Métabolomique du cancer du sein localisé à haut risque de récidive

Caroline Bailleux

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ

THÈSE DE DOCTORAT

Métabolomique du cancer du sein localisé à haut risque de récidive

Caroline BAILLEUX

Laboratoire TIRO-MATOs (UMR E4320)

Présentée en vue de l'obtention du grade de docteur en Sciences de la Vie et de la Santé
Mention : Interactions moléculaires et cellulaires
d'Université Côte d'Azur
Dirigée par : Thierry POURCHER
Soutenue le : 7 Juillet 2023

Devant le jury, composé de :
Jean-Marc FERRERO, PU-PH, Président du Jury
Audrey LEGOUELLEC, MCU-PH, HDR, Rapportrice
Olivier TREDAN, Praticien spécialiste de CLCC, HDR, Rapporteur
Pierre-Etienne HEUDEL, Praticien spécialiste de CLCC, Examinateur
Sonia DAGNINO, Professeure, UCA, Examinatrice
Thierry POURCHER, DR, Directeur de Thèse

cea

Antoine Lacassagne
CENTRE DE LUTTE CONTRE LE CANCER
unicancer NICE

Métabolomique du cancer du sein localisé à haut risque de récidive

Devant le jury composé de :

Président du jury	Pr Jean-Marc FERRERO, Professeur Universitaire-Praticien Hospitalier, Centre Antoine Lacassagne, Nice
Rapporteurs	Dr Audrey LE GOUELLEC, Maître de conférences des Universités-Praticien Hospitalier, HDR, Université Grenoble Alpes
	Dr Olivier TREDAN, Praticien spécialiste de CLCC, HDR, Centre Léon Bérard, Lyon
Examinatrice	Dr Pierre-Etienne HEUDEL, Praticien spécialiste de CLCC, Centre Léon Bérard, Lyon
	Pr Sonia DAGNINO, Professeure, UCA, TIRO, UMR E4320, Nice
Directeur de thèse	Dr Thierry Pourcher, Directeur de recherche, CEA, TIRO, UMR E4320

I. Résumé & Abstract

Métabolomique du cancer du sein localisé à haut risque de récidive.

Le cancer du sein est une maladie hétérogène avec de multiples sous-types histologiques, biologiques et moléculaires. Plusieurs études fondamentales ont mis en évidence l'activation de voies métaboliques spécifiques dans les cancers du sein agressifs. L'objectif de cette thèse était d'identifier une signature ou des marqueurs du métabolome dans le cancer du sein localisé à haut risque de récidive.

Nos premières études se sont basées sur l'inclusion rétrospective de 52 patientes atteintes d'un cancer du sein localisé et traitées au Centre Antoine Lacassagne de Nice. Nous avons également analysé les biopsies diagnostiques issues d'une cohorte de 49 patientes traitées par chimiothérapie néo-adjuvante au Centre Georges-François Leclerc de Dijon pour un cancer du sein localement avancé. Après extraction, séparation et concentration des métabolites, nous avons réalisé un profilage métabolomique par LC-MS/MS pour identifier et quantifier de manière relative les métabolites, suivi d'analyses biologiques et statistiques.

Tout d'abord, nous avons comparé les performances de 5 méthodes de machine learning non supervisées (PCA k-means, sparse k-means, spectral clustering, SIMLR et k-sparse) pour identifier des groupes de patients atteints de cancer du sein. Cette analyse n'a été réalisée que sur la cohorte de Nice.

Les clusters obtenus en utilisant les 5 méthodes de machine learning non supervisées ont été comparés dans l'Article 1. Les cinq méthodes ont identifié trois groupes de patients, distincts par leur pronostic supposé (groupe 1 favorable, groupe 2 intermédiaire, groupe 3 défavorable), avec des profils cliniques et biologiques différents. Les méthodes SIMLR et K-sparse étaient les plus efficaces en termes de clustering. Les voies métaboliques les plus discriminantes étaient la glycolyse, la glutaminolyse et le métabolisme des acides aminés. L'analyse de survie simulée « in-silico » (outil PREDICT) a révélé une différence significative entre les 3 groupes pour la survie spécifique à 5 ans et à 10 ans.

Dans l'Article 2, les analyses de survie ont été réalisées à partir des données de survie réelle des patients. Chaque patient était rattaché à son groupe pronostic comme établi précédemment par les 5 méthodes d'apprentissage automatique non supervisées. Les groupes 1 et 2 ont été regroupés et comparés au groupe 3. Le suivi médian a été prolongé à 85,8 mois. Une optimisation Bootstrap a été appliquée. Les méthodes PCA k-means, K-sparse et Spectral clustering ont obtenu les meilleurs résultats pour prédire la survie sans

progression à 2 ans. La méthode PCA k-means avait les meilleures performances. Les analyses CSS et OS ont révélé cependant des discordances entre les 5 méthodes de machine learning non supervisées.

Parallèlement, une analyse supervisée comparant les tumeurs de haut grade et celles de grade faible/intermédiaire a été réalisée pour déterminer les métabolites entrant en jeu dans l'agressivité tumorale (Article 3). La cohorte niçoise a été utilisée comme cohorte d'entrainement. La cohorte dijonnaise a permis une validation externe en tant que cohorte de validation. La signature métabolomique était composée de 12 métabolites. Les AUC pour la cohorte d'entraînement et la cohorte de validation étaient supérieures à 0,88. Le modèle pouvait donc distinguer les tumeurs de grade élevé et de grade faible/intermédiaire avec une probabilité de près de 90 %. Nous avons identifié plusieurs biomarqueurs de l'agressivité tumorale, tels que la N1, N12 diacétylspermine et les catabolites du tryptophane (la kynurénine et la sérotonine), impliqués dans l'inhibition de la réponse immunitaire.

Ces études ouvrent de nouvelles perspectives sur les mécanismes biologiques sous-jacents à l'agressivité tumorale. De plus, les biomarqueurs identifiés permettront le développement de nouvelles stratégies. Cependant, des analyses sur des populations avec de plus grands effectifs sont nécessaires.

Mots clés : signature métabolomique; cancer du sein; analyse non ciblée; apprentissage automatique non supervisé; agressivité.

Metabolomics of localized breast cancer at high risk of recurrence.

Breast cancer is a heterogeneous disease with multiple histological, biological, and molecular subtypes. Several fundamental studies have highlighted the activation of specific metabolic pathways in aggressive breast cancers. The aim of this thesis was to identify a signature or markers of the metabolome in localized breast cancer at high risk of recurrence.

Our initial studies were based on the retrospective inclusion of 52 patients with localized breast cancer treated at the Antoine Lacassagne Center in Nice. We also analyzed diagnostic biopsies from a cohort of 49 patients treated with neo-adjuvant chemotherapy at the Georges-François Leclerc Center in Dijon for locally advanced breast cancer. After extraction, separation, and concentration of metabolites from diagnostic biopsies and resected tumors, we performed metabolomic profiling using LC-MS/MS to identify and quantify metabolites relatively, followed by biological and statistical analysis.

First, we compared the performance of 5 unsupervised machine learning methods (PCA k-means, sparse k-means, spectral clustering, SIMLR, and k-sparse) to identify groups of breast cancer patients. This analysis was only performed on the cohort from Nice.

In Article 1, the clusters obtained using the 5 unsupervised machine learning methods were compared. The five methods identified three groups of patients, distinguished by their supposed prognosis (favorable group 1, intermediate group 2, unfavorable group 3), with different clinical and biological profiles. The SIMLR and K-sparse methods were the most effective in terms of clustering. The most discriminating metabolic pathways were glycolysis, glutaminolysis, and amino acid metabolism. The simulated "in-silico" survival analysis (PREDICT tool) revealed a significant difference between the 3 groups for 5-year and 10-year specific survival.

In Article 2, survival analyses were performed based on actual patient survival data. Each patient was assigned to his prognostic group established by the 5 unsupervised machine learning methods. Groups 1 and 2 were combined and compared to group 3. The median follow-up was extended to 85.8 months. Bootstrap optimization was applied. The PCA k-means, K-sparse, and Spectral clustering methods achieved the best results for predicting 2-year progression-free survival. The PCA k-means method had the best performance. However, CSS and OS analyses revealed discrepancies between the 5 unsupervised machine learning methods.

Simultaneously, a supervised analysis comparing high-grade tumors to low/intermediate grade tumors was conducted to determine the metabolites involved in tumor aggressiveness (Article 3). The Nice cohort was used as a training cohort, while the Dijon cohort was used for external validation. The metabolomic signature was composed of 12 metabolites. The AUCs for the training and validation cohorts were greater than 0.88. Thus, the model could distinguish high-grade tumors from low/intermediate grade tumors with a probability of nearly 90%. We identified several biomarkers of tumor aggressiveness, such as N1, N12 diacetylspermine and tryptophan catabolites (kynurenine and serotonin), which are involved in inhibiting the immune response.

These studies open up new perspectives on the underlying biological mechanisms of tumor aggressiveness. Furthermore, the identified biomarkers will allow the development of new strategies. However, analyses on larger populations are necessary.

Keywords : metabolomic signature; breast cancer; untargeted analysis; unsupervised machine learning; aggressiveness.

Table des matières

II. Liste des figures

III. Liste des tableaux

IV. Abréviations

ADN [DNA] : acide desoxyribonucléique [DeoxyriboNucleic Acid]

AKT(1) : serine/threonine kinase (1)

AMPK : AMP-activated protein kinase

Anti-LHRH : agoniste de la luteinizing hormone-releasing hormone

ARNm [mRNA] : acide ribonucléique messager [messenger RiboNucleic Acid]

ASCT2 : Alanine, Serine, Cysteine Transporter 2

ATP : adénosine triphosphate

AUC [area under curve]: aire sous la courbe

BCSM [breast cancer-specific mortality]: mortalité spécifique par cancer du sein

BRCA 1/2 : breast cancer 1/2

CE [capillary electrophoresis] : electrophorèse capillaire

CO_2 : dioxyde de carbone

CTNA : chimiothérapie néo-adjuvante

CTS5 : clinical treatment score 5

Cycle TCA : tricarboxylic acid cycle

DiAcSpm : diacétylspermine

DMFS : distant metastasis-free survival

FFPE [formalin-fixed paraffin embedded] : tissus fixés au formol et inclus en paraffine

FISH [fluorescence in situ hybridization] : hybridation in situ en fluorescence

GC [gas chromatography] : chromatographie en phase gazeuse

GLUT 1/3/4 : transporteur GLUT 1/3/4

HBOC syndrome : hereditary breast and ovarian syndrome

HCA : hierarchical cluster analysis

HDAC : histone désacétylase

HER2 [Human Epidermal Growth Factor Receptor-2] : récepteur au facteur de croissance épidermique humain 2

HIF α [Hypoxia-Inducible Factor-1α] : facteur induit par l'hypoxie 1

HK2 : hexokinase 2

HMDB : human metabolome database

HPLC : high performance liquid chromatography

HT : hormonothérapie

IC95% : intervalle de confiance à 95%

IDFS : Invasive disease-free survival

IHC : immunohistochimie

Lasso : least absolute shrinkage and selection operator

LC [liquid chromatography] : chromatographie en phase liquide

LKB1 [Liver kinase B1] : kinase hépatique B1

metPA : metabolomics pathway analysis

MS [mass spectrometry] : spectrométrie de masse

mTORC1 [mTOR1 complex] : complexe mTOR 1

N : node

NADP : nicotinamide adénine dinucléotide phosphate

NADPH : nicotinamide adénine dinucléotide phosphate hydrogène

NCDB : national cancer database

NST : non-special type

O_2 : dioxygène

ODC : ornithine décarboxylase

PCA : principal component analysis

PCR [polymerase chain reaction] : réaction de polymérisation en chaîne

PD1 [programmed cell Death protein 1] : protéine 1 de la mort cellulaire programmée

PDL1 [programmed death-ligand 1] : ligand de PD1

PI3K : phosphatidylinositol 3-kinase

PLS-DA [partial least-squares discriminant analysis] : moindres carrés partiels

qRT-PCR : quantitative reverse transcription PCR

Rapport m/z : rapport masse/charge

Rb : retinoblastome

RE : récepteur aux oestrogènes

RMN [NMR] : résonance magnétique nucléaire [nuclear magnetic resonance]

RNA-seq [RNA sequencing] : séquençage du transcriptome

ROC : receiver operating characteristic

ROS [reactive oxygen species] : espèces réactives de l'oxygène

RP : récepteur aux progestérones

RR : risque relatif

RS : recurrence score

RT [retention time] : temps de rétention

SAT1 : spermine N1 acetyltransférase

SBR : Scarff-Bloom-Richardson

SG : survie globale

SSP : survie sans progression

SIMLR : single-cell interpretation via multikernel learning

SMP : adénylosuccinate

SMPDB : small molecule pathway database

TEM : transition épithélio-mésenchymateuse

TEP [PET] : tomographie par émission de positons [Positron Emission Tomography]

TIL [tumor infiltrating leukocytes]: lymphocytes infiltrant la tumeur

TNBC [triple negative breast cancer] : cancer du sein triple négatif

TP53 : tumor protein 53

V. Introduction générale

1. Introduction à la métabolomique

a) ARRIVEE D'UN NOUVEL OMIQUE

Le concept de métabolomique est apparu à la fin des années 1990, en analogie aux notions de génomique, de transcriptomique et de protéomique. Il fait référence à l'analyse des métabolites contenus dans un système biologique donné : cellules ou fluides biologiques, tels que les urines ou le plasma (1).

Les métabolites sont des composés impliqués dans les processus métaboliques, qu'ils interviennent au niveau du catabolisme ou de l'anabolisme. Le terme métabolite inclut par conséquent toutes les molécules de faible masse moléculaire (taille inférieure à 1500 Dalton) telles que les acides organiques, les nucléotides, les sucres, les acides gras, les acides aminés, certains peptides ou encore les vitamines. Ils peuvent être générés de façon endogène par le système biologique ou présents de façon exogène (i.e. polluants environnementaux, médicaments).

Comme le protéome, le métabolome est dépendant du contexte, c'est-à-dire que les taux de protéines ou de métabolites sont modifiés en fonction de l'état physiologique, développemental, ou pathologique d'une cellule, d'un tissu, d'un organe ou d'un organisme. Les métabolites sont les produits finaux des processus cellulaires et représentent l'ultime réponse d'un organisme à une altération génétique, une pathologie, une exposition environnementale, une exposition toxicologique ou à tout autre facteur susceptible de perturber son fonctionnement(2). Il est par définition caractéristique d'un état physiologique donné et fournit une vue globale sur des évènements biochimiques produits à un instant t. Il permet donc de suivre l'évolution biologique de systèmes complexes en fonction d'interférences extérieures mais aussi dans des contextes pathologiques.

De plus, par rapport à des protéines ou des acides ribonucléiques (ARN), les métabolites possèdent une stabilité relativement élevée, ce qui est un des avantages des analyses en métabolomique. En effet, plusieurs études ont évalué la stabilité des métabolites dans les échantillons biologiques. La majorité des métabolites sont stables à -20°C ou -80°C entre 6 mois et 2 ans(3,4) et restent également stables pendant les cycles de congélation/décongélation (4 à 9 selon les études)(3,5). Les analyses n'ont pas été poursuivies au-delà de 2 ans mais la stabilité observée laisse envisager une stabilité à plus long terme. Ceci est un avantage certain pour la faisabilité des études de métabolomique et permet d'envisager une applicabilité en recherche clinique et en pratique courante en cancérologie.

b) SYSTEME COMPLEXE AU SEIN DES AUTRES OMIQUES

Comme la transcriptomique et la protéomique, la métabolomique s'inscrit dans un contexte post-génomique avec un développement lié à celui des nouvelles technologies. Il a également la particularité d'être fortement interconnectée aux autres OMIQUEs en tant que dernière étape de la chaîne « OMIQUE ». Il est indispensable de comprendre la métabolomique comme un sous-ensemble des «OMIQUES » car elle constitue le résultat d'une cascade d'événements constituant l'ensemble de la post-génomique. En effet, les métabolites explorés par la métabolomique sont l'étape finale, la résultante de l'effet de variations d'expression des gènes et de celui de l'environnement(6).

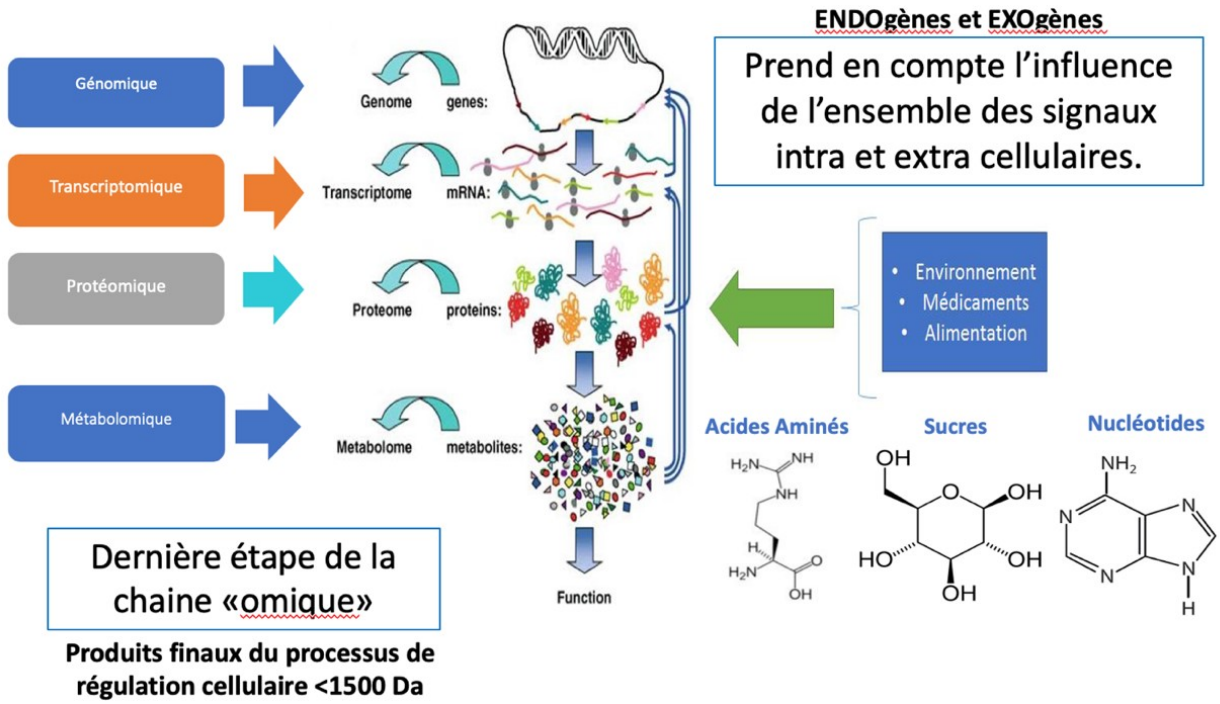


Figure 1. Vision intégrative de la métabolomique *(d'après Courant et al. (6)). Da : dalton ; mRNA : ARN messager.*

En plus d'être le dernier maillon de la cascade -OMIQUE, les métabolites peuvent exercer une action sur les étapes -OMIQUES en amont. Le cancer est causé par des changements au niveau génomique qui entraînent une altération de la transcription des ARNs, ainsi que de l'expression et de la fonction des protéines. Le métabolome est un reflet de ces changements en amont. À leur tour, les métabolites peuvent individuellement affecter l'activité des protéines et modifier la transcription des ARNm et la réplication de l'acide desoxyribonucléique (ADN) (7). Par conséquent, dans le cancer, les métabolites peuvent être, à la fois, la résultante d'une activation oncogénique et l'intermédiaire d'une autre voie d'activation métabolique.

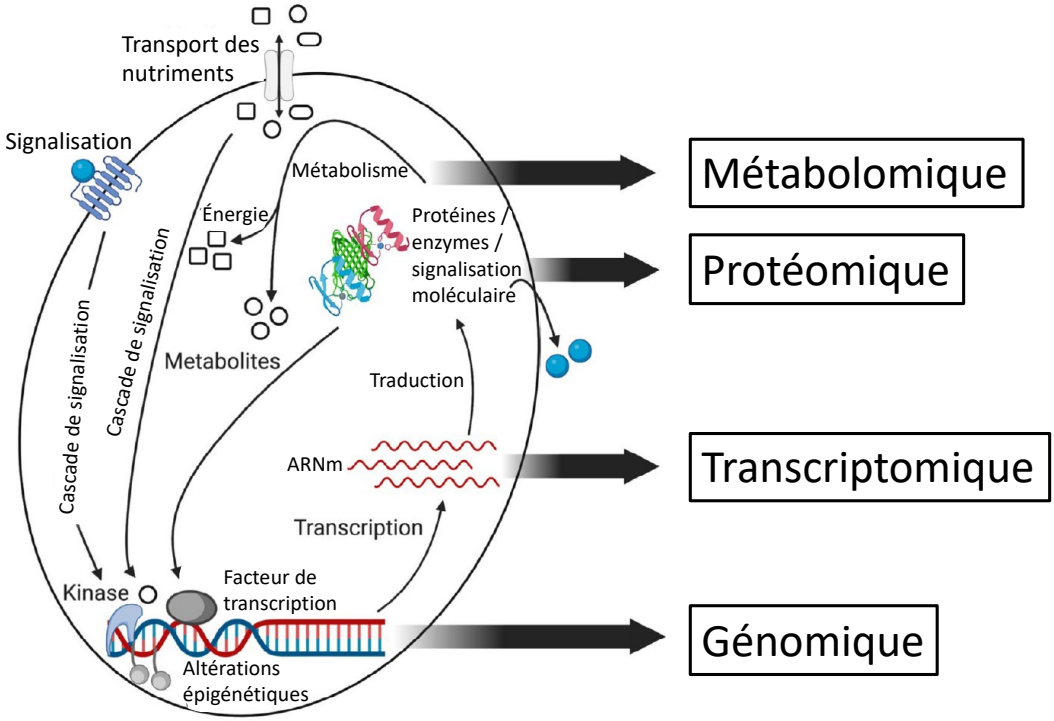


Figure 2. Interactions -omiques autour de la métabolomique *(adapté de Schmidt et al.(7)). ARNm : ARN messager.*

c) INTERET POUR LA COMPREHENSION DE LA BIOLOGIE DES SYSTEMES

L'analyse différentielle des signatures métabolomiques entre différents groupes d'échantillons vise à caractériser les répercussions de la modification d'un facteur externe, et à visualiser la manière dont un système biologique réagit. Elle permet, ainsi, une meilleure compréhension de la biologie des systèmes en mettant en évidence des interrelations métaboliques qui n'auraient pas pu être détectées avec des approches biochimiques traditionnelles. Jusqu'à maintenant, l'intérêt principal de l'analyse des métabolites portait sur la découverte de voies métaboliques fondamentales. Toutefois, au cours des dernières années, il est devenu évident qu'il y avait un besoin d'effectuer des études sur les réponses métaboliques d'un système biologique exposé à des perturbations environnementales, pathologiques et/ou génétiques et sur les variations phénotypiques qui en résultent afin de mieux comprendre les systèmes biologiques, dont le cancer.

Pendant la majeure partie de l'ère génomique, la biologie du cancer était axée sur la façon dont les voies de signalisation et les facteurs de transcription contrôlaient la croissance et la prolifération cellulaire au travers du prisme du cycle cellulaire. Ces dernières années ont été marquées par un intérêt nouveau pour la compréhension du métabolisme altéré intégré à la carcinogénèse, à travers l'étude de nombreux facteurs, tels que l'hypoxie tumorale, la composition stromale, l'infiltration de cellules immunitaires ou encore les altérations génétiques jouant un rôle dans le métabolisme cellulaire cancéreux(8–11). Par exemple, des altérations génétiques et/ou épigénétiques peuvent fournir un avantage crucial pour la survie des cellules cancéreuses dans un environnement carencé en nutriments(12–15). De plus, ces altérations spécifiques intrinsèques, ainsi que les modifications métabolomiques induites par les traitements anti-cancéreux peuvent également avoir un impact sur l'ensemble de l'organisme et interagir de manière complexe avec d'autres facteurs extérieurs comme l'exercice ou le régime alimentaire, et finalement être impliqué dans la survie ou la qualité de vie des patients(7).

Finalement, en oncologie, c'est le caractère global intégré de l'analyse qui différencie la métabolomique des autres analyses post-génomiques et par extension des analyses classiques. Ainsi, K.W. Jordan et L.L. Cheng définissent la métabolomique appliquée au cancer comme étant l'étude des variations métaboliques globales ainsi que la mesure des profils biochimiques liés aux voies métaboliques connues sous l'influence d'un processus oncogène (16,17).

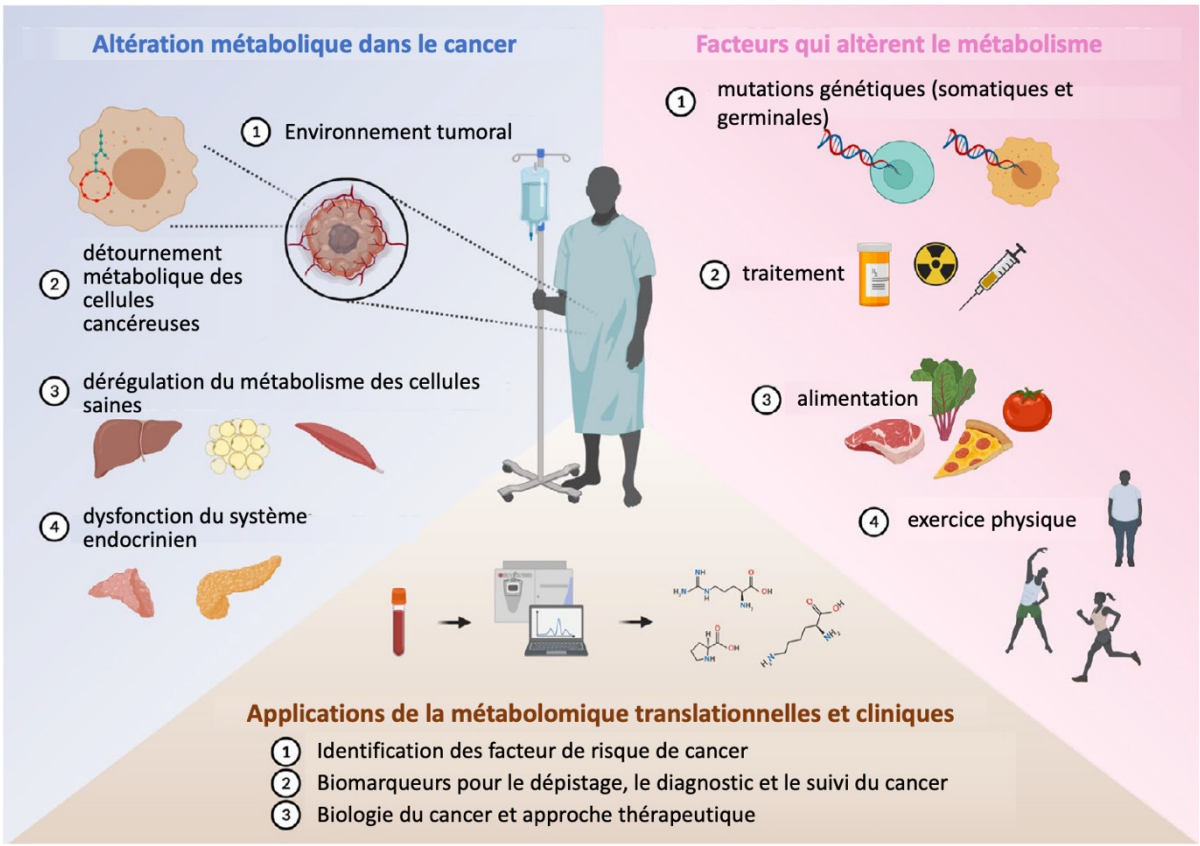


Figure 3. Vision intégrative de la métabolomique spécifique à la cancérologie *(d'après Schmidt et al.(7))*

d) METABOLOMIQUE ET BIOLOGIE DU CANCER

Un certain nombre d'autres travaux ont prouvé le fort potentiel de la métabolomique dans l'optique de la caractérisation et de la compréhension des processus oncologiques(18–27). Étant donné que les cellules cancéreuses ont un taux de croissance et de prolifération soutenu qui nécessite un approvisionnement constant en précurseurs métaboliques, des modifications importantes du métabolisme cellulaire se produisent(28). L'un des changements métaboliques les plus remarquables observés dans les cellules cancéreuses est l'augmentation de la consommation de glucose, qui peut être détectée par la tomographie par émission de positons couplé au scanner (TEP-scanner) lors du diagnostic initial, de l'évaluation thérapeutique ou de la surveillance. Cette modification métabolique permet aux cellules cancéreuses d'assurer leur prolifération et leur survie en modifiant les voies de signalisation et les facteurs de transcription qui contrôlent la

glycolyse, la production d'énergie et la biosynthèse de composés nécessaires à leur prolifération. Les cellules cancéreuses favorisent la métabolisation du glucose par fermentation et non plus par phosphorylation oxydative, produisant une grande quantité de lactate, même en présence d'oxygène. Ce phénomène est connu sous le nom d'effet Warburg(29–31).

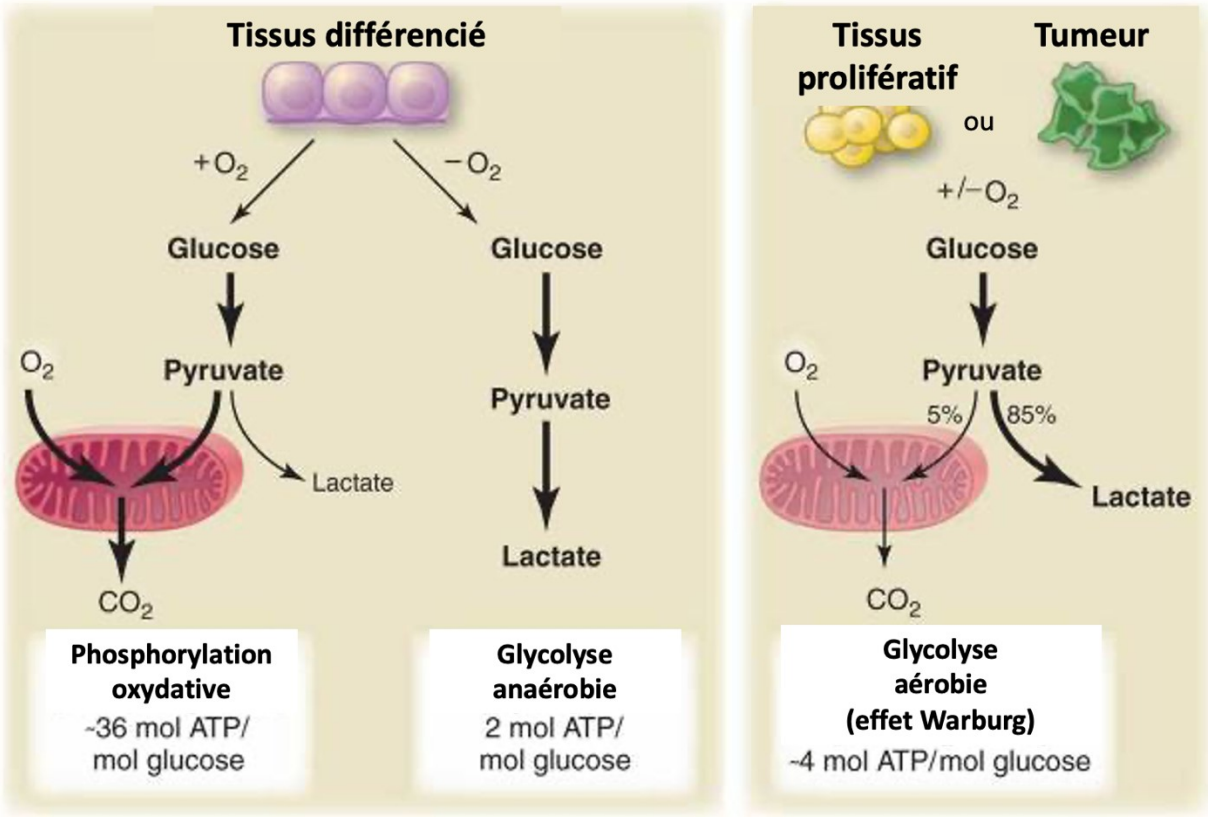


Figure 4. Représentation schématique de la phosphorylation oxydative, de la glycolyse anaérobie et de la glycolyse aérobie (effet Warburg) *(d'après Vander Heiden et al.(30)). CO_2 : dioxyde de carbone; 0_2 : dioxygène; ATP : adénosine triphosphate ; mol : molécule(s) .*

Dans une cellule de tissu différencié, en présence d'oxygène, le glucose est majoritairement transformé en pyruvate via la glycolyse, pyruvate ensuite oxydé en CO_2 dans les mitochondries par phosphorylation oxydative. L'oxygène étant nécessaire en tant qu'accepteur final d'électrons pour oxyder complètement le glucose, il est essentiel à ce processus. En condition d'anaérobie, les cellules peuvent détourner le pyruvate en générant du lactate (glycolyse anaérobie). Cette production de lactate pendant la

glycolyse anaérobie permet à la glycolyse de se poursuivre mais entraîne une production minimale d'adénosine triphosphate (ATP) par rapport à la phosphorylation oxydative (2 mol ATP/mol glucose vs 36 mol ATP/mol glucose). A contrario, les cellules cancéreuses ont tendance à convertir la majeure partie du glucose en lactate (85%), indépendamment de la présence d'oxygène (glycolyse aérobie). Cette propriété est partagée par les tissus prolifératifs normaux. Les mitochondries restent fonctionnelles et la phosphorylation oxydative se poursuit à moindre mesure (environ 5%) à la fois dans les cellules cancéreuses et dans les cellules normales en prolifération. Cependant, la glycolyse aérobie est moins efficiente que la phosphorylation oxydative pour produire de l'ATP (4 mol ATP/mol glucose contre 36 mol ATP/mol glucose). Toutefois, la glycolyse anaérobie présente tout de même un avantage car elle permet la régénération du NAD+ nécessaire à l'équilibre adéquat des cofacteurs d'oxydoréduction et aux processus biosynthétiques. Cette régénération est rendue possible grâce à la transition du pyruvate en lactate, catalysée par l'enzyme lactate déshydrogénase (LDH), qui est régulée par le couple NAD+/NADH(32,33). Pour finir, dans les cellules en prolifération, environ 10 % du glucose est détourné vers des voies de biosynthèse en amont de la production de pyruvate(30).

Par ailleurs, les cellules cancéreuses peuvent également métaboliser d'autres nutriments tels que la glutamine pour soutenir leur survie et leur croissance. L'oncogène MYC a été lié à une augmentation de la glutaminolyse, ce qui entraîne une addiction à la glutamine dans certaines cellules cancéreuses(34). La désactivation de la protéine suppresseur de tumeur du rétinoblastome (Rb) a également été démontrée pour augmenter l'utilisation de la glutamine en raison de la régulation à la hausse du transporteur de glutamine, l' Alanine, Serine, Cysteine Transporter 2 (ASCT2), dans les cellules cancéreuses humaines(35). Bien que le glucose et la glutamine soient les nutriments les plus abondants dans le plasma et les milieux de culture tissulaire, la liste des nutriments consommés par les cellules cancéreuses et les tumeurs en culture est vaste. Il est probable

que, dans la plupart des cas, le besoin en nutriments spécifiques dépende du type de tumeur et de l'environnement nutritif.

Les cellules cancéreuses sont donc impliquées dans un métabolisme modifié par cascades et interactions entre les voies de signalisation cellulaire. Par exemple :

- les récepteurs de la tyrosine kinase induits par l'insuline ou d'autres facteurs de croissance activent la voie de signalisation PI3K-AKT (phosphatidylinositol 3-kinase - serine/threonine kinase) pour stimuler la glycolyse(36). AKT peut augmenter directement l'activité glycolytique en phosphorylant l'hexokinase (l'enzyme qui catalyse la première étape de la glycolyse) ou indirectement en phosphorylant des substrats qui régulent le trafic des transporteurs de glucose 1 et 4 (GLUT1 et GLUT4) vers la membrane plasmique(37–39).

- La voie PI3K-AKT active également le complexe mTORC1, entraînant une augmentation de l'expression du facteur induit par l'hypoxie 1 (HIF1α)(40–43). Les sous-unités α de HIF, normalement dégradées en présence d'oxygène(44), sont stabilisées en présence d'hypoxie et activent une réponse transcriptionnelle qui permet une adaptation au stress hypoxique, notamment une augmentation de l'expression de GLUT1 et GLUT3, d'hexokinase 2 (HK2) et de certaines isoformes de la phosphofructokinase 2(45).

- HIF1α favorise également l'expression de la pyruvate déshydrogénase kinase, qui inhibe l'oxydation du pyruvate et détourne le métabolisme du glucose vers le lactate en adaptation aux conditions hypoxiques(46).

- D’autre part, la perte d'activité de la tumor protein 53 (TP53) peut pousser les cellules vers la voie glycolytique au lieu de la phosphorylation oxydative.

