+H] <sup>+</sup>
721.65							/
723.45							[PA (38:5)+H] <sup>+</sup>
749.55							[PG (34:1)+H] <sup>+</sup>
777.55							[PG (36:1)+H] <sup>+</sup>
779.55							[PG (36:0)+H] <sup>+</sup>
793.55							[PG (38:7)+H] <sup>+</sup>
803.45							/
829.45							[PI (34:5)+H] <sup>+</sup>
831.45							[PI (34:4)+H] <sup>+</sup>
859.55							[PI (36:4)+H] <sup>+</sup>
927.55							[PI (41:5)+H] <sup>+</sup>
995.65							/

À titre d'exemple, plusieurs phosphatidyléthanolamines s'avèrent spécifiques de la souche *E. coli*, notamment les ions  $m/z$  702.55 (PE 33:1) et  $m/z$  688.45 (PE 32:1). Le lipide PE 31:1 est particulièrement spécifique en raison de sa spécificité marquée pour *E. coli* dans les deux modes d'ionisation ( $m/z$  674.45 et 676.45). En ce qui concerne les souches *S. anginosus* et *S. bovis*, la majorité des lipides identifiés sont des PAs et des PIs, tels que le PA 36:4 ( $m/z$  697.45) et le PI 39:4 ( $m/z$  899.55) spécifiques de *S. anginosus*, ainsi que le PA 35:4 ( $m/z$  681.45) et le PI 35:4 ( $m/z$  843.45) spécifiques de *S. bovis*. Les PGs, eux, s'avèrent plus spécifiques des souches *L. lactis* et *L. gasseri*, comme en témoignent respectivement le PG 38:2 ( $m/z$  801.55) et le PG 34:2 ( $m/z$  745.55). Enfin, les marqueurs lipidiques spécifiques pour la souche bactérienne *S. salivarius* sont plus diversifiés, comprenant des PGs, des PIs et des PAs, par exemple le PG 37:1 ( $m/z$  789.55) et le PA 38:5 ( $m/z$  723.45). Il est particulièrement remarquable de noter qu'un ion  $m/z$  869.55, correspondant à un PI 37:5, spécifique de *S. bovis*, présente également une spécificité pour le sous-type cancéreux PCC (**Figure 63**). Par ailleurs, comme attendu pour les lipides des espèces bactériennes, un certain nombre de ces lipides présentent des chaînes à nombre impair de carbone, ce qui n'est pas habituel pour les lipides physiologiques chez l'homme<sup>265</sup>.

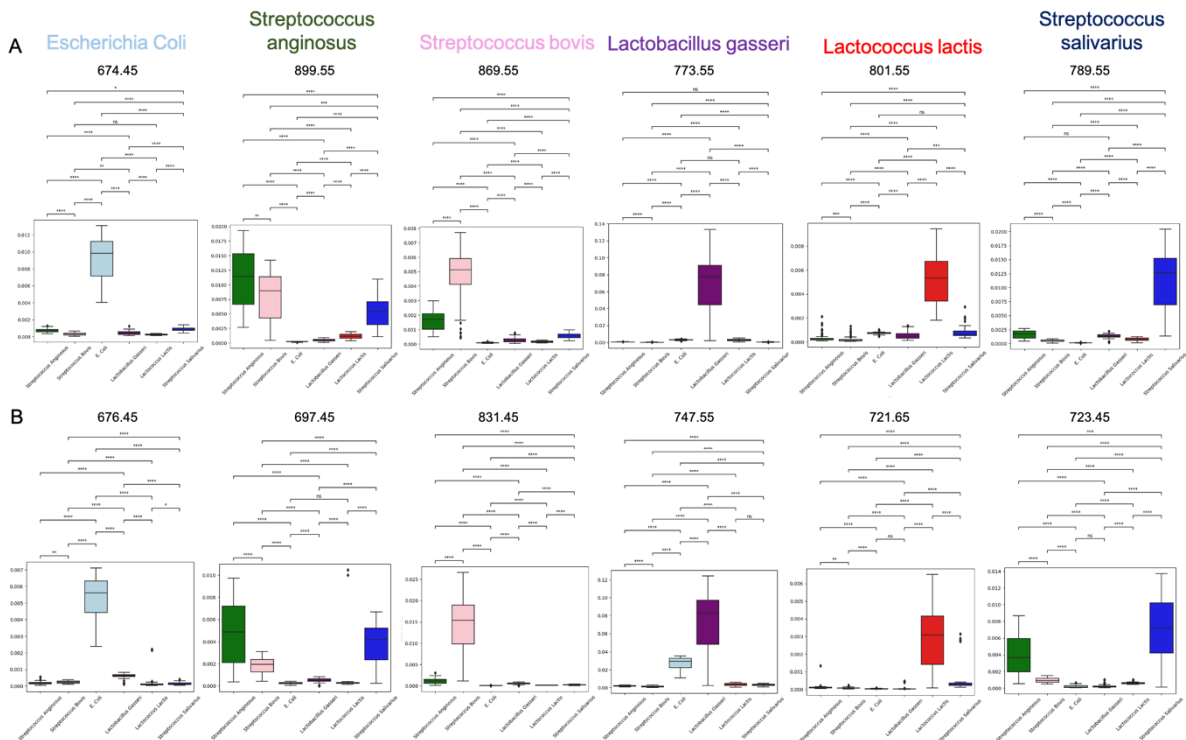


Figure 63 : Exemple d'un biomarqueur fiable pour chaque souche bactérienne avec la boîte à moustaches correspondante en mode négatif (A) et mode positif (B).

*Distribution des scores bactériens sur les images SpiderMass et corrélation avec le type de tissu*

Dans cette étude, nous avons exploré la possibilité de développer un indice bactérien basé sur les données générées par SpiderMass, avec la perspective de futurs diagnostics *in vivo*. Pour développer le modèle de score bactérien, 75 spectres MS en mode négatif ont été utilisés pour chaque souche bactérienne, en exploitant la bibliothèque Python nommée LGBM. Ce modèle permet d'obtenir un taux de classification correcte de 99 % lors de l'entraînement et de la validation croisée. De fait, à partir d'une image SpiderMass, il est possible de prédire la probabilité de présence et de répartition des diverses souches bactériennes dans le tissu. Les résultats obtenus ont permis d'obtenir des scores estimés pour chaque souche bactérienne, et par la suite, des scores de ratio ont été calculés pour déterminer le score relatif de probabilité de présence de chaque souche de bactérie dans l'intégralité de l'image. Ces ratios ont été calculés en additionnant les scores associés à chaque type de bactérie, puis en divisant la somme par les scores totaux de toutes les souches cumulées. Ces ratios offrent un aperçu de la répartition de chaque souche bactérienne sur l'image, ce qui permet une évaluation globale de la population bactérienne dans les tissus œsogastriques, qu'ils soient sains ou cancéreux.

Douze tissus frais congelés ont été analysés par imagerie SpiderMass, comprenant quatre tissus sains et huit tissus cancéreux, dont quatre de chaque sous-type (PCC et ADK). Grâce à cette

nouvelle approche, nous constatons que le ratio prédit de certaines souches bactériennes ne diffère pas de manière significative en fonction de la nature saine ou cancéreuse du tissu. C'est notamment le cas pour *E. Coli*, *S. anginosus* et *S. salivarius*, avec une moyenne de présence avoisinant respectivement 65 %, 6.4 % et 4.5 % pour les tissus sains et tumoraux.

Cependant, la souche *L. gasseri* est, quant à elle, prédite comme étant plus abondante dans les tissus sains (7 %) que dans les deux sous-types cancéreux (2.7 %), tout particulièrement dans les tissus PCC (0.1 %). À l'inverse, le ratio prédit pour la souche *L. lactis* montre une tendance opposée, avec une abondance supérieure dans les tissus cancéreux (12.4 %) par rapport aux tissus sains (9.8 %). En ce qui concerne le ratio prédit pour *S. bovis*, il révèle une association intrigante entre cette population bactérienne et le cancer œsogastrique. Plus précisément, une proportion moindre de *S. bovis* est observée dans les tissus sains (6 %) par rapport aux tissus cancéreux (8 %), avec une légère variation entre les sous-types de cancer où une abondance plus élevée est constatée dans les tissus PCC (8.5 %) par rapport aux tissus ADK (7.5 %) (**Figure 65**).

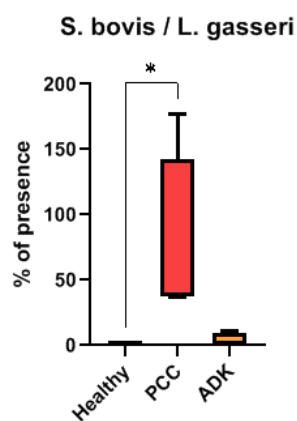


Figure 64 : Box plot du rapport *S. bovis* et *L. gasseri* pour les différents types de tissus.

Le rapport entre *S. bovis* et *L. gasseri* pourrait ainsi servir de marqueur de diagnostic pour les tissus œsogastriques. En effet, lorsque la proportion de *L. gasseri* excède celle de *S. bovis* (rapport < 1), cela indique un tissu sain. À l'inverse, un rapport supérieur à 1 entre *S. bovis* et *L. gasseri* permet de diagnostiquer un tissu cancéreux. De manière plus spécifique, un rapport très élevé (> 40) pourrait même prédire la présence du sous-type PCC. La possibilité d'utiliser l'analyse du microbiote en tant que marqueur prédictif pour les patients atteints de cancer gastrique semble une idée prometteuse<sup>266</sup> (**Figure 64**).

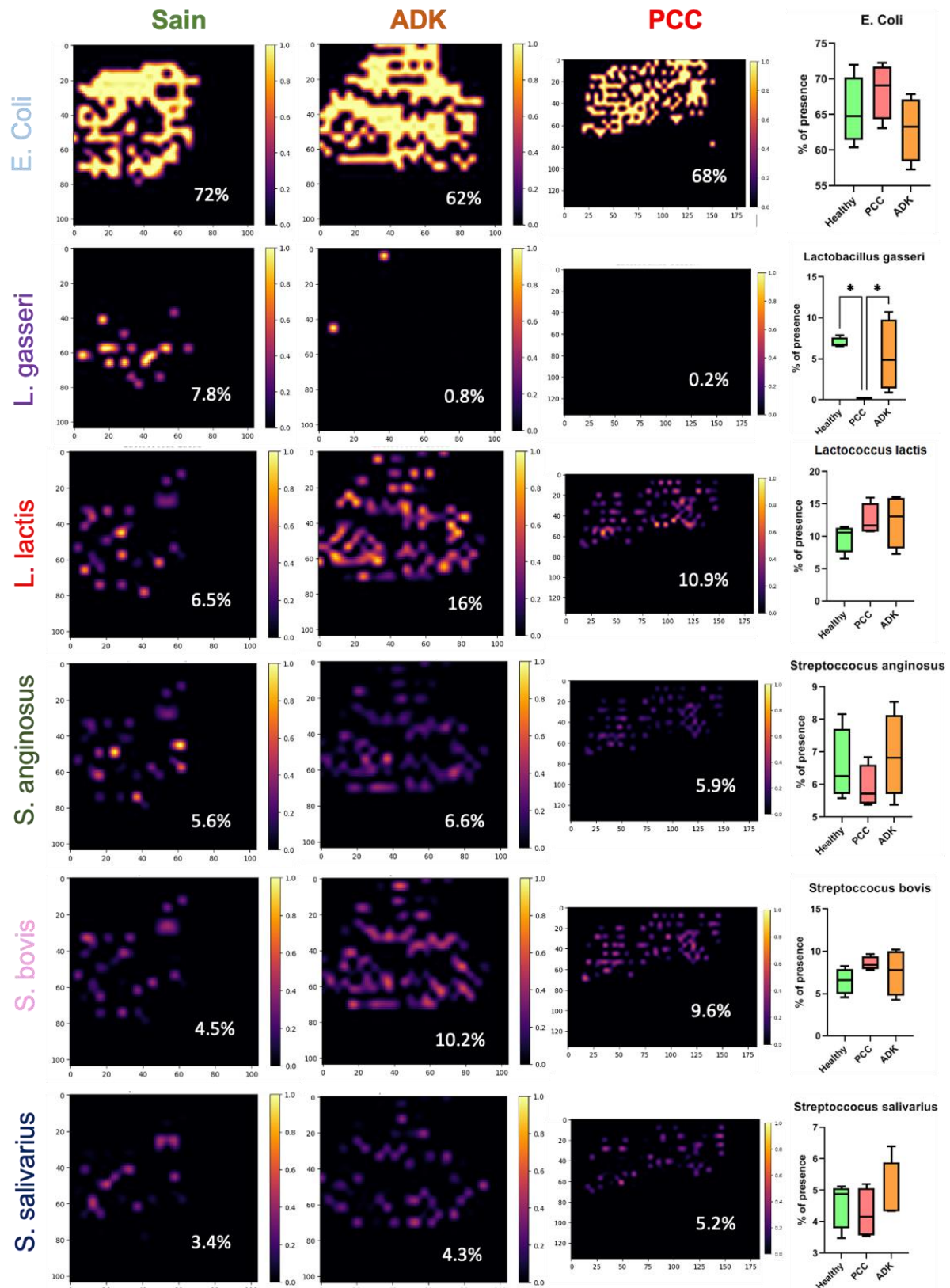


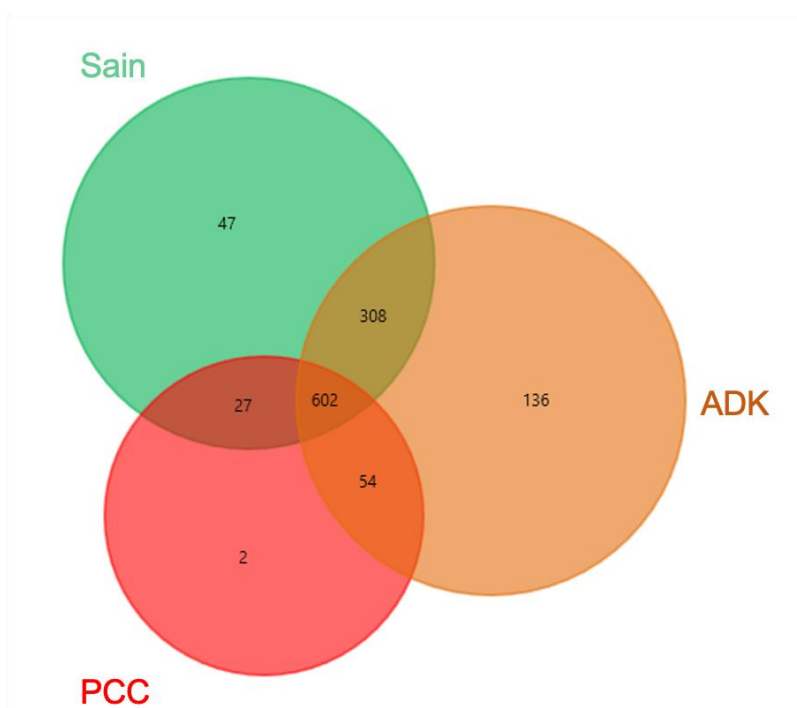
Figure 65 : Distribution des bactérioscopes sur 3 exemples de tissus de cancer oesogastrique pour les 6 souches bactériennes. Pour chaque souche bactérienne, le RPR en pourcentage est indiqué. Le pourcentage moyen des 6 souches bactériennes est représenté dans les tissus sains et cancéreux, ADK et PCC.

## Analyse protéomique spatiale

Une analyse protéomique spatiale via la technologie SpiderMass a permis de récupérer une zone de 350 µm de diamètre par zone d'intérêt sur 30 tissus puis de réaliser une analyse protéomique de type shot-gun après extraction des protéines et digestion enzymatique directement en milieu liquide. Les peptides tryptiques générés pour tous les échantillons ont ensuite analysés par LC-MS/MS en vue de leurs identifications et de leur quantification relative. Après interrogation des banques de données, un total de 2766 protéines a été identifié.

### *Identification des protéines exclusives de chaque type de tissu*

Tout d'abord, nous nous sommes focalisés sur les protéines provenant des différents types de tissus (sain, PCC et ADK). En effet, grâce à l'analyse des données par un diagramme de Venn, il est possible de montrer que les tissus sains présentent 47 protéines exclusives, les tissus cancéreux du sous-type ADK présentent 136 protéines exclusives tandis que les tissus cancéreux de sous-type PCC présentent uniquement 2 protéines exclusives (**Figure 66**).



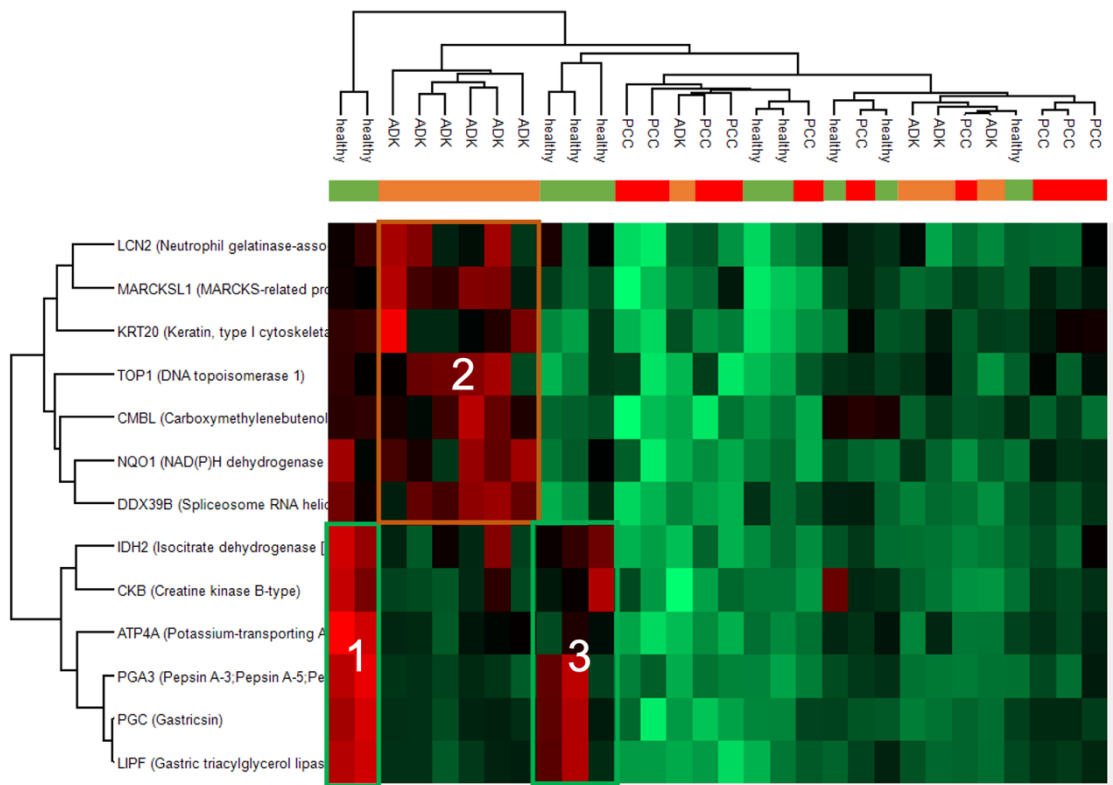
**Figure 66** : Diagramme de Venn mettant en avant le nombre de protéines exclusives pour chaque type de tissu, sains, cancéreux de type PCC et cancéreux de type ADK.

Nous nous sommes d'abord intéressés à ces protéines exclusives, et en particulier celles exclusives des tissus de sous-type cancéreux PCC. Les deux protéines exclusives retrouvées correspondent à la galectine-10 et à la latexine. De façon intéressante, la galectine-10, bien que relativement peu étudiée dans le contexte des cancers, a récemment été associée au cancer colorectal et au cancer gastrique<sup>267</sup>. Dans le cas du cancer gastrique, une forte expression de la galectine-10 a été identifiée comme un marqueur de pronostic défavorable<sup>268</sup>. A l'inverse, dans les

cancers gastriques humains, l'expression de la latexine est diminuée par rapport aux tissus normaux de contrôle. Cela indique que la latexine a un effet inhibiteur sur la croissance et la capacité à former des tumeurs des cellules cancéreuses gastriques humaines<sup>269</sup>. Dans notre étude de protéomique spatiale, nous retrouvons que la latexine est surexprimée dans les tissus cancéreux mais uniquement de sous-type PCC. Ceci démontre une différence indéniable entre les deux sous-types, ADK et PCC, le rôle de la latexine n'ayant jamais été étudiée dans des tissus cancéreux de sous-type PCC. Il serait très intéressant d'étudier plus profondément le rôle de ces deux protéines dans le modèle PCC afin d'envisager leur utilisation en tant que biomarqueur pour le diagnostic ou en tant que cible thérapeutique du sous-type PCC.

*Identification des voies de signalisation spécifiques de chaque type de tissu*

L'étape suivante était d'étudier la régulation des protéines communes. Pour cela, un test ANOVA a été effectué en regroupant les données en fonction des groupes d'annotations histologiques ADK, PCC, et Healthy (sain). Au total, seules 13 protéines ont montré une différence d'expression significative entre les trois groupes, ces protéines ont été représentées sous la forme d'un HeatMap, où les protéines surexprimées sont en rouge et les sous-exprimées en vert (**Figure 67**).



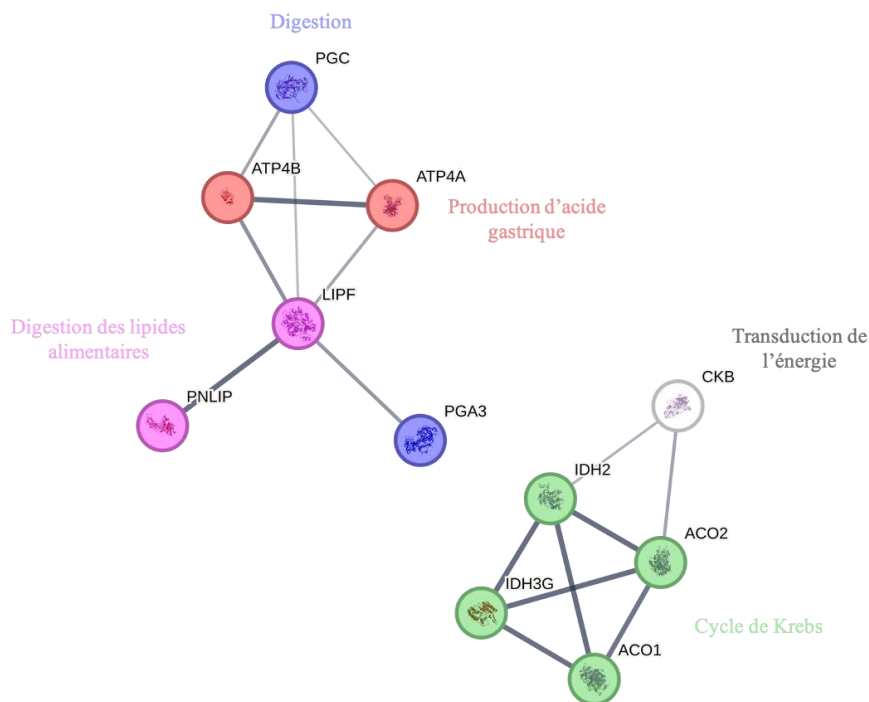
**Figure 67** : Analyse statistique ANOVA des protéines identifiées par irradiation backside correspondant aux différents types de tissus (sains, PCC et ADK).



Le heatmap généré révèle des clusters de protéines avec des niveaux d'expression élevés ou faibles dans trois groupes uniquement. En conséquence, deux ensembles de protéines se distinguent clairement. Le cluster 2, est composé de protéines qui sont surexprimées dans la branche constituée de tissus cancéreux du sous-type ADK. En revanche, les clusters 1 et 3 regroupent des protéines dont l'expression est significativement augmentée dans la branche composée de tissus sains. L'analyse de ces protéines permet finalement d'identifier les processus biologiques spécifiques associés à chaque groupe.

Les protéines différentiellement régulées ont ensuite été analysées sous String<sup>270</sup> afin d'identifier les réseaux spécifiques dans lesquelles ces protéines sont impliquées. String est une base de données qui répertorie les interactions entre les protéines, qu'elles soient déjà connues ou prédites, et comprend des interactions directes (physiques) mais également indirectes (fonctionnelles). Cette analyse aide à simplifier la compréhension des données et fournit un aperçu des voies de signalisation qui sont impliquées dans chaque cluster.

On remarque alors que les principales voies de signalisations mises en évidence dans les tissus sains (clusters 1 et 3) sont associées aux fonctions primitives de l'estomac (**Figure 68**). En effet, plusieurs protéines de ce cluster, majoritairement des protéases aspartiques, sont impliquées dans la digestion, protéines telles que la gastricsine (PGC) et la pepsine A3 (PGA3), ainsi que dans celle des lipides alimentaires, comme la lipase F (LIPF).



**Figure 68** : Voies de signalisation surexprimées dans le cluster 1, représentant les tissus sains. L'épaisseur des traits montre le degré de confiance pour le rapport entre les deux protéines.

De plus, l'une des voies de signalisation mises en évidence dans ces clusters est liée à la production d'acide gastrique, mettant en jeu la protéine ATP4A, ainsi qu'à la transduction de l'énergie (CKB). Celles-ci facilitent la digestion des aliments, l'absorption des minéraux et la lutte contre les bactéries nuisibles, telle que *H. pylori*.

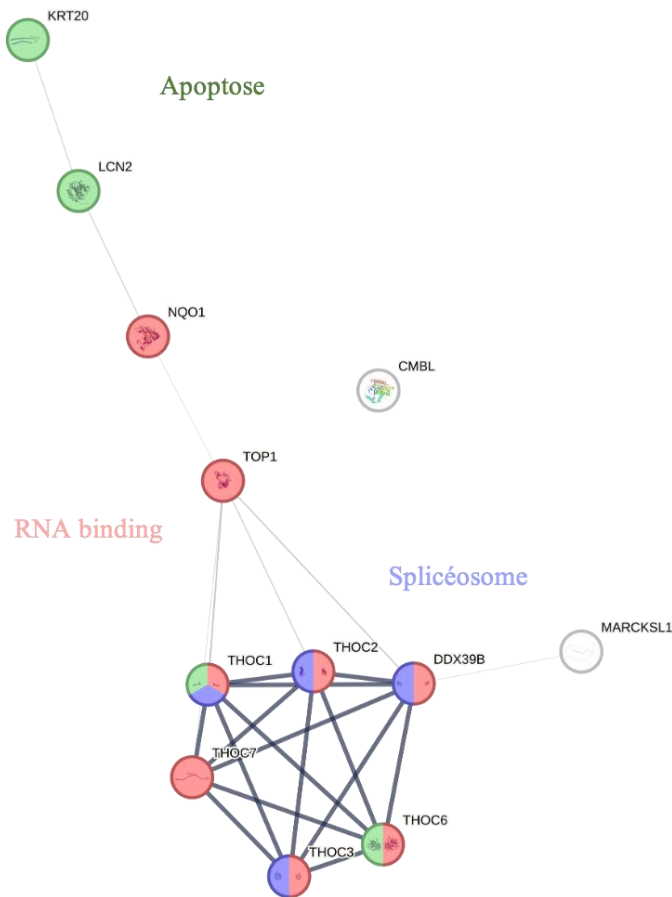
L'analyse des principales voies de signalisation a également validé le rôle des protéines identifiées dans le cycle de Krebs. Sachant que le cycle de Krebs, connu aussi sous le nom de cycle de l'acide citrique (TCA), est une voie métabolique essentiellement conçue pour décomposer les glucides, les lipides et les protéines afin de récupérer de l'énergie sous forme d'électrons à fort potentiel de transfert, ainsi qu'une molécule d'ATP. Le cycle TCA joue un rôle crucial en tant que point de convergence au sein du processus de respiration cellulaire, intégrant diverses sources de carburant provenant de l'alimentation, notamment le glucose, la glutamine et les acides gras. La dérégulation du cycle TCA due à des mutations dans des enzymes spécifiques ou à un métabolisme aérobie défectueux est associée à la croissance des tumeurs<sup>271</sup>.

Parmi les protéines impliquées dans le cycle de Krebs, des protéines de type aconitases sont présentes. De façon intéressante, l'aconitase 2 (ACO2) participe au cycle TCA en convertissant le citrate en isocitrate, et il a été montré que la diminution de l'expression d'ACO2 est associée à un mauvais pronostic dans le cancer gastrique<sup>272</sup>. Il a été constaté que l'expression accrue d'ACO2 entrave la prolifération cellulaire et oblige les cellules à rediriger le pyruvate vers les mitochondries, ce qui affaiblit les caractéristiques bioénergétiques de type Warburg<sup>273</sup>. C'est donc très intéressant de détecter des protéines de type aconitases dans les tissus sains, sachant que leur déficit est un élément favorisant le développement du cancer œsogastrique.

Les tissus cancéreux du sous-type ADK (cluster 1) contiennent des protéines impliquées dans le spliceosome, la formation de complexe ribonucléoprotéique et l'apoptose. En effet, comme le montre la **Figure 69**, trois voies de signalisation sont surexprimées dans le cluster 1.

On retrouve, tout d'abord, une protéine du complexe d'exportation de la transcription, utilisée pour l'exportation de l'ARNm viral à partir du noyau de la cellule hôte (DDX39BX). De façon intéressante, les protéines de liaison à l'ARN peuvent réguler l'expression des gènes en influençant de manière post-transcriptionnelle tous les aspects de la biologie de l'ARN. Des preuves de plus en plus nombreuses renforcent l'idée que le dérèglement ou le dysfonctionnement de ces protéines peuvent être à l'origine de diverses maladies humaines, y compris des cancers. Elles influencent divers phénotypes cellulaires associés au cancer, tels que la prolifération, l'apoptose, la sénescence, la migration, l'invasion et l'angiogenèse, contribuant ainsi à l'initiation et au développement des tumeurs, ainsi qu'au pronostic clinique<sup>274</sup>. De façon très intéressante, la protéine

NQO1 est surexprimée dans les tissus ADK. Celle-ci joue un rôle essentiel dans la protection des cellules normales contre les dommages oxydatifs et les attaques électrophiles. Elle est fortement exprimée dans de nombreuses tumeurs solides, ce qui suggère un rôle dans le développement et la progression du cancer. En effet, le rôle de NQO1 dans le cancer gastrique et sa surexpression est associée à un pronostic significativement défavorable et à une meilleure réponse au 5-FU chez les patients atteints de cancer gastrique<sup>275</sup>.



**Figure 69 :** Voies de signalisation surexprimées dans les clusters 2 et 3, représentant les tissus cancéreux du sous-type ADK. L'épaisseur des traits montre le degré de confiance pour le rapport entre les deux protéines.

De plus, la deuxième voie de signalisation surexprimée dans les tissus ADK est l'apoptose (KRT20 et LCN2). L'apoptose, en tant que mécanisme de défense de l'organisme contre le cancer et les agents infectieux, facilite la suppression des cellules tumorales ou des cellules infectées par des agents pathogènes. Elle contribue également à éliminer les lymphocytes en excès après une réponse immunitaire.

Enfin, la troisième et dernière voie de signalisation surexprimée dans le cluster 1 est lié au splicéosome. En effet, les protéines du complexe THO favorise le couplage entre la transcription et le traitement de l'ARNm tandis que DDX39B est une hélicase ARN du splicéosome. Le splicéosome, connu également sous le nom de particule d'épissage, est une structure en constante évolution composée d'ARN ribonucléoprotéique et de plus de 200 protéines, se trouvant dans le

noyau des cellules. Le spliceosome est essentiel pour l'épissage de l'ARN, une étape cruciale dans la transformation de l'ARN pré-messager en ARN messager mature, prêt pour la traduction en protéines. C'est une structure complexe qui subit des réarrangements pour garantir la conversion précise de l'information génétique en protéines fonctionnelles, et des dysfonctionnements peuvent causer des maladies génétiques ainsi que des cancers. De façon intéressante, des études s'intéressent déjà à la surexpression des facteurs d'épissage dans les tissus cancéreux, tels que U2AF1<sup>276</sup>. Une étude a aussi montré qu'un réseau fonctionnel d'événements d'épissage associés au cancer gastrique était contrôlé par des facteurs d'épissage augmentés<sup>277</sup>.

L'analyse protéomique révèle distinctement les processus biologiques caractéristiques de chaque type de tissu, que ce soit les 2 protéines exclusives détectées dans les tissus cancéreux PCC ou les protéines surexprimées dans le cluster 1 pour les tissus sains et les clusters 2 et 3 pour les tissus cancéreux ADK. Cela met en avant la nouvelle utilisation du SpiderMass en tant qu'outil concurrent de la microdissection laser pour l'analyse protéomique non ciblée. Ces résultats permettent d'améliorer les approches thérapeutiques pour les patients atteints de cancer gastrique.

#### *Identification des protéines alternatives*

La méthode d'analyse protéomique décrit dans le paragraphe précédent se limitait à interroger uniquement la base de données des protéines Uniprot, ce qui signifie que seules les protéines répertoriées pouvaient être détectées. Cependant, ces dernières années, les protéines alternatives issues de séquences codantes différentes ont été décrites<sup>278</sup>. Elles diffèrent des protéines de référence et enrichissent la diversité des protéines. Il devient impératif d'examiner ces protéines alternatives pour une meilleure compréhension des mécanismes pathologiques.

En effet, les protéines alternatives sont des protéines qui ne se trouvent pas répertoriées dans les bases de données protéiques. En réalité, ces protéines émergent à partir de cadres de lecture différents. De manière traditionnelle, on considère que les protéines sont produites à partir des cadres de lecture les plus longs. Ainsi, les protéines alternatives proviennent de cadres de lecture plus courts, de changements de cadre de lecture, de chevauchements avec des régions non codantes, de cadres de lecture situés dans les régions non codantes 3' et 5', et même d'ARNs précédemment annotés comme non codants<sup>279</sup>. Les protéines alternatives représentent un nouveau volet du protéome à explorer.

En utilisant la base de données contenant les protéines alternatives OpenProt<sup>280</sup>, nous avons pu identifier une protéine alternative IP\_2336131 qui est surexprimée dans les tissus cancéreux (**Figure 70**). Cette protéine alternative est issue de l'ARN non codant LOC105375566.

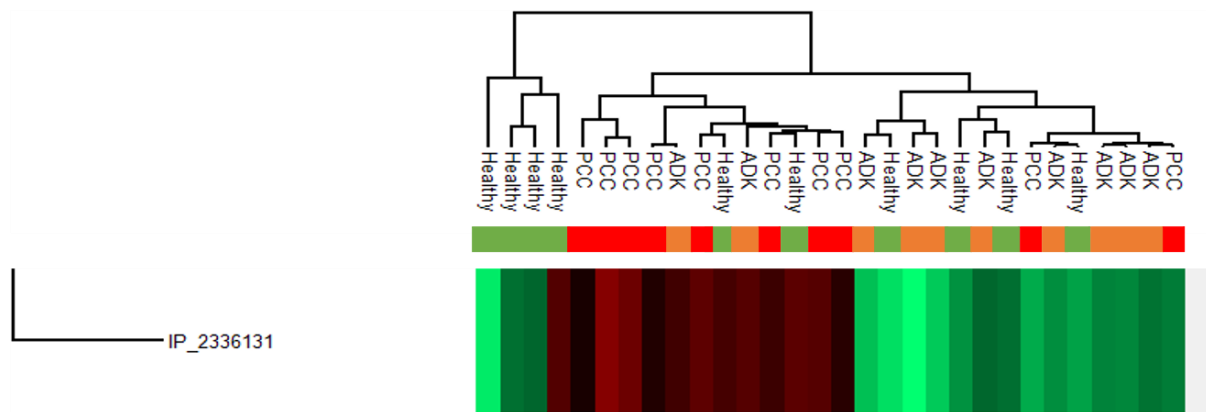


Figure 70 : Analyse statistique ANOVA des protéines alternatives identifiées par irradiation backside correspondant aux différents types de tissus (sains, PCC et ADK).

Il convient de souligner qu'au sein des protéines alternatives, une protéine est significativement surexprimée dans les tissus cancéreux. Cela met en lumière un nouveau potentiel biomarqueur protéique qui mérite une investigation approfondie pour dévoiler sa fonction. Par conséquent, il semble envisageable d'explorer davantage cette protéine en vue d'améliorer le diagnostic et de l'envisager comme une éventuelle cible thérapeutique.

## Conclusion et perspectives

Cette étude constitue une investigation approfondie visant à déterminer la possibilité de diminuer la rechute des patients de cancer œsogastrique grâce à l'utilisation polyvalente de la technologie SpiderMass. En effet, la polyvalence de la technologie SpiderMass a été démontrée par sa capacité à réaliser une classification histologique précise, à effectuer un sous-typage du cancer, à créer un bactérioscope, et à mener une analyse protéomique dans le contexte du cancer œsogastrique.

En premier lieu, le modèle de classification optimal permettant de discriminer les tissus cancéreux des tissus sains a atteint des taux de précision de 88 % et 87 % respectivement, en utilisant les modes d'ionisation négatif et positif. Lors de prédictions à l'aveugle effectuées sur un ensemble de validation indépendant comprenant 36 échantillons, seules 4 erreurs de classification ont été observées dans les deux modes d'ionisation, ce qui équivaut à un taux de classification correct de 90 % et qui est supérieur aux résultats de l'analyse extemporanée. De plus, pour fournir une justification biologique à cette classification basée sur les caractéristiques histologiques, nous avons identifié des marqueurs lipidiques. En effet, nous avons découvert 60 biomarqueurs fiables qui présentent des variations dans les classes de phospholipides en fonction du type de tissu. Les tissus sains se caractérisent par la présence de PIs, de PAs et de PCs, tandis que les tissus tumoraux montrent une concentration significative de DGs, de PSs, de PEs et de Cer.

De plus, nous avons réussi à distinguer de manière nette les deux sous-types de cancer, PCC et ADK, en utilisant un modèle de classification qui a atteint un taux de précision de 94 % dans les deux modes d'ionisation. Lorsque nous avons effectué des prédictions sur un ensemble de validation indépendant, nous avons observé un taux de classification correct de 75 % grâce à la technologie SpiderMass. De manière significative, nous avons identifié 24 marqueurs spécifiques du sous-type ADK et 13 spécifiques du sous-type PCC dans les deux modes d'ionisation. Il semble que les marqueurs moléculaires partagés avec les tissus sains se retrouvent dans les tissus du sous-type ADK, qui est un sous-type moins agressif et associé à une meilleure survie. En outre, des ions spécifiques des lymphocytes et des macrophages de type M1 sont surexprimés dans les tissus cancéreux du sous-type ADK. En revanche, un ion spécifique des macrophages de type M2, est surexprimé dans les tissus du sous-type PCC. Cette découverte est d'autant plus intéressante car les adénocarcinomes à cellules peu cohésives sont particulièrement agressifs tandis que les macrophages de type M2 ont des propriétés anti-inflammatoires et favorisent la croissance tumorale, contrairement aux macrophages de type M1. Ces résultats suggèrent des différences significatives entre les deux sous-types de cancer en termes de marqueurs moléculaires et d'implication de cellules immunitaires spécifiques dans la progression tumorale.

Le troisième aspect novateur de cette étude réside dans l'utilisation du SpiderMass pour diagnostiquer le cancer œsogastrique en se basant sur la présence de certaines souches bactériennes. Nous avons analysé six souches bactériennes afin de déterminer si la technologie SpiderMass pouvait prédire la présence de bactéries dans les tissus œsogastriques en utilisant les données qu'elle génère. Actuellement, les pathologistes se limitent à détecter *Helicobacter pylori*, négligeant ainsi d'autres bactéries potentiellement pertinentes, à partir d'échantillons de tissus conservés, mais cette détection n'est possible qu'après la chirurgie. Cela met en évidence la nécessité de méthodes de détection plus précoces et complètes des bactéries dans les cas de cancer œsogastrique. La création d'un bactérioscope adapté à l'approche SpiderMass pourrait permettre de traduire la présence de bactéries *in vivo* et de prédire les résultats pendant l'intervention chirurgicale, ce qui pourrait aider les chirurgiens à personnaliser leur prise en charge.

C'est ainsi que nous avons pu montrer que la souche *L. gasseri* est moins abondante dans les tissus cancéreux, en particulier dans les tissus PCC, par rapport aux tissus sains, où elle est prédominante, avec des pourcentages respectifs de 7 %, 2.7 % et 0.1 %. De nombreuses études ont montré le lien étroit entre *L. gasseri* et *H. pylori*. En effet, environ la moitié de la population mondiale est infectée par *H. pylori*, une infection qui provoque une inflammation de la muqueuse gastrique et entraîne progressivement la disparition des cellules pariétales de l'estomac responsables de la sécrétion de l'acide chlorhydrique. Cette situation conduit in fine à une condition appelée gastrite atrophique, caractérisée par une inflammation chronique et une réduction de l'acidité gastrique, augmentant ainsi considérablement le risque de développement d'un cancer gastrique<sup>281</sup>. Toutefois, il est important de noter que *L. gasseri*, l'un des probiotiques les plus renommés, semble être efficace contre *H. pylori*. Les principaux mécanismes antibactériens des probiotiques incluent la compétition avec *H. pylori* pour les sites de fixation sur les cellules épithéliales gastriques, le renforcement de la barrière muqueuse et la production d'acides organiques ayant des propriétés bactéricides<sup>282</sup>. Des études *in vitro* ont révélé que *L. gasseri* entraînait une diminution significative de l'expression de certains oncogènes, notamment Bcl-2,  $\beta$ -caténine, intégrine  $\alpha 5$  et intégrine  $\beta 1$ . Dans l'ensemble, les données actuelles suggèrent que certaines souches de *Lactobacillus* peuvent être bénéfiques pour la santé gastrique humaine en raison de leur capacité à éliminer les agents pathogènes, à maintenir l'intégrité de la barrière gastrique, à réduire l'inflammation et à exercer des effets protecteurs contre le cancer<sup>283,284</sup>.

En revanche, une tendance inverse est observée dans le cas de la souche *L. lactis*, avec une présence plus élevée dans les tissus cancéreux par rapport aux tissus sains. En effet, les cellules cancéreuses ont généralement une préférence pour la glycolyse aérobie plutôt que la phosphorylation oxydative, ce qui entraîne une absorption accrue de glucose et une production



importante de lactate, même en présence d'oxygène, phénomène bien connu sous le nom d'effet Warburg<sup>285</sup>. C'est pourquoi des souches bactériennes comme *Lactococcus*, qui contiennent des micro-organismes produisant de l'acide lactique, pourraient théoriquement favoriser la progression des tumeurs, car le lactate pourrait servir de source d'énergie pour la croissance tumorale et l'angiogenèse<sup>286</sup>.

En ce qui concerne le ratio prévu pour *S. bovis*, il met en lumière une association intrigante entre cette population bactérienne et le cancer œsogastrique. Plus précisément, on constate une proportion moins élevée de *S. bovis* dans les tissus sains par rapport aux tissus cancéreux, avec une légère variation entre les sous-types de cancer, où une abondance plus élevée est observée dans les tissus PCC par rapport aux tissus ADK. Des études ont montré que la flore des patients atteints de cancer gastrique est déséquilibrée et que les souches *Streptococcus* sont enrichies dans les tissus du cancer gastrique<sup>287</sup>, ce qui est très différent de la flore des personnes en bonne santé ou des patients souffrant de gastrite chronique<sup>288</sup>. Par conséquent, le streptocoque est déjà considéré comme un marqueur potentiel pour prédire le cancer gastrique. Yu *et al.* ont d'ailleurs utilisé un modèle de forêt aléatoire pour produire des preuves supplémentaires de l'utilisation du streptocoque comme marqueur du cancer gastrique<sup>289</sup>.

