



**PRISM**  
Inserm U1192

**Inserm**  
La science pour la santé  
From science to health

**Université  
de Lille**

Université de Lille - Ecole Doctorale Biologie-Santé

Laboratoire PRISM INSERM U1192

## THESE DE DOCTORAT

En vue de l'obtention du grade Docteur en science de l'université de Lille

Présenté par

Lucas ROUSSEL

# Diagnostiquer le cancer de l'ovaire grâce à la technologie SpiderMass

Soutenue à Lille le 1<sup>er</sup> Décembre 2023

Devant le jury composé de :

Rapporteur	Pr. Gerard Hopfgartner	Université de Genève
Rapporteur	Pr. Laurence Charles	Université de Marseille
Président du jury	Pr David Launay	Université de Lille
Examineur	Dr. Sophie Ayciriex	Université de Lyon
Directrice de thèse	Pr. Isabelle Fournier	Université de Lille
Membre invité	Pr. Michel Salzet	Université de Lille

## Citations :

"Le chercheur est naïf car il croit qu'il peut découvrir la vérité. Mais c'est précisément cette naïveté qui le pousse à avancer dans sa quête du savoir, malgré les difficultés et les incertitudes qui se dressent sur son chemin."

- Michel Foucault

« Quel intérêt y a-t-il à la perfection ? Aucun.

- Là où il y a perfection, il n'y a plus de rien. Elle ne laisse plus de place à l'imagination et à la création. Il n'y a plus de connaissance à améliorer ou de sagesse à atteindre. La perfection est désespérante. Notre but est de concevoir et de créer des choses plus merveilleuses les unes que les autres qui dépassent celles qui existaient avant... Pas d'atteindre la perfection. Il faut améliorer en ne cherchant pas « le meilleur ». »

-Tite Kubo

## Table des matières

Citations : .....	1
Liste des figures : .....	4
Liste des tableaux : .....	5
Listes des abréviations : .....	5
Productions scientifiques .....	7
Publications .....	7
Présentations orales.....	10
Presentation de poster.....	10
Résumé / Abstract .....	11
CHAPITRE I : Introduction générale.....	12
Objectifs :.....	17
CHAPITRE II : État de l’art.....	20
Le cancer de l’ovaire.....	20
Épidémiologie du cancer de l’ovaire .....	21
Origines tissulaires des différents cancers.....	23
Diagnostic et pronostic.....	26
Prévention et traitement.....	29
Nouvelles méthodes de diagnostics peropératoire .....	31
Analyses peropératoire <i>ex vivo</i> .....	31
Analyses peropératoire <i>in vivo</i> .....	34
La spectrométrie de masse.....	35
Principe de la spectrométrie de masse .....	36
Les sources ambiantes en spectrométrie de masse.....	38
L’utilisation des sources ambiantes dans l’imagerie par spectrométrie de masse.....	49
Importance du micro-environnement tumoral dans le pronostic du cancer de l’ovaire.....	53
CHAPITRE III : Analyse des lésions de la trompe de Fallope comme précurseurs potentiels du cancer de l'ovaire précoce.....	64
Introduction.....	64
Objectifs.....	65
Matériels et méthodes .....	66
Résultats .....	75
Discussion .....	91
Conclusion .....	96

CHAPITRE IV : Diagnostic en temps réel et prédiction de l'évolution du cancer de l'ovaire par spectrométrie de masse SpiderMass .....	97
Introduction.....	97
Objectifs.....	98
Matériels et méthodes .....	99
Résultats .....	108
Discussion .....	124
Conclusion .....	127
Figures supplémentaires .....	128
CHAPITRE V : Analyse du TME du glioblastome par SpiderMass pour l'aide à la prise de décision du chirurgien. ....	141
Introduction.....	141
Objectifs.....	142
Matériel et méthodes.....	144
Résultats .....	146
Discussion .....	159
Conclusion .....	162
Figures supplémentaires .....	163
CHAPITRES VI : Conclusions & Perspectives.....	164
Bibliographie.....	167
Remerciements : .....	182
Articles scientifiques :.....	184

## Liste des figures :

<b>Figure 1:</b> Schéma représentant les différentes sources AIMS et leur fonctionnement (extrait de <sup>9</sup> ) .....	15
<b>Figure 2:</b> Cartographie du taux d'incidence des cancers de l'ovaire et leurs mortalités dans le monde.....	22
<b>Figure 3 :</b> La répartition des différents sous-types de cancer de l'ovaire.....	23
<b>Figure 4:</b> Origine tissulaire et cellulaire de chaque sous-type de cancer. (Extrait et adaptée de <sup>35</sup> via l'utilisation de biorender).....	26
<b>Figure 5:</b> Index de carcinose péritonéale de SugarBaker (extrait de <sup>40</sup> ).....	27
<b>Figure 6:</b> Coloration HES des différents sous-types de cancer (extrait de <sup>41</sup> ).....	28
<b>Figure 7:</b> Comparaison de cancer de l'ovaire au stade FIGO précoce et avancé. ....	32
<b>Figure 8:</b> Schéma global du fonctionnement de la source DESI.....	38
<b>Figure 9:</b> Représentation des différentes sources d'ionisation chimique.....	41
<b>Figure 10:</b> Schéma du dispositif LTP et son aspect mini-invasif.....	43
<b>Figure 11 :</b> Représentation schématique de l'installation réalisée pour la désolvatation des clusters en ajoutant un second laser pulsé (Nd:YAG, 1064 nm) (extrait de <sup>99</sup> ).....	44
<b>Figure 12:</b> Schéma explication du fonctionnement du MALDESI.....	45
<b>Figure 13:</b> Schéma récapitulatif des principales ionisations laser.....	47
<b>Figure 14:</b> Schéma représentant la technologie iKnife.....	49
<b>Figure 15:</b> Fonctionnement de l'imagerie MALDI (extrait et traduit de <sup>118</sup> ).....	50
<b>Figure 16:</b> Rôle de la matrice extracellulaire dans le détachement du STIC vers l'ovaire.....	55
<b>Figure 17:</b> Principales fonctions des macrophages dans l'organisme.....	58
<b>Figure 18:</b> Les différentes fonctions des macrophages au sein de la tumeur.....	60
<b>Figure 19:</b> Workflow utilisé pour la protéomique résolue dans l'espace à l'aide d'une section de tissu IHC.....	76
<b>Figure 20:</b> Comparaison des différentes lésions en fonction de leur signature protéique.....	77
<b>Figure 21:</b> Analyse protéomique des lésions de la fimbria.....	81
<b>Figure 22:</b> Analyse des voies signalisation retrouvées dans les protéines différenciantes.....	82
<b>Figure 23:</b> Visualisation des niveaux de protéines dans les différentes lésions pour les protéines connues pour être impliquées dans l'effet Warburg.....	84
<b>Figure 24:</b> Comparaison entre la signature p53 et le tissu normal.....	86
<b>Figure 25:</b> Comparaison des lésions préneoplasique et du tissu sain par IHC.....	87
<b>Figure 26:</b> différencier les lésions préneoplasiques.....	89
<b>Figure 27:</b> Lipides spécifiques identifiés pour chaque lésion.....	91
<b>Figure 28:</b> Workflow globale utilisée dans cette étude.....	107
<b>Figure 29:</b> Modèles basés sur l'analyse statistique multivariée pour décrypter les cancers de l'ovaire.....	109
<b>Figure 30:</b> Validation des modèles de classification sur une cohorte en aveugle de patientes.....	111
<b>Figure 31:</b> Modèle à plusieurs entrées (Dual-input) associant les scans HPS et les spectres MS.....	112

**Figure 32:** Cross-validation des marqueurs lipidiques des différents sous-types d'OC par MALDI-MSI et SpiderMass. .... 114

**Figure 33:** Découverte de marqueurs spécifiques des cellules immunitaires. .... 116

**Figure 34:** Création d'une image à partir du score immunitaire. .... 118

**Figure 35:** Diagnostiquer le cancer de l'ovaire en ciblant les cellules immunitaires. .... 120

**Figure 36:** Pronostiquer le cancer de l'ovaire en ciblant les cellules immunitaires. .... 122

**Figure 37:** Acteurs cellulaires, moléculaires et les potentielles cibles thérapeutiques dans le glioblastome (extrait de <sup>208</sup>) ..... 143

**Figure 38:** Pipeline de diagnostic par IA pour l'obtention d'un modèle de classification optimal et la découverte de marqueurs lipidiques associés aux différentes classes moléculaires. .... 148

**Figure 39:** Liste des m/z discriminants par l'algorithme LIME dans les deux modes d'ionisation. .... 149

**Figure 40:** Boxplot représentant l'intensité des lipides retrouvés discriminants en fonction du type de tissu. .... 151

**Figure 41:** Obtention d'un modèle de classification pour distinguer les patients ayant une durée de survie faible ou longue et découverte de biomarqueurs pour chaque survie. .... 153

**Figure 42:** Classification des tissus basé sur l'immune-score. .... 156

**Figure 43:** Modèle immune score pour la définition de la survie chez les patients GBM <10mois et >36mois.. 158

**Figure 44:** Représentation schématique du microenvironnement tumoral associé au glioblastome. .... 162

## Liste des tableaux :

**Tableau 1:** Les différents stades FIGO du cancer de l'ovaire. .... 20

**Tableau 2:** Caractéristiques majeurs des 5 sous types de cancers épithéliaux de l'ovaire (extrait et adapté de <sup>35</sup>) ..... 24

**Tableau 3:** description des caractéristiques cliniques des patientes présentant une lésion préneoplasique dans la fimbria des trompes de Fallope (cohorte prospective). .... 67

**Tableau 4:** description des caractéristiques cliniques des patientes présentant une lésion préneoplasique dans la fimbria des trompes de Fallope (cohorte rétrospective). .... 68

**Tableau 5:** Immunoréactivité et effet pronostique des protéines surexprimées dans la signature p53 par rapport au tissu normal, selon Protein Atlas. .... 85

## Listes des abréviations :

AIMS	Spectrométrie de masse par ionisation ambiante
AP	Pression ambiante (source MS)
BI	Cancer séreux borderline
CAF	Fibroblastes associés à la tumeur
CCC	Carcinome à cellules claires
CD4	Lymphocyte T CD4
CD68	Marqueurs des macrophages
CD8	Lymphocyte T CD8
CHIP	Chimiothérapie hyperthermique intrapéritonéale
CNN	Réseaux neuronaux
CT	Computed tomography
DART	Direct analysis in real time
DESI	Desorption electrospray ionisation
ESI	Electrospray ionisation
FAB	Fast Atom bombardment MS

FF	Fresh Frozen
FFPE	Tissu fixé au formol et inclus en paraffine
FTE	Epithélium des trompes de Falope
GBM	Glioblastome
HES	Coloration hemalun Eosine Safran
HGSOC	Cancer séreux de haut grade
HS	Imagerie hyperspectrale
IDH	Isocytrate déshydrogénase
IFN- $\gamma$	Interferon- $\gamma$
IHC	Immunohistochimie
IL	Interleukine
Ki67	Marqueur de prolifération cellulaire
LC	Chromatographie liquide
LDA	Linear discriminant analysis
LGBM	Modèle Light Gradient Boosting Machine
LGSOC	Cancer séreux de bas grade
LPS	Lipopolysaccharides
LTP	Low temperature plasma
M1	Macrophage de type M1
M2	Macrophage de type M2
MALDI	Matrix assisted laser desorption and ionisation
ML	Machine Learning
MS	Spectrométrie de masse
MSI	Mass spectrometry imaging
NK	Natural Killer
OC	Cancer de l'ovaire
OS	overall survival
P53	Signature p53
PA	Acide phosphatidique
PARP	Polyadénosine diphosphate ribose
PC	Phosphatidylcholine
PCA	Analyse a composante principale
PE	Phosphatidyléthanolamine
PG	Phosphatidylglycérol
PS	Phosphatidylsérine
SEE-FIM	Sectioning and Extensively Examining the Fimbriated End
SG	Survie globale
STIC	Serous tubale intraepithelial carcinoma
STIL	Serous tubale intraepithelial lesion
TME	Microenvironnement tumoral
VOCs	Espèce volatiles et semi-volatiles

## Productions scientifiques

### Publications

#### **Fallopian tube lesions as potential precursors of early ovarian cancer : A comprehensive proteomic analysis.**

**Lucas Roussel**<sup>1\*</sup>, Maxence Wisztorski<sup>1\*</sup>, Soulaïmane Aboulouard<sup>1\*</sup>, Marie Duhamel<sup>1\*</sup>, Philippe Saudemont<sup>1</sup>, Tristan Cardon<sup>1</sup>, Fabrice Narducci,<sup>1,2</sup> Yves-Marie Robin<sup>1,2</sup>, Anne-Sophie Lemaire<sup>1,2</sup>, Delphine Bertin<sup>1,2</sup>, Nawale Hajjaji<sup>1,3</sup>, Firas Kobeissy<sup>4,5</sup>, Eric Leblanc<sup>1,2‡</sup>, Isabelle Fournier<sup>1,6‡\*</sup>, Michel Salzet<sup>1,6‡\*</sup>

<sup>1</sup>Univ.Lille, Inserm, CHU Lille, U-1192 – Laboratoire Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France

<sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Cancer Center, 59020 Lille, France

<sup>3</sup>Medical Oncology Department, Oscar Lambret Cancer Center, 59020 Lille, France

<sup>4</sup>Department of Neurobiology, Center for Neurotrauma, Multiomics & Biomarkers (CNMB), Morehouse School of Medicine, GA 30310 Atlanta, GA, United States

<sup>5</sup>Department of Biochemistry and Molecular Genetics, Faculty of Medicine, American University of Beirut, Beirut, Lebanon.

<sup>6</sup>Institut Universitaire de France, 75000 Paris

\*Equivalent contribution, Co first author

‡Co-Corresponding author, Co Last author

Ovarian cancer is the leading cause of death from gynecologic cancer worldwide. High-grade serous carcinoma (HGSC) is the most common and deadliest subtype of ovarian cancer. While the origin of ovarian tumors is still debated, it has been suggested that HGSC originates from cells in the fallopian tube epithelium (FTE), specifically the epithelial cells in the region of the tubal-peritoneal junction. Three main lesions, p53 signatures, STILs, and STICs, have been defined based on the immunohistochemistry (IHC) pattern of p53 and Ki67 markers and the architectural alterations of the cells, using the Sectioning and Extensively Examining the Fimbriated End Protocol. In this study, we performed an in-depth proteomic analysis of these pre-neoplastic epithelial lesions guided by mass spectrometry imaging and IHC. We evaluated specific markers related to each preneoplastic lesion. The study identified specific lesion markers, such as CAVIN1, Emilin2, and FBLN5. We also used SpiderMass technology to perform a lipidomic analysis and identified the specific presence of specific lipids signature including dietary Fatty acids precursors in lesions. Our study provides new insights into the molecular mechanisms underlying the progression of ovarian cancer and confirms the fimbria origin of HGSC.

**Publié dans Cell death and disease le 30/09/2023, <https://doi.org/10.1038/s41419-023-06165-5>.**

**MACHINE LEARNING BASED OVARIAN CANCER DIAGNOSTIC AND PROGNOSTIC  
FOR REAL TIME IMMUNE SCORING AT THE OPERATING ROOM**

**Lucas Roussel**<sup>1†</sup>, Yanis Zirem<sup>1†</sup>, Laurine Lagache<sup>1†</sup>, Delphine Hudry<sup>1,2‡</sup>, Bertrand Meresse<sup>3</sup>, Marie Delbeke<sup>3</sup>, Eric Leblanc<sup>1,2</sup>, Gargey Yagnik<sup>4</sup>, Mark J. Lim<sup>4</sup>, Kenneth J. Rothschild<sup>4</sup>, Yves-Marie Robin<sup>1,2</sup>, Camille Pasquesoone<sup>1,2</sup>, Anne-Sophie Lemaire<sup>1,2</sup>, Delphine Bertin<sup>1,2</sup>, Fabrice Narducci<sup>1,2‡</sup>, Michel Salzet<sup>1,5†\*</sup>, Isabelle Fournier<sup>1,5†\*</sup>

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

<sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Center, F-59000 Lille, France

<sup>3</sup>Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000, Lille, France.

<sup>4</sup>AmberGen, Inc., Billerica, MA, United States

<sup>5</sup>Institut Universitaire de France, 75000 Paris

Ovarian cancers are the most lethal gynecological cancers which enumerate more than 200 000 deaths every year in the world. Even with the improvement of both treatment and patient care, the incidence of these pathologies seems to increase. The survival of this pathology is closely linked with the quality of the surgery procedure. In this context, we developed a robust diagnostic method based on the molecular signature of cancer subtype and machine learning which helps the surgeon to detect distinctly the carcinosis around the abdomen of the patient during surgery. Furthermore, the survival of the patient is closely linked with immune cell infiltration. In this study, we introduced a new method that allow direct visualization of the presence and localization of different immune cells in a tissue section. The rate of immune cells in the tissues, called immune score, was linked with both diagnosis and prognosis of the patients. This pipeline has a real potential for surgery guidance in the operative room and improving overall survival of patient.

**Soumis à Experimental & Molecular Medicine le 19/01/2024**

## Real-time Glioblastoma TME Analyses by SpiderMass for Surgery Decision Making and Patient Management

Yanis Zirem<sup>1</sup> $\phi$ , Léa Ledoux<sup>1</sup> $\phi$ , **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\*</sup> $\phi$ , Isabelle Fournier<sup>1,8</sup> $\chi$  $\phi$

<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>Department of 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

Glioblastoma is a highly heterogeneous and infiltrative form of brain cancer associated to a poor outcome with a limited efficiency of therapies. The extend of the surgery is known to be related to the patient survival. Reaching an accurate diagnostic and prognostic by the time of the initial surgery is therefore paramount in the management of glioblastoma. To this end, we have studied the performances of the SpiderMass, an ambient ionization mass spectrometry technology that can be used *in vivo* without invasiveness, coupled to a newly developed AI pipeline. We demonstrated that we can, both stratify IDH-wild type glioblastoma patients into molecular sub-groups and get an accurate diagnostic with over 90% accuracy after cross-validation. Interestingly, the developed method offers the same accuracy for prognosis. Additionally, we tested the potential of an immune-scoring strategy based on SpiderMass fingerprints, showing the association between prognosis and immune cell infiltration, to predict the patient outcome.

**En revue à Cell Report Medicine depuis le 05/12/2023**

## Présentations orales

**Diagnosing ovarian cancer with the SpiderMass technology.** Roussel L, Salzet M, Fournier I. Journée française de la spectrométrie de masse 2021, 14 juin 2021 - 24 juin 2021.

**SpiderMass: a novel tool for immune score for patient prognosis in ovarian cancer.** Roussel L, Lemaire AS, Narducci F, Leblanc E, Hudry D, Salzet M, Fournier I. Analytics 2022 Nantes, 5 septembre au 8 septembre 2022.

**Diagnosing ovarian cancer with the SpiderMass technology.** Roussel L, Lemaire AS, Narducci F, Leblanc E, Hudry D, Salzet M, Fournier I. Journée André Verbert 2022 - Colloque des doctorants (15 novembre 2022 - 15 novembre 2022).

**Diagnosing ovarian cancer with the SpiderMass technology.** Roussel L, Lemaire AS, Narducci F, Leblanc E, Hudry D, Salzet M, Fournier I. Rencontre Club jeune Spectrométrie de Masse. 07 au 11 mars 2022

**Spidermass: a novel tool for immune score for patient prognosis in ovarian cancer.** Roussel L, Lemaire AS, Narducci F, Leblanc E, Hudry D, Salzet M, Fournier I. Oncolille. 02 au 04 Novembre 2022.

## Presentation de poster

**SpiderMass a novel tool for ovarian cancer patient care.** Roussel L, Lemaire AS, Narducci F, Leblanc E, Hudry D, Salzet M, Fournier I. IMSC 2022 Maastricht. 28 au 2 août 2022.

## Résumé / Abstract

Les cancers de l'ovaire (OC) sont les cancers gynécologiques les plus mortels, entraînant plus de 200 000 décès chaque année dans le monde. Le diagnostic de ces cancers reste difficile et un diagnostic tardif entraîne un retard de prise en charge du patient réduisant alors ses chances de survie. Pour répondre à ce besoin clinique, nous avons développé un outil de diagnostic et pronostic en temps réel : le SpiderMass. Dans un premier temps, pour permettre un diagnostic précoce et une action préventive, nous nous sommes intéressés à l'origine du sous-type d'OC le plus agressif : le cancer séreux de haut grade (HGSOC). Suite à la découverte de marqueurs lipidiques et protéiques spécifiques des lésions pré-néoplasiques de la fimbria, nous avons mis en évidence les mécanismes sous-jacents liés à ces lésions et avons confirmé qu'elles étaient à l'origine du HGSOC. Dans un second temps, nous avons étudié l'ensemble des signatures moléculaires lipidiques spécifiques des différents sous-types d'OC afin de bâtir un modèle de classification via la technologie SpiderMass. Ce modèle, associant à la fois les données moléculaires et les données morphologiques des patients, a été capable de reconnaître l'ensemble des sous-types en temps réel *ex vivo*. Nous avons également développé un nouveau modèle d'imagerie par spectrométrie de masse permettant la visualisation directe des différentes cellules immunitaires au sein de tissus. Cela permet de fournir un diagnostic précis des différents types de cancer de l'ovaire et d'y associer un pronostic, étant donné que la survie du patient est étroitement liée à l'infiltration de cellules immunitaires au sein de la tumeur. Nous avons démontré que ce modèle d'imagerie est applicable à plusieurs types de cancer, incluant le cancer de l'ovaire et le glioblastome. Associé à ces modèles novateurs, le SpiderMass guide le chirurgien pendant l'opération pour réduire les marges d'exérèse et apporte un diagnostic et un pronostic fiable pour proposer le meilleur traitement au patient.

Ovarian cancer (OC) is the deadliest gynecological cancer, causing over 200,000 deaths worldwide every year. Diagnosis of OC is extremely difficult and late diagnosis leads to delays in patient management thus reducing the chances of survival. Against this backdrop, we have developed a real-time diagnostic and prognostic tool: SpiderMass. Initially, to enable early diagnosis and preventive action, we focused on the origin of the most aggressive OC subtype: high-grade serous cancer (HGSOC). Following the discovery of lipid and protein markers specific to pre-neoplastic lesions of the fimbria, we highlighted the underlying mechanisms linked to these lesions and confirmed that they were at the origin of HGSOC. Secondly, we studied all the lipid molecular signatures specific to the different OC subtypes to build a classification model using SpiderMass technology for diagnostic. This model, combining both molecular and patient morphological data, was able to recognize all subtypes in real time *ex vivo*. We have also developed a new mass spectrometry imaging model enabling direct visualization of different immune cells within tissues. This model provides an accurate diagnosis of the different types of ovarian cancer, and can associate a prognosis with them, given that patient survival is closely linked to the infiltration of immune cells within the tumor. We have demonstrated that this imaging model is applicable to several types of cancer, including ovarian cancer and glioblastoma. Combined with these innovative models, SpiderMass guides the surgeon during the operation to reduce excision margins and provides a reliable diagnosis and prognosis to propose the best treatment to the patient.

## CHAPITRE I : Introduction générale

En 2023, il est estimé que 3,8 millions de personnes vivent en France avec un cancer diagnostiqué. Dans le monde, le cancer constitue la deuxième cause de décès avec près de 10 millions de morts par an<sup>1</sup>. Le cancer de l'ovaire est le 8<sup>ème</sup> cancer le plus fréquent chez la femme, avec une incidence de 314 000 femmes diagnostiquées en 2020 dans le monde et 207 000 décès par an. En 2023, l'incidence du cancer de l'ovaire en France est estimée à 5348 cas. La forme la plus courante de cancer de l'ovaire est le cancer de l'ovaire épithélial avec 295 000 nouveaux cas par an<sup>1</sup>. Cette proportion ne cesse de croître notamment en raison de différents facteurs de risques tels que l'obésité, des facteurs génétiques (BRCA1, BRCA2...), l'âge, l'utilisation de traitements hormonaux et la nulliparité<sup>2</sup>. De plus, des études montrent l'influence du microbiome dans la synthèse anormale d'œstradiol, augmentant le risque de développer un cancer<sup>3</sup>, et le potentiel impact d'agents infectieux dans la cancérogénèse<sup>4</sup>. Le cancer de l'ovaire occasionne peu de signes cliniques au début de son évolution. En général, la patiente souffre d'inconfort ou de douleurs abdominales associées à des saignements. Ces symptômes étant communs à différentes pathologies, le diagnostic du cancer de l'ovaire reste majoritairement tardif (stade III ou IV), ce qui explique le pronostic souvent défavorable et lui vaut le surnom de « tueur silencieux ». Le cancer de l'ovaire représente donc un enjeu majeur de santé publique. Le traitement de ces tumeurs repose habituellement sur une ablation chirurgicale associée à la chimiothérapie. Cependant, et ce malgré l'évolution des technologies à disposition au bloc opératoire (e.g., examens endoscopiques, ultrasons...), la délimitation précise de l'extension loco-régionale de la tumeur (i.e., invasion de la cavité péritonéale) d'exérèse reste une véritable problématique. En effet, lors de la procédure chirurgicale, les praticiens sont confrontés à la difficulté d'identifier et de retirer tous les tissus cancéreux présents pour d'atteindre le statut de résection complète<sup>5</sup> afin de limiter le risque de rechute, tout en préservant au maximum le tissu sain et ainsi éviter de nuire à la qualité de vie des patients. À défaut d'une délimitation précise, le chirurgien est contraint de réaliser des marges chirurgicales larges autour de la masse tumorale<sup>6</sup>. Ainsi, il est souvent nécessaire de retirer l'entièreté de l'ovaire, des trompes de Fallope et parfois de l'utérus pour les cancers gynécologiques. Selon les cas, les ganglions lymphatiques et l'épiploon doivent également être retirés.

Une analyse immunohistochimique peropératoire du tissu excisé permet la vérification de la qualité de la chirurgie. Mais cette procédure ne répond pas entièrement au besoin du

chirurgien. En effet, le temps nécessaire au pathologiste pour rendre un diagnostic est relativement long, de l'ordre de 20 à 45 minutes (selon les établissements et leur organisation), ce qui limite le nombre de tissus analysés par opération. De plus, la rapidité requise pour cet examen histologique peropératoire diminue sa précision et sa sensibilité par rapport à l'examen post-chirurgicale. Enfin, cet examen repose sur une base visuelle qui est très souvent subjective.

La chirurgie étant l'une des premières lignes de traitement des cancers solides, l'enjeu majeur est de développer des outils permettant la collecte de données moléculaires *in vivo* et en temps réel pour permettre au chirurgien d'exciser totalement les zones tumorales tout en diminuant au maximum la marge chirurgicale.

La spectrométrie de masse (MS) est une technologie qui se révèle particulièrement intéressante pour obtenir le profil moléculaire de tissus biologiques. En particulier, l'imagerie par spectrométrie de masse (MSI) permet de cartographier moléculairement les tissus par une analyse sans marquage et sans *a priori*. Cette technique offre une visualisation de la distribution spatiale moléculaire du tissu analysé. En effet, en une seule acquisition, plusieurs centaines de molécules peuvent être détectées. Il est ainsi possible d'observer différents types de biomolécules tels que les lipides, les métabolites, les glucides et les protéines <sup>7 8</sup>. Cependant, bien que les sources développées pour la MSI soient parfaitement adaptées pour l'analyse de sections histologiques *ex vivo*, elles ne permettent pas de réaliser des analyses *in vivo*. Afin de pouvoir répondre à la problématique de l'analyse peropératoire *in vivo* pour la chirurgie des cancers, des sources particulières, appartenant à la famille des sources d'ionisation ambiante (Ambient Ionization Mass Spectrometry, AIMS) ont été développées <sup>9</sup>.

En 2009, la possibilité de générer des ions suivant un processus purement thermique par évaporation/ionisation rapide (Rapid Evaporative Ionization MS) a été démontré <sup>10</sup>. Ce type de procédé peut être réalisé à l'aide d'un bistouri électrique, couramment utilisé en chirurgie. Cette découverte a permis le développement de la première méthode de diagnostic utilisable au bloc opératoire, le iKnife, qui associe un bistouri électrique à la spectrométrie de masse <sup>11</sup>. Ce scalpel aide le chirurgien à détecter et à retirer les tissus cancéreux à travers une analyse en temps réel de l'aérosol généré par le bistouri électrique lors de la découpe des tissus. Cette analyse permet d'obtenir le profil lipido-métabolique du tissu <sup>12</sup> et ainsi déterminer son caractère sain ou tumoral. Ce dispositif a notamment été utilisé pour la recherche sur le cancer de l'ovaire <sup>13</sup>. Cependant, cette technologie est invasive. En effet, les tissus sont analysés pendant que le bistouri coupe les tissus. De ce fait, les tissus étant détruits, il est impossible pour le pathologiste d'effectuer un second diagnostic pour confirmer les résultats obtenus. De plus, l'iKnife manque de précision, il est impossible de réaliser des micro-analyses sur des

zones très précises. Enfin, le chirurgien collecte l'information au cours de l'excision, ce qui ne lui permet pas d'adapter son geste et d'éviter l'excision de tissus sains.

Une autre technologie introduite en 2016 permettant des analyses *in vivo*, le MassSpec Pen<sup>14</sup>, est également en développement. Elle consiste à micro-extraire les lipides et métabolites par une gouttelette d'eau mise en contact avec l'échantillon qui est ensuite aspirée et transportée dans un tube relié au spectromètre pour analyse. Cette technologie est prometteuse et a permis de mettre en évidence des différences significatives d'abondance de métabolites et lipides tels que les glycerophosphoethanolamine (PE) entre des biopsies d'ovaires sains et tumoraux<sup>15</sup>. Seulement, l'extrémité du système est au contact du tissu, ce qui pose des problèmes de contamination lors de l'analyse et ne permet pas de balayer une zone pour réaliser une marge. Enfin, le transport de liquides à la place d'aérosol rend le transfert plus long et le risque de contamination croisée lors de l'analyse de 2 zones consécutives plus grand (phénomène de carry-over). Les techniques AIMS intra-opératives les plus connues sont illustrées dans la **figure 1**.

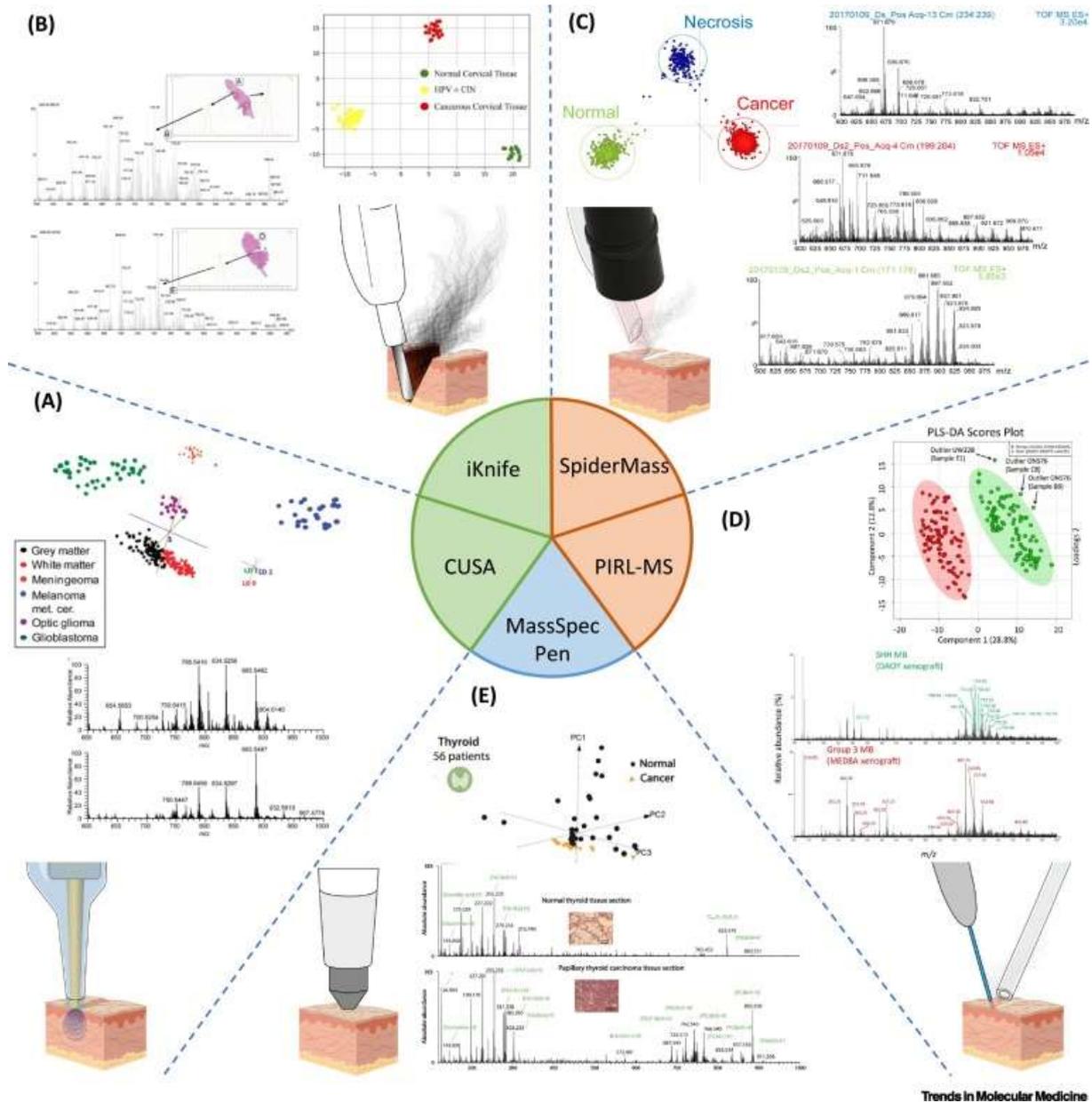


Figure 1: Schéma représentant les différentes sources AIMS et leur fonctionnement (extrait de 9)

Les différentes sondes sont rangées en fonction de leur principe de fonctionnement. En vert, sont représentées les techniques basées sur des outils chirurgicaux existants comme le CUSA (A) et l'iKnife (B). En rouge clair, sont représentées les techniques basées sur l'utilisation d'un laser IR telles que le SpiderMass (C) et le PIRL-MS (D). En bleu, est représenté le MasSpec Pen (E), basé sur l'extraction par solvant. (A) L'échantillonnage par le CUSA consiste en une ablation par ultrason qui est ensuite aspiré via un tube de transfert vers le spectromètre de masse. (B) L'iKnife utilise un bistouri électrique qui brûle les tissus. Les fumées sont ensuite aspirées et amenées au spectromètre de masse. (C) Le SpiderMass utilise un laser IR nano-seconde (5-7 ns) pour micro-échantillonner les biopsies. (D) Le PIRL-MS utilise un laser IR pico-seconde (300 ps) pour ablater la biopsie, limitant les dégâts thermiques causés sur le tissu. (E) Le MassSpec Pen utilise le principe de l'extraction liquide. Une gouttelette d'eau est placée en contact avec le tissu puis aspirée via un tube de transfert.

Depuis 2010, le laboratoire PRISM travaille sur le développement d'une technologie d'analyse *in vivo* basée sur une source laser mini-invasive : le SpiderMass. Cette technologie repose sur le principe du MALDI (Matrix-Assisted Laser Ablation Ionisation). Elle utilise l'eau comme matrice endogène permettant la désorption/ionisation des molécules endogènes au tissu. Cet instrument fonctionne sous l'action d'un laser IR émettant à 2,94  $\mu\text{m}$  envoyé directement sur un tissu via une fibre optique. L'irradiation laser conduit l'excitation résonnante des molécules d'eau endogène des tissus qui engendre la désorption et l'ionisation des biomolécules environnantes (principalement les lipides et les métabolites) et générer un effet WALDI (water-assisted laser desorption/ionization) <sup>16</sup>. Les molécules désorbées/ionisées sont ensuite transmises via un tube de transfert de plusieurs mètres de long pour être analysées par spectrométrie de masse en temps-réel. En comparant le profil moléculaire lipidique obtenu avec un modèle de classification pré-établi à partir d'une banque de données de spectres SpiderMass comprenant une cohorte suffisante de patients (idéalement au moins 20 patients par classe du modèle), la zone irradiée peut être identifier comment étant cancéreuse ou non, le type, le sous-type et le grade peuvent être également préciser, et tout cela en temps réel afin d'aider le chirurgien dans son geste pendant l'opération <sup>17</sup>. Il est également possible de chercher des sites métastatiques et de mieux préciser les marges d'exérèse <sup>17</sup>. Le SpiderMass a donc un intérêt clinique majeur pour l'aide au diagnostic. Les faibles dommages qu'il engendre sur le tissu permettent de minimiser la marge chirurgicale sans abîmer le tissu sain.

La fonctionnalité du système a pu être démontrée dans un contexte clinique réel par une étude réalisée au bloc vétérinaire sur des patients chiens atteints de sarcome <sup>17</sup>, ainsi que par une étude clinique (NCT04472546) réalisée en collaboration avec la société Pierre Fabre Dermo-Cosmétique sur plus de 200 patients atteints de pathologies du derme. De plus, de récentes recherches menées au laboratoire PRISM ont démontré que les profils moléculaires obtenus par SpiderMass étaient suffisamment spécifiques pour permettre de discriminer différentes lignées cellulaires ainsi qu'une même lignée sur laquelle un Knock-Down avait été réalisé ou non <sup>18</sup>. Enfin, cette technologie peut être utilisée non seulement dans la détection de lipides et métabolites, mais aussi dans la détection de peptides et de protéines <sup>19</sup>.

Actuellement, le laboratoire s'intéresse au développement de l'imagerie MS via la technologie SpiderMass. Toujours dans l'objectif d'une application future chez les patients au bloc, l'imagerie SpiderMass repose sur une nouvelle méthode d'imagerie topographique via l'utilisation d'un bras robotisé capable de se déplacer selon 6 axes et permettant de déplacer la sonde laser de façon contrôlée au-dessus du patient. Cette technologie permet le scan intégral de la surface d'un tissu ce qui donne accès à un diagnostic sensible, précis (de l'ordre de 250  $\mu\text{m}$ ) et complémentaire aux techniques d'imagerie conventionnelles utilisées au bloc opératoire <sup>20</sup>.

### Objectifs :

C'est dans ce contexte de développement et de validation de la technologie SpiderMass pour la chirurgie des cancers que se situe mon travail de doctorat.

Plus particulièrement, l'objectif de ma thèse est de démontrer l'apport de la technologie SpiderMass dans le cadre du cancer de l'ovaire. Pour cela, nous nous sommes consacrés à l'étude de l'origine du développement du cancer de l'ovaire de type séreux de haut grade. Par la suite, nous nous sommes focalisés sur la définition d'un diagnostic précis des différents sous-types de cancers de l'ovaire. En parallèle, nous nous sommes intéressés au microenvironnement tumoral associé à ces cancers. Cela a permis d'établir un pronostic basé sur le système immunitaire du patient que nous avons appelé immune-score. Cet immunoscore Spidermass a par la suite été validé sur un autre type de cancer, le glioblastome. Il pourra dans le futur être déployé en temps-réel. Afin de pouvoir développer le SpiderMass pour la définition du diagnostic et du pronostic, la thèse a été articulée autour de 3 objectifs spécifiques :

#### *Etude des lésions pré-néoplasiques de la fimbria pour définir l'origine du cancer séreux de haut grade de l'ovaire.*

Certaines études suggèrent que le carcinome séreux de haut grade (HGSOC) trouve son origine dans les cellules de l'épithélium des trompes de Fallope, la fimbria. Cette zone se situe plus précisément au niveau des cellules épithéliales de la région de la jonction tubopéritonéale. Trois lésions principales ont été définies en se basant sur le profil immunohistochimique des marqueurs p53 et Ki67 et sur les altérations architecturales des cellules. Ces trois lésions comprennent les signatures p53 (p53), les lésions intraépithéliales tubaires séreuses (STIL) et les cancers intraépithéliaux séreux tubaires (STIC). Ces lésions ont été identifiées en utilisant le protocole de sectionnement et d'examen approfondi de l'extrémité des franges appelé (SEE-FIM<sup>21</sup>). Nous avons réalisé une analyse protéomique approfondie de ces lésions en utilisant l'imagerie par spectrométrie de masse, la protéomique shot gun et l'immunohistochimie. Cette étude a été complétée par une analyse lipidomique réalisée au SpiderMass. Il m'a été nécessaire, dans un premier temps, d'optimiser le protocole expérimental afin d'analyser au mieux ces lésions composées d'un faible nombre de cellules et sur tissus FFPE. Puis, un modèle PC-LDA a été créé permettant la distinction de ces différentes lésions. J'ai ensuite relié les lipides discriminants ces lésions aux enzymes impliquées dans le métabolisme lipidique au sein de la tumeur. Ainsi, nous avons mis en évidence le caractère tumoral de ces lésions et décrit la fimbria en tant qu'origine du cancer

séreux de haut grade. Cette étude a démontré l'importance de la détection précoce de ces lésions et l'intérêt de la résection de la fimbria dans la prévention du cancer de l'ovaire.

*Diagnostic du cancer de l'ovaire et détection des différents types cellulaires au sein du microenvironnement tumoral.*

Une fois l'origine du cancer séreux de haut grade caractérisé, je me suis focalisé sur le diagnostic des différents sous-types de cancer de l'ovaire à travers l'analyse d'une cohorte de tissu de patientes au SpiderMass. Pour cela, une cohorte prospective de 79 patientes a été annotée par un pathologiste suivi d'une cohorte de validation de 27 patientes. Suite à ces annotations, chaque échantillon a été analysé par SpiderMass au niveau des zones tumorales. Chacun des spectres obtenus a été intégré à un modèle de classification PC-LDA dans le but de discriminer les sous-types cancéreux. Ces résultats, très prometteurs, ont ensuite été améliorés via 2 méthodes. La première intègre un tout nouveau modèle de classification, le RIDGEClassifier, parfaitement adapté à notre faible quantité d'échantillon et a permis une validation de 97%. De plus, ce modèle ajoute de la clarté dans le rendu du résultat et permet la définition de biomarqueurs spécifiques. La seconde méthode consiste en la création d'un modèle à deux entrées capable d'intégrer à la fois les données moléculaires et les scans des tissus analysés. Ce modèle a permis un diagnostic très précis des tissus par une méthode combinée utilisant des réseaux de neurones et du Deep Learning.

Par ailleurs, les pics permettant la séparation des classes dans ce modèle ont été identifiés par fragmentation (MS/MS). Les lipides ainsi identifiés pourraient être de nouveaux biomarqueurs du cancer de l'ovaire. Les résultats ont été cross-validés par de l'imagerie par spectrométrie de masse MALDI, technique plus conventionnelle. Ce modèle a ensuite pu être testé à l'aveugle sur une cohorte de validation composée de l'ensemble des classes présentes dans ce modèle. La grande majorité des tissus analysés en aveugle a pu être correctement classée, montrant la robustesse de ce modèle. Ainsi, celui-ci peut être utilisé pour une aide au diagnostic en temps réel.

En parallèle, une analyse par SpiderMass de différents types de lignées cellulaires incluant des lignées de cancer, de cellules saines, de macrophages et de lymphocytes a été réalisée. Ces différents types cellulaires sont présents dans le microenvironnement tumoral et l'analyse de ces cellules permettra de préciser la provenance de chaque ion lors de l'acquisition d'un spectre sur des coupes de tissu. Pour cela, chaque lignée a été mise en culture séparément à différents passages puis analysée par SpiderMass en mode négatif. Pour mimer un environnement permettant la différenciation phénotypique des macrophages vers un profil soit pro-inflammatoire (M1) ou pro-tumoral (M2), des stimulations respectivement aux Lipopolysaccharides (LPS) et à l'interleukine 6 (IL6) ont été effectuées. Les profils moléculaires

spécifiques de chaque cellule et phénotype ont donc été enregistrés. Un modèle de classification a ainsi été réalisé permettant de dissocier toutes les sous-populations cellulaires.

*Isolation des cellules immunitaires et détermination de nouveaux marqueurs de survie.*

L'objectif de cette partie est de mettre en place un diagnostic et d'évaluer la survie des patients en fonction de l'infiltration immunitaire sur la base des données SpiderMass. Pour cela, les données obtenues ont été extraites lors de l'analyse de lignées pour définir les pics discriminants de chacune d'entre elles. Pour définir précisément la présence et l'abondance de cellules immunitaires, il a été nécessaire de trouver des signaux exclusifs de ces sous-populations. Ces pics ont donc été déterminés et les lipides associés ont été identifiés par MS<sup>2</sup> par analyse SpiderMass. Après avoir vérifié l'absence de ces lipides au sein des lignées cancéreuses, ces masses ont été recherchées dans les analyses SpiderMass de la cohorte de cancers de l'ovaire et de l'endomètre. Les intensités relatives de ces signaux ont ensuite été comparées selon le sous-type de cancer analysé par analyse statistique univariée Kruskal-Wallis. La confirmation de l'infiltration de cellules inflammatoires au sein des tissus analysés a été réalisée à la fois par immunofluorescence avec marquage à l'anticorps anti-CD68, marqueur spécifique des macrophages, mais également par imagerie MALDI-IHC multiplexe contre les principaux marqueurs de chaque sous-population. Le bilan de cette analyse statistique a été mis en corrélation, dans un premier temps, avec l'agressivité de chaque sous-type de cancer, puis avec les données de survie des patientes analysées. L'identification de l'environnement immunosuppresseur ou pro-inflammatoire a donc permis de mettre en évidence leur impact sur la survie du patient.

En partant de cette analyse, un modèle de classification basé sur le score immunitaire (immune-score) a été mis en place. Pour chaque analyse de tissu, un score de ressemblance avec un spectre lymphocytaire, macrophagique, cancéreux et sain est appliqué à l'échantillon par analyse informatique de type Convolutional neural network (CNN). Grâce à ce score, il est ainsi possible de classer les patients par leur score immunitaire et de corréler ce score à la survie du patient. Ce modèle a été utilisé sur le cancer de l'ovaire dans un premier temps puis a été validé sur un autre type de tumeur, le glioblastome. Nous avons ainsi corrélé l'infiltration lymphocytaire et de macrophages M1 au bon pronostic du patient dans ces deux sujets d'études tandis que l'infiltration de M2 est fortement néfaste pour la survie du patient. Nous avons ainsi prouvé l'intérêt du SpiderMass associé à des méthodes de Deep Learning dans l'évaluation précise de l'infiltration immunitaire. Ce diagnostic rapide permet de dissocier les différents sous-types de cancer et de classer les patients selon leur survie. Ainsi, la combinaison des modèles de classification et de l'immune-score pourrait permettre un diagnostic précis du patient et donc de guider le clinicien dans sa prise de décision dans la prise en charge du patient.

## CHAPITRE II : État de l'art

### Le cancer de l'ovaire.

Les cancers de l'ovaire (OC) sont les cancers gynécologiques les plus mortels. En effet chaque année, environ 313 959 femmes sont diagnostiquées et 207 000 décès sont répertoriés dans le monde<sup>22</sup>. Ces cancers regroupent un grand panel de néoplasmes très hétérogènes. En effet, ils peuvent être issus des cellules germinales de l'ovaire (3%), du stroma des cellules germinales (2%) mais sont beaucoup plus régulièrement issus de l'épithélium ovarien (>90%). Parmi ces cancers épithéliaux, plusieurs sous-types sont représentés. Le plus fréquemment diagnostiqué, mais également le plus agressif, est le cancer séreux de haut grade (>75% des cas), suivi du cancer endométrioïde et du cancer à cellules claires représentant pour chacun d'entre eux 10% des cas. À ces sous-types s'ajoute le cancer séreux de bas grade (5%) et de manière beaucoup moins fréquente le cancer mucineux (3%)<sup>23</sup>.

Malgré les progrès dans la prise en charge des cancers, les OCs restent difficiles à diagnostiquer avant des stades FIGOs très avancés (III/IV) (**Tableau 1**).

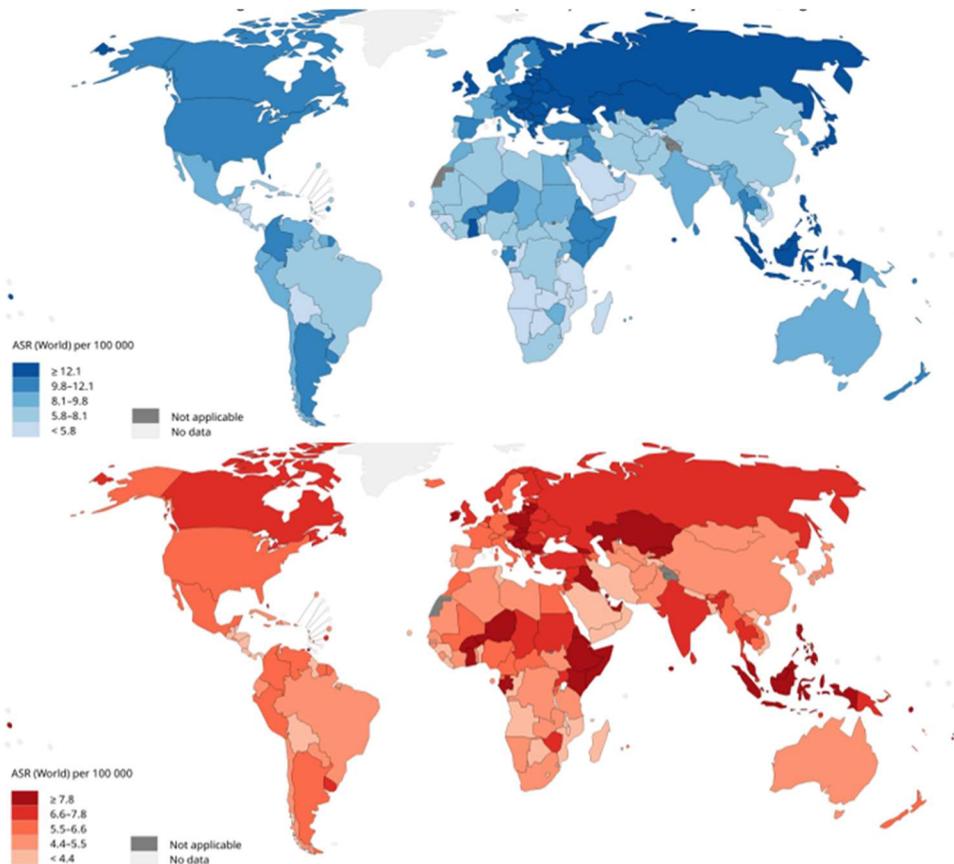
**Tableau 1:** Les différents stades FIGO du cancer de l'ovaire.

Stade	I	II	III	IV
	Tumeur limitée à un seul ovaire	Tumeur sur 1 ou 2 ovaires ou trompes. Extension pelvienne ou cancer péritonéal primitif.	Tumeur limitée aux ovaires (un ou les 2) ou à 1 ou 2 trompes de Fallope ou cancer péritonéal primitif avec diffusion cytologiquement ou histologiquement confirmée au péritoine en dehors du pelvis et/ou adénopathies métastatiques rétropéritonéales	Métastases à distance (autres que les métastases péritonéales)
<b>A</b>	Tumeur limitée à un seul ovaire. Capsule intacte.	Extension et/ou greffe utérine et/ou tubaire et/ou ovarienne	Envahissement péritonéal extrapelvien (au-dessus du pelvis) microscopique avec ou sans adénopathies rétropéritonéales, incluant l'envahissement des intestins	Épanchement pleural avec cytologie positive
<b>B</b>	Tumeur des deux ovaires, capsule intacte.	Extension à d'autres organes pelviens, incluant le côlon sigmoïde et le rectum	Métastases péritonéales macroscopiques au-delà du pelvis ≤ 2 cm dans leur plus grande dimension, incluant l'envahissement des intestins en dehors du pelvis avec ou sans adénopathies rétropéritonéales	Métastases parenchymateuses (foie, rate), infiltration de la muqueuse intestinale et métastases aux organes extra-abdominaux (incluant les adénopathies inguinales et les adénopathies en dehors de la cavité abdominale)
<b>C</b>	Tumeur sur 1 ou 2 ovaires. Rupture de la capsule ou cellule malignes dans les ascites.		Métastases péritonéales au-delà du pelvis > 2 cm dans leur plus grande dimension et/ou adénopathies métastatiques rétropéritonéales (incluant l'extension de la tumeur à la capsule du foie et de la rate sans envahissement parenchymateux de ces organes)	

Pourtant, un diagnostic précoce améliore grandement le pronostic des patientes. Il a été démontré que plus de 90% des patientes prise en charge à des stades précoces ont une survie favorable. Malheureusement, seulement 25% de ces cancers sont diagnostiqués au stade I<sup>24</sup>. Le manque de marqueurs de diagnostic précoces et la difficulté à distinguer les symptômes d'autres pathologies plus bénignes sont les principaux responsables de ce diagnostic tardif<sup>25</sup>. C'est ainsi qu'on leur a octroyé le surnom de « Silent Killers » ou « Tueur silencieux ».

### Épidémiologie du cancer de l'ovaire

Les cancers de l'ovaire ne sont pas les plus fréquents puisqu'ils sont classés au 8<sup>ème</sup> rang en termes d'incidence, représentant 6 900 femmes diagnostiquées en France en 2020<sup>22</sup>. Seulement, ils comptent parmi les cancers les plus mortels chez la femme avec une mortalité l'élevant cette fois au 5<sup>ème</sup> rang<sup>26 27</sup>. La grande majorité de ces cancers est malheureusement diagnostiquée à des stades FIGOs avancés, expliquant un faible taux de survie à 5 ans, de l'ordre de 46%<sup>28</sup>. La **figure 2** représente la mortalité et le taux d'incidence de ces cancers dans le monde. Ces cartes montrent que les pays développés sont les plus impactés par cette pathologie, mais que, malheureusement, sa mortalité est bien plus élevée dans les pays du tiers-monde. Cette différence est explicable principalement par la différence de prise en charge des patientes entre les pays.



**Figure 2:** Cartographie du taux d'incidence des cancers de l'ovaire et leurs mortalités dans le monde.

La première carte en bleu montre l'incidence de la pathologie dans le monde tandis que la seconde indique sa mortalité. (Figure extraite de l'agence internationale de la recherche sur le cancer GLOBOCAN 2020).

Par ailleurs, il a été estimé que d'ici 2040 l'incidence de cette pathologie va croître de manière significative de plus de 20% en France <sup>1</sup>. Cette pathologie est donc un sujet de préoccupation majeur pour les futures recherches cliniques.

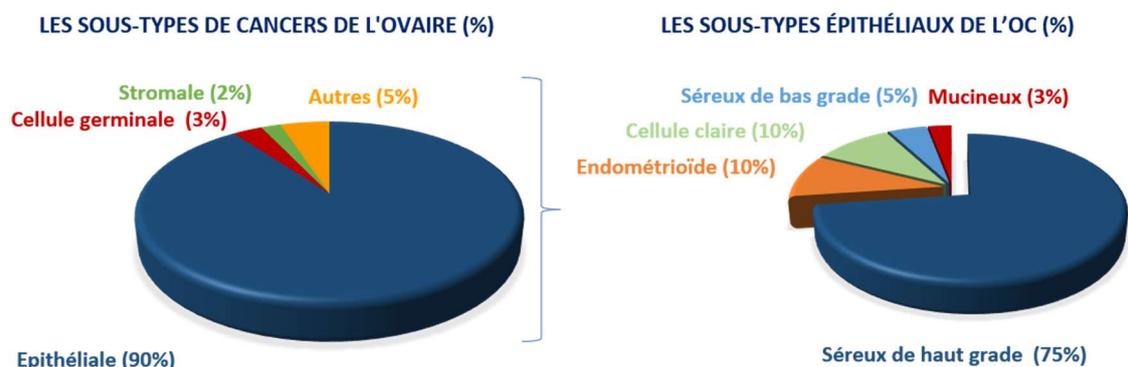
Les facteurs de risques les plus reconnus associés à cette pathologie sont « l'ovulation incessante », une ménarche précoce, une ménopause tardive, l'absence de grossesse et l'absence de contraception œstroprogestative. Bien évidemment, le tabagisme et l'obésité sont également des facteurs de risques reconnus pour certains sous-types histologiques <sup>29</sup>. Des analyses détaillées microscopiques et génétiques ont étayé la compréhension biologique des cancers ovariens. Ceux-ci regroupent en réalité des maladies très différentes et très hétérogènes. Les affections gynécologiques bénignes (y compris l'endométriose, le syndrome des ovaires polykystiques et la maladie inflammatoire pelvienne) peuvent également influencer sur l'apparition d'OC. Il est aussi reconnu que 15% des patientes atteintes d'OC sont associées à des prédispositions génétiques. BRCA1 et BRCA2 sont les deux gènes les plus connus pour être impliqués dans le développement d'OC. Ces deux gènes représentent 65 à 75% des OC héréditaires. En effet, les mutations délétères au sein des gènes BRCA1 et BRCA2, ainsi que d'autres gènes de réparation de cassure double brin de l'ADN, sont largement associées à la

prédisposition au sous-type séreux de haut grade. Le syndrome de Lynch, un syndrome familial de cancer héréditaire autosomique dominant, représente 10 à 15% des OC héréditaires<sup>30,31</sup>, et est généralement associé à des tumeurs endométrioïdes ou à cellules claires<sup>32</sup>. D'autres syndromes génétiques à risque comprennent le syndrome de Peutz-Jeghers et des troubles rares tels que le syndrome de Gorlin.

Décrit par le passé comme étant asymptomatiques, d'où leur surnom de « silent killer », les symptômes des OC sont multiples. Ces derniers sont principalement liés aux ballonnements, à l'augmentation de la taille abdominale, de la difficulté à manger, à une sensation de satiété rapide et à des douleurs abdominales ou pelviennes<sup>25</sup>. Malheureusement, ces symptômes sont communs à un grand nombre d'autres maladies, réduisant les chances de diagnostic précoce. Ces symptômes sont de plus en plus prononcés au fur et à mesure du développement de la tumeur, et c'est souvent à des stades tardifs (III/IV) qu'ils seront associés aux OC. Il est important de noter que le HGSOE (cancer ovarien séreux de haut grade) présente ces symptômes à des stades plus tardifs par rapport aux autres types histologiques comme l'endométriode ou le cancer à cellules claires. Ces deux sous-types développent des symptômes de manière très précoce tels que la distension abdominale, des douleurs liées à la pression du pelvis ou des problèmes liés au système urinaire et intestinal dû à la masse tumorale<sup>33</sup>.

### Origines tissulaires des différents cancers.

Comme expliqué précédemment, les cancers de l'ovaire ont des origines histologiques multiples. Cette diversité va agir directement sur le diagnostic du patient ainsi que sur son pronostic. C'est pourquoi il est nécessaire de comprendre les mécanismes mis en jeu lors de l'apparition de la tumeur. La **figure 3** montre la disposition de ces sous-types épithéliaux pour l'OC.



**Figure 3 :** La répartition des différents sous-types de cancer de l'ovaire.

La répartition des différents types de cancer (en %) ainsi que la répartition des sous-types épithéliaux de l'OC (Adapté et traduit de<sup>34</sup>).

Le sous-type d'OC épithéliale le plus étudié et le plus représenté est le carcinome séreux de haut grade (75% des cas). Il est le plus agressif des cancers et l'un des plus mortels. Il est ensuite suivi des cancers endométrioïdes et à cellule claires (10% chacun) puis du cancer séreux de bas grade (5%) et du cancer mucineux (3%). Outre leurs différences morphologiques distinctes, les différents sous-types présentent des variations dans leurs caractéristiques cliniques ainsi que dans leurs facteurs de risque associés. Le **tableau 2** ci-dessous récapitule les facteurs de risques et l'origine des différents cancers en fonction des types histologiques les plus courants.

**Tableau 2:** Caractéristiques majeurs des 5 sous types de cancers épithéliaux de l'ovaire (extrait et adapté de <sup>35</sup>)

Type de cancer	Origine tissulaire	Possible origine cellulaire	Lésion précurseuse	Risques familiaux	Stratégie de prévention
<b>HGSOC</b>	Trompe de Fallope : Fimbria	Cellule épithéliale sécrétoire des trompes de Fallope	STIC	BRCA1, BRCA2, BRIP1, PALB2, RAD51C and RAD51D	Ligation des trompes, salpingectomie, contraception orale
<b>SBG</b>	Hyperplasie de tube papillaire ou endosalpingiose	Cellule épithéliale sécrétoire des trompes de Fallope	Tumeur séreuse borderline	non	non
<b>CCC</b>	Endométriose ou adénofibrome	Cellule endométriale	Tumeur endométrioïde borderline	Lynch syndrome (MLH1, PMS2, MSH2 and MSH6)	Ligation des trompes, salpingectomie, contraception orale
<b>endométrioïde</b>	endométriose	Cellule endométriale	Tumeur endométrioïde borderline	Lynch syndrome (MLH1, PMS2, MSH2 and MSH6)	Ligation des trompes, salpingectomie, contraception orale
<b>Mucineux</b>	Inconnue : Tératome, endométriose ou jonction tubo-péritonéale	inconnue	Tumeur mucineuse borderline	non	Ligation des trompes et contraception orale

Avant de préciser le précurseur de HGSOC, il est important d'expliquer son origine tissulaire. En effet, de nombreuses études démontrent que ce sous-type, le plus fréquent d'OC, est originaire de lésions de la fimbria<sup>35</sup>. Celle-ci correspond à la partie située à l'extrémité des trompes de Fallope. Des lésions pré-néoplasiques peuvent apparaître au niveau de l'épithélium tubaire de cette fimbria. Ces lésions sont caractérisées par les marqueurs p53 et Ki67, observables en ImmunoHistoChimie (IHC) et par leur altération architecturale, observable grâce à un protocole appelé SEE-FIM (Sectioning and Extensively Examining the Fimbriated End)<sup>21</sup>. Ces lésions sont classées en 3 grands groupes et sont présentées dans la (**Figure 19B**). Premièrement, la lésion p53 qui est caractérisée par une prolifération cellulaire normale et une signature p53 abondante signalant une mutation non-sens en TP53 sur une dizaine de cellules. Ensuite, la lésion STIL (serous tubal intraepithelial les

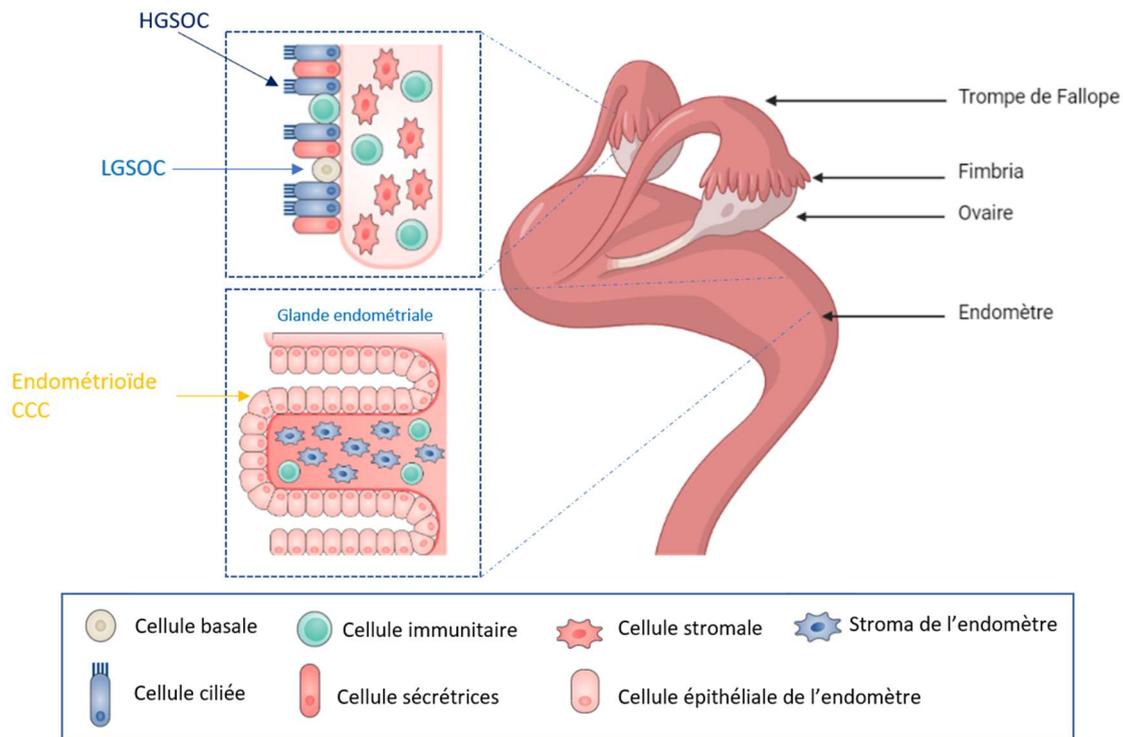
ions), où les cellules prolifèrent également normalement, possède une signature p53 à la fois plus importante et sur un plus grand nombre de cellules auquel sont ajoutées des anomalies morphologiques. Pour finir, la lésion STIC (serous tubal intraepithelial carcinoma), est caractérisée par une prolifération cellulaire importante et par grand nombre de cellules présentant une altération à la fois architecturale et nucléaire ainsi que la mutation TP53<sup>36</sup>. Cette dernière lésion, la lésion STIC, est définie comme l'origine de HGSOV. En effet, il est supposé qu'à la suite de l'apparition du STIC sur les trompes, les cellules cancéreuses vont continuer à se développer par sa proximité avec l'ovaire puis celles-ci vont finir par se détacher de leur membrane basale pour venir se nicher sur l'épithélium ovarien et débuter son développement tumoral et devenir progressivement un carcinome séreux de haut grade. Néanmoins, l'hypothèse selon laquelle la lésion STIC de la fimbria serait à l'origine du cancer séreux de haut grade est à l'heure actuelle très discutée. Pour valider cette hypothèse, nous avons réalisé une étude protéomique spatiale et de SpiderMass des lésions néoplasiques chez des patientes BRCA1 et BRCA2 mutés.

Contrairement au carcinome séreux de haut grade, le carcinome séreux de bas grade, bien que portant un nom similaire, ne se développe pas à partir de lésions de la fimbria, mais prend bien naissance au niveau des trompes de Fallope. Il serait quant à lui issu de tumeurs borderlines originaires des cellules basales de la trompe.

Le carcinome de type endométrioïde se développe, comme son nom l'indique, à partir de cellules endométriales et donc de l'utérus. L'endométriose est très fréquemment liée à ce sous-type de cancer. Il en est de même pour le carcinome à cellules claires.

Le sous-type mucineux n'est que très peu étudié et sa véritable origine est encore méconnue, bien que des hypothèses soient formulées. La première étant que le carcinome se développe à partir d'un épithélium bénin vers une tumeur borderline à un carcinome invasif. La possibilité que l'apparition d'un tératome puisse transformer la cellule ovarienne en une cellule de type gastro-intestinale et ainsi avoir un phénotype totalement différent des autres cancers est également une hypothèse. Il est également possible que ce soit dû à une métaplasie de cellule au niveau de la jonction tubo-péritonéale qui se niche dans l'ovaire<sup>37</sup>.

La **figure 4** schématise l'ensemble de ces sous-types et leurs possibles origines. Le carcinome mucineux, dont l'origine reste très débattue, n'est pas représenté.



**Figure 4:** Origine tissulaire et cellulaire de chaque sous-type de cancer. (Extrait et adaptée de <sup>35</sup> via l'utilisation de biorender).

Le cancer séreux de haut grade (HGSOC) est originaire des cellules épithéliales de la fimbria. Le cancer séreux de bas grade (LGSOC) est originaire des cellules basales des trompes de Fallope. Les cancers endométrioïdes et à cellules claires (CCC) proviennent quant à eux de l'épithélium endométriale.

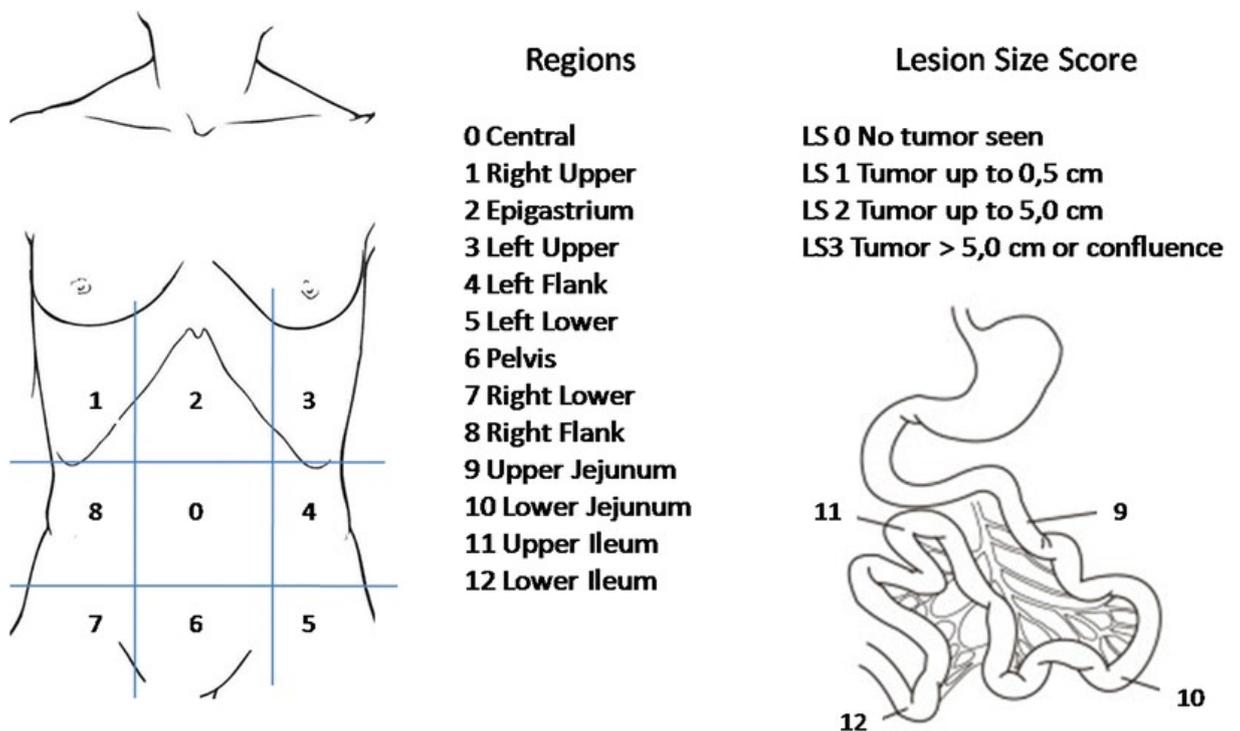
## Diagnostic et pronostic

### Diagnostic

La première étape du diagnostic est réalisée soit par TEP-scanner (tomographie à émission de positron) soit par CT-scan (computed tomography). Le TEP-scanner est un examen scintigraphique détectant un traceur radioactif injecté par intraveineuse qui se fixe sur un marqueur tumoral. Le CT-scan est un scanner rayon X permettant d'imager tranche par tranche le corps selon la densité du tissu<sup>38</sup>. Ces méthodes permettent de déterminer ou non la présence d'une masse tumorale ainsi que la nécessité de réaliser une opération chirurgicale complétant le diagnostic.

Le diagnostic histologique vient faire suite à ces scanners par la réalisation d'une biopsie de la carcinose, un prélèvement du péritoine ou une ablation de l'annexe tumorale par coelioscopie. Cette opération consiste à accéder à l'intérieur de l'abdomen de manière peu invasive via un trocart, une petite incision d'environ 5 à 15mm. L'exploration chirurgicale permet de compléter précisément l'expansion de la maladie dans l'abdomen, et complète la description du TEP-scanner. La coelioscopie apporte la description anatomique précise des lésions et une quantification de la carcinose. Cette description se réalise avec deux types de scores : (i) Le score de Fagotti, ou PI pour « Predictive Index », compris entre 0 et 14 et obtenu

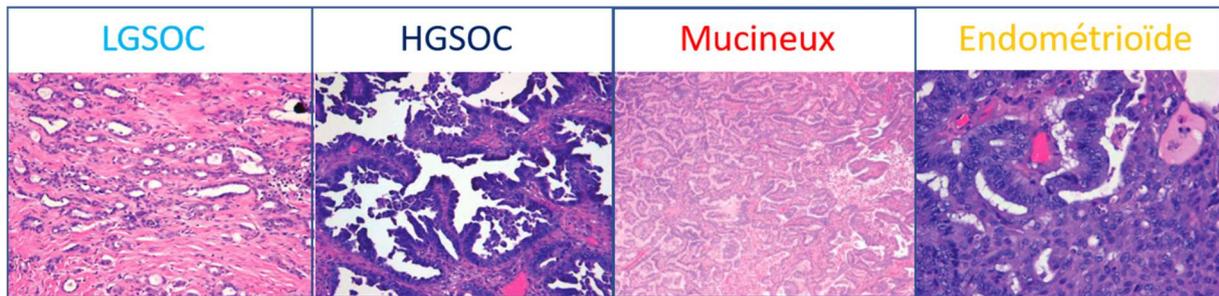
par coelioscopie <sup>39</sup>, et (ii) le score de Sugarbaker, ou PCI pour « Index de carcinose péritonéale », compris entre 0 et 39, obtenu en laparotomie. En résumé, cette quantification consiste à compter le nombre de carcinoses présentes au sein de l'abdomen de la patiente, divisé en différentes zones. Sur chaque zone, le nombre de carcinoses est indiqué et un score lui est attribué. Ce score permet de déterminer si la patiente est opérable ou non<sup>40</sup> (**Figure 5**). Si la patiente présente un score faible, elle pourra être directement prise en charge en chirurgie. En cas de score élevé, des doses de chimiothérapie lui seront prescrites pour réduire les masses tumorales avant l'intervention chirurgicale.



**Figure 5:** Index de carcinose péritonéale de SugarBaker (extrait de <sup>40</sup>).

Lors de la coelioscopie le chirurgien détermine le nombre de carcinose et la masse tumorale de chaque carcinose dans l'abdomen et détermine l'opérabilité de la patiente.

La pièce chirurgicale est ensuite envoyée à un anatomopathologiste, expert en gynécologie, qui déterminera la nature physiopathologique du tissu ainsi que le type de cancer associé. Pour cela, des examens macro et microscopique sont réalisés basés sur la morphologie des cellules observable à la suite d'une coloration HES (Hemalun Eosine Safran). Un marquage immunohistochimique ciblant des marqueurs tels que p53 ou Ki67 complètent cette coloration<sup>41</sup>. La **figure 6** montre les différences morphologiques entre les différents sous-types de cancer.



**Figure 6:** Coloration HES des différents sous-types de cancer (extrait de <sup>41</sup>.)

*La première coloration dévoile un cancer séreux de bas grade. La seconde correspond à un cancer séreux de haut grade. La troisième montre un carcinome mucineux et la dernière un carcinome endométriöïde ressemblant à ceux retrouver au niveau de l'endomètre. Les noyaux sont colorés en bleu, les basophiles en pourpre, les éosinophiles en rouge, le collagène en rose.*

Le LGSOC présente un stroma important (coloration rose) et une faible quantité de noyau (coloration violette) contrairement au HGSOC dont la croissance est beaucoup plus rapide et intense. Les carcinomes endométriöïdes se distinguent par la formation de tubules et de papilles. Les carcinomes mucineux ressemblent un peu plus aux carcinomes endométriöïdes car ils présentent des invaginations ainsi que, comme son nom l'indique, de la mucine, qui n'est malheureusement pas toujours visible. La **figure 6** montre les cas les plus simples et distinguables de ces sous-types de cancers. En réalité, leur distinction est bien plus difficile et peut engendrer, dans les cas les plus complexes, jusqu'à 30% d'erreurs. Ces diagnostics sont souvent associés à une IHC ciblant certains marqueurs de ces sous-types. L'inconvénient majeur de cette technique est le temps de réalisation. En effet, entre l'arrivée de la biopsie et son diagnostic, dépendamment de la disponibilité de l'anatomopathologiste, il peut se dérouler jusqu'à 2 semaines pour un diagnostic complet. Cela ralentit grandement la prise en charge chirurgicale de la patiente, l'adaptation de son traitement et donc sa survie.

Ainsi, il est nécessaire d'apporter de nouvelles technologies dans l'aide au diagnostic, à la fois au chirurgien et à l'anatomopathologiste. Les données moléculaires devraient être utilisées pour définir avec précision et rapidité la zone cancéreuse et ainsi améliorer le diagnostic et la survie globale des patientes.

### Pronostic

Le pronostic de la patiente est étroitement lié à la qualité de la chirurgie<sup>42</sup>. En effet, la survie sans récurrence est corrélée à la capacité du chirurgien à réaliser une chirurgie complète de la masse tumorale. Ainsi, cet acte chirurgical nécessite un chirurgien et une équipe de bloc opératoire entraînés et compétents, mais également des soins per-opératoires adaptés. En effet, une revue récente a montré qu'une gestion per-opératoire adaptée diminuait la morbidité post-opératoire<sup>43</sup>. De plus, ce geste chirurgical est variable en fonction du type histologique. Par exemple, l'ablation de l'appendice est quasiment obligatoire dans le cas d'un carcinome

mucineux tandis qu'il est peu recommandé de réséquer les ganglions sentinelles dans les cas de LGSC <sup>44</sup>.

La survie des patientes dépend énormément du type histologique du cancer et de son diagnostic. En effet, présentant des symptômes très tôt, les carcinomes mucineux et endométrioïdes sont généralement diagnostiqués à des stades précoces (stade I) et ont donc de bons pronostics. De plus, le cancer endométrioïde a pour particularité d'être chimio-sensible ce qui permet d'optimiser son traitement. Le LGSOC a également un bon pronostic pour les mêmes raisons. Le cancer séreux de haut grade a, quant à lui, un très mauvais pronostic. En effet, le taux de mortalité s'élève à 70% à 10 ans<sup>45</sup>. Cela est principalement dû à son diagnostic trop tardif (stade III/IV). À ce stade, le cancer présente déjà des carcinoses et des métastases plus lointaines, notamment au niveau du péritoine. Parmi ces cancers, pour des raisons jusqu'alors ignorées, certains ont pour particularité de récidiver de manière très précoce (< 6 mois). Il est donc nécessaire de développer un outil permettant de diagnostiquer ces cas plus rares et ainsi adapter le traitement de ces patientes.

### Prévention et traitement

#### La prévention

La chirurgie préventive est la première source de traitement. Une option possible est la salpingo-ovariectomie bilatérale, c'est-à-dire la résection totale des trompes de Fallope ainsi que des ovaires. Cela vise à réduire les risques de cancer chez les patientes considérées à haut risque en raison de mutations génétiques identifiées, telles qu'une mutation BRCA1 et BRCA2. Cette opération comporte peu de risques. Seulement, elle peut entraîner des effets secondaires post-opératoires, tels que de la sécheresse vaginale, une diminution de la fonction sexuelle, et même une diminution de la capacité à atteindre l'orgasme<sup>46,47</sup>. Ces chirurgies prophylactiques ont considérablement réduit l'incidence du cancer chez les porteuses de mutations BRCA (jusqu'à 80%). Cependant, l'élimination des tissus à risque, bien qu'étant une stratégie très efficace pour réduire les taux de cancer et de mortalité, doit être envisagée en tenant compte de l'âge de la femme, de sa fertilité et de ses comorbidités.

La ligature des trompes est également une option permettant de réduire significativement le risque de cancer de l'ovaire, les résultats étant comparables à ceux de l'utilisation de contraceptifs oraux.<sup>46</sup>

La contraception orale semble réduire de l'ordre de 40 à 50% les risques de développer un cancer de l'ovaire. En effet, celle-ci va venir bloquer le processus d'ovulation qui serait un des facteurs de risque liés à l'apparition de cancer<sup>47</sup>.

Sachant que le sous-type le plus agressif de cancer est le HGSOC et que celui-ci serait originaire de lésions de la fimbria, une nouvelle stratégie préventive de fimbriectomie a été mis

en place. Elle consiste à retirer uniquement les trompes de Fallope aux patientes atteintes des gènes de prédisposition BRCA1 et BRCA2 pour éviter l'apparition de lésions. Ainsi, il est possible de réduire au maximum l'impact de la chirurgie sur la patiente tout en s'assurant la réduction de risque d'apparition d'un cancer ovarien<sup>48,49</sup>. Sur une étude de 121 patientes, l'analyse anatomopathologique de pièces chirurgicales de patientes ayant subi cette opération a permis de retrouver un carcinome séreux intraépithélial dans le pavillon de la trompe, un STIC et 23 autres lésions tubaires, pouvant donc être originaire d'un possible HGSOc. Par ailleurs, après un suivi médian de 7,3 ans sur les 121 patientes incluses (agées de 28 à 48 ans) aucune n'a développé de carcinome séreux pelvien<sup>48</sup>. D'autres études se poursuivent pour proposer aux femmes à risque une chirurgie prophylactique efficace et moins invalidante. Pour être éligibles à ces chirurgies préventives, un test génétique préalable est demandé aux patientes avec des antécédents familiaux de cancers de l'ovaire et du sein, et/ou des prédispositions génétiques majeures. La patiente a le choix de réaliser cette intervention ou non, selon son souhait de maternité<sup>46</sup>.

### Traitement

Le traitement des cancers de l'ovaire est majoritairement basé sur la chirurgie couplée à la chimiothérapie. Celle-ci tient compte de multiples facteurs afin d'adapter au mieux le processus de soin de la patiente. En effet, la charge tumorale évaluée en imagerie et par coelioscopie, l'état général de la patiente, ainsi que le stade FIGO de la tumeur sont des éléments cruciaux pour la stratégie thérapeutique<sup>50</sup>. Lorsque la masse tumorale est trop importante, une cure de chimiothérapie est réalisée avant la chirurgie pour réduire la tumeur. De manière générale, le traitement correspond à 6 cycles de chimiothérapie avec carboplatine AUC 6 et paclitaxel 175 mg/m<sup>2</sup> toutes les 3 semaines<sup>51</sup>.

Le paclitaxel stimule la formation de microtubule et les stabilise tout en empêchant leur dépolymérisation<sup>52</sup>. Ainsi, cette molécule empêche la croissance tumorale en inhibant totalement la mitose. De plus, elle stimule la réponse immunitaire puisqu'elle est un ligand activateur du TLR4, récepteur exprimé par les cellules immunitaires et impliqué dans la production de facteurs inflammatoires<sup>53</sup>.

Le carboplatine, en revanche, se classe parmi les cytostatiques et partage des propriétés biochimiques comparables à celles du cisplatine. Il se lie à la molécule d'ADN en engendrant des liaisons alkyles qui mènent à la création de ponts entre les deux chaînes, soit au sein de la même molécule, soit entre les chaînes de molécules d'ADN adjacentes. Cette interaction perturbe la réplication de l'ADN ainsi que la séparation de ses brins. En conséquence, les processus de synthèse d'ARN (acide ribonucléique) et des protéines cellulaires sont également affectés.

D'autres thérapies sont également mises en place telles que les inhibiteurs anti-PARP (polyadénosine diphosphate ribose). Les PARP sont des enzymes impliquées dans la réparation des cassures simples brins de l'ADN. Leur inhibition induit ainsi une instabilité génomique qui conduit à la mort cellulaire, notamment sur les patientes présentant un déficit dans la réparation de l'ADN. Son utilisation semble augmenter le taux de survie sans récurrence<sup>54</sup>.

La chimiothérapie hyperthermique intrapéritonéale (CHIP) consiste à déposer durant 90 min à 40°C du cisplatine combiné à du thiosulfate durant l'opération, directement dans la cavité abdominale, pour éliminer le plus de métastase possible à la surface des tissus. Malgré l'efficacité de l'utilisation du cisplatine en intra-veineuse, il semble que cette technique n'apporte pas de meilleurs résultats et pourrait également avoir des effets secondaires sur la qualité de vie de la patiente post-opératoire<sup>55</sup>.

Comme évoqué précédemment, la chimiothérapie n'a pas le plus gros impact sur la survie de la patiente. En effet, elle complète la chirurgie, qui reste le traitement le plus efficace. Ainsi, le chirurgien doit parfaitement définir la zone tumorale afin de réaliser une résection la plus complète possible. Seulement, il doit éviter de retirer trop de tissus sains pour réduire le risque de complications post-opératoire. Ainsi, il est nécessaire d'apporter une aide à la décision au chirurgien durant l'acte chirurgical qui tiendrait compte de l'hétérogénéité de ce cancer particulier. Dans ce contexte, de nouvelles techniques d'analyse peropératoire ont vu le jour pour répondre à ces différents besoins cliniques.

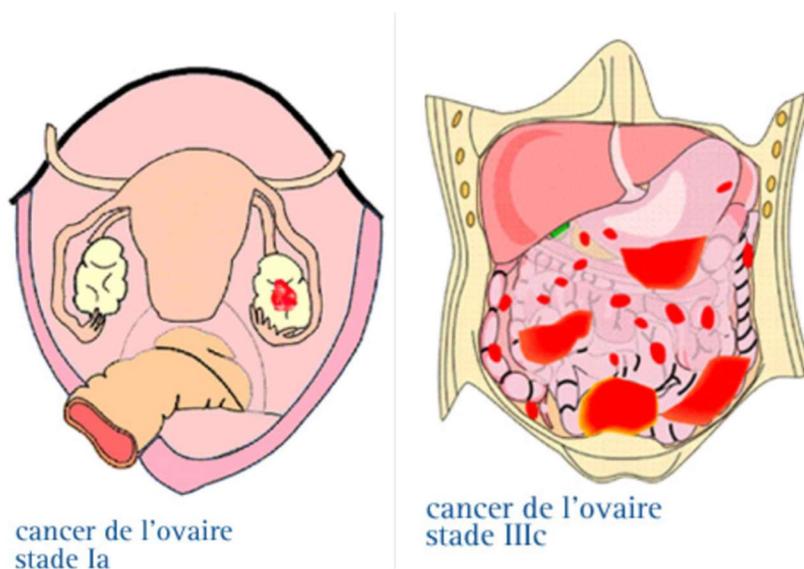
## Nouvelles méthodes de diagnostics peropératoire

Les techniques conventionnelles d'imagerie médicale et de diagnostic ne s'avèrent pas être suffisante pour une détection précise et efficace des cancers de l'ovaire. Dans l'objectif de retirer l'entièreté des cellules tumorale et améliorer la survie des patientes<sup>9</sup>, il est crucial d'apporter *in vivo* et en temps réel des données moléculaires au chirurgien pour lui permettre de prendre des décisions pendant la chirurgie. Pour répondre à cette problématique, de nombreuses techniques d'imagerie permettant une analyse peropératoire des tissus *ex vivo* et/ou *in vivo* ont vu le jour.

### Analyses peropératoire *ex vivo*

La méthode standard d'analyse peropératoire *ex vivo* reste la coloration HES (hématoxyline éosine safran) sur coupe de tissu frais, excisé par le chirurgien. Le pathologiste examine la pièce chirurgicale envoyée afin d'identifier sa nature tumorale ou non. La coloration HES permet une bonne observation de la morphologie cellulaire et tissulaire, c'est ainsi que

le pathologiste peut poser un diagnostic<sup>56</sup>. Le chirurgien peut également désigner la partie de la pièce chirurgicale dont il suspecte la nature pour définir la taille de la marge à réaliser. Cependant, cette méthode est chronophage puisqu'elle nécessite la disponibilité d'un pathologiste expérimenté et peut prendre entre 20 et 50 minutes en moyenne en fonction de l'organisation du centre, voir plus pour des cas compliqués. De plus, le diagnostic est rendu sur les tissus retirés par le chirurgien sur une base qui reste subjective et le pathologiste ne peut donc pas toujours établir son diagnostic sur l'ensemble de la marge ou l'intégralité des tissus qui devraient être étudiés. Une classification des cancers ovariens a ainsi été générée par les pathologistes, la classification FIGO (**Tableau 1 et figure 7**).



*Figure 7: Comparaison de cancer de l'ovaire au stade FIGO précoce et avancé.*

Pour améliorer la précision du diagnostic, des chercheurs ont récemment intégré l'utilisation de l'intelligence artificielle et plus particulièrement, des réseaux neuronaux (CNN) sur la reconnaissance des colorations histologiques. Des résultats prometteurs ont été obtenus et cet outil permettrait un diagnostic à l'aveugle rapide de certaines pathologies<sup>57</sup>. Néanmoins, le diagnostic précis passe souvent par une analyse en immunohistochimie qui est très chronophage. Cette analyse immunohistochimique est systématiquement réalisée sur des marqueurs spécifiques du cancer tels que p53 ou Ki67. Les immunoréactions p53 sont positives dans 41,66 % des tumeurs séreuses malignes, la plupart d'entre elles (90 %) étant des carcinomes de haut grade et 6,66 % des tumeurs « borderlines » et aucune tumeur bénigne. Le marquage Ki67 est positif dans 61,53 % des cas de tumeurs malignes, avec des

pourcentages plus élevés dans les stades cliniques avancés. Les immunoréactions Ki67 sont également positives dans les tumeurs borderlines et bénignes, mais avec des pourcentages plus faibles, 13,3 % et 9,09 % respectivement<sup>58</sup>. Ces marqueurs sont assez sensibles pour déterminer la présence ou non de tumeurs malignes mais cette technique est chronophage. En effet, dans des conditions optimales, l'obtention du résultat d'une coloration IHC nécessite au minimum 24h.

D'autres techniques sont apparues pour permettre la détection de marge *ex vivo*. C'est le cas de l'imagerie par fluorescence. En utilisant des anticorps ciblant des marqueurs tumoraux couplés à des fluorophores, il est possible d'observer, notamment à l'aide d'un microscope confocal, un marquage des tissus avec une résolution spatiale de l'ordre de 1µm. L'avantage de cette technique est la possibilité de travailler en profondeur dans le tissu et de déterminer la marge couche par couche. Il est ainsi possible de poser précisément un diagnostic sur la marge à réaliser. Seulement, cette technique est très chronophage et ne permet qu'un diagnostic post-opératoire *ex vivo*.<sup>59</sup>

Les techniques spectroscopiques, telles que la spectroscopie Raman, ont également été étudiées *ex vivo* pour l'étude des marges chirurgicale. Elles permettent l'analyse des molécules sans utilisation de marquage au préalable. Elle fonctionne sur le principe de diffusion de la lumière et de la matière, chaque molécule possédant un spectre vibrationnelle différent. De ce fait, chaque tissu a un spectre vibrationnelle propre causé par un profil moléculaire distinct. De nombreuses études ont démontré son intérêt dans le diagnostic tumoral<sup>60</sup>. Seulement, les spectres acquis sont complexes et difficiles à analyser. De plus, cette technique présente une sensibilité limitée, ne permet qu'une analyse de surface et ne permet pas d'identifier la nature moléculaire des marqueurs liés à la pathologie.

L'imagerie hyperspectrale (HS) a également montré un potentiel intéressant pour le diagnostic tissulaire non invasif. Elle consiste en une imagerie de réflectance, mesurant la proportion de lumière réfléchi, et est capable de travailler sur une large quantité de bandes spectrales différentes. Ainsi, cette technique permet de définir des empreintes optiques spécifiques d'un état physiologique<sup>61</sup>. Elle a donc été utilisée sur des biopsies de patients pour différencier des échantillons cancéreux d'échantillons sains<sup>62</sup>. Elle est principalement utilisée *ex vivo* de part des limitations techniques, mais peu trouver également son intérêt *in vivo*.

Bien que quelques preuves de concept aient été réalisées, ces technologies sont encore à l'état de recherche, et ne sont pas encore généralisées. Mais elles permettront d'aider le chirurgien dans sa prise de décision une fois utilisées en routine.

### Analyses peropératoire *in vivo*

Des méthodes se sont également mises en place *in vivo* pour aider le chirurgien dans sa prise de décision pendant la chirurgie. C'est le cas du CT-scan (computed tomography) qui permet une analyse directe de la densité tissulaire par rayon X chez le patient <sup>38</sup>. C'est également le cas de l'IRM qui permet la distinction des différents tissus en fonction de leur composition moléculaire. Ces deux techniques, bien en place dans les procédures cliniques, ont certains désavantages. En effet, elles ont une faible résolution spatiale et ne permettent pas une délimitation fine des zones de marges pour le chirurgien. De plus, les images obtenues ne sont pas toujours identiques à ce que peut observer le chirurgien en temps réel, dû à la croissance tumorale entre le scanner et l'opération, à un désalignement de l'image ou encore à la déformation de tissus mous. Il est également à prendre en considération que le CT-scan nécessite une exposition prolongée à des radiations ionisantes qui peuvent être une source d'effets secondaires pour le patient. Enfin, ces techniques nécessitent une installation coûteuse et spécifique pour être utilisées.<sup>9</sup>

La technique intra opératoire à ultrason (US) offre également une aide pré et peropératoire au chirurgien. Elle permet notamment de discerner les différentes densités tissulaires et donc de repérer la masse tumorale. Seulement, cette technologie possède une très faible résolution spatiale qui ne permet pas de définir avec précision la zone d'exérèse. Cependant, elle a pour avantage d'être réalisable en temps réel et de pouvoir aider le chirurgien pendant la chirurgie <sup>63</sup>.

La chirurgie guidée par fluorescence est aussi une belle avancée dans l'aide à la prise de décision du chirurgien. Elle consiste en l'injection par intraveineuse, avant la chirurgie, de marqueurs fluorescents connus pour être plus abondant dans les tissus cancéreux que dans les tissus sains. Ainsi, les tissus marqués émettent un signal fluorescent dans le proche infra-rouge. Ce signal permet la visualisation des tissus à réséquer à l'aide d'une caméra capable de détecter cette fluorescence. Cette technique a été utilisée dans des phases cliniques 1 et 2 mais possède encore des limitations. En effet, il est nécessaire de connaître des biomarqueurs spécifiques et sensibles du type du cancer en question. De plus, il a été observé que l'émission de cette fluorescence est plus large que la zone cancéreuse, ce qui diminue grandement la résolution spatiale de cette aide. De plus, l'émission de cette fluorescence reste encore très faible<sup>64</sup>.

La technologie « optical coherence tomography » (OCT) consiste, quant à elle, à réaliser une image du tissu analysé par rapport à la lumière réfléchiée. Elle est assez similaire à la technique basée sur l'ultrason. Cette technique possède une bonne résolution spatiale (10-30µm) mais des études rapportent sa variabilité en termes de sensibilité et spécificité qui

décroit lorsque la profondeur d'analyse augmente<sup>65</sup>. En effet, cette technique peine à réaliser des analyses en profondeur et se focalise donc uniquement sur des études de surface tels que le cancer du sein<sup>66</sup>.

Les techniques spectroscopiques ont également montré leur intérêt dans l'analyse peropératoire, mais la plupart partagent l'obstacle de l'analyse de surface. C'est le cas de la technologie SERS (Surface Enhanced Raman Spectroscopy). Cette technique consiste en l'analyse du spectre Raman en champs proches. Afin de passer de l'*ex vivo* à l'*in vivo*, il est nécessaire de travailler en surface et d'injecter au préalable des nanoparticules permettant l'amélioration du signal obtenu. Ainsi, cette technique ne peut pas être appliquée en profondeur en raison de sa sensibilité 6 à 10 fois plus faible que la fluorescence. De plus, les spectres obtenus sont très complexes et il est parfois difficile de discerner un spectre tumoral d'un spectre bénin<sup>56</sup>.

L'imagerie hyperspectrale peut être utilisée *in vivo*. Seulement, elle n'est pas encore utilisée couramment du fait de multiples contraintes. La taille de la plateforme d'acquisition et le besoin que le tissu analysé reste statique en sont les principaux responsables. De plus, la technologie éprouve des difficultés à analyser des surfaces non planes, compliquant son utilisation au bloc opératoire. Cependant, des systèmes d'endoscopie hyperspectral sont en cours de développement et pourront potentiellement franchir les différents obstacles évoqués<sup>67</sup>.

Pour finir, la spectrométrie de masse, en développement depuis plusieurs dizaines d'années, pourrait permettre de dépasser les limites des différentes techniques actuelles. La spectrométrie de masse se veut rapide, sensible et spécifique. L'apparition des sources ambiantes intra-opératoire semble donc être une bonne alternative pour un diagnostic en temps réel au bloc opératoire.

## La spectrométrie de masse.

La spectrométrie de masse (MS) a pour but d'identifier (et si besoin quantifier) certaines des molécules présentes dans un échantillon. La MS est caractérisée par une bonne sensibilité qui permet de détecter les variations moléculaires entre différentes conditions pour un même échantillon e.g., les variations liées à un changement dans le phénotype des cellules seront observables en MS. Le développement de l'imagerie par spectrométrie de masse (MSI) a apporté une dimension spatiale qui a ouvert de nombreuses applications, en particulier dans le diagnostic des cancers<sup>68</sup>. En effet, la cellule cancéreuse présente un phénotype particulier qui conduit à des différences notables dans la régulation des différentes familles de molécules. Par exemple, en protéomique des changements dans l'expression des protéines, de leurs

modifications post-traductionnelles ou encore l'observation de mutations sont retrouvées. Cependant, les métabolites comme les lipides présentent une régulation différente dans les cellules cancéreuses. Même si les marqueurs lipidiques ont longtemps été considérés comme moins instructifs il est maintenant reconnu que le métabolisme particulier de la cellule cancéreuse induit des changements importants dans l'expression de ces composés<sup>69</sup>.

### Principe de la spectrométrie de masse

La spectrométrie de masse (MS) est une technique analytique utilisée pour identifier des composés d'origines diverses, de complexité variable (purs ou en mélange complexe) et de poids moléculaire différent (des atomes aux complexes protéiques). La configuration la plus simple d'un instrument de MS se compose de trois parties : une source, un analyseur de masse et un détecteur. Le principe de la MS est de séparer les ions en phase gazeuse en fonction de leur rapport masse-sur-charge ( $m/z$ ). Les ions sont créés dans la source puis dirigés et séparés dans l'analyseur de masse, la partie centrale du spectromètre de masse, qui fonctionne sous vide poussé et utilise des champs électriques ou magnétiques pour séparer les ions. Ces ions sont ensuite convertis par le détecteur en un courant mesurable et le résultat de la mesure représenté sur le spectre de masse qui présente l'intensité du courant en unités arbitraires en fonction du rapport masse/charge ( $m/z$ ).

Les premiers instruments développés par Thomson et Aston en 1919 étaient initialement appelés des « spectrographes » de masse, mesurant les ions avec une plaque photographique ou des écrans phosphorescents<sup>70</sup>. Ils sont maintenant appelés des spectromètres de masse car ils mesurent les ions à l'aide de détecteurs électriques.

La principale difficulté de cette technique a été, dans un premier temps, la création d'ions en phase gazeuse. En effet, lors de l'apparition de la MS, une faible variété d'échantillons pouvait être ionisée et analysée. En impact électronique la source permet l'ionisation des composés à l'état gazeux par l'interaction avec un faisceau d'électrons à 70 eV (généralement). Le passage à l'état gazeux est assuré par chauffage d'un filament sur lequel est déposé le composé, limitant l'analyse aux composés peu polaires et de faible poids moléculaire ( $\leq 500$  u. environ). Malgré les améliorations apportées par l'introduction de l'ionisation chimique (CI) le principe d'introduction reste le même et ne permet pas l'analyse de composés plus polaires et de plus haut poids moléculaire. Ainsi, les protéines, les acides nucléiques et autres biomolécules ne pouvaient pas être analysées.

Par la suite, l'introduction du SIMS où les ions sont obtenus par bombardement d'une surface par un faisceau d'ions primaires fortement accélérés représente un premier changement de paradigme puisque le bombardement entraîne à la fois le passage à l'état gazeux et l'ionisation. Cependant, le SIMS reste limité en routine à l'analyse des composés jusqu'à environ 1000 u.

Par la suite, des méthodes telles que la désorption par plasma (PDMS), le « fast atom bombardment » (FAB) ou la « Field Desorption » (FD) ont commencé à surmonter ces difficultés. En effet, il était possible de détecter de petites protéines grâce à ces méthodes. Seulement, ces protéines devaient être en large excès dans l'échantillon et ces sources présentaient des limitations instrumentales, comprenant notamment la génération de fragments de fission du  $^{252}\text{Cf}$  pour le PDMS. Ces difficultés ont été surmontées par le développement presque simultané des sources de désorption/ionisation laser assistée par matrice (MALDI)<sup>71 72</sup> et de l'ionisation par électronébulisation (ESI)<sup>73</sup>.

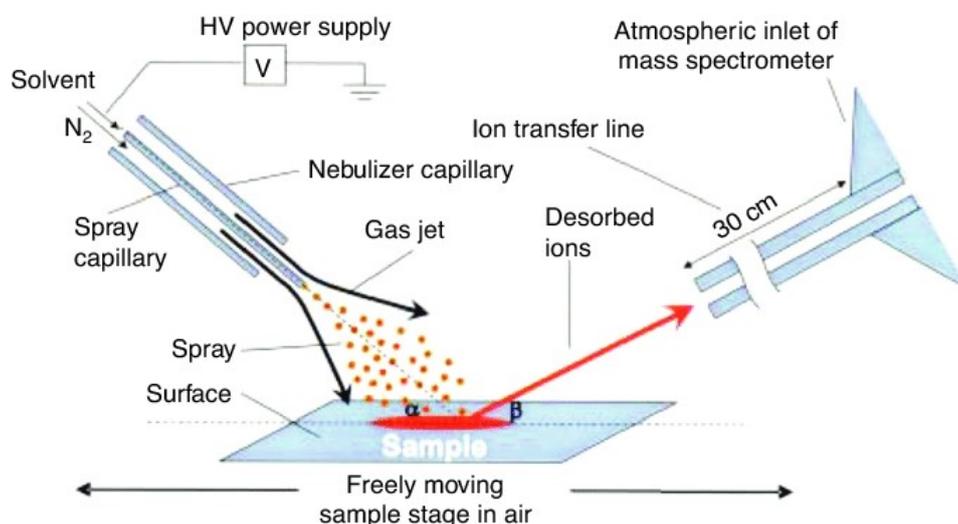
La source MALDI est dérivée de la désorption/ionisation laser (LDI) mais repose sur l'utilisation d'une matrice dont le rôle est d'assurer un transfert d'énergie efficace en absorbant à la longueur du laser d'excitation tout en protégeant l'analyte d'un excès d'énergie et limitant les fragmentations. La matrice MALDI est spécifique au type de biomolécule étudiée car elle dépend des propriétés physico-chimiques des molécules à étudier. Les expériences ont montré qu'elle doit être introduite en large excès pour un maximum d'efficacité. Elle est donc co-cristallisée avec les analytes. L'irradiation par un laser UV pulsé de l'échantillon conduit alors à un processus d'explosion de phase concomitant avec une création d'ions en phase gazeuse (désorption/ionisation)<sup>74</sup>. L'ESI, quant à lui, repose sur un principe différent où des gouttelettes fortement chargées sont produites par un mécanisme d'électro-nébulisation à pression atmosphérique. L'évaporation de ces gouttelettes assistée par différents moyens (gaz séchant, capillaire de transfert chauffé, tension sur les skimmers) conduit à des fissions répétées puis à la formation des ions en phase gazeuse qui sont alors transférés à l'analyseur de masse<sup>75</sup>.

Tant le MALDI que l'ESI permettent l'analyse d'une large gamme de molécules, des petites molécules de faible polarité aux grosses molécules et aux molécules polaires (e.g. les protéines), et sont des méthodes d'ionisation douces où les analytes sont préservés des fragmentations. Le développement de ces deux sources a constitué un véritable tournant pour l'application de la spectrométrie de masse en biologie et ce sont aujourd'hui les deux techniques les plus utilisées. L'un des grands avantages de l'ESI est la possibilité d'être couplée en amont à la chromatographie liquide (LC-MS), ce qui permet une analyse approfondie de mélanges d'échantillons hautement complexes, mais surtout la possibilité de travailler à pression atmosphérique (AP). De son côté le MALDI est plus tolérant aux sels et contaminants et est réalisé très rapidement. Cependant, aucune des deux sources ne permet une analyse sur des échantillons bruts non traités, même si le MALDI s'en rapproche avec l'analyse des sections histologiques de tissus.

### Les sources ambiantes en spectrométrie de masse

Les sources AIMS (spectrométrie de masse par ionisation ambiante) permettent l'analyse d'échantillons bruts, sous leur forme intacte, tout en réduisant, voir en supprimant, les étapes de préparation de l'échantillon<sup>76</sup>. De plus, contrairement aux sources APMS (spectrométrie de masse à pression atmosphérique) classiques telles que l'ESI, qui nécessitait jusqu'alors l'analyse d'un échantillon dans l'état liquide, les sources AIMS permettent l'analyse de tout type d'échantillon, qu'il soit liquide, solide ou gazeux. Il existe actuellement un grand nombre de ces sources qui se distinguent par leur méthode d'extraction et d'ionisation.

*Les sources AIMS reposant sur une extraction de surface :*



**Figure 8:** Schéma global du fonctionnement de la source DESI.

*Un solvant chargé et sous haute tension est sprayé à l'aide de gaz au niveau d'un échantillon. Le spray généré rebondit sur l'échantillon et permet la désorption et l'ionisation des biomolécules qui sont amené au MS par un capillaire de transfert<sup>77</sup>.*

Le DESI (Ionisation et désorption par électronébulisation) est l'une des premières sources d'ionisation ambiante introduite. Elle a été développée par le professeur Cooks<sup>77</sup> en 2004, et depuis elle n'a cessé d'être utilisée et améliorée. Le DESI permet de travailler dans des conditions ambiantes avec un minimum de traitements de l'échantillon. Le fonctionnement de cette source est proche de la source ESI. En effet, à l'aide d'un gaz comprimé et d'un champ électrique, un spray de microgouttelettes de solvant chargées est généré et dirigé vers un échantillon (une coupe de tissu par exemple). Ces gouttelettes vont atterrir sur le tissu puis rebondir après avoir capturé des analytes présents à la surface de l'échantillon<sup>77</sup>. Les gouttelettes sont ensuite transportées, *via* une ligne de transfert particulière (initialement appelée « magic wand », vers l'entrée du spectromètre de masse où elles subissent un processus similaire à celui de l'ESI conventionnel comme schématisé dans **la figure 8**.

Contrairement à l'ESI, seul le solvant est présent dans le capillaire de spray du DESI, et non la solution comprenant l'échantillon. De ce fait, les spectres d'ionisation des molécules sont similaires entre ces sources. Chaque biomolécule ayant des propriétés physico-chimiques différentes, notamment l'hydrophobie, la famille d'analytes ionisée dépendra du solvant choisi. Cette source est parfaitement adaptée à l'analyse de petites molécules tels que les lipides, présents en grande quantité au sein des cellules et donc facilement détectable, et ce, sans préparation de l'échantillon. Cependant, contrairement à la source ESI, peu de peptides/protéines (et souvent en excès dans l'échantillon) ont pu être analysée jusqu'alors. La cause semble être due au manque d'énergie appliqué sur l'échantillon <sup>78,79</sup> empêchant une ionisation complète de celui-ci.

Le DESI, trouve un intérêt tout particulier dans la recherche clinique puisqu'il permet l'analyse rapide et sans traitement préalable des échantillons. En effet, le DESI permet d'étudier les mécanismes lipidiques présents au sein d'un tissu à travers la détection, en une seule analyse, des acides gras, des glycérophospholipides, des sphingolipides ainsi que des glycérolipides <sup>80</sup>. En effet, la signature lipidique d'un tissu est étroitement liée aux phénotype des cellules qui le composent lui-même lié aux mécanismes physiologiques et/ou physiopathologiques. Par exemple, des variations de profils lipidiques au sein de l'ovaire ont été associées à la période d'ovulation chez de nombreuses espèces animales. L'importance de la prostaglandine a également été mis en évidence dans la physiologie de l'ovaire<sup>81</sup>.

Étudier les variations lipidiques au sein d'un tissu permet également une meilleure compréhension des cancers étant donné que la composition lipidique des tissus cancéreux varie énormément par rapport aux tissus sains <sup>82</sup>. Ainsi, dans le contexte de l'oncologie, l'imagerie DESI a montré qu'il était possible, à travers une analyse statistique, de segmenter moléculairement la partie saine d'un tissu par rapport au tissu malin. De nouveaux biomarqueurs tels que les phospholipides et les céramides ont pu être découverts *ex vivo* ajoutant une aide moléculaire à l'anatomopathologie<sup>83</sup>. De nos jours, la source DESI est très présente dans la littérature et démontre son réel potentiel dans la découverte de marqueurs lipidiques en oncologie<sup>84</sup> ainsi que dans le dépistage de drogue au sein d'un tissu<sup>85</sup>.

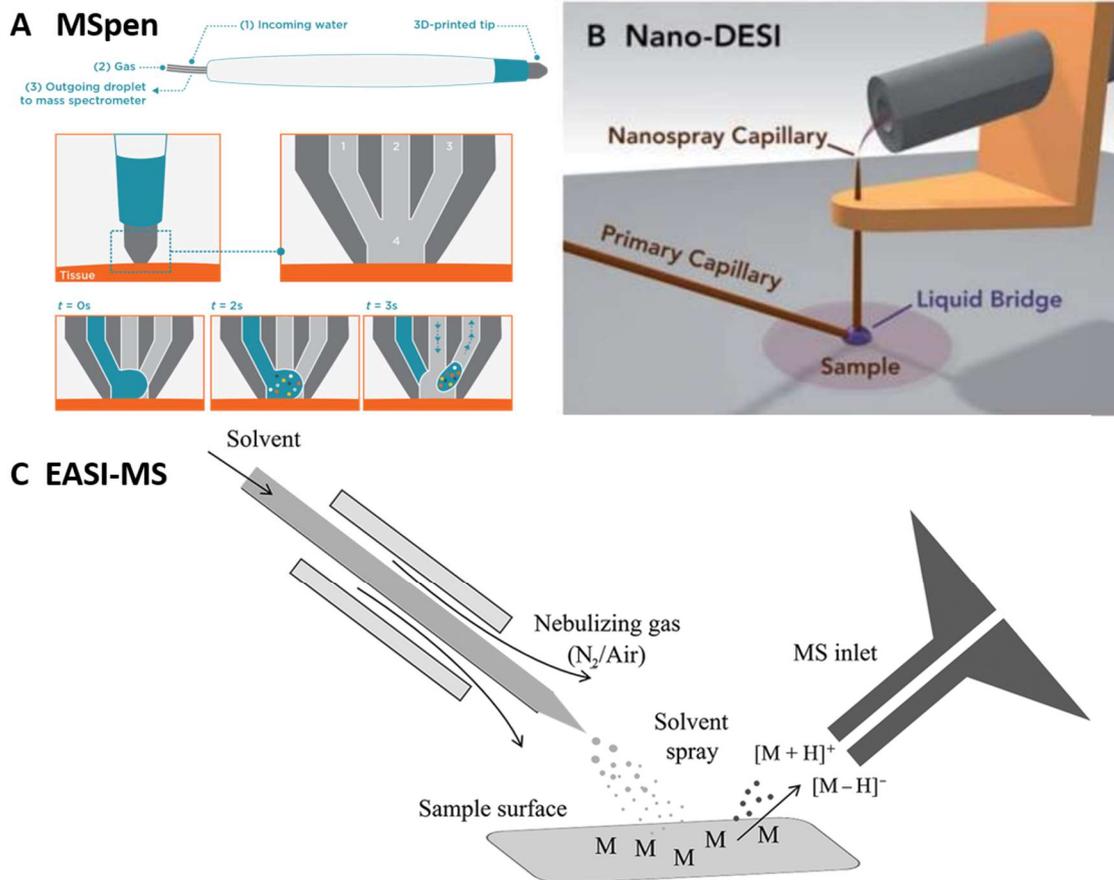
Malheureusement, cette source possède de nombreuses contraintes, dont sa reproductibilité. En effet, afin d'obtenir un spray stable et précis (20-200µm), il est nécessaire d'ajuster la hauteur de l'aiguille selon l'échantillon, la distance entre l'aiguille et le capillaire de transfert, ainsi que l'insertion de l'aiguille de spray au sein de la buse. Ces réglages restent très délicats et les résultats peuvent donc varier selon les laboratoires ou encore selon les expériences<sup>86</sup>. Ce désavantage rend son utilisation clinique difficile et réservée à un petit groupe d'experts. Cette source conserve cependant l'intérêt d'une utilisation en conditions

ambiantes, réduisant la dénaturation des échantillons et donc permettant de mieux comprendre les métabolismes liés aux différentes pathologies. Elle est ainsi la pionnière, avec la source DART (direct analysis in real time), des sources ambiantes et a permis, au même titre que le MALDI, le développement de sources d'ionisation ambiantes connexes.

Directement associée au DESI, la source nano-DESI a fait son apparition en 2010<sup>87</sup> (**Figure 9B**). Leur fonctionnement est proche, mais le nano-DESI s'affranchit de l'utilisation du gaz de nébulisation dans sa création de microgouttelettes chargées. Le nano-DESI repose sur une extraction localisée des analytes via un capillaire de silicate amenant un flux continu de solvant sur la surface de l'échantillon et porté à haute tension. Ce flux est aspiré par une aiguille de nanospray grâce à une technique de micro-jonction liquide<sup>87</sup>. Les molécules ainsi ionisées sont ensuite conduites par le capillaire de transfert vers le spectromètre de masse. Cela permet un prélèvement précis de la surface à haute résolution spatiale (et la création d'images moléculaires, par exemple). Ainsi, le spray de solvant est plus focalisé et les images moléculaires obtenues présentent une meilleure résolution spatiale (jusqu'à la cellule unique).

Une autre mode de fonctionnement dérivée de la source DESI, appelée reactive-DESI, a également vu le jour<sup>88</sup>. Cette technique utilise des réactifs spécifiques, ajoutés au solvant, comme méthode d'ionisation, ce qui améliore la sensibilité de détection pour un type de molécule spécifique. Par exemple, la source DICE (désorption / ionisation par échange de charge) utilise le toluène en tant que solvant pour analyser des espèces plus apolaires<sup>88</sup>.

Enfin, la technique d'EASI permet de s'affranchir de l'utilisation de courant électrique ou de ruban chauffant (**Figure 9C**). Un spray puissant de solvant neutre est appliqué sur la surface de l'échantillon grâce à un simple flux d'azote. Elle a pour avantage de n'appliquer que très peu de modifications à l'échantillon, mais réduit la quantité d'ions détectés<sup>88</sup>.



**Figure 9:** Représentation des différentes sources d'ionisation chimique.

(A) Le MassSpec Pen consiste en un dispositif de la taille d'un stylo. De l'eau est envoyée par une pompe directement au niveau de l'aiguille à l'extrémité du dispositif. L'eau permet l'ionisation des biomolécules au contact de l'échantillon qui sont aspirés dans un second tube en PTFE directement relié au MS. (B) Le nano-DESI consiste en un capillaire en silicate injectant en continu du solvant sur l'échantillon. La micro-jonction entre le premier capillaire et un second au contact de l'échantillon permet la micro-extraction des biomolécules. Celles-ci sont envoyées ensuite dans le MS par une aiguille de nano spray dirigée vers le capillaire de transfert. (C) La source EASI ressemble à la source DESI classique mais n'utilise pas de différence de potentiel. Elle se base sur un spray sous haute pression permettant l'ionisation des molécules, le solvant va ainsi rebondir sur l'échantillon et arracher ses analytes qui sont dirigés vers le capillaire de transfert (extrait de<sup>14,87,89</sup>).