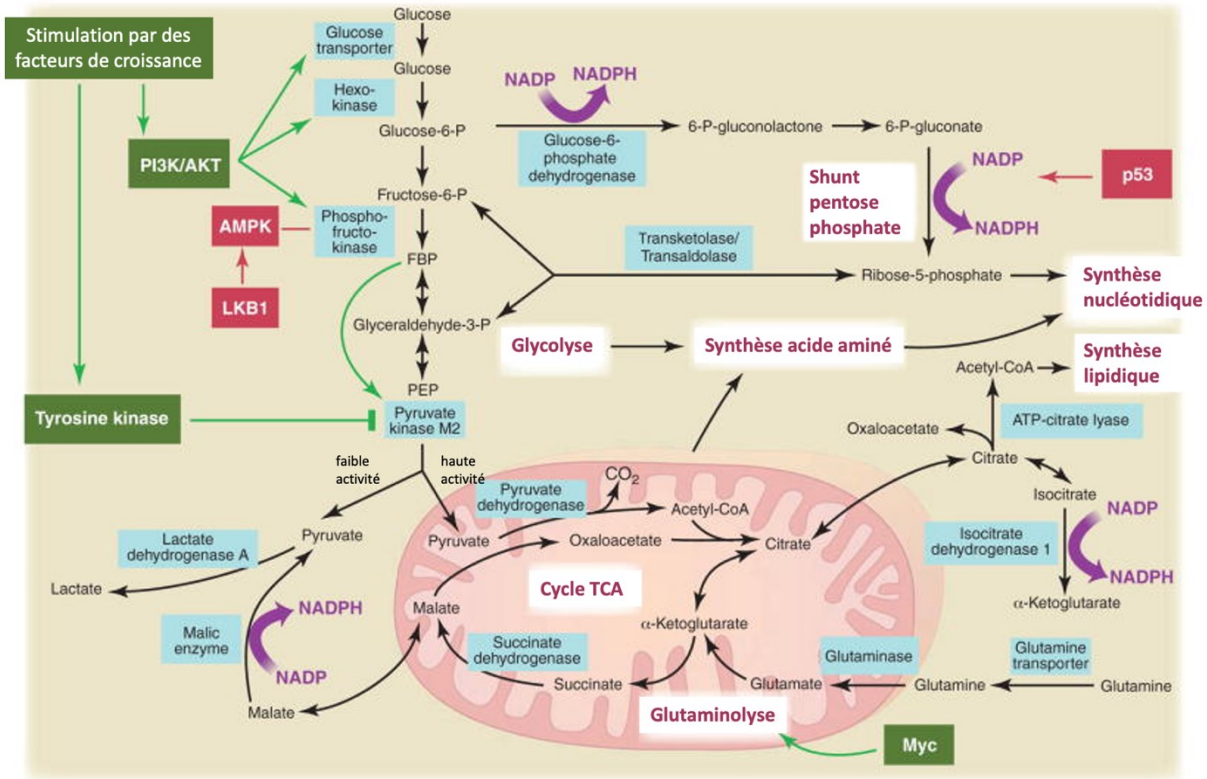


Figure 5. Interaction entre les voies métaboliques et les gènes impliqués dans la carcinogénèse (oncogènes et gènes suppresseurs de tumeur)*. Cycle TCA: tricarboxylic acid cycle (cycle de Krebs); PI3K: phosphatidylinositol 3-kinase; AKT: serine/threonine kinase; AMPK: AMP-activated protein kinase; LKB1: kinase hépatique B1; NADP: nicotinamide adénine dinucléotide phosphate; NADPH: nicotinamide adénine dinucléotide phosphate hydrogène; ATP: adénosine triphosphate; CO_2 : dioxyde de carbone.*

L'étude des métabolites dans le cancer peut également permettre de comprendre comment un métabolisme déficient peut déclencher ou favoriser la prolifération, l'angiogenèse et la transition épithélio-mésenchymateuse (TEM)(47,48). La reprogrammation métabolique des cellules cancéreuses et du stroma adjacent est une étape clé du développement du cancer(49,50). Le modèle biologique actuel de la carcinogenèse met en évidence diverses voies pour ce processus, telles que l'échappement aux mécanismes impliqués de régulation de la croissance cellulaire, la résistance à la mort cellulaire, l'instabilité et les mutations génomiques, la réplication de cellules immortalisées, l'induction de la capacité métastatique, l'inflammation induite par la tumeur et l'échappement au système immunitaire(51,52). La production de cytokines,

de chimiokines et de médiateurs immunitaires, l'activation et l'expansion des cellules immunitaires nécessitent une quantité importante d'énergie. Ceci suggère que des changements dans la disponibilité de l'énergie peuvent affecter la réponse immunitaire(53,54). Ces dépendances constituent de nouvelles cibles thérapeutiques. Par exemple, l'activation de la voie glycolytique dans les lymphocytes T CD4 favorise un phénotype inflammatoire, tandis qu'une augmentation de l'oxydation des acides gras les oriente vers un phénotype régulateur(55). Les récepteurs des lymphocytes B et T activent directement des facteurs de transcription, tels que c-MYC, HIF1α, PI3K, mTOR et FOXO1, qui jouent un rôle clé dans le métabolisme des cellules immunitaires et la réponse immunitaire en aval(56–58). De même, les macrophages de type M1 dépendent des voies de la glycolyse et de la glutaminolyse, tandis que les macrophages de type M2 préfèrent la voie de la phosphorylation oxydative pour répondre à la demande élevée de production d'énergie pendant la phase d'activation(59). Les cellules tumorales et les cellules immunitaires sont donc interconnectées par des facteurs de transcription, mais également par le métabolisme global et les ressources disponibles à un instant déterminé.

En résumé, il existe des interactions multiples et souvent complexes entre les voies métaboliques et les voies de signalisation qui, ensemble, aboutissent à une reprogrammation métabolique, une caractéristique fondamentale du cancer. En outre, des facteurs intrinsèques (altérations génomiques/épigénomiques) et extrinsèques (nutriments, médicaments, hormones et interactions avec les cellules stromales, la matrice extracellulaire et le système immunitaire) contribuent à la reprogrammation métabolique des cellules cancéreuses.

e) METABOLOMIQUE ET APPLICATION EN CANCEROLOGIE CLINIQUE

La métabolomique présente également un fort potentiel applicatif en cancérologie clinique. Cela demanderait un investissement initial financier relativement modeste pour l'achat d'un spectromètre ciblé et la mise en place d'une expertise de routine sur les

différentes plateformes d'analyses. En effet, la recherche de signature nécessite des techniques analytiques plus sophistiquées (spectromètres non ciblés à haute résolution et haute fréquence) et génère une grande quantité de données dont l'interprétation peut être complexe, nécessitant une approche multidisciplinaire. Cependant, une fois les signatures de quelques métabolites pré-identifiées, l'analyse est simple et rapide sans nécessiter un travail important de post-traitement des données. De plus, l'ensemble de la procédure tend vers une automatisation pour y compris dans la préparation des échantillons, ce qui facilitera son applicabilité clinique.

L'approche métabolomique pourrait être utilisée pour le diagnostic précoce des cancers(17,22). Par exemple, Jordan KW et Cheng LL *et al.* ont analysé les profils métaboliques, basés sur la spectroscopie par résonance magnétique nucléaire (RMN), de tissus cancéreux et de tissus sains prélevés chez des patients atteints de cancer de la prostate. Ils ont identifié plusieurs métabolites, tels que les acides aminés et les lipides, qui présentent des différences significatives entre les deux types de tissus. Ces métabolites ont été sélectionnés pour leur potentiel en tant que biomarqueurs pour le diagnostic précoce. Ainsi, en analysant les profils métaboliques de biopsies de prostate de patients, les cliniciens pourraient détecter le cancer à un stade précoce. Le profilage métabolique pourrait être combiné à d'autres outils de diagnostic et de suivi, tel que l'imagerie ou le suivi de l'ADN tumoral circulant. Le même principe a été appliqué dans le cancer du sein. Des résultats intéressants sont développés dans le paragraphe « Apport de la métabolomique dans la biologie du cancer du sein ».

Grâce à l'avènement des biopsies liquides, les praticiens pourraient également surveiller l'efficacité du traitement. Les analyses métabolomiques réalisées sur les fluides (salives, urines, plasma) seraient moins invasives, plus faciles à mettre en place que les biopsies tissulaires itératives, plus fiables, et peu coûteux, une fois établies en routine clinique.

L'approche métabolomique permettrait aussi l'identification des nouvelles voies métaboliques spécifiques impliquées dans l'initiation, la progression ou l'expansion du cancer, ce qui pourrait ouvrir la voie à de nouveaux traitements ciblés. Les thérapies médicamenteuses actuelles reposent souvent sur une approche universelle, non optimale car les patients peuvent avoir des réponses différentes aux médicaments en raison de variations génétiques, épigénétiques et de facteurs environnementaux. De plus, en analysant le profil métabolique d'un patient avant et après le traitement médicamenteux, les cliniciens pourraient identifier les voies métaboliques affectées ou activées par les traitements et utiliser ces informations pour optimiser la thérapie spécifique. Par ailleurs, en comprenant la réponse métabolique aux médicaments, les chercheurs pourraient concevoir des médicaments qui ciblent sélectivement des voies métaboliques spécifiques, réduisant ainsi le risque d'effets secondaires en pratique clinique. Par conséquent, une approche plus personnalisée de la thérapie médicamenteuse serait possible(60).

Pour finir, avec l'arrivée de l'immunothérapie, l'approche métabolomique pourrait permettre l'étude des voies métaboliques spécifiques impliquées dans l'échappement immunitaire. Il serait alors possible de déterminer de nouvelles cibles thérapeutiques et de proposer des association thérapie ciblée/immunothérapie pour restaurer plus efficacement l'immunité anti-tumorale(61).

2. Introduction au cancer du sein localisé

a) LE CANCER DU SEIN

Le cancer du sein est le cancer le plus fréquemment diagnostiqué dans le monde(62). En France, 58 000 nouveaux cas ont été détectés en 2018. Le cancer du sein se situe donc au premier rang des cancers chez la femme. Environ une femme sur onze développera un cancer du sein dans sa vie. Selon l'estimation du Centre International de Recherche sur le

Cancer, pour 2018, l'incidence annuelle, ajustée pour l'âge, pour 100 000 femmes est de près de 100. La maladie peut se voir à tout âge, mais l'âge moyen de découverte est 62 ans. La moitié des cancers du sein sont diagnostiqués entre 50 et 69 ans, 20 % avant 50 ans et 10 % avant 40 ans(63–65).

Actuellement environ 60 % des cancers du sein sont diagnostiqués au stade localisé (stade I), 30 % à un stade localement avancé (stade II-III) et 10% à un stade métastatique (stade IV)(64). Même en contexte localisé et localement avancé, un certain nombre de patientes vont présenter une récidive avec un risque non négligeable de décès spécifique, et ce malgré les progrès récents. Les taux de survie sans maladie (SSM) à 5 ans varient selon le stade de 98 à 100 % pour le stade I, 85 à 98 % pour le stade II à environ 70 à 95% pour le stade III(66). Le cancer du sein est la première cause de décès chez la femme, avec 14 % des décès féminins par cancer en 2018. Il représente par ailleurs près de 8 % de l'ensemble des décès par cancer. La survie à 5 ans, tous stades confondus, est de 88 %(63,67).

Il existe plusieurs sous-types histologiques de cancer du sein. L'histologie la plus fréquente est le carcinome invasif non spécifique (NST, non-special type)(80% des cas), qui se développe dans les canaux mammaires qui transportent le lait vers le mamelon (canaux galactophores). La deuxième histologie la plus fréquente est le carcinome lobulaire invasif (CLI)(10-15% des cas), qui se développe dans les lobules, les structures glandulaires qui produisent le lait. D'autres histologies sont plus rares, comme les carcinomes mucineux, papillaires, tubuleux, inflammatoires, ou médullaires qui représentent environ 10% des cas(68)(69).

Les cancers du sein sont ensuite classés en fonction de la présence ou non de récepteurs hormonaux et HER2.

- Les cancers du sein dit hormonodépendants sont des cancers qui ont des récepteurs hormonaux (récepteur aux œstrogènes RE ou récepteur aux progestérones RP) sur la surface des cellules cancéreuses, ce qui signifie qu'ils ont besoin d'hormones telles que les œstrogènes ou la progestérone pour se développer et se propager. Environ 60%-75 % des cas de cancer du sein sont hormonodépendants(70).
- Le cancer du sein HER2-positif est un type de cancer du sein caractérisé par une surexpression de la protéine HER2. La protéine HER2 est une protéine de surface qui aide à réguler la croissance et la division cellulaire, mais une surexpression de cette protéine peut entraîner une croissance excessive et un comportement agressif des cellules cancéreuses par inhibition de l'apoptose. Le cancer du sein HER2-positif représente environ 15 à 20% de tous les cancers du sein(71).
- Le cancer du sein triple négatif (TNBC) est un type de cancer du sein qui ne présente ni récepteurs hormonaux, ni surexpression de la protéine HER2. D'autres voies d'activation cellulaires rentrent en jeu. Le TNBC représente environ 10 à 15% des cas de cancer du sein et est plus fréquent chez les femmes jeunes(68)(72).

b) LA NECESSITE D'UN NOUVEL OUTIL PREDICTIF

Le taux de mortalité lié au cancer du sein diminue depuis les années 1970(73) grâce à l'amélioration du dépistage du cancer du sein et la mise en place du traitement adjuvant par chimiothérapie, thérapies ciblées, hormonothérapie et radiothérapie(74–77). Par exemple, dans la méta-analyse de l'EBCTCG en 2012, l'utilisation d'un traitement contenant des anthracyclines, comparée à l'absence de traitement, permettait d'obtenir une diminution du risque de récidive de 47 à 39 % (RR 0,73 ; IC95% 0,68-0,79), une diminution de la mortalité due au cancer du sein de 36 à 29 % (RR 0,79 ; IC95% 0,72-0,85) et une diminution de la mortalité globale de 40 à 35 % (RR 0,84 ; IC95% 0,78-0,91)(78).

En parallèle, les risques de la chimiothérapie comprennent des toxicités aiguës telles que les nausées, les vomissements, l'alopécie, la myélosuppression, les troubles cognitifs précoces(79), une perte de fertilité, un risque infectieux et des neuropathie. Parfois, ces neuropathies persistent au long cours et deviennent séquellaires(80). Les toxicités à long terme comprennent également les risques de cardiotoxicité associés aux anthracyclines et le risque rare mais non négligeable de leucémie secondaire liée à la chimiothérapie(81).

Les patientes atteintes d'un cancer du sein triple négatif localement avancé (stade II-III) et traitées par immunochimiothérapie néoadjuvante peuvent également faire face aux effets indésirables de l'immunothérapie. À ce jour, aucun critère prédictif n'a été clairement identifié pour déterminer quels patients sont susceptibles de bénéficier de l'adjonction de l'immunothérapie d'une part, et/ou de développer des effets indésirables d'autre part. L'immunothérapie est donc proposée à toutes les patientes, malgré les risques non négligeables de maladies auto-immunes telles que la thyroïdite, la colite, l'hépatite, l'hypophysite ou la néphrite(82).

L'évaluation des balances bénéfice/risque et efficacité/tolérance est donc primordiale. En effet, l'ampleur du bénéfice de la (immuno-)chimiothérapie dépend du risque initial de récidive, le bénéfice absolu chez les patients présentant un faible risque de récidive pouvant être faible. Ainsi, de nombreux patients sont surtraités (quand la chimiothérapie n'apporte aucun bénéfice supplémentaire) et d'autres sous-traités (quand une chimiothérapie pourtant utile n'a pas été réalisée) par manque de facteurs pronostiques (évaluant l'agressivité potentielle de la maladie) et prédictifs (orientant vers une efficacité attendue de certains traitements) fiables. Il est donc indispensable d'identifier au mieux les patientes les plus susceptibles de tirer un bénéfice de ces traitements afin de permettre une médecine personnalisée à chaque patient.

Le risque de récidive peut être estimé à partir des caractéristiques cliniques, notamment le stade et le grade de la tumeur, et des caractéristiques biologiques de la tumeur, y compris l'expression génétique. Le domaine de la découverte de marqueurs en oncologie est en constante évolution, grâce à une meilleure compréhension de la biologie des tumeurs et à la connaissance du génome humain. Cependant, seule une petite proportion de marqueurs s'est avérée cliniquement utile jusqu'à ce jour, que ce soit à visée pronostique, prédictif ou théranostique.

c) LA CLASSIFICATION CLINICO-PATHOLOGIQUE

Pour les femmes atteintes d'un cancer du sein non métastatique nouvellement diagnostiqué, nous utilisons systématiquement plusieurs facteurs cliniques lors du processus décisionnel de prescription d'une chimiothérapie adjuvante :

1) *L'âge :* l'âge au moment du diagnostic est associé à un pronostic plus défavorable(83). Les patientes âgées de moins de 35 ans ont une survie globale et une survie sans récidive plus faibles(84,85). Ces patientes présentent généralement un stade plus avancé, plus souvent une maladie ER-négative. La mortalité par cancer du sein est également plus élevée chez les patientes plus âgées (>65 ans)(86–88), probablement dû à un stade plus avancé au moment du diagnostic, à des comorbidités et à des insuffisances de traitement (sous-traitement dû à l'âge)(87).

2) *Le statut ménopausique :* le statut ménopausique constitue un facteur favorable. Chez les patientes préménopausées ayant reçu une chimiothérapie adjuvante, l'aménorrhée induite par la chimiothérapie et l'absence de reprise des cycles menstruels après la chimiothérapie seraient associées à une amélioration de la survie, après prise en compte des variables pronostiques standard, en particulier pour les cancers hormonodépendants(89).

3) *Le stade tumoral :* en général, le stade est un facteur pronostique. Le cancer du sein est stadifié en fonction de la taille de la tumeur (T), de la présence ou non d'un envahissement ganglionnaire et son importance (N), et de la présence ou non d'une maladie métastatique (M). La taille de la tumeur a été reconnue très tôt comme un facteur pronostique important dans le cancer du sein(90–92). Le cancer du sein inflammatoire (T4d) est une forme rare mais très agressive de cancer du sein, caractérisée par un tableau clinique d'inflammation locale associée un mauvais pronostic. L'atteinte ganglionnaire est un facteur pronostique négatif fort et indépendant(93)(cf. Annexe 1 TNM et Annexe 2 stade).

4) *Le Grade histopronostique:* le grade est déterminé selon le système de classification d'Elston-Ellis dérivé de celui de Scarff-Bloom-Richardson, qui caractérise le degré de différenciation de la tumeur par le pourcentage de différenciation tubulo-glandulaire, le pléomorphisme nucléaire (degré d'atypie) et l'activité mitotique. Chacun de ces critères est évalué et noté de 1 à 3, où 1 représente un degré élevé de différenciation et un faible niveau d'atypie et d'activité mitotique, tandis que 3 représente un degré faible de différenciation et un niveau élevé d'atypie et d'activité mitotique. Les notes de chaque critère sont ensuite additionnées pour donner un score final permettant d'identifier les tumeurs bien différenciées (grade 1), modérément différenciées (grade 2), ou peu différenciée (grade 3). Le grade histopronostique est utilisé pour aider à prédire le pronostic de la maladie et à guider les choix de traitement. Les tumeurs de grade 1 ont tendance à avoir un meilleur pronostic et à être moins agressives que les tumeurs de grade 3.(94)(cf. Annexe 3 Grade).

5) *L'invasion lymphovasculaire péritumorale* : la présence d'une invasion lymphovasculaire semble être un indicateur de mauvais pronostic, en particulier dans les tumeurs de haut grade(95).

6) *Le Ki-67 :* la relation entre le statut Ki-67 et le pronostic dans le cancer du sein au stade précoce a été largement étudiée(96). Malgré l'hétérogénéité des essais cliniques et des méthodes d'évaluation du Ki-67 utilisées, les résultats de deux grandes méta-analyses sont cohérents avec la valeur pronostique indépendante du Ki-67(97,98).

7) *Les récepteurs hormonaux :* l'expression des récepteurs ER et PR est généralement associée à une amélioration des données de survie, du moins à court terme. La survie globale, la survie sans maladie et le délai avant l'échec du traitement sont tous positivement liés aux niveaux de RE et de RP(99–101). Cependant, bien que le taux annuel de récidive des cancers ER-positifs soit plus faible dans les 5 premières années que celui des cancers ER-négatifs, des études suggèrent qu'il pourrait être plus élevé à plus long terme(102,103). L'absence d'expression du récepteur à la progestérone était associée à un pronostic plus défavorable en survie globale, en survie spécifique et en survie sans maladie. Ces données sont corroborées par le fait que les patientes présentant une maladie ER-positive et PR-négative ont un sous-type plus agressif de cancer du sein à récepteurs hormonaux positifs (104), et sont souvent étiquetées « tumeurs luminales B »(105).

8) *Surexpression de HER2 :* la surexpression de HER2 est associée à un pronostic défavorable, en particulier si les patientes ne sont pas traitées par chimiothérapie et par thérapies ciblées anti-HER2(106,107).

9) *Lymphocytes infiltrant la tumeur (TIL) :* l'augmentation de la concentration de TILs a permis de prédire la réponse à la chimiothérapie néoadjuvante dans tous les sous-types moléculaires évalués, et a été associée à un avantage en termes de survie dans le cancer du sein HER2-positif et le cancer du sein triple négatif. En revanche, ils se sont révélés être un facteur pronostique défavorable pour la survie dans le cancer du sein luminal-HER2-négatif(108), ce qui suggère une biologie différente de l'infiltrat immunologique

dans ce sous-type. Des recommandations pour l'évaluation standardisée des TILs ont été publiées(109).

d) CALCULATEUR DE RISQUE

Plusieurs calculateurs de prédiction du risque clinique existent et permettent d'estimer le risque de récidive à partir de données cliniques et histopronostiques.

Par exemple, l'outil en ligne ESTIMATE a été élaboré à partir d'un registre de plus de 264 000 femmes atteintes d'un cancer du sein invasif non métastatique, et estime le risque cumulatif résiduel de mortalité spécifique par cancer du sein (BCSM) et de mortalité non spécifique par cancer du sein jusqu'à 20 ans après le diagnostic initial(110,111). Cet outil n'estime cependant pas l'apport des thérapeutiques.
Le clinical treatment score 5 (CTS5) classe les patientes qui n'ont pas eu de récidive pendant les cinq premières années de surveillance dans les catégories de risque faible, intermédiaire et élevé de risque absolu de récidives à distance entre 5 et 10 ans(112,113). Cet outil est surtout utilisé pour évaluer les bénéfices d'une prolongation de l'hormonothérapie.

L'outil UK PREDICT a été établi à partir d'une analyse d'environ 6000 patientes et estime l'impact de différentes thérapies adjuvantes telles que la chimiothérapie, l'hormonothérapie, le trastuzumab et les biphosphonates. En utilisant les caractéristiques cliniques et pathologiques de chaque patient, PREDICT est conçu pour fournir des informations pronostiques personnalisées sous la forme d'estimations de la survie globale (SG) à 5 et 10 ans(114). Ce modèle a été validé dans plusieurs séries de cas indépendantes(115–120). Le statut du récepteur HER2 et le statut Ki67 ont également été incorporés dans le modèle, ce qui a permis d'améliorer légèrement la discrimination du modèle(120,121). L'outil UK PREDICT a été utilisé lors de cette thèse (en 2020) pour

prédire les données de survie lorsque le suivi n'était pas suffisant pour calculer les données de survie réelle.

Une validation externe à grande échelle de l'algorithme PREDICT a été réalisée l'année dernière à partir de la national cancer database (NCDB), base de données d'oncologie clinique provenant de données de registres hospitaliers représentant plus de 70 % de tous les cas de cancer du sein nouvellement diagnostiqués aux États-Unis. De 2004 à 2017, près de 3 millions de cas de cancer du sein ont été collectés dans la NCDB. Parmi ces cas, les femmes âgées de 25 à 85 ans atteintes d'un cancer du sein invasif unilatéral au stade précoce et traitées entre 2004 et 2012 ont été identifiées. Au total, 708 652 patients éligibles avec un suivi complet sur 5 ans et 233 455 patients avec un suivi complet sur 10 ans ont été identifiés dans la NCDB, avec des durées médianes de suivi de 97,7 mois et 137,2 mois, respectivement. Les statistiques de l'aire sous la courbe (AUC) pour la prédiction de la SG à 5 et 10 ans étaient respectivement de 0,772 et 0,778. Les différences absolues de mortalité à 5 ans et à 10 ans entre les valeurs prédites et observées étaient respectivement de 0,02 % à 0,09 % et de 0,01 % à 0,11 %, en fonction du nombre de ganglions positifs. Dans la cohorte complète, la différence absolue de mortalité à 5 ans était de 0,05 % et la différence absolue de mortalité à 10 ans était de 0,06 %(122).

Ces calculateurs ne doivent cependant pas être considérés comme un substitut aux tests diagnostiques moléculaires, car ils ne prennent pas en compte tous les biomarqueurs et les relations entre la biologie des tumeurs et les effets du traitement. En effet, des patientes aux mêmes caractéristiques cliniques et pathologiques peuvent avoir des évolutions cliniques très différentes et cela ne peut pas être prédit par des modèles clinico-pathologiques, d'où l'intérêt de classifieur hybride prenant en compte à la fois les données clinico-pathologiques et les données biologiques afin d'établir une meilleure classification individuelle.

e) LA CLASSIFICATION INTRINSEQUE DITE « CLASSIFICATION MOLECULAIRE »

Les études d'expression génique de Perou et Sorlie ont permis d'identifier plusieurs sous-types distincts de cancer du sein qui diffèrent nettement par leur pronostic et par les cibles thérapeutiques qu'ils expriment(123–131). La liste de gènes qui différencie ces sous-types est appelée liste intrinsèque et se compose de plusieurs groupes de gènes liés à l'expression du récepteur des œstrogènes (ER) (le groupe luminal), à l'expression du récepteur 2 du facteur de croissance épidermique humain (HER2) et à un groupe unique de gènes appelé groupe basal.

Les tumeurs luminales A, qui représentent environ 40 % de tous les cancers du sein, présentent généralement une forte expression des gènes liés au RE, une faible expression du groupe de gènes HER2 et une faible expression des gènes liés à la prolifération(132,133). Les tumeurs luminales A sont le sous-type le plus courant et, en général, ont le meilleur pronostic de tous les sous-types de cancer du sein(128–130,134–136).

Les tumeurs luminales B, moins fréquentes (environ 20%), présentent une expression relativement faible (bien que toujours présente) des gènes liés au RE, une expression variable du groupe HER2 (mutations ou amplifications géniques) et une expression plus élevée du groupe de prolifération. Les tumeurs luminales B ont un pronostic plus défavorable que les tumeurs luminales A(135). La plupart des cancers luminaux B ont des scores de récidive élevés, tels qu'évalués par le recurrence score à 21 gènes (Oncotype DX®), et des signatures pronostiques péjoratives à 70 gènes (Mammaprint®)(132).

Le groupe HER2-enrichi représente environ 10 à 15 % des cancers du sein et se caractérise par une forte expression des groupes de gènes de l'amplicon *erbB2* et de prolifération et une faible expression des groupes de gènes luminaux et basaux(137,138).

Environ 50% des tumeurs HER2 positives en immunohistochimie et en hybridation in situ sont ER positives et donc luminales B.

Pour finir, le groupe des ER-négatifs comprend de multiples sous-types, tels que les sous-types basal-like, claudin-low(124), interferon-rich(139), et les sous-types mésenchymateux et luminaux de récepteurs d'œstrogènes(140), entre autres. La plupart d'entre eux entrent dans la catégorie des cancers du sein triple négatifs car ils sont également PR négatifs et HER2 négatifs. Les tumeurs triple-négatives présentent une forte instabilité génétique. Plusieurs études ont montré que cette catégorie de tumeurs englobait la plupart des tumeurs mammaires liées à des mutations breast cancer 1 (BRCA1). Cependant, ce groupe englobe également des tumeurs de types histopathologiques rares, certains étant de bon pronostic comme les carcinomes adénoïdes kystiques(141).

Cependant, cette classification moléculaire intrinsèque n'est pas utilisée en routine clinique. La conférence internationale de consensus sur le cancer du sein de Saint-Gallen de 2013 a défini un substitut anatomo-pathologique, basé sur des modifications d'expression protéique visibles en immunohistochimie (IHC), qui a été largement utilisé(142) et qui sert toujours de base pour la prise en charge des patients ce jour : luminal A (ER+ and/or PR+, HER2− and Ki67% < 30%), luminal B HER2− (ER+ and/or PR+, HER2− and Ki67% ≥ 30%), luminal B HER2+ (ER+ and/or PR+ and HER2+), HER2-neu non-luminal (ER/PR− and HER2+) and basal-like (ER/PR− and HER2−). Cependant, certains sous-types moléculaires, tels que les sous-types Luminaux, ont été difficiles à classer en raison de divergences significatives (jusqu'à 30%) dans l'évaluation des paramètres immunohistochimiques avec la classification moléculaire évaluée par PAM50 [4]. Cette hétérogénéité d'expression moléculaire révèle la nécessité d'élaborer des signatures hybrides multi-omiques, spécifiques, afin d'améliorer le pouvoir discriminatif de l'étude anatomo-pathologique seule.

f) SIGNATURES GENOMIQUES

L'émergence des techniques de génomique et de transcriptomique et la capacité de mesurer simultanément l'expression de milliers de gènes ont conduit à l'identification de profils pronostiques basés sur la biologie. L'expression de plusieurs gènes impliqués dans la caractérisation, la progression et l'agressivité tumorale, parfois associée aux caractéristiques clinico-pathologiques permet d'établir un score de récidive ou d'estimer une survie sans métastases. Plusieurs scores ont été validés et sont utilisés en clinique. Bien que les contributions pronostiques de gènes spécifiques soient inconnues, le panel dans son ensemble est plus efficace pour le pronostic que les caractéristiques clinicopathologiques telles que le grade ou le Ki-67. Ces signatures pronostiques et prédictives ont pour objectif de permettre de diminuer le recours à la chimiothérapie et d'améliorer la qualité de vie des patientes sans risquer de perte de chance ; ou à l'opposé, d'indiquer une agressivité tumorale qui motiverait la mise en place d'une chimiothérapie adjuvante et d'éviter un sous-traitement.

Au niveau biologique, plusieurs OMIQUEs sont explorés :

- Le test Prosigna® analyse 50 gènes (PAM50) par n-counter à barcodes sur lames FFPE. Ce test concerne les tumeurs RH+/HER2– de stade précoce, avec ou sans atteinte ganglionnaire. L'analyse de ces 50 gènes associée à la taille de la tumeur et au nombre de ganglions atteints permet, grâce à un algorithme, de fournir le sous-type intrinsèque (luminal A/B, HER2, basal-like), le risque de récidive à 10 ans (pourcentage individuel et classification du risque) et le score ROR (risk of relapse) allant de 0 à 100(143). Parmi les 50 gènes d'intérêt, 46 permettent de définir les types moléculaires, et 19 gènes de prolifération permettent de définir un score de prolifération. Aucune étude prospective n'a été réalisée.

- Le test Mammaprint® identifie l'expression de 70 gènes par microarray sur FFPE ou prélèvements congelés. Il prédit le risque métastatique à 5 ans et stratifie les patientes en 2 groupes : faible et haut risque. Elle a été étudiée pour des tumeurs RE+/RE– et pN0/N+ dans un essai prospectif international MINDACT(144). Tian *et al.* ont rapporté une analyse d'interactions protéine-protéine sur les gènes impliqués dans cette signature. Les 70 gènes formeraient des réseaux fortement interconnectés et leurs niveaux d'expression seraient régulés par des gènes clés liés à la carcinogénèse tels que TP53, RB1, MYC, JUN et CDKN2A. Les fonctions biologiques des 70 gènes pourraient être associées aux étapes essentielles de la progression tumorale et des métastases : inhibition de l'apoptose, prolifération cellulaire, invasion et angiogenèse(145). Aucune étude prospective n'a été réalisée.

- Le Génomic Grade Index™ analyse 97 gènes par microarray sur FFPE ou prélèvements congelés. Les 97 gènes d'intérêt sont liés au cycle cellulaire et à la prolifération. Parmi les 20 gènes les plus surexprimés figurent UBE2C, KPNA2, TPX2, FOXM1, STK6, CCNA2, BIRC5 et MYBL2. Cette signature devait initialement être utile pour différencier les tumeurs de grade SBR II en 2 groupes de risque de rechute. Cependant, les études ont mis en évidence une catégorie intermédiaire, y compris au niveau génomique (equivocal genomic grade)(130,146). Finalement, comme l'index de prolifération Ki67, le Génomic Grade Index apporte des informations complémentaires pour mieux prédire le risque de rechute à distance. Cette signature a été utilisé dans l'essai prospectif Astrer 70 (sujets âgés > 70 ans).

- Les test Oncotype DX® analyse 21 gènes par quantitative reverse transcription - polymerase chain reaction (qRT-PCR) sur FFPE. Cette signature est utilisée dans deux études prospectives randomisées pour prédire la récidive à 10 ans de tumeurs RE+ sans envahissement ganglionnaire (TAILORx(147)) ou avec 1-3 ganglions atteints (RxPonder(144)). Plusieurs gènes sont étudiés : la voie des récepteurs hormonaux (RE, RP, BCL2, SCUBE2), la voie HER2 (HER2, GRB7), la prolifération (Ki67, STK1,

BIRC5, CCNB1, MYBL2) et l'invasion (MMP11, CTSL2). Selon leur niveau d'expression, le Recurrence Score (RS) est calculé (de 0 à 100).

- Le test EndoPredict® analyse 11 gènes par quantitative reverse transcription - polymerase chain reaction (qRT-PCR) sur FFPE. Les gènes d'intérêt sont DHCR7, AZGP1, MGP, STC2, BIRC5, UBE2C, RBBP8, IL6ST, gènes impliqués dans la carcinogénèse(148). Le score EndoPredict® (EP) a été évalué de façon rétrospective et a montré une valeur pronostique pour la récidive à 10 ans de tumeurs RE+/HER2. Cette signature se compose d'un score génomique associé à la taille tumorale et au statut ganglionnaire (score EPclin, variant de 0 à 6).

- Le test Breast Cancer Index (BCI) analyse 7 gènes par quantitative reverse transcription - polymerase chain reaction (qRT-PCR) sur FFPE. Les gènes d'intérêt sont HOXB13/IL17BR, BUB1, CENPA, NEK2, RACGAP1, RRM2.

De ce fait, au niveau validité et utilité clinique, seules 3 de ces signatures ont été validées par des essais randomisés prospectifs de phase III : Oncotype DX, MammaPrint et Genomic Grade Index. MINDACT(144), TAILORx(147) et RxPonder(149), ont apporté des perspectives supplémentaires sur l'utilisation de MammaPrint et d'Oncotype Dx pour les patientes atteintes d'un cancer du sein de pronostic intermédiaire. ASTER 70 s'est focalisé sur la population > 70 ans pour guider le choix d'une chimiothérapie. Nous ne développerons par la suite que les Test MammaPrint et Oncotype DX, dont les populations d'étude correspondent à la question de thèse explorée.

Les principaux résultats de MINDACT(144), TAILORx(147) et RxPonder(149) sont présentés ci-dessous :

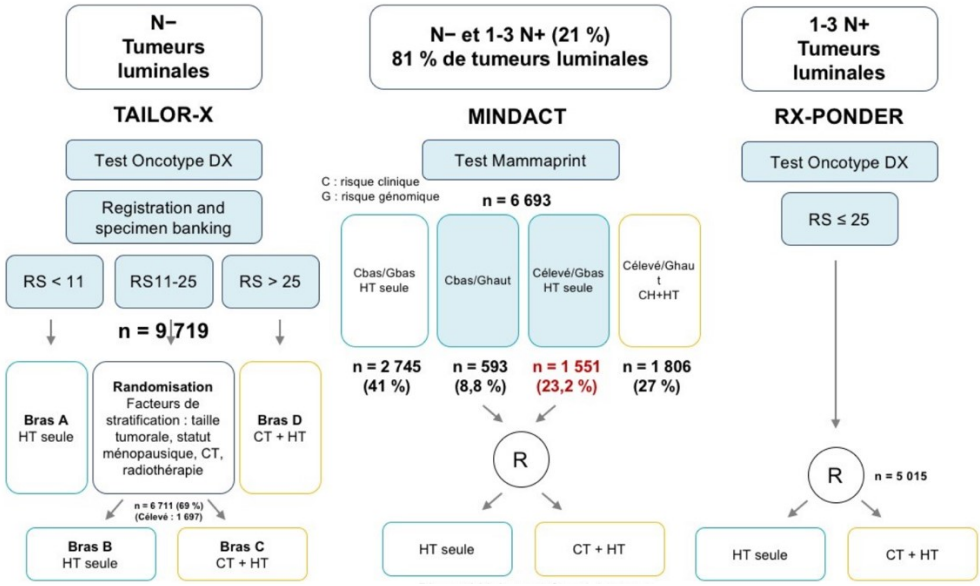


Figure 6. Principales études prospectives de validation de signatures génomiques.
N- : absence d'envahissement ganglionnaire; 1-3N+ : 1 à 3 ganglions envahis en axillaire ; RS : recurrence score ; n : effectif ; R : randomisation ; CT : chimiothérapie ; HT : hormonothérapie ; Cbas : risque clinique bas ; Célevé : risque clinique élevé ; Gbas : risque génomique bas ; Ghaut : risque génomique haut.

	TAILOR-X	MINDACT	RX-PONDER
Signature moléculaire	Oncotype Dx	Mammaprint	Oncotype Dx
Années de recrutement	2006-2010	2007-2011	2011-2015
Patients éligibles	9 719 N0, RH+/HER2−	6696 79% N0 et 21% 1-3 N+ 81% RH+/HER2−	5015 1-3 N+, RH+/HER2−
Critère principal	IDFS	DMFS	IDFS
Suivi médian	9 ans	8,7 ans	5 ans
Type de chimiothérapie	TC : 56% A (±T) : 36 %	T : 24% A (±T) : 64 %	TC : 50% A (±T) : 50 %
Résultats	IDFS HR : 1,08 (0,94-1,24) (limites supérieure requise < 1,32)	DMFS à 5 ans 95,1 % (93,1-96,6) (limites inférieure requise > 92 %)	P 0,30 HR 5 ans 0,81 (IC_{95} : 0,67-0,98)
Non-observance	12 %	13 %	5 %
Références	*Sparano JA, NEJM 2019*	*Cardoso F. ASCO 2020 abstract 506* *Piccart MJ, Lancet Oncol 2021 In press*	*Kalinsky K., SABCS 2020 abstract GS3-00,*

Tableau 2. Principaux résultats des études prospectives de validation de signature génomique. *IDFS: invasive disease-free survival; DMFS: distant metastasis-free survival ; TC : taxane-cyclophosphamide; A(+/-T) : anthracyclines (+/- taxanes); HR : hazard ratio;*

Finalement, les principales signatures disponibles sont résumées dans la table ci-dessous :

Signature		Prosigna®	MammaPrint®	Oncotype DX®	Breast Cancer Index℠	Genomic Grade Index™	EndoPredict®
Population analysée		Tous cancers	Tout RE N0 ou 1-3N+	RE+ N0 et 1-3N+	RE+ N0	Tous cancers	RE+ N0 ou post-ménopause N+
Objectif principal		Survie sans métastase à 10 ans	Survie sans métastase à 10 ans	Survie sans métastase à 10 ans	Survie sans rechute à 10 ans	Survie sans rechute	Survie sans métastase à 10 ans
Valeur pronostique	Rechute précoce (5 ans)	Non	**Oui (bas risque)** *(Cardoso F, N Engl J Med 2016)*	**Oui** *(Sparano JA, N Engl J Med 2015)*	Non	**Oui**	Non
	Rechute tardive (10 ans)	**Oui**	**Oui (bas risque)** *(Cardoso F, J Clin Oncol 2020)*	**Oui** *(Sparano JA, N Engl J Med 2018)*	**Oui**	Non	**Oui**
Valeur prédictive	Hormonothérapie	Oui *(Chia SK, Clin Cancer Res 2012)*	Non *(Esserman LJ, JAMA Oncol 2017)*	Oui *(Paik S, J Clin Oncol 2005)*			
	Chimiothérapie	Anthracyclines + taxanes	**Oui (< 50 ans et bas risque)**	**Oui (< 50 ans et risque intermédiaire)**	Anthracyclines + taxanes	Non	Anthracyclines
Validation prospective		–	MINDACT	TAILORx RxPONDER	–	ASTER 70 (> 70 ans)	–

Tableau 1. Principaux tests moléculaires commercialisés *(adapté de (150).*
RE : récepteur aux œstrogènes; N : node; N0 : absence d'envahissement ganglionnaire; N+ : envahissement ganglionnaire; 1-3N+ : 1 à 3 ganglions axillaires envahis.