Le quatrième point très important de cette étude est l'utilisation du SpiderMass en tant que laser microdissection pour des analyses de protéomique spatiale, ici tout d'abord réalisées par approche bottom-up de type shot-gun. En effet, c'est une nouvelle corde à l'arc des possibilités qu'offrent l'utilisation de la technologie SpiderMass. L'irradiation backside, de 10 tissus sains, 10 tissus cancéreux de type PCC et 10 tissus cancéreux de type ADK, a permis de mettre en exergue des protéines biomarqueurs. C'est le cas de la Galectine-10 et de la Latexin pour les tissus PCC. En effet, de récentes études ont mis en évidence que la Galectine-10, généralement présente dans le cytoplasme des éosinophiles, contribue à la mort des cellules du piège extracellulaire, favorisant ainsi l'inflammation et/ou l'élimination efficace des agents pathogène, influant la progression tumorale<sup>290</sup>. D'ailleurs, plusieurs recherches ont indiqué que les éosinophiles sont capables de phagocyter différentes espèces bactériennes telles que *S. aureus* et *Escherichia coli*. Les éosinophiles possèdent des propriétés bactéricides car ils contribuent à la défense de l'organisme et pourraient donc jouer un rôle essentiel mais souvent méconnu en tant qu'effecteurs cellulaires dans l'immunité antibactérienne<sup>291</sup>. Un lien unique entre les bactéries et les protéines détectées dans les tissus PCC se dévoile et met en avant une potentielle cible thérapeutique.

Les protéines surexprimées dans les tissus cancéreux de sous-type ADK sont pour la plupart impliquées dans le spliceosome et dans la liaison avec l'ARN. De façon intéressante, une récente étude a montré le lien entre la survie globale des patients de cancer œsogastrique avec la présence

d'un groupe de 7 protéines de liaison à l'ARN<sup>292</sup>. De plus, le splicéosome est le complexe ARN multiprotéique chargé de la tâche complexe de l'épissage alternatif. Un épissage aberrant peut provenir de splicéosomes ou de facteurs d'épissage anormaux et conduire au développement et à la progression du cancer. La cartographie récente du splicéosome, des facteurs d'épissage associés et de leur relation avec le cancer a ouvert la voie à de nouvelles approches thérapeutiques qui tirent parti de l'influence généralisée de l'épissage alternatif<sup>293</sup>. La nouvelle méthodologie d'irradiation par l'arrière, ou backside, sur tissus œsogastriques permet de mettre en évidence la possibilité d'analyser les protéines grâce au SpiderMass. Par ailleurs, contrairement aux analyses de microprotéomique spatiales mises en place au laboratoire à partir d'analyses de micro-digestion et d'analyses de micro-extraction par jonction liquide, la technique backside permet d'atteindre une meilleure résolution spatiale, et pourrait être également utilisée pour des analyses Top-down, avec une résolution spatiale équivalente à celle montrée dans les analyses shot-gun et permettant ainsi un accès aux isoformes des protéines et aux modifications post-traductionnelles (PTMs).

Ainsi, notre étude illustre que la technologie SpiderMass pourrait être employée à plusieurs fins pour réaliser une analyse multiomique (lipides, bactéries et protéines) du cancer œsogastrique. Cette approche permet notamment de découvrir des biomarqueurs lipidiques, bactérielles et protéiques. Cela permettrait d'optimiser à la fois le diagnostic ainsi que la gestion thérapeutique approfondie du patient. Cette étude représente une avancée significative dans la compréhension du cancer œsogastrique. Elle démontre l'efficacité de la combinaison entre la spectrométrie de masse ambiante par SpiderMass et l'intelligence artificielle pour diminuer la rechute des patients atteints du cancer œsogastrique.



## Chapitre 6 : Identification des spécificités de la Colibactine dans la reprogrammation du métabolisme des lipides favorisant la progression du cancer du côlon droit

---

En plus de mon projet de thèse, j'ai eu l'occasion de participer à un projet collaboratif porté par le Dr. Matthias Chamillard et sa post-doctorante Dr. Nilmara de Oliveira Alves. Dans cette collaboration, nous avons utilisé le SpiderMass dans le cadre d'analyses métabolomiques des spécificités de la Colibactine dans la reprogrammation du métabolisme des lipides favorisant la progression du cancer du côlon droit. Dans ce chapitre sera introduit le sujet de l'étude, les résultats préalables à l'analyse SpiderMass seront résumés tandis que les résultats de l'analyse SpiderMass qui constituent ma contribution à cette collaboration, seront présentés de manière plus détaillée.

### Introduction

Le cancer colorectal représente la deuxième forme la plus courante de malignité et de décès liés au cancer dans le monde<sup>294</sup>. Les patients atteints de cancer du côlon droit (CCD) connaissent un pronostic moins favorable que ceux souffrant de cancer du côlon gauche (CCG), et il a été signalé précédemment qu'ils ne répondaient pas bien aux traitements de chimiothérapie classiques<sup>295</sup>. Des différences dans les biofilms associés à la tumeur et au métabolisme des lipides par rapport au CCG ont été identifiées dans le CCD. En plus des variations interindividuelles, on a identifié des types de cellules tumorales distincts au sein d'une même tumeur, mais leur origine reste inconnue<sup>296</sup>.

Les bactéries présentes à l'intérieur des tumeurs jouent un rôle dans la diversité des tumeurs par le biais de mécanismes encore mal compris. En effet, récemment, il a été découvert que certaines bactéries occupent des niches spécifiques dans le microenvironnement tumoral, ce qui entrave la réponse immunitaire et peut favoriser la persistance de la tumeur<sup>297</sup>. Il est important de noter que des altérations dans les profils et les communautés microbiennes intestinales sont associées à des modifications du métabolisme des lipides, pouvant contribuer à la résistance aux thérapies antitumorales<sup>298,299</sup>.

Les bactéries *Escherichia coli* productrices de Colibactine (CoPEC) peuvent stimuler la croissance des tumeurs du côlon en altérant le microenvironnement tumoral<sup>300</sup>. La Colibactine est un métabolite secondaire produit par certaines souches d'*E. coli* présentes dans l'intestin humain<sup>301</sup>. Il est bien établi que CoPEC provoque des cassures double-brin dans l'ADN<sup>302</sup> et forme des adduits d'ADN dans les cellules de mammifères<sup>303</sup>. Plus particulièrement, une signature spécifique de

dommages à l'ADN avec une séquence hexamérique riche en AT est observée dans les cellules du cancer colorectal humain lorsqu'elles sont infectées par certaines souches de CoPEC<sup>304</sup>. De manière tout aussi cruciale, cette génotoxine favorise un environnement immunosuppresseur, ce qui entraîne une résistance de la tumeur à l'immunothérapie par des mécanismes encore mal compris<sup>305</sup>.

Cependant, on ne sait toujours pas si et comment les bactéries *Escherichia coli* productrices de Colibactine associées à la tumeur peuvent influencer la reprogrammation du métabolisme des lipides dans le microenvironnement tumoral. C'est pourquoi, cette étude vise à investiguer comment CoPEC sont liées et peuvent potentiellement influencer la réorganisation métabolique des lipides dans le microenvironnement tumoral du cancer côlon droit.

## Résultats préalables à l'analyse SpiderMass

### CCD colonisé par CoPEC : pronostic moins favorable et immunosuppression liée aux bactéries tumorales

Dans cette étude, la présence de Colibactine a été validée dans 47 échantillons de tissus de CCD. Étant donné la corrélation entre la colonisation par des *Escherichia coli* pathogènes et les facteurs de mauvais pronostic pour le CCD, l'influence de la Colibactine sur la survie des patients a été examinée. Les résultats ont montré une moins bonne survie chez les patients atteints de CCD colonisés par des CoPEC positifs (valeur  $p=0.018$ ). En outre, une analyse des sous-types moléculaires consensuels (CMS) sur la base du profil d'expression génique a été effectuée : CMS1 (immunitaire), CMS2 (canonique), CMS3 (métabolique) et CMS4 (mésenchymateux). Plus de 50 % des patients colonisés par le CoPEC ont été classés dans le CMS3. Ces patients présentaient une prédominance de mutation APC. La polypose adénomateuse coli (APC) est largement reconnue comme un gène suppresseur de tumeur fortement muté dans les cancers colorectaux. La mutation et l'inactivation de ce gène est un événement clé et précoce presque uniquement observé dans la tumorigenèse colorectale.

Pour comprendre si la présence de Colibactine influence la composition bactérienne, nous avons analysé les différences entre les groupes CoPEC positifs et négatifs. Bien qu'aucune disparité significative n'ait été observée en termes de diversité alpha, l'analyse de la diversité bêta a révélé 35 taxons bactériens distincts. Sachant que la diversité alpha correspond à la diversité d'un échantillon unique contrairement à la diversité bêta qui permet de mesurer la diversité des espèces entre les échantillons. Parmi ceux-ci, seules les espèces d'*Escherichia-Shigella* sont significativement plus abondantes chez les patients colonisés par le CoPEC, tandis que trois autres espèces (*Treponema denticola*, *Collinsella tanakaei* et *Fusobacterium mortiferum*) sont significativement moins présentes chez ces patients.

Pour explorer les gènes différenciellement exprimés et les mécanismes biologiques liés aux tumeurs positives et négatives pour les CoPEC, les données RNAseq des tissus tumoraux ont été analysées. De façon intéressante, les gènes liés à l'activation des cellules B impliquées dans la réponse immunitaire sont significativement exprimés chez les patients colonisés par les CoPEC, tels que CD19, CCR6, CD40LG et DOCK11. La proportion de cellules T CD8+ cytotoxiques et de cellules B est distincte entre les patients positifs et négatifs pour les CoPEC, ce qui suggère l'hétérogénéité du microenvironnement immunitaire associé à la Colibactine.

Un effet immunosuppresseur local des bactéries associées aux tumeurs a récemment été établi<sup>297</sup>, mais les mécanismes sous-jacents restent insaisissables. En utilisant l'imagerie par hybridation in situ (FISH), la distribution hétérogène des microniches bactériennes dans le tissu du

CCD a été confirmée visuellement. De plus, les microniches bactériennes sont caractérisées par un effet immunosuppresseur avec une diminution des cellules T CD8<sup>+</sup> productrices d'IFN $\gamma$  infiltrant la tumeur, principalement chez les patients colonisés par CoPEC.

### La métabolomique spatiale révèle une augmentation des glycérophospholipides dans les tumeurs associées à la Colibactine

Afin d'explorer si les microniches bactériennes contribuent à une altération métabolique soutenant une immunité bactérienne réduite, une analyse métabolomique spatiale sur 12 échantillons de tissus CCD a été menée en utilisant la technologie de haute résolution spectrale 7T-MALDI-FTICR. Une prédominance des lipides (environ 40 %) chez les patients CoPEC-positifs par rapport aux patients CoPEC-négatifs a été montrée. Parmi ces lipides identifiés, environ 82 % appartenaient à la sous-classe des glycérophospholipides, tels que les lipides PA, PC, PS, PI, PE, PG et d'autres.

Il est notable qu'une corrélation significativement inverse entre la quantité de cellules CD8<sup>+</sup> et l'intensité des PC (valeur  $p=0.04$ ) chez les patients CoPEC-positifs a été observée. Cependant, cette corrélation était absente chez les patients CoPEC-négatifs (valeur  $p=0.81$ ). Ce résultat suggère que l'expression des cellules T CD8<sup>+</sup> est étroitement liée à une voie métabolique spécifique dans le microenvironnement immuno-suppressif induit par la Colibactine dans le cancer du côlon droit. Des indices de cibles dans les tissus CCD ont été découverts par le biais d'une analyse ARNseq. Dans les CoPEC positifs et les tumeurs à statut mutant APC, une augmentation de LIPIN-1, un gène codant pour une enzyme qui catalyse la déphosphorylation de l'AP pour produire du DG, a été observée. En revanche, le gène LPCAT1, qui joue un rôle dans le métabolisme des phospholipides en convertissant la lysophosphatidylcholine (LPC) en PC en présence d'acyl-CoA, était réduit. L'ensemble de ces résultats nous encourage à poursuivre l'étude des mécanismes par lesquels les CoPEC associés aux tumeurs peuvent conduire à une reprogrammation des lipides.

### CoPEC conduit à l'accumulation de gouttelettes lipidiques dans les cellules cancéreuses du côlon

Pour mieux caractériser la corrélation directe entre la Colibactine et l'accumulation de lipides, des cellules humaines de carcinome du côlon HCT116 ont été infectées avec la souche CoPEC (11G5) isolée d'un patient atteint de CCD ou avec une souche mutante qui ne produit pas de Colibactine (11G5 $\Delta$ clbQ)<sup>306</sup>.

Il a été démontré que la présence de CoPEC entraîne une accumulation de gouttelettes de lipides. Comme la production de gouttelettes lipidiques peut agir comme un interrupteur en réponse



à des déséquilibres dans le métabolisme énergétique et l'homéostasie redox<sup>307</sup>, l'accumulation antérieure de ROS dans les cellules HCT116 a été étudiée. Une évaluation semi-quantitative du pourcentage de cellules avec CellROX a été réalisée et une augmentation notable des cellules avec une fluorescence élevée a été observée après l'infection par CoPEC

L'étape suivante était donc de vérifier si les glycérophospholipides contribuaient à la discrimination entre les cellules HCT116 infectées par 11G5-, 11G5ΔclbQ et les cellules non infectées. Pour cela, l'utilisation de la technologie SpiderMass est pertinente.

## Résultats de l'analyse SpiderMass

### Méthodologie

Des cellules de cancer du côlon humain (HCT116) et de souris (MC38) ont été infectées par la souche clinique de CoPEC (11G5) ou par sa souche mutante (11G5 $\Delta$ clbQ). Des cellules non infectées par E. Coli ont été utilisées en tant que témoin négatif.

Après une confluence de près de 70% pour chaque condition des lignées cellulaires MC38 et HCT116, les cellules ont été lavées deux fois avec du DPBS, séchées sous PSM pendant 10 min à température ambiante puis analysées par le SpiderMass directement dans la plaque.

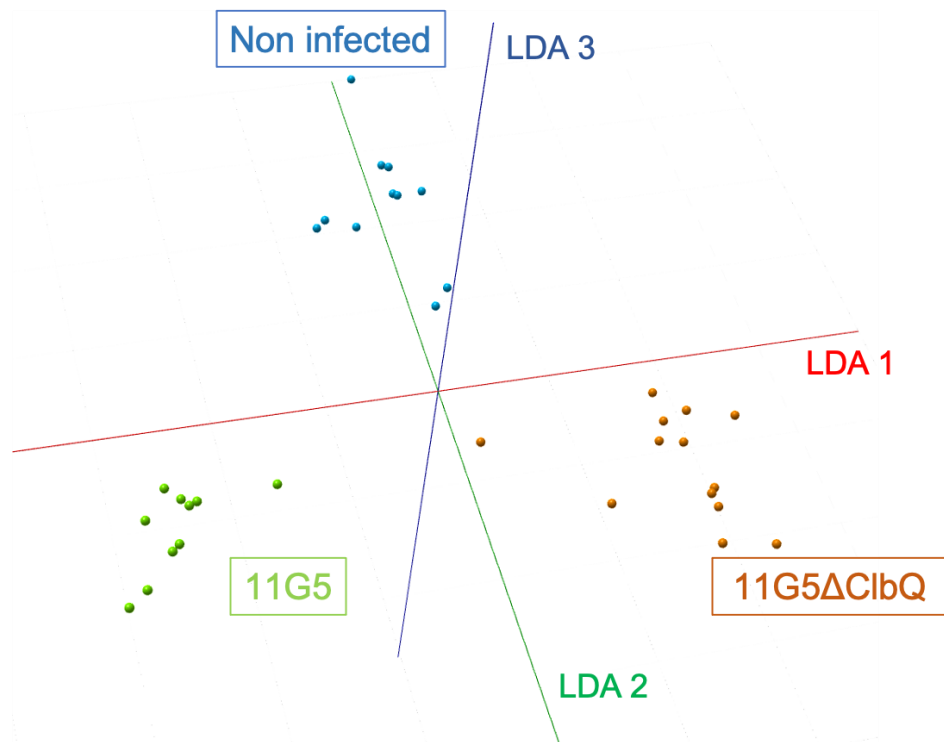
Dans cette étude, l'intensité du laser a été réglée à 4 mJ/impulsion et une infusion de 200  $\mu$ L/min d'isopropanol a été utilisée pendant chaque acquisition. 200  $\mu$ g/mL de leucine encéphaline ont été ajoutés à l'infusion pour jouer le rôle du lockmass. L'acquisition était composée d'une rafale de 10 tirs laser donnant lieu à un spectre individuel moyen. L'acquisition spectrale a été réalisée uniquement en mode négatif et en mode sensibilité du spectromètre (Q-TOF Xevo G2S, Waters) et la gamme de masse a été réglée sur  $m/z$  50-2000.

Les fichiers bruts ont été importés dans "Abstract Model Builder" - AMX pour effectuer des analyses statistiques multivariées à l'aide d'analyses discriminantes linéaires de la gamme de masse  $m/z$  400-1100. Les ions discriminants ont été trouvés grâce à un test non supervisé statistique non paramétrique Kruskal-Wallis avec correction de Bonferroni. Suite à ce test, des boîtes à moustaches pour chaque ion spécifique ont été obtenues.

Finalement, chaque ion correspondant à un biomarqueur a été étudié par MS/MS pour permettre son identification. Ainsi, après isolement de l'ion parent, les ions ont été soumis à une dissociation induite par collision dans la cellule de transfert. Les annotations des lipides ont ensuite été exécutées manuellement à partir des ions fragments observés sur les spectres MS/MS, avec l'aide de différentes banques de données publiques comme LipidMaps, Alex123 ainsi que Metlin.

### Résultats

Les spectres générés, à partir d'un triplicat de chaque condition, ont été soumis à une analyse linéaire discriminante supervisée. Comme le montre la **Figure 71**, une séparation est visible, suivant la LD1, entre les cellules HCT116 infectées par la souche clinique de CoPEC (11G5) ou par sa souche mutante (11G5 $\Delta$ clbQ). Tandis que la discrimination entre les cellules infectées ou non se fait selon la LD2. De plus, le taux de bonne classification, via le test « 20 % out » est de 85 %.



**Figure 71** : Le modèle de classification LDA basé sur trois groupes : les cellules HCT116 infectées par la souche CoPEC (11G5), une souche mutante qui ne produit pas de CoPEC (11G5Δc1bQ) et les cellules non infectées (contrôle négatif).

Après un test de Kruskal-Wallis (avec une valeur  $p \leq 0.05$ ), 23 ions se sont révélés significatifs. Parmi ceux-ci, 18 sont spécifiques de 11G5, 3 spécifiques de 11G5Δc1bQ et 2 spécifiques du contrôle négatif (**Tableau 12**). L'ion  $m/z$  603.45 DG 36:10 est surexprimé dans les cellules HCT166 non infectées tandis que l'ion  $m/z$  666.45 PE O-32:5, lui, y est sous-exprimé. Les 3 ions surexprimés dans les cellules infectées par la souche mutante 11G5Δc1bQ (ne produisant pas de Colibactine) sont  $m/z$  546.45 LPC 20:2,  $m/z$  684.65 et  $m/z$  888.65 PS 44:7. De façon intéressante, parmi les 18 ions spécifiques des cellules infectées par la souche clinique de CoPEC, 10 sont sous-exprimés et 8 sont surexprimés. En effet, une diminution des céramides, des PEs et des LPCs a été identifiée, tel que les ions  $m/z$  570.45 LPC 22:4,  $m/z$  682.55 Cer t42 :0(2OH) et  $m/z$  742.55 PE 36:2. A l'inverse, les glycérophospholipides tels que PC, PS et PI sont régulés à la hausse et présents uniquement ou plus souvent dans les cellules HCT116 infectées par CoPEC, comme les ions  $m/z$  802.55 PC 38:7,  $m/z$  884.45 PS 44:9 et  $m/z$  913.45 PI 40:4. En concordance avec les données observées dans les tumeurs du CCD, ces résultats confirment la prédominance des glycérophospholipides en présence de CoPEC.

**Tableau 12 :** Liste exhaustive des 23 biomarqueurs des cellules HCT116 infectées par la souche mutante 11G5clbQ, par la souche clinique de CoPEC 11G5 et les cellules témoins négatives.

<i>m/z</i>	11G5	11G5ΔclbQ	Témoin négatif	Possible annotation
542.45				[Cer (32:0(04))-H] <sup>-</sup>
544.45				/
546.45				[LPC (20:2)-H] <sup>-</sup>
560.45				[Cer (30:1(O2))-H] <sup>-</sup>
570.45				[LPC (22:4)-H] <sup>-</sup>
572.45				/
574.45				[LPC (22:2)-H] <sup>-</sup>
594.55				[Cer (38:0(O2))-H] <sup>-</sup>
603.45				[DG (36:10)-H] <sup>-</sup>
666.45				[PE (O-32:5)-H] <sup>-</sup>
682.55				[Cer (t42:0(2OH))-H] <sup>-</sup>
684.65				/
694.45				[PE (O-34:5)-H] <sup>-</sup>
740.55				[PE (36:3)-H] <sup>-</sup>
742.55				[PE (36:2)-H] <sup>-</sup>
800.45				[PS (38:9)-H] <sup>-</sup>
802.55				[PC (38:7)-H] <sup>-</sup>
822.55				[PS (O-40:5)-H] <sup>-</sup>
824.55				[PS (O-40:4)-H] <sup>-</sup>
884.45				[PS (44:9)-H] <sup>-</sup>
888.65				[PS (44:7)-H] <sup>-</sup>
910.55				[PS (42:0)-H] <sup>-</sup>
913.45				[PI (40:4)-H] <sup>-</sup>

## Résultats post-analyse SpiderMass

La mutation APC et la lectine de type c collaborent pour maintenir un glycérophospholipide bactérien dans le microenvironnement tumoral du cancer colorectal droit.

Pour définir plus précisément les gènes différentiellement exprimés impliqués dans les tissus tumoraux à forte et faible intensité de PC/PA, l'approche RNAseq a été utilisée. Grâce à cela, une augmentation de l'expression des gènes avec une réponse immunitaire tumorale antimicrobienne médiée par des peptides antimicrobiens, y compris REG3A, un gène qui code pour une lectine de type c qui démontre une activité bactéricide, a été démontrée. À la lumière de ce résultat, les patients ont été classés selon une expression REG3A faible, intermédiaire ou élevée. Curieusement, 100% des tumeurs à faible expression de REG3A sont identifiées comme des patients CoPEC négatifs et que la plupart des patients colonisés par CoPEC se trouvent dans la catégorie intermédiaire. En outre, une corrélation positive significative entre *Escherichia-Shigella* et REG3A a été observée.

Il a été remarqué que REG3A présente une corrélation positive significative avec les glycérophospholipides des PC chez les patients atteints de tumeurs mutantes APC (valeur  $p=0.04$ ). En revanche, l'analyse des tumeurs de type APC sauvage n'a révélé aucune corrélation significative (valeur  $p=0.43$ ). Dans l'ensemble, REG3A pourrait moduler les gènes impliqués dans le métabolisme des PC, contribuant ainsi à la progression tumorale en présence de CoPEC.

### La perte d'une lectine de type c chez la souris réduit la croissance tumorale après une infection par CoPEC

Les patients présentant une forte expression de glycérophospholipides sont associés à une augmentation de l'expression de REG3A, dont l'orthologue est Reg3b chez la souris. Pour étayer cette constatation, la manière dont la croissance tumorale était affectée par l'infection par CoPEC a été examinée en utilisant le modèle de tumeur sous-cutanée MC38. Dans le groupe non infecté traité à la streptomycine, aucune différence significative entre les souris WT et Reg3b<sup>-/-</sup> n'a été observée. En revanche, dans le groupe CoPEC, les souris Reg3b<sup>-/-</sup> avaient un volume tumoral plus faible que les souris WT, ce qui suggère un effet protumoral de Reg3b. Une différence significative a également été observée entre les animaux WT infectés par 11G5 et 11G5 $\Delta$ clbQ, ce qui confirme les études précédentes.

Une faible expression de l'IFN $\gamma$  dans les tissus de CCD de patients colonisés par des CoPEC a aussi été montrée. Par ailleurs, le volume tumoral était négativement corrélé au TNF- $\alpha$ , avec une diminution observée chez les souris WT infectées par le 11G5 par rapport aux souris WT infectées

par le 11G5 $\Delta$ clbQ. Cependant, cette différence n'a pas été observée dans le groupe Reg3b<sup>-/-</sup>. En accord avec ces résultats, l'expression d'IL-1 $\beta$  a été significativement réduite chez les souris WT infectées par 11G5 et aucune différence n'a été observée en l'absence de Reg3b.

Étant donné que REG3A présente une corrélation positive significative avec les glycérophospholipides PC et que l'expression de LPCAT1 était différenciellement exprimée chez les patients atteints de CCD à statut mutant APC, L'analyse de ce gène dans les tumeurs et les tissus du côlon des animaux a été entreprise. De manière intéressante, chez les souris WT infectées par le 11G5, qui avaient une croissance tumorale plus importante, il y a une diminution de l'expression de LPCAT1 par rapport aux souris WT infectées par le 11G5 $\Delta$ ClbQ dans les tumeurs. Cependant, cette différence n'est pas retrouvée dans le groupe de souris Reg3b<sup>-/-</sup>, démontrant que Reg3b a un rôle dans la voie PC en présence de CoPEC. De plus, pour explorer le rôle de la mutation APC, des cellules HT-29 de carcinome du colon humain hébergeant une mutation APC ont été utilisées, et en concordance avec les tumeurs CCD humaines et de souris, cette même différence dans l'expression de LPCAT1 a été observée.

Finalement, ces données démontrent que la présence de Reg3b module l'expression de LPCAT1, contribuant à un microenvironnement qui augmente la croissance tumorale après l'infection par CoPEC.

## Discussion et conclusion

Bien que la grande hétérogénéité du cancer colorectal soit devenue évidente<sup>308,309</sup>, il reste indispensable de définir comment l'hétérogénéité inter-tumorale et les différences entre les cellules cancéreuses au sein d'une tumeur peuvent entraîner des réponses variables aux traitements thérapeutiques. Dans ce contexte, notre compréhension de la façon dont le microbiote de l'hôte est régulé localement dans les tumeurs solides est limité, car cette approche ne peut pas fournir d'indications sur les événements moléculaires qui se sont produits séquentiellement ou parallèlement tout au long du développement de la tumeur<sup>297,310</sup>. En appliquant les technologies de pointe de profilage spatial *in situ* à haut débit, cette étude fournit la première preuve que l'hétérogénéité métabolique induite par les microniches bactériennes peut créer une vulnérabilité potentielle aux immunothérapies, de préférence chez les patients colonisés par des CoPEC positifs.

Les composants microbiens intratumoraux sont étroitement associés à l'initiation et au développement du cancer et affectent le potentiel thérapeutique des approches anticancéreuses<sup>311</sup>. Ici, la colonisation par les CoPEC a été démontrée en tant que facteur pronostique défavorable dans le CCD. En accord avec ce résultat, Bonnet et al.<sup>312</sup> ont rapporté qu'un mauvais pronostic est lié à la présence de CoPEC intra-tumorale, qui est détectée dans environ 50 à 60 % des biopsies de CCD humain, contre environ 20 % des patients souffrant de diverticulose<sup>313</sup>.

De nombreux efforts ont été déployés pour comprendre le mécanisme de la Colibactine dans le CCD humain. Une étude récente a déterminé une signature spécifique des dommages à l'ADN avec un motif de séquence hexamérique riche en AT où les CDB induites par la Colibactine se produisent de préférence, appelé motif de dommage de la Colibactine, qui présente une charge mutationnelle accrue<sup>304</sup>. En outre, des mutations ont été décrites au niveau des CDM, tels que l'APC<sup>314</sup>. Les résultats obtenus ici sont en accord avec ces observations. Par exemple, les patients atteints de RCC colonisés par CoPEC ont montré qu'ils portaient des mutations dans l'APC et sont classés dans le CMS3. En accord, ce sous-type de CMS est défini par une dysrégulation métabolique évidente et est significativement enrichi par des mutations APC<sup>315</sup>.

Outre l'enrichissement des gènes liés à l'activation des cellules B chez les patients colonisés par les CoPEC par rapport aux tumeurs de masse, la proportion de cellules T CD8+ cytotoxiques était distincte entre les patients positifs et négatifs pour les CoPEC, ce qui suggère l'hétérogénéité du microenvironnement immunitaire de la tumeur. Cependant, en appliquant uniquement l'analyse spatiale *in situ*, il a été démontré que les microniches bactériennes locales sont caractérisées par une diminution de l'infiltration des cellules T CD8+ productrices d'IFN $\gamma$ , principalement chez les patients colonisés par les CoPEC. A l'appui, Lopès *et al*<sup>305</sup> ont montré une diminution des



lymphocytes T CD3+ et CD8+ avec une diminution de l'efficacité de l'immunothérapie anti-PD-1 chez des souris infectées par des CoPEC.

L'analyse métabolomique globale est importante pour identifier les biomarqueurs et les voies métaboliques modifiées dans le cancer<sup>298,316</sup>, mais c'est une approche insuffisante pour indiquer l'effet du microbiote associé à la tumeur au niveau local. Sachant cela, en appliquant une approche métabolomique spatiale, cette étude a mis en évidence des schémas métabolomiques spatiaux présents dans des microniches bactériennes caractérisées par une signature lipidique unique chez les patients colonisés par les CoPEC. Conformément à l'altération du métabolisme des glycérophospholipides dans différents types de cancer, y compris le cancer gastrique<sup>317</sup>, des différences de distribution spatiale des PA, PC, PI, PE, PS et PG dans les microniches bactériennes entre les groupes positifs et négatifs pour les CoPEC ont été trouvées. De manière intéressante, l'accumulation de lipides dans les gouttelettes ainsi que l'augmentation des glycérophospholipides après l'infection par les CoPEC *in vitro* a été démontrée, notamment grâce à la technologie SpiderMass. Il est important de noter que les cellules cancéreuses du côlon présentent une accumulation de gouttelettes lipidiques et sont plus résistantes à la chimiothérapie. Parmi les différentes fonctions, les gouttelettes lipidiques peuvent i) protéger les membranes des réactions de peroxydation dans des conditions de stress oxydatif et maintenir l'homéostasie des organites ; ii) réguler l'autophagie par différents mécanismes ; et iii) répondre à une surcharge lipidique exogène pour réduire l'accumulation de lipides lipotoxiques et autres<sup>307</sup>. La présence de PC, PE, PS et PI est en accord avec les données obtenues par l'analyse métabolomique spatiale des tumeurs RCC. Il convient de noter que les PCs sont généralement augmentées dans le CCD et que les PIs ont également été associés au développement et à la progression du cancer<sup>318</sup>. En outre, les glycérophospholipides jouent également un rôle important dans la voie de l'autophagie. Il est clair que l'infection par CoPEC augmente le stress oxydatif dans les cellules cancéreuses du côlon et que l'accumulation de gouttelettes lipidiques est probablement une réponse à ces événements.

Disséquer les effets de la Colibactine dans le CCD par des approches métabolomiques et transcriptomiques permettra de comprendre précisément comment cette génotoxine conduit à la carcinogénèse. Ici, la preuve que l'activité de REG3A est favorable à la colonisation du CoPEC a été apportée. De manière surprenante, 100 % des tumeurs CCD avec une faible expression de REG3A ne sont pas colonisées par CoPEC. Il convient de noter que le rôle des lectines REG3A dans l'infection bactérienne intestinale n'est pas complètement compris, en fonction du contexte du tissu, du type de cancer et de l'initiation ou de la progression du cancer<sup>319</sup>. D'une part, elles ont une activité antimicrobienne. D'autre part, il a été rapporté que Reg3b prolonge la durée de la

colonisation de l'intestin par une bactérie entéropathogène. Ici, REG3A est positivement corrélé à une forte intensité de PC, principalement chez les patients présentant un mutant APC.

Dans cette étude, il a été montré que les tumeurs MC38 injectées par voie sous-cutanée à des souris déficientes en Reg3b entraînaient une adaptation métabolique similaire de la tumeur ainsi qu'une réduction significative de la croissance tumorale après l'infection par CoPEC. Ces données sont étayées par une diminution de l'expression de LPCAT1 observée uniquement chez les souris WT infectées par 11G5, même s'il existe une tendance à l'augmentation de ce gène dans le côlon en relation avec les tumeurs. Les enzymes LPCAT font partie du cycle de Lands et sont impliquées dans la réacylation du LPC pour régénérer le PC. Le rôle de LPCAT1 dans différents types de cancer reste obscur. D'une part, il a été rapporté que l'activité de la LPCAT1 semble être indépendante de toute stimulation inflammatoire, contrairement à la LPCAT2. D'autre part, la LPCAT1 a été associée au microenvironnement tumoral et à l'infiltration de cellules immunitaires<sup>320</sup>. Il est à noter que le cycle de Lands régule la taille des gouttelettes lipidiques tandis que la voie Kennedy régule l'équilibre entre les phospholipides et les lipides neutres<sup>321</sup> et par conséquent, le LPCAT1 est associé au micro-environnement tumoral et à l'infiltration des cellules immunitaires<sup>320</sup>. Par conséquent, il ne faut pas exclure que d'autres gènes impliqués dans la voie des glycérophospholipides puissent également être modulés par la présence de Reg3A, contribuant ainsi à la croissance de la tumeur

En conclusion, ce travail clarifie la façon dont les CoPEC peuvent façonner l'hétérogénéité tumorale par leur influence sur le métabolisme des lipides et permettra d'explorer les mécanismes de reprogrammation lipidique médiée par les CoPEC sur l'efficacité des thérapies antitumorales. En effet, les résultats de cette étude utilisant des approches métabolomiques et transcriptomiques spatialement résolues clarifient comment la présence de bactéries produisant de la Colibactine peut localement établir une hétérogénéité tumorale pour échapper à la surveillance immunitaire. Ces résultats fournissent des informations uniques à la fois pour l'intervention thérapeutique et pour la recherche fondamentale sur les mécanismes de l'interaction entre le microbiote et l'hôte. Ainsi, le ciblage combinatoire des CoPEC associés à REG3A est prometteur en tant que biomarqueur dans la thérapie du cancer afin d'obtenir de meilleurs résultats cliniques.



## Chapitre 7 : Conclusion générale et perspectives

---

L'objectif de ma thèse consistait à optimiser la technologie de spectrométrie de masse SpiderMass en tant qu'outil de diagnostic en temps réel des tumeurs œsogastriques, avec l'objectif d'aider la chirurgie oncologique en délimitant efficacement les marges de résection chirurgicale et l'extension loco-régionale des tumeurs chez les patients.

Pour parvenir à ce résultat, avant d'appliquer la technologie SpiderMass à la problématique du cancer œsogastrique, j'ai entrepris diverses études et développements visant à améliorer le traitement des données et la création des modèles de classification. Plus précisément, j'ai exploré la possibilité d'enrichir les données SpiderMass avec des données d'imagerie MALDI en comparant les spectres lipidiques des deux technologies. J'ai également examiné l'utilisation de nouveaux algorithmes d'apprentissage automatique pour créer des modèles de classification fiables. De plus, j'ai aussi travaillé sur la création d'immunescores et de bactérioscopes pour améliorer le diagnostic et le pronostic des patients.

Tout d'abord, comme détaillé dans le chapitre 2, une analyse approfondie a été effectuée afin de comparer les technologies d'imagerie MALDI et WALDI pour l'analyse des lipides, en utilisant le MALDI comme méthode de référence en recherche biomédicale pour l'analyse moléculaire non marquée. Dans cette étude, les profils lipidiques ont été comparés pour différents tissus, i.e., des cerveaux de rats et des échantillons de cancer œsogastrique. La similitude des pics détectés et l'indice de corrélation de Pearson pour l'intensité et la variance des profils spectraux ont été évalués entre les deux techniques pour différentes matrices MALDI adaptées aux lipides, différentes épaisseurs d'échantillons et résolutions spatiales. L'étude comparative révèle que le WALDI et le MALDI donnent des profils similaires, quelle que soit la résolution spatiale employée, mais que la matrice norharmane est la matrice pour laquelle les spectres présentent le plus de similitudes avec 100 % de similarité et + de 80 % de corrélation d'intensité et de variances. De plus, l'étude a permis de montrer la possibilité d'utiliser les données MALDI-MSI comme référence pour entraîner la technologie SpiderMass. Dans nos perspectives, nous avons tenté d'utiliser l'ensemble des 176 images MALDI ainsi que l'analyse de 190 zones d'intérêt par SpiderMass pour créer un modèle de classification performant et fiable capable de discriminer les tissus sains, les tissus cancéreux de sous-types ADK et PCC. En mode négatif, ce modèle a obtenu un taux de classification correct de 87 % après une validation croisée 20 fois. Lors du diagnostic à l'aveugle, celui-ci a réussi à classer avec succès 30 des 36 régions, atteignant un taux de réussite de 84 %. Cela est particulièrement satisfaisant car cette fois-ci, le modèle est capable de distinguer les sous-types de cancer et les tissus sains en une seule étape.

Par la suite, comme présenté dans le chapitre 3, il a été possible de développer un pipeline d'analyse des données collectées par la technologie SpiderMass pour créer des modèles de classification extrêmement robustes et fiables. En effet, historiquement, la technologie SpiderMass se basait sur la modélisation de l'analyse discriminante linéaire en utilisant l'outil AMX. Cependant, cette approche présentait de nombreuses limitations. Dans le but de surmonter ces limitations, nous avons développé un pipeline d'intelligence artificielle plus avancé et complet, permettant de créer des modèles de classification qui soient non seulement précis et fiables, mais qui permettraient également une identification claire et indiscutable des biomarqueurs lipidiques afin de donner une explicabilité biologique aux modèles de classification créés. Cette étude représente une avancée significative dans la compréhension du glioblastome. Elle met en évidence l'efficacité de la fusion de la spectrométrie de masse ambiante grâce à SpiderMass et de l'intelligence artificielle pour obtenir des classifications histologiques précises et des évaluations pronostiques pertinentes pour cette maladie complexe. En effet, l'intégration du SpiderMass dans ce processus d'apprentissage automatique a ouvert la voie à des possibilités prometteuses dans le domaine de la médecine de précision, que ce soit pour le glioblastome ou pour d'autres types de cancer qui se distinguent par leur diversité intrinsèque.

Par ailleurs, nous avons étudié la possibilité de créer un immunoscore basé sur les données SpiderMass afin de déterminer le pronostic vital précis des patients. Celui-ci représente une avancée importante avec la possibilité de transposer cet immunoscore en environnement *in vivo* et ainsi diagnostiquer et prévoir la survie du patient pendant l'intervention chirurgicale. Nous avons ainsi démontré la capacité du système à prédire la présence d'une plus grande quantité de lymphocytes chez les patients ayant survécu plus de 36 mois avec un taux moyen de lymphocytes de 40 % pour la survie prolongée, contre 36 % pour la survie plus courte. De manière semblable, l'évaluation du ratio des macrophages révèle une relation intéressante entre les cellules immunitaires et la survie globale des patients. Plus précisément, une proportion plus élevée de macrophages de type M1 (3.5 %) est observée dans les tissus des patients à survie prolongée par rapport à ceux des patients à survie courte qui ont une proportion très élevée de macrophages de type M2 (41.3 %).

De la même façon, un bactérioscore a été créé, et expliqué dans le chapitre 4, lors de l'analyse multi-omique effectuée sur le cancer œsogastrique. Les rapports calculés offrent une perspective sur la répartition des différentes souches bactériennes au sein des images SpiderMass, permettant ainsi d'évaluer globalement la composition bactérienne dans les tissus œsogastriques, qu'ils soient sains ou atteints de cancer. Plus précisément, la souche *L. gasseri* est plus abondante dans les tissus sains que dans les deux sous-types de cancer, notamment dans les tissus PCC. À

l'inverse, le ratio prédit pour la souche *S. bovis* révèle une proportion moindre dans les tissus sains par rapport aux tissus cancéreux, avec une légère variation entre les sous-types de cancer. De manière particulièrement intéressante, le rapport entre *S. bovis* et *L. gasseri* pourrait servir d'indicateur diagnostique pour les tissus œsogastriques. L'utilisation de l'analyse du microbiote en tant que marqueur prédictif pour les patients atteints de cancer gastrique semble être une perspective prometteuse.

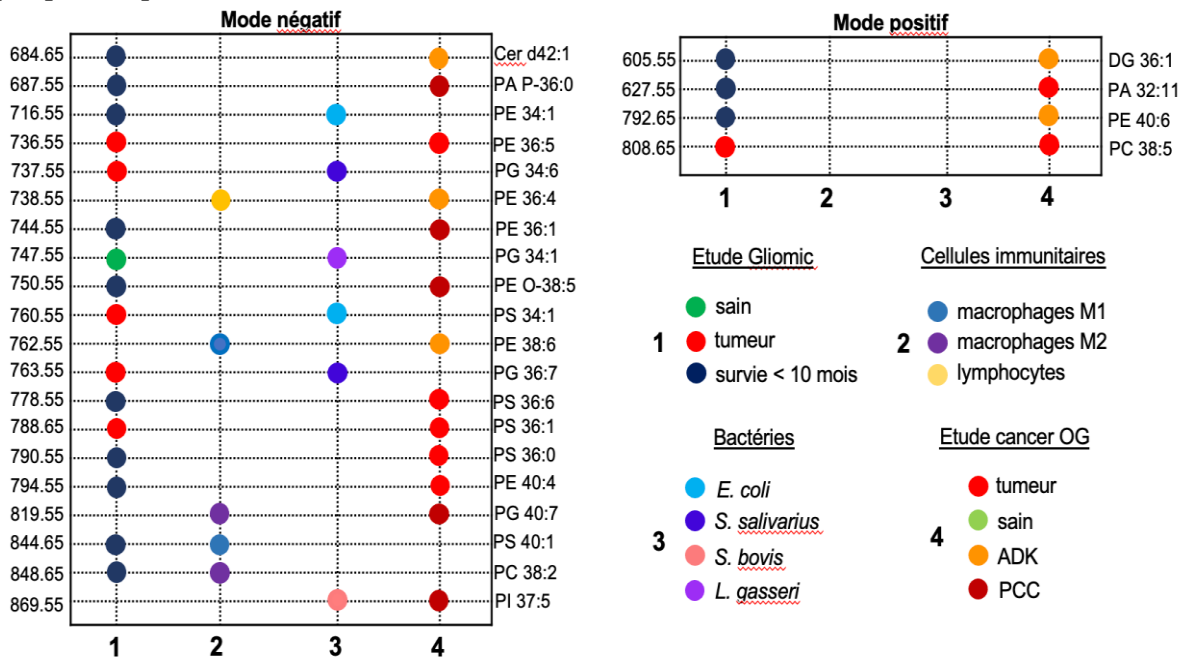


Figure 72 : Vue d'ensemble de l'analyse pan-cancer des marqueurs lipidiques trouvés à l'aide de la découverte supervisée et non supervisée, spécifiques de l'étude sur le glioblastome et sur le cancer œsogastrique, des cellules immunitaires et des bactéries, en mode positif et négatif.