Ainsi, les sources basées sur le principe du DESI permettent des analyses de surface rapides, sans traitement préalable de l'échantillon, et des analyses de tissus *ex vivo*. Elles ont donc un avantage conséquent dans l'analyse de pièces chirurgicales. Seulement, le manque de mobilité de ces sources, leur configuration (en particulier utilisation d'une haute tension) et l'utilisation de solvants organiques, ne sont pas compatibles avec une utilisation *in vivo* au bloc opératoire. C'est pourquoi, le développement d'un dispositif facilement déplaçable sans impacter le signal moléculaire obtenu est nécessaire.

Le MassSpec Pen a été développé pour répondre à l'analyse *in vivo* au bloc opératoire<sup>90</sup>. Le MassSpec Pen repose sur une microextraction localisée réalisée par une pompe qui apporte 5 à 10  $\mu\text{L}$  d'eau au niveau du dispositif, de la taille d'un stylo, en contact

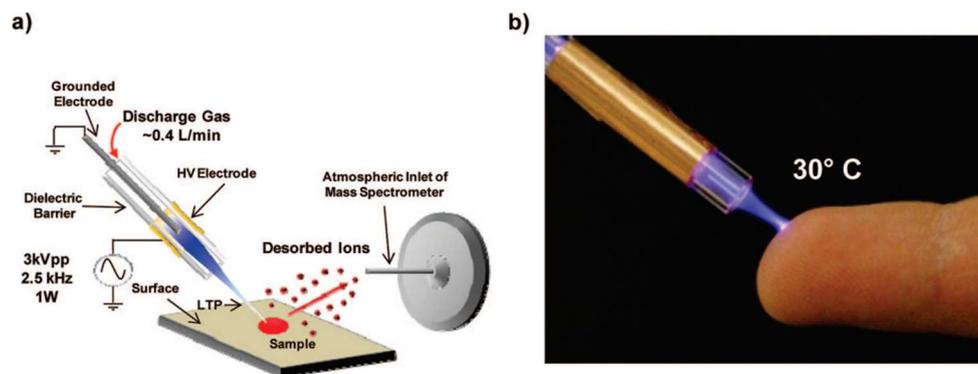
direct avec l'échantillon à analyser. La gouttelette produit est laissée en contact durant 3 secondes avec la surface à étudier, permettant une micro-extraction des espèces solubles mais principalement de lipides et métabolites présents à la surface de l'échantillon (**figure 9A**). La gouttelette d'eau chargée en analytes est ensuite aspirée et transportée par un tube de transfert en PolyTetraFluoroEthylene (PTFE) relié à un capillaire de transfert au spectromètre de masse. Une interface développée et placée à l'entrée du spectromètre de masse permet la création des ions en phase gazeuse via à la fois une hausse rapide de la température, et une baisse drastique de pression. L'avantage majeur de cette innovation technologique est la possibilité de pouvoir échantillonner différents points à la surface d'un échantillon sans invasivité et contrainte de réglages particuliers puisque le système est simplement pacé en contact avec la surface. Ainsi, il est possible de l'utiliser directement sur des échantillons 3D sans aucune préparation d'échantillon. De plus, l'utilisation de l'eau pour l'extraction rend la technique parfaitement biocompatible et utilisable directement sur des patients. Des études ont pu être réalisées sur des pièces chirurgicales *ex vivo*, notamment sur le cancer de l'ovaire<sup>15</sup>, démontrant son grand intérêt dans la classification de cancer séreux de haut grade, avec une très grande sensibilité et spécificité. De plus, la faisabilité au bloc opératoire a été démontrée en couplant le MassSpec Pen au robot chirurgical DaVinci<sup>91</sup> ainsi qu'en étant utilisé par des chirurgiens, *in vivo* et *ex vivo*<sup>92</sup>. Seulement, l'extrémité du système est au contact du tissu. Ce contact pose des problèmes de cross-contamination lors de l'analyse et ne permet pas de balayer une zone pour, par exemple, réaliser une marge. Enfin, le transport de liquides à la place d'aérosol augmente le temps de transfert et le risque de contamination croisée lors de l'analyse de 2 zones consécutives (phénomène de carry-over).

#### Les sources d'ionisation par plasma :

La méthode d'ionisation de ces sources fait intervenir un flux de gaz, tel que l'hélium, l'argon ou l'azote, sous haute température et passant à travers une aiguille à décharge électrique. Ces gaz sont excités afin de former des ions réactifs. Une électrode mise à la terre est utilisée pour sélectionner les espèces positives ou négatives, qui interagissent avec l'air ambiant, conduisant principalement à des agrégats d'eau protonés en mode ion positif et à des radicaux anioniques en mode ion négatif par transfert de charge. Dans le cas d'analytes en phase liquide ou solide, le faisceau de molécules excitées est dirigé vers les échantillons, situés devant l'entrée du spectromètre de masse, pour désorber et ioniser les molécules de la surface où ils sont déposés. Lors de l'analyse de gaz, ce dernier est directement introduit à travers la source pour ionisation. Les techniques de décharge corona ionisent principalement les molécules volatiles et semi-volatiles (VOCs), pour lesquelles elles peuvent constituer une forme robuste d'ionisation. Cette source est commercialisée sous le nom d'analyse direct en temps réel "DART®" par IonSense (IonSense Inc, Saugus, MA) et est utilisée dans des

domaines variés tels que l'identification des agents de guerre et des drogues<sup>93</sup>, l'identification de microorganismes<sup>94</sup> et dans la sécurité alimentaire<sup>95</sup>.

Pour une application *in vivo*, certains ajustements étaient nécessaires pour éliminer le gaz à haute température et le plasma, tous deux toxiques pour le patient. Dans cette optique, le plasma à basse température (LTP)<sup>96</sup> a été développé. Cette source consiste en un plasma froid, d'environ 30°C, créé par deux électrodes permettant l'excitation d'un flux de gaz. Ce plasma percute l'échantillon d'intérêt et induit la désorption/ionisation de celui-ci (illustré en **figure 10**). Le caractère mini-invasif de la technologie est son avantage majeur. Elle n'engendre aucun dommage qui pourrait être causé par une potentielle surchauffe. Seulement, l'énergie apportée à l'échantillon reste faible. De ce fait, seules les molécules de faible poids moléculaire (<1000Da) sont analysables.



**Figure 10:** Schéma du dispositif LTP et son aspect mini-invasif.

(A) Un flux de gaz traverse une électrode mise à la terre et une autre électrode sous haute tension créant ainsi un plasma. Ce plasma percute l'échantillon et permet la désorption des molécules. Ces molécules se dirigent ensuite vers l'entrée de MS. (B) Photo de l'analyse LTP d'un doigt.

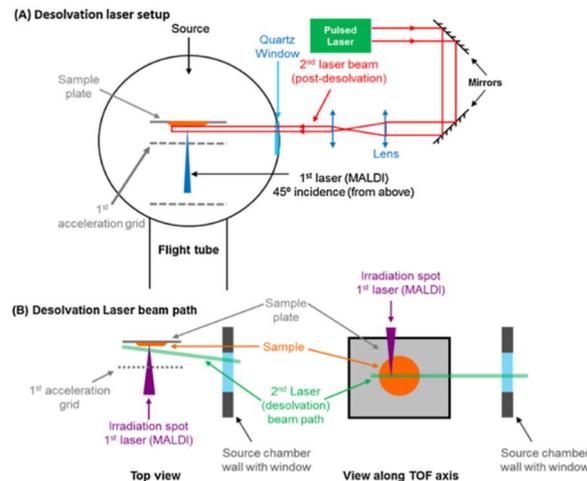
#### Les Sources AIMS basées sur une désorption/Ionisation :

Les sources d'ionisation photonique ont un intérêt majeur : l'utilisation de laser. En effet, contrairement aux méthodes précédentes, les lasers ont pour avantage d'être très résolus permettant des micro-extractions bien plus fines et reproductibles que le spray ou le dépôt d'un solvant.

Pour faire écho à la partie précédente, différentes sources ambiantes développées ces dernières années se basent sur le principe du MALDI. Seulement, deux contraintes majeures liées à cette source, évoquées précédemment, ont dû être surmontées : la nécessité d'utiliser un composé organique comme matrice, et celle de placer l'échantillon dans une chambre à très basse pression avant l'acquisition de données. C'est ainsi qu'est apparu en 2000 la source MALDI à pression atmosphérique (AP-MALDI)<sup>97</sup>. Son principe est similaire au MALDI conventionnel l'exception que la source est à pression atmosphérique et que la plume de

désorption/ionisation s'étendra à PA et nom sous vide poussé. Ainsi, cette méthode répond à une des deux problématiques, mais implique toujours l'utilisation d'une matrice organique, non compatible avec l'analyse *in vivo* chez des patients.

En MALDI, sous vide comme à PA, le rendement de production d'ion reste limité (1 ion pour 10000 molécules environ) même si la technique présente une excellente sensibilité équivalente à celle d'ESI. Ainsi pour augmenter la sensibilité et aider l'ionisation d'espèces dont le rendement d'ionisation est plus faible, des méthodes de post-ionisation peuvent être utilisées. Le principe étant de combiner la désorption/ionisation MALDI ou LDI avec une autre méthode d'ionisation, permettant ainsi d'accroître la détection de certaines molécules. Par exemple, bien qu'elle ne soit pas une source AIMS, le MALDI-2 (développée en 2015) a augmenté la quantité d'ions détectés grâce à l'ajout d'un tir laser, effectué à 266µm, directement sur la plume de matrice créée lors du premier tir laser MALDI<sup>98</sup>. De la même manière, la source de post-désolvation MALDI, créée par l'unité<sup>99</sup> en 2009, permet la désolvation des clusters à l'aide d'un second laser pulsé (Nd:YAG, 1064 nm) qui intercepte la plume MALDI en expansion, et ce, avant l'extraction des ions. Cette technique a permis de doubler le seuil du signal des ions produits.



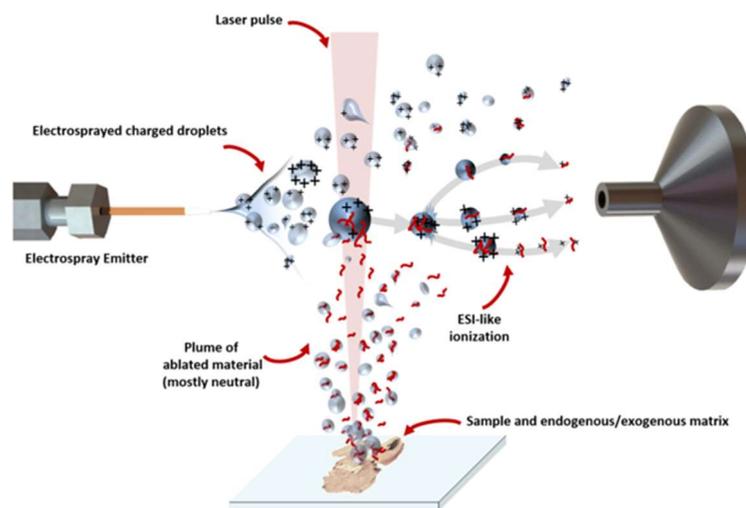
**Figure 11** : Représentation schématique de l'installation réalisée pour la désolvation des clusters en ajoutant un second laser pulsé (Nd:YAG, 1064 nm) (extrait de<sup>99</sup>).

Représentation (A) de l'installation laser et (B) du trajet du faisceau laser à l'intérieur de la chambre de la source d'ions.

Concernant les sources AIMS basées sur une ablation de l'échantillon, l'une des sources les plus connues reste la source LAESI (ablation laser associée à la source ESI<sup>100</sup>). La source LAESI combine une ablation laser avec un principe ESI. Suite à une ablation par un faisceau

laser émettant dans l'infrarouge moyen et favorisant une ablation douce, la plume d'ablation créée, est interceptée par un flux de gouttelette chargée créé par ESI dirigé perpendiculairement à la plume. Les analytes en phase gazeuse sont capturés dans les gouttelette chargés qui entre dans le spectromètre de masse où ces gouttelettes subissent un mécanisme similaire à l'ESI <sup>101</sup> (**Figure 12**)

Le LAESI permet de travailler à partir de coupes de tissus et offre une résolution spatiale jusqu'à 30 µm par suréchantillonnage. Cette technique analyse en grande partie des composés polaires, mais, grâce la post-ionisation ESI, il est possible de faire varier le solvant et détecter des espèces plus apolaires<sup>102</sup>. Il existe aussi une méthode très similaire se nommant le MALDESI reposant sur un principe mixte entre MALDI et ESI, et qui utilise un laser UV associé à une post-ionisation ESI. Malheureusement, contrairement au LAESI, cette source nécessite l'utilisation d'une matrice, comme en MALDI<sup>103</sup>.

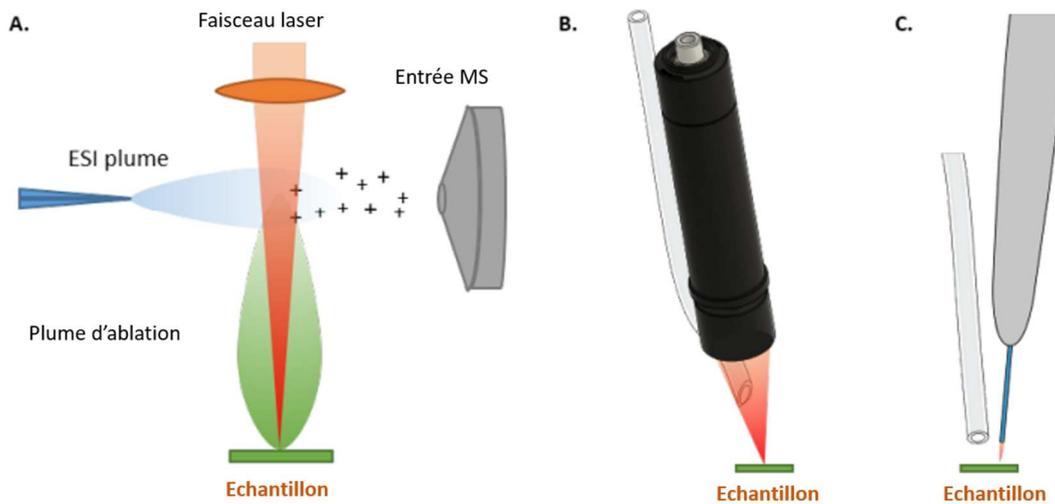


**Figure 12:** Schéma explication du fonctionnement du MALDESI.

*Un tir laser est effectué sur l'échantillon mélangé à la matrice permettant l'ablation de la matière. Un électrospray est ensuite envoyé orthogonalement sur la plume générée et est envoyé au niveau de l'entrée du spectromètre de masse. (CC BY-SA 4.0)*

Depuis 2010, le laboratoire PRISM travaille sur le développement d'une technologie d'analyse *in vivo* basée sur une source laser infrarouge, mini-invasive, initialement appelée ALISE pour Ambient Laser Desorption Source Electrospray (PEPS CNRS) puis devenue SpiderMass. Cette technologie est basée sur le principe du MALDI (Matrix-Assisted Laser Ablation Ionisation). Elle utilise l'eau en tant que matrice endogène, pour induire la désorption/ionisation des molécules. Le faisceau laser utilisé est donc dans le moyen IR à 2,94µm et correspondant à la bande maximum d'absorption des molécules d'eau dans l'IR pour obtenir une excitation

résonnante. Grâce à leur fonctionnement dans le régime de la nanoseconde, les lasers IR favorisent l'analyse *in vivo*. Ainsi, sous l'action d'un laser IR permettant l'excitation des molécules d'eau, un effet WALDI (water-assisted laser desorption/ionization) est généré<sup>16</sup>. Les molécules désorbées/ionisées traversent un tube de transfert de plusieurs mètres de long et sont analysées par spectrométrie de masse. En comparant le profil moléculaire lipidique obtenu avec un modèle de classification pré-établi à partir d'une banque de données de spectres SpiderMass, issue d'une cohorte de patients, il est possible de déterminer si la zone irradiée est cancéreuse ou non. Le type, le sous-type et le grade du cancer peuvent également être précisés, en temps réel, pour faciliter les gestes du chirurgien pendant l'intervention chirurgicale<sup>17</sup>. Il est également possible de rechercher des sites métastatiques et de mieux définir les marges d'exérèse<sup>17</sup>. Le SpiderMass a donc un intérêt clinique majeur. En raison des faibles dommages qu'il engendre, les marges chirurgicales peuvent être minimisées sans abîmer le tissu sain. Ce système a pour avantage de travailler dans les deux modes d'ionisation, permettant ainsi de déterminer le profil global d'un tissu. Le laboratoire PRISM a démontré la fonctionnalité du système dans un contexte clinique réel, à travers une étude au bloc opératoire réalisée sur des patients chiens atteints de sarcome<sup>17</sup>. Aussi, il a été prouvé que le SpiderMass peut discriminer différentes lignées cellulaires entre elles, ainsi qu'une même lignée modifiée génétiquement (Knock-Down) ou non<sup>18</sup>. De plus, cette technologie peut être utilisée, non seulement dans la détection de lipides et métabolites, mais aussi dans la détection de peptides et de protéines<sup>19</sup>. Le SpiderMass permet également l'analyse de tissus fixés et paraffinés (FFPE), sans étape de déparaffinage et après un spray court de glycérol permettant de compenser pour le manque d'eau dans ces échantillons<sup>58</sup>. Les intérêts du SpiderMass sont donc multiples. Il permet d'établir un diagnostic du cancer *in vivo* et en temps réel ainsi qu'une définition précise de la marge de manière mini-invasive. La technologie peut être robotisée pour réaliser des études topographiques<sup>104</sup> facilitant la définition de la marge d'exérèse. À l'heure actuelle, plusieurs développements de la technologie sont en cours en vue d'améliorer son utilisation et le nombre de ses applications. Par exemple, des recherches sont menées pour associer le SpiderMass avec des robots souples afin de réaliser des études d'endoscopie. L'inclusion récente de l'intelligence artificielle (CNN) dans son utilisation a permis de mettre en évidence la possibilité de réduire le nombre de patients dans les cohortes sans réduire la précision et reproductibilité des résultats de classification<sup>105</sup>. Enfin, le SpiderMass est développé en mode imagerie<sup>104</sup> pour réaliser de l'imagerie *in vivo* et intégrera de la réalité augmentée pour permettre une meilleure visibilité pour le chirurgien.



**Figure 13:** Schéma récapitulatif des principales ionisations laser.

(A) Le LAESI consiste à tirer sur l'échantillon puis à envoyer orthogonalement sur la plume générée des microgouttelettes de solvants chargées (ESI) afin d'envoyer les biomolécules post-ionisées vers l'entrée de MS. (B) La technologie SpiderMass consiste à tirer avec un laser à 2940nm directement sur l'échantillon avec un burst de plusieurs tirs à 4mJ d'énergie. La plume générée est aspirée par un tube de transfert en PTFE. Les agglomérats de biomolécule restant sont dissociés par une hélice à haute température présente devant l'entrée de MS. (C) La source PIRL utilise un fonctionnement similaire à la source SpiderMass mais utilise un laser picosecondes (300 picosecondes) qui permet d'éviter le processus de haute température durant l'ablation.

Plus récemment, le laser Infrarouge Picoseconde (PIRL) a fait son apparition. Il présente une utilisation similaire au SpiderMass, il utilise un laser émettant à la même longueur d'onde mais de durée d'impulsion plus courte 0,3 nano seconde (300 picosecondes). L'intérêt d'utiliser des laser dont la durée d'impulsion est plus courte réside dans la limitation des effets thermiques et donc la diminution des dommages aux tissus. C'est pour cela qu'il a, dans un premier temps, servi de moyen d'ablation de matrices biologiques<sup>106</sup>, puis, dans un second temps, comme une méthode ultra-rapide pour atomiser les échantillons et extraire les protéines par Excitation Vibratoire Impulsive de Désorption (PIRL-DIVE)<sup>107</sup>. Il a ensuite été directement couplé à la spectrométrie de masse par le biais d'une fibre optique dans une sonde portable, un tube de transfert et une entrée de spectromètre de masse chauffée<sup>108</sup>. Des développements ont été réalisés sur le couplage du PIRL-MS avec des informations polarimétriques<sup>109</sup>, la tomographie par cohérence optique (OCT)<sup>110</sup> et DESI-MSI<sup>111</sup>, ainsi que l'affichage des informations de classification du MS sur un flux vidéo de l'échantillon analysé en réalité augmentée<sup>112</sup>.

Cependant, ces laser restent encombrants et onéreux. Par ailleurs, à 300 ps les dommages aux tissus ne sont pas drastiquement réduits par rapport au SpiderMass. Les lasers picosecondes posent aussi des défis en ce qui concerne la fibre optique en raison de l'énergie crête au niveau de l'entrée de la fibre qui pourrait l'endommager. De plus, il est connu que la durée de l'impulsion laser influence la pénétration du laser, le seuil d'ablation, ainsi que la

propagation thermique dans les tissus. En effet, l'irradiation "plus froide" du PIRL-MS permet de réduire les dommages thermiques sur les tissus, mais présente également l'inconvénient de produire moins d'ions<sup>109</sup>. Lors d'une ablation laser nanoseconde, le plasma est créé pendant la durée de l'impulsion, et l'impulsion elle-même sert à réchauffer le plasma, réduisant ainsi la taille des particules ablatées<sup>109</sup>. Cependant, le système PIRL utilisé jusqu'à présent, dispose d'un laser de quelques centaines de picosecondes, ce qui est seulement 10 fois plus court que le laser picoseconde SpiderMass (**Figure 13**).

La source d'ionisation par évaporation rapide (REIMS) :

En 2009, une méthode de diagnostic rapide, sensible, précise et *in vivo* a été introduite : le bistouri intelligent, ou iKnife. Ce scalpel a pour but d'aider le chirurgien à détecter et à retirer les tissus cancéreux<sup>11</sup>. Il a notamment été utilisé pour la recherche de nouveaux biomarqueurs lipidiques et de nouvelles méthodes de diagnostic sur le cancer de l'ovaire<sup>113 13</sup>. Cette technologie est basée sur la spectrométrie de masse ambiante, dont le fonctionnement est illustré **figure 14**<sup>10</sup>. Le iKnife repose sur le couplage entre un bistouri électrique qui génère un aérosol lors de la découpe des tissus et un spectromètre de masse permettant l'obtention du profil moléculaire du tissu. Ce processus d'ionisation a été, dans un premier temps, nommé REIMS (rapid evaporative ionization mass spectrometry). Ce nom a, par la suite, été utilisé pour nommer l'interface utilisée dans cette méthode. La technologie permet d'établir un profil lipidique ou métabolique d'un tissu afin de déterminer son caractère sain ou tumoral. Cependant, cette technologie reste très invasive. En effet, les tissus sont analysés pendant que le bistouri coupe les tissus. Les tissus étant détruits, il est impossible pour le pathologiste d'effectuer un second diagnostic pour confirmer les résultats obtenus. De plus, l'iKnife manque de précision, il est impossible de réaliser des micro-analyses sur des zones très précises. Enfin, le chirurgien collecte l'information au cours de l'excision, ce qui ne lui permet pas de prendre la décision nécessaire afin d'adapter son geste avant l'exérèse.

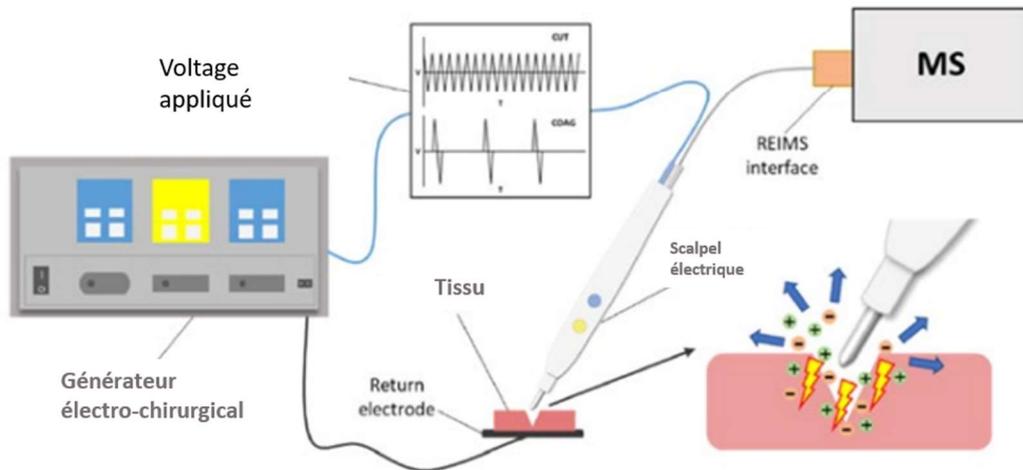


Figure 14: Schéma représentant la technologie iKnife.

*Le chirurgien va reséquer le tissu à l'aide d'un bistouri électrique venant brûler celui-ci. Les fumées générées sont aspirées par un tube de transfert en PTFE dirigé vers le spectromètre de masse (CC BY-SA 4.0).*

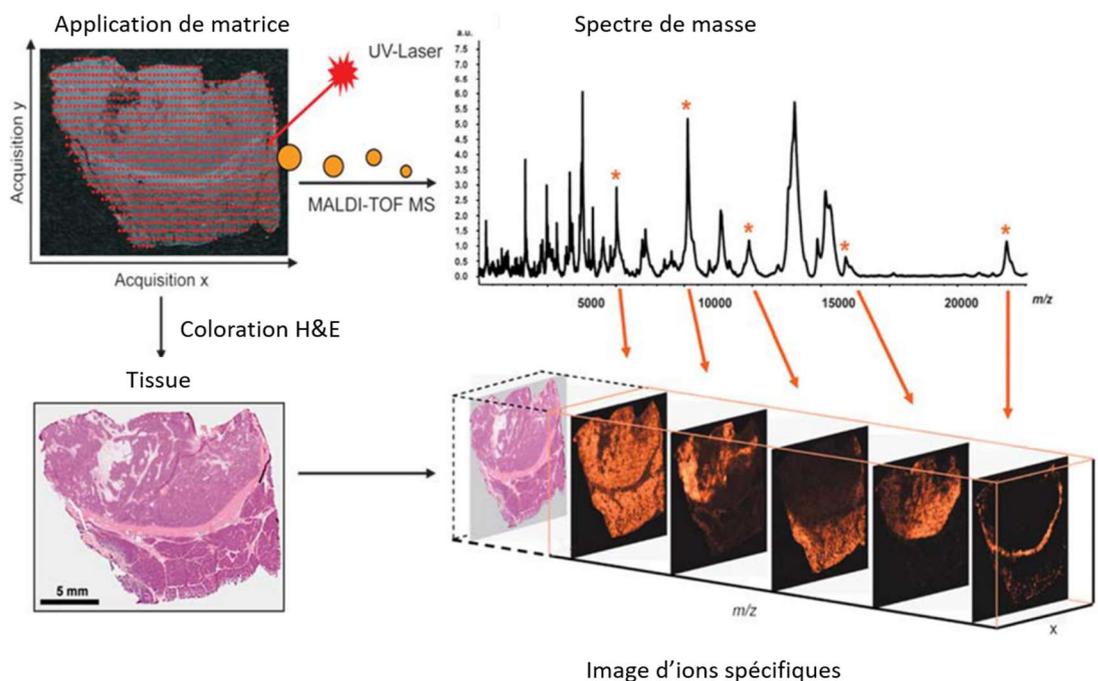
De nos jours, les sources AIMS ont été implémentées dans de nombreux domaines scientifiques. Elles ont apporté de nouvelles méthodes de diagnostic en temps réel et de dépistage de médicaments <sup>114</sup>. Ainsi, ces sources permettent de mieux comprendre les mécanismes moléculaires mis en jeu dans les différentes pathologies telles que le cancer. Beaucoup d'études ont amélioré la détection des cancers de l'ovaire en y apportant de nouveaux marqueurs discriminants.

### L'utilisation des sources ambiantes dans l'imagerie par spectrométrie de masse

L'utilisation de la spectrométrie de masse a apporté une donnée moléculaire supplémentaire à la compréhension des pathologies et notamment du cancer. En effet, l'imagerie par spectrométrie de masse (MSI) a permis de cartographier moléculairement les tissus par une analyse sans marquage<sup>115</sup>. Elle offre ainsi une visualisation de la distribution spatiale moléculaire du tissu analysé. En une seule acquisition, des milliers de molécules peuvent être détectées et leur distribution obtenue. Il est possible d'observer différents types de biomolécules tels que les lipides, les métabolites, les glucides et les protéines <sup>8</sup>. En MSI, chaque pixel scanné correspond à un spectre de masse et, à l'aide d'un logiciel approprié, une image 2D ou 3D du tissu peut être créée en utilisant les données moléculaires de celui-ci. Contrairement aux techniques conventionnelles telles que l'immunohistochimie, l'autoradiographie ou les techniques de fluorescences, la MSI a l'avantage d'offrir une analyse sans marquage, une bonne résolution spatiale (de l'ordre de 1µm à 50µm) ainsi qu'une grande sensibilité et variété de détection (des éléments atomiques jusqu'aux protéines de quelques dizaine de kDa (<30kDa)) en fonction de la source d'ionisation utilisée.

La technique d'imagerie la plus souvent utilisée est basée sur la source MALDI. Le principe est similaire à celui décrit précédemment. Une matrice spécifique au type de molécules

d'intérêt est déposée sur l'échantillon, le plus souvent une section de tissu. Durant l'acquisition en MALDI-MSI, plusieurs centaines de tirs lasers par pixel, sont réalisés et un spectre moyen y est associé. En choisissant un ion spécifique, il est possible de définir son intensité relative pour chaque pixel et de reporter l'intensité sur une échelle de couleur. De ce fait, une cartographie totale de la distribution et de l'abondance de cet ion au sein de l'échantillon est créée (**figure 15**). Par ailleurs, des méthodes d'analyses statistiques ont été développées pour permettre le traitement des données d'imagerie. En particulier, une segmentation peut être réalisée à partir des données d'imagerie pour mettre en évidence les différents phénotypes cellulaires présents dans les tissus. La segmentation, permet, en effet, de regrouper chaque pixel comprenant un spectre commun, représentant ainsi l'organisation moléculaire hétérogène du tissu. Le MALDI-MSI a un réel intérêt clinique par sa complémentarité avec les informations apportées par le pathologiste. Cette technique a notamment permis la découverte de nouveaux marqueurs protéiques et lipidiques dans le cancer de l'ovaire tels que la protéine reg-alpha<sup>116</sup> et l'accumulation de sulfatides dans les cellules cancéreuses<sup>117</sup>. Cette technique présente cependant des limites. En effet, une étape de préparation de l'échantillon est nécessaire, incluant notamment l'application de matrice. Cette étape peut provoquer des changements au sein des mécanismes biologiques et donc biaiser l'analyse. De plus, l'analyse est réalisée sous vide poussée, s'éloignant donc du contexte biologique.



**Figure 15:** Fonctionnement de l'imagerie MALDI (extrait et traduit de <sup>118</sup>).

La matrice est appliquée sur une coupe de tissu selon l'analyte à identifier. Un tir laser UV est réalisé sur chaque pixel du tissu (correspondant à 10-50 $\mu$ m de diamètre). Un ion est ensuite sélectionné sur un spectre de masse et sera associé à une couleur plus ou moins intense en fonction de son intensité relative sur chaque pixel de l'image.

À l'heure actuelle, la source MALDI est la source la plus utilisée en imagerie. Seulement, l'imagerie par sources ambiantes a permis de réaliser des cartographies moléculaires dans des conditions plus physiologiques. L'IR-MALDESI, introduite en 2013, utilise l'eau en tant que matrice et un spray de solvant en méthode de post-ionisation. Cette technique a permis la cartographie de métabolites et de lipides sans préparation d'échantillons et avec une très bonne résolution spatiale<sup>119</sup>.

En parallèle, d'autres sources basées sur l'ablation laser se sont développées en imagerie, notamment la source LAESI, développée par l'équipe de A. Vertes, en 2007<sup>120</sup>. Un laser émettant à 2.94µm sert principalement à l'ablation et une post-ionisation est réalisée par les microgouttelettes ESI. Elle a rapidement été utilisée en mode imagerie et a permis d'atteindre des performances similaires à la source MALDESI introduite en 2006. De manière générale, cette méthode d'imagerie est utilisée avec une résolution spatiale de 250µm. Celle-ci a pu être diminuée à 20µm grâce à l'ajout de fibres optiques ou encore aux techniques de sur-échantillonnage. Ces deux techniques présentent des ions similaires à la source ESI, qui peuvent être multichargés. Un des problèmes majeurs de cette source réside dans l'utilisation d'une longueur d'onde qui excite les molécules d'eau, provoquant ainsi la déshydratation du tissu et donc une perte de l'intensité du signal. Pour pallier à cela, les tissus sont congelés et maintenus à basse température, garantissant une hydratation suffisante à l'analyse. Des progrès ont été réalisés sur cette source et ont permis la réalisation d'imagerie 3D des tissus, notamment au niveau de la surface des échantillons<sup>121</sup>.

Le laboratoire PRISM s'est également intéressé à l'imagerie MS par le SpiderMass. Toujours dans l'objectif d'une application clinique, l'imagerie SpiderMass repose sur une nouvelle méthode d'imagerie topographique moléculaire qui utilise un bras robotisé. Ce bras peut se déplacer selon 6 axes et permet le déplacement de la sonde laser, de façon contrôlée, au-dessus du patient. Cette technologie permet un scan intégral de la surface d'un tissu et donne accès à un diagnostic sensible, précis (de l'ordre de 500 µm) et complémentaire aux techniques d'imagerie conventionnelles utilisées au bloc<sup>20</sup>. Cette résolution peut être réduite à 250µm par sur-échantillonnage et permet un diagnostic très précis pour le chirurgien. Cette méthode présente, contrairement aux techniques précédentes, des spectres similaires à ceux obtenus au MALDI conventionnel. Cette similarité permet de coupler la réponse rapide du SpiderMass à la cartographie précise du MALDI (Publication en cours). De plus, un capteur laser a été placé au niveau du bras robotique pour trianguler la hauteur entre l'échantillon et la pièce à main qui contient la fibre optique. Ainsi, il est possible d'obtenir une image 3D de l'échantillon et de définir précisément un diagnostic à la surface de l'échantillon de manière mini-invasive et en temps réel<sup>104</sup>.

D'autres sources utilisent la désorption et/ou l'ionisation à travers l'application de solvant. C'est le cas de la source DESI détaillée précédemment. Cette source, développée dans les années 2004<sup>77</sup>, est principalement utilisée en imagerie. Elle permet l'analyse de lipides et métabolites de surface grâce à un spray de microgouttelettes de solvant chargées sur l'échantillon. Récemment, elle a permis l'identification de nouveaux biomarqueurs du cancer ovarien<sup>122</sup>. Malheureusement, cette technique n'est pas facile à utiliser pour un utilisateur non expérimenté. En effet, les réglages des différentes distances entre la buse, l'échantillon et le tube de collecte se font manuellement. De plus, le spray se règle lui aussi manuellement en contrôlant l'insertion de l'aiguille de spray dans la buse. Ainsi, le spray n'est pas toujours stable et nécessite d'être adapté à chaque analyse. La résolution spatiale de cette technique peut atteindre les 20µm mais sera variable en fonction de l'utilisateur et sera plus souvent de l'ordre de 50µm.

La source SIMS (Secondary Ion Mass spectrometry) est très utilisée en imagerie par spectrométrie de masse. Elle se résume à bombarder la surface d'un échantillon à l'aide d'un faisceau d'ions qui peuvent être par exemple des gaz nobles comme du  $^{40}\text{Ar}^+$  ou du  $^{16}\text{O}^-$ , ou des molécules ionisées comme du  $\text{SF}_5^+$ . Ce bombardement de particules va générer des ions secondaires à la surface du tissu qui seront analysés par l'analyseur de masse (souvent un time of flight). Cette source a pour avantage de ne pas nécessiter de dépôt de matrice au préalable et possède actuellement la meilleure résolution spatiale pour l'imagerie (200nm). Seulement, cette source a une gamme de masse plus limitée que le MALDI et une sensibilité moindre<sup>123</sup>.

La source APCI a également vu le jour au début des années 2000<sup>124</sup>. Elle consiste en un flux d'azote qui passe le long d'une électrode, créant une décharge corona qui charge le gaz. Le gaz chargé entre en interaction avec l'échantillon et permet son ionisation et sa désorption en surface. Cette technique a pour principal avantage de permettre l'analyse d'une grande variété de biomolécules : des espèces volatiles et non volatiles ainsi que des molécules de faible masse et de faible polarité. L'imagerie MS a été conditionnée par la source DAPCI qui s'affranchit de l'utilisation de gaz<sup>125</sup>. Pour ce faire, la source laisse passer l'air ambiant pour créer la décharge corona. Seulement, l'électrode est cette fois-ci en acier inoxydable et insérée dans un matériau isolant. Cette source a notamment été utilisée pour l'analyse d'encre pour l'authentification de document et la distribution de mélanine sur des coupes d'œufs cuit à l'échelle du millimètre.

Ainsi, l'imagerie par spectrométrie de masse a permis la cartographie de molécules de surface. Elle a pu être utilisée dans le cadre d'études pharmacocinétiques, de détection de biomarqueurs tumoraux au sein de tissus ou encore dans l'analyse de plantes. Elle permet

d'appréhender la localisation des molécules et ainsi comprendre certains mécanismes biologiques. Son utilisation peut-être un atout majeur dans la lutte contre le cancer, non seulement dans le diagnostic, mais également dans la compréhension du micro-environnement tumorale qui est d'une importance primordiale dans l'émission du diagnostic mais aussi dans le diagnostic et le développement thérapeutique. La définition de marqueurs spécifiques de chaque sous-type cellulaire permettrait une cartographie précise des tissus ainsi qu'une compréhension globale du micro-environnement tumorale (TME) et de son rôle dans le pronostic.

### Importance du micro-environnement tumoral dans le pronostic du cancer de l'ovaire

Une tumeur est loin d'être un simple amas de cellules cancéreuses. Ces dernières sont intégrées dans un environnement très hétérogène composé de différents types cellulaires, de molécules variées et d'une matrice extracellulaire. Lors de la cancérogénèse, les cellules cancéreuses induisent des changements moléculaires, cellulaires et physique au sein du tissu pour promouvoir la croissance tumorale. Ces changements induisent la création d'un microenvironnement tumoral (TME) complexe et en constante évolution. La composition de cet environnement varie selon le type de cancer, mais plusieurs caractéristiques sont communes à l'ensemble des cancers tels que l'infiltration de cellules immunitaires, la présence de cellules stromales, d'une matrice extracellulaire et d'une importante vascularisation<sup>126</sup>.

Le microenvironnement tumoral est d'une importance capitale dans le cancer de l'ovaire puisque les cellules qui le composent vont grandement participer à sa progression, sa croissance et sa migration<sup>127</sup>. D'autre part, dans la mesure où les cancers ovariens sont diagnostiqués à des stades tardifs, ce TME sera bien différencié et l'impact des cellules immunitaire le composant, qu'il soit favorable ou défavorable, sera majeur dans le pronostic du patient. Les récentes technologies « omics » ont démontré l'importance du TME dans le développement de nouveaux outils de diagnostic, de traitement ainsi que de pronostic<sup>128</sup>. C'est dans cette optique que l'intérêt des cellules immunitaires est grandissant. En effet, celles-ci ont un impact direct, positif ou négatif, sur le développement tumoral<sup>129</sup>. Elles pourraient représenter à elles seules un outil de pronostic efficace. Ainsi, en étant capable de discriminer les différentes populations cellulaires et leur proportion au sein du TME de la patiente, il sera possible de déterminer la survie relative de celle-ci est d'ajuster le traitement à lui administrer.

### *La composante non cellulaire*

Les différentes populations cellulaires présentent au sein du TME ne sont pas les seules à influencer sur la progression tumorale. En effet, il y a une composante non cellulaire essentielle aux échanges cellulaires qui fournit les molécules indispensables au développement tumoral.

Dans un premier temps, la **matrice extracellulaire** du TME a des propriétés physiques indispensables pour le développement initial de la tumeur. En effet, celle-ci permet un maintien des cellules cancéreuses et donc leur ancrage, favorisant ainsi sa croissance. De plus, elle fournit des facteurs de croissance tels que le VEGF, FGF ou encore le TGF- $\beta$ , sécrétés par les différentes cellules du TME et indispensables à la croissance tumorale<sup>126</sup>. Majoritairement composée de collagène, fibronectine et de laminine, la composition de cette matrice évolue au cours du développement tumoral et facilite également la migration des cellules cancéreuses, responsables de métastases<sup>126</sup>. Dans le cancer de l’ovaire et les lésions STICs, le facteur de croissance transformant bêta induit (TGFBI) est sur-exprimé par rapport à un ovaire sain. Cette protéine, impliquée dans l’adhésion cellulaire, est sécrétée par les macrophages. Elle favorise non seulement la croissance tumorale, mais aussi la dissémination de cellules cancéreuses depuis son site initiale (la fimbria) vers l’ovaire<sup>130</sup>(**Figure 16**).

Les **exosomes** sont un autre exemple de moyen de communication essentiel au TME. Ce sont des microvésicules contenant un matériel nécessaire aux échanges cellulaires. Ils peuvent contenir des molécules très variées telles que de l’ADN, des lipides ou encore des protéines<sup>131</sup>. Au sein du TME, les exosomes jouent un rôle critique dans les échanges cellulaires entre cellules cancéreuses et cellules stromales. Ils sont impliqués dans de nombreux processus du développement tumoral tels que la croissance tumorale, l’angiogenèse ou encore l’invasion.

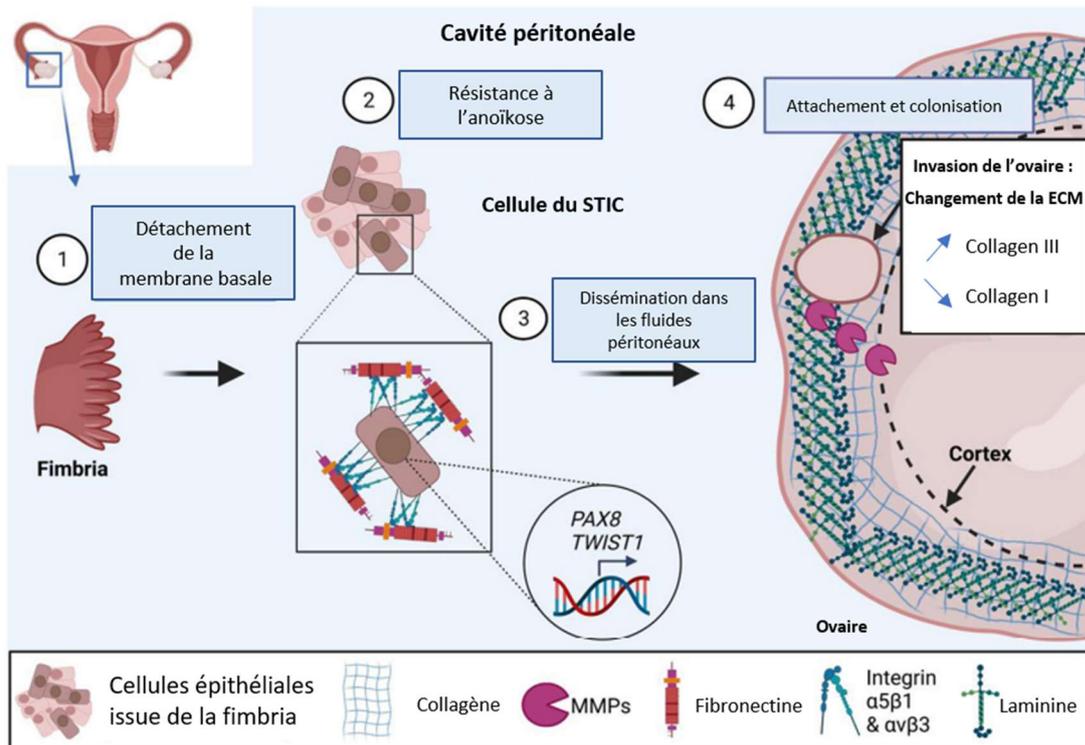


Figure 16: Rôle de la matrice extracellulaire dans le détachement du STIC vers l'ovaire.

### Les cellules stromales

Le stroma fait partie intégrante du microenvironnement tumoral. Les cellules cancéreuses recrutent des cellules de soutien, telles que des cellules endothéliales et des fibroblastes, dans le stroma des tissus endogènes voisins pour favoriser les étapes critiques de la formation de la tumeur et, à ce titre, elles constituent un élément important du microenvironnement tumoral. Une fois recrutées dans le microenvironnement tumoral, les cellules stromales sécrètent de nombreux facteurs qui influencent l'angiogenèse, la prolifération, l'invasion et les métastases.

Les cellules endothéliales se regroupent sous forme de monocouche pour favoriser l'angiogenèse, nécessaire à l'apport d'oxygène et d'autres nutriments essentiels, au maintien de l'homéostasie moléculaire et au recrutement de cellules immunitaires.

Les fibroblastes associés à la tumeur (CAFs) sont souvent dérivés des fibroblastes résidents du tissu, présents avant la cancérogenèse, mais peuvent également être issue des cellules endothéliales, des péricytes ou d'adipocytes environnants. En effet, les cellules stromales, ainsi que les cellules tumorales, libèrent des facteurs de croissance, tels que le TGF- $\beta$ , le PGDF ou encore du FGF2 (fibroblast growth facteur 2), en grande quantité pour différencier les fibroblastes résidents en CAFs. Dans un environnement non tumoral, les fibroblastes sont

impliqués dans la protection des différents organes et dans la cicatrisation, à travers la sécrétion de protéines de la matrice extracellulaire, telles que le collagène ou encore de TGF- $\beta$ , favorisant la prolifération cellulaire. Les cellules tumorales détournent ce processus à leur avantage, on appelle d'ailleurs la tumeur « la blessure qui ne guérit jamais ». Les fibroblastes sont également impliqués dans l'immunosuppression, à travers la sécrétion de cytokines anti-inflammatoires, dans la transition épithéliale-mésenchymateuse, permettant la création de métastases, et dans l'angiogenèse. Ces cellules ont donc un impact direct sur la progression et la croissance tumorale. Certaines études démontrent également l'importance du nombre de ces cellules dans la survie du patient. Par exemple, dans le cancer colorectal, la survie du patient serait liée à la quantité de CAFs présent au sein du TME <sup>132</sup>.

### *Les cellules immunitaires*

Le développement tumoral va de pair avec l'activation du système immunitaire. Celui-ci a pour rôle l'élimination de la tumeur. Des cellules immunitaires effectrices, telles que les lymphocyte T CD4+, les lymphocytes T CD8+, les lymphocytes natural killer (NK) ou encore les macrophages, vont donc être recrutées au niveau du foyer tumoral. Seulement, les cellules cancéreuses ont développé des stratégies efficaces pour échapper à ce système immunitaire et le détourner pour faciliter son développement. En effet, la progression tumorale s'accompagne souvent d'un microenvironnement immunosuppresseur comprenant des cellules T régulatrices (Tregs) et des macrophages de type 2 (M2). Cet environnement réduit considérablement l'activité anti-tumorale du système immunitaire et favorise le développement tumoral.

### *Les Macrophages*

#### *Origine*

Les macrophages dérivent de deux sources différentes. Ils sont, en majorité, produits dans le sac vitellin, avant la naissance, à partir de précurseurs érythro-myéloïdes. Ils s'établissent ensuite dans les différents tissus de l'organisme et sont maintenus pendant la vie adulte<sup>133</sup>. Après la naissance, les macrophages dérivent de cellules souches hématopoïétiques localisées dans la moelle osseuse. Ces cellules se différencient en progéniteurs myéloïdes, en monoblastes et en pro-monocytes avant de devenir des monocytes dans la circulation sanguine. Ces derniers migrent ensuite vers d'autres tissus où, sous l'influence d'interleukines et de facteurs de croissance tels que le CSF-1, ils se différencient en macrophages. Selon leur localisation, les macrophages portent un nom différent. Ils sont nommés cellules de Kupffer dans le foie, ostéoclastes dans les os, macrophages alvéolaires dans les poumons ou encore cellules microgliales dans le système nerveux central <sup>134</sup>. Ces macrophages sont présents de manière endogène en grande quantité dans l'ovaire sain. Ces cellules jouent un rôle majeur dans le maintien de l'homéostasie de

l'ovaire et dans la sensibilité hormonale. De plus, ils sont impliqués dans le cycle ovulatoire puisqu'ils sont plus abondants en période pré-ovulatoire<sup>135</sup>.

#### Fonctions

Les macrophages sont essentiels à la réponse immunitaire innée comme dans la réponse immunitaire adaptative. Leur rôle principal réside dans la phagocytose des débris cellulaires ainsi que des agents pathogènes (bactéries, virus). Pour exercer leurs fonctions immunologiques, les macrophages possèdent des récepteurs membranaires capables de reconnaître des motifs moléculaires caractéristiques des pathogènes, appelés PAMPs (Pathogen Associated Molecular Patterns). Parmi ces motifs, nous retrouvons les lipopolysaccharides (LPS), composants de la membrane des bactéries gram négatives, les ARNs viraux doubles brins, ainsi que les motifs CpG hypométhylés, caractéristique de l'ADN bactérien. Les macrophages sont également capables de reconnaître des cellules en souffrance, grâce à des récepteurs de motifs moléculaires endogènes appelés DAMPs (Damage Associated Molecular Patterns). Lors de la reconnaissance de ces signaux, des cascades de signalisation pro-inflammatoires, telles que la voie NF  $\kappa$ B, sont activées, induisant la sécrétion de cytokines<sup>136,137</sup>.

Enfin, les macrophages jouent un rôle essentiel dans la réponse immunitaire en coordonnant divers processus. Ils interviennent dans la présentation d'antigènes, permettant ainsi la communication entre l'immunité innée et l'immunité adaptative, et alertent le système immunitaire de la présence de pathogènes.<sup>138</sup>

En plus de leur rôle dans la réponse immunitaire, les macrophages sont impliqués dans la régulation de la réponse inflammatoire. Ils modulent la sécrétion de chimiokines et de cytokines pro- et anti-inflammatoires en fonction des besoins de l'organisme. De plus, les macrophages sécrètent divers facteurs de croissance qui participent à des processus tels que l'angiogenèse, le maintien de la matrice extracellulaire et la cicatrisation (**figure 17**)<sup>133</sup>.

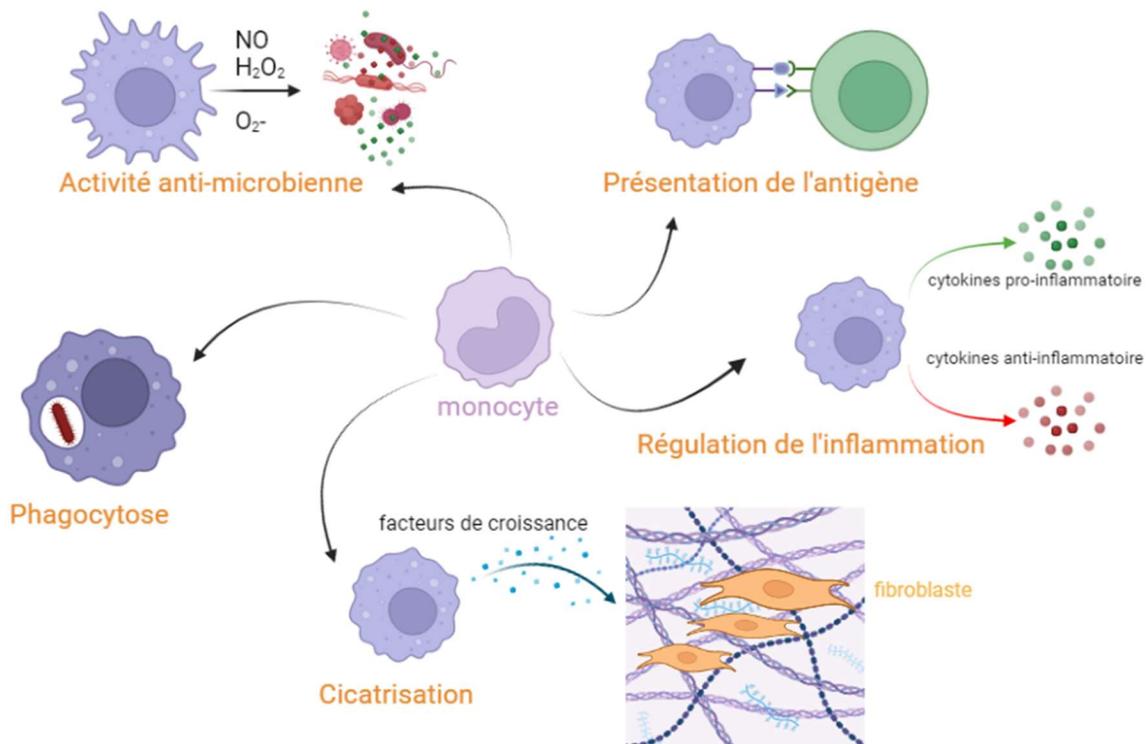


Figure 17: Principales fonctions des macrophages dans l'organisme.

Les macrophages sont impliqués dans la phagocytose, l'activation de la voie anti-microbienne ou encore dans la présentation d'antigènes à d'autres cellules immunitaires effectrices (lymphocytes T, B etc). Ils sécrètent différentes cytokines, pro et anti-inflammatoires, en fonction de leur phénotype et des besoins de l'organisme. Ils jouent également un rôle dans la cicatrisation à travers la sécrétion de facteurs de croissance stimulant l'activité des fibroblastes.

### Les différents phénotypes

Les macrophages représentent une population hétérogène de cellules capables de répondre à différents stimuli. En fonction de leur environnement, ceux-ci peuvent présenter de nombreux phénotypes différents et ainsi varier leur profil de sécrétion, leur fonction et l'expression de protéines de surface. De ce fait, il existe un très large panel de macrophages, décrits selon leurs marqueurs de surfaces et leurs activités biologiques spécifiques. Il est possible de classer l'ensemble de ces phénotypes en deux groupes majeurs : les macrophages de type 1 (M1), considérés comme pro-inflammatoires, et les macrophages de type 2 (M2), considérés comme anti-inflammatoires. Les macrophages se différencient en phénotype M1-like lors de stimulation par des cytokines pro-inflammatoires, telles que l'interféron- $\gamma$  (IFN- $\gamma$ ), le TNF- $\alpha$  ou encore par la stimulation de leur récepteur TLR4 à travers la détection de LPS, par exemple. Ce phénotype se distingue par l'expression de marqueurs tels que iNOS, CD40, CD74 ou encore du CMH II. Ces macrophages sont impliqués, entre autres, dans la réponse anti-tumorale et la sécrétion de cytokines pro-inflammatoires, telles que le TNF-alpha. Quant à lui, le phénotype M2-like, est induit suite à la stimulation par des

cytokines anti-inflammatoires, telles que l'IL4 ou l'IL10. Ils sont identifiables grâce à l'expression des marqueurs arg1 (arginase I), CD206 ainsi que la phosphorylation de la protéine STAT3. Ils sécrètent des chimiokines tels que CCL17 et CCL 22 et des cytokines anti-inflammatoires tels que l'IL10 et le TGF- $\beta$ . Ces macrophages sont impliqués dans la modulation de la réponse inflammatoire, dans les processus de cicatrisation ainsi que dans l'angiogenèse<sup>138</sup>.

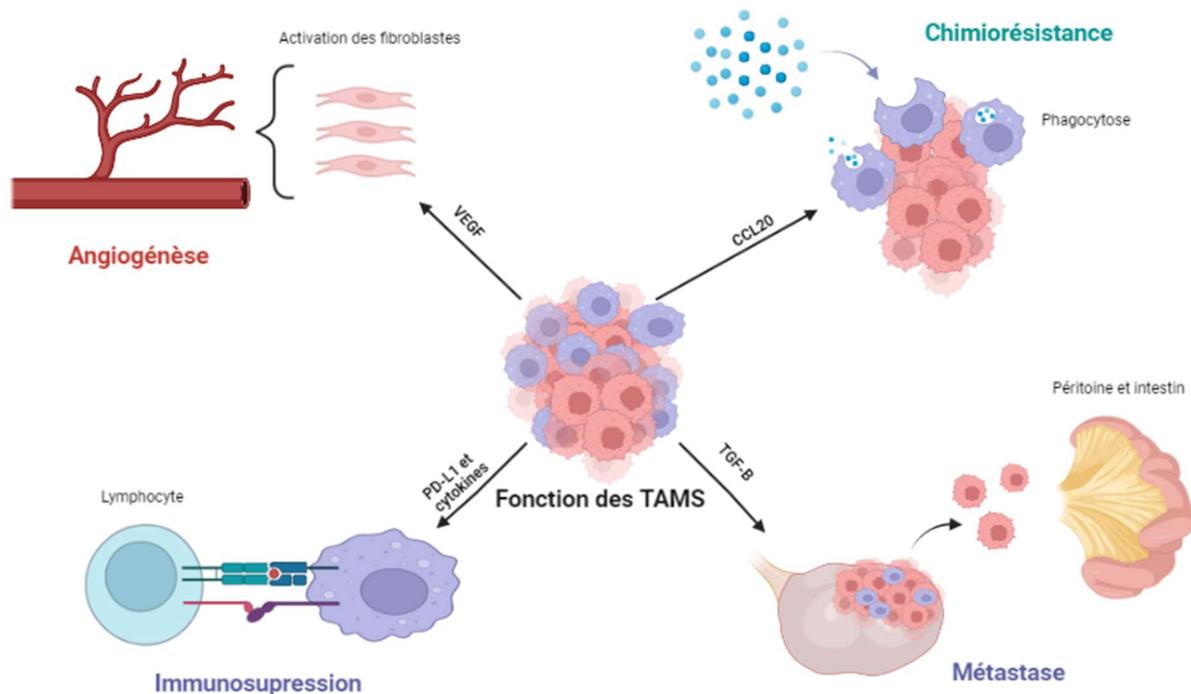
Toutefois, les variations d'expression entre ces deux phénotypes M1 et M2, sont difficiles à discerner. Un macrophage de type M2-like peut, par exemple, exprimer un marqueur caractéristique du phénotype M1-like. De plus, il est important de noter qu'il existe un grand nombre de macrophages, avec des phénotypes qui se chevauchent entre eux. Ainsi, la classification M1-like et M2 like est très simplifiée et, en oncologie, il est plus correct de décrire les phénotypes comme pro- et anti-tumoral.

#### Les macrophages et la tumeur

En condition physiologique, les macrophages sont capables de reconnaître les antigènes tumoraux et d'induire une réponse immunitaire anti-tumorale pour éliminer l'ensemble des cellules cancéreuses. Seulement, malgré la très grande infiltration de macrophages au sein du site tumoral, ces cellules sont incapables d'assurer leur fonction anti-tumorale et, au contraire, ils favorisent la croissance tumorale<sup>139</sup>. Ces derniers ont donc une importance majeure dans le microenvironnement tumoral et sont appelés TAMs (tumor associated macrophages). Les TAMs sont les cellules immunitaires les plus représentées du TME, à la fois en termes de nombre, mais également de volume total dans la tumeur<sup>140</sup>. Dans les cancers de l'ovaire, ils peuvent représenter jusqu'à 39% des cellules immunitaires du TME. De manière générale, 51% de cette population est représentée par des M2, et 20% correspondent à des M1. Plus de 20% de ces TAMs sont indifférenciés, phénotype appelé M0, et 4% sont des monocytes<sup>141</sup>. Ces macrophages peuvent être originaires, à la fois de la différenciation de macrophages résidants de l'ovaire, mais également d'un recrutement par la tumeur elle-même, notamment grâce à la sécrétion de la protéine chimio attractive de monocytes (MCP-1 ou CCL2). Les cytokines et chimiokines ne sont pas les seules molécules utilisées pour recruter et orienter le phénotype des macrophages. En effet, les phosphatidylsérines, présentent de manière anormale au niveau des cellules apoptotiques dans le TME et de la membrane externe des cellules tumorales, semblent également participer activement à l'orientation des macrophages vers un phénotype M2<sup>142,143</sup>. Ainsi, les lipides pourraient devenir une nouvelle cible pour l'étude des mécanismes sous-jacents à l'orientation des macrophages et pour la discrimination de ces phénotypes.

Au sein du TME, les macrophages de phénotype M1 jouent un rôle anti-tumoral et ont pour objectif d'éliminer la tumeur à travers l'action de cytokines pro-inflammatoires, telles que

l'IL-12 ou du TNF alpha, activant la voie NF-KB. Cependant, cette voie semble également être responsable de l'accélération du processus de métastase tumorale<sup>140</sup>. Au contraire, les fonctions du phénotype M2 sont quant à elles multiples et orientées vers une action pro-tumoral. Ils participent à l'angiogénèse à travers la sécrétion du VEGF qui stimule les fibroblastes associés à la tumeur induisant la création de nouveaux vaisseaux sanguins<sup>144</sup>. Ils participent également à la création d'un environnement immunosuppresseur grâce à la sécrétion de facteurs tels que l'IL-10 et le TGF- $\beta$ <sup>139</sup>. Ces cytokines réduisent l'activité cytotoxique des lymphocytes T et inhibent la présentation d'antigène par les macrophages ainsi que leur capacité de phagocytose. De plus, le TGF- $\beta$ , associé à d'autres cytokines comme l'IL-6 ou l'EGF, favorise la croissance tumorale et l'invasion tissulaire. Pour finir, les TAMs produisent également des enzymes protéolytiques, telles que les métalloprotéases, qui dégradent la matrice extracellulaire et donc favorisent le processus de métastase<sup>144</sup>. Les TAMs peuvent également induire l'immunosuppression par contact direct avec les lymphocytes T environnant, notamment à travers l'interaction PD1 / PD-L1. En effet, les TAMs expriment à leur surface la protéine PD-L1 (protéine death ligand 1) capable de se fixer au récepteur PD1 des lymphocytes pour inactiver leur activité<sup>136</sup>. Les différentes activités pro-tumorales de ces TAMs sont résumées dans la **figure 18**.



**Figure 18:** Les différentes fonctions des macrophages au sein de la tumeur.

Les TAMs stimulent les fibroblastes à travers la sécrétion de VEGF, favorisant ainsi l'angiogénèse. Ils contribuent à la résistance aux traitements. Ils modifient l'ECM en produisant des cytokines et des métalloprotéases pour favoriser les processus de métastase. Pour finir, ils induisent un environnement immunosuppresseur par la sécrétion de cytokines anti-inflammatoires et l'interaction entre PD1 et PD-L1. (Réalisé à l'aide de Biorender)

Ils sont donc d'une importance cruciale pour la tumeur et sont de ce fait des marqueurs cellulaires permettant l'élaboration d'un pronostic chez les patientes.

#### Les macrophages et la survie

Les rôles des macrophages au sein du TME sont donc multiples et essentiels à la survie, à la croissance et à la dissémination de la tumeur. Ainsi, il semblerait cohérent que les différences de pronostic selon les sous-types de cancer soient corrélées à la quantité de macrophages au sein de la zone tumorale. Seulement, la quantité n'est pas l'unique facteur responsable du développement tumoral et de la mortalité des cancers. En effet, la polarisation de ces TAMs est une donnée importante pour définir, à la fois leur contribution dans la progression de la pathologie, mais aussi leur lien avec la survie des patients. C'est ainsi qu'en 2014, une étude décrit le lien entre la survie de patientes atteintes du cancer de l'ovaire et le ratio M1/M2 au sein des différents sous-types de cancer<sup>145</sup>. Ainsi, il semblerait qu'un ratio plus important en M1 soit lié à un taux de survie à 5 ans plus grand. Cette étude a été confirmée par l'équipe du docteur Maccio, qui a démontré que les patientes ayant un taux de M1/M2 supérieur à 1.4 étaient plus sensibles à la chimiothérapie et donc présentaient de meilleures chances de survie<sup>146</sup>. De plus, cette étude, ayant été réalisée *ex vivo*, prend en considération la complexité réelle liée aux multiples phénotypes des macrophages. Ainsi, la détection *in vivo* de ces macrophages et la définition de la balance entre leurs phénotypes est un enjeu majeur. Cela permettrait, de manière rapide et directement au bloc opératoire, de réaliser des estimations de la survie de la patiente, mais également d'adapter son traitement.

#### Les lymphocytes infiltrants la tumeur et leur fonction dans le TME :

Les lymphocytes sont des cellules immunitaires responsables d'une réponse immunitaire spécifique et durable. Pour faire face à la grande variété de cibles, il existe plusieurs types de lymphocytes, tels que les lymphocytes B, les natural killers (NK) et les lymphocytes T. Ces derniers sont divisés en plusieurs populations telles que les Treg, les LT CD8+ ou encore les LT CD4+. Ces populations sont identifiables par l'expression différentielle de certaines molécules de surface appelées CD pour Cluster of Differentiation. Durant ma thèse, j'ai principalement travaillé sur les lymphocytes T et les NK, ainsi, cette partie se concentrera principalement sur ceux-ci.

Après les macrophages, la majeure partie des cellules immunitaires est représentée par les lymphocytes T. De manière générale, les lymphocytes T représentent 10% de la masse tumorale<sup>147</sup>. Les lymphocytes infiltrants la tumeur sont appelés TILs (Tumor-infiltrating lymphocytes). Ils représentent une population hétérogène qui exerce à la fois des fonctions pro- et anti-tumorales. Ils sont recrutés par l'action chimio-attractrice de cytokines sécrétées par différents types cellulaires du microenvironnement tumoral (macrophages, fibroblastes, tumeur...).

Les lymphocytes T auxiliaires ou « helper » (TH) CD4+ jouent un rôle d'amplification de la réponse immunitaire dans la mesure où ils produisent des interleukines pro-inflammatoires, telles que l'IL2 et l'IFN- $\gamma$ , et sont responsables du recrutement de lymphocytes effecteurs, tels que les lymphocytes cytotoxiques CD8+ ou encore les natural killers. Ces LT CD4+ se différencient en plusieurs sous-populations (Th1, Th2, etc) sous l'action de différentes cytokines selon les besoins de l'organisme. Selon leur orientation, ces lymphocytes présentent un profil de sécrétion et des fonctions différentes. Par exemple, les LT CD4+ Th1 sont impliqués dans la défense contre les pathogènes, produisent des cytokines pro-inflammatoires et sont associés à un bon pronostic dans la plupart des cancers. Ces lymphocytes ont pour rôle principal la régulation positive de l'inflammation et de la réponse immunitaire et sont donc essentiels pour l'élimination de la tumeur. Seulement, ces lymphocytes sont très rares dans l'environnement tumoral. La majeure partie des lymphocytes T CD4+ présents dans le TME sont orientés vers un autre phénotype, appelé Th2, qui possède un rôle immunosuppresseur et est donc protecteur de la tumeur. En effet, ce phénotype sécrète majoritairement des cytokines telles que l'IL-4, l'IL5 et l'IL13 impliquées dans la croissance tumorale<sup>148</sup>.

Une autre population de lymphocyte T exerce un rôle majeur dans le combat contre la tumeur, les lymphocytes T CD8+ cytotoxiques. Ces cellules jouent un rôle cytotoxique spécifique contre les cibles exprimant des antigènes cibles. Par exemple, dans le contexte du développement tumoral, les cellules présentatrices d'antigènes, telles que les macrophages, présentent un antigène tumoral aux LT CD8+ naïfs au sein du ganglion lymphatique. Cette interaction va induire leur activation en cellules CD8+ effectrices. Une fois activés et envoyés vers le foyer tumoral à travers la circulation sanguine, les LT CD8+ sont capables de reconnaître cet antigène sur la cellule tumorale, de s'y fixer et d'induire la mort cellulaire à travers son activité cytotoxique. Cette mort cellulaire est réalisée par l'action de perforine et de granzyme. La perforine crée des pores dans la membrane plasmique tandis que la granzyme, peut activer la voie de signalisation de l'apoptose (caspase 3)<sup>149</sup>. Ainsi, l'abondance de ces cellules au sein d'un tissu cancéreux est favorable à la survie de la patiente<sup>150</sup>. Malheureusement, leur action est totalement inhibée par l'environnement immunosuppresseur créé par les cytokines sécrétées par d'autres cellules immunitaires (M2, LT CD4+ Th2, Treg ...) ou encore à travers l'interaction PD1 / PDL1.

Les lymphocytes T régulateurs (Tregs) sont une population lymphocytaire très représentée dans l'environnement tumoral. La présence de Tregs traduit un environnement très immunosuppresseur et donc pro-tumoral. Leur action immunosuppressive passe par la sécrétion de cytokines anti-inflammatoires, telles que l'IL10 et le TGF- $\beta$ , et l'inhibition directe de l'activité des lymphocytes CD8+, grâce à la protéine de surface CTLA4, empêchant à la fois la reconnaissance de l'antigène tumoral par le CD8+, mais également son action

cytotoxique<sup>148</sup>. Ainsi, leur présence est très souvent associée à un très mauvais pronostic pour les patients, notamment si le ratio entre les CD8+ et les Treg est faible<sup>150</sup>.

Pour finir, les lymphocytes NK peuvent également être retrouvés au sein du TME. Seulement, ceux-ci ne sont pas localisés au centre de la tumeur, mais plutôt dans le stroma de celle-ci. Ces cellules sont également cytotoxiques, mais sont issues de l'immunité innée (contrairement au CD8+). En effet, après la reconnaissance de l'antigène tumoral, celles-ci vont se fixer et envoyer des protéines cytotoxiques appelées perforines qui vont permettre le désassemblage de la membrane plasmique et donc induire la mort cellulaire<sup>151</sup>. Les NK sont souvent associées à un bon pronostic, mais l'environnement immunosuppresseur, notamment le TGF- $\beta$ , inhibe leur activité cytotoxique<sup>148</sup>.

Ainsi, l'ensemble des cellules présentes dans le TME, qu'elles soient immunitaires ou non, jouent un rôle essentiel dans le pronostic de la patiente. Face à ce constat, il serait intéressant de développer de nouvelles techniques capables de détecter et d'identifier les sous types cellulaires et leurs phénotypes au sein de la tumeur. Cette technologie permettra d'adapter la prise en charge de la patiente de manière optimale. Le temps est un facteur clé pour la survie de la patiente. Être en mesure de détecter ces variations au sein de l'environnement tumoral directement pendant la chirurgie ou *ex vivo* pourrait grandement améliorer la survie sans récurrence et la survie globale de la patiente.

## CHAPITRE III : Analyse des lésions de la trompe de Fallope comme précurseurs potentiels du cancer de l'ovaire précoce.

### Introduction

Les cancers de l'ovaire se classent au 8<sup>ème</sup> rang mondial des cancers chez la femme. Ils ont été responsables d'environ 207 000 décès pour une incidence de 314 000 cas en 2020. Ces carcinomes sont, en grande majorité, issues de l'épithélium ovarien (>90%) et sont divisés en différents sous-types. Le carcinome séreux de haut grade (HGSOC) est le sous-type le plus fréquent (>75%)<sup>23,27</sup>. Bien qu'en général, le risque de développer cette pathologie est faible, de l'ordre de 1.2 à 1.4%, les patientes atteintes d'une mutation familiale BRCA1/BRCA2 ont un risque significativement plus important, pouvant aller jusqu'à 46%. De plus, ces patientes développent un HGSOC à un âge plus précoce (30 - 35 ans) par rapport aux patientes sans mutation (60 ans). L'origine de ce carcinome est très débattue. Certaines hypothèses évoquent l'implication des cellules épithéliales de la jonction entre les trompes et le péritoine<sup>35</sup>. Pour préciser cette origine, des analyses physiopathologiques ont été réalisées sur les prélèvements chirurgicaux de patientes à risque présentant des mutations BRCA par annexectomie bilatérale prophylactique. Le protocole de sectionnement et d'examen exhaustif de l'extrémité en frange (Protocol SEE-FIM) a été réalisé sur ces prélèvements<sup>21</sup>. Ce protocole consiste à analyser en série des coupes consécutives des trompes de Fallope par immunohistochimie sur des marqueurs connus du cancer tels que l'expression de Ki67 et de p53. Une expression inhabituelle de ces marqueurs a été retrouvée au niveau de la jonction entre la surface séreuse de la trompe de Fallope et le péritoine, appelée la fimbria. Ce profil d'expression signale un début de néoplasie, ou tout du moins, une anomalie cellulaire localisée sur l'épithélium tubaire<sup>152,153</sup>. De plus, ces zones sont décrites comme des points chauds du développement tumoral, l'apparition de transformations néoplasiques des cellules y est très fréquente<sup>154</sup>. En effet, les cellules dérivées de cette région semblent présenter un phénotype de cellules souches prédisposées au cancer<sup>49 155</sup>.

Durant cette transformation néoplasique, des modifications morphologiques et phénotypiques apparaissent sur des zones très précises (de l'ordre de la dizaine de cellules) de l'épithélium des trompes de Fallope. Ces lésions sont classées en 3 groupes principaux qui se distinguent par leurs caractéristiques intrinsèques. Le premier type de lésion est appelé la lésion p53. Elle est caractérisée par une prolifération normale, mais avec un motif p53 anormal provenant d'une mutation non-sens sur le gène TP53 sur une dizaine de cellules. Le second type de lésion, appelée STIL (serous tubale intraepithelial lesion) est définie par l'accumulation de p53 dans plus de 20 cellules, dont certaines présentent des anomalies morphologiques, et

un indice de prolifération Ki-67 plus élevé (de 10 à 40 %). Cependant, ces lésions ne présentent pas de prolifération différente de la normale. Ainsi, cette région est considérée comme un potentiel carcinome en dormance. Pour finir, la lésion la plus avancée est la lésion STIC (serous tubal intraepithelial carcinoma). Au sein de cette lésion, la prolifération y est très rapide et anarchique, les cellules atteintes présentent des anomalies morphologiques et des altérations nucléaires importantes. La mutation TP53 y est également présente.

De récentes études suggèrent que ces 3 lésions, et notamment la lésion STIC, pourraient être à l'origine du cancer de l'ovaire séreux de haut grade. Pour valider cette hypothèse, le laboratoire a analysé les mécanismes moléculaires liés à ces lésions pré-néoplasiques. Les pré-lésions néoplasiques P53 seraient donc l'amorce du développement de ce cancer. L'identification de marqueurs spécifiques de ces lésions permettrait de mieux comprendre les mécanismes physio-pathologiques impliqués dans le développement des HGSOE et d'adapter la prise en charge des patientes. Ainsi, l'analyse des pièces opératoires de fimbriectomie orienterait le choix d'une fimbriectomie préventive pour ces patients à risque, mutées BRCA1/BRCA2. La fimbriectomie bilatérale permettrait alors l'élimination des potentielles sources de cellules néoplasiques pour réduire le risque d'apparition d'un cancer de l'ovaire et ce tout en maintenant les fonctions hormonales de cet organe.

### Objectifs

Dans ce contexte, une étude lipidomique a été réalisée à l'aide du SpiderMass dans l'objectif de créer un modèle de classification capable de distinguer ces différentes lésions par rapport à un épithélium tubaire sain. L'enjeu majeur de cette étude a été le faible nombre de cellules (<20 cellules) de ces lésions qui a nécessité une optimisation de l'acquisition des données. De plus, les échantillons analysés ont été fixés au formol et inclus en paraffine (FFPE). Le protocole analytique a donc été optimisé pour améliorer le signal obtenu. Ces développements ont mené à la découverte de potentiels marqueurs lipidiques de ces lésions.

En parallèle, des études de protéomiques spatiales ont été réalisées pour identifier des marqueurs protéiques associées aux régions présentant des lésions pré-néoplasiques, en incluant les protéines alternatives et les voies de signalisations associées spécifiques à ces lésions.

Pour finir, une validation par immunofluorescence (IHC) a été réalisée sur les marqueurs découverts par l'analyse protéomique spatiale shot-gun non ciblée.

Cette étude s'inscrit dans un PHRC « Fimbria » porté par le Dr. Eric Leblanc. Les pièces opératoires viennent du centre de cancérologie Oscar Lambret et les pièces de validation de Toulouse et Lille.

## Matériels et méthodes

### Réactifs et produits chimiques

Des produits chimiques de haute pureté, provenant de divers fournisseurs, ont été achetés pour les différentes expériences et utilisés tels quels, sans aucune modification. L'acétonitrile (ACN) de qualité HPLC, l'éthanol de qualité HPLC, le méthanol de qualité LC-MS et l'alcool isopropylique de qualité LC-MS ont été achetés à CARLO ERBA Reagents SAS (Val de Reuil, France). L'eau UHPLC-MS a été fournie par Fisher Scientific (Illkirch, France). L'acide trifluoroacétique et l'acide formique (FA) ont été obtenus auprès de Biosolve B.V. (Valkenswaard, Pays-Bas). L'acide alpha-cyano-4-hydroxycinnamique (4-HCCA), l'aniline (ANI) et le glycérol proviennent de Sigma Aldrich (Saint-Quentin-Fallavier, France).

### Conception expérimentale, sélection des cas et justification statistique

Des pièces chirurgicales ont été prélevées sur des patients inclus dans un essai clinique multicentrique de phase II, portant sur la fimbriectomie radicale, initié par le Centre Oscar Lambret à Lille, en France (NCT-01608074). Avant le recrutement, le protocole de l'essai a été approuvé par le comité de révision national et institutionnel (IRB) et accepté par le comité d'éthique (CPP Nord-Ouest IV, déc. 2010) conformément à la législation française et européenne. Les patients ont donné leur consentement éclairé par écrit avant de participer à l'essai. Pour protéger la vie privée des patients, aucune information personnelle n'a été utilisée dans ces expériences, et un numéro aléatoire a été attribué à chaque échantillon. Pour plus d'informations sur les critères d'inclusion et le plan d'étude de cet essai clinique, veuillez consulter le site web des essais cliniques (<https://clinicaltrials.gov/ct2/show/NCT01608074>). Nous avons pris les mesures nécessaires pour garantir que les lésions diagnostiques ne soient pas perdues pour le suivi clinique ultérieur. Dans notre approche de protéomique spatiale, nous avons utilisé une lame qui avait déjà été utilisée en IHC. La taille de la cohorte dans cette étude a été limitée en raison de la rareté de la population cible, qui consiste en des patientes porteuses de mutations BRCA subissant une annexectomie prophylactique. Il convient de noter que toutes les patientes recrutées ne présentaient pas de lésions pré-néoplasiques et que la sélection des tissus a été réalisée par les anatomopathologistes sur la base d'un double marquage P53/Ki67. Cependant, il convient également de souligner qu'il n'existe actuellement aucune information sur la prévalence de ces lésions pré-néoplasiques dans la fimbria des patientes présentant des mutations du gène BRCA1 ou des antécédents familiaux et qui subissent une annexectomie prophylactique. Ainsi, notre étude protéomique a porté sur un nombre relativement restreint de patientes. Cependant, des études antérieures ont montré qu'il est possible d'obtenir des résultats statistiquement significatifs, même avec un nombre limité d'échantillons. Par conséquent, dans notre étude, nous avons choisi de donner la priorité

à la qualité des échantillons afin de garantir des résultats précis et fiables pour la partie découverte et d'augmenter le nombre de patients pour la partie validation. Une cohorte prospective de 8 patients a été sélectionnée pour effectuer une analyse protéomique, dont 4 patients avec des lésions de la signature p53, 2 patients avec des lésions STIL, 2 patients avec des lésions STIC, et 1 patient avec HGSC (**voir tableau 3**). Le tissu normal a été analysé à partir de la partie normale de la section de tissu des patients présentant des lésions de la signature p53 et aucune autre lésion concomitante. Ensuite, une cohorte rétrospective de 27 patients a été utilisée (**voir tableau 4**) pour l'analyse SpiderMass.

**Tableau 3:** description des caractéristiques cliniques des patientes présentant une lésion préneoplasique dans la fimbria des trompes de Fallope (cohorte prospective).

Case ID	Lesions sampled	BRCA status	Age at diagnosis	Location in fimbria	Ovary lesion	Other lesions	Personal history of breast cancer	Other remarks
1	Normal, p53 signature	BRCA1	40	Right	No	SCOUT in tube	No	Left ovarian theca cyst
2	Normal, p53 signature	BRCA1	39	Right and left	No	No	Yes	
3	STIL (x3)	unknown	61	left	No	No	No	Uterus Leiomyoma
4	STIC (x3)	BRCA1	36	Right and left	No	Several p53 signature	No	HGSC in the left tube
5	Normal, p53 signature	Family history	38	Left	No	No	No	
6	STIL, STIC	BRCA1	48	Left	No	Several p53 signature	Yes	

7	Normal, p53 signature	BRCA1	35	Left	No	Several p53 signature	yes	HGSC in the left fimbria
8	Patient 8 shows an ovarian high-grade serous carcinoma with no evidence of pre-neoplastic lesion in the fallopian tube. The BRCA status is unknown.							

**Tableau 4:** description des caractéristiques cliniques des patientes présentant une lésion préneoplasique dans la fimbria des trompes de Fallope (cohorte rétrospective).

Case ID	Lesions sampled	BRCA status	Age at diagnosis	Location in Fimbria	Ovary lesion	Other Lesions	Personal history of breast cancer	Other remarks
F1	STIL	BRCA2	65	left	no	no		
F2	Normal, p53 signature	BRCA1	38	right	no	no	yes	
F3	STIC	BRCA1	37	right	no	STIL to the left fimbria	yes	HGSC in the left tube
F4	Normal, p53 signature	unknown	50	left	no	no	no	Follicular cyst in the ovary
F5	Normal, p53 signature	BRCA2	50	Both sides	no	no	Lesion in the breast	
F6	STIC	BRCA1	41	right	no	no	no	cyst in the left ovary
F7	Normal, p53 signature	no	45	right	no	HGSC in the left tube		Family breast and ovarian cancer
F8	Normal, p53 signature	BRCA2	67	left	no	no	no	

<b>F9</b>	Normal, p53 signature	BRCA1	70	Both sides	No	No	no	
<b>F10</b>	STIL, STIC	BRCA1	62	Both sides	HGSC in both ovaries	no	no	
<b>F11</b>	STIC	BRCA1	44	right	No	no	yes	
<b>F12</b>	Normal, p53 signature	no	68	right	no	no	no	
<b>F13</b>	Normal, p53 signature	unknown	62	right	Brenner tumor in right ovary	no	yes	
<b>F14</b>	STIC	BRCA1	44	right	no	HGSC in the right chorion	yes	
<b>F15</b>	Normal, p53 signature	BRCA2	46	left	no	no	no	
<b>F16</b>	Normal, p53 signature	BRCA2	50	left	no	no	yes	
<b>F17</b>	STIC	BRCA1	44	right	no	no	no	
<b>F18</b>	STIL	BRCA2	49	left	no	no	no	
<b>F19</b>	Normal, p53 signature	BRCA2	43	right	no	no	no	
<b>F20</b>	Normal, p53 signature	unknown	51	right	no	no	yes	Family breast cancer
<b>F21</b>	Normal, p53 signature	BRCA2	41	right	no	no	no	
<b>F22</b>	STIL	BRCA1	58	right	no	no	yes	
<b>F23</b>	Normal, p53 signature	BRCA1	60	left	no	HGSC in the right tube	yes	

<b>F24</b>	STIC	BRCA1	42	right	no	no	yes	
<b>F25</b>	STIC, STIL	BRCA1	68	Both sides	no	no	no	
<b>F26</b>	STIL	BRCA1	66	left	no	no	yes	
<b>F27</b>	STIL, STIC	unknown	35	Both sides	no	HGSC in the left tube	yes	

### Traitements des échantillons

Les tissus utilisés dans cette étude sont des tissus FFPE obtenus lors d'annexectomies prophylactiques chez des femmes présentant des mutations BRCA1 ou des antécédents familiaux. Cette étude a été réalisée sur des patientes présentant des lésions pré-néoplasiques sans lésions ovariennes concomitantes. Tous les échantillons ont été examinés selon le protocole SEE-FIM (Sectioning and Extensively Examining the FIMbria<sup>21</sup>). Sur ces échantillons, des coupes de tissus de 7 µm d'épaisseur ont été réalisées à l'aide d'un microtome et déposées sur une lame de verre, HPS (Haematoxylin Phloxine Safran) et des immunomarquages contre p53 (Ventana, 790-4286) et Ki-67 (Ventana, 800-2912) ont été réalisés et examinés par un pathologiste pour identifier des lésions pré-néoplasiques (p53/STIL/STIC/HGSC)<sup>156</sup>.

Protéomique à résolution spatiale sur tissu guidée par l'IHC de p53 et Ki67

### Préparation des échantillons

Pour garantir une puissance statistique adéquate et tenir compte de la variabilité biologique potentielle, nous avons prélevé plusieurs échantillons sur les mêmes patients, mais sur des lésions différentes, ce qui a donné quatre régions d'échantillonnage par lésion (**tableau 3**) pour la protéomique. Cette approche augmente le nombre de réplicats, améliorant ainsi la fiabilité de l'analyse statistique. De plus, elle permet une évaluation plus complète de l'hétérogénéité au sein et entre les différentes lésions. Les mêmes lames que celles utilisées et annotées par le pathologiste ont été démontées. La résine a été enlevée en les trempant pendant une nuit dans du xylène et en les rinçant avec des bains de xylène et d'éthanol. Les tissus ont été réhydratés 5min dans chaque bain successif de degré d'éthanol décroissant (2x95°, 1x30°) et deux bains de tampon NH<sub>4</sub>HCO<sub>3</sub> 10mM. Ensuite, une étape de démasquage des antigènes a été réalisée pour faciliter l'accès de la trypsine aux protéines. Pour ce faire, les lames ont été plongées dans du Tris 20mM pH 9 à 90°C pendant 30 minutes, rincées deux fois avec du NH<sub>4</sub>HCO<sub>3</sub> et séchées sous vide à température ambiante.

### Digestion trypsique in situ

Sur des lésions prénéoplasiques localisées, une digestion trypsique a été réalisée à l'aide d'une imprimante à jet d'encre chimique (CHIP-1000, Shimadzu, Kyoto, Japon). La région a été soigneusement sélectionnée pour s'assurer que l'analyse se limite aux cellules épithéliales marquées par l'IHC, réduisant ainsi le risque de contamination par d'autres types cellulaires. La solution de trypsine (Sequencing grade modified (Promega), 40µg/mL, 50mM NH<sub>4</sub>HCO<sub>3</sub> buffer) a été déposée sur une région définie à 600x600 µm<sup>2</sup> pendant 2h. Pendant ce temps, la trypsine a été changée toutes les demi-heures. Avec 350 cycles et 450 pL par spot, un total de 6.3µg de trypsine a été déposé. Pour arrêter la digestion, 0.1% de TFA a été déposé pendant 25 cycles.

### Extraction liquide

Après la micro-digestion, le contenu du spot a été recueilli par microjonction liquide à l'aide du système TriVersa Nanomate (Advion Biosciences Inc., Ithaca, NY, USA) en utilisant les paramètres d'extraction liquide et d'analyse de surface (LESA) optimisés par le laboratoire ; comprenant 3 compositions de solvants différentes respectivement 0,1% TFA, ACN/0,1% TFA (8:2, v/v), et MeOH/0,1% TFA (7:3, v/v). Une séquence LESA complète effectuée 2 cycles pour chaque mélange composé d'une aspiration (2µL), d'un mélange sur le tissu, et d'une distribution dans des tubes à faible liaison. Pour chaque spot intéressant, 2 séquences sont regroupées dans le même flacon.

### Protéomique basée sur la MS

Une description complète de NanoLC-ESI-MS/MS est fournie dans les informations complémentaires. Nous avons mis en œuvre un processus de randomisation pour minimiser l'impact des effets de lot dans notre analyse.

### Identification des protéines

Les spectres MS/MS bruts ont été traités à l'aide de MaxQuant (version 1.5.6.5, 29). Une base de données leurre a été utilisée pour identifier les faux positifs contre la banque de données Homo sapiens Uniprot (version de 2017\_02, 20 172 entrées) et une base de données contenant 262 contaminants couramment détectés. La base de données UniProt humaine a été utilisée comme base de données directe et une base de données inverse pour la recherche de leurre a été automatiquement générée dans MaxQuant. Les tolérances de masse ont été fixées à 10 ppm et 20 ppm respectivement pour la mesure des parents et des fragments. La spécificité de l'enzyme a été réglée sur "trypsine" avec un maximum de 2 clivages manqués autorisés. L'oxydation de la méthionine et l'acétylation de l'extrémité N-terminale de la protéine ont été définies comme des modifications variables. La FDR < 1 % a été fixée pour l'identification des

peptides et des protéines. Deux peptides dont un unique étaient nécessaires pour évaluer l'identification des protéines. Pour la label-free quantification, l'algorithme MaxLFQ a été utilisé. L'option "Match Between Runs" a été activée pour maximiser le nombre d'événements de quantification entre les échantillons. Cette option a permis de quantifier les caractéristiques MS1 à haute résolution qui n'ont pas été identifiées dans chaque mesure. Les données générées par MaxQuant ont été analysées à l'aide de Perseus (version 1.6.2.3,154). Les valeurs LFQ ont été utilisées et les protéines ont été supprimées si elles se trouvaient dans la catégorie identifiée uniquement par des modifications de site, dans la base de données de leurres inversés ou identifiées dans la base de données de contaminants.

### Analyses statistiques protéomiques

Les protéines qui n'étaient pas présentes dans au moins trois des quatre réplicats de chaque lésion ont été supprimées. Nous avons pris l'expression moyenne par groupe pour effectuer une comparaison et l'avons visualisée à l'aide d'un diagramme de Venn. Les valeurs individuelles du LFQ ont été utilisées pour réaliser un diagramme de dispersion multiple et calculer une corrélation de Pearson entre les échantillons. Les valeurs manquantes ont été imputées sur la base de la distribution normale (largeur = 0,3, décalage vers le bas = 1,8). Une analyse en composantes principales (ACP) a été effectuée pour comparer la teneur en protéines de chaque échantillon. Un test ANOVA multi-échantillon a été réalisé et consolidé par un FDR basé sur la permutation ( $FDR < 0,05$ , 250 randomisations). Une comparaison spécifique entre les échantillons normaux et les échantillons portant la signature p53 a été effectuée à l'aide d'un test T de Student. Les protéines différentiellement exprimées présentant des différences significatives ont été sélectionnées (les valeurs étaient des z-score). Les échantillons protéomiques ont ensuite été regroupés en fonction d'une moyenne euclidienne comme mesure de distance pour le regroupement des colonnes et des lignes. Les protéines surrégulées ou sous-régulées dans les différents groupes ont été utilisées pour effectuer une analyse d'annotation des termes de l'ontologie des gènes à l'aide de Funrich (v3.1.3, 31). Un test hypergéométrique a été effectué sur toutes les listes de gènes/protéines annotées en comparant les différents ensembles de données. L'enrichissement pour les processus biologiques, les facteurs de transcription et les composants cellulaires a été présenté sous forme de diagramme à barres.

### Analyses immunohistochimiques de confirmation

La validation du pronostic précoce a été réalisée à l'aide d'anticorps dirigés contre CAVIN2 (Rockland, 600-401-J31), Emilin2 (Thermo Scientific, pa584657), CAVIN1 (Abnova, H00284119-M02), SPTAN1 (OriGene, TA812019), et FBLN5 (Invitrogen, MA5-42544). Après déparaffinage et récupération de l'antigène avec un tampon au citrate, les tissus ont été incubés avec un anticorps primaire à 4°C pendant une nuit, suivie de l'application d'un

anticorps secondaire conjugué au fluor Alexa, (dilution 1/200) pendant 1 heure à température ambiante. Nous avons utilisé les anticorps primaires aux dilutions suivantes : Cavin-1 (1/300), Cavin-2 (1/100), SPTAN1 (1/2000), Emilin-2 (1/500), Fibulin-5 (1/500). Toutes les lames ont été imagées au microscope confocal Zeiss LSM700. Trois images ont été prises pour chaque section de tumeur.

#### Analyses lipidomiques grâce à la technologie SpiderMass Instrument

Le principe de la technologie SpiderMass et la configuration de l'instrument ont été décrits dans une étude précédente<sup>16</sup>. Brièvement, le setup est composé de trois parties, un système laser permettant l'ablation des tissus analysés à distance, une ligne de transfert permettant le transfert de l'aérosol vers le spectromètre de masse qui représente la troisième partie. La sonde laser est composée d'un système OPO accordable entre 2,8  $\mu\text{m}$  et 3,1  $\mu\text{m}$  (Radiant version 1.0.1, OPOTEK Inc., Carlsbad, USA) pompé par un laser Nd:YAG pulsé (durée d'impulsion : 5 ns,  $\lambda=1064$  nm, Quantel, Les Ulis, France) et présentant un taux de répétition de 10 Hz. Une fibre laser multimode GeG (diamètre du cœur de 450  $\mu\text{m}$  ; longueur de 1 m ; Le verre Fluoré, France) est connectée à la sortie du système laser et une pièce à main comprenant une lentille de focalisation de 4 cm est fixé à l'extrémité de la fibre laser. La pièce à main permet à l'utilisateur de tenir le système et de passer au crible la surface des tissus bruts ou de fixer le système sur un support. À la distance de focalisation, la taille du spot laser sur l'échantillon est de 500  $\mu\text{m}$  de diamètre. Dans ces expériences, l'appareil a été fixé sur un bras robotisé (bras robotisé MECA 500, Mecademic, Montréal, Canada) pour obtenir un positionnement précis et le temps d'irradiation a été fixé à 3 secondes (c'est-à-dire 30 tirs laser) à une énergie laser de 4 mJ/impulsion correspondant à une fluence laser de 2,03 J/cm<sup>2</sup>. L'énergie laser a été mesurée au point focal à l'aide d'un wattmètre (ThorLabs, Maisons-Laffitte, France). La deuxième partie du système correspond à une ligne de transfert de 3 m de long fabriquée à partir d'un tube Tygon ND 100-65 (2,4 mm de diamètre intérieur, 4 mm de diamètre extérieur, Akron, États-Unis). La ligne de transfert est fixée au côté de la pièce à main et positionnée près de la sortie de la fibre laser afin d'aspirer le matériau ablaté provenant de l'irradiation laser. De l'autre côté, le tube de transfert est directement fixé sur le spectromètre de masse (Xevo G2-XS, Waters, Wilmslow, Royaume-Uni) dont la source d'électrospray conventionnelle a été retirée et remplacée par une interface à pression atmosphérique (REIMS). Les acquisitions MS ont été réalisées en mode résolution de l'instrument en ion positif, avec un temps de balayage de 1 seconde. Pour l'analyse des tissus, la séquence d'acquisition spectrale était composée d'une acquisition avec des périodes d'irradiation de 3 seconds comptes tenus de la superficie réduite des lésions.

#### Analyse SpiderMass de tissus FFPE

Les tissus ont été sectionnés à une épaisseur de 7  $\mu\text{m}$  et déposés sur une lame de verre en polylysine. Ensuite, les coupes de tissus FFPE ont été soumises à deux déparaffinages au xylène pendant 5 minutes et ont été manuellement pulvérisées avec une solution de glycérol/alcool isopropylique (IPA) (8:2, v/v) en deux passages successifs à l'aide d'un sprayeur manuel (Agilent). Le pousse-seringue (74 900 series Cole Parmer Instrument Company) a été réglé à un débit de 700  $\mu\text{L}/\text{min}$ . Les 2 passages successifs correspondaient à 5  $\mu\text{L}$  déposés sur 1  $\text{cm}^2$  et duraient  $\sim 10$  s. Les échantillons ont été analysés dans les 10 min après le dépôt de glycérol

#### Modèle de classification à partir des données SpiderMass

Pour l'analyse des données, tous les fichiers de données brutes produits avec l'instrument SpiderMass ont été importés dans le logiciel Abstract Model Builder (AMX v.0.9.2092.0). Après l'importation, les spectres ont d'abord été prétraités. Les étapes de prétraitement comprennent la soustraction du bruit de fond, la normalisation du nombre total d'ions, l'utilisation d'un pic de référence pour corriger la masse (« lockmass ») et le rétrécissement à une fenêtre de 0,1Da pour le binning des spectres. Tous les spectres MS traités obtenus à partir des 27 échantillons validés histologiquement ont ensuite été utilisés pour construire un modèle de classification par analyse en composantes principales et analyse discriminante linéaire (ACP-LDA). La première étape a consisté en une analyse non-supervisée de type ACP pour réduire la multidimensionnalité des données en générant des caractéristiques qui expliquent la majeure partie de la variance observée. Ces caractéristiques ont ensuite été soumises à une analyse supervisée à l'aide de la LDA en définissant les classes sur lesquelles le modèle sera basé. La LDA permet de construire un modèle à partir des classes données aux spectres de l'échantillon puis d'évaluer le modèle par une validation croisée. La validation croisée a été effectuée en utilisant soit la méthode "20% out", soit la méthode "leave one patient out". Pour la première méthode, 20 % des spectres MS sont prélevés au hasard sur l'ensemble des spectres, et le modèle est reconstruit à partir des 80 % restants. Les 20 % de spectres restants sont utilisés pour interroger le modèle reconstruit. La permutation est automatiquement répétée pendant 5 cycles avant de présenter les résultats de la validation croisée. Pour la deuxième méthode, les spectres sont regroupés par patient et éliminés un par un ; à chaque étape, le modèle sans le patient est interrogé par rapport à ce modèle.

#### Identification des lipides

Les intervalles m/z correspondant aux signaux ayant la plus grande contribution à la variance expliquée observée dans les différents groupes ont été sélectionnés pour l'identification par MS/MS. Pour ces expériences, les paramètres étaient les mêmes que ceux décrits dans la

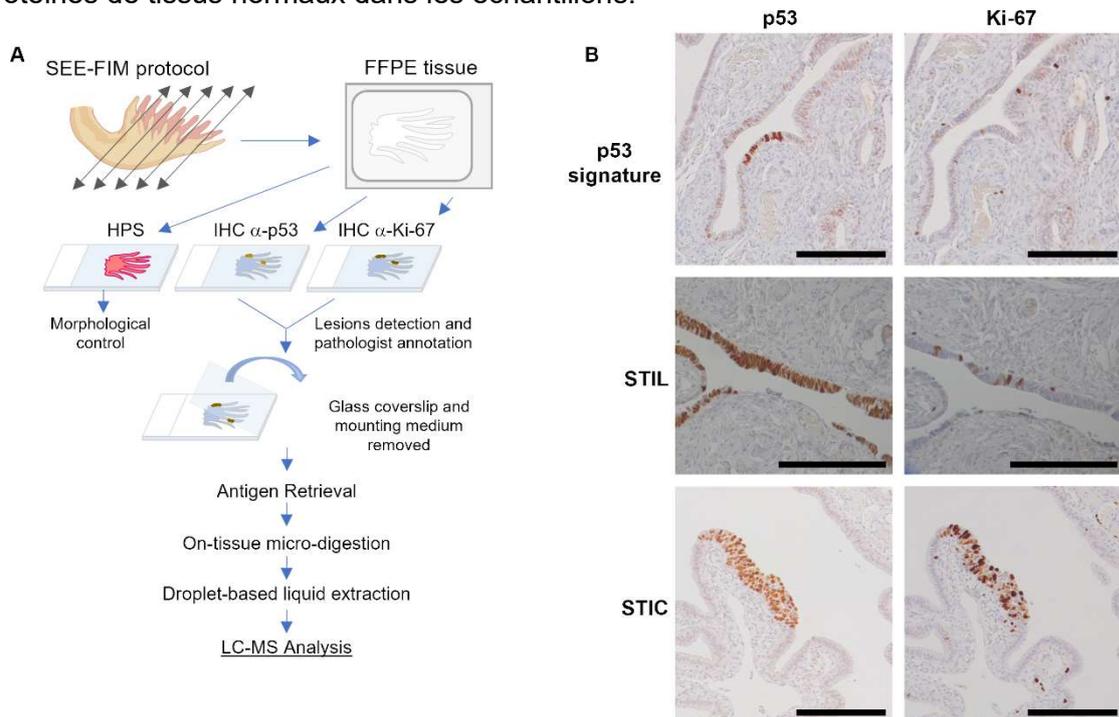
section SpiderMass. Les scans complets ont été acquis dans le Xevo G2-XS. Les identifications ont été réalisées directement sur le tissu en effectuant d'abord un balayage complet pour vérifier la présence des masses ciblées. Ensuite, après avoir basculé en mode MS/MS, les ions ont été sélectionnés puis fragmentés par dissociation induite par collision (CID) avec une fenêtre d'isolement de 1 m/z. Les spectres MS/MS ont été acquis pour un temps d'irradiation continu de 5 s. Les annotations lipidiques ont été réalisées manuellement à travers les bases de données Metfrag, LipidMaps et Alex 123.

## Résultats

### *Spatially resolved protéomique in situ des lésions du FTE*

L'objectif principal de l'essai clinique était d'étudier l'efficacité de la fimbriectomie radicale par laparoscopique bilatérale pour éliminer la source potentielle de cellules dysplasiques dans les trompes de Fallope tout en maintenant la fonction hormonale ovarienne<sup>48</sup>. Une deuxième partie visait à identifier une chronologie protéomique potentielle associée à la progression et au développement du cancer de l'ovaire. Compte tenu de la rareté de la lésion pré-néoplasique. Seul un nombre limité de patientes a été recruté pour l'essai clinique. Pour optimiser l'utilisation des échantillons de tissus, nous avons décidé de réutiliser les coupes de tissus FFPE préalablement immuno-colorées en utilisant des méthodes hautement précises basées sur la spectrométrie de masse (**Figure 19A**). Une cohorte de 8 patientes présentant des mutations BRCA 1 ou une susceptibilité héréditaire a été sélectionnée pour les expériences protéomiques spatialement résolues (**tableau 3**). Ces expériences visaient à cibler les lésions pré-néoplasiques, qui ne sont présentes que dans un très petit nombre de cellules déjà en cours de transformation dans de petites zones des tissus. Pour confirmer la présence de ces lésions, une double IHC contre les marqueurs p53 et Ki-67 est nécessaire (**Figure 19B**). Pour permettre les expériences de protéomique spatiale, nous avons développé un nouveau protocole qui utilise la même lame d'IHC que le pathologiste, qui avait précédemment servi à la détection des lésions (**Figure 19A**). Le ciblage de ces lésions

spécifiques permet une localisation précise des cellules affectées et diminue la présence de protéines de tissus normaux dans les échantillons.

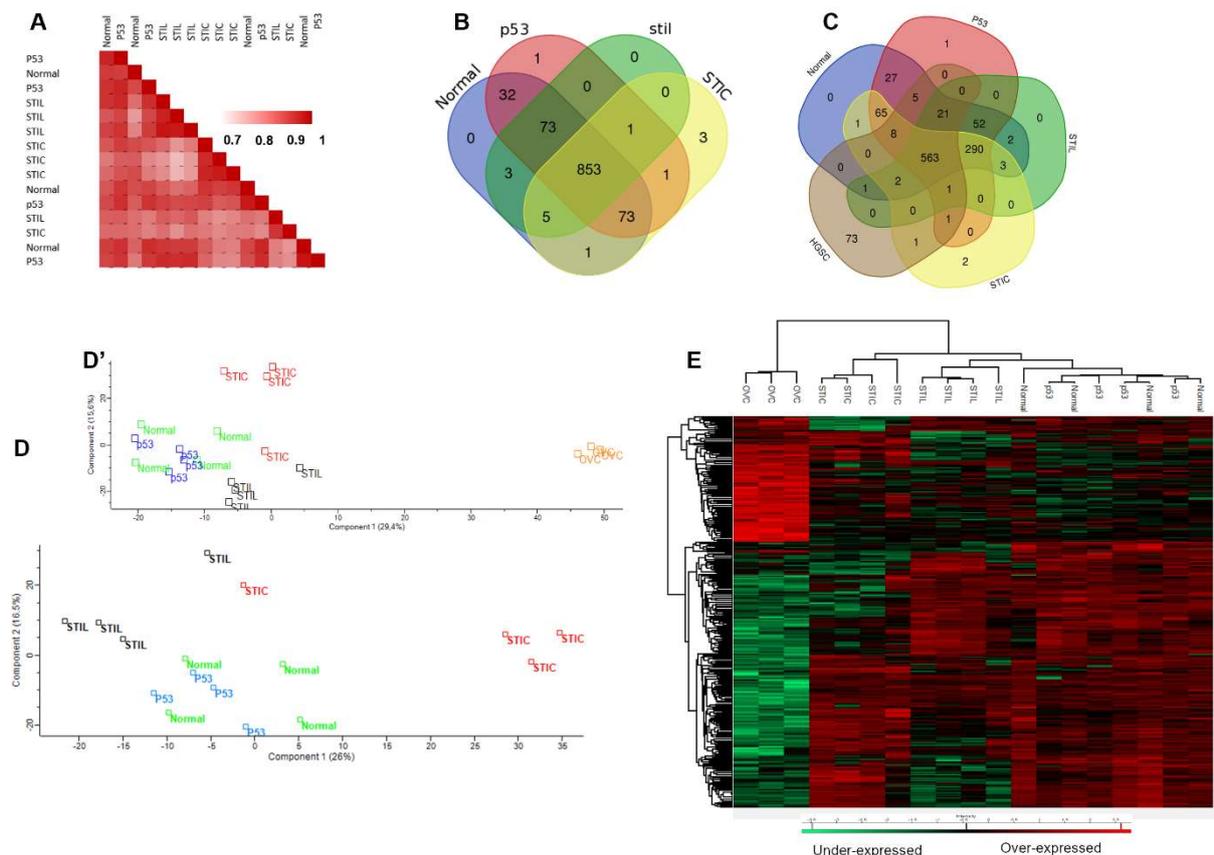


**Figure 19:** Workflow utilisé pour la protéomique résolue dans l'espace à l'aide d'une section de tissu IHC.

(A) Protocole basé sur les tissus du protocole SEE-FIM. Après l'IHC contre p53 et Ki-67, la lamelle et le milieu de montage sont retirés pour accéder à la section de tissu. La digestion de la lésion est effectuée directement sur la section de tissu et une extraction liquide en gouttelettes est réalisée pour récupérer les peptides avant l'analyse protéomique par MS. (B) Les lésions pré-néoplasiques trouvées dans la trompe de Fallope sont définies par un marqueur positif de p53 et l'index Ki-67. Pour la signature p53 : accumulation de p53 dans au moins 12 cellules sans anomalies morphologiques et faible indice Ki-67 ; STIL : même accumulation de p53 dans plus de 20 cellules avec quelques anomalies morphologiques et un indice Ki-67 prolifératif plus élevé (10-40%) ; STIC : p53 et indice Ki-67 élevés et cellules de morphologie atypique (ressemblant à un carcinome).

Les lames IHC colorées pour p53 et Ki67 ont été directement utilisées pour réaliser les expériences de protéomique spatiale à partir de tous les patients et de toutes les coupes de tissus. Après marquage de la zone et démontage de la lamelle, les tissus ont été soumis à une micro-digestion à la trypsine et d'une micro-extraction (**Figure 19A**). Notre workflow a permis d'identifier 10 375 séquences peptidiques uniques correspondant à 1 617 groupes de protéines distincts dans les lésions pré-néoplasiques et les tissus non pathologiques. La label-free quantification (LFQ) des protéines a permis de quantifier un total de 1571 protéines. Après avoir filtré les protéines en fonction d'un nombre minimum de valeurs (2/3 des valeurs valides) dans au moins un des quatre groupes définis (normal, signature p53, STIL et STIC), nous avons retenu 1 046 protéines pour la suite de l'analyse. Le nombre de protéines identifiées est cohérent avec notre expérience précédente sur les tissus FFPE, compte tenu du fait que les tissus ont été stockés pendant de nombreuses années après la fixation. L'analyse protéomique spatiale n'a pas été affectée par le retrait de la lamelle de verre et du milieu de montage des sections IHC évaluées par le pathologiste. Pour valider ces résultats, nous avons effectué une analyse de corrélation de toutes les protéines quantifiées. Les profils protéomiques des

échantillons d'un même groupe comparés aux variations entre groupes étaient similaires (corrélations de Pearson moyenne de 0,92). Les principales différences ont été observées entre les lésions STICs et STILs (corrélations de Pearson moyenne de 0,84) (**figure 20A**). Bien que la taille de notre échantillon soit limitée, ces résultats sont cohérents avec la variabilité inter-patients généralement observée, ce qui indique que notre protocole n'a pas modifié de manière significative le contenu protéomique des coupes de tissus. Cela nous a permis d'effectuer en toute confiance une analyse comparative afin d'obtenir une signature protéomique spécifique des différentes lésions.



**Figure 20:** Comparaison des différentes lésions en fonction de leur signature protéomique.

(A) Analyse de corrélation basée sur le protéome entier. Diagrammes de Ven montrant la distribution des protéines identifiées parmi B) le tissu normal et les différentes lésions et C) avec l'ajout des données HGSC. Analyse en composantes principales utilisant D) le tissu normal et les lésions préneoplasiques et D') avec l'ajout des données HGSC. E) Regroupement hiérarchique des protéines les plus variables (ANOVA avec FDR basé sur la permutation  $< 0,05$ ) incluant les données HGSC.

### Comparaison des profils protéomiques de la signature p53, du STIL, du STIC et de la fimbria normale

La teneur en protéines de chaque groupe a été comparée en calculant la moyenne des réplicats et en utilisant un diagramme de Venn. Il a été observé que 853 protéines sont communes (81,5 % de chevauchement) entre les tissus normaux et les lésions préneoplasiques (**Figure 20B**). Trois protéines se sont avérées exclusives aux lésions STIC et une protéine n'a été identifiée que dans la lésion signée p53. Parmi ces protéines, la protéine RNA binding protein 10 (RBM10) a été trouvée dans les lésions de signature p53, mais pas

dans le tissu normal ni dans les autres lésions. La RBM10 est une protéine de liaison à l'ARN située dans le noyau qui joue un rôle crucial dans l'épissage de l'ARN et qui est impliquée dans plusieurs maladies humaines, y compris le cancer. De plus, trois protéines ont été trouvées exclusivement dans les lésions du STIC : la glucosamine (N-acétyl)-6-sulfatase (GNS), l'upstream binding transcription factor, la RNA polymerase I (UBTF), et l'ATPase family, AAA domain containing 3 (ATAD3A/B). Selon les données du Human Protein Atlas (HPA), RBM10, qui a été trouvé exclusivement dans la lésion de signature p53, est surexprimé dans les cancers du col de l'utérus et de l'endomètre et a une valeur pronostique favorable. D'autre part, l'expression élevée de GNS et d'ATAD3A/B est associée à de mauvais taux de survie dans le cancer du foie, tandis qu'une forte expression de GNS dans les gliomes et le cancer de l'ovaire est un marqueur de mauvais pronostic. En revanche, l'expression d'UBTF ne semble pas avoir d'impact sur la survie. En ce qui concerne les données HPA sur les protéines identifiées, il convient de noter que si ces protéines sont indiquées comme étant fortement exprimées dans les trompes de Fallope, cette expression est principalement détectée dans les cellules glandulaires. Ces cellules se trouvent en plus grande proportion près de l'utérus, alors que les cellules ciliaires sont plus représentées dans la fimbria. Par conséquent, les niveaux d'expression de ces protéines peuvent varier en fonction de la région spécifique de la trompe de Fallope analysée. L'inclusion des données HGSOc dans l'analyse (**Figure 20C**) a révélé que l'UBTF était exprimée à la fois dans les lésions STIC et HGSOc. La fibrillarine (FBL) n'est présente que dans les lésions précancéreuses et non dans les tissus normaux. Cette protéine est également observée dans les HGSOc (**Figure 20C**).

Un nombre limité de protéines a été exclusivement identifié dans chaque groupe, ce qui indique que la variation concerne principalement l'abondance des protéines plutôt que leur composition spécifique. Pour mieux comprendre la modulation enregistrée dans les différentes lésions, nous avons effectué une analyse en composantes principales (ACP) (**Figure 20D'**). Les composantes 1 et 2 représentent 42,5 % de la variation totale des données. Les échantillons de tissus normaux et la signature p53 sont difficiles à distinguer. Cependant, les échantillons des lésions STIL et STIC étaient différents des échantillons normaux et de la signature p53 dans la composante 2. Il est intéressant de noter que les lésions STIL étaient plus proches des lésions normales et de la signature p53 dans la composante 1 que les échantillons STIC. Nous avons également confirmé que le protéome des tumeurs ovariennes HGSOc est très différent de celui des tissus STILs, STICs, de la signature p53 et des tissus normaux (**Figure 20D**). Nous avons ensuite procédé à un regroupement hiérarchique pour analyser les différences significatives dans l'expression des protéines entre les cinq groupes (**figure 20E**). Les résultats ont montré que les tumeurs HGSOc étaient séparées des autres tissus par une branche principale. La deuxième branche est divisée en deux sous-branches,

l'une séparant les lésions STIC des lésions p53 signature/Normal/STIL, et l'autre séparant les lésions STIL des tissus p53/Normal (**Figure 20E**). Comme observé précédemment, le dernier groupe (p53/Normal) était difficile à différencier. Néanmoins, les différences moléculaires dans la composition cellulaire entre l'ovaire et la trompe de Fallope ne nous ont pas permis de tirer une conclusion directe sur la relation entre HGSC et les lésions pré-néoplasiques de la trompe de Fallope. Cependant, nous avons démontré que le profil protéomique des lésions STIC était plus proche de HGSC que des autres lésions, et que chaque lésion avait une signature protéomique spécifique.

#### *Analyse protéomique descriptive des lésions préneoplasiques*

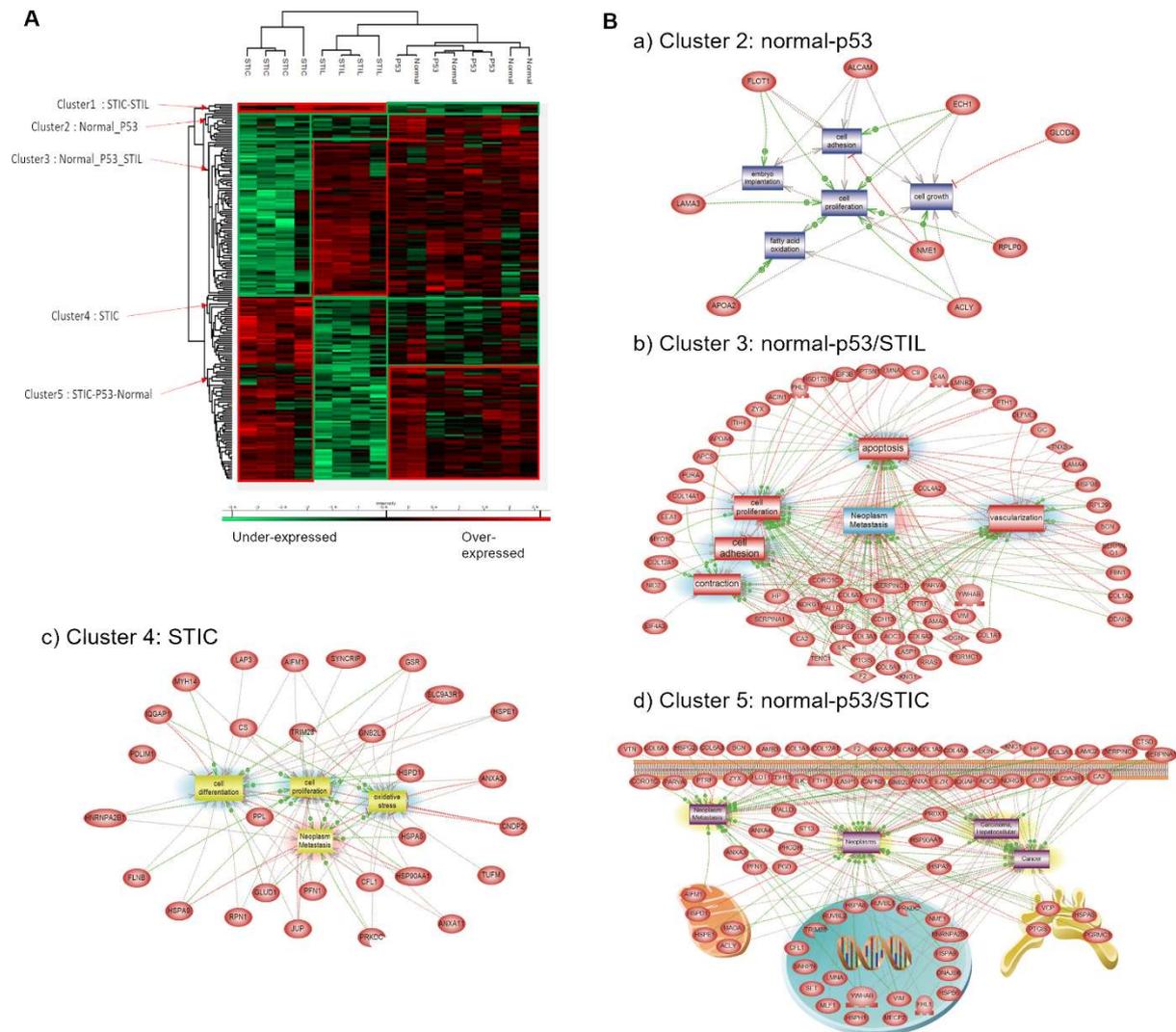
Pour une analyse plus approfondie des lésions préneoplasiques du FTE, un test ANOVA a été réalisé sans les données de l'HGSC. Au total, 197 protéines ont été exprimées de manière différentielle (significativement différente) dans les quatre groupes. Un regroupement hiérarchique basé sur l'expression de ces protéines a donné lieu à deux groupes principaux (**Figure 21A**). Comme précédemment observé, les signatures normales et p53 n'ont pas été différenciées. Ce groupe ainsi que les lésions STIL ont été regroupés (**figure 21A**). Le profil protéomique des lésions STIC diffère significativement des deux autres groupes. La Heat map présente cinq groupes principaux de protéines. Les protéines enrichies dans les régions STIL et STIC sont représentées dans le groupe 1. Celles qui étaient plus abondantes à la fois dans le tissu normal et dans les lésions signées p53 étaient représentées dans le groupe 2.