Au vu de l'ensemble de ces données, une revue bibliographique actualisée a été réalisée par André *et al.*(151) dans le cadre de la mise à jour des recommandations de l'ASCO en 2022, reprenant l'ensemble des essais cliniques publiés entre janvier 2016 et octobre 2021 et comprenant des données de survie globale, de survie sans maladie ou sans récidive. Ci-dessous sont résumées les indications actuelles en fonction des contextes cliniques :

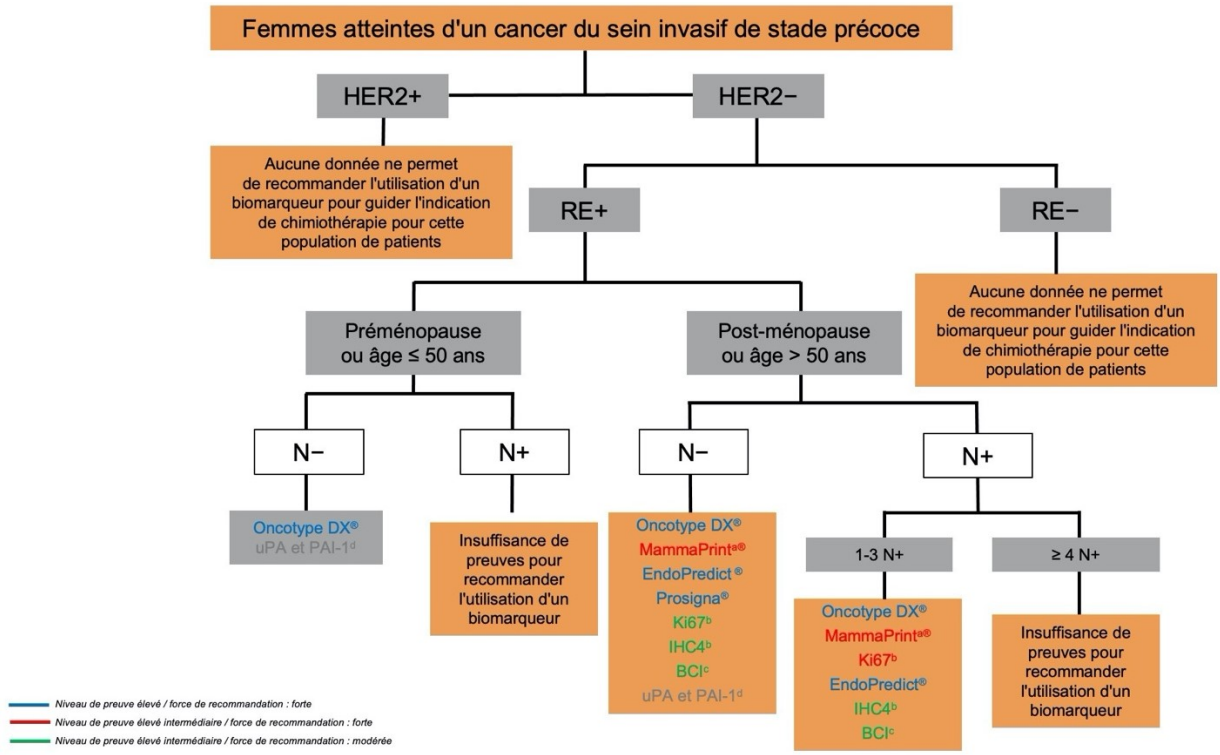


Figure 7. Algorithme d'utilisation des signatures génomiques pour guider les décisions sur l'hormonothérapie adjuvante et la chimiothérapie adjuvante*. a) Uniquement chez les patientes présentant un risque clinique élevé selon la catégorisation MINDACT. b) Uniquement si validé localement et avec d'autres paramètres chez les patientes n'ayant pas accès aux tests génomiques. c) Peut également être proposé aux patientes ayant reçu 5 ans d'hormonothérapie sans preuve de récidive. d) Ce biomarqueur n'est plus utilisé. BC : Breast Cancer Index ; ER : récepteur des œstrogènes ; HER2 : récepteur 2 du facteur de croissance épidermique humain ; IHC4 : immunohistochimie 4 ; NEG : négatif ; PAI-1 : inhibiteur de l'activateur du plasminogène-1 ; POS : positif ; uPA : activateur de l'Urokinase plasminogène ; N0 : absence d'envahissement ganglionnaire ; N+ : envahissement ganglionnaire ; node : ganglion.*

Finalement, ces signatures bien qu'informatives montrent de nombreuses limites :

(1) elle ne sont utiles que pour les cancers du sein RH+/HER2- à risque intermédiaire, permettant souvent une désescalade thérapeutique. Aucun de ces tests n'est recommandé pour guider le traitement des personnes atteintes d'un cancer du sein HER2-positif ou triple-négatif.

(2) les indications d'utilisation de ces signatures varient régulièrement et restent régulées du fait du coût financier pour chaque établissement prescripteur. En effet, ces signatures

ne sont pour le moment pas remboursées et restent à la charge de l'établissement prescripteur dans l'enveloppe RIHN (référentiel des actes innovants hors nomenclature de biologie et d'anatomopathologie).

(3) les signatures moléculaires ne possèdent que peu de gènes en commun parmi les milliers de gènes analysés. Toutes présentent des gènes impliqués dans la prolifération cellulaire. Les études de concordances donnent des résultats discordants. La concordance pour Mammaprint® et Oncotype DX® serait comprise entre 77 et 81%. D'autres études retrouvent des discordances dans plus de 50% des échantillons (κ variant de 0,33 à 0,60) en cas d'analyses par plusieurs signatures dont Oncotype DX®, Prosigna®, Mammaprint® et deux signatures immunohistochimiques IHC4. Cette controverse alimente toujours la question du possible impact du choix du test dans la décision thérapeutique ultime.

3. Introduction à la métabolomique dans le cancer du sein

Comme conclus précédemment, les signatures génomiques visant la classification des tumeurs possèdent quelques limitations. De plus, ces signatures génomiques n'ont pas pour objectif d'identifier des cibles thérapeutiques spécifiques nouvelles pouvant permettre d'accéder à des thérapies ciblées. L'idéal serait de pouvoir déterminer les tumeurs avec critères d'agressivité tumorale et de pouvoir déterminer les voies d'activation finale, d'où l'idée d'utiliser la métabolomique.

a) APPORT DE LA METABOLOMIQUE DANS LA BIOLOGIE DU CANCER DU SEIN

Les premières études ont été réalisées fin des années 90, début des années 2000 (21,152–156) en comparant les profils métaboliques sur tumeurs malignes(21,152,154,155), sur adénopathies métastatiques(156) ou sur cytoponction tumorale diagnostique(153) aux tissus sains ou aux tumeurs bénignes. Par exemple, Sitter *et al.* ont utilisé les rapports des composés de la choline pour la détermination d'éventuelles infiltrations cancéreuses des

tissus adjacents à la tumeur, avec une sensibilité de 82% et une spécificité de 100%(154). Aboud *et al.* ont confirmé une élévation de la choline, une faible taux de glycérophosphocholine et un faible taux de glucose dans les tissus cancéreux du sein par rapport aux tissus sains ou aux tumeurs bénignes(157). D'autres études plus récentes ont été réalisées sur biopsies liquides. Da Cunha *et al.* ont mis en évidence, grâce à une analyse non ciblée sur plasma en ultra-HLPC/MS, une signature métabolique spécifique, composée de 16 métabolites, capable de différencier les patientes porteuses d'un cancer du sein des patients du groupe contrôle avec une performance de 0,9279(158). D'autres études de détection précoce ont également été réalisées sur le plasma(159,160), l'urine (161,162) ou encore la salive(163,164). Ainsi, Murata *et al.* ont montré sur des prélèvements salivaires que les polyamines étaient significativement plus élevées dans la salive des patients atteints d'un cancer du sein par rapport aux patientes du groupe témoin. Par exemple, la spermine avait une AUC à 0,766 (IC95 % : 0,671-0,840, $p < 0,0001$)(163).

Une étude de cohortes sur plasma (3248 plasmas analysés) a même décrit des métabolites prédictifs du risque de développer un cancer du sein, ou du risque de décès. Une corrélation inverse était observée entre l'augmentation des concentrations d'asparagine et de certaines phosphatidylcholines avec le risque de cancer du sein, et une corrélation positive était retrouvée entre les concentrations d'acylcarnitine et le risque de décès(165).

Le métabolisme énergétique a également été exploré. La glycolyse dans les cellules cancéreuses du sein s'est révélée caractérisée par une diminution du taux de glucose(166–168) et une augmentation du taux de lactate(163,169,170), révélateur de l'effet Warburg. La glycolyse, la glycogénolyse, le cycle de l'acide tricarboxylique (cycle de Krebs), la prolifération et les voies d'oxydoréduction, étaient significativement altérés dans le cancer du sein, avec des taux accrus en cas de formes agressives(47,171,172). Par ailleurs, plus de 30 métabolites endogènes ont été identifiés dans les tissus mammaires.

Plusieurs voies qui influencent les niveaux de glutamine(173), de lipides(174), de sérine(175), la traduction des protéines(176) et le métabolisme du cholestérol(177) ont été démontrées être régulées à la hausse dans le cancer du sein(47). Ces changements étaient rattachés à l'activation de voie de transduction sous-jacente, comme l'activation des facteurs inductibles par l'hypoxie (HIF) (178,179), de mTOR(180–182), des récepteurs aux œstrogènes(183), phosphatidyl-inositol-3 kinase (PI3 kinase)(184,185), AMP-activated protein kinase (AMPK)(186,187).

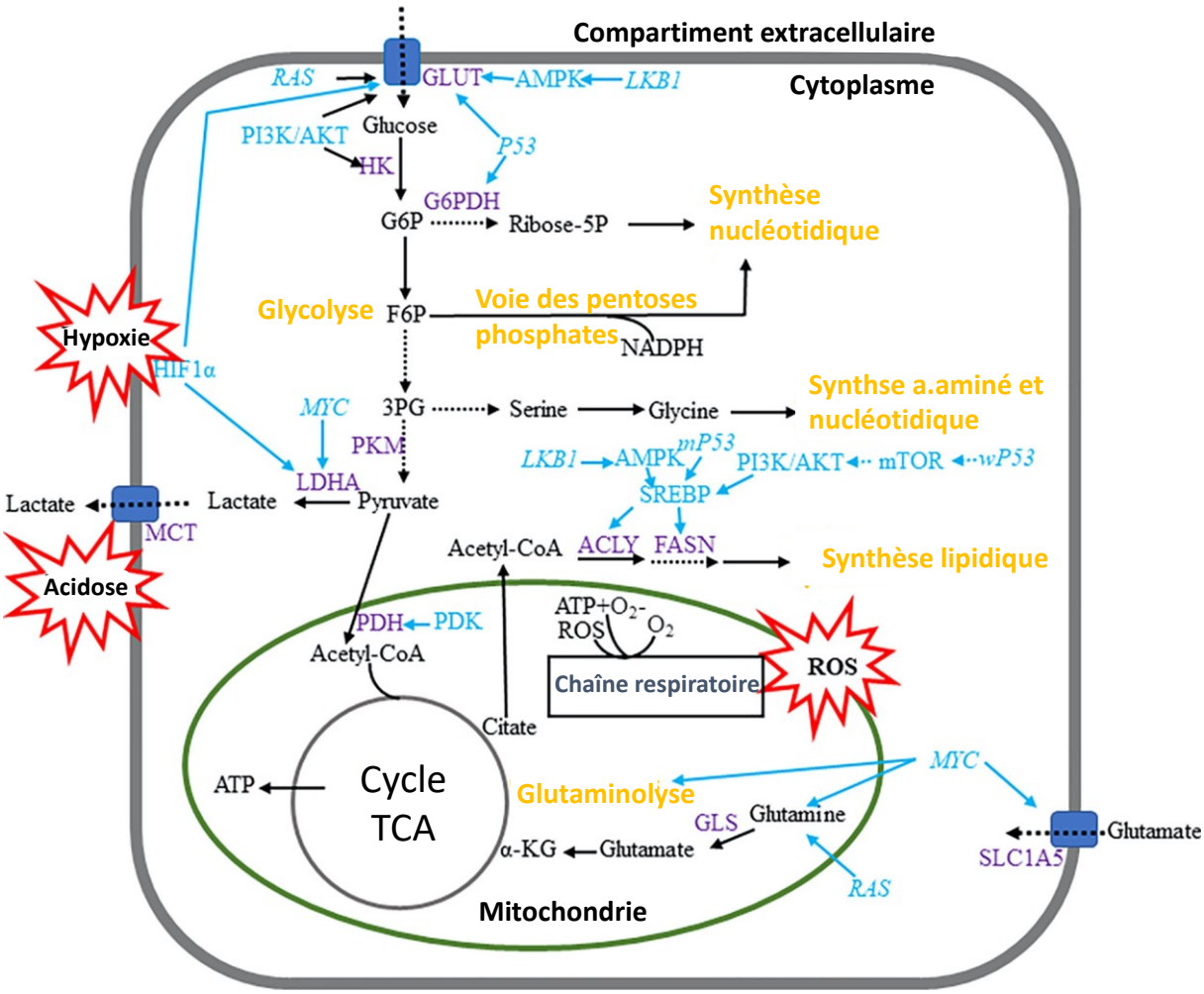


Figure 8. Voies métaboliques modifiées dans les cellules cancéreuses du sein *(adapté de Wang et al.(188)). Cycle TCA: tricarboxylic acid cycle (cycle de Krebs); PI3K: phosphatidylinositol 3-kinase; AKT: serine/threonine kinase; AMPK: AMP-activated protein kinase; LKB1: kinase hépatique B1; NADP: nicotinamide adénine dinucléotide phosphate; NADPH: nicotinamide adénine dinucléotide phosphate hydrogène; ATP: adénosine triphosphate; CO_2 : dioxyde de carbone; O2 : dioxygène; ROS: reactive oxygen species; mTOR: mammalian target of rapamycin; GLS: glutaminase; Acetyl-CoA: acétyl-coenzyme A; SREBP: sterol regulatory element binding protein.*

Les données de la métabolomique ont par la suite permis de distinguer les sous-types moléculaires ER et HER2 en utilisant le ratio glutamate/glutamine et la glycolyse aérobie comme biomarqueurs(189), ces métabolites pouvant être utilisés comme indicateur potentiel de l'agressivité du cancer du sein(190). De même, les métabolites des voies énergétiques telles que la glycolyse, le cycle de KREBS et la bêta-oxydation se sont révélés plus élevés dans les cancers du sein à récepteurs hormonaux négatifs et triple-négatifs par rapport aux cancers du sein à récepteurs hormonaux positifs, ce qui était positivement corrélé à l'agressivité du cancer du sein(171). Des études similaires ont été réalisées sur le plasma. Fan *et al.* ont montré que comparativement au groupe HER2-négatif, le groupe HER2-positif avait une glycolyse aérobie élevée, une gluconéogenèse et une biosynthèse accrue d'acides gras avec un cycle de Krebs réduit. Comparé au groupe RE-négatif, le groupe RE-positif avait une augmentation du métabolisme de l'alanine, de l'aspartate et du glutamate, une diminution du catabolisme des glycérolipides et une augmentation du métabolisme des purines(191). Les métabolites spécifiques de la mutations BRCA1 ont également été recherchées à partir de lignées cellulaires de cancer du sein puis validés sur des prélèvements de plasma de patientes porteuses d'un cancer du sein triple-négatif héréditaire (HBOC syndrome), BRCA1 muté ou non. Les niveaux plasmatiques d'adénine, de N6-méthyladénosine et de 1-méthylguanine ont pu distinguer les patientes atteintes d'un syndrome HBOC TN BRCA1 muté des patientes atteintes d'un syndrome HBOC TN BRCA non muté(192).

Plus récemment, la métabolique a été utilisée comme facteur prédictif et pronostique. Gong *et al.* ont étudié la dérégulation métabolique dans les cancers du sein triple négatif (TNBC) en utilisant une base de données multi-omiques (n = 465). Ils ont classé les échantillons de TNBC en trois sous-types hétérogènes basés sur les voies métaboliques (lipogénique, glycolytique ou mixte), avec des pronostics, des distributions de sous-types moléculaires et des altérations génomiques distincts. En outre, ils ont montré que le sous-type lipogénique était plus sensible aux inhibiteurs métaboliques ciblant la synthèse des

acides gras, tandis que le sous-type glycolytique s'est révélé plus sensible aux inhibiteurs ciblant la glycolyse et avec une meilleure réponse tumorale à l'immunothérapie anti-protéine 1 de la mort cellulaire programmée (anti-PD1) avec inhibition de la lactate déshydrogénase. Leur étude démontre l'hétérogénéité métabolique des cancers et souligne l'importance des thérapies personnalisées ciblant des profils métaboliques spécifiques(193). Autre exemple, le ciblage de GLUT-1, un transporteur de glucose présent à des niveaux élevés dans le cancer du sein, a permis l'inhibition sélective de la croissance des lignées cellulaires de cancer du sein(194). Par ailleurs, les métabolites issus du métabolisme secondaire des acides biliaires, de la dégradation des acides aminés, de la production d'acides gras à chaîne courte et des hormones déconjuguées ont pu être mesurés pour prédire la résistance au tamoxifène, l'apoptose induite par les hormones, l'agressivité du cancer et l'inhibition de l'histone désacétylase (HDAC)(195–197). Au niveau pronostique, des taux de survie à 5 ans plus faibles ont été associés à des niveaux plus élevés de lactate(198).

Pour finir, l'arrivée des biopsies liquides, avec l'analyse des fluides (urine, salive, plasma), a ouvert le champ des applications avec des possibilités de détection de progression métastatique(199), et de suivi en cours de traitement(200,201).

b) FOCUS SUR LE METABOLISME DES ACIDES AMINES DANS LE CANCER DU SEIN

Plusieurs acides aminés ont un intérêt prouvé dans le cancer du sein. Des niveaux élevés de cystéine sont associés à des dommages oxydatifs et à une surproduction de radicaux libres qui entraînent des mutations génétiques(202). Les changements métaboliques majeures des niveaux de choline et de proline sont connus pour être caractéristiques du cancer du sein métastatique(203). L'altération des taux d'arginine et d'asparagine est également corrélée au cancer du sein. Il a été démontré que les cancers du sein sont fortement dépendants de la L-arginine(204). Une supplémentation en L-arginine renforcerait les réponses immunitaires innées et adaptatives et inhiberait la croissance

du cancer du sein(205). Une étude menée chez des femmes préménopausées a révélé qu'une augmentation des taux plasmatiques d'arginine entraînait une diminution des taux plasmatiques d'insuline, de facteur de croissance 1 et d'œstradiol(206). La diminution de la biodisponibilité de l'asparagine, soit en réduisant l'asparagine alimentaire, soit en inhibant l'asparagine synthétase, a réduit les métastases du cancer du sein(207).

Parmi les dérivés polyamines (synthétisées à partir de deux acides aminés : la L-méthionine et la L-ornithine), les N1,N12-diacétylspermines ont récemment attiré beaucoup d'attention en oncologie, et les diacétylspermines urinaires ont été décrites comme des marqueurs tumoraux très sensibles dans de nombreux cancers, notamment le cancer du sein(208–211). Des études précédentes ont montré que des niveaux élevés de polyamines acétylées sont trouvés dans le cancer du sein en association avec une augmentation simultanée de l'activité de la spermine et de la spermine N1 acétyltransférase (SAT1), et une diminution de l'activité de la polyamine oxydase(212).

Une étude fonctionnelle a examiné les effets de la spermine sur le récepteur d'œstrogène (ER) (213). Les résultats obtenus suggèrent que la spermine joue un rôle important dans la régulation de la liaison du ligand au RE et dans l'activation des gènes, et donc également dans la résistance aux hormones. Le diacétylspermine (DiAcSpm) a été étudié par Fahrmann *et al.* chez des patientes atteintes d'un cancer du sein triple négatif (TNBC)(214). Les échantillons de sérum de patientes TNBC ont montré un niveau de DiAcSpm plus élevé que les échantillons de patientes non TNBC et de volontaires sains.

Dans une cohorte prospective, les auteurs ont observé que les taux sériques de DiAcSpm avaient significativement augmenté chez les patientes présentant une récidive précoce (<1 an). Les taux sériques plus élevés de DiAcSpm étaient également associés à une survie sans métastase à distance et à une survie globale à 5 ans plus faible. En outre, Fahrmann

et al. ont fourni des preuves que des niveaux élevés de DiAcSpm dans le plasma sont associés à plusieurs sous-types de tumeurs TNBC caractérisés par un faible infiltrat immunitaire, des signatures génétiques liées à l'immunité réduites, une faible survie globale et des métastases.

Concernant la voie du tryptophane, Tang *et al.* ont montré que les niveaux de kynurénine étaient significativement plus élevés dans les tumeurs ER-négatives par rapport aux tumeurs ER-positives(215). De plus, l'altération des profils d'expression de la sérotonine et des récepteurs de la sérotonine a été rattachée à une dérégulation de l'homéostasie épithéliale, qui a été associée aux événements initiaux du développement du cancer du sein, à la cancérogénèse et à la progression tumorale(216–219).

4. Introduction au travail de thèse

a) OBJECTIF DE THESE

L'objectif initial de cette thèse était de déterminer des marqueurs métabolomiques qui soient prédictifs et/ ou pronostiques dans le cancer du sein non-métastatique.

Pour cela, les marqueurs métabolomiques devaient remplir trois niveaux de validité clinique (220,221):

●Validité analytique : la validité analytique fait référence aux aspects techniques d'un test, notamment la précision, la reproductibilité et la fiabilité.

●Validité clinique : la validité clinique est la capacité d'un facteur à séparer une population d'intérêt en deux sous-groupes, ou plus, qui diffèrent en termes de résultats biologiques ou cliniques. Cependant, la validité clinique n'implique pas qu'un facteur soit utilisé dans les soins directs aux patients.

● Utilité clinique : l'utilité clinique implique qu'un facteur soit utile pour les soins aux patients, qu'il ait une utilité clinique basée sur la preuve qu'il a un impact sur les résultats lorsqu'il est comparé à une utilisation clinique sans lui. L'utilité clinique nécessite l'hypothèse d'une utilisation ou d'un contexte spécifique dans lequel le facteur est pertinent.

Actuellement, aucun marqueur métabolomique n'est validé d'utilité clinique. L'objectif premier était donc de confirmer la validité analytique et clinique d'un marqueur métabolomique dans le cancer du sein non-métastatique avant de pouvoir envisager une étude d'utilité clinique.

b) ÉTAPE TRANSLATIONNELLE

Pour cela, nous avions besoin d'avoir accès à une cinquantaine de tumeurs congelées provenant de patientes prises en charge pour un cancer du sein localisé ou localement avancé, de toute histologie confondu, naïves de tout traitement néoadjuvant, avec indication potentielle de chimiothérapie adjuvante (présence de critères d'agressivité).

La première étape de ce projet a donc été la création d'un réseau de collaboration entre oncologues médicaux, médecins nucléaires, statisticiens et anatomopathologistes pour pouvoir identifier, au sein de la tumorothèque du Centre Antoine Lacassagne, les prélèvements permettant de répondre à la question posée.

Les étapes clés ont été :

- Recherche de financements et soutien institutionnel du Centre Antoine Lacassagne
- Présentation du projet aux différents acteurs
- Rédaction d'une demande de mise à disposition des prélèvements
- Recherche des tumeurs correspondant aux critères d'inclusion principaux
- Stabilisation de la base de données cliniques
- Déclarations réglementaires CNIL
- Vérification de la qualité des prélèvements avant destockage
- Destockage des tumeurs par le Centre de Recherche Biologique
- Acheminement des tumeurs au laboratoire de métabolomique TIRO
- Préparation des échantillons
- Analyse par LC-MS/MS
- Étape pré-analytique : filtrage, validation des données de métabolomiques
- Étape analytique : biologique et statistique

Dans un second temps, le Centre Georges-François Leclerc a accepté de participer au projet en mettant à disposition les prélèvements de biopsies diagnostiques de patientes suivies pour un cancer du sein localement avancé, avec une indication de chimiothérapie

néoadjuvant avant chirurgie. Cette deuxième cohorte était composée de tumeurs plus avancées, avec un profil majoritairement plus agressif, permettant de constituer une cohorte de validation externe. Les prélèvements ont été acheminés puis traités en LC-MS/MS indépendamment de ceux de la cohorte niçoise. Afin de pouvoir comparer les bases de données, seules les métabolites communs ont été gardés par fusion des bases. Un travail d'identification a permis d'enrichir la base de métabolites d'intérêt et des étapes de filtrage et de suppression des doublons et de fusion des lignes dédoublées ont été ajoutées à la procédure.

Ces étapes seront développées par la suite dans les parties méthodologie générale et spécifique.

c) ÉTAPE ANALYTIQUE

Avec les données métabolomiques obtenues, nous avons réalisé une analyse non supervisée de faisabilité et avons comparé la performance de la classification de 5 méthodes de machine learning non supervisées (sans question posée) aux données de survie simulée par PREDICT tool (Article 1) et aux données de vie réelle (Article 2). Cette première étape avait pour but de me familiariser avec la métabolomique, les méthodes d'analyse et de vérifier la précision, la reproductibilité et la fiabilité de la métabolomique pour ce qui est de prédire les données de survie. Ces analyses n'ont été réalisées que sur la cohorte niçoise.

Nous avons également réalisé en parallèle une analyse supervisée comparant des tumeurs invasives de haut grade et des tumeurs de grade bas/intermédiaire en utilisant une cohorte d'entrainement (Nice) et une cohorte de validation (Dijon) (Article 3) pour déterminer les métabolites d'agressivité tumorale. Cette deuxième partie avait pour but d'élaborer une signature métabolomique validée des cancers du sein non-métastatiques de haut grade pour ensuite analyser les métabolites d'intérêt, afin de trouver de nouvelles cibles thérapeutiques.

VI. Méthodologie

La méthode globale d'analyse métabolomique optimale dépend du matériel source et de l'application. Diverses technologies et méthodes de séparation et de détection peuvent être utilisées pour acquérir des données brutes, qui constituent ensuite le point de départ de l'analyse informatique.

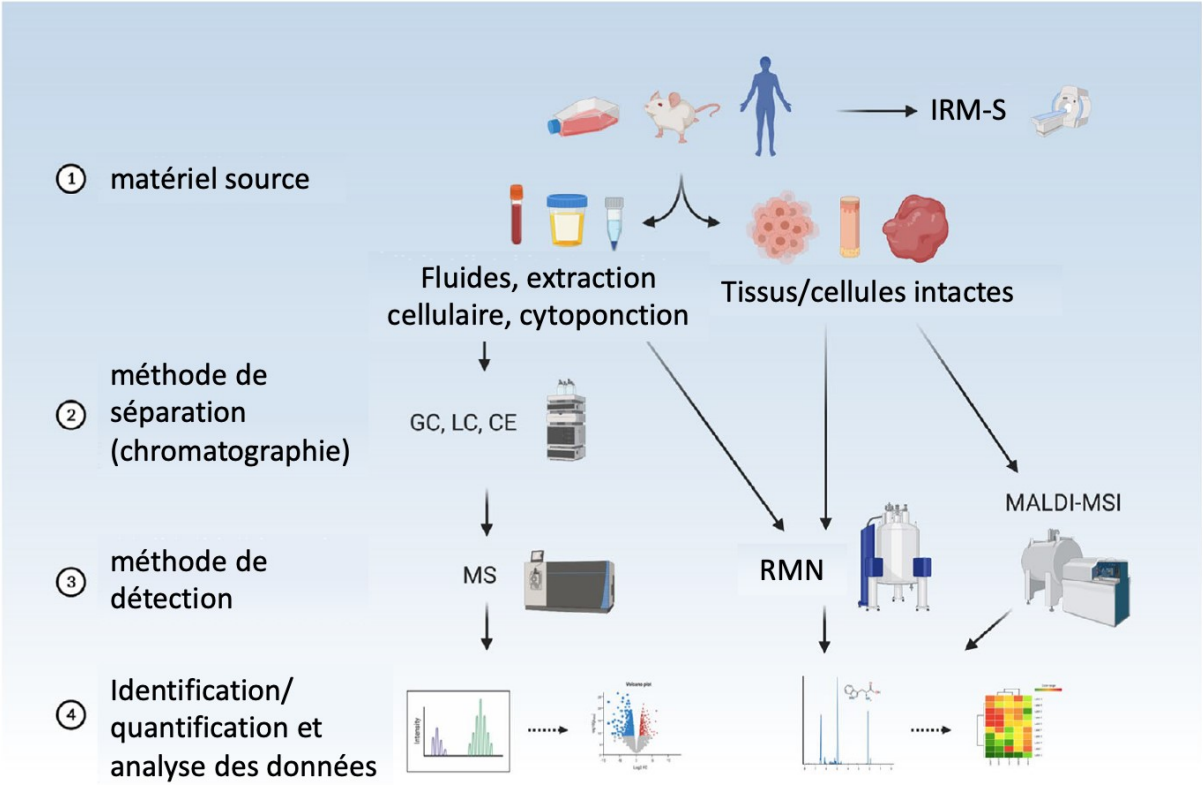


Figure 9. Les principales étapes d'une analyse métabolomique *(d'après Schmidt et al.(7)). IRM-S : spectroscopie par résonance magnétique; MS : mass spectrometry : spectrométrie de masse; GC : gas chromatography: chromatographie en phase gazeuse; LC: liquid chromatography: chromatographie en phase liquide; CE: capillary electrophoresis : électrophorèse capillaire; RMN: résonance magnétique nucléaire; MALDI-MSI: Matrix-assisted laser desorption/ionization mass spectrometry imaging: imagerie par spectrométrie de masse de type MALDI(désorption/ionisation laser assistée par matrice).*

Après une introduction générale sur la LC-MS, la méthodologie spécifique pré-analytique de cette thèse sera développée.

1. Méthodologie générale LC-MS

a) ENTRE RMN ET LC-MS

La métabolomique se base sur l'obtention d'empreintes métaboliques par différentes méthodologies analytiques. Pour être utile dans un projet de métabolomique, une méthode analytique doit présenter une haute résolution permettant l'identification et l'analyse qualitative et/ou quantitative d'un grand nombre de métabolites. Elle doit également être rapide pour permettre l'analyse de nombreux échantillons nécessaires pour le traitement statistique, ce qui la qualifie comme méthode d'analyses à haut débit. Les deux plateformes analytiques les plus couramment utilisées sont : la spectroscopie en résonance magnétique nucléaire (RMN ou NMR en anglais) et la spectrométrie de masse (MS).

La spectroscopie RMN donne des informations structurales sur les métabolites en utilisant les propriétés magnétiques de certains noyaux atomiques, comme l'hydrogène (1H) et le carbone (13C). La RMN est une technique non-destructive, hautement reproductible et quantitative. La préparation des échantillons est simple et une large gamme de métabolites peut être analysée. Cependant, la RMN est limitée par sa faible sensibilité, ce qui restreint son application à la mesure des métabolites les plus abondants dans un échantillon(222,223).

La spectrométrie de masse (MS) est une méthode analytique qui sépare les molécules chargées (ions) en fonction de leur rapport masse sur charge (m/z) permettant ainsi la détection et l'identification des métabolites. Les spectromètres de masse sont constitués de quatre parties fondamentales :

1) un système de manipulation destiné à introduire l'échantillon dans le dispositif. L'échantillon est injecté soit directement dans le spectromètre de masse (infusion

directe), soit via un système de séparation, par exemple une chromatographie en phase liquide (LC)* ou en phase gazeuse (GC)*.

2) une source d'ions. Les molécules à analyser contenues dans l'échantillon sont ionisées par une source d'ionisation.

3) un analyseur qui sépare les particules en fonction de leur rapport m/z. Elles sont ensuite fragmentées dans une cellule de collision pour permettre l'identification de leur structure chimique (cf. paragraphe MS1 et MS2)(224).

4) un détecteur dans lequel les composants des ions séparés sont récupérés et caractérisés.

* LC/GC : La séparation des métabolites repose sur la différence d'affinité de ces composés pour la phase mobile et pour la phase stationnaire. Plus la molécule a d'affinité pour la phase stationnaire, moins elle est entraînée par la phase mobile, et donc plus elle est retenue sur la colonne. La chromatographie en phase liquide (LC) est une méthode de séparation qui utilise une phase mobile liquide pour entraîner les composants du mélange à travers une phase stationnaire solide ou liquide. Les composants du mélange se séparent en fonction de leur affinité pour la phase stationnaire et de leur solubilité dans la phase mobile liquide. La chromatographie en phase gazeuse (GC), quant à elle, utilise une phase mobile gazeuse pour entraîner les composants d'un mélange à travers une phase stationnaire solide ou liquide. Les composants se séparent en fonction de leur affinité pour la phase stationnaire et de leur volatilité dans la phase mobile gazeuse. Si les métabolites d'un échantillon ont des coefficients de partage différents, alors leurs durées de parcours dans la colonne seront différentes. Ainsi, les métabolites se séparent puis sortent de la colonne les uns après les autres. La durée entre le temps d'injection et le temps de sortie de colonne d'un métabolite est son « temps de rétention »(225,226).

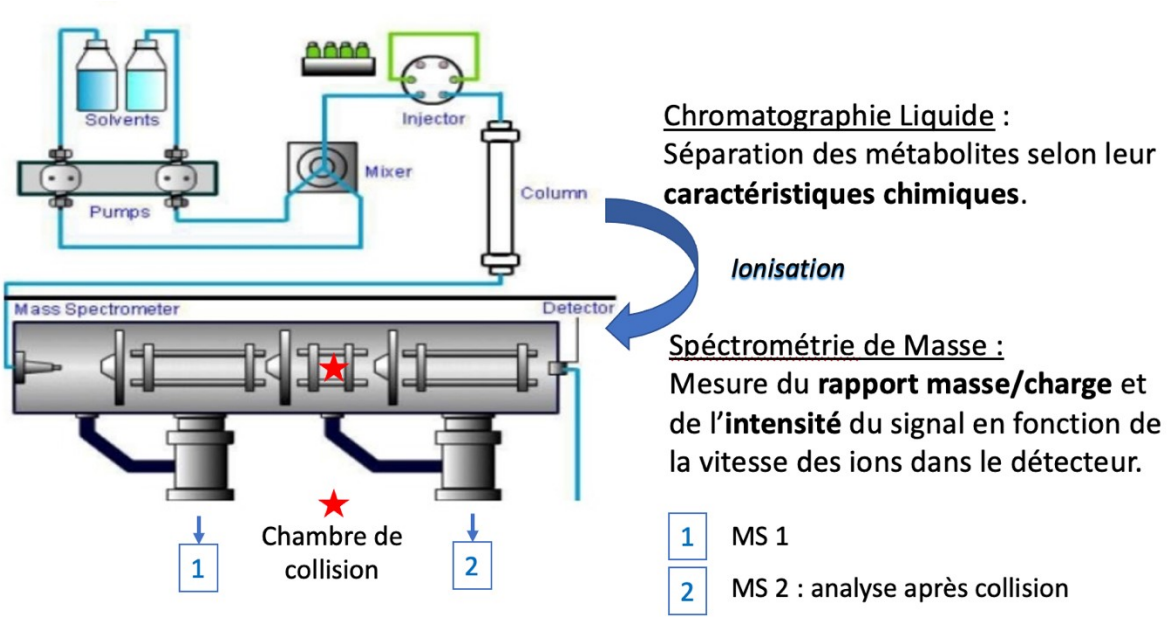


Figure 10. Principe de LC-MS/MS*. MS 1 : spectre de masse initial ; MS2 : MS/MS spectre de masse après collision.*

L'analyse en LC-MS/MS génère ensuite des données spectrales (MS1/MS2) qui sont traitées par différents outils informatiques et statistiques pour donner la liste des métabolites ayant significativement variés. Ces métabolites sont ensuite étudiés dans les voies métaboliques pour faciliter l'interprétation biologique.

Par rapport à la RMN, la MS est une technique d'une grande sensibilité permettant la détection de variations de métabolites présents en quantités moins importantes. En revanche, la MS est parfois confrontée à des problèmes de reproductibilité et la quantification reste relative dans le cas de l'approche non-ciblée(227,228). La MS est une technique destructive, nécessitant une préparation de l'échantillon.

La spectroscopie RMN, la GC-MS et la LC-MS sont 3 approches qui présentent chacune des avantages et des limitations (229–231). Ces trois techniques, les plus utilisées aujourd'hui en métabolomique, sont souvent complémentaires et de nombreuses études ont été réalisées en les combinant dans le but d'élargir la couverture des métabolites détectables et quantifiables(232,233).

b) DONNEES OBTENUES PAR LC-MS/MS : MS1 ET MS2

Les métabolites à analyser sont séparés par chromatographie en phase liquide en fonction de leur polarité. Tout au long de la séparation chromatographique, ces métabolites sont détectés par spectrométrie de masse, produisant des spectres de masse (ou spectres MS1). L'abondance relative des métabolites, représentée par l'intensité des courants ioniques des métabolites détectés à un certain temps de rétention (RT), est ensuite visualisée sur le chromatogramme.