Enfin, un point très important, abordé dans les différents chapitres de ce manuscrit de thèse, est la découverte de biomarqueurs robustes et fiables. En effet, l'utilisation d'analyses supervisées et non supervisées ont mis en exergue la possibilité de découvrir des marqueurs lipidiques qu'ils soient spécifiques des différents types histologiques ou du pronostic dans le glioblastome, des différentes cellules immunitaires, de chaque souche bactérienne analysée ainsi que des types et sous-types de tissus de cancer œsogastrique. En perspectives, il a été intéressant de faire une analyse pan-cancer des biomarqueurs découverts dans le glioblastome ainsi que dans le cancer œsogastrique (**Figure 72**). En mode négatif et positif, 10 et 4 ions sont retrouvés spécifiques et dans le glioblastome et dans le cancer œsogastrique. Neuf ions sont spécifiques du mauvais pronostic chez les patients atteints de glioblastome ainsi que du cancer œsogastrique. C'est le cas des ions  $m/z$  687.55 PA P-36:0 et  $m/z$  750.55 PE O-38:5 qui sont surexprimés dans le sous-type PCC contrairement aux ions  $m/z$  684.65 Cer d42 :1,  $m/z$  605.55 DG 36:1 et  $m/z$  792.65 PE 40:6 présents en grande proportion dans les tissus ADK. De plus, certains lipides ont été retrouvés spécifiques de certaines souches bactériennes ainsi que du glioblastome. C'est le cas des ions  $m/z$  716.55 PE 34:1 et  $m/z$  760.55 PS 34:1 surexprimés dans *E. coli* alors que chacun n'apparaît

respectivement que dans le pronostic < 10 mois ou dans les tissus tumoraux pour les patients atteints de glioblastome. Deux ions surexprimés  $m/z$  737.55 PG 34:6 et  $m/z$  763.55 PG 36:7 dans la souche *S. salivarius* sont aussi présents dans les tissus tumoraux du glioblastome. Tandis que les tissus sains de glioblastome et la souche bactérienne *L. gasseri* ont un ion biomarqueur en commun, qui est le  $m/z$  747.55 PG 34:1. En outre, nous avons examiné la détection des lipides biomarqueurs liés à la présence ou à l'absence de cancer dans les tissus de glioblastome ou de cancer œsogastrique, dans le but de mieux comprendre les processus biologiques impliqués dans leur distinction.

En conclusion, ces recherches ont permis d'apporter de nombreuses améliorations à la technologie SpiderMass dans l'objectif d'en faire un dispositif médical efficace pour la chirurgie de précision. Par ailleurs, ces investigations offrent diverses perspectives. La confirmation de la similitude des spectres MALDI et SpiderMass ouvre la possibilité d'appliquer des stratégies basées sur des réseaux de neurones avec apprentissage par transfert. De manière similaire, ces méthodes pourraient être employées pour pré-entraîner les modèles de classification à partir de vastes cohortes de tissus FFPE rétrospectifs, avec un transfert d'apprentissage applicable aux tissus frais congelés dans des cohortes prospectives plus restreintes. Concernant l'immunoscore, les résultats montrent la fonctionnalité de cette approche, laquelle devra être approfondie par l'étude de plusieurs sous-types de cellules immunitaires pour améliorer davantage la précision des prédictions de survie et de rechute des patients. D'autres aspects, tels que la résistance aux traitements, pourraient également être explorés grâce à ces méthodes. Il sera également nécessaire de développer ces approches pour les déployer en temps réel au bloc opératoire. En ce qui concerne le bactérioscore, il ouvre une voie particulièrement prometteuse. Des études devront être menées pour comprendre les mécanismes biologiques sous-tendant les différences observées entre les différentes régions histologiques d'un même tissu, ainsi qu'entre les tissus sains et les divers sous-types de cancer. Ces informations pourraient se révéler précieuses pour le dépistage par endoscopie. La technologie SpiderMass, étant peu invasive, pourrait permettre de détecter la dysbiose à l'origine des cancers œsogastriques, une fois miniaturisée et adaptée pour une utilisation en endoscopie, et de mettre en place des traitements visant à rétablir un équilibre bactérien, potentiellement préventif du développement de ces cancers. Enfin, la prochaine étape majeure pour la technologie SpiderMass sera sa validation en chirurgie humaine, avec le déploiement des développements réalisés en temps réel et *in vivo*.



## Références

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J Clin* **71**, 209–249 (2021).
2. Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017. Rapport technique. 80.
3. Renaud, F. *et al.* FREGAT : base clinicobiologique nationale française dédiée à la recherche sur les cancers œsogastriques. *Annales de Pathologie* **37**, 457–466 (2017).
4. Piessen, G., Messager, M., Leteurtre, E., Jean-Pierre, T. & Mariette, C. Signet Ring Cell Histology is an Independent Predictor of Poor Prognosis in Gastric Adenocarcinoma Regardless of Tumoral Clinical Presentation. *Annals of Surgery* **250**, 878–887 (2009).
5. Okines, A., Verheij, M., Allum, W., Cunningham, D. & Cervantes, A. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* **21**, v50–v54 (2010).
6. Wang, S.-Y. *et al.* Clinical Impact of Positive Surgical Margin Status on Gastric Cancer Patients Undergoing Gastrectomy. *Ann Surg Oncol* **16**, 2738–2743 (2009).
7. Spicer, J. *et al.* Diagnostic accuracy and utility of intraoperative microscopic margin analysis of gastric and esophageal adenocarcinoma. *Ann Surg Oncol* **21**, 2580–2586 (2014).
8. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
9. Cristescu, R. *et al.* Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**, 449–456 (2015).
10. Ogrinc, N., Saudemont, P., Takats, Z., Salzet, M. & Fournier, I. Cancer Surgery 2.0: Guidance by Real-Time Molecular Technologies. *Trends in Molecular Medicine* **27**, 602–615 (2021).
11. Fatou, B. *et al.* In vivo Real-Time Mass Spectrometry for Guided Surgery Application. *Sci Rep* **6**, 25919 (2016).
12. Ogrinc, N. *et al.* Water-assisted laser desorption/ionization mass spectrometry for minimally invasive in vivo and real-time surface analysis using SpiderMass. *Nat Protoc* **14**, 3162–3182 (2019).
13. Fatou, B. *et al.* Real time and in vivo pharmaceutical and environmental studies with SpiderMass instrument. *Journal of Biotechnology* **281**, 61–66 (2018).
14. Saudemont, P. *et al.* Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption/Ionization Mass Spectrometry Technology. *Cancer Cell* **34**, 840–851.e4 (2018).
15. Ogrinc, N. *et al.* Direct Water-Assisted Laser Desorption/Ionization Mass Spectrometry Lipidomic Analysis and Classification of Formalin-Fixed Paraffin-Embedded Sarcoma Tissues without Dewaxing. *Clin Chem* **67**, 1513–1523 (2021).
16. Ogrinc, N. *et al.* Mass Spectrometry-Based Differentiation of Oral Tongue Squamous Cell Carcinoma and Nontumor Regions With the SpiderMass Technology. *Frontiers in Oral Health* **3**, (2022).
17. Ogrinc, N., Schneider, S., Bourmaud, A., Salzet, M. & Fournier, I. Direct in vivo Analysis of CBD- and THC-Acid Type Cannabinoids and Classification of Cannabis Cultivars by SpiderMass. (2021) doi:10.20944/preprints202112.0493.v1.
18. Ogrinc, N. *et al.* Robot-Assisted SpiderMass for In Vivo Real-Time Topography Mass Spectrometry Imaging. *Anal. Chem.* **93**, 14383–14391 (2021).
19. Awad, H., Khamis, M. M. & El-Aneed, A. Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **50**, 158–175 (2015).
20. Gowda, G. A. N. & Djukovic, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. *Methods Mol Biol* **1198**, 3–12 (2014).
21. Munson, B. Chemical Ionization Mass Spectrometry: Theory and Applications. in *Encyclopedia of Analytical Chemistry* (ed. Meyers, R. A.) a6004 (John Wiley & Sons, Ltd, Chichester, UK, 2000). doi:10.1002/9780470027318.a6004.

22. Lattimer, R. P. & Schulten, H.-R. AND FIELD DESORPTION MASS SPECTROMETRY: PAST, PRESENT, AND FUTURE.
23. Vestal, M. L. Methods of Ion Generation. *Chem. Rev.* **101**, 361–376 (2001).
24. Barber, M., Bordoli, R. S., Elliott, G. J., Sedgwick, R. D. & Tyler, A. N. Fast atom bombardment mass spectrometry. *ACS Publications* <https://pubs.acs.org/doi/pdf/10.1021/ac00241a002> (2002) doi:10.1021/ac00241a002.
25. Cremin, P., Donnelly, D. M. X., Wolfender, J.-L. & Hostettmann, K. Liquid chromatographic-thermospray mass spectrometric analysis of sesquiterpenes of *Armillaria* (Eumycota: Basidiomycotina) species. *Journal of Chromatography A* **710**, 273–285 (1995).
26. Karas, M., Bachmann, D., Bahr, U. & Hillenkamp, F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *International Journal of Mass Spectrometry and Ion Processes* **78**, 53–68 (1987).
27. Tanaka, K. *et al.* Protein and polymer analyses up to  $m/z$  100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).
28. Berkenkamp, S., Karas, M. & Hillenkamp, F. Ice as a matrix for IR-matrix-assisted laser desorption/ionization: mass spectra from a protein single crystal. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7003–7007 (1996).
29. Dreisewerd, K. *et al.* Molecular Profiling of Native and Matrix-Coated Tissue Slices from Rat Brain by Infrared and Ultraviolet Laser Desorption/Ionization Orthogonal Time-of-Flight Mass Spectrometry. *Anal. Chem.* **79**, 2463–2471 (2007).
30. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64–71 (1989).
31. Parr, M. K., Wüst, B., Teubel, J. & Joseph, J. F. Splitless hyphenation of SFC with MS by APCI, APPI, and ESI exemplified by steroids as model compounds. *Journal of Chromatography B* **1091**, 67–78 (2018).
32. Chen, H., Gamez, G. & Zenobi, R. What Can We Learn from Ambient Ionization Techniques? *Journal of the American Society for Mass Spectrometry* **20**, 1947–1963 (2009).
33. Feider, C. L., Krieger, A., DeHoog, R. J. & Eberlin, L. S. Ambient Ionization Mass Spectrometry: Recent Developments and Applications. *Anal. Chem.* **91**, 4266–4290 (2019).
34. Venter, A. R., Douglass, K. A., Shelley, J. T., Hasman, G. & Honarvar, E. Mechanisms of Real-Time, Proximal Sample Processing during Ambient Ionization Mass Spectrometry. *Anal. Chem.* **86**, 233–249 (2014).
35. Laskin, J. & Lanekoff, I. Ambient Mass Spectrometry Imaging Using Direct Liquid Extraction Techniques. *Anal. Chem.* **88**, 52–73 (2016).
36. Takáts, Z., Wiseman, J. M. & Cooks, R. G. Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J. Mass Spectrom.* **40**, 1261–1275 (2005).
37. Takáts, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass Spectrometry Sampling under Ambient Conditions with Desorption Electrospray Ionization. *Science, New Series* **306**, 471–473 (2004).
38. Haddad, R., Sparrapan, R. & Eberlin, M. N. Desorption sonic spray ionization for (high) voltage-free ambient mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 2901–2905 (2006).
39. Chen, H., Venter, A. & Cooks, R. G. Extractive electrospray ionization for direct analysis of undiluted urine, milk and other complex mixtures without sample preparation. *Chem. Commun.* 2042 (2006) doi:10.1039/b602614a.
40. Wu, C., Siems, W. F. & Hill, H. H. Secondary Electrospray Ionization Ion Mobility Spectrometry/Mass Spectrometry of Illicit Drugs. *Anal. Chem.* **72**, 396–403 (2000).
41. Chan, C.-C., Bolgar, M. S., Miller, S. A. & Attygalle, A. B. Desorption ionization by charge exchange (DICE) for sample analysis under ambient conditions by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **21**, 1554–1560 (2010).

42. Luo, Z. *et al.* Air Flow-Assisted Ionization Imaging Mass Spectrometry Method for Easy Whole-Body Molecular Imaging under Ambient Conditions. *Anal. Chem.* **85**, 2977–2982 (2013).
43. Van Berkel, G. J., Sanchez, A. D. & Quirke, J. M. E. Thin-Layer Chromatography and Electrospray Mass Spectrometry Coupled Using a Surface Sampling Probe. *Anal. Chem.* **74**, 6216–6223 (2002).
44. Roach, P. J., Laskin, J. & Laskin, A. Nanospray desorption electrospray ionization: an ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst* **135**, 2233 (2010).
45. Kertesz, V. & Van Berkel, G. J. Fully automated liquid extraction-based surface sampling and ionization using a chip-based robotic nanoelectrospray platform. *J. Mass Spectrom.* **45**, 252–260 (2010).
46. Wisztorski, M. *et al.* Microproteomics by liquid extraction surface analysis: Application to FFPE tissue to study the fimbria region of tubo-ovarian cancer. *Prot. Clin. Appl.* **7**, 234–240 (2013).
47. Meurs, J. *et al.* Improved Extraction Repeatability and Spectral Reproducibility for Liquid Extraction Surface Analysis–Mass Spectrometry Using Superhydrophobic–Superhydrophilic Patterning. *Anal. Chem.* **90**, 6001–6005 (2018).
48. Wang, H., Liu, J., Cooks, R. G. & Ouyang, Z. Paper Spray for Direct Analysis of Complex Mixtures Using Mass Spectrometry. *Angew. Chem.* **122**, 889–892 (2010).
49. Hiraoka, K., Nishidate, K., Mori, K., Asakawa, D. & Suzuki, S. Development of probe electrospray using a solid needle. *Rapid Commun. Mass Spectrom.* **21**, 3139–3144 (2007).
50. Zhang, J. *et al.* Nondestructive tissue analysis for ex vivo and in vivo cancer diagnosis using a handheld mass spectrometry system. *Sci Transl Med* **9**, eaan3968 (2017).
51. Yue, H., He, F., Zhao, Z. & Duan, Y. Plasma-based ambient mass spectrometry: Recent progress and applications. *Mass Spectrometry Reviews* **42**, 95–130 (2023).
52. Cody, R. B., Laramée, J. A. & Durst, H. D. Versatile New Ion Source for the Analysis of Materials in Open Air under Ambient Conditions. *Anal. Chem.* **77**, 2297–2302 (2005).
53. Shelley, J. T. *et al.* Characterization of direct-current atmospheric-pressure discharges useful for ambient desorption/ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **20**, 837–844 (2009).
54. Pitman, C. N. & LaCourse, W. R. Desorption atmospheric pressure chemical ionization: A review. *Analytica Chimica Acta* **1130**, 146–154 (2020).
55. McEwen, C. N., McKay, R. G. & Larsen, B. S. Analysis of Solids, Liquids, and Biological Tissues Using Solids Probe Introduction at Atmospheric Pressure on Commercial LC/MS Instruments. *Anal. Chem.* **77**, 7826–7831 (2005).
56. Harper, J. D. *et al.* Low-Temperature Plasma Probe for Ambient Desorption Ionization. *Anal. Chem.* **80**, 9097–9104 (2008).
57. Na, N., Zhao, M., Zhang, S., Yang, C. & Zhang, X. Development of a dielectric barrier discharge ion source for ambient mass spectrometry. *J. Am. Soc. Mass Spectrom.* **18**, 1859–1862 (2007).
58. Ratcliffe, L. V. *et al.* Surface Analysis under Ambient Conditions Using Plasma-Assisted Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **79**, 6094–6101 (2007).
59. Nemes, P. & Vertes, A. Laser Ablation Electrospray Ionization for Atmospheric Pressure, in Vivo, and Imaging Mass Spectrometry. *Anal. Chem.* **79**, 8098–8106 (2007).
60. McMillan, K. S., McCluskey, A. G., Sorensen, A., Boyd, M. & Zagnoni, M. Emulsion technologies for multicellular tumour spheroid radiation assays. *Analyst* **141**, 100–110 (2016).
61. Sampson, J. S., Hawkridge, A. M. & Muddiman, D. C. Generation and detection of multiply-charged peptides and proteins by matrix-assisted laser desorption electrospray ionization (MALDESI) fourier transform ion cyclotron resonance mass spectrometry. *J. Am. Soc. Mass Spectrom.* **17**, 1712–1716 (2006).
62. Ekelöf, M. & Muddiman, D. C. IR-MALDESI method optimization based on time-resolved measurement of ion yields. *Anal Bioanal Chem* **410**, 963–970 (2018).

63. Yung, Y. P. *et al.* Solid Sampling with a Diode Laser for Portable Ambient Mass Spectrometry. *Analytical Chemistry*.
64. Benham, K., Hodyss, R., Fernández, F. M. & Orlando, T. M. Laser-Induced Acoustic Desorption Atmospheric Pressure Photoionization via VUV-Generating Microplasmas. *J. Am. Soc. Mass Spectrom.* **27**, 1805–1812 (2016).
65. Woolman, M. *et al.* Optimized Mass Spectrometry Analysis Workflow with Polarimetric Guidance for ex vivo and in situ Sampling of Biological Tissues. *Sci Rep* **7**, 468 (2017).
66. Schäfer, K.-C. *et al.* In Vivo, In Situ Tissue Analysis Using Rapid Evaporative Ionization Mass Spectrometry. *Angewandte Chemie International Edition* **48**, 8240–8242 (2009).
67. Liu, Y. & Adewuyi, Y. G. A review on removal of elemental mercury from flue gas using advanced oxidation process: Chemistry and process. *Chemical Engineering Research and Design* **112**, 199–250 (2016).
68. Li, X., Attanayake, K., Valentine, S. J. & Li, P. Vibrating Sharp-edge Spray Ionization (VSSI) for voltage-free direct analysis of samples using mass spectrometry. *Rapid Commun Mass Spectrom* **35**, (2021).
69. McEwen, C. N., Larsen, B. S. & Trimpin, S. Laserspray Ionization on a Commercial Atmospheric Pressure-MALDI Mass Spectrometer Ion Source: Selecting Singly or Multiply Charged Ions. *Anal. Chem.* **82**, 4998–5001 (2010).
70. Pagnotti, V. S., Chubatyi, N. D. & McEwen, C. N. Solvent Assisted Inlet Ionization: An Ultrasensitive New Liquid Introduction Ionization Method for Mass Spectrometry. *Anal. Chem.* **83**, 3981–3985 (2011).
71. Horan, A. J., Apsokardu, M. J. & Johnston, M. V. Droplet Assisted Inlet Ionization for Online Analysis of Airborne Nanoparticles. *Anal. Chem.* **89**, 1059–1062 (2017).
72. Hänel, L., Kwiatkowski, M., Heikau, L. & Schlüter, H. Mass spectrometry-based intraoperative tumor diagnostics. *Future Science OA* **5**, FSO373 (2019).
73. Balog, J. *et al.* Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Sci. Transl. Med.* **5**, (2013).
74. Balog, J. *et al.* Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry. *Anal. Chem.* **82**, 7343–7350 (2010).
75. Balog, J. *et al.* Identification of the Species of Origin for Meat Products by Rapid Evaporative Ionization Mass Spectrometry. *J. Agric. Food Chem.* **64**, 4793–4800 (2016).
76. Phelps, D. L. *et al.* The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). *Br J Cancer* **118**, 1349–1358 (2018).
77. St John, E. R. *et al.* Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. *Breast Cancer Res* **19**, 59 (2017).
78. Vaysse, P.-M. *et al.* Stromal vapors for real-time molecular guidance of breast-conserving surgery. *Sci Rep* **10**, 20109 (2020).
79. Tzafetas, M. *et al.* The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. *PNAS* **117**, 7338–7346 (2020).
80. Marcus, D. *et al.* Point-of-Care Diagnosis of Endometrial Cancer Using the Surgical Intelligent Knife (iKnife)—A Prospective Pilot Study of Diagnostic Accuracy. *Cancers* **14**, 5892 (2022).
81. Alexander, J. *et al.* A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife. *Surg Endosc* **31**, 1361–1370 (2017).
82. Balog, J. *et al.* In Vivo Endoscopic Tissue Identification by Rapid Evaporative Ionization Mass Spectrometry (REIMS). *Angew. Chem.* **127**, 11211–11214 (2015).
83. Guitton, Y. *et al.* Rapid evaporative ionisation mass spectrometry and chemometrics for high-throughput screening of growth promoters in meat producing animals. *Food Additives & Contaminants: Part A* **35**, 900–910 (2018).



84. Song, G. *et al.* In Situ Method for Real-Time Discriminating Salmon and Rainbow Trout without Sample Preparation Using iKnife and Rapid Evaporative Ionization Mass Spectrometry-Based Lipidomics. *J. Agric. Food Chem.* **67**, 4679–4688 (2019).
85. Barlow, R. S. *et al.* Rapid Evaporative Ionization Mass Spectrometry: A Review on Its Application to the Red Meat Industry with an Australian Context. *Metabolites* **11**, 171 (2021).
86. Strittmatter, N. *et al.* Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chem. Commun.* **49**, 6188 (2013).
87. Genangeli, M., Heeren, R. M. A. & Porta Siegel, T. Tissue classification by rapid evaporative ionization mass spectrometry (REIMS): comparison between a diathermic knife and CO<sub>2</sub> laser sampling on classification performance. *Anal Bioanal Chem* **411**, 7943–7955 (2019).
88. Saudemont, P. *et al.* Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption/Ionization Mass Spectrometry Technology. *Cancer Cell* **34**, 840–851.e4 (2018).
89. Keung, E. Z. & Roland, C. L. Accurate and Reproducible Diagnosis of Canine Soft Tissue Sarcoma Using Mass Spectrometry: A Step in the Right Direction. *Cancer Cell* **34**, 697–699 (2018).
90. Duhamel, M. *et al.* Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. *Nat Commun* **13**, 6665 (2022).
91. Fatou, B., Salzet, M. & Fournier, I. Real time human micro-organisms biotyping based on Water-Assisted Laser Desorption/Ionization. *The EuroBiotech Journal* **3**, 97–104 (2019).
92. Ogrinc, N. *et al.* Sarcoma Molecular Pathological Diagnosis of FFPE Tissues without Dewaxing Based on Lipidomic Profiles Using SpiderMass Technology. <https://www.researchsquare.com/article/rs-157940/v1> (2021) doi:10.21203/rs.3.rs-157940/v1.
93. these philippe.pdf.
94. Zhang, J. *et al.* Nondestructive tissue analysis for ex vivo and in vivo cancer diagnosis using a handheld mass spectrometry system. *Sci. Transl. Med.* **9**, eaa3968 (2017).
95. Sans, M. *et al.* Performance of the MasSpec Pen for Rapid Diagnosis of Ovarian Cancer. *Clinical Chemistry* **65**, 674–683 (2019).
96. Lu, H., Zhang, H., Wei, Y. & Chen, H. Ambient mass spectrometry for the molecular diagnosis of lung cancer. *Analyst* **145**, 313–320 (2020).
97. Zhang, J. *et al.* Direct Molecular Analysis of In Vivo and Freshly Excised Tissues in Human Surgeries with the MasSpec Pen Technology. <http://medrxiv.org/lookup/doi/10.1101/2020.12.14.20248101> (2020) doi:10.1101/2020.12.14.20248101.
98. King, M. E. *et al.* Rapid diagnosis and tumor margin assessment during pancreatic cancer surgery with the MasSpec Pen technology. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104411118 (2021).
99. Keating, M. F. *et al.* Integrating the MasSpec Pen to the da Vinci Surgical System for In Vivo Tissue Analysis during a Robotic Assisted Porcine Surgery. *Anal. Chem.* **92**, 11535–11542 (2020).
100. Gatmaitan, A. N., Lin, J. Q., Zhang, J. & Eberlin, L. S. Rapid Analysis and Authentication of Meat Using the MasSpec Pen Technology. *J. Agric. Food Chem.* **69**, 3527–3536 (2021).
101. Feider, C. L. *et al.* Integrating the MasSpec Pen with Sub-Atmospheric Pressure Chemical Ionization for Rapid Chemical Analysis and Forensic Applications. *Anal. Chem.* **93**, 7549–7556 (2021).
102. Garza, K. Y. *et al.* Rapid Screening of COVID-19 Directly from Clinical Nasopharyngeal Swabs Using the MasSpec Pen. *Anal. Chem.* **93**, 12582–12593 (2021).
103. Krieger, A. C., Povilaitis, S. C., Gowda, P., O’Connell, L. A. & Eberlin, L. S. Noninvasive Detection of Chemical Defenses in Poison Frogs Using the MasSpec Pen. *ACS Meas. Sci. Au* **2**, 475–484 (2022).
104. Franjic, K., Cowan, M. L., Kraemer, D. & Miller, R. J. D. Laser selective cutting of biological tissues by impulsive heat deposition through ultrafast vibrational excitations. *Opt. Express* **17**, 22937 (2009).

105. Woolman, M. *et al.* Rapid determination of medulloblastoma subgroup affiliation with mass spectrometry using a handheld picosecond infrared laser desorption probe. *Chem. Sci.* **8**, 6508–6519 (2017).
106. Zhang, L. *et al.* Caged circular siRNAs for photomodulation of gene expression in cells and mice. *Chem. Sci.* **9**, 44–51 (2018).
107. Katz, L. *et al.* Dual Laser and Desorption Electrospray Ionization Mass Spectrometry Imaging Using the Same Interface. *Anal. Chem.* **92**, 6349–6357 (2020).
108. Woolman, M. *et al.* Picosecond Infrared Laser Desorption Mass Spectrometry Identifies Medulloblastoma Subgroups on Intraoperative Timescales. *Cancer Res* **79**, 2426–2434 (2019).
109. Wurlitzer, M. *et al.* Mass Spectrometric Lipid Profiles of Picosecond Infrared Laser-Generated Tissue Aerosols Discriminate Different Brain Tissues. *Lasers Surg. Med.* **52**, 228–234 (2020).
110. Katz, L. *et al.* Picosecond Infrared Laser Mass Spectrometry Identifies a Metabolite Array for 10 s Diagnosis of Select Skin Cancer Types: A Proof-of-Concept Feasibility Study. *Anal. Chem.* **94**, 16821–16830 (2022).
111. Ifa, D. R. & Eberlin, L. S. Ambient Ionization Mass Spectrometry for Cancer Diagnosis and Surgical Margin Evaluation. *Clinical Chemistry* **62**, 111–123 (2016).
112. Paraskevaidi, M. *et al.* Laser-assisted rapid evaporative ionisation mass spectrometry (LA-REIMS) as a metabolomics platform in cervical cancer screening. *EBioMedicine* **60**, 103017 (2020).
113. Katz, L., Tata, A., Woolman, M. & Zarrine-Afsar, A. Lipid Profiling in Cancer Diagnosis with Hand-Held Ambient Mass Spectrometry Probes: Addressing the Late-Stage Performance Concerns. *Metabolites* **11**, 660 (2021).
114. Seddiki, K. *et al.* Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun* **11**, 5595 (2020).
115. Amstalden Van Hove, E. R., Smith, D. F. & Heeren, R. M. A. A concise review of mass spectrometry imaging. *Journal of Chromatography A* **1217**, 3946–3954 (2010).
116. Buchberger, A. R., DeLaney, K., Johnson, J. & Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal. Chem.* **90**, 240–265 (2018).
117. Vaysse, P.-M., Heeren, R. M. A., Porta, T. & Balluff, B. Mass spectrometry imaging for clinical research – latest developments, applications, and current limitations. *Analyst* **142**, 2690–2712 (2017).
118. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
119. Fournier, I., Wisztorski, M. & Salzert, M. Tissue imaging using MALDI-MS: a new frontier of histopathology proteomics. *Expert Review of Proteomics* **5**, 413–424 (2008).
120. Franck, J., Arafah, K., Elayed, M., Bonnel, D. & Vergara, D. MALDI Imaging Mass Spectrometry.
121. Strupat, K., Karas, M. & Hillenkamp, F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry.
122. Beavis, R. C., Chait, B. T. & Fales, H. M. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **3**, 432–435 (1989).
123. Beavis, R. C., Chaudhary, T. & Chait, B. T.  $\alpha$ -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **27**, 156–158 (1992).
124. Lee, P. Y. *et al.* Molecular tissue profiling by MALDI imaging: recent progress and applications in cancer research. *Critical Reviews in Clinical Laboratory Sciences* **58**, 513–529 (2021).

125. Korte, A. R. & Lee, Y. J. MALDI-MS analysis and imaging of small molecule metabolites with 1,5-diaminonaphthalene (DAN): Diaminonaphthalene as small molecule MALDI matrix. *J. Mass Spectrom.* **49**, 737–741 (2014).
126. Juhasz, P. & Costello, C. E. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry of underivatized and permethylated gangliosides. *J. Am. Soc. Mass Spectrom.* **3**, 785–796 (1992).
127. Chandler, C. J. *et al.* Fast atom bombardment mass spectrometry of some anthracycline and bisanthracycline derivatives. *Biol. Mass Spectrom.* **17**, 21–25 (1988).
128. Buchberger, A. R., DeLaney, K., Johnson, J. & Li, L. Mass Spectrometry Imaging: A Review of Emerging Advancements and Future Insights. *Anal. Chem.* **90**, 240–265 (2018).
129. Yamagaki, T. & Nakanishi, H. A new technique distinguishing a2-3 sialyl linkage from a2-6 linkage in sialyllactoses and sialyl-N-acetyllactosamines by post-source decay fragmentation method of MALDI-TOF mass spectrometry.
130. Ellis, S. R. *et al.* More from less: high-throughput dual polarity lipid imaging of biological tissues. *Analyst* **141**, 3832–3841 (2016).
131. Angerer, T. B., Bour, J., Biagi, J.-L., Moskovets, E. & Frache, G. Evaluation of 6 MALDI-Matrices for 10  $\mu$ m Lipid Imaging and On-Tissue MSn with AP-MALDI-Orbitrap. *J. Am. Soc. Mass Spectrom.* **33**, 760–771 (2022).
132. Lemaire, R. *et al.* Solid Ionic Matrixes for Direct Tissue Analysis and MALDI Imaging. *Anal. Chem.* **78**, 809–819 (2006).
133. Juhasz, P. & Costello, C. E. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry of underivatized and permethylated gangliosides. *J. Am. Soc. Mass Spectrom.* **3**, 785–796 (1992).
134. Strupat, K., Karas, M. & Hillenkamp, F. 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption-ionization mass spectrometry.
135. Beavis, R. C., Chaudhary, T. & Chait, B. T.  $\alpha$ -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Org. Mass Spectrom.* **27**, 156–158 (1992).
136. Yamagaki, T. & Nakanishi, H. A new technique distinguishing a2-3 sialyl linkage from a2-6 linkage in sialyllactoses and sialyl-N-acetyllactosamines by post-source decay fragmentation method of MALDI-TOF mass spectrometry.
137. Beavis, R. C., Chait, B. T. & Fales, H. M. Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Commun. Mass Spectrom.* **3**, 432–435 (1989).
138. Powers, T. W. *et al.* MALDI Imaging Mass Spectrometry Profiling of N-Glycans in Formalin-Fixed Paraffin Embedded Clinical Tissue Blocks and Tissue Microarrays. *PLoS ONE* **9**, e106255 (2014).
139. Lemaire, R. *et al.* Direct Analysis and MALDI Imaging of Formalin-Fixed, Paraffin-Embedded Tissue Sections. *J. Proteome Res.* **6**, 1295–1305 (2007).
140. Buck, A. *et al.* Round robin study of formalin-fixed paraffin-embedded tissues in mass spectrometry imaging. *Anal Bioanal Chem* **410**, 5969–5980 (2018).
141. Denti, V. *et al.* Antigen Retrieval and Its Effect on the MALDI-MSI of Lipids in Formalin-Fixed Paraffin-Embedded Tissue. *J. Am. Soc. Mass Spectrom.* **31**, 1619–1624 (2020).
142. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018. Étude à partir des registres des cancers du réseau Francim. Résultats préliminaires. Rapport. (1990).
143. Zaanan, A. *et al.* Gastric cancer: French intergroup clinical practice guidelines for diagnosis, treatments and follow-up (SNFGE, FFCD, GERCOR, UNICANCER, SFCD, SFED, SFRO). *Digestive and Liver Disease* **50**, 768–779 (2018).
144. Hu, B. *et al.* Gastric cancer: Classification, histology and application of molecular pathology. *Journal of Gastrointestinal Oncology* **3**, (2012).



145. Nagtegaal, I. D. *et al.* The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 182–188 (2020).
146. Drubay, V. *et al.* Poorly cohesive cells gastric carcinoma including signet-ring cell cancer: Updated review of definition, classification and therapeutic management. *WJGO* **14**, 1406–1428 (2022).
147. European Chapter of International Gastric Cancer Association *et al.* Consensus on the pathological definition and classification of poorly cohesive gastric carcinoma. *Gastric Cancer* **22**, 1–9 (2019).
148. Dos Santos, N., Seruca, R., Constancia, M., Seixas, M. & Sobrinho-Simoes, M. Microsatellite instability at multiple loci in gastric carcinoma: Clinicopathologic implications and prognosis. *Gastroenterology* **110**, 38–44 (1996).
149. Lei, Z. *et al.* Identification of Molecular Subtypes of Gastric Cancer With Different Responses to PI3-Kinase Inhibitors and 5-Fluorouracil. *Gastroenterology* **145**, 554–565 (2013).
150. Wenk, M. R. The emerging field of lipidomics. *Nat Rev Drug Discov* **4**, 594–610 (2005).
151. McNamara, J. R., Warnick, G. R. & Cooper, G. R. A brief history of lipid and lipoprotein measurements and their contribution to clinical chemistry. *Clinica Chimica Acta* **369**, 158–167 (2006).
152. The nomenclature of lipids (recommendations 1976). *Journal of Lipid Research* **19**, 114–128 (1978).
153. Lipides chole-doc principaux-constituants - numéro SPÉCIAL 132 & 133 2012 PRINCIPAUX - Studocu. <https://www.studocu.com/fr/document/universite-daix-marseille/biologie/lipides-chole-doc-principaux-constituants/50912782>.
154. Fahy, E. *et al.* A comprehensive classification system for lipids. *Journal of Lipid Research* **46**, 839–861 (2005).
155. Andrikopoulos, N. K. Triglyceride species compositions of common edible vegetable oils and methods used for their identification and quantification. *Food Reviews International* **18**, 71–102 (2002).
156. Hirabayashi, Y. & Furuya, S. Roles of l-serine and sphingolipid synthesis in brain development and neuronal survival. *Progress in Lipid Research* **47**, 188–203 (2008).
157. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
158. Swinnen, J. V., Brusselmans, K. & Verhoeven, G. Increased lipogenesis in cancer cells: new players, novel targets. *Current Opinion in Clinical Nutrition & Metabolic Care* **9**, 358–365 (2006).
159. Han, X. Lipidomics for studying metabolism. *Nature Reviews Endocrinology* **12**, 668–679 (2016).
160. Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M. & Kulbacka, J. Lipid composition of the cancer cell membrane. *J Bioenerg Biomembr* **52**, 321–342 (2020).
161. Pradas, I. *et al.* Lipidomics Reveals a Tissue-Specific Fingerprint. *Front. Physiol.* **9**, 1165 (2018).
162. Casares, D., Escribá, P. V. & Rosselló, C. A. Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues. *IJMS* **20**, 2167 (2019).
163. Putta, P. *et al.* Phosphatidic acid binding proteins display differential binding as a function of membrane curvature stress and chemical properties. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1858**, 2709–2716 (2016).
164. Han, S. *et al.* Phospholipase D activates HIF-1-VEGF pathway via phosphatidic acid. *Exp Mol Med* **46**, e126–e126 (2014).
165. Liu, L.-K., Choudhary, V., Toulmay, A. & Prinz, W. A. An inducible ER–Golgi tether facilitates ceramide transport to alleviate lipotoxicity. *Journal of Cell Biology* **216**, 131–147 (2017).

166. Rockenfeller, P. *et al.* Phosphatidylethanolamine positively regulates autophagy and longevity. *Cell Death Differ* **22**, 499–508 (2015).
167. Patel, D. & Witt, S. N. Ethanolamine and Phosphatidylethanolamine: Partners in Health and Disease. *Oxidative Medicine and Cellular Longevity* **2017**, 1–18 (2017).
168. Podo, F. *et al.* Activation of Phosphatidylcholine-Specific Phospholipase C in Breast and Ovarian Cancer: Impact on MRS-Detected Choline Metabolic Profile and Perspectives for Targeted Therapy. *Front. Oncol.* **6**, (2016).
169. Bailey, L. J., Choudhary, V. & Bollag, W. B. Possible Role of Phosphatidylglycerol-Activated Protein Kinase C- $\beta$ II in Keratinocyte Differentiation. *TODJ* **11**, 59–71 (2017).
170. Shlomovitz, I., Speir, M. & Gerlic, M. Flipping the dogma – phosphatidylserine in non-apoptotic cell death. *Cell Commun Signal* **17**, 139 (2019).
171. Belzile, O. *et al.* Antibody targeting of phosphatidylserine for the detection and immunotherapy of cancer. *ITT Volume* **7**, 1–14 (2018).
172. Muftuoglu, Y., Xue, Y., Gao, X., Wu, D. & Ha, Y. Mechanism of substrate specificity of phosphatidylinositol phosphate kinases. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8711–8716 (2016).
173. Ogretmen, B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer* **18**, 33–50 (2018).
174. Fatou, B. *et al.* Remote Atmospheric Pressure Infrared Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry (Remote IR-MALDI MS) of Proteins \*. *Molecular & Cellular Proteomics* **17**, 1637–1649 (2018).
175. Murphy, R. C., Hankin, J. A. & Barkley, R. M. Imaging of lipid species by MALDI mass spectrometry. *J Lipid Res* **50 Suppl**, S317-322 (2009).
176. Bowman, A. P., Heeren, R. M. A. & Ellis, S. R. Advances in mass spectrometry imaging enabling observation of localised lipid biochemistry within tissues. *TrAC Trends in Analytical Chemistry* **120**, 115197 (2019).
177. Guo, S., Wang, Y., Zhou, D. & Li, Z. Significantly increased monounsaturated lipids relative to polyunsaturated lipids in six types of cancer microenvironment are observed by mass spectrometry imaging. *Sci Rep* **4**, (2014).
178. Andersen, M. K. *et al.* Spatial differentiation of metabolism in prostate cancer tissue by MALDI-TOF MSI. *Cancer & Metabolism* **9**, 9 (2021).
179. Mallah, K. *et al.* Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Imaging of Lipids in Experimental Model of Traumatic Brain Injury Detecting Acylcarnitines as Injury Related Markers. *Anal. Chem.* **91**, 11879–11887 (2019).
180. Mulder, I. A. *et al.* Distinguishing core from penumbra by lipid profiles using Mass Spectrometry Imaging in a transgenic mouse model of ischemic stroke. *Sci Rep* **9**, 1–10 (2019).
181. Longuespée, R. *et al.* MALDI mass spectrometry imaging: A cutting-edge tool for fundamental and clinical histopathology. *PROTEOMICS – Clinical Applications* **10**, 701–719 (2016).
182. Holzlechner, M., Eugenin, E. & Prideaux, B. Mass spectrometry imaging to detect lipid biomarkers and disease signatures in cancer. *CANCER REPORTS* **2**, e1229 (2019).
183. Moyer, S. C. *et al.* Atmospheric pressure matrix-assisted laser desorption/ionization (AP MALDI) on a quadrupole ion trap mass spectrometer. *International Journal of Mass Spectrometry* **226**, 133–150 (2003).
184. Laiko, V. V., Baldwin, M. A. & Burlingame, A. L. Atmospheric Pressure Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Chem.* **72**, 652–657 (2000).
185. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012).
186. Race, A. M., Styles, I. B. & Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J Proteomics* **75**, 5111–5112 (2012).
187. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther Drug Monit* **27**, 747–751 (2005).

188. Sud, M. *et al.* LMSD: LIPID MAPS structure database. *Nucleic Acids Res* **35**, D527-532 (2007).
189. Pauling, J. K. *et al.* Proposal for a common nomenclature for fragment ions in mass spectra of lipids. *PLoS One* **12**, e0188394 (2017).
190. Fagerer, S. R., Nielsen, S., Ibáñez, A. & Zenobi, R. Matrix-Assisted Laser Desorption/Ionization Matrices for Negative Mode Metabolomics. *Eur J Mass Spectrom (Chichester)* **19**, 39–47 (2013).
191. Eveque-Mourroux, M. R. *et al.* Spatially resolved endogenous improved metabolite detection in human osteoarthritis cartilage by matrix assisted laser desorption ionization mass spectrometry imaging. *Analyst* **144**, 5953–5958 (2019).
192. Tucker, L. H. *et al.* Untargeted Metabolite Mapping in 3D Cell Culture Models Using High Spectral Resolution FT-ICR Mass Spectrometry Imaging. *Anal. Chem.* **91**, 9522–9529 (2019).
193. Schaff, L. R. & Mellinshoff, I. K. Glioblastoma and Other Primary Brain Malignancies in Adults: A Review. *JAMA* **329**, 574 (2023).
194. Hassn Mesrati, M., Behrooz, A. B., Y. Abuhamad, A. & Syahir, A. Understanding Glioblastoma Biomarkers: Knocking a Mountain with a Hammer. *Cells* **9**, 1236 (2020).
195. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803–820 (2016).
196. Fernandes, C. *et al.* Current Standards of Care in Glioblastoma Therapy. in *Glioblastoma* (ed. De Vleeschouwer, S.) (Codon Publications, Brisbane (AU), 2017).
197. Thakkar, J. P. *et al.* Epidemiologic and Molecular Prognostic Review of Glioblastoma. *Cancer Epidemiol Biomarkers Prev* **23**, 1985–1996 (2014).
198. Stupp, R. *et al.* Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N Engl J Med* **352**, 987–996 (2005).
199. Van Tellingen, O. *et al.* Overcoming the blood–brain tumor barrier for effective glioblastoma treatment. *Drug Resistance Updates* **19**, 1–12 (2015).
200. Brown, T. J. *et al.* Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis. *JAMA Oncol* **2**, 1460 (2016).
201. Ma, R., Taphoorn, M. J. B. & Plaha, P. Advances in the management of glioblastoma. *J Neurol Neurosurg Psychiatry* **92**, 1103–1111 (2021).
202. Livermore, L. J. *et al.* Raman spectroscopy to differentiate between fresh tissue samples of glioma and normal brain: a comparison with 5-ALA–induced fluorescence-guided surgery. *Journal of Neurosurgery* 1–11 (2020) doi:10.3171/2020.5.JNS20376.
203. Livermore, L. J. *et al.* Rapid intraoperative molecular genetic classification of gliomas using Raman spectroscopy. *Neuro-Oncology Advances* **1**, vdz008 (2019).
204. Jermyn, M. *et al.* Intraoperative brain cancer detection with Raman spectroscopy in humans. *Sci. Transl. Med.* **7**, (2015).
205. Agar, N. Y. R. *et al.* Development of Stereotactic Mass Spectrometry for Brain Tumor Surgery. *Neurosurgery* **68**, 280–290 (2011).
206. Eberlin, L. S. *et al.* Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1611–1616 (2013).
207. Rose, M. *et al.* Surfaceome Proteomic of Glioblastoma Revealed Potential Targets for Immunotherapy. *Front Immunol* **12**, 746168 (2021).
208. Lee, E. S. & Durant, T. J. S. Supervised machine learning in the mass spectrometry laboratory: A tutorial. *Journal of Mass Spectrometry and Advances in the Clinical Lab* **23**, 1–6 (2022).
209. Seddiki, K. *et al.* Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun* **11**, 5595 (2020).
210. Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)* **23**, 18 (2020).
211. Gunning, D. & Aha, D. W. DARPA’s Explainable Artificial Intelligence Program.