Les protéines surexprimées dans les tissus normaux, les lésions à signature p53 et les lésions STIL, étaient représentées dans le groupe 3, mais pas les lésions STIC. Les protéines surexprimées dans les lésions STIC étaient représentées dans le groupe 4, tandis que celles spécifiques aux tissus normaux, aux lésions à signature p53 et aux lésions STIC étaient regroupées dans le groupe 5. Le groupe 1 (STIL/STIC) ne comprenait que 5 protéines, à savoir la protéine MARCKS-related protein (MARCKSL1), la myosin-10 (MYH10), la protéine SET (SET), la Double-stranded RNA-binding staufer homolog-1 (STAU1), et la D-3-phosphoglycerate dehydrogenase (PHGDH). La protéine SET participe à de nombreuses fonctions cellulaires, dont la réparation de l'ADN, la transcription, la survie cellulaire et la prolifération. Cette protéine est également impliquée dans de nombreux processus cancéreux tels que la métastase, le développement de la résistance thérapeutique et joue un rôle clé dans la tumorigenèse<sup>157</sup>. MARCKSL1 est connue pour favoriser la progression de l'adénocarcinome pulmonaire en régulant la transition épithélio-mésenchymateuse (MTE)<sup>158</sup>. STAU1 a déjà été identifiée dans le cancer colorectal<sup>159</sup>, tandis que la protéine PHGDH a été identifiée dans le cancer de l'ovaire<sup>160</sup>. Dans le groupe 2 (normal-p53), 15 protéines ont montré une surexpression pour les tissus normaux et les lésions à signature p53. Parmi ces protéines,

l'antigène CD166 (ALCAM/CD166) est considéré comme un marqueur potentiel de cellules souches cancéreuses. Il a été identifié comme un marqueur pronostique défavorable dans le cancer du pancréas<sup>161</sup>. Cette protéine est connue pour contribuer à l'invasion locale et à la progression tumorale en agissant sur le détachement des cellules tumorales<sup>162</sup>. De plus, les protéines TNPO1, ACLY, NME1, FLOT1 et KLC4 sont déjà connues pour être impliquées dans le cancer ovarien épithélial<sup>163</sup>.

Les analyses sur Gene Ontology (GO) ont confirmé que ces protéines sont impliquées dans la prolifération cellulaire, l'adhérence et la croissance cellulaire (**Figure 21B,b**). Le groupe 3 (normal-p53/STIL) contient 80 protéines. Des protéines impliquées dans le complément et de la cascade de coagulation sont présentes dans ce groupe (la SERPIN A1, SERPIN C1, SERPIN G1, C4a, C9, la vitamin-D binding protein (GC), pro-thrombine (F2), vitronectine (VTN), kininogène (KNG1)) ainsi que 10 membres de la famille du collagène. La protéine PGRMC1 (Membrane-associated progesterone receptor component 1) est connue pour être impliquée dans le cancer de l'ovaire<sup>164</sup>, de même que PTGIS (Prostacyclin synthase) et la protéine NDRG1, qui est connue pour moduler les gènes impliqués dans la métastase du cancer de l'ovaire<sup>165,166</sup>. L'analyse GO a confirmé que ces protéines sont impliquées dans la néoplasie, la métastase, l'angiogenèse et l'apoptose (**Figure 21B,b**). Le groupe 4 (STIC uniquement) contenait 36 protéines impliquées dans la migration cellulaire, telles que CFL1 (Cofilin-1), PFN1 (profiline-1), PPL (periplakine), FLNB (filamine B), PDLM1 (protéine à domaine PDZ et LIM 1), IQGAP1 (protéine similaire à la GTPase Ras), MYH14 (Myosine-14), ainsi que dans le métabolisme (CS [Citrate synthase], DECR1 [2,4-diénoyl-CoA réductase], GSR [Glutathion réductase], LAP3 [Aminopeptidase cytosolique], GLUD1 [Glutamate déshydrogénase 1]). Les protéines de ce groupe étaient impliquées dans la différenciation cellulaire, la prolifération, ainsi que dans la néoplasie/métastase (**Figure 21B,c**). Le dernier groupe (normal-p53/STIC) était composé de 61 protéines. Plusieurs d'entre-elles ont été identifiées et sont impliquées dans les exosomes (annexines A1, A2, A4, A13, EZR [ezrine], HSPA8 [Heat shock 71kDa protein 1A]). D'autres protéines sont liées au stress, comme PRDX1 (Peroxyrédoxine-1), HSP8, HSPA1, HSPA2, ST13 (Hsc70-interacting protein), DNAJB6 (DnaJ homolog subfamily B member 6), VCP (Transitional endoplasmic reticulum ATPase), TUBB4B (chaîne bêta-4B de la tubuline). Ce groupe présentait une surreprésentation de protéines impliquées dans la néoplasie et les transitions cancéreuses (**Figure 21B,d**). L'analyse d'enrichissement fonctionnel souligne des différences entre les groupes. Différents processus biologiques ont été identifiés, tels que la réponse immunitaire, la régulation de la réponse immunitaire, la croissance et/ou le maintien cellulaire, le métabolisme, les voies énergétiques, la transduction de signaux et le métabolisme des protéines (**Figure 22A**). Des facteurs de transcription enrichis (FT) régulant les protéines surexprimées dans les différents groupes ont

également été obtenus (**Figure 22B**). Le groupe contenant les protéines surexprimées dans la signature normale-p53 a montré une forte abondance de protéines impliquées dans le métabolisme et les voies énergétiques, ainsi qu'un enrichissement des facteurs de transcription SP1, SP4 et TEAD1 qui ciblaient la plupart des gènes de ce groupe.



**Figure 21:** Analyse protéomique des lésions de la fimbria.

A) Classification hiérarchique des protéines les plus variables entre le tissu normal, la signature p53, STIL et STIC ( $n=4$  pour chaque catégorie, ANOVA avec un FDR basé sur les permutations  $< 0.05$ ) ; B) Une Analyse de Sous-Réseau Enrichi a été réalisée pour mettre en évidence les voies biologiques et fonctionnelles altérées dans les différents groupes de protéines.

Pour le groupe normal-p53/STIL, les protéines étaient principalement impliquées dans la croissance et/ou le maintien cellulaire ainsi que le métabolisme des protéines. Un enrichissement des facteurs de transcription NFIC et EGR1 est observé. En ce qui concerne Normal-p53/STIC, les protéines identifiées sont impliquées dans la croissance et/ou le maintien cellulaire, le métabolisme, les voies énergétiques et le métabolisme des protéines,

tandis que la protéine KLF7 a été identifiée comme facteur de transcription. Pour le groupe STIL/STIC, un enrichissement des facteurs de transcription ZEB1 et ETS1 est observé.

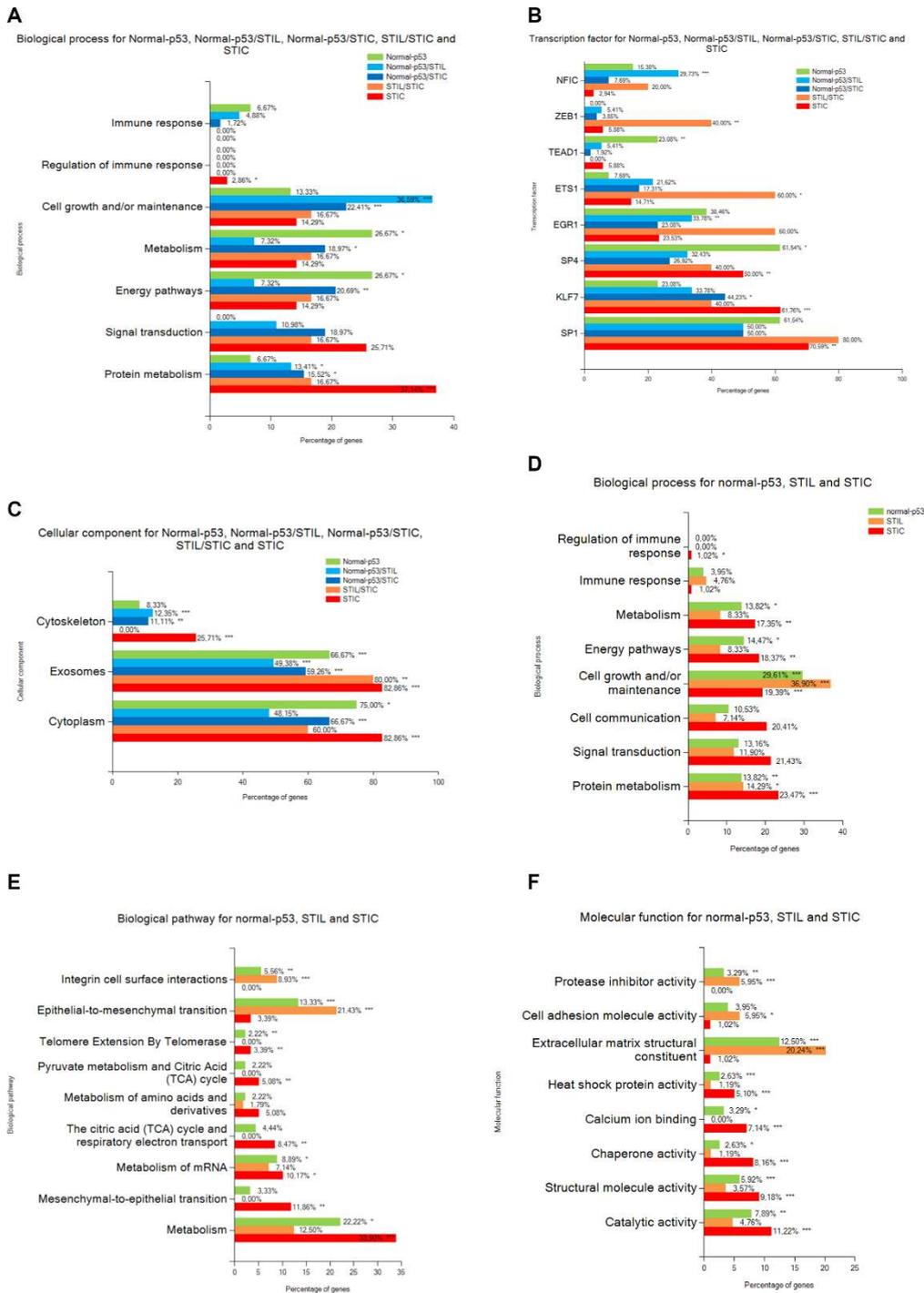
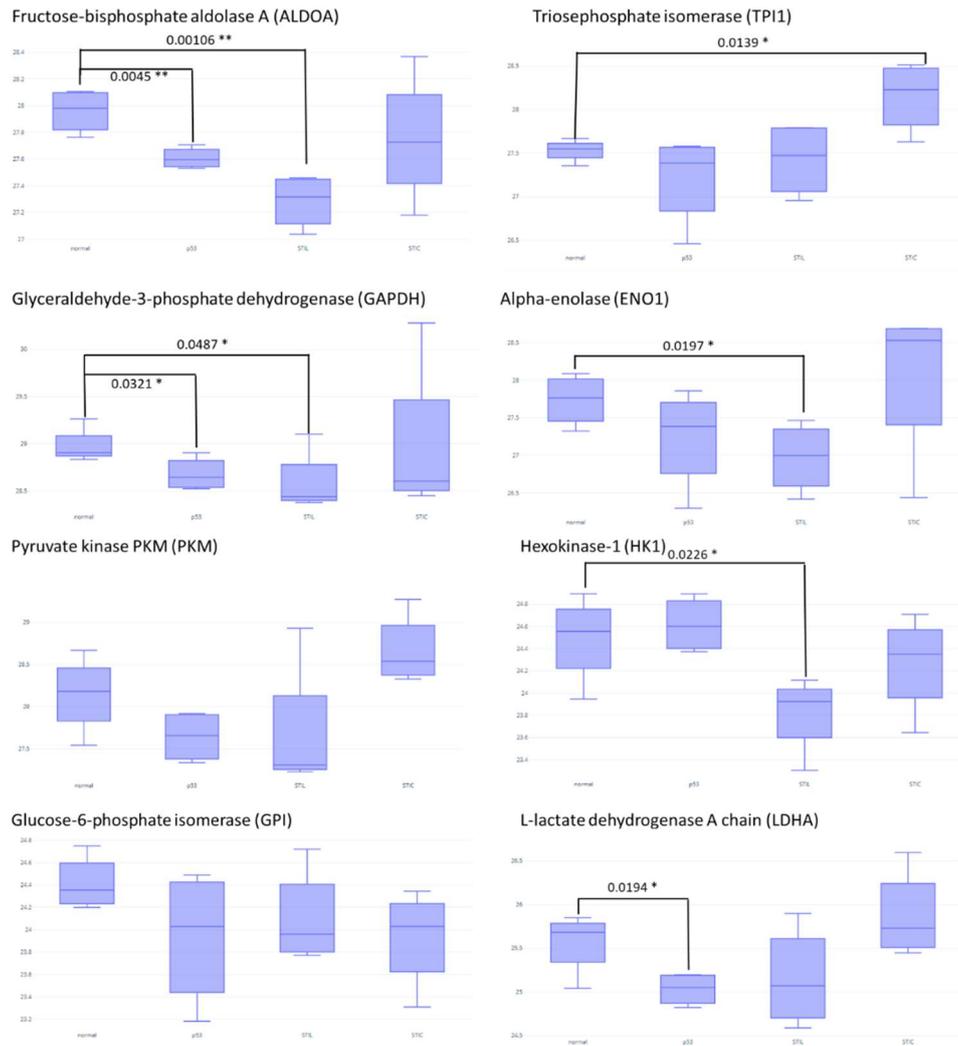


Figure 22: Analyse des voies de signalisation retrouvées dans les protéines différenciantes.

A) Processus biologiques, B) Facteur de transcription et C) Composant cellulaire pour les groupes de protéines. D) Processus biologiques, E) Voies biologiques et F) Fonction moléculaire des protéines surexprimées dans chaque lésion. (Test hypergéométrique par rapport à l'ensemble des gènes/protéines annotés de la base de données Funrich, la valeur p est représentée par des étoiles : \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$  pas d'étoile pour  $p > 0.05$ ).

Le groupe STIC présente un fort enrichissement en protéines impliquées dans le métabolisme des protéines et la régulation de la réponse immunitaire, avec les facteurs de transcription SP1, SP4 et KLF7. Nous observons également une réduction des protéines impliquées dans la réponse immunitaire dans les lésions STIC. L'analyse des composants cellulaires montre que toutes les protéines des 5 groupes se trouvent dans le cytoplasme ou dans les exosomes (**Figure 22C**), mais avec quelques différences. En effet, dans les groupes STIL-STIC et STIC, un fort enrichissement en exosomes est observé, tandis que la signature p53 et les STIC impliquent des protéines cytoplasmiques. Un enrichissement élevé en protéines du cytosquelette est également spécifique aux lésions STIC. Dans l'ensemble, ces analyses des transitions moléculaires entre la signature normale-p53, la signature p53 vers STIL et STIL vers STIC, avec un changement dans les protéines impliquées dans la croissance et/ou le maintien cellulaire ainsi que dans les différents processus métaboliques. Pour confirmer ces observations, d'autres analyses ont été réalisées, regroupant cette fois toutes les protéines surexprimées dans chaque lésion, c'est-à-dire la signature normale-p53, STIL et STIC (**Figure 22D, E et F**).

Dans les lésions STIL, un enrichissement en protéines impliquées dans la croissance et/ou le maintien cellulaire (**Figure 22D**), dans les interactions de surface cellulaire des intégrines (**Figure 22E**), ainsi que dans les constituants structurels de la matrice extracellulaire (ECM) et l'activité des molécules d'adhésion cellulaire (**Figure 22F**) est retrouvé. Les protéines impliquées dans divers processus métaboliques et voies énergétiques (**Figure 22D et E**) ainsi que l'activité catalytique (**Figure 22F**) étaient moins nombreuses que pour les autres transitions cellulaires. Nous avons également voulu déterminer si les protéines impliquées dans l'effet Warburg<sup>167</sup> étaient activées pour compenser la diminution globale des processus métaboliques classiques observée (**Figure 23**).



**Figure 23:** Visualisation des niveaux de protéines dans les différentes lésions pour les protéines connues pour être impliquées dans l'effet Warburg.

Les valeurs de Log(LFQ) ont été utilisées pour l'expression de la Fructose-bisphosphate aldolase A (ALDOA), la Triosephosphate isomérase (TPI1), la Glycerinaldehyde-3-phosphate dehydrogenase (GAPDH), l'Alpha-enolase (ENO1), la Pyruvate kinase PKM (PKM), l'Hexokinase-1 (HK1), la Glucose-6-phosphate isomerase (GPI) et la Chaîne A de la L-lactate déshydrogénase (LDHA). Un test t a été utilisé pour comparer les valeurs entre le tissu normal et chaque lésion. La valeur p est représentée par des étoiles (\*\*\*)  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , pas d'étoile pour  $p > 0.05$ .

Dans l'ensemble, l'expression de ces protéines n'est pas altérée et une tendance à la diminution suggère qu'aucun effet Warburg ne semble se produire. Dans les lésions STIC, les protéines impliquées dans le métabolisme, les voies énergétiques, la communication cellulaire, la transduction des signaux (**Figure 22D et E**), la liaison aux ions calcium, l'activité catalytique et les molécules structurales (**Figure 22F**) sont plus abondantes que dans les autres lésions de transition. Parallèlement, la réponse immunitaire (**Figure 22D**), les interactions des intégrines (**Figure 22E**), l'activité des molécules d'adhésion cellulaire et les constituants structurels de la matrice extracellulaire (**Figure 22F**) sont fortement réprimés.

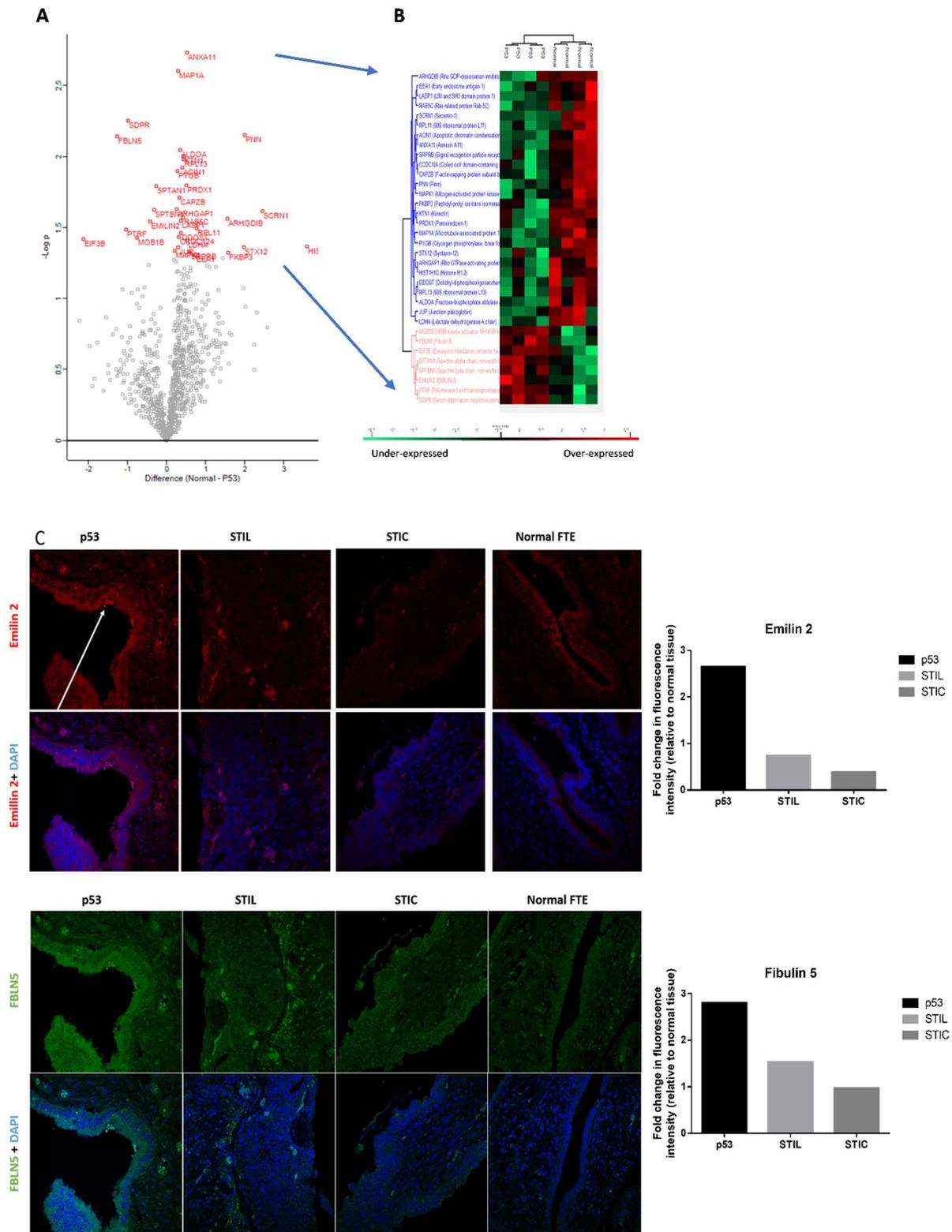
La signature p53 et le tissu normal sont difficiles à distinguer à l'aide de l'analyse de groupe. Seulement, grâce à un test t de Student, huit protéines ont été trouvées fortement

discriminantes entre les tissus normaux et les lésions p53, avec une surexpression spécifique dans la signature p53 (**Figure 24A, 24B**). Le **Tableau 5** présente l'immunoréactivité dans l'épithélium tubaire normal (FTE) et l'effet pronostique sur la survie globale (OS) dans le cancer de l'ovaire pour ces huit protéines extraites du HPA. PTRF, également connue sous le nom de CAVIN1, est un marqueur de pronostic défavorable pour les cancers de l'ovaire, de l'urothélium et du côlon, tandis que SDPR (connue sous le nom de CAVIN2) est associée à un pronostic favorable dans le cancer rénal et à un pronostic défavorable dans le cancer de l'estomac. La validation des données d'immunohistochimie (IHC) a indiqué que ces protéines ne sont pas détectées dans l'épithélium tubaire normal. Elle a également montré une positivité modérée dans le cancer de l'ovaire pour CAVIN1 et une positivité forte dans un cas rare de carcinome endométrioïde de l'ovaire pour CAVIN2. SPTAN1 et FBLN5 sont des marqueurs de pronostic favorable dans le cancer rénal mais défavorables dans le cancer de l'ovaire (pour un seuil de  $p < 0.05$ ). Emilin2 est déjà connue comme un marqueur de pronostic défavorable pour le cancer du foie et de pronostic défavorable pour les cancers du rein et de la tête et du cou.

**Tableau 5:** Immunoréactivité et effet pronostique des protéines surexprimées dans la signature p53 par rapport au tissu normal, selon Protein Atlas.

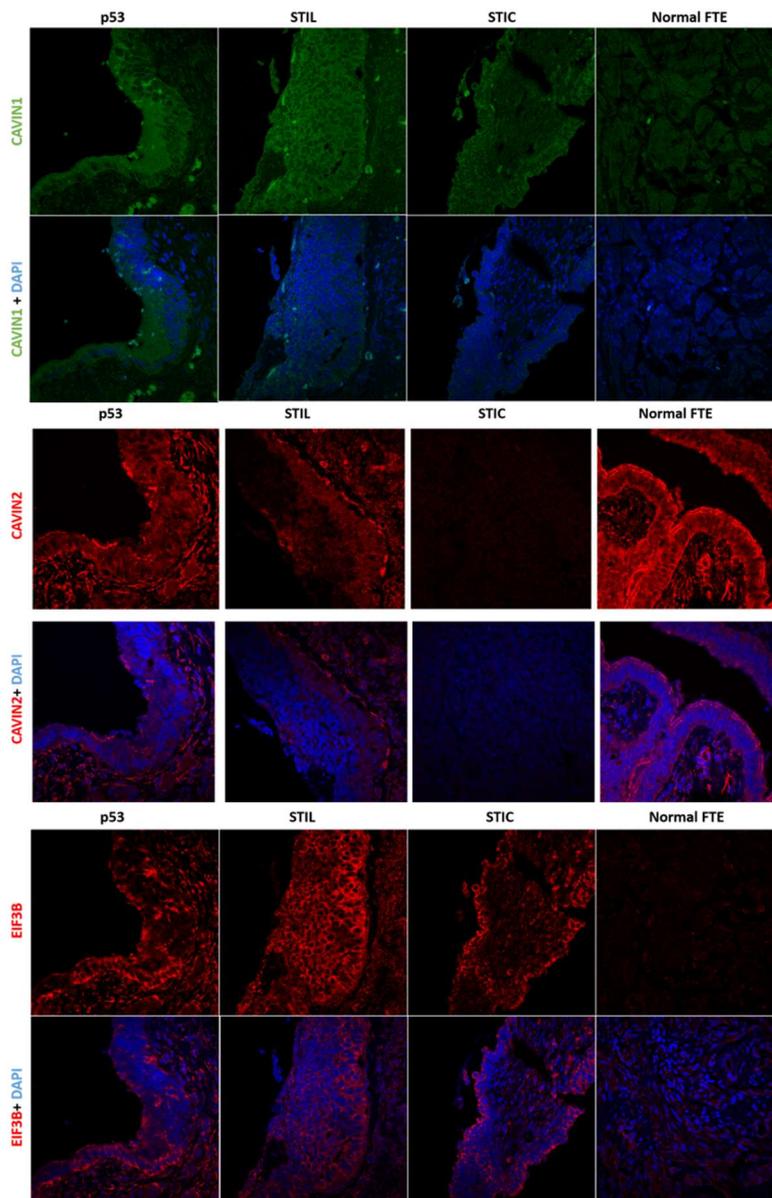
Protéine	Immunoréactivité dans le FTE	Effet pronostic dans l'OC
CAVIN1	Pas détecté	Pas favorable ***
SPTAN1	Positivité modérée à forte	Pas favorable *
FBLN5	Pas détecté	Pas favorable *
EIF3B	Positivité forte	Non significatif
MOB1B	Positivité modérée	Non significatif
EMILIN2	Positivité modérée	Non significatif
SPTBN1	Positivité modérée à forte	Non significatif
CAVIN2	Pas détecté	Non significatif

La coloration immunohistochimique des protéines dans le tissu normal de l'épithélium tubaire normal (FTE) et l'effet pronostique d'une expression élevée du gène correspondant dans le cancer de l'ovaire (OVC) ont été extraites du Human Protein Atlas (crédit des données : Human Protein Atlas disponible sur [v19.proteinatlas.org](http://v19.proteinatlas.org), la valeur  $p$  est représentée par des étoiles \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$  et pas d'étoile pour  $p > 0.05$ ).



**Figure 24:** Comparaison entre la signature p53 et le tissu normal.

(A) Volcano plot comparant le tissu normal de la lésion p53 par un test t (les protéines avec une valeur de  $p < 0.05$  sont en rouge). (B) Classification hiérarchique des protéines significativement exprimées entre ces deux conditions. (C) IHC du marquage anti-FBLN5 et anti-Emilin2 des lésions p53, STIL et STIC. Les flèches indiquent les lésions p53. Le graphique représente la quantification de l'intensité de fluorescence de chaque marqueur comparé à l'intensité de fluorescence retrouvé dans le tissu normal.



**Figure 25:** Comparaison des lésions préneoplasique et du tissu sain par IHC.

IHC du marquage anti-CAVIN1, anti-CAVIN2 et anti-EIF3B des lésions p53, STIL et STIC. Les flèches indique les lésions p53. En vert et rouge correspond le marquage tandis que le DAPI marque les noyaux (bleu).

Par immunohistochimie (IHC), nous avons validé une expression plus élevée de CAVIN1 dans les lésions p53 et STIL par rapport à l'épithélium tubaire normal (FTE). FBLN-5 et Emilin-2 ont été trouvés surexprimées dans les lésions p53 et, dans une moindre mesure, dans les lésions STIL. Ces protéines n'ont pas été détectées dans les lésions STIC par rapport à l'épithélium tubaire normal (FTE). SPTAN1 est surexprimée dans certaines cellules des lésions p53 et STIL. CAVIN2 est absent des lésions STIC (**Figure 24C, figure 25**).

*Etude lipidomique par analyse SpiderMass*

La première difficulté de cette étude a été d'extraire le contenu lipidique des tissus conservés en bloc FFPE sur de longues périodes. En effet, le protocole analytique a dû être adapté pour optimiser l'intensité du spectre obtenu par le SpiderMass. Ainsi, dans un premier temps, plusieurs méthodes de préparation ont été testées i.e., analyse des tissus sans déparaffinage avec déparaffinage mais sans réhydratation ou avec déparaffinage et réhydratation. La méthode s'avérant être la plus efficace a été de déparaffiner 2 fois les lames par des bains de 5min au xylène, suivi d'un spray de glycérol/IPA (2:8) sur celles-ci. Cette technique a permis d'obtenir un spectre suffisamment intense, tout en retirant les signaux parasites relargués par la paraffine. Contrairement à l'analyse protéomique, cette analyse a été réalisée sur des lames miroirs de celles utilisées pour l'immunohistochimie. La deuxième difficulté consistait à réaliser une ablation laser uniquement sur ces lésions de faible cellularité sans toucher le tissu sain environnant. Pour vérifier la qualité des spectres obtenus, sur chaque tissu présentant une lésion p53, 3 tirs ont été réalisés sur l'épithélium tubaire éloigné de cette lésion. Ces derniers ont été considérés comme normal dans notre modèle de classification.

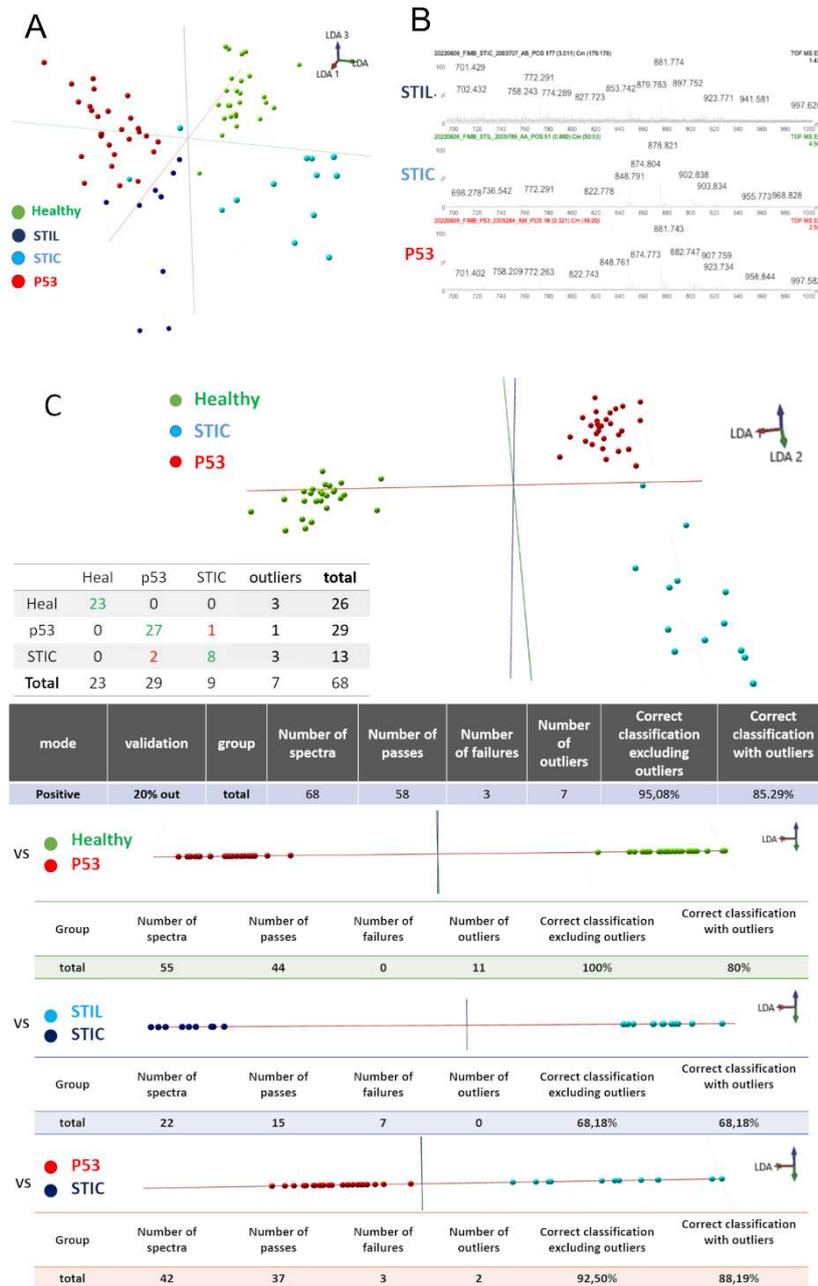


Figure 26: différencier les lésions préneoplasiques.

) Spectres de masse obtenus par SpiderMass à partir des lésions préneoplasiques p53, STIL et STIC dans la fimbria. b) Le modèle de classification PCA-LDA construit à partir des 3 stades de lésions préneoplasiques ; p53, STIL et STIC c) Représentation LDA de la PCA-LDA à 3 classes (à droite). Le tableau représente les résultats de la validation croisée "20%-out" du modèle de classification construit.

L'acquisition des spectres MS de cette cohorte de 27 patients a été réalisée en mode ion positif. Les spectres obtenus sont proches, mais présentent des différences dans leur contenu moléculaire (Figure 26B). Par la suite, une analyse ACP des spectres générés à partir de tissus sains, des lésions p53, STILs et STICs, a été réalisée. Cette ACP a ensuite été soumise à une analyse supervisée à l'aide d'une analyse discriminante linéaire (LDA). Ce modèle de classification a permis la différenciation de l'ensemble des lésions et du tissu sain (figure 26A). D'après la figure 26A, l'analyse discriminante linéaire 1 a permis de distinguer

le groupe sain des groupes p53 et STIC. De manière intéressante la signature p53 est plus proche du groupe STIL que du groupe sain provenant des mêmes échantillons. Ce résultat s'oppose à ceux obtenus en protéomique qui peinait à différencier la signature p53 du tissu sain. Cette différenciation de la lésion p53 du tissu sain a également été observé lorsque nous avons comparé les spectres. En effet, les spectres de la signature p53 et de STIL sont beaucoup plus proches que ceux de STIC (**Figure 26B**). Pour vérifier la robustesse du modèle, une méthode de classification croisée basée sur le « 20% out » a été réalisée. Celle-ci consiste à retirer 20% des spectres du modèle et d'essayer de reclasser ceux-ci à partir du modèle construit avec les 80% restant. Cette méthode est réalisée avec 5 itérations afin de vérifier l'ensemble du modèle. Malheureusement, la classe STIL est sous-représentée par rapport au reste des tissus. Ce manque d'homogénéité au sein des échantillons entraîne des taux de classification de 83% sans valeurs aberrantes (outliers) et 77% avec outliers. Cette validation est correcte mais insuffisante pour une prédiction avec certitude. Ainsi, un autre modèle a été généré en excluant les lésions STILs. Les résultats de validation croisée obtenus à partir de la signature p53, des STICs et des tissus sains à l'aide de la méthode "20% out" ont cette fois démontrés d'excellents taux de classification avec 95,08% et 85,29% sans et avec outliers respectivement. Ces résultats ont mis en avant la capacité de la technologie à prédire la présence de ces lésions de faible cellularité et de les dissocier du tissu sain environnant.

Pour identifier de potentiels marqueurs spécifiques de ces lésions, des modèles PC-LDA comparant 2 à 2 les classes ont été réalisés. Une excellente classification entre les lésions p53 et les tissus sains, avec un taux de 100% avec outliers et 80% sans outliers, a démontré une discrimination significative de ces deux classes provenant de la même section de tissu. Nous avons également observé une séparation correcte de 75 % avec outliers et 72,93 % sans outliers entre p53 et STIL. Enfin, un excellent taux de classification, de 92,5 % et 88,1 % avec et sans outliers, a été obtenu entre p53 et STIC (**figure 26C**). Ces différents modèles ont permis d'isoler certains ions discriminants en fonctions des classes. L'identification de ces ions ( $m/z$ ) correspondant à des lipides, a été réalisée en analysant de nouveau les lames en mode MS<sup>2</sup>. Leur intensité normalisée est représentée sous forme de boxplot pour chaque type de lésion (**figure 27**). Nous avons ainsi identifié des marqueurs lipidiques spécifiques des différentes lésions tels que le TG (16:0/18:1/20:4) ( $m/z$  881.7) ou le PE (P-18:0/22:4) ( $m/z$  780.5), sous-exprimés dans les tissus sains. En plus de ces deux lipides, la surexpression du PC (P-16:0/22:6) ( $m/z$  790.5) et du DG (37:6) ( $m/z$  627.5) dans les STICs permet de différencier les phénotypes sains et lésés. Il est intéressant de noter que nous avons distingué les lésions p53 des tissus sains grâce à la sous-représentation des PS (22:1/20:1) ( $m/z$  874,9) et PS (22:0/20:0) ( $m/z$  876,9). Enfin, nous avons distingué les STIL

des STIC par le PG(O-16:0/20:1) ( $m/z$  763.5), sous-représenté dans les STIL par rapport aux autres phénotypes.

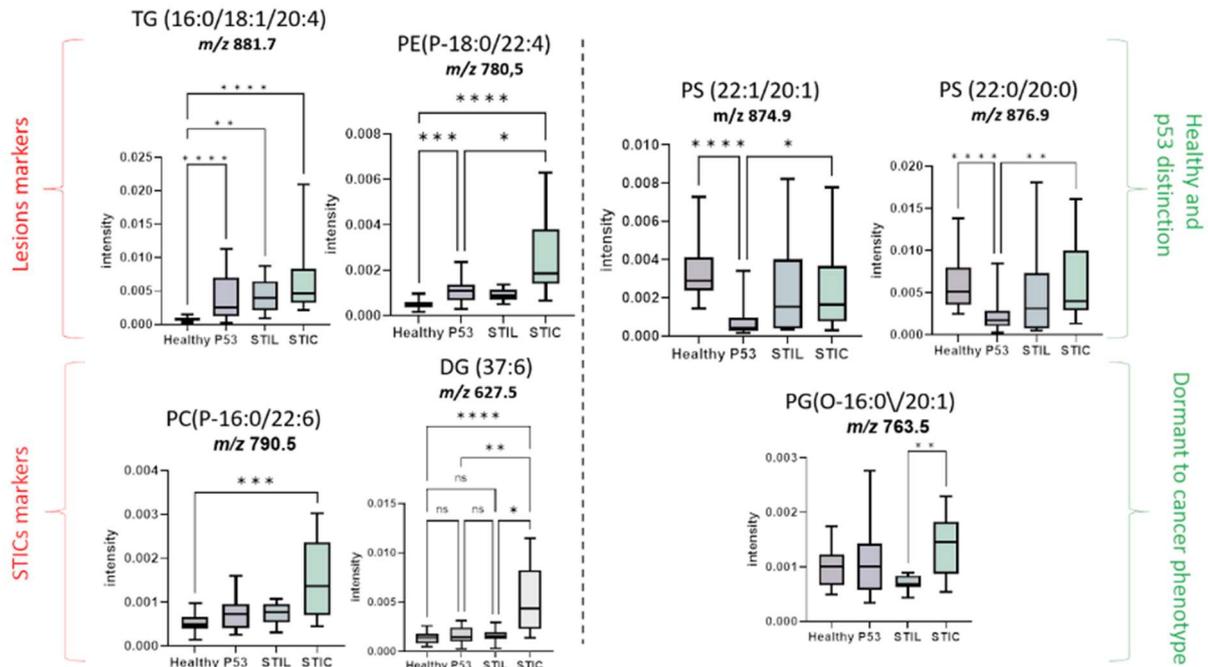


Figure 27: Lipides spécifiques identifiés pour chaque lésion.

Boxplots des intensités relatives des marqueurs spécifiques identifiés par MS<sup>2</sup> au SpiderMass en fonction du type de lésion. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) no star for  $p > 0.05$ ).

## Discussion

L'objectif principal de cette étude était de définir les mécanismes associés à l'apparition des lésions pré-néoplasiques des cancers ovariens sévères de haut grade. En effet, la signature p53 semble être à l'origine des HGSOV, mais elle reste très difficile à diagnostiquer par de simples marqueurs morphologiques. La découverte de nouveaux marqueurs associés aux lésions pré-néoplasiques pourrait faciliter la détection de ces lésions et donc d'anticiper l'apparition du HGSOV à travers des mesures préventives. Cela est d'autant plus impactant chez les patientes BRCA mutés dont les risques de développer un cancer sont plus importants.

Pour répondre à cet objectif, nous avons initié une étude multi-omics de ces lésions. Dans un premier temps, le protocole a dû être optimisé pour analyser ces lésions directement sur lame IHC. Ainsi, les analyses IHC et anatomopathologiques des coupes de tissu ont permis de diriger les étapes de micro-digestion sur des zones de lésions précises. Cette étude protéomique a été réalisée sur un nombre relativement restreint de patientes car ces lésions pré-néoplasiques. Cette contrainte est liée à la rareté des patientes présentant à la fois une mutation BRCA, un début de lésion et ayant subi une annexectomie prophylactique avec un protocole SEE-FIM. En effet, celui-ci demande de couper rigoureusement les tissus réséqués

des trompes de Fallope et de chercher minutieusement les potentielles lésions cellulaires. Ainsi, nous avons favorisé une quantité limitée d'échantillons ayant des caractéristiques cliniques similaires afin d'améliorer la robustesse des résultats. De plus, nous nous sommes focalisés sur la limitation des pertes d'échantillon en analysant directement les lames annotées par le pathologiste. Cela a également pu garantir la fiabilité des résultats puisque la localisation de ces lésions étaient parfaitement annotées. Ce choix a conduit à la distinction de protéines significativement discriminantes qui pourraient être validé sur des cohortes plus importantes. Cette analyse protéomique a été complétée par une analyse lipidomique par SpiderMass sur une cohorte plus importante de 27 patientes en sélectionnant parmi l'ensemble de la cohorte nationale du PHRC Fimbria, les lésions P53, les STIL, les STIC.

Les analyses protéomiques ont permis l'identification de 10 375 séquences peptidiques uniques correspondant à 1 617 protéines distinctes et la découverte de changement d'expressions protéiques dans ces lésions. De manière intéressante, en comparant le protéome global de ces lésions, il est possible de retrouver la chronologie de développement du HGSOE. En effet, la signature p53 est très proche du tissu sain et possède des marqueurs communs avec la lésion STIL. La lésion STIC se distingue des autres lésions et présente un profil se rapprochant de celui du cancer séreux de haut grade.

Des marqueurs spécifiques de chaque lésion ont été identifiés. La RNA-binding protein 10 (RBM10) est retrouvée spécifiquement dans la lésion p53 tandis que 3 protéines spécifiques sont retrouvées pour la lésion STIC telles que la glucosamine (N-acetyl)-6-sulfatase (GNS), l'ARN polymérase I (UBTF) et des protéines de la famille des ATPases (ATAD3A/B). Certaines de ces protéines, telles que CAVIN1, SPTAN1 et FBLN5, sont liées à un mauvais pronostic dans le cancer de l'ovaire. La différence d'expression de ces dernières a notamment pu être validée par immunofluorescence qui montre une nette surexpression dans cette région contrairement à un épithélium tubaire sain. Bien que l'implication de certaines protéines dans le pronostic du cancer de l'ovaire ne soit pas encore décrite, celles-ci sont liés à différents pronostics dans d'autres type de cancers. La protéine EIF3B, semble être de mauvais pronostics pour les patients atteints de cancers rénaux. Pour finir, les protéines SPTBN1 et CAVIN 2 ont également été retrouvées spécifiques de la signature p53 et semblent être, contrairement aux précédentes, liés à un bon pronostic, comme cela a été mis en évidence dans les cancers rénaux. Cette analyse globale a également permis de déterminer les différents mécanismes intervenant dans ce processus de lésion tels que la réponse immunitaire, l'activation de la traduction et la régulation du trafic à travers les protéines CAVINs. De plus, la modification de la matrice extracellulaire semble aussi être impactée par la surexpression des protéines Emilines, ce qui corrèle d'autres études évoquant le rôle des fibroblastes et de la matrice extracellulaire dans la carcinogenèse de l'ovaire<sup>168</sup>.

Les protéines surexprimées dans la lésion STIL sont liées principalement à la maintenance et non pas à la prolifération cellulaire. Une réduction de l'activité métabolique est observée dans cette lésion, en particulier du cycle de Krebs. Ce cycle est primordial dans la synthèse d'acétyl-coA et donc d'énergie. Cependant, aucun effet Warburg, qui pourtant distingue les cellules cancéreuses<sup>169</sup>, n'est observé dans ces lésions. Pour finir, la croissance cellulaire est régulée négativement. Ainsi, l'ensemble de ces résultats semblent mettre en évidence un phénotype « dormant » de cette lésion et donc de STIC en devenir.

Contrairement au profil dormant des STILs, les STICs montrent une dérégulation très agressive dans leur profil protéique. Une sous-expression importante des protéines participant au processus d'adhésion cellulaire et à la structuration de la matrice extracellulaire est observée. Ces lésions présentent également des caractéristiques communes aux cellules cancéreuses telles que la surexpression de protéines impliquées dans les voies énergétiques ou dans l'activité des télomères et la maturation des centrosomes. Ces processus sont en lien direct avec la tumorigenèse<sup>170</sup>. La lésion STIC est supposée comme étant à l'origine du HGSOc, cela pourrait refléter le début du processus d'invasion vers l'ovaire<sup>169</sup>. De plus, une surexpression de protéines impliquées dans la sécrétion d'exosomes a été détectée. Ces résultats suggèrent suggère une communication cellulaire qui pourrait être en lien avec l'invasion d'organe distant tel que l'ovaire.

Les protéines retrouvées différenciantes des lésions ont un lien avec les voies de signalisation lipidiques. Par exemple, les CAVIN1 et CAVIN2 sont des protéines spécifiques des cavéoles, un radeau lipidique particulier. Une étude lipidomique pourrait permettre un degré de compréhension supplémentaire des mécanismes cellulaires en jeu dans ces lésions. Par ailleurs, le SpiderMass a l'avantage majeure de pouvoir identifier rapidement une lésion sans marquage préalable par IHC. C'est pourquoi l'analyse protéomique a été complétée par une étude lipidomique *ex vivo* sur des tissus FFPE issus d'une cohorte de 27 patientes.

La fimbria étant un tissu peu gras, le signal lipidique contenu dans les tissus était masqué par la présence de paraffine. Ainsi, dans un premier temps, le protocole analytique a été optimisé pour obtenir un signal moléculaire exploitable. Une étape de déparaffinage au xylène a été ajoutée suivie d'un spray de glycérol/IPA (2:8) garantissant un spectre suffisamment intense et des résultats robustes et fiables en mode ion positif. Grâce à ce protocole, nous avons pu démontrer l'efficacité du SpiderMass dans la détection de changements lipidiques très localisés en créant un modèle de classification capable de discriminer toutes les lésions du tissu sain. Un modèle global basé sur l'analyse statistique multivariée PC-LDA a été réalisé comprenant l'ensemble de ces lésions avec un bon score de classification de 83%. Ce modèle nous a permis de mettre en évidence les différences de profils moléculaires entre ces lésions

pré-néoplasiques et le tissu sain environnant. Ces différences ont donné lieu à une classification qui a démontré que le profil moléculaire de la lésion p53 est bien plus proche de la lésion STIL que du tissu sain présent autour de cette lésion et présentent des caractéristiques communes à un phénotype de STIC en dormance. Ainsi, cette dizaine de cellules formant la lésion semblent se rapprocher d'une transformation néoplasique. Cette conclusion est intéressante mais provient d'un modèle dont la sensibilité pourrait être améliorée. En effet, la classe STIL est sous représentée dans cette cohorte, ce qui a conduit à la baisse de sensibilité du modèle. Pour améliorer notre classification et augmenter la précision et la robustesse de nos conclusions, un second modèle a été créé. Ce modèle ne se base que sur les lésions p53, STICs et le tissu sain. Ainsi, ce modèle présente un taux de validation plus élevé, de 95% sans et 85% avec outliers. De plus, l'ensemble de ces lésions semblent se classer en fonction de la gravité de celles-ci, laissant penser que la lésion p53 évolue vers un profil dormant STIL qui évolue progressivement en un phénotype cancéreux (STIC).

Grâce à la technologie SpiderMass, nous avons démontré qu'il est possible de distinguer les tissus sains des lésions pré-néoplasiques sans marquage préalable. De plus, un profil évolutif de ces lésions a été mis en évidence. Après nous être intéressés au profil moléculaire globale de chacune des lésions, nous avons souhaité déterminer de potentiels marqueurs lipidiques. Ces marqueurs donnent un degré supplémentaire de compréhension de ces lésions et des voies de signalisations associées à celles-ci. De plus, Ces marqueurs pourraient apporter une aide au diagnostic de celles-ci et engendrer des mesures préventives pour ralentir la progression vers un cancer de l'ovaire.

Dans la recherche de marqueurs spécifiques de ces lésions, nous avons identifié deux phosphatidylsérines PS(22:1/20:1) et PS(22:0/20:0) en tant que marqueurs discriminants entre la signature p53 et les tissus sains, tandis que les PE(P-18:0/22:4) et les TG (16:0/18:1/20:4) ont été trouvés surexprimés dans l'ensemble des lésions. De plus, nous avons spécifiquement observé la présence de phosphatidylcholine (P-16:0/22:6) et de diglycéride DG (37:6) dans les STICs, permettant leur identification distincte. Il est important de noter que la surexpression du phosphatidylglycérol PG (O-16:0/20:1) dans la STIC a permis la distinction entre ces derniers et les STILs. Dans le carcinome ovarien à cellules claires (CCC), étudié à partir de tissus FFPE, le nombre d'espèces lipidiques insaturées augmente tandis que le nombre d'espèces lipidiques saturées diminue par rapport aux tissus témoins. Les PE (32:1), PE (34:1), PI (34:1) et PS (36:1) présentent une signature commune dans les CCC<sup>171</sup>. Dans le cancer ovarien sévère, la surexpression de la protéine fatty acid binding protein 4 entraîne une augmentation des glycérolipides, des glycérophosphoéthanolamines et des glycérophosphoinositols tels que LysoPE, LysoPG et LysoPI<sup>172</sup>. Les cellules cancéreuses dépendent de la lipogenèse pour s'adapter au stress cytotoxique dans le microenvironnement

tumoral. Cela est particulièrement important dans les zones tumorales où l'apport exogène en acides gras est limité, comme dans les régions hypovasculaires et hypoxiques. De plus, il est intéressant de noter la présence de plusieurs précurseurs d'acides gras (par exemple, 20:4, 22:4) issus des acides gras oméga-6. Ces acides gras polyinsaturés sont connus pour renforcer le processus carcinogène dans différents cancers<sup>173</sup>. Les acides gras alimentaires sont utilisés par les cellules cancéreuses comme source d'énergie par lipogenèse pour la biosynthèse membranaire, les processus de signalisation et l'inhibition des processus ROS.

Les observations que nous avons faites concernant la modulation lipidique sont étayées par la présence de plusieurs enzymes métaboliques dérégulées dans le cancer. Notamment, nous avons constaté que la PHGDH est surexprimée dans les lésions STILs et STICs, ce qui est cohérent avec sa forte expression dans d'autres types de cancer, notamment le cancer du sein, du côlon et de l'endomètre. De plus, il existe des preuves liant les niveaux de triglycérides à la méthylation de l'ADN de la PHGDH, suggérant un rôle régulateur potentiel de cette enzyme dans le métabolisme des triglycérides dans le cancer<sup>174,175</sup>. Notre étude a également identifié d'autres protéines liées au métabolisme lipidique surexprimées dans les STIC, telles que la MARCKS-related protein, la 2,4-diényl-CoA réductase<sup>176</sup>, la Trifunctional enzyme subunit beta et la 3-cétoacyl-CoA thiolase. En revanche, certaines protéines du métabolisme lipidique, comme l'Aldehyde dehydrogenase family 3 member B1 (ALDH3B1), impliquée dans l'oxydation des aldéhydes lipidiques à chaînes moyennes et longues de la membrane plasmique, se sont révélées sous-exprimées dans les STILs.

Une étude récente d'Ackerman et al démontre l'importance du stockage des acides gras saturés sous forme de triglycérides dans la prévention de la toxicité et le maintien de l'homéostasie dans un environnement hypoxique dans le carcinome à cellules rénales à cellules claires en libérant de l'oléate (18:1) qui sera métabolisé au sein des phospholipides<sup>177</sup>. De manière intéressante, le TG (16:0/18:1/20:4) a été trouvé discriminatif des lésions de la fimbria et encore plus abondant dans les STICs. Les STICs sont des lésions dotées d'une forte prolifération qui utilisent les triglycérides pour préserver leur homéostasie pendant leur croissance. De plus, il a été observé que les patients porteurs de mutations BRCA1 et BRCA2 présentent des niveaux élevés de triglycérides<sup>178</sup>. Cette abondance en triglycérides pourrait contribuer à la survie accrue des cellules lésionnelles en préservant l'homéostasie et en facilitant le développement des STICs et donc le potentiel développement d'HGSOC dans l'ovaire.

### Conclusion

En conclusion, cette étude a permis une meilleure compréhension des mécanismes impliqués dans les processus lésionnels de l'épithélium tubaire à l'origine du développement du cancer de l'ovaire. Nous avons décrit l'évolution chronologique de ces lésions allant de simple mutation p53 sur certaines cellules des trompes jusqu'à l'apparition d'un cancer capable de réduire son ancrage et envahir des tissus voisins tels que l'ovaire. De plus, nous avons mis en avant la capacité de la technologie SpiderMass à détecter des variations phénotypiques présentes sur une faible quantité de cellules à travers l'identification de marqueurs spécifiques de ces lésions. Cependant, la cohorte analysée était limitée du fait de la difficulté de détection de ces lésions sur des patientes BRCA muté. Ainsi, cette étude exploratoire doit-être complétée par une cohorte plus importante afin de consolider les résultats et vérifier la sensibilité de ces nouveaux marqueurs dans la détection de lésion précoces. Ces marqueurs pourront également aider dans le développement de stratégies thérapeutiques et améliorer la prise en charge des patientes.

## CHAPITRE IV : Diagnostic en temps réel et prédiction de l'évolution du cancer de l'ovaire par spectrométrie de masse SpiderMass

Après m'être focalisé sur l'origine du cancer séreux de haut grade, j'ai entrepris l'analyse par SpiderMass de différents sous-types de cancer de l'ovaire. L'intérêt principal était de bâtir un modèle de classification capable de diagnostiquer le plus efficacement possible cette pathologie au bloc opératoire en temps-réel. Cette étude a également pour objectif d'améliorer la compréhension de l'environnement tumoral et de son importance dans le diagnostic et pronostic. Ces travaux pourront apporter au clinicien une aide à la décision pour la prise en charge des patients après chirurgie.

### Introduction

Les cancers de l'ovaire, et surtout le HGSOC, sont souvent détectés à des stades FIGO tardifs (III/IV). Ceci est dû au manque de marqueurs diagnostiques précoces et de symptômes spécifiques<sup>25</sup>. Ces stades sont associés à des carcinoses multiples au sein de l'ensemble de l'abdomen, notamment au niveau du péritoine et, pour des stades très avancés, à des métastases distantes (foie, rate, cerveau). Ces métastases engendrent, malheureusement, de très mauvais pronostics, avec une mortalité de 70% à 10 ans<sup>45</sup>. De plus, la survie de ces patientes est très intimement liée à la capacité du chirurgien à réaliser une résection complète, sans résidu, de la masse tumorale<sup>42</sup>. Cette procédure est relativement complexe, notamment à cause du manque de données dont il dispose durant l'acte chirurgical pour identifier le caractère malin ou non du tissu<sup>9</sup>. Ainsi, il est souvent contraint de réséquer une grande partie de l'appareil reproducteur, affectant la qualité de vie du patient post-chirurgie. De plus, le protocole opératoire diffère selon le sous-type cancéreux. En effet, les tissus à réséquer dépendent du phénotype tumoral<sup>44</sup>. L'utilisation du SpiderMass sur une cohorte de patientes peut permettre l'élaboration d'un modèle de classification, basé sur le contenu lipidique des tissus, qui faciliterait le diagnostic. En comparant le spectre d'un tissu avec ce modèle de classification, et ce, en temps réel, il est possible de définir le type de cancer analysé. Le chirurgien peut ainsi être guidé et adapter son geste en temps réel<sup>17</sup>.

Seulement, la plupart des études actuelles de spectrométrie de masse ambiante sur le cancer de l'ovaire ne prennent pas en compte le micro-environnement tumoral pour le diagnostic de la pathologie<sup>13,15</sup>. Ce micro-environnement tumoral, notamment les cellules immunitaires, jouent un rôle clef dans l'évolution de la maladie. En effet, la présence de macrophages de phénotype M1-like de lymphocytes CD8+ et de natural killers est associée à un bon pronostic. Au contraire, les macrophages M2-like sont connus pour être impliqués dans la croissance

tumorale et l'immunosuppression. Des études précédentes, comparant des macrophages PC1/3 KD et des macrophages Non targeted (NT) ShRNA<sup>18</sup>, ont démontré la possibilité de différencier des macrophages ayant des phénotypes différents par SpiderMass. Ainsi, il serait possible d'associer des marqueurs lipidiques à différents types cellulaires grâce à la technologie SpiderMass. Le ratio entre les macrophages M1 et les macrophages M2, est un des facteurs pronostic du cancer de l'ovaire<sup>139,145,146</sup>. De plus, ce ratio peut être associé à la survie du patient afin de détecter rapidement des patientes prédisposées à une récurrence précoce et donc de permettre aux cliniciens de leur proposer de thérapies différenciées.

### Objectifs

L'objectif principal de cette étude est la description précise des différences moléculaires au sein des sous-types de cancer de l'ovaire. Chaque tissu ayant son propre profil moléculaire, il sera possible de les différencier par des approches de machine learning. Un diagnostic pourra être posé suite à l'interrogation de ces modèles lors d'une analyse en temps réel. Dans un second temps, la création de ces modèles d'apprentissage a permis la distinction d'espèces lipidiques spécifiques de chacun des sous-types cancéreux menant à une meilleure compréhension des mécanismes intervenant dans le processus d'oncogénèse.

Nous avons également étudié le micro-environnement tumoral, et plus particulièrement les cellules immunitaires, telles que les macrophages (M1 et M2) et les lymphocytes (NK, CD4+ et CD8+) afin d'associer le taux d'infiltration immunitaire aux différences moléculaires présentes dans les différents sous-types tumoraux. Ainsi, des modèles de classifications basés uniquement sur ces types cellulaires (lignées cellulaires et cellules primaires) ont été réalisés pour identifier leur profil au sein d'un spectre moléculaire de tissus cancéreux. Ce modèle permet d'obtenir un score de présence de chacun des types cellulaires (lymphocytes, macrophages, cellules cancéreuses, cellules saines) à partir des pics discriminants retrouvés au préalable.

Pour finir, une analyse en imagerie par SpiderMass de chacun des sous-types cancéreux a été réalisée. Étant donné la corrélation entre les données obtenues en l'imagerie MALDI et le SpiderMass. Chaque pixel analysé a été interrogé par notre modèle de classification, conduisant à une image, non plus moléculaire, mais basée sur le score de présence de chacun des types cellulaires. On peut donc désormais prédire le taux de présence de ces types cellulaires au sein des tissus, et déterminer un score de probabilité d'infiltration immunitaire et phénotypique qui sera corrélé à la survie des patients. Afin de valider cette prédiction, chacune des images a été associée à des images obtenues par l'imagerie MALDI ciblée, basée sur la technologie Tag-Mass développée par le laboratoire PRISM<sup>179</sup>, encore appelée MALDI-IHC. Celle-ci est réalisée sur des marqueurs associés aux macrophages (CD68), au cancer (Ki67)

ou encore aux lymphocytes (CD8). Cette technique a permis d'obtenir la cross-validation des résultats et une image plus résolue du tissu (20µm de résolution spatiale). Ces résultats ont ensuite été associés au diagnostic et pronostic de chacun des types cancéreux. Dans ce contexte, l'imagerie basée sur le score immunitaire a été réalisée sur des patientes atteinte de HGSOC divisée en deux groupes distincts. L'un regroupe les patientes ayant une survie longue et l'autre, les patientes avec une survie globale courte. Ainsi, il a été possible de comparer le pronostic des patientes en prenant en compte leur score immunitaire.

## Matériels et méthodes

### *Prise en charge des échantillons*

#### **Echantillons de patientes**

En tant que centre de référence pour la prise en charge des tumeurs ovariennes rares, le Centre Oscar Lambret (COL) est à nouveau certifié par la Société européenne de gynécologie oncologique (ESGO) pour la chirurgie du cancer de l'ovaire avancé. Ainsi, des échantillons de tissus ont été prélevés sur des patientes par le COL à Lille, en France, dans le cadre d'une cohorte de tissu frais congelés (FF). Les patientes ont donné leur consentement éclairé par écrit avant de participer à l'essai. Pour protéger la vie privée des patientes, aucune information personnelle n'a été utilisée dans ces expériences, et un numéro aléatoire a été attribué à chaque échantillon. Une cohorte de 78 échantillons frais a été traitée entre septembre 2020 et janvier 2021, dont 41 ovaires normaux contre 37 OC comprenant 10 ovaires séreux borderline (SBL), 13 ovaires séreux de haut grade (HGSOC), 8 ovaires mucineux et 6 endométrioïdes. De plus, 79 échantillons congelés d'endomètre provenant du même centre ont été traités en septembre 2021 dont 42 normaux, 31 carcinomes endométrioïdes et 6 carcinomes séreux de haut grade. Une cohorte de validation de 24 échantillons a été utilisée pour valider le modèle. Enfin, une cohorte rétrospective de composée de 47 patientes a été analysée. Cette cohorte de 83 échantillons FFPE a été fournie par la banque de tissus du centre Oscar-Lambret.

#### **Examen de la pathologie et contrôle de l'histologie**

Pour l'histologie, un pathologiste spécialisé dans les cancers gynécologiques a lu et annoté les tissus colorés au HPS. Deux pathologistes certifiés pour le cancer de l'ovaire humain (Dr Anne-Sophie Lemaire et Dr Camille Pasquesonne, Centre Oscar Lambret) ont commenté et validé ces annotations. Les pathologistes n'ont eu connaissance d'aucune information sur l'acquisition des données MS. Les annotations ont été effectuées sur des tissus colorés à l'Hémalun/Phloxine/Saffran (HPS). Pour la coloration HPS, une tranche de tissu de 7 µm d'épaisseur consécutive au tissu analysé par SpiderMass a été traitée avec une solution d'hémalun pendant 1 minute et rincée à l'eau du robinet. Ensuite, la section de tissu a été colorée dans une solution de phloxine à 0,1 % pendant 10 secondes et rincée à l'eau du robinet avant d'être déshydratée dans des bains d'éthanol à 70 % et à 100 %. Enfin, les sections ont

été trempées dans du safran pendant 5 secondes, rincées deux fois dans de l'alcool, nettoyées au xylène et montées avec des lamelles couvre-objet et le milieu de montage pour diapositives EUKITT®. Après coloration HPS, les noyaux ont été colorés en bleu foncé, le cytoplasme en rose et le tissu conjonctif en orange. La lame colorée a été scannée pour l'acquisition d'images numériques à l'aide du scanner de lames Panoramic MIDI (3DHISTECH LTD. Budapest, Hongrie) et les images ont été visualisées et exportées à l'aide de Panoramic viewer 1.15.

### Préparation des tissus

Les échantillons ont été congelés et conservés à  $-80^{\circ}\text{C}$ . Les échantillons ont ensuite été sectionnés à  $7\mu\text{m}$  d'épaisseur, colorés au HPS et envoyés à des pathologistes pour être annotés à l'aide d'un cryotome CryoStats (Leica Microsystems, Nanterre, France). La coupe miroir à ce tissu a également été sectionnée à  $20\mu\text{m}$  pour l'analyse SpiderMass et une autre coupe de  $12\mu\text{m}$  a été réalisée pour les analyses MALDI-MSI. Pour ces dernières expériences les tissus ont été déposés sur des lames de verre revêtues d'une couche conductrice d'Oxyde Indium Etain (ou ITO) (LaserBioLabs, Valbonne, France) et stockées à  $-80^{\circ}\text{C}$ . Les lames annotées ont été utilisées pour identifier les différentes zones à analyser avec SpiderMass.

### Culture des lignées cellulaires

Des cellules ovariennes épithéliales immortalisées (dites saines) ont été cultivées dans le milieu Prigrow I. Plusieurs lignées cellulaires de cancer de l'ovaire de différentes origines ont été utilisées et cultivées. PA-1 (RRID:CVCL\_0479) est une lignée cellulaire de tératocarcinome ovarien provenant d'une patiente de 12 ans présentant une aberration chromosomique unique ( $t(15;20)(p11.2;q11.2)$  réciproque). Les cellules PA-1 ont été cultivées dans du DMEM. SK-OV-3 (RRID:CVCL\_0532) est une lignée cellulaire de cancer de l'ovaire de morphologie épithéliale, résistante au facteur de nécrose tumorale et à d'autres médicaments cytotoxiques (par exemple, la toxine diphtérique, le cisplatine et l'adriamycine), dérivée des ascites d'une femme caucasienne de 64 ans atteinte d'un cystadénocarcinome séreux. Les cellules de carcinome ovarien SK-OV-3 ont été cultivées dans le milieu McCoy. La lignée cellulaire PEO4 (RRID:CVCL\_2690) est dérivée des ascites d'une patiente, présente une mutation BRCA2 et est multirésistante à la chimiothérapie. THP-1 (RRID:CVCL\_0006) est une lignée cellulaire de monocytes isolée à partir du sang périphérique d'un enfant d'un an atteint de leucémie monocyttaire aiguë. Les cellules PEO4 et THP-1 ont été cultivées dans un milieu RPMI. Tous les milieux ont été supplémentés avec 10% de sérum bovin fœtal et 100 U/mL de pénicilline-streptomycine dans un incubateur à air humidifié à  $37^{\circ}\text{C}$  sous une atmosphère de 5% de  $\text{CO}_2$ . Après confluence à 70%, les cellules ont été lavées deux fois avec du DPBS, séchées sous PSM pendant 10min à RT puis analysées directement dans la plaque cellulaire.

### Isolement des monocytes primaires

Du sang de donneur a été collecté par l'EFS (établissement français du sang). Les poches de sang (environ 600mL au total) ont été diluées deux fois dans du PBS-EDTA. Les leucocytes ont été isolés par une centrifugation de 25 minutes à 2200 rpm avec un gradient de Ficoll. Les leucocytes ont ensuite été lavés trois fois avec du PBS-EDTA, remis en suspension dans du milieu RPMI et incubés dans une plaque de culture 1H30 à 37°C. Afin de ne conserver que les macrophages collés à la plaque, celle-ci a été lavée trois fois avec du PBS. Les macrophages ont été cultivés pendant 7 jours avec du milieu RPMI contenant 10 % de sérum bovin fœtal, 100U/mL de pénicilline-streptomycine et du MCSF.

### Différenciation des macrophages

Après l'isolement des monocytes, les cellules primaires, ainsi que les cellules THP-1, ont été stimulées avec 10ng/mL de PMA pendant 24h pour la différenciation en macrophages. Les macrophages obtenus ont été stimulés pendant 48h avec 0,5mg/mL de LPS et 20ng/mL d'IFN- $\gamma$  pour les induire vers un phénotype de type M1 ou avec 20ng/mL d'IL4 pour un phénotype de type M2.

### Isolement des lymphocytes primaires

Les cellules mononucléaires du sang périphérique (PBMC) ont été isolées à partir d'échantillons de sang total par centrifugation en gradient de densité (Ficoll Paque Plus (GE Healthcare)). Les cellules ont ensuite 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 20 minutes à 4°C dans l'obscurité. Après lavage, les cellules CD3+ CD4+, les cellules D3+CD8+ et les cellules CD3-CD7+ ont été triées à l'aide du BD FACS ARIA II SORP. Un million de chaque population a été transféré sur des lames de verre à l'aide d'une centrifugeuse Cytospin™ (Thermo Shandon Cytospin) et conservé à -80°C.

### Analyse SpiderMass

#### Tissue

Le système SpiderMass est similaire à celui décrit dans l'étude précédente (**page 73**). Les acquisitions MS ont été réalisées en mode résolution de l'instrument, à la fois en ion positif et négatif, avec un temps de balayage de 1 seconde. Pour l'analyse des tissus, les échantillons ont été sortis du congélateur (-20°C) et décongelés à température ambiante pendant 1 minute avant l'analyse SpiderMass. La séquence d'acquisition spectrale était composée de 3 acquisitions avec des périodes d'irradiation de 3 secondes. En tenant compte du fait que les modes d'ions positifs et négatifs ont été enregistrés, un total de 6 acquisitions par biopsie a

été réalisé. La section de tissu miroir de celle analysée a été utilisée comme contrôle morphologique pour s'assurer que chaque acquisition a été réalisée dans la même zone histologique et prendre en compte l'hétérogénéité du tissu si nécessaire. Les données SpiderMass ont été validées par MALDI-MSI et les lipides significativement discriminants ont été identifiés à l'aide d'expériences MS/MS.

### Cellules

À 70% de confluence, les plaques de culture ont été lavées 2 fois avec du PBS et séchées sous le PSM pendant 10 minutes à température ambiante. L'acquisition des spectres a ensuite été réalisée directement sur les plaques. La séquence d'acquisition spectrale était composée de 12 acquisitions avec des périodes d'irradiation de 3 secondes. En tenant compte du fait que seul le mode d'ions négatif a été enregistré, un total de 12 acquisitions par plaque a été réalisé.

Les lymphocytes ont été déposés sur des lames de verre par cytopspin au PBS. Les acquisitions ont ensuite été réalisées directement sur cette lame de verre après séchage de 5min au dessiccateur à température ambiante.

### Imagerie par spectrométrie de masse MALDI (MALDI-MSI)

Les tissus ont été sectionnés à 20µm d'épaisseur sur un cryostat (Leica Microsystems, Nanterre, France). Ces sections ont été déposées sur des lames de verre recouvertes d'ITO (LaserBioLabs, Valbonne, France) et séchées sous vide pendant 15 minutes. La norharmane a été utilisée comme matrice MALDI pour l'imagerie en mode ionique négatif et positif. La matrice a été sprayée à l'aide du système M5 TM (HTX Technologies, US) avec les paramètres suivants : température de la buse de 30°C, 12 passages, débit de 0,1 ml/min, vitesse de la buse de 1200 mm/min, espacement de la largeur de la piste de 2 mm et pression du gaz de 10 psi. Les expériences MALDI-MSI ont été réalisées sur un RapifleX MALDI TissueTyper™ (Bruker Daltonik GmbH) et les images ont été acquises dans les deux modes d'ionisation en utilisant du phosphore rouge comme calibrant externe avant l'analyse. La résolution spatiale a été fixée à 50 µm et la gamme de masse à  $m/z$  400-1200. L'énergie a été réglée entre 70 et 80 % de l'énergie du laser. Le porte-échantillon MALDI et la première lentille d'extraction de la source ont été réglées respectivement à une tension de 20 kV et 11 kV, la deuxième et dernière lentille étant à la terre. Le laser a été utilisé à sa plus grande fréquence de répétition, soit 10 kHz, et les spectres acquis en mode réflectron. Les données MSI ont été acquises avec une résolution spatiale de 50 µm.

### Imagerie par spectrométrie de masse en immunohistochimie (Tagmass)

L'imagerie multiplex a été réalisée à partir des mêmes lames que celles utilisées pour le l'imagerie MALDI des lipides. Le protocole utilisé était le protocole recommandé pour les sondes Ambergem Miralys. La matrice a été éliminée grâce à des bains d'EtOH de 5 secondes chaque. Ensuite, les tissus ont été séchés sous vide pendant 10 minutes et fixés avec du PFA 1%. Deux bains de PBS ont ensuite été réalisés pendant 10 min chacun pour laver les tissus, puis deux bains de 1 min dans l'acétone et un de 3 min dans une solution de Carnoy ont été effectués. Les tissus ont ensuite été réhydratés par deux bains dans de l'EtOH à 100% pendant 2 minutes chacun, suivis de trois bains consécutifs de 3 minutes dans de l'EtOH à 95%, de l'EtOH à 70% et de l'EtOH à 50%. Un bain de 10 minutes dans du TBS (50 mM Tris, pH 7,5, 200 mM NaCl) a été appliqué avant la récupération de l'antigène, qui a été effectuée dans du Tris 20mM pH9 pendant 30 minutes à 95°C. Les lames ont été lavées 10 minutes dans du TBS afin d'être traitées avec un tampon de blocage des tissus 2% (v/v) de sérum normal, 5% (w/v) de BSA dans du TBS-OG pendant 1 h. Après avoir retiré ce dernier, les tissus ont été incubés à 4°C pendant une nuit dans un tampon de blocage, dans lequel les CD3ε, CD68, CD8, CD4, Ki67, Vimentine et collagène ont été ajoutés à une concentration de 2,5 µg/mL. Chaque lame a été lavée séparément par trois bains de 5 minutes dans du TBS et de 2 minutes dans du NH<sub>4</sub>HCO<sub>3</sub> 50mM, toujours à l'abri de la lumière. Les tissus ont été séchés pendant 1h30 à température ambiante dans une chambre de dessiccation sous vide avant le photoclavage des sondes par une exposition UV de 365 nm (LED Cube 100 IC de Honle UV Technology, Marlboro, MA) pendant 10min. La matrice DHB à 20 mg/mL dans MeOH:TFA 0,1% (70:30, v/v) a été nébulisée sur les tissus à l'aide du système HTX M5 (HTX technologies, LLC), suivie d'une étape de recristallisation de la matrice dans une chambre humide avec 5% d'IPA à 55°C pendant 1 min. Ensuite, les lames ont été soumises à l'analyse MALDI-MSI à l'aide d'un instrument RapifleX MALDI-TOF-MS (Bruker Daltonics, Allemagne) selon les paramètres suivants : mode réflecteur, mode ions positifs, résolution spatiale de 20 µm, 500 tirs laser par pixel. Les spectres ont été normalisés au nombre total d'ions (TIC). L'analyse des spectres et des images a été réalisée à l'aide des logiciels FlexImaging (Bruker Daltonics, Billerica, MA) et SCiLS Lab (SCiLS Lab 2023, SCiLS GmbH).

### Imagerie SpiderMass-Mass

La pièce à main du SpiderMass a été couplée à un bras robotique rigide décrit dans un travail antérieur<sup>9</sup>. La résolution spatiale a été fixée à 250  $\mu\text{m}$  et les images acquises en mode « oversampling » sur une gamme de masse fixée entre  $m/z$  600-1100. La séquence d'acquisition était composée de 3 tirs laser consécutifs et de 3 secondes entre chaque pixels. Le pilotage de acquisition est géré automatiquement via une interface développée en interne sous MATLAB et qui permet de synchroniser les tirs laser, le déplacement du bras robotisé ainsi que l'acquisition des données par le spectromètre de masse. Les données ont été acquises en mode ion négatif et le Xevo G2-XS Q-TOF (Waters) a été réglé sur le mode « sensible ».

### Identification des lipides

Les intervalles  $m/z$  correspondant aux charges ayant la plus grande contribution à la variance observée dans les différents groupes ont été sélectionnés pour l'identification par MS/MS. Pour ces expériences, les paramètres étaient les mêmes que ceux décrits dans l'étude précédente au SpiderMass (Erreur ! Source du renvoi introuvable. Erreur ! Signet non défini.). Les scans complets ont été acquis dans le Xevo G2-XS. Les identifications ont été réalisées directement sur le tissu en effectuant d'abord un balayage complet pour vérifier la présence des masses ciblées. Ensuite, après avoir basculé en mode MS/MS, les ions ont été sélectionnés avec une fenêtre d'isolement de 0,1  $m/z$  pour réaliser une fragmentation de type dissociation induite par collision (CID). Les spectres MS/MS ont été acquis pour un temps d'irradiation continu de 5 s. Les annotations lipidiques ont été réalisées manuellement à travers les bases de données Metfrag, LipidMaps et Alex 123.

### Traitement des données

#### Classification AMX

Pour l'analyse des données, toutes les données brutes produites avec l'instrument SpiderMass ont été importés dans le logiciel AMX. Après l'importation, les spectres ont d'abord été prétraités. Les étapes de prétraitement comprennent la soustraction du bruit de fond, la normalisation du nombre total d'ions et un binning de 0,1  $m/z$ . Tous les spectres MS traités obtenus à partir des 78 échantillons validés histologiquement ont ensuite été utilisés pour construire un modèle de classification par analyse en composantes principales et analyse discriminante linéaire (ACP-LDA). La première étape a consisté en une ACP visant à réduire la multidimensionnalité des données en générant des caractéristiques qui expliquent la majeure partie de la variance observée. Ces caractéristiques ont ensuite été soumises à une

analyse supervisée à l'aide de la LDA en définissant les classes sur lesquelles le modèle sera basé. LDA tente de classer les spectres de l'échantillon et d'évaluer le modèle par validation croisée. La validation croisée a été réalisée à l'aide des méthode « 20%out » (décrite précédemment) et "Full group out". Dans cette méthode, les spectres sont regroupés par patient et éliminés un par un ; à chaque étape, les spectres sont reclassés dans ce modèle.

### Analyses statistiques

Les intervalles m/z correspondant aux scores ayant la plus grande contribution aux premières composantes principales (c'est-à-dire où idéalement 80 % de la variance est expliquée) ont été obtenus et leurs intensités normalisées pour les différentes classes ont été représentées à l'aide de GraphPad Prism 9. Pour tenir compte du déséquilibre dans le nombre d'échantillons par classe, nous avons utilisé une ANOVA bilatérale non paramétrique (Kruskal-Wallis).

### Nouveau pipeline de classification

Les données SpiderMass peuvent être traitées de deux façon. Soit les données sont importées dans la suite "Abstract Model Builder" - AMX pour permettre la réalisation de classification par LDA. Dans ce cas, les fichiers bruts sont directement importés dans "Abstract Model Builder" - AMX (version 1.0 1972.0, Waters, Hongrie) et l'extraction des données matricielles est réalisée sous forme de fichiers csv. Soit les données brutes (fichier RAW Waters) sont exportées vers un format opensource où elles peuvent être réutilisées pour des traitements à façon. Dans ce second cas, les fichiers bruts sont convertis en fichiers mzML à l'aide de MSConvert (Proteowizard) et une importation des fichiers mzML est ensuite réalisée dans Python à l'aide de la bibliothèque pyopenMS. La bibliothèque pandas est également utilisée pour importer les fichiers csv dans Python. La bibliothèque Lazy Predict a été utilisée pour construire plusieurs modèles à partir de la bibliothèque scikit-learn en formant et en testant une gamme de 24 classifier. L'état aléatoire a toujours été maintenu à 1. Ensuite, le modèle optimal a été reconstruit individuellement à l'aide de la bibliothèque scikit-learn, ce qui a permis un réglage fin de ses paramètres pour l'optimisation et l'évaluation de sa précision. Pour mieux évaluer les performances du modèle, une validation croisée 5 fold a été effectuée à l'aide des fonctions KFold et cross\_val\_score, et le rapport de classification a été généré à l'aide de la fonction classification\_report. En outre, la fonction ConfusionMatrixDisplay de la bibliothèque matplotlib a été utilisée pour afficher la matrice de confusion. Le modèle optimal a ensuite été sauvegardé et chargé pour la prédiction en aveugle à l'aide de la bibliothèque joblib, avec les fonctions joblib.dump et joblib.load, respectivement. L'algorithme LIME (Local Interpretable Model-agnostic Explanations) a été utilisé pour expliquer la prise de décision du modèle de classification. Cet algorithme calcule les contributions des caractéristiques qui peuvent être positives ou négatives. La bibliothèque ELI5 a été utilisée pour générer un tableau LIME

contenant le poids des contributions des caractéristiques à l'aide de la fonction `explain_prediction`. Ensuite, un test statistique non paramétrique Kruskal-Wallis avec correction de Bonferroni, utilisant la fonction `stats.kruskal` de la bibliothèque Scipy, a été utilisé pour évaluer l'importance de chaque contribution de caractéristique élevée. Seules les caractéristiques significatives ayant une valeur p égale ou inférieure à 0,05 ont été retenues. Enfin, une étape de filtrage a été ajoutée pour ne conserver que les pics mono-isotopiques correspondant aux molécules de la liste finale. Les box plots correspondants ont ensuite été générés à partir de la bibliothèque seaborn.

### Multi input neuronal network

Les bibliothèques TensorFlow et Keras ont été utilisées pour construire et entraîner un modèle de réseau neuronal avec des spectres et des images HPS à entrées multiples. Les spectres au format csv et les images HPS au format png ont été combinés en les convertissant en tableaux numpy. L'architecture du modèle est définie avec des branches séparées pour les données d'image (2D-CNN) et de spectres (MLP). Les branches sont concaténées et des couches supplémentaires sont ajoutées pour la classification. Le modèle est compilé avec les paramètres appropriés. Le script divise les données en ensemble d'apprentissage, de validation et de test (formation : 0,6, validation : 0,2 et test : 0,2).

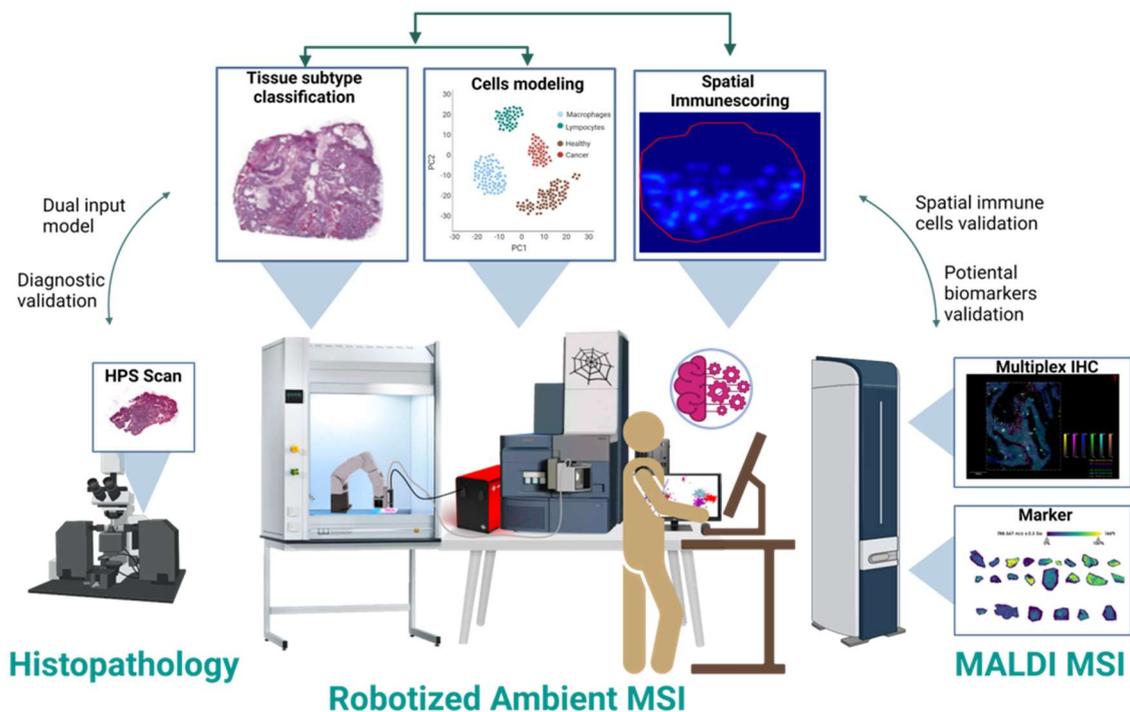
### Modélisation du score immunitaire

La bibliothèque LightGBM Python a été utilisée pour former un modèle LightGBM, qui est un cadre de renforcement du gradient développé par Microsoft. Les modèles de score immunitaire ont été construits sur la base des spectres cellulaires dans la gamme de masse m/z [600-1100] en mode ion négatif, y compris les catégories suivantes et les nombres de spectres correspondants : Macrophages (M1 & M2) : 216 spectres, lignées cellulaires saines : 163 spectres, lignées cellulaires cancéreuses : 146 spectres, lymphocytes (NK, CD8, CD4) : 114 spectres. Pour récupérer les scores de prédiction pour chaque type cellulaire, la fonction `predict_proba` du modèle LightGBM a été utilisée à la place de la fonction `predict`. Cette approche est privilégiée car elle fournit des estimations de probabilité pour chaque classe en obtenant les scores pour chaque type de cellule présent.

### Immune-score par SpiderMass-MSI

Un code Matlab a été développé pour extraire les scans MS de chaque pixel des fichiers raw Waters. Les fichiers SpiderMass bruts ont été convertis en mzML à l'aide de MSConvert (Proteowizard). Le convertisseur imzML a été utilisé pour reconstruire les fichiers d'imagerie. Un code Python a été développé pour une analyse plus approfondie et utilise la classe `ImzMLParser` de la bibliothèque `pyimzml`. Il importe et analyse le fichier imzML dans

l'environnement Python. Le script itère sur chaque coordonnée (x, y, z) dans le fichier imzML et récupère les spectres de masse correspondants à l'aide de la méthode `getspectrum()`. Il applique un filtre de plage de masse souhaitée pour sélectionner des valeurs de masse spécifiques et stocke les spectres avec leurs coordonnées respectives. Plusieurs étapes de prétraitement sont effectuées dans le code Python, notamment l'identification des pixels non nuls, la normalisation et l'application d'un lissage gaussien à l'aide de la fonction `gaussian_filter` du module `scipy.ndimage.filters`. La résolution spatiale est également augmentée d'un facteur 2 en utilisant l'interpolation avec la fonction `zoom` de `scipy.ndimage`. Pour les prédictions, un modèle LightGBM pré-entraîné est chargé à partir d'un fichier `joblib` à l'aide de la fonction `joblib.load`. Le script itère sur chaque spectre de l'image et les prépare pour la prédiction. La méthode `predict_proba` du modèle est utilisée pour obtenir les scores et les étiquettes prédits pour chaque spectre. Après avoir obtenu les étiquettes prédites, le script visualise les cartes d'étiquettes pour chaque classe prédite par le modèle. Il crée une carte de couleurs en utilisant le schéma de couleurs "jet" et itère sur chaque classe pour afficher les cartes d'étiquettes à l'aide de la fonction `imshow` du module `matplotlib.pyplot`. Une barre de couleur des scores est également ajoutée à titre de référence. L'ensemble de la méthodologie développée pour l'étude du cancer de l'ovaire est présentée **Figure 28**.



**Figure 28:** Workflow globale utilisée dans cette étude.

Un modèle de classification robuste basé sur la combinaison de la coloration H&E et des spectres MS par SpiderMass a été réalisé pour le diagnostic du cancer de l'ovaire. Un autre modèle composé de différents types de cellules de la TME a été construit pour les déterminer dans un échantillon de tissu. Ensuite, l'analyse MSI de la section de tissu a été effectuée et chaque

pixel a été comparé au modèle précédemment construit pour déterminer la présence des différentes cellules immunitaires sur le tissu.

## Résultats

### *Sous-typage histologique du cancer de l'ovaire basé sur les données moléculaires de SpiderMass*

Nous avons d'abord cherché à déterminer si la technologie SpiderMass pouvait permettre le typage moléculaire, en temps réel, des différents sous-types cancéreux par rapport au tissu ovarien normal. Pour cela, les spectres obtenus avec le SpiderMass ont été recueillis à partir de coupes histologiques de la cohorte prospective du cancer de l'ovaire. Ces spectres MS montrent des différences dans leur empreinte moléculaire en fonction du type de cancer à la fois en mode ion négatif (**Figure 29, A,B**) et en mode ion positif (**supp data 1**). Certains lipides, identifiés par MS/MS, tels que  $m/z$  750.55 PE (36 :4) ou  $m/z$  885.5 PI (20 :4/18 :0), présentent une distribution similaire dans les différents sous-types, alors que d'autres, sont clairement spécifiques à l'un d'entre eux (**supp data 2**). Pour mieux comprendre ces modèles moléculaires distincts et créer un modèle de classification permettant de séparer chaque sous-type, nous avons entraîné des algorithmes de Machine Learning (ML) sur l'ensemble de la cohorte prospective de patientes atteintes d'OC. Différents algorithmes de ML ont été évalués. Tout d'abord, un modèle d'analyse multivariée conventionnel basé sur la PC-LDA a été construit en utilisant le logiciel AMX dans les deux modes d'ionisation. Ceux-ci sont basés sur l'acquisition des 79 échantillons de la cohorte prospective (13 HGSC, 8 Mucineux, 10 séreux borderline (BL), 6 Endométrioides, 44 ovaires normaux) de 59 patientes. Ce modèle a permis la discrimination avec succès de ces cinq classes dans les deux modes d'ionisation. Une méthode de cross-validation de ce modèle a ensuite été réalisée par "20%-out". Cette méthode consiste à retirer 20% des spectres et à les classer dans le modèle comprenant les 80% restants. Elle a permis d'obtenir des classifications correctes allant de 93,78% sans à 92,89% avec outliers en mode ions négatifs (**Figure 29C**), et de 96,10% sans à 95,17% avec outliers en mode ions positifs (**Figure supp data 3, A,B**). La méthode "full group out" a ensuite été utilisée. Elle consiste à retirer patient par patient les spectres du modèle et de reclasser ceux-ci au sein du modèle généré sans ces patients. Il en résulte 72,11% de classification correcte en mode négatif sans et 64,93 % avec outliers. En mode positif, 74,19% de classification correcte sans et 66,76 % avec outliers.

Pour améliorer la classification, nous avons procédé à une augmentation des données en combinant les tissus déjà analysés à 83 échantillons supplémentaires provenant d'une cohorte rétrospective de tissus FFPE. Le modèle de classification PC-LDA généré à partir des données FFPE seules a conduit à une classification correcte de 83% avec et sans outliers par « 20%-out » et 70.24% sans et 68.90% avec outliers par la méthode "full group out" en mode négatif. Le modèle étendu au deux cohortes (FFPE+FF), quant à lui, a obtenu un taux de

cross-validation de 89% avec et sans outliers (**Figure supp data 4**) et 76.30% avec et 75,71% sans outliers avec la méthode “full group out”. Ce modèle n'a donc pas amélioré la précision de la prédiction. La raison en est probablement le manque de tissus ovariens normaux dans la cohorte FFPE. De plus, le logiciel AMX présente certaines limites. La principale est l'utilisation de la classification LDA qui manque d'évaluation du modèle tel que le F1-score. De plus, on peut reprocher à ces modèles leur manque d'évaluation et de cross-validation des résultats.

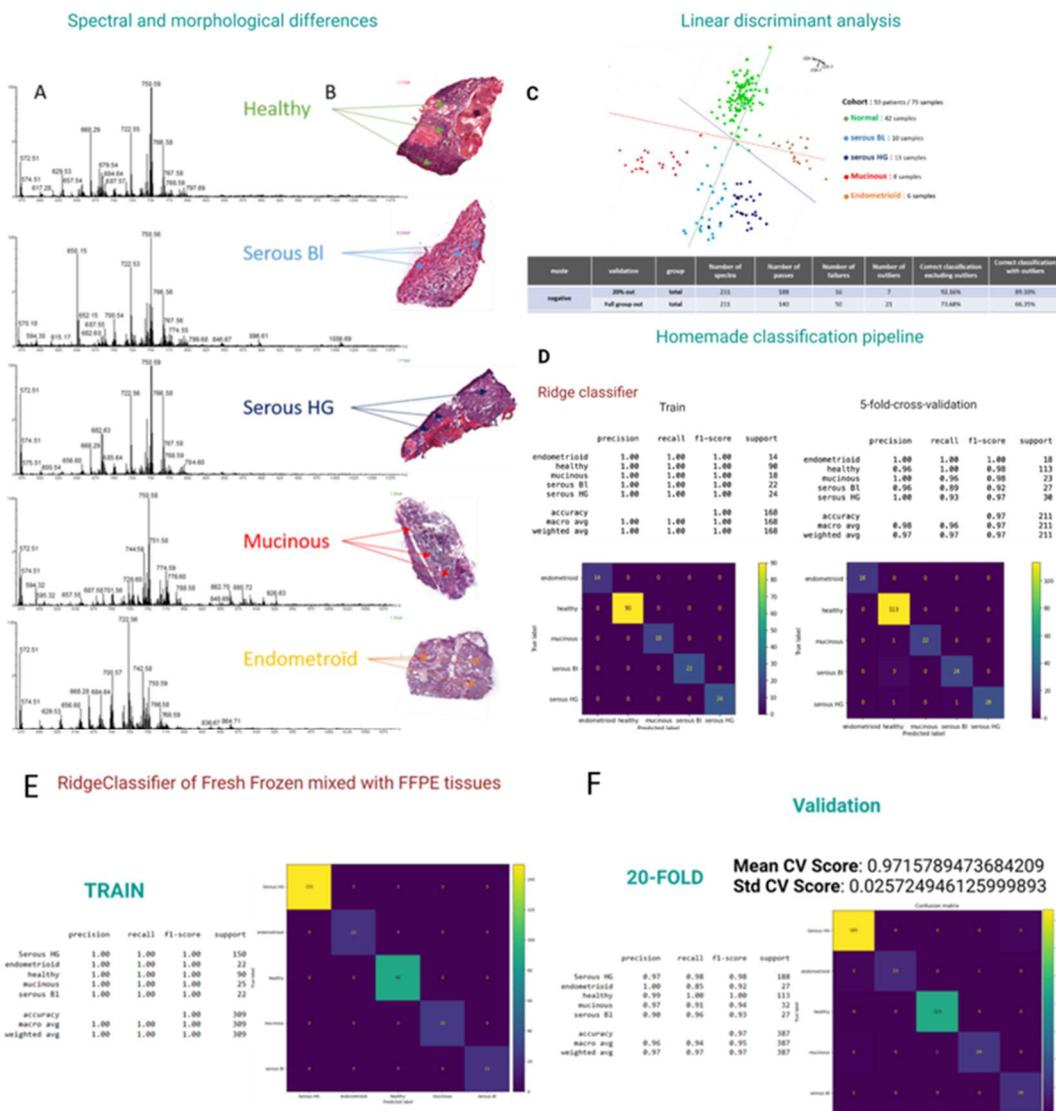


Figure 29: Modèles basés sur l'analyse statistique multivariée pour décrypter les cancers de l'ovaire.

(A) Spectre moyen obtenu pour chaque sous-type de cancer de l'ovaire en mode ion négatif (burst de 10 tirs, 4mJ). (B) coloration HPS de ces tissus. Les flèches correspondent aux régions analysées du tissu. (C) Modèle PC-LDA de classification de la cohorte de 79 échantillons de cancer de l'ovaire en mode ion négatif. Le point vert correspond à un spectre de tissu sain moyen. Le point bleu clair correspond à un spectre SBL moyen. Le point bleu foncé correspond à un spectre moyen HGSOc. Les points orange et rouge correspondent respectivement aux spectres des tissus endométrioïdes et mucineux. (D) Modèle basé sur le ridge Classifier pour les sous-types de cancer de l'ovaire (sain, SBL, HGSOc, mucineux, endométrioïde). Le rapport de classification et la matrice de confusion ont été obtenus sur l'ensemble d'apprentissage pour les deux ensembles de training et de cross-validation 5-fold. (E) Modèle de classification RIDGE basé sur l'ajout de données FFPE et FF. Le rapport de

*classification et la matrice de confusion ont été obtenus sur pour les deux ensembles d'apprentissage et (F) de validation croisée 20-fold.*

Pour remédier aux limites de la classification des classes à faible prévalence dans la classification LDA et pour mieux gérer les différences spectrales entre les tissus ayant des méthodes de conservation différentes (frais congelés et FFPE), d'autres algorithmes de classification ont été testés tels que le SVM, le Random Forest et l'algorithme Perceptron. Au total 24 algorithmes de classification avec cross-validation k-fold ont été appliqués sur le même jeu de données. Le modèle optimal a été sauvegardé et testé par la suite pour le diagnostic en aveugle. De plus, la cross-validation « 20 % out » choisi au hasard les données retirées du modèle et ne représente pas l'ensemble des données (généralisation des données) tandis que la méthode "full group out" est connue pour augmenter l'écart-type de la précision moyenne. Par conséquent, la cross-validation k-fold a été utilisée pour l'évaluation du modèle pour le reste de l'analyse. Le classificateur RIDGE<sup>180</sup> a obtenu la plus grande précision, soit 100 % en entraînement et 97 % après une validation croisée 5-fold en mode ion négatif sur la cohorte de tissus frais (**figure 29D**). Ce modèle convient parfaitement aux ensembles de données à haute dimension comme les nôtres (5000 dimensions  $m/z$ ). C'est un classificateur linéaire qui utilise la régularisation L2 pour éviter l'ajustement excessif. Il a notamment démontré de solides performances pour les données MS provenant de tissus<sup>181</sup>. Ce modèle a également été utilisé en mode ion positif, ce qui a permis d'obtenir une précision de 100 % à l'entraînement et de 94 % après une validation croisée 5-fold. La cohorte de tissus FFPE a également été utilisée pour construire un autre modèle qui a donné une précision de 100 % à l'entraînement et de 94,2 % après une validation croisée de 20-fold (**supp data 3, D, E**). Ce nouveau modèle, qui combine à la fois les tissus FFPE et les tissus frais congelés (FFPE+FF), a obtenu un excellent taux de classification avec une précision de 100 % à l'entraînement et de 97 % après une validation croisée de 20-fold (**figure 29, E-F**).

Les différents modèles ont ensuite été évalués en analysant en aveugle 24 nouvelles patientes d'une cohorte de validation. Pour chacun des 24 tissus, 3 zones différentes au moins ont été analysés au hasard dans le tissu, pour un total de 72 analyses en aveugle dans la cohorte. Pour chaque analyse en aveugle, le logiciel AMX a interrogé la classification (PC-LDA) précédemment établie et a prédit la classe en quelques secondes, comme le montre la **figure 30A-B** pour l'HGSOC et l'ovaire normal. Premièrement, l'analyse en aveugle a montré que les tissus HGSOC, SBL et normaux sont correctement diagnostiqués. Seulement, le modèle LDA peine à différencier le carcinome endométrioïde du carcinome mucineux (**Figure 30C**). Plus précisément, les trois tissus endométrioïdes testés ont été mal classés et un des tissus mucineux n'a été que partiellement classé. Cela peut s'expliquer par le manque de tissus

endométrioides dans la cohorte d'entraînement par rapport aux autres sous-types. En retirant la classe endométrioides de la cohorte en aveugle, le taux de classification correcte passe à 85%. Ensuite, le modèle créé avec le RIDGE Classifier basé sur le mélange entre les tissus FFPE et FF a été testé. Ce modèle a permis une meilleure prédiction que le précédent, mais les tissus endométrioides n'ont toujours pas été correctement classés (**Figure 30C**). Toutefois, ce modèle a atteint 95,25 % de classification correcte en supprimant les tissus endométrioides.

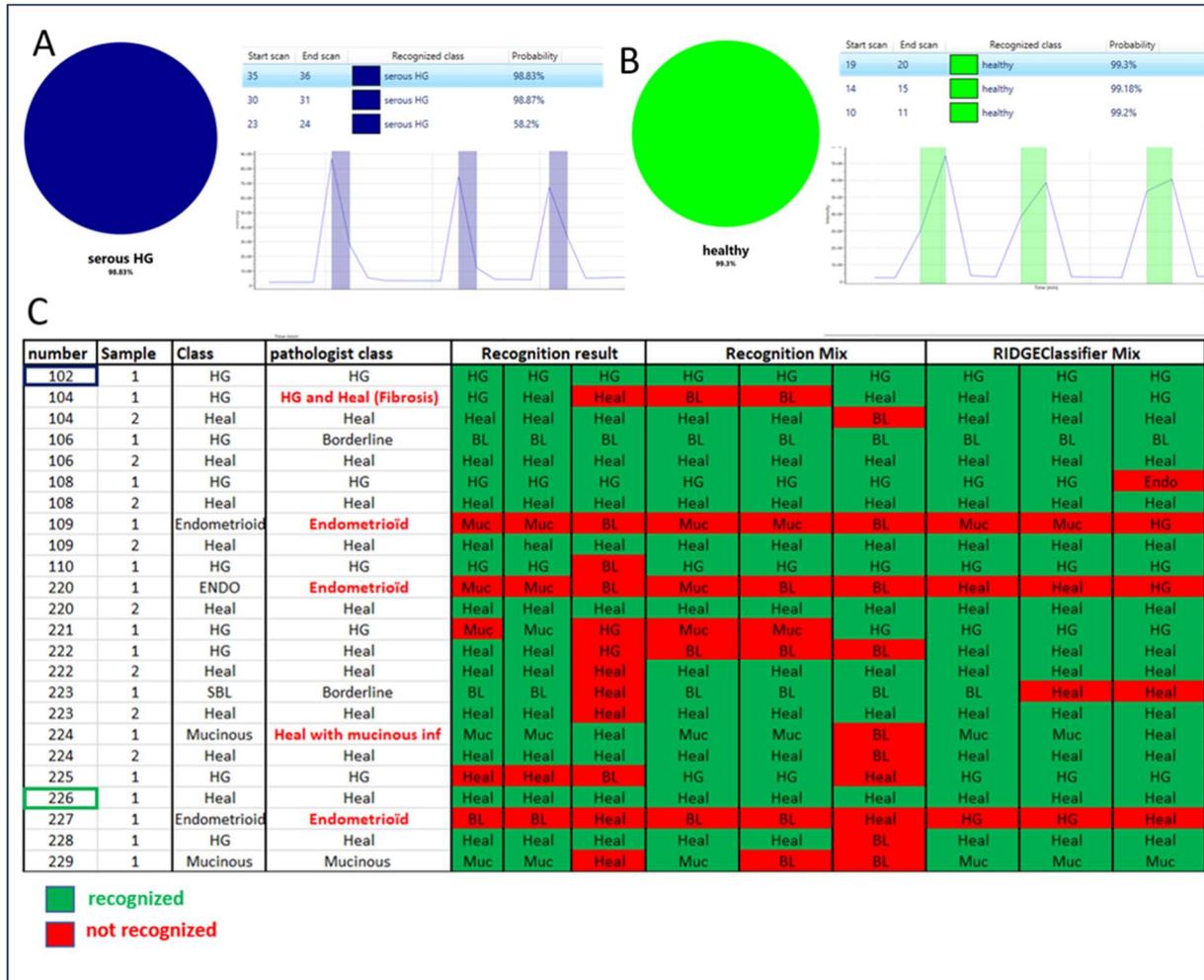
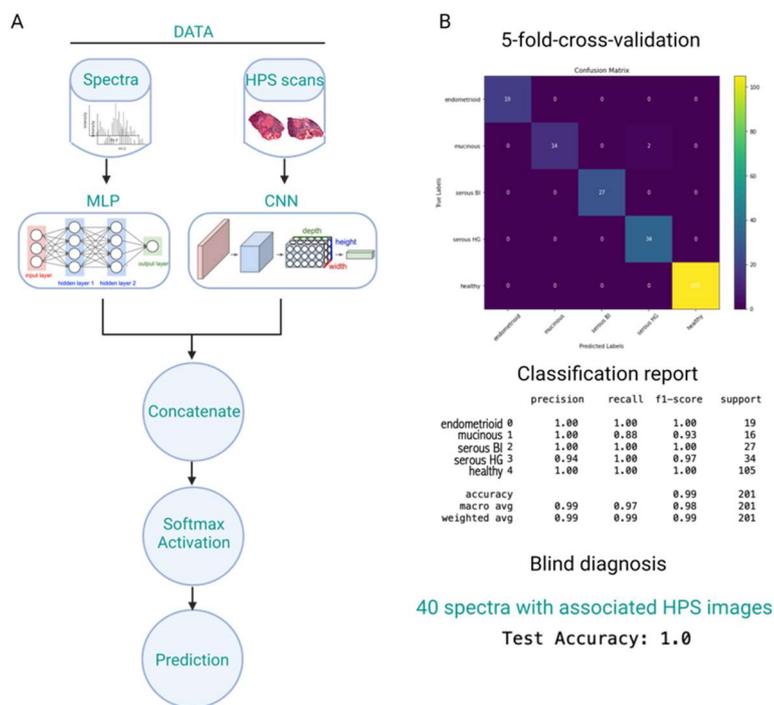


Figure 30: Validation des modèles de classification sur une cohorte en aveugle de patientes.

(A) Exemple de HGSOC analysé en aveugle et reconnu par le logiciel AMX recognition. (B) Exemple de tissu sain analysé en aveugle et reconnu par le logiciel AMX recognition. (C) Tableau représentant toute la cohorte de validation analysée et comparée aux modèles LDA, RIDGE FF et RIDGE FFPE/FF. Chaque échantillon entièrement reconnu est coloré en vert. Les non reconnus sont surlignés en rouge.

Pour déterminer le modèle le plus performant, les tissus colorés utilisés pour l'annotation ont été scannés et intégrés aux spectres MS afin de créer un modèle à double entrée par Deep Learning. Pour ce faire, nous avons développé un modèle de réseau neuronal à entrées multiples (**Figure 31A**) qui combine un Multilayer Perceptron (pour les données moléculaires)

et un 2D-CNN (pour les données morphologiques). Cette nouvelle approche semble produire un modèle plus sensible et plus robuste. En effet, le modèle a atteint une précision de 100 % en mode ion négatif après une cross-validation 5-Fold. Ce modèle a également démontré une précision de 100 % en diagnostic aveugle lorsqu'il a été testé sur 40 spectres de la cohorte (**figure 31B**). En outre, ce modèle a présenté une sensibilité et une spécificité de 100 % dans les ensembles de validation et le test. L'intégration des informations histologiques et des données moléculaires améliore considérablement la capacité à diagnostiquer divers carcinomes épithéliaux ovariens. Ce modèle est particulièrement convaincant car il intègre des informations moléculaires qui faisaient défaut dans d'autres modèles uniquement axés sur les caractéristiques morphologiques<sup>182</sup>.

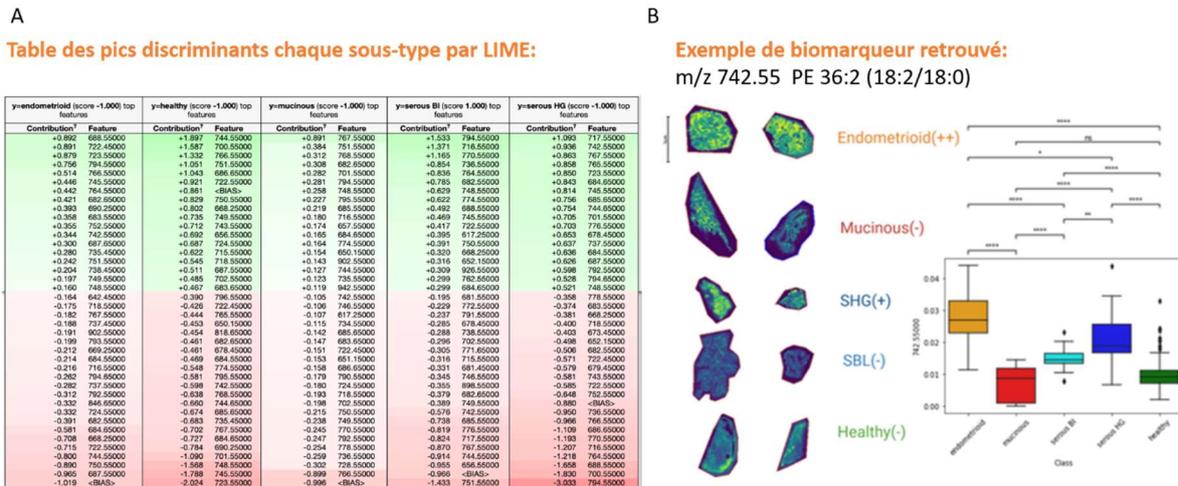


**Figure 31:** Modèle à plusieurs entrées (Dual-input) associant les scans HPS et les spectres MS.

(A) Le processus d'élaboration du modèle à entrées multiples comprend deux étapes principales : l'entraînement sur les scans des tissus à l'aide d'un 2D-CNN (réseau neuronal convolutif) et des données SpiderMass à l'aide du MLP (Multilayer perceptron). Après l'entraînement des modèles CNN et MLP, les sorties des deux réseaux sont concaténées. Le processus de concaténation combine les caractéristiques extraites des données de numérisation et de SpiderMass, créant ainsi une représentation complète qui incorpore les informations des deux sources. Pour établir des prédictions, la représentation des caractéristiques concaténées est soumise à une fonction d'activation SoftMax. La fonction SoftMax convertit les caractéristiques concaténées en scores de probabilité, ce qui permet de prédire les étiquettes de classe. (B) Les performances du modèle à entrées multiples ont été évaluées à l'aide des paramètres suivants : Matrice de confusion et Rapport de classification. Pour garantir la robustesse et la fiabilité, une technique de cross-validation 5-fold a été utilisée pendant l'évaluation. L'ensemble de données a été divisé en cinq sous-ensembles (fold) et le modèle a été entraîné et validé cinq fois, chaque servant une fois d'ensemble de validation et les quatre plis restants de données d'entraînement. Les mesures de performance finales ont été calculées en faisant la moyenne des résultats des 5-Fold. Pour mieux évaluer la capacité de généralisation du modèle, 40 spectres ont été analysés en aveugle. L'ensemble de test est constitué de ces 40 échantillons non vus, qui n'ont pas été utilisés lors de la formation ou de la validation croisée. La précision du modèle sur cet ensemble de test fournit une mesure fiable de la performance du modèle sur de nouvelles données inédites.

*Définition de potentiels biomarqueurs du cancer de l'ovaire.*

L'étape suivante a consisté à cross-valider les résultats de la classification et de fournir une explication biologique de la distinction des classes. Pour cela, nous avons développé un pipeline complet de traitement des données qui, en parallèle de la classification, utilise l'algorithme LIME pour déterminer la contribution positive ou négative de chaque  $m/z$  à la classification de chaque type de cancer (**Figure 32A**). Dans le pipeline, les  $m/z$  qui contribuent de manière significative à la prédiction d'une classe particulière, si leur pertinence est confirmée par des tests statistiques, peuvent être considérées comme des biomarqueurs potentiels. De plus, chaque biomarqueur potentiel identifié a été cross-validé à l'aide de la MALDI-MSI en vérifiant que ces marqueurs soient bien localisés dans les régions cancéreuses. Par exemple, l'ion  $m/z$  742,5, identifié par  $MS^2$  comme PE (18:0/18:2), s'est avéré plus discriminant dans les HGSOE et les EC. Cela est confirmé par la distribution très abondante de ces 2 lipides dans les images MALDI et par la variation de l'abondance relative de ces lipides dans les données SpiderMass de la cohorte entière (**Figure 32B**). Tous les lipides correspondant aux biomarqueurs potentiels ont été identifiés par  $MS^2$  (**Figure 32, supp data 5 et 6**). Des lipides discriminants ont été observés pour les deux modes d'ions (positif et négatif) et ont présenté une diversité remarquable. Dans le mode d'ion négatif, des acides gras saturés et insaturés, ainsi que plusieurs classes de glycérophospholipides (GP) tels que les glycérophosphoéthanolamines (PE), les glycérophosphoglycérols (PG), les glycérophosphosérines (PS) et les glycérophosphoinositols (PI) ont été observés (**Figure 32, supp data 6**). En mode ion positif, les PE et PS ont été identifiés comme différentiellement exprimés. Par rapport au tissu ovarien normal, chaque sous-type présente des lipides spécifiques. Une abondance relative élevée de  $m/z$  747.516,  $m/z$  725.5,  $m/z$  744.5 et  $m/z$  718.5 a été observée dans les régions à forte densité de cellules tumorales dans les HGSOE. Dans le mode d'ion positif, les espèces PC, y compris PC (20:1) et PC (18:4) sont trouvées à des abondances relatives élevées dans les carcinomes mucineux, tandis que PS et PE sont plus représentatifs de HGSOE et SBL. De plus, plusieurs lipides se distinguent entre HGSOE et SBL tels que PI (22:1/18:0), TG 16:0/18:4/20:4), PE (18:0/22:4) dans SBL contre PE (18:0/16:0), PA (18:1/20:2), PE (18:0/18:1) dans l'HGSOE.