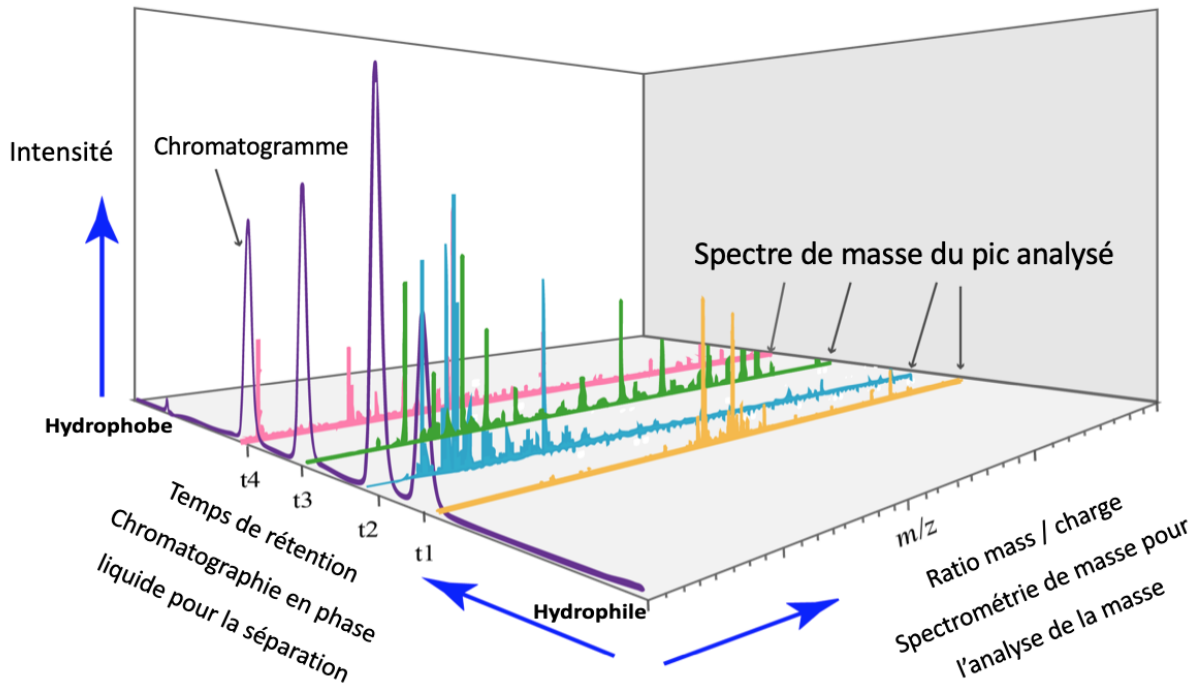


Figure 11. Exemple de chromatogramme*. m/z : rapport mass/charge ; t : temps de rétention.*

Après cela, chaque métabolite est fragmenté dans la chambre de collision par rupture des liaisons chimiques. Des fragments plus petits sont formés et sont également détectés, ce qui permet d'obtenir un spectre MS/MS (ou spectre MS2) pour chaque fragmentation.

Ces spectres MS2 permettent d'identifier la structure de la molécule principale en déterminant les fragments spécifiques générés. En effet, plusieurs métabolites peuvent avoir le même rapport m/z et le même temps de rétention. Cependant, la fragmentation sera différente selon les caractéristiques structurelles propres de la molécule chimique (cf. Figure 12 et 13).

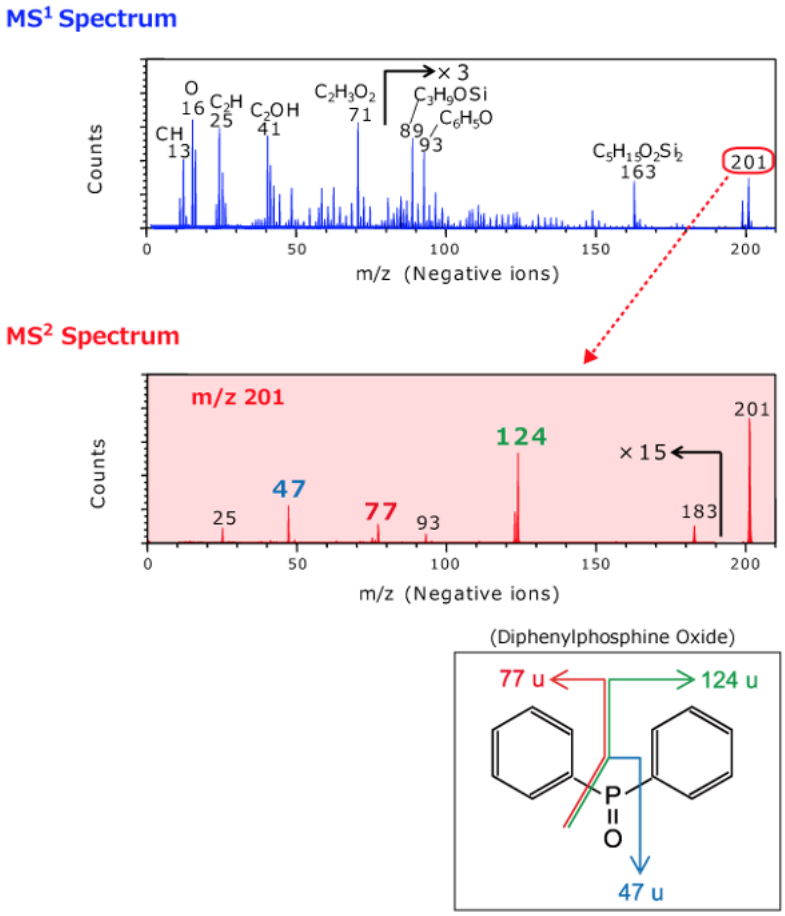


Figure 12. Principe de fragmentation pour analyse des spectres MS et MS/MS (MS2) *(https://www.ulvac-phi.com/en/surface-analysis/topics/msms/). Counts : intensité de signal; m/z : rapport masse/charge; MS¹ : spectre de masse initial ; MS² : spectre de masse après collision pour m/z 201.*

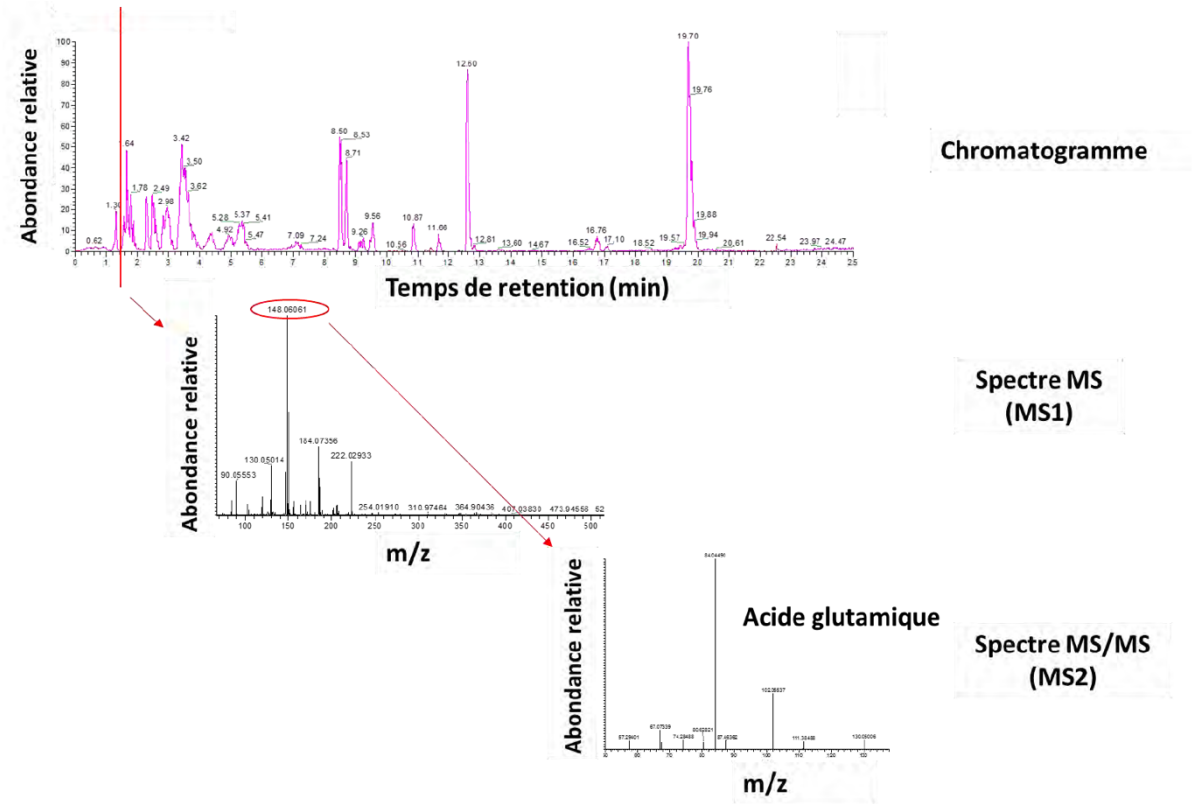


Figure 13. Données spectrales générées par l'analyse en LC-MS*. m/z : rapport masse/charge; MS¹ : spectre de masse initial ; MS² : spectre de masse après collision pour m/z 148,06.*

c) ANALYSES CIBLEES OU NON CIBLEES

Les approches métabolomiques sont classiquement classées en plusieurs niveaux d'études : l'analyse ciblée et l'analyse non ciblée.

L'analyse ciblée est centrée sur un petit nombre de métabolites, soit par le biais d'un profilage métabolique, basé sur l'analyse de tous les composés appartenant à une voie ou à une famille chimique donnée, soit par l'établissement d'une empreinte métabolique, visant à la comparaison de spectres. Ici, les métabolites sont prédéfinis, typiquement associés à des voies métaboliques d'intérêt. La méthode d'analyse est optimisée par rapport aux métabolites définis et des gammes de calibration peuvent être établies préalablement pour chacun des métabolites(234). Cette quantification est alors absolue

et a pour objectif de rechercher les variations des métabolites connus. Cette approche ciblée possède une sensibilité considérablement plus élevée mais plus restreinte, car sur une série de métabolites préalablement connus.

L'analyse non ciblée, aussi nommée l'approche métabolomique (ou la métabonomique) a elle pour ambition l'identification et la quantification non biaisée d'un maximum de métabolites présents dans un échantillon biologique prélevé dans des conditions données, sans a priori(235,236). Ici, les métabolites ne sont pas présélectionnés. Cette quantification est relative et a pour objectif de rechercher de nouveaux biomarqueurs et de nouveaux mécanismes d'action. L'approche non-ciblée fournit une vue globale sur l'état métabolique de l'échantillon permettant la découverte de nouvelles perturbations métaboliques associées à une maladie, un médicament ou un changement environnemental. En revanche, la sensibilité est plus faible par rapport à l'approche ciblée et l'identification des métabolites nécessite une validation supplémentaire. L'information générée est complexe et nécessite un traitement de données approprié.

2. Méthodologie spécifique

a) COHORTES

La cohorte niçoise était constituée de 52 patientes traitées au Centre Antoine Lacassagne, Centre de lutte contre le cancer de Nice, entre mars 2013 et septembre 2016 pour un cancer du sein invasif de stade clinique I à IIIB prouvé par biopsie, avec une indication de chimiothérapie adjuvant après chirurgie.

La cohorte dijonnaise était constituée de 49 patientes traitées au Centre Georges-François Leclerc, Centre de lutte contre le cancer de Dijon, entre février 2007 et juillet 2012 pour

un cancer du sein invasif prouvé par biopsie de stade clinique IIA à IV, avec une indication de chimiothérapie néoadjuvant avant chirurgie.

Toutes les patientes ont été traitées conformément aux recommandations de bonnes pratiques, avec une chimiothérapie séquentielle comprenant des anthracyclines (épirubicine et cyclophosphamide) et des taxanes avant ou après la chirurgie et la radiothérapie. Le statut HER2-positif était défini comme IHC3+ ou IHC2+/FISH+. Les patients présentant des tumeurs HER2-positives ont été traités par trastuzumab et taxanes simultanément puis par trastuzumab seul, pour une durée totale de trastuzumab de 1 an. Les patientes présentant un cancer hormonodépendant ont ensuite été traitées par hormonothérapie avec du tamoxifène (+/- agoniste de la luteinizing hormone-releasing hormone (anti-LHRH)) ou un inhibiteur de l'aromatase, en fonction du statut ménopausique.

b) RECUEIL DES DONNEES

Les données cliniques, histologiques, radiologiques et thérapeutiques ont été extraites rétrospectivement des dossiers numériques ou collectées manuellement.

c) COLLECTE ET PREPARATION DES ECHANTILLONS

Pour la cohorte niçoise, les échantillons analysés provenaient de la pièce d'exérèse obtenue lors de la prise en charge chirurgicale première. Pour la cohorte dijonnaise, les biopsies diagnostiques prélevées avant la chimiothérapie néo-adjuvante étaient utilisées. Dans ce travail de thèse, seuls des tissus congelés ont été utilisés. Les analyses ont été effectuées séparément sur chacune des deux cohortes.

Tous les échantillons étaient rapidement congelés et transférés dans les biobanques de des établissements respectifs où ils étaient conservés à -80°C jusqu'à leur analyse. Les échantillons de Dijon ont été transportés à Nice à -80°C avant l'analyse métabolomique.

Tous les échantillons ont été préparés et analysés sur la plateforme de métabolomique du laboratoire TIRO.

L'objectif de l'étape de préparation des échantillons était de récupérer les métabolites à partir des tissus congelés, tout en concentrant l'échantillon et en éliminant les impuretés telles que les protéines et les sels qui interfèrent avec l'analyse par spectrométrie de masse. En métabolomique non-ciblée, la méthode idéale de préparation des échantillons devrait être non-sélective, simple, rapide avec un minimum d'étapes et reproductible.

La préparation des tissus a été réalisée en trois étapes :

* L'EXTRACTION : précipitation par solvant organique : transfert des échantillons dans des tubes Eppendorf de 1,5 ml contenant 1 ml de méthanol, broyage mécanique au piston pellet puis stockage à -20°C pendant la nuit.
* La SÉPARATION des métabolites, des protéines précipités et autres impuretés par centrifugation à 13 000 rpm pendant 15 minutes à 0°C.
* La CONCENTRATION : les surnageants étaient transférés dans de nouveaux tubes et placés dans un Speed Vac jusqu'à évaporation complète du liquide. Les échantillons étaient ensuite stockés à -80°C jusqu'aux analyses en LC-MS. Avant l'analyse LC-MS, ils étaient resuspendus dans 100µL d'un mélange 50% acétonitrile et 50% eau.

En métabolomique non-ciblée, la précipitation par solvant organique est souvent privilégiée en raison de sa simplicité et de sa faible sélectivité. L'ajout direct d'un solvant organique permet d'extraire les métabolites et de précipiter les protéines en même temps. La métabolomique classique cible principalement les métabolites polaires, tels que les sucres, les acides aminés, les nucléotides et certains lipides relativement polaires tels que les phospholipides et les acides gras. Le méthanol et l'acétonitrile sont les solvants les plus couramment utilisés en raison de leur efficacité pour la précipitation des protéines, de leur capacité à couvrir une gamme large de classes de métabolites et de leur facilité de préparation.

d) ANALYSE LC-MS/MS

Dans cette thèse, la méthode utilisée a été la chromatographie en phase liquide couplée à la spectrométrie de masse (LC-MS). Cette méthode a été choisie car elle permet la détection d'une large gamme de métabolites et une forte performance analytique en résolution, en sensibilité, en rapidité et en robustesse(237).

L'analyse par chromatographie liquide a été réalisée à l'aide d'un système high performance liquid chromatography (HPLC) DIONEX Ultimate 3000 (Thermo Fisher Scientific). 10 µL de chaque échantillon ont été injectés sur une colonne Synergi 4 µm Hydro-RP 80 Å, 250 x 3,0 mm (Phenomenex, Le Pecq, France). Les phases mobiles étaient composées d'acide formique à 0,1% (Thermo Fisher Scientific) dans l'eau (A) et d'acide formique à 0,1% dans l'acétonitrile (B). Le gradient était réglé comme suit avec un débit de 0,9 ml/min : phase B à 0 % de 0 à 5 min, B à 0 - 95 % de 5 à 21 min, maintien à 95 % de B jusqu'à 21,5 min, B à 95 - 0 % de 21,5 à 22 min, maintien à 0 % de B jusqu'à 25 min pour l'équilibrage de la colonne.

L'analyse par spectrométrie de masse a été effectuée sur un spectromètre de masse Q Exactive Plus Orbitrap (Thermo Scientific) avec une source d'ionisation par électrospray chauffée, HESI II, fonctionnant en mode positif et négatif. Le MS full-scan à haute résolution et à masse précise et les 5 spectres MS2 les plus élevés ont été collectés en fonction des données à un pouvoir de résolution de 70 000 et 35 000 à m/z 400, respectivement.

3. Étape pré-analytique spécifique

Le traitement des données brutes est une étape cruciale dans l'analyse en métabolomique. Les pics des métabolites contenus dans les spectres obtenus par analyse LC-MS doivent être détectés, filtrés, alignés, quantifiés, identifiés et normalisés pour pouvoir être utilisés pour les analyses statistiques en aval. Le résultat final doit contenir, pour chaque

métabolite identifié, la masse, le temps de rétention et l'intensité dans chacun des échantillons.

De nombreux logiciels existent pour le traitement des données métabolomiques, dont certains sont mis à disposition par les vendeurs d'instruments (par exemple, Compound Discoverer de Thermo) et d'autres sont en accès libre, dit « open source », par exemple, MZmine. Une liste exhaustive est donnée dans (238).

Dans cette thèse, le principal logiciel utilisé a été MZmine, dont la procédure est résumée sur la figure 12 ci-dessous en 3 étapes :

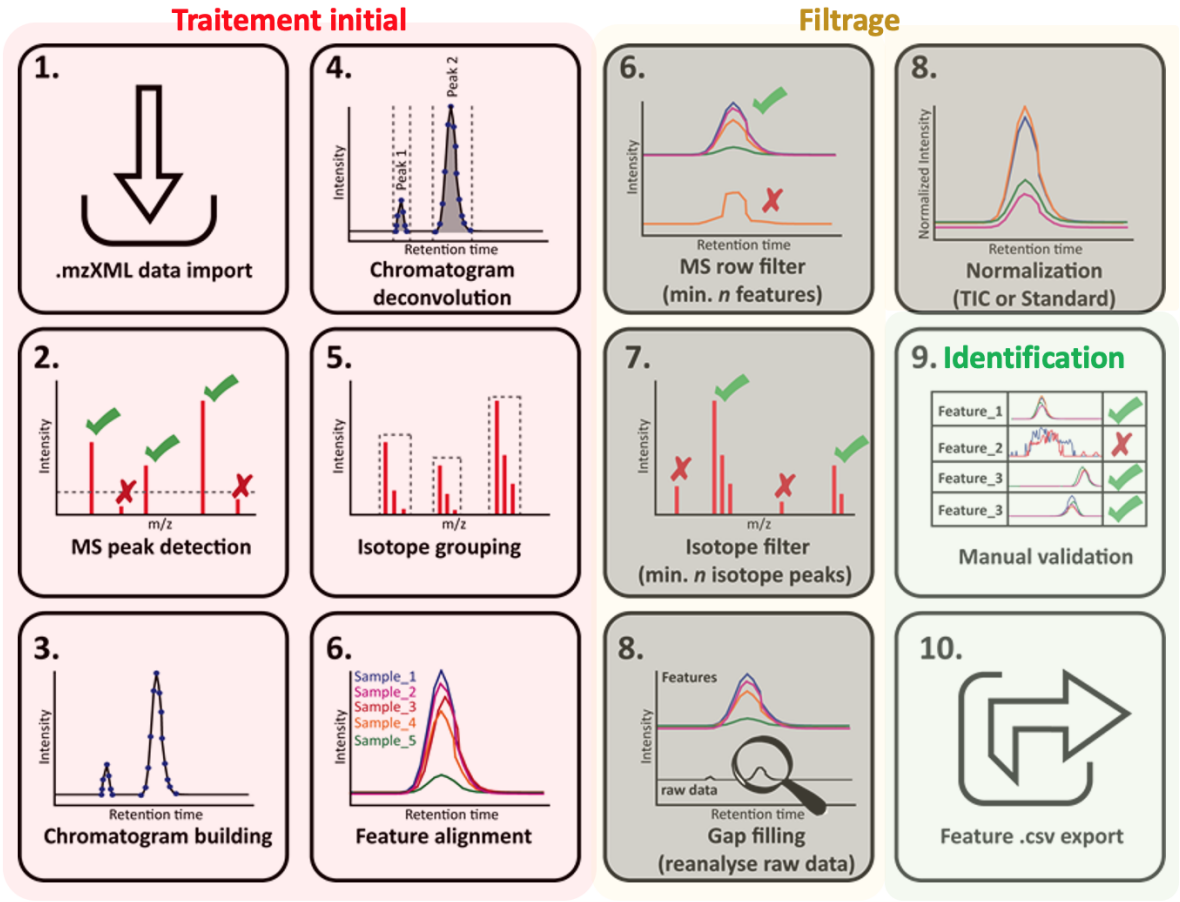


Figure 14. Procédure MZmine *(adapté de GNPS documentation https://gnps.ucsd.edu/). m/z : rapport masse/charge ; MS : spectre de masse ; n : nombre ; TIC : total ion current/count.*

a) TRAITEMENT DES DONNEES PAR MZMINE

Plusieurs étapes sont nécessaires :

1. *Import des données brutes au format .mzXML* : les données brutes sont d'abord converties d'un format instrument-dépendant à un format libre(239).

2. *Détection des MS peak* : l'étape de la détection des pics a pour but de distinguer les signaux correspondant aux vrais métabolites par rapport à de faux positifs(240). Les critères de sélection peuvent être posés sur la hauteur minimale du pic, le ratio signal sur bruit, la forme du pic ou encore le profil isotopique. Une fois le pic détecté, l'aire sous le pic et la hauteur maximale du pic sont calculées et pourront servir pour la quantification relative des métabolites par la suite.

3. *Construction des chromatogrammes*

4. *Déconvolution de chaque chromatogramme* : les données LC-MS contiennent un bruit de fond causé par des molécules contenues dans le solvant et celui causé par le détecteur lui-même. L'étape de filtrage a pour but de supprimer ou du moins réduire ces bruits de fond tout en préservant les pics. Plusieurs algorithmes sont disponibles : « baseline cut-off », « noise amplitude », « Savitsky-Golay » et « local minimum Search ».

5. *Regroupement des isotopes*

6. *Alignement :* pour comparer la variation d'un métabolite entre les différents échantillons, les pics détectés dans chacun des échantillons doivent être alignés au niveau du rapport m/z et du temps de rétention. L'alignement des pics peut être effectué sur le chromatogramme complet (le cas de Compound Discoverer) ou sur les chromatogrammes extraits (le cas de MZmine).

Lors de cette thèse, les données brutes obtenues en modes d'ionisation positive et négative ont été analysées séparément avec MZmine (Version 2.38)(241,242). Des chromatogrammes individuels ont été construits pour chaque masse avec un seuil de 10^5. Un algorithme de recherche de « local minimum Search » a été utilisé pour sélectionner

les pics validés. Les pics ont ensuite été alignés par l'algorithme RANSAC (random sample consensus) avec une tolérance de 10 ppm en m/z et un temps de rétention de 1 min.

b) FILTRAGE DES DONNEES

Plusieurs méthodes de filtrage sont communément utilisées :

1. *MS row filter qui ne garde que les pics retrouvés dans un nombre d'échantillons suffisant et prédéfini.*

2. *Filtrage des isotopes* pour ne garder que les « vrais » pics. Seules les molécules accompagnées de leurs isotopes sont gardées afin d'éliminer les pics artefactuels sans réelles structures biochimiques associées.

3. *Gap filling* pour retrouver les pics « manquants », non détectés ou perdus dans le processus. Certains pics peuvent ne pas être alignés dans tous les échantillons. Ces valeurs manquantes peuvent être comblées en cherchant une deuxième fois l'existence potentielle d'un pic à m/z et temps de rétention présumés (241).

4. *Normalisation TIC ou standard :* à la suite du traitement des données métabolomiques, une matrice de données, contenant les métabolites en colonne et les échantillons en ligne, est générée. L'étape de normalisation a pour but d'enlever les biais systématiques dans l'intensité des ions entre les mesures tout en gardant les variations biologiques intéressantes. La normalisation peut être effectuée entre les différents échantillons (autrement dit normalisation par ligne) ou être effectuée entre les différents métabolites (normalisation par colonne)(240). Il n'existe pas de méthode standardisée ou unifiée de normalisation en métabolomique non-ciblée. Le choix de la méthode de normalisation doit être adapté à la nature de l'échantillon (urine, sang, tissu, culture cellulaire, etc.) et doit être considéré dès l'étape de la collection et de l'extraction des échantillons(243).

Les métabolites correspondant aux m/z ont ensuite été prédits à l'aide de la Human Metabolome DataBase (HMDB, version 3.0) en recherchant les formes ioniques M+H+ et M-H+ en modes positif et négatif, respectivement, avec une tolérance de masse de 15 ppm.

Seuls les pics prédits ont été inclus dans l'analyse finale. Une normalisation linéaire a été effectuée en utilisant l'intensité moyenne de chaque échantillon comme facteur de normalisation. Seuls les métabolites ne présentant aucune valeur nulle après le prétraitement ont été sélectionnés pour l'analyse finale. Si un métabolite était détecté à la fois dans des modes positifs et négatifs, seul le mode présentant l'intensité moyenne la plus élevée a été pris en compte. Enfin, une fonction de filtrage a été appliquée avant l'analyse statistique en sélectionnant uniquement les métabolites présentant l'intensité moyenne la plus élevée. Cette étape nous a permis d'éliminer les métabolites qui pouvaient être considérés comme des signaux de fond ou pour lesquels la quantification n'était pas assez robuste.

c) IDENTIFICATION PAR VERIFICATION DE LA MS2

Quelques soient les méthodes d'analyses et de filtrage des données utilisées, une validation manuelle reste nécessaire pour éliminer les faux pics et vérifier l'identification des métabolites d'intérêt. Le laboratoire TIRO possède une banque de données construites et validées avec les principaux métabolites d'intérêt. Plusieurs banques de donnée de métabolites existent également : HMDB (Human Metabolome Database)(244), MZCloud (ThermoFisher), METLIN(245), ChemSpider(246). La vérification de l'annotation est indispensable, parce que même à un m/z avec une forte précision de masse de 1 ppm, plusieurs annotations peuvent tout de même être attribuées(247). La vérification de l'annotation peut se baser sur la fragmentation MS2, le temps de rétention et la distribution isotopique(238). L'automatisation de l'étape de vérification est difficile parce que certains paramètres dépendent de l'instrument.

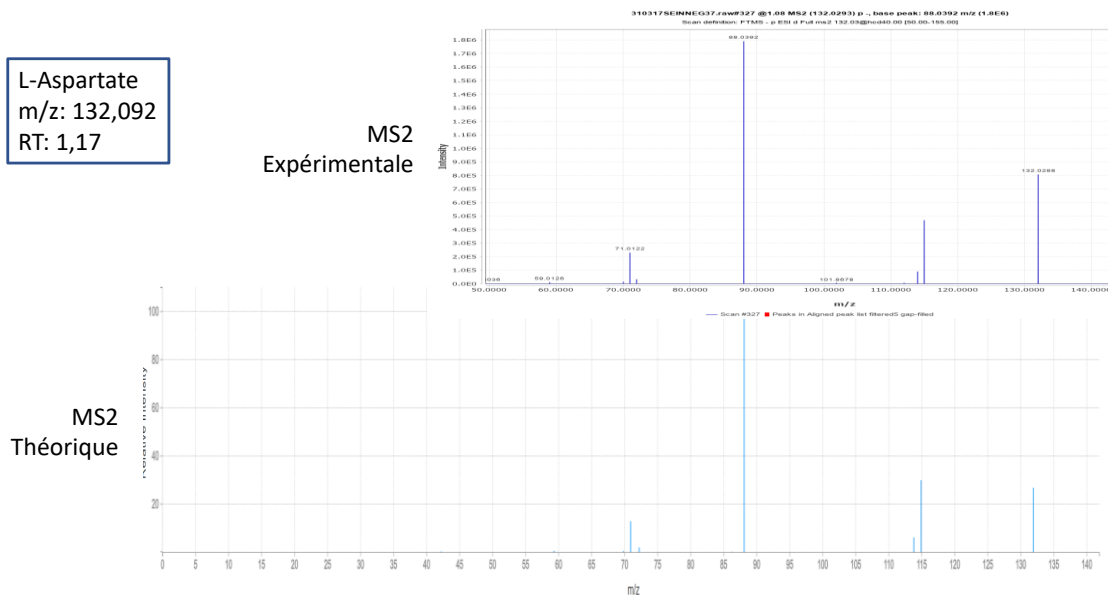


Figure 15. Comparaison manuelle des MS2 expérimentale et théorique*. m/z : rapport masse/charge.*

Pour l'étude EMME-A (Article 1&2), cette étape de vérification n'a pas été réalisée de manière systématique, mais uniquement pour les métabolites d'intérêt principaux mis en évidence. Les analyses métabolomiques principales étaient des analyses de pathway qui ne nécessitaient pas une étape de validation complète car l'ensemble d'un pathway devait être enrichi ou surreprésenté pour que l'analyse sorte positive. Une erreur ou une approximation d'annotation n'aurait donc pas eu de conséquence majeure sur le résultat. Les analyses métabolomiques ont donc été réalisées sur des métabolites prédits.

En revanche, dans l'étude du grade (Article 3), le but premier était la mise en évidence d'une signature métabolomique validée pour pouvoir faire une interprétation biologique précise. Tous les métabolites de la top-liste ont été vérifiés au niveau MS/MS et au niveau du sens biologique. Les analyses métabolomiques ont donc été réalisées sur des métabolites validés. Si la validation n'était pas possible, le métabolite était retiré.

Une fois l'ensemble de ces étapes réalisées, les données étaient ensuite analysées au niveau statistique pour déterminer les différences significatives entre plusieurs groupes d'échantillons.

VII. Étape Analytique spécifique

1. Analyses statistiques

Une fois que la matrice de données générée à partir des données brutes normalisées, différents types d'analyses statistiques peuvent être réalisés dans le but d'identifier des métabolites variants entre les différents groupes d'échantillons. Le domaine des statistiques appliquées aux données de grande dimension et à la métabolomique est très riche. Nous ne présenterons ici que les méthodes utilisées lors de cette thèse.

a) ANALYSES UNIVARIEES

Les analyses univariées telles que le test de Student, l'analyse de la variance (ANOVA) et leur équivalent non-paramétrique permettent d'évaluer la significativité des différences observées entre les échantillons des différents groupes ou classes(248). Chaque test n'évalue qu'une seule variable, et la significativité est indiquée par la valeur p. Cependant, les données par métabolomique non-ciblée en LC-MS contiennent souvent des centaines de variables (métabolites) corrélées, et la réalisation répétée de ces tests peut entraîner un taux de faux positifs élevé. Pour limiter cela, diverses méthodes de correction de la valeur p (par exemple la correction de Bonferroni), ou de calcul du taux de fausses découvertes peuvent être appliquées après l'analyse univariée(249,250).

b) ANALYSES MULTIVARIEES

Contrairement aux analyses univariées, les analyses multivariées impliquent l'observation et l'analyse simultanément de plusieurs variables. Étant donné le nombre de variables générées par une analyse métabolomique non-ciblée, les analyses multivariées permettent une vue globale de l'ensemble des données et sont les outils principalement utilisés. Les métabolites discriminants identifiés à partir des analyses

multivariées sont ensuite vérifiés par les analyses univariées dans un deuxième temps(238). Nous pouvons distinguer deux approches d'analyses multivariées très distinctes : les méthodes descriptives (analyses non supervisées) et les méthodes explicatives (analyses supervisées). Ces deux approches fournissent des informations complémentaires et sont généralement effectuées séquentiellement.

i. CLASSIFICATIONS NON SUPERVISEES

La classification non supervisée est une analyse sans a priori, également appelée analyse multivariée descriptive. Les classes ne sont pas connues au moment de l'analyse des métabolites. Elle utilise des variables descriptives, ici les données de métabolomique, pour effectuer une analyse sans a priori. Cette approche permet de définir des groupes de patients (clusters) en analysant les métabolites « en aveugle », de rechercher les caractéristiques propres à chaque cluster et d'analyser les métabolites et les voies métaboliques d'intérêt. Les méthodes les plus couramment utilisées en métabolomique non-ciblée sont l'analyse en composantes principales (PCA) et la classification ascendante hiérarchique (HCA, Hierarchical Cluster Analysis)(238,251).

La PCA ou *Principal Component Analysis,* représente chaque métabolite comme une dimension et chaque échantillon comme un point dans un espace multidimensionnel. Elle utilise une combinaison linéaire pour regrouper les métabolites entre elles et réduire le nombre de dimensions. L'analyse en composantes principales projette les données dans un nouvel espace représenté par les composantes principales grâce à une projection orthogonale (en général 2 ou 3)(252,253). Cela permet de réduire le nombre de variables et ainsi d'observer la variation globale des données de manière concise. Plus le profil métabolomique des échantillons est similaire, plus les points qui les représentent dans le graphe de score se rapprochent les uns des autres. La contribution de chaque métabolite à chacune des composantes principales est représentée par le graphe de poids (Loading

plot). Plus un métabolite est discriminant pour la construction des composantes principales, plus le point qui le représente dans le graphe de poids s'éloignera du point d'origine. La PCA est utilisée pour observer la variation globale dans les échantillons, par exemple pour détecter la présence de points aberrants ou un effet batch (effet spécifique d'une série d'échantillons analysés au même moment).

L'HCA, ou l'analyse par Classification Ascendante Hiérarchique, est une méthode qui utilise la similarité entre deux échantillons en calculant un algorithme spécifique. Ces étapes sont répétées jusqu'à ce que tous les échantillons soient regroupés ensemble(254). Contrairement à la PCA, l'HCA ne nécessite pas de spécification préalable du nombre de clusters et ne se base que sur une mesure de similarité pour regrouper les échantillons. Les regroupements sont souvent représentés sous forme d'un dendrogramme, un arbre binaire dont les feuilles représentent les échantillons et les branches représentent la hiérarchie du regroupement.

ii. CLASSIFICATIONS SUPERVISEES

Pour les classifications supervisées, aussi appelées analyses multivariées explicatives, chaque échantillon est considéré comme appartenant à une classe. Ces classes peuvent être des types de tumeurs, des conditions de traitements ou encore des marqueurs biologiques. Elle utilise des variables descriptives, ici les données de métabolomique pour réaliser une analyse orientée. Cette modalité permet de répondre à une question clinique, de définir des groupes de patients en fonction de la classe connue, de rechercher les métabolites spécifiques à chaque classe et d'analyser les métabolites et les voies métaboliques d'intérêt.

Parmi les méthodes de classifications supervisées, la PLS-DA ou *Partial Least Square Discriminant Analysis* est la plus utilisée. Chaque métabolite représente une dimension,

chaque échantillon représente un point dans un espace multidimensionnel. L'algorithme recherche des vecteurs permettant de séparer au mieux les 2 classes d'échantillons dans cet espace multidimensionnel. La régression des moindres carrés partiels (PLS) cherche à maximiser la corrélation entre les variables X (= données métabolomiques) et les variables à expliquer Y (= classes). Pour ceci, l'algorithme construit un espace de faible dimension basé sur les combinaisons linéaires des variables X initiales tout en ajustant le modèle pour y capturer les variations liées à Y(255). La régression PLS est particulièrement adaptée aux situations où le nombre d'observations (= nombre d'échantillons) est largement inférieur au nombre de variables (= nombre de métabolites)(256,257).

L'analyse discriminante PLS (PLS-DA) est une extension de la PLS, qui est adaptée aux variables Y qui sont catégoriques. Cette méthode cherche à obtenir une séparation maximale entre les différents groupes ou classes. La discrimination des différentes classes est visualisée sur le graphe de score et le pouvoir discriminatif de chaque métabolite est représenté sur le graphe de poids.

c) ÉTAPE DE VALIDATION

Test de validation interne

Pour estimer le pouvoir prédictif d'un modèle de discrimination et évaluer sa performance, la réalisation de tests de validation est nécessaire, particulièrement en cas de petit effectif avec un nombre d'échantillons limité(240). Les deux outils de validation statistique les plus utilisés sont le test de permutation (258,259) et la validation croisée.

Dans le test de permutation, l'ordre des classes (Y) est mélangé de façon aléatoire et un nouveau modèle de discrimination, basé sur la matrice Y permutée, est construit. Le

pouvoir prédictif des modèles permutés doit être inférieur au modèle de discrimination initial, sinon cela signifie que le modèle initial n'est pas performant.

Dans la validation croisée, une partie des échantillons est retirée du modèle pour servir de jeu de validation. Le modèle de discrimination est construit avec les échantillons restants (= jeu d'apprentissage). Le pouvoir prédictif du modèle est estimé sur le jeu de validation. Ce processus est répété plusieurs fois jusqu'à ce que tous les échantillons aient jouer le rôle de jeu de validation(260). Les performances sont finalement estimées en moyennant les taux de mauvaises classifications obtenus. Cette méthode de rééchantillonnage porte le nom de k-folds cross-validation, k correspondant au nombre de sous-groupes initialement formés et testés en tant de jeu de validation(261).

Test de Validation externe

Idéalement pour qu'un modèle soit validé, il faudrait que le modèle construit sur le jeu de données d'apprentissage, soit ensuite validé sur un jeu de données indépendant. Les capacités discriminantes de la règle de décision sont alors estimées sur ce jeu de données en calculant des indices de performance comme l'AUC, le taux de mauvaises classifications pour un critère binaire, la sensibilité et la spécificité ainsi que les intervalles de confiance associés. C'est ce qui a été réalisé dans l'article 3 Grade avec la cohorte de validation externe (biopsies diagnostiques provenant de Dijon)(cf. Paragraphe introduction au travail de thèse et paragraphe Cohortes).