212. Ogrinc, N. *et al.* Water-assisted laser desorption/ionization mass spectrometry for minimally invasive in vivo and real-time surface analysis using SpiderMass. *Nat Protoc* **14**, 3162–3182 (2019).
213. Ribeiro, M. T., Singh, S. & Guestrin, C. ‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier. Preprint at <http://arxiv.org/abs/1602.04938> (2016).
214. King, M. E. *et al.* Long-chain polyunsaturated lipids associated with responsiveness to anti-PD-1 therapy are colocalized with immune infiltrates in the tumor microenvironment. *Journal of Biological Chemistry* **299**, 102902 (2023).
215. Gerke, V. & Moss, S. E. Annexins: From Structure to Function. *Physiological Reviews* **82**, 331–371 (2002).
216. Lizarbe, M., Barrasa, J., Olmo, N., Gavilanes, F. & Turnay, J. Annexin-Phospholipid Interactions. Functional Implications. *IJMS* **14**, 2652–2683 (2013).
217. Kiss, A., Erdódi, F. & Lontay, B. Myosin phosphatase: Unexpected functions of a long-known enzyme. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1866**, 2–15 (2019).
218. Saddoughi, S. A. & Ogretmen, B. Diverse Functions of Ceramide in Cancer Cell Death and Proliferation. in *Advances in Cancer Research* vol. 117 37–58 (Elsevier, 2013).
219. Galon, J. *et al.* Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* **313**, 1960–1964 (2006).
220. Mlecnik, B. *et al.* Multicenter International Society for Immunotherapy of Cancer Study of the Consensus Immunoscore for the Prediction of Survival and Response to Chemotherapy in Stage III Colon Cancer. *JCO* **38**, 3638–3651 (2020).
221. Marliot, F. *et al.* Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer. *J Immunother Cancer* **8**, e000272 (2020).
222. Hijazi, A., Antoniotti, C., Cremolini, C. & Galon, J. Light on life: immunoscore immune-checkpoint, a predictor of immunotherapy response. *Onc Immunology* **12**, 2243169 (2023).
223. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree.
224. Eberlin, L. S. *et al.* Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging. *Proceedings of the National Academy of Sciences* **111**, 2436–2441 (2014).
225. Dijkstra, T. K. Ridge regression and its degrees of freedom. *Qual Quant* **48**, 3185–3193 (2014).
226. Cortes, C., Mohri, M. & Rostamizadeh, A. L2 Regularization for Learning Kernels. (2009).
227. Blanco, V. M. *et al.* Optical and nuclear imaging of glioblastoma with phosphatidylserine-targeted nanovesicles. *Oncotarget* **7**, 32866–32875 (2016).
228. Sakane, F., Hoshino, F., Ebina, M., Sakai, H. & Takahashi, D. The Roles of Diacylglycerol Kinase  $\alpha$  in Cancer Cell Proliferation and Apoptosis. *Cancers* **13**, 5190 (2021).
229. Chen, Z. *et al.* Cellular and molecular identity of tumor-associated macrophages in glioblastoma. *Cancer Res* **77**, 2266–2278 (2017).
230. Zhang, H. *et al.* The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Computational and Structural Biotechnology Journal* **19**, 4603–4618 (2021).
231. Vidyarthi, A. *et al.* Predominance of M2 macrophages in gliomas leads to the suppression of local and systemic immunity. *Cancer Immunol Immunother* **68**, 1995–2004 (2019).
232. Michiba, A. *et al.* Morphologic Analysis of M2 Macrophage in Glioblastoma: Involvement of Macrophage Extracellular Traps (METs). *Acta Histochem. Cytochem.* **55**, 111–118 (2022).
233. Projection de l’incidence et de la mortalité par cancer en France métropolitaine en 2017. Rapport technique.
234. Choi, J. *et al.* Comparison of endoscopic ultrasonography and conventional endoscopy for prediction of depth of tumor invasion in early gastric cancer. *Endoscopy* **42**, 705–713 (2010).
235. Jiang, Z. *et al.* Impact of Surgical Margin Status on Survival in Gastric Cancer: A Systematic Review and Meta-Analysis. *Cancer Control* **28**, 107327482110436 (2021).



236. Berlth, F. *et al.* Prognostic Impact of Frozen Section Investigation and Extent of Proximal Safety Margin in Gastric Cancer Resection. *Annals of Surgery* **272**, 871–878 (2020).
237. Drubay, V. *et al.* Poorly cohesive cells gastric carcinoma including signet-ring cell cancer: Updated review of definition, classification and therapeutic management. *WJGO* **14**, 1406–1428 (2022).
238. Liu, W. *et al.* Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. *Nat Commun* **12**, 4961 (2021).
239. Kim, H. K. *et al.* Gastric Cancer-Specific Protein Profile Identified Using Endoscopic Biopsy Samples via MALDI Mass Spectrometry. *J. Proteome Res.* **9**, 4123–4130 (2010).
240. Smith, A. *et al.* Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry Imaging in the Study of Gastric Cancer: A Mini Review. *IJMS* **18**, 2588 (2017).
241. Eberlin, L. S. *et al.* Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2436–2441 (2014).
242. Uehara, T. *et al.* Overexpression of Lysophosphatidylcholine Acyltransferase 1 and Concomitant Lipid Alterations in Gastric Cancer. *Ann Surg Oncol* **23**, 206–213 (2016).
243. Niu, P.-H., Zhao, L.-L., Wu, H.-L., Zhao, D.-B. & Chen, Y.-T. Artificial intelligence in gastric cancer: Application and future perspectives. *WJG* **26**, 5408–5419 (2020).
244. Liu, D.-Y. *et al.* Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process. *Medical Image Analysis* **32**, 281–294 (2016).
245. Sakai, Y. *et al.* Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 4138–4141 (IEEE, Honolulu, HI, 2018). doi:10.1109/EMBC.2018.8513274.
246. Guimarães, P., Keller, A., Fehlmann, T., Lammert, F. & Casper, M. Deep-learning based detection of gastric precancerous conditions. *Gut* **69**, 4–6 (2020).
247. Li, L. *et al.* Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging. *Gastric Cancer* **23**, 126–132 (2020).
248. Duraipandian, S. *et al.* Real-time Raman spectroscopy for in vivo, online gastric cancer diagnosis during clinical endoscopic examination. *J. Biomed. Opt.* **17**, 1 (2012).
249. Balog, J. *et al.* In Vivo Endoscopic Tissue Identification by Rapid Evaporative Ionization Mass Spectrometry (REIMS). *Angew. Chem. Int. Ed.* **54**, 11059–11062 (2015).
250. Rodriguez, J. P. M. *et al.* Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: A systematic review. *Journal of Pathology Informatics* **13**, 100138 (2022).
251. Garcia, E. *et al.* Automatic Lymphocyte Detection on Gastric Cancer IHC Images Using Deep Learning. in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)* 200–204 (IEEE, Thessaloniki, 2017). doi:10.1109/CBMS.2017.94.
252. Sun, M. *et al.* Accurate Gastric Cancer Segmentation in Digital Pathology Images Using Deformable Convolution and Multi-Scale Embedding Networks. *IEEE Access* **7**, 75530–75541 (2019).
253. Huang, Z. *et al.* Retrospective imaging studies of gastric cancer: Study protocol clinical trial (SPIRIT Compliant). *Medicine* **99**, e19157 (2020).
254. Gao, Y. *et al.* Deep neural network-assisted computed tomography diagnosis of metastatic lymph nodes from gastric cancer. *Chinese Medical Journal* **132**, 2804–2811 (2019).
255. Tyanova, S. *et al.* Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**, 1453–1456 (2015).
256. Cox, J. *et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
257. Alexandrov, T. *et al.* Spatial Segmentation of Imaging Mass Spectrometry Data with Edge-Preserving Image Denoising and Clustering. *J. Proteome Res.* **9**, 6535–6546 (2010).

258. Rajilic-Stojanovic, M. *et al.* Systematic review: gastric microbiota in health and disease. *Aliment Pharmacol Ther* **51**, 582–602 (2020).
259. Wong, S. H. & Yu, J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol* **16**, 690–704 (2019).
260. Zi, M. *et al.* A literature review on the potential clinical implications of streptococci in gastric cancer. *Front. Microbiol.* **13**, 1010465 (2022).
261. Lertpiriyapong, K. *et al.* Gastric colonisation with a restricted commensal microbiota replicates the promotion of neoplastic lesions by diverse intestinal microbiota in the *Helicobacter pylori* INS-GAS mouse model of gastric carcinogenesis. *Gut* **63**, 54–63 (2014).
262. Plottel, C. S. & Blaser, M. J. Microbiome and Malignancy. *Cell Host & Microbe* **10**, 324–335 (2011).
263. Liang, Z. *et al.* Hesperidin Reversed Long-Term N-methyl-N-nitro-N-Nitroguanidine Exposure Induced EMT and Cell Proliferation by Activating Autophagy in Gastric Tissues of Rats. *Nutrients* **14**, 5281 (2022).
264. Ravegnini, G. *et al.* Gastric Adenocarcinomas and Signet-Ring Cell Carcinoma: Unraveling Gastric Cancer Complexity through Microbiome Analysis—Deepening Heterogeneity for a Personalized Therapy. *IJMS* **21**, 9735 (2020).
265. Zhang, L.-S., Liang, S., Zong, M.-H., Yang, J.-G. & Lou, W.-Y. Microbial synthesis of functional odd-chain fatty acids: a review. *World J Microbiol Biotechnol* **36**, 35 (2020).
266. Zhang, Y. *et al.* Gut microbiome analysis as a predictive marker for the gastric cancer patients. *Appl Microbiol Biotechnol* **105**, 803–814 (2021).
267. Peng, F. Dissecting characteristics and dynamics of differentially expressed proteins during multistage carcinogenesis of human colorectal cancer. *WJG* **22**, 4515 (2016).
268. Hashimoto, I. *et al.* Identification of a Biomarker Combination for Survival Stratification in pStage II/III Gastric Cancer after Curative Resection. *Cancers* **14**, 4427 (2022).
269. Li, Y. *et al.* Latexin expression is downregulated in human gastric carcinomas and exhibits tumor suppressor potential. *BMC Cancer* **11**, 121 (2011).
270. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **47**, D607–D613 (2019).
271. Anderson, N. M., Mucka, P., Kern, J. G. & Feng, H. The emerging role and targetability of the TCA cycle in cancer metabolism. *Protein Cell* **9**, 216–237 (2018).
272. Wang, P. *et al.* Decreased expression of the mitochondrial metabolic enzyme aconitase (ACO2) is associated with poor prognosis in gastric cancer. *Med Oncol* **30**, 552 (2013).
273. Ciccarone, F. *et al.* Aconitase 2 inhibits the proliferation of MCF-7 cells promoting mitochondrial oxidative metabolism and ROS/FoxO1-mediated autophagic response. *Br J Cancer* **122**, 182–193 (2020).
274. Kang, D., Lee, Y. & Lee, J.-S. RNA-Binding Proteins in Cancer: Functional and Therapeutic Perspectives. *Cancers* **12**, 2699 (2020).
275. Jiang, Z.-N., Ahmed, S. M. U., Wang, Q.-C., Shi, H.-F. & Tang, X.-W. Quinone oxidoreductase 1 is overexpressed in gastric cancer and associated with outcome of adjuvant chemotherapy and survival. *WJG* **27**, 3085–3096 (2021).
276. Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* **44**, 53–57 (2012).
277. Cheng, S. *et al.* A functional network of gastric-cancer-associated splicing events controlled by dysregulated splicing factors. *NAR Genomics and Bioinformatics* **2**, lqaa013 (2020).
278. Delcourt, V. *et al.* The Protein Coded by a Short Open Reading Frame, Not by the Annotated Coding Sequence, Is the Main Gene Product of the Dual-Coding Gene MIEF1. *Molecular & Cellular Proteomics* **17**, 2402–2411 (2018).
279. Vanderperre, B. *et al.* Direct Detection of Alternative Open Reading Frames Translation Products in Human Significantly Expands the Proteome. *PLoS ONE* **8**, e70698 (2013).

280. Brunet, M. A. *et al.* OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Research* **49**, D380–D388 (2021).
281. Koga, Y. Microbiota in the stomach and application of probiotics to gastroduodenal diseases. *World J Gastroenterol* **28**, 6702–6715 (2022).
282. Li, Z.-P. *et al.* Overgrowth of *Lactobacillus* in gastric cancer. *WJGO* **13**, 1099–1108 (2021).
283. Zuo, F., Appaswamy, A., Gebremariam, H. G. & Jonsson, A.-B. Role of Sortase A in *Lactobacillus gasseri* Kx110A1 Adhesion to Gastric Epithelial Cells and Competitive Exclusion of *Helicobacter pylori*. *Front. Microbiol.* **10**, 2770 (2019).
284. Yarmohammadi, M., Yadegar, A., Ebrahimi, M. T. & Zali, M. R. Effects of a Potential Probiotic Strain *Lactobacillus gasseri* ATCC 33323 on *Helicobacter pylori*-Induced Inflammatory Response and Gene Expression in Coinfected Gastric Epithelial Cells. *Probiotics & Antimicro. Prot.* **13**, 751–764 (2021).
285. Vinasco, K., Mitchell, H. M., Kaakoush, N. O. & Castaño-Rodríguez, N. Microbial carcinogenesis: Lactic acid bacteria in gastric cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1872**, 188309 (2019).
286. Sonveaux, P. *et al.* Targeting the Lactate Transporter MCT1 in Endothelial Cells Inhibits Lactate-Induced HIF-1 Activation and Tumor Angiogenesis. *PLoS ONE* **7**, e33418 (2012).
287. Dai, D. *et al.* Interactions between gastric microbiota and metabolites in gastric cancer. *Cell Death Dis* **12**, 1104 (2021).
288. Coker, O. O. *et al.* Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut* **67**, 1024–1032 (2018).
289. Yu, D. *et al.* Fecal *Streptococcus* Alteration Is Associated with Gastric Cancer Occurrence and Liver Metastasis. *mBio* **12**, e02994-21 (2021).
290. Tomizawa, H. *et al.* Galectin-10 as a Potential Biomarker for Eosinophilic Diseases. *Biomolecules* **12**, 1385 (2022).
291. Liu, D., Zhu, H. & Li, C. Galectins and galectin-mediated autophagy regulation: new insights into targeted cancer therapy. *Biomark Res* **11**, 22 (2023).
292. Wang, Y., Ye, W., Tian, G. & Zhang, Y. Identification of a new RNA-binding proteins-based signature for prognostic prediction in gastric cancer. *Medicine* **101**, e28901 (2022).
293. Murphy, A. J., Li, A. H., Li, P. & Sun, H. Therapeutic Targeting of Alternative Splicing: A New Frontier in Cancer Treatment. *Front. Oncol.* **12**, 868664 (2022).
294. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J Clinicians* **71**, 209–249 (2021).
295. Baran, B. *et al.* Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterol Res* **11**, 264–273 (2018).
296. Sagaert, X., Vanstapel, A. & Verbeek, S. Tumor Heterogeneity in Colorectal Cancer: What Do We Know So Far? *Pathobiology* **85**, 72–84 (2018).
297. Galeano Niño, J. L. *et al.* Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* **611**, 810–817 (2022).
298. Snaebjornsson, M. T., Janaki-Raman, S. & Schulze, A. Greasing the Wheels of the Cancer Machine: The Role of Lipid Metabolism in Cancer. *Cell Metabolism* **31**, 62–76 (2020).
299. Zhao, S.-S. *et al.* Altered Gut Microbial Profile Accompanied by Abnormal Fatty Acid Metabolism Activity Exacerbates Endometrial Cancer Progression. *Microbiol Spectr* **10**, e02612-22 (2022).
300. Dalmaso, G., Cougnoux, A., Delmas, J., Darfeuille-Michaud, A. & Bonnet, R. The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. *Gut Microbes* **5**, 675–680 (2014).
301. Faïs, T., Delmas, J., Barnich, N., Bonnet, R. & Dalmaso, G. Colibactin: More Than a New Bacterial Toxin. *Toxins* **10**, 151 (2018).
302. Nougayrède, J.-P. *et al.* *Escherichia coli* Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science, New Series* **313**, 848–851 (2006).



303. Wilson, M. R. *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).
304. Dziubańska-Kusibab, P. J. *et al.* Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med* **26**, 1063–1069 (2020).
305. Lopès, A. *et al.* Colibactin-positive *ESCHERICHIA COLI* induce a procarcinogenic immune environment leading to immunotherapy resistance in colorectal cancer. *Intl Journal of Cancer* **146**, 3147–3159 (2020).
306. Coughnoux, A. *et al.* Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **63**, 1932–1942 (2014).
307. Petan, T., Jarc, E. & Jusović, M. Lipid Droplets in Cancer: Guardians of Fat in a Stressful World. *Molecules* **23**, 1941 (2018).
308. Molinari, C. *et al.* Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *IJMS* **19**, 3733 (2018).
309. Zheng, Z. *et al.* Intratumor heterogeneity: A new perspective on colorectal cancer research. *Cancer Medicine* **9**, 7637–7645 (2020).
310. Gilson, P., Merlin, J.-L. & Harlé, A. Deciphering Tumour Heterogeneity: From Tissue to Liquid Biopsy. *Cancers* **14**, 1384 (2022).
311. Sadrekarimi, H. *et al.* Emerging role of human microbiome in cancer development and response to therapy: special focus on intestinal microflora. *J Transl Med* **20**, 301 (2022).
312. Bonnet, M. *et al.* Colonization of the Human Gut by *E. coli* and Colorectal Cancer Risk. *Clinical Cancer Research* **20**, 859–867 (2014).
313. Buc, E. *et al.* High Prevalence of Mucosa-Associated *E. coli* Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS ONE* **8**, e56964 (2013).
314. Pleguezuelos-Manzano, C. *et al.* Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli*. *Nature* **580**, 269–273 (2020).
315. Inamura, K. Colorectal Cancers: An Update on Their Molecular Pathology. *Cancers* **10**, 26 (2018).
316. Vasseur, S. & Guillaumond, F. Lipids in cancer: a global view of the contribution of lipid pathways to metastatic formation and treatment resistance. *Oncogenesis* **11**, 46 (2022).
317. Bellon, E. *et al.* Increased lysophosphatidylcholine acyltransferase 1 expression is unrelated to prognosis of esophageal cancer patients. *J Cancer Res Clin Oncol* **147**, 2879–2884 (2021).
318. Kurabe, N. *et al.* Accumulated phosphatidylcholine (16:0/16:1) in human colorectal cancer; possible involvement of LPCAT4. *Cancer Sci* **104**, 1295–1302 (2013).
319. Wang, L. *et al.* The regenerating protein 3A: a crucial molecular with dual roles in cancer. *Mol Biol Rep* **49**, 1491–1500 (2022).
320. Li, L. *et al.* LPCAT1 acts as an independent prognostic biomarker correlated with immune infiltration in hepatocellular carcinoma. *Eur J Med Res* **27**, 216 (2022).
321. Moessinger, C., Kuerschner, L., Spandl, J., Shevchenko, A. & Thiele, C. Human Lysophosphatidylcholine Acyltransferases 1 and 2 Are Located in Lipid Droplets Where They Catalyze the Formation of Phosphatidylcholine. *Journal of Biological Chemistry* **286**, 21330–21339 (2011).



Cite this: DOI: 10.1039/d3an01096a

Received 30th June 2023,  
 Accepted 15th September 2023  
 DOI: 10.1039/d3an01096a

rsc.li/analyst

## Comparing MS imaging of lipids by WALDI and MALDI: two technologies for evaluating a common ground truth in MS imaging†

Léa Ledoux,<sup>a</sup> Yanis Zirem,<sup>a</sup> Florence Renaud,<sup>b</sup> Ludovic Duponchel,<sup>c</sup> Michel Salzet,<sup>a,d</sup> Nina Ogrinc<sup>†‡</sup> and Isabelle Fournier<sup>†‡</sup>

In this study, we conducted a direct comparison of water-assisted laser desorption ionization (WALDI) and matrix-assisted laser desorption ionization (MALDI) mass spectrometry imaging, with MALDI serving as the benchmark for label-free molecular tissue analysis in biomedical research. Specifically, we investigated the lipidomic profiles of several biological samples and calculated the similarity of detected peaks and Pearson's correlation of spectral profile intensities between the two techniques. We show that, overall, MALDI MS and WALDI MS present very close lipidomic analyses and that the highest similarity is obtained for the norharmane MALDI matrix. Indeed, for norharmane in negative ion mode, the lipidomic spectra revealed 100% similarity of detected peaks and over 0.90 intensity correlation between both technologies for five samples. The MALDI-MSI positive ion lipid spectra displayed more than 83% similarity of detected peaks compared to those of WALDI-MSI. However, we observed a lower percentage (77%) of detected peaks when comparing WALDI-MSI with MALDI-MSI due to the rich WALDI-MSI lipid spectra. Despite this difference, the global lipidomic spectra showed high consistency between the two technologies, indicating that they are governed by similar processes. Thanks to this similarity, we can increase datasets by including data from both modalities to either co-train classification models or obtain cross-interrogation.

Mass spectrometry imaging (MSI) is a robust label-free molecular technique that enables the distribution of a diverse range of endogenous and exogenous molecules in biological tissues without compromising sample integrity or morphology.<sup>1</sup> Due to years of instrumentation, sample preparation and bioinformatics advances, MSI has been developed widely

and its range of applications broadened in the biomedical and pharmaceutical fields.<sup>2</sup> In particular, due to a wide range of accessible analytes, matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI-MSI) has demonstrated its vast potential for clinical applications including biomarker discovery, diagnostics, prognosis and patient stratification.<sup>3</sup> While MALDI-MSI remains the most popular and democratized MSI technique, it requires sample preparation that is not trivial and time-consuming especially for proteins, peptides and glycans. Analysis of lipids, particularly glycerophospholipids (GPLs), present in tissues, has gradually gained importance in the course of the last 15 years and lipids are now recognized as important mediators in pathophysiological mechanisms.<sup>4</sup> In this context, MALDI-MSI has quickly emerged as one of the universal tools for studying lipid biochemistry across tissue sections.<sup>5,6</sup> Many lipids are readily detected due to their higher abundance in cells. The variations of lipid composition within tissues was already being studied in oncology<sup>7,8</sup> and neurodegenerative diseases,<sup>9,10</sup> among others. As the distribution and abundance of metabolites and lipids follow the changes of cell phenotypes, lipid MSI has gradually emerged as an alternative to histopathology.<sup>11,12</sup> Several ambient ionization mass spectrometry (AIMS) technologies have become widespread for imaging samples under their near-native conditions since they require minimal sample preparation for analysis, unlike traditional vacuum-based MS approaches. The most commonly used techniques include desorption electrospray ionization (DESI),<sup>13</sup> atmospheric pressure matrix-assisted laser desorption ionization (AP-MALDI),<sup>14,15</sup> laser-ablation electrospray ionization (LAESI)<sup>16–18</sup> or matrix-assisted laser desorption electrospray ionization (MALDESI).<sup>19</sup>

SpiderMass AIMS employs contactless micro-invasive analysis, using a laser desorption/ionization process in the mid-infrared range at 2.94  $\mu\text{m}$  to excite the strongest vibrational bond of water molecules (O–H stretching).<sup>20</sup> This enables a MALDI-like process, utilizing endogenous water as the MALDI matrix; hence the name water-assisted laser desorption ionization (WALDI).<sup>20,21</sup> Additionally, the system is designed to

<sup>a</sup>Univ. Lille, Inserm, CHU Lille, U1192 – Protéomique Réponse Inflammatoire Spectrométrie de Masse – PRISM, F-59000 Lille, France.

E-mail: isabelle.fournier@univ-lille.fr

<sup>b</sup>Unité mixte SIRIC CURAMUS et 938, Inserm. Université Paris Sorbonne, France

<sup>c</sup>Univ. Lille, CNRS, UMR 8516 – LASIRE, F-59000 Lille, France

<sup>d</sup>Institut Universitaire de France (IUF), Paris, France

†Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3an01096a>

‡These authors contributed equally to this work.



enable remote analysis of desorbed materials, which is achieved through aspiration using a tubing line connected directly to the MS instrument that creates differential pressure. Like other AIMS technologies, SpiderMass provides direct analysis of metabolites and lipids, and has also been demonstrated on protein standards.<sup>22</sup> The system has already been applied in different fields, including *ex vivo* analysis of glioblastoma<sup>3</sup> and oral squamous cell carcinoma,<sup>23</sup> and *in vivo* analysis of human skin<sup>20</sup> as well as cultured cells without preparation.<sup>24</sup> To enable MSI, the system was recently coupled to a robotic arm equipped with a distance sensor.<sup>25</sup> This way, the WALDI-MSI can provide 2D as well as 3D molecular topographic images. Interestingly, because the laser microprobe is moved above the sample surface, both *ex vivo* imaging of flat tissue sections and *in vivo* imaging of human tissues are possible.<sup>25</sup> The molecular profiles from WALDI-MS reveal undeniable resemblance and similar characteristics to those observed during MALDI-MS processes. Similar collective behaviour to MALDI was previously observed in WALDI while tuning the laser wavelength to achieve resonant excitation.<sup>22</sup> In this study, we aimed to compare WALDI-MSI and MALDI-MSI by calculating the similarity of detected peaks and correlation indexes of lipidomic profiles from several biological samples, including those from rat brain and human gastric cancer.

## Individual assessment of peak similarity and intensity correlations of MALDI-MSI and WALDI-MSI lipidomic profiles

First, we evaluated the effectiveness of our newly developed correlation and similarity calculations (ESI<sup>+</sup>) by examining triplicates of MALDI and WALDI MSI analyses of rat brain (RB) tissues in both positive and negative ion modes. In both ion modes, 100% of the same peaks were detected in each replicate for MALDI (M) and WALDI (W) (ESI<sup>+</sup>). The intensity correlations were well above 94% with a coefficient of variance no higher than 3% between the MALDI data sets in both modes (ESI<sup>+</sup>). A slightly higher intensity correlation, above 97%, was obtained for WALDI-MSI triplicates with a coefficient of variance no higher than 1%. Regardless of different anatomical regions of the rat brain tissue sections, the techniques display high analytical and molecular reproducibility. The obtained results, therefore, highlight the robustness and the precision of our methods for further evaluation of the correlation and similarity between the two technologies.

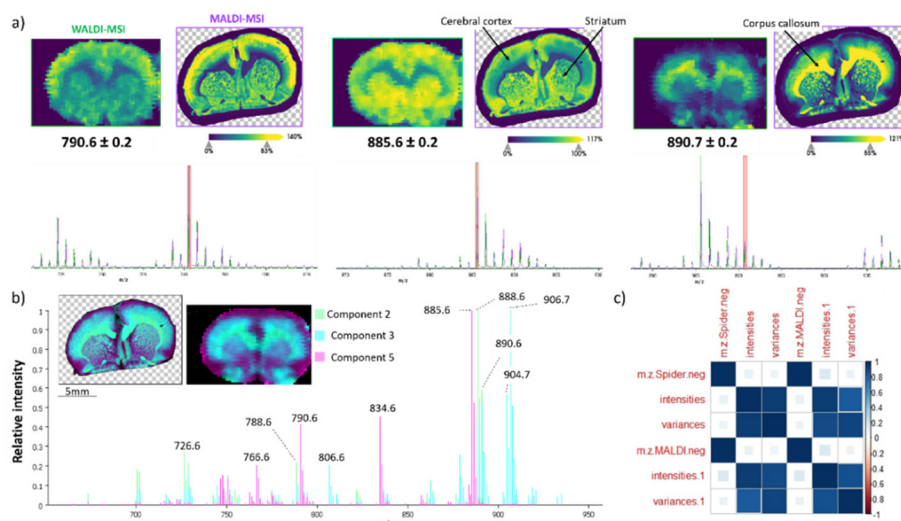
## Similarity of peak detection and correlation of intensities and variances between MALDI-MSI and WALDI-MSI

In WALDI, endogenous water plays the role of the matrix, but in MALDI, it is well known that a different family or subfamily

of molecules are better detected by changing the matrix. Therefore, the experiments were carried out using the four most commonly utilized matrices for lipid MALDI MSI analysis, namely norharmane, 9-AA, 2,5-DHB, and DAN. The glycerophospholipid species are typically observed in the 500–1000 mass range with exceptions of lyso-glycerophospholipids. 2,5-DHB analysis was only performed in the positive ion mode and conversely 9-AA only in the negative ion mode, since these two matrixes are known to be more specific to one of the two modes.<sup>26,27</sup> The norharmane matrix negative ion mode spectrum displayed high overlap with WALDI-MSI in the lipid 600–1000 *m/z* range (Fig. 1). This is further supported by the zoom and selected ion images at *m/z* 790.6 [PE (18:0\_22:6)-H]<sup>-</sup>, *m/z* 885.6 [PI (18:0\_20:4)-H]<sup>-</sup> and *m/z* 890.7 [SHexCer (d18:1/24:0)-H]<sup>-</sup> (Fig. 1a). pLSA was used to further elucidate the spectral signals resulting from a particular spatial feature within the two brain sections (Fig. 1b). The pLSA components 2, 3 and 5 outline only the major contributors to the grey and white matter by both modalities. The same 3 clusters are observed with bisecting *k*-means segmentation (ESI<sup>+</sup>). The ions at *m/z* 885.6, 888.6 [SHexCer (d18:1/24:1)-H]<sup>-</sup> and *m/z* 906.7 [SHexCer (42:1;3)-H]<sup>-</sup> display the highest contribution to each component. There was no dissimilarity of peaks detected and the intensity correlations were well above 91% and above 88% for variance (Fig. 1c and ESI<sup>+</sup>) with only 2 and 6% coefficient of variance between the triplicates. The 9-AA matrix revealed 84% peak similarity for MALDI vs. WALDI while only 48% peak similarity of WALDI compared to MALDI (ESI<sup>+</sup>). This is in line with the 9-AA matrix being conventionally used for metabolite imaging in the lower mass range. The DAN matrix showed 77% peak similarity and only 7% intensity correlation (ESI<sup>+</sup>).

In positive ion mode, the WALDI-MSI spectrum shows strong visual differences compared to other MALDI matrices. It exhibits more diverse lipid content particularly in the 600–700 *m/z* range (ESI<sup>+</sup>). This is also confirmed by the similarity calculations with a low of 77% for norharmane, 51% for DHB and 75% for DAN (ESI<sup>+</sup>). The dissimilarity could be due to several additional diglycerides found in the WALDI-MSI spectrum. DGs are natively present in the tissue; however, the specific mass range can also indicate fragments from fatty acid neutral losses of PCs or TGs. To refute this assumption, a PC standard (18:1/18:1) was analysed by MALDI and by WALDI. The spectrum showed no fragmentation with both technologies (ESI<sup>+</sup>). It has already been demonstrated that DHB displays lower lipid coverage than norharmane<sup>28</sup> and consequently has lower similarity index in rat brain samples. The sensitivity and the molecular coverage achievable by MALDI-MSI are ultimately governed by the ionization efficiencies (fraction of ions compared to the desorbed neutrals), which is relatively low and differs with each matrix. Despite this fact, the data clearly demonstrate that over 84.52% of all lipid peaks detected by MALDI-MSI are also detected by WALDI-MSI (ESI<sup>+</sup>) with norharmane being the most comparable matrix in both positive and negative ion modes.<sup>28,29</sup>





**Fig. 1** WALDI-MSI and MALDI-MSI (norharmane matrix) of M1 and W1 brain tissue in the negative ion mode. (a) Selected ion images and spectral zoom at  $m/z$  790.6, 885.6 and 890.7. (b) Probabilistic latent semantic analysis (pLSA) loading plots and spectra of the resulting components in the negative ion mode (components 2, 3 and 5) for WALDI-MSI and MALDI-MSI. The loading spectra are annotated with the ions that show the highest contribution to each component. (c) Pearson's intensity correlation matrix for intensities and variances of the WALDI-MSI and MALDI-MSI data.

## Effects of the laser spot size and tissue thickness on molecular profiles

So far, the two technologies cannot be performed with the same laser spot size with WALDI-MSI resolution limited to the incident laser beam diameter ( $\sim 400$ – $500$   $\mu\text{m}$ ). Thus, it could be speculated that the comparison between MALDI-MSI and WALDI-MSI may be affected by differences in the amount of analysed materials. To assess the impact of laser spot size on similarity and correlation calculations, MALDI-MSI analysis was conducted at a reduced spatial resolution of  $150$   $\mu\text{m}$ . While there is 100% peak similarity, there is only a 2% difference in the intensity correlations between MALDI- and WALDI-MSI when MALDI data are collected at either  $150$   $\mu\text{m}$  or  $50$   $\mu\text{m}$  spatial resolution (ESI $^\dagger$ ). Similar findings are observed when comparing different thicknesses of the tissue. The peak similarity is equal and intensity correlations are above 92% between  $12$   $\mu\text{m}$  and  $20$   $\mu\text{m}$  tissue thicknesses (ESI $^\dagger$ ). According to this, the spatial resolution and tissue thickness have only a limited influence on the peak similarity and the correlation of intensities and variances between the two imaging modalities.

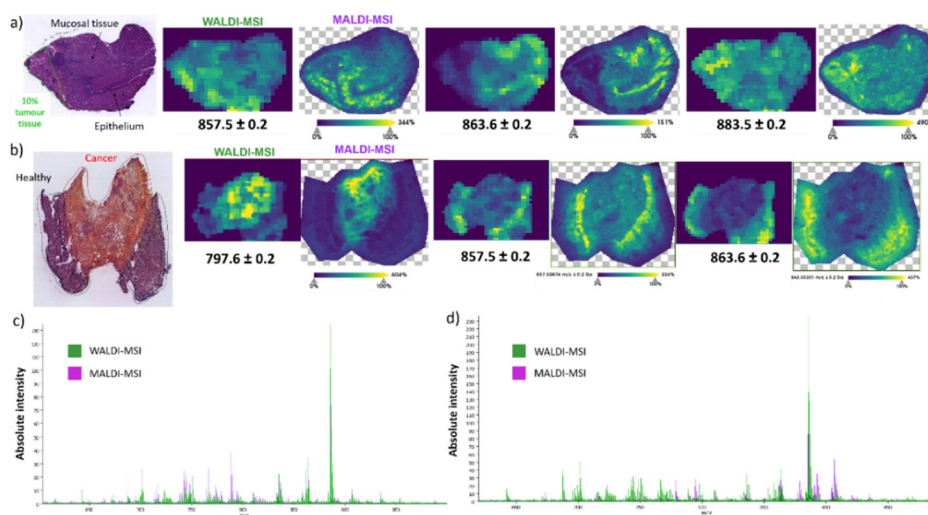
## Comparative analysis and the interrogation model for gastric cancer

While rat brain tissue sections serve effectively as the standard for MALDI and AIMS MSI, we also wanted to confirm our hypothesis for other biological tissues, such as human gastric cancer. Building on the above, we exclusively used the norharmane matrix since it displayed the highest similarity to the WALDI-MSI analysis. Two heterogeneous tissue sections of

poorly cohesive esogastric carcinoma (PCC) were used for analysis. The first tissue section is composed of healthy tissue with 10% tumor as indicated in the optical image (Fig. 2a). Visual comparison of the selected ion images (Fig. 2a) and spectra (Fig. 2c) shows high similarity between the WALDI- and MALDI-MSI. Specifically, ions at  $m/z$  857.5 [PI (36:4)-H] $^-$ ,  $m/z$  863.6 [PI (18:1\_18:0)-H] $^-$  and  $m/z$  883.5 [PI (18:1\_20:4)-H] $^-$  reveal not only the same distribution in the two subzones but also localized sub-features in the tissue. Indeed, the ion at  $m/z$  863.6 is specific to the epithelium, in contrast to the ion at  $m/z$  857.5, which is preferentially observed in mucosal tissue. The extracted overview spectra revealed a 100% peak similarity, 91% correlation of intensities and 75% correlation of variances (ESI $^\dagger$ ). Furthermore, exploration of the imaging data by ROC analysis (AUC above 0.85 thresholds) showed 2 discriminatory peaks in MALDI-MSI and 5 discriminatory peaks in WALDI-MSI. The peaks at  $m/z$  766.5 [PE (18:0\_20:4)-H] $^-$  and  $m/z$  788.6 [PS (18:1\_18:0)-H] $^-$  were only found with lower intensities in the WALDI data (ESI $^\dagger$ ). On the other hand, peaks at  $m/z$  642.5 [HexCer 30:1;2-H] $^-$  and  $m/z$  687.6 [PA (P-20:0\_16:0) or (O-18:0\_18:1)-H] $^-$  were not found in the MALDI-MSI data while those at  $m/z$  701.5 [PA (18:1\_18:0)-H] $^-$ ,  $m/z$  750.5 [PE (16:0\_20:4)-H] $^-$  and  $m/z$  853.6 [PI (36:6)-H] $^-$  were found but only with lower intensities (ESI $^\dagger$ ). The second-analysed gastric cancer tissue section shows important tumoral and peritumoral areas (Fig. 2b). Similarly, in the other tissue section, the selected ion distributions are uniform in both the WALDI and MALDI images (Fig. 2b). Indeed, ions at  $m/z$  857.5 and  $m/z$  863.6 are specific to the healthy tissue, in contrast to the ion at  $m/z$  797.7, which is attributed to [PG (18:1\_20:3)-H] $^-$  or [PG (18:0\_20:4)-H] $^-$  and is only present in the cancerous tissue. The overview spectra, however, do reveal more pronounced differences in intensities (Fig. 2d). Although there is 100%





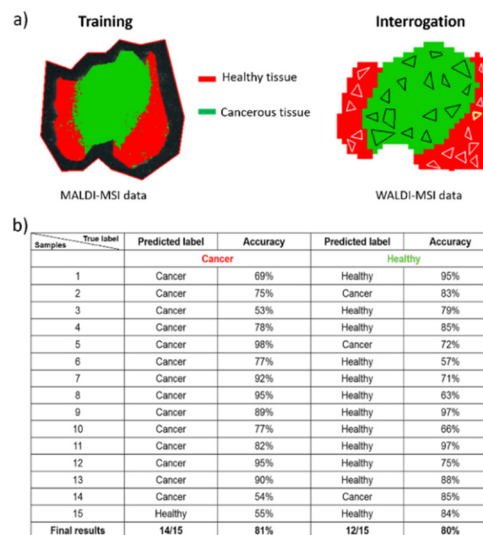


**Fig. 2** WALDI-MSI and MALDI-MSI (norharmane matrix) gastric cancer imaging data in the negative ion mode. (a) Optical image of the HPS-stained gastric tissue section A with the annotated cancer region, mucosal tissue and epithelium with the selected ion images at  $m/z$  857.5, 863.6 and 883.5. (b) Optical image of the HPS-stained gastric tissue section B with the annotated cancer region and two healthy regions, accompanied to the right by the selected ion and spectral zoom images at  $m/z$  797.6, 857.5 and 863.5. (c and d) WALDI-MSI (green) and MALDI-MSI (purple) overview spectra of (c) gastric tissue section A and (d) gastric tissue section B.

similarity of the detected peaks, there is only 77% correlation of intensities and 67% correlation of variances ( $ESI^\dagger$ ).

Finally, we assessed whether the high similarity enables one to train with MALDI data and then interrogate the WALDI data against the MALDI training. For this purpose, a two-class CNN model was built with MALDI-MSI data that is able to distinguish the cancerous tissue and the healthy tissue of gastric tissue section B ( $ESI^\dagger$ ). Then, thanks to this, the label of 30 random samples from WALDI-MSI data was predicted. As shown in Fig. 3, we obtained good classification with 93% prediction of cancerous tissue with 81% mean accuracy and 80% prediction of healthy tissue with a mean accuracy of 80%. This confirmed that MALDI-MSI and WALDI-MSI data show the same ground truth in MS imaging and that WALDI can be trained from MALDI data.

In summary, we have demonstrated high resemblance between the newly developed WALDI-MSI and MALDI-MSI for lipi-domic analysis. Our findings show not only the detection of the same species but also a strong correlation between the spectral intensities obtained through both modalities. This correlation is particularly notable in the negative ion mode and when using norharmane<sup>25,26</sup> as the MALDI matrix. Despite differences in resolution between the two techniques, we were able to identify the same tissue heterogeneity. In gastric cancer, it was possible to differentiate not only the tumoral and peritumoral regions but also the dense cancer cells, epithelium, and mucosal tissue. In the future, we intend to further increase the spatial resolution to 100  $\mu\text{m}$  through mechanical improvements. Our data clearly demonstrate that the WALDI-MSI technique is a promising new ambient ionization imaging method. Moving forward, we intend to leverage MALDI-MSI data to expedite and enhance WALDI-MSI training, particularly in the context of data intensive supervised machine learning.



**Fig. 3** CNN model based on MALDI-MSI data with WALDI-MSI data for the prediction of cancerous and healthy tissues from gastric tissue section B. (a) Regions of interest based on segmentation used for the training (MALDI-MSI data) and the interrogation (WALDI-MSI data) of the model. (b) Predicted labels and their accuracies obtained for 30 random samples from WALDI-MSI data (black and white triangles).

## Author contributions

L.L., I.F., N.O. and Y.Z. wrote the manuscript; I.F. and N.O. designed the experiments. L.L. and Y.Z. performed the experiments. L.L. and Y.Z. analyzed the data. Y.Z. built the similarity and correlation code. F.R. provided the gastric cancer samples and HPS annotations. L.L., Y.Z., I.F., N.O., M.S. and L.D.



corrected the manuscript. I.F. and N.O. supervised the project. I.F. and M.S. provided the funding.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The project was supported by the Ministère de l'Enseignement Supérieur et de la Recherche (I.F.), Inserm (I.F.) and Excellence Initiative ISite ERC Generator (I.F.). The authors would like to thank the Centre for Studies and Research on Lasers and Applications (CERLA) for support.