**Figure 32:** Cross-validation des marqueurs lipidiques des différents sous-types d'OC par MALDI-MSI et SpiderMass.

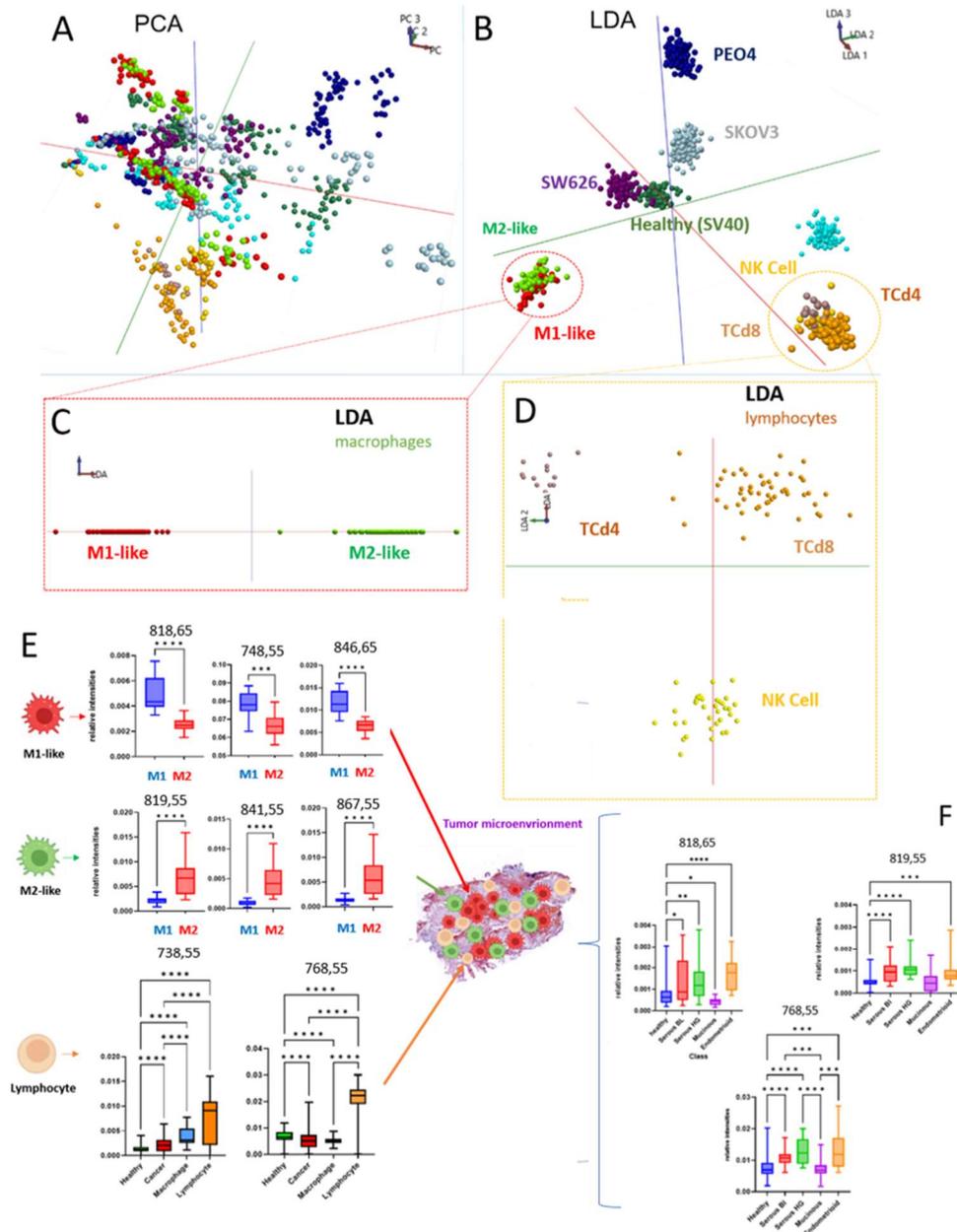
(A) L'algorithme LIME a été utilisé pour déterminer la contribution, positive ou négative, de chaque pic dans la différenciation de chaque classe (rouge : négativement, vert : positivement). La contribution positive de l'ion par LIME est représentée par un (+) et la contribution négative par un (-). (B) Exemple d'un marqueur lipidique, PE 36:2 (18:2/18:0) (m/z 742.55), différentiellement exprimé entre les différents histotypes de CO. Ce marqueur a été validé par MALDI-MSI et par l'algorithme LIME. En outre, un test de Kruskal-Wallis a été réalisé sur les données d'analyse SpiderMass de cet ion et représenté sous forme de boxplot. (\*\*\*)  $p < 0,0001$ , (\*\*\*)  $p < 0,001$ , (\*\*)  $p < 0,01$ , (\*)  $p < 0,05$  NS pour  $p > 0,05$ ).

### Décrypter l'environnement tumoral par la technologie SpiderMass.

Le TME est connu pour jouer un rôle clé dans la progression du cancer, la survie et la migration<sup>183</sup>. Pour distinguer directement par SpiderMass les types cellulaires les plus représentés dans le TME, nous avons d'abord dû obtenir le profil moléculaire de ces différentes cellules. Ainsi, une analyse SpiderMass a été réalisée sur différentes lignées cellulaires cancéreuses (SKOV3, POE4, PA1, SW626), une lignée cellulaire d'épithélium ovarien (considérée comme saine), une lignée cellulaire de macrophages différenciés, des macrophages primaires triés (M1-like et M2-like) et des lymphocytes (NK, CD4, CD8) afin d'obtenir leur signature moléculaire spécifique. Une analyse statistique multivariée non supervisée par ACP (**Figure 33, A**) puis un modèle de classification LDA (**Figure 33, B**) ont ensuite été réalisés à partir de toutes ces données.

Les différents types de cellules étaient déjà séparés les uns des autres dans le modèle PCA et certains groupes cellulaires étaient déjà visibles. Il en résulte un modèle LDA de 96,83 % de classification correcte sans outliers et 92,48 % avec outliers, représentant seulement 43 outliers sur 957 spectres (**figure 33, B**). Toutes ces erreurs d'affectation ont été causées par la discrimination des sous-populations de macrophages. En effet, 10 % de toute la cohorte de macrophages, incluant les THP1 et les cellules primaires, a été mal assigné entre les sous-classes M1- et M2-like. Cependant, en assignant à ces deux phénotypes la classe "macrophage", ce modèle passe à 99,78% et 95,30% avec et sans outliers respectivement. En isolant uniquement les macrophages sur un autre modèle PC-LDA, il a été possible de les

discriminer uniquement par leur phénotype M1-like ou M2-like avec 100% de bonne classification (**Figure 33,C**). De plus, les macrophages sont bien séparés des autres cellules immunitaires (lymphocytes) et encore plus des cellules cancéreuses ou saines (**Figure 33,B**). Il est intéressant de noter que 2 lipides à  $m/z$  818.65 et  $m/z$  846.65 se sont révélés très abondants dans les spectres MS des macrophages et significativement différents des lignées cellulaires du cancer de l'ovaire. Ces lipides ont été identifiés par MS<sup>2</sup> et correspondent respectivement au GlcCer (d18:1/22:0) et au GlcCer (d18:1/24:0). Ceux-ci se sont révélés spécifiques des macrophages, à la fois dans la lignée cellulaire THP1 et dans les macrophages primaires issus de prélèvements sanguins. D'autres marqueurs sont exprimés différemment par les macrophages primaires en fonction de leur phénotype M1- et M2-like. Pour les macrophages de type M2, une augmentation significative des abondances relatives a été observée pour  $m/z$  819.55,  $m/z$  841.55,  $m/z$  843.55 et  $m/z$  867.55, identifiés respectivement comme PG (18:1/22 :6), PG (22 :6/20 :4), PG (20 :3/22 :6) et PI (0-16 :0/22 :6) (**Figure 33, E**). D'autre part, les M1-like ont été fortement discriminés par  $m/z$  818.65 et  $m/z$  846.65 correspondant aux GlcCer(d18:1/22:0) et GlcCer(d18:1/24:0) précédemment identifiés. L'ion  $m/z$  748.55 qui était déjà abondant dans les tissus ovariens normaux et identifié comme un PE (16:0/22:5) a également permis la discrimination des M1-like. Il est également intéressant de noter qu'une augmentation relative des PGs d'acides gras polyinsaturés a été observée pour les macrophages de type M2.



**Figure 33:** Découverte de marqueurs spécifiques des cellules immunitaires.

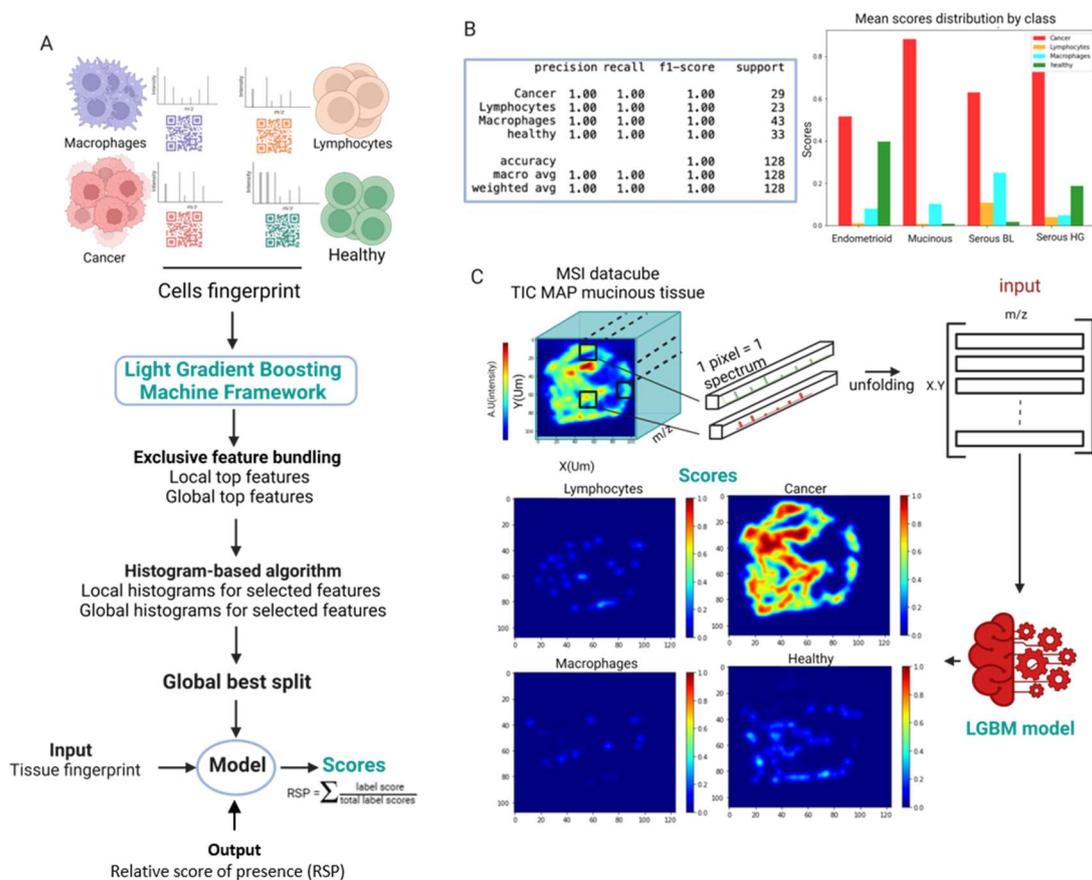
Modèle PC-LDA de différentes lignées cellulaires de cancer de l'ovaire avec différents macrophages (M1-like, M2-like, NK, CD4+, CD8+). (A) Modèle PCA de lignées cellulaires de cancer de l'ovaire et de cellules immunitaires (B) Modèle LDA de ces lignées cellulaires avec un binning de 0.1Da et une normalisation des spectres. (C) Modèle LDA des macrophages type M1 vs. macrophages de type M2. (D) Modèle LDA des lymphocytes primaires incluant NK, CD4+ et CD8+ avec les mêmes paramètres statistiques. (E) Boxplot représentant les intensités relatives des ions marqueurs des cellules immunitaires (M1, M2 et lymphocytes) avec le test statistique Kruskal-Wallis. (\*\*\*\*  $p < 0,0001$ , \*\*\*  $p < 0,001$ , \*\*  $p < 0,01$ , \*  $p < 0,05$ ) NS pour  $p > 0,05$ ). (F) Boxplot représentant l'intensité relative de ces mêmes m/z au sein de la cohorte OC en fonction des sous-types de cancer.

Différents lipides ont été identifiés comme étant discriminants des lymphocytes. Parmi eux, les ions  $m/z$  768.55,  $m/z$  885.55 et  $m/z$  738.55 qui correspondent respectivement à PE (38:3), PI (20:4/18:0) et PE (20:4/16:0). De plus, la discrimination des sous-populations de lymphocytes (NK, CD8+ et CD4+) a été possible grâce à la création d'un modèle PC-LDA contenant uniquement des spectres lymphocytaires. Ce modèle sépare tous les phénotypes de lymphocytes avec 100% de bonne classification (**Figure 33, D**). Ceci conduit à identifier  $m/z$  752.55 (PE (0-18:0/20:4)) et  $m/z$  794.55 (PE(18:0;22:4)) comme spécifiques des cellules NK, et  $m/z$  722.55 (PE(P-16:0/20:4)) et  $m/z$  885.55 (PI(20:4/18:0)) spécifiques des CD4+ (**Figure 33, supp. Data 7**). Les intensités de chacun de ces lipides, jugés discriminants pour les différentes sous-populations de cellules immunitaires, ont été collectées dans l'ensemble de la cohorte d'OC afin d'estimer leur abondance dans les différents sous-types (**Figure 33, E**). Il est intéressant de noter que l'abondance des  $m/z$  818.65 et 819.55, spécifiques des macrophages, est plus élevée dans les HGSOE, alors qu'elle est très limitée dans les tissus ovariens sains. Cela indique une corrélation potentielle entre la présence des macrophages et l'agressivité du cancer. Cependant, le carcinome endométrioïde présente également une quantité importante de macrophages malgré sa plus faible agressivité. Pour mieux comprendre cette observation, une cohorte d'endomètres de 79 patientes a été analysée. En recherchant ces lipides spécifiques des macrophages dans les tissus sains de l'endomètre, une abondance de ces lipides associés aux macrophages a été constatée. En effet, ceux-ci sont significativement plus élevés dans les endomètres sains plutôt que le tissu ovarien sain (**Figure 33, supp. 8,B**).

#### *Diagnostic et pronostic basés sur l'immuno-scoring par SpiderMass-MSI*

L'analyse SpiderMass MS fournit un moyen rapide et direct de récupérer des informations moléculaires qui pourraient être transférées en analyse peropératoire. Pour répondre aux besoins de la chirurgie, le système est conçu pour analyser des zones de tissu de 500  $\mu\text{m}$  par tir. Le spectre MS recueilli après une seule analyse correspond à un mélange potentiel des profils moléculaires de plusieurs types de cellules constituant le TME. Cependant, nous souhaitons obtenir plus d'informations sur la composition de celui-ci. Plus précisément, nous souhaitons discerner la présence de sous-populations immunitaires infiltrées par rapport aux autres types de cellules dans les tissus analysés. À cette fin, un traitement permettant de prédire l'abondance tissulaire des différents types cellulaires à partir de leur empreinte moléculaire globale a été développé. Un modèle de score immunitaire a donc été entraîné à l'aide de la bibliothèque Python LGBM (Light Gradient Boosting Machine) qui est un puissant modèle développé par Microsoft<sup>184</sup>. Ce modèle s'est entraîné sur les spectres MS des différentes cellules dans la gamme  $m/z$  [600-1100] en mode ion négatif (**Figure 34A**), ce qui a permis une discrimination réussie avec une précision de 100 %. Ce

modèle a été formé à partir de toutes les données obtenues à partir des lignées cellulaires et des cellules immunitaires primaires précédemment analysées et a été testé sur la cohorte OC. Tous les tissus ont reçu une probabilité de présence pour chaque type de cellule. Ainsi, en ajoutant un score relatif de présence (RSP) à ce nouveau modèle de classification puissant, les tissus ont pu être discriminés en fonction de leur score immunitaire (**Figure 34B**). En effet, des différences de signatures cellulaires ont été mis en évidence selon le sous-type de cancer. Par exemple, les carcinomes mucineux présentent une plus grande quantité de cellules cancéreuses alors que les SBL montrent plus de lymphocytes et de macrophages. Les endométrioïdes, quant à eux, présentent plus de cellules normales. Ce dernier point pourrait être lié au meilleur pronostic de ces carcinomes.

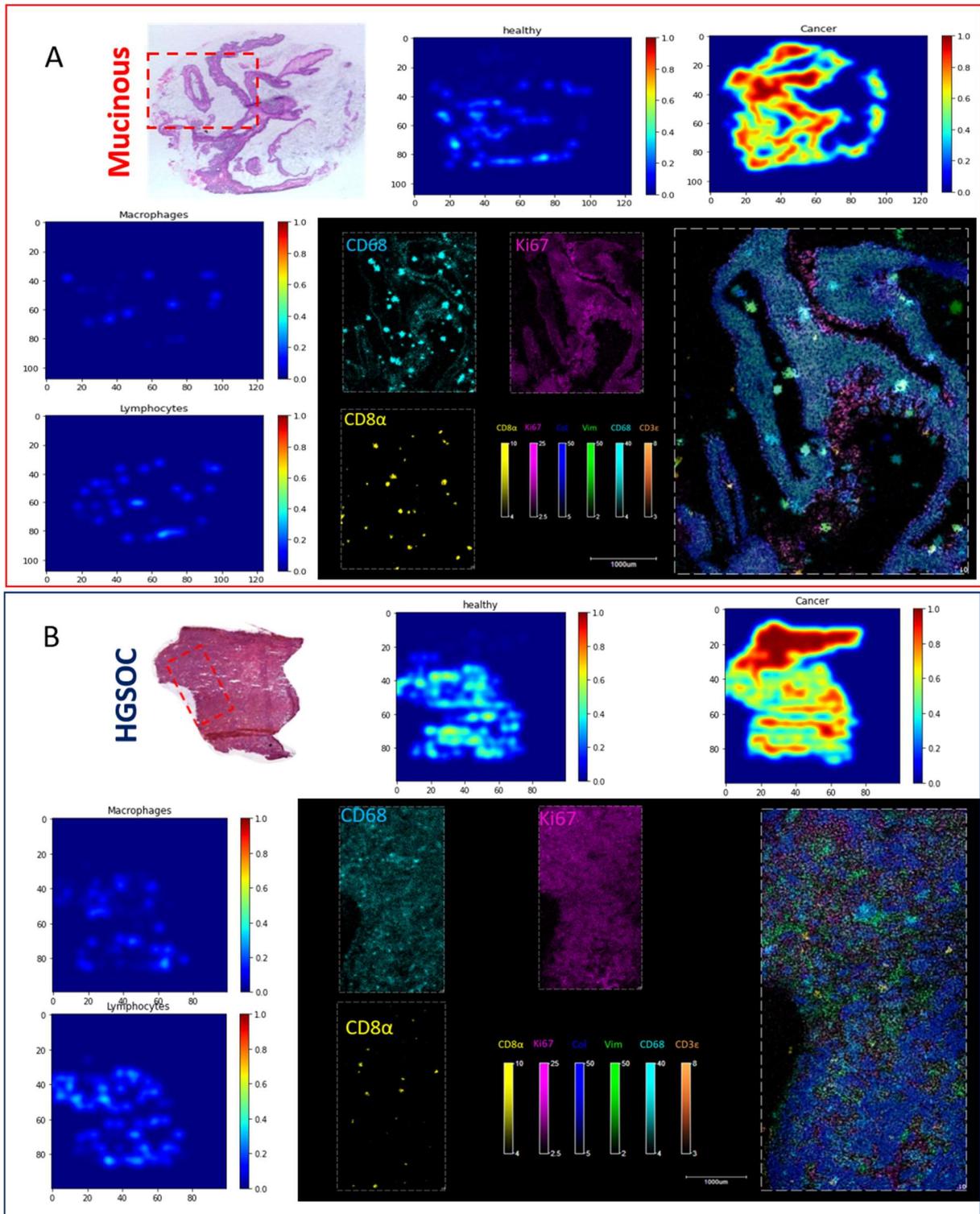


**Figure 34:** Création d'une image à partir du score immunitaire.

(A) La structure du modèle LGBM a été construite sur la base de l'empreinte moléculaire de chaque cellule. Tous les types cellulaires ont été analysés et les pics contribuant à la fois positivement et négativement à la distinction des classes ont été utilisés pour les discriminer. Le spectre d'un tissu est ensuite interrogé par LGBM pour définir un score de présence de chaque type de cellules. (B) Tableau représentant la performance du modèle. Histogramme représentant la moyenne des scores obtenus pour chaque type cellulaire en fonction du type de cancer analysé par SpiderMass. (C) Exemple d'une visualisation d'une image SpiderMass basée sur le score immunitaire. Chaque pixel (correspondant à un spectre moyen) est testé par le modèle LGBM pour donner le score de présence de cellules cancéreuses, de lymphocytes, de macrophages et de cellules saines. Pour mieux visualiser la présence de chaque classe, une image est construite pour chaque classe (type de cellules).

Pour une vision globale de la composition cellulaire des différents tissus, l'imagerie SpiderMass (SpiderMass-MSI) a été réalisée sur 3 échantillons pour chaque classe (**Figure 34C**). Cette opération a été réalisée après l'acquisition des images. Chaque pixel (correspondant à un spectre moyen) de chaque image a été extrait puis interrogé à l'aide du modèle LGBM pour créer une autre image, cette fois-ci, basée sur les résultats LGBM. Ces scores ont été calculés en additionnant les scores fournis pour chaque type de cellule par le modèle RSP et en divisant la somme par les scores totaux pour toutes les classes du modèle.

Pour valider les distributions basées sur les prédictions LGBM et proposer un diagnostic précis de ces cancers, une immunohistochimie MALDI MS (MALDI-IHC) a été réalisée. Nous avons utilisé les marqueurs CD68, CD8 et Ki67 qui sont des cibles spécifiques des macrophages, des lymphocytes et de la prolifération des cellules cancéreuses respectivement (**figure 34**). Comme prévu, une différence dans la distribution des macrophages et des lymphocytes est observée en fonction du sous-type cancéreux (**figure 34, supp data 9 à 12**). Les échantillons SBL et HGSOE montrent une distribution assez homogène des macrophages dans tous les tissus avec quelques îlots où les macrophages sont légèrement plus abondants (**figure 34 B et données complémentaires 10,11**). Au contraire, les carcinomes endométrioïdes et mucineux présentent des distributions plus hétérogènes avec des îlots présentant des signaux très intenses alors que la présence de macrophages est très faible dans le tissu environnant (**figure 34A, supp data 9,12**). Dans l'ensemble, chaque tissu semble présenter la même quantité de cellules cancéreuses prolifératives, à l'exception des échantillons séreux borderline. Cependant, l'abondance des macrophages et des lymphocytes dans les différents tissus est différente. En effet, les HGSOE et les EC présentent une plus grande abondance de macrophages. Il est intéressant de noter que la signature lymphocytaire des différents tissus est en corrélation avec ce qui a été montré précédemment pour l'abondance de l'ion à  $m/z$  768.55 (**Figure 33, E**). La distribution des différentes populations de cellules obtenues par SpiderMass combiné à LGBM a été confirmées par MALDI-IHC, qui permet une résolution spatiale plus élevée (20  $\mu$ m). Il est intéressant de noter que l'immune-score offre un moyen plus rapide de diagnostiquer l'OC sur la base de la présence de cellules immunitaires dans la TME. De plus, le SpiderMass-MSI, basé sur la notation immunitaire, permet de distinguer les différentes sous-populations présentes dans le TME sans avoir recours à une technique basée sur l'utilisation d'anticorps tels que l'IHC. Il peut être, par la suite être complété par une analyse post-opératoire, basée sur le MALDI-IHC, permettant un diagnostic plus résolu.



**Figure 35:** Diagnostiquer le cancer de l'ovaire en ciblant les cellules immunitaires.

(A) Analyse du tissu mucineux avec le modèle immune-score. Celui-ci prédit une probabilité de présence de cellules cancéreuses, de cellules saines, de lymphocytes et de macrophages pour chaque pixel de l'image réalisée par SpiderMass MSI. Une MALDI-IHC de la zone (carré rouge) a été réalisée avec l'anti-Ki67 (cancer), l'anti-CD8α (LTCD8+) et l'anti-CD68 (macrophages) pour valider le résultat de la SpiderMass-MSI. (B) Analyse du tissu HGSOc avec ce même modèle.

Comme le score immunitaire MSI a permis de déchiffrer les différents sous-types, nous avons décidé d'approfondir cette étude. En effet, nous nous sommes intéressés à la prédiction de la survie des patientes à partir des tissus HGSOC. Pour déterminer le potentiel pronostic du score immunitaire SpiderMass, la signature des macrophages a été subdivisée en sous-types M1 et M2. Des travaux antérieurs ont mis en évidence la relation entre le pronostic des patients et le rapport entre les macrophages de type M1 et M2, connus pour être respectivement pro-inflammatoires et pro-tumoraux<sup>139,145,146</sup>. De plus, l'immune-score basé sur l'IHC a été démontré dans plusieurs cancers comme étant un test robuste pour le pronostic des patients<sup>185,186</sup>. Le score immunitaire SpiderMass a été obtenu à partir de 8 patients HGSC différents répartis en 2 groupes. Le premier groupe est composé de patients sans récurrence, avec une OS >50 mois et toujours en cours de suivi. Le second est composé de patients avec récurrence et OS < 42 mois. Afin d'éviter tout biais lié au traitement reçu par le patient, nous avons réalisé cette analyse uniquement à partir des pièces chirurgicales excisées avant la chimiothérapie. Les résultats montrent une réelle différence entre la présence de macrophages (M1 et M2), de lymphocytes et de cellules cancéreuses. Les tissus des patients dont la survie est plus longue présentent une abondance élevée de lymphocytes et de macrophages de type M1 (**Figure 36, A**) par rapport aux patients dont la survie est plus courte qui, quant à eux, présentent un score plus élevé de cellules cancéreuses et de macrophages de type M2 (**Figure 36, B**). Les scores obtenus pour les macrophages de type M1 et M2 pour chaque patient sont présentés en **Figure 36, C**. Nous avons également calculé le ratio M1/M2 pour chaque patient afin de comparer nos résultats avec les données de la littérature. Un ratio moyen de 6 a été obtenu pour les patients ayant une survie plus longue et de 4 pour ceux ayant une survie plus courte. Pour 3 patients sur 4 ayant une survie courte, le ratio calculé se situe entre 2 et 3. Le dernier patient, quant à lui, présente un ratio anormal proche de 9 augmentant drastiquement la moyenne de ces ratios.

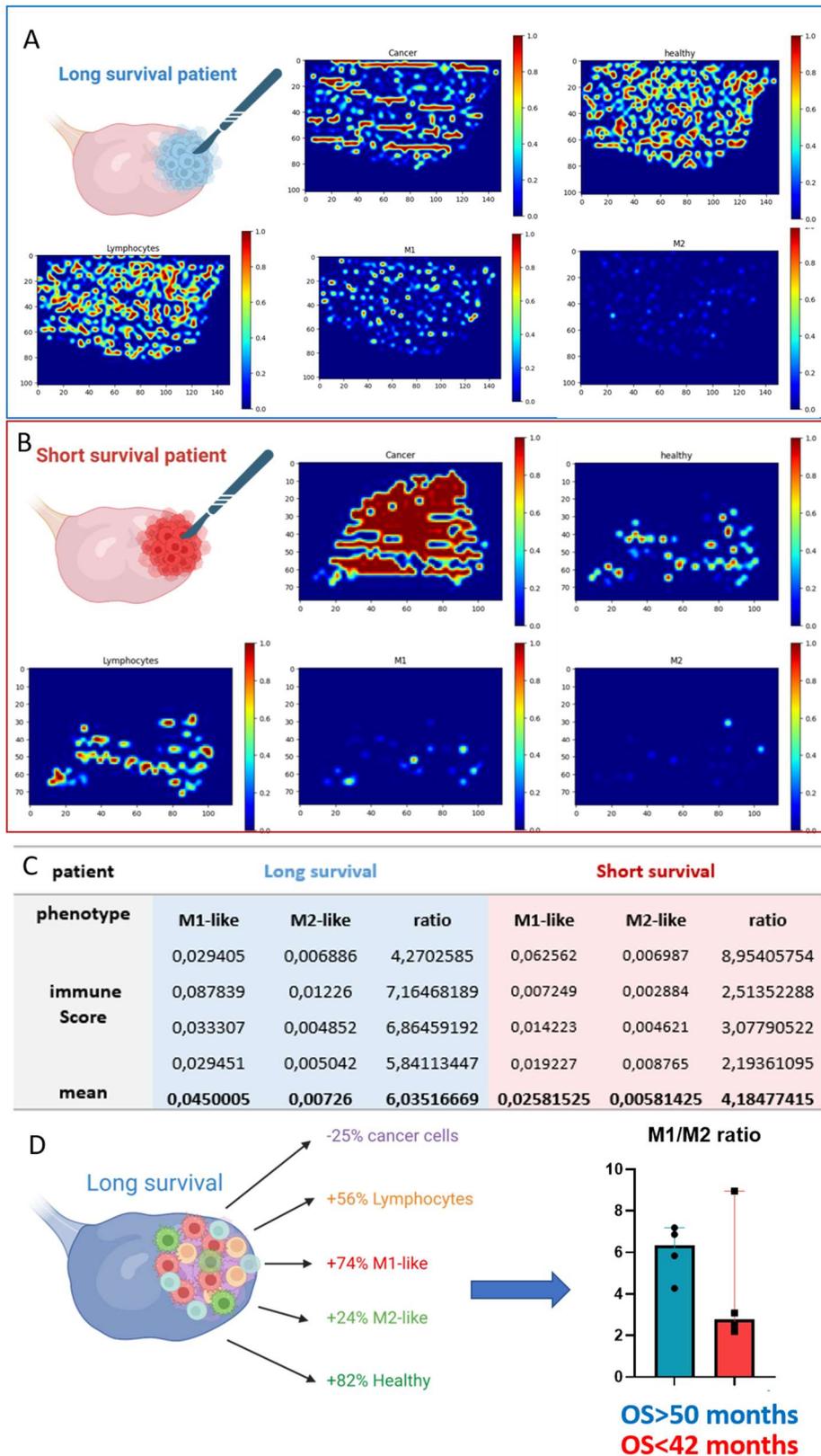


Figure 36: Pronostiquer le cancer de l'ovaire en ciblant les cellules immunitaires.

(A) Exemple de score immunitaire MSI d'un patient avec survie globale <50 mois et toujours en suivi. Les scores de présences évalués sont ceux des lymphocytes, des macrophages de type M1 et M2, des cellules cancéreuses et des cellules saines. (B) Exemple de score immunitaire MSI d'un patient ayant une survie globale <42mois. (C) Tableau représentant le score de présence de macrophages M1-like et M2-like obtenu pour chaque tissu et leur rapport M1/M2. (D) Schéma représentant les

*différences entre la survie à long terme et la survie à court terme en fonction des différents scores. L'histogramme permet de visualiser les différences de ratio entre ces deux catégories de patients.*

En ce qui concerne les scores moyens des différentes populations cellulaires du TME, nous constatons une augmentation de 25 % de la présence de cellules cancéreuses chez les patients ayant une survie plus courte. Nous observons également chez les patients ayant une survie plus longue une augmentation de 56 % de l'infiltration lymphocytaire et une augmentation considérable de 74 % de l'infiltration de macrophages M1. De plus, ceux-ci présentent une abondance très élevée de cellules saines (82 %) (**Figure 36,D**). Cette différence est évidente lorsque l'on compare le rapport M1/M2, ce qui montre qu'un meilleur pronostic est associé à une forte infiltration de macrophages pro-inflammatoires. Les données histopathologiques issues de la littérature sur le CO comprennent 39 % de macrophages répartis en 25 % de macrophages M0, 20 % de macrophages M1, 50 % de macrophages M2 et 4 % de monocytes<sup>140</sup>. Les résultats obtenus dans cette étude sont en ligne avec des données publiées. En effet, une présence de 21,44% de macrophages (M1, M2) en moyenne a été détecté dans les différents sous-types d'OC dans nos expériences. De plus, la prédiction du ratio moyen de lymphocytes/M2 était de 32,10 pour les patients avec une OS > 50 mois contre 24,03 pour les patients avec une OS < 42 mois. Cela met une évidence une infiltration lymphocytaire plus importante chez les patients avec OS > 50 par rapport à l'OS < 42, ce qui est en accord avec ce qui est connu dans la réponse immunitaire liée au cancer, c'est à dire un environnement immunosuppresseur important dans les tumeurs de mauvais pronostic. Cette différence est évidente lorsque l'on compare le rapport M1/M2. En effet, un meilleur pronostic est associé à une forte infiltration de macrophages pro-inflammatoires. Ainsi, nous pouvons utiliser le rapport M1/M2 pour établir le pronostic des patients. Ce rapport peut aider les cliniciens à déterminer le profil d'un patient et à adapter le traitement et la chirurgie en fonction des résultats. Étant donné que le SpiderMass est capable d'effectuer une analyse directe chez l'homme et qu'il est non invasif, cette évaluation pourrait être réalisée en temps réel, pendant l'opération, pour permettre au chirurgien d'effectuer une chirurgie de précision. Si ce modèle est encore simple, il peut être amélioré en augmentant le nombre de différentes sous-populations de cellules immunitaires présente dans le TME. Cependant, il représente une première approche de définition du pronostic du patient qui pourrait être utilisée lors d'une intervention chirurgicale.

## Discussion

La première étape de cette étude a consisté en l'analyse par SpiderMass d'une cohorte exploratoire de 79 pièces chirurgicales composée de 41 tissus sains, 13 HGSOC, 10 SBL, 8 mucineux et 6 endométrioïdes. Cette analyse a conduit à la distinction de profils moléculaires spécifiques à chacun de ces sous-types. Ainsi, il a été possible de bâtir des modèles de classification, basés sur l'analyse statistique multivariée PC-LDA, capables de distinguer, à la fois en mode positif et en mode négatif, l'ensemble de ces classes. Le mode négatif possède un taux de validation de 93% tandis que le mode positif atteint les 95%. Pour valider les capacités de diagnostic de ces modèles, les spectres d'une cohorte en aveugle, constituée de 24 patients, ont été interrogés par le modèle. Comme attendu, les échantillons HGSOC, SBL, sains et mucineux de cette cohorte ont été correctement identifiés selon leur sous-type. Cependant, les échantillons endométrioïdes, sous-représentés, n'ont pas pu être reconnus et ont été identifiés en tant qu'échantillon mucineux par le modèle. C'est pour cela que nous nous sommes focalisés sur l'amélioration de ce modèle afin d'être capable d'identifier l'ensemble de ces sous-types, malgré la quantité limitée d'échantillons. Ainsi, nous avons développé notre propre modèle de classification basé sur la classification RIDGE. Cette classification est basé sur un modèle de régression capable de distinguer des classes sans surentraîner le modèle. Il fonctionne en attribuant des points de pénalités aux pics contribuant négativement à la distinction des différentes classes. L'utilisation de RIDGE a permis l'élaboration d'un modèle encore plus performant, capable de distinguer l'ensemble des sous-types avec une efficacité de 100% en mode négatif et en entraînement et de 97% après une cross-validation à 5fold (20%-out). En mode positif, ce modèle démontre une efficacité de 94%. En associant la cohorte précédemment analysée à une nouvelle cohorte FFPE, il a été possible de générer un modèle RIDGE avec 97% de bonne classification après la cross-validation 20-Fold. Ce modèle a permis une meilleure reconnaissance de la cohorte en aveugle de l'ordre de 95% (en ne prenant pas compte des tissus endométrioïdes).

Afin de mieux comprendre la prise de décision de l'algorithme, nous avons, en parallèle, utilisé l'algorithme LIME. Celui-ci permet de déterminer les ions contribuant positivement à la distinction de la classe, mais également ceux contribuant négativement. En comparant à la fois cette contribution, la distribution de l'intensité des pics dans la cohorte et l'analyse en imagerie MALDI et SpiderMass, nous avons pu isoler une liste de lipides caractéristiques pour chacune des classes. Ces lipides ont été validés par un grand nombre d'analyse et peuvent être considérés comme de potentiels marqueurs de ces sous-types.

Ces lipides ont été identifiés par fragmentation MS/MS au SpiderMass afin de mieux comprendre les métabolismes lipidiques associés à ces différences d'abondance relative. Il convient alors de noter que certains lipides ont également été identifiés comme discriminants

par d'autres études utilisant des outils tels que le MassSpec Pen ou encore le DESI<sup>15,83</sup>. C'est le cas des ions  $m/z$  885.5 (PI (20:4/18:1)), ou  $m/z$  788.54 (PS (18:0/18:1), retrouvés dans les HGSO<sup>83</sup> ou encore des ions 701.51 et 750.5, respectivement identifiés en tant que (PA (18:0/18 :1)) et (PE(P-16:0/20:4)) par le MassSpec Pen<sup>15</sup>. L'identification de la plupart de ces lipides a permis de mettre en évidence certains métabolismes associés aux cancers. En effet, l'augmentation de PS est associée à une rigidification de la bicouche lipidique retrouvée dans différents cancers<sup>187-189</sup>. De plus, l'augmentation de PS dans la membrane externe est liée à aux processus de migration cellulaire et de l'immunosuppression du TME. La présence de PS dans la membrane externe protège la cellule cancéreuse de l'activité des cellules immunitaires<sup>190,191</sup>. Nous avons également observé une augmentation de la quantité d'acide phosphatidique (PA), ce qui corrèle avec son rôle d'activateur d'enzyme kinases de la voie MAPK, connue pour intervenir dans la voie métabolique liée au stress cellulaire<sup>192</sup>.

Ce modèle, déjà très performant, a pu être amélioré grâce à l'association des données moléculaire aux données morphologiques des tissus. Pour ce faire, un modèle à double entrée a été créé. Celui-ci combine les spectres moléculaires générés par SpiderMass et les scans des colorations HPS des mêmes coupes de tissus. Ce nouveau type de modèle de classification a été créé en combinant du Multi-Layer perceptron pour les données moléculaire à un réseau de neurones convolutif 2D (2D-CNN) des scans HPS. Ce nouveau type de modèle a permis l'élaboration d'une classification bien plus robuste et sensible avec une efficacité de 100% à la fois en entraînement et en test. Ainsi, un degré de complexité et de compréhension supplémentaire des données a été ajouté au modèle en y intégrant les données morphologiques aux données moléculaires conventionnelles<sup>57</sup>.

Seulement, ces modèles ainsi générés ne prennent pas en compte la complexité cellulaire au sein de la tumeur. En effet, les cellules présentent dans ces tissus ne sont pas exclusivement des cellules cancéreuses. Celles-ci sont insérées dans un microenvironnement tumoral complexe et hétérogène. Ainsi, la suite de l'étude s'est consacrée à étudier cet environnement afin de préciser ces différentes distributions lipidiques. Dans un premier temps, un modèle PC-LDA a été réalisé sur un grand nombre d'analyse de différentes lignées cellulaires de cancer de l'ovaire, de macrophages primaires et de THP1 orientés vers les phénotypes M1 ou M2, de cellules saines et de lymphocytes primaires (CD8+, CD4+ et NK). Les résultats ont permis de prouver la capacité du SpiderMass à discriminer l'ensemble de ces types cellulaires. De plus, le modèle a défini des marqueurs spécifiques pour l'ensemble de ces cellules tels que la surexpression des ions  $m/z$  818.65 et 846.65 dans les macrophages par rapport aux cellules cancéreuses. Ces ions ont été identifiés en tant que glucosylcéramides GlcCer (d18:1/22:0) et GlcCer(d18:1/24 :0). De manière intéressante, ces lipides sont connus pour être responsable de la régulation de l'inflammation notamment lors d'une infection bactérienne<sup>193</sup>.

Les lipides discriminants les macrophages M2-like des macrophages M1-like sont, pour la plupart, des phosphatidylglycérides polyinsaturée à longues chaînes. Cette observation est corrélée par d'autres études démontrant une augmentation de ces lipides dans les cellules immunitaires associées aux cancers<sup>194</sup>. Les lymphocytes ont également pu être discriminés du reste des types cellulaires par d'autres marqueurs lipidiques spécifiques tels que les ions  $m/z$  768.55,  $m/z$  885.55 et  $m/z$  738.55 correspondant aux PE (38:3), PI (20:4/18:0) et au PE (20:4/16:0). De plus, la distribution des lipides spécifiques des cellules immunitaire diffère selon les sous-types et est en corrélation avec l'agressivité de ces cancers. En effet, l'intensité de ces pics semble plus intense dans les HGSOE.

Cependant, les échantillons endométrioïdes montrent également une forte abondance dans les masses associées aux macrophages M1 like, traduisant une grande infiltration de ces derniers dans le tissu tumoral. En effet, le système immunitaire est très présent dans l'endomètre dont est issu ce sous-type cancéreux. Celui-ci est lié notamment à l'endométriose où beaucoup de lésion tissulaire en résulte et donc où le système immunitaire est extrêmement sollicité.

Afin d'avoir une vision plus précise de cette problématique nous avons développé une méthode capable de distinguer, non pas de simples pics, mais le profil moléculaire de chaque sous-type cellulaire au sein d'un tissu. Pour ce faire, nous avons utilisé le modèle Light Gradient Boosting Machine (LGBM). Il a permis de distinguer, dans un premier temps, l'ensemble des types cellulaires avec une efficacité de 100%. Il a ensuite été possible d'interroger la cohorte de tissus analysés par ce modèle afin de définir, pour chacun des sous-types d'OC, les scores de présence de l'ensemble de ces cellules. Nous avons donc pu créer un profil immunitaire localisé au niveau des tirs laser effectués au SpiderMass. L'étape suivante était d'obtenir la distribution globale de l'ensemble de ces cellules immunitaires au sein des différents tissus. Pour ce faire, chaque sous-type a été analysé en triplicat par imagerie SpiderMass. Chaque pixel de ces images a été interrogé par ce modèle. Il en résulte de nouvelles images basées uniquement sur le score de présence des cellules cancéreuses, des cellules saines, des macrophages M1 et M2 et des lymphocytes. Nous avons appelé ce score « immune-score ». Cet immune-score a été cross-validé par une autre technique, l'imagerie MALDI-IHC. Celle-ci consiste en une imagerie réalisée sur des anticorps spécifiques tels que CD68 (macrophages), Ki67 (cancer) ou CD8 (lymphocytes TCD8+) liés à des sondes photo-clivables<sup>179</sup>. La combinaison de ces deux techniques a pu confirmer la robustesse des résultats. De plus, elle a permis une compréhension globale de la distribution de ces cellules (SpiderMass) et la localisation précise de ces dernières avec une résolution de 20 $\mu$ m (MALDI-IHC). Il a été mis en évidence des différences visibles entre les différents sous-types cancéreux. En effet, nous avons constaté une proportion très importante de macrophages

dans les tissus HGSOC et SBL avec une disposition très homogène de ceux-ci dans tout le tissu. Au contraire les échantillons mucineux présentent une hétérogénéité importante dans la distribution des macrophages. En effet, il est possible de voir apparaître des « hot spots » macrophagique dans le tissu avec une très grande concentration de macrophages, très localisés et peu de signal autour de ceux-ci. Cette technique peut donc permettre un diagnostic visuel rapide des sous-types cancéreux à l'aide de leur score immunitaire.

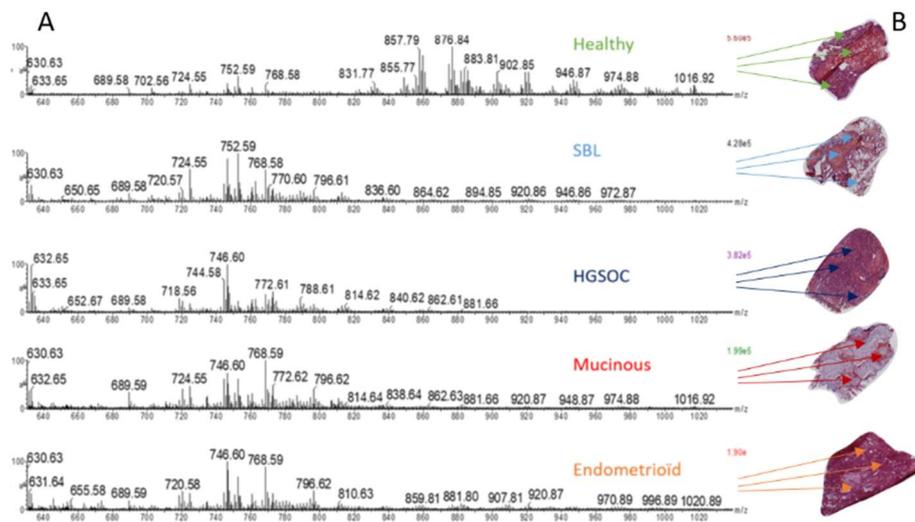
La dernière partie de notre étude a associé ce score immunitaire à la survie des patients. Pour ce faire, 8 échantillons, dont 4 avec une survie <42 mois et 4 une survie >50mois. Leur score immunitaire a été établi et comparé. La différence entre les scores immunitaires permet la distinction des groupes de survie. En effet, nous avons démontré une augmentation de 25 % de cellules cancéreuses chez les patients avec une OS <42 mois. Tandis que les patients avec une OS>50 mois présentent une augmentation de 56 % de l'infiltration lymphocytaire, une augmentation de 74 % de l'infiltration de macrophages type M1 et enfin, une ressemblance plus marquée avec les cellules saines. Afin de corroborer ces résultats avec les études réalisées sur ce domaine, le ratio entre la quantité de M1 et de M2 a été étudié<sup>139,145,146</sup>. Il en résulte un rapport moyen de 6 pour les patients avec une survie prolongée et de 4 pour les patients avec une survie courte. Ainsi, nous avons mis au point une nouvelle méthode de pronostic rapide qui pourrait être utilisée au bloc opératoire durant la chirurgie. En travaillant sur de plus grandes cohortes et en améliorant ce modèle, il serait alors possible de guider directement les cliniciens sur le traitement à apporter à la patiente directement après la chirurgie.

### Conclusion

Au cours de cette étude, nous avons développé une stratégie efficace pour diagnostiquer les sous-types les plus abondants du cancer de l'ovaire en se basant sur le profilage moléculaire et le profilage des cellules immunitaires. Un modèle à double entrée a été créé en couplant les informations moléculaires et morphologiques, ce qui a conduit à une excellente précision de 99 % et a été complété par la découverte de certains marqueurs lipidiques spécifiques des différents sous-types. Pour mieux comprendre le microenvironnement tumoral dans l'OC, un modèle LGBM a été construit en se basant sur le profil lipidique de chaque type de cellules (CD4+, CD8+, M1, M2, cancer). Cela a ouvert une nouvelle voie pour différencier les tissus cancéreux en fonction de la distribution des cellules immunitaires. Enfin, nous sommes allés plus loin dans l'analyse en testant ce modèle sur les données de survie des tumeurs HGSOC. Les résultats ont démontré une infiltration lymphocytaire et de macrophages M1 chez les patientes ayant une survie prolongée. Cette preuve de concept ouvre de nouvelle opportunité dans le guidage du chirurgien. Certains patients atteints d'OC ont une survie sans récurrence <6mois conduisant à une OS très faible

(<1ans). Jusqu'alors, il est impossible de prédire ce l'agressivité de ce type de tumeur avant la première récurrence. Ainsi, l'imagerie basée sur le score immunitaire pourrait devenir un nouvel outil de diagnostic de ces patientes. En effet, elle permettrait une définition du score immunitaire rapide en démontrant notamment une forte activité immunosuppressive chez ces patientes. De ce fait, les cliniciens pourraient établir un stratégie thérapeutique différente pour améliorer la prise en charge de ces patientes.

Figures supplémentaires



**Figure 29, Supp data 1** : Empreinte morphologique et spectrale des sous-types de cancer de l'ovaire en mode ionique positif. (A) Spectres de chaque sous-type de cancer de l'ovaire en mode ionique négatif avec une taille de spot de 500µm de diamètre avec un burst de 10tirs à 4mJ. (B) Coloration HPS de ces tissus spécifiques avec le spot utilisé pour l'acquisition.



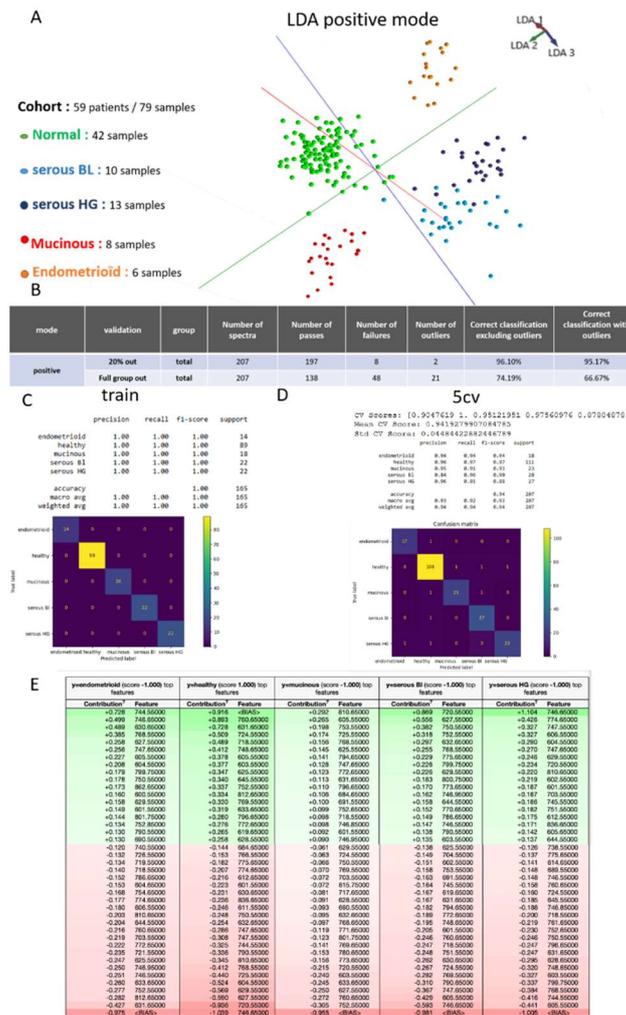


Figure 29, Supp data 3 : Diagnostic de l'OC en mode positif.

(A) Modèle LDA des différents sous-types de cancer en mode ion positif. (B) Tableau représentant le score de validation de ce modèle avec et sans outliers. (C) Entraînement du modèle basé sur le classificateur RIDGE. (D) Rapport de la validation croisée 5-fold du classificateur RIDGE. (E) Analyse LIME montrant l'ion discriminante pour chaque classe du modèle.

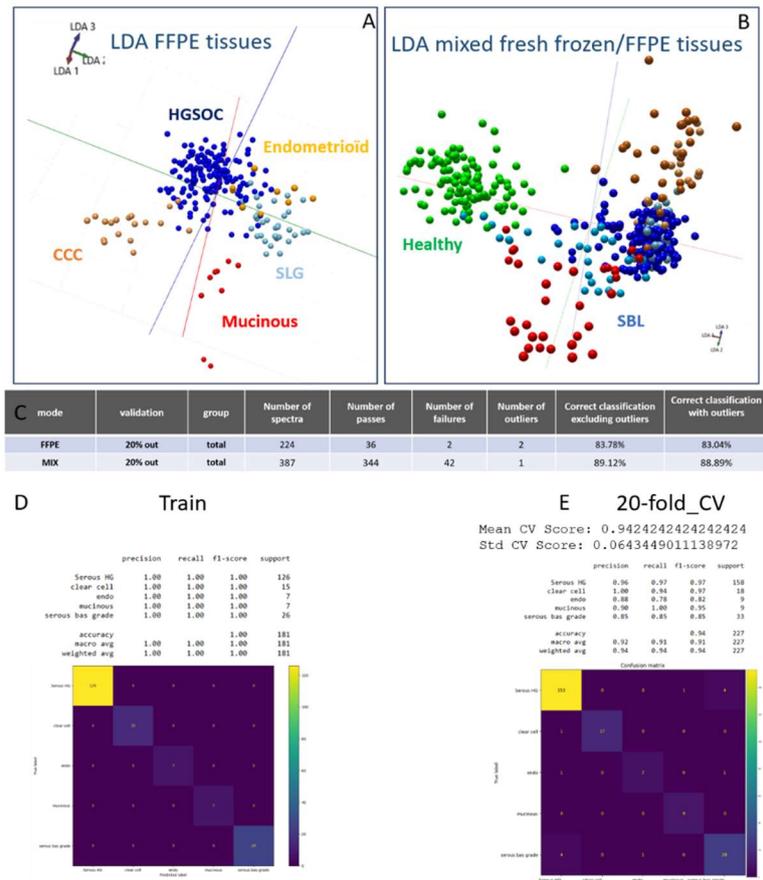
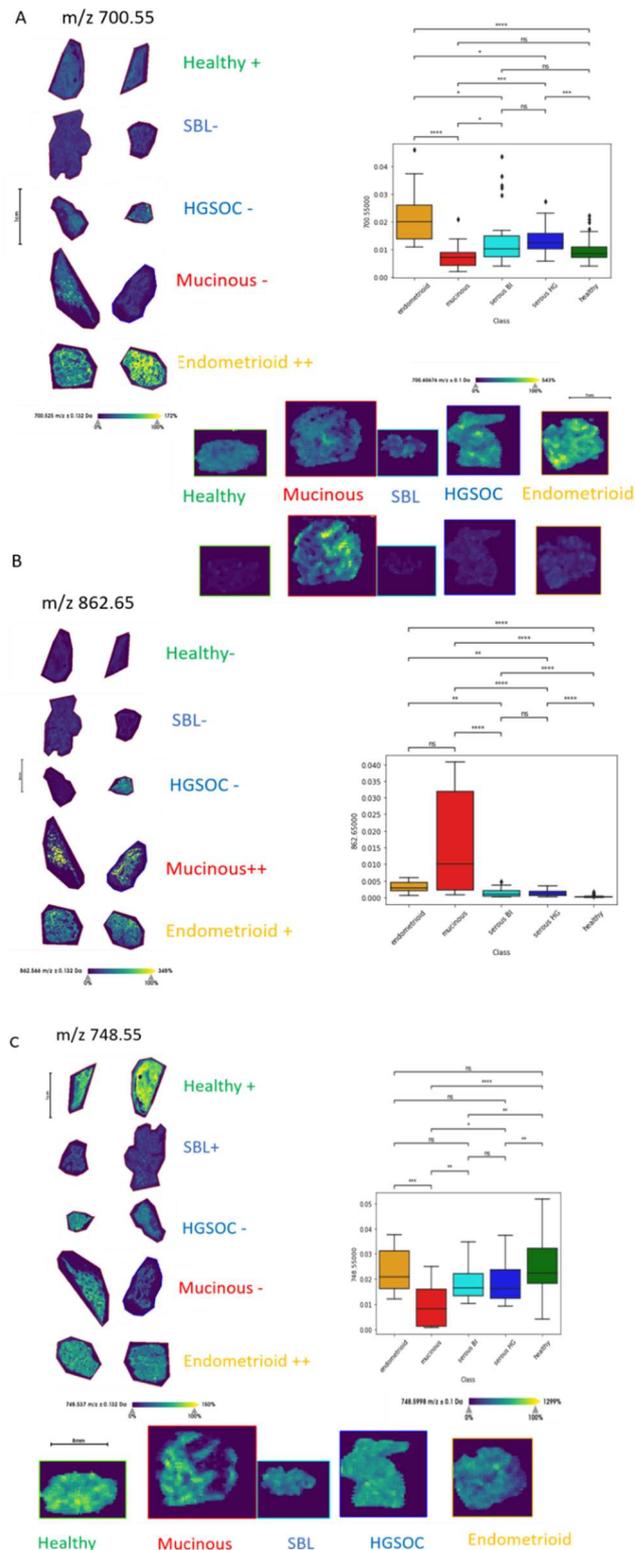


Figure 29, supp data 4 : Modèle de classification réalisé sur la cohorte FFPE.

(A) Analyse discriminante linéaire basée sur l'acquisition de tissu FFPE sur les tissus HGSOE, SBL, SLG, CCC, mucineux et endométrioïdes en mode ion négatif. (B) Modèle précédent mélangé avec le modèle LDA de la cohorte de tissus frais congelés. (C) Tableau représentant le score de validation du modèle (A) et (B) avec et sans outliers. (D) Entraînement du modèle RIDGEClassifier sur la cohorte FFPE. (E) Cross-validation 20-Fold du modèle.



**Figure 32, supp data 5 :** Exemples de potentiels biomarqueurs.

(A) Exemple d'un bon marqueur de cancer avec m/z 700,55. La contribution positive de l'ion par LIME est représentée par un (+) et la contribution négative par un (-). Un test de Kruskal-wallis a été effectué sur les données de l'analyse SpiderMass pour cet ion et représenté sous forme de diagramme en boîte. (\*\*\*  $p < 0,0001$ , \*\*  $p < 0,001$ , \*  $p < 0,01$ , \*  $p < 0,05$ ) NS pour  $p > 0,05$ ). Cet ion a ensuite été validé par MALDI-MSI avec une matrice de norharmane et par SpiderMass-MSI. (B) Distribution de l'ion 862.65 m/z, marqueurs des tissus mucineux. (C) Distribution de l'ion 748.55 m/z.

Mode	sample	Potential identification	Peak m/z	
Negative	Normal	PE (16:0-22:5)	748,55	
	Normal	PS (18:0/18:1)	788,55	
	Normal	PI (20:4/18:0)	885,55	
	Normal	PA (18:1:18:0)	701,55	
	Serous BL	PI (22:1/18:0)	919,79	
	Serous BL	TG (56:12)	893,63	
	Serous BL	PE (18:0/22:4)	794,55	
	Serous HG	PE (18:0/16:0)	718,55	
	Serous HG	PE(18:0/18:2)	742,55	
	[M-H]-	Serous HG	PA (18:1/20:2)	725,55
	[M-H]-	Serous HG	PE (18:0/18:1)	744,55
	[M-H]-	Serous HG	PA (18:0/22:6)	747,55
	[M-H]-	Mucinous	PA (18:0/20:2)	727,55
	[M-H]-	Mucinous	PS (22:2/20:4)	862,65
	[M-H]-	Endometrioid	PS (22:0/22:2) or PS (22:1/22:1)	898,55
	Positive	Serous BL/HG	PE (16:0/20:0)	748,55
		Serous BL/HG	PE (18:2/22:4)	792,55
[M-H]-		Serous BL	PS (P-16:0/18:2)	744,55
[M-H]-		Serous BL	PS (18:4/18:1)	782,55
[M-H]-		Serous HG	PS (0-16:0V20:1)	776,55
[M-H]-		Serous HG	PE (0-18:0/18:2)	730,55
[M-H]-		Mucinous	PS (0-18:0/18:1)	776,55
[M-H]-		Mucinous	PC (20:1/16:1)	788,65
[M-H]-		Mucinous	PE (0-18:0/20:4)	752,55
[M-H]-		Mucinous	PC (18:4/12:0)	698,55
[M-H]-		Endometrioid	PE (16:0/20:5)	724,55

Figure 32, supp data 6 : Tableaux récapitulatifs des ions marqueurs identifiés.

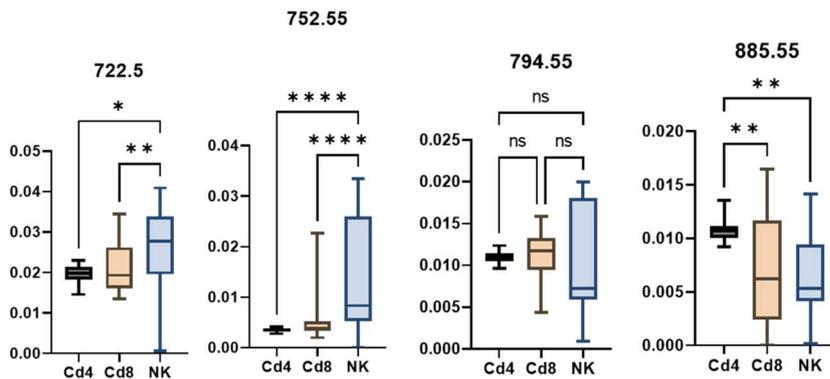


Figure 33, supp data 7 : Listes des ions discriminants les différents sous-types lymphocytaires.

Boxplot des intensités relatives des ions m/z 722.55, 752.55, 794.55 et 885.55 en fonction du phénotype des lymphocytes (NK, CD8 ou CD4) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS pour  $p > 0.05$ ).

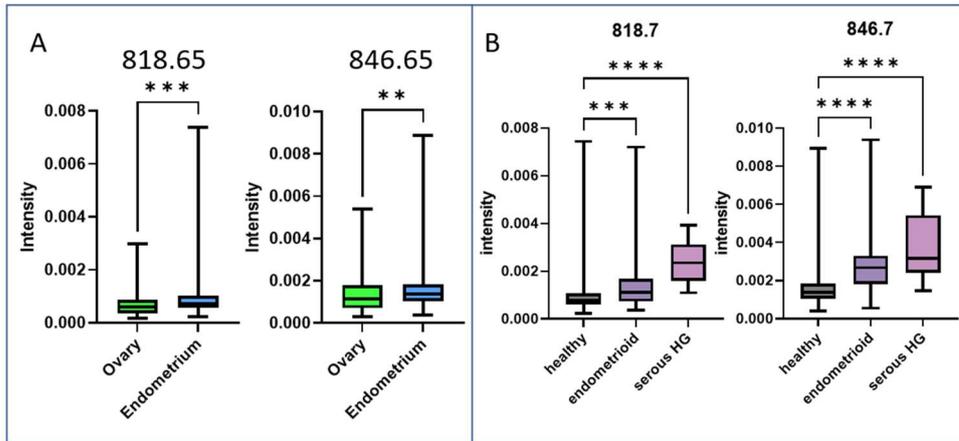
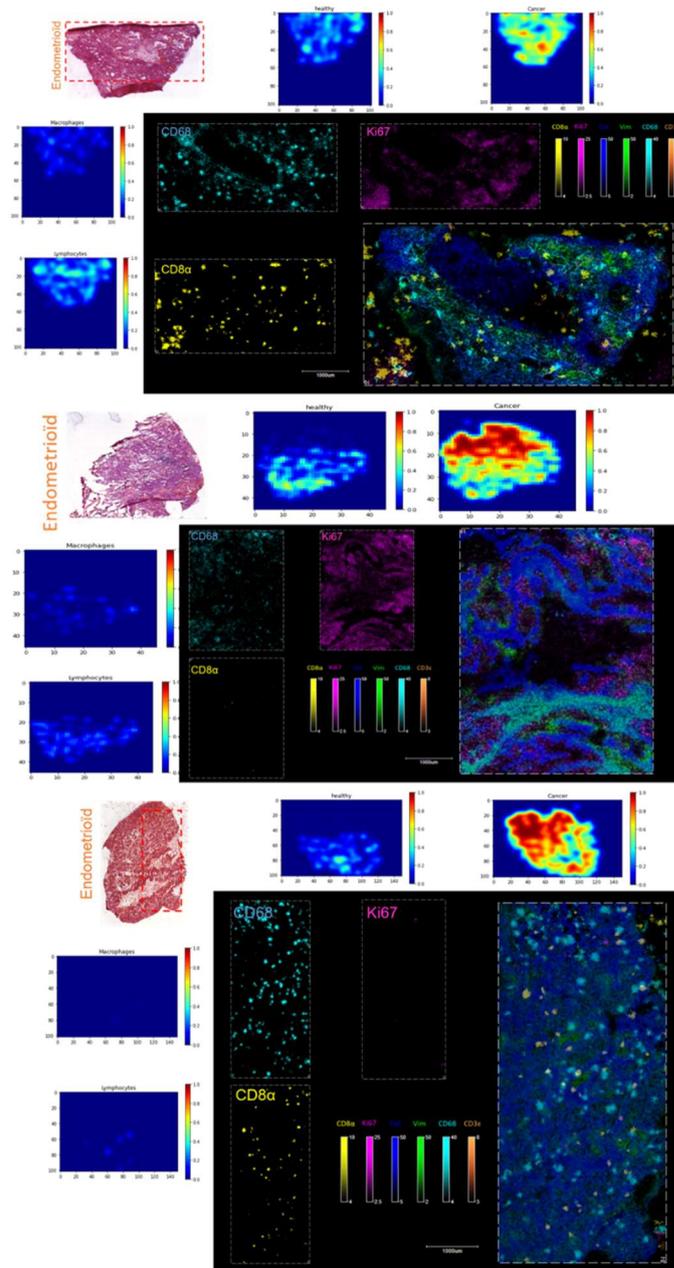


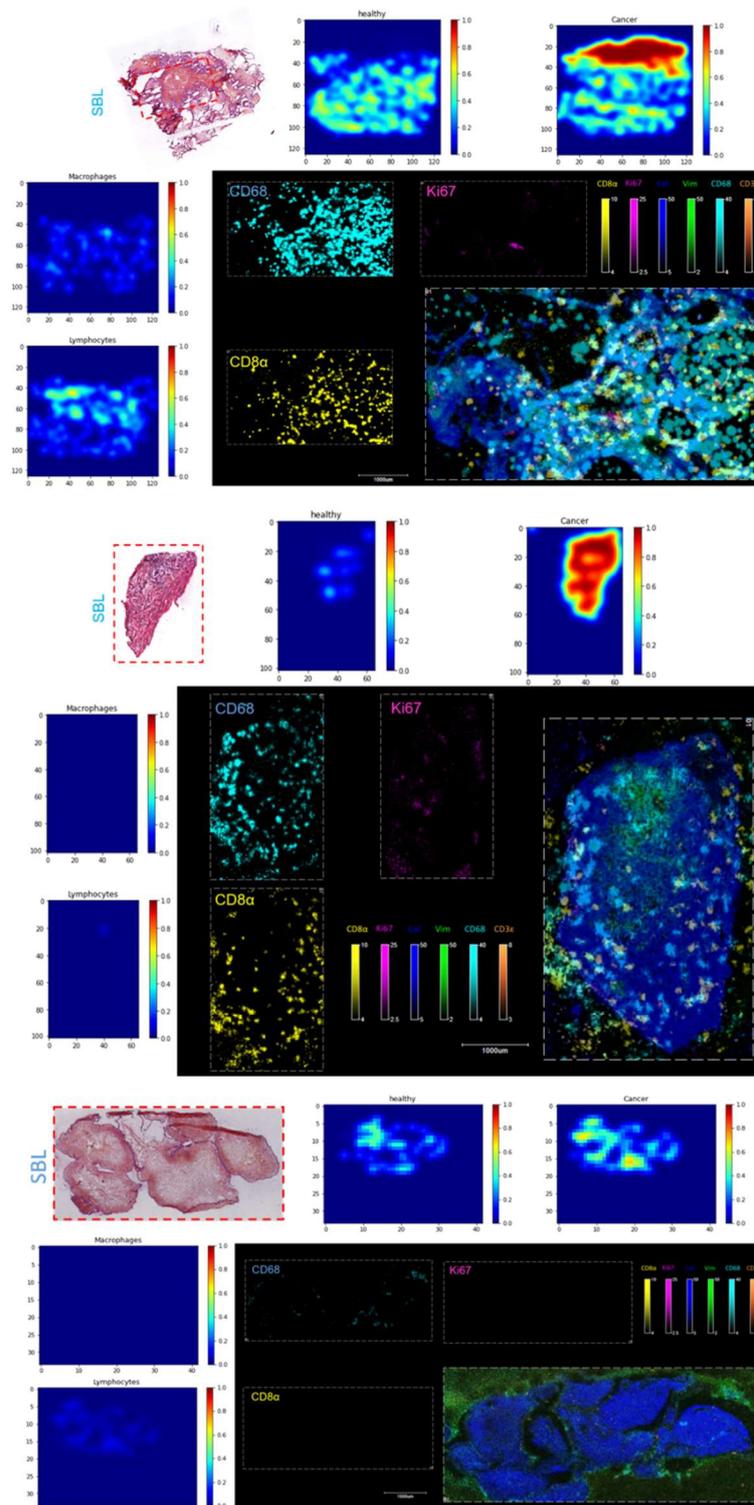
Figure 33, supp data 8 : Distribution des marqueurs macrophagiques dans le cancer de l'endomètre.

(A) Boxplot des intensités relatives des ions m/z 818.65 et 846.65 en fonction du tissu (endomètre ou ovaire) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ). NS pour  $p > 0.05$ ). (B) Boxplot des intensités relatives des ions m/z 818.65 et 846.65 en fonction du phénotype de l'endomètre (sain, HGSOE ou endométrioïde)



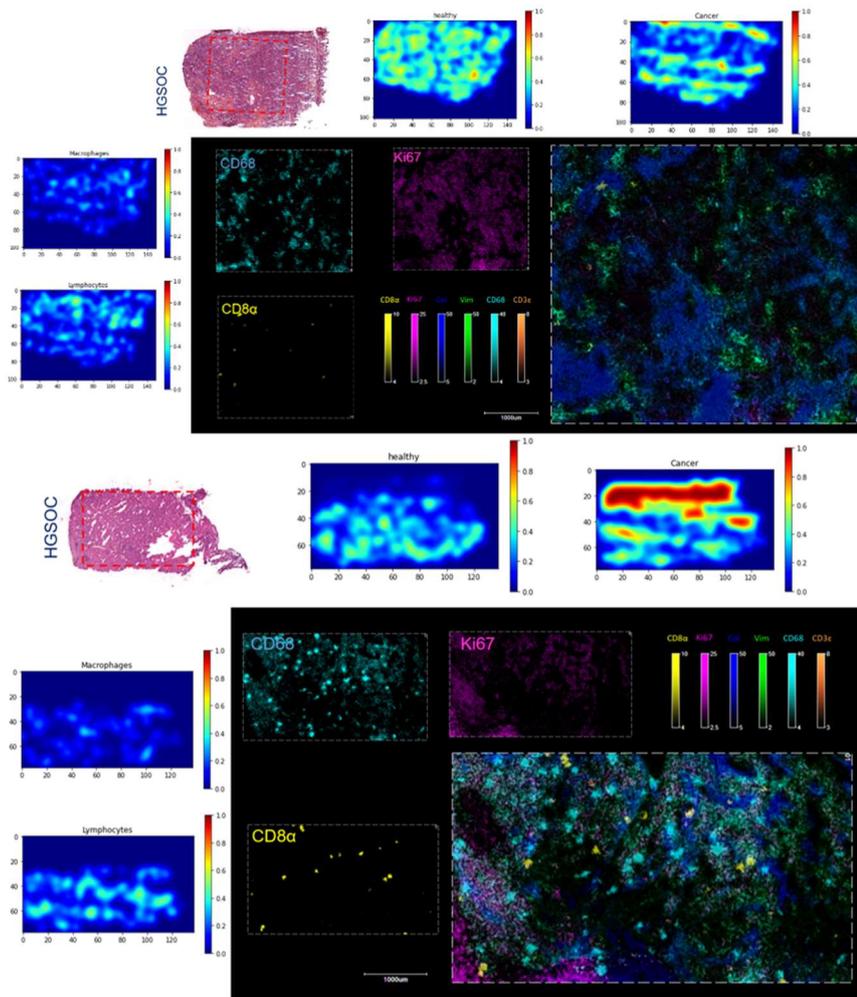
**Figure 35, supp data 9 :** Immune-score SpiderMass-MSI sur les tissus endométriaux.

Analyse des tissus endométriaux avec le modèle immune-score. Un score de présence de cellules cancéreuses, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Une MALDI-IHC de la zone (carré rouge) a été réalisée avec l'anti-Ki67 (cancer), l'anti-CD8 $\alpha$  (LTCD8) et l'anti-CD68 (macrophages) pour valider le résultat de la MSI de SpiderMass.



**Fig 35, supp data 10 :** Immune-score SpiderMass-MSI sur les tissus SBL.

Analyse des tissus SBL avec le modèle immune-score. Un score de présence de cellules cancéreuse, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Une MALDI-IHC de la zone (carré rouge) a été réalisée avec l'anti-Ki67 (cancer), l'anti-CD8α (LTCD8) et l'anti-CD68 (macrophages) pour valider le résultat de la MSI de SpiderMass.



**Figure 35, supp data 11** : Immune-score SpiderMass-MSI sur les tissus HGSOC.

Analyse des tissus HGSOC avec le modèle immune-score. Un score de présence de cellules cancéreuse, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Une MALDI-IHC de la zone (carré rouge) a été réalisée avec l'anti-Ki67 (cancer), l'anti-CD8α (LTCD8) et l'anti-CD68 (macrophages) pour valider le résultat de la MSI de SpiderMass.

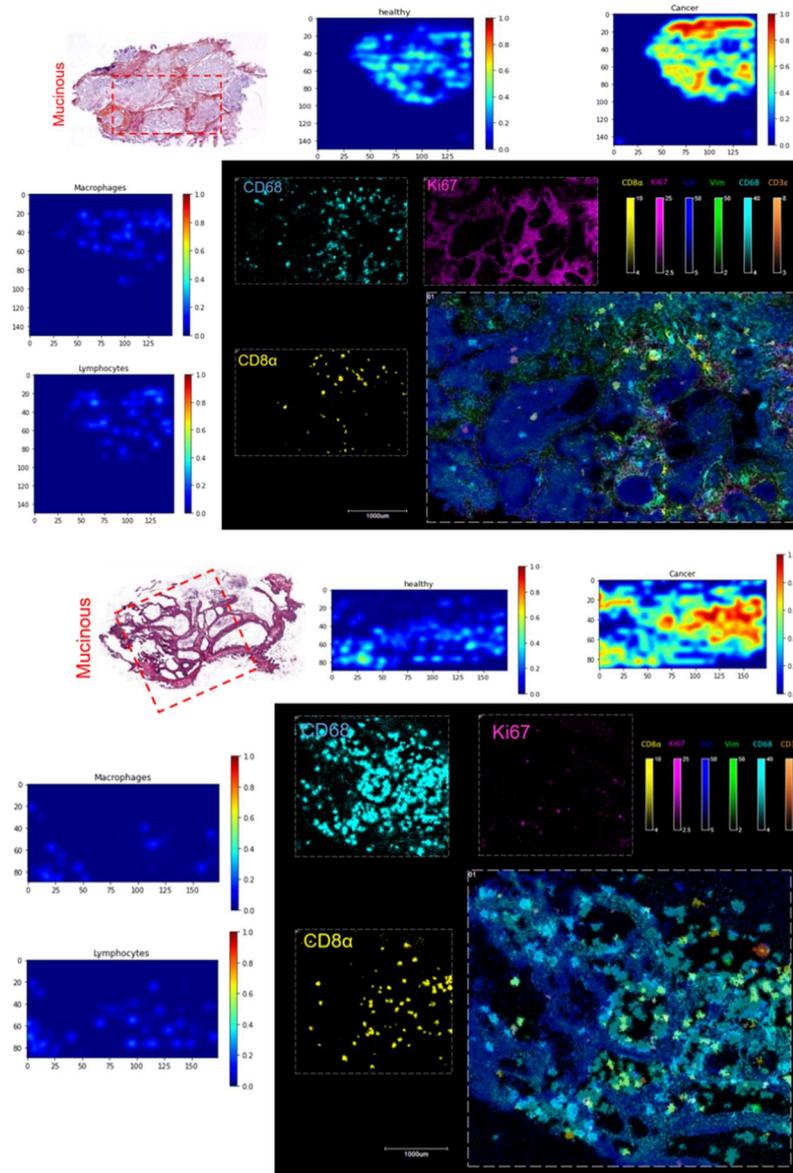


Figure 35, supp data 12 : : Immune-score SpiderMass-MSI sur les tissus mucineux.

Analyse des tissus mucineux avec le modèle immune-score. Un score de présence de cellules cancéreuse, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Une MALDI-IHC de la zone (carré rouge) a été réalisée avec l'anti-Ki67 (cancer), l'anti-CD8α (LTCD8) et l'anti-CD68 (macrophages) pour valider le résultat de la MSI de SpiderMass.

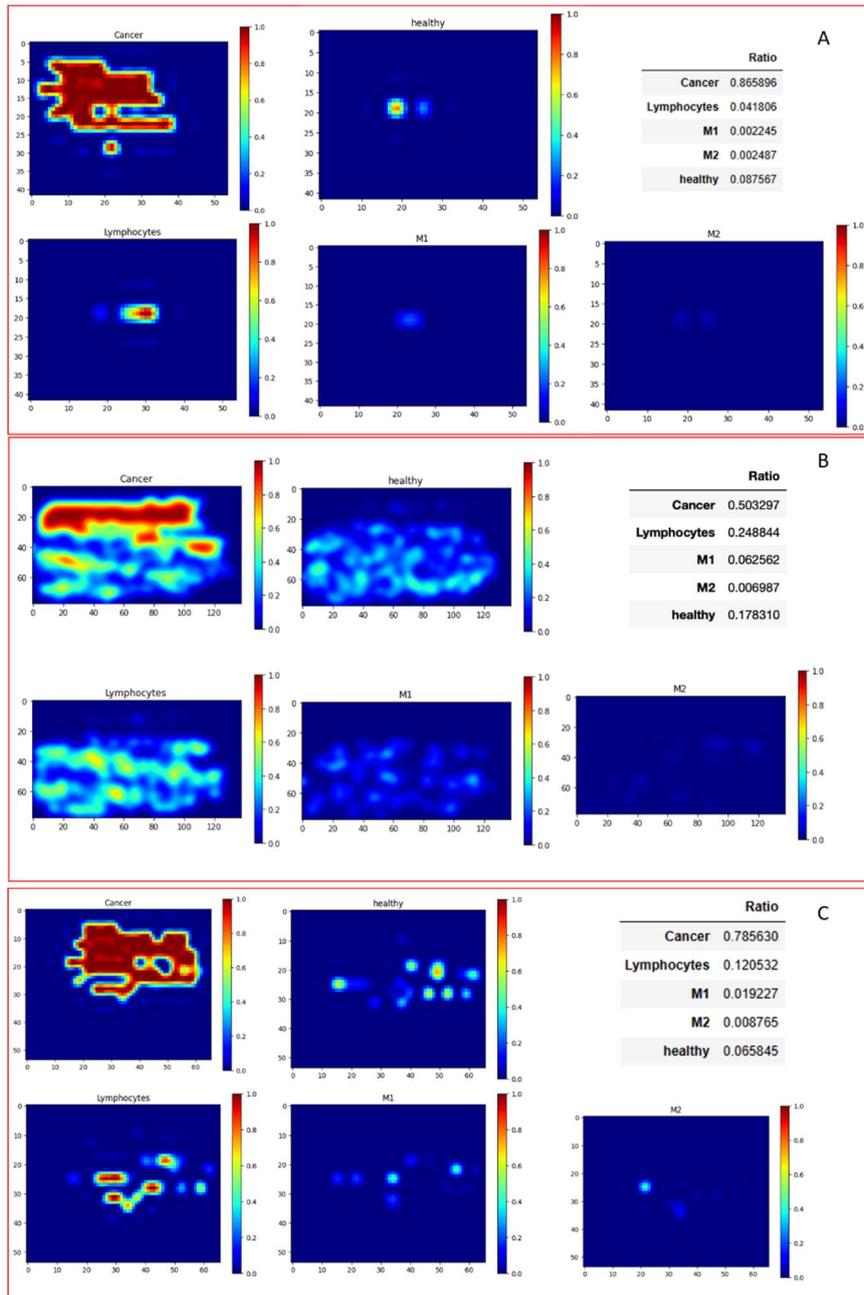


Figure 36, supp data 13 : Immune-score SpiderMass-MSI sur les tissus HGSOc de mauvais pronostic.

Analyse des tissus HGSOc de bon pronostic avec le modèle immune-score. Un score de présence de cellules cancéreuse, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Le score de présence RSP pour chaque classe est indiqué. A,B,C correspondent à des patients distinctes.

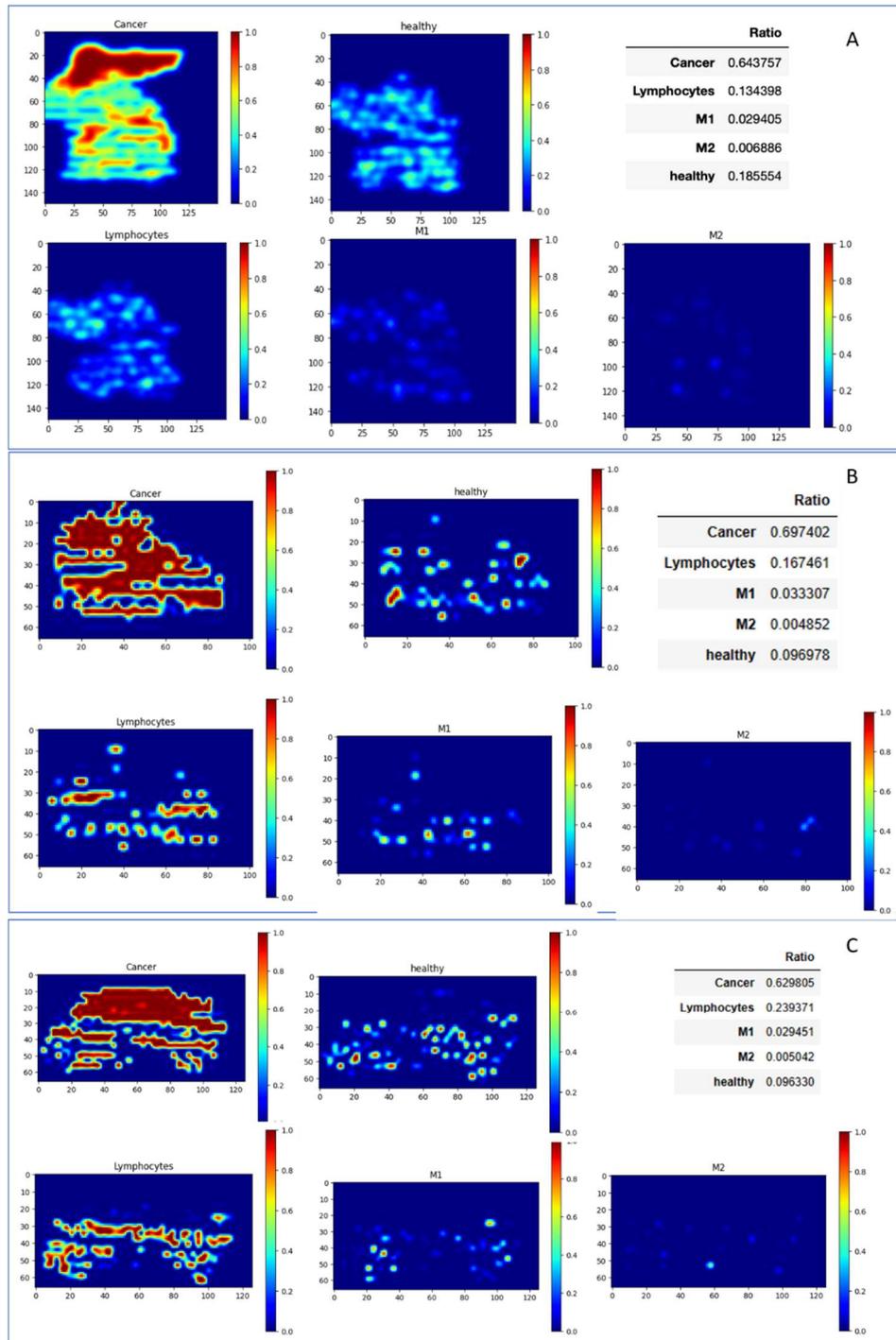


Figure 36, supp data 14 : Immune-score SpiderMass-MSI sur les tissus HGSOc de bon pronostic.

Analyse des tissus HGSOc de mauvais pronostic avec le modèle immune-score. Un score de présence de cellules cancéreuse, de cellules saines, de lymphocytes et de macrophages pour chaque pixel a été attribué sur l'image réalisée par SpiderMass MSI. Le score de présence RSP pour chaque classe est indiqué. A,B,C correspondent à des patients distinctes.

## CHAPITRE V : Analyse du TME du glioblastome par SpiderMass pour l'aide à la prise de décision du chirurgien.

La notion d'immune-score que nous avons pu introduire dans notre étude sur le cancer de l'ovaire serait un outil d'aide à la décision des plus utiles pour les neurochirurgiens. En effet, le glioblastome est une tumeur très hétérogène avec un pronostic très sombre. Le SpiderMass, associé à l'immune-score, pourrait être un atout indéniable au bloc opératoire pour stratifier les patients et permettre une prise en charge personnalisée avec la mise en place de thérapies spécifiques telles que la thérapie photodynamique<sup>195</sup> ou ce que l'unité a récemment mis en place, la photothérapie directe, appelée IROSMASS (INPI, DSO2023006838). Le laboratoire PRISM a pu démontrer que les patients ayant la moins bonne survie étaient ceux présentant une infiltration immunitaire, notamment macrophagique, plus importante. C'est pour cela qu'une détection précoce de l'abondance de macrophage est un enjeu majeur pour la survie du patient.

### Introduction

Les tumeurs du système nerveux central (SNC) et du cerveau sont les cancers les plus mortels. En effet, 308 102 nouveaux cas sont diagnostiqués dans le monde et 251 329 décès sont recensés chaque année<sup>196</sup>. Ces cancers sont avant tout représentés par les gliomes qui représentent 80% des tumeurs malignes du SNC<sup>197</sup>. Différents grades sont décrits en fonction de la morphologie, du degré de malignité et, plus récemment du profil moléculaire. Les glioblastomes sont le grade le plus commun et le plus agressif des gliomes<sup>198</sup>. Ce sont des tumeurs très hétérogènes, hautement invasives, immunosuppressives et incurables. La classification actuelle des gliomes repose sur la cinquième édition de la Classification des tumeurs du système nerveux central de l'OMS (WHO CN5, publiée en 2021)<sup>199</sup>. Le premier critère de cette classification est la mutation IDH. Les gliomes IDH non mutés, ou sauvage, sont classifiés en tant que glioblastomes. Les gliomes IDH mutés sont ensuite stratifiés en astrocytomes ou oligodendrogliomes de différents grades. L'enzyme IDH joue un rôle dans les processus métaboliques liés à l'énergie tel que le cycle de Krebs<sup>200</sup>. Le statut IDH est également important pour le pronostic des patients<sup>201</sup>. Jusqu'à présent, la prise en charge du glioblastome de type IDH sauvage implique une approche multimodale comprenant une résection chirurgicale maximale en première ligne, suivie d'une radiothérapie associée à une chimiothérapie adjuvante avec du témozolomide (connue sous le nom de protocole Stupp<sup>202</sup>). Ce traitement est malheureusement resté inchangé pendant de nombreuses années. Malgré les différents traitements liés à cette pathologie, la médiane de survie de ces patients n'excède

pas 17 mois<sup>203</sup>. Il est rare que les nouvelles stratégies thérapeutiques dépassent le stade d'essai clinique, du fait de leur manque d'efficacité<sup>204</sup>. Les raisons de ces difficultés sont multiples. Premièrement, les traitements doivent être capables de traverser la barrière hémato-encéphalique afin de pouvoir agir au niveau de la tumeur. Deuxièmement, ces tumeurs sont très hétérogènes et ont donc toutes des voies de signalisation tumorales différentes. Cela implique donc une connaissance précise et personnalisée du phénotype tumoral pour adapter la prise en charge. Pour finir, l'environnement tumoral a un très grand impact dans le glioblastome. En effet, ce type de tumeur est considéré comme très immunosuppresseur. Les TAMs (macrophages associés à la tumeur) représentent une proportion très importante de la masse tumorale. Ils sont représentés par deux types cellulaires : les cellules microgliales et les macrophages. À eux seuls, ils représentent entre 30 et 50% du microenvironnement tumoral dans le glioblastome<sup>205</sup>. Cette densité macrophagique permet la défense de la tumeur contre le système immunitaire. Elle permet également la croissance tumorale et favorise, à la fois l'invasion des tissus environnants, et la vascularisation au sein de la tumeur<sup>206</sup>. De plus, les TAMs ont un impact direct sur le traitement des patients. En effet, ils sont impliqués dans les processus de chimiorésistance. C'est pour cela que la chirurgie reste le traitement le plus efficace pour l'augmentation de la survie globale des patients. Il a été observé que la survie globale était plus élevée lorsque la résection était totale par rapport à une résection subtotale<sup>207</sup>. L'intérêt est donc de réaliser une résection de plus de 95% de la tumeur tout en préservant au maximum le tissu environnant pour garantir l'activité fonctionnelle de celui-ci. Seulement, le glioblastome est de nature très diffuse et il est donc difficile de définir précisément la zone tumorale. C'est dans ce contexte précis que le SpiderMass, associé à de l'intelligence artificielle (nouvelles méthodes de Machine Learning), a été développé et mis en situation dans cette étude.

### Objectifs

L'objectif global de cette étude est d'améliorer la définition de la marge en stratifiant les patients selon leur profil moléculaire. L'objectif est également d'améliorer le pronostic de ceux-ci en fonction de marqueurs spécifiques et de leur profil immunitaire afin d'aider le clinicien sur les stratégies de prise en charge des patients en fonction de la nature de la tumeur (**Figure 37**).

L'objectif de cette étude est, dans un premier temps, de créer de nouvelles méthodes d'analyses permettant une classification sensible et robuste des glioblastomes. Ces méthodes d'analyses ont été rendus plus facile d'utilisation afin d'aider le chirurgien dans sa prise de décision. De plus, elles ont permis la caractérisation de nouveaux biomarqueurs en lien avec les différents sous-types de glioblastome. Ces marqueurs ont été corrélés à la survie globale du patient.

En parallèle, nous avons étudié l'immune score des tissus de GBM pour préciser l'infiltration immunitaire en fonction des sous-groupes. L'impact du système immunitaire a été corrélé à l'évolution de la tumeur chez le patient, en incluant, de manière quantitative, le ratio des macrophages et des lymphocytes.

Dans ce projet collaboratif, je me suis concentré sur l'adaptation du modèle immune-score pour l'étude du glioblastome. J'ai également participé à l'utilisation de ce modèle dans la définition du pronostic chez les patients atteints de GBM.

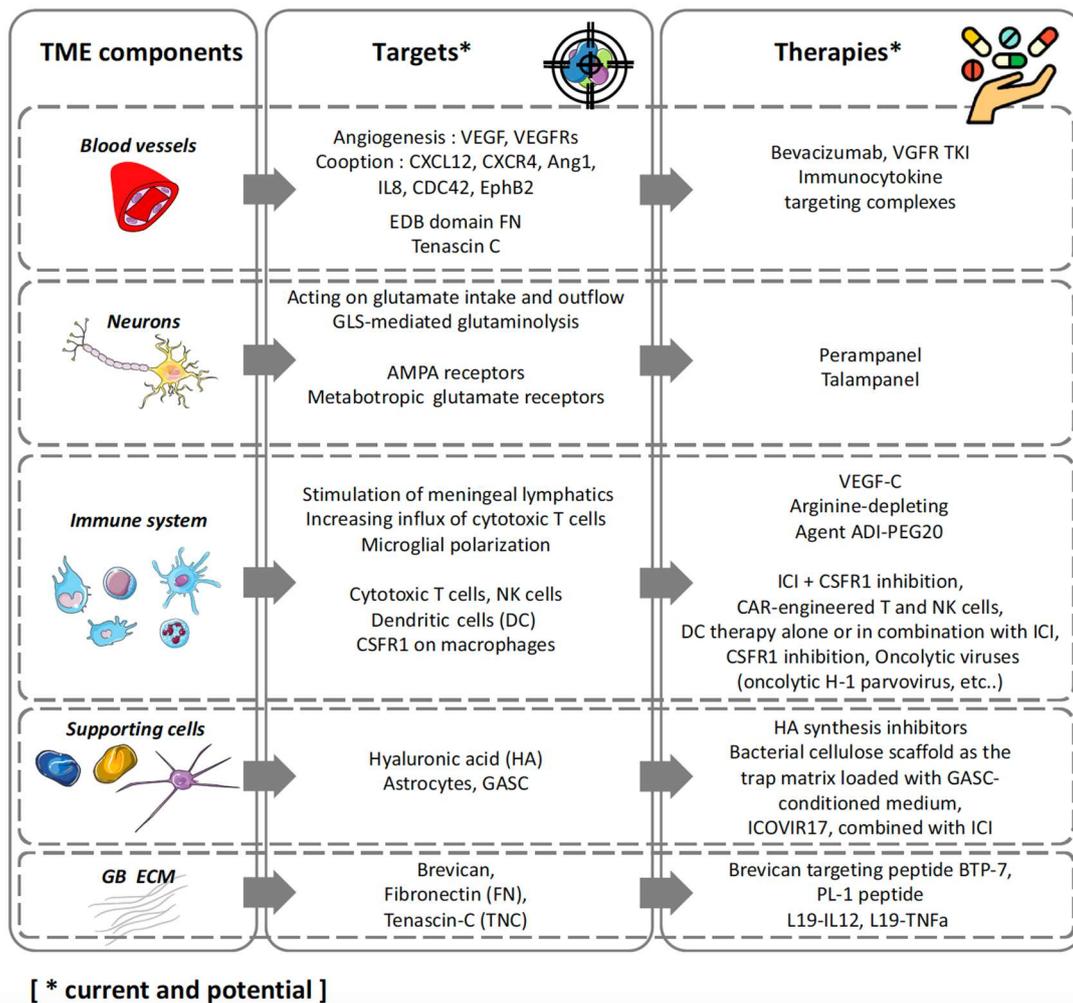


Figure 37: Acteurs cellulaires, moléculaires et les potentielles cibles thérapeutiques dans le glioblastome (extrait de <sup>208</sup>)

## Matériel et méthodes

### Prise en charge des échantillons

#### Patients

Une cohorte rétrospective de 50 tissus de glioblastome FFPE a été obtenue auprès du service de pathologie de l'hôpital de Lille, en France. Une cohorte prospective de 31 tissus de glioblastome frais congelés a également été incluse dans cette étude. 31 patients atteints de glioblastome nouvellement diagnostiqué ont été recrutés prospectivement entre septembre 2014 et novembre 2018 au CHU de Lille, France (NCT02473484). Tous les patients ont donné leur consentement éclairé écrit avant l'inscription.

#### Coloration histologique et annotations

Cette étude a suivi le même procédé que l'étude précédente (Prise en charge des échantillons 99).

#### Préparation des échantillons

Tous les blocs FFPE ont été sectionnés en coupes de 8  $\mu\text{m}$  à l'aide d'un microtome (Leica Biosystems, Wetzlar, Allemagne) à température ambiante. Une étape de déparaffinage a été réalisée grâce à deux passages de 5 min dans une solution de xylène. Ensuite, chaque coupe de tissu déparaffinée a été sprayée avec une solution de glycérol/alcool isopropylique (IPA) (2:8, v/v) en deux passages successifs à l'aide d'un sprayer manuel (Agilent). Le pousse-seringue (série 74900 Cole Parmer Instrument Company) a été réglé sur un débit de 500  $\mu\text{L}/\text{min}$ . Pour les tissus frais congelés, un cryostat Leica CM1510S (Leica Microsystems, Nanterre, France) a été utilisé pour couper des sections de 20  $\mu\text{m}$ . Leur analyse par le SpiderMass ne nécessite aucune préparation de l'échantillon.

#### Analyse SpiderMass

L'analyse est similaire à l'étude précédente (99). La position d'échantillonnage a été déterminée sur la base des annotations histopathologiques. L'acquisition était composée d'un burst de 10 tirs laser résultant en un spectre individuel. L'acquisition spectrale a été réalisée en mode ion positif et négatif en mode sensibilité avec un temps de balayage de 1s. La gamme de masse a été réglée sur  $m/z$  50-2000.

#### Lignées cellulaires et analyse des cellules immunitaires

Les cellules humaines de glioblastome NCH82 ont été obtenues auprès du Dr Regnier-Vigouroux. Ces cellules ont été cultivées dans du DMEM. Le milieu a été supplémenté avec 10% de sérum bovin fœtal et 100 U/mL de pénicilline-streptomycine dans un incubateur à air humidifié à 37°C sous une atmosphère de 5% de CO<sub>2</sub>. À 70% de confluence, les cellules ont

été lavées deux fois avec du DPBS, séchées sous PSM pendant 10min à RT puis analysées par SpiderMass, directement dans la plaque cellulaire. La lignée de monocytes humains THP1 et les cellules primaires de macrophages et lymphocytes, utilisées pour l'élaboration du modèle immune score GBM, proviennent de l'étude précédente sur l'ovaire.

#### Analyse des données

##### **Modèle de classification optimal**

La bibliothèque Lazy Predict a été utilisée pour construire plusieurs modèles à partir de la bibliothèque scikit-learn en entraînant et en testant une série de 24 classificateurs. L'état aléatoire a toujours été maintenu à 1. Ensuite, le modèle optimal a été reconstruit individuellement à l'aide de la bibliothèque scikit-learn, ce qui a permis un réglage fin de ses paramètres pour l'optimisation et l'évaluation de sa précision. Pour mieux évaluer les performances du modèle, une validation croisée 20 fois a été effectuée à l'aide des fonctions KFold et cross\_val\_score, et le rapport de classification a été généré à l'aide de la fonction classification\_report. En outre, la fonction ConfusionMatrixDisplay de la bibliothèque matplotlib a été utilisée pour afficher la matrice de confusion. Le modèle optimal a ensuite été enregistré et chargé pour la prédiction en aveugle à l'aide de la bibliothèque joblib, avec les fonctions joblib.dump et joblib.load.

##### **Explications de la prédiction**

L'algorithme LIME (Local Interpretable Model-agnostic Explanations) a été utilisé pour expliquer le modèle de classification. Cet algorithme calcule les contributions des caractéristiques qui peuvent être positives ou négatives. La bibliothèque ELI5 a été utilisée pour générer un tableau LIME contenant le poids des contributions des caractéristiques à l'aide de la fonction explain\_prediction.

##### **Caractéristiques significatives**

Un algorithme de sélection des pics, grâce à la fonction find\_peaks\_cwt de la bibliothèque scipy, a été appliqué pour éliminer le bruit des instruments. Une carte thermique de regroupement a ensuite été générée sur la base de la liste des pics. La Heat map utilise un regroupement hiérarchique entre les variables m/z avec une distance euclidienne, affichant les ions qui étaient plus ou moins abondants dans chaque groupe. À cette fin, la fonction clustermap de la bibliothèque seaborn a été utilisée. Ensuite, un test statistique non paramétrique Kruskal-Wallis avec correction de Bonferroni, utilisant la fonction stats.kruskal de la bibliothèque Scipy, a été utilisé pour évaluer l'importance de chaque caractéristique. Seules les caractéristiques significatives ayant une p value inférieure ou égale à 0,05 ont été retenues. Enfin, une étape de filtrage a été ajoutée pour ne conserver que les pics mono-

isotopiques. Les boxplots correspondants ont ensuite été générés à partir de la bibliothèque seaborn.

### Immunoscores

Le modèle d'immunoscoring a été entraîné à l'aide de la bibliothèque Python LGBM (Light Gradient Boosting Machine), un puissant cadre de gradient boosting développé par Microsoft. Ce modèle s'appuie sur les spectres cellulaires dans la gamme  $m/z$  de 600 à 1100 en mode ion négatif. Les spectres ont été classés en différents types de cellules : macrophages (M1 et M2) avec 107 spectres, lignée cellulaire cancéreuse (NCH82) avec 107 spectres et lymphocytes (NK, CD8+, CD4+) avec 107 spectres. En raison du manque de spectres pour chaque type de lymphocyte, et pour éviter de créer un modèle incohérent avec des classes déséquilibrées, les lymphocytes ont tous été regroupés. Pour prédire les types de cellules sur les images SpiderMass, la fonction `predict_proba` du modèle LGBM a été utilisée. Cette fonction fournit des estimations de probabilité pour chaque type de cellule, ce qui permet une compréhension plus nuancée de la probabilité de présence de chaque type de cellule dans l'environnement local. En outre, des scores de ratio ont été calculés pour estimer la présence relative de chaque type de cellule sur l'ensemble de l'image. Ces scores ont été calculés en additionnant les scores fournis pour chaque type de cellule et en les divisant par la somme des scores totaux de toutes classes. Les ratios ont fourni des indications sur la distribution des types de cellules formés dans l'ensemble de l'image, offrant ainsi une évaluation complète du paysage cellulaire dans l'échantillon analysé.

## Résultats

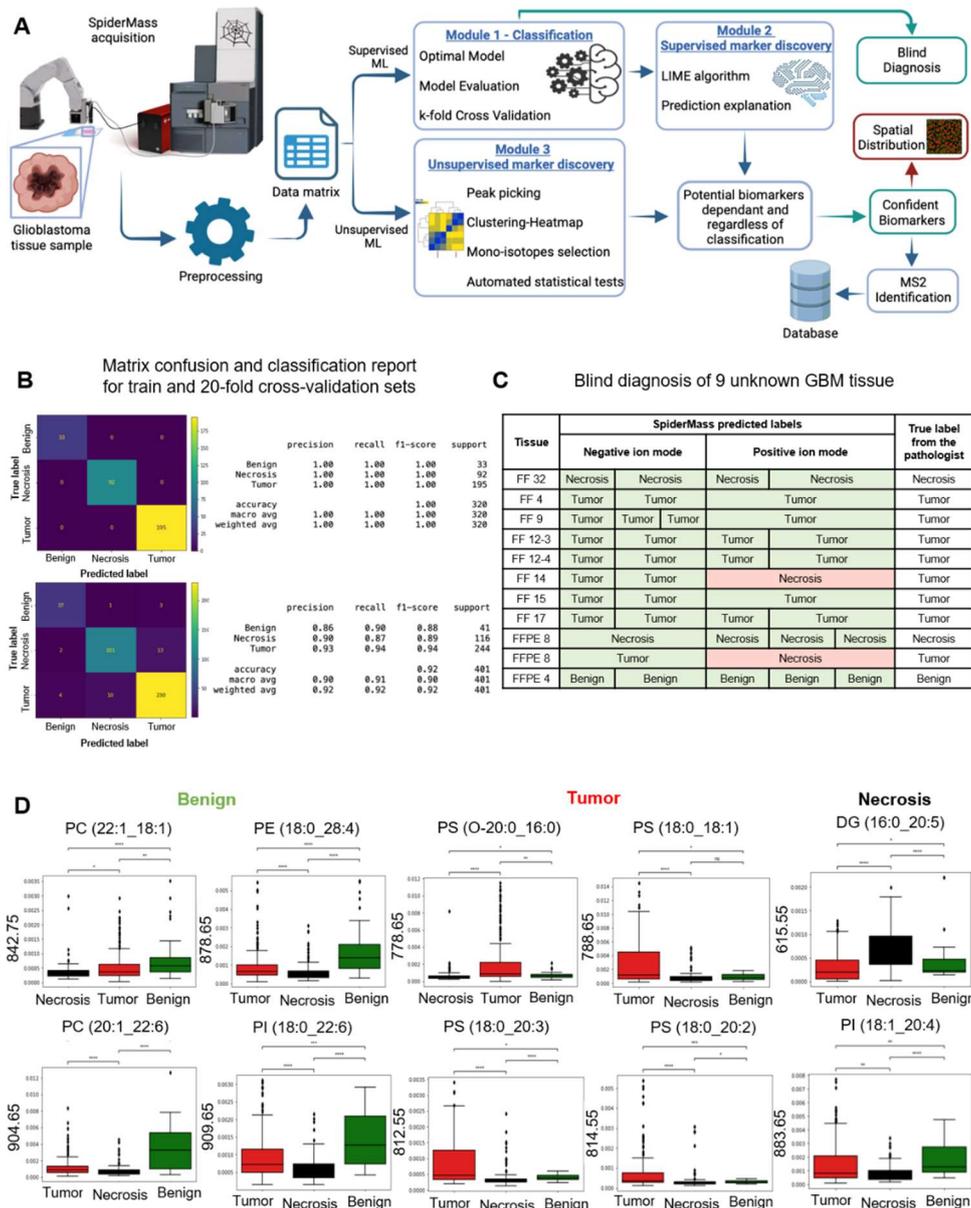
Afin de mieux comprendre l'intérêt de l'utilisation du modèle immune score pour le glioblastome, je vais brièvement évoquer les recherches menées sur ce dernier.

### *Diagnostic du glioblastome*

#### Modèle de classification optimal et pipeline de traitement de l'IA.

Le but de cette étude est, tout d'abord, de créer de nouvelles méthodes d'analyse permettant une classification sensible et robuste des glioblastomes. Pour réaliser cet objectif une cohorte de tissus FFPE (50 tumeurs, 42 nécroses et 18 tissus bénins) a été analysée dans les deux modes d'ionisation à l'aide du SpiderMass. Ainsi, 24 différents classificateurs ont été testés sur l'ensemble des spectres acquis en suivant la même méthode que pour le cancer de l'ovaire (**Figure 38A**). En mode ions négatifs et positifs, la classification optimale a été obtenue en utilisant le Ridge Classifier avec une précision de 93 % et 92 % respectivement. Toutefois, pour garantir la robustesse et la généralisation des données, une cross-validation 20k-fold a été effectuée. Cela a montré que la précision réelle du modèle en mode ion négatif était de 88 % et 87 % en mode positif. Le faible nombre d'échantillons dans la cohorte rétrospective

pourrait expliquer cette faible précision finale. Pour résoudre ce problème, une extension des données a été réalisée en utilisant une cohorte prospective de 31 tissus frais congelés composé de 30 régions tumorales et de 7 régions de nécrose. Cette fois, l'ensemble d'entraînement a atteint 100 % dans les deux modes d'ionisation (**figure 38B**). Dans le mode ionique négatif, la précision a augmenté et s'est stabilisée à 92 % après une cross-validation 20k-Fold (**figure 38B**). Le mode positif, quant à lui, s'est stabilisé à 88% de cross-validation. Ainsi, de bons modèles de classification ont été générés permettant la détection des marges à la fois en mode positif et négatif. Ces deux modèles ont été testés sur une cohorte de 9 patients dans chacun des modes d'ionisation (**Figure 38C**). Il en a résulté une reconnaissance de l'ensemble de la cohorte en mode négatif. Le mode positif a cependant commis 2 erreurs de classification. En effet, 2 échantillons tumoraux ont été prédits comme nécrosés.



**Figure 38:** Pipeline de diagnostic par IA pour l'obtention d'un modèle de classification optimal et la découverte de marqueurs lipidiques associés aux différentes classes moléculaires.

A) Pipeline global pour le diagnostic moléculaire par spectrométrie de masse ambiante combiné à la prédiction par intelligence artificielle. Le chemin bleu est obligatoire pour obtenir les résultats finaux (flèches vertes) qui sont le diagnostic à l'aveugle des tissus et l'identification des biomarqueurs confiants alors que le chemin marron est optionnel. En effet, des étapes facultatives (flèche rouge) peuvent être ajoutées au pipeline comme la distribution spatiale des biomarqueurs confiants grâce au Spidermass-MSI. (B) Rapport de classification et matrice de confusion du Ridge Classifier avec l'ensemble d'apprentissage et après cross-validation 20-Fold, obtenus pour le modèle de classification après extension des données en mode ion négatif. (C) Tableau montrant la classe prédite par le diagnostic en aveugle par SpiderMass comparée à l'annotation du pathologiste pour deux tissus GBM inconnus FFPE et huit tissus GBM inconnus frais congelés (FF), résultant en 95 % de bons diagnostics dans les deux modes d'ionisation. (D) Boxplot correspondant à 10 exemples de biomarqueurs confiants spécifiques de la nécrose, des tissus bénins et tumoraux en mode ionique négatif et positif avec leur tentative d'identification.

## Marqueurs lipidiques associés aux différentes classes moléculaires.

La suite de cette étude s'est concentrée sur l'identification de marqueurs lipidiques spécifiques du glioblastome. Pour ce faire, l'algorithme LIME a été utilisé. Celui-ci inclus dans le pipeline de traitement des explications interprétables sur l'identification potentielle de biomarqueurs en calculant le poids de chaque  $m/z$  pour classer chaque type de tissu. Par exemple, en mode ionique négatif, le rapport  $m/z$  qui contribue le plus positivement à la classification des tissus bénins est le  $m/z$  766,55. Le  $m/z$  751,55 contribue positivement à la classification des tissus nécrotiques mais négativement à celle des tissus tumoraux, tandis que l'inverse est visible pour le  $m/z$  794,55. En mode ions positifs, les pics qui contribuent le plus négativement à la classification des tissus tumoraux, bénins et nécrotiques sont respectivement  $m/z$  603,55,  $m/z$  810,75 et  $m/z$  663,55. Les  $m/z$  qui contribuent le plus positivement à leur classification sont les  $m/z$  810,75,  $m/z$  605,65 et  $m/z$  603,55 (Figure 39).

Listes des  $m/z$  retrouvés discriminant en (A) mode négatif, (B) mode positif.

A Liste des $m/z$ et leur contribution par l'algorithme LIME en mode négatif			B Liste des $m/z$ et leur contribution par l'algorithme LIME en mode positif		
Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature	Contribution <sup>†</sup>	Feature
+5412.239	766.55000	+5215.585	751.55000	+14391.356	605.65000
+4728.161	797.55000	+4204.671	778.55000	+10019.944	828.75000
+4311.434	834.55000	+2578.476	654.55000	+8417.347	603.45000
+3243.889	715.55000	+2247.142	737.55000	+7459.211	842.75000
+3102.389	790.55000	+2183.105	835.55000	+6811.577	629.35000
+3070.118	763.55000	+2118.639	752.55000	+6631.405	790.65000
+3016.019	718.55000	... 2525 more positive ...		+6430.528	814.75000
+2707.719	818.55000	... 2436 more negative ...		+5634.563	724.65000
+2475.288	767.55000	-2010.554	715.55000	+5348.013	1022.85000
+2294.061	716.55000	-2050.669	700.55000	+5238.042	725.55000
+2019.951	856.55000	-2055.321	726.55000	+5082.151	762.65000
+1978.888	747.55000	-2066.985	739.55000	+5004.316	606.65000
+1896.468	717.55000	-2188.433	745.55000	+4449.352	615.45000
+1881.941	753.55000	-2380.984	763.55000	+4410.884	619.35000
+1738.089	752.55000	-2394.559	600.45000	+4198.877	791.65000
+1697.404	820.55000	-2562.633	852.55000	+4010.903	800.75000
+1683.352	854.55000	-2855.512	790.55000	... 2409 more positive ...	
+1668.450	754.55000	-2903.019	886.55000	... 2532 more negative ...	
+1614.210	719.55000	-2961.432	744.55000	-3986.286	812.65000
+1441.449	819.55000	-3059.090	812.55000	-3792.661	602.55000
+1371.089	700.55000	-3187.444	742.55000	-3177.404	835.75000
+1317.073	850.65000	-3911.557	887.55000	-2414.591	609.35000
+1284.273	779.55000	-4166.995	775.55000	-4489.018	1006.85000
+1276.648	798.55000	-4515.737	792.55000	-4636.422	972.85000
+1248.840	938.55000	-4514.203	842.55000	-4627.973	834.75000
+1229.167	848.55000	-5665.635	750.55000	-4962.488	601.35000
... 2408 more positive ...		-5789.564	789.55000	-5025.318	650.95000
... 2533 more negative ...		-6237.321	824.55000	-5342.335	794.75000
-1215.061	806.55000	-7147.062	722.55000	-5397.399	833.75000
-1311.534	809.55000	-7905.779	844.55000	-5516.260	630.65000
-1491.028	744.55000	-8069.601	716.55000	-5622.760	703.65000
-1581.850	792.55000	-8153.465	816.55000	-6090.346	990.85000
-1604.603	822.55000	-10299.142	785.55000	-6373.792	797.75000
-1669.870	746.55000	-10630.938	774.55000	-7364.651	890.75000
-1858.529	775.55000	-11929.719	820.55000	-8652.207	832.75000
-2054.256	722.55000	-12645.402	767.55000	-8641.303	611.25000
-2139.656	824.55000	-12996.554	788.55000	-10572.408	612.25000
-3167.022	775.55000	-14557.721	788.55000	-10690.043	610.25000
-3721.509	796.55000	-15580.072	796.55000	-11700.957	667.25000
-4222.629	774.55000	-21055.690	746.55000	-14029.800	813.75000
-7418.603	768.55000	-22862.133	766.55000	-16062.897	811.75000
-19406.164	794.55000	-42305.082	794.55000	-17649.049	812.75000
				-28783.504	810.75000

Figure 39: Liste des  $m/z$  discriminants par l'algorithme LIME dans les deux modes d'ionisation.

Au total, 41 biomarqueurs fiables ont été identifiés, 27 en mode ion négatif et 14 en mode ion positif, pour les tissus tumoraux, bénins et nécrotiques. Ceux-ci ont été identifiés par MS/MS et leur variation d'abondance relative sont présentée dans les boxplots (figure 38D et 40). En mode ion négatif, sur les 27 biomarqueurs confirmés, neuf ions sont surexprimés dans le tissu tumoral. La majorité d'entre eux étant des phosphatidylsérines (PS) et des acides phosphatidiques (PA). Un total de sept ions sont spécifiques des tissus bénins. Parmi ceux-ci six présentent une forte abondance dans les tissus sains et un DG 40:9 ( $m/z$  661,55) est, quant à lui, peu abondant. Les tissus cérébraux normaux sont composés de différentes espèces lipidiques, mais l'absence de diglycérides (DG) semble être un marqueur potentiel d'absence

de tumeur. Il est intéressant de noter que tous les ions très abondants trouvés dans les tissus bénins sont des phospholipides avec des acides gras à longues chaînes carbonées contenant plus de 3 insaturations. Dans les cellules de gliome, la forte concentration d'acides gras polyinsaturés semble inhiber la division cellulaire et la formation de tumeurs<sup>209</sup>. Dans les régions nécrotiques, trois ions sont surexprimés à de faibles masses et huit ions sont sous-exprimés à des masses élevées. Elles sont également caractérisées par l'absence de molécules plutôt que par leur présence. Plus précisément, les phosphatidylinositols (PI) sont sous-exprimés dans ces tissus. En mode ions positifs, parmi les 14 biomarqueurs confirmés, deux sont spécifiques des tissus bénins,  $m/z$  842,65 et  $m/z$  850,65. Pour les tissus nécrotiques, trois biomarqueurs fiables sont faiblement représentés, à savoir les PI et les phosphatidylcholines (PC). Pour les régions tumorales, huit biomarqueurs sont surexprimés, dont la majorité sont des phosphatidyléthanolamines (PE). Il est intéressant de noter que les glycérophospholipides (GPL) fortement exprimés dans les tissus tumoraux sont principalement des lipides saturés tels que PC (34:2) ( $m/z$  758,65) et PS (38:2) ( $m/z$  814,55).