2. Logiciel MetaboAnalyst

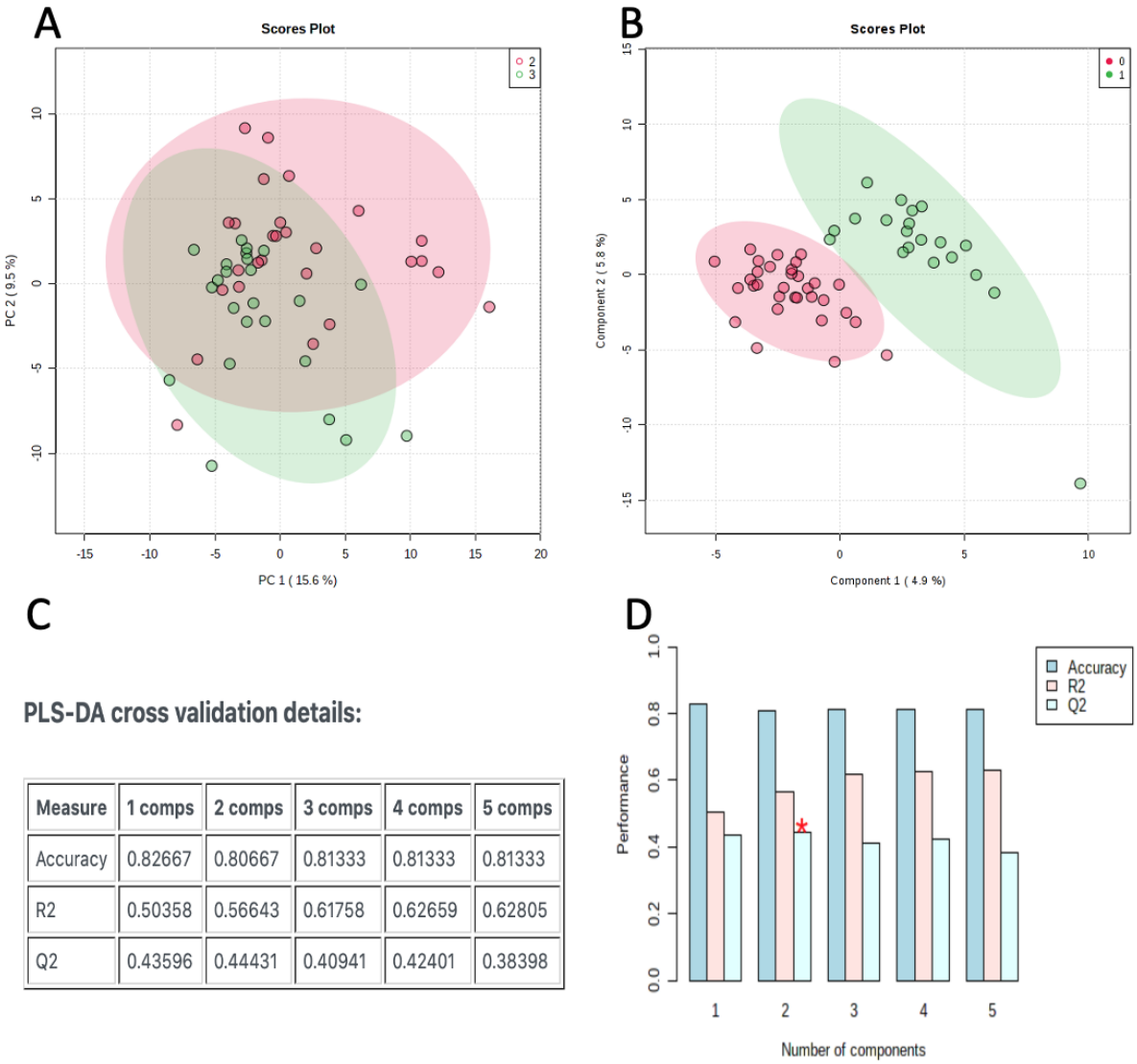
MetaboAnalyst est un classifieur utilisé dans le cadre de cette thèse pour les analyses supervisées. Toutes les analyses statistiques ont été effectuées en ligne à l'aide de MetaboAnalyst (https://www.metaboanalyst.ca/) version 5.0(262).

La seule méthode de normalisation, de transformation des données et de centrage des données utilisée était la transformation logarithmique. Les normalisations par la somme ou par la médiane n'ont pas amélioré les performances de l'analyse. L'analyse PLS-DA a été utilisée pour établir les score plots, les loading plots et pour réaliser les tests de validations croisées (précision de la performance, R2, Q2). Les courbes ROC (Receiver Operating Characteristic), les Heatmaps, l'exploration de l'enrichissement et l'analyse des voies métaboliques ont été générés en ligne à l'aide de MetaboAnalyst (https://www.metaboanalyst.ca/). La voie du tryptophane a été interprétée à l'aide des données de SMPDB (small molecule pathway database) et Kegg.

Ce chapitre introductif permet la compréhension des résultats présentés dans les articles de cette thèse

a) EXEMPLE DE PCA ET PLSDA

La méthode non supervisée PCA et la méthode supervisée PLSDA sont représentées sur la figure ci-dessous (Figure 16). Une analyse de la PCA était réalisée au préalable de toute analyse supervisée pour éliminer un facteur de confusion évident au sein des échantillons. De plus, une étape de cross-validation était réalisée afin d'obtenir les performances des tests en PLS-DA (Accuracy, R2, Q2).



Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.82667	0.80667	0.81333	0.81333	0.81333
R2	0.50358	0.56643	0.61758	0.62659	0.62805
Q2	0.43596	0.44431	0.40941	0.42401	0.38398

Figure 16. Exemple de PCA et de PLS-DA obtenues*. (A) : PCA ; (B) : PLS-DA ; (C) : PLS-DA cross validation avec accuracy (précision), R2 et Q2 en fonction du nombre de composantes prises en compte; (D) : représentation graphique par histogrammes des données de cross-validation en PLS-DA. PC : principal component ;*

L'accuracy mesure l'efficacité d'un modèle à prédire correctement à la fois les vrais positifs et les vrais négatifs. L'accuracy est une métrique pour évaluer la performance des modèles de classification à 2 classes ou plus. Le R^2, ou R-carré est appelé coefficient de détermination. C'est un indicateur utilisé en statistiques pour juger de la qualité d'une régression linéaire. Le Q2 est une estimation de la capacité prédictive du modèle. Plus précisément, le Q2 est une mesure de la capacité du modèle à prédire les valeurs pour les

échantillons qui n'ont pas été utilisés pour le calibrage du modèle. Le Q2 est calculé par validation croisée. Dans chaque validation croisée, les données prédites sont comparées aux données originales et la somme des carrés des erreurs est calculée. L'erreur de prédiction est ensuite additionnée sur tous les échantillons (somme des carrés des résidus prédits ou PRESS). Pour des raisons de commodité, le PRESS est divisé par la somme des carrés initiale et soustrait de 1 pour ressembler à l'échelle du R2. Les bonnes prédictions auront un PRESS faible ou un Q2 élevé. Un Q2 faible peut être révélateur d'un effectif insuffisant, ce qui est souvent le cas dans les études cliniques avec un nombre limité de patients. En effet, si le nombre de patients inclus dans l'étude est insuffisant, cela peut entraîner une sous-représentation de certaines caractéristiques ou une variabilité insuffisante dans les données. Ces facteurs peuvent affecter la qualité de la prédiction du modèle statistique et donc conduire à un Q2 faible. Il est également possible d'avoir un Q2 négatif, ce qui signifie que le modèle n'est pas du tout prédictif ou qu'il est surajusté (overfitting).

b) EXEMPLE D'ANALYSE DE METABOLITES D'INTERET

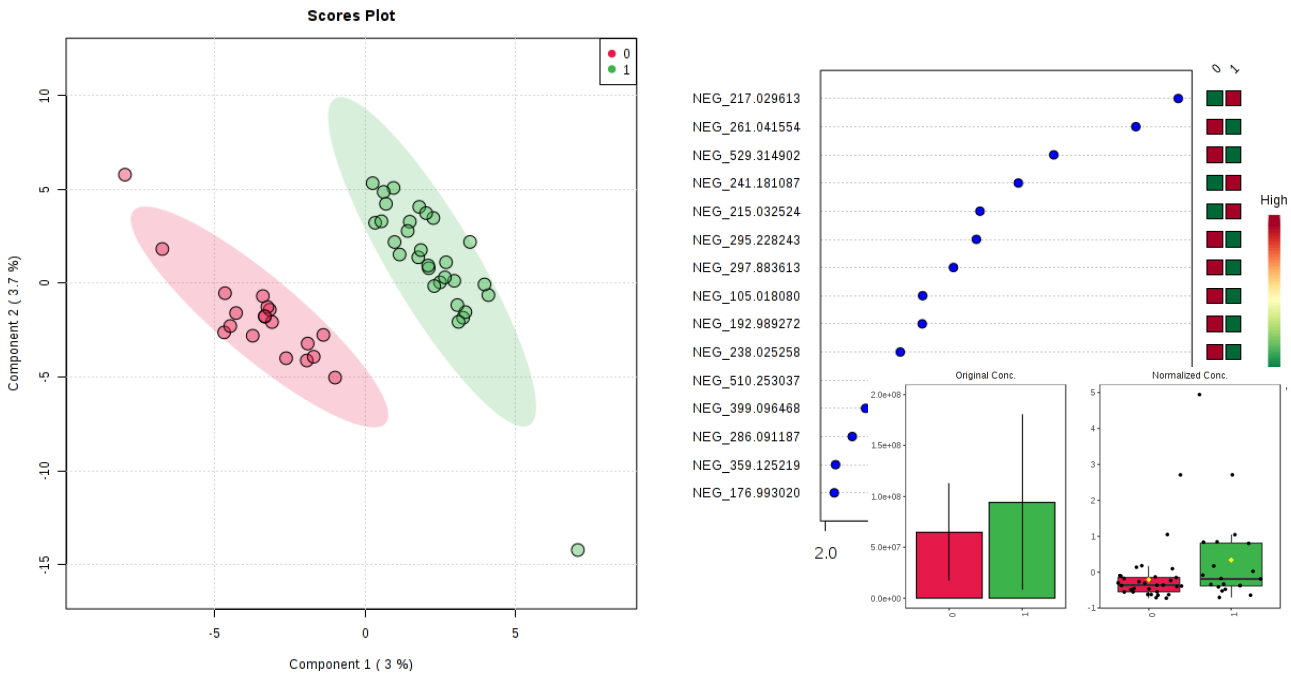


Figure 17. Exemple d'analyse de métabolites d'intérêt*. Conc. : concentration.*

Il existe deux mesures d'importance dans le PLS-DA : l'une est l'importance de la variable dans la projection (VIP) et l'autre est la somme pondérée des coefficients de régression absolus (coef.). Les cases colorées à droite indiquent les concentrations relatives du métabolite correspondant dans chaque groupe étudié. Les loading plots sont une visualisation de la répartition des concentrations d'un métabolite d'intérêt au sein d'une même classe et entre les classes étudiées.

c) HEATMAP

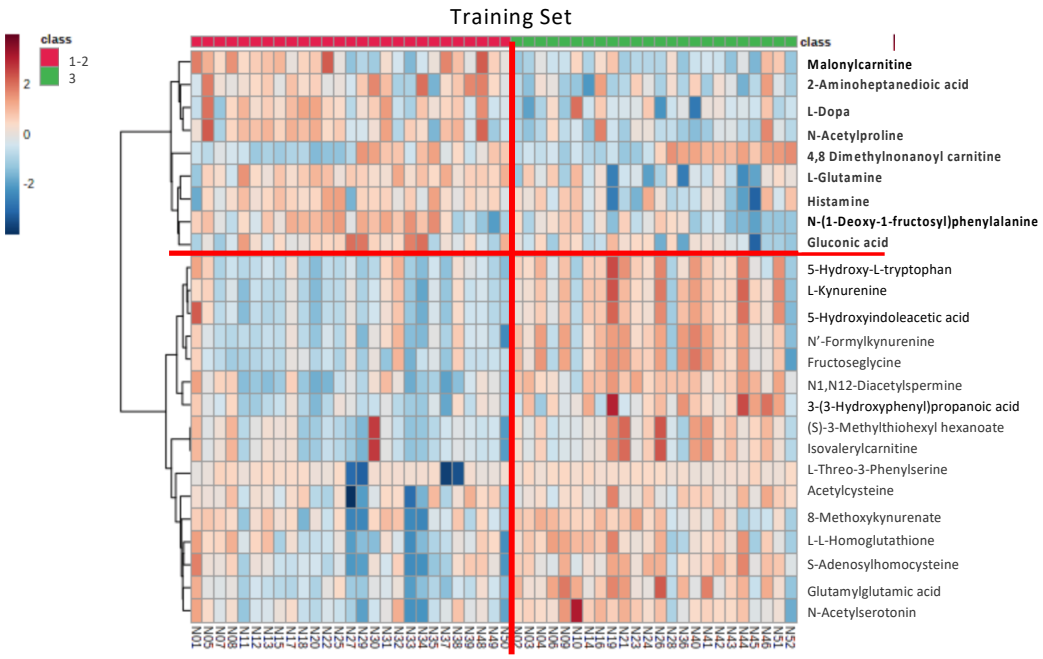


Figure 18. Heatmap ou carte thermique*. N : échantillon de la cohorte niçoise.*

La Heatmap ou carte thermique permet une visualisation intuitive d'un tableau de données. Chaque cellule colorée de la carte correspond à une valeur de concentration de la table de données, avec les échantillons en colonnes et les métabolites en ligne. Cette représentation permet d'avoir une vue d'ensemble sur les variations de concentration entre les échantillons au sein d'une même classe et entre les classes étudiées. Cette représentation a été très utile pour analyser l'hétérogénéité de concentrations des différents métabolites de la voie du tryptophane (Article Grade).

d) COURBE ROC/AUC

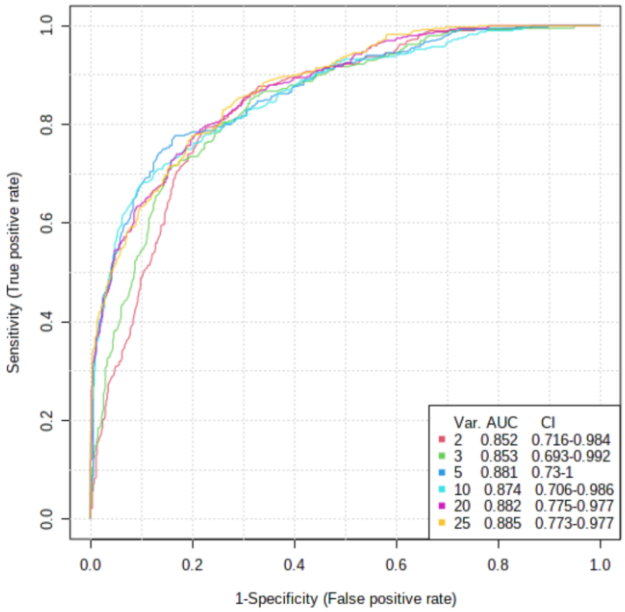


Figure 19. Exemple de courbe ROC avec valeur AUC*. AUC : aire sous la courbe ; CI: intervalle de confiance à 95%.*

La performance de la signature générée a été évaluée à l'aide de courbes ROC(263) qui permettent d'analyser les variations de la sensibilité et de la spécificité, afin de visualiser la performance globale du modèle. Ces courbes ROC sont représentées sur un graphique où l'axe des ordonnées correspond à la « sensibilité » et l'axe des abscisses à « 1 – spécificité ». La performance du modèle est donc déterminée par le meilleur compromis entre ceux deux critères. À partir de cette courbe, l'indice numérique utilisé afin d'évaluer la performance du modèle est l'aire sous la courbe ROC (AUC, Area Under the Curve) qui représente la probabilité du modèle à prédire correctement la catégorie de variable analysée. Cet indice varie entre 0.5 (performance liée au hasard) et 1 (performance parfaite).

e) ÉTUDE D'ENRICHISSEMENT

Pour évaluer l'importance biologique des voies métaboliques dans la condition expérimentale observée, la distribution des métabolites de chaque voie dans les

échantillons analysés est comparée à celle de l'ensemble des métabolites dans une banque de donnée, telle que KEGG ou SMPDB (The Small Molecule Pathway Database). Si les métabolites d'une voie donnée sont observés plus fréquemment et en plus grande quantité dans le set échantillon par rapport au set référence, la voie métabolique est considérée comme enrichie et est donc jugée importante sur le plan biologique pour la condition expérimentale observée(264,265). Pour chaque voie du métabolisme, un taux d'enrichissement et une valeur-p qui correspond à la probabilité d'un tel enrichissement est indiqué par MetaboAnalyst(262) (https://www.metaboanalyst.ca/).

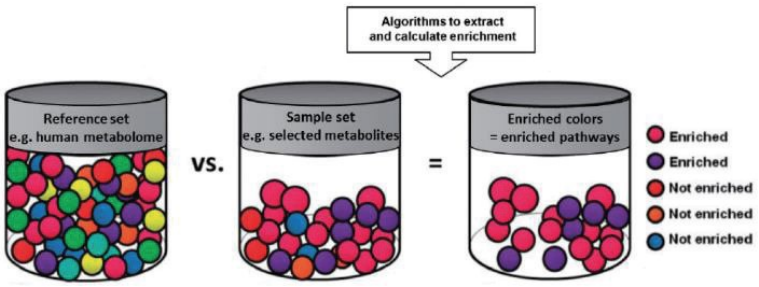


Figure 20. Principe de l'analyse par enrichissement*(264).*

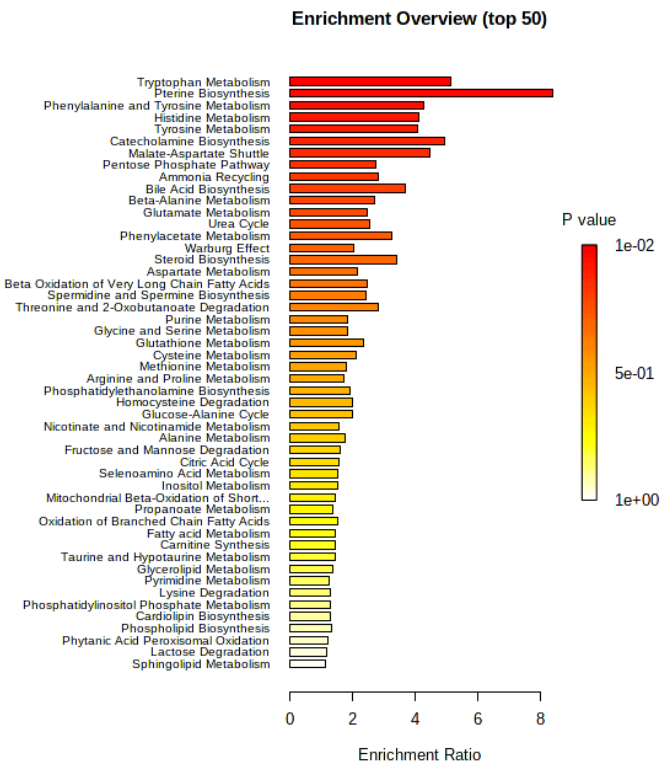


Figure 21. Exemple de résultat d'analyse d'enrichissement.

a) ANALYSE DES VOIES D'ACTIVATION METABOLIQUE

L'outil MetPA (Metabolomics Pathway Analysis) de MetaboAnalyst est mis à disposition pour l'analyse des voies métaboliques(266). A partir d'une liste de métabolites d'intérêt, MetPA permet d'identifier les voies métaboliques les plus pertinentes impliquées dans une étude métabolomique en utilisant une analyse de surreprésentation (over-representation analysis) ou bien une analyse d'enrichissement des voies (pathway enrichment analysis) qui peut être combinée à une analyse des caractéristiques topologiques des voies (pathway topological analysis).

L'analyse de surreprésentation vise à déterminer si le nombre de métabolite associé à une voie métabolique est dû au hasard ou non en se basant sur la liste des métabolites d'intérêt. Cependant, cette méthode ne prend en compte que le nombre de métabolites par voie métabolique et n'intègre pas l'amplitude du changement de leur abondance entre deux conditions testées. Ainsi les métabolites dont le changement d'amplitude est plus significatif auront le même poids que des métabolites moins significatifs.

Pathway Name	Match Status	p
Porphyrin and chlorophyll metabolism	2/30	1.9833E-4
Histidine metabolism	6/16	2.3132E-4
Cysteine and methionine metabolism	9/33	0.0014316
Pyruvate metabolism	3/22	0.0024176
Glycine, serine and threonine metabolism	12/33	0.0024495
Purine metabolism	15/65	0.0055832
Pyrimidine metabolism	9/39	0.0080538
Nitrogen metabolism	2/6	0.021183
Terpenoid backbone biosynthesis	1/18	0.031479
Tryptophan metabolism	5/41	0.035568

Tableau 3. Analyse de surreprésentation avec le nombre de métabolites concordants retrouvés. *P : p-value.*

Il est recommandé de se référer à l'analyse d'enrichissement des voies lorsque les concentrations de métabolites sont disponibles (cf. paragraphe précédent). Cependant, utilisée seule, ni l'analyse de surreprésentation, ni l'analyse d'enrichissement des voies, ne prennent en compte la structure globale de la voie métabolique pour déterminer les voies les plus impliquées. L'analyse des voies métaboliques génère finalement différents graphiques représentés ci-dessous :

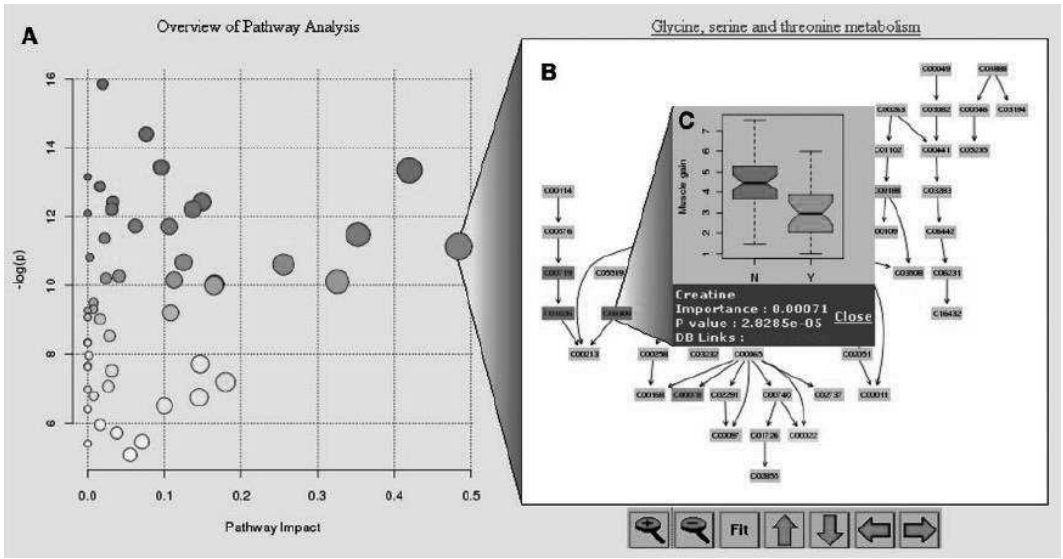


Figure 22. Interprétation intégrative des voies métabolique*. (A) Visualisation de l'analyse globale, (B) visualisation des voies métaboliques, (C) visualisation des métabolites(266).*

Le graphique interactif présenté dans l'analyse permet de classer les voies métaboliques en fonction de leur significativité (axe des ordonnées ; avec la valeur de p obtenue grâce à l'analyse de surreprésentation ou d'enrichissement des voies) et de leur impact (axe des abscisses ; calculé à partir de l'importance des métabolites mesurée dans l'analyse topologique des voies). Chaque cercle du graphique est associé à une voie métabolique spécifique, que l'on peut découvrir en interagissant avec le graphique. Si des données quantitatives ont été jointes à l'analyse, il est également possible de visualiser les différences de concentrations entre deux groupes d'échantillons pour un métabolite sélectionné.

VIII. Articles

1. Article EMMEA – Analyse non supervisée initiale

a) INTRODUCTION PREALABLE AUX ANALYSES NON SUPERVISEES

Le premier article EMMEA écrit en collaboration avec GAL Jocelyn avait pour objectif de comparer 5 méthodes de clustering de machine learning. Les méthodes de PCA et de HCA ont été décrites précédemment (Partie VII, Étape analytique spécifique). Quelques éléments complémentaires sont nécessaires à la compréhension spécifique de cet article.

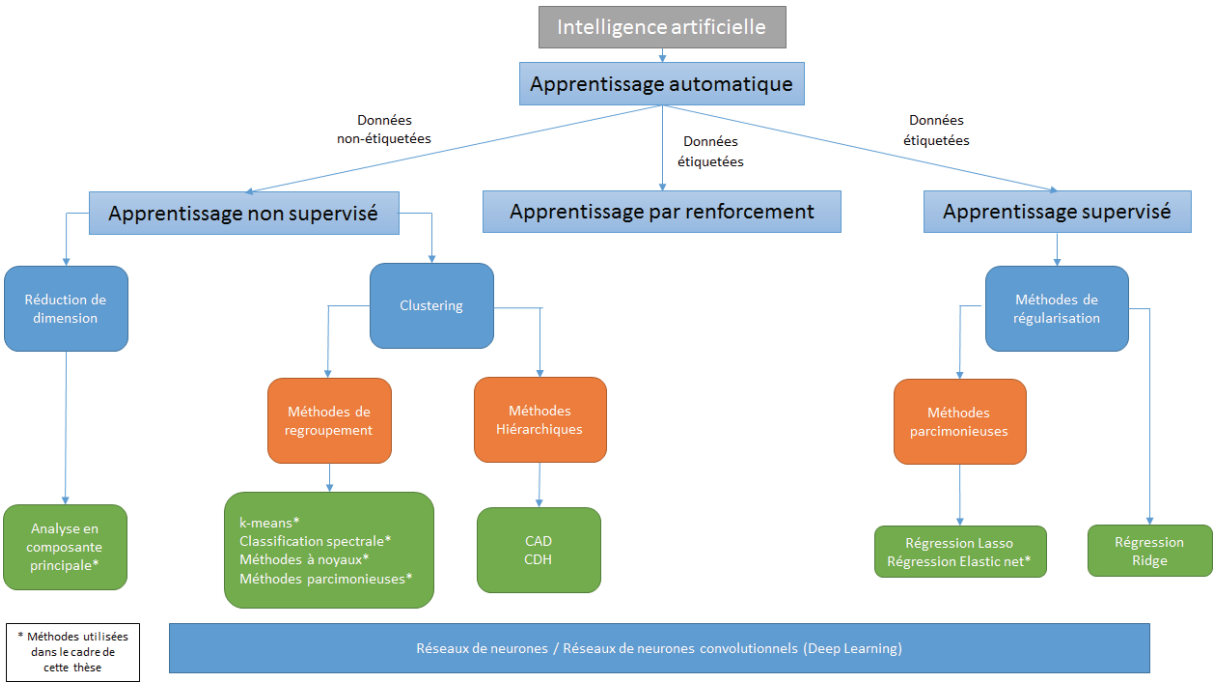


Figure 23. Schématisation des différentes méthodes d'apprentissage automatique.

Comme le montre la figure ci-dessus, il existe des méthodes de partitionnement, distinctes des méthodes hiérarchiques au sein des méthodes de clustering. Cette famille de méthodes propose une partition des données plutôt qu'une structure du type

“dendrogramme”. Le principe de ces méthodes est alors de comparer plusieurs partitionnements jusqu’à sélectionner le meilleur.

La méthode des k-means est la plus connue (267,268). C’est une méthode itérative de classification largement utilisée en biologie qui vise à minimiser la somme des distances entre les points d'un même groupe et le centroïde, désigné comme le centre d'un cluster. Le résultat final est conditionné par le choix initial des centroïdes, chaque cluster se définissant en fonction de son centroïde. Il s'agit donc de l'élément central de l'algorithme.

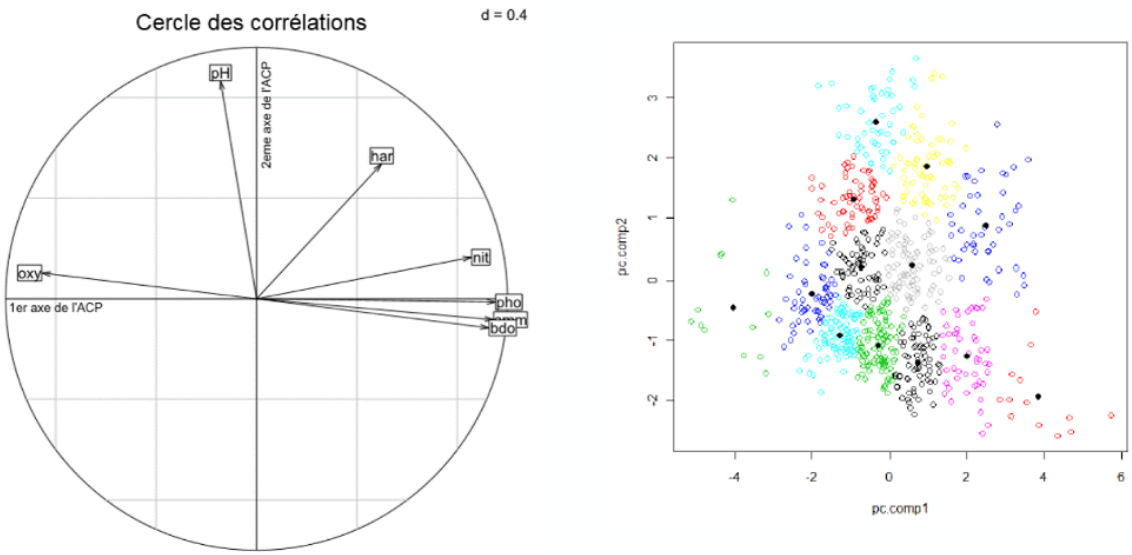


Figure 24. Schématisation de la méthode des k-means.

La classification spectrale (ou clustering spectral) est utilisée pour son efficacité et sa simplicité d’implémentation qui se résume en l’extraction des valeurs et vecteurs propres d’une matrice de similarités créée à partir d’un ensemble de données.

Les méthodes à noyaux (269,270) permettent de trouver des fonctions non linéaires, tout en s’appuyant sur des méthodes linéaires (figure 25). Une fonction noyau correspond à

un produit scalaire dans un espace de re-description des données et est souvent de grande dimension. De nombreuses applications requièrent des modèles non linéaires pour rendre compte des dépendances et des régularités sous-jacentes dans les données.

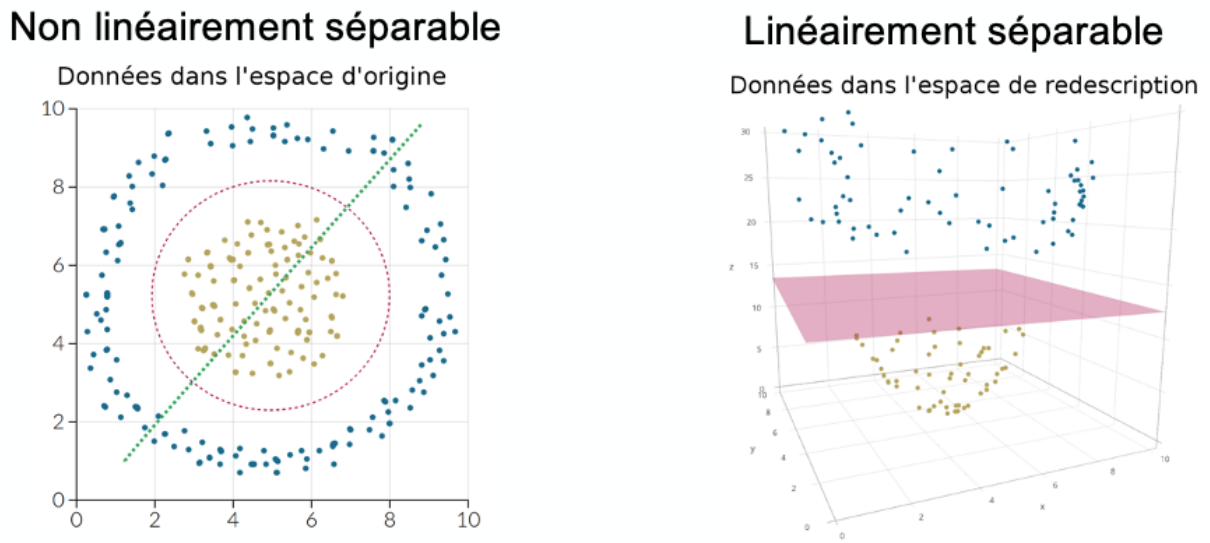


Figure 25. Schématisation de la méthode à noyaux*. Passage où les données sont non linéairement séparables vers un espace de description où les données sont linéairement séparables*

En outre, afin de prendre en compte la nature hétérogène de certaines données, il est intéressant de combiner plusieurs noyaux, afin d'obtenir un modèle plus souple(271). L'approche appelée multinoyaux (Multiple Kernel Learning) a ainsi été introduite (272) pour généraliser l'approche mono-noyau. Récemment, la méthode single-cell interpretation via multikernel learning (SIMLR) (273), basée sur une approche multi noyaux a été développée et appliquée sur un ensemble de données de séquençage du transcriptome (RNA-seq). Cette méthode combine à la fois une approche multi-noyaux, une réduction de dimension et une représentation graphique des données. Ces trois étapes sont représentées en figure 26.

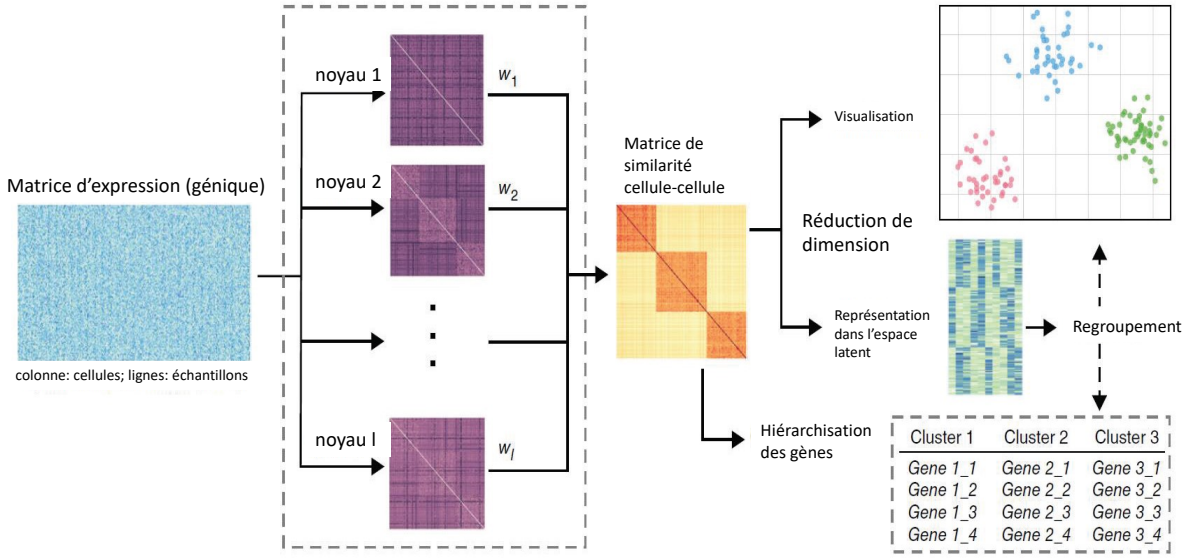


Figure 26. Schématisation de l'approche multinoyaux*. Visualisation des 3 étapes de la méthode SIMLR appliquée à des données de RNA-seq(273)*

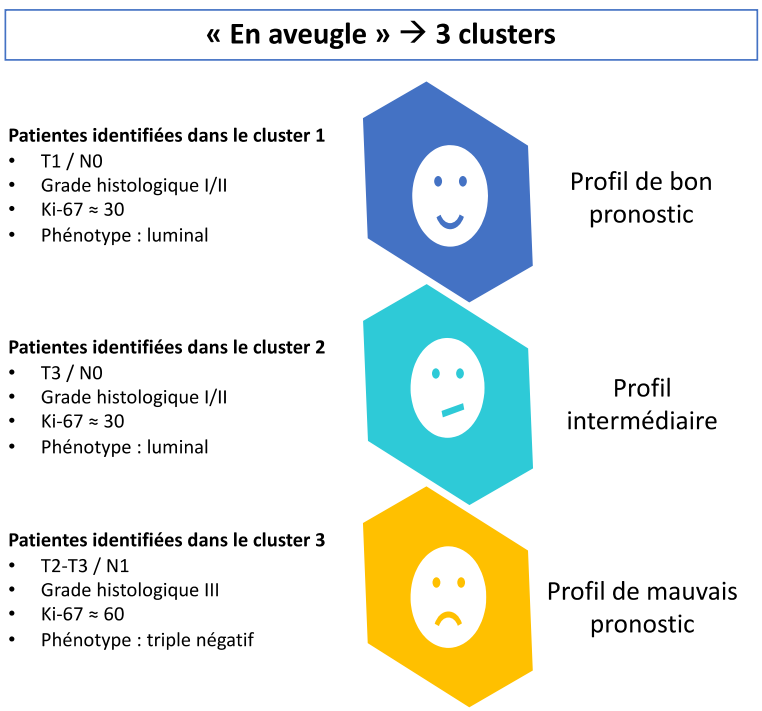
Les méthodes de classification parcimonieuses (sparse) développées plus récemment, permettent de former des clusters en utilisant d'une part les algorithmes de clustering classiques (classification hiérarchique, K-means) et d'autre part, en ajoutant une pénalisation de type Lasso(274) (Least Absolute Shrinkage and Selection Operator) à la fonction objectif pour sélectionner des biomarqueurs. La méthode de clustering sparse k-means(275) est un algorithme amélioré du k-means permettant de partitionner les observations lorsque la base de données contient un grand nombre de données. Gilet *et al.* en 2017(276) ont proposé une nouvelle méthode alternative nommée K-sparse. Plutôt qu'une pénalité, les auteurs proposent de définir une contrainte en norme *l1* (277). Cette méthode combine à la fois la méthode de clustering k-means, une réduction de dimension et une sélection de variables.

b) RESUME DES PRINCIPAUX RESULTATS & ARTICLE

L'objectif de l'étude EMMEA était de comparer les signatures métabolomiques du cancer du sein obtenues par 5 méthodes différentes d'apprentissage automatique non supervisé.