## References

- 1 A. R. Buchberger, K. DeLaney, J. Johnson and L. Li, *Anal. Chem.*, 2018, **90**, 240–265.
- 2 P.-M. Vaysse, R. M. A. Heeren, T. Porta and B. Balluff, *Analyst*, 2017, **142**, 2690–2712.
- 3 M. Duhamel, L. Drelich, M. Wisztorski, S. Aboulouard, J.-P. Gimeno, N. Ogrinc, P. Devos, T. Cardon, M. Weller, F. Escande, F. Zairi, C.-A. Maurage, É. Le Rhun, I. Fournier and M. Salzset, *Nat. Commun.*, 2022, **13**, 6665.
- 4 X. Han, *Nat. Rev. Endocrinol.*, 2016, **12**, 668–679.
- 5 R. C. Murphy, J. A. Hankin and R. M. Barkley, *J. Lipid Res.*, 2009, **50**(Suppl), S317–S322.
- 6 A. P. Bowman, R. M. A. Heeren and S. R. Ellis, *TrAC, Trends Anal. Chem.*, 2019, **120**, 115197.
- 7 S. Guo, Y. Wang, D. Zhou and Z. Li, *Sci. Rep.*, 2014, **4**, 5959.
- 8 M. K. Andersen, T. S. Høiem, B. S. R. Claes, B. Balluff, M. Martin-Lorenzo, E. Richardsen, S. Krossa, H. Bertilsson, R. M. A. Heeren, M. B. Rye, G. F. Giskeødegård, T. F. Bathen and M.-B. Tessem, *Cancer Metab.*, 2021, **9**, 9.
- 9 K. Mallah, J. Quanico, A. Raffo-Romero, T. Cardon, S. Aboulouard, D. Devos, F. Kobeissy, K. Zibara, M. Salzset and I. Fournier, *Anal. Chem.*, 2019, **91**, 11879–11887.
- 10 I. A. Mulder, N. O. Potočnik, L. a. M. Broos, A. Prop, M. J. H. Wermer, R. M. A. Heeren and van den A. M. J. M. Maagdenberg, *Sci. Rep.*, 2019, **9**, 1–10.
- 11 R. Longuespée, R. Casadonte, M. Kriegsmann, C. Pottier, de G. P. Muller, P. Delvenne, J. Kriegsmann and E. D. Pauw, *Proteomics: Clin. Appl.*, 2016, **10**, 701–719.
- 12 M. Holzlechner, E. Eugenin and B. Prideaux, *Cancer Rep.*, 2019, **2**, e1229.
- 13 Z. Takáts, J. M. Wiseman, B. Gologan and R. G. Cooks, *Science*, 2004, **306**, 471–473.
- 14 V. V. Laiko, N. I. Taranenko, V. D. Berkout, M. A. Yakshin, C. R. Prasad, H. S. Lee and V. M. Doroshenko, *J. Am. Soc. Mass Spectrom.*, 2002, **13**, 354–361.
- 15 V. V. Laiko, M. A. Baldwin and A. L. Burlingame, *Anal. Chem.*, 2000, **72**, 652–657.
- 16 P. Nemes and A. Vertes, *J. Visualized Exp.*, 2010, e2097.
- 17 P. Nemes and A. Vertes, *Methods Mol. Biol.*, 2010, **656**, 159–171.
- 18 P. Nemes, A. S. Woods and A. Vertes, *Anal. Chem.*, 2010, **82**, 982–988.
- 19 J. S. Sampson, A. M. Hawkridge and D. C. Muddiman, *J. Am. Soc. Mass Spectrom.*, 2006, **17**, 1712–1716.
- 20 B. Fatou, P. Saudemont, E. Leblanc, D. Vinatier, V. Mesdag, M. Wisztorski, C. Focsa, M. Salzset, M. Ziskind and I. Fournier, *Sci. Rep.*, 2016, **6**, 25919.
- 21 N. Ogrinc, P. Saudemont, J. Balog, Y.-M. Robin, J.-P. Gimeno, Q. Pascal, D. Tierny, Z. Takats, M. Salzset and I. Fournier, *Nat. Protoc.*, 2019, **14**, 3162–3182.
- 22 B. Fatou, M. Ziskind, P. Saudemont, J. Quanico, C. Focsa, M. Salzset and I. Fournier, *Mol. Cell. Proteomics*, 2018, **17**, 1637–1649.
- 23 N. Ogrinc, C. Attencourt, E. Colin, A. Boudahi, R. Tebbakha, M. Salzset, S. Testelin, S. Dakpé and I. Fournier, *Front. Oral Health*, 2022, **3**, 827360.
- 24 P. Saudemont, J. Quanico, Y.-M. Robin, A. Baud, J. Balog, B. Fatou, D. Tierny, Q. Pascal, K. Minier, M. Pottier, C. Focsa, M. Ziskind, Z. Takats, M. Salzset and I. Fournier, *Cancer Cell*, 2018, **34**, 840–851.
- 25 N. Ogrinc, A. Kruszewski, P. Chaillou, P. Saudemont, C. Lagadec, M. Salzset, C. Duriez and I. Fournier, *Anal. Chem.*, 2021, **93**, 14383–14391.
- 26 W. J. Perry, N. H. Patterson, B. M. Prentice, E. K. Neumann, R. M. Caprioli and J. M. Spraggins, *J. Mass Spectrom.*, 2020, **55**, e4491.
- 27 M. R. Eveque-Mourroux, P. J. Emans, R. R. M. Zautsen, A. Boonen, R. M. A. Heeren and B. Cillero-Pastor, *Analyst*, 2019, **144**, 5953–5958.
- 28 T. B. Angerer, J. Bour, J.-L. Biagi, E. Moskovets and G. Frache, *J. Am. Soc. Mass Spectrom.*, 2022, **33**, 760–771.
- 29 S. R. Ellis, J. Cappell, N. O. Potočnik, B. Balluff, J. Hamaide, A. Van der Linden and R. M. A. Heeren, *Analyst*, 2016, **141**, 3832–3841.



# Real-time Glioblastoma Tumor MicroEnvironment Assessment by SpiderMass for Improved Patient Management

**Yanis Zirem<sup>1,9</sup>, Léa Ledoux<sup>1,9</sup>, Lucas Roussel<sup>1</sup>, Claude Alain Maurage<sup>2</sup>, Pierre Tirilly<sup>3</sup>, Émilie Le Rhun<sup>1,4</sup>, Bertrand Meresse<sup>5</sup>, Gargey Yagnik<sup>6</sup>, Mark J. Lim<sup>6</sup>, Kenneth J. Rothschild<sup>6,7</sup>, Marie Duhamel<sup>1</sup>, Michel Salzet<sup>1,8,9,\*</sup>, Isabelle Fournier<sup>1,8,9,10,\*</sup>**

<sup>1</sup>Univ. Lille, Inserm, CHU Lille, U1192 - Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France

<sup>2</sup>CHU Lille, Service de biochimie et biologie moléculaire, CHU Lille, F-59000 Lille, France

<sup>3</sup>Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

<sup>4</sup>Departments of Neurosurgery and Neurology, Clinical Neuroscience Center, University Hospital of Zurich, University of Zurich, Zurich, Switzerland

<sup>5</sup>Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000 Lille, France

<sup>6</sup>AmberGen, Inc., Billerica, MA, United States

<sup>7</sup>Department of Physics and Photonics Center, Boston University, Boston, MA, United States

<sup>8</sup>Institut Universitaire de France (IUF), Paris, France

<sup>9</sup>These authors contributed equally

<sup>10</sup>Lead contact

\*Correspondence : [michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr) (M.S.), [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr) (I.F.)



## Summary

Glioblastoma is a highly heterogeneous and infiltrative form of brain cancer associated with a poor outcome and limited therapeutic effectiveness. The extent of the surgery is related to survival. Reaching an accurate diagnosis and prognosis assessment by the time of the initial surgery is therefore paramount in the management of glioblastoma. To this end, we are studying the performance of SpiderMass, an ambient ionization mass spectrometry technology that can be used *in vivo* without invasiveness, coupled to our recently established artificial intelligence (AI) pipeline. We demonstrate that we can both stratify IDH-wild type glioblastoma patients into molecular sub-groups and achieve an accurate diagnosis with over 90% accuracy after cross-validation. Interestingly, the developed method offers the same accuracy for prognosis. Additionally, we are testing the potential of an immunoscore strategy based on SpiderMass fingerprints, showing the association between prognosis and immune cell infiltration, to predict patient outcome.

## Keywords

Mass Spectrometry • SpiderMass • Glioblastoma • Diagnosis • Lipids • Prognosis • Machine Learning • Immunoscore • Imaging • MALDI-IHC

## Introduction

With an incidence of 308,102 new cases and 251,329 deaths worldwide in 2020<sup>1</sup>, brain and central nervous system (CNS) tumors are among the deadliest cancer. Gliomas are the most commonly occurring tumors of the CNS, which account for almost 80% of all malignant primary tumors of brain<sup>2</sup>. Gliomas can originate from different cell types and their classification represents a challenge. However, precise classification of gliomas is paramount to estimate the patient prognosis and foresee the best possible care and personalized treatment options<sup>3</sup>. Since 2016 the World Health Organization (WHO) recommendation is to integrate histological features and molecular alterations for the classification of tumors of the central nervous system<sup>4</sup>. Yet, the classification is based on the fifth edition of the WHO Classification of Tumors of the CNS<sup>5</sup> which considers further the importance of the molecular features while keeping the histology. The first criterion of classification is the isocitrate dehydrogenase (IDH) mutation status which separates the glioma into the IDH-wild type versus the IDH1- or IDH2- mutants groups. IDH-mutant status is not only important for the classification but it is also associated with a better prognosis<sup>6</sup>. The tumors classified as IDH-wild type are further categorized into glioblastoma (GBM) and astrocytoma/not elsewhere classifiable. Within the adult-type diffuse gliomas, glioblastoma, histone-mutant gliomas, several rare entities is the most prevalent and aggressive form with an overall survival (OS) of about 15 months and a median of 14-24 months (as opposed to lower-grade gliomas which can be controlled for decades) and is classified as a high-grade malignant glioma (grade 4). While glioblastoma shares some astrocytic histological features with astrocytoma, it also shows its own histological features including a high cellular density, a marker nuclear atypia, a high mitotic activity and necrosis and/or microvascular proliferation, or at least one of the following specific molecular characteristics i.e., EGFR amplification, TERT promoter mutation, a concurrent gain of chromosome 7 and loss of chromosome 10 (+7/-10).

Glioblastoma is a highly heterogeneous disease, making difficult to develop effective treatments<sup>7,8</sup>. So far, management of IDH-wildtype glioblastoma involves a multimodal approach with maximal safe surgical resection surgery in frontline followed by temozolomide chemoradiotherapy and maintenance (known as Stupp protocol<sup>9</sup>). Treatment of glioblastoma clearly meets different challenges. Indeed, novel immunotherapies, vaccination or immune checkpoint inhibition, have shown disappointing results against GBM. One of the hurdles of glioblastoma treatment is associated to the blood-brain-barrier (BBB) which limits the passage of drugs into the brain, thus preventing the administration of many chemotherapies. Another difficulty is the heterogeneity of the tumor and signaling pathway associated which make targeting therapeutic goals more complicated<sup>10</sup> and also from the heterogeneity of the tumor; with the tumor microenvironment (TME) promoting the resistance and decreasing the efficiency of the patient immunological response. Indeed, the development of brain tumors disrupts the homeostasis of the CNS. The vasculature, the extracellular matrix and the BBB are hijacked and integrated into the processes of tumor development<sup>11</sup>. It is known that TME is immunosuppressive and that infiltration of different subpopulations of immune cells are correlated to both the response to therapy<sup>12</sup> and patient OS<sup>13</sup>.

On the other hand, the extent of resection (EOR) has been demonstrated to correlate with OS such that, higher OS is observed for gross total resection (GTR) over subtotal resection or biopsy<sup>14</sup>. In particular, increasing the EOR to >95% while sparing tissue to preserve functional integrity may improve surgical outcomes and OS<sup>3</sup>. The importance of the EOR is thus recognized and has been incorporated into the European guidelines for the management of patients with glioblastoma<sup>15</sup>. However, GTR remains difficult due to the diffuse nature of glioblastoma which makes the accurate delineation of the tumor margins an issue. Different tools have been developed to help with finding the resection margins<sup>16</sup>. Neuronavigation by magnetic resonance imaging (MRI) with addition of 5-Ala has become a standard procedure to visualize margins and was shown to enable better EOR. Awake surgery has also been shown to limit the damages to functional tissues while increasing the EOR. Despite margins being more distinguishable for glioblastoma, it is still infiltrating, hence awake surgery has proven its usefulness. More

recently, Raman spectroscopy combined with machine learning (ML) has been developed for intraoperative assessment of glioma. Raman was shown to discriminate tumor from non-tumor tissue<sup>17</sup> but also to distinguish between IDH-1 and -2 mutant from IDH-wild type. This opens the possibility of a tailored EOR based on the aggressivity of the glioma<sup>18</sup>. A handheld Raman probe is also available and was shown to achieve a sensitivity of 93% to classify normal brain from dense cancer<sup>19</sup>. Therefore, both a real-time molecular analysis of the brain tissues during the surgery to achieve tailored resection and a better understanding of the microenvironment, could become essential to improve the management of GBM patients<sup>20</sup>. Local therapies are increasingly seen as a unique option to deliver higher dose of therapeutics while limiting the adverse effects of systemic administration. In this context, there is a need for stratifying patients based on their molecular features at the time of the surgery and correlate this stratification with patient survival to provide the best therapeutic option and start the treatment as soon as possible.

GBM has been the subject of numerous omics studies using mass spectrometry (MS)-based approaches and specifically using either shot gun proteomics or mass spectrometry imaging (MSI). Different MSI modalities, notably MALDI and DESI, have been used to follow known markers (e.g., IDH-mutant associated metabolites 2-hydroxyglutarate)<sup>21</sup>, identify potential biomarkers and therapeutic targets for the disease<sup>22,23</sup>. However, more recently MS has also emerged as novel technology for intraoperative assessment thanks to the development of novel ambient ionization mass spectrometry (AIMS) techniques<sup>24</sup>. Intra-operative AIMS is achieved from freshly resected tissues<sup>25,26</sup> or *in vivo* to advise neurosurgeons on the extent of tumor resection, with real-time feedback<sup>27-29</sup>. SpiderMass is one of these AIMS technology that provides a minimally invasive solution for *in vivo* surgical applications<sup>30</sup>. Very interestingly the technology is micro-invasive thanks to the high absorption of endogeneous water present in the tissues at 2.94  $\mu\text{m}$  (most intense absorption band of water) leading to a low penetration depth of the laser beam. Thus, there is only a white trace of dehydration left by the system with a few  $\mu\text{m}$  depth of tissues remove. It is also a very efficient process that equally work for all types of tissues because the water content of tissues is  $>70\%$ <sup>27,30</sup>. Additionally, the technology is based on MS and thus has the potential to become an essential tool for precise tissue characterization during excision surgery, because of the high specificity of the molecular profiles where the different markers are separated according to their mass (and charge). The MS spectra generated by the SpiderMass are indeed used to build a classification model which associates the MS molecular fingerprint obtained with a specific cell type (such as cancer). This association is achieved thanks to the use of a tissue cohort that has been annotated by a pathologist post-surgery. The classification model after validation can then be implemented in the instrument and is interrogated in real-time during the surgery to provide an instant feedback, so far using a code, to the physician. The MS spectra are recorded in the millisecond range as well as the interrogation of the classification with a result that can be provided within less than a second. Moreover, we recently demonstrated that if SpiderMass can be used to analyze fresh or fresh frozen tissues, it is also suitable for retrospective analysis in the pathology lab from Formalin Fixed and Paraffin Embedded (FFPE) samples<sup>31</sup>. Interestingly, previous studies conducted intraoperatively both *ex vivo* and *in vivo* at the veterinary operating room on dog patients with sarcoma have also shown that the chemical background is not higher *in vivo* than *ex vivo*<sup>27</sup>. In cancer diagnosis-based MS, supervised ML algorithms play a critical role in building accurate classification models based on labelled datasets<sup>32</sup>. Choosing and fine-tuning the appropriate classification model can have a significant impact on the accuracy of the results<sup>33</sup>. Cross-analysis by non-supervised ML strategies as well as interpretability of the results are also paramount and must be carefully considered.

Hence SpiderMass technology reveals a potential to help with the management of patients with solid tumors. Here we have explored this potential for patients with GBM. More specifically, we have identified two main aspects for which the technology would be useful. The first is to better delineate the surgical margin to move towards precision surgery. The second is to provide a stratification of GBM and be able to get a prognosis to help the oncologist with refining the therapy accordingly. Ultimately, if in the future some therapies could be administrated intraoperatively, we could offer all this information

available by the time of the surgery. To perform SpiderMass intraoperatively, we must rely on a strong classification model. To address this, we have developed a more comprehensive AI pipeline to provide accurate and reliable classification models and confident biomarker identification. The application of the pipeline to GBM demonstrates an accurate diagnosis and applicability to prognosis with more than 90% accuracy to distinguish patients with longer survival from those with shorter survival. We also show that specific lipid markers are associated with the prognosis of patients and are more specific to the molecular stratification than the tumor localization or histological class. Additionally, we study the importance of the immune cell infiltration within the tissues and demonstrate that the ratio of macrophages subpopulation can be used to predict patient outcome. Ultimately, in the present paper we demonstrate that SpiderMass could be used to get a sub-stratification of patients with glioblastoma as well as the prognosis in real-time to afford improved management of the patient therapy.

## Results

### Glioblastoma IDH-wild type lipidomic status assessment

#### *Optimal classification model and AI processing pipeline.*

To undertake surgery, SpiderMass is combined to machine learning. Different ML algorithms are yet available, ranging from traditional ML to deep learning (DL) and the choice of the modelling will influence the accuracy of the built models. Traditionally, molecular diagnosis-based SpiderMass has relied on linear discriminative analysis (LDA) modelling using the Abstract Model Builder - AMX software, which has some limitations such as using only LDA classification, missing model evaluation and cross-validation, providing no explanation for predictions, and lacking statistical tests. In addition, ML models have been criticized for being opaque and providing limited insight into how or why they arrive at a particular outcome<sup>34,35</sup>. To address these limitations, we have developed a more comprehensive AI pipeline (**Figure 1A**) to provide reliable classification models and confident biomarker identification. Thus, we compared the performances of 24 different classifiers. This was performed from a retrospective cohort of FFPE tissues (**Figure S1, Table S1**). All histologic regions (50 tumor, 42 necrosis and 18 benign tissues) were analyzed using SpiderMass in both negative (138, 102 and 42 MS spectra respectively) and positive ion mode (121, 88 and 39 MS spectra respectively) (**Table S2**). Classification models were constructed to discriminate between tumor, necrotic and benign tissue for which class assignment has been provided by pathologist and hence considered "the ground truth". In negative and positive ion mode, the optimal classification was obtained using the RidgeClassifier, based on the highest accuracy on the 20% validation set, with 93% and 92% accuracy respectively (**Figure S2A and D**). The corresponding accuracy of the training set was 100% for both (**Figure S2B and E**). After obtaining the classification report and confusion matrix for the optimal classifier, its performance was evaluated by using both accuracy and F1 scores, the latter considering both precision and recall and provide a more balanced measure of performance across classes. However, to ensure robustness and generalization of the data, a 20-fold cross-validation was performed. This showed that the actual accuracy of the model was 88% with a standard deviation (std) of 0.03 and 87% with a std of 0.02 in negative and positive ion mode respectively. The model achieved F1 scores of 85%, 87% and 89% for the benign, necrotic and tumor tissue in negative ion mode, whereas the corresponding F1 scores in positive ion mode were 84%, 86% and 88% (**Figure S2C and F**). The lower final accuracy obtained after 20-fold cross-validation could come from the small number of samples in the retrospective cohort, making the result less reliable. To address this issue, we added a second dataset (D2) based on a prospective cohort of 31 fresh-frozen tissues (**Figure S1 and Table S1**) to get an extended dataset (DE). The added dataset D2 consists of 30 tumor regions and 7 necrosis regions, corresponding to 111 and 20 MS spectra in negative ion mode and 110 and 16 MS spectra in positive ion mode (**Table S2**). Despite some spectral differences (**Figure S3**) between the spectra of the FFPE and fresh tissues, the used of DE improved both the accuracy and F1 scores after cross-validation. In negative ion mode, the accuracy increased and stabilized at 92% after 20-fold cross-

validation (**Figure 1B**). In addition, the std decreased from 0.03 to 0.02. In positive ion mode, there was a slight improvement in performance (**Table S3**). Importantly, the accuracy remained stable even after cross-validation, allowing an actual accuracy of 88% with a std of only 0.02.

#### ***Automatic blind prediction of histological class from the built classification.***

To further test the performance of the statistical model in both ion modes, blind predictions were made on an independent test set consisting of two FFPE tissues and seven fresh frozen tissues. These nine tissues were not used to construct or evaluate the previous built model. The regions were subjected to SpiderMass analysis in a blind fashion and evaluated against the classification model trained on DE. In negative mode, all the analyzed tissue were correctly attributed to their respective region after comparing the SpiderMass result to the pathologist exam. In positive mode, 2 out of 19 regions were misclassified, resulting in 90% of correct assignments (**Figure 1C**). All necrotic regions were correctly predicted, while two tumor regions were classified as necrotic regions. These two classification errors can be attributed to the fact that the blind analysis point resides at the border between the tumor and necrotic zones.

#### ***Lipid markers associated to the different molecular classes.***

We were then interested to obtain the identification of the lipids associated to the different histological classes to better understand the biological pathways responsible for the discrimination and address the issue of most models which remain black box. The LIME<sup>36</sup> algorithm included in the developed processing pipeline provides interpretable explanations with potential biomarker identification by calculating the weight of each  $m/z$  feature to classify each tissue type. The top 120 positively and negatively contributing  $m/z$  features and their contribution are shown in **Figure S4A-C** for negative and positive ion mode. For example, in negative ion mode, the  $m/z$  feature that contributes most positively to the classification of benign tissue is  $m/z$  766.55. The  $m/z$  751.55 contributes positively to classify necrotic tissue but negatively to tumor tissue, while the opposite is true for the  $m/z$  794.55. To identify the potential lipid biomarkers independently of classification, ion abundance variations between the different regions of interest (ROI) were investigated using multivariate statistical analysis on the same data. In negative ion mode, peak picking, considering signal with signal under noise  $> 10$ , yielded 509 peaks out of a total of 5000  $m/z$  features. Hierarchical clustering was used to construct a heatmap and allowed us to identify three ion clusters, each of which was more over-expressed in a specific tissue type (**Figure S4D**). After heatmap visualization, a Kruskal-Wallis test was performed to determine whether each ion was statistically significant. Of these 509 peaks, 154 were found to be significant ( $p$ -value  $\leq 0.05$ ). After filtering out unresolved peaks and isotopes, 57 potential biomarkers remained. The same strategy was used in positive ion mode (**Figure 5B**), resulting in 100 significant peaks and a total of 37  $m/z$  features were retained as potential biomarkers.

To uncover lipid biomarkers with high confidence, we combined the data from the supervised and unsupervised analyses. Thus, for an  $m/z$  feature to be considered as confident biomarker, it should be significant (either over-expressed or under-expressed) in a given ROI and should also have a positive contribution if over-expressed and a negative contribution if under-expressed in the same ROI. A total of 41 confident biomarkers were identified, 27 in negative ion mode and 14 in positive ion mode, which are listed in **Table S4** with their corresponding identification, obtained by MS/MS thanks to the SpiderMass analysis, and their relative abundance variation presented in box plots (**Figure 1D** and **Figure S5**). In negative ion mode, of the 27 confident biomarkers, nine ions are overexpressed in the tumor tissue, the majority of which are phosphatidylserines (PS) and phosphatidic acids (PA). Seven ions are specific for benign tissue, of which six are highly abundant and one DG 40:9 ( $m/z$  661.55) is poorly abundant. Normal brain tissues are composed of different lipid species, but the absence of diglycerides (DGs) appears to be a potential marker of health. Interestingly, all the highly abundant ions found in benign tissue are large phospholipids with fatty acids containing more than 3 unsaturation. In glioma cells, the high concentration of polyunsaturated fatty acids appears to inhibit cell division and tumor formation<sup>37</sup>. In necrotic

regions, three ions are over-expressed at low masses and eight ions are under-expressed at high masses. More specifically, phosphatidylinositol (PI) are found to be under-expressed in these tissues. In positive ion mode, among the 14 confident biomarkers, two are specific for benign tissue,  $m/z$  842.65 and  $m/z$  850.65. For necrotic tissue, three confident biomarkers are poorly represented, again PIs and phosphatidylcholines (PC). For tumor regions, eight were found to be over-expressed, the majority of which are phosphatidylethanolamines (PE). Interestingly, the glycerophospholipids (GPL) highly expressed in tumor tissues are mostly saturated lipids such as PC 34:2 ( $m/z$  758.65) and PS 38:2 ( $m/z$  814.55).

### ***Unsupervised and supervised molecular sub-stratification of glioblastoma IDH-wild type based on lipidomics.***

In a previous study using spatially resolved shot gun proteomics guided by MALDI-MSI, we have shown that grade 4 glioblastoma could be stratified in 3 specific molecular sub-groups<sup>23</sup>. Interestingly, these were shown to be involving different pathways respectively related to i) neuro-developmental genes, characteristic of neuronal/glial lineages and neural progenitor cells (red region), ii) immune status with macrophages infiltration (yellow region) and iii) antiviral immune response and viral infection (blue region). Additionally, a correlation was observed between these groups and the OS of the patients, revealing that those patients with identified inflammation markers, macrophage infiltration, and antibodies exhibited a shorter OS. We thus firstly performed MALDI-MSI in positive ion mode from the remaining tissues of our prospective cohort and performed a global segmentation from all the tissues (**Figure 2A**). As previously demonstrated for peptides, we found 3 main molecular groups with a good correlation with the global segmentation from the protein MALDI MSI. Indeed, it gives the same sub-stratification of IDH-wild type based on lipidomic and proteomic MALDI-MSI data. There would be a high value to obtain such a stratification *in vivo* and in real time during the surgery to adapt the EOR according to the prognosis of the patient. MALDI-MSI cannot be used *in vivo* but SpiderMass can. Interestingly, MALDI and SpiderMass lipid profiles are very close because SpiderMass is based on WALDI-MS which recapitulate a similar process to MALDI using endogenous water as a MALDI matrix<sup>38</sup>. Thanks to the AI processing pipeline developed, classification models were obtained to differentiate the 3 specific molecular sub-groups with a correct classification rate, after 20-fold cross-validation, of 94% and 88% in negative and positive ion mode respectively (**Figure 2B-C**). The automatic blind prediction of the different subgroups made it possible to classify well six under the eight unknown tissue. Next, 18 confident biomarkers were discovered in both modes (**Figure 2D**). Interestingly, some ions were in common with the lipid markers of the different histological zones found above. For example, PS 34:1 ( $m/z$  760.55) specific to the tumor area is also specific of the proteomic blue region related to tumor growth. In addition, two ions  $m/z$  878.65 (PE 46:4) and 902.65 (PS 44:0) were found specific to benign tissue but also here of the red region related to neuro-developmental genes. This enables multiple correlations between proteomic and lipid data, as well as the possibility of using SpiderMass as an *in vivo* tool for stratifying glioblastoma tissue to adapt the surgery response according to the survival time of the patient.

### **Prediction of patient outcome from SpiderMass lipidomic data**

#### ***SpiderMass classification of patient according to the survival.***

On top of the validated diagnosis and stratification based on SpiderMass, we were also interested in investigating the potential to predict the patient outcome. A classification was built using the previously developed processing pipeline, according to a 15-month median of OS of glioblastoma patients, separating the cohort into  $<$  and  $>$  15-month OS. We used the MS spectra recorded from the tissues independently from their localization to find the lipids discriminative of the outcome rather than the histological annotation. Both in positive and negative ion mode, we could build a classification model with a correct classification rate of 79% and 87% after 20-fold cross-validation (**Figure 3A**). To improve the specificity of the outcome-based classification, we then looked to the patients with the more severe survivals i.e.,  $<10$  and  $>36$  months. By narrowing our model to these cases, we obtained, in negative ion

mode, a classification rate of 100% for the training set, 94% for the training set of the 20% validation and 93% after the 20-fold cross-validation (**Figure 3B**). In positive ion mode, the classification model reaches 87% of correct classification rate after cross-validation (**Figure 3B**).

#### ***Discriminative markers associated to the patient outcome.***

When searching for the biological explanation and potential biomarkers associated to the shortest and longest OS, some ions were found to be specific of one or the other. Indeed, with both ion modes, 48 statistically confident prognosis biomarkers were found (**Table S5**). In summary, PCs and PIs are more present in patients with a longer OS, as PC 42:4 and PC 44:6 in positive ion mode (**Figure 3D**) and PI 38:4 and PC 42:6 in negative ion mode (**Figure 3C**). Moreover, ions  $m/z$  850.75 (PE 44:5) and  $m/z$  920.75 (PS 46:5) are more abundant in benign tissues and were found also to be associated with longer survival. It would appear that tumor tissue from a patient with long survival shares molecular markers with benign tissue. This could explain why some patients live longer. On the contrary, PSs, PAs and PEs, such as PA 38:1 and PE 38:0, are more highly expressed in poor prognosis tumor tissue (**Figure 3C-D**). In particular, PS 40:6 was found to be associated with lower survival in both ion modes ( $m/z$  834.55 and  $m/z$  836.55) with a high significance and a high contribution. In addition, sphingolipids, specifically ceramides (Cer), were characteristic of a worsened prognosis in our study, since none were found for the longer survivals. It was also shown that the abundance of 5 Cer lipids, including Cer d42:2 ( $m/z$  682.65) and CerP 34:0 ( $m/z$  654.55), was significantly higher in patients with poor survival, regardless of the acyl chain length or the degree of unsaturation. Interestingly, our findings at the lipid level are closely related to those at the proteomic level. Indeed, the previously identified three prognosis markers<sup>23</sup> are closely linked to the prognosis lipid markers found here. Tumors from patients with longer survival showed increased expression levels of proteins RPS14 and PPP1R12A, whereas tumors from patients with shorter survival showed higher expression level of ANXA11. ANXA11 is a calcium-dependent phospholipid-binding protein<sup>37</sup>, that bind to negatively charged phospholipids in the presence of calcium ions<sup>39</sup>. The phospholipids identified as biomarkers of shorter survival, PSs, PAs, and PEs, are negatively charged, suggesting a specific lipid and protein pathway associated with a poorer prognosis. Targeting this pathway in the future may improve the survival of glioblastoma patients. In addition, PPP1R12A is a subunit of myosin phosphatase, which regulates anti-tumor signaling pathways such as activating gene expression of tumor suppressors like Rb, Rap, and c-Myc<sup>40</sup>. On the other hand, there is a known ceramide-protein interaction between ceramides and I2PP2A, leading to the degradation of c-Myc<sup>41</sup>. This highlights a balance between the presence of PPP1R12A protein, which activates c-Myc and leads to a better prognosis, and ceramides, which inhibit c-Myc and are associated to a shorter survival.

## **Immunoscore of glioblastoma by identification of SpiderMass signature of immune cells**

### ***SpiderMass classification of immune cells.***

Infiltration of various population of immune cells was found to be associated to the patient outcome in colon cancer<sup>42-44</sup> and is now recognized in many cancers<sup>45</sup>. Nowadays, the immunoscore is obtained from excised tissues or biopsies by immunohistochemistry (IHC) using antibodies specific of the different immune cell populations. However, the immunoscore is obtained post-surgery. Here, we explored the possibility to create an immunoscore based on SpiderMass data that could in the future be exploited *in vivo*. To this end, we analyzed with SpiderMass different population of immune cells namely macrophages (M1-like and M2-like), CD4 et CD8 T lymphocytes and NK cells (all regrouped as lymphocytes) versus cancer cells (NCH82 glioblastoma cells). Immune cells were analyzed directly from culture well-plates after cell sorting by flow cytometry to get the specific molecular profiles of each cell type. Indeed, since SpiderMass is not designed to offer single cell analysis, as its spatial resolution is only 250  $\mu\text{m}$ , hundreds of cells are analyzed at the same time from one analytical spot. To compensate for this more



limited spatial resolution by comparison to IHC, we developed a solution to predict the ratio of one cell type from a SpiderMass MS spectrum. A Python library called LGBM (Light Gradient Boosting Machine)<sup>46</sup> was used to train an immunoscore model using 107 MS spectra for each cell type for which a correct classification rate of 100% was obtained in both training and cross-validation (**Figure 4A**). Using our developed AI pipeline, 10 lipid markers of M1-like macrophages, 8 lipid markers of M2-like macrophages, 6 lipid markers of lymphocytes and 3 lipid markers of NCH82 cancer cells were obtained (**Figure 4B and Table S5**). For example, the ions  $m/z$  818.65 (GlcCer d18:1\_22:0) and 819.55 (PG 18:1\_22:6) were found specific to M1-like and M2-like macrophages respectively. For the lymphocytes, PEs were reported as lipid markers as the PE (20:4\_16:0) ( $m/z$  738.55) and PE 38:3 ( $m/z$  768.55). Interestingly, two ions ( $m/z$  844.65,  $m/z$  848.65), specific to M1-like and M2-like macrophages, were also found specific of tumor tissue from patient with an OS below 10 months.

### *Correlation of immunoscores with patient prognosis.*

From a SpiderMass image, we were able to get the predicted distribution of the different immune cells across the tissue. Indeed, a second pipeline was dedicated to predict the probability of presence of each cell type based on SpiderMass images of fresh frozen glioblastoma tissues of the prospective cohort. The results obtained provide estimated scores for each cell type, and ratio scores were computed to determine the relative score presence (RSP) of each cell type across the entire image. These scores were calculated by summing the scores provided for each cell type and dividing the sum by the total scores across all labels (**Figure 4A-C**). A total of six fresh frozen tissues were analyzed, three from patients with a <10-month survival and three from patients with >36-month survival. The ratios provide insight into the distribution of the trained cell types across the image, allowing a comprehensive assessment of the cellular landscape in patient with shorter versus longer survival (**Figure 5A-B**). Interestingly, the predicted ratio of cancer cells is not significantly different according to the survival. However, the lymphocytes were predicted to have a higher abundance in tissue from patients with >36-month survival (**Figure 5C**). Indeed, the mean RSP of lymphocytes was above 40% and below 36% for longer versus shorter survival. The predicted ratio of macrophages also revealed an interesting association between immune cells and the patient outcome. Specifically, a higher proportion of M1-like macrophages (3.5%) was found in the tissues of patients with a better outcome compared to those with a worse (2.8%) (**Figure 5D**). Conversely, M2-like macrophages had a mean RSP of 35.8% and 41.3% in patients with longer and shorter survival (**Figure 5E**). The ratio of M1-like to M2-like macrophages could serve as a prognosis marker. In fact, a proportion of M1-like macrophages below 7%, in comparison to M2-like macrophages, signifies a shorter OS. Conversely, a composition exceeding 10% of M1-like macrophages relative to M2-like macrophages is indicative of an extended OS (**Figure 5F**). Furthermore, a correlation appears to exist between the prognosis and the balance between the presence of lymphocytes and macrophages. Specifically, in cases of unfavorable prognosis, macrophages tend to outnumber lymphocytes (+ 22%), while the converse holds true for patients with a favorable prognosis (**Figure 5G**).

A 5-plex MALDI-IHC panel was constructed to corroborate previous results from the recently developed immunoscore pipeline. MALDI-IHC relies on the use of antibodies conjugated with novel photocleavable mass-tags and enables highly multiplexed MALDI mass spectrometric imaging of targeted protein biomarkers<sup>47-49</sup>. For this purpose, the five biomarkers chosen were two controls, vimentin and a collagen cellular matrix marker; the Ki67 cancer-related biomarker (proliferation/cancer cells); and two biomarkers for immune-cells CD8 (cytotoxic T-cells) and CD68 (macrophages). The distribution of distinct cell populations (importantly immune cells), as achieved through the integration of SpiderMass and LGBM, was validated by MALDI-IHC (**Figure 4D**). Indeed, a higher concentration of macrophages were detected in tissue from a patient with a poor prognosis. In comparing immunoscores with the MALDI-IHC analysis, it is important to note that immunoscore uses a probabilistic approach where each pixel indicates the likelihood of specific cell types being present, spanning a gradient from 0 to 1.

In contrast, the MALDI-IHC analysis follows a binary strategy, assigning pixels a value of 0 (absence) or 1 (presence) for the considered cell types.

Of particular interest is the fact that the SpiderMass-MSI approach, based on immunoscore, allows the differentiation of various subpopulations within the TME without necessitating techniques reliant on probe utilization. In conclusion, this innovative approach not only offers insights into the presence of immune cells within the TME but also presents the potential for a more rapid prognostication of survival time among GBM patients.

## Discussion

This study represents a comprehensive investigation of the possibility to get an accurate diagnosis, sub-stratification and prognosis of glioblastoma using AIMS by SpiderMass combined with an AI pipeline. We explored various classification methods to discover the most appropriate one for our dataset. Given the lack of consensus in the community and the utilization of different algorithms (such as LDA<sup>50</sup>, Lasso<sup>51</sup>) for similar purposes, our goal was to identify the most suitable approach. A total of 24 classifiers were evaluated and an optimal model was obtained using the RidgeClassifier<sup>52</sup>. It is a linear classifier that uses L2 regularization to avoid overfitting and demonstrate strong performances for MS data from tissues (high-dimensional datasets)<sup>53</sup>. To address the issue of small sample size which is responsible for the limited performances of the classification model, we investigated data extension by aggregating SpiderMass data from both fresh frozen tissues and FFPE tissues thanks to the ability of the technology to provide good quality data independently of the conservation of the sample. Data extension led to an increase model accuracy with 92% and 88% after 20-fold cross-validation for tissue histological type classification in both ion modes. Blind predictions resulted in no misclassified spectra in negative ion mode and only two misclassified spectra in positive ion mode, with an average calculated sensitivity and specificity of 95% and 98%. By comparing the specificity and sensitivity obtained in our study for pre-operative delimitation of brain tumors with other techniques in use<sup>54</sup>, SpiderMass technology shows to have an interesting potential by comparison with the other modalities such as fluorescence-guided surgery or even preoperative image guidance. SpiderMass also appears to be in upper range of success compared to other MS-based modalities like REIMS, while being far less invasive (**Table S6**).

Additionally, to validate further our results and provide biological explainability, we included a secondary data processing pipeline. LIME algorithm combined with unsupervised machine learning techniques and statistical tests was used to find confident lipid markers. The pipeline's predictive explanations and statistical tests can increase the transparency and usefulness of machine learning models for biomedical applications. The combination of SpiderMass and ML pipeline represents a promising approach for advancing precision medicine for glioblastoma. Interestingly, the study identified 41 reliable biomarkers that showed different phospholipid classes based on tissue type. Benign tissue was characterized by the presence of polyunsaturated fatty acids and the absence of DGs. Necrotic tissue showed the absence of PIs, while tumor tissue had a significant presence of PSs and PAs among the detected phospholipids<sup>55,56</sup>. These robust biomarkers warrant further investigation for their therapeutic potential.

Here, we have extended our previous work by showing that IHD-wild-type glioblastoma could be stratified in three main molecular sub-groups with similar results based on lipidomic data as those obtained by spatially resolved proteomics. The accuracy of classification for these subgroups was found to be 94% and 88% for negative and positive ion mode, respectively. Notably, within the set of 19 robust biomarkers uncovered, three biomarkers were already identified as specific to distinct histological tissue types. For instance, one biomarker exhibited specificity to the tumor region and also within the blue region linked to tumor growth. Two other biomarkers demonstrated specificity to both benign regions and the red region, which, is associated with neuro-developmental genes. This not only facilitates intricate correlations between proteomic and lipid data but also opens the door to employing SpiderMass as an *in vivo*

tool for stratifying glioblastoma tissue. This stratification can significantly contribute to tailoring surgical responses in alignment with a patient's survival timeline.

Interestingly, our developed ML pipeline not only serves as a diagnosis tool but also for prognosis. The classification model developed in this study successfully distinguished tumors from patients with less or more than the median survival of 15 months. However, more specific results were obtained when comparing patient with more extreme survivals, i.e., <10 months versus > 36 months with achieving accuracies of 93% and 87% in negative and positive ion mode respectively. The presence of PCs and PIs was associated with a longer survival, while PSs, PAs, PEs, and Cer were more highly expressed in tumor tissue with a shorter survival. A recent study also found significantly more Cer in control tumor tissue compared to tissue treated with anti-PD-1 therapy, supporting our findings that ceramides are markers of a poor prognosis<sup>37</sup>.

A graphical depiction showcasing all identified biomarker ions from our study, alongside their corresponding groups of specificity, is presented in **Figure 6**. This visual representation facilitates an effortless assessment of whether a given ion exhibits specificity toward more than one group. Notably, a compelling correlation emerges between tumor methylation, tissue benignity, OS exceeding 36 months, and the segmentation clusters denoted in red and yellow. Indeed, a direct association between ions specific to tumor methylation and those specific to survival beyond 36 months ( $m/z$  725.55 PA 38:3 and 906.65 PC 42:6), as well as benign tissue ( $m/z$  904.65 PC 42:7 and 932.75), becomes apparent. Furthermore, ions exclusive to benign tissue also exhibit specificity toward survival exceeding 36 months ( $m/z$  850.65 PE 44:5, 883.65 PI 38:5, and 920.75 PS 46:5), in addition to the red and yellow sub-stratification clusters ( $m/z$  874.85,  $m/z$  878.65 PE 46:4,  $m/z$  881.75, and  $m/z$  902.55 PS 44:0). In contrast, the blue sub-stratification cluster demonstrates a distinct association with tumoral tissue ( $m/z$  760.55 PS 34:1) and survival periods of less than 10 months. Notably, ions specific with no-methylation also exhibit specificity toward tumor tissue ( $m/z$  780.65 PE O-40:5 and 802.55 PC 38:7) and shorter survival durations ( $m/z$  744.55 PE 36:1). Intriguingly, tumor tissue's connection to short survival is underscored by the presence of three ions ( $m/z$  749.55 PA 40:6, 778.55 PS 36:6, and 788.65 PE P-40:0). Considering that some ions were found to be specific of the methylation status, obtaining the methylation status by the SpiderMass in a less time-consuming and resource manner than currently achieved by genomics is something that could further be investigated in the future.

The third innovative aspect is the use of SpiderMass to predict the survival time of glioblastoma patients based on immune cell infiltration. Here, we studied macrophages and lymphocytes as a first approach to assess the possibility of predicting the presence of immune cell infiltration based on SpiderMass data. The M1 phenotype of macrophages refers to pro-inflammatory and anti-tumor macrophages, while the M2 phenotype refers to anti-inflammatory and pro-tumor macrophages<sup>57</sup>. Creating an immunoscore for SpiderMass offers a way to translate the immunoscore *in vivo* and predict the patient outcomes already at the time of the surgery. This is offering a way for surgeons to tailor the therapy if a local intraoperative therapy (e.g, photodynamic therapy) is planned. So far, SpiderMass analysis, based on surgeon needs, was set to a spatial resolution of 400  $\mu\text{m}$  which is obtained by focusing the laser beam at the exit of the optical fiber. Thus, a recorded MS spectrum in an area of the tissue showing immune cells infiltration corresponds to a mix spectrum of different cell types. To address this problem, we have tested the performance of the LGBM algorithm to predict the ratio of different cells within the MS profile. However, this required training the LGBM algorithm to recognize the profile of the lymphocytes and macrophages though. This was achieved by analyzing the different immune cells with the SpiderMass to feed the algorithm. Remarkably, based on the prediction, we showed that the macrophage type (M1-like versus M2-like) were associated with the prognosis with different ratios of M1-like and M2-like according to the patient survival, patient with lower survival presenting more M2-like and less M1-like and vice versa for patient with longer survival. This is well-known for cancer and it validate the approach. Additionally, the ratio of M1 and M2 fall within the values previously calculated by histology and IHC<sup>58-60</sup>. Besides, the predictions were validated using MALDI-IHC using specific antibodies directed against immune

cells. These results indicate that the immunoscore has the potential to become a prognosis prediction tool for glioblastoma patients without the need of probes, allowing for personalized treatment for the patients based on their personal immunoscores. Additionally, considering that the approach employed for the immunoscore enables to predict the percentages of presence of each cell type in the local environment analyzed, it can effectively be also translated for the definition of surgical margins. Indeed, for accurate margin assessment it is needed to get the percentage of cancer cells versus normal cells. The same approach as we developed could be used to predict these percentage of these cell types. While none of the tissues in the present cohort were suitable for establishing this boundary, preliminary findings from an ongoing study from transgenic TgC3(1)-Tag mice models of breast cancer demonstrate its successful application for establishing ratio of cancer/normal cells for improving the delineation of the tumor margins (**Figure S6**).