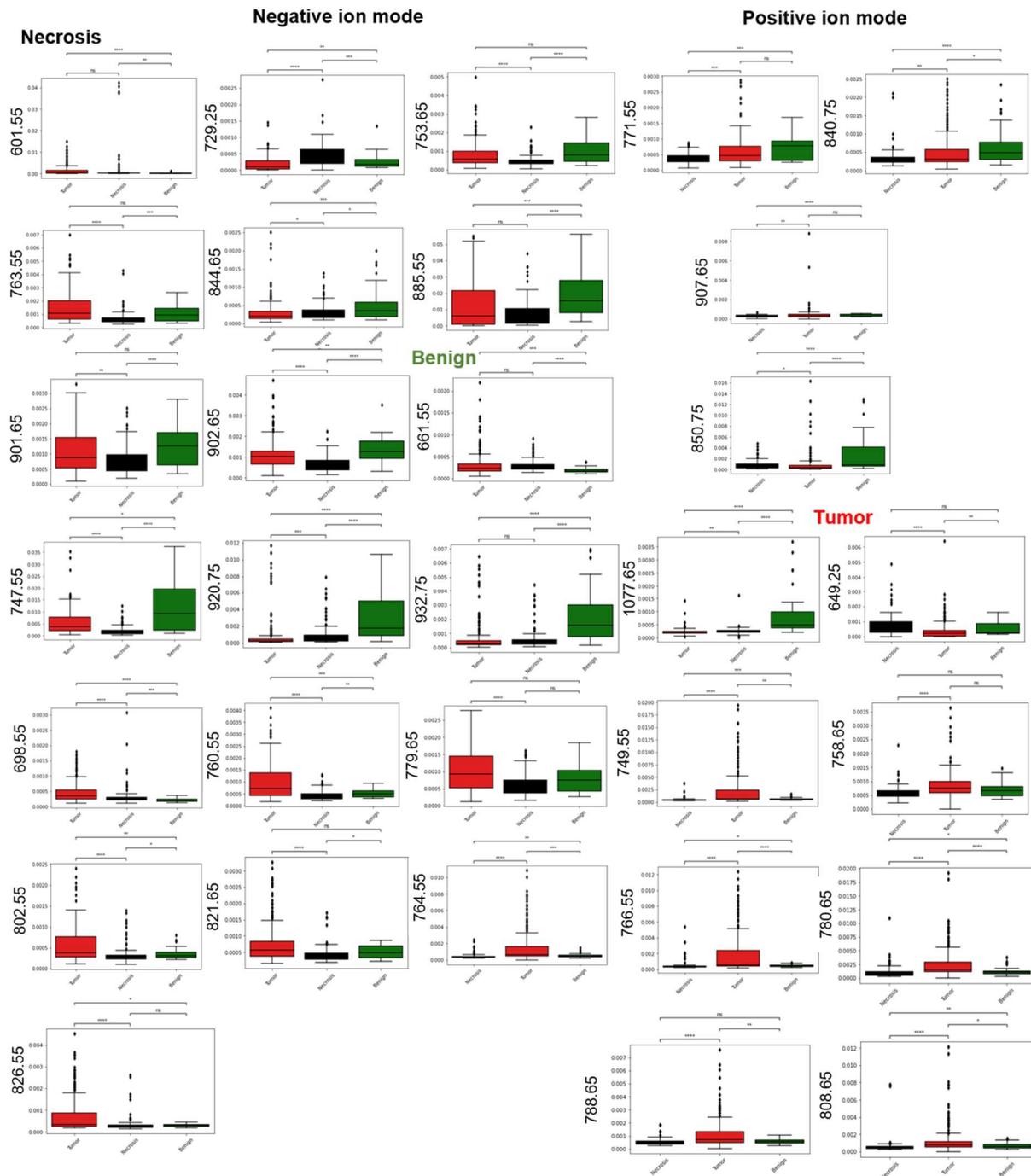


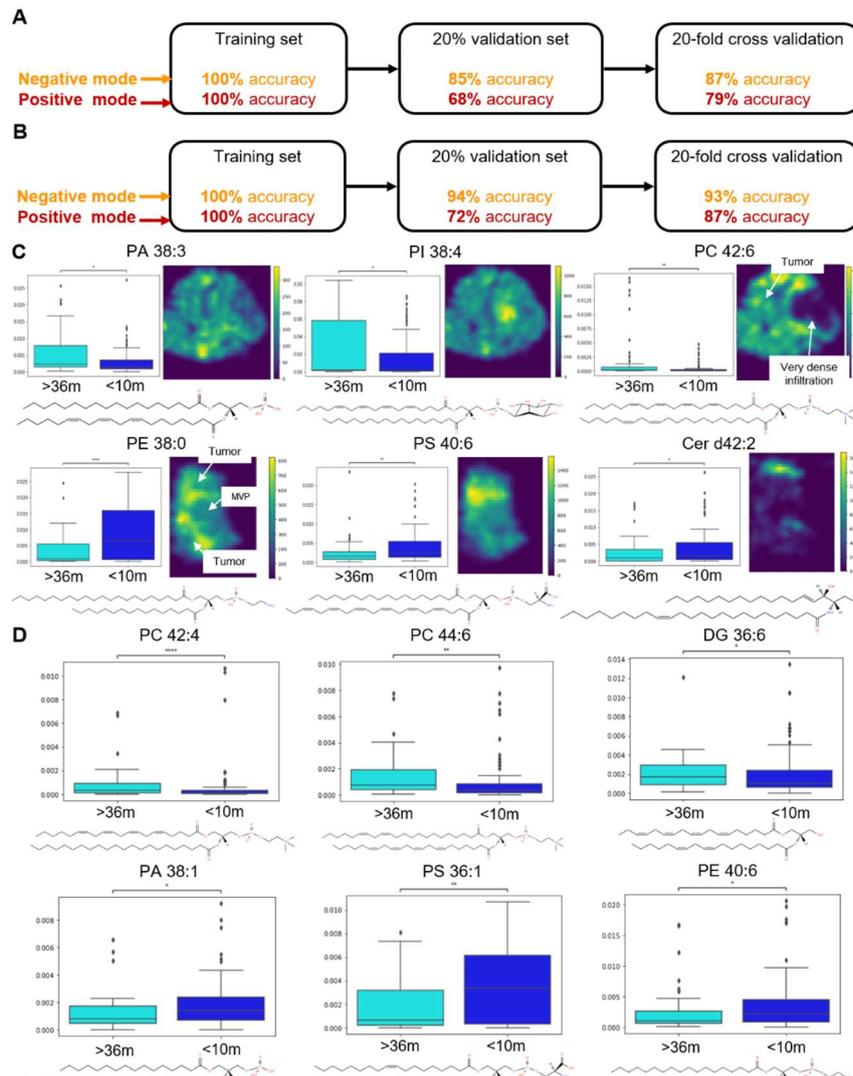
Figure 40: Boxplot représentant l'intensité des lipides retrouvés discriminants en fonction du type de tissu.

Le test statistique Kruskal-Wallis a été réalisé. (\*\*\*\*  $p < 0,0001$ , \*\*\*  $p < 0,001$ , \*\*  $p < 0,01$ , \*  $p < 0,05$ ) NS pour  $p > 0,05$ ).

*Prédiction du devenir des patients à partir des données lipidomiques de SpiderMass.*

Classification des patients en fonction de leur survie par analyse SpiderMass.

En plus du diagnostic validé et de la stratification basée sur SpiderMass, l'équipe s'est également intéressés à l'étude de la prédiction de la survie du patient. Les patients ont été séparés sur la base d'une médiane de survie globale (SG) de 15 mois. Ainsi, deux groupes ont été établis à l'aide du pipeline de traitement développé précédemment, un groupe avec une SG <15 mois et l'autre avec une SG >15 mois, 15 mois étant la médiane de survie des patients GBM. Contrairement aux analyses précédentes qui cherchaient à différencier les types histologiques, les spectres ont été séparés en fonction de la survie globales entre les patients. En mode ionique positif et négatif, un modèle de classification a été bâti avec un taux de classification correcte de 79 % et 87 % après une cross-validation de 20-fold (**Figure 41A**). Pour améliorer la spécificité de la classification basée sur les résultats, nous nous sommes ensuite intéressés aux patients présentant les survies les plus extrêmes, c'est-à-dire <10 mois et >36 mois. En restreignant notre modèle aux cas les plus extrêmes, nous avons obtenu, en mode ion négatif, un taux de classification de 100 % pour l'ensemble d'entraînement, de 94 % pour l'ensemble d'entraînement de la validation de 20 % et de 93 % après la validation croisée de 20 fois (**figure 41B**). En mode ion positif, le modèle de classification atteint 87% de taux de classification correcte après la validation croisée (**Figure 41B**).



**Figure 41:** Obtention d'un modèle de classification pour distinguer les patients ayant une durée de survie faible ou longue et découverte de biomarqueurs pour chaque survie.

(A) Précisions obtenues pour le modèle de classification réalisé avec RidgeClassifier, pour distinguer les patients ayant une durée de survie < ou >15mois, avec l'ensemble d'entraînement, après validation à 20% et après validation croisée à 20 Fold en mode ions négatifs (orange) et positifs (rouge). (B) Précision obtenue pour le modèle de classification réalisé avec RidgeClassifier, après avoir restreint notre modèle aux cas les plus extrêmes (<10 mois et >36 mois), avec l'ensemble d'apprentissage, après une validation de 20% et après une validation croisée de 20 fois en mode négatif (orange) et en mode ion positif (rouge). (C) Six biomarqueurs potentiels, spécifiques du pronostic <10 mois et >36 mois en mode ion négatif, leur boxplot correspondant, leur structure chimique et leur localisation dans un tissu. (D) Six biomarqueurs potentiels, spécifiques du pronostic <10 mois et >36 mois, leur boxplot correspondant et leur structure chimique en mode ion positif.

#### Marqueurs discriminants associés au pronostic du patient

Lors de la recherche de biomarqueurs potentiels de la survie globale et de la corrélation biologique qui leur sont associée, certains ions se sont révélés spécifiques de chacun des groupes de survie. En effet, les deux modes d'ionisation ont permis de trouver 48 biomarqueurs pronostiques statistiquement fiables. En résumé, les PC et les PI sont plus présents chez les patients dont la durée de vie est plus longue, comme PC 42 :4 et PC 44 :6 en mode ionique positif (**Figure 39D**) et PI 38 :4 et PC 42 :6 en mode ionique négatif (**Figure 39C**). Aussi,  $m/z$  850.75 identifié comme PE (22 :1\_22 :4) et  $m/z$  920.75 comme PS 46 :5, qui sont plus abondants dans les tissus bénins, ont également été associés à une survie plus longue. Il semblerait que le tissu tumoral d'un patient ayant une SG>36 mois partage des marqueurs moléculaires avec le tissu bénin. Cela pourrait expliquer pourquoi certains patients vivent plus longtemps. Au contraire, les PS, PA et PE, tels que PA 38 :1, PE 38 :0, PE 40 :6 et PS 36 :1, sont plus fortement exprimés dans les tissus tumoraux de mauvais pronostic (**figure 39C-D**). En particulier, le PS 40 :6 a été associé à une survie plus faible dans les deux modes d'ionisation ( $m/z$  834,55 et  $m/z$  836,55), avec une signification élevée et une contribution importante (algorithme LIME). En outre, les sphingolipides, en particulier les céramides (Cer), étaient caractéristiques d'un pronostic aggravé dans notre étude, puisqu'aucun n'a été trouvé pour les survivants les plus âgés. Il a également été montré que l'abondance de 5 lipides Cer, y compris Cer d42 :2 ( $m/z$  682.65) et CerP 34 :0 ( $m/z$  654.55), était significativement plus élevée chez les patients ayant une mauvaise survie que chez ceux ayant une survie plus longue, indépendamment de la longueur de la chaîne acyle ou du degré d'insaturation. En outre, le  $m/z$  844.65 identifié comme PS 40 :1, spécifique des tissus tumoraux, est associé à une survie plus courte. Les phospholipides identifiés comme biomarqueurs d'une survie plus courte, les PS, PA et PE, sont chargés négativement, ce qui suggère une voie lipidique et protéique spécifique, associée à un pronostic plus défavorable. D'autre part, il existe une interaction céramide-protéine connue entre les céramides et I2PP2A, conduisant à la dégradation de c-Myc<sup>210</sup>. Cela met en évidence un équilibre entre la présence de la protéine PPP1R12A, qui active c-Myc et conduit à un meilleur pronostic, et les céramides, qui inhibent c-Myc et sont associées à une survie plus courte.

#### Utilisation du modèle immune score sur le glioblastome

C'est suite aux découvertes faites sur des marqueurs pronostic que l'intérêt de l'immune score dans le glioblastome a émergé. En effet, certains lipides retrouvés discriminants du cancer ont été retrouvés discriminants des macrophages dans l'études précédentes. De plus, le glioblastome est l'un des tumeurs présentant le plus grand nombre de macrophages avec jusqu'à 50% du nombre totale de cellule. Le glioblastome peut servir de modèle d'étude permettant sa cross-validation tandis que l'immune score peut servir à

mieux définir la survie des patients en se basant sur l'infiltration des cellules immunitaires pour le pronostic.

L'infiltration de diverses populations de cellules immunitaires s'est avérée associée à l'évolution du patient<sup>211,212</sup> dans le cas du cancer du côlon et est désormais reconnue dans un nombre croissant de cancers<sup>213</sup>. De nos jours, l'immune-score est obtenu à partir de tissus excisés ou de biopsies par immunohistochimie (IHC) en utilisant des anticorps spécifiques des différentes populations de cellules immunitaires. Cependant, il est obtenu après l'intervention chirurgicale. Ici, nous avons exploré la possibilité de créer un immune score basé sur les données de SpiderMass qui pourrait, à l'avenir, être réalisé *in vivo* et en temps-réel. À cette fin, nous avons analysé avec SpiderMass différentes populations de cellules immunitaires, à savoir les macrophages (M1-like et M2-like), les lymphocytes CD8+, CD4+ et NK (tous regroupés sous le terme de lymphocytes) par rapport aux cellules cancéreuses (cellules de glioblastome NCH82) afin de trouver des lipides qui sont spécifiques aux cellules immunitaires. La bibliothèque Python appelée LGBM (Light Gradient Boosting Machine), précédemment utilisée pour l'ovaire, a été adaptée pour former un modèle de notation immunitaire du glioblastome. Ainsi, à l'aide de 107 spectres MS pour chaque type de cellule, le modèle a obtenu un taux de classification correcte de 100 %, à la fois en entraînement et en cross-validation (**figure 40A**). En utilisant le nouveau pipeline d'IA, 10 marqueurs lipidiques de macrophages de type M1, 8 marqueurs lipidiques de macrophages de type M2, 6 marqueurs lipidiques de lymphocytes et 3 marqueurs lipidiques de cellules cancéreuses NCH82 ont été obtenus. Par exemple, les ions  $m/z$  818.65 (GlcCer d18 :1\_22 :0) et 819.55 (PG 18 :1\_22 :6) ont de nouveau été retrouvés comme étant spécifiques des macrophages de type M1 et M2 respectivement. Pour les lymphocytes, les PE ont été signalés comme marqueurs lipidiques tels que PE (20:4\_16:0) ( $m/z$  738,55) et PE 38:3 ( $m/z$  768,55). Il est intéressant de noter que deux ions,  $m/z$  844.65 et  $m/z$  848.65, spécifiques des macrophages de type M1 et M2 respectivement, ont également été trouvés spécifiques du tissu tumoral de patients dont la durée de vie était inférieure à 10 mois.

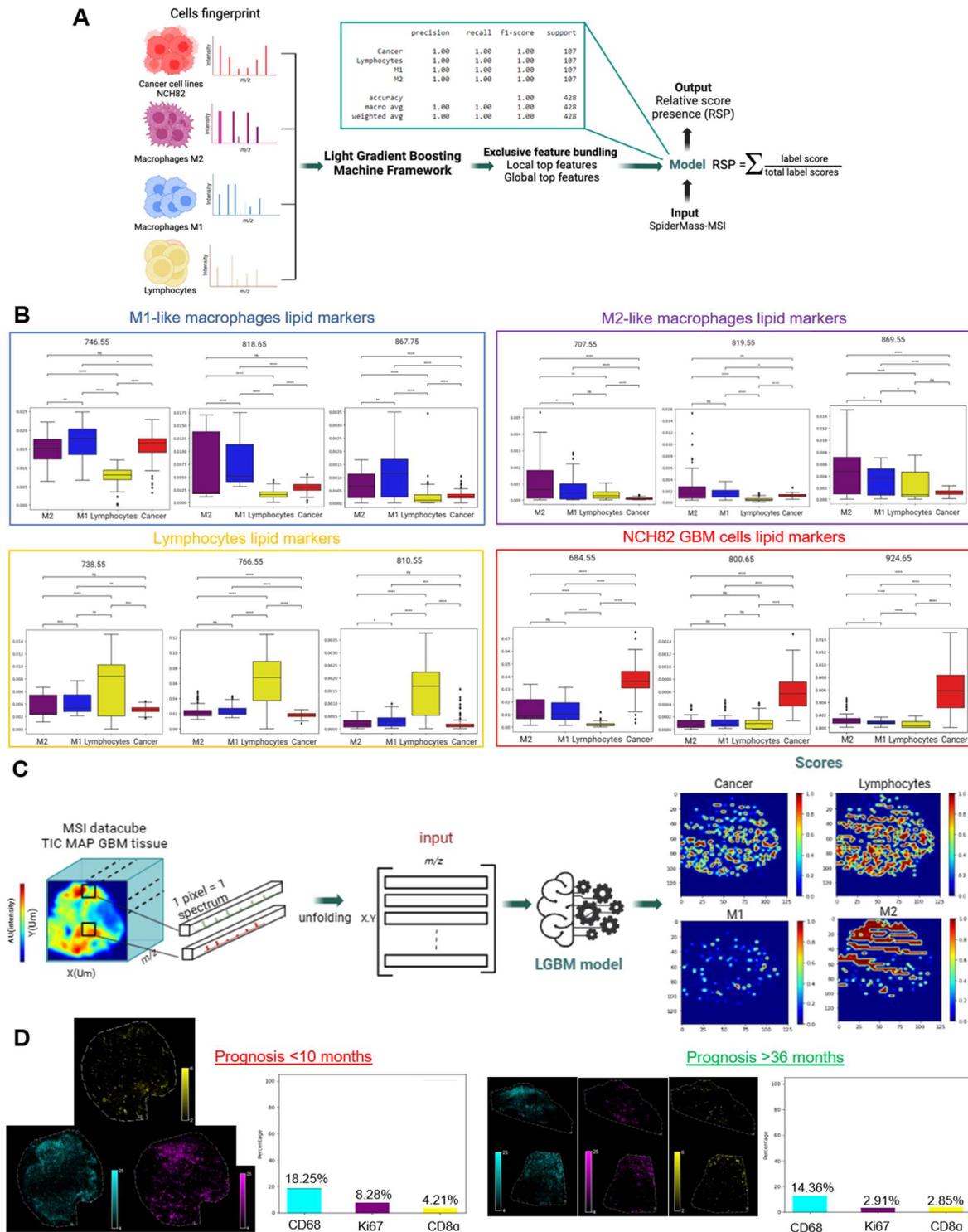


Figure 42: Classification des tissus basé sur l'immune-score.

(A) Modèle LGBM entraîné pour les cellules cancéreuses, les lymphocytes, les macrophages de type M1 et M2 qui ont été analysés précédemment par la technologie SpiderMass. Un taux de classification correct de 100 % a été obtenu pour le modèle LGBM pour l'ensemble de test, les 20 % restants et également après une validation croisée de 20 fois. (B) Trois exemples de biomarqueurs et leur boxplot correspondant pour les cellules cancéreuses du GBM, les lymphocytes, les macrophages de type M1 et M2. (C) Le pipeline global pour voir la distribution des cellules cancéreuses et immunitaires (macrophages M1-like, M2-like et lymphocytes) dans le tissu GBM analysé par SpiderMass-MSI. (D) Analyse par TagMass sur les marqueurs Cd68, Ki67 et CD8 d'une coupe de tissu d'un patient avec un survie globale <10 mois et >36mois.

#### Corrélation entre l'immune score basé sur les images SpiderMass et le pronostic du patient

A partir d'une image SpiderMass, nous pouvons obtenir la distribution prédite des différentes cellules immunitaires dans le tissu comme nous l'avons montré dans la précédente étude. En effet, un second pipeline a été dédié à la prédiction de la présence de chaque type de cellule à partir d'images SpiderMass de tissus de glioblastomes frais congelés de la cohorte prospective. Les résultats obtenus fournissent des scores estimés pour chaque type de cellule, et des scores de rapport ont été calculés pour déterminer le score relatif de présence (RSP) de chaque type de cellule sur l'ensemble de l'image. Ces scores ont été calculés en additionnant les scores fournis pour chaque type de cellule et en divisant la somme par les scores totaux de toutes les classes (**Figure 42A-C**). Au total, six tissus frais congelés ont été analysés, trois provenant de patients ayant une survie <10 mois et trois provenant de patients ayant une survie >36 mois. Les ratios donnent un aperçu de la distribution des types de cellules formés sur l'image. Ainsi, une évaluation complète du paysage cellulaire chez les patients dont la survie est plus courte ou plus longue a pu être réalisée (**Figure 42A-B**). Il est intéressant de noter que le ratio prédit des cellules cancéreuses n'est pas significativement différent en fonction de la survie. Cependant, les lymphocytes ont été prédits pour avoir une plus grande abondance dans les tissus des patients avec une survie de plus de 36 mois (**Figure 43C**). En effet, le RSP moyen des lymphocytes était supérieur à 40 % et inférieur à 36 % pour les survies plus longues et plus courtes respectivement. Le ratio prédit des macrophages a également révélé une association intéressante entre les cellules immunitaires et le résultat du patient. Plus précisément, une proportion plus élevée de macrophages de type M1 (3,5 %) a été trouvée dans les tissus des patients dont l'issue était meilleure que dans ceux dont l'issue était moins bonne (2,8 %) (**figure 43D**). À l'inverse, les macrophages de type M2 avaient un RSP moyen de 35,8 % et 41,3 % chez les patients dont la survie était respectivement plus longue et plus courte (**Figure 43E**). Le ratio M1/M2 semble également intéressant dans le glioblastome. En effet, lorsque le ratio M1/M2 est inférieure à 7%, cela signifie une durée de survie plus courte. A l'inverse, un ratio supérieur à 10% indique une durée de survie prolongée (**Figure 43F**). Par ailleurs, une corrélation semble exister 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 être plus nombreux que les lymphocytes (+ 22%), alors que l'inverse est vrai pour les patients dont le pronostic est plus favorable (**Figure 43G**).

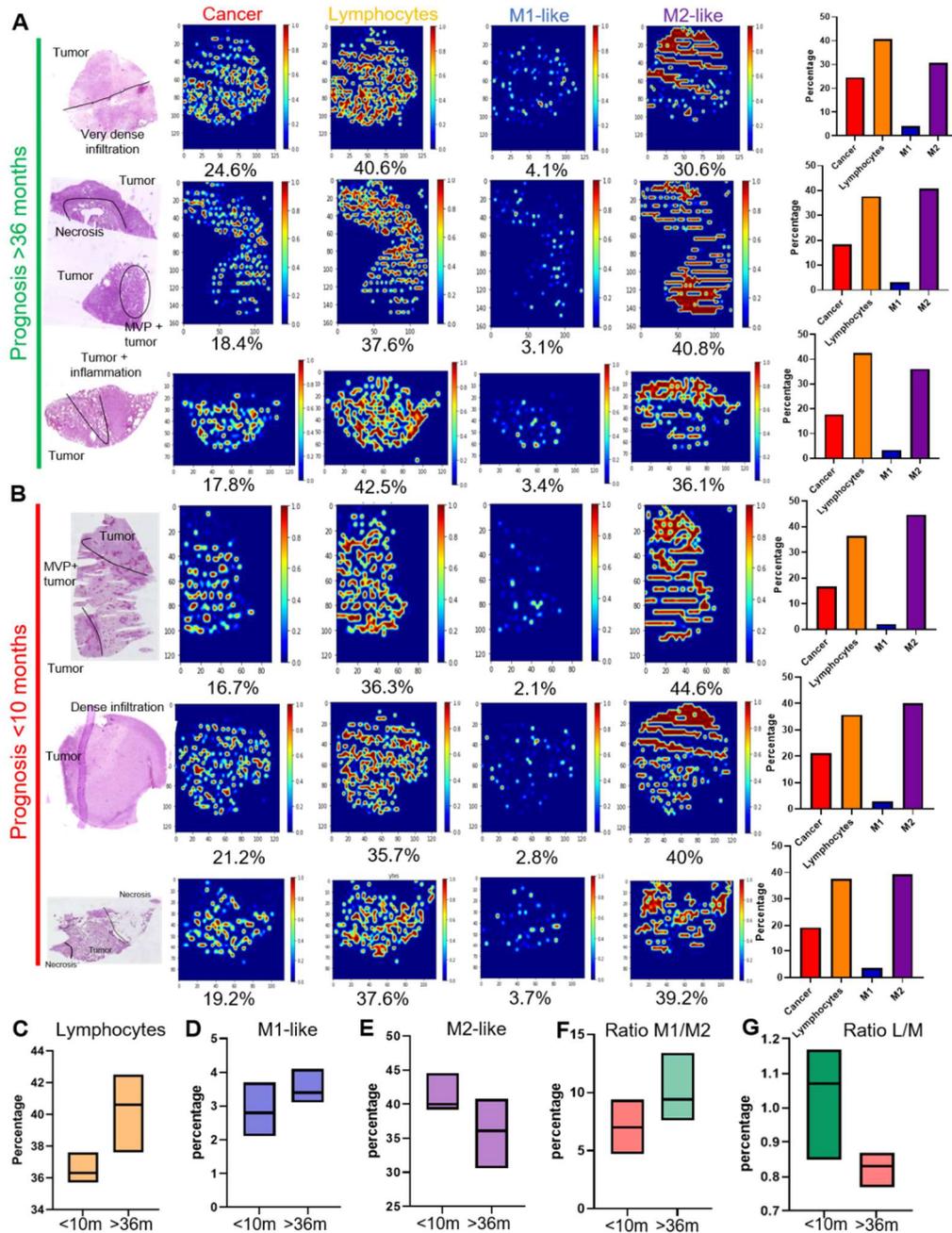


Figure 43: Modèle immune score pour la définition de la survie chez les patients GBM <10mois et >36mois.

(A) Scans H&E et score relatif de présence de cellules cancéreuses, de lymphocytes, de macrophages de type M1 et de macrophages de type M2 dans trois tissus de patients dont le pronostic est >36 mois. Pour chaque type cellulaire, le rapport en pourcentage est indiqué. En outre, un histogramme a été construit pour comparer le pourcentage de chaque type cellulaire dans chaque tissu. (B) Scans H&E et score relatif de présence de cellules cancéreuses, de lymphocytes, de macrophages de type M1 et de macrophages de type M2 dans trois tissus d'un patient dont le pronostic est > 10 mois. Pour chaque cellule, le rapport en pourcentage est indiqué. En outre, un histogramme a été construit pour comparer le pourcentage de chaque type de cellules dans chaque tissu. (C) Le pourcentage moyen de lymphocytes dans les tissus dont le pronostic est <10 mois et >36 mois. (D) Le pourcentage moyen de macrophages de type M2 dans les tissus dont le pronostic est <10 mois et >36 mois. (E) Le pourcentage moyen de macrophages de type M2 dans les tissus dont le pronostic est <10 mois et >36 mois. (F) Le rapport moyen M1/M2 macrophages dans les tissus avec un pronostic <10 mois et >36 mois. (G) Rapport moyen lymphocytes (L)/macrophages (M) dans les tissus dont le pronostic est <10 mois et >36 mois.

Le MALDI-IHC a été construit pour corroborer les résultats antérieurs du pipeline d'immunoscore nouvellement développé. À cette fin, cinq biomarqueurs ont été choisis, dont deux témoins, la vimentine, et un marqueur de matrice cellulaire, le collagène ; un biomarqueur lié au cancer, le Ki67 (prolifération/cellules cancéreuses) ; et des biomarqueurs pour les cellules immunitaires CD8 (cellules T cytotoxiques) et CD68 (macrophages). La distribution de populations cellulaires distinctes (notamment les cellules immunitaires), obtenue grâce à l'intégration de SpiderMass et de LGBM, a ensuite été cross-validée par MALDI-IHC. Il est particulièrement intéressant de noter que l'approche SpiderMass-MSI, basée sur l'immunoscore, permet la différenciation de diverses sous-populations au sein du microenvironnement tumoral sans nécessiter de techniques reposant sur l'utilisation de sondes, telles que l'IHC. En définitive, cette approche innovante permet non seulement de mieux comprendre la présence de cellules immunitaires dans le microenvironnement tumoral, mais elle offre également la possibilité de pronostiquer plus rapidement le temps de survie des patients atteints de GBM.

### Discussion

Cette étude représente une investigation approfondie de la possibilité d'obtenir une classification histologique, une sous-stratification et un pronostic précis du glioblastome en utilisant la spectrométrie de masse par ionisation ambiante via la technologie SpiderMass, combinée à une chaîne de prédiction associée à l'intelligence artificielle. Une précédente étude a permis de mettre en évidence que le glioblastome de type sauvage IDH était stratifié en trois principaux sous-groupes moléculaires<sup>214</sup>. Dans cette étude, nous avons étendu, par une approche par lipidomique via le SpiderMass, nos précédents travaux avec une très bonne corrélation.

De plus, nous avons exploré différentes méthodes de classification pour déterminer le modèle le plus efficace pour la discrimination des profils moléculaires. Jusqu'à présent, divers algorithmes, tels que le LDA<sup>215</sup> ou le Lasso<sup>216</sup>, ont été utilisés pour les données obtenues par l'iKnife, le MasSpec Pen ou le DESI<sup>217</sup>. Un total de 24 classificateurs a été évalué et le modèle optimal a été obtenu en utilisant le RidgeClassifier. Le problème lié à cette pathologie, considérée comme une tumeur rare et de très mauvais pronostic, est la limite du nombre d'échantillons disponibles permettant l'apprentissage du modèle et donc, son efficacité. Pour contourner ce problème, nous avons combiné les spectres issus des analyses de tissus frais

(31 patients, cohorte prospective) et FFPE (50 patients, cohorte rétrospective) avec le SpiderMass. En effet, les spectres obtenus à partir de ces deux types d'échantillons sont assez proches pour être combinés, permettant pour une détection efficace de signatures moléculaires spécifiques des différentes classes. L'augmentation du nombre de données a entraîné une augmentation de la précision du modèle, avec 92 % et 88 % de bonne classification du type histologique des tissus (tumeur, nécrose, tissu bénin) en modes ioniques négatifs et positifs, respectivement. Les prédictions en aveugle à l'aide d'un ensemble de validation indépendantes n'ont généré aucune erreur de classification en mode ionique négatif et seulement deux spectres n'ont pas été correctement assignés en mode ionique positif. Le RidgeClassifier est adapté aux ensembles de données multidimensionnels et démontre de bonnes performances pour les données MS issues de tissus.

L'intérêt était également de fournir une corrélation biologique liée à cette classification. Celle-ci est basée à la fois sur les caractéristiques histologiques et les données moléculaires. Pour cela, nous avons inclus une deuxième chaîne de traitement des données, basée sur l'algorithme LIME. L'algorithme LIME, combiné à des techniques d'apprentissage automatique non supervisées et à des tests statistiques, permet d'identifier des biomarqueurs potentiels indépendants de la classification. La combinaison de SpiderMass et de cette chaîne de traitement d'apprentissage automatique représente une approche prometteuse pour faire progresser la médecine de précision pour le glioblastome et d'autres cancers hétérogènes. De manière intéressante, l'étude a identifié 41 biomarqueurs fiables. Des variations au sein de la proportion des classes de phospholipides ont permis la discrimination des différents types de tissu. Le tissu bénin se caractérisait par la présence d'acides gras polyinsaturés et l'absence de DG. Le tissu nécrotique se distingue par l'absence de PI, tandis que le tissu tumoral présentait une présence significative de PS et de PA parmi les phospholipides détectés. Ceux-ci sont en lien direct avec les constatations faites sur le cancer de l'ovaire. En effet, l'augmentation de PS est associée à une rigidification de la bicouche lipidique notable dans certains cancers<sup>187-189</sup>. De plus, l'augmentation de PS dans la membrane externe est liée aux processus de migration cellulaire et d'immunosuppression<sup>190,191</sup>. Ces biomarqueurs méritent d'être davantage étudiés dans un objectif thérapeutique futur dans le cadre des GBM. Il est néanmoins nécessaire de réaliser une validation ultérieure sur une cohorte plus conséquente.

De manière intéressante, la nouvelle chaîne de traitement d'apprentissage automatique développée ne sert pas seulement d'outil de diagnostic, mais a également été testée pour le pronostic lors de l'utilisation de SpiderMass. En effet, il a été possible de différencier les patients ayant une survie supérieure à la médiane de survie globale selon leur profil moléculaire. Les lipides discriminants ont été identifiés par MS<sup>2</sup>. Il apparaît que la présence de PCs et de PIs soit associée à une survie prolongée. Au contraire, les PSs, PAs, PEs et les

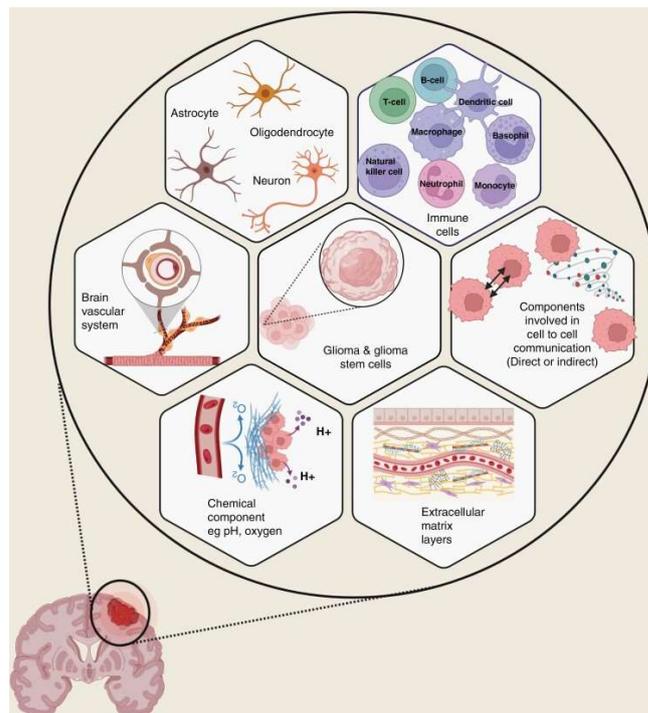
céramides semblent surexprimés dans les tumeurs ayant un mauvais pronostic. Cela est en lien direct avec une étude évoquant la surexpression de céramides dans les tissus n'ayant pas subi de thérapie anti-PD1<sup>209</sup>.

Le glioblastome étant l'une des tumeurs les plus immunosuppressive, il était intéressant d'analyser le TME pour compléter le pronostic. Ainsi, les tissus de GBM ont été interrogés par le modèle immune score précédemment développé dans le contexte des cancers de l'ovaire. Pour ce faire, ce modèle a été entraîné sur la lignée cellulaire NCH82, sur les cellules primaires immunitaires M1-like et M2-like ainsi que sur les différentes sous-populations lymphocytaires (CD4+, CD8+ et NK). L'infiltration immunitaire a pu être déterminée en interrogeant chaque pixel obtenu par SpiderMass dans sa version imagerie sur des coupes de GBM. Ces prédictions ont mis en évidence une différence dans la répartition des phénotypes macrophagiques en fonction de la survie globale des patients. En effet, comme attendu, les patients ayant un mauvais pronostic possédaient une quantité de macrophages M2-like très importante. Au contraire, une quantité importante de M1-like est associée à une survie prolongée des patients. Ces résultats sont en parfaite adéquation avec ceux obtenus lors de nos précédentes études sur le cancer de l'ovaire. De plus, la probabilité de présence de ces cellules immunitaires semble fiable, le score de TAMs se situe entre 37% et 41% ce qui est relativement proche des données décrites dans la littérature, estimant la quantité de TAMs dans les glioblastomes entre 30% et 50% de la masse tumorale<sup>205</sup>. De plus, le ratio entre les phénotypes M1 et M2 a son importance dans le pronostic du glioblastome. En effet, les patients ayant une survie supérieure à 36 mois montrent un ratio M1/M2 de l'ordre de plus de 9% tandis que les patients ayant une survie inférieure à 10 mois montrent un ratio plus faible, inférieur à 7%. La prédiction de la probabilité de présence des lymphocytes est également à prendre en compte pour le pronostic. En effet, l'abondance de lymphocytes est associée à un bon pronostic. Dans ce cas, on a pu observer une légère augmentation de ceux-ci au sein des patients ayant un meilleur pronostic. Pour finir, les scores obtenus par le modèle LGBM ont été validés par imagerie ciblée en utilisant la technologie TagMass (MALDI-IHC) sur les marqueurs macrophagiques, lymphocytaire et cancéreux.

Cette étude a donc pu valider sur un autre type de cancer la preuve de concept de l'immune score réalisée sur le cancer de l'ovaire. De plus, cette nouvelle méthode permet la distinction de sous-groupes de patients de pronostics différents. Ainsi, en travaillant sur des cohortes plus larges de patients, cette technique pourrait définir, *in vivo*, l'infiltration immunitaire au sein d'un tissu. Selon la prédiction, il serait alors possible d'adapter la prise en charge du patient afin de ralentir au mieux la progression de cette pathologie.

## Conclusion

L'apport de l'intelligence artificielle à la technologie SpiderMass a permis la création d'excellents modèles de classification, robustes, souples, et très clairs pour le clinicien. Grâce à ces modèles, des biomarqueurs spécifiques de la pathologie ont été identifiés, permettant la sous-stratification des patients, jusqu'alors considérés comme un seul groupe par le pathologiste. Ces marqueurs protéiques et lipidiques ont également permis une meilleure compréhension des voies de signalisation importante pour le pronostic des patients. Ces marqueurs ont été complétés par l'immune-score, permettant une prise en considération de l'infiltration immunitaire dans les cancers. La capacité de déterminer la nature et la proportion des principales cellules immunitaires au sein du TME est un bon outil pour le diagnostic en temps réel des marges et pour le pronostic de survie des patients. Seulement, il ne prend pas encore en compte de l'ensemble des acteurs du micro-environnement. Ainsi, l'amélioration de ce modèle immune-score passera par l'ajout de nouvelles populations cellulaires telles que les cellules microgliales (**figure 44**). L'ajout de données à ce modèle permettra alors l'amélioration de sa phase d'entraînement et une meilleure compréhension du TME.



**Figure 44:** Représentation schématique du microenvironnement tumoral associé au glioblastome.

*L'environnement microscopique d'un gliome est un système complexe, hétérogène et interactif composé de cellules cancéreuses, de cellules souches, de cellules immunitaires, de cellules résidentes du système nerveux, des cellules du système vasculaire cérébral et de couches de matrice extracellulaire. Les facteurs impliqués dans la communication cellulaire directe et indirecte, ainsi que dans l'environnement chimique de la tumeur, tels que le pH et l'oxygène, jouent également un rôle significatif dans la modulation de l'environnement microscopique des gliomes (extrait de <sup>218</sup>).*

## Figures supplémentaires

<i>m/z</i>	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>
729.25	-	+		[PG (34:10)-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>
<i>m/z</i>	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>

**Supp data 1** : Tableau représentant les 41 ions discriminants pour le mode négatif et positif et leur identification.

En rouge sont représentés les espèces sous-exprimées et en vert surexprimées. Les signes (+,-) correspondent à la contribution des pics pour la classe d'après l'algorithme LIME.

## CHAPITRES VI : Conclusions & Perspectives

Le cancer de l'ovaire est un des cancers les plus mortels dans le monde. Bien qu'il n'ait pas l'incidence la plus élevée, la survie associée à cette pathologie est très faible<sup>219</sup>. Comme précisé durant le second chapitre, cela est principalement dû à un diagnostic trop tardif des patientes. Celles-ci sont malheureusement diagnostiquées à des stades très avancés (III/IV) où les carcinomes sont multiples au sein de l'abdomen et notamment au niveau du péritoine<sup>220</sup>. C'est pourquoi, il est primordial de définir des marqueurs spécifiques qui pourraient être détectés à des stades plus précoces de la maladie. Pour cela, une meilleure compréhension du développement tumoral est nécessaire. Dans cette optique, nous nous sommes focalisés dans notre première étude sur l'origine du cancer de l'ovaire le plus agressif, le HGSOC. Cette étude a mis en évidence des marqueurs multi-omics présents sur des lésions très localisées de l'épithélium tubaire. Ces marqueurs, à la fois protéiques et lipidiques, ont été associés à des voies de signalisation connue chez l'OC. Ces marqueurs ont souligné le caractère malin de ces lésions et la nécessité de les détecter avant leur progression. Une détection précoce préviendrait l'invasion de ces lésions vers l'épithélium ovarien et donc, de l'apparition d'un cancer de l'ovaire. De plus, nous avons démontré la faisabilité l'analyse SpiderMass sur de très petites lésions, visibles uniquement par IHC. Cette technique a permis de discriminer l'ensemble des lésions selon leur profil moléculaire et a mis en avant les bienfaits de la fimbriectomie bilatérale sur les patientes atteintes des mutations BRCA. En effet, cette chirurgie de prévention permet de conserver l'environnement hormonal de la patiente et réduit drastiquement les risques d'apparition d'un HGSOC<sup>48</sup>.

Malgré les progrès scientifiques quant à la découverte de nouveaux traitements contre le cancer, la chirurgie reste, dans la plupart des cas, le traitement le plus efficace pour l'amélioration de la survie des patients<sup>42</sup>. Elle l'est d'autant plus dans les cancers gynécologiques et cérébrales tels que le glioblastome. Seulement, le chirurgien éprouve des difficultés dans la définition des marges d'exérèse. La tumeur est souvent diffuse et les carcinomes dans l'abdomen ne sont, malheureusement, pas visibles à l'œil nu. Ces contraintes empêchent le chirurgien de retirer l'entièreté des cellules tumorales, notamment au niveau du péritoine, source majeure de rechute pour les patientes atteintes d'OC. Les sources AIMS ont permis un bond en avant dans l'aide au diagnostic per-opératoire. En effet, elles apportent une aide à la prise de décision en apportant des données moléculaires<sup>9</sup>. Ces techniques couplées à la protéomique sont des outils puissants dans la compréhension des mécanismes moléculaires associés aux cancers. À l'aide du SpiderMass, nous avons mis en place une méthode de diagnostic précise, basée sur le contenu moléculaire et le phénotype morphologique du tissu. Cette technique a démontré une très grande sensibilité et a permis

de discriminer les sous-types de cancers efficacement. La nature tissulaire peut ainsi être identifiés en aveugle, en temps réel, sur des pièces chirurgicales afin d'accélérer le retour fait au chirurgien par le pathologiste. De plus, grâce à la mini-invasivité du SpiderMass, le diagnostic peut être établi *ex vivo*, mais également *in vivo*, directement au bloc opératoire.

L'apport de l'intelligence artificielle couplée au SpiderMass a permis la création d'excellents modèles de classification, robustes, souples, et clairs. En utilisant le Ridge Classifier, de nouveaux biomarqueurs ont été identifiés à la fois dans le cancer de l'ovaire et dans le glioblastome. Ces marqueurs ont permis de stratifier des sous-groupes de patients jusqu'alors considérés comme un groupe homogène et de les dissocier selon leur survie. Ce nouveau pipeline a rendu le diagnostic plus précis et plus compréhensible pour le clinicien.

Plus récemment, l'étude du microenvironnement tumoral a démontré l'importance du système immunitaire dans la lutte contre le cancer<sup>221</sup>. Les macrophages associés à la tumeur (TAMs) présentent un grand intérêt pour le diagnostic et pronostic. En effet, leur plasticité phénotypique, simplifiée en 2 phénotypes majeurs : les M1-like (pro-inflammatoires) et les M2-like (pro-tumoraux), joue un rôle majeur dans le développement tumoral<sup>144</sup>. Les macrophages M2-like sont étroitement liés à la progression tumorale en participant à l'angiogenèse, l'immunosuppression, la prolifération et à l'invasion tumorale<sup>136,144</sup>. Ainsi, l'abondance des TAMs et leur phénotype sont corrélés à la survie des patients. En effet, l'abondance de M1 semble être associée à une survie prolongée. Au contraire, une grande quantité de M2 inhibe totalement l'action du système immunitaire et est associée à un mauvais pronostic. C'est pourquoi, le rapport entre ces deux phénotypes peut aider dans l'estimation de survie des patientes<sup>139,145,146</sup>. Les lymphocytes ont également une importance majeure dans le pronostic. En effet, les lymphocytes CD8+, du fait de leur activité cytotoxique, permettent l'élimination des cellules tumorales. Ainsi, leur abondance est souvent liée à un bon pronostic<sup>222</sup>. Durant cette thèse, j'ai construit un modèle capable de discriminer l'ensemble des types cellulaires du TME. Ce modèle peut donc identifier les macrophages M1, les M2 ou encore les lymphocytes. En ajoutant un score relatif de présence à ce modèle immune score, l'élaboration d'un score de présence de chacune de ces cellules sur une coupe de tissu a été rendu possible par imagerie SpiderMass. Ainsi, nous avons pu créer une méthode de diagnostic visuelle basée sur la présence et la distribution de ces cellules dans les tissus. En prenant en compte la plasticité phénotypique des macrophages, cette méthode a également permis de discriminer les patients selon leur pronostic, bon ou mauvais.

Ce tout nouveau modèle, développé dans le contexte du cancer de l'ovaire, est très prometteur. Il peut être appliqué à différentes pathologies telles que le glioblastome. En effet, nous avons entraîné ce modèle sur une lignée de glioblastome, NCH82, et sur les cellules

immunitaires, préalablement analysées. Ainsi, notre modèle a été parfaitement capable de discriminer deux sous-groupes de survie de patients atteints de glioblastome. Étant en parfaite adéquation avec les études menées sur les glioblastomes<sup>205</sup>, les probabilités de présence liée au TAMs semblent robustes. Nous avons donc associé ces différences immunitaires à des marqueurs moléculaires spécifiques du pronostic. De manière intéressante, ces groupes ont pu être retrouvés à l'aide de l'immune-score et aussi de notre nouveau modèle de classification basé sur le Ridge Classifier, prouvant la robustesse de notre pipeline.

Nos différentes études ont donc prouvé le potentiel très prometteur du SpiderMass dans l'élaboration du diagnostic précis, robustes, *in vivo* et en temps réel. Cet outil mini-invasif a un très grand intérêt pour le chirurgien. La survie des patientes dépend exclusivement de la capacité du chirurgien à réaliser une résection totale de la tumeur<sup>223</sup>. Seulement, à ces stades, la tumeur n'est pas restreinte à une seule zone. De nombreuses carcinomes se propagent dans l'abdomen. Les carcinomes péritonéales sont la plus grosse source de mauvais pronostic pour les patientes atteintes d'OC<sup>220</sup>. En effet, ces carcinomes sont très difficilement perceptibles pour le chirurgien. Il ne peut donc pas, à l'heure actuelle, réséquer l'entièreté de celles-ci, par manque de visibilité. C'est pour cela, que l'utilisation du SpiderMass pourrait, à l'aide d'un simple tir, déterminer l'état physiopathologique du péritoine et donc permettre aux chirurgiens de retirer l'entièreté des cellules tumorales. L'étude des métastases péritonéales est donc un enjeu majeur. La poursuite de cette étude sera donc d'évaluer la compétence de ces nouveaux modèles dans la détection de celles-ci. L'origine des différents cancers de l'ovaire a un poids important dans distinction des classes (sous-type). C'est pourquoi notre instrument devrait être très facilement en mesure de distinguer une carcinome de cancer de l'ovaire dans le tissu péritonéal. Le chirurgien pourra donc retirer efficacement la totalité des carcinomes situé dans le péritoine et permettre une survie prolongée de la patiente.

Les marqueurs spécifiques des cellules immunitaires et l'immune score permettent une meilleure compréhension du TME. Seulement, ce modèle reste encore incomplet. En effet, le TME est constitué non seulement de cellules immunitaires, mais également de cellules de soutien essentielles à l'angiogenèse et à l'ancrage tumoral. Afin d'améliorer la compréhension des mécanismes cellulaires, il est nécessaire de parfaire ce modèle à travers un entraînement sur d'autres de types cellulaires présents au sein du TME. À l'heure actuelle, ce modèle donne des résultats très prometteurs. En effet, il a permis d'évaluer rapidement l'infiltration immunitaire et le caractère immunosuppresseur de la tumeur. À l'avenir, une meilleure compréhension de chaque tumeur pourrait être possible grâce à une cartographie complète du TME. Ainsi, la définition du profil immunitaire du patient aidera la prise de décision du clinicien vis-à-vis du traitement à lui administrer. Cette avancée est donc en lien avec les recherches thérapeutiques où l'immune-score aidera à estimer précisément l'efficacité d'une

chimiothérapie, ou d'une immunothérapie, en indiquant la présence ou non de cellules immunitaires cibles. Pour le moment, ce modèle exprime une probabilité de présence d'un type cellulaire dans le tissu. Cette probabilité semble proche des estimations réalisées sur l'abondance dans les tumeurs. Cependant, cette approche n'est pas quantitative car elle est basée sur des probabilités. Cette problématique pourrait être réduite en analysant un nombre plus conséquent de chacun des types cellulaires. En effet, plus le nombre d'échantillon permettant l'apprentissage sera grand, plus le modèle sera capable de reconnaître avec précision l'ensemble des spectres. Par ailleurs, cette problématique est également liée à la diversité des types cellulaires ayant servi à l'entraînement du modèle. En effet, plus la diversité de type cellulaire ayant servi à l'entraînement du modèle sera proche de celle observé dans le TME, plus les performances de celui-ci seront grandes.

## Bibliographie

1. Bray, F., Ferlay, J. & Soerjomataram, I. Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA CANCER J CLIN* 394–424 (2018) doi:10.3322/caac.21492.
2. Matulonis, U. A., Sood, A. K., Fallowfield, L., Howitt, B. E. & Sehouli, J. Ovarian cancer. *Nat Rev Dis Primers* (2016) doi:10.1038/nrdp.2016.61.
3. Cheng, H. *et al.* Opportunities and Challenges of the Human Microbiome in Ovarian Cancer. *Front Oncol* **10**, (2020).
4. Idahl, A. *et al.* Chlamydia trachomatis and mycoplasma genitalium plasma antibodies in relation to epithelial ovarian tumors. *Infect Dis Obstet Gynecol* **2011**, (2011).
5. Bristow, R. E., Duska, L. R. & Montz, F. J. The role of cytoreductive surgery in the management of stage IV uterine papillary serous carcinoma. *Gynecol Oncol* **81**, 92–99 (2001).
6. Shin, D. Clinical importance and surgical decision-making regarding proximal resection margin for gastric cancer. *World J Gastrointest Oncol* **5**, 4 (2013).
7. Neumann, E. K., Djambazova, K. V., Caprioli, R. M. & Spraggins, J. M. Multimodal Imaging Mass Spectrometry: Next Generation Molecular Mapping in Biology and Medicine. *J Am Soc Mass Spectrom* **31**, 2401–2415 (2020).
8. 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).
9. Ogrinc, N., Saudemont, P., Takats, Z., Salzert, M. & Fournier, I. Cancer Surgery 2.0: Guidance by Real-Time Molecular Technologies. *Trends Mol Med* **27**, 602–615 (2021).
10. 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).
11. Balog, J. *et al.* Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci Transl Med* **5**, (2013).

12. Balog, J. *et al.* Identification of biological tissues by rapid evaporative ionization mass spectrometry. *Anal Chem* **82**, 7343–7350 (2010).
13. 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**, 118, 1349–1358 (2018).
14. 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**, (2017).
15. Sans, M. *et al.* Performance of the MasSpec pen for rapid diagnosis of Ovarian cancer. *Clin Chem* **65**, 674–683 (2019).
16. 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**, (2019).
17. Saudemont, P. *et al.* Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption / Ionization Mass Spectrometry Technology Article Real-Time Molecular Diagnosis of Tumors Using Water-Assisted Laser Desorption / Ionization Mass Spectrometry Technology. *Cancer Cell* **34**, 840–851 (2018).
18. Fatou, B. *et al.* Real time and in vivo pharmaceutical and environmental studies with SpiderMass instrument. *J Biotechnol* **281**, 61–66 (2018).
19. 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).
20. Ogrinc, N. *et al.* Robot-assisted SpiderMass for in vivo real-time topography mass spectrometry imaging. *bioRxiv* **33**, 1–24 (2020).
21. Mingels, M. J. J. M. *et al.* Müllerian precursor lesions in serous ovarian cancer patients: Using the SEE-Fim and SEE-End protocol. *Modern Pathology* **27**, 1002–1013 (2014).
22. Ferlay, J. *et al.* Global Cancer Observatory: Cancer Today. <https://gco.iarc.fr/today> (2023).
23. Lisio, M. A., Fu, L., Goyeneche, A., Gao, Z. H. & Telleria, C. High-grade serous ovarian cancer: Basic sciences, clinical and therapeutic standpoints. *International Journal of Molecular Sciences* vol. 20 Preprint at <https://doi.org/10.3390/ijms20040952> (2019).
24. Badgwell, D. & Bast, R. C. *Early detection of ovarian cancer. Disease Markers* vol. 23 (2007).
25. Goff, B. Symptoms Associated With Ovarian Cancer. *Clin Obstet Gynecol* **55**, 36–42 (2012).
26. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J Clin* **73**, 17–48 (2023).
27. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
28. Lheureux, S., Gourley, C., Vergote, I. & Oza, A. M. Epithelial ovarian cancer. *The Lancet* vol. 393 1240–1253 Preprint at [https://doi.org/10.1016/S0140-6736\(18\)32552-2](https://doi.org/10.1016/S0140-6736(18)32552-2) (2019).
29. Wentzensen, N. *et al.* Ovarian cancer risk factors by histologic subtype: An analysis from the Ovarian Cancer Cohort Consortium. *Journal of Clinical Oncology* **34**, 2888–2898 (2016).

30. Hans F. *et al.* New Clinical Criteria for Hereditary Nonpolyposis Colorectal Cancer (HNPCC, Lynch Syndrome) Proposed by the International Collaborative Group on HNPCC. *Gastroenterology* **116**, 1453–1456 (1999).
31. Reid, B. M., Permuth, J. B. & Sellers, T. A. Epidemiology of ovarian cancer: a review. *Cancer Biology and Medicine* vol. 14 9–32 Preprint at <https://doi.org/10.20892/j.issn.2095-3941.2016.0084> (2017).
32. Folkins, A. K. & Longacre, T. A. Hereditary gynaecological malignancies: Advances in screening and treatment. *Histopathology* vol. 62 2–30 Preprint at <https://doi.org/10.1111/his.12028> (2013).
33. Goff, B. A., Mandel, L., Muntz, H. G. & Melancon, C. H. Ovarian carcinoma diagnosis: Results of a national ovarian cancer survey. *Cancer* **89**, 2068–2075 (2000).
34. Reavis, H. D. & Drapkin, R. The tubal epigenome – An emerging target for ovarian cancer. *Pharmacology and Therapeutics* vol. 210 Preprint at <https://doi.org/10.1016/j.pharmthera.2020.107524> (2020).
35. Karnezis, A. N., Cho, K. R., Gilks, C. B., Pearce, C. L. & Huntsman, D. G. The disparate origins of ovarian cancers: Pathogenesis and prevention strategies. *Nature Reviews Cancer* vol. 17 65–74 Preprint at <https://doi.org/10.1038/nrc.2016.113> (2017).
36. Visvanathan, K. *et al.* Diagnosis of serous tubal intraepithelial carcinoma based on morphologic and immunohistochemical features: A reproducibility study. *American Journal of Surgical Pathology* vol. 35 1766–1775 Preprint at <https://doi.org/10.1097/PAS.0b013e31822f58bc> (2011).
37. Babaier, A. & Ghatage, P. Mucinous cancer of the ovary: Overview and current status. *Diagnostics* vol. 10 Preprint at <https://doi.org/10.3390/diagnostics10010052> (2020).
38. Weissleder, R. & Pittet, M. J. Imaging in the era of molecular oncology. *Nature* vol. 452 580–589 Preprint at <https://doi.org/10.1038/nature06917> (2008).
39. Fagotti, A. *et al.* A laparoscopy-based score to predict surgical outcome in patients with advanced ovarian carcinoma: A pilot study. *Ann Surg Oncol* **13**, 1156–1161 (2006).
40. Sugarbaker Paul.H. Successful management of microscopic residual disease in large bowel cancer. *Cancer Chemother Pharmacology* **43**, 15–25 (1999).
41. Soslow, R. A. Histologic subtypes of ovarian carcinoma: An overview. *International Journal of Gynecological Pathology* **27**, 161–174 (2008).
42. Bois, A. Du *et al.* Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: A combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the arbeitgemeinschaft gynaekologische onkologie studien-gruppe ovarialkarzinom (AGO-OVAR) and the groupe d'investigateurs nationaux pour les études des cancers de l'Ovaire (GINECO). *Cancer* **115**, 1234–1244 (2009).
43. Fotopoulou, C. *et al.* European Society of Gynaecological Oncology guidelines for the peri-operative management of advanced ovarian cancer patients undergoing debulking surgery. *International Journal of Gynecological Cancer* **31**, 1199–1206 (2021).

44. Hacker, K. E., Uppal, S. & Johnston, C. *Principles of Treatment for Borderline, Micropapillary Serous, and Low-Grade Ovarian Cancer NCCN: Continuing Education Accreditation Statement. JNCCN-Journal of the National Comprehensive Cancer Network* | vol. 14 <http://education.nccn.org/> (2016).
45. Stewart, C., Ralyea, C. & Lockwood, S. Ovarian Cancer: An Integrated Review. *Seminars in Oncology Nursing* vol. 35 151–156 Preprint at <https://doi.org/10.1016/j.soncn.2019.02.001> (2019).
46. Mallen, A. *et al.* Surgical prevention strategies in ovarian cancer. *Gynecologic Oncology* vol. 151 166–175 Preprint at <https://doi.org/10.1016/j.ygyno.2018.08.005> (2018).
47. Walker, J. L. *et al.* Society of Gynecologic Oncology recommendations for the prevention of ovarian cancer. *Cancer* vol. 121 2108–2120 Preprint at <https://doi.org/10.1002/cncr.29321> (2015).
48. Leblanc, E. *et al.* Prophylactic Radical Fimbriectomy with Delayed Oophorectomy in Women with a High Risk of Developing an Ovarian Carcinoma: Results of a Prospective National Pilot Study. *Cancers (Basel)* **15**, (2023).
49. Leblanc, E. *et al.* Radical fimbriectomy: A reasonable temporary risk-reducing surgery for selected women with a germ line mutation of BRCA 1 or 2 genes? Rationale and preliminary development. *Gynecol Oncol* **121**, 472–476 (2011).
50. Armstrong, D. K. *et al.* Ovarian cancer, version 1.2019 featured updates to the nccn guidelines. *JNCCN Journal of the National Comprehensive Cancer Network* vol. 17 896–909 Preprint at <https://doi.org/10.6004/jnccn.2019.0039> (2019).
51. Blagden, S. P. *et al.* Weekly platinum-based chemotherapy versus 3-weekly platinum-based chemotherapy for newly diagnosed ovarian cancer (ICON8): quality-of-life results of a phase 3, randomised, controlled trial. *Lancet Oncol* **21**, 969–977 (2020).
52. PETER B. SCHIFF, JANE FANT & SUSAN B. HORWITZ. Promotion of microtubule assembly by taxol. *Nature* **277**, 665–667 (1979).
53. Byrd-Leifer, C. A., Block, E. F., Takeda, K., Akira, S. & Ding, A. The role of MyD88 and TLR4 in the LPS-mimetic activity of Taxol. *Eur J Immunol* **31**, 2448–2457 (2001).
54. Colombo, N. *et al.* ESMO-ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *International Journal of Gynecological Cancer* **29**, 728–760 (2019).
55. Monk, B. J. & Chan, J. K. Is intraperitoneal chemotherapy still an acceptable option in primary adjuvant chemotherapy for advanced ovarian cancer? *Annals of Oncology* **28**, viii40–viii45 (2017).
56. Pradipta, A. R. *et al.* Emerging Technologies for Real-Time Intraoperative Margin Assessment in Future Breast-Conserving Surgery. *Advanced Science* vol. 7 Preprint at <https://doi.org/10.1002/adv.201901519> (2020).
57. Allaupe, P. *et al.* Artificial Intelligence-Based Opportunities in Liver Pathology—A Systematic Review. *Diagnostics* vol. 13 Preprint at <https://doi.org/10.3390/diagnostics13101799> (2023).

58. Nicoleta Giurgea, L., Ungureanu, C., Sultana Mihailovici, M. & County Hospital, E. The immunohistochemical expression of p53 and Ki67 in ovarian epithelial borderline tumors. Correlation with clinicopathological factors. *Rom J Morphol Embryol* **2012**, 967–973.
59. Abeytunge, S. *et al.* Evaluation of breast tissue with confocal strip-mosaicking microscopy: a test approach emulating pathology-like examination. *J Biomed Opt* **22**, 034002 (2017).
60. Orosco, R. K. *et al.* Positive Surgical Margins in the 10 Most Common Solid Cancers. *Sci Rep* **8**, (2018).
61. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J Biomed Opt* **19**, 010901 (2014).
62. Fei, B. *et al.* Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients. *J Biomed Opt* **22**, 1 (2017).
63. Veeratterapillay, R. *et al.* Intraoperative and surgical specimen (ex vivo) ultrasound in the assessment of margins at partial nephrectomy. *Int Urol Nephrol* **47**, 1665–1669 (2015).
64. Lauwerends, L. J. *et al.* Real-time fluorescence imaging in intraoperative decision making for cancer surgery. *The Lancet Oncology* vol. 22 e186–e195 Preprint at [https://doi.org/10.1016/S1470-2045\(20\)30600-8](https://doi.org/10.1016/S1470-2045(20)30600-8) (2021).
65. Hamdoon, Z., Jerjes, W., McKenzie, G., Jay, A. & Hopper, C. Optical coherence tomography in the assessment of oral squamous cell carcinoma resection margins. *Photodiagnosis Photodyn Ther* **13**, 211–217 (2016).
66. Zysk, A. M. *et al.* Intraoperative Assessment of Final Margins with a Handheld Optical Imaging Probe During Breast-Conserving Surgery May Reduce the Reoperation Rate: Results of a Multicenter Study. *Ann Surg Oncol* **22**, 3356–3362 (2015).
67. Yoon, J. Hyperspectral Imaging for Clinical Applications. *Biochip Journal* vol. 16 Preprint at <https://doi.org/10.1007/s13206-021-00041-0> (2022).
68. Franck, J. *et al.* MALDI imaging mass spectrometry: State of the art technology in clinical proteomics. *Molecular and Cellular Proteomics* vol. 8 2023–2033 Preprint at <https://doi.org/10.1074/mcp.R800016-MCP200> (2009).
69. Zhao, G., Cardenas, H. & Matei, D. Ovarian cancer—why lipids matter. *Cancers* vol. 11 Preprint at <https://doi.org/10.3390/cancers11121870> (2019).
70. Squires, G. *Francis Aston and the mass spectrograph*. *J. Chem. Soc., Dalton Trans* (1998).
71. Karas, M. & H. F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal Chem* **60**, 2299–2301 (1988).
72. Tanaka, K. *et al.* Protein and Polymer Analyses up to *m/z* 100 000 by Laser Ionization Time-of-flight Mass Spectrometry. *COMMUNICATIONS IN MASS SPECTROMETRY* vol. 2 (1988).
73. Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* (1979) **246**, 64–71 (1989).
74. Amstalden van Hove, E. R., Smith, D. F. & Heeren, R. M. A. A concise review of mass spectrometry imaging. *J Chromatogr A* **1217**, 3946–3954 (2010).

75. Konermann, L., Ahadi, E., Rodriguez, A. D. & Vahidi, S. Unraveling the mechanism of electrospray ionization. *Anal Chem* **85**, 2–9 (2013).
76. Perez, C. J., Bagga, A. K., Prova, S. S., Yousefi Taemeh, M. & Ifa, D. R. Review and perspectives on the applications of mass spectrometry imaging under ambient conditions. *Rapid Communications in Mass Spectrometry* **33**, 27–53 (2019).
77. Wiseman, J. M., Gologan, B. & Cooks, R. G. Mass Spectrometry Sampling Under Ambient Conditions with Desorption Electrospray Ionization. **306**, 471–474 (2004).
78. Javanshad, R., Maser, T. L., Honarvar, E. & Venter, A. R. The Addition of Polar Organic Solvent Vapors During the Analysis of Proteins by DESI-MS. (2019) doi:10.1007/s13361-019-02345-w.
79. Takats, Z. *et al.* Desorption Electrospray Ionization ( DESI ) Analysis of Tryptic Digests / Peptides Desorption Electrospray Ionization ( DESI ) Analysis of Tryptic Digests / Peptides. (2014) doi:10.1101/pdb.prot4993.
80. Eberlin, L. S., Ferreira, C. R., Dill, A. L., Ifa, D. R. & Cooks, R. G. Desorption electrospray ionization mass spectrometry for lipid characterization and biological tissue imaging. *Biochim Biophys Acta Mol Cell Biol Lipids* **1811**, 946–960 (2011).
81. Cordeiro, F. B. *et al.* Mammalian ovarian lipid distributions by desorption electrospray ionization – mass spectrometry ( DESI-MS ) imaging. (2020).
82. Cowart, L. A. Sphingolipids: players in the pathology of metabolic disease. *Trends in Endocrinology and Metabolism* **20**, 34–42 (2009).
83. Sans, M. *et al.* Metabolic markers and statistical prediction of serous ovarian cancer aggressiveness by ambient ionization mass spectrometry imaging. *Cancer Res* **77**, 2903–2913 (2017).
84. Krieger, A. C., Macias, L. A., Goodman, J. C., Brodbelt, J. S. & Eberlin, L. S. Mass Spectrometry Imaging Reveals Abnormalities in Cardiolipin Composition and Distribution in Astrocytoma Tumor Tissues. *Cancers (Basel)* **15**, (2023).
85. Dannhorn, A. *et al.* Targeted Desorption Electrospray Ionization Mass Spectrometry Imaging for Drug Distribution, Toxicity, and Tissue Classification Studies. *Metabolites* **13**, 377 (2023).
86. Tillner, J. *et al.* Faster, More Reproducible DESI-MS for Biological Tissue Imaging. *J Am Soc Mass Spectrom* **28**, 2090–2098 (2017).
87. Roach, P. J., Laskin, J. & Laskin, A. Nanospray desorption electrospray ionization: An ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst* **135**, 2233–2236 (2010).
88. Haddad, R., Milagre, H. M. S., Catharino, R. R. & Eberlin, M. N. Easy ambient sonic-spray ionization mass spectrometry combined with thin-layer chromatography. *Anal Chem* **80**, 2744–2750 (2008).
89. Garrett, R. *et al.* Ambient mass spectrometry employed for direct analysis of intact Arabica coffee beans. *J Braz Chem Soc* **25**, 1172–1177 (2014).
90. Jialing Zhang *et al.* Nondestructive tissue analysis for ex vivo and in vivo cancer diagnosis using a handheld mass spectrometry system. *Sci Transl Med* **9**, (2017).

91. 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).
92. Zhang, J. *et al.* Direct Molecular Analysis of In Vivo and Freshly Excised Tissues in Human Surgeries with the MasSpec Pen Technology. doi:10.1101/2020.12.14.20248101.
93. Pavlovich, M. J., Musselman, B. & Hall, A. B. Direct analysis in real time—Mass spectrometry (DART-MS) in forensic and security applications. *Mass Spectrometry Reviews* vol. 37 171–187 Preprint at <https://doi.org/10.1002/mas.21509> (2018).
94. Pierce, C. Y. *et al.* Ambient generation of fatty acid methyl ester ions from bacterial whole cells by direct analysis in real time (DART) mass spectrometry. *Chemical Communications* 807–809 (2007) doi:10.1039/b613200f.
95. Busman, M., Liu, J., Zhong, H., Bobell, J. R. & Maragos, C. M. Determination of the aflatoxin AFB1 from corn by direct analysis in real time-mass spectrometry (DART-MS). *Food Additives and Contaminants - Part A* **31**, 932–939 (2014).
96. Harper, J. D. *et al.* Low-temperature plasma probe for ambient desorption ionization. *Anal Chem* **80**, 9097–9104 (2008).
97. Laiko, V. V., Baldwin, M. A. & Burlingame, A. L. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* **72**, 652–657 (2000).
98. Jens Soltwisch *et al.* Mass spectrometry imaging with laser-induced postionization. *Scienceexpress* 1–10 Preprint at <https://doi.org/10.1126/science.aaa1051> (2015).
99. Diologent, L., Bolbach, G., Focsa, C., Ziskind, M. & Fournier, I. Laser induced post-desolvation of MALDI clusters. *Int J Mass Spectrom* **416**, 29–36 (2017).
100. Nemes, P. & Vertes, A. Laser Ablation Electrospray Ionization for Atmospheric Pressure, in Vivo, and Imaging Mass Spectrometry. *Anal Chem* **79**, 8098–8106 (2007).
101. Nemes, P. & Vertes, A. Laser ablation electrospray ionization for atmospheric pressure molecular imaging mass spectrometry. *Methods in Molecular Biology* **656**, 159–171 (2010).
102. Li, H. & Vertes, A. Solvent gradient electrospray for laser ablation electrospray ionization mass spectrometry. *Analyst* **142**, 2921–2927 (2017).
103. Sampson, J. S., Hawkrige, 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).
104. Ogrinc, N. *et al.* Robot-Assisted SpiderMass for In Vivo Real-Time Topography Mass Spectrometry Imaging. *Anal Chem* **2**, 14383–14391 (2021).
105. Seddiki, K. *et al.* Cumulative learning enables convolutional neural network representations for small mass spectrometry data classification. *Nat Commun* **11**, (2020).
106. Jowett, N. *et al.* Bone ablation without thermal or acoustic mechanical injury via a novel picosecond infrared laser (PIRL). *Otolaryngology - Head and Neck Surgery (United States)* **150**, 385–393 (2014).

107. Kwiatkowski, M. *et al.* Ultrafast extraction of proteins from tissues using desorption by impulsive vibrational excitation. *Angewandte Chemie - International Edition* **54**, 285–288 (2015).
108. 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).
109. 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**, (2017).
110. Matthias Schluter, M. M. F. *et al.* Towards OCT-Navigated Tissue Ablation with a Picosecond Infrared Laser (PIRL) and Mass-Spectrometric Analysis. *Annu Int Conf IEEE Eng Med Biol Soc* 158–161 (2019).
111. Katz, L. *et al.* Dual Laser and Desorption Electrospray Ionization Mass Spectrometry Imaging Using the Same Interface. *Anal Chem* **92**, 6349–6357 (2020).
112. Woolman, M. *et al.* In situ tissue pathology from spatially encoded mass spectrometry classifiers visualized in real time through augmented reality. *Chem Sci* **11**, 8723–8735 (2020).
113. Phelps, D. L., Borley, J. V., Brown, R., Takáts, Z. & Ghaem-Maghami, S. The use of biomarkers to stratify surgical care in women with ovarian cancer: Scientific Impact Paper No. 69 May 2022. *BJOG* **129**, e66–e74 (2022).
114. Jannetto, P. J. & Fitzgerald, R. L. Effective use of mass spectrometry in the clinical laboratory. *Clin Chem* **62**, 92–98 (2016).
115. Amstalden van Hove, E. R., Smith, D. F. & Heeren, R. M. A. A concise review of mass spectrometry imaging. *Journal of Chromatography A* vol. 1217 3946–3954 Preprint at <https://doi.org/10.1016/j.chroma.2010.01.033> (2010).
116. Ayed, M. El. MALDI imaging mass spectrometry in ovarian cancer for tracking, identifying, and validating biomarkers %5. *Med Sci Monit* **16**, 233–245 (2010).
117. Liu, Y. *et al.* Elevation of sulfatides in ovarian cancer: An integrated transcriptomic and lipidomic analysis including tissue-imaging mass spectrometry. <http://www.molecular-cancer.com/content/9/1/186> (2010).
118. Aichler, M. & Walch, A. MALDI Imaging mass spectrometry: Current frontiers and perspectives in pathology research and practice. *Laboratory Investigation* **95**, 422–431 (2015).
119. Robichaud, G., Barry, J. A., Garrard, K. P. & Muddiman, D. C. Infrared matrix-assisted laser desorption electrospray ionization (IR-MALDESI) imaging source coupled to a FT-ICR mass spectrometer. *J Am Soc Mass Spectrom* **24**, 92–100 (2013).
120. Nemes, P. & Vertes, A. Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry. *Anal Chem* **79**, 8098–8106 (2007).
121. Nemes, P., Barton, A. A. & Vertes, A. Three-dimensional imaging of metabolites in tissues under ambient conditions by laser ablation electrospray ionization mass spectrometry. *Anal Chem* **81**, 6668–6675 (2009).

122. Handley, K. F. *et al.* Classification of High-Grade Serous Ovarian Cancer Using Tumor Morphologic Characteristics. *JAMA Netw Open* **5**, E2236626 (2022).
123. Massonnet, P. & Heeren, R. M. A. A concise tutorial review of TOF-SIMS based molecular and cellular imaging. *J Anal At Spectrom* **34**, 2217–2228 (2019).
124. Song, Y. & Cooks, R. G. Atmospheric pressure ion/molecule reactions for the selective detection of nitroaromatic explosives using acetonitrile and air as reagents. *Rapid Communications in Mass Spectrometry* **20**, 3130–3138 (2006).
125. Chen, H. *et al.* Surface desorption atmospheric pressure chemical ionization mass spectrometry for direct ambient sample analysis without toxic chemical contamination. *Journal of Mass Spectrometry* **42**, 1045–1056 (2007).
126. Anderson, N. M. & Simon, M. C. *The tumor microenvironment*. *Current Biology* vol. 30 (2020).
127. Sjöberg, E., Augsten, M., Bergh, J., Jirstrom, K. & Östman, A. Expression of the chemokine CXCL14 in the tumour stroma is an independent marker of survival in breast cancer. *Br J Cancer* **114**, 1117–1124 (2016).
128. Yang, Y., Yang, Y., Yang, J., Zhao, X. & Wei, X. Tumor Microenvironment in Ovarian Cancer: Function and Therapeutic Strategy. *Frontiers in Cell and Developmental Biology* vol. 8 Preprint at <https://doi.org/10.3389/fcell.2020.00758> (2020).
129. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* vol. 144 646–674 Preprint at <https://doi.org/10.1016/j.cell.2011.02.013> (2011).
130. Brown, Y., Hua, S. & Tanwar, P. S. Extracellular matrix in high-grade serous ovarian cancer: Advances in understanding of carcinogenesis and cancer biology. *Matrix Biology* vol. 118 16–46 Preprint at <https://doi.org/10.1016/j.matbio.2023.02.004> (2023).
131. Minciacchi, V. R., Freeman, M. R. & Vizio, D. Di. Extracellular Vesicles in Cancer: Exosomes, Microvesicles and the Emerging Role of Large Oncosomes. *Semin Cell Dev Biol* **40**, 41–51 (2015).
132. Crusio, W. E., Lambris, J. D. & Radeke, H. H. *Advances in Experimental Medicine and Biology*. *Advances in Experimental Medicine and Biology* vol. 1234 (2020).
133. Franklin, R. A. & Li, M. O. Ontogeny of Tumor-Associated Macrophages and Its Implication in Cancer Regulation. *Trends in Cancer* vol. 2 20–34 Preprint at <https://doi.org/10.1016/j.trecan.2015.11.004> (2016).
134. Wynn, T. a., Chawla, A. & Pollard, J. W. Origins and Hallmarks of Macrophages: Development, Homeostasis, and Disease. *Nature* **496**, 445–455 (2013).
135. Zhang, Z., Huang, L. & Brayboy, L. Macrophages: An indispensable piece of ovarian health. *Biology of Reproduction* vol. 104 527–538 Preprint at <https://doi.org/10.1093/biolre/ioaa219> (2021).
136. Schweer, D. *et al.* Tumor-Associated Macrophages and Ovarian Cancer: Implications for Therapy. *Cancers* vol. 14 Preprint at <https://doi.org/10.3390/cancers14092220> (2022).
137. Chambers, S. E. J., O’Neill, C. L., O’Doherty, T. M., Medina, R. J. & Stitt, A. W. The role of immune-related myeloid cells in angiogenesis. *Immunobiology* **218**, 1370–1375 (2013).

138. Shapouri-Moghaddam, A. *et al.* Macrophage plasticity, polarization, and function in health and disease. *Journal of Cellular Physiology* vol. 233 6425–6440 Preprint at <https://doi.org/10.1002/jcp.26429> (2018).
139. Mantovani, A., Schioppa, T., Porta, C., Allavena, P. & Sica, A. Role of tumor-associated macrophages in tumor progression and invasion. *Cancer and Metastasis Reviews* vol. 25 315–322 Preprint at <https://doi.org/10.1007/s10555-006-9001-7> (2006).
140. Wang, J., Li, D., Cang, H. & Guo, B. Crosstalk between cancer and immune cells: Role of tumor-associated macrophages in the tumor microenvironment. *Cancer Medicine* vol. 8 4709–4721 Preprint at <https://doi.org/10.1002/cam4.2327> (2019).
141. El-Arabey, A. A. *et al.* GATA3 as a master regulator for interactions of tumor-associated macrophages with high-grade serous ovarian carcinoma. *Cell Signal* **68**, (2020).
142. Chang, W., Fa, H., Xiao, D. & Wang, J. Targeting phosphatidylserine for Cancer therapy: Prospects and challenges. *Theranostics* vol. 10 9214–9229 Preprint at <https://doi.org/10.7150/thno.45125> (2020).
143. Liang, X. *et al.* Phosphatidylserine released from apoptotic cells in tumor induces M2-like macrophage polarization through the PSR-STAT3-JMJD3 axis. *Cancer Commun* **42**, 205–222 (2022).
144. Chen, Y. *et al.* Tumor-associated macrophages: An accomplice in solid tumor progression. *Journal of Biomedical Science* vol. 26 Preprint at <https://doi.org/10.1186/s12929-019-0568-z> (2019).
145. Zhang, M. *et al.* A high M1/M2 ratio of tumor-associated macrophages is associated with extended survival in ovarian cancer patients. *J Ovarian Res* **7**, (2014).
146. Macciò, A. *et al.* Role of M1-polarized tumor-associated macrophages in the prognosis of advanced ovarian cancer patients. *Sci Rep* **10**, (2020).
147. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *J Cell Sci* **125**, 5591–5596 (2012).
148. Fridman, W. H., Pagès, F., Sauts-Fridman, C. & Galon, J. The immune contexture in human tumours: Impact on clinical outcome. *Nature Reviews Cancer* vol. 12 298–306 Preprint at <https://doi.org/10.1038/nrc3245> (2012).
149. Voskoboinik, I., Whisstock, J. C. & Trapani, J. A. Perforin and granzymes: Function, dysfunction and human pathology. *Nature Reviews Immunology* vol. 15 388–400 Preprint at <https://doi.org/10.1038/nri3839> (2015).
150. Knutson, K. L. *et al.* Regulatory T cells, inherited variation, and clinical outcome in epithelial ovarian cancer. *Cancer Immunology, Immunotherapy* **64**, 1495–1504 (2015).
151. Di Santo, J. P. Functionally distinct NK-cell subsets: Developmental origins and biological implications. *European Journal of Immunology* vol. 38 2948–2951 Preprint at <https://doi.org/10.1002/eji.200838830> (2008).
152. Crum, C. P. *et al.* Lessons from BRCA: The tubal fimbria emerges as an origin for pelvic serous cancer. *Clinical Medicine and Research* vol. 5 35–44 Preprint at <https://doi.org/10.3121/cmr.2007.702> (2007).

153. Medeiros, F. *et al.* *The Tubal Fimbria Is a Preferred Site for Early Adenocarcinoma in Women With Familial Ovarian Cancer Syndrome.*
154. Gan, C. *et al.* Persistence of fimbrial tissue on the ovarian surface after salpingectomy. *Am J Obstet Gynecol* **217**, 425.e1-425.e16 (2017).
155. Seidman, J. D., Yemelyanova, A., Zaino, R. J. & Kurman, R. J. The fallopian tube-peritoneal junction: A potential site of carcinogenesis. *International Journal of Gynecological Pathology* **30**, 4–11 (2011).
156. Visvanathan, K. *et al.* Diagnosis of serous tubal intraepithelial carcinoma based on morphologic and immunohistochemical features: A reproducibility study. *American Journal of Surgical Pathology* vol. 35 1766–1775 Preprint at <https://doi.org/10.1097/PAS.0b013e31822f58bc> (2011).
157. Bayarkhangai, B. *et al.* A comprehensive and perspective view of oncoprotein SET in cancer. *Cancer Medicine* vol. 7 3084–3094 Preprint at <https://doi.org/10.1002/cam4.1526> (2018).
158. Liang, W. *et al.* MARCKSL1 promotes the proliferation, migration and invasion of lung adenocarcinoma cells. *Oncol Lett* **19**, 2272–2280 (2020).
159. Bonnet-Magnaval, F. & DesGroseillers, L. The Staufen1-dependent cell cycle regulon or how a misregulated RNA-binding protein leads to cancer. *Biological Reviews* **96**, 2192–2208 (2021).
160. Zhang, X. Y., Hong, S. S., Cai, Q. Q., Zhang, M. X. & Xu C J. Proteomic alterations of fibroblasts induced by ovarian cancer cells reveal potential cancer targets. *Neoplasma* **65**, 104–112 (2018).
161. Kahlert, C. *et al.* Increased expression of ALCAM/CD166 in pancreatic cancer is an independent prognostic marker for poor survival and early tumour relapse. *Br J Cancer* **101**, 457–464 (2009).
162. Lunter, P. C. *et al.* Activated leukocyte cell adhesion molecule (ALCAM/CD166/MEMD), a novel actor in invasive growth, controls matrix metalloproteinase activity. *Cancer Res* **65**, 8801–8808 (2005).
163. Creekmore, A. L. *et al.* Changes in gene expression and cellular architecture in an ovarian cancer progression model. *PLoS One* **6**, (2011).
164. Peluso, J. J. Progesterone signaling mediated through progesterone receptor membrane component-1 in ovarian cells with special emphasis on ovarian cancer. *Steroids* vol. 76 903–909 Preprint at <https://doi.org/10.1016/j.steroids.2011.02.011> (2011).
165. Zhao, G. *et al.* Identification of NDRG1-regulated genes associated with invasive potential in cervical and ovarian cancer cells. *Biochem Biophys Res Commun* **408**, 154–159 (2011).
166. Santin, A. D. *et al.* Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: identification of candidate molecular markers for ovarian cancer diagnosis and therapy. *Int J Cancer* **112**, 14–25 (2004).
167. Scatena, R., Bottoni, P., Pontoglio, A. & Giardina, B. Revisiting the Warburg effect in cancer cells with proteomics. The emergence of new approaches to diagnosis, prognosis and therapy. *Proteomics Clin Appl* **4**, 143–158 (2010).

168. Brown, Y., Hua, S. & Tanwar, P. S. Extracellular matrix in high-grade serous ovarian cancer: Advances in understanding of carcinogenesis and cancer biology. *Matrix Biology* vol. 118 16–46 Preprint at <https://doi.org/10.1016/j.matbio.2023.02.004> (2023).
169. Wu, R. C. *et al.* Genomic landscape and evolutionary trajectories of ovarian cancer precursor lesions. *Journal of Pathology* **248**, 41–50 (2019).
170. Kuhn, E. *et al.* Shortened telomeres in serous tubal intraepithelial carcinoma: An early event in ovarian high-grade serous carcinogenesis. *American Journal of Surgical Pathology* **34**, 829–836 (2010).
171. Mir, S. A. *et al.* Lipidomic analysis of archival pathology specimens identifies altered lipid signatures in ovarian clear cell carcinoma. *Metabolites* **11**, (2021).
172. Gharpure, K. M. *et al.* FABP4 as a key determinant of metastatic potential of ovarian cancer. *Nat Commun* **9**, (2018).
173. Wang, W., Bai, L., Li, W. & Cui, J. The Lipid Metabolic Landscape of Cancers and New Therapeutic Perspectives. *Frontiers in Oncology* vol. 10 Preprint at <https://doi.org/10.3389/fonc.2020.605154> (2020).
174. Li, M. *et al.* 3-Phosphoglycerate dehydrogenase: a potential target for cancer treatment. *Cellular Oncology* vol. 44 541–556 Preprint at <https://doi.org/10.1007/s13402-021-00599-9> (2021).
175. Truong, V. *et al.* Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci Rep* **7**, (2017).
176. Blomme, A. *et al.* 2,4-dienoyl-CoA reductase regulates lipid homeostasis in treatment-resistant prostate cancer. *Nat Commun* **11**, (2020).
177. Ackerman, D. *et al.* Triglycerides Promote Lipid Homeostasis during Hypoxic Stress by Balancing Fatty Acid Saturation. *Cell Rep* **24**, 2596-2605.e5 (2018).
178. Ramadan, S. *et al.* Lipid and metabolite deregulation in the breast tissue of women carrying BRCA1 and BRCA2 genetic mutations. *Radiology* **275**, 675–682 (2015).
179. Lemaire, R. *et al.* Tag-Mass: Specific molecular imaging of transcriptome and proteome by mass spectrometry based on photocleavable tag. *J Proteome Res* **6**, 2057–2067 (2007).
180. Dijkstra, T. K. Ridge regression and its degrees of freedom. *Qual Quant* **48**, 3185–3193 (2014).
181. Cortes, C., Mohri, M. & Rostamizadeh, A. L2 Regularization for Learning Kernels. *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009* 109–116 (2012).
182. Allaume, P. *et al.* Artificial Intelligence-Based Opportunities in Liver Pathology—A Systematic Review. *Diagnostics* vol. 13 Preprint at <https://doi.org/10.3390/diagnostics13101799> (2023).
183. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine* vol. 19 1423–1437 Preprint at <https://doi.org/10.1038/nm.3394> (2013).
184. Ke, G. *et al.* *LightGBM: A Highly Efficient Gradient Boosting Decision Tree.* <https://github.com/Microsoft/LightGBM>.

185. Bruni, D., Angell, H. K. & Galon, J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nature Reviews Cancer* vol. 20 662–680 Preprint at <https://doi.org/10.1038/s41568-020-0285-7> (2020).
186. Marliot, F., Pagès, F. & Galon, J. Usefulness and robustness of Immunoscore for personalized management of cancer patients. *OncImmunity* vol. 9 Preprint at <https://doi.org/10.1080/2162402X.2020.1832324> (2020).
187. Koundouros, N. & Poulogiannis, G. Reprogramming of fatty acid metabolism in cancer. *British Journal of Cancer* vol. 122 4–22 Preprint at <https://doi.org/10.1038/s41416-019-0650-z> (2020).
188. Munir, R., Lisec, J., Swinnen, J. V. & Zaidi, N. Lipid metabolism in cancer cells under metabolic stress. *British Journal of Cancer* vol. 120 1090–1098 Preprint at <https://doi.org/10.1038/s41416-019-0451-4> (2019).
189. Pöyry, S. & Vattulainen, I. Role of charged lipids in membrane structures — Insight given by simulations. *Biochim Biophys Acta Biomembr* **1858**, 2322–2333 (2016).
190. Skipper, K. Surface Tension Drives Cancer Cell Migration. *Physics (College Park Md)* **16**, (2023).
191. Lankry, D., Rovis, T. L., Jonjic, S. & Mandelboim, O. The interaction between CD300a and phosphatidylserine inhibits tumor cell killing by NK cells. *Eur J Immunol* **43**, 2151–2161 (2013).
192. Putta, P. *et al.* Phosphatidic acid binding proteins display differential binding as a function of membrane curvature stress and chemical properties. *Biochim Biophys Acta Biomembr* **1858**, 2709–2716 (2016).
193. Yeom, M. *et al.* Glucosylceramide attenuates the inflammatory mediator expression in lipopolysaccharide-stimulated RAW264.7 cells. *Nutrition Research* **35**, 241–250 (2015).
194. 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**, (2023).
195. Miretti, M., González Graglia, M. A., Suárez, A. I. & Prucca, C. G. Photodynamic therapy for glioblastoma: A light at the end of the tunnel. *J Photochem Photobiol* **13**, (2023).
196. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209–249 (2021).
197. Schaff, L. & Mellinghoff, I. Glioblastoma and Other Primary Brain Malignancies in Adults: A Review. *JAMA* **329**, 574–587 (2023).
198. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathologica* vol. 131 803–820 Preprint at <https://doi.org/10.1007/s00401-016-1545-1> (2016).
199. Louis, D. N. *et al.* The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro Oncol* **23**, 1231–1251 (2021).
200. Krell, D. *et al.* Screen for IDH1, IDH2, IDH3, D2HGDH and I2HGDH mutations in glioblastoma. *PLoS One* **6**, (2011).
201. Yan, H. *et al.* IDH1 and IDH2 Mutations in Gliomas *Abstract. N Engl J Med* vol. 360 (2009).

202. Stupp, R. *et al.* *Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma*. [www.nejm.org](http://www.nejm.org).
203. Molinaro, A. M., Taylor, J. W., Wiencke, J. K. & Wrensch, M. R. Genetic and molecular epidemiology of adult diffuse glioma. *Nature Reviews Neurology* vol. 15 405–417 Preprint at <https://doi.org/10.1038/s41582-019-0220-2> (2019).
204. Taylor, O. G., Brzozowski, J. S. & Skelding, K. A. Glioblastoma multiforme: An overview of emerging therapeutic targets. *Frontiers in Oncology* vol. 9 Preprint at <https://doi.org/10.3389/fonc.2019.00963> (2019).
205. Chen, Z. *et al.* Cellular and molecular identity of tumor-associated macrophages in glioblastoma. *Cancer Res* **77**, 2266–2278 (2017).
206. Roesch, S., Rapp, C., Dettling, S. & Herold-Mende, C. When immune cells turn bad—tumor-associated microglia/macrophages in glioma. *International Journal of Molecular Sciences* vol. 19 Preprint at <https://doi.org/10.3390/ijms19020436> (2018).
207. Brown, T. J. *et al.* Association of the extent of resection with survival in glioblastoma a systematic review and meta-Analysis. *JAMA Oncol* **2**, 1460–1469 (2016).
208. Bikfalvi, A. *et al.* Challenges in glioblastoma research: focus on the tumor microenvironment. *Trends in Cancer* vol. 9 9–27 Preprint at <https://doi.org/10.1016/j.trecan.2022.09.005> (2023).
209. 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**, (2023).
210. Salustiano, E. J., da Costa, K. M., Freire-De-Lima, L., Mendonça-Previato, L. & Previato, J. O. Inhibition of glycosphingolipid biosynthesis reverts multidrug resistance by differentially modulating ABC transporters in chronic myeloid leukemias. *Journal of Biological Chemistry* **295**, 6457–6471 (2020).
211. 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).
212. 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. *Journal of Clinical Oncology* **38**, 3638–3651 (2020).
213. Hijazi, A., Antoniotti, C., Cremolini, C. & Galon, J. Light on life: immunoscore immune-checkpoint, a predictor of immunotherapy response. *Oncol Immunology* vol. 12 Preprint at <https://doi.org/10.1080/2162402X.2023.2243169> (2023).
214. Duhamel, M. *et al.* Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. *Nat Commun* **13**, (2022).
215. Balog, J. *et al.* Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci Transl Med* **5**, (2013).
216. 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).

217. Pirro, V., Alfaro, C. M., Jarmusch, A. K., Hattab, E. M. & Cohen-gadol, A. A. Intraoperative assessment of tumor margins during glioma resection by desorption electrospray ionization-mass spectrometry. **114**, (2017).
218. Sharma, P., Aaroe, A., Liang, J. & Puduvalli, V. K. Tumor microenvironment in glioblastoma: Current and emerging concepts. *Neurooncol Adv* **5**, (2023).
219. Kocarnik, J. M. *et al.* Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019 A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncol* **8**, 420–444 (2022).
220. van Baal, J. O. A. M. *et al.* Development of Peritoneal Carcinomatosis in Epithelial Ovarian Cancer: A Review. *Journal of Histochemistry and Cytochemistry* vol. 66 67–83 Preprint at <https://doi.org/10.1369/0022155417742897> (2018).
221. Finn, O. J. Immuno-oncology: Understanding the function and dysfunction of the immune system in cancer. in *Annals of Oncology* vol. 23 (Oxford University Press, 2012).
222. Santoiemma, P. P. & Powell, D. J. Tumor infiltrating lymphocytes in ovarian cancer. *Cancer Biology and Therapy* vol. 16 807–820 Preprint at <https://doi.org/10.1080/15384047.2015.1040960> (2015).
223. Chi, D. S. *et al.* Guidelines and selection criteria for secondary cytoreductive surgery in patients with recurrent, platinum-sensitive epithelial ovarian carcinoma. *Cancer* vol. 106 1933–1939 Preprint at <https://doi.org/10.1002/cncr.21845> (2006).

## Remerciements :

La thèse est un marathon enrichissant. Elle m'a permis d'apprendre constamment et d'assouvir ma curiosité. Seulement, celle-ci s'avère par moments éprouvante et difficile. Tout au long de cette expérience, j'ai eu l'occasion de m'entourer de personnes qui m'ont épaulé à la fois professionnellement et personnellement. Je voudrais terminer cette thèse en les remerciant.

Mes premiers remerciements sont destinés à ma directrice de thèse, le **Pr. Isabelle Fournier**, et au **Pr. Michel Salzet** qui m'ont fait confiance dès mon stage. Grâce à vous, j'ai pu progresser depuis le concours de l'école doctorale. Vos précieux conseils m'ont permis d'évoluer et d'avancer pour le meilleur du projet de recherche.

J'aimerais ensuite remercier l'université de Lille et l'école doctorale biologie santé pour avoir financé ces 3 années de doctorat et m'avoir permis de vivre cette expérience inoubliable.

Je tiens également à remercier mon jury de thèse. L'apport de votre expérience dans l'évaluation de mes travaux est une réelle opportunité. Merci d'avoir accepté de lire et d'évaluer mon travail.

Je tiens ensuite à remercier tous les membres du laboratoire avec qui j'ai pu travailler de près ou de loin et plus particulièrement les 3 doctorants avec qui j'ai partagé cette aventure :

**Kamel Bachiri** : J'ai découvert durant ma deuxième année de master un véritable ami en qui je peux toujours compter. On a eu l'opportunité de réaliser notre thèse et de se serrer les coudes jusqu'au bout. Parfois trop bavard, parfois trop bruyant, on a pu s'épauler tout au long de cette expérience. Cette thèse n'aurait pas été la même sans nos fous rires, nos discussions pseudo-scientifiques, et nos blagues nulles incessantes. Merci beaucoup.

**Diego Fernando Garcia Del Rio** : No necesito un traductor para agradecerte. Eres lo mejor que he conocido en esta tesis. Nunca me cansaré de llamarte el conquistador y de intentar hablarte en español a pesar de mi mediocre nivel. Ha sido un placer poder hacer esta tesis al mismo tiempo que tú. Nunca olvidaré: "Semper tiempore, Diversis Locis." Gracias amigo.

**Léa Ledoux** : On a commencé cette thèse ensemble sur cette belle technologie qu'est le SpiderMass. C'était un réel plaisir d'évoluer dans cette thèse à tes côtés à la fois sur le plan professionnel et personnel. On a pu se serrer les coudes jusqu'à la fin sans jamais entrer en compétition. Nos discussions et chamailleries ont toujours été un plaisir. Merci à toi.

**Laurine** : Merci pour ta bonté et ta gentillesse. J'ai adoré ces petits moments de debriefing d'après-match le lundi matin, en se plaignant de nos courbatures. Reste telle que tu es, tu es une bouffée d'oxygène. Et je n'oublie pas le "Baaaaaarbeeeeeuc."

**Yanis** : On restera toujours en éternel débat sur à peu près tout et son contraire. C'est un plaisir de discuter avec toi. Merci beaucoup.

I would like to thanks also **Nina**. Thanks you for all the fun we had and the knowledge you shared. I hope we will keep in touch.

Je tiens également à remercier **Adrien**. Même si j'ai passé mon temps à charrier ta gentillesse, ne change rien. Merci pour toute ton aide et ces bons moments.

Merci à toi **Jean Pascal** pour tout le temps que tu as consacré à m'apprendre la spectrométrie de masse. J'adore ton franc parlé et ton énergie. Merci beaucoup.

Merci **Maxence** pour cette belle collaboration qui a abouti une très belle étude. Merci pour ta gentillesse et ta pédagogie. C'était un plaisir de travailler avec toi.

Je tiens également à remercier **Philippe**. Tu as, malgré toi, été mon mentor pendant ma première année de thèse. C'est toi qui m'as tout appris sur le Spider et qui as su trouver les mots quand je perdais confiance en moi. Tu es un vrai couteau suisse et j'ai beaucoup appris à tes côtés. En dehors du travail, rien de tel que de contempler les balles réelles envoyées à tout être vivant dans ton périmètre. Et tout ça, avec un grand sourire qui fait qu'on ne t'en tient pas rigueur, bravo.

Merci également à **Delphine Hudry**. Malgré ton emploi du temps chargé, tu as toujours pris le temps de répondre à mes questions et de m'aider durant cette thèse, avec joie et bonne humeur. Tu m'as également fait comprendre les réelles problématiques liées à mon sujet de thèse.

Je tiens également à remercier **Estelle Laurent** et **Etienne Coyaud** qui ont su être là quand j'en avais le plus besoin. Vous avez le cœur sur la main. Vous m'avez même accueilli chez vous alors que vous ne me connaissiez pas. C'est grâce à vous que j'ai pu aller de l'avant, alors, encore une fois, merci.

Je souhaite évidemment remercier mes proches sans qui rien ne serait possible. Merci à tous mes amis d'enfance. Je ne pourrais pas citer tous vos noms et tout ce que vous m'avez apporté car ce paragraphe serait plus long que ma thèse. Mais grâce à vous, je sais que j'ai une deuxième (trop grande) famille sur qui je pourrais toujours compter.

**Maman** : Tu as toujours tout donné pour tes enfants, parfois même en t'oubliant. Je ne saurais comment te remercier tel que tu le mérites. C'est avant toutes choses toi qui m'as donné l'envie d'aider mon prochain. C'est toi qui m'as donné envie de participer à l'avancée de la recherche médicale. Et c'est grâce à toi que j'en suis là aujourd'hui. Tu as partagé mes joies et mes stress tout au long de cette thèse. Je ne te remercierai jamais assez. Avant le début de ma soutenance, quand mon anxiété sera la plus élevée, je sais que ta présence me rassurera et me permettra de me surpasser, comme toujours.

**Chris** : Tu as tout d'un merveilleux papa même si je ne te l'ai pas toujours fait ressentir. Merci de m'avoir élevé et fait grandir. C'est également grâce à toi que je suis devenu qui je suis aujourd'hui. Merci d'avoir été là pour moi.

**Lisa** : On n'a pas toujours été d'accord sur tout. Après réflexion, on ne l'a sûrement jamais été d'ailleurs. Mais même si nos chemins sont très différents, sans t'en rendre compte, tu m'as apporté énormément. Merci beaucoup.

Tout a commencé par un « **à toi** ». Tu espérais que nos chemins coïncident encore longtemps et ça a été le cas. C'est à mon tour d'espérer me lever chaque matin à tes côtés encore longtemps. Tu as été mon pilier, mon moteur et mon inspiration dans la vie comme dans cette thèse. Tu es l'une des plus belles choses qui me soit arrivée. Merci beaucoup **Mélanie**.

## **Articles scientifiques :**

## ARTICLE OPEN



# Fallopian tube lesions as potential precursors of early ovarian cancer: a comprehensive proteomic analysis

Maxence Wisztorski<sup>1,7</sup> , Soulimane Aboulouard<sup>1,7</sup> , Lucas Roussel<sup>1,7</sup>, Marie Duhamel<sup>1,7</sup>, Philippe Saudemont<sup>1</sup>, Tristan Cardon<sup>1</sup> , Fabrice Narducci<sup>1,2</sup>, Yves-Marie Robin<sup>1,2</sup>, Anne-Sophie Lemaire<sup>1,2</sup>, Delphine Bertin<sup>1,2</sup>, Nawale Hajjaji<sup>1,3</sup>, Firas Kobeissy<sup>4,5</sup> , Eric Leblanc<sup>1,2,8</sup> , Isabelle Fournier<sup>1,6,7,8</sup> and Michel Salzet<sup>1,6,7,8</sup>

© The Author(s) 2023

Ovarian cancer is the leading cause of death from gynecologic cancer worldwide. High-grade serous carcinoma (HGSC) is the most common and deadliest subtype of ovarian cancer. While the origin of ovarian tumors is still debated, it has been suggested that HGSC originates from cells in the fallopian tube epithelium (FTE), specifically the epithelial cells in the region of the tubal-peritoneal junction. Three main lesions, p53 signatures, STILs, and STICs, have been defined based on the immunohistochemistry (IHC) pattern of p53 and Ki67 markers and the architectural alterations of the cells, using the Sectioning and Extensively Examining the Fimbriated End Protocol. In this study, we performed an in-depth proteomic analysis of these pre-neoplastic epithelial lesions guided by mass spectrometry imaging and IHC. We evaluated specific markers related to each preneoplastic lesion. The study identified specific lesion markers, such as CAVIN1, Emilin2, and FBLN5. We also used SpiderMass technology to perform a lipidomic analysis and identified the specific presence of specific lipids signature including dietary Fatty acids precursors in lesions. Our study provides new insights into the molecular mechanisms underlying the progression of ovarian cancer and confirms the fimbria origin of HGSC.

*Cell Death and Disease* (2023)14:644; <https://doi.org/10.1038/s41419-023-06165-5>

## FACTS

1. The better understanding of the biology of pre-neoplastic lesions in the fallopian tube and their role in the development and progression of high-grade serous ovarian cancer.
2. The definition and diagnostic criteria for preneoplastic lesions in the fallopian tube need to be more explicitly outlined to improve consistency across studies and facilitate accurate diagnosis.
3. The development and optimization of on-tissue digestion and microextraction strategies for peptide identification on tissue should continue to be explored to improve spatially resolved proteomics approaches.
4. Future studies should focus on identifying additional biomarkers and developing new analytical techniques to improve the sensitivity and specificity of early detection and diagnosis of high-grade serous ovarian cancer.

## INTRODUCTION

Ovarian carcinoma (OC) ranks as the eighth leading cause of cancer in women worldwide in 2020, responsible for approximately 207,000 deaths out of 314,000 cases in 2020. Amongst the different classes, high-grade serous carcinoma (HGSC), is the most represented (> 75%) and within this class, 10% of the patients are mutated for *BRCA1/2* or present with Lynch syndrome [1–3]. While the general population has a 1.2–1.4% risk of developing OC, women with familial *BRCA* mutations have a significantly higher risk of up to 46%, even at a younger age (between 30 and 35 years old).

Various theories have been proposed regarding the origin of HGSC, some of which suggest the involvement of the epithelial cells in the region of the tubal-peritoneal junction [4–8]. During the last 10 years, special attention has been geared towards pathological examination of operative specimens of early OC and prophylactic bilateral adnexectomy performed in women at risk of OC due to an inherited mutation in *BRCA* genes. Based on Sectioning and Extensively Examining the Fimbriated End Protocol (SEE-FIM) [6], a systematic serial examination of the fallopian tubes

<sup>1</sup>Univ.Lille, Inserm, CHU Lille, U-1192 – Laboratoire Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France. <sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Cancer Center, 59020 Lille, France. <sup>3</sup>Medical Oncology Department, Oscar Lambret Cancer Center, 59020 Lille, France. <sup>4</sup>Department of Neurobiology, Center for Neurotrauma, Multiomics & Biomarkers (CNMB), Morehouse School of Medicine, Atlanta, GA 30310, USA. <sup>5</sup>Department of Biochemistry and Molecular Genetics, Faculty of Medicine, American University of Beirut, Beirut, Lebanon. <sup>6</sup>Institut Universitaire de France, 75000 Paris, France. <sup>7</sup>These authors contributed equally: Maxence Wisztorski, Soulimane Aboulouard, Lucas Roussel, Marie Duhamel, Isabelle Fournier, Michel Salzet. <sup>8</sup>These authors jointly supervised this work: Eric Leblanc, Isabelle Fournier, Michel Salzet. ✉email: e-leblanc@o-lambret.fr; isabelle.fournier@univ-lille.fr; michel.salzet@univ-lille.fr

Edited by Professor Nickolai Barlev

Received: 24 April 2023 Revised: 8 September 2023 Accepted: 20 September 2023

Published online: 30 September 2023

is performed, coupled with IHC evaluation of p53 and Ki-67 expression. In this group of patients, an unusual rate of certain cancer or at least cellular abnormalities was observed in the fallopian tube epithelium (FTE), especially at its terminal end, the fimbria [9, 10].

This junction occurs where the peritoneum covers the serosal surface of the fallopian tube and meets the specialized epithelium of the tubal fimbria and is marked by a transition between the Müllerian and the mesothelial cells [11]. Junctional sites between different types of epithelia are known to be cancer hot spots where neoplastic transformations can occur. This is supported by evidence that cells derived from this transition region exhibit a cancer-prone stem cell phenotype [4, 8, 11]. During the neoplastic transformation in the fallopian tube cells, mutations and/or aberrant morphological features can be observed. These neoplastic lesions have been classified into three main groups: p53 signatures, serous tubal intraepithelial lesions (STIL), and serous tubal intraepithelial carcinoma (STIC). These lesions are defined according to the immunohistochemical (IHC) pattern of p53 and Ki67 markers and the architectural alterations of the cells. The p53 signature is identified by a small number of epithelial cells (10–20) with a proliferation activity like the adjacent normal epithelium, but with a p53 staining pattern corresponding to a missense TP53 mutation. STIC (serous tubal intraepithelial carcinoma) involves many cells with architectural and nuclear alterations, TP53 mutations, and high proliferative activity. STILs (serous tubal intraepithelial lesions) are defined by the accumulation of p53 in more than 20 cells, some of which exhibit morphological abnormalities, and a higher Ki-67 proliferating index (10–40%). STIL is characterized by a lower level of abnormalities compared to STIC along with normal proliferative activity [12, 13].

Series of morphological changes concomitant with multi-step accumulation of molecular and genetic alterations on pre-neoplastic lesions of FTE hypothesized that HGSCs derive their origin, not in the ovary, but in the fimbria part of the fallopian tube [14–16]. PAX8, a Müllerian marker is expressed in most HGSC but not calretinin (mesothelial marker) [17, 18]. This view has not been universally accepted, primarily as it conflicts with traditional theories on the origins of OC, and secondarily owing to variation in the detection of tubal lesions in association with HGSC, which may in turn be due to differences in sampling or difficulties in diagnostic interpretation [19, 20]. Recent studies have provided a stepwise progression of FTE to precursor lesions to carcinoma, with the aid of - p53 signature - STIL - STIC - HGSC sequence's model [15, 19–22].

Our study aims to investigate the molecular mechanisms associated with the spectrum of pre-neoplastic lesions found in the fallopian tube epithelium (FTE). Due to the limited cellularity of these lesions, we performed a spatially resolved proteomics analysis on a discovery cohort consisting of eight patients. Some of the markers identified were validated by immunofluorescence. We also conducted a lipidomic study using SpiderMass technology on a cohort of 27 patients. Finally, we correlated the proteomic and lipidomic markers to better understand the underlying mechanisms. To identify the molecular pathways associated with the different lesions, we have implemented a state-of-the-art spatially resolved proteomics workflow previously validated on different types of cancers [23–25]. However, this work has a novel component by using a pathologist-guided IHC slide of formalin-fixed paraffin-embedded (FFPE) tissue section of patients presented p53 signature lesions, STIL lesions, and STIC lesions which will be compared to non-pathological fallopian tissues and HGSC. Here, we investigate the different pre-neoplastic lesions, their proteome, the mutations detected in proteins, the proteome translated from alternative ORF, so-called the Ghost proteome, and finally construct proteome pathways indicative of underlying mechanisms [26, 27].

We identified several early-stage markers known to be associated in OC with poor (Emilin2, and CAVIN1) or good prognosis (CAVIN2, SPTAN1, FBLN5). Using IHC, we validated the higher expression of CAVIN1/2, SPTAN1, FBLN-5, and Emilin2 in p53, STIL lesions compared to normal FTE and STIC. Finally, we used ambient ionization mass spectrometry with Spidermass technology to perform an analysis of lipidomic profiles on FFPE tissues after dewaxing. The lipidomic data cross-validate the proteomic results, and we identified several discriminative lipids of the p53 signatures, STILs, and STICs, which were related to previously identified pathways of FTE lesions.

## EXPERIMENTAL PROCEDURES

### Reagents and chemicals

High-purity chemicals from various suppliers were purchased for the different experiments and used as received without any further modifications. Acetonitrile (ACN) HPLC grade, Ethanol HPLC grade, Methanol LC-MS grade, and Isopropyl alcohol LC-MS were purchased from CARLO ERBA Reagents SAS (Val de Reuil, France). Water UHPLC-MS was supplied by Fisher Scientific (Illkirch, France). Trifluoroacetic acid and Formic Acid (FA) were obtained from Biosolve B.V. (Valkenswaard, the Netherlands). Additionally, alpha-Cyano-4-hydroxycinnamic acid (4-HCCA), Aniline (ANI), and Glycerol were sourced from Sigma Aldrich (Saint-Quentin-Fallavier, France).

### Experimental design, case selection and statistical rationale

Tissue samples were obtained from patients who participated in a multicentric phase II clinical trial of radical fimbriectomy initiated by Oscar Lambret Cancer Centre in Lille, France. Before enrollment, the trial protocol (NCT-01608074) was approved by the national and institutional review boards (IRB) and accepted by the ethics committee (CPP Nord Ouest IV, Dec. 2010) following the French and European legislation.

### Ethics approval and consent to participate

Patients provided written informed consent before participating in the trial. To protect patient privacy, no personal information was used in these experiments, and a random number was assigned to each sample. For more information about the participation criteria and study plan for this clinical trial, please refer to the Clinical Trials website (<https://clinicaltrials.gov/ct2/show/NCT01608074>). We took necessary measures to ensure that diagnostic lesions were not lost for subsequent clinical follow-up. In our spatially resolved proteomics approach, we used a slide that had already been used in IHC for our analyses. The sample size for this study was limited due to the rarity of the target population, which consists of patients with BRCA mutations undergoing prophylactic adnexectomy, as well as the need for a sufficient sample size to evaluate the feasibility and safety of the surgical procedure. It is also worth noting that not all recruited patients presented with preneoplastic lesions. However, it should be emphasized that there is currently no available information on the prevalence of preneoplastic lesions in the fimbria of patients with *BRCA1 mutations* or family histories who undergo prophylactic adnexectomy. In our study, we performed proteomic analysis using a relatively small number of patients. It is worth noting that increasing the number of samples is not always the optimal solution, as other important factors such as ethical considerations, resource availability, and sustainability need to be considered. Moreover, previous studies have shown that even with a limited number of samples, statistically significant results can still be obtained. Therefore, in our study, we chose to prioritize the quality of the samples to ensure accurate and reliable results for the discovery part and increase the number of patients for the validation part. A prospective cohort of 8 patients was selected to perform proteomics analysis, including 4 patients with

**Table 1.** a: Description of clinical characteristics of patients with a preneoplastic lesion in the fallopian tube fimbria (prospective cohort). b: Description of clinical characteristics of patients with a preneoplastic lesion in the fallopian tube fimbria (validation cohort).

Case ID	Lesions sampled	BRCA status	Age at diagnosis	Location in fimbria	Ovary lesion	Other lesions	Personal history of breast cancer	Other remarks
<b>a</b>								
1	Normal, p53 signature	BRCA1	40	Right	No	SCOUT in tube	No	Left ovarian theca cyst
2	Normal, p53 signature	BRCA1	39	Right and left	No	No	Yes	
3	STIL (x3)	unknown	61	Left	No	No	No	Uterus Leiomyoma
4	STIC (x3)	BRCA1	36	Right and left	No	Several p53 signature	No	HGSC in the left tube
5	Normal, p53 signature	Family history	38	Left	No	No	No	
6	STIL, STIC	BRCA1	48	Left	No	Several p53 signature	Yes	
7	Normal, p53 signature	BRCA1	35	Left	No	Several p53 signature	Yes	HGSC in the left fimbria
8	Patient 8 shows an ovarian high-grade serous carcinoma with no evidence of pre-neoplastic lesion in the fallopian tube. The BRCA status is unknown.							
<b>b</b>								
F1	STIL	BRCA2	65	Left	No	No	No	
F2	Normal, p53 signature	BRCA1	38	Right	No	No	Yes	
F3	STIC	BRCA1	37	Right	No	STIL to the left fimbria	Yes	HGSC in the left tube
F4	Normal, p53 signature	unknown	50	Left	No	No	No	Follicular cyst in the ovary
F5	Normal, p53 signature	BRCA2	50	Both sides	No	No	Lesion in the breast	
F6	STIC	BRCA1	41	Right	No	No	No	cyst in the left ovary
F7	Normal, p53 signature	No	45	Right	No	HGSC in the left tube	No	Family breast and ovarian cancer
F8	Normal, p53 signature	BRCA2	67	Left	No	No	No	
F9	Normal, p53 signature	BRCA1	70	Both sides	No	No	No	
F10	STIL, STIC	BRCA1	62	Both sides	HGSC in both ovaries	No	No	
F11	STIC	BRCA1	44	Right	No	No	Yes	
F12	Normal, p53 signature	No	68	Right	No	No	No	
F13	Normal, p53 signature	unknown	62	Right	Brenner tumor in right ovary	No	Yes	
F14	STIC	BRCA1	44	Right	No	HGSC in the right chorion	Yes	
F15	Normal, p53 signature	BRCA2	46	Left	No	No	No	
F16	Normal, p53 signature	BRCA2	50	Left	No	No	Yes	

Table 1. continued

Case ID	Lesions sampled	BRCA status	Age at diagnosis	Location in fimbria	Ovary lesion	Other lesions	Personal history of breast cancer	Other remarks
F17	STIC	BRCA1	44	Right	No	No	No	
F18	STIL	BRCA2	49	Left	No	No	No	
F19	Normal, p53 signature	BRCA2	43	Right	No	No	No	
F20	Normal, p53 signature	unknown	51	Right	No	No	Yes	Family breast and ovary cancers
F21	Normal, p53 signature	BRCA2	41	Right	No	No	No	
F22	STIL	BRCA1	58	Right	No	No	Yes	
F23	Normal, p53 signature	BRCA1	60	Left	No	HGSC in the right tube	Yes	
F24	STIC	BRCA1	42	Right	No	No	Yes	
F25	STIC, STIL	BRCA1	68	Both sides	No	No	No	
F26	STIL	BRCA1	66	Left	No	No	Yes	
F27	STIL, STIC	unknown	35	Both sides	No	HGSC in the left tube	Yes	

p53 signature lesions, 2 patients with STIL lesions, 2 patients with STIC lesions, and 1 patient with HGSC (see Table 1a). Normal tissue was analyzed from the normal part of the tissue section from patients with p53 signature lesions and no other concomitant lesions. Afterward, a retrospective cohort of 27 patients was used (see Table 1b) for Spidermass analysis.

### Sample processing

Our study analyzed FFPE tissue obtained from prophylactic adnexectomies of women with *BRCA1* mutations or family history. Our study was performed on patients presenting with pre-neoplastic lesions without any concomitant ovarian lesions. All samples were examined following SEE-FIM (Sectioning and Extensively Examining the Fimbria) protocol [28]. On these samples, tissues sections of 7 µm thickness were cut using a microtome and were deposited on a glass slide, HPS (Haematoxylin Phloxine Saffron) and immunostaining against P53 (Ventana,790–4286) and Ki-67 (Ventana, 800–2912) were performed and examined by a pathologist to find preneoplastic lesions (p53/STIL/STIC/HGSC) [13].

### MALDI-mass spectrometry analysis

To ensure reproducibility and transparency, we have provided a detailed description of the sample preparation protocol for MALDI MSI and the data analysis methodology in the Supporting Information.

### On-tissue spatially resolved proteomics guided by p53 and Ki67 IHC

The complete workflow for sample preparation is illustrated in Fig. 1A.