Cette étude a été permise grâce à l'inclusion rétrospectives de 52 patientes atteintes d'un cancer du sein localisé avec une indication de chimiothérapie adjuvante entre 2013 et 2016. Nous avons réalisé le profilage métabolomique des tumeurs réséquées par LC-MS, avec 449 métabolites sélectionnés.

Les clusters obtenus à l'aide de 5 méthodes ML non supervisées (PCA k-means, sparse k-means, spectral clustering, SIMLR et k-sparse) ont été comparés en termes de caractéristiques cliniques et biologiques. Avec un paramètre de partitionnement optimal k = 3, les cinq méthodes ont permis d'identifier trois groupes de patientes se différenciant par leur pronostic (favorable, intermédiaire, défavorable), présentant des profils cliniques et biologiques différents. Les méthodes SIMLR et K-sparse ont été les plus efficaces en termes de regroupement. L'analyse de survie in-silico (PREDICT tool) a révélé une différence significative entre les 3 groupes en ce qui concerne la survie spécifique à 5 ans et à 10 ans (données de survie simulée).



PREDICT Tool	Survie spécifique à 5 ans	Survie spécifique à 10 ans
	p=0.002	p=0.004
Cluster 1	87% [80 - 91]	80% [73 - 86]
Cluster 2	81% [69 - 90]	75% [60 - 85]
Cluster 3	68% [60 - 74]	62% [53 - 69]

Méthode K-Sparse

Figure 27. Principaux résultats cliniques du clustering de l'étude EMMEA.
Classification TNM en Annexe 1.

Une analyse plus poussée des voies métaboliques utilisant les 449 métabolites prédits sélectionnés a montré des différences significatives dans le métabolisme des acides aminés, du glucose et de la glutaminolyse.

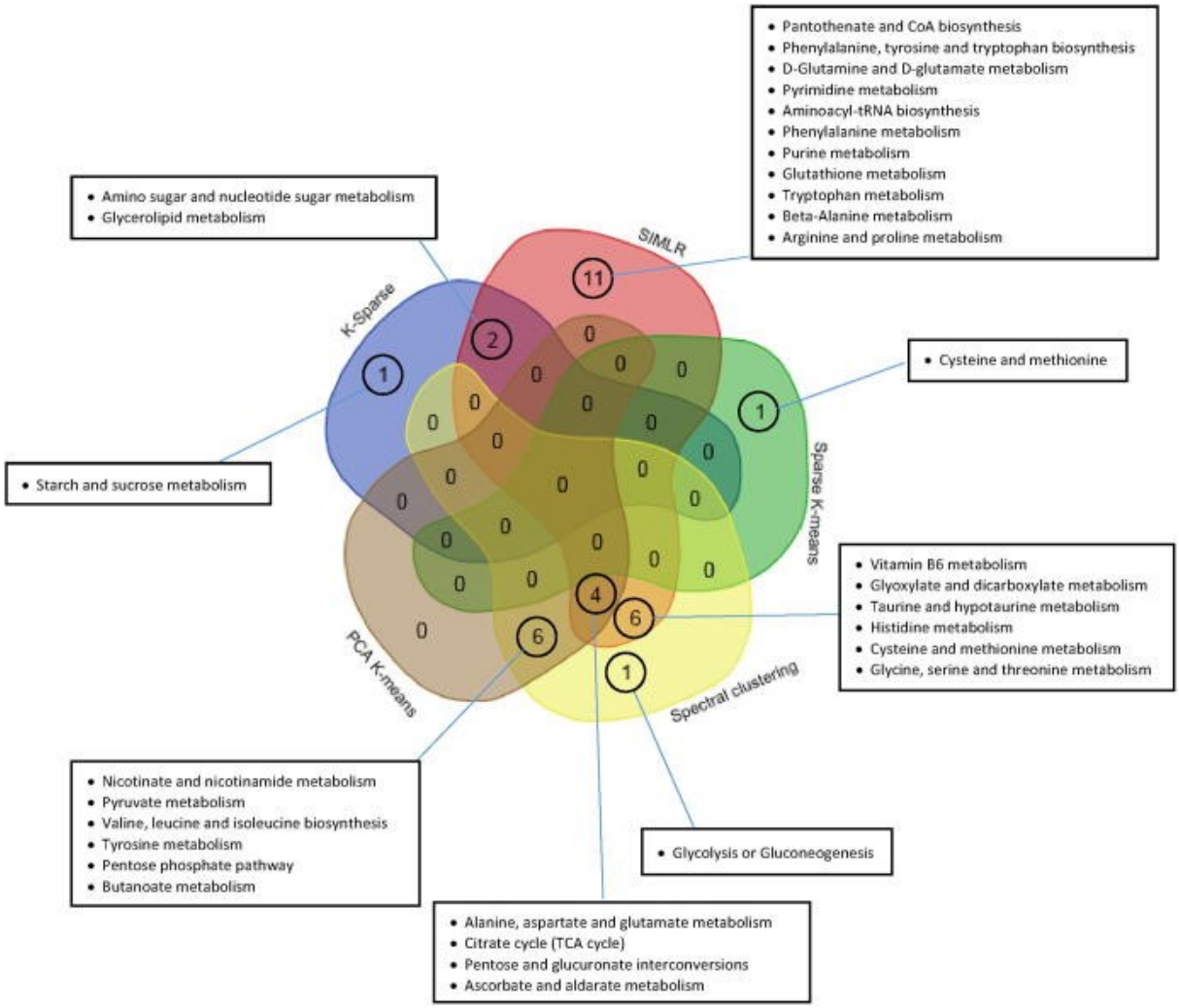


Figure 28. Principaux résultats métabolomiques de l'étude EMMEA.

Computational and Structural Biotechnology Journal 18 (2020) 1509–1524

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Comparison of unsupervised machine-learning methods to identify metabolomic signatures in patients with localized breast cancer

Jocelyn Gal [a,1,*], Caroline Bailleux [b,1], David Chardin [c,d,1], Thierry Pourcher [d], Julia Gilhodes [e], Lun Jing [d], Jean-Marie Guigonis [d], Jean-Marc Ferrero [b], Gerard Milano [f], Baharia Mograbi [g], Patrick Brest [g], Yann Chateau [a], Olivier Humbert [c,d], Emmanuel Chamorey [a]

[a] University Côte d'Azur, Epidemiology and Biostatistics Department, Centre Antoine Lacassagne, Nice F-06189, France
[b] University Côte d'Azur, Medical Oncology Department Centre Antoine Lacassagne, Nice F-06189, France
[c] University Côte d'Azur, Nuclear Medicine Department, Centre Antoine Lacassagne, Nice F-06189, France
[d] University Côte d'Azur, Commissariat à l'Energie Atomique, Institut de Biosciences et Biotechnologies d'Aix-Marseille, Laboratory Transporters in Imaging and Radiotherapy in Oncology, Faculty of Medicine, Nice F-06100, France
[e] Department of Biostatistics, Institut Claudius Regaud, IUCT-O Toulouse, France
[f] University Côte d'Azur, Centre Antoine Lacassagne, Oncopharmacology Unit, Nice F-06189, France
[g] University Côte d'Azur, CNRS UMR7284, INSERM U1081, IRCAN TEAM4 Centre Antoine Lacassagne FHU-Oncoage, Nice F-06189, France

ARTICLE INFO

Article history:
Received 11 February 2020
Received in revised form 15 May 2020
Accepted 16 May 2020
Available online 3 June 2020

Keywords:
Unsupervised machine learning
Metabolomics
Breast neoplasms
Computer simulation

ABSTRACT

Genomics and transcriptomics have led to the widely-used molecular classification of breast cancer (BC). However, heterogeneous biological behaviors persist within breast cancer subtypes. Metabolomics is a rapidly-expanding field of study dedicated to cellular metabolisms affected by the environment. The aim of this study was to compare metabolomic signatures of BC obtained by 5 different unsupervised machine learning (ML) methods. Fifty-two consecutive patients with BC with an indication for adjuvant chemotherapy between 2013 and 2016 were retrospectively included. We performed metabolomic profiling of tumor resection samples using liquid chromatography-mass spectrometry. Here, four hundred and forty-nine identified metabolites were selected for further analysis. Clusters obtained using 5 unsupervised ML methods (PCA k-means, sparse k-means, spectral clustering, SIMLR and k-sparse) were compared in terms of clinical and biological characteristics. With an optimal partitioning parameter k = 3, the five methods identified three prognosis groups of patients (favorable, intermediate, unfavorable) with different clinical and biological profiles. SIMLR and K-sparse methods were the most effective techniques in terms of clustering. *In-silico* survival analysis revealed a significant difference for 5-year predicted OS between the 3 clusters. Further pathway analysis using the 449 selected metabolites showed significant differences in amino acid and glucose metabolism between BC histologic subtypes. Our results provide proof-of-concept for the use of unsupervised ML metabolomics enabling stratification and personalized management of BC patients. The design of novel computational methods incorporating ML and bioinformatics techniques should make available tools particularly suited to improving the outcome of cancer treatment and reducing cancer-related mortalities.

1. Introduction

Breast cancer (BC) is the most common type of cancer in women worldwide and the second leading cause of cancer-associated deaths [1]. The treatment strategy may be guided by two classifications indicating the aggressiveness of the tumor. The anatomy-clinical classification is based on age, TNM, histological factors (histological grade, Ki-67) as well as on hormonal-receptor status and Her-2 expression. The molecular classification resulting from genomic [2], transcriptomic [3] and proteomic [4] analyses introduced the concept of luminal A, luminal B, Her-2 and basal-like BC [5–7]. This latter classification from Perou and Sorlie was assessed using unsupervised analyses [6,8]. Efforts have been made to develop multivariate prognostic models such as, AdjuvantOnline®,

* Corresponding author at: Department of Epidemiology and Biostatistics, Centre Antoine Lacassagne, University Côte d', Azur 33 avenue de Valombrose, 06189 Nice, France.
E-mail address: jocelyn.gal@nice.unicancer.fr (J. Gal).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2020.05.021

PREDICT Tool [9,10] and multigene predictors [11,12]. The use of biomarker-based tests, including omics-based tests, has steadily increased over the last decade as a result of the need for personalized treatment strategies designed to optimize outcomes [13–18]. Several genomic prognostic markers have been described for BC such as OncotypeDX®, Prosigna®, MammaPrint®, Endopredict® Genomic grade index® and BC Index® [19]. Two markers are commercially available and are increasingly used in clinical practice (21-gene recurrence score OncotypeDX® and 70-gene prognostic signature MammaPrint®). However, heterogeneity persists in biological features within BC subtypes, thus highlighting the need to improve the taxonomy [20]. This heterogeneity may be related to specific combinations of genetic, pathological and environmental factors leading to specific metabolic alterations and interactions [21,22].

Metabolomics is a new and growing field dedicated to the study of metabolism at overall level that promises to provide new insights into disease mechanisms and drug effects. Indeed, metabolomics may offer a complementary approach to genomics and could be used to better understand the influence of the environment on tumor phenotype [23]. Two distinct approaches characterize metabolomics: a targeted approach aimed at quantifying as accurately as possible a limited number of predefined metabolites of interest [24] and an untargeted approach aimed at measuring, without any a priori, as many metabolites as possible in a sample [25,26]. As with other omics approaches, metabolomics generates high-dimensional data. The processing of these data can be done by applying supervised or unsupervised machine learning (ML) algorithms that are increasingly used for medical diagnosis and therapeutic strategy guidance [27–29]. Unsupervised ML, in which no a priori class label information is given to guide the algorithm [30], seems a suitable alternative to analyze these data and address the problem of BC heterogeneity [6]. The aim of this study was to compare metabolomic signatures of BC obtained using five different unsupervised ML methods. To evaluate the consistency of our results, the clusters obtained by unsupervised ML methods were compared with patients' clinical characteristics and identified metabolic pathways.

2. Material and methods

2.1. Patients

This is a retrospective cohort study based on data and samples from 52 patients already available in the Centre Antoine Lacassagne tumor bank and collected during routine practice between 2013 and 2016. Patient tumor characteristics were: clinical stages I to III_B biopsy-proven BC, with an indication for post-surgery adjuvant therapy. Tumor phenotypes were classified into three subtypes: triple-negative (estrogen receptor, progesterone receptor and Her-2 non-over-expressed); luminal (estrogen receptor and/or progesterone receptor positive and Her-2 non-over-expressed); Her-2 over-expressed (Her-2 over-expressed, estrogen receptor and progesterone receptor either positive or negative) [31]. After surgery, all patients were treated according to current guidelines, with sequential chemotherapy including anthracyclines (epirubicin and cyclophosphamide) and taxanes followed by radiotherapy. Patients with Her-2 over-expressed tumors were treated with trastuzumab concurrently with taxanes and continued for one year. Patients with luminal BC were then treated by endocrine therapy with tamoxifen or an aromatase inhibitor, based on menopausal status. Clinical, histological, radiological and therapeutic data were retrospectively extracted from our facility's digital records or collected by a clinical data monitor. Follow-up data were either extracted from our facility's digital records or retrieved by telephone if patients had changed facilities during surveillance. Written informed consent was obtained from all study participants. All procedures performed in this study involving tissue collection and analyses were following the ethical standards of the institutional and/or national research committee (French National Commission for Informatics and Liberties N°17003 and National Institute Health data N°1515251018).

2.2. Data-preprocessing, metabolite identification, statistical and pathway analysis

Sample collection, preparation and data-processing using MZmine [32,33] are shown in Supplementary Material S1 and Supplementary Fig. 1 Metabolites obtained from positive and negative ionization modes were combined. Only metabolites with no null values after pre-processing were selected for analysis. When a metabolite was detected in both positive and negative modes, only the mode offering the highest average intensity was considered. After these steps, 1271 metabolites were identified. To eliminate noisy data, a filtering function was applied before statistical analysis. Finally, statistical analysis was performed on 449 metabolites. The identification of metabolic pathways was performed using MetaboAnalyst database sources [34]. The impact score was determined by the relative pathway topological effect of the metabolites, and $-\log(p)$ was used as the enrichment score, reflecting the probability of the pathway being identified at random; the number of "hits" was the actual number of matched metabolites in the pathway. For the selection of the most relevant pathways, we applied the following criteria: Impact >0, FDR < 0.25 and $p < 0.05$ [35].

A Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to display all possible logical relations between the metabolites or pathways identified by the clustering methods. Differences between clusters regarding the most active metabolites were plotted using boxplots.

2.3. Clustering algorithms

Five unsupervised clustering methods were selected and compared: Principal Component Analysis (PCA) k-means, Sparse k-means, Single-cell Interpretation via Multi-kernel LeaRning (SIMLR), k-sparse and Spectral clustering. Many clustering approaches exist, among which two of the most popular are K-means and spectral clustering [36]. PCA k-means and Sparse k-means are two well established, K-means based methods frequently used in computational. SIMLR and K-sparse are two recently developed k-means based methods of particular interest for omics data. These methods use different dimension reduction steps with k-means. In order to apply these five unsupervised clustering methods, the optimal number of clusters was determined in advance using five criteria: gap [37], silhouette [38,39], Davies-Bouldin [40], Calinski-Harabasz [41] and SIMLR method [42]. PCA k-means clustering, combines PCA to reduce the number of dimensions of a dataset and the k-means method to minimize the intra-cluster variance for a chosen number of k clusters [43–45]. Spectral clustering [46,47] is based on graph theory. It consists of identifying dense regions in a multidimensional dataset, i.e. observations that can form a non-convex set but are close to each other. Sparse k-means clustering was developed in 2010 by Witten and Tibshirani [8]. This method is based on a Least Absolute Shrinkage and Selection Operator (LASSO) approach [48] and combines the LASSO approach and the k-means method which simultaneously find the clusters and select features. SIMLR clustering [42] was developed to analyze scRNA-seq data. This method searches for appropriate cell-to-cell similarity metrics to perform dimension reduction and clustering. In multiple-kernel learning frameworks, this

method may be especially beneficial for data containing no identifiable clusters. K-sparse clustering [49] is an algorithm combining dimension reduction and relevant feature selection using a constraint in L1-norm rather than a lasso-type penalty to select the features. The performance of an unsupervised clustering method is measured by its ability to partition data. Partitioning is considered optimal when it minimizes the average distance between patients within a cluster (homogeneity) and maximizes cluster distances 2 by 2 (separability). The performances of the five methods were compared using the silhouettes index (SI) [39]. The SI ranges between −1 and 1 and assesses whether a patient belongs to the "right" cluster. The closer the index is to 1, the more satisfactory the assignment of a patient to a cluster. The t-SNE method was used for data visualization [50]. Processing times were obtained on a computer using an i5 processor (3.1 GHz).

2.4. Clinical evaluation

The relevance of the discovered clusters was assessed by comparing the clinical and survival characteristics between clusters using χ^2 or Fisher's exact tests for categorical data, analysis of variance or Mann-Whitney's test for continuous variables and log-rank test for censored data. Overall survival (OS) was defined as the time between diagnosis and death due to any cause. Specific survival (SS) was determined by the time between diagnosis and death due to BC. Recurrence-Free Survival (RFS) was defined as the time between diagnosis and the first recurrence (local, regional and metastasis). Patients showing no event (death or recurrence) or lost to follow-up were censored at the date of their last contact. OS, SS, and RFS were estimated using the Kaplan-Meier method. Median follow-up with a 95% confidence interval was calculated by reverse Kaplan–Meier method. All analyses were performed with Matlab® R2018b for PCA k-means, Spectral clustering, SIMLR (https://github.com/BatzoglouLabSU/SIMLR/tree/SIMLR/MATLAB) and k-sparse clustering and R [51] using package Sparcl [52] for sparse k-means clustering. The difference between clusters regarding the most biologically significant metabolites was plotted using boxplots. For clinical and biological analyses, all p-values <0.05 (two-sided) were considered statistically significant.

2.5. Prediction for 5- and 10-year overall and specific survival

Web-based prognostication PREDICT tool (https://breast.predict.nhs.uk/tool) [9,10,53] was used to estimate predicted OS (pOS) and predicted SS (pSS) at 5 and 10 years, based on several patient and tumor characteristics. For each patient, ten characteristics were entered manually: age at diagnosis, menopausal status, estrogen receptor status, Her-2 status, Ki-67 status, tumor stage, histological grade, mode of detection, number of positive nodes and presence of micrometastases. PREDICT tool can be used to estimate expected overall survival at 5 years and 10 years in the absence of available survival data due to short follow-up. If information was missing for detection, bisphosphonate therapy or menopausal status, patients were not excluded but the "unknown" category was used. Only one patient was excluded because of missing tumor grade data. A 1000 resamples bootstrap was used to estimate the 95% confidence interval.

3. Results

3.1. Patient characteristics

Tumor and treatment features of the 52 patients were described in Table 1. Median age was 63 years (range: 37–88). The main histological type was invasive ductal carcinoma (92%), and the main tumor stages were T1 (40.5%) and T2 (46%). Twenty-four patients (46%) presented axillary lymph node invasion. Two patients (4%) were oligometastatic at diagnosis. Forty-three percent of patients had histological grade II tumors and 47% had grade III tumors. Half of the patients had negative hormone receptor status (48%) and 24% of patients had Her-2 over-expression. Median follow–up was 48.5 months (95%CI [43–54.5]). Twenty-one patients presented a recurrence: 4 local recurrences (7.5%), 6 regional recurrences (11.5%) and 11 metastatic recurrences (21%). Three-year OS was 90% [82–99], 3-year SS was 92% [85–100] and 3-year RFS was 82% [72–93] (Supplementary Fig. 2). Median OS, SS, and RFS were not reached.

Table 1
Patients' demographics and treatment characteristics.

Clinical characteristic	No. of patients	%
Age (median min – max)	63.2 (37–88)	
Histology type		
Invasive ductal carcinoma	48	92
Invasive lobular carcinoma	3	6
Microinvasive carcinoma	1	2
Tumor stage		
T1	21	40.5
T2	24	46
T3	7	13.5
Axillary lymph node status		
N0	28	54
N+	24	46
Metastasis		
M0	50	96
M1	2	4
Histological grade		
I	5	10
II	22	43
III	24	47
Hormonal receptors status*		
Negative	25	48
Positive	27	52
Her-2 status		
Non-over-expressed	40	74
Over-expressed	12	24
Triple-negative status		
No	37	71
Yes	15	29
Tumor phenotype		
Her2	12	23
Luminal	25	48
Triple-Negative	15	29
Adjuvant Chemotherapy		
No	13	25
Yes	39	75
Adjuvant Radiotherapy		
No	9	17
Yes	43	83
Adjuvant Hormonotherapy		
No	24	46
Yes	28	54

* Oestrogen and/or progesterone.

3.2. Clustering results

3.2.1. Estimated number of clusters

Using four methods (Gap statistic, Calinski-Harabasz, Silhouette and SIMLR criterion), the optimal number of clusters was equal to three (k = 3) (Supplementary Fig. 3). Only for Davies-Bouldin criterion, the optimal number of clusters was equal to four (k = 4). It

seems reasonable, therefore, to conclude that the optimal number of clusters is equal to 3.

3.2.2. Patient distribution

Three clusters were identified with each of the five clustering methods, (Fig. 1). In terms of processing times, PCA k-means was the fastest and K-sparse was the longest (Supplementary Table 1). SIMLR and k-sparse methods were the most discriminants with an average silhouette value of 0.85 and 0.91, respectively (Fig. 2). Seventy-three percent of patients (38/52) were ranked in the same clusters by the five methods, 17.5% of patients (9/52) were classified in the same clusters by 4 methods and 9.5% of patients (5/52) were classified in the same clusters by 3 methods.

3.2.3. Comparison of clinical characteristics between clusters

As shown in Table 2, the 5 methods revealed significant inter-cluster differences. Patients in cluster 3 had mainly unfavorable prognostic factors: tumor stage T2/T3, histological grade III, high mitotic score and triple-negative phenotype. In contrast, patients in cluster 1 had mainly favorable prognosis factors: tumor stage T1, histological grade I/II, lower mitotic score and luminal phenotype, whereas patients in cluster 2 constitute an intermediate group presenting both good and poor prognostic factors. Clusters defined by PCA k-means were significantly different for 5 characteristics: tumor stage, mitosis, tumor phenotype, Her-2 status and luminal. Clusters defined by Spectral Clustering were significantly different for 6 characteristics: tumor stage, histological grade, mitosis, Ki67, tumor phenotype and luminal. Clusters defined by Sparse k-means were significantly different for 4 characteristics: histological grade, tumor phenotype, Her-2 status and luminal. Clusters defined by SIMLR were significantly different for 6 characteristics: tumor stage, histological grade, mitosis, Ki67, tumor phenotype and luminal. Clusters defined by K-Sparse were significantly different for 6 characteristics: tumor stage, histological grade, mitosis, Ki67, tumor phenotype and luminal. From a strictly clinical point of view, Spectral clustering, SIMLR and K-sparse are the 3 most discriminating methods. Indeed, for these 3 methods, six prognostic factors (tumor stage, histological grade, mitosis score, Ki-67, tumor phenotype and luminal) were distributed significantly different between the 3 clusters.

3.2.4. Comparison of survival and predicted survival between clusters

None of the methods created clusters showing significant differences for OS, SS or RFS. Analysis of patients' simulated survival data

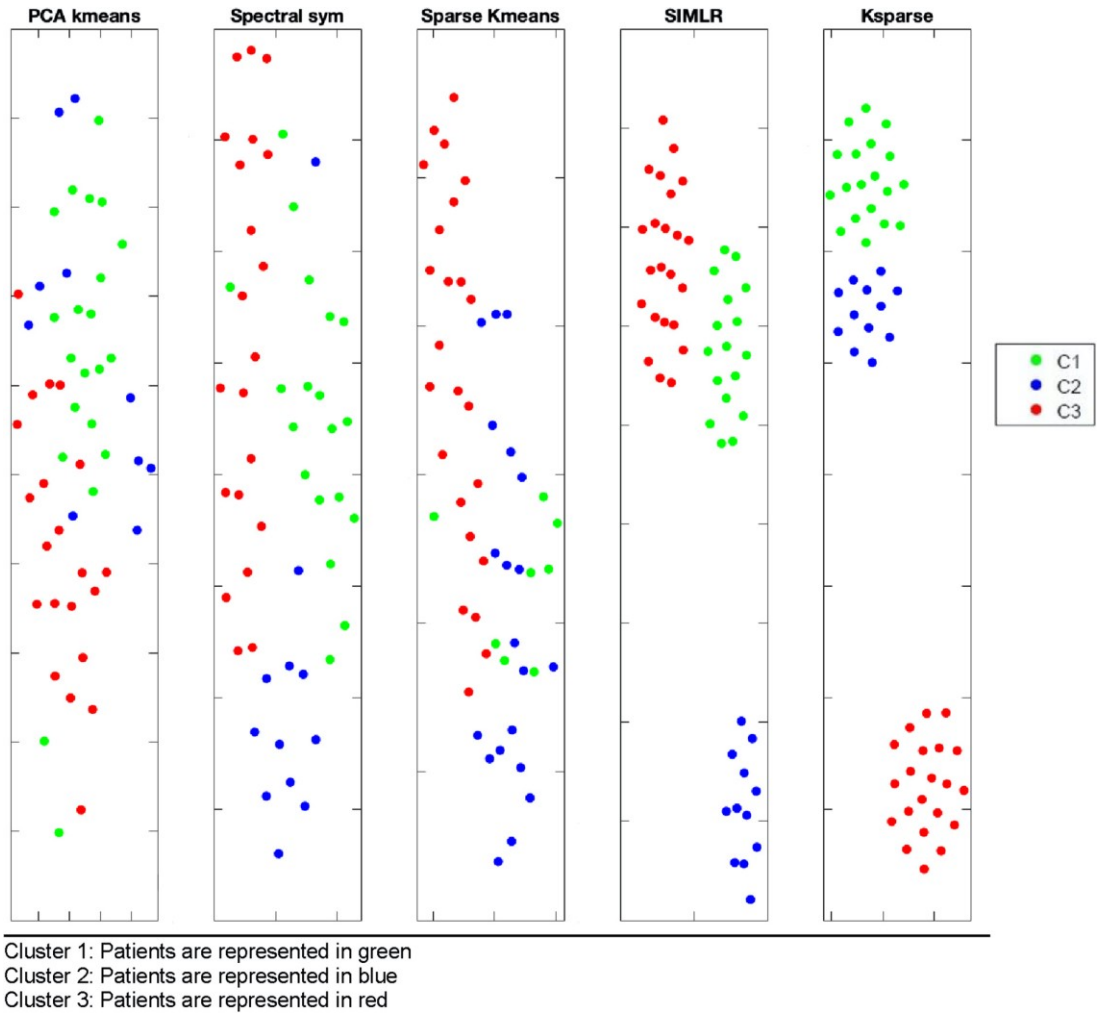


Fig. 1. Visualization of each cluster by clustering method using T-sne.

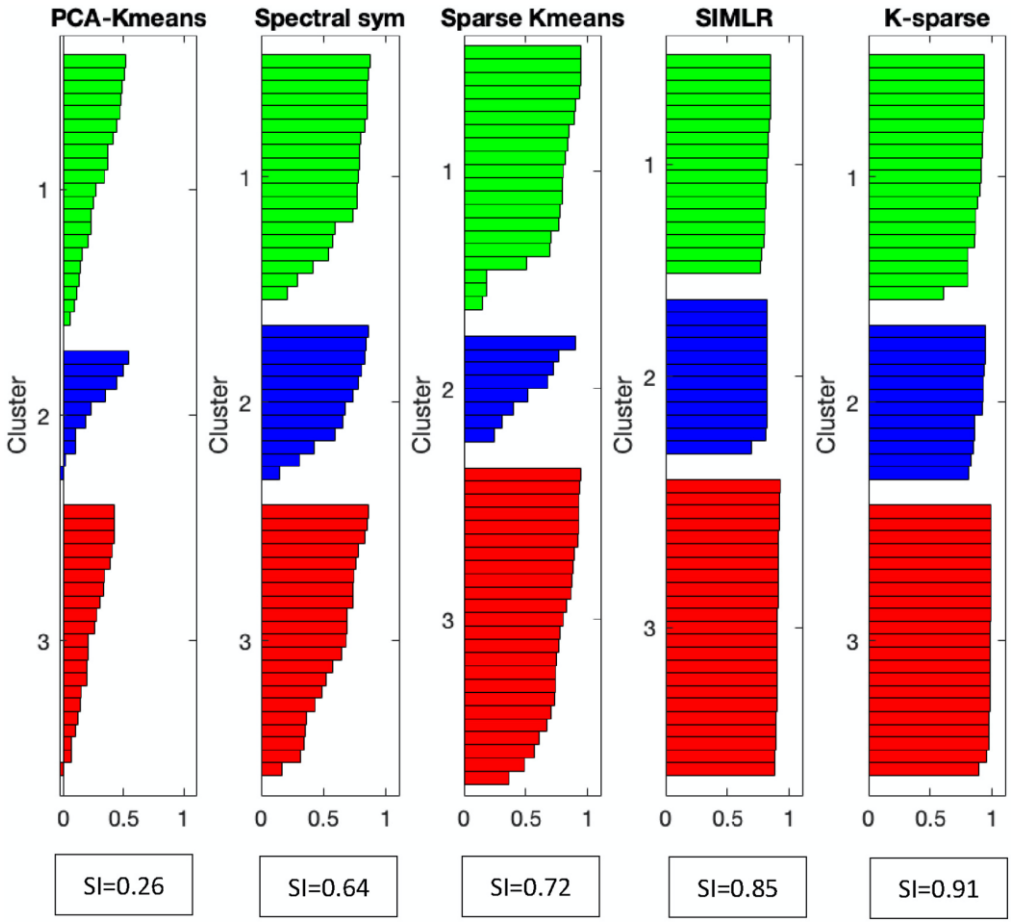


Fig. 2. Silhouette value (SI) representation for each patient by clustering method.

using PREDICT tool are presented in Table 3 and show a predicted survival gradient for clusters obtained with the 5 methods for OS and SS. There were significant differences for 5-year pOS between clusters obtained with K-sparse ($p = 0.021$), Sparse K-means ($p = 0.049$), Spectral and clustering ($p = 0.021$). The five methods showed a significant difference for 5-year pSS between clusters. In terms of 10-year pOS, there were no significant differences between clusters obtained by any of the 5 methods. In contrast, for 10-year pSS, the 5 methods showed significant differences between clusters. Patients in cluster 3 clearly showed the poorest predicted survival.

3.2.5. Comparison of the most impactful metabolites according to the five methods

To relate the impact of 449 metabolites to cluster construction, we ranked these metabolites extracted from each of the five methods based on their functional contributions to outputs. With this approach, we classified the relative impact of metabolites on cluster construction and on the identification of metabolic signatures. The highest-ranked metabolites were those that provided relevant information to the signature versus those that provided redundant information or no information. Among a total of 449 metabolites, 116 (26%) were selected by K-sparse clustering and 69 (15%) by Sparse K-means clustering. As for the three other methods, which don't select sparse features, the number of metabolites remained equal to 449. The 50 most effective metabolites identified by the five methods are presented in Supplementary Table 2. Furthermore, a comparison of the top 50 metabolites in each of the 5 methods is presented using a Venn diagram (Fig. 3). Two metabolites were shared by the 5 methods (Creatine, L-Proline), 9 were shared by 4 methods (Betaine, Glutathione, Humulinic Acid A, Isoleucyl-Methionine, L-Carnitine, L-Methionine, L-Phenylalanine Triethanolamine, Alnustone), 28 were shared by 3 methods and 38 were shared by 2 methods (Table 4).

3.2.6. Comparison between 5 methods of identified metabolic pathways

For a better understanding of metabolic dysregulation among BC subtypes, pathway analysis was performed. Identification of all the metabolic pathways highlighted by each of the 5 methods as shown in Supplementary Table 3. The most relevant pathways for each of the 5 methods are shown in Table 5. Sparse K-means identified only one statistically significant pathways, "cysteine and methionine metabolism", involved in amino acid metabolism. K-Sparse identified 3 different pathways: "glycerolipid metabolism", "Starch and sucrose metabolism" involved in carbohydrates metabolic pathway and "Aminoacyl-tRNA biosynthesis" involved in translation pathway. Spectral clustering identified 17 pathways, the 3 most important being "Glycine, serine and threonine metabolism", "Alanine, aspartate and glutamate metabolism" and "Histidine metabolism and glutathione metabolism" involved in amino acid metabolic pathway. PCA K-

Table 2
Clinical comparison of 52 patients between clusters.

Clinical characteristic	PCA-K-means				Spectral Clustering				Sparse K-means				SIMLR				K-Sparse			
	C1 (N = 21)	C2 (N = 10)	C3 (N = 21)	P-value	C2 (N = 19)	C1 (N = 12)	C3 (N = 21)	P-value	C1 (N = 24)	C2 (N = 8)	C3 (N = 20)	P-value	C1 (N = 17)	C2 (N = 12)	C3 (N = 23)	P-value	C1 (N = 19)	C2 (N = 12)	C3 (N = 21)	P-va
Age [a]	62.7 (15.2)	64.8(16)	62.9(15)	0.93	64.8 (14.3)	62.5 (16.5)	62 (15.3)	0.8	64.1(15)	60.5 (17.2)	63 (14.9)	0.85	64.3 (14.1)	64.9 (16.1)	61.4 (15.6)	0.755	64.8 (14.3)	62.5 (16.5)	62(15.3)	0.8
Histology type				1				0.392				0.106				0.752				0.3
Ductal carcinoma	19(90.5)	10(1 0 0)	19(90.5)		17(89.5)	11(91.7)	20(95.2)		21(87.5)	7(87.5)	20(1 0 0)		15(88.2)	12(1 0 0)	21(91.3)		17(89.5)	11(91.7)	20(95.2)	
Lobular carcinoma	2(9.5)	0(0)	1(4.8)		2(10.5)	1(8.3)	0(0)		3(12.5)	0(0)	0(0)		2(11.8)	0(0)	1(4.3)		2(10.5)	1(8.3)	0(0)	
Microinvasive carcinoma	0(0)	0(0)	1(4.8)		0(0)	0(0)	1(4.8)		0(0)	1(12.5)	0(0)		0(0)	0(0)	1(4.3)		0(0)	0(0)	1(4.8)	
Tumor stage				0.005				**0.018**				0.063				**0.045**				**0.0**
T1	14(66.7)	3(30)	4(19)		12(63.2)	5(41.7)	4(19)		14(58.3)	2(25)	5(25)		10(58.8)	6(50)	5(21.7)		12(63.2)	5(41.7)	4(19)	
T2/T3	7(33.3)	7(70)	17(81)		7(36.8)	7(58.3)	17(81)		10(41.7)	6(75)	15(75)		7(41.2)	6(50)	18(78.3)		7(36.8)	7(58.3)	17(81)	
Axillary lymph node				0.162				0.075				0.526				0.387				0.0
N0	14(66.7)	6(60)	8(38.1)		14(73.7)	6(50)	8(38.1)		15(62.5)	4(50)	9(45)		11(64.7)	7(58.3)	10(43.5)		14(73.7)	6(50)	8(38.1)	
N+	7(33.3)	4(40)	13(61.9)		5(26.3)	6(50)	13(61.9)		9(37.5)	4(50)	11(55)		6(35.3)	5(41.7)	13(56.5)		5(26.3)	6(50)	13(61.9)	
Metastasis				0.667				1				1				0.497				1
M0	21(1 0 0)	10(1 0 0)	19(90.5)		18(94.7)	12(1 0 0)	20(95.2)		23(96)	8(1 0 0)	19(95)		17(1 0 0)	12(1 0 0)	21(86.9)		18(94.7)	12(1 0 0)	20(95.2)	
M1	0(40)	0(0)	2(9.5)		1(5.3)	0(0%)	1(4.8)		1(4)	0(0%)	1(5)		0(0%)	0(0%)	2(13.1)		1(5.3)	0(0)	1(50)	
Histological grade				0.109				**0.025**				**0.008**				**0.007**				**0.0**
I/II	13(61.9)	7(70)	7(35)		12(63.2)	9(75)	6(30)		15(62.5)	5(71.4)	7(35)		11(64.7)	9(75)	7(31.8)		12(63.2)	9(75)	6(30)	
III	8(38.1)	3(30)	13(75)		7(36.8)	3(25)	14(70)		9(37.5)	2(28.6)	13(65)		6(35.3)	3(25)	15(68.2)		7(36.8)	3(25)	14(70)	
Mitosis				**0.024**				**0.016**				0.133				**0.005**				**0.0**
1	11(52.4)	4(40)	2(10)		10 (52.6)	5 (41.7)	2 (10)		11 (45.8)	2 (28.6)	4 (20)		10 (58.8)	5 (41.7)	2 (9.1)		10 (52.6)	5 (41.7)	2 (10)	
2	3(14.3)	4(40)	7(35)		3 (15.8)	5 (41.7)	6 (30)		4 (16.7)	4 (57.1)	6 (30)		2 (11.8)	5 (41.7)	7 (31.8)		3 (15.8)	5 (41.7)	6 (30)	
3	7(33.3)	2(20)	11(55)		6 (31.6)	2 (16.7)	10 (60)		9 (37.5)	1 (14.3)	10 (50)		5 (29.4)	2 (16.7)	13 (59.1)		6 (31.6)	2 (16.7)	12 (60)	
Ki67 [a]	25 (5,100)	27.5 (10,90)	60 (10,90)	0.066	41.1 (30.6)	33(22.6)	58.8 (27.2)	**0.027**	30 (19.2, 80)	35 (23.8, 45)	60 (28.8, 90)	0.196	38 (31)	32.8 (22.7)	59.7 (25.9)	**0.009**	41.1 (30.6)	33 (22.6)	58.8 (27.2)	**0.0**
Tumour phenotype				**0.024**				**0.012**				**0.006**				**0.018**				**0.0**
Her-2 over-expressed	1(4.8)	4(40)	7(33.3)		1(5.3)	4(33.3)	7(33.3)		2(8.3)	4(50)	6(30)		1(5.9)	4(33.3)	7(30.4)		1(5.3)	4(33.3)	7(33.3)	
Luminal	14(66.7)	5(50)	6(28.6)		13(68.4)	7(58.3)	5(23.8)		16(66.7)	4(50)	5(25)		12(70.6)	7(58.3)	6(26.1)		13(68.4)	7(58.3)	5(23.8)	
Triple-Negative	6(28.6)	1(10)	8(38.1)		5(26.3)	1(8.3)	9(42.9)		6(25)	0(0)	9(45)		4(23.5)	1(8.3)	10(43.5)		5(26.3)	1(8.3)	9(42.9)	
Hormonal receptors status				0.178				0.075				0.112				0.071				0.0
Negative	7(33.3)	5(50)	13(61.9)		6(31.6)	5(41.7)	14(66.7)		8(33.3)	4(50)	13(65)		5(29.4)	5(41.7)	15(65.2)		6(31.6)	5(41.7)	14(66.7)	
Positive	14(66.7)	5(50)	7(38.1)		13(68.4)	7(58.3)	7(33.3)		16(66.7)	4(50)	7(35)		12(70.6)	7(58.3)	8(34.8)		13(68.4)	7(58.3)	7(33.3)	
Her-2 status				**0.028**				0.061				**0.031**				0.115				0.0
Non-over-expressed	20(95.2)	6(60)	13(66.7)		18(94.7)	8(66.7)	14(66.7)		22(91.7)	4(50)	14(70)		16(94.1)	8(66.7)	16(69.6)		18(94.7)	6(66.7)	14(66.7)	
Over-expressed	1(4.8)	5(40)	6(33.3)		1(5.3)	4(33.3)	7(33.3)		2(8.3)	4(50)	6(30)		1(5.9)	4(33.3)	7(30.4)		1(5.3)	4(33.3)	7(33.3)	
Triple-Negative status				0.272				0.104				0.051				0.087				0.1
No	15(71.4)	9(90)	13(61.9)		14(73.7)	11(91.7)	12(57.1)		18(75)	8(1 0 0)	11(55)		13(76.5)	11(91.7)	13(56.5)		14(73.7)	11(91.7)	12(57.1)	
Yes	6(28.6)	1(10)	8(38.1)		5(26.3)	1(8.3)	9(42.9)		6(25)	0(0)	9(45)		4(23.5)	1(8.3)	10(43.5)		5(26.3)	1(8.3)	9(42.9)	
Luminal				**0.047**				**0.014**				**0.018**				**0.015**				**0.0**
No	7(33.3)	5(50)	15(71.4)		6(31.6)	5(41.7)	16(76.2)		8(33.3)	4(50)	15(75)		5(29.4)	5(41.7)	17(73.9)		6(31.6)	5(41.7)	16(76.2)	
Yes	14(66.7)	5(50)	6(28.6)		13(68.4)	7(58.3)	5(23.8)		16(66.7)	4(50)	5(25)		12(70.6)	7(58.3)	6(26.1)		13(68.4)	7(58.3)	5(28.8)	
Adjuvant Chemotherapy				0.52				0.423				0.459				0.459				0.4
No	7(33.3)	3(30)	4(19)		7(36.8)	2(16.7)	4(19)		6(25)	2(25)	5(25)		6(35.3)	3(25)	4(17.4)		7(36.8)	2(16.7)	4(19)	
Yes	14(85.7)	7(70)	17(81)		12(63.2)	10(83.3)	17(81)		18(75)	6(75)	1575)		11(64.7)	9(75)	19(82.6)		12(63.2)	10(83.3)	17(81)	
Adjuvant Radiotherapy				0.561				0.803				0.69				1				0.8
No	3(14.3)	3(30)	3(14.3)		3(15.8)	3(25)	3(14.3)		3(12.5)	2(25)	4(20)		3(17.6)	2(16.7)	4(17.4)		3(15.8)	3(25)	3(14.3)	
Yes	18(85.7)	7(70)	18(85.7)		16(84.2)	9(75)	18(85.7)		21(87.5)	6(75)	16(80)		14(82.4)	10(83.3)	19(82.6)		16(84.2)	9(75)	18(85.7)	

C1: cluster 1; C2: cluster 2; C3: cluster 3; [a]: mean (sd) or median (min, max).