In conclusion, this study demonstrates the potential of SpiderMass combined with AI predictions for accurate classification of tumors and tissue subtypes. The ML pipeline developed in this study provides accurate classification results, identifies potential biomarkers, and provides interpretability. The reliable biomarkers identified reveal the distinct phospholipid profiles associated with different tissue types. In addition, the lipid and proteomic findings suggest a specific lipid-protein pathway associated with both poor and good prognosis in glioblastoma patients. The use of SpiderMass as a prognosis tool and the association between macrophage phenotypes and patient prognosis further contribute to the advancement of precision medicine for glioblastoma and other heterogeneous cancers. Future investigations should focus on validating these findings in larger cohorts and exploring targeted therapies based on the identified pathways and biomarkers. Overall, the combination of SpiderMass and the ML pipeline holds great promise for possible application in the future for accurate diagnosis and prognosis as well as improved EOR to be achieved during the surgery directly from the patients and is even compatible with awake surgery considering the analysis is painless.

## Limitations of the Study

The diagnosis and prognosis aspects, rooted with Ridge model classification, yield highly favorable results, and the corresponding markers identified through both supervised and unsupervised approaches demonstrate robustness. From a molecular standpoint, this reliability is underpinned by the well-established presence of glyceropholipids, primarily situated in the bilayer of cell membranes.

While the LGBM model performs admirably across various fingerprints of four cell types (NHC82, M1, M2, lymphocytes), it is crucial to acknowledge that the interpretability of scores on the spidermass image remains relative. To comprehensively capture the diverse landscape of the TME, the inclusion of additional cell types is warranted. To enhance the generalizability of the link between immunoscore and patient prognosis, there is a need to expand the cohort of tissues subjected to imaging analysis.

Moreover, considering the promising potential targets identified in this study, future investigations should explore these findings through drug screening and brainoid models. This proactive approach could shed light on the translational implications of the study and facilitate the development of targeted interventions.

## Acknowledgments

This research was supported by grants from Ministère de l'Enseignement Supérieur et de la Recherche (MESR), Inserm specific funding for SpiderMass project (I.F.), Inserm and Institut Universitaire de France (I.F.). L.L. PhD was cofounded by University of Lille Excellence Initiative, Région Haut de France (EU Feder funds) and OCR. We thank OCR for its contribution to the project. The work using MALDI-IHC was supported in part by grants to AmberGen, Inc. from the National Institutes of Health including the following: R44CA236097, R44AG078097 and R44MH132196.

## **Author contributions**

Y.Z., L.L., M.S. and I.F. wrote the manuscript original draft. I.F. and M.S. designed the experiments. L.L. and M.D. performed the experiments. L.R. performed cells SpiderMass analysis. E.L.R. is the coordinating investigator of the clinical study and gathered patient clinical data. C.A.M. collected the glioma samples and performed the histopathology examination. B.M. performed the purification of immune cells. G.Y., K.J.R. and M.J.L. provided the Miralys<sup>TM</sup> probes and protocols. Y.Z. developed the machine learning pipelines. Y.Z. and L.L. analyzed the data. I.F., M.S., Y.Z., L.L., M.D. and P.T. corrected the manuscript. I.F. and M.S. supervised the project and provided the funding.

## **Declaration of interests**

L.L., Y.Z., C.A.M., L.R., M.D., B.M. and P.T. declare they have no competing interests. E.L.R. has received grant research from Bristol Meyer Squibb and honoraria for lectures or advisory board from Bayer, Janssen, Leo Pharma, Pierre Fabre, Roche, Seattle Genetics and Servier. M.S. and I.F. are inventors on a patent (priority number WO2015IB57301 20150922) related to part of the described protocol. D.Y., K.J.R. and M.J.L. are current employees of AmberGen, Inc., 44 Manning Road, Billerica, MA, USA. AmberGen, Inc. has filed patent applications on different aspects of MALDI-IHC.

## Main figure legends

**Figure 1. AI pipeline for optimal lipid classification model and discovery of molecular class-associated lipid markers.** (A) Overall pipeline for molecular diagnosis. The continuous arrows are mandatory to obtain the final results (arrow ending with a dot) while facultative step (dotted arrow) can be added to the pipeline, like the spatial distribution of confident biomarkers thanks to WALDI-MSI. (B) Classification report and confusion matrix of RidgeClassifier with the training set and after 20-fold cross-validation, obtained for the classification model using DE in negative ion mode. (C) Table with the predicted label by SpiderMass blind diagnosis compared to the pathologist true label of two FFPE and seven fresh frozen (FF) unknown GBM tissue. (D) Corresponding boxplot of 10 confident biomarkers specific for necrosis, benign and tumor tissue in both ion mode with their tentative annotation. Related to Figures S1-S6 and Tables S1-S6.

**Figure 2. Unsupervised and supervised molecular sub-stratification of IDH-wild type based on lipidomics.** (A) Global segmentation based on lipidomic and proteomic data for nine FF GBM tissue. (B) Classification report and confusion matrix of Ridge Classifier with the training set and after 20-fold cross-validation, obtained for the classification model that distinguish the 3 segmentation clusters. (C) Boxplots of two lipid biomarkers obtained for each segmentation clusters.

**Figure 3. Obtainment of a classification model to distinguish patient with a low or a long survival time and the discovery of biomarkers for each survival.** (A-B) Accuracies obtained for the classification model made with RidgeClassifier, with the training set, after 20% validation and after 20-fold cross-validation in negative (orange) and positive ion mode (red) to distinguish patient with a survival time < or >15months and after narrowing our model to most extreme cases (<10 months and >36 months). (C-D) Six potential biomarkers, their corresponding boxplot and their chemical structure, found specific of prognosis <10 and >36 months in both ion mode. Related to Table S7.

**Figure 4. SpiderMass classification based on immune cells lipid profiles.** (A) Trained LGBM model for cancer cells, lymphocytes, M1-like and M2-like macrophages that were analyzed previously by SpiderMass technology after isolation. (B) Boxplot of 3 biomarkers for GBM cancer cells, lymphocytes, M1-like and M2-like macrophages. (C) The overall pipeline for seeing the distribution of cancer and immune cells in GBM tissue analyzed by SpiderMass-MSI. (D) Multiplex MALDI-IHC on 3 biomarkers in two GBM fresh frozen tissue sections. The display scale (arbitrary peak intensity units) is as follows (minimum intensity/full intensity threshold): 4/25 (Ki67), 2/6 (CD8 $\alpha$ ) and 4/25 (CD68). Related to Table S8.

**Figure 5. Immunoscoring distribution from GBM FF tissues for cancer cells, lymphocytes, M1-like and M2-like macrophages.** (A) and (B) H&E scans and relative score of presence for cancer cells, lymphocytes, M1-like and M2-like macrophages in three tissues of patient with a good or a poor prognosis. For each cell, the ratio in percentage is indicated. In addition, a histogram was built to compare the percentage in each tissue. (C-D-E) The mean percentage of lymphocytes, M1-like and M2-like macrophages in tissue with a prognosis <10 months and >36 months. (F-G) The mean ratio M1/M2 macrophages and lymphocytes (L)/macrophages (M) in tissue with a prognosis <10 months and >36 months.

**Figure 6. Overview of all lipid markers identified using supervised and unsupervised discovery specific of the tumor methylation, tissue histology, sub-stratification of GBM tumor and the survival time of the patient.** (A) negative ion mode. (B) positive ion mode. Related to Figure S6 and Tables S6-S8.

## STAR METHODS

### KEY RESOURCE TABLE

See separate Word with the key resource table.

### RESOURCE AVAILABILITY

#### *Lead Contact*

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Isabelle Fournier ([isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr)).

#### *Materials availability*

This study did not generate new unique reagents.

#### *Data and code availability*

All SpiderMass raw data (MS spectra, MSMS spectra and MS imaging data) have been deposited at Harvard dataverse and are publicly available as of the date of publication. DOI is listed in the key resources table.

All original code has been deposited at GitHub ([https://github.com/yanisZirem/Pipeline\\_SpiderMass1D](https://github.com/yanisZirem/Pipeline_SpiderMass1D) and [https://github.com/yanisZirem/Pipeline\\_SpiderMass2D](https://github.com/yanisZirem/Pipeline_SpiderMass2D)) and was archived in Zenodo. The DOIs are available in the key resources table. If you have any questions or feedback, please contact [yanis.zirem2016@gmail.com](mailto:yanis.zirem2016@gmail.com).

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Study design

According to the French Public Health Code and in application of the General Data Protection Regulations, all patients had been informed at the time of care that their standard clinical and biological data could be used for research purposes regarding the retrospective analysis of FFPE samples, and none had expressed his opposition. Regarding the prospective collection of samples, each patient's informed consent for the collection and publication of clinical and biological data was obtained at the time of hospitalization prior to surgical intervention.

### Patients

A retrospective cohort of 50 FFPE glioblastoma tissues was obtained from the Pathology department of Lille Hospital, France. A prospective cohort of 31 fresh frozen glioblastoma tissues were also included in this study. 31 patients with newly diagnosed glioblastoma were prospectively enrolled between September 2014 and November 2018 at Lille University Hospital, France (NCT02473484). All patients gave written informed consent before enrollment. The demographic and clinical data of the cohorts are reported in Table S1.

### Cell line

Human NCH82 stage 4 glioblastoma cells were obtained from Dr Regnier-Vigouroux.



## METHOD DETAILS

### Histological staining and annotations

The paraffin blocks and the fresh-frozen tissues were sliced into 7  $\mu\text{m}$  sections and placed on poly-lysine coated slides. This tissue slice consecutive to the SpiderMass-analyzed tissue was treated with hemalum solution for 1 min, then rinsed with tap water. The tissue section was dehydrated in 70% and 100% ethanol baths after being dyed in phloxine 0.1% solution for 10 seconds. The section was cleaned in xylene, cleaned twice in alcohol, dipped in saffron for 5 seconds then mounted with cover slips and the EUKITT slide mounting media. Using the Panoramic MIDI slide scanner, the stained slide was scanned to capture a digital image (3DHISTECH LTD. Budapest, Hungary) and the images were viewed and exported using QuPath 0.4.4.<sup>61</sup>. On tissues that had been stained, annotations were made.

### Sample preparation

All the FFPE blocks were sectioned into 8  $\mu\text{m}$  tissue sections using a microtome (Leica Biosystems, Wetzlar, Germany) at room temperature. A dewaxing step was performed through two 5 min washes in a xylene solution. Then, each dewaxed tissue section of the retrospective glioblastoma cohort was manually sprayed with a glycerol/isopropyl alcohol (IPA) (8:2, v/v) solution in two successive passes using a manual sprayer (Agilent). The syringe pump (74900 series Cole Parmer Instrument Company) was set to a 500  $\mu\text{L}/\text{min}$  flow rate. As for fresh frozen tissues, a Leica CM1510S cryostat (Leica Microsystems, Nanterre, France) was used to cut 20  $\mu\text{m}$  sections. Their analysis by the SpiderMass does not require any sample preparation.

### SpiderMass analysis

The overall layout of the instrument setup has already been covered elsewhere<sup>30</sup>. In brief, the system is made up of three parts: the mass spectrometer itself, a laser system for remote micro-sampling of tissues and a transfer line allowing for the transfer of the micro-sampled material. The first component consists of a pulsed Nd:YAG laser (pulse duration: 5 ns, = 1064 nm, Quantel, Les Ulis, France) pumping a tunable wavelength OPO (Radiant version 1.0.1, OPOTEK Inc., Carlsbad, CA, USA). A handpiece with a 4 cm focusing lens is attached to the end of the biocompatible laser fiber, which is connected to the laser system output and has an inner diameter of 450 microns and a length of 1 m. In these studies, the laser intensity was set to 4 mJ/pulse for a fixed irradiation time of 10 s, resulting in a laser fluence of approximately 3 J/cm<sup>2</sup>. The second component of the system is a 2 m transfer line made of Tygon ND 100-65 tubing (Akron, Ohio, USA, 2.4 mm inner diameter, 4 mm outer diameter). The transfer line is directly connected to the mass spectrometer (Xevo G2-XS, Waters, Manchester, United Kingdom) from which the conventional electrospray source was removed and replaced by a REIMS interface on one side and is attached to the laser handpiece on the other. A 200  $\mu\text{L}/\text{min}$  infusion of isopropanol was administered before each acquisition. 200  $\mu\text{g}/\text{mL}$  of Leucine enkephalin was added to the infusion to play the role of a lockmass. The sampling position was determined based on the histopathological annotations. The acquisition was composed of a burst of 10 laser shots resulting in an individual spectrum. Spectral acquisition was performed both in positive and negative ion mode in sensitivity mode with a scan time of 1 s. The mass range was set to  $m/z$  50-2000.

### SpiderMass MSI

The SpiderMass setup was described in the previous section. To perform imaging analysis, the SpiderMass microprobe was coupled to a stiff robotic arm described in a previous work<sup>62</sup>. The spatial step size was set to 250  $\mu\text{m}$  for fresh frozen tissue to achieve oversampling. The mass-range was fixed between  $m/z$  100-1500. The acquisition sequence was composed of 3 consecutive laser shots and 3 seconds between each step. The laser bursts and the spectrometer acquisition were automatically triggered through a MATLAB in-house user interface developed for the robotic WALDI-MSI<sup>62</sup>. The data was acquired in negative and sensitivity ion mode.

### **Cell line and immune cell analysis**

Human NCH82 stage 4 glioma cells were obtained from Dr Regnier-Vigouroux. These cells were cultured in DMEM. The medium was supplemented with 10% fetal bovine serum and 100 U/mL penicillin-streptomycin in a humidified air incubator at 37°C under an atmosphere of 5% CO<sub>2</sub>. After 70% confluence, cells were washed two times with DPBS, dried under PSM during 10min at RT than analyzed by SpiderMass directly into cell plate. First, to obtain primary macrophages, 50 mL of blood, from Etablissement Français du Sang (EFS), were diluted two times into PBS-EDTA than leucocytes were isolated thanks to 25 min centrifugation 2200 rpm with a Ficoll gradient. Leucocytes were washed three times with PBS-EDTA than resuspended into RPMI medium and incubated into a cell plate 1h30 at 37°C. Cell plate were washed three times with PBS. Macrophages were than grown 7 days with RPMI medium with 10% fetal bovine serum, 100 U/mL penicillin-streptomycin and MCSF. Next, primary macrophages were stimulated into two different conditions. M1-like macrophages were stimulated with 0.5 mg/mL of LPS and 20 ng/mL of IFN- $\gamma$  during 48h in contrary to M2-like macrophages that were stimulated with 20 ng/mL of Il4 during 48h. Like NCH82 cells, macrophages were washed two times with DPBS, dried under PSM during 10min at RT than analyzed by SpiderMass directly into cell plate. In addition, primary lymphocytes were isolated. For that, peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples using Ficoll density gradient centrifugation. Then, cells were labeled with mix antibodies (Sony): anti-CD3 PE (clone SK7), anti-CD4 FITC (clone A161A1), anti-CD8 APC (clone HIT8a) and anti-CD7 PE-Cy5 (clone CD7-6B7) for 20 minutes at 4°C in dark. After washing, CD3+CD4+ cells, D3+CD8+ cells and CD3-CD7+ cells were sorted using the BD FACS ARIA II SORP. One million of each population was transferred onto glass slides using a Cytospin™ centrifuge (Thermo Shandon Cytospin) and conserved at -80°C. The SpiderMass analysis were made directly onto the glass slide.

### **MALDI-MSI**

Nine prospective tumors were analyzed using MALDI-MSI. For this purpose, 12  $\mu$ m sections were cut using a Leica CM1510S cryostat (Leica Microsystems, Nanterre, France), and these sections were then placed on ITO-coated glass slides from LaserBioLabs (Valbonne, France). The application of the 2,5-dihydroxybenzoic acid (DHB) matrix was made thanks to a manual sprayer. The MALDI-MSI analyzes were carried out utilizing an Ultraflex II MALDI-TOF/TOF instrument (Bruker) operating in positive ion mode. The spatial resolution was adjusted at 70  $\mu$ m, and the mass range was fixed at  $m/z$  60-1000. Subsequently, the MALDI-MSI data was subjected to analysis employing SCiLS Lab software (SCiLS Lab 2022a PPro, SCiLS GmbH). The data was normalized using Total Ion Count (TIC) normalization. The segmentation of the images was then performed using the bisecting  $k$ -means algorithm, facilitating global and individual segmentation across the nine images. This comprehensive spatial segmentation enabled the identification of regions of interest, which were found to correspond with those obtained through proteomic data.

### **MALDI-IHC**

Multiplex imaging was conducted on two fresh-frozen GBM tissue samples that were previously analyzed using SpiderMass-MSI. One tissue sample had a survival rate of less than 10 months, while the other had an OS of more than 36 months. The MALDI-IHC analysis was made on an adjacent tissue section from the same tumor analysed by SpiderMass. The tissue preparation and imaging protocol utilized was the recommended one for AmberGen (BillERICA, MA) Miralys™ probes. Initially, the tissues were vacuum-dried for 10 minutes and then fixed with 1% PFA for 30 minutes. Subsequently, the tissues underwent a series of baths: one bath in PBS for 10 minutes, two baths in acetone for 3 minutes each, and one bath in Carnoy solution for 3 minutes. Following this, the tissues were rehydrated through two baths in 100% ethanol for 2 minutes each, succeeded by three consecutive baths in 95% EtOH, 70% EtOH, and 50% EtOH, each for 3 minutes. A 10-minute TBS bath (50 mM Tris, pH 7.5, 200 mM NaCl)

preceded antigen retrieval, which occurred in 20 mM Tris buffer at pH 9 for 30 minutes at 95°C. After a 10-minute TBS wash, the tissues were treated with a tissue blocking buffer (2% each of normal mouse and rabbit serum and 5% BSA in TBS-OG [TBS with 0.05% w/v Octyl  $\beta$ -D-glucopyranoside]) for 1 hour. Following this, the tissues were incubated at 4°C overnight in the same blocking buffer, which contained CD68, CD8 $\alpha$ , Ki67, Vimentin, and collagen probes at a concentration of 2.5  $\mu$ g/mL. Each slide was individually washed with three 5-minute TBS baths and three 2-minute baths in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, all conducted in darkness. The tissues were then vacuum-dried for 1 hour and 30 minutes at room temperature before subjecting them to a 365 nm UV exposure (Miralys™ Light Box from AmberGen, Inc., Billerica, MA) for 10 minutes to cleave the probes. DHB matrix at a concentration of 20 mg/mL in MeOH:TFA 0.1% (70:30, v/v) was sprayed onto the tissues using the HTX sprayer M5 from HTX technologies, LLC (Chappel Hill, NC). The two slides were subjected to MALDI-MSI analysis using a rapifleX MALDI-TOF-MS instrument (Bruker Daltonics, Germany) in reflector mode, positive ion mode, with a laser spot size of 20  $\mu$ m and continuous raster scanning of 20  $\mu$ m. Each pixel underwent 500 laser shots, and a TIC normalization was employed for multiplex image processing. The resulting images were analyzed using flexImaging (Bruker Daltonics, Billerica, MA).

## Data analysis

### *Preprocessing and data importation.*

Before obtaining the matrix data for each dataset, several preprocessing steps were applied. First, the data was binned to 0.1 Da to reduce the number of data points. Then, all the spectra were aligned based on the lockmass ( $m/z$  554.2 and 556.2 in negative and positive ion mode respectively) and the mass range was set between 600 and 1100  $m/z$ . Finally, the total ion current (TIC) normalization was applied to remove any intensity variations between spectra. All final data sets contain 5000  $m/z$  data points. For the data importation, there are two methods to obtain the matrix data from Waters RAW files. The first method involved a conversion of the raw files to mzML files using the MSConvert (Proteowizard)<sup>63</sup> and an importation of the mzML files to Python using the pyopenMS library. The second method, mainly used, included the importation of the raw files into “Abstract Model Builder” - AMX (version 1.0.2053.0, Waters research center, Hungary) and the extraction of the matrix data as csv files. The pandas library is used to import the csv files into Python.

### *Optimal classification model, cross-validation and blind prediction.*

The Lazy Predict library (<https://lazypredict.readthedocs.io/en/latest/>) was used to build multiple models from the scikit-learn library by training and testing a range of 24 classifiers. The random state was always kept at 1. Subsequently, the optimal model was reconstructed individually using the scikit-learn library, which enabled tuning of its parameters for optimization and evaluation of its accuracy. To further evaluate the model's performance, 20-fold cross-validation was performed using KFold and cross\_val\_score functions, and the classification report was generated using the classification\_report function. Additionally, the ConfusionMatrixDisplay function from the matplotlib library was used to display the confusion matrix. The optimal model was then saved and loaded for blind prediction using the joblib library, with the joblib.dump and joblib.load functions.

### *Prediction explanations.*

The Local Interpretable Model-agnostic Explanations (LIME) algorithm<sup>36</sup> was employed to explain the classification model. This algorithm calculates feature contributions that can be positive or negative. The ELI5 library (<https://eli5.readthedocs.io/en/latest/overview.html>) was utilized to generate a LIME table containing the weight of feature contributions using the explain\_prediction function.

### *Significant features.*

A peak picking algorithm, the `find_peaks_cwt` function from the `scipy` library was applied to remove instrument noise. Then a clustering heatmap was generated based on the peak list. The heatmap used hierarchical clustering between  $m/z$  variables with euclidean distance, displaying the ions that were more or less abundant in each target clustered together. For this purpose, the `clustermap` function from the `seaborn` library was utilized. Next, a non-parametric statistical test Kruskal-Wallis with Bonferroni correction, employing the `stats.kruskal` function from the `Scipy` library, was used to evaluate the significance of each feature. Only significant features with a p-value of equal or less than 0.05 were retained. Finally, a step of filtering was added to only keep the mono-isotopic peaks. The corresponding box plots were then generated from the `seaborn` library.

### **Immunoscores**

The immunoscore model was trained using the LGBM (Light Gradient Boosting Machine)<sup>46</sup> Python library, a gradient boosting framework developed by Microsoft. This model leveraged cell spectra within the  $m/z$  range of 600 to 1100 in negative ion mode. The spectra were categorized into different cell types: macrophages (M1 & M2) with 107 spectra, cancer cell line (NCH82) with 107 spectra and lymphocytes (NK, CD8, CD4) with 107 spectra. Due to the lack of spectra for each type of lymphocyte and to prevent creating an inconsistent (unbalanced classes) model, the lymphocytes were all regrouped. To predict the cell types on SpiderMass images, the `predict_proba` function of the LGBM model was utilized. This function provided probability estimates for each cell type, allowing for a more nuanced understanding of the likelihood of each cell type's presence in the local environment. Furthermore, ratio scores were computed to estimate the probability of presence over the whole sample, which is interpreted as a relative presence of each cell type across the entire image. These scores were calculated by summing the scores provided for each cell type and dividing it by the sum of the total scores across all labels. The ratios provided insights into the distribution of the trained cell types throughout the image, offering a comprehensive assessment of the cellular landscape in the analyzed sample.

### **MALDI-IHC quantification**

The Python PIL library was employed to bring in the IHC image in PNG format. Subsequently, all the images underwent a grayscale conversion and were transformed into NumPy arrays with the help of the NumPy library. An algorithm was designed and applied to compute the count of black pixels in the image using the NumPy array, where each channel ranges from 0 (representing black) to 255 (representing white). The quantification of each IHC marker was determined by subtracting the count of black pixels from the total image pixels. To visualize the image and generate a bar plot, the `matplotlib.pyplot` library was utilized.

### **MS/MS analysis**

SpiderMass technology facilitated the MS/MS investigation using the Xevo G2-XS instrument from Waters. MS/MS spectra were recorded after the isolation of the precursor ion and they were subjected to collision-induced dissociation (CID) in the transfer cell with a transfer collision energy ranging between 30 to 40 V, depending on the selected precursor ion. The lipid annotations were performed manually through fragmentation spectra guidelines and compared to LipidMaps database, Alex123, MetFrag database<sup>64</sup> and literature.

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

Three datasets were employed for training, cross-validation, and testing of all classification models, including those for tissue type, tissue region, prognosis, and immunoscore. Evaluation involved a 20% validation split and 20-fold cross-validation, with a classification report providing metrics such as accuracy, recall, precision, and F1 score. To assess the statistical significance of biomarkers, a non-parametric Kruskal-Wallis test was employed. Bonferroni corrections were applied to adjust p-values for multiple comparisons. Values are presented as medians and visualized through scatter box plots.

## Supplemental figure legends

**Figure S1.** H&E scans of the fifty FFPE and the 31 fresh frozen glioblastoma tissue from the retrospective cohort and their histological annotations, related to Figure 1.

**Table S1.** Demographic and clinical data of the retrospective and prospective cohort. Related to Figure 1.

**Table S2.** Number of MS spectra for each histological region (tumor, necrosis and benign brain tissue) for the 50 FFPE and 31 fresh frozen tissues of glioblastoma from the retrospective and prospective cohort. Related to Figure 1.

**Figure S2.** Identification of the optimal classification model on the retrospective FFPE glioblastoma cohort only, before data augmentation, in both ion modes, related to Figure 1. **(A)** Table of the 24 classifiers that were trained and evaluated in negative ion mode, showing their accuracies, balanced accuracy, F1 Score and time taken. RidgeClassifier came out on top with 93% of accuracy. **(B)** Classification report and confusion matrix of RidgeClassifier with the training set in negative ion mode. **(C)** Classification report and confusion matrix of RidgeClassifier after 20-fold cross-validation in negative ion mode. **(D)** Table of the 24 classifiers that were trained and evaluated in positive ion mode, showing their accuracies, balanced accuracy, F1 score, and time taken. RidgeClassifier came out on top with 92% of accuracy. **(E)** Classification report and confusion matrix of RidgeClassifier with the training set in positive ion mode. **(F)** Classification report and confusion matrix of RidgeClassifier after 20-fold cross-validation in positive ion mode.

**Figure S3.** Averaged SpiderMass MS spectra obtained from the glioblastoma tissue in the lipid mass range ( $m/z$  600-1100), related to Figure 1. **(A)** Averaged spectra for benign, necrosis and tumor tissue in negative ion mode for FFPE tissue. **(B)** Averaged spectra for benign, necrosis and tumor tissue in positive ion mode for FFPE tissue. **(C)** Averaged spectra for necrosis and tumor tissue in negative ion mode for fresh frozen tissue. **(D)** Averaged spectra for necrosis and tumor tissue in positive ion mode for fresh and frozen tissue.

**Table S3.** Accuracy, standard deviation and F1 scores obtained with Ridge classifier before and after data augmentation in positive ion mode, related to Figure 1.

**Figure S4.** Supervised and unsupervised marker discovery result in negative and positive ion mode, related to Figure 1. **(A)** and **(C)** Top 120  $m/z$  features that contribute positively (green) or negatively (red) for the classification of each tissue in the model in positive and negative ion mode respectively. **(B)** and **(D)** Heatmap of picked peaks for three different ROIs in positive and negative ion mode. The heatmap displays the intensity of each ion, with red indicating overexpression and green indicating under-expression.

**Table S4.** Total of 41 confident biomarkers for benign, necrotic and tumor tissue in both negative and positive ion mode, Related to Figure 1. Red and green are showing the under-expressed and over-expressed ions respectively. Signs (++, +, -, --) corresponding to the contribution weight of the  $m/z$  in each type of tissue found with LIME algorithm.

**Figure S5.** Corresponding box plot of the 31 remaining biomarkers specific for necrosis, benign and tumor tissue in negative and positive ion mode, related to Figure 1.

**Table S5.** Annotation of the 48 biomarkers for poor and good prognosis in addition to the 27 biomarkers specific of cancer and immune cells (lymphocytes, M1-like and M2-like macrophages). Green indicates for which class the ion is specific. Related to Figures 3, 4 and 6.

**Table S6.** Comparison of the calculated specificity and sensitivity for different technics used for intraoperative analysis in brain tumors. Related to Figure 1.

**Figure S6.** Preliminary results from an ongoing study on transgenic TgC3(1) mice models spontaneously developing breast cancer and demonstrating the application of the cell scoring prediction pipeline for determining the ratio of cancer versus normal cells in the objective to improve tumor margin delineation. The final predicted labels are based on a score threshold  $> 0.5$ . Related to Figure 4.



## References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* 71, 209–249. 10.3322/caac.21660.
2. Schaff, L.R., and Mellingshoff, I.K. (2023). Glioblastoma and Other Primary Brain Malignancies in Adults: A Review. *JAMA* 329, 574. 10.1001/jama.2023.0023.
3. Hassn Mesrati, M., Behrooz, A.B., Y. Abuhamad, A., and Syahir, A. (2020). Understanding Glioblastoma Biomarkers: Knocking a Mountain with a Hammer. *Cells* 9, 1236. 10.3390/cells9051236.
4. Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 131, 803–820. 10.1007/s00401-016-1545-1.
5. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* 23, 1231–1251. 10.1093/neuonc/noab106.
6. Yan, H., McLendon, R., Kos, I., Riggins, G.J., Reardon, D., Velculescu, V.E., and Bigner, D.D. (2009). IDH1 and IDH2 Mutations in Gliomas. *The new england journal of medicine*.
7. Fernandes, C., Costa, A., Osório, L., Lago, R.C., Linhares, P., Carvalho, B., and Caeiro, C. (2017). Current Standards of Care in Glioblastoma Therapy. In *Glioblastoma*, S. De Vleeschouwer, ed. (Codon Publications).
8. Thakkar, J.P., Dolecek, T.A., Horbinski, C., Ostrom, Q.T., Lightner, D.D., Barnholtz-Sloan, J.S., and Villano, J.L. (2014). Epidemiologic and Molecular Prognostic Review of Glioblastoma. *Cancer Epidemiol Biomarkers Prev* 23, 1985–1996. 10.1158/1055-9965.EPI-14-0275.
9. Stupp, R., Mason, W.P., Van Den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *N Engl J Med* 352, 987–996. 10.1056/NEJMoa043330.
10. Le Rhun, E., Preusser, M., Roth, P., Reardon, D.A., Van Den Bent, M., Wen, P., Reifenberger, G., and Weller, M. (2019). Molecular targeted therapy of glioblastoma. *Cancer Treatment Reviews* 80, 101896. 10.1016/j.ctrv.2019.101896.
11. Van Tellingen, O., Yetkin-Arik, B., De Gooijer, M.C., Wesseling, P., Wurdinger, T., and De Vries, H.E. (2015). Overcoming the blood–brain tumor barrier for effective glioblastoma treatment. *Drug Resistance Updates* 19, 1–12. 10.1016/j.drug.2015.02.002.
12. Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., Hedrick, C.C., et al. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med* 24, 541–550. 10.1038/s41591-018-0014-x.
13. González- Tablas Pimenta, M., Otero, Á., Arandia Guzman, D.A., Pascual- Argente, D., Ruíz Martín, L., Sousa- Casasnovas, P., García- Martín, A., Roa Montes De Oca, J.C., Villaseñor- Ledezma, J., Torres Carretero, L., et al. (2021). Tumor cell and immune cell profiles in primary human glioblastoma: Impact on patient outcome. *Brain Pathology* 31, 365–380. 10.1111/bpa.12927.
14. Brown, T.J., Brennan, M.C., Li, M., Church, E.W., Brandmeir, N.J., Rakszawski, K.L., Patel, A.S., Rizk, E.B., Suki, D., Sawaya, R., et al. (2016). Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis. *JAMA Oncol* 2, 1460. 10.1001/jamaoncol.2016.1373.
15. Weller, M., Van Den Bent, M., Preusser, M., Le Rhun, E., Tonn, J.C., Minniti, G., Bendszus, M.,

- Balana, C., Chinot, O., Dirven, L., et al. (2021). EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. *Nat Rev Clin Oncol* 18, 170–186. 10.1038/s41571-020-00447-z.
16. Ma, R., Taphoorn, M.J.B., and Plaha, P. (2021). Advances in the management of glioblastoma. *J Neurol Neurosurg Psychiatry* 92, 1103–1111. 10.1136/jnnp-2020-325334.
  17. Livermore, L.J., Isabelle, M., Bell, I.M., Edgar, O., Voets, N.L., Stacey, R., Ansorge, O., Vallance, C., and Plaha, P. (2020). Raman spectroscopy to differentiate between fresh tissue samples of glioma and normal brain: a comparison with 5-ALA–induced fluorescence-guided surgery. *Journal of Neurosurgery*, 1–11. 10.3171/2020.5.JNS20376.
  18. Livermore, L.J., Isabelle, M., Bell, I.M., Scott, C., Walsby-Tickle, J., Gannon, J., Plaha, P., Vallance, C., and Ansorge, O. (2019). Rapid intraoperative molecular genetic classification of gliomas using Raman spectroscopy. *Neuro-Oncology Advances* 1, vdz008. 10.1093/oaajnl/vdz008.
  19. Jermyn, M., Mok, K., Mercier, J., Desroches, J., Pichette, J., Saint-Arnaud, K., Bernstein, L., Guiot, M.-C., Petrecca, K., and Leblond, F. (2015). Intraoperative brain cancer detection with Raman spectroscopy in humans. *Sci. Transl. Med.* 7. 10.1126/scitranslmed.aaa2384.
  20. Agar, N.Y.R., Golby, A.J., Ligon, K.L., Norton, I., Mohan, V., Wiseman, J.M., Tannenbaum, A., and Jolesz, F.A. (2011). Development of Stereotactic Mass Spectrometry for Brain Tumor Surgery. *Neurosurgery* 68, 280–290. 10.1227/NEU.0b013e3181ff9cbb.
  21. Eberlin, L.S., Norton, I., Orringer, D., Dunn, I.F., Liu, X., Ide, J.L., Jarmusch, A.K., Ligon, K.L., Jolesz, F.A., Golby, A.J., et al. (2013). Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1611–1616. 10.1073/pnas.1215687110.
  22. Rose, M., Cardon, T., Aboulouard, S., Hajjaji, N., Kobeissy, F., Duhamel, M., Fournier, I., and Salzet, M. (2021). Surfaceome Proteomic of Glioblastoma Revealed Potential Targets for Immunotherapy. *Front Immunol* 12, 746168. 10.3389/fimmu.2021.746168.
  23. Duhamel, M., Drelich, L., Wisztorski, M., Aboulouard, S., Gimeno, J.-P., Ogrinc, N., Devos, P., Cardon, T., Weller, M., Escande, F., et al. (2022). Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. *Nat Commun* 13, 6665. 10.1038/s41467-022-34208-6.
  24. Ogrinc, N., Saudemont, P., Takats, Z., Salzet, M., and Fournier, I. (2021). Cancer Surgery 2.0: Guidance by Real-Time Molecular Technologies. *Trends in Molecular Medicine* 27, 602–615. 10.1016/j.molmed.2021.04.001.
  25. Pirro, V., Alfaro, C.M., Jarmusch, A.K., Hattab, E.M., Cohen-Gadol, A.A., and Cooks, R.G. (2017). Intraoperative assessment of tumor margins during glioma resection by desorption electrospray ionization-mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6700–6705. 10.1073/pnas.1706459114.
  26. Tzafetas, M., Mitra, A., Paraskevaidi, M., Bodai, Z., Kalliala, I., Bowden, S., Lathouras, K., Rosini, F., Szasz, M., Savage, A., et al. (2020). The intelligent knife (iKnife) and its intraoperative diagnostic advantage for the treatment of cervical disease. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7338–7346. 10.1073/pnas.1916960117.
  27. Saudemont, P., Quanico, J., Robin, Y.-M., Baud, A., Balog, J., Fatou, B., Tierny, D., Pascal, Q., Minier, K., Pottier, M., et al. (2018). Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption/Ionization Mass Spectrometry Technology. *Cancer Cell* 34, 840–851.e4. 10.1016/j.ccell.2018.09.009.
  28. Calligaris, D., Norton, I., Feldman, D.R., Ide, J.L., Dunn, I.F., Eberlin, L.S., Cooks, R.G., Jolesz, F.A., Golby, A.J., Santagata, S., et al. (2013). Mass Spectrometry Imaging as a Tool for Surgical Decision-Making. *J Mass Spectrom* 48, 1178–1187. 10.1002/jms.3295.
  29. Ifa, D.R., and Eberlin, L.S. (2016). Ambient Ionization Mass Spectrometry for Cancer Diagnosis and Surgical Margin Evaluation. *Clinical Chemistry* 62, 111–123. 10.1373/clinchem.2014.237172.
  30. Ogrinc, N., Saudemont, P., Balog, J., Robin, Y.-M., Gimeno, J.-P., Pascal, Q., Tierny, D., Takats,

- Z., Salzet, M., and Fournier, I. (2019). Water-assisted laser desorption/ionization mass spectrometry for minimally invasive in vivo and real-time surface analysis using SpiderMass. *Nat Protoc* *14*, 3162–3182. 10.1038/s41596-019-0217-8.
31. Ogrinc, N., Caux, P.-D., Robin, Y.-M., Bouchaert, E., Fatou, B., Ziskind, M., Focsa, C., Bertin, D., Tierny, D., Takats, Z., et al. (2021). Direct Water-Assisted Laser Desorption/Ionization Mass Spectrometry Lipidomic Analysis and Classification of Formalin-Fixed Paraffin-Embedded Sarcoma Tissues without Dewaxing. *Clin Chem* *67*, 1513–1523. 10.1093/clinchem/hvab160.
  32. Lee, E.S., and Durant, T.J.S. (2022). Supervised machine learning in the mass spectrometry laboratory: A tutorial. *Journal of Mass Spectrometry and Advances in the Clinical Lab* *23*, 1–6. 10.1016/j.jmsacl.2021.12.001.
  33. Seddiki, K., Saudemont, P., Precioso, F., Ogrinc, N., Wisztorski, M., Salzet, M., Fournier, I., and Droit, A. (2020). Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun* *11*, 5595. 10.1038/s41467-020-19354-z.
  34. Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)* *23*, 18. 10.3390/e23010018.
  35. Gunning, D., and Aha, D.W. DARPA’s Explainable Artificial Intelligence Program. *AI Magazine* *40*(2), 44–58. 10.1609/aimag.v40i2.2850.
  36. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. *KDD’16*, 1135–1144. 10.1145/2939672.2939778
  37. King, M.E., Yuan, R., Chen, J., Pradhan, K., Sariol, I., Li, S., Chakraborty, A., Ekpenyong, O., Yearley, J.H., Wong, J.C., et al. (2023). Long-chain polyunsaturated lipids associated with responsiveness to anti-PD-1 therapy are colocalized with immune infiltrates in the tumor microenvironment. *Journal of Biological Chemistry* *299*, 102902. 10.1016/j.jbc.2023.102902.
  38. Ledoux, L., Zirem, Y., Renaud, F., Duponchel, L., Salzet, M., Ogrinc, N., and Fournier, I. (2023). Comparing MS imaging of lipids by WALDI and MALDI: two technologies for evaluating a common ground truth in MS imaging. *Analyst* *148*, 4982–4986. 10.1039/D3AN01096A.
  39. Lizarbe, M., Barrasa, J., Olmo, N., Gavilanes, F., and Turnay, J. (2013). Annexin-Phospholipid Interactions. Functional Implications. *IJMS* *14*, 2652–2683. 10.3390/ijms14022652.
  40. Kiss, A., Erdödi, F., and Lontay, B. (2019). Myosin phosphatase: Unexpected functions of a long-known enzyme. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* *1866*, 2–15. 10.1016/j.bbamcr.2018.07.023.
  41. Saddoughi, S.A., and Ogretmen, B. (2013). Diverse Functions of Ceramide in Cancer Cell Death and Proliferation. In *Advances in Cancer Research* (Elsevier), pp. 37–58. 10.1016/B978-0-12-394274-6.00002-9.
  42. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* *313*, 1960–1964. 10.1126/science.1129139.
  43. Mlecnik, B., Bifulco, C., Bindea, G., Marliot, F., Lugli, A., Lee, J.J., Zlobec, I., Rau, T.T., Berger, M.D., Nagtegaal, I.D., et al. (2020). Multicenter International Society for Immunotherapy of Cancer Study of the Consensus Immunoscore for the Prediction of Survival and Response to Chemotherapy in Stage III Colon Cancer. *JCO* *38*, 3638–3651. 10.1200/JCO.19.03205.
  44. Marliot, F., Chen, X., Kirilovsky, A., Sbarrato, T., El Sissy, C., Batista, L., Van Den Eynde, M., Haicheur-Adjouri, N., Anitei, M.-G., Musina, A.-M., et al. (2020). Analytical validation of the Immunoscore and its associated prognostic value in patients with colon cancer. *J Immunother Cancer* *8*, e000272. 10.1136/jitc-2019-000272.
  45. Hijazi, A., Antoniotti, C., Cremolini, C., and Galon, J. (2023). Light on life: immunoscore immune-

- checkpoint, a predictor of immunotherapy response. *OncoImmunology* *12*, 2243169. 10.1080/2162402X.2023.2243169.
46. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in neural information processing systems* *30*,
  47. Yagnik, G., Liu, Z., Rothschild, K.J., and Lim, M.J. (2021). Highly Multiplexed Immunohistochemical MALDI-MS Imaging of Biomarkers in Tissues. *J. Am. Soc. Mass Spectrom.* *32*, 977–988. 10.1021/jasms.0c00473.
  48. Lim, M.J., Yagnik, G., Henkel, C., Frost, S.F., Bien, T., and Rothschild, K.J. (2023). MALDI HiPLEX-IHC: multiomic and multimodal imaging of targeted intact proteins in tissues. *Front. Chem.* *11*, 1182404. 10.3389/fchem.2023.1182404.
  49. Gigante, E., Cazier, H., Albuquerque, M., Laouirem, S., Beaufrère, A., and Paradis, V. (2023). MALDI Imaging, a Powerful Multiplex Approach to Decipher Intratumoral Heterogeneity: Combined Hepato-Cholangiocarcinomas as Proof of Concept. *Cancers* *15*, 2143. 10.3390/cancers15072143.
  50. Balog, J., Sasi-Szabó, L., Kinross, J., Lewis, M.R., Muirhead, L.J., Veselkov, K., Mirnezami, R., Dezső, B., Damjanovich, L., Darzi, A., et al. (2013). Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Sci. Transl. Med.* *5*. 10.1126/scitranslmed.3005623.
  51. Eberlin, L.S., Tibshirani, R.J., Zhang, J., Longacre, T.A., Berry, G.J., Bingham, D.B., Norton, J.A., Zare, R.N., and Poultsides, G.A. (2014). Molecular assessment of surgical-resection margins of gastric cancer by mass-spectrometric imaging. *Proceedings of the National Academy of Sciences* *111*, 2436–2441. 10.1073/pnas.1400274111.
  52. Dijkstra, T.K. (2014). Ridge regression and its degrees of freedom. *Qual Quant* *48*, 3185–3193. 10.1007/s11135-013-9949-7.
  53. Cortes, C., Mohri, M., and Rostamizadeh, A. (2009). L2 Regularization for Learning Kernels.
  54. Van Hese, L., De Vleeschouwer, S., Theys, T., Rex, S., Heeren, R.M.A., and Cuypers, E. (2022). The diagnostic accuracy of intraoperative differentiation and delineation techniques in brain tumours. *Discover Oncol.* *13(1)*, 123. 10.1007/s12672-022-00585-z.
  55. Blanco, V.M., Chu, Z., LaSance, K., Gray, B.D., Pak, K.Y., Rider, T., Greis, K.D., and Qi, X. (2016). Optical and nuclear imaging of glioblastoma with phosphatidylserine-targeted nanovesicles. *Oncotarget* *7*, 32866–32875. 10.18632/oncotarget.8763.
  56. Sakane, F., Hoshino, F., Ebina, M., Sakai, H., and Takahashi, D. (2021). The Roles of Diacylglycerol Kinase  $\alpha$  in Cancer Cell Proliferation and Apoptosis. *Cancers* *13*, 5190. 10.3390/cancers13205190.
  57. Chen, Z., Feng, X., Herting, C.J., Garcia, V.A., Nie, K., Pong, W.W., Rasmussen, R., Dwivedi, B., Seby, S., Wolf, S.A., et al. (2017). Cellular and molecular identity of tumor-associated macrophages in glioblastoma. *Cancer Res* *77*, 2266–2278. 10.1158/0008-5472.CAN-16-2310.
  58. Zhang, H., Luo, Y.-B., Wu, W., Zhang, L., Wang, Z., Dai, Z., Feng, S., Cao, H., Cheng, Q., and Liu, Z. (2021). The molecular feature of macrophages in tumor immune microenvironment of glioma patients. *Computational and Structural Biotechnology Journal* *19*, 4603–4618. 10.1016/j.csbj.2021.08.019.
  59. Vidyarthi, A., Agnihotri, T., Khan, N., Singh, S., Tewari, M.K., Radotra, B.D., Chatterjee, D., and Agrewala, J.N. (2019). Predominance of M2 macrophages in gliomas leads to the suppression of local and systemic immunity. *Cancer Immunol Immunother* *68*, 1995–2004. 10.1007/s00262-019-02423-8.
  60. Michiba, A., Shiogama, K., Tsukamoto, T., Hirayama, M., Yamada, S., and Abe, M. (2022). Morphologic Analysis of M2 Macrophage in Glioblastoma: Involvement of Macrophage Extracellular Traps (METs). *Acta Histochem. Cytochem.* *55*, 111–118. 10.1267/ahc.22-00018.

61. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. *Sci Rep* 7, 16878. 10.1038/s41598-017-17204-5.
62. Ogrinc, N., Kruszewski, A., Chaillou, P., Saudemont, P., Lagadec, C., Salzet, M., Duriez, C., and Fournier, I. (2021). Robot-Assisted SpiderMass for In Vivo Real-Time Topography Mass Spectrometry Imaging. *Anal. Chem.* 93, 14383–14391. 10.1021/acs.analchem.1c01692.
63. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30, 918–920. 10.1038/nbt.2377.
64. Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., and Neumann, S. (2016). MetFrag re-launched: incorporating strategies beyond in silico fragmentation. *J Cheminform* 8, 3. 10.1186/s13321-016-0115-9.

## Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-CD68	Miralys™ PC-MT Probe	Cat#AP1001124
Anti-CD8 $\alpha$	Miralys™ PC-MT Probe	Cat#AP100152
Anti-Ki67	Miralys™ PC-MT Probe	Cat#AP1001184
Anti-VIM	Miralys™ PC-MT Probe	Cat#AP1001122
Anti-Collagen-1A1	Miralys™ PC-MT Probe	Cat#AP100153
Anti-CD3 $\epsilon$	Miralys™ PC-MT Probe	Cat#AP1001181
Anti-CD4	Miralys™ PC-MT Probe	Cat#AP100173
Anti-CD3 PE (clone SK7)	Sony Biotechnology	Cat#2324030
Anti-CD4 FITC (clone A161A1)	Sony Biotechnology	Cat#2387030
Anti-CD8 Alexa Fluor 647 (clone HIT8a)	Sony Biotechnology	Cat#2104590
Anti-CD7 PE Cy5 (clone CD7-6B7)	Sony Biotechnology	Cat#2315550
<b>Biological samples</b>		
Healthy and cancerous glioma tissues	University Hospital Center of Lille	<a href="https://www.chu-lille.fr">https://www.chu-lille.fr</a>
Blood samples for macrophages and lymphocytes	French blood agency (EFS)	<a href="https://dondesang.efs.sante.fr">https://dondesang.efs.sante.fr</a>
<b>Chemicals, peptides, and recombinant proteins</b>		
Hemalun	Merck	Cat#109249
Phloxin	VWR	Cat#10047229
Saffron	MM France	Cat#F/SAFRAN
Ethanol (EtOH)	Carlo Erba Reagents	Cat#4127012
Xylene	Carlo Erba Reagents	Cat#492301
Glycerol	Reidel de Haen	Cat#15523
Isopropanol (IPA)	Carlo Erba Reagents	Cat#415183
DMEM	Thermo Fischer Scientific	Cat#12-491-015
RPMI Medium	Thermo Fischer Scientific	Cat#31870025
DPBS, no calcium, no magnesium	Thermo Fischer Scientific	Cat#14190-094
Ficoll	Dominique Dutscher	Cat#17-5442-02
EDTA	Sigma Aldrich	Cat#03690
Leucine enkephalin	Waters	Cat#70002456
Phorbol-12-myristate-13-acetate (PMA)	Sigma Aldrich	Cat#P8139
Lypopolysachharide (LPS)	Invivogen	Cat#tlrl-3pelps
Macrophage colony stimulating factor (MCSF)	Ozyme	Cat#BLE574802
IFN $\gamma$	PeproTech	Cat#300-02
IL4	PeproTech	Cat#200-04
2,5-dihydroxybenzoic acid	Sigma Aldrich	Cat#149357

Methanol (MeOH)	Carlo Erba Reagents	Cat#4148831
Chloroform	Carlo Erba Reagents	Cat#508320
Trifluoroacetic (TFA)	Sigma Aldrich	Cat#T6508
Paraformaldehyde (PFA)	Alfa Aesar	Cat#43368
Acetone	VWR	Cat#20067.320
Acetic acid	Sigma Aldrich	Cat#320099
TRIS-HCl	Promega	Cat#H5125
TRIS-Base	Promega	Cat#H5135
NaCl	Fischer chemicals	Cat#S/3161/60
Octyl $\beta$ -D-glucopyranoside (OG)	Sigma Aldrich	Cat#03757
Ammonium bicarbonate (NH <sub>4</sub> HCO <sub>3</sub> )	Sigma Millipore	Cat#149357
<b>Deposited data</b>		
The raw data from SpiderMass analysis has been deposited to the Harvard dataverse site.	This paper	Harvard dataverse : <a href="https://doi.org/10.7910/DVN/RUULD8">https://doi.org/10.7910/DVN/RUULD8</a>
<b>Experimental models: Cell lines</b>		
Human NCH82 stage IV glioma	Collaboration with Dr Regnier-Vigouroux	N/A
<b>Software and algorithms</b>		
GraphPad Prism v9.5.1	GraphPad software	RRID:SCR_000306 <a href="https://www.graphpad.com">https://www.graphpad.com</a>
Abstract Model Builder (AMX) version 1.0.2053.0	Waters research center, Hungary	
Python 3.11.5	Open source	RRID:SCR_008394 <a href="https://www.python.org">https://www.python.org</a>
Jupyter Notebook 6.5.4	Open source	<a href="https://jupyter.org">https://jupyter.org</a>
FlexImaging 5.0	Bruker	<a href="https://www.bruker.com/fr/products-and-solutions/mass-spectrometry/maldi-tof/rapiflex-maldi-tissuetypeper.html">https://www.bruker.com/fr/products-and-solutions/mass-spectrometry/maldi-tof/rapiflex-maldi-tissuetypeper.html</a>
SCiLS Lab 2022a Pro	Bruker and SCiLS GmbH	RRID:SCR_014426 <a href="http://scils.de/software/">http://scils.de/software/</a>
QuPath 0.4.4	Bankhead et al., 2017. <sup>56</sup>	RRID:SCR_018257 <a href="https://qupath.github.io">https://qupath.github.io</a>
Matlab R2019a	MathWorks	RRID:SCR_001622 <a href="https://matlab.mathworks.com">https://matlab.mathworks.com</a>
MetFrag	Ruttkies et al., 2016. <sup>59</sup>	<a href="https://msbi.ipb-halle.de/MetFrag/">https://msbi.ipb-halle.de/MetFrag/</a>

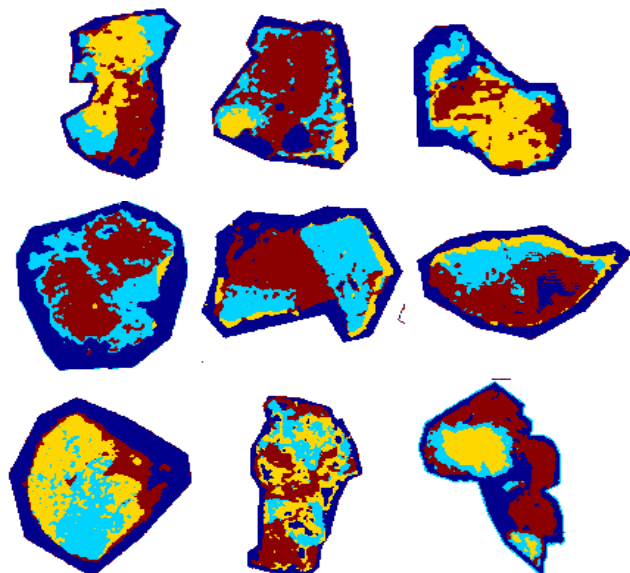


MassLynx 4.1	Waters	RRID:SCR_014271 <a href="https://www.waters.com/waters/fr_FR/Logiciel-de-Spectrometrie-de-Masse-MassLynx/nav.htm?cid=513662&amp;locale=fr_FR">https://www.waters.com/waters/fr_FR/Logiciel-de-Spectrometrie-de-Masse-MassLynx/nav.htm?cid=513662&amp;locale=fr_FR</a>
The original code has been deposited on GitHub. The O-DAPIA script for one dimensional SpiderMass data were released and archived in Zenodo	This paper	<a href="https://zenodo.org/doi/10.5281/zenodo.10656851">https://zenodo.org/doi/10.5281/zenodo.10656851</a>
The original code has been deposited on GitHub. The T-DAPIA script for SpiderMassMSI data were released and archived in Zenodo	This paper	<a href="https://zenodo.org/doi/10.5281/zenodo.10656830">https://zenodo.org/doi/10.5281/zenodo.10656830</a>

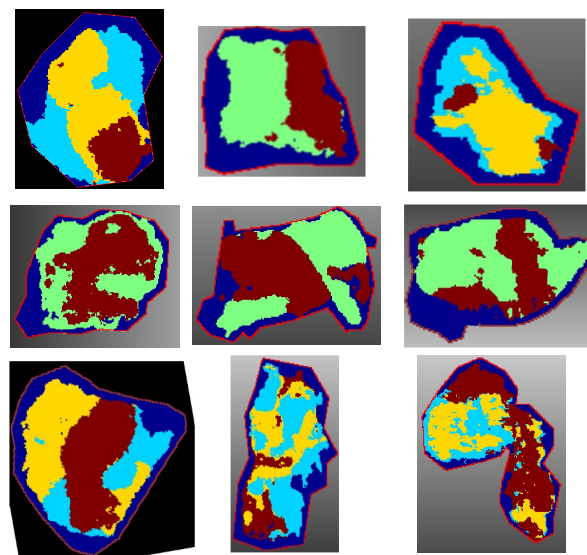


**A**

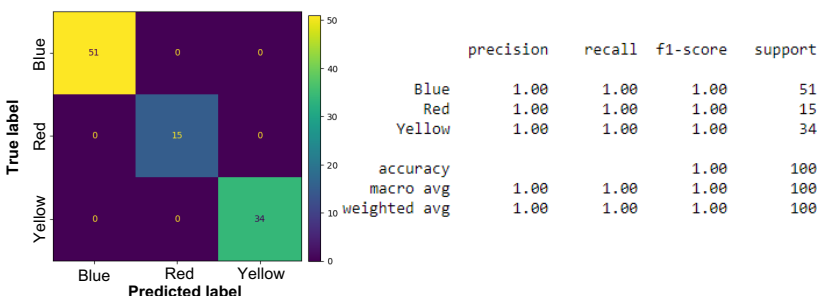
Segmentation based on lipids



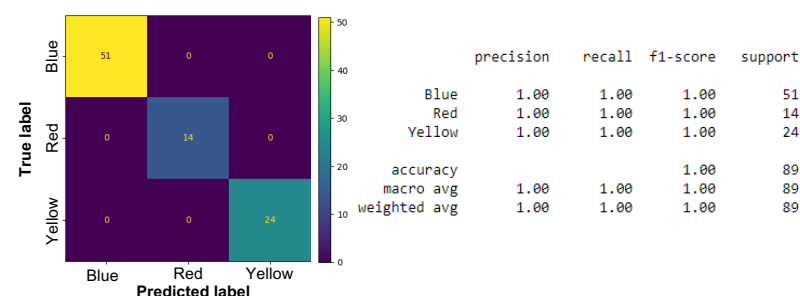
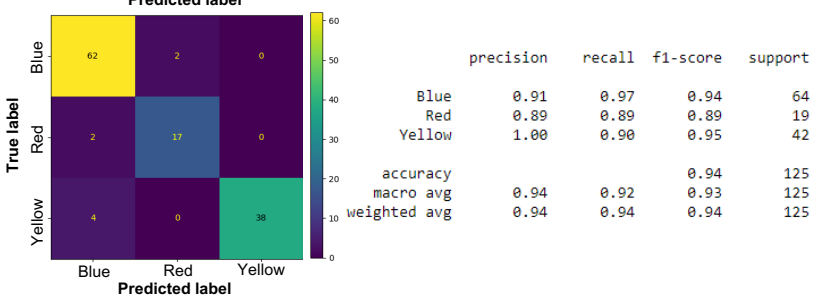
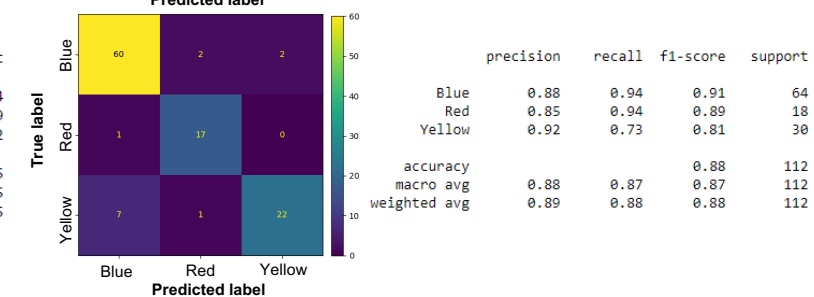
Segmentation based on peptides

**B**

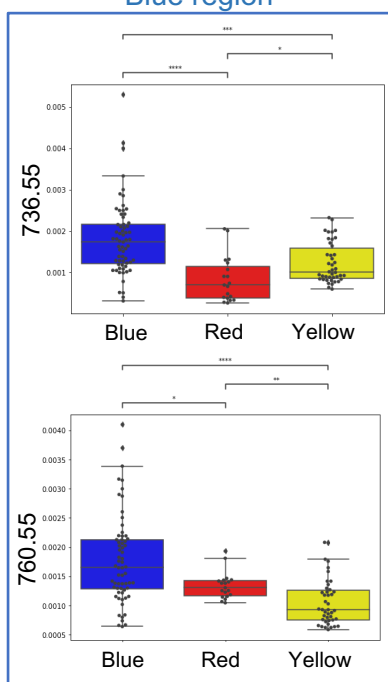
Negative ion mode

**C**

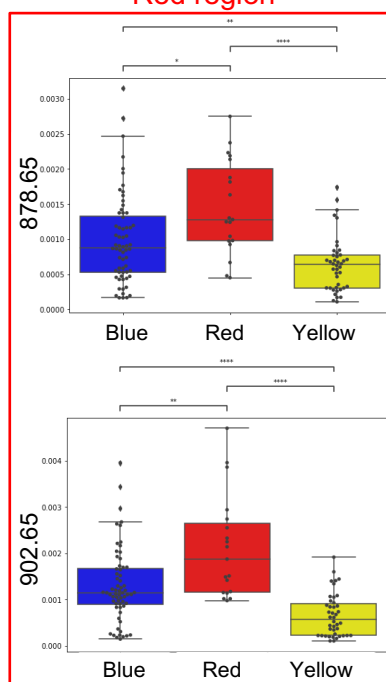
Positive ion mode

**B****C****D**

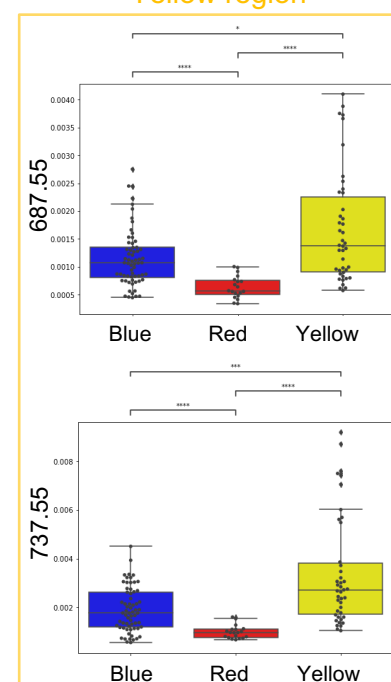
Blue region

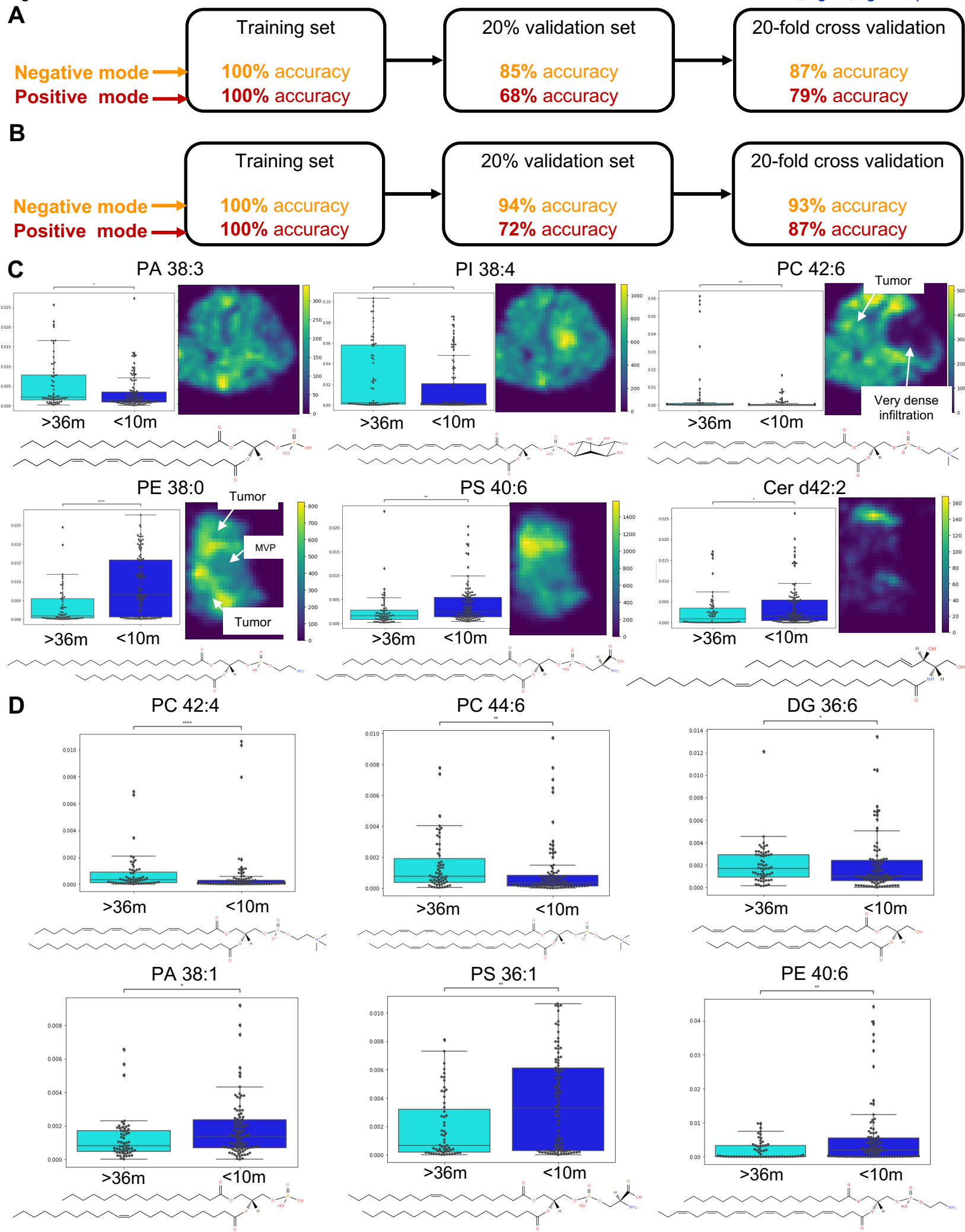


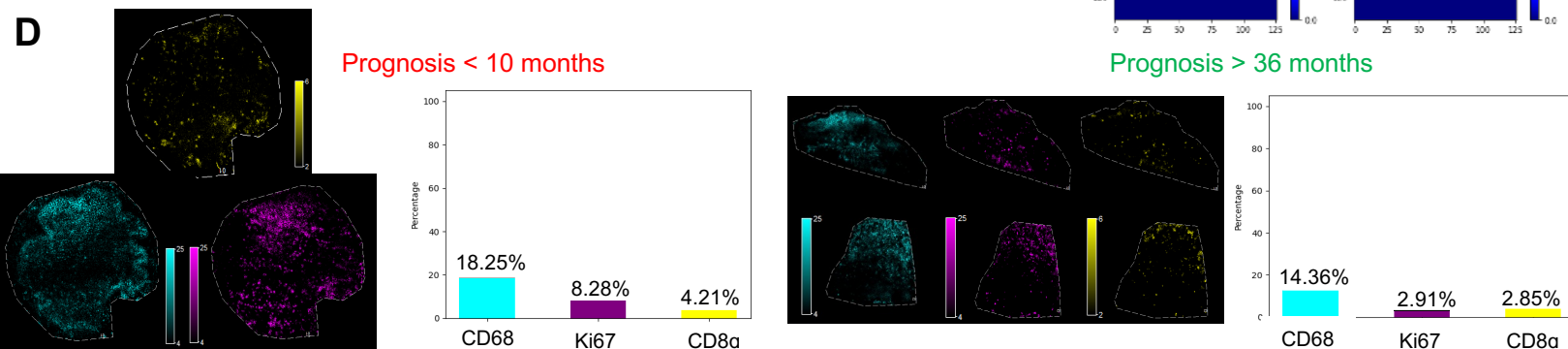
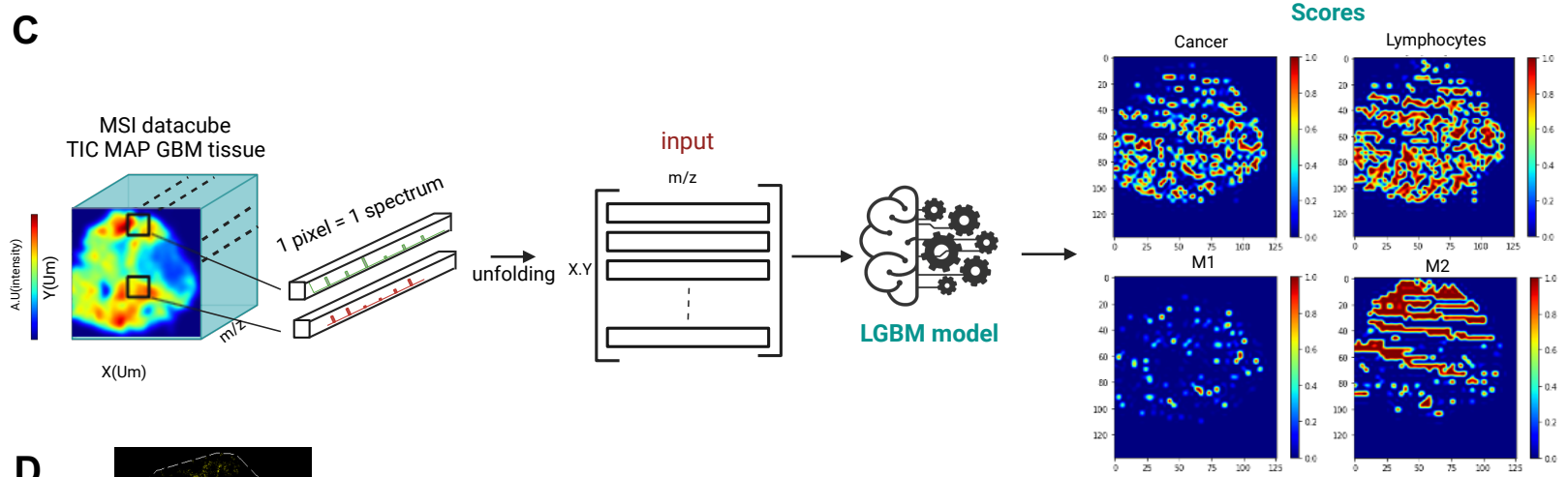
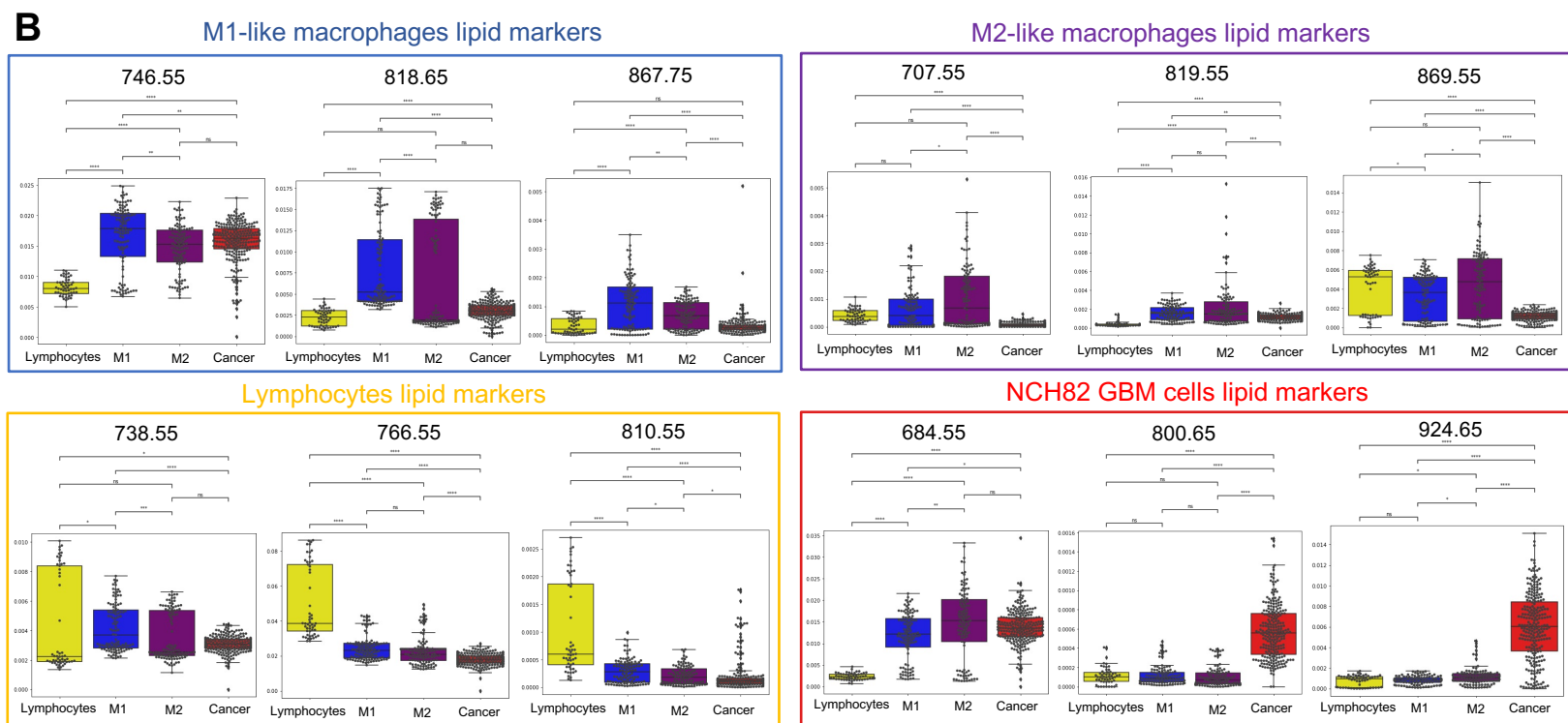
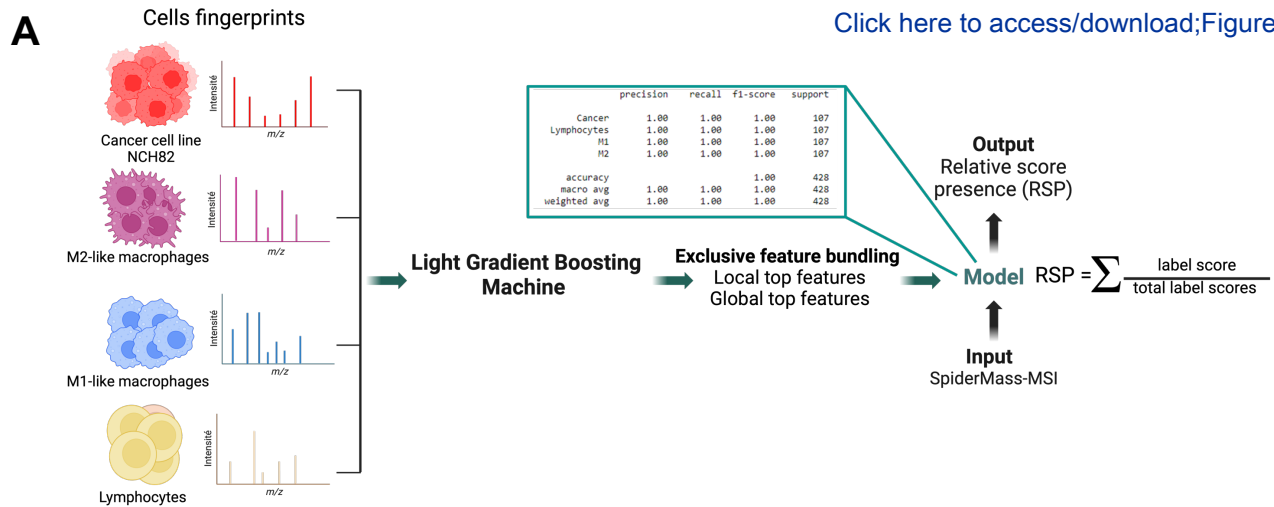
Red region



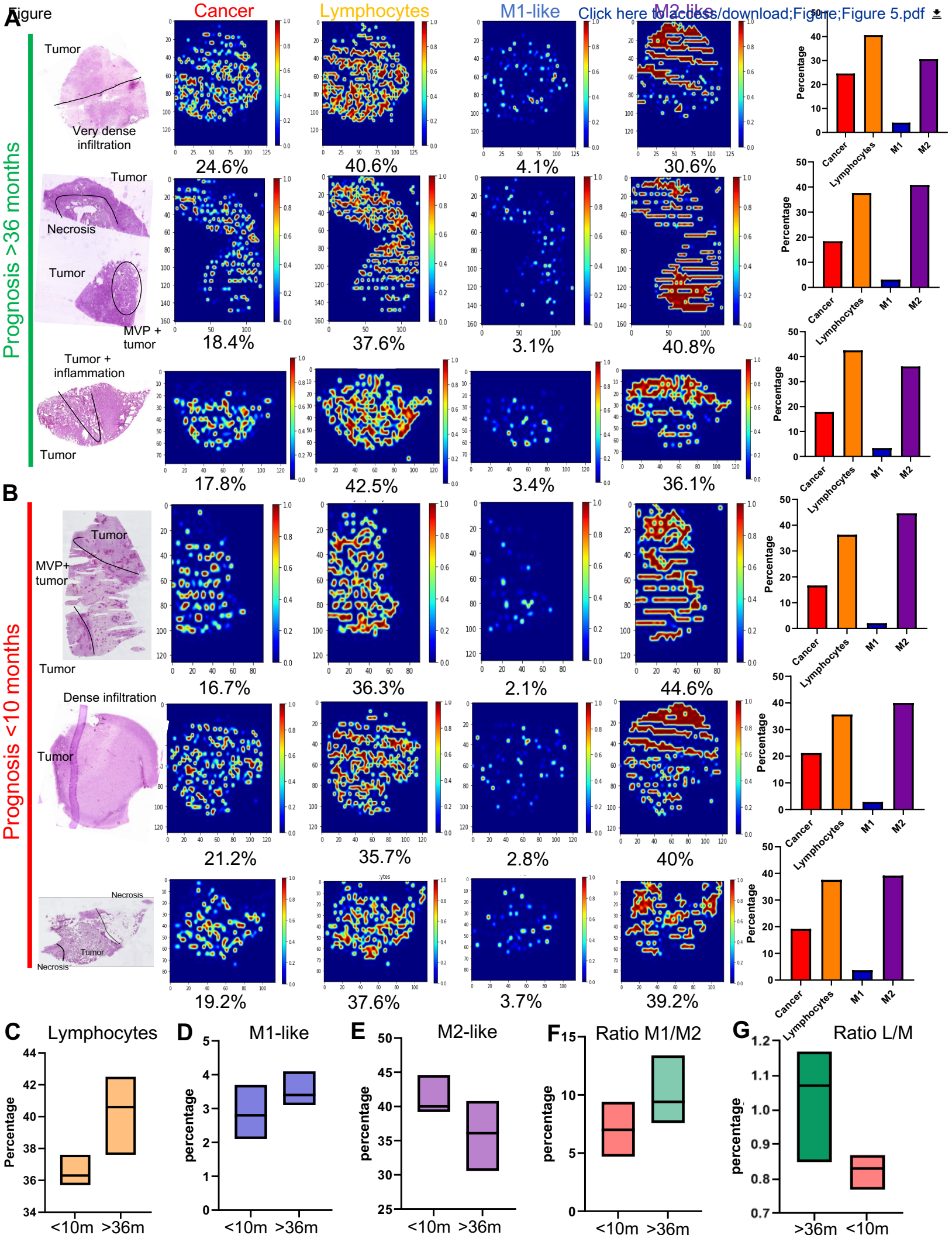
Yellow region

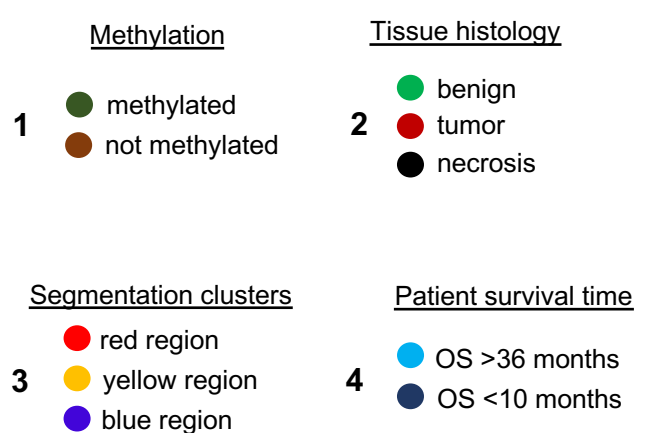
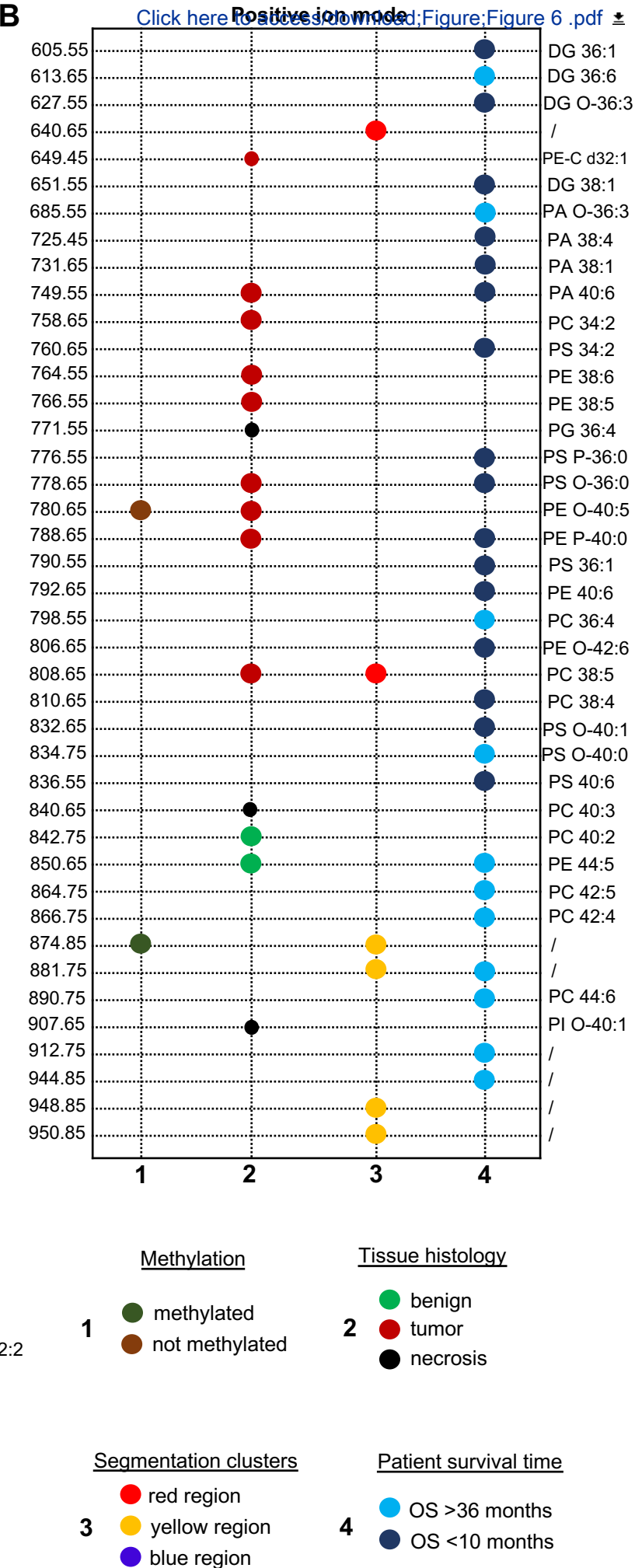
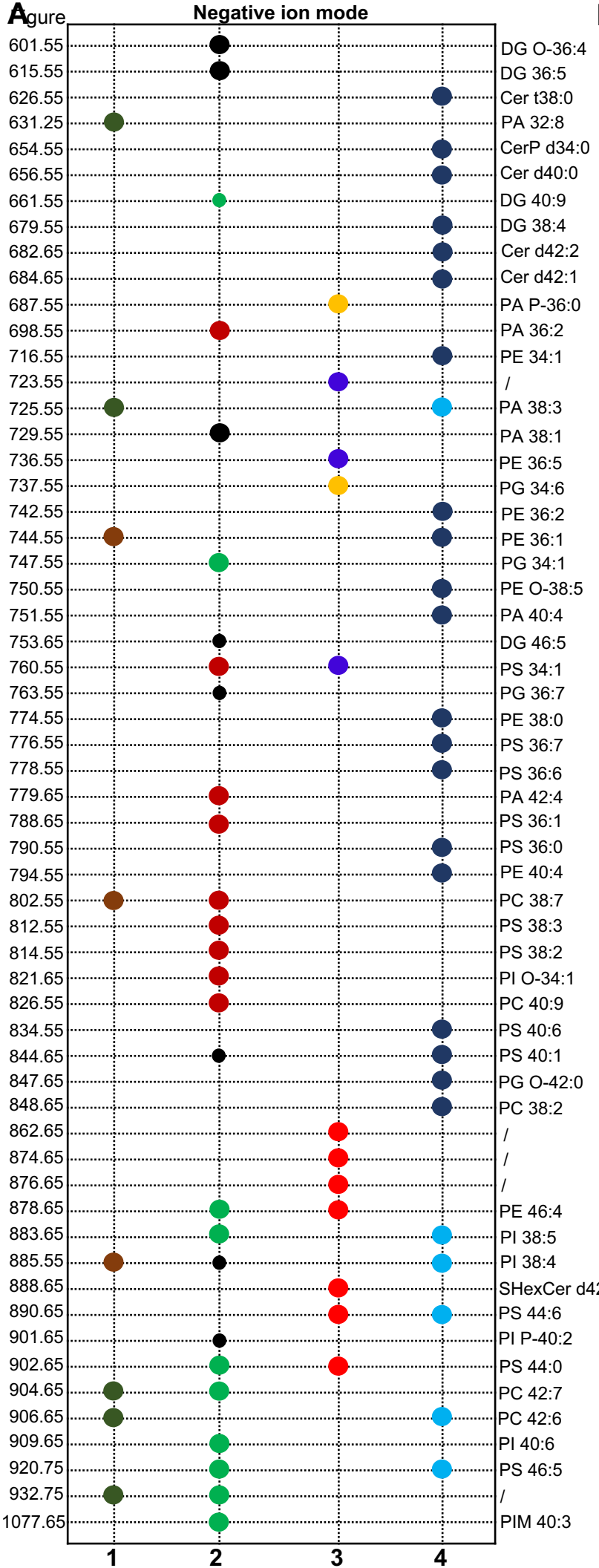




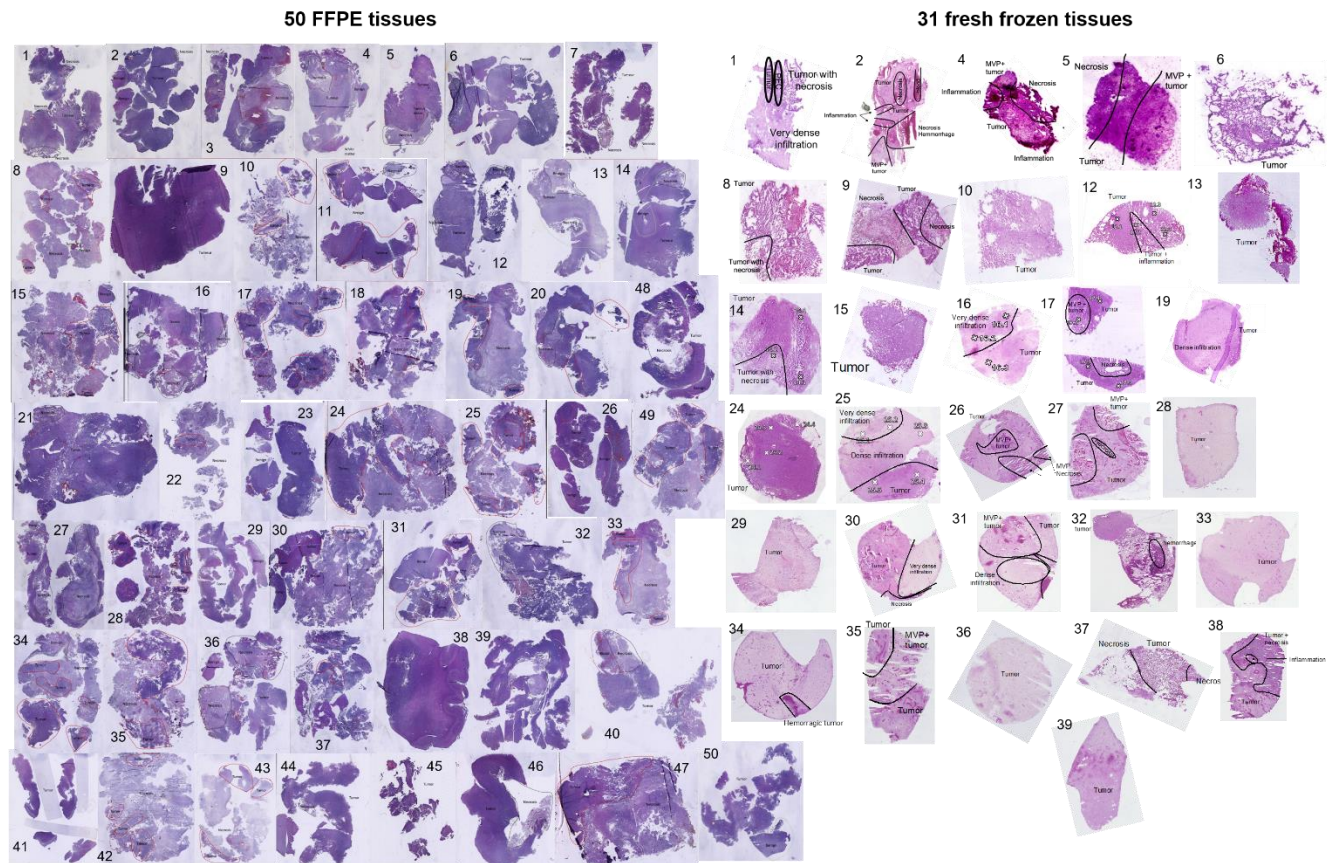












**Figure S1.** H&E scans of the fifty FFPE and the 31 fresh frozen glioblastoma tissue from the retrospective cohort and their histological annotations. Related to Figure 1.