### Sample preparation

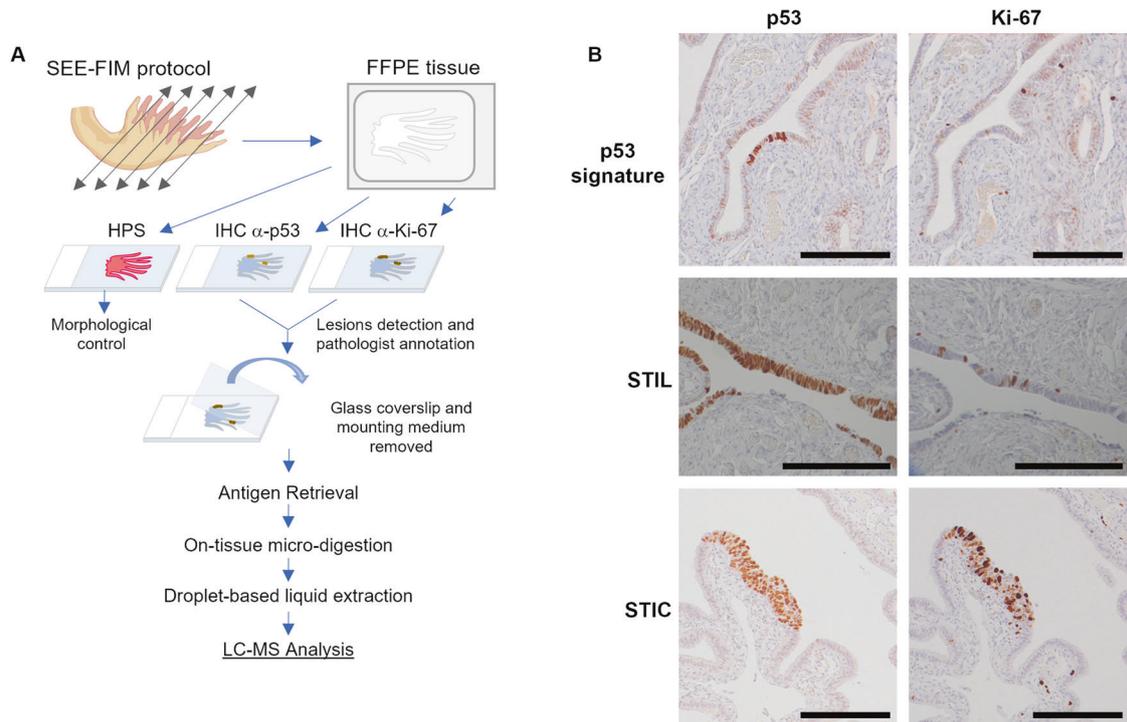
To ensure adequate statistical power and account for potential biological variability, we collected multiple samples from the same patients but on different lesions, resulting in four sampling regions per lesion (Table 1) for proteomics. This approach increases the number of replicates, thus improving the reliability of the statistical analysis. In addition, it allows for a more comprehensive assessment of the heterogeneity within and between different lesions. The same slides as the one used and annotated by the pathologist were unmounted; resin was removed by soaking them overnight in xylene and rinsing them with xylene and ethanol baths. The tissues were rehydrated using 5' each successive bath of decreasing ethanol degree (2 × 95°, 1 × 30°) and two baths of 10 mM NH<sub>4</sub>HCO<sub>3</sub> buffer. Then, an antigen retrieval step was performed to increase trypsin access to biomolecules. For this, the slides were dipped in 90 °C pH9 20 mM Tris for 30 min, rinsed twice with NH<sub>4</sub>HCO<sub>3</sub>, and dried under vacuum at room temperature.

### In situ trypsin digestion

On localized pre-neoplastic lesions, tryptic digestion was performed using a Chemical Inkjet Printer (CHIP-1000, Shimadzu, Kyoto, Japan). The region was carefully selected to ensure that the analysis was restricted to the epithelial cells marked by IHC, thus reducing the potential for contamination from other cell types. The trypsin solution (Sequencing grade modified (Promega), 40 µg/mL, 50 mM NH<sub>4</sub>HCO<sub>3</sub> buffer) was deposited on a region defined to 600 × 600 µm<sup>2</sup> during 2 h. During this time, the trypsin was changed every half-hour. With 350 cycles and 450 pL per spot, a total of 6.3 µg of trypsin was deposited. To stop digestion, 0.1% TFA was spotted during 25 cycles.

### Liquid extraction

After micro-digestion, the spot content was gathered by liquid microjunction using the TriVersa Nanomate (Advion Biosciences Inc., Ithaca, NY, USA) using Liquid Extraction and Surface Analysis



**Fig. 1 Workflow for spatially resolved proteomic using IHC tissue section.** **A** Protocol based on tissue from SEE-FIM protocol. After IHC against p53 and Ki-67, the coverslip glass and the mounting medium are removed to access the tissue section. Digestion of the lesion is performed directly on the tissue section and a droplet-based liquid extraction is performed to recover the peptides before MS-based proteomics analysis. **B** Pre-neoplastic lesions found in the fallopian tube are defined by p53 positivity and Ki-67 index. For p53 signature: accumulation of p53 in at least 12 cells without morphological abnormalities and low Ki-67 index; STIL: same accumulation of p53 in more than 20 cells with some morphological abnormalities and a higher Ki-67 proliferating index (10–40%); STIC: high p53 and Ki-67 index and cells atypical morphology (carcinoma-like).

(LESA) settings. With 3 different solvents composed of 0.1% TFA, ACN/0.1% TFA (8:2, v/v), and MeOH/0.1% TFA (7:3, v/v). A complete LESA sequence runs 2 cycles for each mixture composed of an aspiration (2  $\mu$ L), a mixing onto the tissue, and a dispensing into low-binding tubes. For each interesting spot, 2 sequences are pooled into the same vial.

### MS-based proteomic

A complete description of NanoLC-ESI-MS/MS is provided in supporting information. We implemented a randomization process to minimize the impact of batch effects in our analysis.

### Proteins identification

Raw MS/MS spectra were processed using MaxQuant (version 1.5.6.5, [29]). Peptide identifications were obtained according to a target decoy search against the Homo sapiens Uniprot database (version of 2017\_02, 20,172 entries) and a database containing 262 commonly detected contaminants. The human UniProt database was used as the forward database and a reverse one for the decoy search was automatically generated in MaxQuant. Mass tolerances were set to 10 ppm and 20 ppm respectively for parent and fragments measurement. Enzyme specificity was set to “trypsin” with a maximum of 2 missed cleavages allowed. Methionine oxidation and acetylation of protein N-terminal were set as variable modifications. FDR < 1% was set for peptides and proteins identification. Two peptides with one unique were necessary to assess protein identification. For label-free quantification, the MaxLFQ algorithm was used. The option “Match Between Runs” was enabled to maximize the number of quantification events across samples. This option allowed the quantification of high-resolution MS1 features not identified in every single measurement. Data generated by MaxQuant were analyzed using Perseus (version

1.6.2.3, [30]). LFQ values were used and proteins were removed if found in the category only identified by site modifications, in the decoy reverse database, or identified in the contaminant database.

### Proteomics statistical analyses

We removed proteins that were not presented in at least three of four replicates of each lesion. We took the average expression per group to perform a comparison and visualized using a Venn diagram. Individual LFQ values were used to perform a multi-scatter plot and calculate a Pearson Correlation between samples. Missing values were imputed based on the normal distribution (width = 0.3, down-shift = 1.8). Principal component analysis (PCA) was done to compare the protein content of each sample. An ANOVA Multi-sample test was performed and consolidated by a Permutation-based FDR (FDR < 0.05, 250 randomizations). A specific comparison between normal and p53 signature samples was performed using a student’s T-test. The differentially expressed proteins with significant differences were selected (values were z-score). The proteomic samples were then clustered according to a Euclidean average as a distance measure for column and row clustering. Up-and-down-regulated proteins in the different groups were utilized to perform an annotation analysis of gene ontology terms by using Funrich (v3.1.3, [31]). A hypergeometric test was performed against all annotated gene/protein lists by comparing the multiple datasets. Enrichment for biological process, transcription factor, and cellular component were presented as a bar chart. PANTHER Classification System (v14.1) was also used. PANTHER Overrepresentation test (downloaded on 20190701) was performed using each list of up-or down-regulated proteins as “analyzed list” and Homo sapiens as “reference list”. Fisher’s Exact test with false discovery rate (FDR) correction was also used. Subnetwork Enrichment Analysis (SNEA)

from Elsevier's PathwayStudio version 10.0 (Elsevier). was used to extract statistically significant altered biological and functional pathways in the different clusters of proteins. The immunohistochemical (IHC) data of different proteins of interest were investigated from the Human Protein Atlas (HPA) database (<http://www.proteinatlas.org>) [32]. Evaluation of the prognostic effects of these proteins on overall survival (OS) was also extracted from HPA. Comparisons are performed with the 20 genes of the highest significance associated with unfavorable prognosis for ovarian cancer, cervix cancer, endometrial cancer, and breast cancer.

### Mutation identification

The raw MS data were also processed to search for potential mutated peptides using the XMAN database [33]. This database contains information on mutations observed in cancers as well as other diseases. Proteome Discoverer 2.1 (PD 2.1) was used to query the data using MS Amanda as a search node against the XMAN database, the human Uniprot database, and a database containing potential contaminants. Peptides identified only in the XMAN database at a high level of confidence were selected, and the MSMS spectra were manually inspected to confirm the presence of the mutation.

### Alternative proteins (AltProts) identification

RAW data obtained by nanoLC-MS/MS analysis were analyzed with Proteome Discoverer V2.3 (Thermo Scientific) using the Label-Free Quantification node, the protein database is downloaded from Openprot (<https://openprot.org/>, [34]). This database included RefProt, novel isoforms, and AltProts predicted from both Ensembl and RefSeq annotations (Ensembl: GRCh38.83, RefSeq: GRCh38p7) for a total of 658,263 entries. The following processing and consensus parameters are used: Trypsin as an enzyme, 2 missed cleavages, methionine oxidation as variable modification, precursor mass tolerance: 10 ppm, and fragment mass tolerance: 0.1 Da. The validation was performed using Percolator with a protein-strict FDR set to 0.001. A consensus workflow was then applied for the statistical arrangement, using the high-confidence protein identification. Results are filtered to keep the master protein and high-confidence protein FDR. Then results are extracted in a table, to use the LFQ values in PERSEUS, where an ANOVA statistical test is performed, and the results are represented by a diagram as the heatmap described above.

### Confirmatory immunohistochemistry analyses

Early prognosis validation was performed using antibodies directed against CAVIN2 (Rockland, 600-401-J31), Emilin2 (Thermo Scientific, pa584657), CAVIN1 (Abnova, H00284119-M02), SPTAN1 (OriGene, TA812019), and FBLN5 (Invitrogen, MA5-42544). After dewaxing and antigen retrieval with citrate buffer, the tissues were incubated with a primary antibody at 4 °C overnight, followed by the application of a secondary antibody (Alexa fluor conjugated antibody, 1/200 dilution) for 1 h at RT. We used the following primary antibodies and dilutions: Cavin-1 (1/300), Cavin-2 (1/100), SPTAN1 (1/2000), Emilin-2 (1/500), Fibulin-5 (1/500). All slides were imaged on the Zeiss LSM700 confocal microscope. Three pictures were taken for each tumor section.

### Lipidomic analyses through Spidermass technology

A complete description of Spidermass is provided in the Supporting information.

### SpiderMass analysis of FFPE tissues

Tissues were sectioned at 7 μm thickness and deposited onto a polylysine glass slide. Then, the FFPE tissue sections were submitted to dewaxing two times in xylene for 5 min and were manually sprayed with a glycerol/isopropyl alcohol (IPA) (8:2, v/v) solution in 2 successive passes using a manual sprayer (Agilent).

The syringe pump (74 900 series Cole Parmer Instrument Company) was set to a 700 μL/min flow rate. The 2 successive passes were equal to 5 μL deposited on 1 cm<sup>2</sup> and took ~10 s. The samples were analyzed within 10 min after the glycerol deposition [52].

### Classification model from SpiderMass data

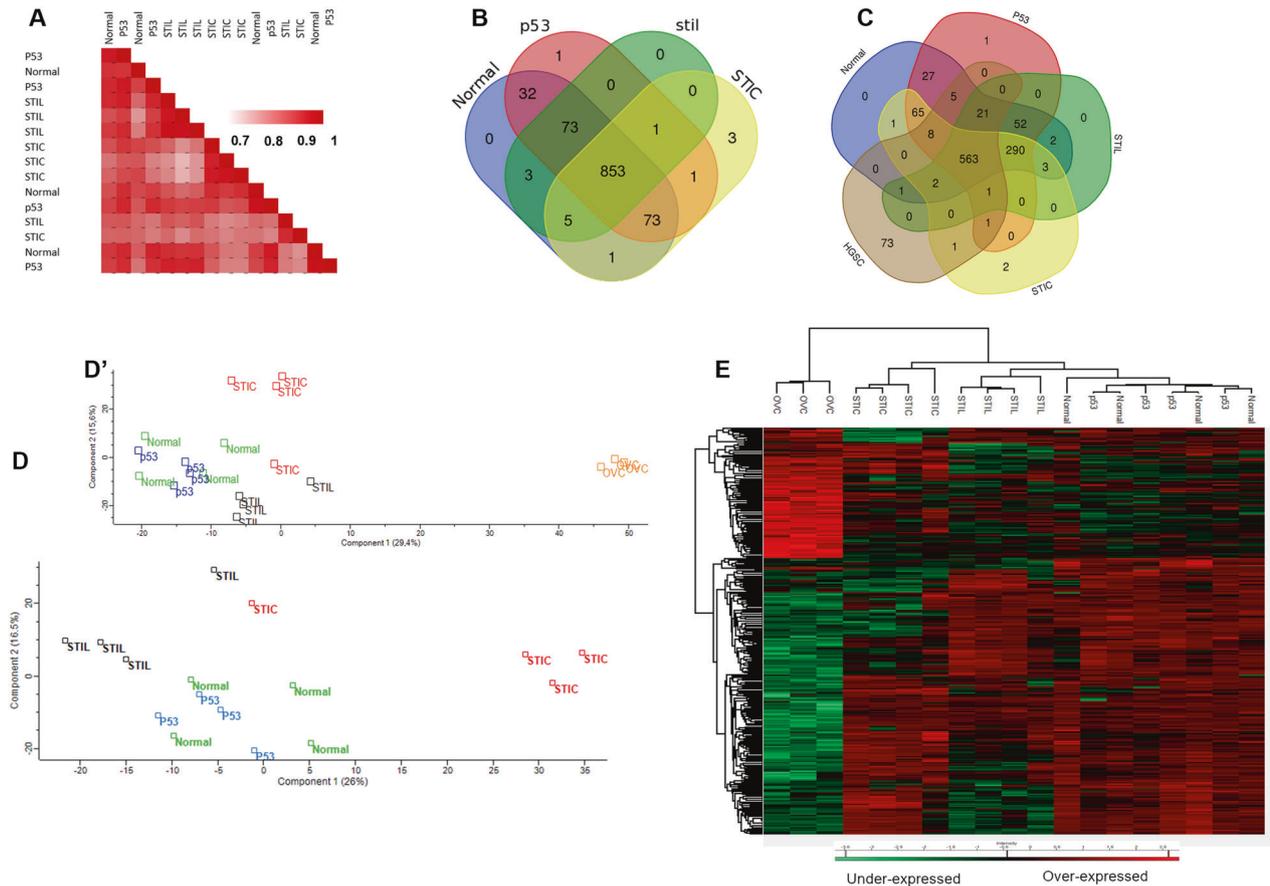
For data analysis, all raw data files produced with the SpiderMass instrument were imported into the Abstract Model Builder (AMX v.0.9.2092.0) software. After importation, spectra were first pre-processed. The pre-processing steps include background subtraction, total ion count normalization, lock mass correction, and re-binning to a 0.1 Da window. All processed MS spectra obtained from the 27 histologically validated samples were then used to build a principal component analysis and linear discriminant analysis (PCA-LDA) classification model [51]. The first step consisted of PCA to reduce data multidimensionality by generating features that explain most of the variance observed. These features were then subjected to supervised analysis using LDA by setting the classes that the model will be based upon. LDA attempts to classify the sample spectra and assess the model by cross-validation. Cross-validation was carried out by either using the "20% out" or the "leave one patient out" methods. For the first method, 20% of MS spectra are randomly taken from the total spectra, and the model is reconstructed from the remaining 80%. The remaining 20% of spectra are used to interrogate the reconstructed model. The permutation is automatically reiterated for 5 cycles before reporting the cross-validation results. For the second method, the spectra are grouped by patients and left out one by one; at each step, the model without the patient is interrogated against this model.

## RESULTS

### In situ spatially resolved proteomics of FTE lesions

The primary goal of the clinical trial investigated the effectiveness of bilateral laparoscopic radical fimbriectomy in eliminating the potential source of dysplastic cells in the fallopian tubes while maintaining ovarian hormonal function [35]. A second part aimed to identify a potential proteomic timeline associated with ovarian cancer progression and development. Given the scarcity of the preneoplastic lesion, only a limited number of patients were recruited for the clinical trial. To optimize the use of tissue samples, we decided to reuse the previously immunostained FFPE tissue sections using highly refined mass spectrometry-based methods (Fig. 1A). A cohort of 8 patients presenting *BRCA 1* mutations or hereditary susceptibility, was selected for the spatially resolved proteomic experiments (Table 1a). These experiments aimed to target the preneoplastic lesions, which are only present in very few cells already under transformation in small areas of the tissues. To confirm the presence of these lesions, a double IHC against p53 and Ki-67 markers is needed (Fig. 1B). To enable spatially resolved proteomics, we developed a new protocol that uses the same IHC slide as the pathologist, which had previously served for the detection of the lesions (Fig. 1A). Targeting these specific lesions allows for precise localization of the affected cells and decreases the presence of proteins for normal tissue in the samples.

To investigate the possibility of identifying specific protein signals from different lesions, we conducted MALDI mass spectrometry imaging (MALDI-MSI) experiments. Our findings, as shown in Supplementary Fig. S1, revealed the presence of specific ions corresponding to proteins that were specific to the STIC lesion, such as *m/z* 3019.19 and *m/z* 2639.76, which were only found in the tumor area. To highlight the main molecular regions of the tissue, we performed a spatial segmentation, which revealed a statistically significant difference between the tumor region and the non-tumoral region of the tissue. Specifically, we



**Fig. 2 Comparison of the different lesions based on their proteomic signature.** **A** Correlation analysis based on the whole proteome. Ven diagrams showing the distribution of identified proteins among **(B)** normal tissue and the different lesions and **(C)** with the addition of HGSC data. Principal component analysis using **(D)** normal tissue and the pre-neoplastic lesions and **(D')** with addition to HGSC. **E** Hierarchical clustering of the most variable proteins (ANOVA with permutation-based FDR < 0.05) including HGSC data.

identified a specific cluster (light blue) in the tumor region (Supplementary Fig. S1B). Moreover, our segmentation of MALDI-MSI data from two tissues, one with no lesion and one with a p53 signature, also showed the presence of a specific cluster (orange) with a different molecular profile within the area where the lesion was identified by IHC (Supplementary Fig. S1C). These findings suggest that specific protein signals can be identified from different lesions, providing a foundation for spatial proteomics analysis.

Then, the IHC slides stained for p53 and Ki67 were directly used to perform the spatial shot gun proteomic experiments from all the patients and tissue sections. After marking the area and unmounting the coverslip, the tissues were submitted once more to AR followed by a micro-digestion with trypsin and a micro-extraction (see picture, Fig. 1A and Supplementary Fig. S1D). Our workflow identified 10,375 unique peptide sequences corresponding to 1,617 distinct protein groups across pre-neoplastic lesions and non-pathological tissues. Label-free quantification (LFQ) of proteins resulted in a total of 1571 proteins quantified (Fig. 1C). After filtering the proteins according to a minimum number of values (2/3 of valid values) in at least one group of the four defined groups (normal, p53 signature, STIL, and STIC), we retained 1,046 proteins for further analysis. The number of identified proteins is consistent with our previous experience on FFPE tissue, considering that the tissues were stored over many years after fixation [36]. Notably, the spatial proteomic analysis was unaffected by the removal of the coverslip glass and mounting medium from the IHC sections evaluated by the pathologist. To validate these findings, we conducted a correlation

analysis of all quantified proteins. The protein patterns of samples from the same group compared with inter-group variation were similar (mean Pearson correlation 0.92). The main differences were observed between STIC and STIL lesions (mean Pearson correlation 0.84) (Fig. 2A). Although our sample size was limited, these findings are consistent with the inter-patient variability typically observed, indicating that our protocol did not significantly alter the proteomic content of the tissue sections. This allows us to confidently perform a comparative analysis to obtain a specific proteomic signature of the different lesions.

### Comparison of the proteomic profiles of p53 signature, STIL, STIC, and normal fimbria

The protein content of each group was compared by averaging the replicates and using a Venn diagram representation. It was observed that 853 common proteins (81.5% of overlap) were presented between normal tissues and pre-neoplastic lesions (Fig. 2B, Supplementary Data 1). Three proteins were found to be exclusive to the STIC lesions and one protein was identified only in the p53 signature lesion. Among these, RNA-binding protein 10 (RBM10) was found in the p53 signature lesions and not in the normal tissue or other lesions. RBM10 is an RNA-binding protein located in the nucleus that plays a crucial role in RNA splicing and is implicated in several human diseases, including cancer. Three proteins were found exclusively in the STIC lesions: Glucosamine (N-acetyl)-6-sulfatase (GNS), Upstream binding transcription factor, RNA polymerase I (UBTF), and ATPase family, AAA domain containing 3 (ATAD3A/B). According to data from the Human Protein Atlas (HPA), RBM10, which was found exclusively in the

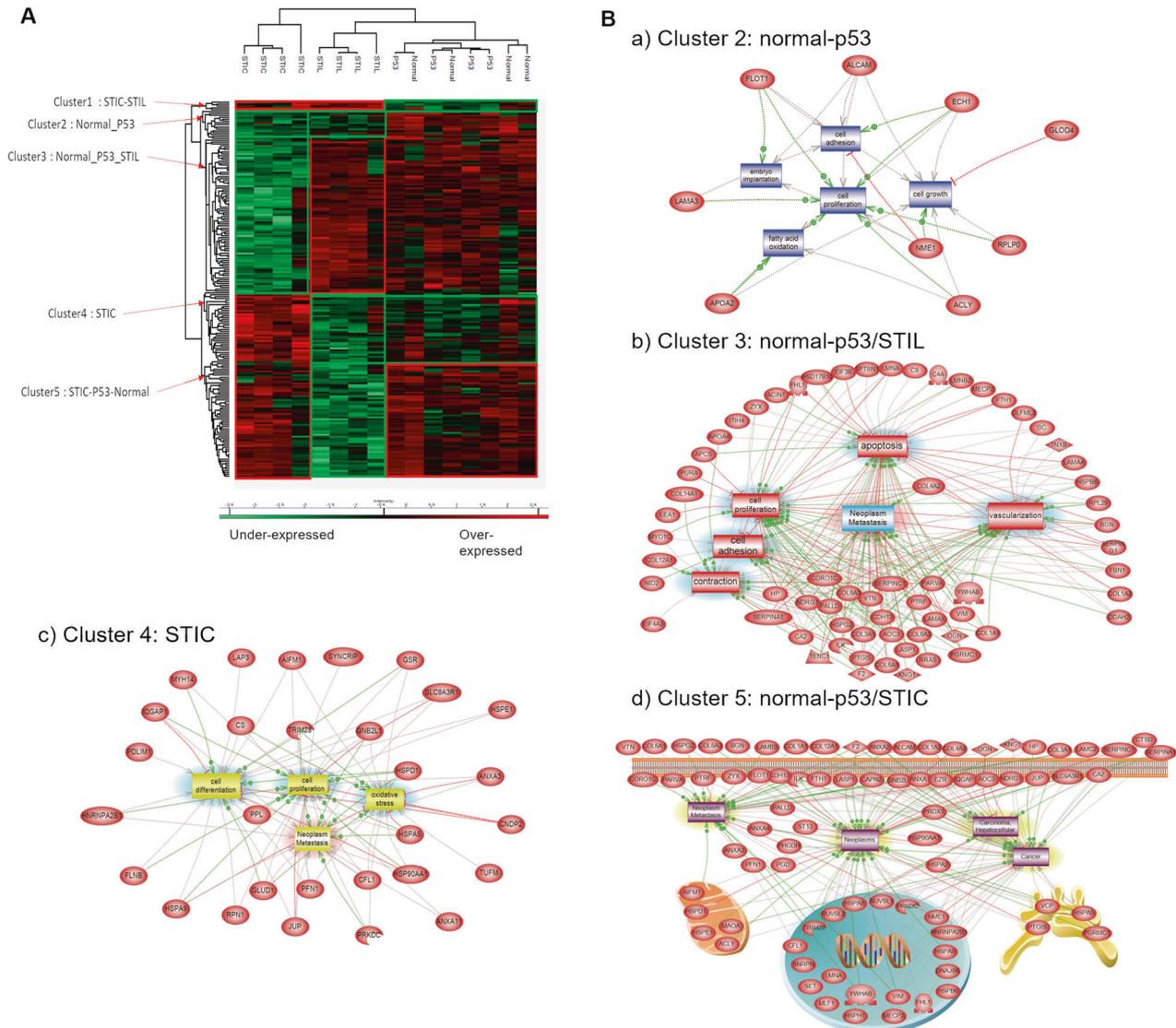
p53 signature lesion, is overexpressed in cervical and endometrial cancers and has a favorable prognostic value. On the other hand, the elevated expression of GNS and ATAD3A/B is associated with poor survival rates in liver cancer, while high expression of GNS in glioma and ovarian cancer is a marker of poor prognosis. However, the expression of UBTF appears to have no impact on survival. Regarding the HPA data on the identified proteins, it should be noted that while these proteins are indicated as highly expressed in the fallopian tube, this expression is primarily detected in the glandular cells. These cells are found in higher proportions near the uterus, whereas cilia cells are more represented in the fimbria part. Therefore, the expression levels of these proteins may differ depending on the specific region of the fallopian tube being analyzed. The inclusion of HGSC data in the analysis (Fig. 2C, Supplementary data 2) revealed that UBTF was expressed both in STIC lesions and HGSC. Fibrillarin (FBL) is found only in pre-cancerous lesions and not in normal tissue. This protein is also observed in HGSC (Fig. 2C, Supplementary Data 2).

There were a limited number of proteins exclusively identified in each group, indicating that the variation is primarily in protein abundance rather than specific protein composition. To better understand the modulation registered across the different lesions, we performed principal component analysis (PCA) (Fig. 2D). Components 1 and 2 represent 42.5% of the total data variation. Samples from normal tissue and p53 signature are difficult to discriminate. However, samples from STIL and STIC lesions were different from normal and p53 signature samples in component 2. Interestingly, STIL lesions were closer to normal and p53signature lesions in component 1, compared to STIC samples. We further confirmed that the proteome of ovarian HGSC tumors is highly different compared to STIL, STIC, p53 signature, and normal tissues (Fig. 2D'). Then, we performed hierarchical clustering to analyze the significant differences in protein expression between the five groups (Fig. 2E). The results showed that HGSC tumors were separated from the other tissues by one main branch. The second branch is further divided into two sub-branches with one separating STIC lesions from p53 signature/Normal/STIL lesions, and the other separating STIL lesions from p53 signature/Normal tissues (Fig. 2E). As previously observed, the last group (p53/Normal) was challenging to differentiate. Nevertheless, the molecular differences in cell composition between the ovary and the fallopian tube did not allow us to draw a direct conclusion on the relationship between HGSC and the pre-neoplastic lesions of the fallopian tube. However, we demonstrated that the proteomic profile of the STIC lesions was closer to HGSC compared to the other lesions, and each lesion had a specific proteomic signature.

### Descriptive proteomics analysis of preneoplastic lesions

For a deeper analysis of the pre-neoplastic lesions of the FTE, ANOVA testing was performed without the HGSC data. A total of 197 proteins were differentially expressed (significantly different) within the four groups (Supplementary Data 3). A hierarchical clustering based on the expression of these proteins resulted in two main clusters (Fig. 3A, Supplementary Data 3). As observed previously, normal and p53 signatures were not resolved. This group as well as STIL lesions clustered together (Fig. 3A). The proteomic profile of STIC lesions significantly differed from the two other groups. The heatmap displays five main clusters of proteins. Proteins that were enriched in STIL and STIC regions were represented in **cluster 1**. Those that were more abundant in both normal tissue and p53 signature lesions were represented in **cluster 2**. Proteins overexpressed in normal, p53 signature lesion and STIL but not in the STIC were represented in **cluster 3**. Proteins that were overexpressed in the STIC lesions were represented in **cluster 4** and those specific to normal tissue, p53 signature lesion, and STIC were represented in **cluster 5**. **Cluster 1** (STIL/STIC) was composed of only 5 proteins i.e.,

MARCKS-related protein (MARCKSL1), Myosin-10 (MYH10), Protein SET (SET), Double-stranded RNA-binding stauferin homolog-1 (STAU1), and D-3-phosphoglycerate dehydrogenase (PHGDH). The Protein SET participates in numerous cellular functions including DNA repair, transcription, cell survival, and proliferation. This protein is also involved in many cancer processes such as metastasis, the development of therapeutic drug resistance, and play a key role in tumorigenesis [37]. MARCKSL1 is known to promote the progression of lung adenocarcinoma by regulating epithelial-to-mesenchymal transition (EMT) [38]. STAU1 has been previously identified in colorectal cancer [39], whereas PHGDH protein has been identified in ovarian cancer [40]. In **cluster 2** (normal-p53), 15 proteins showed an overexpression for normal and p53 signature lesions. Among these proteins, CD166 antigen (ALCAM/CD166) is shown to be a potential cancer stem cell marker [41]. It has been found as a poor prognostic marker in pancreatic cancer [42]. This protein is known to contribute to local invasion and tumor progression by acting on the detachment of tumor cells [43]. Moreover, TNPO1, ACLY, NME1, FLOT1, and KLC4 are proteins already known to be involved in epithelial ovarian cancer [44, 45]. Gene Ontology (GO) analyses confirmed that these proteins are involved in cell proliferation, adhesion, and growth (Fig. 3Ba). **Cluster 3** (normal-p53/STIL) contains 80 proteins. Proteins involved in complement and coagulation cascades are present in this cluster (SERPIN A1, SERPIN C1, SERPIN G1, C4a, C9, Vitamin D-binding protein (GC), pro-thrombin (F2), Vitronectin (VTN), kininogen (KNG1)) and 10 members of the collagen family. The PGRMC1 (Membrane-associated progesterone receptor component 1) is known to be involved in ovarian cancer [46] as well as PTGIS (Prostacyclin synthase) and the Protein NDRG1 known to modulate genes involved in ovarian cancer metastasis [47, 48]. GO analysis confirmed that these proteins are involved in neoplasia/metastasis, angiogenesis, and apoptosis (Fig. 3Bb). **Cluster 4** (STIC only) contained 36 proteins involved in cell migration such as CFL1 (Cofilin-1), PFN1 (profilin-1), PPL (periplakin), FLNB (filamin B), PDLIM1 (PDZ and LIM domain protein 1), IQGAP1 (Ras GTPase-activating-like protein), MYH14 (Myosin-14) and in metabolism (CS (Citrate synthase), DECR1 (2,4-dienoyl-CoA reductase), GSR (Glutathione reductase), LAP3 (Cytosol aminopeptidase), GLUD1 (Glutamate dehydrogenase 1). Proteins of this cluster were involved in cell differentiation, proliferation, and in neoplasia/metastasis (Fig. 3Bc). The last cluster (normal-p53/STIC) was composed of 61 proteins. Several proteins involved in exosomes have been identified (Annexin A1, A2, A4, A13, EZR (ezrin), HSPA8 (Heat shock 71 kDa protein 1 A)). Other proteins are involved in stress such as PRDX1 (Peroxiredoxin-1), HSP8, HSPA1 (Heat shock 70 kDa protein 1), HSPA2 (Heat shock 70 kDa protein 2), ST13 (Hsc70-interacting protein), DNAJB6 (DnaJ homolog subfamily B member 6), VCP (Transitional endoplasmic reticulum ATPase), TUBB4B (Tubulin beta-4B chain). This cluster showed an over-representation of proteins involved in neoplasia and cancer transitions (Fig. 3Bd). Functional enrichment analysis points out differences between clusters. Different biological processes have been identified i.e. immune response, regulation of immune response, cell growth and/or maintenance, metabolism, energy pathways, signal transduction and protein metabolism (Fig. 4A). Enriched transcription factors (TFs) that regulate the over-expressed proteins in the different clusters were also obtained (Fig. 4B). The cluster containing proteins overexpressed in normal-p53 signature showed a high abundance of proteins involved in metabolism and energy pathways as well as an enrichment of SP1, SP4 and TEAD1 transcription factors that targeted the most genes of this cluster. For the cluster normal-p53/STIL, proteins were mostly involved in cell growth and/or maintenance and protein metabolism. Enrichment of the TFs NFIC and EGR1 was observed. Concerning Normal-p53/STIC, identified proteins were implicated in cell growth and/or maintenance, metabolism, energy pathways and protein metabolism, while KLF7 protein was identified as a TF.

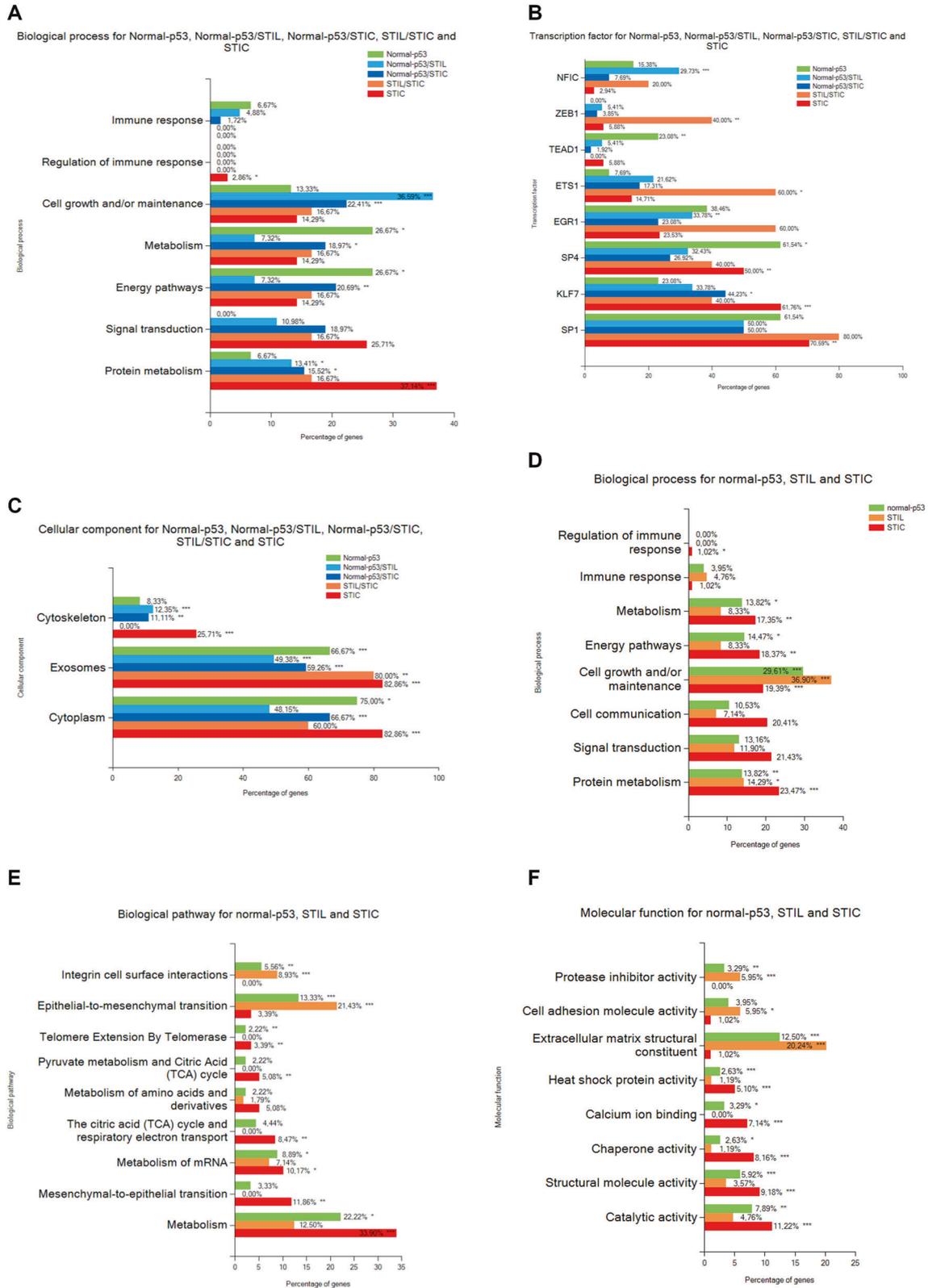


**Fig. 3** Proteomic analysis of the pre-neoplastic lesions. **A** Hierarchical clustering of the most variable proteins between normal tissue, p53 signature, STIL and STIC ( $n = 4$  for each category, ANOVA with permutation-based FDR < 0.05); **B** Subnetwork Enrichment Analysis was done to highlight altered biological and functional pathways in the different clusters of proteins.

For STIL/STIC cluster, enrichment of the TFs ZEB1 and ETS1 was observed. STIC cluster presented a strong enrichment of proteins involved in protein metabolism and regulation of immune response and SP1, SP4 and KLF7 as TFs. We also observed a reduction of the proteins involved in immune response in STIC lesions. Cellular components analysis reflected that all proteins of the 5 clusters were in the cytoplasm or in exosomes (Fig. 4C) but with some differences. In fact, in STIL-STIC and STIC, we observed a strong enrichment of exosomes, whereas p53 signature and STIC involved cytoplasmic proteins. We also observed a high enrichment of cytoskeletal proteins in STIC lesions. Altogether, these analyses showed molecular transitions occurring between Normal-p53 signature, p53 to STIL and STIL to STIC with a change in proteins involved in cell growth and/or maintenance and the different metabolism processes. To confirm these observations, further analysis were carried out, this time grouping together all proteins overexpressed in each lesion, i.e., normal-p53 signature, STIL, and STIC (Fig. 4D, E, F and Supplementary Data 3). In STIL, an enrichment of proteins involved in cell growth and/or maintenance (Fig. 4D), in integrin cell surface interactions (Fig. 4E), in extracellular matrix (ECM) structural constituent and cell adhesion

molecule activity (Fig. 4F) were observed. The ones in various metabolism and energy pathways (Fig. 4D, E) and catalytic activity (Fig. 4F) were lower than for the other cellular transition. We also investigated whether the proteins involved in the Warburg effect [49] were activated to compensate for the overall decrease in the classical metabolism processes observed (Supplementary Fig. S2). These proteins were not altered overall and a trend toward decreasing suggest that no Warburg effect seemed to occur. In STIC, proteins involved in metabolism, energy pathways, cell communication, signal transduction (Fig. 4D, E), calcium ion binding, catalytic and structural molecule activities (Fig. 4F) were higher than in the other transition lesion. In the same time, immune response (Fig. 4D), integrin interactions (Fig. 4E), cell adhesion molecule activity and ECM structural constituent (Fig. 4F) were highly repressed. A statistical overrepresentation test using PANTHER classification system analyses (Supplementary Data 4) confirmed such transition.

Regarding the p53 signature and normal tissue, which could not be distinguished using group analysis, but using a Student's *T* test, 8 proteins were found to be highly discriminant between Normal and p53 lesions with a specific over-expression in p53 signature



**Fig. 4 Annotation analysis of gene ontology terms. A Biological process, B Transcription factor and C Cellular component for the clusters of proteins. D Biological process, E Biological pathway and F Molecular function of the proteins overexpressed in each lesion. (hypergeometric test against all annotated gene/protein list of Funrich database, the  $p$ -value is represented by stars: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$  no star for  $p > 0.05$ ).**



**Table 2.** Immunoreactivity and prognostic effect of proteins overexpressed in p53 signature compared to normal tissue according to Protein Atlas.

Protein name	Immunoreactivity in FTE	Prognostic effect in OVC
CAVIN1	Not detected	Unfavorable***
SPTAN1	Moderate to strong positivity	Unfavorable*
FBLN5	Not detected	Unfavorable*
EIF3B	Strong positivity	Not significant
MOB1B	Moderate positivity	Not significant
EMILIN2	Moderate positivity	Not significant
SPTBN1	Moderate to strong positivity	Not significant
CAVIN2	Not detected	Not significant

Immunohistochemical staining of proteins in normal FTE tissue and prognostic effect of high expression of the corresponding gene in OVC were extracted from Human protein atlas (data credit: Human Protein Atlas available from v19.proteinatlas.org).

*p*-value is represented by stars \*\*\**p* < 0.001; \**p* < 0.05 and no star for *p* > 0.05.

(Fig. 5A, B, Supplementary Data 5). Table 2 presents the immunoreactivity in normal FTE and the prognostic effect on overall survival (OS) in OC for these 8 proteins extracted from the HPA. PTRF, also named CAVIN1, is an unfavorable prognosis marker for ovarian, urothelial, and colorectal cancers, whereas SDRP (known as CAVIN2) is associated with a favorable prognosis in renal cancer and a poor prognosis in stomach cancer. The IHC data validation indicated that these proteins are not detected in normal FTE. It also showed a moderate positivity in OC for CAVIN1 and a strong positivity in a rare case of endometrioid carcinoma of the ovary for CAVIN2. SPTAN1 and FBLN5 are favorable prognostic markers in renal cancer but unfavorable in OC (considering a *p* < 0.05). Emilin2 is already known as a poor prognostic marker for liver cancer and a poor prognostic marker for renal and head-neck cancers. By IHC, we validated a higher expression of CAVIN1 in p53 and STIL lesions compared to normal FTE. FBLN-5 and Emilin-2 were found higher expressed in p53 and to a lesser extent in STIL and were not detected in STIC lesions compared to normal FTE while SPTAN1 is higher expressed in some cells of the P53 and STIL lesions. CAVIN2 is absent of STIC lesions. (Fig. 5C, Supplementary Data 6).

### Protein mutation

The linear relationship between the different preneoplastic lesions and HGSC has been suggested by the presence of genomic changes and mutations [50]. To identify specific modifications that may have occurred across the spectrum of the pre-cancerous lesions, we explored the presence of protein mutations in each lesion. For that purpose, we used the human database combined with the XMAN database [33]. This database contains information concerning mutated peptides that could be found in some cancers extracted from the COSMIC database. This database integration resulted in the identification of 83 peptide sequences containing possible mutations (Table 3, Supplementary Data 7). For four of them, it was possible to determine the amino acid modification directly via the tandem mass spectra (Fig. 6A). These peptides were derived from four proteins i.e., the vitamin D binding protein (GC), the polyubiquitin-C (UBC), the Histone H2B (H2B1C), and Histone H3.1 (H31). The mutations found in the protein GC (1296T>G p.D432E), lead to the sequence LPEATPTE-LAK in the protein and have been identified in stomach cancer studies according to the COSMIC database. It was observed in data

**Table 3.** Mutation detected in proteins from proteomic data obtained by spatially resolved proteomic using XMAN database.

Found in samples	Patients <sup>a</sup> Number	Protein	Mutation	Cancers
p53 signature	2	Histone H4	c.239A>Gp.K80R AKRRVT Missense	Endometrium
	2	Histone H2B type 1-C/E/F/G/I	c.181G>T p.G61C KAMC MN Missense	Endometrium
STIL	3	60S ribosomal protein L14	p.A159_K160insAA AAAAAKVP Insertion - in frame	Endometrium, Ovary, Breast
STIC	2	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	c.767C>A p.P256H GPLHVPF Missense	Large intestine
Normal-p53 signature	2/2	Hemoglobin subunit delta	c.427G>T p.A143S NALSHKY Missense	Lung
	2/2	Histone H2B type 1-C/E/F/G/I	c.133G>C p.V45L VYKLLKQ Missense	Breast
	2/2/1 HGSC	Actin, alpha cardiac muscle 1	c.166G>A p.V56I DSYIGDE Missense	Endometrium
	1/3	Synaptic vesicle membrane protein VAT-1 homolog	c.278T>G p.L93R FADRMAR Missense	Lung
	3/3	Isoform 2 of Drebrin-like protein	c.746G>A p.R249Q SRNQNEQ Missense	Endometrium
	2/3/1 STIL	Isoform 2 of Synaptopodin-2	c.3392C>T p.P1131L RPPLWEA Missense	Skin
	2/2/1 STIL	polyubiquitin c	c.368G>C p.G123A FAAKQL Missense	Breast
p53signature/STIL	2/3	Isoform 2 of Tropomyosin beta chain	c.835C>A p.L279M QTLMELN Missense	Autonomic ganglia
Normal-p53/STIL	2/3/2/1 STIC	Histone H1.4	c.262G>T p.V88L KSLLSKG Missense	Esophagus
Normal-p53/STIL/STIC	1/3/3/2/1 HGSC	Vimentin	c.1313C>G p.S438 <sup>a</sup>  DTH Nonsense	Endometrium
	4/4/3/2	Laminin subunit gamma-1	c.2033C>T p.P678L SARLPGF Missense	Breast
	4/4/4/4; HGSC	Histone H3.1	c.368A>G p.K123R IMPRDQ Missense	Endometrium

<sup>a</sup>Only mutations found in more than two patients.



untranslated regions of the mRNA (3' and 5'UTR), from a shift in the open reading frame, or non-coding RNA (ncRNA). Using our raw proteomic data against the OpenProt database, 59 AltProts have been identified. More than 64% of these proteins are associated with sequences from non-coding RNA (ncRNA) and from the mRNA coding for RefProt. 23% are from 5'UTR, 33% from the 3'UTR, and 42% from a shift in the CDS (Supp. Data 8). Among these proteins (Table 4, Supplementary. Data 8), we identified IP\_270787 (AltZNF709) in normal tissues. IP\_2326985 is issued from an ncRNA (LOC105376003) and is only found in STIC. The AltProt IP\_232994 (AltMYEF2), coded by the 3'UTR part of the transcript mRNA referenced to translate the MYEF2, is identified only in HGSC samples. IP\_572777 (from the ncRNA: SDR16C6P), IP\_774783 (from the ncRNA: RP11-358N4.6) and IP\_285024 (from the ncRNA: HNF4A) are specific to Normal, p53 signature and STIL samples. In both Normal/p53/STIL and STIC samples, 17 AltProts were identified (Table 4). Using the ANOVA test, 12 AltProts displayed significant variations in the different lesion stages. Nine AltProts are derived from ncRNA, one from 3'UTR, and two from a shift in the CDS. The hierarchical clustering and heatmap representation (Fig. 6Ba) revealed a clear separation between STIC and the other lesions. The second branch separates STIL from the normal-p53 signature. The pseudogene KRT8P11 (IP\_563986) is overexpressed in STIC (Fig. 6Ba), whereas COL18A1 (IP\_290397) pseudogene is overexpressed mainly in the STIL samples and slightly in P53 and is underexpressed in carcinoma whereas LOC105376924 is present in STIL and in carcinoma and absent in P53 samples (Fig. 6Bb). The other AltProts are differentially expressed in p53 lesions and Normal tissue but are significantly down-regulated in the STIC or non-expressed in STIL (Fig. 6Ba). When data from HGSC were added, the separation of STIC from other samples was observed, carcinoma samples composed the second branch, and a third branch is divided into subbranches. These subdivisions group included three of the fourth STIL samples and three of the fourth Normal samples (Fig. 6Bb). Based on our findings, it is proposed that KRT8P11 (IP\_563986) is a specific marker of STIC, B4GALN (IP\_191334) is specific of HGSC, and KRT8P32 (IP\_602534) is overexpressed in two p53 signature samples and one of the normal tissue samples but also highly decreased in HGSC. Ultimately, COL18A1 (IP\_290397) is specific to STIL. The other AltProts were observed between Normal-p53 signatures with most of them significantly down-regulated in STIC (Fig. 6Bb).

### SpiderMass analysis

To validate the classification obtained by MALDI-MSI, we analyzed samples with SpiderMass technology. Following the acquisition of the MS spectra in positive ion mode, a PCA analysis of the generated spectra acquired from healthy, p53 signature, STIL and STIC tissues was performed. The PCA features were subjected to a supervised analysis using linear discriminant analysis (LDA) [51] which resulted in 4 different groups (Fig. 7A). According to Fig. 7A, LDA 1 discriminated the Healthy group from the p53 and STIC groups. However, p53 signature are less separated from STIL than from Healthy. This was also observed when we compared spectra. The p53 signature and STIL spectra are much closer than STIC ones (Fig. 7B). Cross-validation results obtained from the p53 signature, Healthy and STIC groups using the "20% out method" shows excellent classification rates with 95.08% and 85.29% without and with outliers respectively. An excellent classification between P53 and Healthy with 100 and 80% including or not outliers respectively showed significant discrimination of both classes taken from the same tissue section. A correct separation of 75% with and 72.93% without outliers between P53 and STIL. Finally, a better classification rate of 92.5% and 88.1% with and without outliers was performed between P53 and STIC (Fig. 7C).

Some examples of discriminant ions ( $m/z$ ) between the four groups corresponding to lipids, are presented with their normalized intensities in (Fig. 7D) and were identified by MS/MS. Thanks to the generated boxplot we were able to find lesions' lipidic markers such as TG (16:0/18:1/20:4) (881.7  $m/z$ ) or PE (P-18:0/22:4) (780.5  $m/z$ ) that are underexpressed in healthy tissues. In addition to these 2 lipids, the overexpression of PC(P-16:0/22:6) (790.5  $m/z$ ) and DG (37:6) (627.5  $m/z$ ) in STICs enables the differentiation between healthy and lesion phenotypes from the cancerous one. Interestingly, we were able to discriminate p53 lesion from healthy tissue by PS (22:1/20:1) (874.9  $m/z$ ) and PS (22:0/20:0) (876.9  $m/z$ ) which are underrepresented in a p53 lesion. Finally, we were able to distinguish STILs from STICs by the PG(O-16:0/20:1) (763.5  $m/z$ ). In fact, this lipid is underrepresented in STILs compared with other phenotypes.

### DISCUSSION

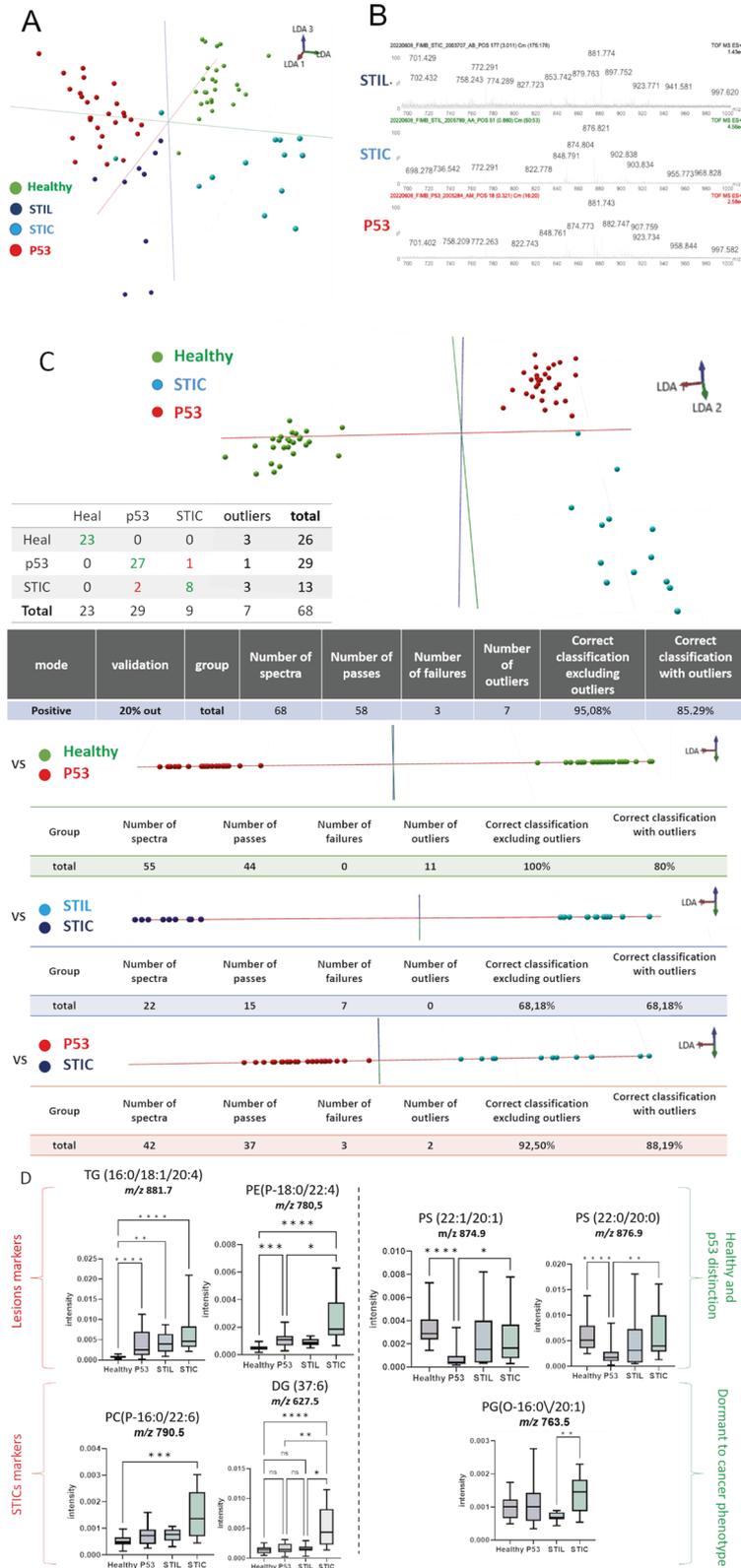
High-grade serous ovarian carcinoma (HGSC) may originate at the tubal-peritoneal junction where the tubal fimbriae's specialized epithelium meets the peritoneum covering the serosal surface of the fallopian tube, according to recent theories [20, 52]. This region represents a cancer hotspot due to the transition between different types of epithelia, and there is evidence supporting the presence of cancer-prone stem cells in this area. STIC (serous tubal intraepithelial carcinoma) has emerged as a precursor lesion recognized by its preferential localization in the fimbriae end of the fallopian tube, which has led to the recommendation of standardized and detailed sampling of tubal specimens using the SEE-FIM protocol. However, diagnosing STIC based on morphology alone has low reproducibility. The histological examination of fallopian tubes has revealed a variety of pre-neoplastic lesions, including p53 signature and STIL (serous tubal intraepithelial lesion) lesions, which have not been previously characterized. The temporal relationship between these lesions and HGSCs is still debated.

It should be noted that pre-neoplastic lesions of the FTE are not commonly used for cancer detection or diagnosis in clinical practice, as they are not easily detectable. Although the goal of our study was not to identify diagnostic biomarkers due to the limited number of patients, we aimed to provide a comprehensive description of the proteomic content of pre-neoplastic lesions in BRCA1 mutation carriers. This description can help to better understand the transition between different lesion types and their potential link to ovarian cancer. Future studies with larger cohorts are necessary to identify potential biomarkers for early detection of these lesions. Our study provides the first in-depth proteomics and system biology analysis of epithelial cells in the fimbria end of the fallopian tube to better understand the different pre-neoplastic lesions. The robust and unique protocol established for tissue section sampling provides a strong foundation for future research in this area. We developed a novel strategy based on pathology routine protocol to investigate pre-cancerous lesions. Obtaining several tissue sections containing these lesions is challenging due to their limited cell numbers. The SEE-FIM protocol allows for their localization and differentiation but on a limited number of slides. To better understand the transitions of the lesions, we performed IHC-guided spatially resolved proteomic analyses on each identified lesion. Our workflow is compatible with proteomic analysis, allowing for targeted digestion of each region highlighted by IHC. The tissue section surface recovery protocol does not alter the proteomic content and enables an accurate comparison of quantified proteins between each lesion. Our protocol has the advantage of reusing the diagnostic IHC slides, which provide the localization and categorization of each lesion. It allows for the in-depth study of pre-cancerous lesions using a limited number of tissue sections.

**Table 4.** Summary of the repartition to the Venn diagram result of the 45 AltProts identified.

Altprot	Normal	P53	STIL	STIC	Carcinoma
IP_270787	X				
IP_572777	X	X	X		
IP_774783	X	X	X		
IP_285024	X	X	X		
IP_669028	X	X	X	X	
IP_563986	X	X	X	X	
IP_621829	X	X	X	X	
IP_595290	X	X	X	X	
IP_730500	X	X	X	X	
IP_2322038	X	X	X	X	
IP_763981	X	X	X	X	
IP_708166	X	X	X	X	
IP_737172	X	X	X	X	
IP_290397	X	X	X	X	
IP_652773	X	X	X	X	
IP_689518	X	X	X	X	
IP_721145	X	X	X	X	
IP_060976	X	X	X	X	
IP_632428	X	X	X	X	
IP_274301	X	X	X	X	
IP_138273	X	X	X	X	
IP_716829	X		X	X	
IP_2326985				X	
IP_591310		X	X	X	X
IP_2285716	X		X	X	X
IP_572435	X	X		X	X
IP_2323408	X			X	X
IP_079312		X	X		X
IP_667690		X	X		X
IP_076499		X	X		X
IP_601870	X		X		X
IP_755940	X	X			X
IP_232994					X
IP_602534	X	X	X	X	X
IP_667001	X	X	X	X	X
IP_755869	X	X	X	X	X
IP_658355	X	X	X	X	X
IP_724315	X	X	X	X	X
IP_591792	X	X	X	X	X
IP_631742	X	X	X	X	X
IP_688853	X	X	X	X	X
IP_2371754	X	X	X	X	X
IP_191334	X	X	X	X	X
IP_773656	X	X	X	X	X
IP_794359	X	X	X	X	X

14 AltProt have been deleted because they are identified but without a sufficient abundance to be quantified. The AltProt identification accession ID is according to the OpenProt Database. Venn diagram representation is based on the abundance obtain after analysis, function to the triplicate samples type (normal, p53, STIL, STIC, Carcinoma).



**Fig. 7 SpiderMass ex vivo real-time analyses of tissues section. A** Mass spectra obtained by SpiderMass from P52, STIL, and STIC preneoplasia lesion in fimbria, **B** The built PCA-LDA classification model based on 3 preneoplasia lesion stages; P53, STIL and STIC. **C** LDA representation of the 3-class PCA-LDA (right). The table (right) represents the “leave-one-patient-out” cross-validation results of the built classification model. **D** Specific markers detected in P53, or STIC. (\*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ ) no star for  $p > 0.05$ ).

We initially verified, using MALDI-MSI, that there are discernible differences in the proteomic content of pre-neoplastic lesions. With this information, we performed a spatially resolved proteomics analysis of each lesion in comparison to normal tissue. By comparing the proteomic content of HGSC and pre-neoplastic lesions found in the fallopian tube, we observed intriguing groupings that may correspond to the chronology of HGSC development proposed by different studies. Our analysis revealed that the normal tissue and p53 signature lesions cluster together and are near the STIL. In contrast, the STIC lesion presents a distinct proteomic signature and is closer to HGSC than the other lesions. Given the differences in composition between the fallopian tube tissue and the ovary, we focused our analysis on the pre-neoplastic lesions to gain a deeper understanding of the biological processes involved in their molecular transition.

By comparing the global content of each lesion, we successfully identified specific markers for p53 signature, the protein RNA-binding protein 10 (RBM10)- 3 specific proteins in STIC lesions -i.e., Glucosamine (N-acetyl)-6-sulfatase (GNS), Upstream binding transcription factor, RNA polymerase I (UBTF) and ATPase family, AAA domain containing 3 (ATAD3A/B). Specific mutated proteins from each lesion stage were also characterized. These mutations (mutated Histone H4, Histone H2B type 1-C/E/F/G/I) included in p53 signature, mutated 60S ribosomal protein L14 in STIL, and mutated (Na(+)/H(+)) exchange regulatory cofactor NHERF1 in STIC).

The p53 signature lesion is only characterized by a limited number of epithelial cells (~10) presenting a p53 overexpression. In our analysis, the p53 signature samples are close in their protein content to the normal tissue. Nevertheless, some specific proteins of these lesions could be obtained. Indeed, a signature emerged and demonstrated that among the 8 proteins overexpressed in the p53 signature, some of them are linked with a poor prognosis as follows: i) CAVIN1, SPTAN1 and FBLN5 for OC; ii) (EIF3B) for renal, head and neck cancers iii) MOB1B, EMILIN-2, EIF3B for liver cancer. On the contrary, 3 are linked with a favorable prognosis for renal cancer (SPTAN1, SPTBN1, CAVIN2).

We also dissected the underlying molecular mechanisms. In addition, the whole proteome analysis showed that the transition from normal to p53 signature was characterized by a residual immune response, translation activation, regulation of trafficking through cavins proteins, and ECM modifications via emilins.

From p53 signature to STIL, we showed that upregulated proteins were involved in the adhesion and anchoring of cells in the ECM, allowing stabilization of cells, along with an increase in maintenance activity instead of cell proliferation. Immune response and inflammation processes are still maintained, but an overall decrease of all metabolic processes, especially the TCA cycle, is observed without any Warburg effect. These observations are in line with a recent study suggesting that STIL could be considered "Dormant STICs" and take a prolonged time of more than one decade to develop into STIC [21]. We observed that mainly cell growth and maintenance metabolism pathways are downregulated in STIL, which could explain this "dormant" characteristic.

In contrast to the dormant profile of STILs, STICs show a more aggressive dysregulation with a decrease in the protein involved in cell adhesion. This could reflect the start of a cancer invasion step. STIC's upregulated proteins show various functions inside the cells and appear to be involved in different steps of cancer processes. An overexpression of proteins involved in energy pathways is observed. If we compare the molecular functions involved in the three different groups, calcium-dependent protein binding is overrepresented. In addition, heat shock protein activity is highly represented in this dataset. Moreover, proteins involved in telomere activities and centrosome maturation are specific to STICs compared to other lesions showing their cancerous specificity. Interestingly,

these two processes are linked with early tumorigenesis [53], and STIC precedes the development of many HGSCs [54]. A low expression of proteins involved in the extracellular structure organization and cell adhesion molecule activity was observed in STIC lesions. This would then suggest a breaching of the basement membrane and an increase in cell mobility that could induce an escape of these pre-cancerous cells to a distant organ. Finally, we observed a high level of proteins involved in extracellular vesicles including exosome production which could consequently modify the microenvironment and promote distant carcinogenic processes.

From pre-neoplastic lesions to HGSC, one protein was found common in the pre-neoplastic lesions and HGSC, the Fibrillarin (FBL) protein. FBL is one of the core proteins of box C/D small nucleolar ribonucleoprotein complexes (snoRNP). This complex is involved in the first steps of pre-rRNA processing [55]. It has also been shown that dysregulation of ribosome biogenesis plays key roles in oncogenesis [56]. Overactivation in cancer cells of ribosome biogenesis could be due to a loss of function of RNA polymerase repressors such as p53 [57]. Particularly FBL expression has been demonstrated to be correlated with p53 activity [58]. High levels of FBL protein were associated with the expression of mutant p53 and contributed to tumorigenesis by altering translational control of key cancer genes. UBTF is only present in STIC and HGSC. This protein is known to be a transcription factor involved in the regulation of the RNA polymerase I (Pol I), implicated in the regulation of cell cycle checkpoint, DNA damage response and regulated by MYC [59, 60]. Different studies have shown a link between Pol I initiation factor and chemoresistance of ovarian cancer [61].

Similarities between these pre-neoplastic lesions and HGSC have been found especially concerning *TP53* mutations and other genomic changes, but without excessive cell proliferation [62]. In that context, we decided to explore the presence of mutations at the protein level in the different steps of tubal tumor development. The use of a specific database for protein mutation, XMAN, resulted in the identification of 83 mutated peptide sequences, of which 4 were observed directly in the tandem mass spectra. It is interesting to observe that most of the proteins mutated in the p53 signature have been initially identified in endometrium cancer and are linked to histones and ribosomes suggesting that epigenetic regulation may be involved in the translation of the pre-neoplastic lesions.

Polyubiquitin C is related to proteasome and self-antigen presentation. Synaptopodin-2 and the Isoform 2 of Drebrin-like protein are actin-binding proteins known to be respectively invasive cancer biomarkers [63] and potential markers for breast, lung, and colorectal cancers [64]. Finally, Actin, alpha cardiac muscle 1 is involved in cisplatin ovarian cancer cell resistance [65] and the synaptic vesicle membrane protein VAT-1 homolog is a marker of epithelial cells [66].

These mutated proteins were only found in histones from STIL and STIC. Most of the mutated peptides observed in STIL, STIC, p53/STIL, or in Normal/p53 signature/STIL/STIC are linked to the cytoskeleton and cell migration. These mutations could be key drivers of cancer's process from the epigenetic modification step, and then on cytoskeleton modifications for cell migration and cancer development.

Using a dedicated database, identification of significantly variable AltProts, provides us with a new view of unknown markers. Some of the identified AltProts or transcripts in the present work have been described in other cancer studies. For example, the pseudogene KRT8P32 has been identified in breast cancer [67] and the pseudogene ACTBP11 was observed in GM12878 (B cells) cell lines according to Diana-LncBase V3 [68]. Additionally, it has recently been shown that AltANKRD28 can act as a novel BRCA1-interacting protein in breast and ovarian cancer [69]. Furthermore, RP11-13K12.2 is described to be overexpressed

in pediatric ovarian fibrosarcoma [70]. COL18A1 and B4GALNT4 are also overexpressed in endometrial and ovarian cancers according to the HPA and are unfavorable markers for overall survival for both cancers. Some of these AltProts are specific to the type of lesion - i.e. LOC105376924 and COL18A1 in STIL, KRT8P11 in STIC, and ANKRD28 in STIL-STIC. Altogether, these Altprots may be clinically relevant, and further investigation may improve the understanding of the disease.

Finally, our results identified AltProts and many mutations affecting histones and cytoskeleton proteins observed between normal to p53 signatures compared to the other lesions. These findings are in line with the hypothesis of epigenetic reprogramming toward carcinogenesis and cell transformation [71]. A recent study established that epigenetic reprogramming occurs specifically in the proximal end of the fallopian tubes in BRCA mutation carriers. This epigenetic reprogramming event is driven by aberrantly high AICDA (also named AID, activation-induced cytosine deaminase) expression and is an integral early pre-malignant event in HGSC development. In this context, our results would provide an interesting starting point for further studies, especially concerning the potential link to the endometrial carcinoma cells' exfoliation due to epigenetic factors leading to carcinogenesis [6]. It is also hypothesized that STIL represents exfoliated precursor cells that eventually undergo malignant transformation within the peritoneal cavity [72].

Using SpiderMass technology on FFPE tissue, we have successfully demonstrated the ability to distinguish healthy tissues from pre-neoplastic conditions in a lipidomic study. Specifically, we identified two phosphatidylserines (PS (22:1/20:1) and PS (22:0/20:0)) as discriminatory markers between p53 signature and healthy tissue, while PE(P-18:0/22:4) and TG (16:0/18:1/20:4) were found to be overexpressed in the lesions. Additionally, we specifically observed the presence of phosphatidylcholine (P-16:0/22:6) and diglyceride (DG (37:6)) in STICs, enabling their distinct identification. Notably, the inclusion of phosphatidylglycerol (PG (O-16:0/20:1)) allowed us to differentiate between STILs and STICs. This significant variability in lipid composition distinguishes our study from previous investigations in ovarian cancer. In Ovarian Clear Cell Carcinoma (OCCC) studied from FFPE tissues, the number of unsaturated lipid species increases whereas the number of saturated lipid species decreases in OCCC compared to the controls. PE (32:1), PE (34:1), PI (34:1), and PS (36:1), present a common signature in OCCC [73]. In serous ovarian cancer, the overexpression of fatty acid binding protein 4 leads to an increase of glycerolipids, glycerophosphoethanolamines, glycerophosphoinositols such as LysoPE, LysoPG, and LysoPI [74]. Cancer cells rely on lipogenesis to adapt to cytotoxic stress in the tumor microenvironment. This is particularly important in tumor areas where the exogenous supply of fatty acids is scarce, such as in hypovascular and hypoxic regions. Moreover, it is also interesting to note the presence of several dietary fatty acid (FA) precursors (e.g., 20:4, 22:4) issued from omega-6 FA to give polyunsaturated FA known to enhance the carcinogenic process in different cancers [75]. The dietary FA are used by cancer cells as energy supply by lipogenesis for membrane biosynthesis, signaling processes and ROS process inhibition. The observations we made regarding lipid modulation are substantiated by the presence of several dysregulated metabolic enzymes in cancer. Notably, we found that PHGDH is overexpressed in STIL and STIC lesions, which is consistent with its high expression in other cancer types, including breast, colon, and endometrial cancer. Moreover, there is evidence linking triglyceride levels to PHGDH DNA methylation, suggesting a potential regulatory role of this enzyme in triglyceride metabolism in cancer [76, 77]. Furthermore, our study identified other proteins related to lipid metabolism that are overexpressed in STICs, such as MARCKS-related protein, 2,4-dienoyl-CoA reductase [78], Trifunctional enzyme subunit beta [79], and 3-ketoacyl-CoA thiolase. Conversely, certain lipid metabolism proteins, like Aldehyde

dehydrogenase family 3 member B1 (ALDH3B1), implicated in the oxidation of medium and long-chain lipid-derived aldehydes generated in the plasma membrane [80], were found to be underexpressed in STILs. Finally, the study of Ackerman et al. [81] shows the importance of saturated fatty acid storage into triglycerides to prevent toxicity and preserve homeostasis in a hypoxic environment on clear cell renal cell carcinoma by releasing oleate (18:1) that will embed into phospholipids. Interestingly, TG (16:0/18:1/20:4) was found discriminative of fimbria lesions and even more abundant in STICs. STIC is a high-proliferative lesion that exhibits the potential to utilize triglycerides as a means to safeguard its homeostasis during growth. Furthermore, it has been observed that patients with BRCA1 and BRCA2 mutations demonstrate elevated levels of triglycerides [82]. This finding implies that individuals with BRCA1 and BRCA2 mutations may have heightened triglyceride levels, which could contribute to the enhanced survivability of lesional cells by preserving homeostasis and facilitating the development of STICs.

## CONCLUSION

In conclusion, this study provides valuable insights into the molecular events that lead to the development of ovarian cancer. By analyzing the proteomic content of pre-neoplastic lesions in the fimbriated end of the fallopian tube, we were able to identify the underlying mechanisms and potential timeline of events that lead to the formation of ovarian cancer. Our findings support previous studies proposing a sequence of molecular events that start with p53 signature lesions and progress through STIL to STIC and ultimately to HGSC. The results of this study contribute to a better understanding of the etiology of ovarian cancer and may have important implications for the development of early-detection biomarkers and therapeutic strategies.

## DATA AVAILABILITY

The MS data sets and Perseus result files used for analysis were deposited at the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>, [83]) via the PRIDE partner repository with the data set identifier PXD020024 (for reviewer access only, Username: reviewer86596@ebi.ac.uk; Password: x88jqaDV).

## REFERENCES

- Braicu E-I, Sehoul J, Richter R, Pietzner K, Denkert C, Fotopoulou C. Role of histological type on surgical outcome and survival following radical primary tumour debulking of epithelial ovarian, fallopian tube and peritoneal cancers. *Br J Cancer*. 2011;105:1818–24.
- Pal T, Permeth-Wey J, Betts JA, Krischer JP, Fiorica J, Arango H, et al. BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer*. 2005;104:2807–16.
- Seidman JD, Horkayne-Szakaly I, Haiba M, Boice CR, Kurman RJ, Ronnett BM. The histologic type and stage distribution of ovarian carcinomas of surface epithelial origin. *Int J Gynecol Pathol*. 2004;23:41–4.
- Leblanc E, Narducci F, Farre I, Peyrat JP, Taieb S, Adenis C, et al. Radical fimbriectomy: a reasonable temporary risk-reducing surgery for selected women with a germ line mutation of BRCA 1 or 2 genes? Rationale and preliminary development. *Gynecologic Oncol*. 2011;121:472–6.
- Lee Y, Miron A, Drapkin R, Nucci MR, Medeiros F, Saleemuddin A, et al. A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J Pathol*. 2007;211:26–35.
- Mingels MJ, Van Ham MAPC, De Kievit IM, Snijders MP, Van Tilborg AAG, Bulten J, et al. Müllerian precursor lesions in serous ovarian cancer patients: using the SEE-Fim and SEE-End protocol. *Mod Pathol*. 2014;27:1002–13.
- Labidi-Galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M et al. High grade serous ovarian carcinomas originate in the fallopian tube. *Nat Commun*. 2017;8. <https://doi.org/10.1038/s41467-017-00962-1>.
- Seidman JD, Yemelyanova A, Zaino RJ, Kurman RJ. The fallopian tube-peritoneal junction: a potential site of carcinogenesis. *Int J Gynecol Pathol*. 2011;30:4–11.
- Crum CP, Drapkin R, Kindelberger D, Medeiros F, Miron A, Lee Y. Lessons from BRCA: the tubal fimbria emerges as an origin for pelvic serous cancer. *Clin Med Res*. 2007;5:35–44.

10. Medeiros F, Muto MG, Lee Y, Elvin JA, Callahan MJ, Feltmate C, et al. The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am J Surgical Pathol.* 2006;30:230–6.
11. Gan C, Chenoy R, Chandrasekaran D, Brockbank E, Hollingworth A, Vimplis S, et al. Persistence of fimbrial tissue on the ovarian surface after salpingectomy. *Am J Obstet Gynecol.* 2017;217:425.e1–e16.
12. Vang R, Visvanathan K, Gross A, Maambo E, Gupta M, Kuhn E, et al. Validation of an algorithm for the diagnosis of serous tubal intraepithelial carcinoma. *Int J Gynecol Pathol.* 2012;31:243–53.
13. Visvanathan K, Vang R, Shaw P, Gross A, Soslow R, Parkash V, et al. Diagnosis of serous tubal intraepithelial carcinoma based on morphologic and immunohistochemical features: a reproducibility study. *Am J Surgical Pathol.* 2011;35:1766–75.
14. Auersperg N. The origin of ovarian cancers -hypotheses and controversies. *Front Biosci Scholar.* 2013;5:709–19.
15. Kobayashi H, Iwai K, Niiro E, Morioka S, Yamada Y, Ogawa K, et al. The conceptual advances of carcinogenic sequence model in high-grade serous ovarian cancer. *Biomed Rep.* 2017;7:209–13.
16. Roh MH, Yassin Y, Miron A, Mehra KK, Mehrad M, Monte NM, et al. High-grade fimbrial-ovarian carcinomas are unified by altered p53, PTEN and PAX2 expression. *Mod Pathol.* 2010;23:1316–24.
17. Bowtell DDL. The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer.* 2010;10:803–8.
18. Folkins AK, Jarboe EA, Roh MH, Crum CP. Precursors to pelvic serous carcinoma and their clinical implications. *Gynecologic Oncol.* 2009;113:391–6.
19. Singh N, Gilks CB, Wilkinson N, McCluggage WG. The secondary Müllerian system, field effect, BRCA, and tubal fimbria: our evolving understanding of the origin of tubo-ovarian high-grade serous carcinoma and why assignment of primary site matters. *Pathology.* 2015;47:423–31.
20. Bachert SE, McDowell A, Piccoro D, Branch LB. Serous tubal intraepithelial carcinoma: A concise review for the practicing pathologist and clinician. *Diagnostics.* 2020;10. <https://doi.org/10.3390/diagnostics10020102>.
21. Wu RC, Wang P, Lin SF, Zhang M, Song Q, Chu T, et al. Genomic landscape and evolutionary trajectories of ovarian cancer precursor lesions. *J Pathol.* 2019;248:41–50.
22. Zeppernick F, Meinhold-Heerlein I, Shih IM. Precursors of ovarian cancer in the fallopian tube: Serous tubal intraepithelial carcinoma - an update. *J Obstet Gynaecol Res.* 2015;41:6–11.
23. Aboulouard S, Wisztorski M, Duhamel M, Saudemont P, Cardon T, Narducci F et al. In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer. *Cell Rep Med.* 2021;2. <https://doi.org/10.1016/j.xcrm.2021.100318>.
24. Delcourt V, Franck J, Leblanc E, Narducci F, Robin YM, Gimeno JP, et al. Combined mass spectrometry imaging and top-down microproteomics reveals evidence of a hidden proteome in ovarian cancer. *EBioMedicine.* 2017;21:55–64.
25. Duhamel M, Drelich L, Wisztorski M, Aboulouard S, Gimeno JP, Ogrinc N et al. Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. *Nat Commun.* 2022;13. <https://doi.org/10.1038/s41467-022-34208-6>.
26. Cardon T, Hervé F, Delcourt V, Roucou X, Salzet M, Franck J, et al. Optimized sample preparation workflow for improved identification of ghost proteins. *Anal Chem.* 2020;92:1122–9.
27. Vergara D, Verri T, Damato M, Trerotola M, Simeone P, Franck J, et al. A hidden human proteome signature characterizes the epithelial mesenchymal transition program. *Curr Pharm Des.* 2020;26:372–5.
28. Mingels MJJM, van Ham MA, de Kievit IM, Snijders MPML, van Tilborg AA, Bulten J, et al. Müllerian precursor lesions in serous ovarian cancer patients: using the SEE-Fim and SEE-End protocol. *Mod Pathol.* 2014;27:1002–13.
29. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008;26:1367–72.
30. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat Methods.* 2016;13:731–40.
31. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CYJ, Williamson NA, et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics.* 2015;15:2597–601.
32. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, et al. A pathology atlas of the human cancer transcriptome. *Science.* 2017;357. <https://doi.org/10.1126/science.aan2507>.
33. Flores MA, Lazar IM. XMAN v2-a database of Homo sapiens mutated peptides. *Bioinformatics.* 2020;36:1311–3.
34. Brunet MA, Brunelle M, Lucier JF, Delcourt V, Levesque M, Grenier F, et al. OpenProt: a more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Res.* 2019;47:D403–D410.
35. Leblanc E, Narducci F, Ferron G, Mailliez A, Charvolin JY, Houssein EH, et al. Prophylactic radical fimbriectomy with delayed oophorectomy in women with a high risk of developing an ovarian carcinoma: results of a prospective national pilot study. *Cancers.* 2023;15. <https://doi.org/10.3390/cancers15041141>.
36. Wisztorski M, Quanco J, Franck J, Fatou B, Salzet M, Fournier I. Droplet-based liquid extraction for spatially-resolved microproteomics analysis of tissue sections. *Methods Mol Biol.* 2017;1618:49–63. [https://doi.org/10.1007/978-1-4939-7051-3\\_6](https://doi.org/10.1007/978-1-4939-7051-3_6).
37. Bayarkhangai B, Noureldin S, Yu L, Zhao N, Gu Y, Xu H, et al. A comprehensive and perspective view of oncoprotein SET in cancer. *Cancer Med.* 2018;7:3084–94.
38. Liang W, Gao R, Yang M, Wang X, Cheng K, Shi X, et al. MARCKSL1 promotes the proliferation, migration and invasion of lung adenocarcinoma cells. *Oncol Lett.* 2020;19:2272–80.
39. Bonnet-Magnaval F, DesGroseillers L. The Staufin1-dependent cell cycle regulon or how a misregulated RNA-binding protein leads to cancer. *Biol Rev.* 2021;96:2192–208.
40. Zhang XY, Hong SS, Zhang M, Cai QQ, Zhang MX, Xu CJ. Proteomic alterations of fibroblasts induced by ovarian cancer cells reveal potential cancer targets. *Neoplasma.* 2018;65:104–12.
41. Shimamura M, Nagayama Y, Matsuse M, Yamashita S, Mitsutake N. Analysis of multiple markers for cancer stem-like cells in human thyroid carcinoma cell lines. *Endocr J.* 2014;61:481–90.
42. Kahlert C, Weber H, Mogler C, Bergmann F, Schirmacher P, Kennigott HG, et al. Increased expression of ALCAM/CD166 in pancreatic cancer is an independent prognostic marker for poor survival and early tumour relapse. *Br J Cancer.* 2009;101:457–64.
43. Lunter PC, Van Kilsdonk JWJ, Van Beek H, Cornelissen IMHA, Bergers M, Willems PHGM, et al. Activated leukocyte cell adhesion molecule (ALCAM/CD166/MEMD), a novel actor in invasive growth, controls matrix metalloproteinase activity. *Cancer Res.* 2005;65:8801–8.
44. Creekmore AL, Silkworth WT, Cimini D, Jensen RV, Roberts PC, Schmelz EM. Changes in gene expression and cellular architecture in an ovarian cancer progression model. *PLoS One.* 2011;6. <https://doi.org/10.1371/journal.pone.0017676>.
45. Li J, Zuo X, Shi J, Zhang J, Duan X, Xu G. Flotillin 1 is differentially expressed in human epithelial ovarian tumors. *Neoplasma.* 2018;65:561–71.
46. Peluso JJ. Progesterone signaling mediated through progesterone receptor membrane component-1 in ovarian cells with special emphasis on ovarian cancer. *Steroids.* 2011;76:903–9.
47. Santin AD, Zhan F, Bellone S, Palmieri M, Cane S, Bignotti E, et al. Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: Identification of candidate molecular markers for ovarian cancer diagnosis and therapy. *Int J Cancer.* 2004;112:14–25.
48. Zhao G, Chen J, Deng Y, Gao F, Zhu J, Feng Z, et al. Identification of NDRG1-regulated genes associated with invasive potential in cervical and ovarian cancer cells. *Biochem Biophys Res Commun.* 2011;408:154–9.
49. Scatena R, Bottoni P, Pontoglio A, Giardina B. Revisiting the Warburg effect in cancer cells with proteomics. The emergence of new approaches to diagnosis, prognosis and therapy. *Proteom Clin Appl.* 2010;4:143–58.
50. Kindelberger DW, Lee Y, Miron A, Hirsch MS, Feltmate C, Medeiros F, et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am J Surgical Pathol.* 2007;31:161–9.
51. Balog J, Sasi-Szabó L, Kinross J, Lewis MR, Muirhead LJ, Veselkov K, et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Scie Transl Med.* 2013;5. <https://doi.org/10.1126/scitranslmed.3005623>.
52. Zhang S, Dolgalev I, Zhang T, Ran H, Levine DA, Neel BG. Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. *Nat Commun.* 2019;10:5367.
53. Kuhn E, Meeker A, Wang T-L, Sehdev AS, Kurman RJ, Shih I-M. Shortened telomeres in serous tubal intraepithelial carcinoma: an early event in ovarian high-grade serous carcinogenesis. *Am J Surg Pathol.* 2010;34:829–36.
54. Kuhn E, Wang TL, Doberstein K, Bahadiri-Talbot A, Ayhan A, Sehdev AS, et al. CCNE1 amplification and centrosome number abnormality in serous tubal intraepithelial carcinoma: further evidence supporting its role as a precursor of ovarian high-grade serous carcinoma. *Mod Pathol.* 2016;29:1254–61.
55. Newton K, Petfalski E, Tollervy D, Cáceres JF. Fibrillarlin is essential for early development and required for accumulation of an intron-encoded small nucleolar RNA in the mouse. *Mol Cell Biol.* 2003;23:8519–27.
56. Truitt ML, Ruggero D. New frontiers in translational control of the cancer genome. *Nat Rev Cancer.* 2016;16:288–304.
57. Zhai W, Comai L. Repression of RNA polymerase I transcription by the tumor suppressor p53. *Mol Cell Biol.* 2000;20:5930–8.
58. Marcel V, Ghayad SE, Belin S, Therizols G, Morel AP, Solano-González E, et al. P53 acts as a safeguard of translational control by regulating fibrillarlin and rRNA methylation in cancer. *Cancer Cell.* 2013;24:318–30.

59. Sanij E, Diesch J, Lesmana A, Poortinga G, Hein N, Lidgerwood G, et al. A novel role for the pol I transcription factor ubtf in maintaining genome stability through the regulation of highly transcribed pol II genes. *Genome Res.* 2015;25:201–12.
60. Rossetti S, Wierzbicki AJ, Sacchi N. Undermining ribosomal RNA transcription in both the nucleolus and mitochondrion: An offbeat approach to target MYC-driven cancer. *Oncotarget.* 2018;9:5016–31.
61. Cornelison R, Dobbin ZC, Katre AA, Jeong DH, Zhang Y, Chen D, et al. Targeting RNA-polymerase I in both chemosensitive and chemoresistant populations in epithelial ovarian cancer. *Clin Cancer Res.* 2017;23:6529–40.
62. Yamamoto Y, Ning G, Howitt BE, Mehra K, Wu L, Wang X, et al. In vitro and in vivo correlates of physiological and neoplastic human Fallopian tube stem cells. *J Pathol.* 2016;238:519–30.
63. Kai FB, Fawcett JP, Duncan R. Synaptopodin-2 induces assembly of peripheral actin bundles and immature focal adhesions to promote lamellipodia formation and prostate cancer cell migration. *Oncotarget.* 2015;6:11162–74.
64. Chen W, Liu G, Jin M, Ju P, Xu J, Zhang Y, et al. Drebrin is a potential diagnostic biomarker for breast cancer. *Int J Clin Exp Med.* 2016;9:23598–604.
65. Pan S, Cheng L, White JT, Lu W, Utleg AG, Yan X, et al. Quantitative proteomics analysis integrated with microarray data reveals that extracellular matrix proteins, catenins, and P53 binding protein 1 are important for chemotherapy response in ovarian cancers. *OMICS A J Integr Biol.* 2009;13:345–54.
66. Koch J, Foekens J, Timmermans M, Fink W, Wirzbach A, Kramer MD, et al. Human VAT-1: a calcium-regulated activation marker of human epithelial cells. *Arch Dermatological Res.* 2003;295:203–10.
67. Elaine Hardman W, Primerano DA, Legenza MT, Morgan J, Fan J, Denvir J. mRNA expression data in breast cancers before and after consumption of walnut by women. *Data Brief.* 2019;25. <https://doi.org/10.1016/j.dib.2019.104050>.
68. Karagkouni D, Paraskevopoulou MD, Tastsoglou S, Skoufos G, Karavangeli A, Pierros V, et al. DIANA-LncBase v3: Indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res.* 2020;48:D101–D110.
69. Vincent A, Berthel E, Dacheux E, Magnard C, Dalla Venezia NL. BRCA1 affects protein phosphatase 6 signalling through its interaction with ANKRD28. *Biochem J.* 2016;473:949–60.
70. Melendez-Zajgla J, Mercado-Celis GE, Gaytan-Cervantes J, Torres A, Gabiño NB, Zapata-Tarres M, et al. Genomics of a pediatric ovarian fibrosarcoma. Association with the DICER1 syndrome. *Sci Rep.* 2018;8. <https://doi.org/10.1038/s41598-018-21663-9>.
71. Bartlett TE, Chindera K, McDermott J, Breeze CE, Cooke WR, Jones A, et al. Epigenetic reprogramming of fallopian tube fimbriae in BRCA mutation carriers defines early ovarian cancer evolution. *Nat Commun.* 2016;7. <https://doi.org/10.1038/ncomms11620>.
72. Kuhn E, Kurman RJ, Vang R, Sehdev AS, Han G, Soslow R, et al. TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma-evidence supporting the clonal relationship of the two lesions. *J Pathol.* 2012;226:421–6.
73. Mir SA, Wong SBJ, Narasimhan K, Esther CWL, Ji S, Burla B, et al. Lipidomic analysis of archival pathology specimens identifies altered lipid signatures in ovarian clear cell carcinoma. *Metabolites.* 2021;11. <https://doi.org/10.3390/metabo11090597>.
74. Gharpure KM, Pradeep S, Sans M, Rupaimoole R, Ivan C, Wu SY, et al. FABP4 as a key determinant of metastatic potential of ovarian cancer. *Nat. Commun.* 2018;9. <https://doi.org/10.1038/s41467-018-04987-y>.
75. Wang W, Bai L, Li W, Cui J. The lipid metabolic landscape of cancers and new therapeutic perspectives. *Front Oncol.* 2020;10. <https://doi.org/10.3389/fonc.2020.605154>.
76. Truong V, Huang S, Dennis J, Lemire M, Zwingerman N, Aissi D, et al. Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci Rep.* 2017;7. <https://doi.org/10.1038/s41598-017-09552-z>.
77. Li M, Wu C, Yang Y, Zheng M, Yu S, Wang J, et al. 3-Phosphoglycerate dehydrogenase: a potential target for cancer treatment. *Cell Oncol.* 2021;44:541–56.
78. Blomme A, Ford CA, Mui E, Patel R, Ntala C, Jamieson LE, et al. 2,4-dienoyl-CoA reductase regulates lipid homeostasis in treatment-resistant prostate cancer. *Nat Commun.* 2020;11. <https://doi.org/10.1038/s41467-020-16126-7>.
79. Sekine Y, Yamamoto K, Kurata M, Honda A, Onishi I, Kinowaki Y, et al. HADHB, a fatty acid beta-oxidation enzyme, is a potential prognostic predictor in malignant lymphoma. *Pathology.* 2022;54:286–93.
80. Chen J, Liu Z, Wu Z, Li W, Tan X. Identification of a chemoresistance-related prognostic gene signature by comprehensive analysis and experimental validation in pancreatic cancer. *Front Oncol.* 2023;13. <https://doi.org/10.3389/fonc.2023.1132424>.
81. Ackerman D, Tumanov S, Qiu B, Michalopoulou E, Spata M, Azzam A, et al. Triglycerides promote lipid homeostasis during hypoxic stress by balancing fatty acid saturation. *Cell Rep.* 2018;24:2596–2605.e5.
82. Ramadan S, Arm J, Silcock J, Santamaria G, Buck J, Roy M, et al. Lipid and metabolite deregulation in the breast tissue of women carrying BRCA1 and BRCA2 genetic mutations. *Radiology.* 2015;275:675–82.
83. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol.* 2014;32:223–6.

## ACKNOWLEDGEMENTS

The authors thanks Annie Desmons for her technical help in this project.

## AUTHOR CONTRIBUTIONS

IF, EL and MS participated in experimental design. IF, and MS supervised the study. MW, SA, MD, LR, PS performed the experiments. Samples were provided by EL and FN. The pathology analyses were done by YMR, DB, ASL, FK, MS performed the systemic biology analyses. MW, MS, MD, LR, and PS performed data analyses. IF, MS and MW wrote the manuscript with contributions from all co-authors. LR, MD and NH participated in writing and reviewing the manuscript. MS, EL, and IF found funds for the project.

## FUNDING

This research was supported by grants from Ministère de de l'Enseignement Supérieur et de la Recherche (MESR), INSERM (IF, MS), the PHRC FIMBRIA (EL), University of Lille (LR).

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41419-023-06165-5>.

**Correspondence** and requests for materials should be addressed to Eric Leblanc, Isabelle Fournier or Michel Salzet.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

# **MACHINE LEARNING BASED OVARIAN CANCER DIAGNOSTIC AND PROGNOSTIC FOR REAL TIME IMMUNE SCORING AT THE OPERATING ROOM**

**Lucas Roussel<sup>1†</sup>, Yanis Zirem<sup>1†</sup>, Laurine Lagache<sup>1†</sup>, Delphine Hudry<sup>1,2†</sup>, Bertrand Meresse<sup>3</sup>, Marie Delbeke<sup>3</sup>, Eric Leblanc<sup>1,2</sup>, Gargey Yagnik<sup>4</sup>, Mark J. Lim<sup>4</sup>, Kenneth J. Rothschild<sup>4,6</sup>, Yves-Marie Robin<sup>1,2</sup>, Camille Pasquesoone<sup>1,2</sup>, Anne-Sophie Lemaire<sup>1,2</sup>, Delphine Bertin<sup>1,2</sup>, Fabrice Narducci<sup>1,2‡</sup>, Michel Salzet<sup>1,5†\*</sup>, Isabelle Fournier<sup>1,5†\*</sup>**

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

<sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Center, F-59000 Lille, France

<sup>3</sup>Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000, Lille, France.

<sup>4</sup>AmberGen, Inc., Billerica, MA, United States

<sup>5</sup>Institut Universitaire de France, 75000 Paris

<sup>6</sup>Department of Physics, Photonics Center, Boston University, Boston, MA, United States

† equal contribution

\*Co-Corresponding author

‡Lead contact

Prof. Michel Salzet, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - Inserm U1192 - Bât SN3, 1<sup>er</sup> étage, Campus Cité Scientifique, Université de Lille, F-59655 Villeneuve d'Ascq Cedex. Phone: +33 (0)320 434 194; email: [michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr)

Prof. Isabelle Fournier, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - Inserm U1192 - Bât SN3, 1<sup>er</sup> étage, Campus Cité Scientifique, Université de Lille, F-59655 Villeneuve d'Ascq Cedex. Phone: +33 (0)320 434 194; email: [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr)

Dr. Fabrice Narducci, Department of Gynecology Oncology, Oscar Lambret Center, F-59000 Lille, France and Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - INSERM U1192 - Université Lille 1, Bât SN3, 1<sup>er</sup> étage, Cité Scientifique, F-59655 Villeneuve d'Ascq Cedex. Phone: + (33)320 295 959, email: [f-narducci@o-lambret.fr](mailto:f-narducci@o-lambret.fr)



## **SUMMARY**

Ovarian cancers are the most lethal gynecological cancers which enumerate more than 200 000 deaths every year in the world. Despite advancements in treatment modalities and patient care, this cancer is still a silent killer. The success of ovarian cancer management is intricately intertwined with the precision of surgical interventions. In response to this critical need, we have developed a robust diagnostic approach leveraging the molecular signatures of cancer subtypes and machine learning techniques. This innovative method empowers surgeons to discern carcinosis with unprecedented clarity within the abdominal cavity during surgical procedures. Additionally, patient survival in the context of ovarian cancer is intimately associated with the extent of immune cell infiltration. In this comprehensive study, we introduce a novel method for direct visualization and precise localization of diverse immune cell populations within tissue sections. The prediction of presence of immune cells in tissues, called immune score, emerges as a pivotal factor with profound implications for both diagnosis and prognosis in ovarian cancer patients. This integrated diagnostic and prognostic pipeline holds substantial promise for guiding surgical interventions in real-time within the operative setting. By harnessing the power of molecular signatures, machine learning, and immune cell assessment, our approach not only enhances surgical precision but also offers a tangible opportunity to improve the overall survival outcomes of ovarian cancer patients.

## **IN BRIEF**

In this study Roussel *et al.* demonstrate that molecular fingerprint and machine learning enables accurate molecular diagnostic of ovarian cancer. This analysis can be performed in real-time, while the technology also offers immune scoring based on the molecular profiles of different immune cell populations. This immune score has the potential for predicting the outcome of the patients based on the infiltration of the immune cell populations.

## **HIGHLIGHTS**

- A robust classification model of OC sub-types for real-time analysis was established.
- Association of both morphological and molecular data helped the diagnosis of OC sub-types;
- Specific molecular pathways of OC sub-types were deciphered by lipid identification.
- An immune score based on the molecular signature of the immune cells predicts the patient outcome.

## **KEYWORDS**

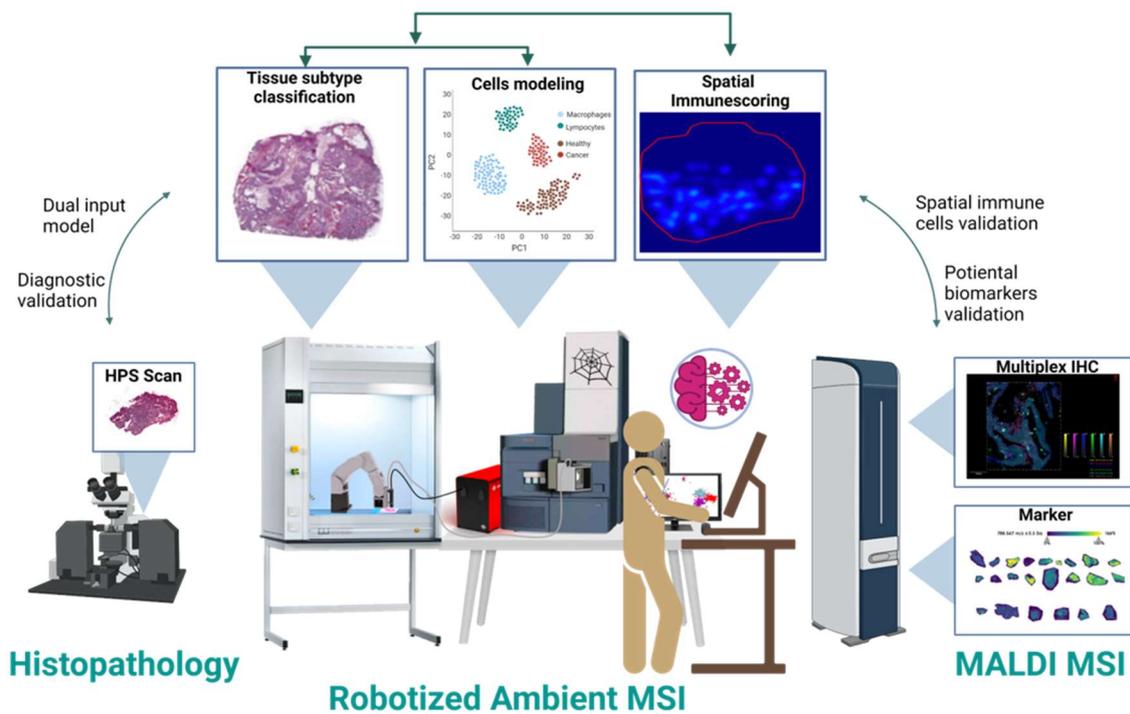
Ovarian cancer, diagnosis, prognosis, machine learning, ambient ionization mass spectrometry, SpiderMass, lipidomics, immune score, macrophages, T cells, MALDI-IHC

## INTRODUCTION

Ovarian carcinoma (OC) is one of the deadliest cancers, with 198000 deaths<sup>1</sup> out of 294 000 cases worldwide<sup>2</sup>. Most of patients (75%) are unfortunately diagnosed at advanced stages<sup>3</sup>. Therefore, the 5-year survival rate is estimated to be only of 46% for all stages combined and 26% for stage IV. There are different types of OC of different origins but the majority of them is epithelial (90%)<sup>4</sup>. Within epithelial ovarian cancer, four primary histologic subtypes are found; namely the serous, endometrioid, mucinous, and clear cell. Serous tumors are themselves categorized into high-grade serous carcinomas (HGSOC) and low-grade serous carcinomas (LGSC). HGSOCs account for 70% to 80% of all subtypes of epithelial ovarian cancer, while LGSCs account for less than 5%. Endometrioid, mucinous, and clear cell subtypes are less prevalent and account for only 10%, 3%, and 10%, respectively.

Regarding the patient management with OC, the therapeutic sequence is determined by a multidisciplinary evaluation based on an estimation of the tumor burden, staging according to FIGO classification(Prat, 2014) and the histological diagnostic by laparotomy<sup>6</sup>. For advanced stages, both American and European recommendations is to combine surgery and chemotherapy in expert centers as first-line therapy (Armstrong et al., 2019; Colombo et al., 2019). The reference systemic chemotherapy is based on 6 cycles of chemotherapy with carboplatin AUC 6 and paclitaxel 175 mg/m<sup>2</sup> every 3 weeks. PARP inhibitors can also be used as maintenance treatment for patients tested positive for homologous recombination deficiency<sup>7</sup>. Considering the management patients, complete resection of macroscopic tumors without residual disease is a major prognostic factor in advanced ovarian cancer. Cytoreduction surgery (CRS), which aims at resection without residual disease, is the cornerstone of OC treatment and thus complete CRS is significantly associated with improved survival<sup>8</sup>. Indeed, overall survival is dependent on the quality of CRS and complete resection without residual disease (classified as CC-0)<sup>9</sup> is a major prognostic factor. The disease-free survival is 22.3 months in case of CC-0, against 12.3 and 6.3 months in case of CC-1 (residual disease less than 2.5mm) or CC-2 (residual disease between 2.5mm and 25mm) respectively<sup>10</sup>. Median overall survival is comparably affected with 106 months with no gross residual, 66 months with residual disease ≤5mm, 34 months >2cm, (p<0.01)<sup>11</sup>. If the goal is to achieve a complete resection through radical surgery, it is preferable to limit the extent of resection to only what is strictly necessary. However, a more aggressive surgical approach results in potential major morbidity. The surgical complexity score is a scoring system developed to resume the extent of surgical procedures in ovarian cancer surgery<sup>12</sup> and increased complexity is associated with increased surgical morbidity. Similarly, as the operating time increases, the risk of postoperative complications rises<sup>13</sup>. However, achieving an optimal surgery for each patient through complete resection still presents challenges because surgeons lack tools to assess the nature of the tissue. In ovarian carcinoma, the assistance by the intraoperative frozen section histopathological analysis leads to a sensitivity on suspicious pelvic masses estimated at 90% with specificity at 99.5%<sup>14</sup>. But this exam in addition to requiring an expert gynecologic pathologist, remains time-consuming (15-20min), and cannot be repeated more than 2 to 3 times per patient, which is insufficient. Due to this difficult context of OC, with high rate of late diagnostic with possible spread of cancer in the abdominal cavities added to intrinsic issues

related to the intraoperative pathology exam, there is a clear need for a technology enabling one to track residual tumor in real-time during OC surgery, especially for later stages of surgical cytoreduction to improve patient survival and decrease post-operative morbidity. The future of surgery for OC is, thus, to: i) be able to remove the abdominal organs, the peritoneum, affected by carcinosis, to obtain a complete surgery and thus improve the survival of the patient, while; ii) not removing the peritoneum or organ free of carcinosis to decrease the morbidity of the procedure. These objectives cannot be met without the introduction of new technologies enabling a real-time diagnostic. In this context, we investigate the impact of the SpiderMass technology associated with deep learning to perform real-time diagnosis and prognosis of patient survival with OC. We introduce as well, associated with SpiderMass technology, the creation of a real-time immune scoring from patients excised tissues. This enabled a more accurate diagnostic associated with patient outcome. The fast detection of immune cells could increase the quality of patient management considering that this immune scoring could be performed intraoperatively in the future (**Figure 1**).



**Figure 1: Overall workflow developed in the study.** A robust classification model based on the combination of H&E staining and MS spectra by SpiderMass was created for the diagnosis of ovarian cancer. A second model was also built to detect the presence of the different cell types in the TME (including the immune cells). This model was applied to MS Imaging data to predict the presence of the immune cells across the tissue and create an immune score for prognosis.

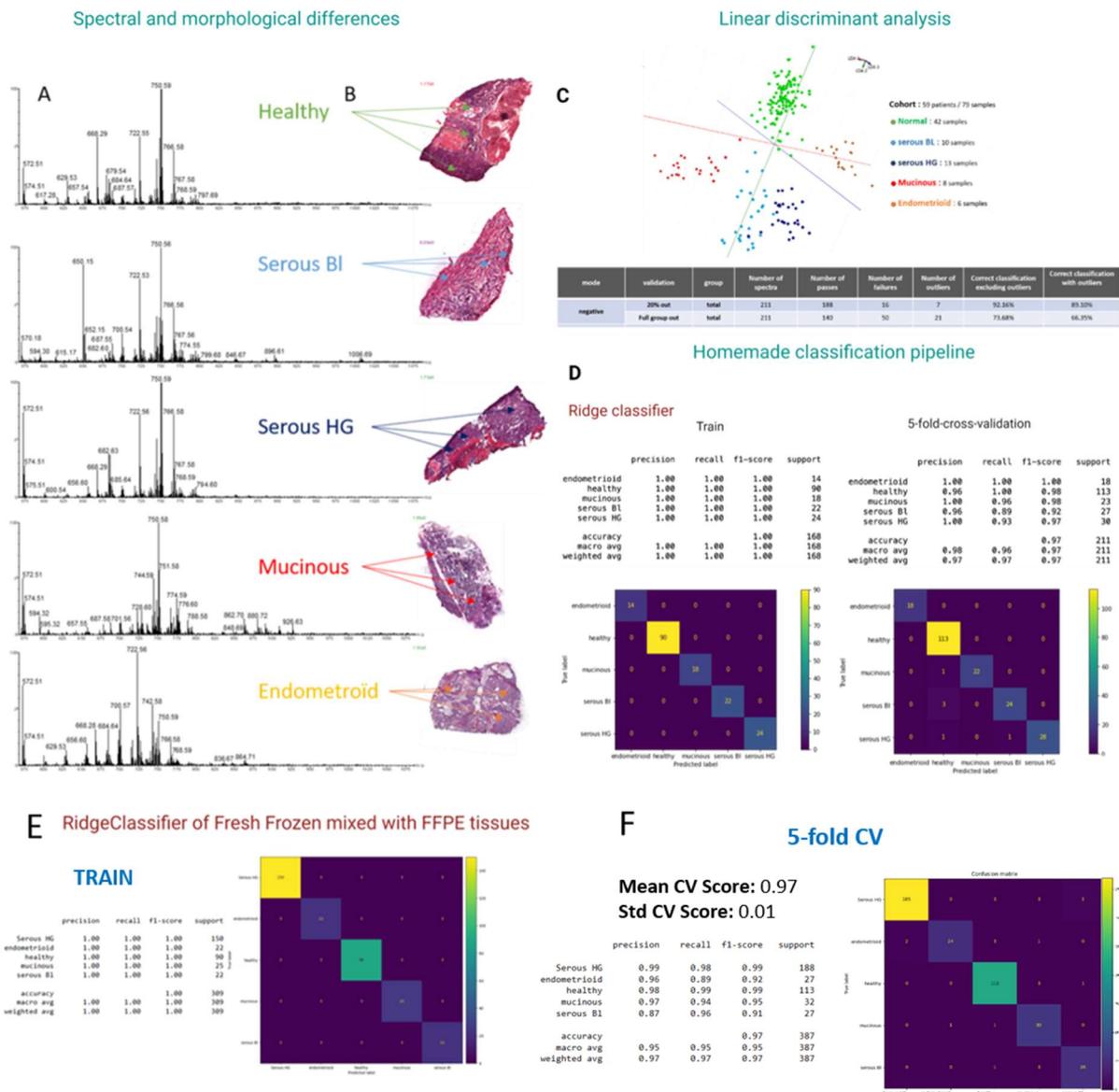
## RESULTS

### **Ovarian cancer histological subtyping based on molecular SpiderMass data**

Five principal histotypes are described for ovarian cancer (OC): high-grade serous carcinoma (HGSOC), low-grade serous carcinoma (LGSC), mucinous carcinoma (MC), endometrioid carcinoma (EC), and clear cell carcinoma (CCC)<sup>15</sup>. We aimed first at determining if SpiderMass technology could enable real-time molecular typing of these different histotypes versus normal ovary tissue and between themselves. To assess SpiderMass histotyping we collected data with the SpiderMass from histological sections on the OC prospective cohort (**Supp. Table 1**). The obtained MS spectra demonstrate differences in their molecular fingerprint according to the cancer type both in negative (**Figure 2A-B**) and positive ion mode (**Figure S1A-B**). Some lipids identified by MS/MS, *e.g.*,  $m/z$  750.55 (PE 36:4) or  $m/z$  885.5 (PI 20:4/18:0) showed a similar distribution across the different histotypes while others are clearly specific to one (**Figure S2**). To gain a better understanding of these distinct molecular patterns, we trained ML algorithms on the entire OC prospective patients cohort to create a classification model which allows for separating each histotype. Various ML algorithms were evaluated. First, we constructed a conventional multivariate analysis models based on PC-LDA using AMX software in both ion modes (**Figure 2C**). These models were built using the 79 samples of the prospective cohort (13 HGSOC, 8 MC, 10 borderline serous carcinoma (BL), 6 EC, 44 normal ovary) from 59 patients, and successfully discriminates the five classes in both ion modes (**Figure 2C**). The models were then validated using the "20%-out" validation. This method consists in training the model with 80% of the data and evaluate it with the 20% remaining. It resulted in correct classifications ranging from 93.78% without outliers to 92.89% with outliers in negative ion mode (**Figure 2C**), and from 96.10% without outliers to 95.17% with outliers in positive ion mode (**Figure S3A-B**). The "leave one patient" cross validation was then used and resulted it 72.11 % correct classification without outliers and 64.93% with outliers in negative mode. In positive mode, 74.19% correct classification excluding outliers and 66.76% with outliers.

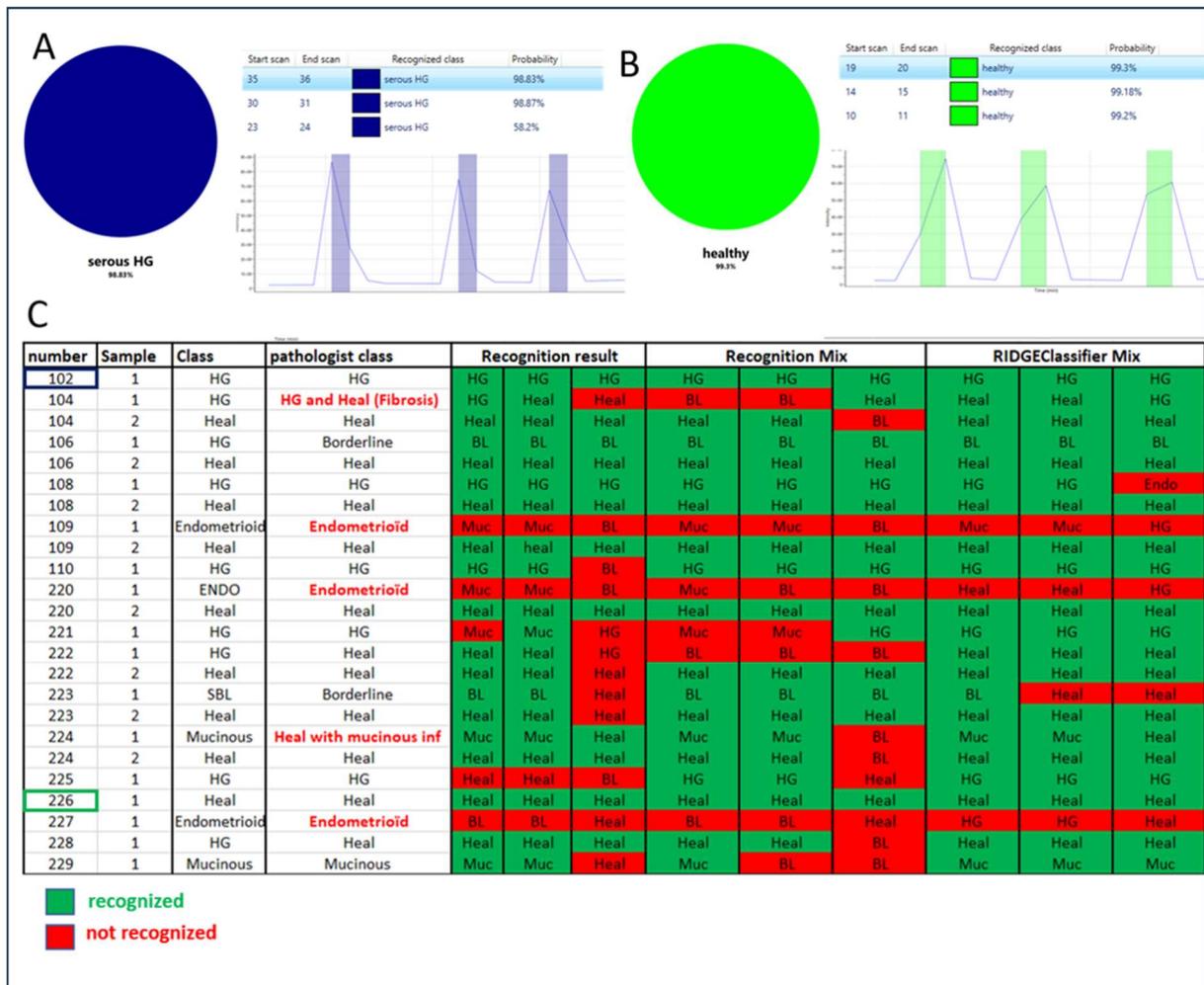
To improve the classification and to ensure robustness, we performed data-extension by combining the already analyzed tissue to an additional 83 samples from a retrospective cohort of FFPE tissues (**Supp. Table 3**). The classification LDA model generated from this cohort led to 83% correct classification without and with outliers in "20% out" and 70.24% without and 68.90% with outliers in "leave one patient out" out for FFPE alone in negative ion mode. (**Figure S4**). The extended data set did not improve the accuracy of the prediction compared with FF tissue model. Nevertheless, the mix up of FFPE and frozen tissues compared to FFPE alone increase the classification rate from 83.78% to 89.12% (**Figures S4 A, B, C**) and 76.30% with and 75.71% without outliers with "leave one patient" method in negative ion mode. The accuracy obtained is not sufficient since AMX software present limitations. The main one is that it is only based on LDA classification. This model misses some cross-validation and provides no explanation for predictions. Besides, LDA show\*s limited performances when the data of a class are more heterogeneous.

To address both the limitations in classifying the histotypes with lower prevalence in the LDA classification, we tested other classification solutions such as SVM, random forest and Perceptron algorithms (**Supp. Table 4**). Twenty-four classification algorithms with k-fold cross-validation were applied on the same dataset (**Supp. Table 4**). In 5-fold cross-validation method, SVM reached 95%, Random Forest 84% and Ridge Classifier 100%. For the evaluation methods, k-fold cross-validation was used. The two previous one (20% out and leave one patient) lacked robustness. In fact, 20% validation is randomly chosen and does not represent the full data set. The “leave one patient” method is known to increase the standard deviation of the mean accuracy. Thus, the k-fold cross validation was used for model evaluation for the rest of the analysis. The ridge classifier achieved the highest accuracy of 100% in training and 97% after 5-fold cross-validation in negative ion mode on the FF cohort (**Figure 2D, Supp table 4**). The ridge classifier<sup>16</sup>, which is suitable for high-dimensional datasets such as ours with 5000 m/z dimensions, is a linear classifier that uses L2 regularization to avoid overfitting and demonstrated strong performances for MS data from tissues<sup>17</sup>. This model was also used for positive ion mode that resulted also in 100% in training and 94% after 5-fold cross-validation (**Figure S3C-D**). The FFPE tissues cohort was also used to build another model resulting in 100% accuracy in training and 94.2% after 5-fold cross-validation (**Figure S4D-E**). As this model was more robust than LDA classification we decided to performed data-extension on ridge Classifier. This new model which combines both FFPE and FF tissues resulted in an excellent classification with 100% of accuracy in training and 97% after 5-fold cross-validation (**Figure 2E-F**).