Table 3
Comparison of prediction for overall and specific survival between clusters at 5 and 10-year.

		Predict 5-year				Predict 10-year			
		Overall Survival		Specific Survival		Overall Survival		Specific Survival	
Methods	No. of patients	% [95% CI]	P-value	% [95% CI]	P-value	% [95% CI]	P-value	% [95% CI]	P-value
K-sparse			**0.021**		**0.002**		**0.077**		**0.004**
	Cluster 1 (n = 19)	77% [67–82]		87% [80–91]		58% [48–65]		80% [73–86]	
	Cluster 2 (n = 12)	71% [57–82]		81% [69–90]		53% [38–66]		75% [60–85]	
	Cluster 3 (n = 20)	59% [47–69]		68% [60–74]		41% [29–52]		62% [53–69]	
SIMLR			0.1		**0.011**		0.241		**0.009**
	Cluster 1 (n = 17)	75% [64–82]		85% [77–91]		55% [45–64]		77% [65–84]	
	Cluster 2 (n = 12)	72% [56–82]		83% [69–91]		55% [40–67]		79% [65–87]	
	Cluster 3 (n = 22)	61% [50–70]		71% [63–77]		43% [32–53]		64% [55–70]	
Sparse K-means			**0.049**		**0.027**		0.203		**0.024**
	Cluster 1 (n = 24)	74% [64–80]		84% [76–89]		54% [43–63]		80% [73–86]	
	Cluster 2 (n = 7)	72% [58–87]		83% [70–94]		56% [37–72]		75% [60–85]	
	Cluster 3 (n = 20)	61% [49–69]		70% [61–78]		42% [32–52]		62% [53–69]	
Spectral clustering			**0.021**		**0.002**		0.077		**0.004**
	Cluster 1 (n = 19)	77% [68–83]		77% [80–91]		58% [48–65]		82% [73–86]	
	Cluster 2 (n = 12)	71% [57–81]		71% [69–90]		52% [32–64]		75% [60–85]	
	Cluster 3 (n = 20)	59% [47–68]		69% [60–76]		41% [29–52]		62% [53–69]	
PCA K-means			0.055		**0.009**		0.085		**0.008**
	Cluster 1 (n = 21)	77% [67–81]		86% [79–91]		58% [48–65]		79% [71–85]	
	Cluster 2 (n = 10)	69% [53–81]		80% [66–90]		52% [32–64]		77% [63–86]	
	Cluster 3 (n = 20)	60% [47–69]		69% [61–78]		41% [29–52]		63% [54–70]	

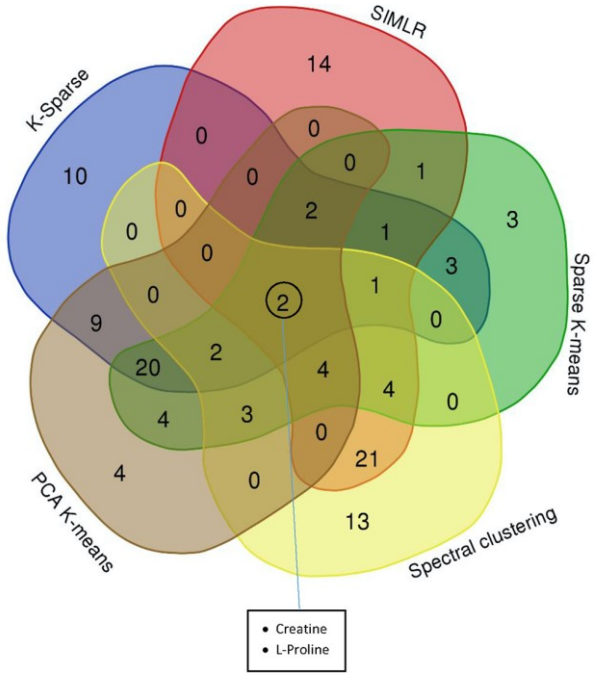


Fig. 3. Venn diagram of metabolic that were in common or unique to the five clustering methods.

means identified 10 pathways the 3 most important of which are "Alanine, aspartate and glutamate metabolism" involved in amino acid metabolic pathway, "Pyruvate metabolism" involved in carbohydrates metabolic/glucose oxidation pathway and "Citrate cycle (TCA cycle)" involved in energy metabolic pathway.

Finally, with 30 identified pathways, SIMLR is the method that identified the most metabolic pathways. Of these, the 3 most important highlighted metabolic pathways are "arginine and proline metabolism", "glycine, serine and threonine metabolism" and "alanine, aspartate and glutamate metabolism", involved in amino acid metabolic pathways. The Venn diagram (Fig. 4) shows the overlap of pathways detected by the five methods. Amino acid metabolism appeared to be the most frequently modified pathway. Enrichment and pathway analyses also showed modifications in glucose metabolism. From the biological point of view, SIMLR and spectral clustering are the two methods that identified the most relevant metabolic pathways.

3.2.7. Comparison of intensity of metabolites between the 5 methods

Among amino acid and glucose metabolisms, fourteen related metabolites were selected as potential biomarkers in BC [54–57]. As shown in Supplementary Fig. 4, the intensities of these 14 metabolites were compared between the 3 clusters for each of the 5 methods. The intensity of Uridine diphosphate (UDP) glucose, Guanine, L-Glutamine, L-Glutamic acid, L-Isoleucine, L-Proline, L-Methionine, L-Phenylalanine, Pyruvic acid, Spermine, Glutathione, Creatine, L-Carnitine and L-Acetylcarnitine were statistically significant between at least one of the clusters. The five methods agree that cluster 3 patients have low levels of Creatine, L-acetylcarnitine, L-Glutamic acid and high levels of Guanine, L-Isoleucine, L-Phenylalanine, Pyruvic acid and Spermine (Fig. 5). These metabolite levels seem to be predictive of poor prognosis [57–59].

4. Discussion

4.1. From a machine learning perspective

To the best of our knowledge, this proof-of-concept study is the first to compare different unsupervised ML methods to identify metabolomics-based prognostic signatures in BC. Analyses were performed intentionally without any prior clinical or biological assumptions. Clinical and biological interpretations were performed only after cluster identification. The objective of our study was to compare different unsupervised ML algorithms for feature selection from untargeted metabolomic data and to evaluate the capacity of these methods to select relevant features for further use in prediction models. This study did not seek to highlight significant differences but rather to assess how unsupervised methods might behave with high-dimension metabolic data and to open up new perspectives in the particularly active domain of BC

Table 4
Table indicating which metabolites are in each intersection or are unique to a certain list.

	Clustering Methods	Nbr	Metabolites
5	K-Sparse PCA K-means SIMLR Sparse K-means Spectral clustering	2	Creatine; L-Proline;
4	K-Sparse SIMLR Sparse K-means Spectral clustering	1	Triethanolamine;
	K-Sparse PCA K-means SIMLR Sparse K-means	2	L-Methionine; L-Phenylalanine
	K-Sparse PCA K-means Sparse K-means Spectral clustering	2	L-Carnitine; Betaine;
	PCA K-means SIMLR Sparse K-means Spectral clustering	4	Glutathione; Isoleucyl-Methionine; Humulinic acid A; Alnustone;
3	K-Sparse SIMLR Sparse K-means	1	Hydroxyprolyl-Valine;
	K-Sparse PCA K-means Sparse K-means	20	Aminoadipic acid; Methylmalonic acid; 1b-Furanoeudesm-4(15)-en-1-ol acetate; Glycerophosphocholine; Lidocaine; Adenosine monophosphate; 2-Methyl-3-ketovaleric acid; Liqcoumarin; p-Cresol sulfate; 2-Methylbutyroylcarnitine; Methoxsalen; Citramalic acid; Hypoxanthine; L-Acetylcarnitine; Ethyl aconitate; Guanine; L-Glutamic acid; Uridine 5′-monophosphate; N1,N12-Diacetylspermine; 5-Aminoimidazole ribonucleotide
	SIMLR Sparse K-means Spectral clustering	4	2,5-Dichloro-4-oxohex-2-enedioate; Histidinyl-Isoleucine; 3-(4-Methyl-3-pentenyl)thiophene; (−)-Epigallocatechin
	PCA K-means Sparse K-means Spectral clustering	3	L-Isoleucine; Ascorbic acid; Neurine;
2	K-Sparse Sparse K-means	3	5-Hydroxyisourate; Hexanoylcarnitine; L-Glutamine;
	K-Sparse PCA K-means	9	Creatinine; Proline; betaine; Erythronic acid; Garcinia acid; Thiolutin; 4-Chloro-1H-indole-3-acetic acid; Niacinamide 3-Dehydroxycarnitine; Dihydrothymine;
	SIMLR Spectral clustering	21	5b-Cyprinol sulfate; 2′,4-Dihydroxy-4′,6′-dimethoxychalcone; Propenoylcarnitine; 5-Hydroxyindoleacetic acid; Phaseolic acid Lisuride; 2-Bromophenol; (alpha-D-mannosyl)7-beta-D-mannosyl-diacetylchitobiosyl-L-asparagine isoform B (protein); Plastoquinone 3; 2,2,4,4,-Tetramethyl-6-(1-oxopropyl)-1,3,5-cyclohexanetrione; 1-Pyrroline; Gingerol; Prehumulinic acid; 1-Methylpyrrolo[1,2-a]pyrazine; 5-(methylthio)-2,3-Dioxopentyl phosphate; Propionic acid; Isosakuranin; Phenmetrazine; Methionine sulfoxide; Glycerol; Carboxyphosphamide
	SIMLR Sparse K-means	1	Phosphoric acid;
	PCA K-means Sparse K-means	4	I(−); L-Tyrosine; Gravelliferone; Valganciclovir;
1	K-Sparse	10	Prolylhydroxyproline; Guanidoacetic acid; Histamine; PC-M6; L-Histidine; N-Acetyl-L-aspartic acid; 3-Mercaptohexyl hexanoate; Trimethylamine N-oxide; Pantothenic acid; Flunitrazepam
	SIMLR	14	3-Hydroxy-6,8-dimethoxy-7(11)-eremophilen-12,8-olide; Glycerol tripropanoate; Alanyl-Isoleucine; 1-(2,4,6-Trimethoxyphenyl)-1,3-butanedione; 1-Oxo-1H-2-benzopyran-3-carboxaldehyde; 1,3,11-Tridecatriene-5,7,9-triyne; N-Acetyl-L-methionine; 3-Methyl sulfolene; 5-(4-Acetoxy-3-oxo-1-butynyl)-2,2′-bithiophene; Ac-Ser-Asp-Lys-Pro-OH; Cyclic AMP; Benzothiazole; (±)-2-Methylthiazolidine; 2-Methylcitric acid
	Spectral clustering	13	2,3-diketogulonate; 2,5-Furandicarboxylic acid; Pyrrolidine; Piperidine; Beta-Alanine; Aspartyl-L-proline; Erythro-5-hydroxy-L-lysinium(1 +); Acrylamide; 5-Hydroxylysine; S-Nitrosoglutathione; 2,2-dichloro-1,1-ethanediol; Valerenic acid; Dichloromethane
	Sparse K-means	3	Erinapyrone C; Ergothioneine; N-Methylethanolaminium phosphate
	PCA K-means	4	Dimethylglycine; Pipecolic acid; Methyl (9Z)-10′-oxo-6,10′-diapo-6-carotenoate; N-Desmethylvenlafaxine

phenotype predictors. We demonstrated that the K-sparse and SIMLR methods have a higher clustering performance compared with the three other popular unsupervised ML methods in detecting groups of patients with BC using metabolomic data. Interestingly, even though the spectral method is a little less clinically efficient than the k-sparse and SIMLR methods, it identified relevant metabolic pathways.

Our study suffers from various limitations, namely the relatively small number of patients and the monocentric and retrospective nature of the study. Besides, our results could not be validated on an external cohort. The clustering performances were assessed only by internal validation based on silhouette value. Indeed, we could not compare the labels obtained from our classification with the true labels to calculate the accuracy of the classification since the true labels were unknown.

Other unsupervised ML methods such as model-based clustering, bi-clustering and deep learning may be of value in this analysis and should be further explored. Yet it is worth noting that, even

Table 5
List of significant relevant pathways identified by 5 methods.

K-Sparse method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd[a]	Match Status[b]	Raw P^c	-log(p)	Impact[d]
C1 vs C3	UDP – glucose	Starch and sucrose metabolism	50	1	0,0107	4,5388	0,1390
	UDP – glucose	Amino sugar and nucleotide sugar metabolism	88	1	0,0107	4,5388	0,0928
	UDP - glucose; Glyceric acid	Glycerolipid metabolism	32	2	0,0153	4,1831	0,0206
SIMLR method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	P Value	-log(p)	Impact
C1 VS C2	Glutathione; Oxidized glutathione; Glycine; L-Glutamic acid; Pyroglutamic acid; Spermidine; Ornithine; Putrescine; Spermine; Cadaverine; Aminopropylcadaverine; Ascorbic acid	Glutathione metabolism	38	12	0	12,826	0,3628
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0	12,469	0,1383
	L-Tryptophan; N-Acetylserotonin; 5-Hydroxyindoleacetic acid; 2-Aminomuconic acid semialdehyde; 3-Hydroxyanthranilic acid; L-Kynurenine; Acetyl-N-formyl-5-methoxykynurenamine; Isophenoxazine;	Tryptophan metabolism	79	8	0,0001	9,1233	0,2741
	5′-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-phosphate; Pyruvic acid;	Cysteine and methionine metabolism	56	9	0,0008	7,1674	0,2509
	L-Glutamine; Phosphoribosylformylglycineamidine; Cyclic AMP; Adenosine monophosphate; Adenosine; Inosine; Adenine; Hypoxanthine; Guanine; Uric acid; 5-Hydroxyisourate; Guanosine; Adenosine diphosphate ribose; 5-Aminoimidazole ribonucleotide; Glyoxylic acid; Glycine; Adenosine 3′,5′-diphosphate;	Purine metabolism	92	17	0,0011	6,8091	0,2048
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0027	5,9281	0,268
	L-Glutamine; Ornithine; Citrulline; L-Arginine; L-Glutamic acid; N-Acetylornithine; L-Proline; Hydroxyproline; Guanidoacetic acid; Creatine; 4-Guanidinobutanoic acid; N2-Succinyl-L-ornithine; Putrescine; Spermidine; N-Acetylputrescine; Pyruvic acid; Glyoxylic acid; Spermine;	Arginine and proline metabolism	77	19	0,0053	5,238	0,6514
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,0075	4,8991	0,176
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,0076	4,8821	0,0394
	2-Hydroxyethanesulfonate; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0154	4,1754	0,0324
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; Pyruvic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,018	4,0154	0,46986
	Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; N-Acetyl-D-Glucosamine 6-Phosphate; Uridine diphosphate-N-acetylglucosamine; Cytidine monophosphate N-acetylneuraminic acid; D-Glucose; D-Xylose	Amino sugar and nucleotide sugar metabolism	88	7	0,0187	3,9783	0,1417
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0412	3,1903	0,3705
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0412	3,1898	0,0773
C1 VS C3	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0139	4,2752	0,3705
	Phenylpyruvic acid; L-Phenylalanine; L-Tyrosine; 3-Dehydroquinate; L-Tryptophan;	Phenylalanine, tyrosine and tryptophan biosynthesis	27	5	0,0189	3,9687	0,099
	L-Tryptophan; N-Acetylserotonin; 5-Hydroxyindoleacetic acid; 2-Aminomuconic acid semialdehyde; 3-Hydroxyanthranilic acid; L-Kynurenine; Acetyl-N-formyl-5-methoxykynurenamine; Isophenoxazine;	Tryptophan metabolism	79	8	0	16,409	0,2741
C2 VS C3	Glutathione; Oxidized glutathione; Glycine; L-Glutamic acid; Pyroglutamic acid; Spermidine; Ornithine; Putrescine; Spermine; Cadaverine; Aminopropylcadaverine; Ascorbic acid;	Glutathione metabolism	38	12	0	16,133	0,3628
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid	Ascorbate and aldarate metabolism	45	5	0	13,096	0,1383
	5′-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-	Cysteine and methionine	56	9	0,0001	9,8548	0,2509

(*continued on next page*)

Table 5 (*continued*)

SIMLR method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	P Value	-log(p)	Impact
	phosphate; Pyruvic acid;	metabolism					
	Phenylpyruvic acid; L-Phenylalanine; L-Tyrosine; 3-Dehydroquinate; L-Tryptophan;	Phenylalanine, tyrosine and tryptophan biosynthesis	27	5	0,0001	8,9814	0,099
	L-Histidine; L-Phenylalanine; L-Arginine; L-Glutamine; Glycine; L-Methionine; L-Lysine; L-Isoleucine; L-Threonine; L-Tryptophan; L-Tyrosine; L-Proline; L-Glutamic acid; Phosphoserine;	Aminoacyl-tRNA biosynthesis	75	14	0,0002	8,758	0,1127
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0004	7,7271	0,268
	L-Glutamine; Phosphoribosylformylglycineamidine; Cyclic AMP; Adenosine monophosphate; Adenosine; Inosine; Adenine; Hypoxanthine; Guanine; Uric acid; 5-Hydroxyisourate; Guanosine; Adenosine diphosphate ribose; 5-Aminoimidazole ribonucleotide; Glyoxylic acid; Glycine; Adenosine 3′,5′-diphosphate;	Purine metabolism	92	17	0,0007	7,306	0,2048
	Malonic acid; Beta-Alanine; Spermine; Spermidine; Dihydrouracil; Pantothenic acid; Uracil; L-Histidine	beta-Alanine metabolism	28	8	0,0012	6,7568	0,3577
	Uridine 5′-monophosphate; L-Glutamine; Dihydrouracil; Cytidine monophosphate; Cytidine; Cytosine; Uracil; Dihydrothymine; Uridine diphosphate glucose; Malonic acid; Ureidosuccinic acid; Beta-Alanine; Methylmalonic acid;	Pyrimidine metabolism	60	13	0,0014	6,5817	0,2756
	Pantothenic acid; Dihydrouracil; Beta-Alanine; Pyruvic acid; Adenosine 3′,5′-diphosphate; Uracil;	Pantothenate and CoA biosynthesis	27	6	0,0023	6,0879	0,2736
	L-Phenylalanine; Phenylpyruvic acid; Benzoic acid; Hippuric acid; Pyruvic acid; L-Tyrosine;	Phenylalanine metabolism	45	6	0,0072	4,9364	0,2468
	L-Glutamic acid; L-Glutamine; Oxoglutaric acid	D-Glutamine and D-glutamate metabolism	11	3	0,0124	4,39	0,139
	L-Glutamine; Ornithine; Citrulline; L-Arginine; L-Glutamic acid; N-Acetylornithine; L-Proline; Hydroxyproline; Guanidoacetic acid; Creatine; Creatinine; 4-Guanidinobutanoic acid; N2-Succinyl-L-ornithine; Putrescine; Spermidine; N-Acetylputrescine; Pyruvic acid; Glyoxylic acid; Spermine;	Arginine and proline metabolism	77	19	0,0169	4,082	0,6514
	2-Hydroxyethanesulfonate; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0215	3,8411	0,0324
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,0221	3,8108	0,4122
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0267	3,6235	0,0773
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid	Citrate cycle (TCA cycle)	20	3	0,0302	3,5015	0,176
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,0372	3,2914	0,4699
	Uridine diphosphate glucose; Glycerol 3-phosphate; Glycerol; Glyceric acid; Galactosylglycerol;	Glycerolipid metabolism	32	5	0,0427	3,1546	0,2162
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,0427	3,1536	0,0394

Sparse K-means method							
Clusters Comparaison	Interaction metabolite		Total Cmpd	Match Status	Raw p	-log(p)	Impact
C1 VS C2	L-Methionine; Glutathione	Cysteine and methionine metabolism	56	2	0.007	4.9	0.0454
C1 VS C3	L-Methionine; Glutathione;	Cysteine and methionine metabolism	56	2	0.0020	6.2	0.00454

Spectral clustering method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-log(p)	Impact
C1 VS C3	Iminoaspartic acid; Quinolinic acid; Niacinamide; Pyruvic acid; Propionic acid;	Nicotinate and nicotinamide metabolism	44	5	0,0024	6,0206	0,0712
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,0040	5,5100	0,4699

Table 5 (*continued*)

Spectral clustering method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-log(p)	Impact
	5′-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-phosphate; Pyruvic acid;	Cysteine and methionine metabolism	56	9	0,0098	4,6232	0,2509
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0101	4,5961	0,3705
	xoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,0171	4,0710	0,1760
	Pyruvic acid; L-Threonine; L-Isoleucine;	Valine, leucine and isoleucine biosynthesis	27	3	0,0178	4,0277	0,0350
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,0210	3,8609	0,0394
	D-Glucose; Glyceric acid; Pyruvic acid;	Pentose phosphate pathway	32	3	0,0232	3,7622	0,0218
	Pyruvic acid; L-Lactic acid; D-Glucose;	Glycolysis or Gluconeogenesis	31	3	0,0249	3,6928	0,0953
	Pyruvic acid; L-Lactic acid;	Pyruvate metabolism	32	2	0,0274	3,5955	0,3201
	L-Glutamic acid; Pyruvic acid; Butyric acid; Oxoglutaric acid;	Butanoate metabolism	40	4	0,0283	3,5644	0,0852
	2-Hydroxyethanesulfonate; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0287	3,5525	0,0324
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0303	3,4966	0,2680
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0,0330	3,4104	0,1383
	Epinephrine; Dopamine; L-Tyrosine; Homovanillic acid; Pyruvic acid;	Tyrosine metabolism	76	5	0,0385	3,2580	0,1750
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,0390	3,2431	0,4122
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0447	3,1074	0,0773
PCA K-means method							
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-log(p)	Impact
C1 vs C3	Iminoaspartic acid; Quinolinic acid; Niacinamide; Pyruvic acid; Propionic acid;	Nicotinate and nicotinamide metabolism	44	5	0,003	5,9412	0,0712
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,011	4,4865	0,1760
	Epinephrine; Dopamine; L-Tyrosine; Homovanillic acid; Pyruvic acid;	Tyrosine metabolism	76	5	0,024	3,7311	0,1750
	Pyruvic acid; L-Lactic acid;	Pyruvate metabolism	32	2	0,043	3,1507	0,3201
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,044	3,1214	0,0394
	Pyruvic acid; L-Threonine; L-Isoleucine;	Valine, leucine and isoleucine biosynthesis	27	3	0,045	3,1107	0,0350
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0,045	3,0926	0,1383
	L-Glutamic acid; Pyruvic acid; Butyric acid; Oxoglutaric acid;	Butanoate metabolism	40	4	0,046	3,0843	0,0852
	D-Glucose; Glyceric acid; Pyruvic acid;	Pentose phosphate pathway	32	3	0,046	3,0769	0,0218
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid	Alanine, aspartate and glutamate metabolism	24	7	0,048	3,0446	0,4122

[a] Total cmpd is the total number of compounds in the pathway.
[b] Hits is the actual matched number from the uploaded data.
[c] Raw p is the original *p*-value calculated from the pathway analysis.
[d] Impact is the pathway impact value calculated from pathway topology analysis.

though deep learning methods are of particular interest in many fields, they necessitate a very large number of patients to be efficiently trained and may therefore not be suitable for small metabolomics datasets obtained on real life patients, such as the one we have used. While obtaining imaging or clinical data concerning several thousands of patients seems achievable, obtaining metabolomics data for that many patients is currently much more complicated. Furthermore, even though some efforts are being made to tackle this issue [60], it is currently impossible to understand which features are responsible for the outcome when using deep-learning clustering techniques. It would therefore be impossible to understand the metabolic differences underlying different patient clusters if deep learning clustering was used.

These considerations raise important questions: in the future, on what basis should decisions be made? On results from a single method? Or on results provided by several methods? In view of the

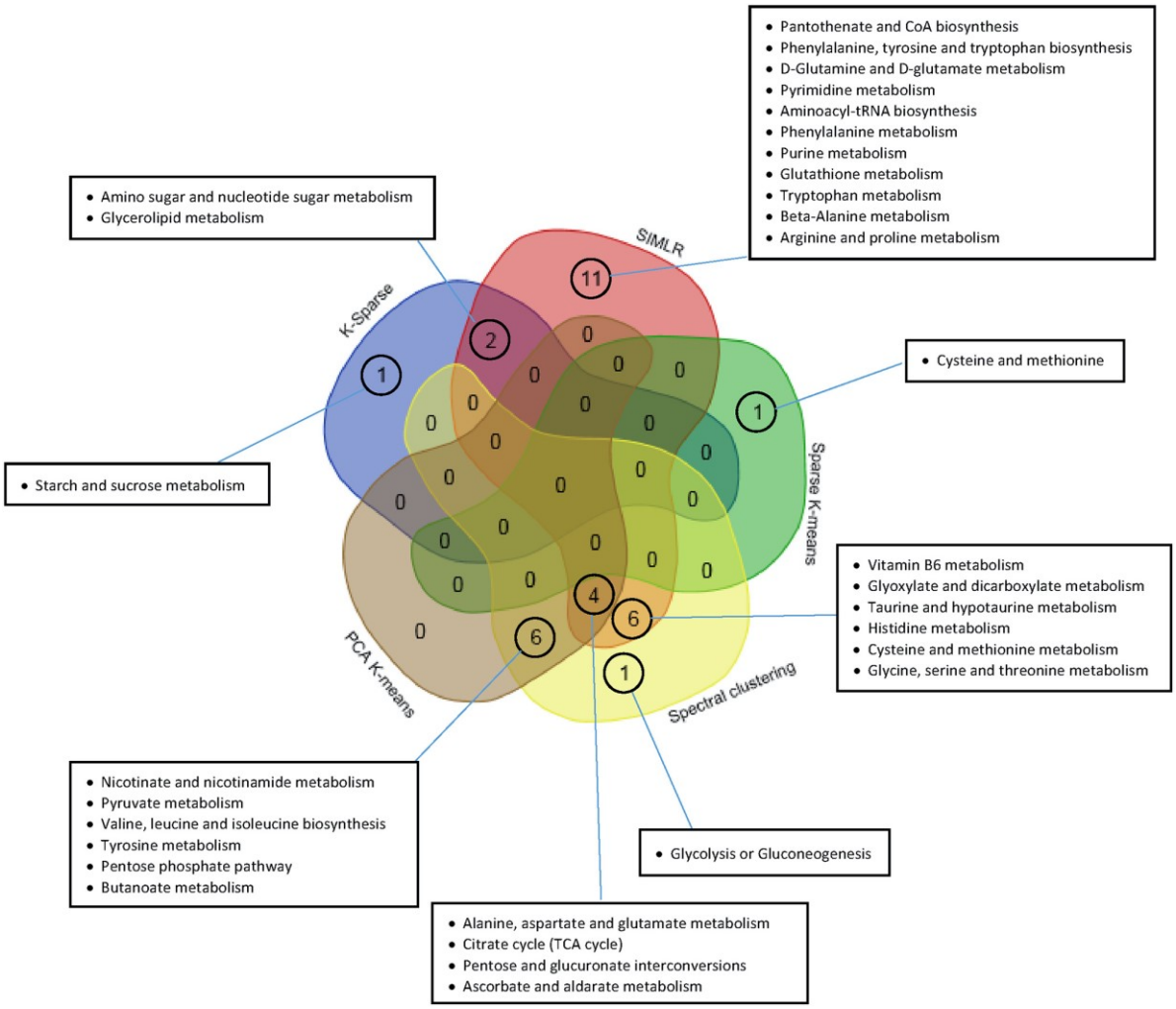


Fig. 4. Venn diagram of pathways that were in common or unique to the five clustering methods.

findings we have highlighted, it seems that decisions should be taken collegially, i.e. based on the results of a set of methods, as at multidisciplinary consultation meetings involving health professionals from different disciplines and whose skills are essential to take decisions ensuring patients the best possible care according to the state of the science.

4.2. From a clinical perspective

From a clinical point of view, the methods were able to highlight three distinct groups of patients with different clinical profiles. Patients identified in cluster 1 may be considered to have the best prognosis, patients in cluster 2 an intermediate prognosis, while patients in cluster 3 may be considered to have the worst prognosis. The results in Table 2 show that the tumors of patients in cluster 1 were predominantly non-invasive and non-proliferative, whereas the tumors of cluster 3 patients were mainly invasive and proliferative. Tumors in cluster 2 were rather invasive but not proliferative, hence the intermediate prognosis. We hypothesize that these patients would have an intermediate (atypical) biological profile, which is why the methods are discordant. We further evidence heterogeneity within the triple-negative BC subpopulation with most of the patients classified in cluster 3. However, a third of the triple-negative patients were in cluster 1 Recent molecular profiling studies of triple-negative BC using parallel sequencing and other "omics" technologies have also uncovered an unexpectedly high level of heterogeneity as well as a number of common features [61,62].

In addition, no significant difference between clusters could be demonstrated in terms of age, histologic type, lymph node involvement, metastasis or survival (OS, SS or RFS). Indeed, with a median follow-up of only 48.5 months, this duration is insufficient to demonstrate a significant difference in terms of OS, SS, or RFS. Nevertheless, it is quite easy to predict that patients in cluster 3 have the highest risk of progression and that, conversely, patients in cluster 1 have the lowest risk of progression. To confirm this intuition and try to reduce this short follow-up limitation, we analyzed simulated survival data obtained with the PREDICT tool. With a 5-year pOS rate at around 75% for cluster 1, 70% for cluster 2 and 60% for cluster 3, *in-silico* analyses have demonstrated their high potential value [28,63,64] and confirmed that patients in cluster 3 have a poorer prognosis [65,66]. One limitation of our study could be the

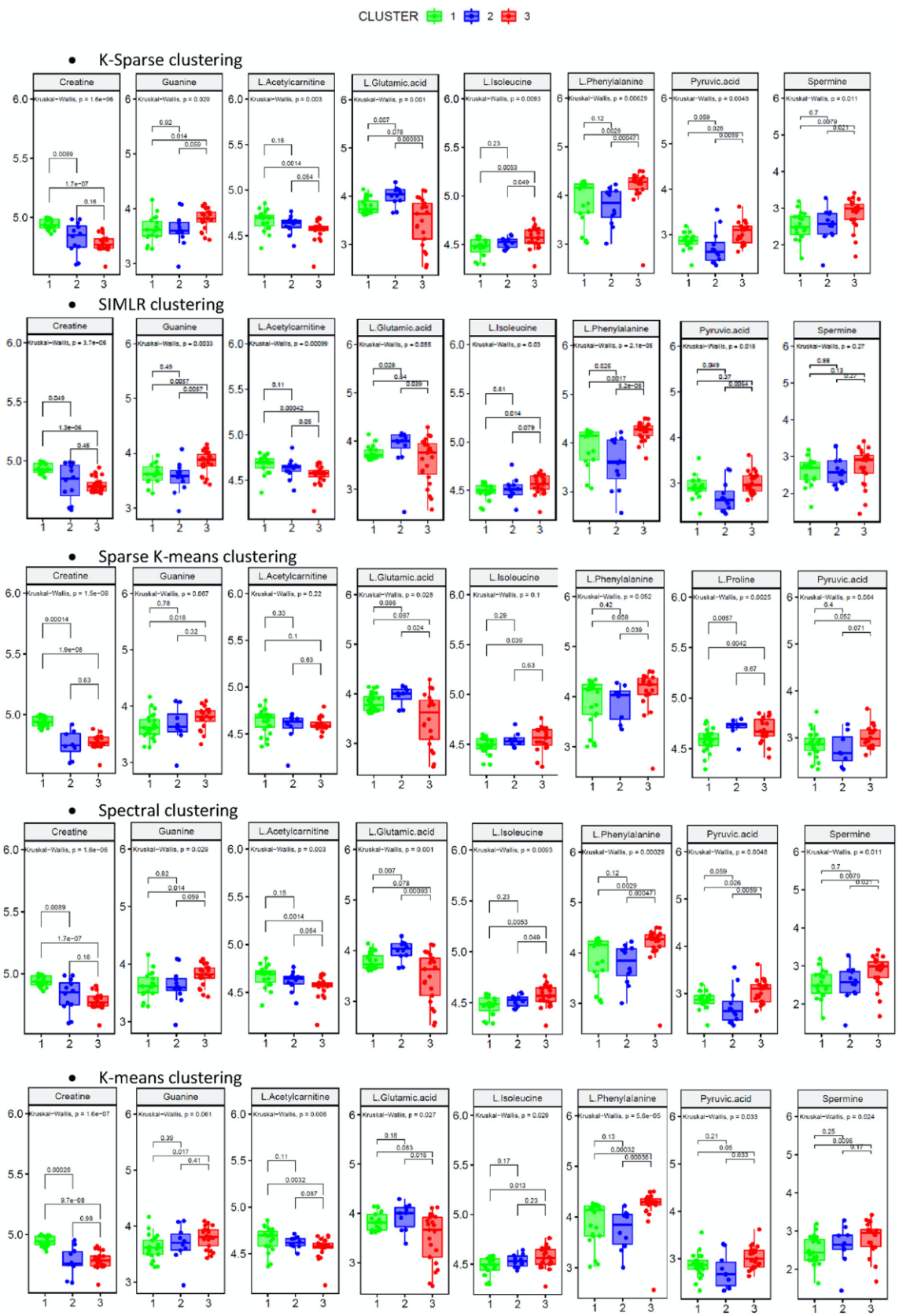


Fig. 5. Boxplot of the 8 metabolites extracted from 5 ML methods.

representativity of our population, e.g. it is recognized that BCs in younger patients (<40 years) are more aggressive [67]. Our study did not include a large number of young patients, which could explain why no significant difference was demonstrated in terms of age between clusters. Similarly, with only three patients with invasive lobular carcinoma (6%), our results did not identify a metabolic signature associated with this phenotype. Previous studies have shown a survival benefit in favor of invasive lobular carcinoma [68,69] and metabolomic studies focused on this particular type of BC could provide valuable biological information. Furthermore, due to the over-representation of hormonal-receptor negative tumors (48%) in our population compared to the literature [70], our population could have had unfavorable prognosis. This bias may result from our method of tumor selection. We decided to analyze frozen samples available in our biobank. Obviously, hormonal-receptor negative, triple-negative, Her-2-positive tumors are more often frozen and stored for further molecular testing and inclusion in clinical trials. In the present study, it is interesting to note that the five methods classified 73% of the patients in the same cluster. Among the 27% of patients classified differently by at least one of the methods, 9.5% of patients were classified heterogeneously by the five methods. Indeed, for each of these 5 patients, three methods classified them in one cluster and 2 others in another cluster without any connection between the types of methods used. Moreover, it is interesting to note that the different methods classified patients, on the one hand, in either the good prognostic cluster or the intermediate prognostic cluster or, on the other, in either the intermediate prognostic cluster or the poor prognostic cluster, but never in the good prognostic cluster or the poor prognostic cluster. A clinical analysis of these 5 patients showed that they had atypical clinical profiles, probably due to particular biological profiles. These atypical profiles would explain why no classification consensus could be highlighted. Overall, ML methods must remain a decision-making tool for the clinician, especially in cases where patients have particular clinical and biological characteristics. To avoid possible medical errors, the final responsibility for the decision lies with the clinician [71].