**Table S1.** Demographic and clinical data of the retrospective and prospective cohort. Related to Figure 1.

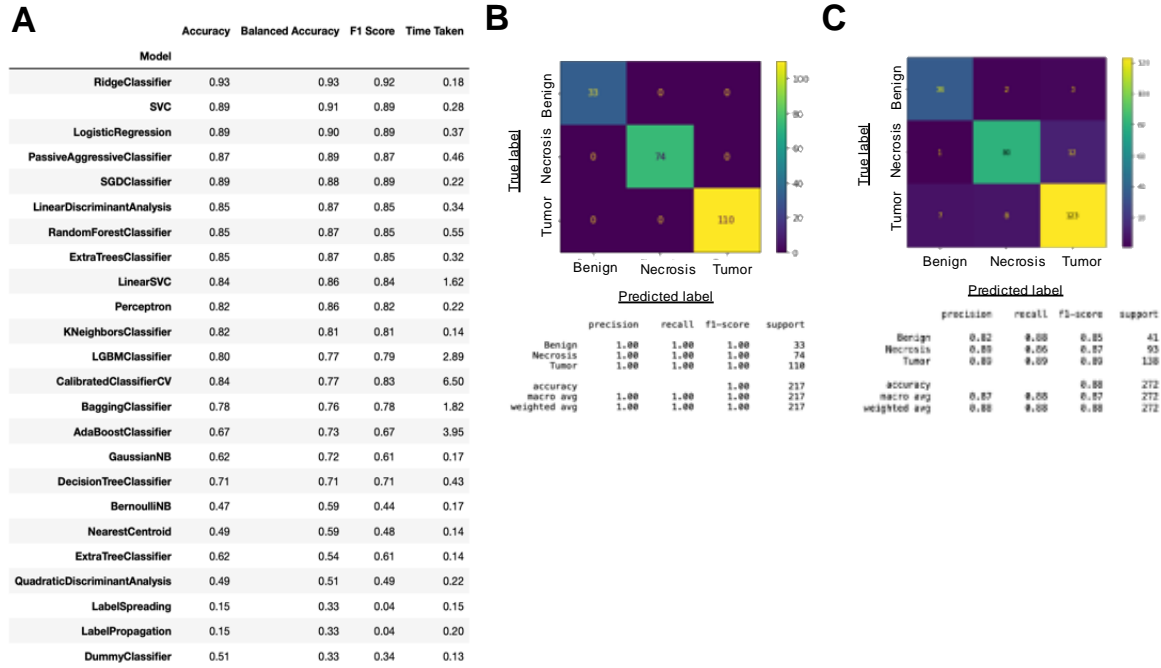
<b>Retrospective group (n = 50)</b>		
Age, median (range)		64.1 (45-82)
Sex	Females, <i>n</i>	20
	Males, <i>n</i>	30
Main location of the tumor	Frontal, <i>n</i>	15
	Occipital, <i>n</i>	3
	Parietal, <i>n</i>	10
	Temporal, <i>n</i>	22
MGMT promoter methylation status	Not methylated, <i>n</i>	36
	Methylated, <i>n</i>	14
Death	Yes, <i>n</i>	49
	No, <i>n</i>	1
Survival	OS < 15 months, <i>n</i>	24
	OS > 15 months, <i>n</i>	26
	OS < 10 months, <i>n</i>	14
	OS > 36 months, <i>n</i>	9
<b>Prospective group (n = 31)</b>		
Death	Yes, <i>n</i> (%)	30 (97)
	No, <i>n</i> (%)	1 (3)
Survival	OS < 15 months, <i>n</i>	10
	OS > 15 months, <i>n</i>	21
	OS < 10 months, <i>n</i>	10
	OS > 36 months, <i>n</i>	21

**Table S2.** Number of MS spectra for each histological region (tumor, necrosis and benign brain tissue) for the 50 FFPE and 31 fresh frozen tissues of glioblastoma from the retrospective and prospective cohort. Related to Figure 1.

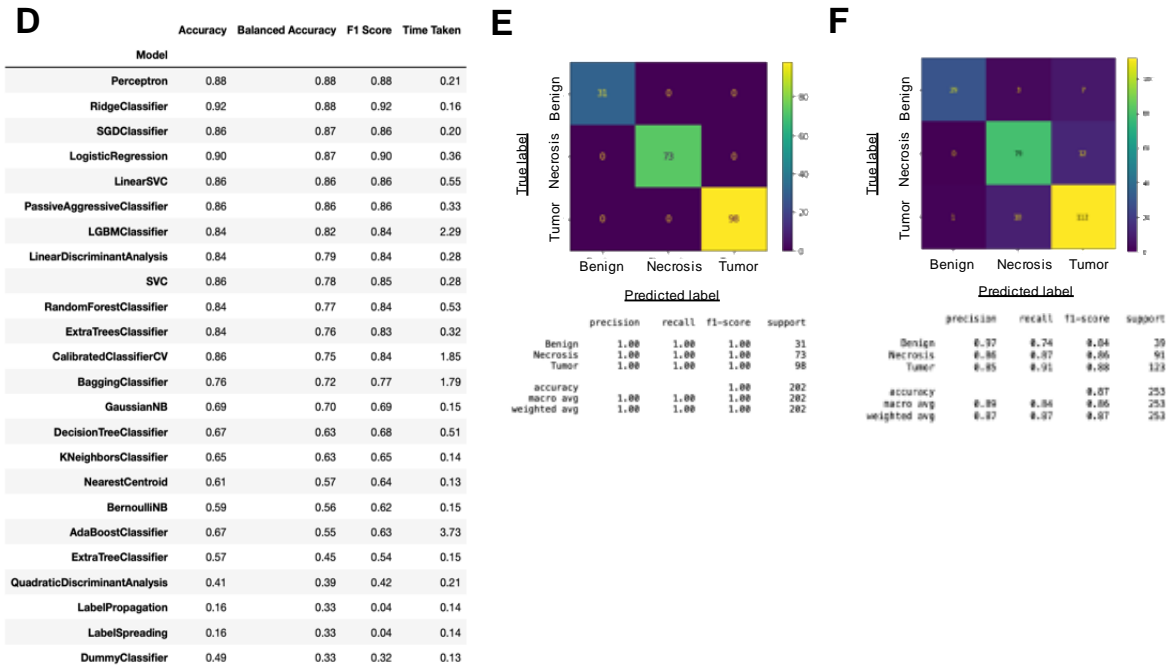
<b>Retrospective cohort (n=50)</b>			
<b>Tissue number</b>	<b>Tumor region (neg/pos)</b>	<b>Necrosis region (neg/pos)</b>	<b>Benign region (neg/pos)</b>
1	4/1	1/1	
2	3/3	1/1	2/3
3	3/3	3/3	
4	3/2	2/2	2/3
5	3/3	2/2	
6	3/2		
7	4/3	2/2	
8	1/1	3/4	
9	3/3		
10	2/2	2/2	
11	2/2	2/2	2/2
12	3/3	2/1	
13	3/3	3/3	3/3
14	3/3	1/2	2/2
15	3/3	3/3	2/2
16	4/1	2/2	
17	2/2	3/2	
18	2/2	3/2	
19	2/2	3/3	
20	3/2	2/1	1/1
21	3/3	2/2	
22	1/3	3/3	
23	4/1		3/2
24	3/3	3/2	2/1
25	3/2	3/3	
26	3/2	1/1	3/3
27	3/3	3/2	1/2
28	3/3	4/3	
29	3/2	1/1	3/3
30	3/3	3/2	½
31	1/2		3/3
32	3/3	2/3	
33	3/2	3/1	
34	3/3	2/3	2/2
35	2/2	3/2	
36	3/2	3/1	3/2
37	1/1	3/1	3/2
38	3/3	1/1	
39	3/3	2/3	
40	3/1	2/2	
41	3/4		
42	2/2	3/3	
43	2/3	3/1	
44	3/3	1/1	
45	2/2		
46	4/3	3/1	
47	3/3	3/3	
48	3/2	3/1	2/2
49	3/3	3/4	
50	2/1		3/2

<b>Prospective cohort (n=31)</b>		
<b>Tissue number</b>	<b>Tumor region (neg/pos)</b>	<b>Necrosis region (neg/pos)</b>
1		4/3
2	3/3	3/3
3	5/4	
4	3/2	
5	1/2	
6	5/5	
7	6/4	3/2
8	2/2	
9	8/8	
10	5/5	
11	4/2	2/2
12	6/4	
13	2/2	
14	6/6	
15	2/1	
16	8/6	
17	3/3	
18	2/3	
19	2/2	2/1
20	4/4	
21	6/5	
22	2/2	
23	2/2	
24	2/2	4/3
25	7/8	
26	6/4	
27	2/3	
28	6/7	
29	4/3	
30	3/3	2/2
31	4/3	

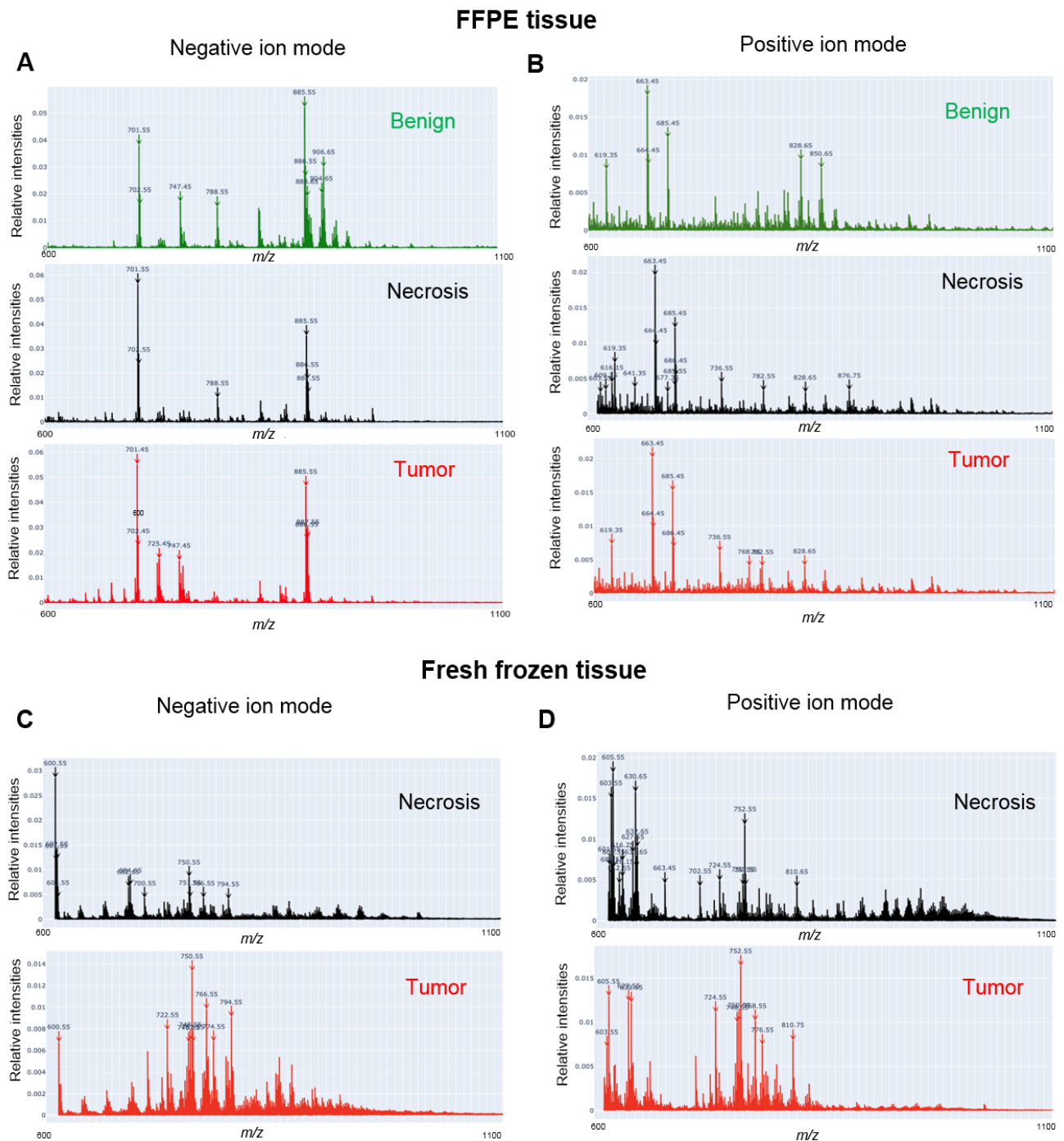
### Negative ion mode



### Positive ion mode



**Figure S2.** Identification of the optimal classification model on the retrospective FFPE glioblastoma cohort only, before data augmentation, in both ion modes. Related to Figure 1. **(A)** Table of the 24 classifiers that were trained and evaluated in negative ion mode, showing their accuracies, balanced accuracy, F1 Score and time taken. RidgeClassifier came out on top with 93% of accuracy. **(B)** Classification report and confusion matrix of RidgeClassifier with the training set in negative ion mode. **(C)** Classification report and confusion matrix of RidgeClassifier after 20-fold cross-validation in negative ion mode. **(D)** Table of the 24 classifiers that were trained and evaluated in positive ion mode, showing their accuracies, balanced accuracy, F1 score, and time taken. RidgeClassifier came out on top with 92% of accuracy. **(E)** Classification report and confusion matrix of RidgeClassifier with the training set in positive ion mode. **(F)** Classification report and confusion matrix of RidgeClassifier after 20-fold cross-validation in positive ion mode.



**Figure S3.** Averaged SpiderMass MS spectra obtained from the glioblastoma tissue in the lipid mass range ( $m/z$  600-1100). Related to Figure 1. **(A)** Averaged spectra for benign, necrosis and tumor tissue in negative ion mode for FFPE tissue. **(B)** Averaged spectra for benign, necrosis and tumor tissue in positive ion mode for FFPE tissue. **(C)** Averaged spectra for necrosis and tumor tissue in negative ion mode for fresh frozen tissue. **(D)** Averaged spectra for necrosis and tumor tissue in positive ion mode for fresh and frozen tissue.

**Table S3.** Accuracy, standard deviation and F1 scores obtained with Ridge classifier before and after data augmentation in positive ion mode. Related to Figure 1.

		Accuracy (%) and standard deviation	F1 scores		
			Benign	Necrosis	Tumor
<b>Retrospective cohort</b>	20 % out	92	86	89	96
	Cross-validation	87 +/- 0.02	84	86	88
<b>Prospective and retrospective cohort</b>	20 % out	87	67	86	91
	Cross-validation	88 +/- 0.02	76	85	90

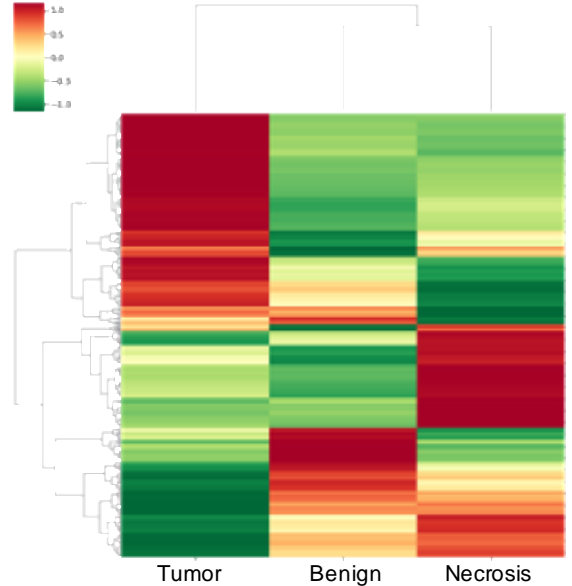


Positive ion mode

**A** Top 120 *m/z* and their contributions from LIME algorithm

y=Benign (score -141036.366) top features		y=Necrosis (score -239946.612) top features		y=Tumor (score 591791.096) top features	
Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature
+14361.356	605.650000	+20344.056	603.550000	+58964.813	810.750000
+13019.944	828.750000	+18759.806	601.550000	+40062.559	663.550000
+8417.347	603.450000	+15793.225	990.850000	+28352.044	811.750000
+7426.211	842.750000	+15703.218	886.850000	+23889.260	664.550000
+6811.577	629.350000	+15463.407	605.550000	+23246.629	812.750000
+6521.425	790.650000	+14864.456	630.650000	+18991.014	790.750000
+6426.528	814.750000	+12547.890	604.550000	+17831.813	667.350000
+5634.563	724.650000	+10756.424	620.650000	+15362.012	794.750000
+5348.013	1022.850000	+10612.214	810.650000	+14896.105	954.850000
+5238.042	725.550000	+10467.160	606.550000	+14308.301	648.650000
+5062.151	762.650000	+9836.305	634.950000	+12659.726	860.750000
+5004.316	606.850000	+8866.437	860.650000	+12004.983	813.750000
+4446.352	615.450000	+8487.462	702.850000	+11434.748	925.650000
+4410.884	619.350000	+8192.925	611.250000	+12568.879	808.750000
+4186.877	791.650000	+6124.968	1006.850000	+11184.001	627.500000
+4010.862	620.750000	+6047.467	854.850000	+10427.714	792.350000
... 2409 more positive ...		+5736.946	632.650000	+10402.533	637.250000
... 2522 more negative ...		+5519.848	617.250000	+8047.865	911.650000
-3986.286	812.650000	+7992.661	602.550000	+6021.045	748.500000
-1777.404	835.750000	+7636.558	663.450000	+7957.045	908.650000
-4214.891	609.350000	+7439.618	616.250000	+7790.649	679.250000
-4488.016	1006.850000	... 2424 more positive ...		... 2517 more positive ...	
-4886.422	872.850000	... 2522 more negative ...		... 2444 more negative ...	
-4827.973	834.750000	-7771.912	637.250000	-7908.556	617.250000
-4602.488	621.550000	-7627.811	748.550000	-7815.961	620.650000
-5025.318	650.350000	-7832.645	925.650000	-8069.445	602.550000
-5342.335	794.750000	-8033.935	726.550000	-8062.811	663.450000
-5397.389	833.750000	-8216.028	796.750000	-8274.962	806.550000
-5516.280	625.650000	-8492.224	702.750000	-8282.885	623.600000
-5822.760	753.650000	-8426.624	734.650000	-8118.568	603.450000
-6060.346	990.850000	-10019.877	794.750000	-8135.853	702.500000
-6372.387	792.750000	-10481.059	808.750000	-8428.187	630.800000
-7364.651	890.750000	-10467.353	606.650000	-8567.796	622.650000
-8662.862	632.750000	-12387.147	811.750000	-8702.888	990.850000
-8641.303	611.250000	-12865.158	627.550000	-10258.913	610.650000
-10572.408	612.250000	-12907.610	648.850000	-10334.104	634.850000
-10660.043	610.250000	-13217.617	790.750000	-10998.823	854.850000
-11760.967	627.250000	-13683.903	654.850000	-11128.363	604.500000
-14029.800	813.750000	-20439.625	664.550000	-15132.125	886.850000
-16062.897	811.750000	-20933.216	605.850000	-17968.016	605.550000
-17648.049	812.750000	-29880.909	810.750000	-18524.971	601.550000
-38793.504	810.750000	-40824.751	663.550000	-27494.320	603.550000

**B** Unsupervised clustering heatmap

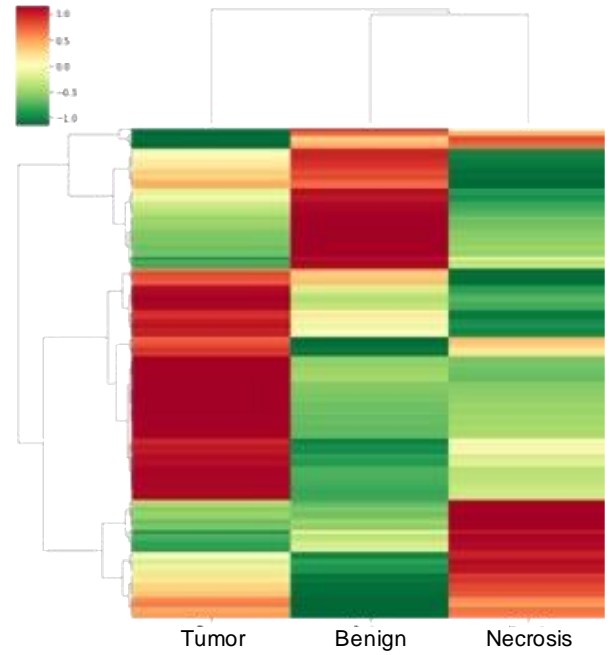


Negative ion mode

**C** Top 120 *m/z* and their contributions from LIME algorithm

y=Benign (score 10276.074) top features		y=Necrosis (score -208076.772) top features		y=Tumor (score 240099.703) top features	
Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature
+5412.239	796.550000	+3215.580	751.550000	+61711.266	794.650000
+4728.161	797.550000	+4204.671	778.550000	+22725.560	746.550000
+4371.434	834.650000	+2578.476	694.550000	+21976.324	798.550000
+3243.889	716.550000	+2247.742	737.550000	+19301.581	796.550000
+3102.309	790.550000	+2163.105	835.550000	+17449.894	796.550000
+3070.116	763.550000	+2118.629	752.550000	+14853.566	774.550000
... 2522 more positive ...		... 2522 more positive ...		+13126.948	798.550000
... 2522 more negative ...		... 2436 more negative ...		+10204.594	790.550000
-2075.019	718.550000	-2010.354	716.550000	+10232.315	825.550000
+2707.719	816.550000	-2000.869	790.550000	+10170.117	767.550000
+2475.286	767.550000	-2058.321	726.550000	+9801.321	722.550000
+2294.061	716.550000	-2066.985	739.550000	+9243.062	816.550000
+2019.951	806.550000	-2186.420	745.550000	+8400.845	844.550000
+1978.888	747.550000	-2380.864	763.550000	+8276.203	789.550000
+1896.498	717.550000	-2394.308	693.450000	+8037.886	806.650000
+1881.941	753.550000	-2362.633	862.550000	+6625.524	775.650000
+1738.089	792.550000	-2088.512	790.550000	+5877.143	842.550000
+1697.404	820.550000	-2902.019	886.550000	+5775.539	716.550000
+1683.352	864.650000	-2961.432	744.550000	+5353.500	804.650000
+1668.430	764.550000	-3036.260	812.550000	+5090.311	790.550000
+1614.210	718.550000	-3187.444	742.550000	+4452.480	744.550000
+1447.449	816.550000	-3611.357	887.550000	+4371.250	742.550000
+1371.699	700.550000	-4166.895	725.550000	+3453.261	852.550000
+1317.573	840.650000	-4018.737	792.550000	+3383.927	887.550000
+1284.273	778.550000	-3142.203	842.550000	+3033.721	812.550000
+1276.448	798.550000	-2886.835	790.550000	+2966.424	776.550000
+1248.840	938.650000	-5789.564	789.550000	+2449.094	886.550000
+1229.167	848.550000	-6227.327	854.550000	+2424.813	824.550000
... 2409 more positive ...		-7747.669	722.550000	+2374.557	726.550000
... 2522 more negative ...		-7905.779	844.550000	+2316.400	847.650000
-1375.061	806.550000	-8068.601	716.550000	... 2449 more positive ...	
-1311.534	809.650000	-8153.485	816.550000	... 2572 more negative ...	
-1487.028	744.550000	-10296.142	785.550000	-2486.416	604.650000
-1587.850	792.550000	-10926.528	774.550000	-2531.115	753.550000
-1804.603	822.550000	-11929.791	800.550000	-2688.896	737.550000
-1869.870	746.550000	-12845.452	780.550000	-2772.778	836.550000
-1885.918	778.550000	-12998.554	788.550000	-2878.329	856.550000
-2054.296	720.550000	-14657.721	768.550000	-3396.834	778.550000
-2138.656	824.550000	-15880.272	796.550000	-3604.586	797.550000
-2167.022	776.550000	-21038.690	746.550000	-3856.728	732.550000
-2721.509	796.550000	-27862.133	786.550000	-4672.702	834.550000
-3202.829	774.550000	-42308.892	794.550000	-6400.054	791.550000
-7416.603	768.550000				
-76498.184	794.550000				

**D** Unsupervised clustering heatmap

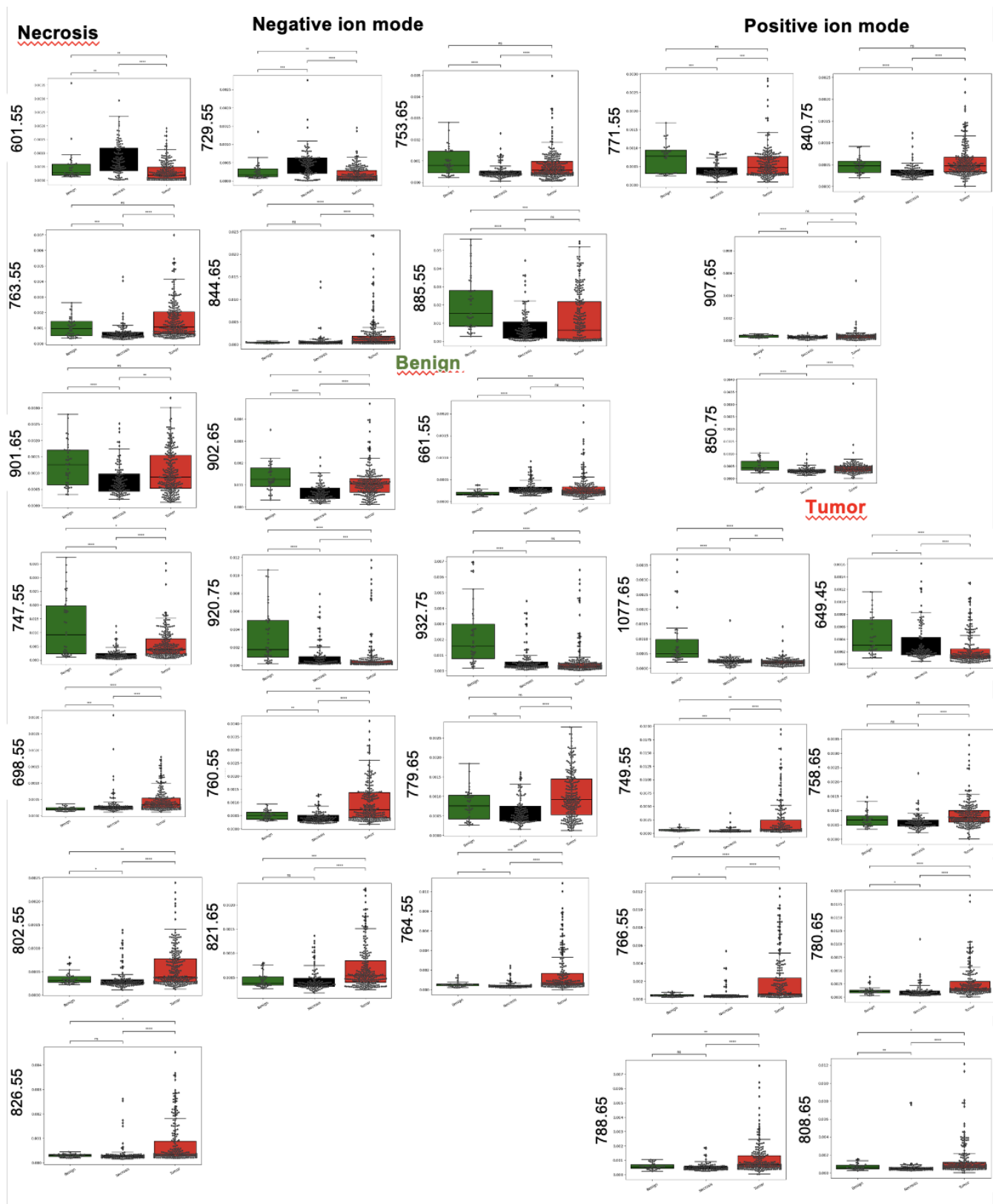


**Figure S4.** Supervised and unsupervised marker discovery result in negative and positive ion mode. Related to Figure 1. **(A)** and **(C)** Top 120 *m/z* features that contribute positively (green) or negatively (red) for the classification of each tissue in the model in positive and negative ion mode respectively. **(B)** and **(D)** Heatmap of picked peaks for three different ROIs in positive and negative ion mode. The heatmap displays the intensity of each ion, with red indicating overexpression and green indicating under-expression.

**Table S4.** Total of 41 confident biomarkers for benign, necrotic and tumor tissue in both negative and positive ion mode. Related to Figures 1 and 6. Red and green are showing the under-expressed and over-expressed ions respectively. Signs (++, +, -, -) corresponding to the contribution weight of the  $m/z$  in each type of tissue found with LIME algorithm.

$m/z$	Negative ion mode			Potential lipids
	Tumor	Necrosis	Benign	
601.55	-	+		[DG (O-14:1_22:3)-H] <sup>-</sup>
615.55	--	++		[DG (16:0_20:5)-H] <sup>-</sup>
661.55	+		-	[DG (40:9)-H] <sup>-</sup>
698.55	+	-		[PA (36:2)-H] <sup>-</sup>
729.55	-	+		[PA (38:1)-H] <sup>-</sup>
747.55	--		++	[PG (16:1_18:0)-H] <sup>-</sup>
753.65	++	--		[DG (18:1_28:4)-H] <sup>-</sup>
760.55	+		--	[PS (18:1_16:0)-H] <sup>-</sup>
763.55		--	++	[PG (22:6_14:1)-H] <sup>-</sup>
779.65	++		--	[PA (42:4)-H] <sup>-</sup>
788.65	++	-		[PS 18:0_18:1)-H] <sup>-</sup>
802.55	+	-		[PC (22:6_16:1)-H] <sup>-</sup>
812.55	++	--		[PS (18:0_20:3)-H] <sup>-</sup>
814.55	+	-		[PS (20:2_18:0)-H] <sup>-</sup>
821.65	++	--		[PI (O-16:0_18:1)-H] <sup>-</sup>
826.55	+	-		[PC (20:5_20:4)-H] <sup>-</sup>
844.65	+	-		[PS 40:1)-H] <sup>-</sup>
878.65		-	+	[PE (18:0_28:4)-H] <sup>-</sup>
883.65	+	-		[PI (18:1_20:4)-H] <sup>-</sup>
885.55	++	--		[PI (18:0_20:4)-H] <sup>-</sup>
901.65		--	++	[PI (P-18:0_22:2)-H] <sup>-</sup>
902.65		--	++	[PS (22:0_22:0)-H] <sup>-</sup>
904.65	--		++	[PC (20:1_22:6)-H] <sup>-</sup>
909.65		--	++	[PI (18:0_22:6)-H] <sup>-</sup>
920.75		--	++	[PS (46:5)+H] <sup>-</sup>
932.75	--		++	/
1077.65	-		+	[PIM (18:0_22:3)-H] <sup>-</sup>
$m/z$	Positive ion mode			Potential lipids
	Tumor	Necrosis	Benign	
649.45	-	+		[PE-Cer (d14:1_18:0)+H] <sup>+</sup>
749.55	+	-		[PA (16:5_24:1)+H] <sup>+</sup>
758.65	++	-		[PC (34:2)+H] <sup>+</sup>
764.55	+	-		[PE (14:6_24:0)+H] <sup>+</sup>
766.55	+	-		[PE (18:3_20:2)+H] <sup>+</sup>
771.55		-	+	[PG (36:4)+H] <sup>+</sup>
778.65	+	-		[PS (O-20:0_16:0)+H] <sup>+</sup>
780.65	+	-		[PE (O-20:0_20:5)+H] <sup>+</sup>
788.65	+	-		[PE (P-20:0_20:0)+H] <sup>+</sup>
808.65	++	-		[PC (18:0_20:5)+H] <sup>+</sup>

840.65		-	+	[PC (18:2_22:1)+H] <sup>+</sup>
842.65		-	+	[PC (22:1_18:1)+H] <sup>+</sup>
850.65		-	+	[PE (22:1_22:4)+H] <sup>+</sup>
907.65	+	-		[PI (O-20:1_20:0)+H] <sup>+</sup>



**Figure S5.** Corresponding box plot of the 31 remaining biomarkers specific for necrosis, benign and tumor tissue in negative and positive ion mode. Related to Figures 1 and 6.

**Table S5.** Annotation of the 48 biomarkers for poor and good prognosis in addition to the 27 biomarkers specific of cancer and immune cells (lymphocytes, M1-like and M2-like macrophages). Green indicates for which class the ion is specific. Related to Figures 3, 4 and 6.

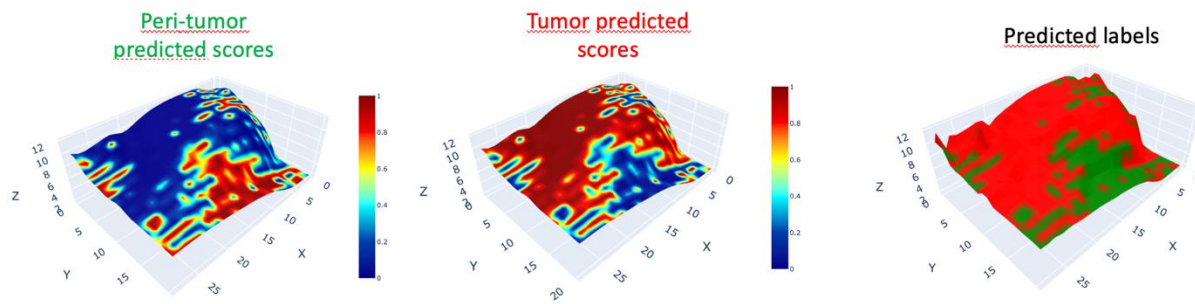
<i>m/z</i>	Negative ion mode		Potential lipids
	Bad prognosis	Good prognosis	
626.55			[Cer (t18:0_20:0)-H] <sup>-</sup>
654.55			[CerP (d18:0_16:0)+Cl] <sup>-</sup>
656.55			[Cer (d40:0)+Cl] <sup>-</sup>
679.55			[DG (18:0_20:4)+Cl] <sup>-</sup>
682.65			[Cer (d42:2)+Cl] <sup>-</sup>
684.65			[Cer (d42:1)+Cl] <sup>-</sup>
716.55			[PE (34:1)-H] <sup>-</sup>
725.55			[PA (38:3)-H] <sup>-</sup>
742.55			[PE (18:1_18:1)-H] <sup>-</sup>
744.55			[PE (36:1)-H] <sup>-</sup>
750.55			[PE (O-38:5)-H] <sup>-</sup>
751.55			[PA (22:4_18:0)-H] <sup>-</sup>
774.55			[PE (16:0_22:0)-H] <sup>-</sup>
776.55			[PS (36:7)-H] <sup>-</sup>
778.55			[PS (36:6)-H] <sup>-</sup>
790.55			[PS (36:0)-H] <sup>-</sup>
794.55			[PE (18:0_22:4)-H] <sup>-</sup>
834.55			[PS (40:6)-H] <sup>-</sup>
844.65			[PS (40:1)-H] <sup>-</sup>
847.65			[PG (O-42:0)-H] <sup>-</sup>
848.65			[PC (18:0_20:2)+Cl] <sup>-</sup>
885.65			[PI (18:0_20:4)-H] <sup>-</sup>
906.65			[PC (42:6)-H] <sup>-</sup>
920.75			[PS (46:5)-H] <sup>-</sup>
<i>m/z</i>	Positive ion mode		Potential lipids
	Bad prognosis	Good prognosis	
605.55			[DG (36:1)+H] <sup>+</sup>
613.65			[DG (36:6)+H] <sup>+</sup>
627.55			[DG O-38:6)+H] <sup>+</sup>
651.55			[DG (38:1)+H] <sup>+</sup>
685.55			[PA (O-36:3)+H] <sup>+</sup>
725.45			[PA (38:4)+H] <sup>+</sup>
731.65			[PA (22:1_16:0)+H] <sup>+</sup>
760.65			[PS (34:2)+H] <sup>+</sup>
776.55			[PS (P-18:0_18:0)+H] <sup>+</sup>
788.65			[PE (P-20:0_20:0)+H] <sup>+</sup>
790.55			[PS (36:1)+H] <sup>+</sup>
792.65			[PE (40:6)+H] <sup>+</sup>
798.55			[PC (36:4(OH))+H] <sup>+</sup>
806.65			[PE (O-20:0_22:6)+H] <sup>+</sup>
810.65			[PC (16:0_22:4)+H] <sup>+</sup>
832.65			[PS (O-40:1)+H] <sup>+</sup>

834.75					[PS (O-40:0)+H] <sup>+</sup>
836.55					[PS (40:6)+H] <sup>+</sup>
850.65					[PE (22:1_22:4)+H] <sup>+</sup>
864.75					[PC (42:5)+H] <sup>+</sup>
866.75					[PC (42:4)+H] <sup>+</sup>
890.75					[PC (44:6)+H] <sup>+</sup>
912.75					/
944.85					/
<i>m/z</i>	M1-like macrophages	M2-like macrophages	Lymphocytes	NCH82 cancer cell line	Potential lipids
655.55					/
668.55					/
684.65					[Cer (d42:1)+Cl] <sup>-</sup>
707.55					/
722.55					/
734.55					/
738.55					[PE (20:4_16:0)-H] <sup>-</sup>
746.55					/
762.55					/
766.55					[PE (38:4)-H] <sup>-</sup>
768.55					[PE (38:3)-H] <sup>-</sup>
800.65					/
808.55					/
810.55					[PS (38:4)-H] <sup>-</sup>
818.65					[GlcCer (d18:1_22:0)+Cl] <sup>-</sup>
819.55					[PG (18:1_22:6)-H] <sup>-</sup>
832.65					/
844.65					[PS (40:1)-H] <sup>-</sup>
848.65					[PC (18:0_20:2)+Cl] <sup>-</sup>
867.75					/
869.55					/
891.55					/
895.55					/
924.65					/
941.75					/
965.75					/
967.75					/

**Table S6.** Comparison of the calculated specificity and sensitivity for different technics used for intraoperative analysis in brain tumors. Related to Figure 1.

		<b>Sensitivity (= recall)</b> <b>VP / (VP + FN)</b>	<b>Specificity</b> <b>VN / (VN + FP)</b>
<b>SpiderMass (blind prediction)</b>	Negative ion mode	100%	100%
	Positive ion mode	90%	95%
<b>Average</b>		<b>95%</b>	<b>98%</b>
<b>Fluorescent guided surgery</b>	5-ALA	89%	85%
	Fluorescein	85%	87%
	Indocyanine green	95%	55%
<b>Average</b>		<b>90%</b>	<b>76%</b>
<b>Intraoperative image guidance</b>	Neuronavigation	58%	57%
	iMRI	72%	75%
	iUS	54%	87%
<b>Average</b>		<b>61%</b>	<b>73%</b>
<b>Experimental techniques</b>	DESI	86%	93%
	REIMS	96%	100%
	Raman spectroscopy	94%	96%
	Hyperspectral imaging	61%	92%
<b>Average</b>		<b>84%</b>	<b>95%</b>





**Figure S6.** Preliminary results from an ongoing study on transgenic TgC3(1) mice models spontaneously developing breast cancer and demonstrating the application of the cell scoring prediction pipeline for determining the ratio of cancer versus normal cells in the objective to improve tumor margin delineation. The final predicted labels are based on a score threshold  $> 0.5$ . Related to Figure 4.

## Résumé

Dans cette thèse, trois études importantes sont présentées. La première met en évidence une forte similarité entre les techniques d'analyse lipidique WALDI-MSI et MALDI-MSI, montrant une détection similaire des espèces moléculaires avec une corrélation significative entre les intensités et les variances des spectres obtenus. La deuxième étude représente une avancée majeure dans la compréhension du glioblastome. Elle démontre l'efficacité de l'utilisation conjointe de la spectrométrie de masse ambiante via SpiderMass et de l'intelligence artificielle pour obtenir des classifications histologiques précises et des évaluations pronostiques utiles pour cette maladie complexe. Tandis que la troisième partie suggère que la technologie SpiderMass peut être employée de manière polyvalente pour réaliser une analyse multi-omique du cancer œsogastrique, incluant lipides, bactéries et protéines. Cette approche permet la découverte de biomarqueurs dans ces différentes catégories, ce qui pourrait améliorer à la fois le diagnostic et la gestion thérapeutique approfondie des patients atteints de ce cancer. L'étude montre également que l'intégration de la spectrométrie de masse ambiante et de l'intelligence artificielle contribue à réduire les rechutes chez les patients atteints de cancer œsogastrique, représentant ainsi une avancée significative dans la lutte contre cette maladie. Finalement, cette thèse permet de montrer le potentiel de l'implantation du SpiderMass dans un centre hospitalier, suite à toutes les optimisations faites tout au long de celle-ci.

## Abstract

In this thesis, three important studies are presented. The first highlights a strong similarity between the WALDI-MSI and MALDI-MSI lipid analysis techniques, showing similar detection of molecular species with a significant correlation between the intensities and variances of the spectra obtained. The second study represents a major advance in our understanding of glioblastoma. It demonstrates the effectiveness of using SpiderMass ambient mass spectrometry in conjunction with artificial intelligence to obtain accurate histological classifications and useful prognostic assessments for this complex disease. While the third part suggests that SpiderMass technology can be versatilely employed to perform a multiomic analysis of esogastric cancer, including lipids, bacteria and proteins. This approach enables the discovery of biomarkers in these different categories, which could improve both the diagnosis and in-depth therapeutic management of patients with this cancer. The study also shows that the integration of ambient mass spectrometry and artificial intelligence helps to reduce relapse in patients with esogastric cancer, representing a significant advance in the fight against this disease. Finally, this thesis shows the potential of implementing SpiderMass in a hospital, following all the optimizations made throughout.