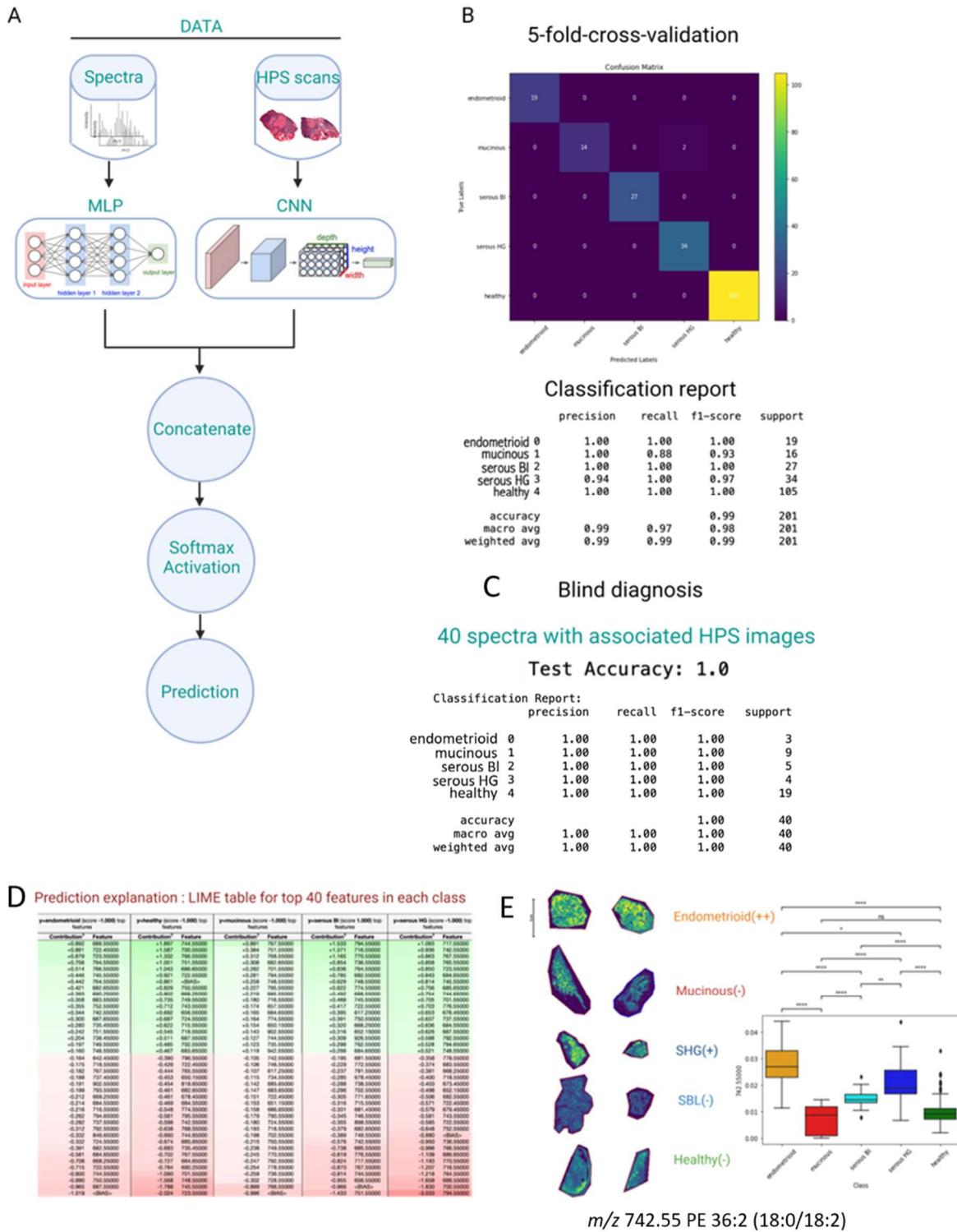
**Figure 2: Multivariate statistical analysis-based model for ovarian cancer diagnosis. (A-B)** Morphological and spectral fingerprint of the different OC subtypes. **(A)** MS spectra recorded in the negative ion mode with a 4mJ/shot burst of 10 shots are presented for each OC subtype. **(B)** HPS staining of these specific tissues with the spot used for acquisition. **(C)** LDA classification model based on the MS spectra recorded in the negative ion mode from an OC cohort of 79 samples. Green dots correspond to spectra from normal ovary tissues (healthy); clear blue to SBL; bark blue to HGSOc; and orange and red dots respectively to endometrioid and mucinous. **(D)** Results from the RIDGE-based classification for OC subtypes (healthy, SBL, HGSOc, mucinous, endometrioid). Classification report and confusion matrix were obtained on the training set for both training and 5-fold cross-validation sets. **(E)** RIDGE-based classification of the extended dataset comprising both the FFPE and FF tissues. Classification report and the matrix confusion were obtained on the training set for both training and 5-fold cross-validation sets. (See also figure S1-S4, Supp table 1-4)

The different models were then evaluated by analyzing blindly a validation cohort composed of 24 independent samples analyzed by a pathologist (**Supp Table 5**). For each of the 24 tissues, 3 different locations at least were analyzed randomly across the tissue for a total of 72 blind analyses over the cohort. For each blind analysis, the AMX software interrogate the previously built classification model (LDA) and provides the class within a few seconds as exemplified in **Figure 3A-B** for HGSOE and normal ovary. Firstly, blind analysis shows that HGSOE, SBL and normal tissues are correctly diagnosed but the model struggles to differentiate the endometrioid carcinoma from the mucinous (**Figure 3C**). More specifically, the 3 endometrioid tissues tested are mis-classified and 1 of the mucinous was only partially correctly classified. This can be explained by the lack of endometrioid tissues in the training cohort compared to the other subtypes. By removing the endometrioid class from the blind cohort, the correct classification goes up to 85%. Then, the model created with ridge classifier based on the mix between the FFPE and FF tissues was tested. This model allowed better prediction than the previous one but endometrioid tissue were still not correctly classified (**Figure 3C**). However, this model went up to 95.25% of correct classification by removing endometrioid tissues.



**Figure 3: Validation of the classification models in blind. (A-B)** Example of 2 patient tissues analyzed in blind and result after interrogation of the classification model. The first tissue (**A**) was identified to be a HGSOE and the second (**B**) to be attributed to normal ovary tissue (“healthy”). (**C**) Blind diagnostic based on the validation cohort analyzed (3 different spots / tissue) and compared to the models LDA, RIDGE and RIDGE for the extended dataset comprising the FFPE and FF tissues. Each correctly recognized sample are in green and failures in red and sample. (See also table 5)

To further enhance the performance of the model, stained tissues used for annotation were scanned and integrated with the MS spectra to create a dual-input dataset for deep learning. In order to achieve this, we developed a multi-input neural network model (**Figure 4A**) that combines a multi-layer perceptron (for molecular data) and a 2D-Convolutional neural network (2D-CNN) (for morphological data). The architecture and summary of the model are provided (**Figure S5**). This novel approach appears to yield a more sensitive and robust model. In fact, the model achieved 99% accuracy in negative ion mode after 5-fold cross-validation, and it also demonstrated 100% accuracy in blind diagnosis when tested on 40 spectra-images (**Figure 4B**). Furthermore, this model exhibited a sensitivity and specificity of 100% in both the validation and testing sets which is more than the molecular branch only (perceptron) that resulted in a accuracy 88% (**supp table 4**). The integration of histological information with molecular data significantly improves the ability to diagnose various OC epithelial carcinomas. This model is particularly compelling because it incorporates molecular information that was lacking in other models that solely focused on morphological features<sup>18</sup>.



**Figure 4: Multi-input model based on HPS staining and SpiderMass data.** (A) The multi-input model building process consists of two main steps: training the Scans data using CNN (Convolutional Neural Network) and the SpiderMass data using MLP (Multi-Layer Perceptron). Subsequently, the outputs of both models are concatenated to enable a unified

prediction of classes. The CNN is adept at capturing morphological patterns and features in images, making it suitable for analyzing scan data. The SpiderMass data is processed and trained using a Multi-Layer Perceptron (MLP). Unlike CNNs, MLPs are designed for processing structured data. After training both the CNN and MLP models, the outputs from the two networks are concatenated together. The concatenation process combines the extracted features from both the scan and SpiderMass data, creating a comprehensive representation that incorporates information from both sources. To make predictions, the concatenated feature representation undergoes a SoftMax activation function. The SoftMax function converts the concatenated features into probability scores, enabling the prediction of class labels. **(B)** The performance of the multi-input model was evaluated using the following metrics: confusion matrix and classification report. To ensure robustness and reliability, a 5-fold cross-validation technique was employed during the evaluation. The dataset was divided into five subsets (folds), and the model was trained and validated five times, with each fold serving as the validation set once and the remaining four folds as the training data. The final performance metrics were computed by averaging the results of all five folds. **(C)** The test of the model was performed on 40 blind spectra. The test set consists of these 40 unseen samples, which were not used during training or cross-validation. The accuracy of the model on this test set provides a reliable measure of how well the model can perform on new, previously unseen data. **(D)** Lime algorithm was used to determine the contribution, positively or negatively, of each peak in the differentiation of each class (red: negatively, green: positively). The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). **(E)** Example of a lipid marker, PE 18:2/18:0 ( $m/z$  742.55), differentially expressed between the different OC histotypes. This marker was validated by MALDI-MSI and LIME algorithm. Furthermore, a Kruskal-Wallis test was performed on the SpiderMass analysis data on this ion and represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (See also figure S5-S6).

### **Lipids biomarkers associated to the different OC subtypes**

To cross-validate the results from the classification by providing biological justification, we developed a complete data processing pipeline. In parallel to the classification model, the LIME algorithm (Local Interpretable Model-agnostic Explanations) was used to determine the positive or negative contribution of each  $m/z$  feature towards the classification of each cancer type (**Figure 4C**). In the pipeline,  $m/z$  features that contributes significantly to the prediction of a particular class can be considered as potential biomarkers, if their significance is confirmed through statistical testing. Furthermore, each identified potential biomarkers are cross-validated using MALDI-MSI by checking that these markers are localized in their corresponding specific histological tissue region. For example, ion  $m/z$  742.5, identified by MS<sup>2</sup> as PE 36:2 (18:0/18:2) was found to be more discriminative to HGSOE and EC. This was confirmed by the highly abundant distribution in these 2 histotypes by MALDI-MSI and the relative abundance variation

of this lipid in the SpiderMass data from the whole cohort (**Figure 4D**). Twenty-six lipids corresponding to potential biomarkers were identified by MS<sup>2</sup> (**Figure S6-Supp Table 2**). Discriminative lipids were observed for both ion modes (positive and negative) and presented a remarkable diversity of species. In the negative ion mode, saturated and unsaturated fatty acids, and several classes of glycerophospholipids (GP) such as glycerophosphoethanolamines (PE), glycerophosphoglycerols (PG), glycerophosphoserines (PS) and glycerophosphoinositols (PI) were observed (**Figure S6- Supp Table 2**). In the positive ion mode, mostly PE and PS were identified as differentially expressed. Compared to normal ovarian tissue, each OC histotype presents specific lipids. High relative abundance of  $m/z$  747.516,  $m/z$  725.5,  $m/z$  744.5, and  $m/z$  718.5 were observed in regions with high density of tumor cells in HGSOc. In the positive ion mode, PC species including PC 20:1, PC 18:4 are found at high relative abundances in MCs whereas PS and PE are more representative of HGSOc and SBL. Additionally, several lipids separate between HGSOc and SBL such as PI (22:1/18:0), TG 54:8 16:0/18:4/20:4), PE (18:0/22:4) in SBL *versus* PE 34:0 (18:0/16:0), PA 38:3 (18:1/20:2), PE 36:1 (18:0/18:1) in HGSOc. These results are in line with these previously obtained in DESI-MS for the HGSOc<sup>19</sup>, Masspec Pen<sup>20</sup> and MALDI-MSI<sup>21</sup> (**supp table 2**). For example, normal stromal ovarian tissue displayed a lower overall signal intensity for lipid species when compared to HGSOc tissue in the negative ion mode in DESI-MS (Sans et al., 2017). Moreover, a higher relative abundance of  $m/z$  885.55 attributed to PI 38:5 (20:4/18:1) in comparison to  $m/z$  788.55 corresponding to PS 36:1 (18:0/18:1) is observed in HGSOc when compared to normal ovarian tissues (Sans et al., 2017). In the MasSpec Pen study, glycerophosphoethanolamine PE 38:4 (18:0/20:4) ( $m/z$  750.55), glycerophosphoserine PS 36:1 (18:0/18:1) ( $m/z$  788.55), and glycerophosphoinositol PI 38:4 (18:0/20:4) ( $m/z$  885.55) were detected in the HGSOc (Sans et al., 2019). Comparison between HGSOc and LGSC for the MasSpec Pen also revealed a higher relative abundance of plasmalogen PE species, such as  $m/z$  722.51 and  $m/z$  750.55 in the LGSC, while glycerophosphatidic acid PA 36:1 (18:0/18:1) ( $m/z$  701.51) were at higher abundances in HGSOc<sup>20</sup> which is also in line with our findings. Regarding the nature of the lipids identified in the different OC subtypes, it is interesting to note that OC tissues contained more PS, PA, and PE than normal tissue as it was already shown for other cancers such as NSCLC, melanoma and prostate cancers<sup>22,23</sup>. Our hypothesis is that PS lipids may have a strong effect on the mechanical properties of the mixed PS/PC bilayers. An increase in the PS content would increase the interfacial tension and stiffen the bilayer contributing to a change in the mechanical property of the cancer cells and hence drives cancer cell migration<sup>24</sup>. Dysregulation of PS in cancers is also linked to its exposure at the surface of tumor cells and immunosuppression<sup>25</sup> (Stoica et al., 2022). Furthermore, the exposition of PS in cancer cells protect them from NK activity and others cytotoxic immune cells<sup>26</sup>. Fatty acid remodeling of phospholipids may be an adaptive response to the acidic microenvironment observed in most tumors<sup>27</sup> and cancer cells require these membrane changes for rapid cell proliferation leading to an increased phospholipid metabolism. Indeed, increased phospholipid membrane expression renders the cancer cell membranes less susceptible to peroxidation- and oxidative stress-mediated cell death. Besides, modifications in the phospholipid levels can also alter crucial cellular signaling pathways, e.g., cell proliferation and survival, and promote tumorigenesis<sup>28</sup>. Finally, it seems that PA are more abundant in several cancer types compared to normal tissues. This correlates with the findings of<sup>29</sup> which explains the role of PA on the activation of kinase

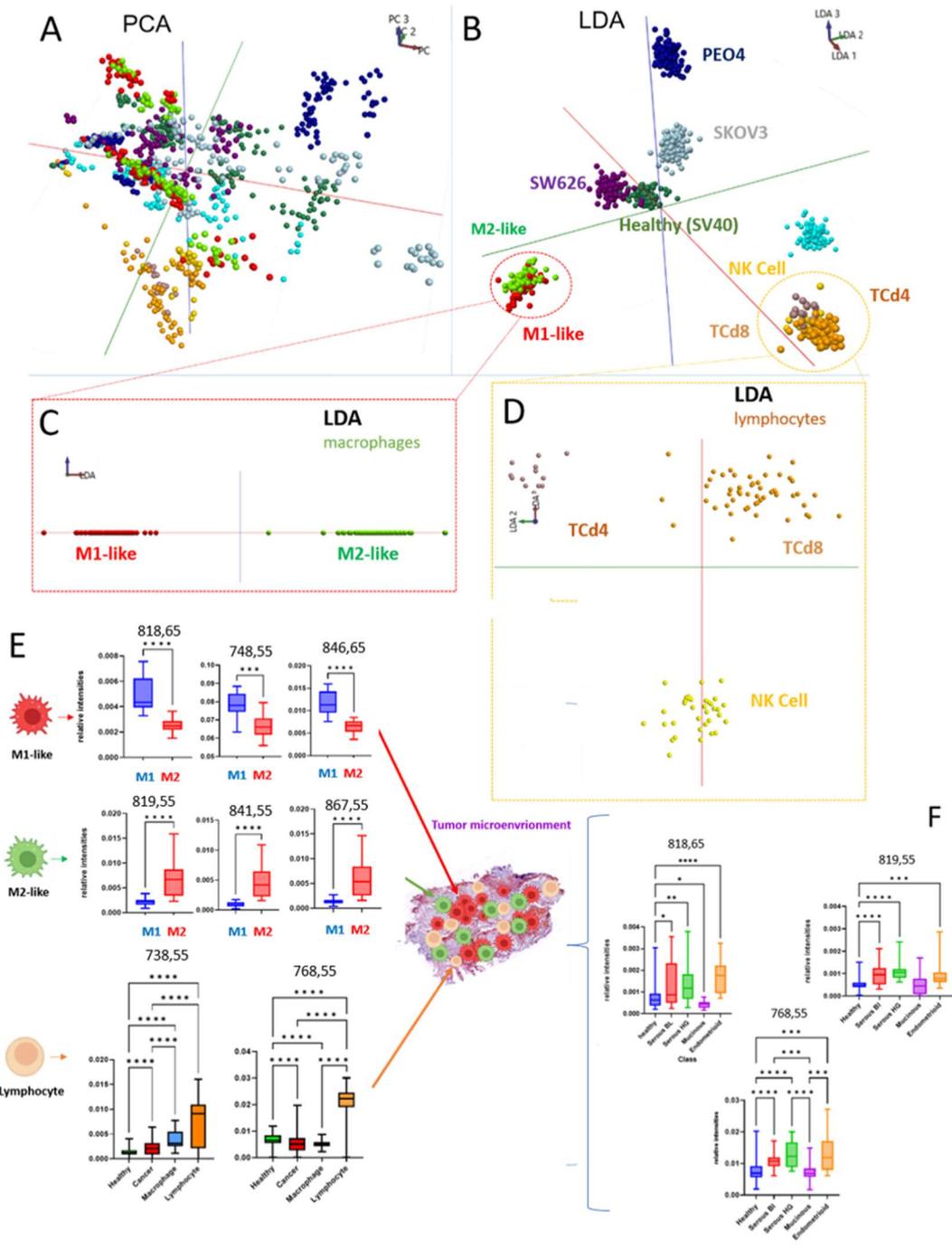
such as MAPK, which is known to be involved in stress signaling pathway. Altogether this shows that the real-time molecular diagnostic reflects the underlying change of lipidic metabolism in the different cancer subtypes. The different abundancy of some phospholipids is closely linked with the tumoral microenvironment (TME) progression and the inflammatory system. Consequently, we decided to investigate more this TME to better understand its impact for the diagnostic of all patients.

### **Deciphering the TME by SpiderMass**

The TME is known to play a key role in cancer progression, survival, and migration<sup>30</sup>. To directly distinguish by SpiderMass the most represented cell-types within in the TME we had to first get the molecular profile of these different cells. Thus, a SpiderMass analysis was performed on different cancer cell lines (SKOV3, POE4, PA1, SW626), ovarian epithelium cell line (referred as healthy), dedifferentiated macrophages cell lines, sorted primary macrophages (M1-like and M2-like) and lymphocytes (NK, CD4, CD8) to get their specific molecular signature. A non-supervised multivariate statistical analysis by PCA (**Figure 5A**) and then a LDA classification model (**Figure 5B**) were then realized from all these data. The different cell-type already separate from each other in the PCA model with some groups of cells visible with 60.6% of variability explained by the 3 first PCs. This resulted into a LDA model of 96.83% correct classification without outliers and 92.48% with outliers representing only 43 outliers over 957 spectra (**Figure 5A**). The few misclassifications are caused by the macrophage subpopulation discrimination. Indeed, 10% of all the macrophage data were miss-assigned between the M1- and M2-like subclasses. However, by mixing the two phenotypes on the “macrophage” class this model goes up to 99.78% et 95.30% with and without outliers respectively. Furthermore, by isolating only the macrophages MS spectra to build another PC-LDA model, it was possible to discriminate them by their M1-like or M2-like phenotype with 100% good classification (**Figure 5C**). Moreover, macrophages are well separated from the other immune cells (lymphocytes) and even more separated from the cancer or normal cells (**Figure 5A**). Curiously, 2 lipids at  $m/z$  818.65 and  $m/z$  846.65 were found to be highly abundant in the macrophages MS spectra and significantly different from the ovarian cancer cell lines (**Figure S7A**) which were identified by MS<sup>2</sup> as GlcCer d40:1 (d18:1/22:0) and GlcCer d42:1 (d18:1/24:0) respectively (**Figure S7B**). These lipids were found to be specific to both, the THP1 cell line and the primary macrophages sorted from blood, after dedifferentiation into M1- and M2-like. Interestingly, glucosylceramides are known to be involved in inflammation as regulation of immune system. In fact, these lipids have been studied for their ability to reduce the inflammatory response upon bacterial infection<sup>31</sup>.

Additionally, others markers were found to be differently expressed in primary macrophages according to their phenotypes (M1- and M2-like). For M2-like, significant increase in relative abundances were found for  $m/z$  819.55,  $m/z$  841.55,  $m/z$  843.55 and  $m/z$  867.55 identified respectively as PG 40:7 (18:1/22 :6), PG 42:10 (22 :6/20 :4), PG 42:9 (20:3/22:6) and PI 0-28:6 (0-16 :0/22 :6) (**Figure 5E**). On the other hand, M1-like were highly discriminated by  $m/z$  818.65 and  $m/z$  846.65 corresponding to the previously identified GlcCer d40:1 (d18:1/22:0) and GlcCer d42:1 (d18:1/24:0) as well as  $m/z$  748.55 which were already abundant in normal ovarian tissues and identified as a PE 38:5 (16:0/22:5). Thus, a relative increase in polyunsaturated fatty acid PGs were observed for M2-like macrophages. Previously, Zhang et al. have studied smaller fatty acid chain PGs in mice and human cell line and showed a lower concentration of these lipids on M2 macrophages<sup>32</sup>. This indicates that M2-like macrophages would present both a higher concentration of small LysoPGs and PGs with long fatty acid chains. Besides, other studies demonstrated the colocalization of polyunsaturated PG with

inflammatory cells in cancer<sup>33</sup>. For lymphocytes, different lipids were found to separate the LT populations such as *m/z* 768.55, *m/z* 885.55 and *m/z* 738.55 which correspond to PE (38:3), PI 38:4 (20:4/18:0) and PE 36:4 (20:4/16:0) respectively. Furthermore, discrimination of subpopulation of lymphocytes (NK, CD8 and CD4) was possible thanks to the creation of a PC-LDA model containing only lymphocytes MS spectra. This model separates all lymphocyte phenotypes with 100% good classification (**Figure 5D**). This leads to identify *m/z* 752.55 PE 0-38:4 (0-18:0/20:4) and *m/z* 794.55 PE 40:4 (18:0/22:4) as specific of NK cells, and *m/z* 722.55 PE P-36:4 (P-16:0/20:4) and *m/z* 885.55 PI 48:4 (20:4/18:0) of CD4 (**Figure S8**). The intensities of each of these lipids found to be discriminative of the different immune cell subpopulations were collected from the entire OC cohort to spot their abundance across the different OC histotypes (**Figure 5E**). Interestingly, the abundance of *m/z* 818.65 and 819.55 which are specific of macrophages indicates a higher expression in the HGSOC notably while they have a very limited expression in the normal ovary tissues indicating a potential correlation with the aggressiveness of the cancer. However, EC also present a significant quantity of macrophages despite its lower aggressiveness. To better understand this observation, we analyzed an endometrium cohort of 79 patients. By checking these specific macrophage lipids in healthy endometrium tissues, we deduced that macrophages were already resident in normal endometrium. In fact, normal ovarian tissues do only present a few signals of macrophages (**Figure S9A**). Of note, the abundance of these signals associated to macrophages is significantly higher in EC (**Figure S9B**). Even if macrophages have a key role in the ovary homeostasis<sup>34</sup>, endometrium is more subjected to inflammation. Interestingly, it was recently demonstrated that according to their origin macrophages can promote endometriosis which is inflammatory<sup>35</sup>. This can be explained by the function of the endometrium in cyclical changes, including proliferation and shedding of the uterine lining which creates favorable environment to inflammation. For example, during menstruation, the expel of endometrium tissue can trigger a natural inflammatory response. Moreover, EC usually originate from endometriosis which is exemplified in our study. This explains the increased number of macrophages in EC by comparison to other types such as HGSOC which originates from the fallopian tube and is therefore less exposed to inflammation or potential infections than the endometrium being not directly in contact with the external environment. Interestingly, the *m/z* 768.55 which is associated with lymphocyte shows the same distribution as the 2 lipids from macrophages showing a dependance between lymphocyte and macrophage abundancies.

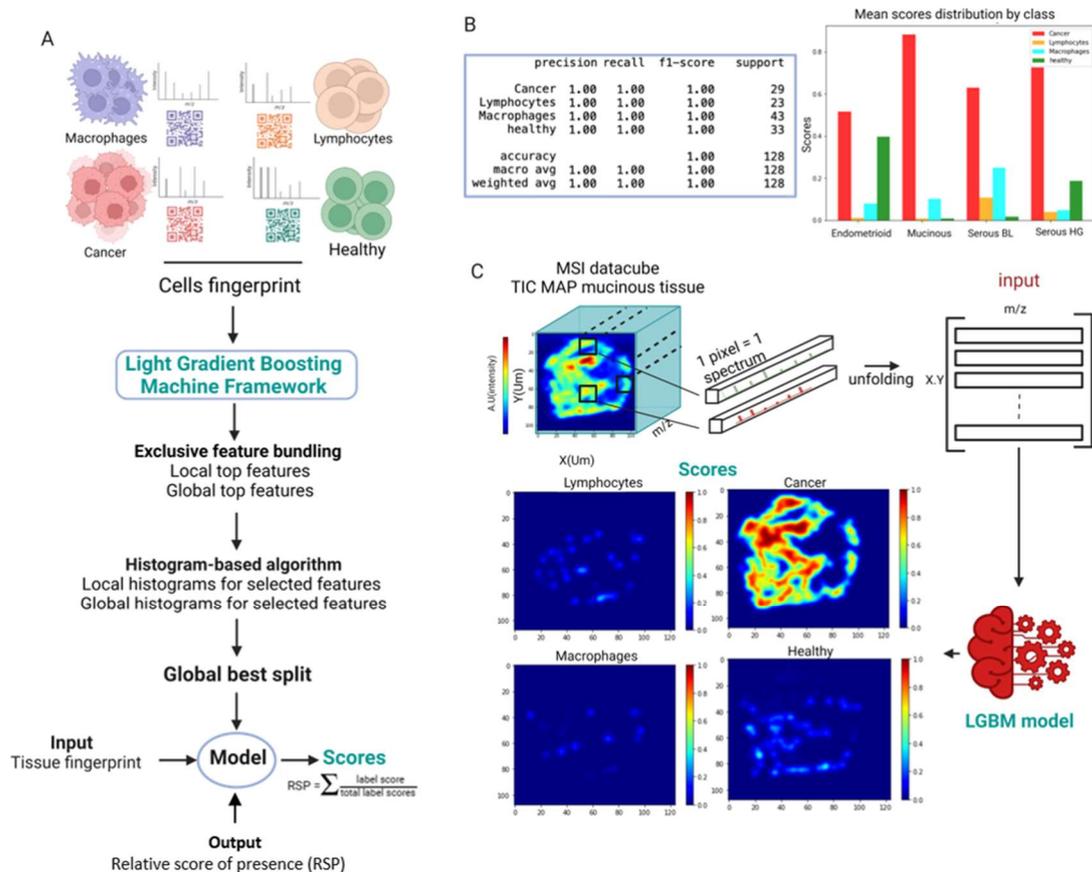


**Figure 5: Classification of the different immune cell types and their associated specific markers.** (A-B) PCA (A) and LDA (B) models of the different ovarian cancer cell lines and immune cell types (M1-like, M2-like, NK, CD4, CD8) with 0.1Da binning and normalization of MS spectra. (C) LDA model of M1-like versus M2-like macrophages. The model combines data from THP1 cell line and primary cells after isolation. (D) LDA model of primary lymphocytes including NK, CD4 and CD8. (E) Boxplots representing the relative intensities of the *m/z* 742.55 (PE 36:2 (18:2/18:0)) found to be markers of the immune cells (M1, M2 and lymphocytes) with the Kruskal-Wallis statistical test. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (F) Boxplot representing the relative intensity of these same *m/z* within the OC cohort as a function of cancer subtypes. (See also figure S7-9)

### **Immuno-scoring based diagnostic and prognostic by SpiderMass-MSI**

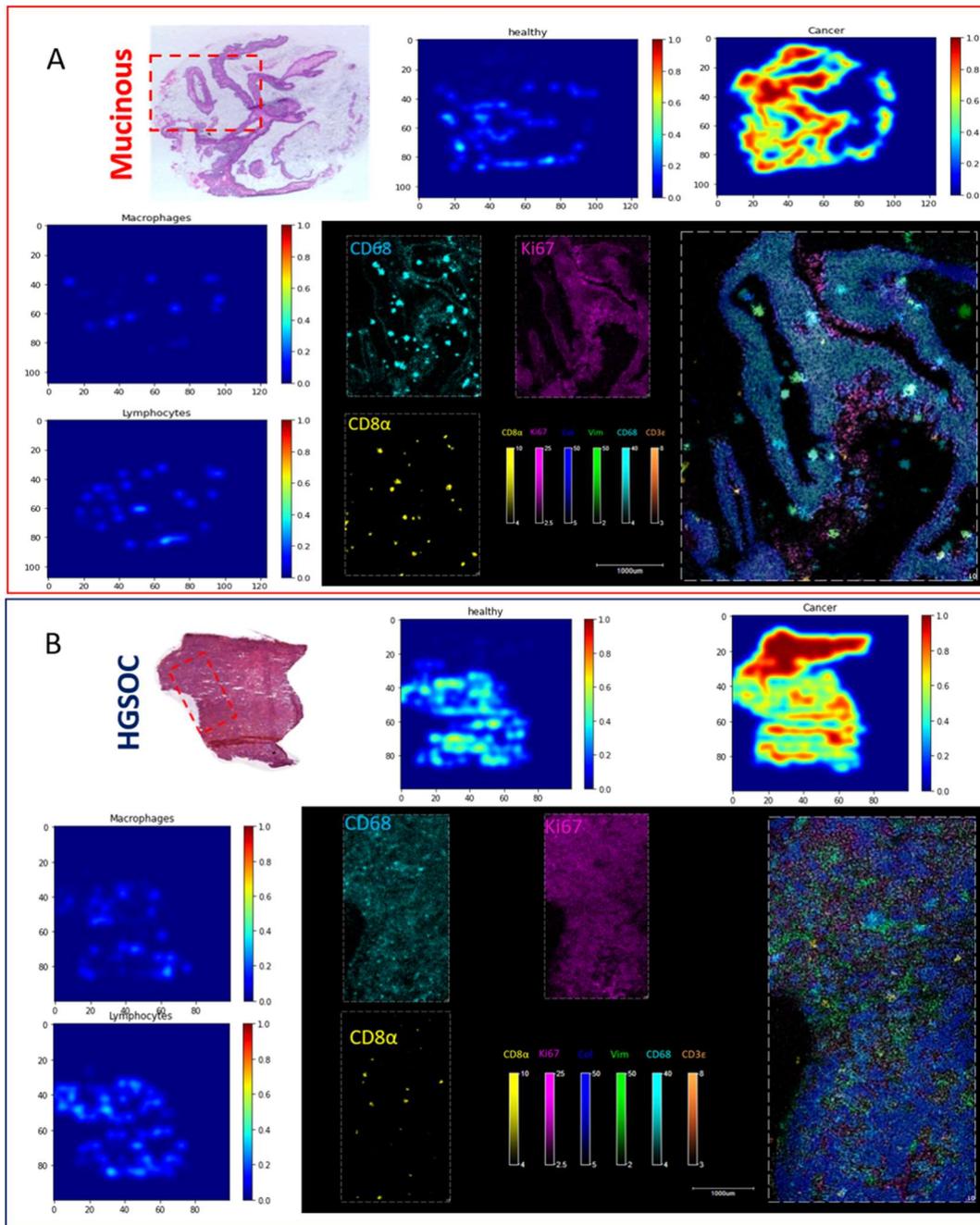
SpiderMass MS analysis provides a fast and direct way to retrieve molecular informations whose could be translated to intraoperative analysis. To meet the need of the surgery, the system is designed to analyze areas of tissue of 400  $\mu\text{m}$  per position. The MS spectrum collected after a single analysis therefore corresponds to a potential mix in the molecular profiles of several cell types constituting the TME. However, we wanted to get more information on the TME composition, and more specifically, on the presence of infiltrated immune cell subpopulations versus the other cell types in the different tissue areas analyzed. To this end, we developed a process to learn from the global molecular fingerprint of the different cell types and predict their probability of presence in the tissues. An immune score model was, thus, trained using the LGBM (Light Gradient Boosting Machine) Python library, a gradient boosting framework developed by Microsoft<sup>36</sup>. This model leveraged MS spectra of the different cells within the  $m/z$  [600-1100] range in the negative ion mode (**Figure 6A**) leading to a successful discrimination with an accuracy of 100%. This model was trained from all the cell line and primary immune cells data previously analyzed and was tested on the OC cohort. By adding a relative score of presence (RSP) to this new effective classification model, all the tissues were given a probability of presence for each cell type enabling tissues to be discriminated according to their immune score (**Figure 6B**). This highlights different cellular signatures according to the cancer subtype. For example, MC present a higher probability of presence of cancerous cells while SBL show a higher probability of presence of LT and macrophages whereas EC shows higher probability of presence of normal cells. The latter could be linked to the better prognosis of EC.

To understand more globally the cellular composition of the different tissues, SpiderMass-MS Imaging (SpiderMass-MSI) was performed on 3 samples for each class to allow the visualization of the different immune cell distribution in the tissues (**Figure 6C**).



**Figure 6: Processing workflow to create an immune score based on the SpiderMass MS images. (A)** A LGBM model structure was built based on cell fingerprint. All cells were analyzed separately and the top and excluding features of each cell type were used to discriminate them. The MS spectra of a tissue is then interrogated with LGBM to define a score of presence of each cell types. **(B)** Table representing the performance of the model. Histogram of the mean of each cancer-type analyzed by SpiderMass based on the LGBM model to define the cellular population. **(C)** Example of a SpiderMass MSI visualization based on immune score. Each pixel (corresponding to a mean spectrum) is tested through LGBM model to give the score of presence of cancer, lymphocyte, macrophages and healthy cells. To better visualize the presence of each class, an image is built for each class label.

This was performed after image acquisition by unfolding and processing each pixel of each image (corresponding to one medium spectrum) using the LGBM model to recreate another image based on the LGBM results. These scores were calculated by summing the scores provided for each cell type and dividing the sum by the total scores across all labels (RSP). To validate the distributions based on the LGBM predictions, we also performed multiplex MALDI MS-based immunohistochemistry (MALDI-IHC<sup>37,39,41</sup>) using CD68, CD8 and Ki67 markers which are specific targets of macrophages, lymphocytes, and proliferation of cancer cells respectively (**Figure 7**).

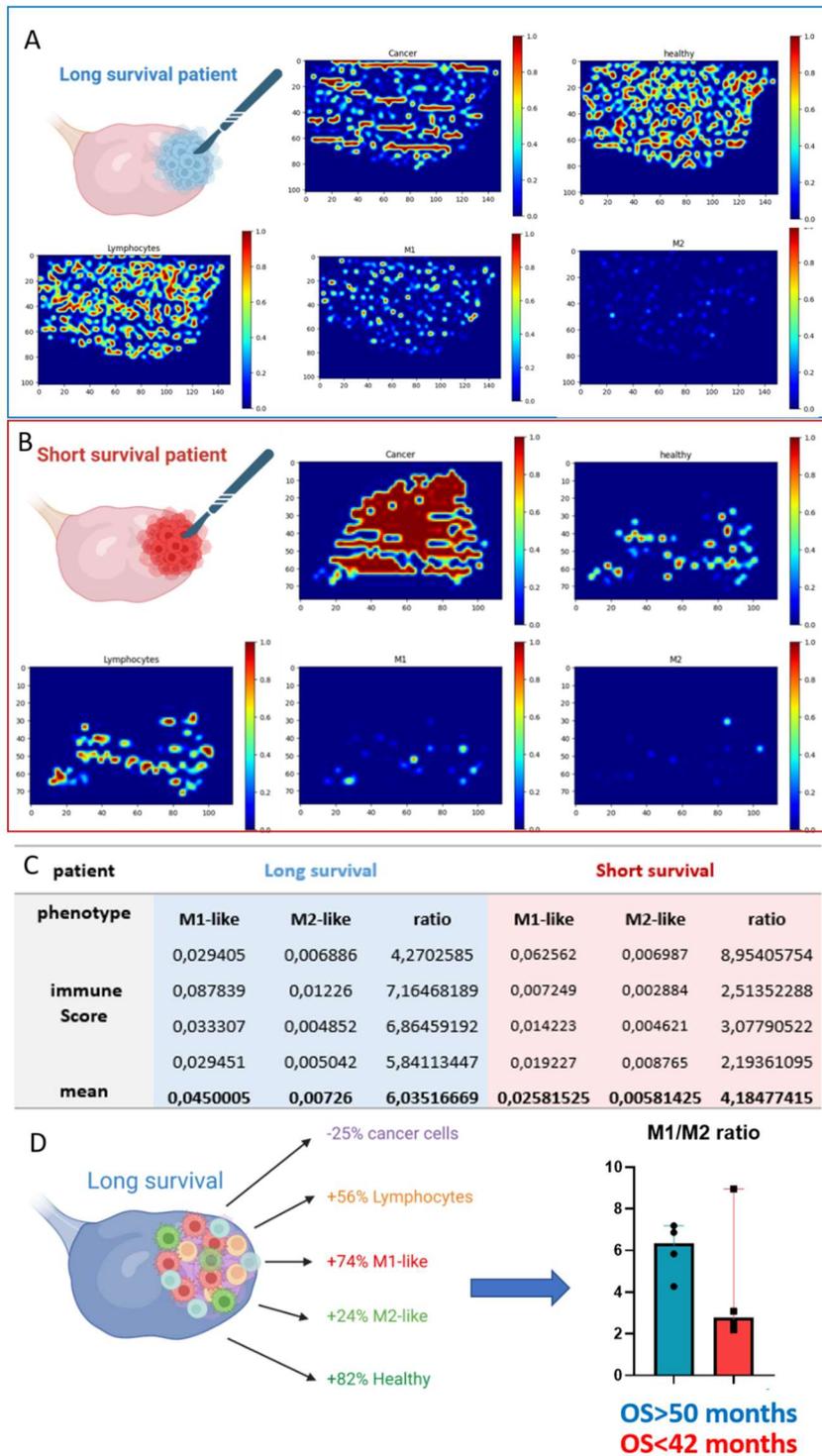


**Figure 7: SpiderMass-based immune score and its validation by MALDI-IHC for the diagnosis of OC. (A)** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTC8) and anti-CD68 (macrophages). **(B)** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTC8) and anti-CD68 (macrophages). (See also figure S10-S22)

As expected, a difference in the probability of presence of macrophages and lymphocytes is observed according to the histotype of OC (**Figure S10-22**). SBL and HGSOC show a quite homogenous spatial distribution of macrophages across all the tissue with some islets of macrophages with higher abundance (**Figure S13-15,16-17**). In contrast, EC and MC show more heterogenous distributions of macrophages with islets of highly intense signal while sparingly present in the surrounding tissues (**Figure S10-18,19**). Overall, each tissue shows about the same amount of cancer cells. However, the abundance of macrophages and lymphocytes across the different tissues is different. Indeed, HGSOC and EC show higher abundance of macrophages. Interestingly, the lymphocyte signature of the different tissues is correlated with what was previously shown for the abundance of the ion at  $m/z$  768.55 (**Figure 5E**). The distribution of the different cell population obtained by SpiderMass combined with LGBM was confirmed by the MALDI-IHC, although a higher spatial resolution (20  $\mu$ m) was used for MALDI-IHC. This approach offers a faster way to diagnose OC based on the immune cell presence in the TME. Interestingly, the SpiderMass-MSI based on immune score allows for the distinction of the different subpopulation present in the TME without the need of any technique based on the use of probes such as IHC.

Since the immune-score MSI made possible to decipher the different OC histotypes, we also investigated using the HGSOC tissues its correlation with patient survival. To dig deeper into the prognostic potential of the SpiderMass immune-scoring, the macrophage signature was subdivided into the M1- and M2-like components. Previous work has highlighted the relation between patient prognosis and the ratio of M1- versus M2-like macrophages, which are known to be respectively pro-inflammatory and pro-tumoral<sup>38,40,42</sup>. Moreover, immuno-scoring based on IHC was shown in several cancers to be a robust test for patient prognosis<sup>43,44</sup>. SpiderMass immune-score was achieved from 8 different HGSOC patients divided into 2 groups (**Supp Table 5**). The first group is composed of patients without recurrence and with OS >50 months being followed up. The second is composed of patient with recurrence and OS < 42 months. To avoid any bias from the therapy received by the patient, we performed this analysis only from surgical pieces excised before chemotherapy. The result shows a real difference in score of probability of presence of macrophages (both M1 and M2), lymphocytes and cancer cells. Tissues from patients with longer survival shows a high abundance of both lymphocytes and M1-like macrophages (**Figure 8A-S23**) compared to patients with shorter survival which show a higher score of cancer cells and M2-macrophages (**Figure 8B-S24**)(**Supp table 6**). The score obtained for M1- and M2-like macrophages for each patient are presented (**Figure 8C**). We also calculated the ratio of M1/M2 for each patient to compare our results with what is known from the literature. A mean ratio of 6 was obtained for the patients with longer survival and 4 for those with shorter survival. For 3 out of the 4 patients with short survival, the calculated ratio falls between 2 and 3. The only one patient shows an abnormal ration that was close to 9. Regarding the mean RSP for the different cell population of the TME between shorter versus longer survival we find an increase of 25% of the probability of presence of cancer cells for the patients with the shorter survival, while we have an increase of 56% of the lymphocyte infiltration, a huge increase of 74% in the M1-like infiltration and finally more normal cells

(82%) in the longer-term prognosis (**Figure 8D**). This difference is clear when comparing the ratio of M1/M2. This illustrates that better prognosis is associated with a high infiltration of pro-inflammatory macrophages. Our results agree with earlier published histopathological data, showing OC includes 39% of macrophages which are divided into subclassifications (25% of M0, 20% of M1, 50% of M2) macrophages and 4% of monocytes<sup>45</sup>. In fact, we detected 21.44% on average of Macrophages (M1, M2) in OC in our experiments (**Supp. Table 7**).



**Figure 8: SpiderMass-based immune score and for the prognosis of OC. (A)** Example of the immune score MSI found in a long-time survival patient based on cancer, normal ovary tissue (“healthy”), lymphocyte, M1- and M2-like macrophages. **(B)** Example of the immune score MSI found in a short time survival patient based on cancer, healthy, lymphocyte, M1 and M2-like macrophages. **(C)** M1- and M2-like macrophages scores obtained for each tissue and their M1/M2 ratio. **(D)** Compared M1/M2 ratio for patient with OS <42 months and those with OS >50 months. Patient with longer survival show higher M1/M2 ratio. (See also figure S23-S24, supp table 5-6).

To go deeper in the understanding of the TME, the goal is to train this model on additional different cell type such as monocyte, M0, Tumor infiltrated macrophages (TAM), Treg and tumor infiltrated lymphocytes (TIL). This training will avoid the potential miss-detection of immune cells. Furthermore, the prediction for the mean ratio of lymphocytes/M2 was 32.10 for patient with OS > 50 months compared to 24.03 from patient with OS<42 months. This shows that there is more lymphocyte infiltration in patients with OS>50 compared to the OS<50. This is linked with what is known from onco-immune response with an important immunosuppressive environment in poor prognosis tumor. This prognostic was successful only on patients who did surgery as the first treatment. Indeed, when comparing the immune score for the patients who received neo adjuvant chemotherapy, the number of macrophages and lymphocytes detected were drastically decreased leading to uncertain ratio of both Lymphocyte/Macrophages and M1/M2. Chemotherapy for OC is an aggressive treatment with systemic administration that is not only targeted to cancer cell but also affects immune cells of the TME. Accordingly, the effect of the chemotherapy on the immune cells is clearly observed in our study. Nonetheless, as previously described we can use the ratio of M1/M2 and lymphocyte/M2 to prognose patients which can help clinicians to determine the profile of a patient and adapt the treatment and the surgery based on the result. Considering that the SpiderMass is capable of direct analysis and is non-invasive, this scoring could be performed in real-time during the surgery for the surgeon to achieve precision surgery. While this model is already simple it can be further improved by increasing the number of different immune cell subpopulation measured in the TME but as presented represents the first approach to patient prognostic during the surgery.

## CONCLUSION

Accurate diagnostic and prognostic methods are paramount to the patient management with OC considering the generally late discovery of this cancer in patients. Getting an accurate diagnostic and prognostic by the time of surgery, intraoperatively, directly in real-time would greatly help the surgeon in providing a more tailored surgery and patient management. Here, we have assessed the ability of a recently MS-based developed technology, the SpiderMass, to achieve such a goal by the combination of the collection of non-target molecular data to appropriate machine learning processing. We have thus conducted a study from a prospective and retrospective cohort of patients with OC representative of the most prevalent subtypes of epithelial OC with the SpiderMass technology to train and test the technology. We can show correct classification of 100% in blind after training and 97% after validation. Higher performances with a 100% accuracy were obtained using a dual input model which couples molecular and morphological information. We also were interested in providing biological interpretation to our findings and the lipids differentially regulated between the different OC subtypes were identified by MS<sup>2</sup>. Interestingly we found that the lipids markers found were mostly PS and PA. These phospholipids are related to tumor growth. In fact, presence of PS in inner membrane is related to immunosuppressive response and PA stimulate MAPK which is evolved in the stress signaling pathway. Furthermore, these discriminative lipids were in line with other studies using DESI or the MassPec Pen. To get a better description of the TME considering its importance

for diagnostic and prognostic cancer, we developed a workflow based on the LGBM algorithm to predict the quantity of the different cell populations in each analytical pixel. This was achieved by training the algorithm on the lipid molecular profile of the different cell population analyzed separately either from cell cultures or from primary cells sorted by FACS (CD4, CD8, M1, M2, NKT, normal cells, cancer cells). This led to a new way to differentiate tissue areas presenting cancer cells based on the immune cell distribution within the tissue calculated from the SpiderMass MS Imaging data. Finally, we went deeper into the analysis by testing this model for prognosis on the HGSOC patients. The results showed an increased quantity of lymphocytes and M1-macrophages infiltrations for patients with a longer OS. As some OC patients have a shorter survival without recurrence, immune-score imaging opens up a new way to tailor the treatment and better manage these patients. The immune scoring from the SpiderMass shows a strong agreement with the literature and was validated by MALDI-IHC<sup>37,39,41</sup>. So far, our study was conducted from tissue section post-surgery, but we wish in the near future to translate such a data processing to data collected in real-time during the surgery directly on the patients enabling for real-time immune-scoring to emerge.

## Abbreviation

CCC	Clear cell carcinoma
CD8	Lymphocyte TCD8+
CD4	Lymphocyte TCD4+
Ki67	Proliferation marker protein Ki-67
CD68	Macrosialin protein in the lysosome of macrophages
EMT	Epithelial-mesenchymal transition
FFPE	Formalin Fixed and Paraffin-Embedded
FTE	Fallopian tube epithelium
HGSOC	High Grade Serous Carcinoma
IHC	Immunohistochemistry
LDA	Linear discriminant analysis
MALDI	Matrix assisted Laser Desorption Ionization
MSI	Mass spectrometry imaging
M1	Macrophages M2-like phenotype
M2	Macrophages M2-like phenotype
MC	Mucinous ovarian cancer
NK	Natural killers
OC	Ovarian cancer
OS	Overall survival
PA	Phosphatidic Acid
PC	Phosphatidyl choline
PCA	Principal component analysis
PE	Phosphatidyl ethanolamin
PI	Phosphatidyl inositol
PS	Phosphatidyl serine

SBL	Serous borderline carcinoma
TF	Transcription factor
TG	Triglyceride

**Competing interest:** L.R., Y.Z., L.L., B.M., M.D., F.N., D. H., Y-M.R., C.P., A-S.L., D.B., E.L. declare they have no competing interest in this work. M.S. and I.F. are inventors on a patent (priority number WO2015IB57301 20150922) related to part of the described experimental methods. 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.

**Acknowledgements:** 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.R. PhD was funded by University of Lille. 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:** I.F., E.L., N.D. and M.S. designed the in experiments. IF and MS supervised the study. L.R., Y.Z., L.L. performed the experiments. The pathology analyses were done by YMR, DB, ASL, CP and the clinical interpretation was conducted by D.H. and E.L. G.Y., K.J.R. and M.J.L. provided the Miralys<sup>TM</sup> probes and protocols. L.R., Y.Z. performed the data analyses. I.F., M.S., Y.Z. and L.R. wrote the manuscript with contributions from all co-authors. I.F. and M.S. funded the project.

## FIGURE LEGENDS

**Figure 1: Overall workflow developed in the study.** A robust classification model based on the combination of H&E staining and MS spectra by SpiderMass was created for the diagnosis of ovarian cancer. A second model was also built to detect the presence of the different cell types in the TME (including the immune cells). This model was applied to MS Imaging data to predict the presence of the immune cells across the tissue and create an immune score for prognosis.

**Figure 2: Multivariate statistical analysis-based model for ovarian cancer diagnosis. (A-B)** Morphological and spectral fingerprint of the different OC subtypes. **(A)** MS spectra recorded in the negative ion mode with a 4mJ/shot burst of 10 shots are presented for each OC subtype. **(B)** HPS staining of these specific tissues with the spot used for acquisition. **(C)** LDA classification model based on the MS spectra recorded in the negative ion mode from an OC cohort of 79 samples. Green dots correspond to spectra from normal ovary tissues (healthy); clear blue to SBL; bark blue to HGSOC; and orange and red dots respectively to endometrioid and mucinous. **(D)** Results from the RIDGE-based classification for OC subtypes (healthy, SBL, HGSOC, mucinous, endometrioid). Classification report and confusion matrix were obtained on the training set for both training and 5-fold cross-validation sets. **(E)** RIDGE-based classification of the extended dataset comprising both the FFPE and FF tissues. Classification report and the matrix confusion were obtained on the training set for both training and 5-fold cross-validation sets. (See also figure S1-S4, Supp table 1-4)

**Figure 3: Validation of the classification models in blind. (A-B)** Example of 2 patient tissues analyzed in blind and result after interrogation of the classification model. The first tissue **(A)** was identified to be a HGSOC and the second **(B)** to be attributed to normal ovary tissue (“healthy”). **(C)** Blind diagnostic based on the validation cohort analyzed (3 different spots / tissue) and compared to the models LDA, RIDGE and RIDGE for the extended dataset comprising the FFPE and FF tissues. Each correctly recognized sample are in green and failures in red and sample. (See also table 5)

**Figure 4: Multi-input model based on HPS staining and SpiderMass data. (A)** The multi-input model building process consists of two main steps: training the Scans data using CNN (Convolutional Neural Network) and the SpiderMass data using MLP (Multi-Layer Perceptron). Subsequently, the outputs of both models are concatenated to enable a unified prediction of classes. The CNN is adept at capturing morphological patterns and features in images, making it suitable for analyzing scan data. The SpiderMass data is processed and trained using a Multi-Layer Perceptron (MLP). Unlike CNNs, MLPs are designed for processing structured data. After training both the CNN and MLP models, the outputs from the two networks are concatenated together. The concatenation process combines the extracted features from both the scan and SpiderMass data, creating a comprehensive representation that incorporates information from both sources. To make predictions, the concatenated feature representation undergoes a SoftMax activation function. The SoftMax function converts the concatenated features into probability scores, enabling the prediction of class labels. **(B)** The performance of the multi-input model was evaluated using the following metrics: confusion

matrix and classification report. To ensure robustness and reliability, a 5-fold cross-validation technique was employed during the evaluation. The dataset was divided into five subsets (folds), and the model was trained and validated five times, with each fold serving as the validation set once and the remaining four folds as the training data. The final performance metrics were computed by averaging the results of all five folds. **(C)** The test of the model was performed on 40 blind spectra. The test set consists of these 40 unseen samples, which were not used during training or cross-validation. The accuracy of the model on this test set provides a reliable measure of how well the model can perform on new, previously unseen data. **(D)** Lime algorithm was used to determine the contribution, positively or negatively, of each peak in the differentiation of each class (red: negatively, green: positively). The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). **(E)** Example of a lipid marker, PE 18:2/18:0 ( $m/z$  742.55), differentially expressed between the different OC histotypes. This marker was validated by MALDI-MSI and LIME algorithm. Furthermore, a Kruskal-Wallis test was performed on the SpiderMass analysis data on this ion and represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (See also figure S5-S6).

**Figure 5: Classification of the different immune cell types and their associated specific markers.** **(A-B)** PCA **(A)** and LDA **(B)** models of the different ovarian cancer cell lines and immune cell types (M1-like, M2-like, NK, CD4, CD8) with 0.1Da binning and normalization of MS spectra. **(C)** LDA model of M1-like versus M2-like macrophages. The model combines data from THP1 cell line and primary cells after isolation. **(D)** LDA model of primary lymphocytes including NK, CD4 and CD8. **(E)** Boxplots representing the relative intensities of the  $m/z$  742.55 (PE 36:2 (18:2/18:0) found to be markers of the immune cells (M1, M2 and lymphocytes) with the Kruskal-Wallis statistical test. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). **(F)** Boxplot representing the relative intensity of these same  $m/z$  within the OC cohort as a function of cancer subtypes. (See also figure S7-9)

**Figure 6: Processing workflow to create an immune score based on the SpiderMass MS images.** **(A)** A LGBM model structure was built based on cell fingerprint. All cells were analyzed separately and the top and excluding features of each cell type were used to discriminate them. The MS spectra of a tissue is then interrogated with LGBM to define a score of presence of each cell types. **(B)** Table representing the performance of the model. Histogram of the mean of each cancer-type analyzed by SpiderMass based on the LGBM model to define the cellular population. **(C)** Example of a SpiderMass MSI visualization based on immune score. Each pixel (corresponding to a mean spectrum) is tested through LGBM model to give the score of presence of cancer, lymphocyte, macrophages and healthy cells. To better visualize the presence of each class, an image is built for each class label.

**Figure 7: SpiderMass-based immune score and its validation by MALDI-IHC for the diagnosis of OC.** (A) Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (B) Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (See also figure S10-S22)

**Figure 8: SpiderMass-based immune score and for the prognosis of OC.** (A) Example of the immune score MSI found in a long-time survival patient based on cancer, normal ovary tissue (“healthy”), lymphocyte, M1- and M2-like macrophages. (B) Example of the immune score MSI found in a short time survival patient based on cancer, healthy, lymphocyte, M1 and M2-like macrophages. (C) M1- and M2-like macrophages scores obtained for each tissue and their M1/M2 ratio. (D) Compared M1/M2 ratio for patient with OS <42 months and those with OS >50 months. Patient with longer survival show higher M1/M2 ratio. (See also figure S23-S24, supp table 5-6).

## SUPPLEMENTARY FIGURE LEGENDS

**Figure S1: Morphological and spectral fingerprint of ovarian cancer subtypes in positive ion mode.** (A) Mean Spectra for each ovarian cancer subtype in negative ion mode (4mJ/shot, burst mode, 10 shots, 1s/spectrum). (B) HPS staining of the corresponding histological sections. The arrows indicate the location where the laser was fired. (Related to figure 2)

**Figure S2: Example of annotated MS/MS spectra used for lipid identification.** (A)  $m/z$  885.5 corresponding to PI 38:4 (20:4/18:0). (B)  $m/z$  718.55 corresponding to PE 34:0 (18:0/16:0). (C)  $m/z$  744.55 corresponding to PE 36:1 (18:0/18:1). (D)  $m/z$  788.55 corresponding to PS 36:1 (18:0/18:1). (Related to figure 2)

**Figure S3: Classification models and their cross-validation for OC subtypes in the positive ion mode.** (A) Linear discriminant analysis model for the different OC subtypes. (B) Cross-validations of the LDA model by “20out” and “leave one patient” methods with and without outliers. (C) Training of the model based on the ridge classifier. (D) Cross-validation report of the ridge model with 5-fold method. (related to figure 2)

**Figure S4: Classification models based on FFPE alone and extended to FF dataset in negative ion mode.** (A) LDA model based on the FFPE dataset alone (HGSOC, SBL, SLG, CCC, mucinous and endometrioid tissues) in negative ion mode. (B) LDA model based on the extended dataset comprising both FFPE and FF. (C) Cross-validations of the LDA models (FFPE, FFPE/FF) by the “20out” method with and without outliers. (D) Training of the ridge model on the FFPE dataset. (E) 5-fold cross validation of the ridge model on the FFPE dataset. (Related to figure 2)

**Figure S5: Architecture of the multi-input model.** (A) Dual-input model architecture. (B) Accuracy and loss of training and validation sets according to epoch. (Related to figure 4)

**Figure S6: Cross-validation of the lipid markers of the different OC subtypes by MALDI-MSI and SpiderMass.** (A) Example of a cross-validated marker at  $m/z$  700.55 which is specific to endometrioid. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the endometrioid tissues. (B) Example of a cross-validated marker at  $m/z$  862.65 which is specific to mucinous carcinoma. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the mucinous tissues. (C) Example of a cross-validated marker at  $m/z$  748.55 which is specific to endometrioid and normal tissues. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the endometrioid and healthy tissues. (related to figure 4)

**Figure S7: MS/MS identification and variation of abundance of two Hexosylceramides between cancer cells and macrophages.** (A) SpiderMass mean spectra obtained from the THP1 macrophage versus SKOV3 cancer cell lines. The red frames indicate the ions  $m/z$  818.65 and  $m/z$  846.65 specific to the macrophages. (B) Boxplot based on the relative intensities of these two ions for SKOV3 vs. THP1 (\*\*\*\*= $p$ value $<0.0001$ ) showing the higher abundance of these two markers in the macrophages. (C) MS/MS spectrum of the ion  $m/z$  818.65 identified as GlcCer d40:1 (d18:1/22:0). (D) MS/MS spectrum of the ion  $m/z$  846.65 identified as GlcCer d42:1 (d18:1/24:0). (related to figure 5)

**Figure S8: Discriminative lipid markers of lymphocyte cells.** Boxplots based on the relative intensities of ions  $m/z$  722.55, 752.55, 794.55 and 885.55 plotted for the different subpopulation of lymphocytes (NK, CD8 or CD4) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (Related to figure 5)

**Figure S9: Abundance of lipid markers specific to macrophages in endometrium cancers vs normal ovary and endometrium tissues. Boxplots based on relative intensities of ion m/z 818.65, 846.65. (A)** In normal endometrium and ovary tissues (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). **(B)** In the different endometrium cancer subtypes (Healthy, HGSOC or endometrioid) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (Related to figure 5)

**Figure S10: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S11: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S12: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S13: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S14: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S15: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S16: Diagnosis of the serous high-grade carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S17: Diagnosis of the serous high-grade carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S18: Diagnosis of the mucinous carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S19: Diagnosis of the mucinous carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S20: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue

indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S21: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S22: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)

**Figure S23: Predicted probability of immune cell presence using LGBM model based on SpiderMass-MSI for patients with OS<42 months.** Immune score calculated from SpiderMass-MSI data for three patients with OS<42 months (A) Patient 89-3. (B) Patient 62-1. (C) Patient 114.1. (Related to figure 8)

**Figure S24: Predicted probability of immune cell presence using LGBM model based on SpiderMass-MSI for patients with OS>50 months.** Immune score calculated from SpiderMass-MSI data for three patients with OS>50 months (A) patient 102-1. (B) Patient 108-1. (C) Patient 44-2. (Related to figure 8)

## SUPPLEMENTARY TABLE LEGENDS

**Supp table 1: Clinical data of patients included in the FF cohort.** (Related to figure 2)

**Supp table 2: Discriminative lipids identified by MS/MS for the different OC subtypes.** (Related to figure 2)

**Supp table 3: Clinical data of patients included in the FFPE cohort.** (Related to figure 2)

**Supp table 4: List of classifiers tested on the FF dataset with their F1-score and 5-fold cross-validation.** (Related to figure 2)

**Supp table 5: List of the samples used for the validation test in blind.** (Related to figure 3)

**Supp table 6: Samples used to SpiderMass-MSI and create the immune scores.** (Related to figure 8)

**Supp table 7: Quantification of macrophages in the different OC subtypes calculated from MALDI-IHC. (Related to figure 7)**

## METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
<b>Anti-CD68</b>	Miralys™ PC-MT Probe	Cat#AP1001124
<b>Anti-CD8<math>\alpha</math></b>	Miralys™ PC-MT Probe	Cat#AP100152
<b>Anti-Ki67</b>	Miralys™ PC-MT Probe	Cat#AP1001184
<b>Anti-VIM</b>	Miralys™ PC-MT Probe	Cat#AP1001122
<b>Anti-Collagen-1A1</b>	Miralys™ PC-MT Probe	Cat#AP100153
<b>Anti-CD3<math>\epsilon</math></b>	Miralys™ PC-MT Probe	Cat#AP1001181
<b>Anti-CD4</b>	Miralys™ PC-MT Probe	Cat#AP100173
<b>Anti-CD3 PE (clone SK7)</b>	Sony Biotechnology	Cat#2324030
<b>Anti-CD4 FITC (clone A161A1)</b>	Sony Biotechnology	Cat#2387030
<b>Anti-CD8 Alexa Fluor 647 (clone HIT8a)</b>	Sony Biotechnology	Cat#2104590
<b>Anti-CD7 PE Cy5 (clone CD7-6B7)</b>	Sony Biotechnology	Cat#2315550
<b>Biological samples</b>		
<b>Healthy and cancerous ovarian tissues</b>	Oscar Lambret Center	<a href="https://www.centreoscarlambret.fr">https://www.centreoscarlambret.fr</a>
<b>Blood samples for macrophages and lymphocytes</b>	French blood agency (EFS)	<a href="https://dondesang.efs.sante.fr">https://dondesang.efs.sante.fr</a>

<b>Chemicals, peptides, and recombinant proteins</b>		
<b>Hemalun</b>	Merck	Cat#109249
<b>Phloxin</b>	VWR	Cat#10047229
<b>Saffron</b>	MM France	Cat#F/SAFRAN
<b>Ethanol (EtOH)</b>	Carlo Erba Reagents	Cat#4127012
<b>Xylene</b>	Carlo Erba Reagents	Cat#492301
<b>Glycerol</b>	Reidel de Haen	Cat#15523
<b>Isopropanol (IPA)</b>	Carlo Erba Reagents	Cat#415183
<b>DMEM</b>	Thermo Fischer Scientific	Cat#12-491-015
<b>Prigrow 1 Medium</b>	Applied Biological Materials	Cat#TM001
<b>Mc Coy's 5A Medium</b>	Gibco	Cat#26600023
<b>RPMI Medium</b>	Thermo Fischer Scientific	Cat#31870025
<b>DPBS, no calcium, no magnesium</b>	Thermo Fischer Scientific	Cat#14190-094
<b>Ficoll</b>	Dominique Dutscher	Cat#17-5442-02
<b>EDTA</b>	Sigma Aldrich	Cat#03690
<b>Leucine enkephalin</b>	Waters	Cat#70002456
<b>Phorbol-12-myristate-13-acetate (PMA)</b>	Sigma Aldrich	Cat#P8139
<b>Lypopolysachharide (LPS)</b>	Invivogen	Cat#tlrl-3pelps

<b>Macrophage colony stimulating factor (MCSF)</b>	Ozyme	Cat#BLE574802
<b>IFN<math>\gamma</math></b>	PeptoTech	Cat#300-02
<b>IL4</b>	PeptoTech	Cat#200-04
<b>2,5-dihydroxybenzoic acid</b>	Sigma Aldrich	Cat#149357
<b>Norharmane</b>	Sigma Aldrich	Cat#N6252
<b>Methanol (MeOH)</b>	Carlo Erba Reagents	Cat#4148831
<b>Chloroform</b>	Carlo Erba Reagents	Cat#508320
<b>Trifluoroacetic (TFA)</b>	Sigma Aldrich	Cat#T6508
<b>Paraformaldehyde (PFA)</b>	Alfa Aesar	Cat#43368
<b>Acetone</b>	VWR	Cat#20067.320
<b>Acetic acid</b>	Sigma Aldrich	Cat#320099
<b>TRIS-HCl</b>	Promega	Cat#H5125
<b>TRIS-Base</b>	Promega	Cat#H5135
<b>NaCl</b>	Fischer chemicals	Cat#S/3161/60
<b>Octyl<math>\beta</math>-D-glucoopyranoside (OG)</b>	Sigma Aldrich	Cat#03757
<b>Ammonium bicarbonate (NH<math>_4</math>HCO<math>_3</math>)</b>	Sigma Millipore	Cat#149357
<b>Deposited data</b>		
<b>The raw data from SpiderMass analysis has been deposited to the Harvard dataverse site.</b>	This paper	Harvard dataverse :

		<a href="https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IMKIN9">https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IMKIN9</a> )
<b>The original code has been deposited on GitHub.</b>	This paper	GitHub  <a href="https://github.com/yanisZirem/SpiDerMass_AI_pipelines.git">https://github.com/yanisZirem/SpiDerMass_AI_pipelines.git</a>
<b>Experimental models: Cell lines</b>		
<b>Human Immortalized Ovarian Epithelial Cell Line (SV40)</b>	Applied Biological Materials	Cat#T1074
<b>SKOV-3</b>	ATCC	Cat#HTB-77
<b>POE4</b>		RRID:CVCL_2690
<b>PA-1</b>	ATCC	Cat#CRL-1572
<b>SW 626</b>	ATCC	Cat#HTB-78
<b>THP-1</b>	ATCC	Cat#TIB-202
<b>Software and algorithms</b>		
<b>GraphPad Prism v9.5.1</b>	GraphPad software	RRID:SCR_000306  <a href="https://www.graphpad.com">https://www.graphpad.com</a>
<b>Abstract Model Builder (AMX) Beta version 1.0.2053.0</b>	Waters research center, Hungary	
<b>Abstract Model Builder (AMX) Recognition Beta version 1.0.2053.0</b>	Waters research center, Hungary	
<b>Python 3.11.5</b>	Open source	RRID:SCR_008394  <a href="https://www.python.org">https://www.python.org</a>
<b>Jupyter Notebook 6.5.4</b>	Open source	<a href="https://jupyter.org">https://jupyter.org</a>

<b>FlexImaging 5.0</b>	Bruker	<a href="https://www.bruker.com/fr/products-and-solutions/mass-spectrometry/maldi-tof/rapiflex-maldi-tissue typer.html">https://www.bruker.com/fr/products-and-solutions/mass-spectrometry/maldi-tof/rapiflex-maldi-tissue typer.html</a>
<b>SCiLS Lab 2022a Pro</b>	Bruker and SCiLS GmbH	RRID:SCR_014426 <a href="http://scils.de/software/">http://scils.de/software/</a>
<b>QuPath 0.4.4</b>	Bankhead et al., 2017. <sup>52</sup>	RRID:SCR_018257 <a href="https://qupath.github.io">https://qupath.github.io</a>
<b>Matlab R2019a</b>	MathWorks	RRID:SCR_001622 <a href="https://matlab.mathworks.com">https://matlab.mathworks.com</a>
<b>MetFrag</b>	Ruttkies et al., 2016. <sup>51</sup>	<a href="https://msbi.ipb-halle.de/MetFrag/">https://msbi.ipb-halle.de/MetFrag/</a>
<b>MassLynx 4.1</b>	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>

## RESSOURCE 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 publications, DOI is listed in the key resources table.

All generated code (processing, pipeline, immune score and IHC quantification) has been deposited at GitHub [https://github.com/yanisZirem/SpiderMass\\_AI\\_pipelines.git](https://github.com/yanisZirem/SpiderMass_AI_pipelines.git) and is available from the lead contact upon request. If you have any questions or feedback, please contact [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr).

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Ovarian cancer cohort

This is a single cohort study conducted at the Oscar Lambret Center in Lille, France. This is an observational study. The Institutional Review Board of the Oscar Lambret Center has confirmed that no ethical approval is required (number 2022-08). The study complies with the “reference methodology” MR004 adopted by the French Data Protection Authority (CNIL), and we checked that patients did not object to the use of their anonymized clinical data for the research purpose. Data was obtained through electronic medical record review. Oscar Lambret is a European Society of Gynecologic Oncology (ESGO)-accredited center for advanced ovarian cancer surgery.

### Patient samples

As a reference center for the management of rare ovarian tumors, the Centre Oscar Lambret (COL) is once again certified by the European Society of Gynecologic Oncology (ESGO) for surgery on advanced ovarian cancer. Thus, tissue samples were obtained from patients by COL in Lille, France as a prospective cohort. Patients provided written informed consent before participating in the trial. To protect patient privacy, no personal information was used in these experiments, and a random number was assigned to each sample. A cohort of 78 Fresh Frozen (FF) samples were processed between September 2020 and January 2021 including 41 normal ovaries versus 37 OC including (10 serous borderline ovaries, 13 serous high-grade ovaries, 8 mucinous ovaries, 6 endometrioid) (**Supp Table 1**). Moreover 79 frozen samples of endometrium issued by the same center were processed in September 2021 including 42 normal, 31 endometrioid carcinoma and 6 serous high-grade carcinomas. A validation cohort of 24 samples has been used to validate the model. Finally, a FFPE retrospective cohort of 47 patients has been analyzed. Retrospective cohort of 83 FFPE specimens was issued by Oscar-Lambret center tissue bank including 33 HGSOC, 5 CCC and 12 SBL. (**Supp Table 2**)

### Pathology Review and Histology Control

For histology, one gynecologist pathologist (QP) read and annotated HPS stained tissues, and two human Ovarian cancer board certified pathologist (Dr. Anne-Sophie Lemaire and Dr. Camille Pasquesonne, Oscar Lambret Center) commented and validated these annotations. The

pathologists were blind to any information about the acquisition from MS studies. The annotations were done on Hemalun/Phloxine/Saffron (HPS) stained tissues. For HPS staining, a 10 µm thick tissue slice consecutive to the SpiderMass-analyzed tissue was treated with hemalun solution for 1 minute and rinsed with tap water. Then the tissue section was stained in phloxine 0.1% solution for 10 sec and rinsed with tap water before dehydration in 70% and 100% ethanol baths. Finally, the sections were dipped in saffron for 5sec, rinsed twice in alcohol, cleaned in xylene, and mounted with cover slips and the EUKITT® slide mounting medium. After HPS staining, the nuclei were colored dark blue, the cytoplasm pink, and the conjunctive tissue orange. The stained slide was scanned for digital image acquisition using the Panoramic MIDI slide scanner (3DHISTECH LTD. Budapest, Hungary) and the images were viewed and exported using Panoramic viewer 1.15.

### Cell lines culture

Immortalized epithelial ovarian cell (referred as healthy) were cultured in Prigrow I medium. Several ovarian cancer cell lines of different origins were used and cultured. PA-1 (RRID:CVCL\_0479) is an ovarian teratocarcinoma cell line from a 12 years old patient with a single chromosomal aberration (reciprocal t(15;20)(p11.2;q11.2)). PA-1 cells were cultured in DMEM. SK-OV-3 (RRID:CVCL\_0532) is an ovarian cancer cell line with epithelial-like morphology and which is resistant to tumor necrosis factor and to other cytotoxic drugs (e.g., diphtheria toxin, cisplatin, and Adriamycin) derived from the ascites of a 64-year-old Caucasian female with serous cystadenocarcinoma. SK-OV-3 ovarian carcinoma cells were cultured in McCoy medium. PEO4 (RRID:CVCL\_2690) cell line is derived from patient ascites, has a BRCA2 mutation and is multi-resistant to chemotherapy. THP-1 (RRID:CVCL\_0006) is a monocyte cell line isolated from peripheral blood from a 1 year old child with acute monocytic leukemia. PEO4 and THP-1 cells were cultured with RPMI medium. All the medium were 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 directly into cell plate.

## **METHOD DETAILS**

### Experimental design

The data were collected and processed in blind after anonymization of the patients using an internal laboratory labelling. All samples of the cohort were analyzed and since ovarian cancer is an heterogeneous class of cancer originating from various types of tissues, we took great care that features discriminating the different types of tissues contributed less weight to the classification model than the other features of interest such as grading or typing. No patients or generated data were excluded from the study and outliers were defined per se when their value was more than 5 time the mean's SD. All cross-validation results are shown with and without the outliers considered. The tissues were analyzed in 5 different locations to replicate the measure and in both mode of analysis.

### Tissue Preparation

Frozen samples had been snap-frozen and stored at  $-80^{\circ}\text{C}$ . Samples were cut in slice of  $7\mu\text{m}$ , stained with HPS and sent to pathologists for annotations using CryoStats (Leica Microsystems, Nanterre, France). The mirror of this tissue was also cut at  $20\mu\text{m}$  for SpiderMass analysis and cross-validated with another cut of  $12\mu\text{m}$  for MALDI-MSI analysis on ITO-coated glass slides (LaserBioLabs, Valbonne, France) and stored at  $-80^{\circ}\text{C}$ . Annotated slide were used to identify the different areas to shot with SpiderMass.

### Primary monocyte isolation

Blood was collected from patient donator by the French Blood agency. An amount of 600mL was collected for this study. This blood was diluted two times into PBS-EDTA. Leucocytes were isolated by 25min centrifugation 2200 rpm with a Ficoll gradient. Leucocytes were than washed three times with PBS-EDTA, resuspended into RPMI medium and incubated into a cell plate 1H30 at  $37^{\circ}\text{C}$ . To only keep the macrophages that are stuck to the plate, the cell plate was then washed three times with PBS. Macrophages were grown 7 days with RPMI medium with 10% fetal bovine serum, 100U/mL penicillin-streptomycin and MCSF.

### Macrophage differentiation

After the isolation of monocyte, THP-1 cells were stimulated with 10ng/mL PMA for the macrophage differentiation during 24h. THP-1 as well as primary macrophages were stimulated into two different conditions. M1-like macrophages were stimulated with 0.5mg/mL of LPS and 20ng/mL of IFN- $\gamma$  for 48h and M2-like macrophages were stimulated with 20ng/mL of IL4 for 48h.

### Primary lymphocyte isolation

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood samples using density gradient centrifugation (Ficoll Paque Plus (GE Healthcare)). Cells were then 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  $4^{\circ}\text{C}$  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<sup>TM</sup> centrifuge (Thermo Shandon Cytospin) and conserved at  $-80^{\circ}\text{C}$ .

### SpiderMass mass spectrometry analysis

The global design of the instrument setup was described in a previous study<sup>46 47</sup>. Briefly, the setup is composed of three parts, a laser system for micro-sampling of tissues set remotely, a transfer line allowing for transfer of the micro-sampled material to the mass spectrometer which represents the third part. The laser probe is composed of an OPO which is tunable between  $2.8\mu\text{m}$  to  $3.1\mu\text{m}$  (Radiant version 1.0.1, OPOTEK Inc., Carlsbad, USA) pumped by a pulsed Nd:YAG laser (pulse duration: 5 ns,  $\lambda=1064\text{ nm}$ , Quantel, Les Ulis, France) and a repetition rate of 10 Hz. A GeG multimode laser fiber ( $450\mu\text{m}$  core diameter; length of 1 m; Le verre Fluoré, France) is connected to the laser system output and a handheld including a 4 cm focusing lens is attached to the end of the laser fiber. The handheld allows the user to hold the system and screen the surface of raw tissues or the system to be fixed on a holder. A the focusing distance the laser spot size on the sample is  $500\mu\text{m}$  diameter. In these experiments the handheld

was fixed on a robotic arm (commercially available stiff 3D-axis precision MECA 500 robotic arm, Mecademic, Montreal, Canada) to obtain an accurate positioning and the irradiation time was fixed to 10 sec (i.e., 100 laser shots) at 4 mJ/pulse laser energy corresponding to a laser fluence of 2.03 J/cm<sup>2</sup>. The laser energy was measured at the focal point using a power meter (ThorLabs, Maisons-Laffitte, France). The second part of the system corresponds to a 3 m length transfer line made from a Tygon ND 100-65 tubing (2.4 mm inner diameter, 4 mm outer diameter, Akron, USA). The transfer line is attached on one side onto the laser handheld and positioned close to the exit of the laser fiber in order to aspire the ablated material coming from the laser irradiation. On its other side, the transfer tube is directly attached onto the mass spectrometer (Xevo G2-XS, Waters, Wilmslow, United Kingdom) from which the conventional electrospray source was removed and replaced by an atmospheric pressure interface described elsewhere (Balog et al., 2015). MS acquisitions were performed in the resolution mode of the instrument both in positive and negative ion with a scan time of 1 sec. For tissue analysis, the samples were taken out of the -20°C freezer and thawed to RT for 1 min prior to SpiderMass analysis. The spectral acquisition sequence was composed of 5 acquisitions using 10-sec irradiation periods. Taking into consideration that both positive and negative ion modes were recorded, a total of 10 acquisitions per biopsy were performed. The mirror tissue section to the one analyzed was used as a morphological control to ensure that each acquisition was performed in the same histological area and consider the tissue heterogeneity when needed. SpiderMass data were validated by MALDI-MSI and significantly discriminant lipids were identified using MS/MS experiments.

At 70% of confluence cell line were washed with PBS 2 times and dried under the PSM during 10min at RT. Then MS acquisition were performed directly on the cell plate. The spectral acquisition sequence was composed of 12 acquisitions using 3-sec irradiation periods. Taking into consideration that only negative ion modes were recorded, a total of 12 acquisitions per plate were performed.

#### MALDI Mass Spectrometry Imaging (MALDI-MSI)

The tissues were sectioned at 20µm thickness using CryoStats (Leica Microsystems, Nanterre, France). These sections were deposited on ITO-coated glass slides (LaserBioLabs, Valbonne, France) and vacuum-dried for 15 min. Norharmane was used as MALDI matrix for dual polarity imaging in both negative and positive ion mode. The matrix was sprayed using a M5 TM-Sprayer (HTX Technologies, US). The parameters were set at 30°C nozzle temperature, 12 passes with a CC pattern, 0.1 mL/min as flow rate, the nozzle velocity at 1200 mm/min, the track width spacing at 2 mm and 10 psi for the gas pressure. The MALDI-MSI experiments were performed on a RapifleX MALDI Tissuetyper™ (Bruker Daltonik GmbH) and images acquired in both ion modes using red phosphorus as an external calibrant before the analysis. The spatial resolution was set to 50 µm and the mass-range was fixed to *m/z* 400-1200. The energy was set between 70 and 80% of the laser energy. The sample probe and the first lens of the ion source were set to a voltage of 20 kV and 11 kV respectively, the last one being grounded. System operating at 10 kHz in reflectron mode (Bruker Daltonik GmbH) with a nominal acceleration potential of ± 20 kV. MSI data were acquired by using a 50 × 50 µm<sup>2</sup>.

### MALDI-Immunochemistry: MALDI-IHC

Multiplex MALDI-IHC imaging was performed from the same slides that were used for MALDI-MSI of lipids. The protocol used was the recommended protocol for AmberGen (Billerica, MA) Miralys™ probes. The matrix was removed using 5 seconds EtOH baths. Then, the tissues were dried under vacuum for 10 min and fixed with PFA 1%. A PBS bath was performed for 10 min each to wash the tissues, then two baths of 1 min in acetone and one for 3 min in a Carnoy solution was performed. Then tissues were re-hydrated with two baths in 100% EtOH for 2 min each, followed by three consecutive bath of 3 min in EtOH 95%, EtOH 70%, and EtOH 50%. A 10 min bath in TBS (50 mM Tris, pH 7.5, 200 mM NaCl) was applied before the antigen retrieval, which was performed in Tris 20mM pH9 during 30 min at 95°C. The slides were washed 10 min in TBS in order to be treated with a tissue blocking buffer (2% (v/v) normal serum, 5% (w/v) BSA in TBS-OG) for 1 h. After removal of the later, tissues incubated at 4°C overnight in tissue blocking buffer, in which: CD3ε, CD68, CD8, CD4, Ki67, Vimentin and collagen probes were added in a 2.5 µg/mL concentration. Each slide was separately washed in a 10 cm diameter glass petri dish, with the tissue slide face-up in the horizontal position, and with gentle shaking with three baths of 5 min in TBS and 2 min in NH<sub>4</sub>HCO<sub>3</sub> 50mM, always protected from the light. The tissues were dried for 1h30 at RT in a vacuum desiccation chamber prior the photocleavage of the probes by a 365 nm UV exposition (AmberGen (Phrozen) UV box, Billerica, MA) for 10 min. DHB matrix at 20 mg/mL in MeOH:TFA 0,1% (70:30, v/v) was sprayed on the tissues using the HTX sprayer M5 (HTX technologies, LLC), followed by a matrix recrystallization step in a humid chamber with 5% IPA at 55°C for 1 min. Next, the slides were subjected to MALDI-MSI analysis using a rapiflex MALDI-TOF-MS instrument (Bruker Daltonics, Germany) according to the parameters: reflector mode, positive ion mode, laser spot size 20 µm with 20µm continuous raster scanning, 500 laser shots per pixel, and normalization to the total ions count (TIC) for the multiplex images processing. The analysis of spectrums and images were performed using flexImaging (Bruker Daltonics, Billerica, MA) and SCiLS Lab software (SCiLS Lab 2023, SCiLS GmbH).

### SpiderMass-Mass Spectrometry Imaging

The Spider-Mass microprobe was coupled to a stiff robotic arm described in a previous work<sup>50</sup>. The spatial step size was set to 250 µm to achieve oversampling and the mass-range was fixed between  $m/z$  600-1100. The acquisition sequence was composed of 3 consecutive laser shots and 3 seconds between each step. The laser bursts were automatically fired by the MATLAB in house user interface developed for the robotic WALDI-MSI. The data were acquired negative, sensitivity ion mode on the Xevo mass analyzer.

### Lipid Identification

$m/z$  intervals corresponding to loadings with the largest contribution to the explained variance observed in the different groups were selected for MS/MS-based identification. For these experiments, the settings were the same as described in the SpiderMass section. Full scans were acquired in the Xevo G2-XS. The identifications were performed directly on the tissue by doing a full scan first to verify the presence of the targeted masses. Then after switching to MS/MS mode, the ions were selected “on the fly” for with collision induced dissociation (CID) with an isolation window of 0.1  $m/z$ . MS/MS spectra were acquired for a continuous irradiation time of

5 s. The lipid annotations were performed manually through Metfrag<sup>51</sup>, LipidMaps and Alex 123 databases.

### Data processing

**AMX classification.** For data analysis, all raw data files produced with the SpiderMass instrument were imported into the Abstract Model Builder (AMX) Recognition Beta version 1.0.2053.0. After importation, spectra were first pre-processed. The pre-processing steps include background subtraction, total ion count normalization, and re-binning to a 0.1 *m/z* window. All the processed MS spectra obtained from the 78 histologically validated samples were then used to build a principal component analysis and linear discriminant analysis (PCA-LDA) classification model. The first step consisted of PCA to reduce data multidimensionality by generating features that explain most of the variance observed. These features were then subjected to supervised analysis using LDA by setting the classes that the model will be based upon. LDA attempts to classify the sample spectra and assess the model by cross validation. Cross validation was carried out by the “leave one patient out” methods. In this method, the spectra are grouped by patient and left out one by one; at each step the model without the patient is interrogated against this model.

**Statistical analysis from classification.** *m/z* intervals corresponding to loading scores with the largest contribution to the first principal components (i.e., where ideally 80% of the variance is explained) were obtained and their normalized intensities across different classes were plotted using GraphPad Prism v9.5.1. To consider the imbalance in the numbers of sample per class we used non-parametrical two-sided ANOVA (Kruskal-Wallis) followed by Dunn’s test and adjusted P-value to account for the multiple comparisons with a family-wise significance and confidence level of 0.05.

**Homemade classification pipeline.** Before interrogating the homemade classification pipeline, 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) 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, Hungary) and the extraction of the matrix data as csv files. The pandas library is used to import the csv files into Python. The Lazy Predict library was used to build multiple models from the scikit-learn library by training and testing a range of 24 classifiers (<https://lazypredict.readthedocs.io/en/latest/>). The random state was always kept at 1. Subsequently, the optimal model was reconstructed individually using the scikit-learn library, which enabled fine-tuning of its parameters for optimization and evaluation of its accuracy. To further evaluate the model's performance, 5-fold cross-validation was performed using K-fold 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, respectively. The Local Interpretable Model-agnostic Explanations (LIME) algorithm was employed to explain decision making of the classification model. This algorithm calculates feature contributions that can be positive or negative. The ELI5 library was utilized

to generate a LIME table containing the weight of feature contributions using the `explain_prediction` function (<https://eli5.readthedocs.io/en/latest/overview.html>). 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 high feature contribution. 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 corresponding to molecules in the final list. The corresponding box plots were then generated from the seaborn library.

**Multi input neuronal network.** To build and train a neuronal network model with multiple input spectra and HPS images (by using QuPath<sup>52</sup>), the TensorFlow and Keras libraries were used. spectra in csv format and HPS images in png format were combined by converting them into NumPy arrays. The model architecture is defined with separate branches for image (2D-CNN) and spectra (MLP) data inputs. Branches are concatenated and additional layers are added for classification. The model is compiled with the appropriate parameters. The script divides the data into training, validation, and testing sets (train:0.6, validation :0.2 and test: 0.2).

#### Immune-score modelling

The LightGBM Python library was employed to train a LightGBM model, which is a gradient boosting framework developed by Microsoft. The immune score models were built based on cell spectra in the  $m/z$  [600-1100] mass range in negative ion mode including the following categories and corresponding spectra counts: Macrophages (M1 & M2): 216 spectra, healthy cell line: 163 spectra, cancer cell lines: 146 spectra, Lymphocytes (NK, CD8, CD4): 114 spectra. To retrieve the prediction scores for each cell type, the `predict_proba` function from the LightGBM model was used instead of the `predict` function. This approach is favored because it provides probability estimates for each class by obtaining the scores for each cell type being present.

#### SpiderMass MSI immunescoring

A Matlab code was developed to extract the MS scans in each pixel from the Waters Raw file. The raw SpiderMass files were converted to mzML using MSConvert (Proteowizard). The imzML converter was used to reconstruct the imaging files. A Python code was developed for further analysis and utilizes the `ImzMLParser` class from the `pyimzml` library. It imports and parses the imzML file within the Python environment. The script iterates over each coordinate (x, y, z) in the imzML file and retrieves the corresponding mass spectra using the `getspectrum()` method. It applies a desired mass range filter to select specific mass values and stores the spectra along with their respective coordinates. Several preprocessing steps are performed in the Python code. This includes identifying non-zero pixels, normalization, and applying Gaussian smoothing using the `gaussian_filter` function from the `scipy.ndimage.filters` module. The spatial resolution is also increased by a factor 2 using interpolation with the `zoom` function from `scipy.ndimage`. For predictions, a pre-trained LightGBM model is loaded from a joblib file using the `joblib.load` function. The script iterates over each spectrum in the image and prepares them for prediction. The `predict_proba` method of the model is used to obtain the predicted scores and labels for each spectrum. After obtaining the predicted labels, the script visualizes

the label maps for each class predicted by the model. It creates a colormap using the 'jet' color scheme and iterates over each class to display the label maps using the `imshow` function from the `matplotlib.pyplot` module. A color bar of scores is also added for reference.

## REFERENCES

1. Kocarnik, J.M., Compton, K., Dean, F.E., Fu, W., Gaw, B.L., Harvey, J.D., Henrikson, H.J., Lu, D., Pennini, A., Xu, R., et al. (2022). Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019 A Systematic Analysis for the Global Burden of Disease Study 2019. *JAMA Oncol* 8, 420–444. 10.1001/jamaoncol.2021.6987.
2. 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 Cancer J Clin* 71, 209–249. 10.3322/caac.21660.
3. Lheureux, S., Braunstein, M., and Oza, A.M. (2019). Epithelial ovarian cancer: Evolution of management in the era of precision medicine. *CA Cancer J Clin*, caac.21559. 10.3322/caac.21559.
4. Lisio, M.-A., Fu, L., Goyeneche, A., Gao, Z., and Telleria, C. (2019). High-Grade Serous Ovarian Cancer: Basic Sciences, Clinical and Therapeutic Standpoints. *Int J Mol Sci* 20, 952. 10.3390/ijms20040952.
5. Prat, J. (2014). Staging classification for cancer of the ovary, fallopian tube, and peritoneum. *International Journal of Gynecology and Obstetrics* 124, 1–5. 10.1016/j.ijgo.2013.10.001.
6. Hudry, D., Bécourt, S., Scambia, G., and Fagotti, A. (2022). Primary or Interval Debulking Surgery in Advanced Ovarian Cancer: a Personalized Decision—a Literature Review. *Curr Oncol Rep* 24, 1661–1668. 10.1007/s11912-022-01318-9.

7. Alvarez Secord, A., O'Malley, D.M., Sood, A.K., Westin, S.N., and Liu, J.F. (2021). Rationale for combination PARP inhibitor and antiangiogenic treatment in advanced epithelial ovarian cancer: A review. *Gynecol Oncol* 162, 482–495. 10.1016/j.ygyno.2021.05.018.
8. Barakat, A., Ismail, A., and Chattopadhyay, S. (2022). Management of advanced ovarian cancer in Leicester: The benefits of a paradigm shift in surgical approach. *Journal of Obstetrics and Gynaecology Research* 48, 3233–3241. 10.1111/jog.15433.
9. Brennan, D.J., and Moran, B.J. (2021). Time to Evolve Terminology from “Debulking” to Cytoreductive Surgery (CRS) in Ovarian Cancer. *Ann Surg Oncol* 28, 5805–5807. 10.1245/s10434-021-10490-4.
10. du Bois, A., Reuss, A., Pujade-Lauraine, E., Harter, P., Ray-Coquard, I., and Pfisterer, J. (2009). Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: A combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials. *Cancer* 115, 1234–1244. 10.1002/cncr.24149.
11. Chi, D.S., Eisenhauer, E.L., Lang, J., Huh, J., Haddad, L., Abu-Rustum, N.R., Sonoda, Y., Levine, D.A., Hensley, M., and Barakat, R.R. (2006). What is the optimal goal of primary cytoreductive surgery for bulky stage IIIc epithelial ovarian carcinoma (EOC)? *Gynecol Oncol* 103, 559–564. 10.1016/j.ygyno.2006.03.051.
12. Aletti, G.D., Dowdy, S.C., Podratz, K.C., and Cliby, W.A. (2007). Relationship among surgical complexity, short-term morbidity, and overall survival in primary surgery for advanced ovarian cancer. *Am J Obstet Gynecol* 197, 676.e1-676.e7. 10.1016/j.ajog.2007.10.495.

13. Gerestein, C.G., Nieuwenhuyzen-de Boer, G.M., Eijkemans, M.J., Kooi, G.S., and Burger, C.W. (2010). Prediction of 30-day morbidity after primary cytoreductive surgery for advanced stage ovarian cancer. *Eur J Cancer* 46, 102–109. 10.1016/j.ejca.2009.10.017.
14. Ratnavelu, N.D., Brown, A.P., Mallett, S., Scholten, R.J., Patel, A., Founta, C., Galaal, K., Cross, P., and Naik, R. (2016). Intraoperative frozen section analysis for the diagnosis of early stage ovarian cancer in suspicious pelvic masses. *Cochrane Database of Systematic Reviews* 2016. 10.1002/14651858.CD010360.pub2.
15. Köbel, M., and Kang, E.Y. (2022). The Evolution of Ovarian Carcinoma Subclassification. *Cancers (Basel)* 14. 10.3390/cancers14020416.
16. Dijkstra, T.K. (2014). Ridge regression and its degrees of freedom. *Qual Quant* 48, 3185–3193. 10.1007/s11135-013-9949-7.
17. Cortes, C., Mohri, M., and Rostamizadeh, A. (2012). L2 Regularization for Learning Kernels. *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009*, 109–116.
18. Allauame, P., Rabilloud, N., Turlin, B., Bardou-Jacquet, E., Loréal, O., Calderaro, J., Khene, Z.E., Acosta, O., De Crevoisier, R., Rioux-Leclercq, N., et al. (2023). Artificial Intelligence-Based Opportunities in Liver Pathology—A Systematic Review. *Diagnostics* 13. 10.3390/diagnostics13101799.
19. Sans, M., Gharpure, K., Tibshirani, R., Zhang, J., Liang, L., Liu, J., Young, J.H., Dood, R.L., Sood, A.K., and Eberlin, L.S. (2017). Metabolic markers and statistical prediction of serous ovarian cancer aggressiveness by ambient ionization mass spectrometry imaging. *Cancer Res* 77, 2903–2913. 10.1158/0008-5472.CAN-16-3044.

20. Sans, M., Zhang, J., Lin, J.Q., Feider, C.L., Giese, N., Breen, M.T., Sebastian, K., Liu, J., Sood, A.K., and Eberlin, L.S. (2019). Performance of the MasSpec pen for rapid diagnosis of Ovarian cancer. *Clin Chem* 65, 674–683. 10.1373/clinchem.2018.299289.
21. Meriaux, C., Franck, J., Wisztorski, M., Salzet, M., and Fournier, I. (2010). Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids. *J Proteomics* 73, 1204–1218. 10.1016/j.jprot.2010.02.010.
22. Koundouros, N., and Poulogiannis, G. (2020). Reprogramming of fatty acid metabolism in cancer. *Br J Cancer* 122, 4–22. 10.1038/s41416-019-0650-z.
23. Munir, R., Lisec, J., Swinnen, J. V., and Zaidi, N. (2019). Lipid metabolism in cancer cells under metabolic stress. *Br J Cancer* 120, 1090–1098. 10.1038/s41416-019-0451-4.
24. Pöyry, S., and Vattulainen, I. (2016). Role of charged lipids in membrane structures — Insight given by simulations. *Biochim Biophys Acta Biomembr* 1858, 2322–2333. 10.1016/j.bbamem.2016.03.016.
25. Skipper, K. (2023). Surface Tension Drives Cancer Cell Migration. *Physics (College Park Md)* 16. 10.1103/physics.16.4.
26. Lankry, D., Rovis, T.L., Jonjic, S., and Mandelboim, O. (2013). The interaction between CD300a and phosphatidylserine inhibits tumor cell killing by NK cells. *Eur J Immunol* 43, 2151–2161. 10.1002/eji.201343433.
27. Urbanelli, L., Buratta, S., Logozzi, M., Mitro, N., Sagini, K., Raimo, R. Di, Caruso, D., Fais, S., and Emiliani, C. (2020). Lipidomic analysis of cancer cells cultivated at acidic pH reveals phospholipid fatty acids remodelling associated with transcriptional reprogramming. *J Enzyme Inhib Med Chem* 35, 963–973. 10.1080/14756366.2020.1748025.

28. Stoica, C., Ferreira, A.K., Hannan, K., and Bakovic, M. (2022). Bilayer Forming Phospholipids as Targets for Cancer Therapy. *Int J Mol Sci* 23. 10.3390/ijms23095266.
29. Putta, P., Rankenberg, J., Korver, R.A., van Wijk, R., Munnik, T., Testerink, C., and Kooijman, E.E. (2016). Phosphatidic acid binding proteins display differential binding as a function of membrane curvature stress and chemical properties. *Biochim Biophys Acta Biomembr* 1858, 2709–2716. 10.1016/j.bbamem.2016.07.014.
30. Quail, D.F., and Joyce, J.A. (2013). Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 19, 1423–1437. 10.1038/nm.3394.
31. Yeom, M., Park, J., Lim, C., Sur, B., Lee, B., Han, J.J., Choi, H.D., Lee, H., and Hahm, D.H. (2015). Glucosylceramide attenuates the inflammatory mediator expression in lipopolysaccharide-stimulated RAW264.7 cells. *Nutrition Research* 35, 241–250. 10.1016/j.nutres.2015.01.001.
32. Zhang, C., Wang, Y., Wang, F., Wang, Z., Lu, Y., Xu, Y., Wang, K., Shen, H., Yang, P., Li, S., et al. (2017). Quantitative profiling of glycerophospholipids during mouse and human macrophage differentiation using targeted mass spectrometry. *Sci Rep* 7. 10.1038/s41598-017-00341-2.
33. 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. 10.1016/j.jbc.2023.102902.
34. Zhang, Z., Huang, L., and Brayboy, L. (2021). Macrophages: An indispensable piece of ovarian health. *Biol Reprod* 104, 527–538. 10.1093/biolre/ioaa219.

35. Hogg, C., Panir, K., Dhimi, P., Rosser, M., Mack, M., Soong, D., Pollard, J.W., Jenkins, S.J., Horne, A.W., and Greaves, E. Macrophages inhibit and enhance endometriosis depending on their origin. *10.1073/pnas.2013776118/-/DCSupplemental*.
36. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree.
37. 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*. *10.3389/fchem.2023.1182404*.
38. Mantovani, A., Schioppa, T., Porta, C., Allavena, P., and Sica, A. (2006). Role of tumor-associated macrophages in tumor progression and invasion. *Cancer and Metastasis Reviews 25*, 315–322. *10.1007/s10555-006-9001-7*.
39. Claes, B.S.R., Krestensen, K.K., Yagnik, G., Grgic, A., Kuik, C., Lim, M.J., Rothschild, K.J., Vandenbosch, M., and Heeren, R.M.A. (2023). MALDI-IHC-Guided In-Depth Spatial Proteomics: Targeted and Untargeted MSI Combined. *Anal Chem 95*, 2329–2338. *10.1021/acs.analchem.2c04220*.
40. Macciò, A., Gramignano, G., Cherchi, M.C., Tanca, L., Melis, L., and Madeddu, C. (2020). Role of M1-polarized tumor-associated macrophages in the prognosis of advanced ovarian cancer patients. *Sci Rep 10*. *10.1038/s41598-020-63276-1*.
41. 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*.

42. Zhang, M., He, Y., Sun, X., Li, Q., Wang, W., Zhao, A., and Di, W. (2014). A high M1/M2 ratio of tumor-associated macrophages is associated with extended survival in ovarian cancer patients. *J Ovarian Res* 7. 10.1186/1757-2215-7-19.
43. Bruni, D., Angell, H.K., and Galon, J. (2020). The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer* 20, 662–680. 10.1038/s41568-020-0285-7.
44. Marliot, F., Pagès, F., and Galon, J. (2020). Usefulness and robustness of Immunoscore for personalized management of cancer patients. *Oncoimmunology* 9. 10.1080/2162402X.2020.1832324.
45. Wang, J., Li, D., Cang, H., and Guo, B. (2019). Crosstalk between cancer and immune cells: Role of tumor-associated macrophages in the tumor microenvironment. *Cancer Med* 8, 4709–4721. 10.1002/cam4.2327.
46. Fatou, B., Saudemont, P., Leblanc, E., Vinatier, D., Mesdag, V., Wisztorski, M., Focsa, C., Salzet, M., Ziskind, M., and Fournier, I. (2016). In vivo Real-Time Mass Spectrometry for Guided Surgery Application. *Nature Publishing Group*, 1–14. 10.1038/srep25919.
47. Ogrinc, N., Saudemont, P., Balog, J., Robin, Y., Gimeno, J., 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. 10.1038/s41596-019-0217-8.
48. Schäfer, K.C., Dénes, J., Albrecht, K., Szaniszló, T., Balogh, J., Skoumal, R., Katona, M., Tóth, M., Balogh, L., and Takáts, Z. (2009). In vivo, in situ tissue analysis using

- rapid evaporative ionization mass spectrometry. *Angewandte Chemie - International Edition* 48, 8240–8242. 10.1002/anie.200902546.
49. Balog, J., Kumar, S., Alexander, J., Golf, O., Huang, J., Wiggins, T., Abbassi-Ghadi, N., Enyedi, A., Kacska, S., Kinross, J., et al. (2015). In Vivo Endoscopic Tissue Identification by Rapid Evaporative Ionization Mass Spectrometry (REIMS). *Angewandte Chemie* 127, 11211–11214. 10.1002/ange.201502770.
50. 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* 2, 14383–14391.
51. Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., and Neumann, S. (2016). MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J Cheminform* 8. 10.1186/s13321-016-0115-9.
52. 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. 10.1038/s41598-017-17204-5.

# **MACHINE LEARNING BASED OVARIAN CANCER DIAGNOSTIC AND PROGNOSTIC FOR REAL TIME IMMUNE SCORING AT THE OPERATING ROOM**

**Lucas Roussel<sup>1†</sup>, Yanis Zirem<sup>1†</sup>, Laurine Lagache<sup>1†</sup>, Delphine Hudry<sup>1,2†</sup>, Bertrand Meresse<sup>3</sup>,  
Marie Delbeke<sup>3</sup>, Eric Leblanc<sup>1,2</sup>, Gargey Yagnik<sup>4</sup>, Mark J. Lim<sup>4</sup>, Kenneth J. Rothschild<sup>4,6</sup>,  
Yves-Marie Robin<sup>1,2</sup>, Camille Pasquesoone<sup>1,2</sup>, Anne-Sophie Lemaire<sup>1,2</sup>, Delphine Bertin<sup>1,2</sup>,  
Fabrice Narducci<sup>1,2‡</sup>, Michel Salzet<sup>1,5†\*</sup>, Isabelle Fournier<sup>1,5†\*</sup>**

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

<sup>2</sup>Department of Gynecology Oncology, Oscar Lambret Center, F-59000 Lille, France

<sup>3</sup>Univ. Lille, Inserm, CHU Lille, U1286 - INFINITE - Institute for Translational Research in Inflammation, F-59000, Lille, France.

<sup>4</sup>AmberGen, Inc., Billerica, MA, United States

<sup>5</sup>Institut Universitaire de France, 75000 Paris

<sup>6</sup>Department of Physics, Photonics Center, Boston University, Boston, MA, United States

† equal contribution

\*Co-Corresponding author

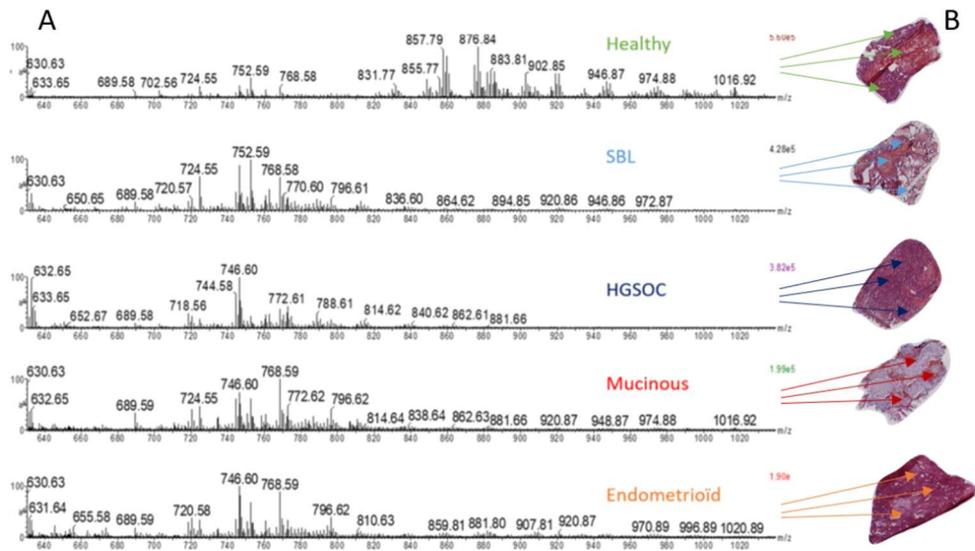
‡Lead contact

Prof. Michel Salzet, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - Inserm U1192 - Bât SN3, 1<sup>er</sup> étage, Campus Cité Scientifique, Université de Lille, F-59655 Villeneuve d'Ascq Cedex. Phone: +33 (0)320 434 194; email: [michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr)

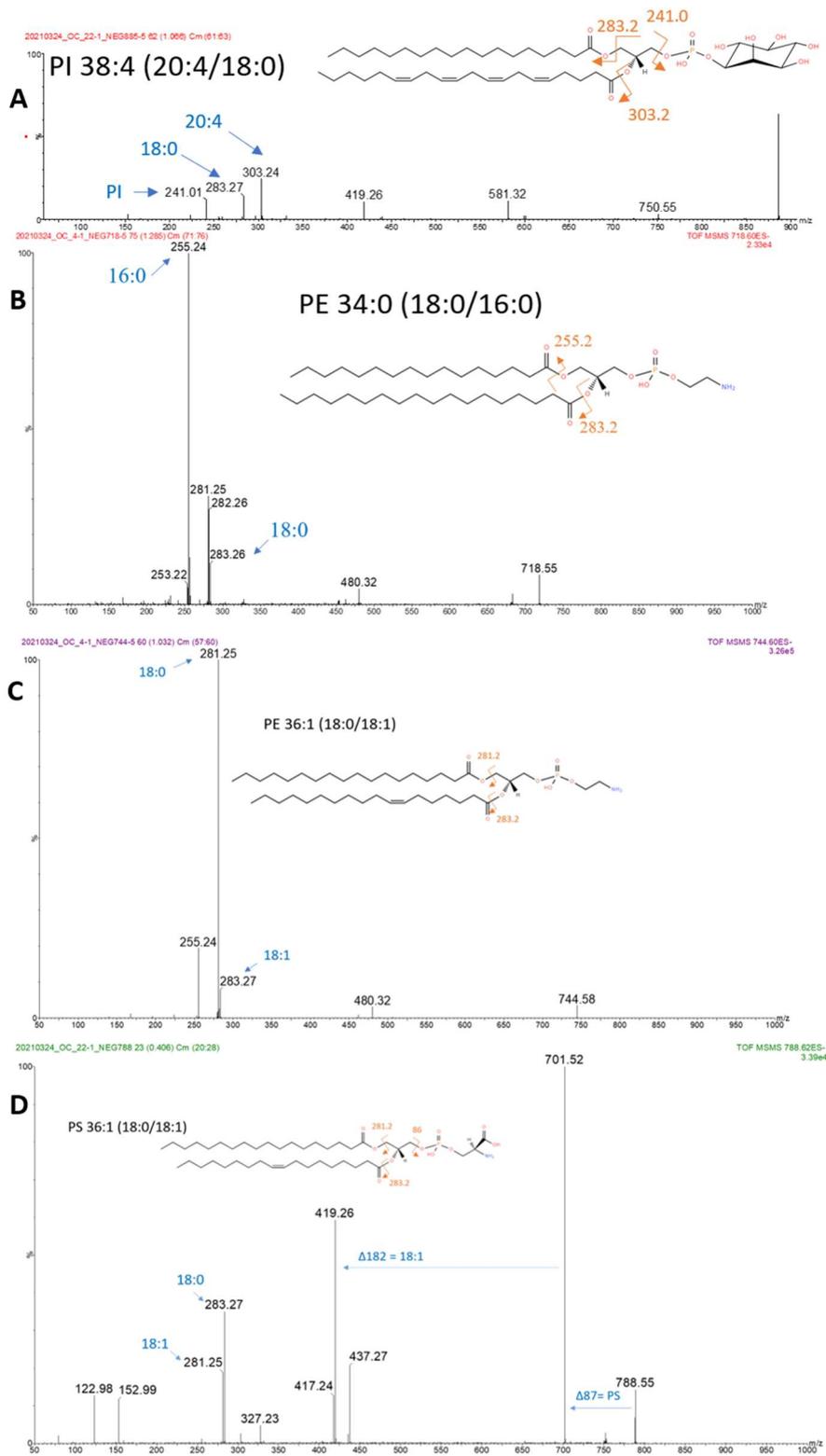
Prof. Isabelle Fournier, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - Inserm U1192 - Bât SN3, 1<sup>er</sup> étage, Campus Cité Scientifique, Université de Lille, F-59655 Villeneuve d'Ascq Cedex. Phone: +33 (0)320 434 194; email: [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr)

Dr. Fabrice Narducci, Department of Gynecology Oncology, Oscar Lambret Center, F-59000 Lille, France and Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - INSERM U1192 - Université Lille 1, Bât SN3, 1<sup>er</sup> étage, Cité Scientifique, F-59655 Villeneuve d'Ascq Cedex. Phone: + (33)320 295 959, email: [f-narducci@o-lambret.fr](mailto:f-narducci@o-lambret.fr)

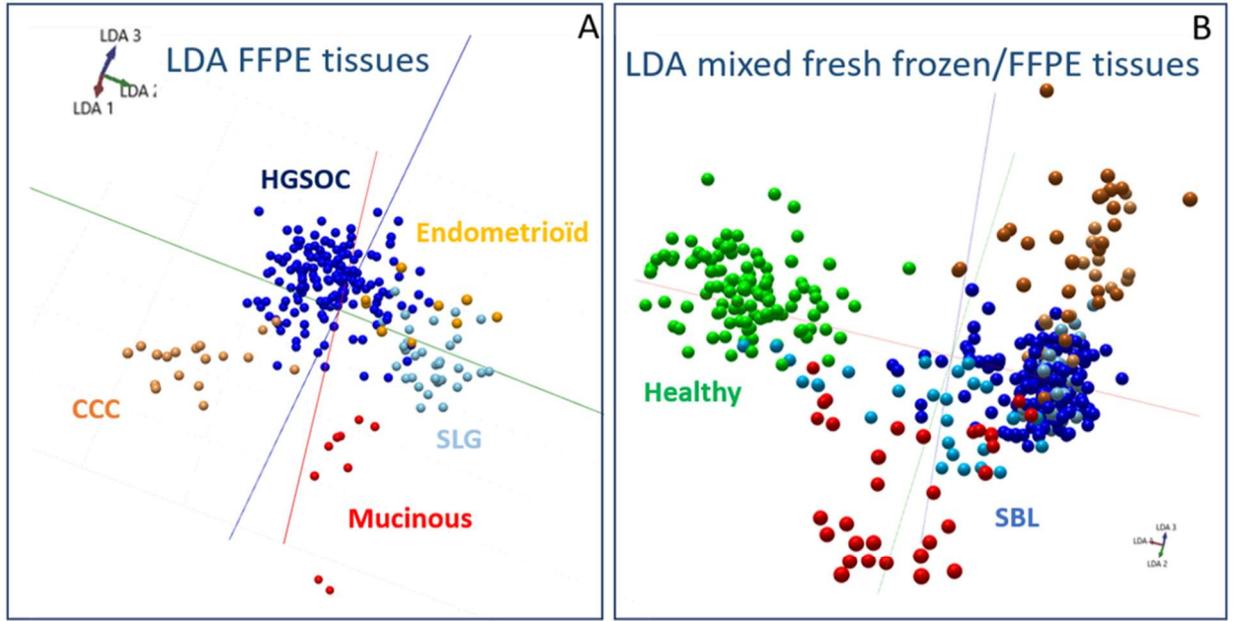
Supplementary figures:



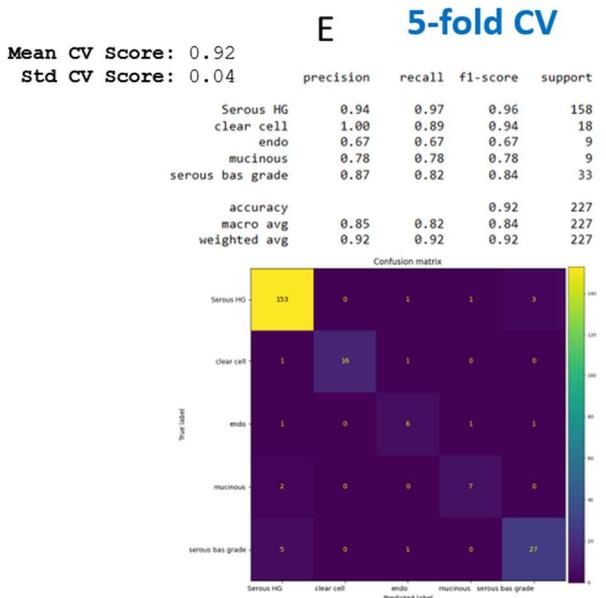
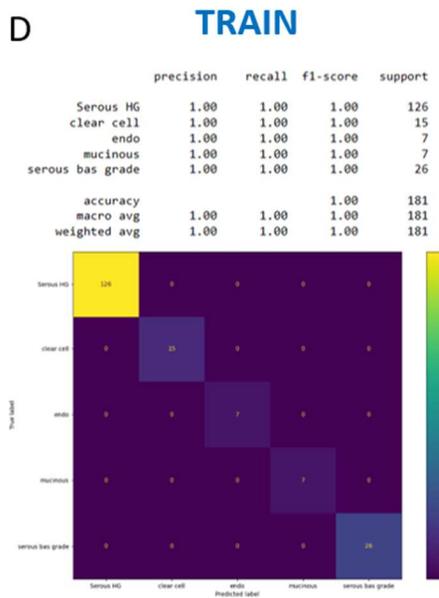
**Figure S1: Morphological and spectral fingerprint of ovarian cancer subtypes in positive ion mode. (A)** Mean Spectra for each ovarian cancer subtype in negative ion mode (4mJ/shot, burst mode, 10 shots, 1s/spectrum). **(B)** HPS staining of the corresponding histological sections. The arrows indicate the location where the laser was fired. (Related to figure 2)



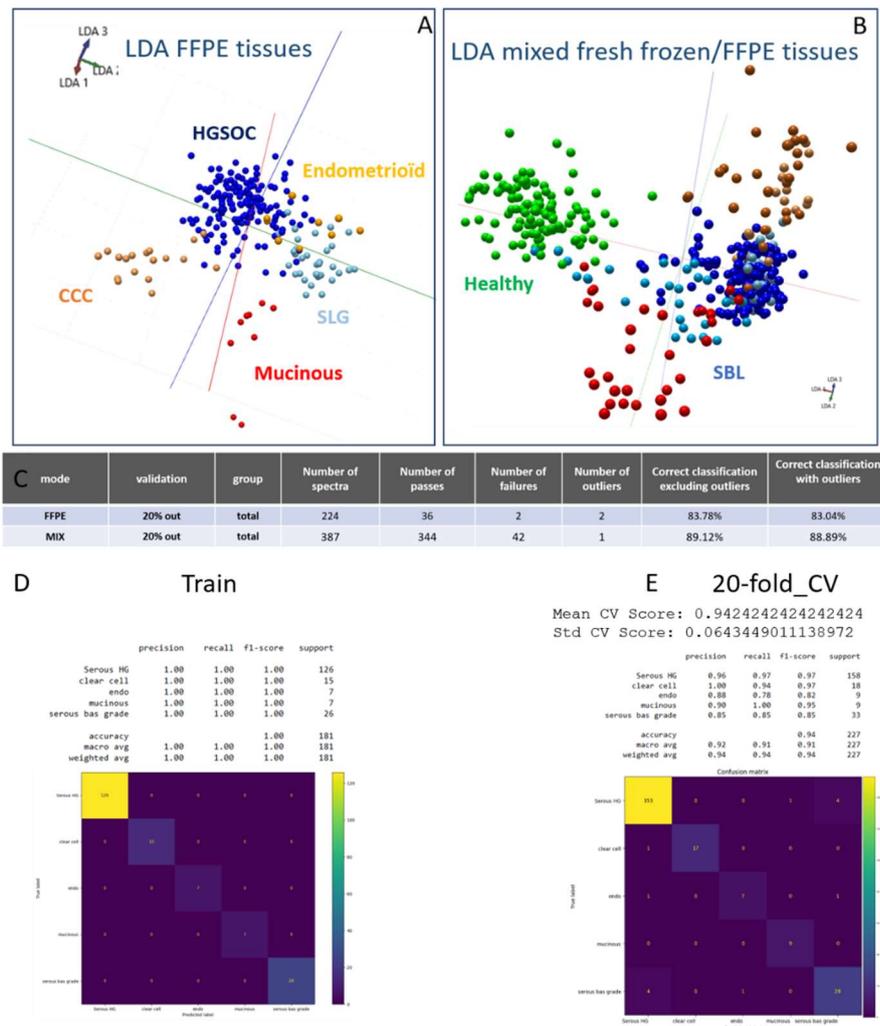
**Figure S2: Example of annotated MS/MS spectra used for lipid identification. (A)  $m/z$  885.5 corresponding to PI 38:4 (20:4/18:0). (B)  $m/z$  718.55 corresponding to PE 34:0 (18:0/16:0). (C)  $m/z$  744.55 corresponding to PE 36:1 (18:0/18:1). (D)  $m/z$  788.55 corresponding to PS 36:1 (18:0/18:1). (Related to figure 2)**



C mode	validation	group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct classification excluding outliers	Correct classification with outliers
FFPE	20% out	total	224	36	2	2	83.78%	83.04%
MIX	20% out	total	387	344	42	1	89.12%	88.89%



**Figure S3: Classification models and their cross-validation for OC subtypes in the positive ion mode. (A)** Linear discriminant analysis model for the different OC subtypes. **(B)** Cross-validations of the LDA model by “20out” and “full group out” methods with and without outliers. **(C)** Training of the model based on the ridge classifier. **(D)** Cross-validation report of the ridge model with 5-fold. (related to figure 2)



**Figure S4: Classification models based on FFPE alone and extended to FF dataset in negative ion mode. (A)** LDA model based on the FFPE dataset alone (HGSOC, SBL, SLG, CCC, mucinous and endometrioid tissues) in negative ion mode. **(B)** LDA model based on the extended dataset comprising both FFPE and FF. **(C)** Cross-validations of the LDA models (FFPE, FFPE/FF) by the “20out” method with and without outliers. **(D)** Training of the ridge model on the FFPE dataset. **(E)** 5-fold cross-validation of the ridge model on the FFPE dataset. (Related to figure 2)

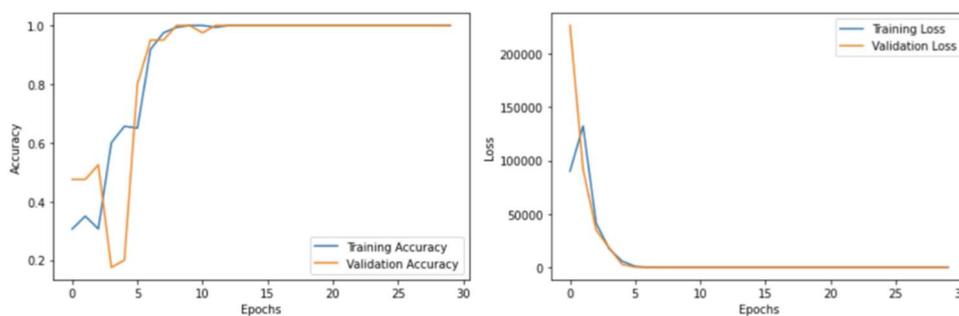
**A**

Model: "model"

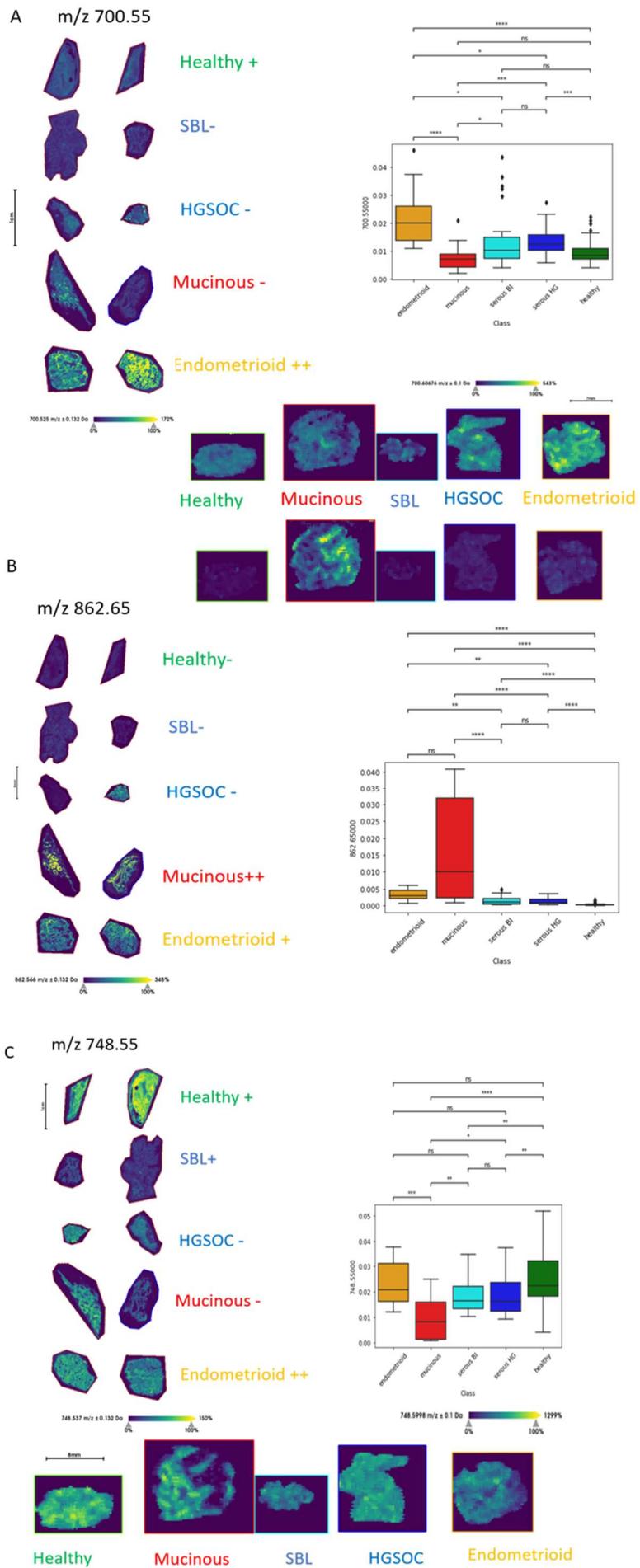
Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 224, 224, 3)]	0	[]
conv2d (Conv2D)	(None, 222, 222, 32)	896	['input_1[0][0]']
input_2 (InputLayer)	[(None, 5000)]	0	[]
flatten (Flatten)	(None, 1577088)	0	['conv2d[0][0]']
dense (Dense)	(None, 64)	320064	['input_2[0][0]']
concatenate (Concatenate)	(None, 1577152)	0	['flatten[0][0]', 'dense[0][0]']
dense_1 (Dense)	(None, 128)	201875584	['concatenate[0][0]']
dense_2 (Dense)	(None, 5)	645	['dense_1[0][0]']

=====  
Total params: 202,197,189  
Trainable params: 202,197,189  
Non-trainable params: 0  
=====

**B**



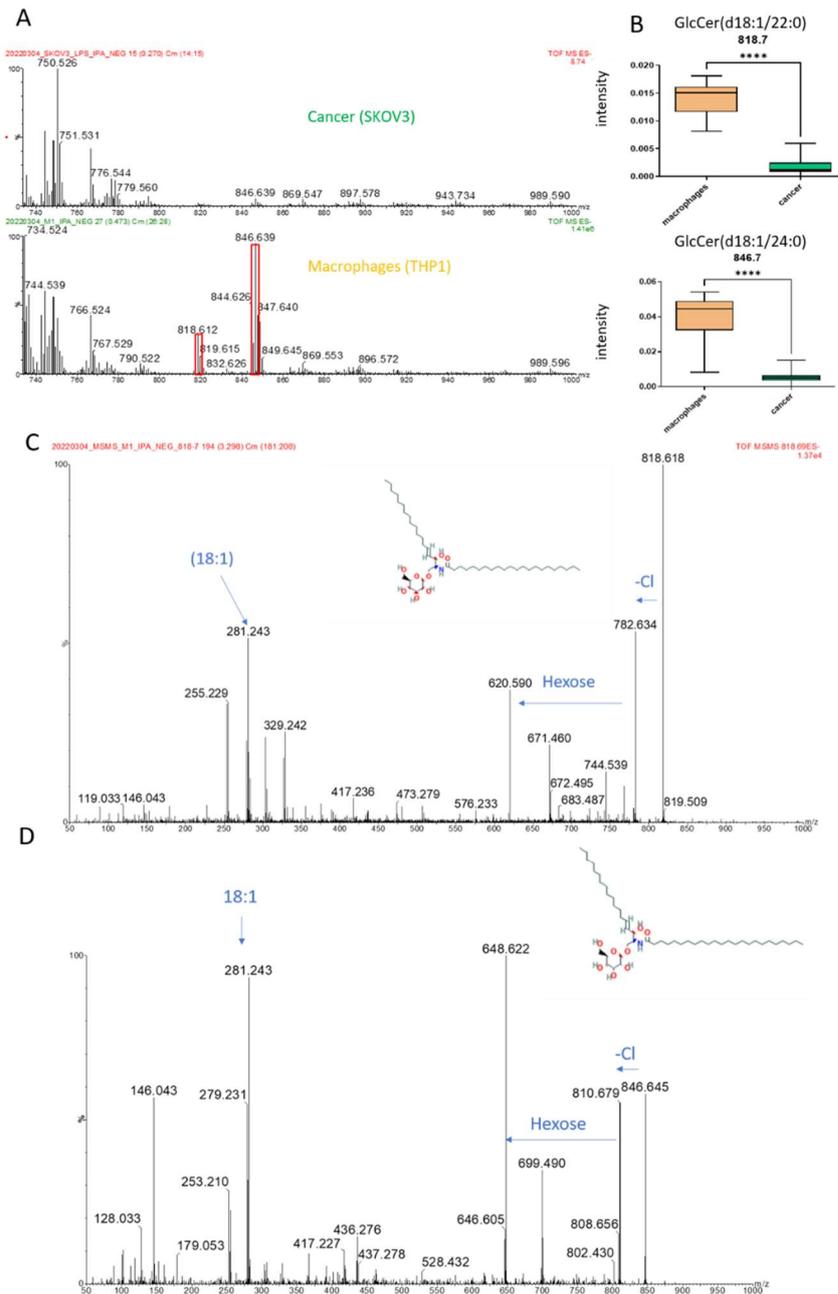
**Figure S5: Architecture of the multi-input model. (A) Dual-input model architecture. (B) Accuracy and loss of training and validation sets according to epoch. (Related to figure 4)**



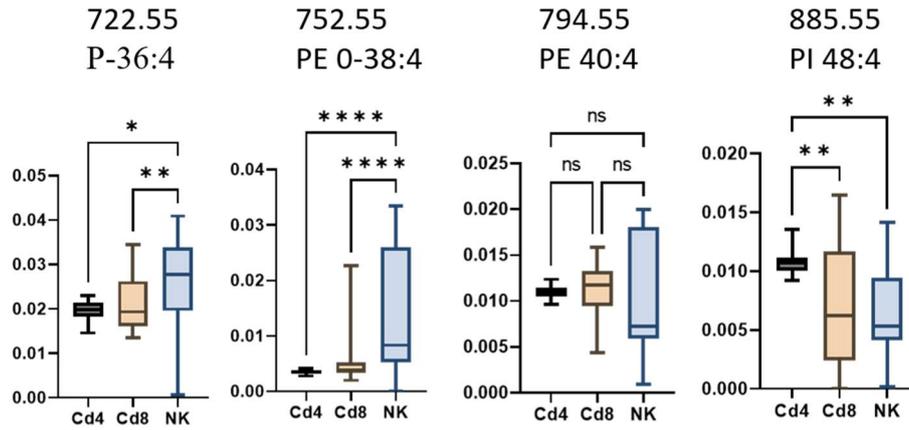
**Figure S6: Cross-validation of the lipid markers of the different OC subtypes by MALDI-MSI and SpiderMass.** (A) Example of a cross-validated marker at  $m/z$  700.55 which is specific to endometrioid. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the endometrioid tissues.

(B) Example of a cross-validated marker at  $m/z$  862.65 which is specific to mucinous carcinoma. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the mucinous tissues.

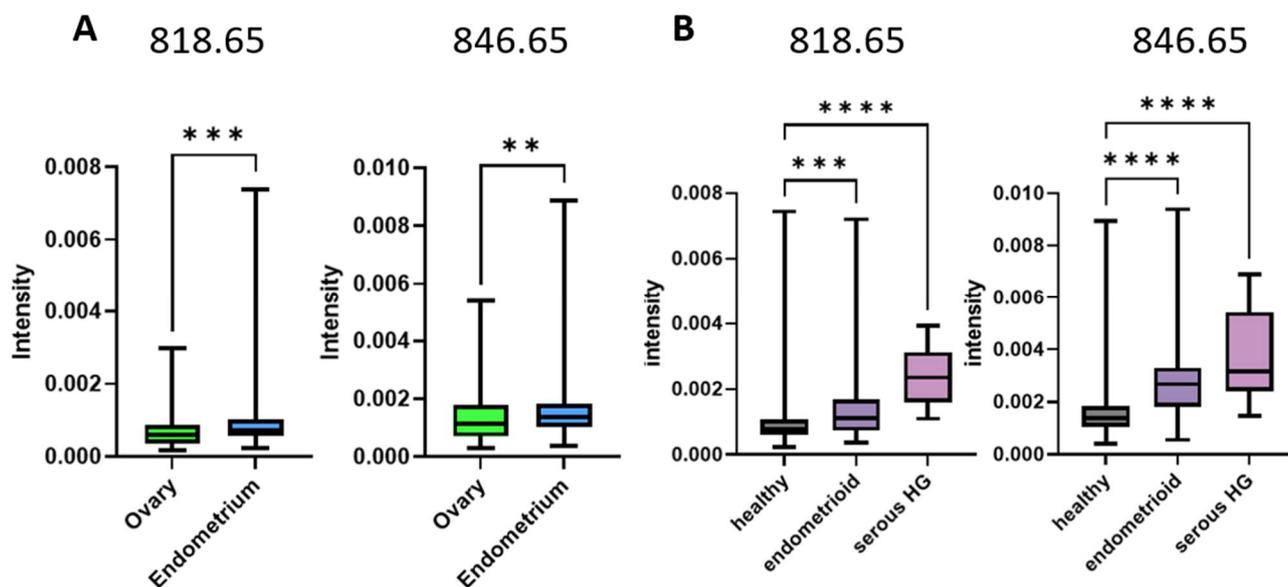
(C) Example of a cross-validated marker at  $m/z$  748.55 which is specific to endometrioid and normal tissues. The ion positive contribution by LIME is represented by a (+) and the negative contribution by a (-). A Kruskal-wallis test was performed on the SpiderMass data for this ion and the relative intensities are represented as a boxplot. (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). MALDI- and SpiderMass-MSI confirmed the over-regulation of this ion in the endometrioid and healthy tissues. (related to figure 4)



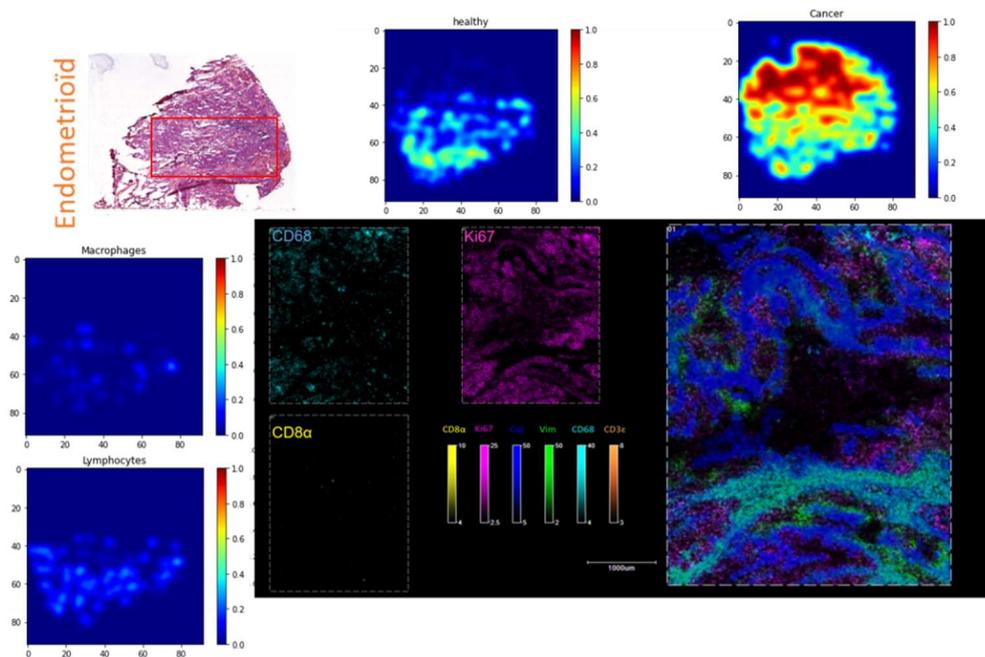
**Figure S7: MS/MS identification and variation of abundance of two Hexosylceramides between cancer cells and macrophages.** (A) SpiderMass mean spectra obtained from the THP1 macrophage versus SKOV3 cancer cell lines. The red frames indicate the ions  $m/z$  818.65 and  $m/z$  846.65 specific to the macrophages. (B) Boxplot based on the relative intensities of these two ions for SKOV3 vs. THP1 (\*\*\*\*= $p$ value $<0.0001$ ) showing the higher abundance of these two markers in the macrophages. (C) MS/MS spectrum of the ion  $m/z$  818.65 identified as GlcCer d40:1 (d18:1/22:0). (D) MS/MS spectrum of the ion  $m/z$  846.65 identified as GlcCer d42:1 (d18:1/24:0). (related to figure 5)



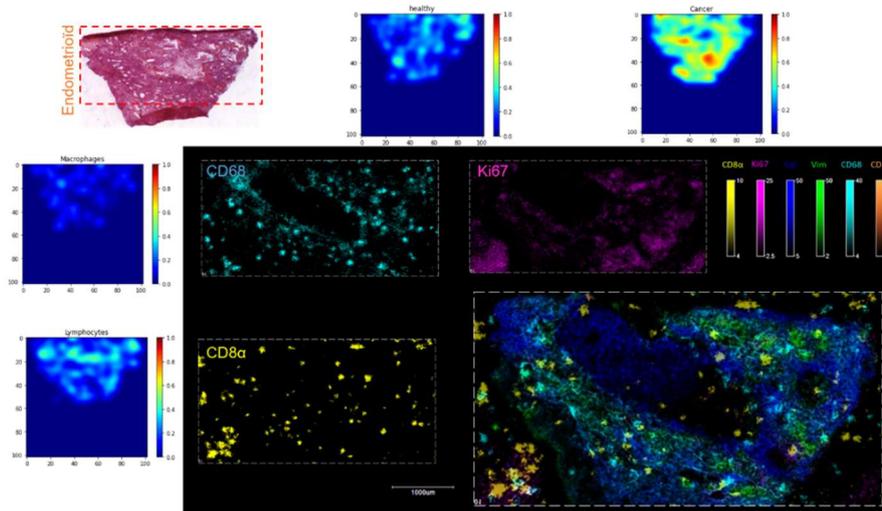
**Figure S8: Discriminative lipid markers of lymphocyte cells.** Boxplots based on the relative intensities of ions  $m/z$  722.55, 752.55, 794.55 and 885.55 plotted for the different subpopulation of lymphocytes (NK, CD8 or CD4) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (Related to figure 5)



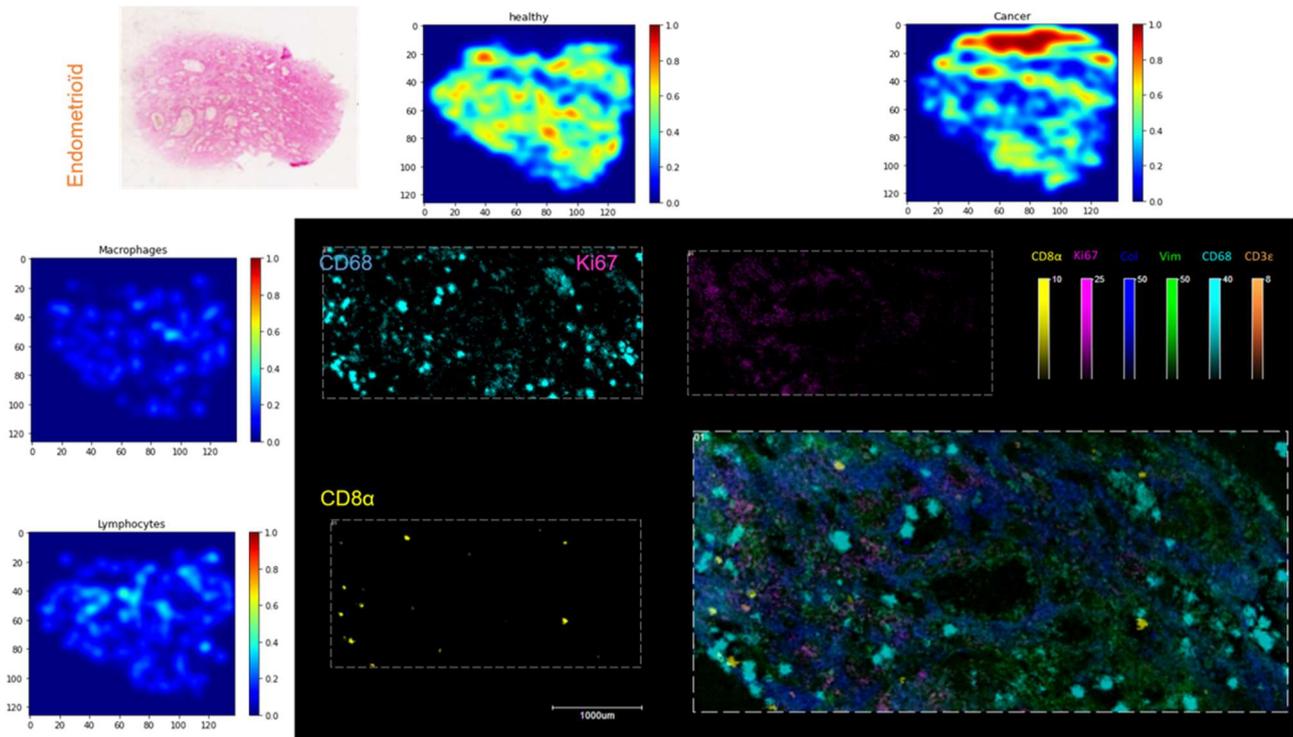
**Figure S9: Abundance of lipid markers specific to macrophages in endometrium cancers vs normal ovary and endometrium tissues. Boxplots based on relative intensities of ion m/z 818.65, 846.65. (A) In normal endometrium and ovary tissues (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (B) In the different endometrium cancer subtypes (Healthy, HGSOE or endometrioid) (\*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ) NS for  $p > 0.05$ ). (Related to figure 5)**



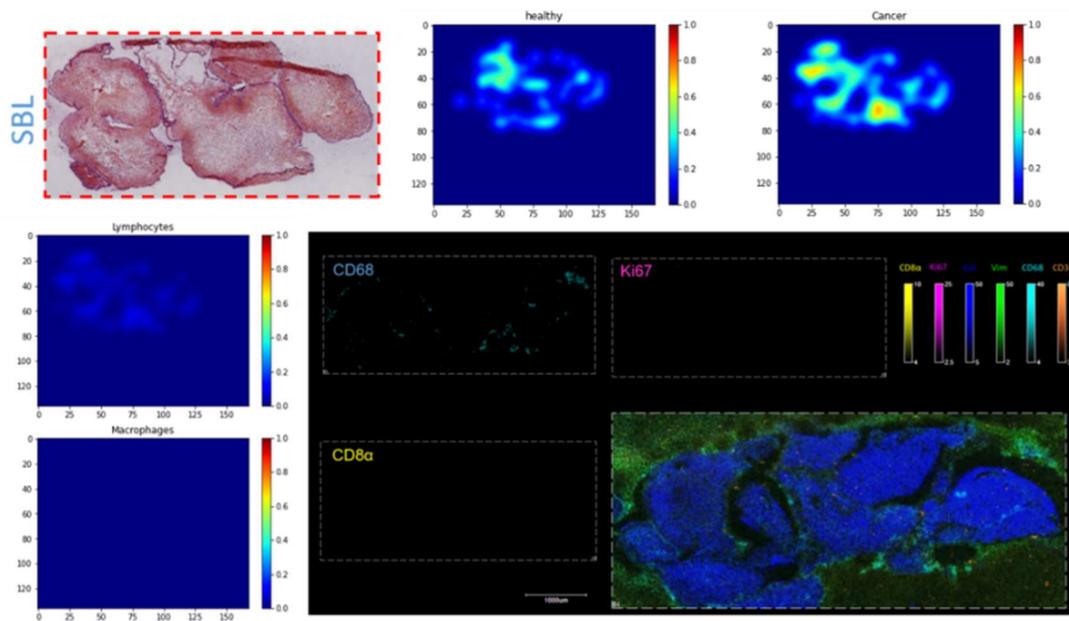
**Figure S10: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



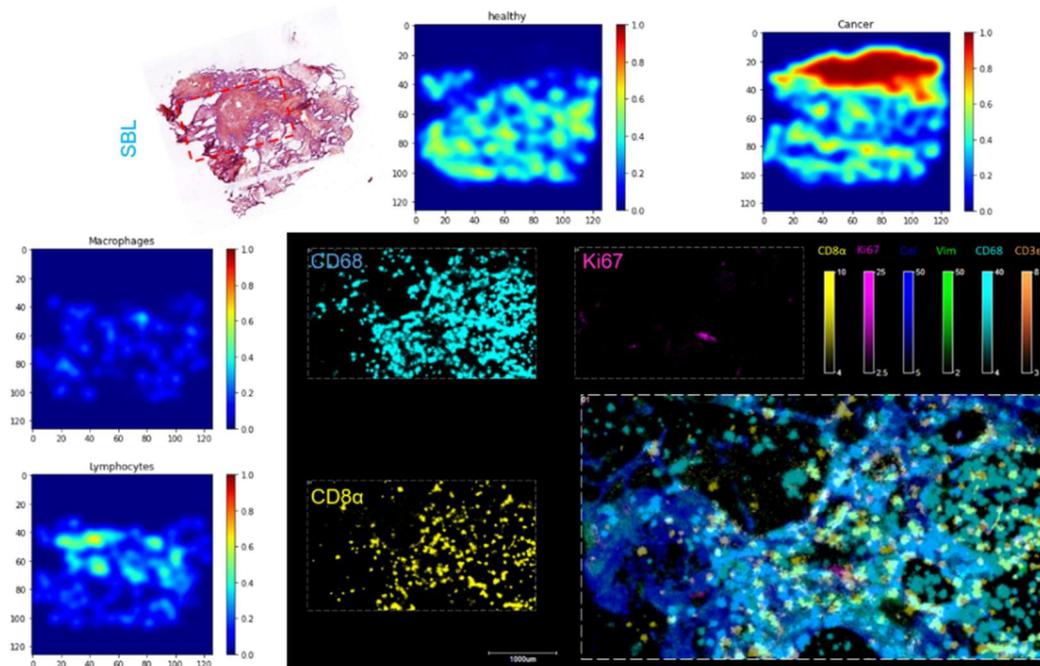
**Figure S11: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



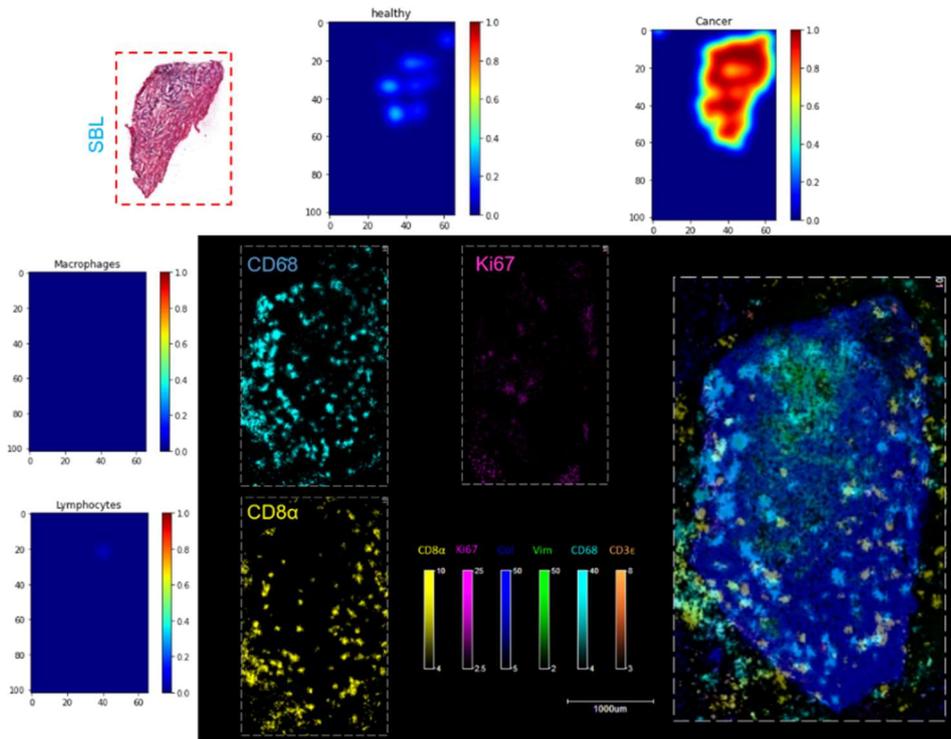
**Figure S12: Diagnosis of the endometrioid carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in endometrioid carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTC8) and anti-CD68 (macrophages). (Related to figure 7)



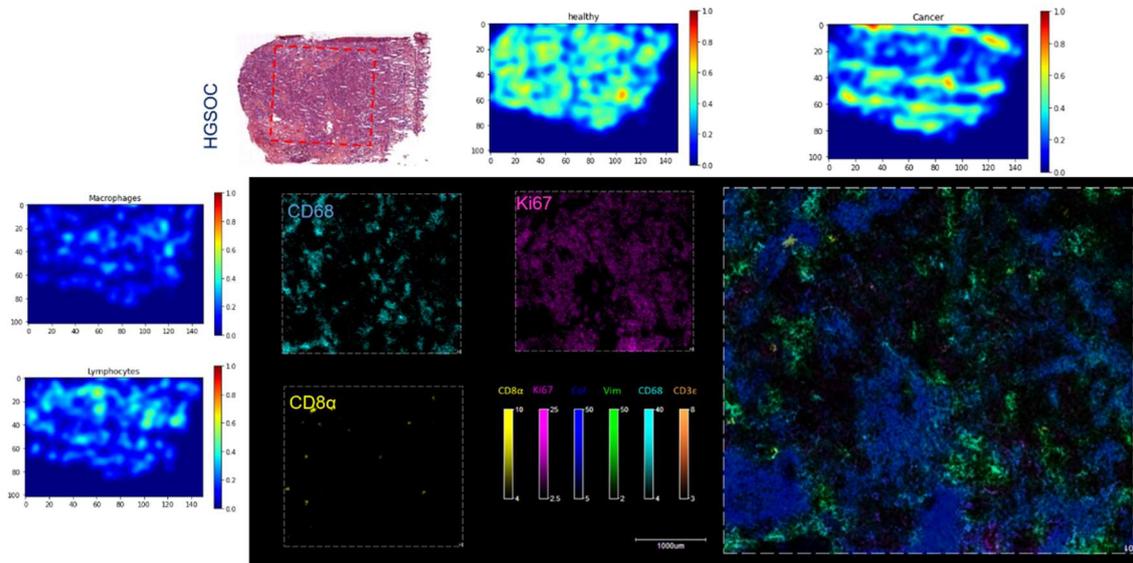
**Figure S13: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



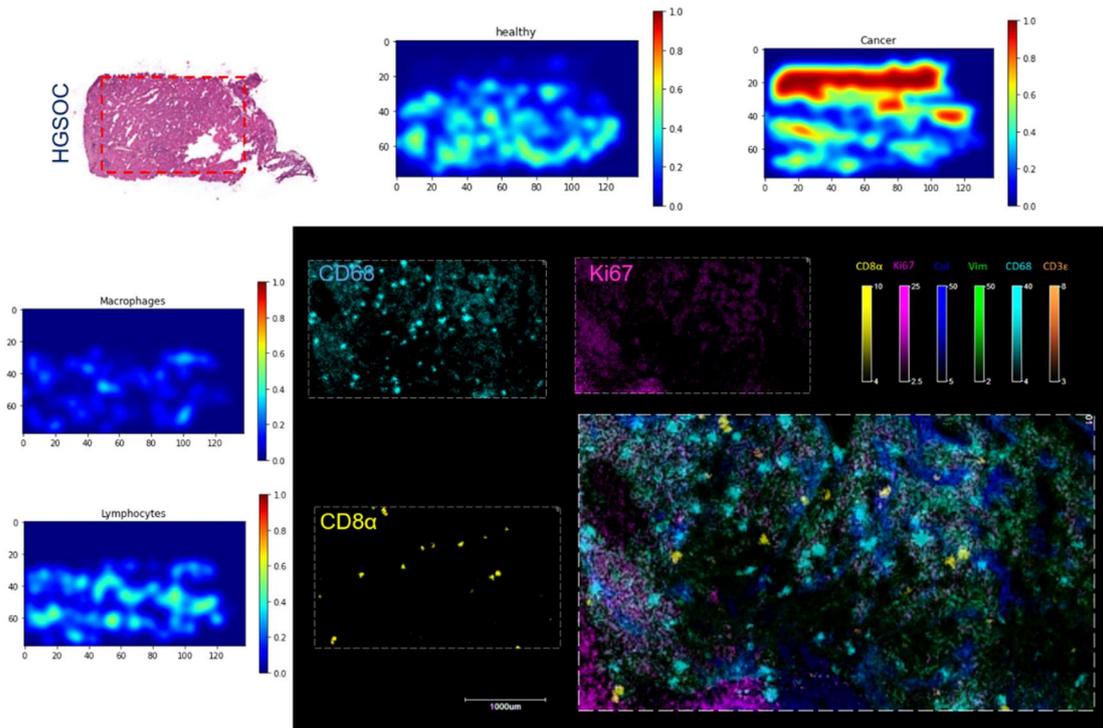
**Figure S14: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



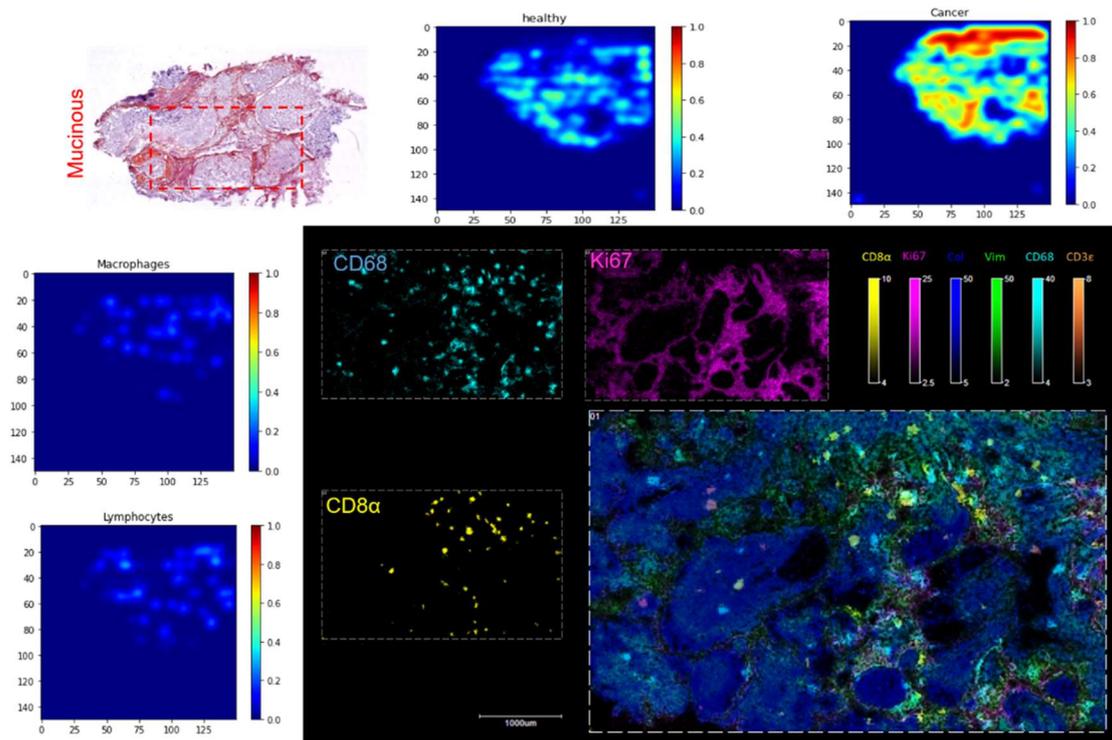
**Figure S15: Diagnosis of the serous borderline carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous borderline carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



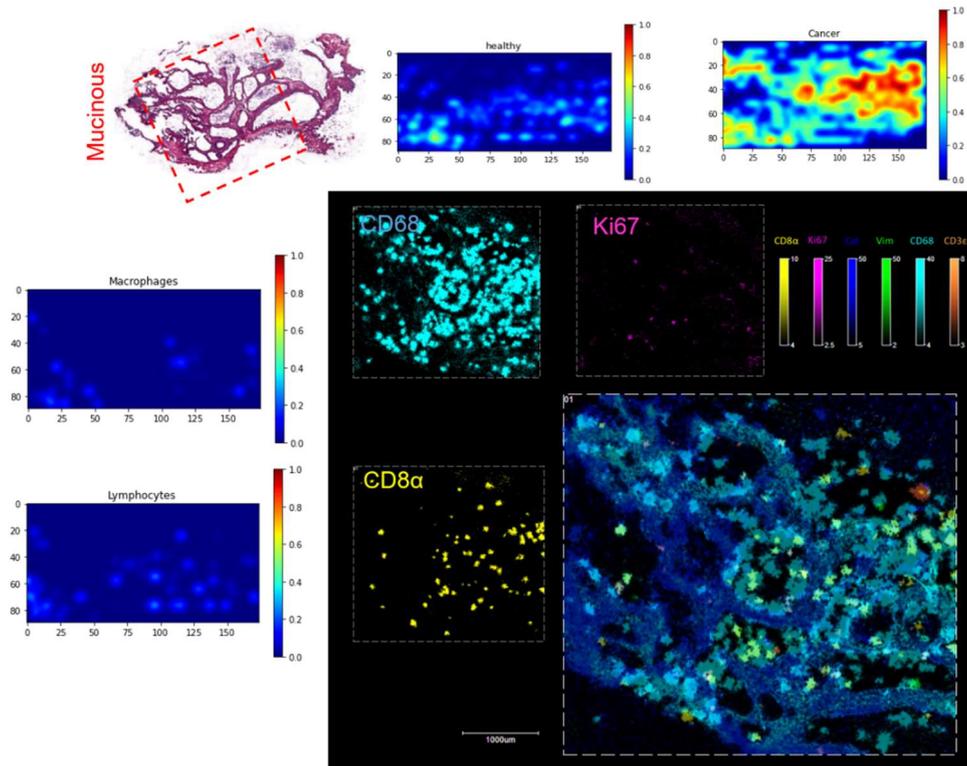
**Figure S16: Diagnosis of the serous high-grade carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



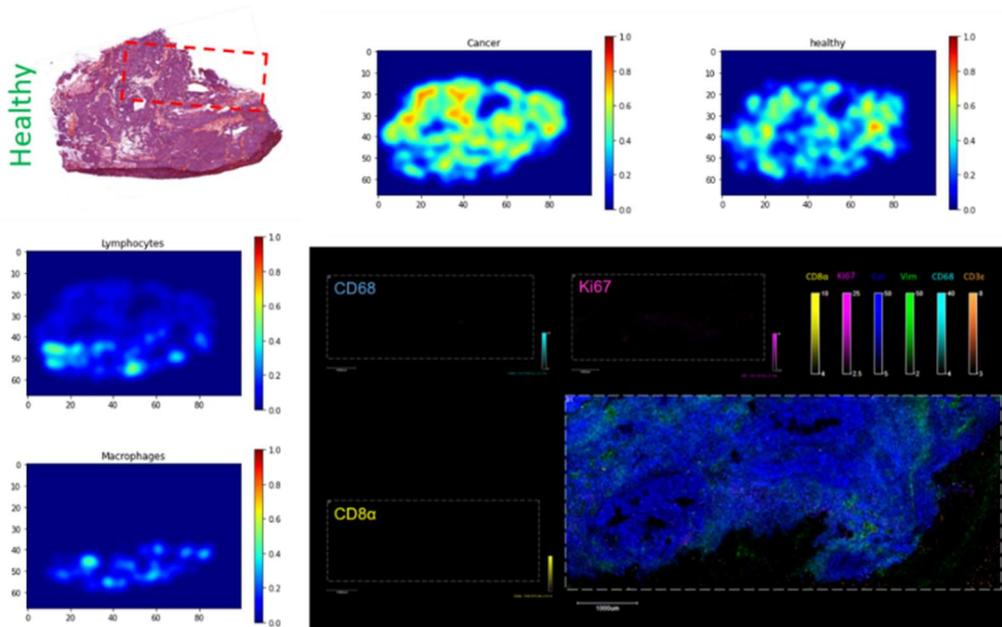
**Figure S17: Diagnosis of the serous high-grade carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in serous high-grade carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTC8) and anti-CD68 (macrophages). (Related to figure 7)



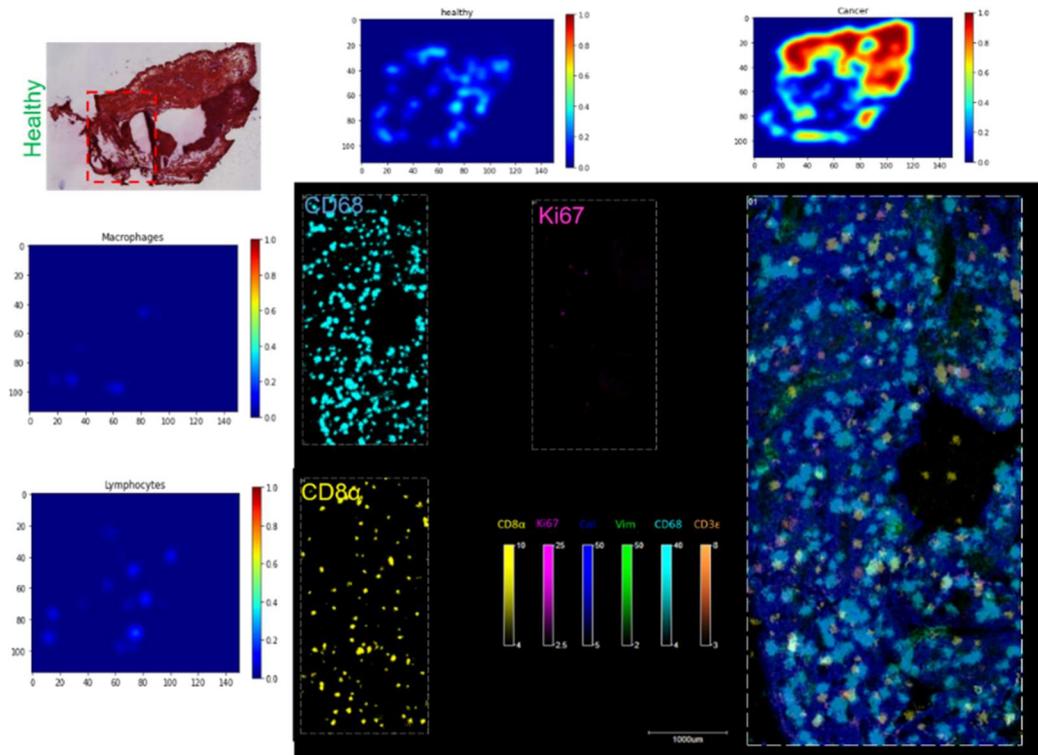
**Figure S18: Diagnosis of the mucinous carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



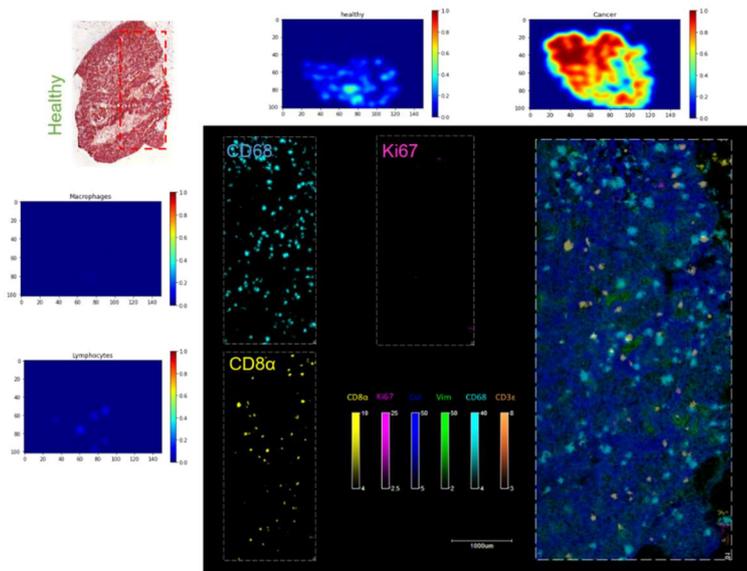
**Figure S19: Diagnosis of the mucinous carcinoma based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in mucinous carcinoma obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



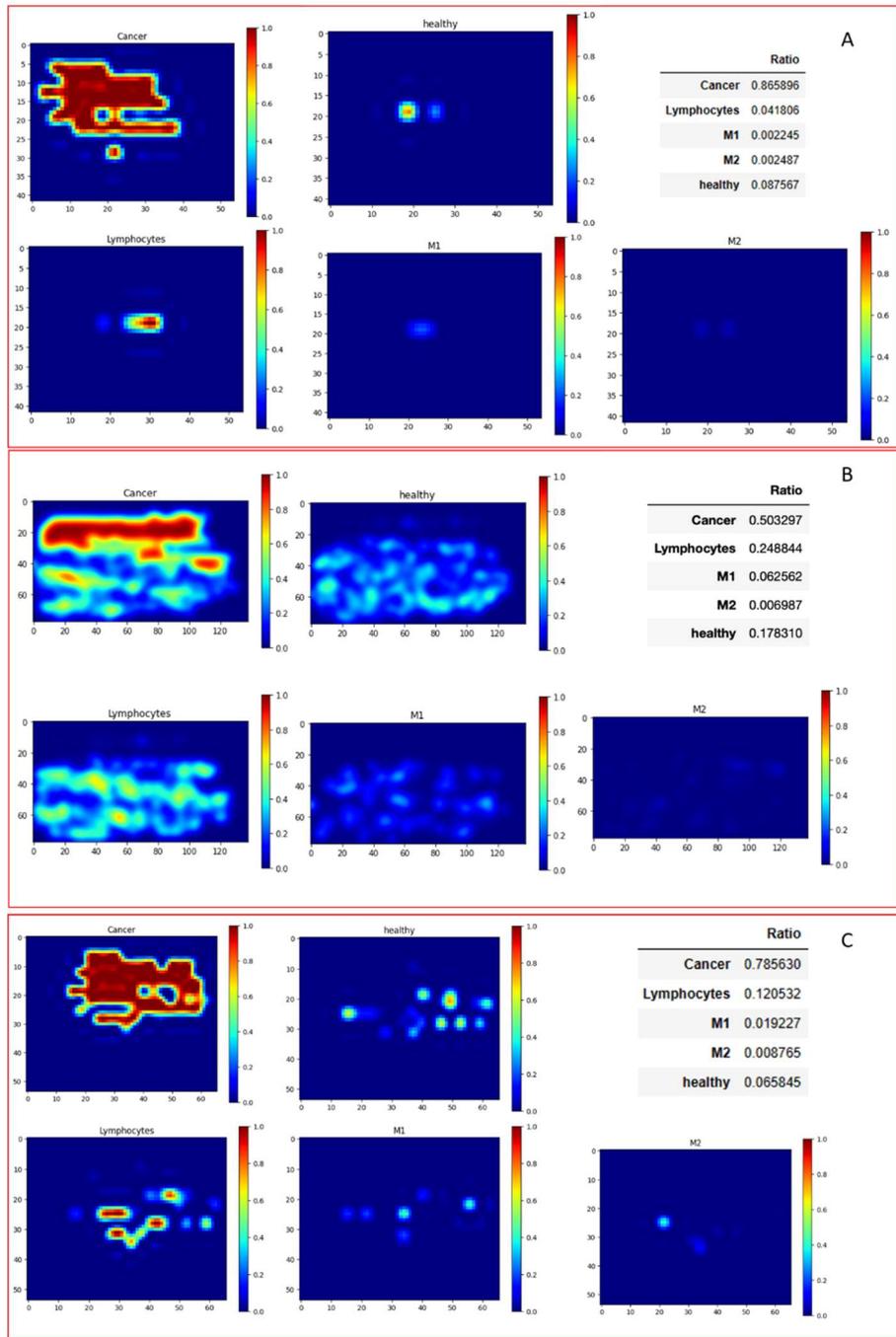
**Figure S20: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



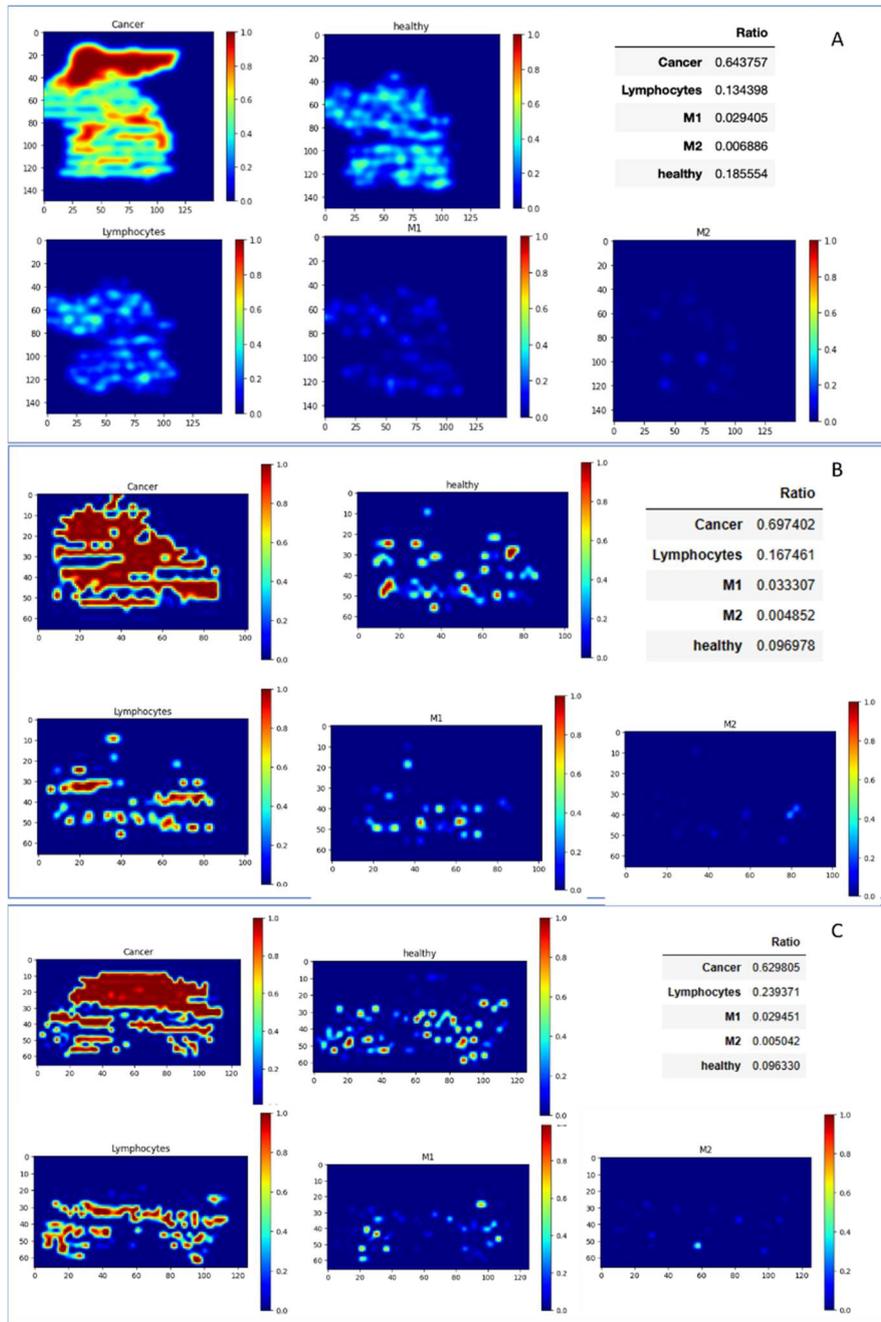
**Figure S21: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



**Figure S22: Diagnosis of the normal tissue based on predicted presence of based on the different immune cell populations and cross-validation by MALDI-IHC.** Predicted presence of the different cell populations (Lymphocyte, macrophages, cancer and normal cells) in normal tissue obtained using the immune score model (LGBM) from SpiderMass-MSI data. The predictions were cross-validated by MALDI-IHC performed from the region of this tissue indicated by the red square. MALDI-IHC was performed using anti-Ki67 (cancer), anti-CD8 $\alpha$  (LTCD8) and anti-CD68 (macrophages). (Related to figure 7)



**Figure S23: Predicted probability of the immune cell presence using LGBM model based on SpiderMass-MSI for patients with OS<42 months.** Immune score calculated from SpiderMass-MSI data for three patients with OS<42 months (A) Patient 89-3. (B) Patient 62-1. (C) Patient 114.1. (Related to figure 8)



**Figure S24: Predicted probability of immune cell presence using LGBM model based on SpiderMass-MSI for patients with OS>50 months.** Immune score calculated from SpiderMass-MSI data for three patients with OS>50 months (A) patient 102-1. (B) Patient 108-1. (C) Patient 44-2. (Related to figure 8)

**Supplementary Table:**

**Supp table 1: Clinical data of patients included in the FF cohort. (related to figure 2)**

<b>Serous High Grade</b>	<b>age</b>	<b>recurrence</b>	<b>survival</b>	<b>Death</b>	<b>First Treatment</b>
47	37	5	28	yes	chemiotherapy
48	61	33	86	no	chemiotherapy
44	58	52	52	no	surgery
62	60	7	42	yes	surgery
63	41	20	43	yes	chemiotherapy
86	59	17	63	yes	surgery
89	69	8	42	yes	chemiotherapy
95	65	8	60	yes	surgery
114	64	25	40	yes	surgery
42	50	84	84	no	surgery
187	75	12	24	yes	chemiotherapy
183	41	24	54	no	chemiotherapy
<b>Serous Borderline</b>					
13	38	17	63	yes	surgery
22	52	92	92	no	surgery
64	51	70	70	no	surgery
65	52	63	63	no	surgery
113	26	6	66	no	surgery
174	67	6	48	no	surgery
<b>Mucinous</b>					
6	70	94	94	no	surgery
8	54	7	7	no	surgery
23	43	17	31	no	chemiotherapy
55	86	21	21	no	surgery
68	52	4	4	no	surgery
84	54	51	51	no	surgery
103	19	64	64	no	surgery
135	53	59	59	no	surgery
<b>Endometrioid</b>					
108	58	18	57	non	surgery
179	67	16	38	yes	surgery
181	49	43	43	non	surgery
<b>Healthy</b>					
5	24	no	no	no	surgery
20	53	6	22	yes	chemiotherapy
22	52	no	no	no	surgery
48	61	33	no	no	surgery
51	65	no	72	yes	surgery
52	66	39	68	yes	surgery
53	57	no	no	no	surgery
54	38	25	no	no	chemiotherapy
56	81	no	no	no	chemiotherapy
59	53	no	no	no	surgery

61	56	12	24	yes	chemiotherapy
63	41	20	43	yes	chemiotherapy
65	52	no	no	no	surgery
66	74	8	42	yes	surgery
67	78	no	no	no	surgery
69	66	18	no	no	surgery
71	78	6	no	no	chemiotherapy
73	63	no	no	no	surgery
74	69	4	11	yes	chemiotherapy
76	76	23	61	yes	chemiotherapy
102	73	no	no	no	surgery
114-b	64	25	60	yes	surgery
116	62	13	no	no	chemiotherapy
137	72	no	no	no	surgery
139	73	no	no	no	surgery
141	74	6	no	no	surgery
145	78	no	45	yes	surgery
146	75	no	no	no	surgery
177	52	no	no	no	surgery
179	67	16	38	yes	surgery
181	49	43	43	non	surgery
183	78	24	46	yes	chemiotherapy
187	77	12	30	yes	chemiotherapy
81	49	no	no	no	surgery
82	42	no	no	no	surgery
87	26	47	no	no	surgery
97	60	no	no	no	surgery
99	76	6	29	yes	surgery

**Supp table 2: Discriminative lipids identified by MS/MS for the different OC subtypes. (Related to figure 2)**

<b>Mode</b>	<b>sample</b>	<b>Potential identification</b>	<b>Peak m/z</b>
<b>Negative</b>	Normal	PE (16:0-22:5)	748,55
<b>[M-H]-</b>	Normal	PS (18:0/18:1)	788,55
	Normal	PI (20:4/18:0)	885,55
	Normal	PA (18:1:18:0)	701,55
	Serous BL	PI (22:1/18:0)	919,79
	Serous BL	TG (56:12)	893,63
	Serous BL	PE (18:0/22:4)	794,55
	Serous HG	PE (18:0/16:0)	718,55
	Serous HG	PE(18:0/18:2)	742,55
	Serous HG	PA (18:1/20:2)	725,55
	Serous HG	PE (18:0/18:1)	744,55
	Serous HG	PA (18:0/22:6)	747,55
	Mucinous	PA (18:0/20:2)	727,55
	Mucinous	PS (22:2/20:4)	862,65
	Endometrioid	PS(22:0/22:2) or PS (22:1/22:1)	898,55
<b>Positive</b>	Serous BL/HG	PE (16:0/20:0)	748,55
<b>[M-H]-</b>	Serous BL/HG	PE (18:2/22:4)	792,55
	Serous BL	PS (P-16:0/18:2)	744,55
	Serous BL	PS (18:4/18:1)	782,55
	Serous HG	PS (0-16:0V20:1)	776,55
	Serous HG	PE (0-18:0/18:2)	730,55
	Mucinous	PS (0-18:0/18:1)	776,55
	Mucinous	PC (20:1/16:1)	788,65
	Mucinous	PE (0-18:0/20:4)	752,55
	Mucinous	PC (18:4/12:0)	698,55
	Endometrioid	PE (16:0/20:5)	724,55

**Supp table 3: Clinical data of patients included in the FFPE cohort. (Related to figure 2)**

FFPE tissues cohort						
Class	Patient	Age	recurrence	recurrence	Overall survival	treatment
CARCINOSARCOME	4	57	YES	1	15	surgery
CCC	26	NA	NO	36	36	surgery
CCC	32	NA	NO	31	31	surgery
CCC	40	NA	YES	13	15	surgery
CCC	42	68	YES	1	2	surgery
CCC	74	74	NO	5	5	surgery
ENDOMETRIOIDE	51	51	NO	48	48	surgery
ENDOMETRIOIDE	55	55	YES	16	32	surgery
ENDOMETROID	13	NA	NO	48	48	surgery
ENDOMETROIDE	37	59	NO	22	22	surgery
HGSOC	1	66	YES	39	50	surgery
HGSOC	2	67	NO	27	27	surgery
HGSOC	3	NA	NO	14	14	surgery
HGSOC	5	NA	NO	25	25	surgery
HGSOC	6	48	NO	36	36	surgery
HGSOC	7	NA	NO	3	3	surgery
HGSOC	8	66	YES	13	56	surgery
HGSOC	9	67	YES	28	45	surgery
HGSOC	10	NA	NO	55	55	surgery
HGSOC	11	61	YES	15	48	surgery
HGSOC	12	NA	NO	36	36	surgery
HGSOC	14	48	YES	22	46	surgery
HGSOC	15	74	NO	47	47	surgery
HGSOC	16	58	YES	39	45	surgery
HGSOC	19	NA	NO	32	32	surgery
HGSOC	20	74	NO	41	41	surgery
HGSOC	22	73	YES	14	41	surgery
HGSOC	23	NA	NO	38	38	surgery
HGSOC	24	NA	NO	39	39	surgery
HGSOC	27	78	NO	37	37	surgery
HGSOC	28	58	YES	10	30	surgery
HGSOC	29	NA	NO	38	38	surgery
HGSOC	30	71	YES	8	38	surgery
HGSOC	33	65	YES	24	31	surgery
HGSOC	34	NA	NO	31	31	surgery
HGSOC	35	62	NO	27	27	surgery
HGSOC	38	61	NO	20	20	surgery
HGSOC	39	62	NO	18	18	surgery
HGSOC	41	77	NO	12	12	surgery
HGSOC	43	69	NO	5	5	surgery
HGSOC	44	77	NO	3	3	surgery

HGSOC	45	65	NO	3	3	surgery
HGSOC	46	89	NO	2	2	surgery
HGSOC	49	80	NO	1	1	surgery
HGSOC	50	50	YES	9	48	surgery
HGSOC	52	52	NO	40	40	surgery
HGSOC	53	53	YES	31	57	surgery
HGSOC	54	54	YES	14	17	surgery
HGSOC	56	56	YES	29	30	chemotherapy
HGSOC	58	58	NO	6	6	chemotherapy
HGSOC	59	59	NO	20	20	surgery
HGSOC	60	60	NO	41	41	surgery
HGSOC	61	61	NO	6	6	surgery
HGSOC	62	62	YES	19	35	surgery
HGSOC	67	67	NO	26	26	surgery
HGSOC	69	70	NO	19	19	surgery
HGSOC	72	72	YES	13	22	surgery
HGSOC	73	73	NO	35	35	surgery
HGSOC	77	77	YES	24	34	surgery
HGSOC	82	82	NO	3	6	chemotherapy
HGSOC	84	84	YES	5	32	surgery
HGSOC	85	85	NO	14	14	surgery
Mucinous	57	57	YES	2	4	surgery
Mucinous	75	75	NO	36	36	surgery
Mucinous	76	76	YES	2	4	surgery
SBL	18	55	YES	20	28	surgery
SBL	68	68	NO	44	44	surgery
SLG	21	25	NO	37	37	surgery
SLG	48	16	NO	1	1	surgery
SLG	65	65	YES	3	6	surgery
SLG	66	66	YES	3	6	surgery
SLG	70	69	NO	6	6	surgery
SLG	71	71	NO	6	6	surgery
SLG	78	78	YES	6	54	surgery
SLG	79	79	YES	6	54	surgery
SLG	80	80	NO	42	42	surgery
SLG	81	81	NO	25	25	surgery
SLG	83	83	YES	13	32	chemotherapy

Supp table 4: List of classifiers tested on the FF dataset with their F1-score and 5K-Fold cross-validation. (Related to figure 2)

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
RidgeClassifierCV	1.00	1.00	None	1.00	0.12
RidgeClassifier	1.00	1.00	None	1.00	0.12
LGBMClassifier	0.98	0.96	None	0.98	2.02
LinearSVC	0.95	0.95	None	0.96	0.95
SGDClassifier	0.95	0.95	None	0.96	0.16
PassiveAggressiveClassifier	0.95	0.95	None	0.96	0.33
Perceptron	0.88	0.90	None	0.89	0.15
LogisticRegression	0.93	0.88	None	0.93	0.36
DecisionTreeClassifier	0.86	0.86	None	0.87	0.29
ExtraTreesClassifier	0.91	0.84	None	0.91	0.27
GaussianNB	0.88	0.82	None	0.88	0.11
RandomForestClassifier	0.84	0.75	None	0.83	0.42
LinearDiscriminantAnalysis	0.84	0.72	None	0.83	0.23
BaggingClassifier	0.79	0.72	None	0.79	0.96
CalibratedClassifierCV	0.84	0.71	None	0.83	2.69
NearestCentroid	0.67	0.67	None	0.69	0.10
BernoulliNB	0.56	0.67	None	0.57	0.13
KNeighborsClassifier	0.74	0.57	None	0.73	0.10
SVC	0.74	0.56	None	0.70	0.16
ExtraTreeClassifier	0.63	0.49	None	0.61	0.09
AdaBoostClassifier	0.67	0.48	None	0.63	2.48
QuadraticDiscriminantAnalysis	0.37	0.27	None	0.38	0.13
LabelSpreading	0.09	0.20	None	0.02	0.11
LabelPropagation	0.09	0.20	None	0.02	0.11
DummyClassifier	0.53	0.20	None	0.37	0.09

**Supp table 5: List of the samples used for the validation test in blind. (Related to figure 3)**

<b>Sample List</b>	<b>Age</b>	<b>Class</b>	<b>recurrence</b>	<b>OS</b>	<b>Death</b>	<b>FIGO</b>	<b>First treatment</b>
102.1	64	HGSOC	52	52	no	IIC	Surgery
102.2	64	Normal			no		Surgery
104.1	44	HGSOC	24	60	no	IIC	chemotherapy
104.2	44	Normal			no		chemotherapy
106.1	38	HGSOC	48	48	no	IIC	chemotherapy
106.2	38	Normal			no		chemotherapy
108.1	46	HGSOC	50	50	no	IIC	surgery
108.2	46	Normal			no		surgery
109.1	48	endometrioid	20	20	no	IA	surgery
109.2	48	Normal			no		surgery
110.1	61	HGSOC	33	33	no	IIC	chemotherapy
110.2	61	Normal			no		chemotherapy
220.1	62	Endometrioid			no		chemotherapy
220.2	62	Normal			no		chemotherapy
221.1	72	HGSOC	6	6	no	IIC	surgery
222.1	55	Endometrioid			no		surgery
222.2	55	Normal			no		surgery
223.1	65	SBL	7	7	no	IIC	surgery
223.2	65	Normal			no		surgery
224.1	68	Mucinous	8	8	no	IIC	chemotherapy
224.2	68	Normal			no		chemotherapy
225.1	60	HGSOC	6	6	no	IIC	surgery
226.1	64	Normal			no		surgery
227.1	56	endometrioid	11	11	no	IIC	chemotherapy
228.1	67	HGSOC	11	11	no	IVB	chemotherapy
229.1	65	Mucinous	2	2	no	IA	surgery

**Supp table 6: Samples used to SpiderMass-MSI and create the immune scores. (Related to figure 8)**

<i>Survival</i>	<i>Sample</i>	<i>age</i>	<i>recurrence</i>	<i>OS</i>	<i>Death</i>	<i>First treatment</i>
<i>Long survival &lt;50 months</i>	89-3	59	17	42	yes	surgery
	66-4	74	8	42	yes	surgery
	114.1	64	25	40	yes	surgery
	62.1	60	7	42	yes	surgery
<i>Short survival &gt;50 months</i>	102.1	64	52	52	no	Surgery
	108.1	46	50	50	no	surgery
	44-2	50	51	51	no	surgery
	42-1	50	84	84	no	surgery

**Supp table 7: Quantification of macrophages in the different OC subtypes calculated from MALDI-IHC. (Related to figure 7)**

Ki67 marker	SBL	HGSOC	Mucinous	Endo	Heal
Quantification	19,59	47,25	17,64	8,08	0,02
	42,01	22,04	42,64	15,2	6
	1,1	15,05	10,38	16,31	26,18
Mean	20,9	28,11333333	23,55333333	13,19666667	10,73333333
Median	19,59	22,04	17,64	15,2	6

# Real-time Glioblastoma TME Analyses by SpiderMass for Surgery Decision Making and Patient Management

Yanis Zirem<sup>1</sup>ϕ, Léa Ledoux<sup>1</sup>ϕ, 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\*ϕ</sup>, Isabelle Fournier<sup>1,8</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 CRISAL, F-59000 Lille, France

<sup>4</sup>Department of 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

ϕThese authors contributed equally.

\*Corresponding author : Prof. Michel Salzet, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) – Inserm U1192 – Université de Lille, Bât SN3, 1<sup>er</sup> étage, Cité Scientifique, F-59655 Villeneuve d’Ascq Cedex, France. Phone: +33 (0)3 20 43 41 94; Fax: +33 (0)3 20 43 40 54; email: [michel.salzet@univ-lille.fr](mailto:michel.salzet@univ-lille.fr)

⌘Lead contact : Prof. Isabelle Fournier, Laboratoire Protéomique, Réponse Inflammatoire et Spectrométrie de Masse (PRISM) - Inserm U1192 - Université de Lille, Bât SN3, 1<sup>er</sup> étage, Cité Scientifique, F-59655 Villeneuve d’Ascq Cedex, France. Phone: +33 (0)3 20 43 41 94; Fax: +33 (0)3 20 43 40 54; email: [isabelle.fournier@univ-lille.fr](mailto:isabelle.fournier@univ-lille.fr)

## Summary

Glioblastoma is a highly heterogeneous and infiltrative form of brain cancer associated to a poor outcome with a limited efficiency of therapies. The extent of the surgery is known to be related to the patient survival. Reaching an accurate diagnostic and prognostic by the time of the initial surgery is therefore paramount in the management of glioblastoma. To this end, we have studied the performances of the SpiderMass, an ambient ionization mass spectrometry technology that can be used *in vivo* without invasiveness, coupled to a newly developed AI pipeline. We demonstrated that we can, both stratify IDH-wild type glioblastoma patients into molecular sub-groups and get an accurate diagnostic with over 90% accuracy after cross-validation. Interestingly, the developed method offers the same accuracy for prognosis. Additionally, we tested the potential of an immune-scoring strategy based on SpiderMass fingerprints, showing the association between prognosis and immune cell infiltration, to predict the patient outcome.

## Keywords

Mass Spectrometry • Water-Assisted Laser Desorption/Ionization • SpiderMass • Glioblastoma • Cancer diagnostic • Lipids • Prognosis • Machine Learning • Immunoscore • Imaging • MALDI-IHC

## Highlights

- SpiderMass combined to AI provides GBM diagnosis with over 90% accuracy;
- 99 lipid markers are identified in relation to diagnosis, prognosis and immune cells;
- An immuno-score based on SpiderMass data is developed;
- Immuno-score applied to SpiderMass MS Imaging data provides access to patient prognosis.

## • 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. With a prevalence of 80 to 85% gliomas remains the most common form of all adult brain tumors<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 possible best 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 Central Nervous System<sup>5</sup> which considers further the importance of the molecular features while keeping the histology. The first criterion of classification is between the isocitrate dehydrogenase (IDH)-wild type versus the IDH1- or IDH2- mutants which defines 2 classification groups. IDH-mutant status is not only important to classification but it is also associated with a better prognosis<sup>6</sup>. The IDH-wild type are themselves sub-classified between glioblastoma (GBM) and astrocytoma/not elsewhere classifiable (NEC). Within the adult-type diffuse gliomas, glioblastoma is the most prevalent and aggressive form with an overall survival (OS) of about 15 months and a median of 14-24 months (contrarily to lower-grade gliomas which can be controlled for decades) and is classified as a high-grade malignant glioma (grade IV). If 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 (MVP) or at least one of the following specific molecular characteristics i.e., EGFR amplification, TERT promoter mutation and a concurrent gain of chromosome 7 and loss of chromosome 10 (+7/-10). GBM 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 adjuvant long-course chemotherapy by temozolomide combined with radiotherapy (known as Stupp protocol<sup>9</sup>) which has remained unchanged for many years. Treatment of glioblastoma clearly meets different challenges. Indeed, novel immunotherapies, e.g., PDL-1, have shown disappointing results in glioblastoma. One of the hurdles of glioblastoma treatment is associated to the blood-brain-barrier (BBB) which limits the passage of drugs into the brain preventing the administration of many chemotherapies without serious secondary effects. Another is the heterogeneity of the tumor and signaling pathway associated which make targeting therapeutic targets more complicated. Besides, the tumor microenvironment (TME) promotes the resistance and decrease the efficiency of the patient immunological response. Indeed, the development of brain tumors disrupts the homeostasis of the central nervous system. The vasculature, the extracellular matrix and the blood-brain barrier are hijacked and integrated into the processes of tumor development<sup>10</sup>. It is known that the tumor environment is immunosuppressive and that infiltration of different sub-population of immune cells are correlated to both the response to the therapy and the patient overall survival.

On the other hand, extent of resection (EOR) has been demonstrated to be associated with OS and higher OS is observed for gross total resection (GTR) over subtotal resection or biopsy<sup>11</sup>. In particular, increasing the EOR to >95% while sparing tissue to preserve functional integrity is proven to 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>12</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 developing to help with finding the resection margins<sup>13</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 because of their more aggressive nature, 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 with higher performances over 5-Ala<sup>14</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>15</sup>. A handheld Raman probe is also available and was shown to achieve a sensitivity of 93% and a specificity of 91% to classify normal brain from dense cancer and normal brain invaded by cancer cells<sup>16</sup>.

Therefore, both a molecular characterization of the brain tissues during the surgery to achieve tailored resection and a better understanding of the microenvironment could become essential in the neurosurgical management of patients with glioblastoma<sup>17</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.

Glioblastoma 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>18</sup>, identify potential biomarkers and therapeutic targets for the disease<sup>19,20</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>21</sup>. Intra-operative AIMS is achieved from freshly resected tissues<sup>22,23</sup> or *in vivo* to advise neurosurgeons on the extent of tumor resection, with real-time feedback<sup>24-26</sup>. SpiderMass is one of these AIMS technology that provides a minimally invasive solution for *in vivo* surgical applications<sup>27</sup>. It has the potential to become an essential tool for precise tissue characterization during excision surgery, because of the high specificity of its molecular profiles where the different markers are separated according to their mass (and charge) and can be identified. Interestingly, 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>28</sup>. As for Raman spectroscopy, MS should be combined with ML to provide an interpretable result to the practitioner. Thus, in cancer diagnosis-based MS, supervised ML algorithms play a critical role in building accurate classification models based on labelled datasets<sup>29</sup>. Choosing and fine-tuning the appropriate classification model can have a significant impact on the accuracy of the results<sup>30</sup>. Cross-analysis by non-supervised ML strategies as well as interpretability of the results are also paramount and must be carefully considered.

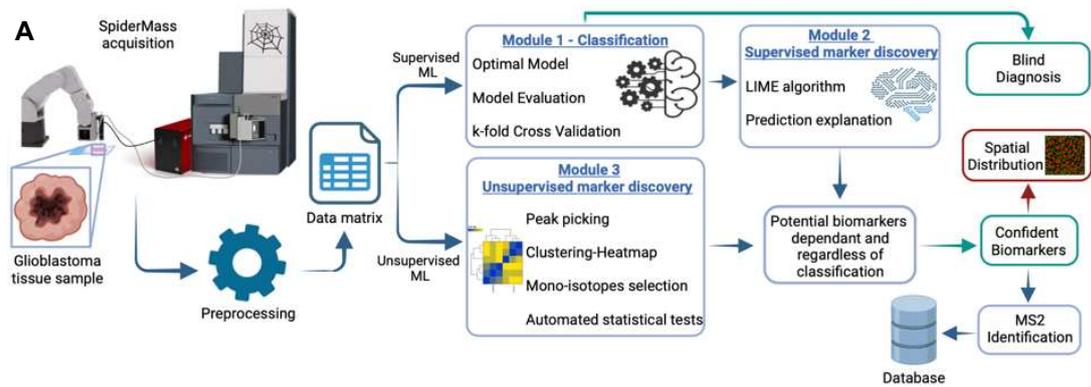
We have previously demonstrated that IDH-wild type glioblastoma could be stratified with three molecular groups using spatially resolved proteomics guided by MSI<sup>20</sup>. However, MALDI-MSI cannot be performed intraoperatively, and we aim at enabling similar molecular characterization in real-time to improve the EOR. We show that the same stratification is obtained by lipidomics in MALDI-MSI and that the lipid markers found are like those obtained with the SpiderMass. To operate SpiderMass intraoperatively, we must rely on a strong classification model. To address this, we have developed a new and more comprehensive artificial intelligence (AI) pipeline to provide accurate and reliable classification models and confident biomarker identification. The application of the pipeline to glioblastomas demonstrates an accurate discrimination between molecular classes as well as histological classes and applicability to prognosis with more than 90% accuracy to distinguish patients with longer survival (> 36 months) from those with shorter (< 10 months). 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 as well as NKT cells can be used to predict the evolution of patient. With this we demonstrate that SpiderMass could be used to get a sub-stratification of patients with glioblastoma as well as the prognostic in real-time to afford both for optimal EOR and further management of the patient therapy.

## Results

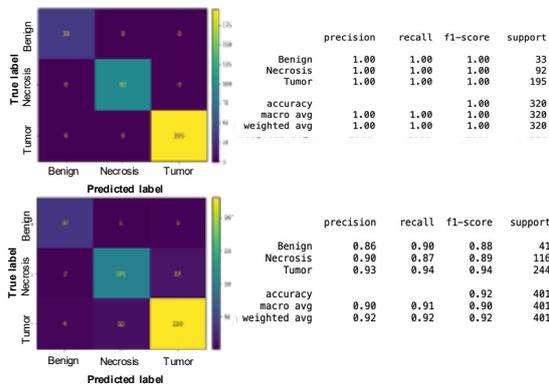
### Glioblastoma IDH-wild type lipidomic status assessment

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

To operate for 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 build 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 for feature variations. 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>31,32</sup>. To address these limitations, we have developed a new and more comprehensive AI pipeline (**Figure 1A**) to provide accurate and reliable classification models and confident biomarker identification not only to get a classification but also to validate the classifications through biological understanding. Thus, we compared the performances of 24 different classifiers. This was performed from the retrospective cohort of FFPE tissues (**Figure S1, Table S1**). All histologic regions (50 tumor, 42 necrosis and 18 benign tissue) were analyzed using SpiderMass in both negative (138, 102 and 42 MS spectra respectively per region) and positive ion mode (121, 88 and 39 MS spectra respectively per region) (**Table S2**). Classification models were constructed to discriminate between tumor, necrotic and benign tissue which class assignment is undoubtful as given by an expert pathologist. The optimal classifier was selected based on the highest accuracy on the 20% validation set. After obtaining the classification report and confusion matrix for the optimal classifier, its performance was evaluated by using both accuracy and F1 scores, which considers both precision and recall and provides a more balanced measure of performance across classes. In negative and positive ion mode, the optimal classification was obtained using the RidgeClassifier 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**). 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 in the negative ion mode was 88% with a standard deviation (std) of 0.03. For the positive ion mode, the model has an actual accuracy of around 87% with a std of 0.02. 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**). A possible explanation for the lower final accuracy and F1 scores previously obtained after 20-fold cross-validation could be 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 S3**) to the retrospective dataset (D1) to get an extended dataset (DE). The added dataset D1 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 S3**). Despite some spectral differences (**Figure S4**) between the analysis of the FFPE (D1) and fresh tissues (D2), the used of DE improved both the accuracy and F1 scores of the RidgeClassifier, after cross-validation. Very interestingly, MS spectra gathered from fresh frozen and FFPE tissues were sufficiently close to be used together to build and improve the classification. Indeed, as before the training set achieved 100% in both ion modes (**Figure 1B**). In the 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 the positive ion mode, there was a slight improvement in performance (**Table S4**). Importantly, the accuracy remained stable even after cross-validation, allowing an actual accuracy of 88% with a std of only 0.02.

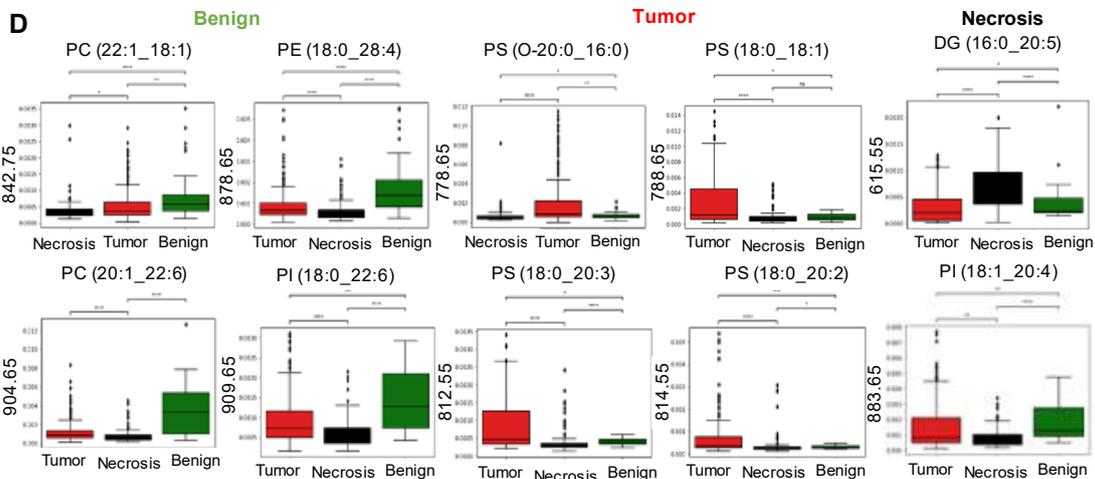


**B** Matrix confusion and classification report for train and 20-fold cross-validation sets



**C** Blind diagnosis of 9 unknown GBM tissue

Tissue	SpiderMass predicted labels				True label from the pathologist
	Negative ion mode		Positive ion mode		
FF 32	Necrosis	Necrosis	Necrosis	Necrosis	Necrosis
FF 4	Tumor	Tumor	Tumor		Tumor
FF 9	Tumor	Tumor	Tumor	Tumor	Tumor
FF 12-3	Tumor	Tumor	Tumor	Tumor	Tumor
FF 12-4	Tumor	Tumor	Tumor	Tumor	Tumor
FF 14	Tumor	Tumor	Necrosis		Tumor
FF 15	Tumor	Tumor	Tumor		Tumor
FF 17	Tumor	Tumor	Tumor	Tumor	Tumor
FFPE 8	Necrosis		Necrosis	Necrosis	Necrosis
FFPE 8	Tumor		Necrosis		Tumor
FFPE 4	Benign	Benign	Benign	Benign	Benign



**Figure 1. AI diagnostic pipeline for the optimal classification model obtainment and lipid markers associated to the different molecular class discovery.** (A) Overall pipeline for molecular diagnosis. The blue path is mandatory to obtain the final results (green arrows) while the brown path is optional. Indeed, facultative step (red 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 negative and positive ion mode with their tentative annotation. (See also Figures S1-S6 and Tables S1-S5)

### ***Automatic blind prediction of the histologic 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 in the construction and evaluation of the previous model with subgroups of benign, necrotic, and tumor regions. The regions were subjected to SpiderMass analysis in a blinded manner and evaluated against the classification model trained on DE. In negative mode, all the analyzed tissue were correctly attributed to its respective region after confronting the SpiderMass result to the pathologist exam. In positive mode, 2 out of 19 regions were misclassified in positive ion mode, 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.***

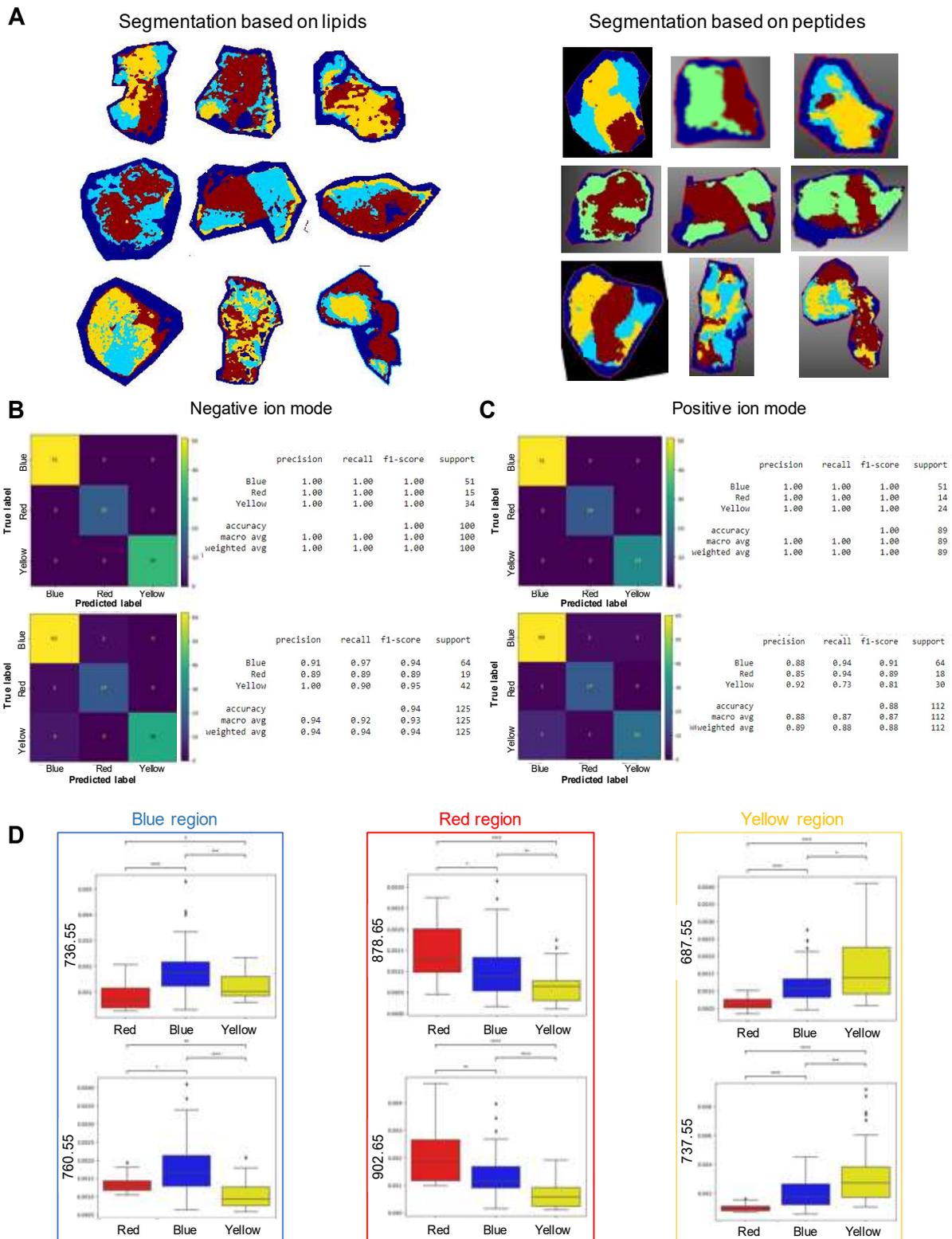
We were then interested to get the identification of the lipids associated to the different histological and stratification classes to better understand the biological pathways responsible for the discrimination and address the issue of most models which remain black box. The LIME<sup>33</sup> algorithm 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 S5A-C** for negative ion mode and for 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. In positive ion mode, the feature with the most negative contribution to the classification of tumor, benign and necrotic tissue are  $m/z$  603.55,  $m/z$  810.75 and  $m/z$  663.55 respectively. While the  $m/z$  that contribute most positively to their classification are  $m/z$  810.75,  $m/z$  605.65 and  $m/z$  603.55. 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 SNR > 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 S5D**). 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. In positive ion mode, 636 peaks were selected. The resulting heatmap is shown in **Figure S5B**. Of these peaks, 100 were significant and a total of 37  $m/z$  features were retained as potential biomarkers after filtering out isotopes and unresolved peaks.

To uncover lipid biomarkers with high confidence, we crossed the data from the supervised and unsupervised analyses. Thus, for an  $m/z$  feature to be considered a 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, for tumor, benign and necrotic tissue, which are listed in **Table S5** with their corresponding identification by MS/MS and their relative abundance variation presented in the box plots (**Figure 1D and S6**). 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 unsaturations. In glioma cells, the high concentration of polyunsaturated fatty acids appears to inhibit cell division and tumor formation<sup>34</sup>. In necrotic regions, three ions are over-expressed at low masses and eight ions are under-expressed at high masses. They are also characterized by the absence of molecules rather than their presence. More specifically, phosphatidylinositols (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 IDH-wild type based on lipidomics.***

In a previous study using spatially resolved shot gun proteomics guided by MALDI-MS Imaging, we have shown that despite not associated with peculiar mutation status, grade IV glioblastoma could be stratified in 3 specific molecular sub-groups. 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 as well as RNA processing (blue region)<sup>20</sup>. Additionally, a correlation between these groups and the OS of the patient was observed with patients for whom inflammation markers, macrophage infiltration and antibodies were identified were found with the shorter OS. We thus firstly performed the 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 considering that the lipidomic and proteomic data allowed to differentiate the same area in multiple tissue. 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 (unpublished data). Thanks to the AI processing pipeline newly 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 for which three are specific to the blue regions, ten to the red regions and six to the yellow regions (**Figure 2D**). Interestingly, some ions were in common with the lipid markers of the different histological zones found above. For example, the PS 34:1 ( $m/z$  760.55) specific of 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 of 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.



**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 distinguishes the 3 segmentation clusters. **(C)** Boxplots of two lipid biomarkers obtained for each segmentation clusters.

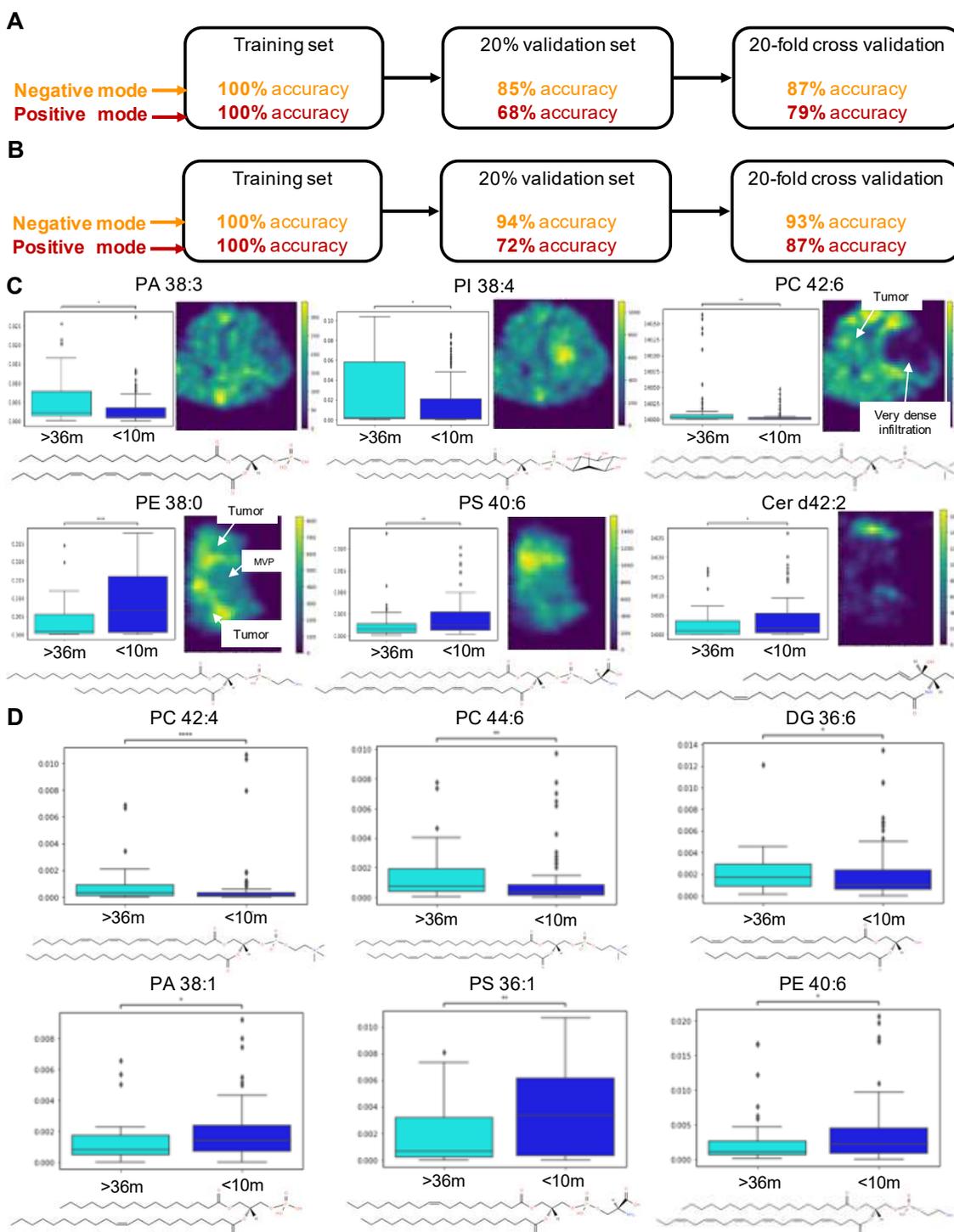
## Prediction of patient outcome from SpiderMass lipidomic data.

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

On top of the validated diagnostic and stratification based on SpiderMass, we were also interested in investigating the potential to predict the patient outcome. According a 15-month median of OS of glioblastoma patients, we build a classification using the previously developed processing pipeline, separating our cohort in <15-month and >15-month OS. We used the MS spectra recorded from the tissues independently from their localization in a histologic region 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 extreme survivals i.e., <10 and >36 months. By narrowing our model to most extreme cases, we obtained, in the 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 the 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 prognostic biomarkers were found (**Table S6**). 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,  $m/z$  850.75 identified as PE (22:1\_22:4) and  $m/z$  920.75 as PS 46:5 which are more abundant in the benign tissues 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, PE 38:0, PE 40:6 and PS 36:1, 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 high significance and a high contribution (LIME algorithm). 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 than in those with longer ones, regardless of acyl chain length or degree of unsaturation. In addition,  $m/z$  844.65 identified as PS 40:1 and which is specific to the tumor is associated with a shorter survival. Interestingly, our findings at the lipid level are closely related to those at the proteomic level. Indeed, previous identified three prognostic markers<sup>20</sup> are closely linked to the prognosis lipid markers previously found. 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 levels of ANXA11. ANXA11 is a calcium-dependent phospholipid-binding protein<sup>34</sup>, and annexins are soluble proteins that bind to negatively charged phospholipids in the presence of calcium ions<sup>35</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 (also known as MYPT1) is a subunit of myosin phosphatase (MP), which regulates anti-tumor signaling pathways such as activating gene expression of tumor suppressors such as Rb, Rap, and c-Myc<sup>36</sup>. On the other hand, there is a known ceramide-protein interaction between ceramides and I2PP2A, leading to the degradation of c-Myc<sup>37</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.

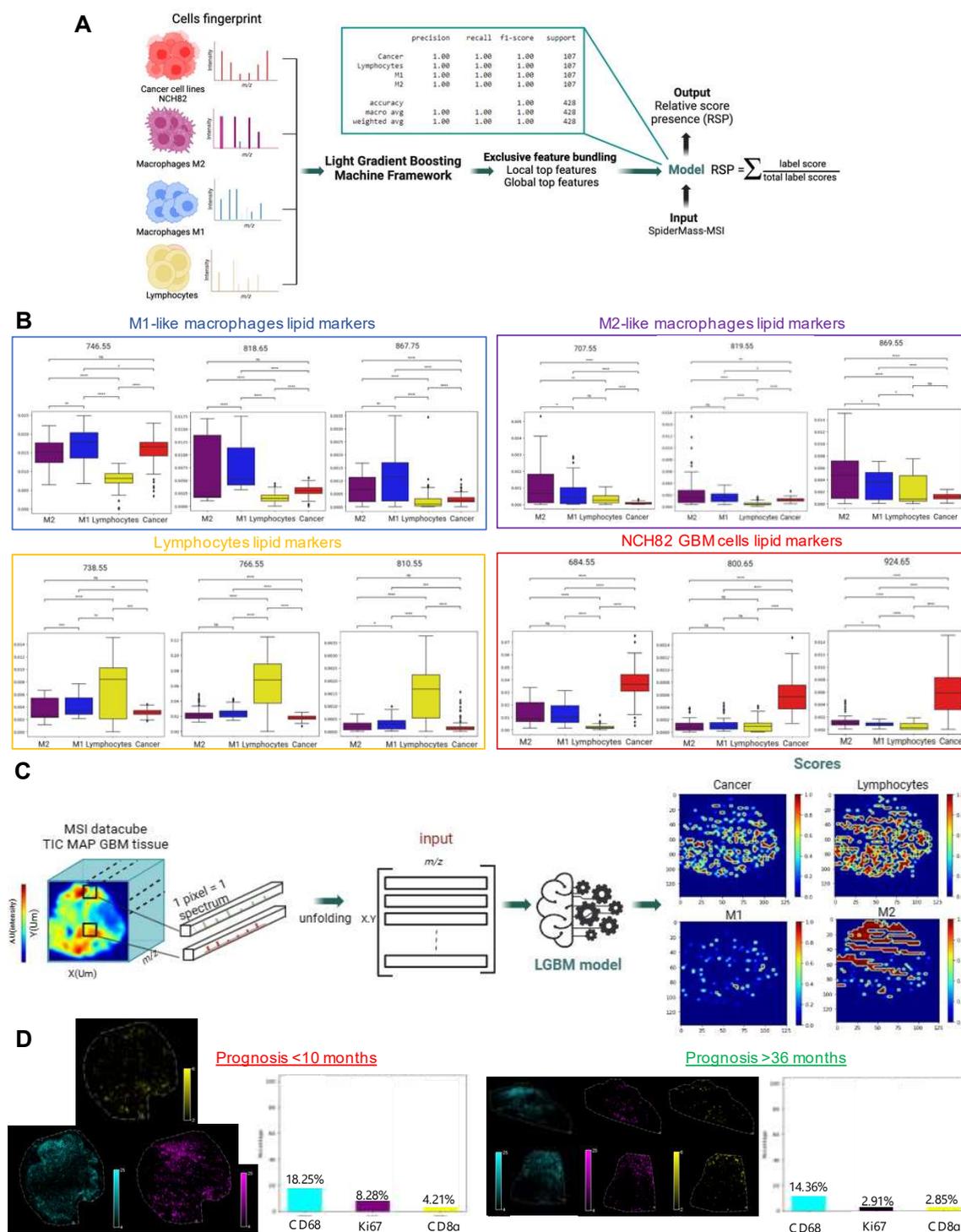


**Figure 3. Obtainment of a classification model to distinguish patient with a low or a long survival time and the discovery of biomarkers for each survival.** (A) and (B) Accuracies obtained for the classification model made with RidgeClassifier, with the training set, after 20% validation and after 20-fold cross-validation in negative (orange) and positive ion mode (red) to distinguish patient with a survival time  $<$  or  $>$  15 months and after narrowing our model to most extreme cases ( $<$  10 months and  $>$  36 months). (C) and (D) Six potential biomarkers, their corresponding boxplot and their chemical structure, that were found specific of prognosis  $<$  10 months and  $>$  36 months in negative and positive ion mode, respectively. (See also Table S6)

## 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<sup>38-40</sup> in colon cancer and is now recognized in more and more cancers<sup>41</sup>. Nowadays, the immunoscore is obtained from excised tissues or biopsies by immunohistochemistry (IHC) using antibodies specific of the different immune cell populations. However, it is obtained post-surgery. Here, we explored the possibility to create an immunoscore based on SpiderMass data that could in the future be realized *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) to find lipids which are specific to the immune 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 the SpiderMass is not designed to offer single cell analysis, 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 develop 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>42</sup> was used to train an immune scoring 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 the newly developed AI pipeline, 10 lipids 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 S7**). For example, the ions  $m/z$  818.65 (GlcCer d18:1\_22:0) and 819.55 (PG 18:1\_22:6) were found specific of 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 and  $m/z$  848.65, specific of M1-like and M2-like macrophages respectively were also found specific of tumor tissue from patient with an OS below 10 months.



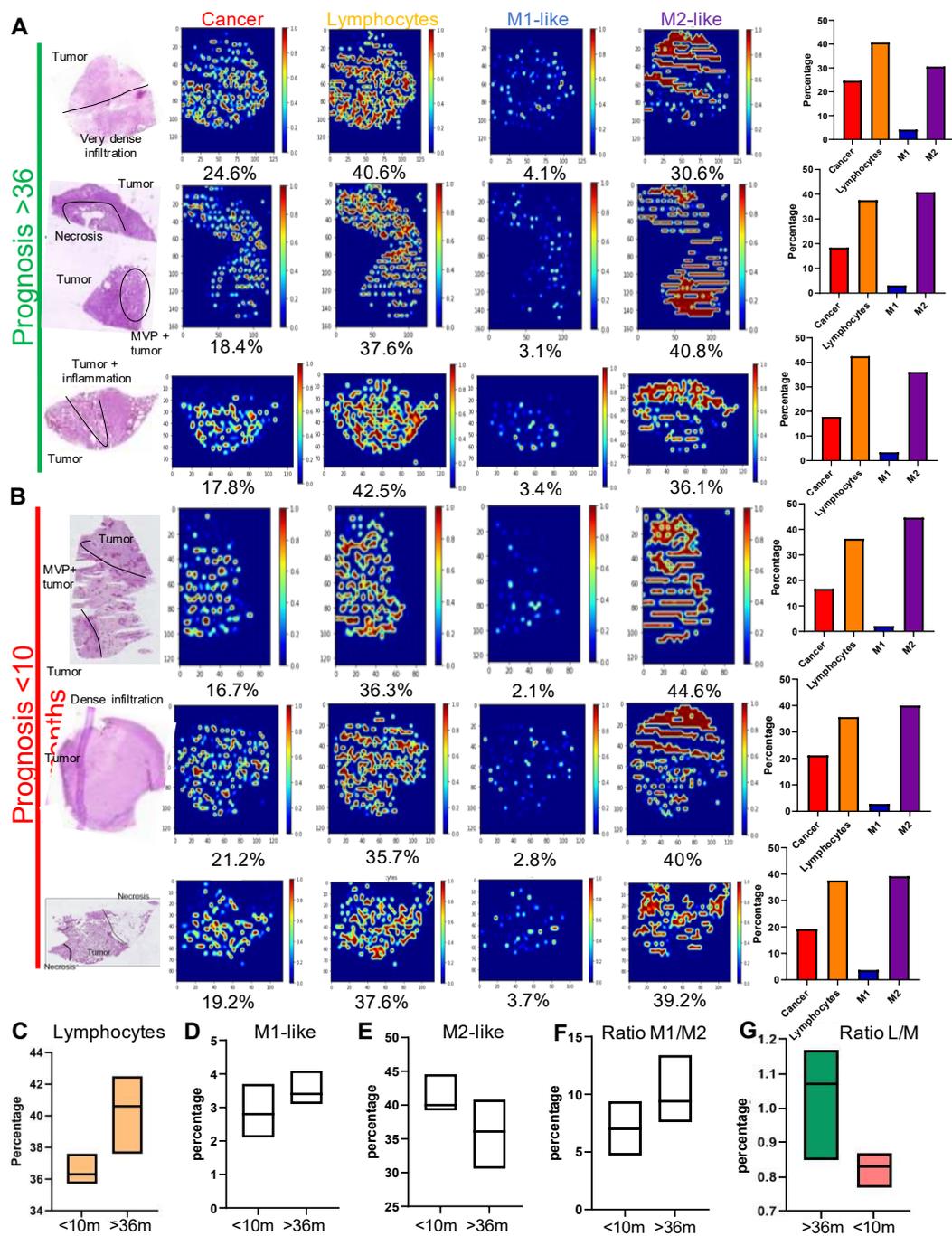
**Figure 4. SpiderMass classification based on immune cells lipid profiles. (A)** Trained LGBM model for cancer cells, lymphocytes, M1-like and M2-like macrophages that were analyzed previously by SpiderMass technology after isolation. **(B)** Boxplot of 3 biomarkers for GBM cancer cells, lymphocytes, 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). (See also Table S7)

### *Correlation of immunoscores based on SpiderMass images with patient prognosis*

From a SpiderMass image, we can 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 respectively. 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 respectively (**Figure 5E**). The ratio of M1-like to M2-like macrophages could serve as a prognosis marker. In fact, when the proportion of M1-like macrophages is below 7% in comparison to M2-like macrophages, it signifies a shorter survival time. Conversely, a composition exceeding 10% of M1-like macrophages relative to M2-like macrophages is indicative of an extended survival duration (**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 patient prognosis, macrophages tend to outnumber lymphocytes (+ 22%), while the converse holds true for patients with a more favorable prognosis (**Figure 5G**).

A 5-plex MALDI-IHC panel was constructed to corroborate previous results from the newly 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>43-45</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 subsequently 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 use a probabilistic approach where each pixel signifies 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, indicating full presence.

Of particular interest, the SpiderMass-MSI approach, based on immunoscore, allows the differentiation of various subpopulations within the tumor microenvironment without necessitating techniques reliant on probe utilization, such as IHC. In culmination, 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.



**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, respectively. For each cell, the ratio in percentage is indicated. In addition, a histogram was built to compare the percentage of each type of cells in each tissue. (C) The mean percentage of lymphocytes in tissue with a prognosis <10 months and >36months. (D) The mean percentage of M1-like macrophages in tissue with a prognosis <10 months and >36months. (E) The mean percentage of M2-like macrophages in tissue with a prognosis <10 months and >36months. (F) The mean ratio M1/M2 macrophages in tissue with a prognosis <10 months and >36months. (G) The mean ratio lymphocytes (L)/ macrophages (M) in tissue with a prognosis <10 months and >36months.

## Discussion

This study represents a comprehensive investigation of the possibility to get an accurate histological classification (diagnosis), sub-stratification and prognosis of glioblastoma using ambient ionization mass spectrometry by SpiderMass combined with an artificial intelligence prediction pipeline. In this work, 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 than those obtained by spatially resolved proteomics. Besides, we investigated different classification methods to find which is best suited to our MS-based molecular dataset since there is no consensus in the community and various algorithms (e.g., LDA<sup>46</sup>, Lasso<sup>47</sup>) have been used for similar purposes from very close data.

A total of 24 classifiers were evaluated and optimal model was obtained using the ridge classifier. The RidgeClassifier<sup>48</sup>, which is suitable for high-dimensional datasets such as ours with 5000  $m/z$  dimensions, is a linear classifier that uses L2 regularization to avoid overfitting and demonstrate strong performances for MS data from tissues<sup>49</sup>. To address the issue of small sample size in the accessible cohort which is responsible for the more limited performances of the classification model, we investigated data extension by aggregating the MS data collected with the SpiderMass both from the fresh frozen tissues of the prospective cohort (31 patients) and the FFPE tissue samples from the retrospective cohort (50 patients) 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 of model accuracy with 92% and 88% after 20-fold cross-validation for tissue histological type (tumor, necrosis and benign tissue) classification in negative and positive ion modes, respectively. Blind predictions using an independent validation set of ten tissues resulted in no misclassified spectra in negative ion mode and only two misclassified spectra in positive ion mode.

Additionally, to validate further our results and provide biological explainability to the classification by histological features and molecular sub-types, we included a secondary data processing pipeline based on LIME algorithm. LIME algorithm combined with unsupervised machine learning techniques and statistical tests to identify potential biomarkers independent of classification 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 this machine learning pipeline represents a promising approach for advancing precision medicine for glioblastoma and other heterogeneous cancers. 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>50,51</sup>. These robust biomarkers warrant further investigation for their therapeutic potential in GBM.

Furthermore, this innovative AI processing pipeline has enabled the clear differentiation of the three molecular subgroups resulting from both the segmentation of lipids and peptides through MALDI-MSI. The accuracy of classification for these subgroups was standing at 94% and 88% for both ion modes. Notably, within the set of 19 robust biomarkers uncovered through a combination of supervised and unsupervised marker discovery approaches, 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, based on prior proteomic analyses, 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, the newly developed machine learning pipeline not only serves as a diagnostic tool but was also tested for prognosis when using SpiderMass. 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-month versus >36-month with achieving accuracies of 93% and 87% in negative and positive ion modes 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>34</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, survival duration exceeding 36 months, and the segmentation clusters denoted in red and yellow. This correlation is readily discernible regardless of the employed ionization mode. 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). The link between methylation status and survival time is also corroborated by the fact that the median survival time for methylated tissue is 31 months, while for unmethylated tissue it's 14 months.

The third innovative aspect of the new strategy 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>52</sup>. Immunoscoring was introduced already long ago for colon cancer and since then has been demonstrated for other cancers<sup>40,41</sup>. Immunoscoring is based on IHC and can be performed during the post-surgery pathology exam. Creating an immunoscore for SpiderMass would offer a way to translate the immunoscore *in vivo* and already predict by the time of surgery the patient outcome. This is offering a way for surgeons to tailor the EOR according to the predicted evolution and adapt the therapy if a local intraoperative therapy (e.g, photodynamic therapy) is planned. Considering that an analysis at the single cell level, would it be possible with the SpiderMass, would be far too long to achieve during the surgery, we cannot simply use a classification to find to which type of cell corresponds to the MS spectrum. 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>53-55</sup>. Besides, the predictions were validated using mass spectrometry-based IHC (MALDI-IHC) using specific antibodies directed against immune cells. These results indicate that immunoscore has the potential to become a prognostic prediction tool for glioblastoma patients without the need of probes, allowing for personalized treatment based on their immune score.

In conclusion, this study demonstrates the potential of SpiderMass combined with artificial intelligence predictions for accurate classification of tumors and tissue subtypes. The machine learning 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 prognostic 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 machine learning pipeline holds great promise for possible application in the future for accurate diagnostic and prognostic to be achieved during the surgery directly from the patients and is even compatible with awake surgery considering the analysis is painless.

## **Acknowledgments**

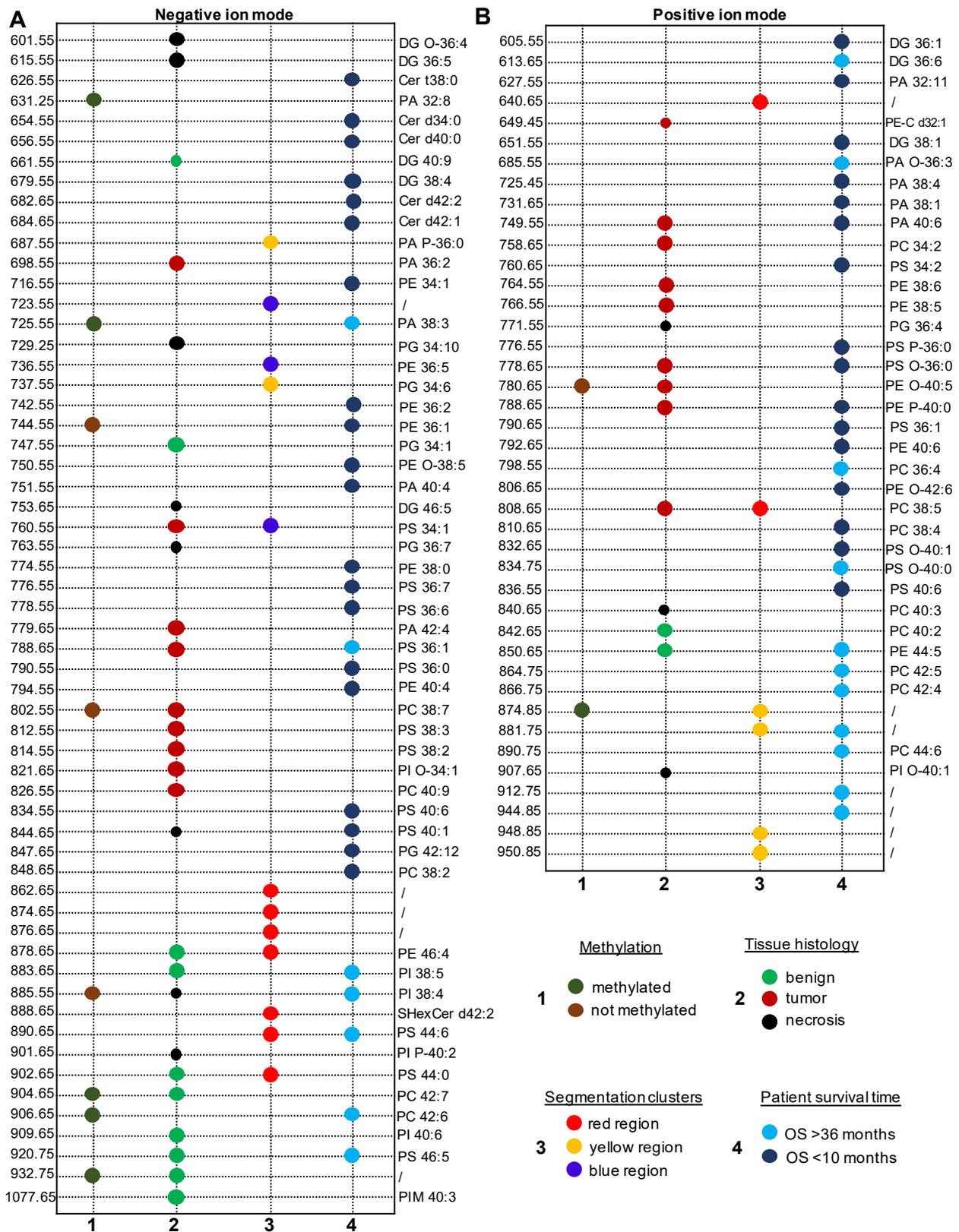
This research was supported by grants from Ministère de l'Enseignement Supérieur et de la Recherche (MESR), Inserm specific funding for SpiderMass project (I.F.), Inserm and Institut Universitaire de France (I.F.). L.L. PhD was cofounded by University of Lille Excellence Initiative, Région Haut de France (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 cell line and immune cells SpiderMass analysis. E.L. is promotor 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 the 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. and M.D. 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., E.L., B.M. and P.T. declare they have no competing interests. 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.



**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.**

## Supplemental figures legends

**Figure S1. H&E scans of the fifty FFPE glioblastoma tissue from the retrospective cohort and their histological annotations, related to Figure 1.**

**Table S1. Demographic and clinical data of the retrospective cohort, related to Figure 1.**

**Table S2. Number of MS spectra for each histological region (tumor, necrosis and normal brain tissue) for the 50 FFPE tissues of glioblastoma from the retrospective 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. H&E scans of the 31 fresh frozen glioblastoma tissue from the prospective cohort and their histological annotations, related to Figure 1.**

**Table S3. Number of MS spectra for each histological region (tumor and necrosis) for the 31 fresh frozen glioblastoma tissues from the prospective cohort, related to Figure 1.**

**Figure S4. 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 S4. Accuracy, standard deviation and F1 scores obtained with Ridge classifier before and after data augmentation in positive ion mode, related to Figure 1.**

**Figure S5. 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 S5. 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 S6. 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 S6. Total of 48 biomarkers for poor and good prognosis in both negative and positive ion modes, related to Figure 3.** Green indicates for which prognosis the ion is specific.

**Table S7. Total of 27 biomarkers specific of cancer and immune (lymphocytes, M1-like and M2-like macrophages) cells in negative ion modes, related to Figure 4.** Green indicates for which cell types the ion is specific. Signs (++ and +) corresponding to the contribution weight of the  $m/z$  in each type of tissue found with LIME algorithm.

## 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 (processing pipeline, immunoscores and IHC quantification) has been deposited at GitHub ([https://github.com/yanisZirem/SpiderMass\\_AI\\_pipelines.git](https://github.com/yanisZirem/SpiderMass_AI_pipelines.git)) and is available from the lead contact upon request. 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 SUBJECT 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 cohort are reported in Table S1.

### **Cell line**

Human NCH82 stage IV glioma 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>56</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 has been done thanks to two passages of 5 min 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. 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, 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 lock-mass. 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 MS Imaging

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>57</sup>. The spatial step size was set to 250  $\mu\text{m}$  for fresh frozen tissue to achieve oversampling. For each one, the final spatial resolution was divided by two thanks to interpolation algorithm. 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>57</sup>. The data was acquired in negative and sensitivity ion mode.

### **Cell line and immune cell analysis**

Human NCH82 stage IV 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 then 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<sup>TM</sup> centrifuge (Thermo Shandon Cytospin) and conserved at -80°C. The SpiderMass analysis were made directly onto the glass slide.

### **MALDI-MS Imaging**

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 PRO, 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-Immunohistochemistry (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<sup>TM</sup> 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>58</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>33</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>42</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**

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>59</sup> and literature.

## 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 Mellinghoff, 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. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
  18. 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.
  19. 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.
  20. 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.
  21. 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.
  22. 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.
  23. Tzafetas, M., Mitra, A., Paraskevas, 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.
  24. 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.
  25. 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.
  26. 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.
  27. 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.
  28. 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.
  29. 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.
  30. 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.

31. Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy (Basel)* 23, 18. 10.3390/e23010018.
32. Gunning, D., and Aha, D.W. DARPA's Explainable Artificial Intelligence Program.
33. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier.
34. 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.
35. Lizarbe, M., Barrasa, J., Olmo, N., Gavilanes, F., and Turnay, J. (2013). Annexin-Phospholipid Interactions. Functional Implications. *IJMS* 14, 2652–2683. 10.3390/ijms14022652.
36. Kiss, A., Erdődi, 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.
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.
48. Dijkstra, T.K. (2014). Ridge regression and its degrees of freedom. *Qual Quant* *48*, 3185–3193. 10.1007/s11135-013-9949-7.
  49. Cortes, C., Mohri, M., and Rostamizadeh, A. (2009). L2 Regularization for Learning Kernels.
  50. 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.
  51. 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.
  52. 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.
  53. 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.
  54. 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.
  55. 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.
  56. 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.
  57. 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.
  58. 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.
  59. 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.