Finally, the initial clinical objective of this study was to define a metabolomic signature to refine the current classification and help the clinician in his chemotherapy prescription. This paper is the result of methodological research analyzing the best ML methods to develop this new tool. The patients selected were therefore patients eligible for adjuvant chemotherapy. An analysis of the metastatic population could help define a specific signature of metastatic status and/or a signature associated to survival. However, the use of biopsy faces two practical difficulties: 1) the intratumoral and inter-site heterogeneity that could be overcome through the analysis of blood or urine samples; and 2) the amount of material available once the pathologic analyses essential for patient management have been performed. Metabolomic analysis on paraffin slides could facilitate access to specimens and limit the amount of material required.

4.3. From a biological perspective

From a physiological point-of-view, this study extends the molecular stratification of BC to metabolomic profiles. Indeed, our results suggest that dysregulation of metabolic pathways exists between BC subtypes and that a particular amino acid profile characterizes the different BC histologic subtypes. Dysregulations of amino acid metabolism are well-known key events during cancer development [72] and are emerging hallmarks of cancers [73,74]. Amino acids serve not only as building blocks in protein synthesis but also as energy sources favoring cancer cell proliferation and growth [75]. Of interest, we identified significant differences between the BC subtypes of three metabolic pathways (i.e. Glycolysis and lactate production, Glutaminolysis, and amino acid) that play a pivotal role in BC growth [76,77]. Using the five methods, we consistently found that patients in cluster 3 showed higher levels of Guanine, L-Isoleucine, L Methionine, L-Phenylalanine, Pyruvic acid, Spermine and low levels of Creatine, L-Acetylcarnitine and L-Glutamic acid. Our results suggested that these metabolites could be candidate biomarker predictors of poorer prognosis [78–82]. All these results are consistent with the literature [57,83–86].

Given the exploratory nature of our study, we decided to use an FDR rate of 0.25 as a threshold in order to identify relevant candidate pathways (https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/FAQ).

A validation of these pathways, during a study whose main objective will be to evaluate the usefulness of our metabolomics signatures for decision-making, will need to be established with the use of a lower False Discovery Rate or Family Wise Error Rate (<0.05).

Indeed, to meet the biosynthetic needs associated with rapid proliferation, cancer cells must increase the import of nutrients. Two main metabolites are essential for biosynthesis and survival in mammalian cells, and particularly in cancer cells: glucose [87] and glutamine [88]. The increased glucose uptake in tumors compared to other healthy and non-proliferative tissues was first described more than 90 years ago by Otto Warburg [89]. Glucose is the primary energy source of all cells because of its involvement in many processes such as glycolysis or the Krebs cycle [90] in mitochondria. Unlike healthy cells that adapt to available substrates (glucose/fatty acids/proteins), some tumor cells are addicted to glucose. The other important point is that, once metabolized, tumor cells will prefer lactic fermentation to the Krebs cycle.

Lastly, the precise etiology of BC is still unknown even though some genetic, epigenetic and environmental factors have been identified [91]. It has been conclusively demonstrated that cancer cell metabolism is heavily influenced by microenvironmental factors, including nutrient availability. Sullivan and coworkers [92] found that diet affects local nutrient availability. This effect can lead to substantial changes in the metabolism of tumor cells, thereby modifying the response of these cells to drugs targeting metabolism. Drugs capable of inhibiting tumor proliferation may then become ineffective. Therefore, knowledge of microenvironmental nutrient levels is essential to a better understanding of tumor metabolism.

Outcomes for cancer patients vary greatly. The classification of BC into subtypes has been was defined in the literature on the basis of molecular characterization of proteomics (single omic). This has helped improve prognosis and personalized treatment. These considerations have motivated efforts to produce large amounts of multi-omic data such as TCGA [93] and ICGC [94]. However, current algorithms still face challenges and need to integrate omic data [95–98]. Defining BC subtypes using multi-omic data could help to better understand some of the dark areas that still persist in the field of tumor mechanisms in order to offer even more personalized treatments.

5. Conclusion

In the era of personalized medicine, OMICS science (genomics, transcriptomics, proteomics, and metabolomics) must contribute to the quest for cancer-specific biomarkers. The present study argues in favor of further research in this domain. Metabolomics is emerging as a relevant and promising tool for the classification of BC to enable more precise diagnosis [54,99–101]. Even though it is less accurate than the targeted approach, untargeted metabolomics nevertheless permits identification and quantification of a

vast number of major metabolites. Thus, this approach presents a particular interest in the search for new candidate biomarkers [102–104] and could be applied in everyday medical practice given that the cost and duration of metabolomic analyses are relatively low. However, due to the retrospective design of our study and the small number of patients recruited, our results need to be validated in a larger cohort and in the context of a prospective clinical trial.

Funding

The authors declare no competing financial interests.

CrediT authorship contribution statement

Jocelyn Gal: Methodology, Formal analysis, Writing - original draft. **Caroline Bailleux:** Writing - original draft. **David Chardin:** Software, Writing - original draft. **Thierry Pourcher:** Conceptualization, Writing - review & editing. **Julia Gilhodes:** . **Lun Jing:** . **Jean-Marie Guigonis:** Methodology, Writing - review & editing. **Jean-Marc Ferrero:** Data curation. **Gerard Milano:** Writing - review & editing. **Baharia Mograbi:** Writing - review & editing. **Patrick Brest:** Writing - review & editing. **Yann Chateau:** . **Olivier Humbert:** Conceptualization, Writing - review & editing. **Emmanuel Chamorey:** Supervision, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge support from the Centre Antoine Lacassagne, TIRO Unit, University Côte d'Azur and the Departmental Council of the Alpes Maritimes, France.

The authors sincerely thank Mrs. Clair Della Vedova for her help in developing the figures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.05.021.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin 2017;67:7–30.
[2] Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 1999;96:9212–7.
[3] Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000;405:827–36.
[4] Pandey A, Mann M. Proteomics to study genes and genomes. Nature 2000;405:837–46.
[5] Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
[6] Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
[7] Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100:8418–23.
[8] Witten DM, Tibshirani R. A framework for feature selection in clustering. J Am Stat Assoc 2010;105:713–26.
[9] Candido Dos Reis FJ, Wishart GC, Dicks EM, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. Breast Cancer Res 2017;19:58.
[10] Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res 2010;12:R1.
[11] Ross JS. Multigene predictors in early-stage breast cancer: moving in or moving out?. Expert Rev Mol Diagn 2008;8:129–35.
[12] Ross JS, Hatzis C, Symmans WF, et al. Commercialized multigene predictors of clinical outcome for breast cancer. Oncologist 2008;13:477–93.
[13] Buyse M, Loi S, van't Veer L,, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98:1183–92.
[14] Cao Y, DePinho RA, Ernst M, Vousden K. Cancer research: past, present and future. Nat Rev Cancer 2011;11:749–54.
[15] Ehmann F, Caneva L, Prasad K, et al. Pharmacogenomic information in drug labels: European Medicines Agency perspective. Pharmacogenomics J 2015;15:201–10.
[16] McShane LM, Polley MY. Development of omics-based clinical tests for prognosis and therapy selection: the challenge of achieving statistical robustness and clinical utility. Clin Trials 2013;10:653–65.
[17] van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.
[18] Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671–9.
[19] Wesolowski R, Ramaswamy B. Gene expression profiling: changing face of breast cancer classification and management. Gene Expr 2011;15:105–15.
[20] Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer?. Nat Rev Cancer 2012;12:323–34.
[21] Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. Cell 2008;134:703–7.
[22] McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141:210–7.
[23] Cannon WB. The wisdom of the body. 2nd ed. Oxford, England: Norton & Co.; 1939.
[24] Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted metabolomics. Curr Protoc Mol Biol 2012. Chapter 30: Unit 30 32 31-24.
[25] Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted metabolomics strategies-challenges and emerging directions. J Am Soc Mass Spectrom 2016;27:1897–905.
[26] Vinayavekhin N, Saghatelian A. Untargeted metabolomics. Curr Protoc Mol Biol 2010. Chapter 30: Unit 30 31 31-24.
[27] Camacho DM, Collins KM, Powers RK, et al. Next-generation machine learning for biological networks. Cell 2018;173:1581–92.
[28] Gal J, Milano G, Ferrero JM, et al. Optimizing drug development in oncology by clinical trial simulation: why and how? Brief Bioinform 2017.
[29] Yu MK, Ma J, Fisher J, et al. Visible machine learning for biomedicine. Cell 2018;173:1562–5.
[30] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015;349:255–60.
[31] Tang P, Tse GM. Immunohistochemical surrogates for molecular classification of breast carcinoma: A 2015 update. Arch Pathol Lab Med 2016;140:806–14.
[32] Katajamaa M, Oresic M. Processing methods for differential analysis of LC/MS profile data. BMC Bioinf 2005;6:179.
[33] Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinf 2010;11:395.
[34] Xia J, Mandal R, Sinelnikov IV, et al. MetaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis. Nucleic Acids Res 2012;40:W127–133.
[35] Irizarry RA, Wang C, Zhou Y, Speed TP. Gene set enrichment analysis made simple. Stat Methods Med Res 2009;18:565–75.
[36] Saxena A, Prasad M, Gupta A, et al. A review of clustering techniques and developments. Neurocomputing 2017;267:664–81.
[37] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J Royal Stat Soc: Series B (Statistical Methodol) 2001;63:411–23.
[38] Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons; 2009.
[39] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[40] Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979:224–7.
[41] Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat-Theory Methods 1974;3:1–27.
[42] Wang B, Zhu J, Pierson E, et al. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nat Methods 2017;14:414–6.
[43] Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics; 2007. p. 1027–35.
[44] Lloyd S. Least squares quantization in PCM. IEEE Trans. Inform. Theory 1982;28(2):129–37. https://doi.org/10.1109/TIT.1982.1056489.
[45] Steinhaus H. Sur la division des corps materiels en parties. Bull. Acad. Polon. Sci., C1. III 1956;IV:801–4.
[46] Ng AY, Jordan MI, Weiss Y. Analysis and an algorithm. In: Advances in neural information processing systems. On spectral clustering; 2002. p. 849–56.
[47] Von Luxburg U. A tutorial on spectral clustering. Stat Comput 2007;17:395–416.

[48] Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc: Ser B (Methodol) 1996:267–88.
[49] Gilet C, Deprez M, Caillau J-B, Barlaud M. Clustering with feature selection using alternating minimization, Application to computational biology. arXiv preprint arXiv:1711.02974 2017.
[50] Lvd Maaten, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–605.
[51] Team RCR. A language and environment for statistical. Computing 2013.
[52] Witten DM, Tibshirani R. sparcl: Perform sparse hierarchical clustering and sparse k-means clustering. R package version 2013;1.
[53] Wishart GC, Bajdik CD, Azzato EM, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. Eur J Surg Oncol 2011;37:411–7.
[54] Beger RD. A review of applications of metabolomics in cancer. Metabolites 2013;3:552–74.
[55] Gunther UL. Metabolomics biomarkers for breast cancer. Pathobiology 2015;82:153–65.
[56] McCartney A, Vignoli A, Biganzoli L, et al. Metabolomics in breast cancer: a decade in review. Cancer Treat Rev 2018;67:88–96.
[57] Silva C, Perestrelo R, Silva P, et al. Breast cancer metabolomics: from analytical platforms to multivariate data analysis. A Review. Metabolites 2019:9.
[58] Asiago VM, Alvarado LZ, Shanaiah N, et al. Early detection of recurrent breast cancer using metabolite profiling. Cancer Res 2010;70:8309–18.
[59] Cardoso MR, Santos JC, Ribeiro ML, et al. A Metabolomic approach to predict breast cancer behavior and chemotherapy response. Int J Mol Sci 2018;19.
[60] Karim MR, Beyan O, Zappa A, et al. Deep learning-based clustering approaches for bioinformatics. Brief Bioinform 2020.
[61] Bianchini G, Balko JM, Mayer IA, et al. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol 2016;13:674–90.
[62] Mills MN, Yang GQ, Oliver DE, et al. Histologic heterogeneity of triple negative breast cancer: A national cancer centre database analysis. Eur J Cancer 2018;98:48–58.
[63] Belkacemi Y, Hanna NE, Besnard C, et al. Local and regional breast cancer recurrences: salvage therapy options in the new era of molecular subtypes. Front Oncol 2018;8:112.
[64] Buonaguro FM, Caposio P, Tornesello ML, et al. Cancer diagnostic and predictive biomarkers 2018. Biomed Res Int 2019;2019:3879015.
[65] Ponde NF, Zardavas D, Piccart M. Progress in adjuvant systemic therapy for breast cancer. Nat Rev Clin Oncol 2018.
[66] Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26(Suppl 5):v8–30.
[67] Assi HA, Khoury KE, Dbouk H, et al. Epidemiology and prognosis of breast cancer in young women. J Thorac Dis 2013;5(Suppl 1):S2–8.
[68] Wang K, Zhu GQ, Shi Y, et al. Long-term survival differences between T1–2 invasive lobular breast cancer and corresponding ductal carcinoma after breast-conserving surgery: A propensity-scored matched longitudinal cohort study. Clin Breast Cancer 2019;19:e101–15.
[69] Wasif N, Maggard MA, Ko CY, Giuliano AE. Invasive lobular vs. ductal breast cancer: a stage-matched comparison of outcomes. Ann Surg Oncol 2010;17:1862–9.
[70] Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. World J Clin Oncol 2014;5:412–24.
[71] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
[72] Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab 2016;23:27–47.
[73] Hainaut P, Plymoth A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. Curr Opin Oncol 2013;25:50–1.
[74] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
[75] Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cell Mol Life Sci 2016;73:377–92.
[76] DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. Sci Adv 2016;2:e1600200.
[77] Haukaas TH, Euceda LR, Giskeodegard GF, Bathen TF. Metabolic portraits of breast cancer by HR MAS MR spectroscopy of intact tissue samples. Metabolites 2017:7.
[78] Jeon H, Kim JH, Lee E, et al. Methionine deprivation suppresses triple-negative breast cancer metastasis in vitro and in vivo. Oncotarget 2016;7:67223–34.
[79] Melone MAB, Valentino A, Margarucci S, et al. The carnitine system and cancer metabolic plasticity. Cell Death Dis 2018;9:228.
[80] Thomas TJ, Thomas T. Cellular and animal model studies on the growth inhibitory effects of polyamine analogues on breast cancer. Med Sci (Basel) 2018:6.
[81] Xiao F, Wang C, Yin H, et al. Leucine deprivation inhibits proliferation and induces apoptosis of human breast cancer cells via fatty acid synthase. Oncotarget 2016;7:63679–89.
[82] Zuo Y, Ulu A, Chang JT, Frost JA. Contributions of the RhoA guanine nucleotide exchange factor Net1 to polyoma middle T antigen-mediated mammary gland tumorigenesis and metastasis. Breast Cancer Res 2018;20:41.
[83] Lecuyer L, Dalle C, Lyan B, et al. Plasma metabolomic signatures associated with long-term breast cancer risk in the SU.VI.MAX prospective cohort. Cancer Epidemiol Biomarkers Prev 2019.
[84] Oikari S, Kettunen T, Tiainen S, et al. UDP-sugar accumulation drives hyaluronan synthesis in breast cancer. Matrix Biol 2018;67:63–74.
[85] Pan H, Xia K, Zhou W, et al. Low serum creatine kinase levels in breast cancer patients: a case-control study. PLoS One 2013;8:e62112.
[86] Phannasil P, Ansari IH, El Azzouny M, et al. Mass spectrometry analysis shows the biosynthetic pathways supported by pyruvate carboxylase in highly invasive breast cancer cells. Biochim Biophys Acta Mol Basis Dis 2017;1863:537–51.
[87] Mason EF, Rathmell JC. Cell metabolism: an essential link between cell growth and apoptosis. Biochim Biophys Acta 2011;1813:645–54.
[88] Hensley CT, Wasti AT, DeBerardinis RJ. Glutamine and cancer: cell biology, physiology, and clinical opportunities. J Clin Invest 2013;123:3678–84.
[89] Warburg O, Wind F, Negelein E. The metabolism of tumors in the body. J Gen Physiol 1927;8:519–30.
[90] Anderson NM, Mucka P, Kern JG, Feng H. The emerging role and targetability of the TCA cycle in cancer metabolism. Protein Cell 2018;9:216–37.
[91] Fernandez MF, Reina-Perez I, Astorga JM, et al. Breast Cancer and Its Relationship with the Microbiota. Int J Environ Res Public Health 2018;15.
[92] Sullivan MR, Danai LV, Lewis CA, et al. Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. Elife 2019:8.
[93] Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8.
[94] Zhang J, Baran J, Cros A, et al. International Cancer Genome Consortium Data Portal–a one-stop shop for cancer genomics data. Database (Oxford) 2011;2011:bar026.
[95] Mitra S, Saha S. A multiobjective multi-view cluster ensemble technique: Application in patient subclassification. PLoS One 2019;14:e0216904.
[96] Ramazzotti D, Lal A, Wang B, et al. Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival. Nat Commun 2018;9:4453.
[97] Rappoport N, Shamir R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res 2018;46:10546–62.
[98] Wu C, Zhou F, Ren J, et al. A selective review of multi-level omics data integration using variable selection. High Throughput 2019:8.
[99] Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. J Pharm Biomed Anal 2014;87:1–11.
[100] Bennett DA, Waters MD. Applying biomarker research. Environ Health Perspect 2000;108:907–10.
[101] Vermeersch KA, Styczynski MP. Applications of metabolomics in cancer research. J Carcinog 2013;12:9.
[102] Jacob M, Lopata AL, Dasouki M, Abdel Rahman AM. Metabolomics toward personalized medicine. Mass Spectrom Rev 2017.
[103] Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: The future of metabolomics in a personalized world. New Horiz Transl Med 2017;3:294–305.
[104] Wishart DS. Emerging applications of metabolomics in drug discovery and precision medicine. Nat Rev Drug Discov 2016;15:473–84.

Appendix A. Supplementary data

Supplementary File S1: Sample collection, preparation, and liquid chromatography-mass spectrometry analysis.

Tumor samples were collected during breast surgery and quickly stored at -80°C until analysis in our facility's biobank. Freeze-dried samples were processed by methanol extraction and analyzed by liquid chromatography-mass spectrometry (LC-MS) for metabolite characterization [1]. Liquid chromatographic analysis was performed using a DIONEX Ultimate 3000 HPLC system (Thermo Fisher Scientific). 10µL of each sample was injected onto a Synergi 4µm Hydro-RP 80 Å, 250 x 3.0 mm column (Phenomenex, Le Pecq, France). The mobile phases were composed of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The gradient was set as follows with a flow rate of 0.9 mL/min: 0% phase B from 0 to 5 min, 0-95% B from 5 to 21min, holding at 95% B to 21.5min, 95-0% B from 21.5 to 22min, holding at 0% B until 25min for column equilibration. Mass spectrometry analysis was carried out on a Q Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific) with a heated electrospray ionization source, HESI II, operating in both positive and negative modes. High-resolution accurate-mass full-scan MS and top 5 MS2 spectra were collected in a data-dependent fashion at a resolving power of 70 000 and 35 000 at m/z 400, respectively.
MSconvert (Version2.1, ProteoWizard) was used to convert raw data files obtained from LC-MS/MS to centroided mzXML files. The data collected from positive and negative ionization modes were analyzed separately using MzMine® (Version 2.38) [2, 3]. Isolated chromatograms were built for each mass with a noise threshold of 10^5. A local minimum search algorithm was used to select the validated peaks. Peaks were then aligned by RANSAC (random sample consensus) algorithm with a tolerance of 10 ppm in m/z and 1 min of retention time. Missing values were filled in using the same m/z and RT range as observed in detected samples, where possible. Only peaks with no missing values after gap-filling were kept. Peaks were then identified using the Human Metabolome DataBase (HMDB, version 3.0) by searching for $M+H^+$ and $M-H^+$ ion forms in positive and negative mode, respectivey, with 15ppm of mass tolerance. Linear normalization was performed using the average intensity in each sample as a normalization factor.

[1] Maharjan RP, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli, Anal Biochem 2003;313:145-154.
[2] Katajamaa M, Oresic M. Processing methods for differential analysis of LC/MS profile data, BMC Bioinformatics 2005;6:179.
[3] Pluskal T, Castillo S, Villar-Briones A et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics 2010;11:395.

Supplementary Table 1: Processing time's comparison between 5 clustering methods

	PCA K-means	SIMLR	Spectral clustering	Sparse K-means	K-sparse
Times (s)	0.04	0.25	0.30	0.31	0.48

Supplementary Table 2: The 50 most effective metabolites identified by the 5 ML methods

Importance	K-sparse	SIMLR	Sparse K-means	Spectral clustering	PCA K-means
1	Thiolutin	2,2,4,4,-Tetramethyl-6-(1-oxopropyl)-1,3,5-cyclohexanetrione	Creatine	Creatine	Creatine
2	L-Glutamic acid	Histidinyl-Isoleucine	Humulinic acid A	1-Methylpyrrolo[1,2-a]pyrazine	Ascorbic acid
3	Lidocaine	Humulinic acid A	Ascorbic acid	Humulinic acid A	L-Acetylcarnitine
4	1b-Furanoeudesm-4(15)-en-1-ol acetate	5b-Cyprinol sulfate	L-Proline	Histidinyl-Isoleucine	L-Proline
5	Citramalic acid	Isosakuranin	L-Carnitine	Ascorbic acid	Glutathione
6	2-Methyl-3-ketovaleric acid	Lisuride	Methoxsalen	Phaseolic acid	Humulinic acid A
7	Betaine	Prehumulinic acid	Glutathione	1-Pyrroline	L-Carnitine
8	Hexanoylcarnitine	Phosphoric acid	L-Phenylalanine	L-Proline	Lidocaine
9	L-Proline	5-(methylthio)-2,3-Dioxopentyl phosphate	L-Isoleucine	5b-Cyprinol sulfate	L-Phenylalanine
10	Flunitrazepam	Cyclic AMP	Betaine	2,5-Dichloro-4-oxohex-2-enedioate	N-Desmethylvenlafaxine
11	Liqcoumarin	Carboxyphosphamide	L-Acetylcarnitine	2,2,4,4,-Tetramethyl-6-(1-oxopropyl)-1,3,5-cyclohexanetrione	Methoxsalen
12	Methoxsalen	L-Phenylalanine	Lidocaine	Glycerol	I(-)
13	Ethyl aconitate	Plastoquinone 3	Neurine	Propionic acid	L-Isoleucine
14	Methylmalonic acid	Gingerol	Hypoxanthine	Triethanolamine	Betaine
15	Niacinamide	(-)-Epigallocatechin	Ethyl aconitate	2',4-Dihydroxy-4',6'-dimethoxychalcone	Aminoadipic acid
16	Guanine	Hydroxyprolyl-Valine	L-Glutamic acid	Isoleucyl-Methionine	Liqcoumarin
17	Pantothenic acid	Triethanolamine	Aminoadipic acid	Alnustone	Dihydrothymine
18	L-Methionine	Phaseolic acid	Isoleucyl-Methionine	Plastoquinone 3	Hypoxanthine
19	L-Glutamine	Glycerol	Alnustone	S-Nitrosoglutathione	Glycerophosphocholine
20	Uridine 5'-monophosphate	Propionic acid	5-Aminoimidazole ribonucleotide	Prehumulinic acid	1b-Furanoeudesm-4(15)-en-1-ol acetate
21	L-Carnitine	Propenoylcarnitine	1b-Furanoeudesm-4(15)-en-1-ol acetate	2-Bromophenol	Proline betaine
22	Histamine	Isoleucyl-Methionine	Triethanolamine	5-(methylthio)-2,3-Dioxopentyl phosphate	Citramalic acid
23	Dihydrothymine	Alnustone	Citramalic acid	Propenoylcarnitine	5-Aminoimidazole ribonucleotide
24	L-Phenylalanine	2,5-Dichloro-4-oxohex-2-enedioate	Liqcoumarin	Phenmetrazine	2-Methylbutyroylcarnitine
25	Triethanolamine	2',4-Dihydroxy-4',6'-dimethoxychalcone	L-Tyrosine	(-)-Epigallocatechin	Niacinamide
26	L-Histidine	1,3,11-Tridecatriene-5,7,9-triyne	Methylmalonic acid	Lisuride	L-Glutamic acid
27	N1,N12-Diacetylspermine	N-Acetyl-L-methionine	2-Methylbutyroylcarnitine	Acrylamide	Guanine
28	Glycerophosphocholine	1-Pyrroline	Guanine	Betaine	Erythronic acid
29	3-Dehydroxycarnitine	L-Methionine	p-Cresol sulfate	3-(4-Methyl-3-pentenyl)thiophene	L-Tyrosine
30	Adenosine monophosphate	Methionine sulfoxide	L-Methionine	(alpha-D-mannosyl)7-beta-D-mannosyl-diacetylchitobiosyl-L-asparagine, isoform B (protein)	Adenosine monophosphate
31	2-Methylbutyroylcarnitine	L-Proline	I(-)	Carboxyphosphamide	Methylmalonic acid
32	Creatine	3-Methyl sulfolene	L-Glutamine	L-Isoleucine	Neurine
33	N-Acetyl-L-aspartic acid	2-Methylcitric acid	2-Methyl-3-ketovaleric acid	Methionine sulfoxide	p-Cresol sulfate
34	5-Hydroxyisourate	1-Oxo-1H-2-benzopyran-3-carboxaldehyde	N1,N12-Diacetylspermine	Dichloromethane	Creatinine
35	Hydroxyprolyl-Valine	(±)-2-Methylthiazolidine	Gravelliferone	5-Hydroxylysine	Dimethylglycine
36	Aminoadipic acid	3-Hydroxy-6,8-dimethoxy-7(11)-eremophilen-12,8-olide	(-)-Epigallocatechin	Pyrrolidine	2-Methyl-3-ketovaleric acid
37	5-Aminoimidazole ribonucleotide	1-Methylpyrrolo[1,2-a]pyrazine	Glycerophosphocholine	Gingerol	Ethyl aconitate
38	4-Chloro-1H-indole-3-acetic acid	1-(2,4,6-Trimethoxyphenyl)-1,3-butanedione	Adenosine monophosphate	2,3-diketogulonate	N1,N12-Diacetylspermine
39	Creatinine	5-(4-Acetoxy-3-oxo-1-butynyl)-2,2'-bithiophene	Uridine 5'-monophosphate	2,2-dichloro-1,1-ethanediol	Thiolutin
40	PC-M6	Ac-Ser-Asp-Lys-Pro-OH	5-Hydroxyisourate	L-Carnitine	Uridine 5'-monophosphate
41	Trimethylamine N-oxide	Alanyl-Isoleucine	2,5-Dichloro-4-oxohex-2-enedioate	Erythro-5-hydroxy-L-lysinium(1+)	L-Methionine
42	3-Mercaptohexyl hexanoate	5-Hydroxyindoleacetic acid	Ergothioneine	Neurine	Gravelliferone
43	Prolylhydroxyproline	Benzothiazole	Hexanoylcarnitine	Isosakuranin	Pipecolic acid
44	p-Cresol sulfate	Glutathione	Histidinyl-Isoleucine	Aspartyl-L-proline	Methyl (9Z)-10'-oxo-6,10'-diapo-6-carotenoate
45	L-Acetylcarnitine	Phenmetrazine	Valganciclovir	Valerenic acid	4-Chloro-1H-indole-3-acetic acid
46	Hypoxanthine	Glycerol tripropanoate	Phosphoric acid	5-Hydroxyindoleacetic acid	Isoleucyl-Methionine
47	Garcinia acid	Creatine	Hydroxyprolyl-Valine	Beta-Alanine	Alnustone
48	Erythronic acid	2-Bromophenol	3-(4-Methyl-3-pentenyl)thiophene	Glutathione	3-Dehydroxycarnitine
49	Guanidoacetic acid	3-(4-Methyl-3-pentenyl)thiophene	Erinapyrone C	Piperidine	Garcinia acid
50	Proline betaine	(alpha-D-mannosyl)7-beta-D-mannosyl-diacetylchitobiosyl-L-asparagine, isoform B (protein)	N-Methylethanolaminium phosphate	2,5-Furandicarboxylic acid	Valganciclovir

Supplementary Table. 3: List of all pathways identified 5 methods

Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-*log*(p)	Impact
K-Sparse method							
C1 vs C2		Porphyrin and chlorophyll metabolism	104	2	0,0101	4,5996	0,0000
	L-Histidine; L-Phenylalanine; L-Glutamine; L-Methionine; L-Isoleucine; L-Threonine; L-Tyrosine; L-Proline; L-Glutamic acid; Phosphoserine;	Aminoacyl-tRNA biosynthesis	75	10	0,0453	3,0950	0,0563
C1 vs C3	UDP- glucose	Starch and sucrose metabolism	50	1	0,0107	4,5388	0,1390
	UDP- glucose	Amino sugar and nucleotide sugar metabolism	88	1	0,0107	4,5388	0,0928
	UDP- glucose	Galactose metabolism	41	1	0,0107	4,5388	0,0009
	UDP-glucose; Glyceric acid	Glycerolipid metabolism	32	2	0,0153	4,1831	0,0206
	Isoleucine; Methylmalonic acid	Valine, leucine and isoleucine degradation	40	2	0,0264	3,6350	0,0000
C2 vs C3	Formiminoglutamic acid; L-Glutamic acid; L-Histidine; Histamine; Ergothioneine;	Histidine metabolism	44	5	0,0299	3,5114	0,1977
	Triethanolamine; Glycerylphosphorylethanolamine; Glycerophosphocholine;	Glycerophospholipid metabolism	39	3	0,0366	3,3065	0,0263
	Pipecolic acid; Aminoadipic acid;	Lysine degradation	47	2	0,0490	3,0160	0,0161
SIMLR method							
Clusters Comparaison	**Interaction metabolite**	**Pathway Name**	**Total Cmpd**	**Match Status**	**Raw p**	**-*log*(p)**	**Impact**
C1 VS C2	Glutathione; Oxidized glutathione; Glycine; L-Glutamic acid; Pyroglutamic acid; Spermidine; Ornithine; Putrescine; Spermine; Cadaverine; Aminopropylcadaverine; Ascorbic acid;	Glutathione metabolism	38	12	0	12,826	0,3628
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0	12,469	0,1383
	L-Tryptophan; N-Acetylserotonin; 5-Hydroxyindoleacetic acid; 2-Aminomuconic acid semialdehyde; 3-Hydroxyanthranilic acid; L-Kynurenine; Acetyl-N-formyl-5-methoxykynurenamine; Isophenoxazine;	Tryptophan metabolism	79	8	0,0001	9,1233	0,2741
	5'-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-phosphate; Pyruvic acid;	Cysteine and methionine metabolism	56	9	0,0008	7,1674	0,2509
	L-Glutamine; Phosphoribosylformylglycineamidine; Cyclic AMP; Adenosine monophosphate; Adenosine; Inosine; Adenine; Hypoxanthine; Guanine; Uric acid; 5-Hydroxyisourate; Guanosine; Adenosine diphosphate ribose; 5-Aminoimidazole ribonucleotide; Glyoxylic acid; Glycine; Adenosine 3',5'-diphosphate;	Purine metabolism	92	17	0,0011	6,8091	0,2048
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0027	5,9281	0,268
	L-Glutamine; Ornithine; Citrulline; L-Arginine; L-Glutamic acid; N-Acetylornithine; L-Proline; Hydroxyproline; Guanidoacetic acid; Creatine; 4-	Arginine and proline metabolism	77	19	0,0053	5,238	0,6514

	Guanidinobutanoic acid; N2-Succinyl-L-ornithine; Putrescine; Spermidine; N-Acetylputrescine; Pyruvic acid; Glyoxylic acid; Spermine; Arginine; Ornithine;	D-Arginine and D-ornithine metabolism	8	2	0,0054	5,2304	0
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,0075	4,8991	0,176
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid; Glycine;5b-Cyprinol sulfate;	Pentose and glucuronate interconversions	53	4	0,0076	4,8821	0,0394
		Primary bile acid biosynthesis	47	2	0,0123	4,4004	0,0082
	2-Hydroxyethanesulfonate; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0154	4,1754	0,0324
	Glycerylphosphorylethanolamine; Glycerophosphocholine;	Ether lipid metabolism	23	2	0,0162	4,1223	0
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; Pyruvic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,018	4,0154	0,46986
	Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; N-Acetyl-D-Glucosamine 6-Phosphate; Uridine diphosphate-N-acetylglucosamine ; Cytidine monophosphate N-acetylneuraminic acid; D-Glucose; D-Xylose	Amino sugar and nucleotide sugar metabolism	88	7	0,0187	3,9783	0,1417
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0412	3,1903	0,3705
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0412	3,1898	0,0773
C1 VS C3	Isoleucine; Methylmalonic acid	Valine, leucine and isoleucine degradation	40	2	0,0067	4,9992	0
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0139	4,2752	0,3705
	L-Phenylalanine; L-Tyrosine; L-Tryptophan; L-Glutamic acid; L-Glutamine; L-Homocysteine; L-Histidine; Glycine; Adenosine monophosphate;	Nitrogen metabolism	39	9	0,0152	4,1851	0
	Phenylpyruvic acid; L-Phenylalanine; L-Tyrosine; 3-Dehydroquinate; L-Tryptophan;	Phenylalanine, tyrosine and tryptophan biosynthesis	27	5	0,0189	3,9687	0,099
	L-Histidine; L-Phenylalanine; L-Arginine; L-Glutamine; Glycine; L-Methionine; L-Lysine; L-Isoleucine; L-Threonine; L-Tryptophan; L-Tyrosine; L-Proline; L-Glutamic acid; Phosphoserine;	Aminoacyl-tRNA biosynthesis	75	14	0,0245	3,7085	0,1127
	L-Glutamine; Phosphoribosylformylglycineamidine; Cyclic AMP; Adenosine monophosphate; Adenosine; Inosine; Adenine; Hypoxanthine; Guanine; Uric acid; Guanosine; Adenosine diphosphate ribose; 5-Aminoimidazole ribonucleotide; Glyoxylic acid; Glycine; Adenosine 3',5'-diphosphate;	Purine metabolism	92	17	0,0328	3,4159	0,2048
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0434	3,1375	0,0773

	Glycerol 3-phosphate; Triethanolamine; Citicoline; O-Phosphoethanolamine; Glycerylphosphorylethanolamine; CDP-Ethanolamine; Glycerophosphocholine;	Glycerophospholipid metabolism	39	7	0,0491	3,0137	0,2464
C2 VS C3	L-Tryptophan; N-Acetylserotonin; 5-Hydroxyindoleacetic acid; 2-Aminomuconic acid semialdehyde; 3-Hydroxyanthranilic acid; L-Kynurenine; Acetyl-N-formyl-5-methoxykynurenamine; Isophenoxazine;	Tryptophan metabolism	79	8	0	16,409	0,2741
	Glutathione; Oxidized glutathione; Glycine; L-Glutamic acid; Pyroglutamic acid; Spermidine; Ornithine; Putrescine; Spermine; Cadaverine; Aminopropylcadaverine; Ascorbic acid;	Glutathione metabolism	38	12	0	16,133	0,3628
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid	Ascorbate and aldarate metabolism	45	5	0	13,096	0,1383
	5'-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-phosphate; Pyruvic acid;	Cysteine and methionine metabolism	56	9	0,0001	9,8548	0,2509
	L-Phenylalanine; L-Tyrosine; L-Tryptophan; L-Glutamic acid; L-Glutamine; L-Homocysteine; L-Histidine; Glycine; Adenosine monophosphate;	Nitrogen metabolism	39	9	0,0001	9,1978	0
	Phenylpyruvic acid; L-Phenylalanine; L-Tyrosine; 3-Dehydroquinate; L-Tryptophan;	Phenylalanine, tyrosine and tryptophan biosynthesis	27	5	0,0001	8,9814	0,099
	L-Histidine; L-Phenylalanine; L-Arginine; L-Glutamine; Glycine; L-Methionine; L-Lysine; L-Isoleucine; L-Threonine; L-Tryptophan; L-Tyrosine; L-Proline; L-Glutamic acid; Phosphoserine;	Aminoacyl-tRNA biosynthesis	75	14	0,0002	8,758	0,1127
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0004	7,7271	0,268
	Glycine;Cyprinol sulfate;	Primary bile acid biosynthesis	47	2	0,0006	7,4259	0,0082
	L-Glutamine; Phosphoribosylformylglycineamidine; Cyclic AMP; Adenosine monophosphate; Adenosine; Inosine; Adenine; Hypoxanthine; Guanine; Uric acid; 5-Hydroxyisourate; Guanosine; Adenosine diphosphate ribose; 5-Aminoimidazole ribonucleotide; Glyoxylic acid; Glycine; Adenosine 3',5'-diphosphate;	Purine metabolism	92	17	0,0007	7,306	0,2048
	Malonic acid; Beta-Alanine; Spermine; Spermidine; Dihydrouracil; Pantothenic acid; Uracil; L-Histidine	beta-Alanine metabolism	28	8	0,0012	6,7568	0,3577
	Uridine 5'-monophosphate; L-Glutamine; Dihydrouracil; Cytidine monophosphate; Cytidine; Cytosine; Uracil; Dihydrothymine; Uridine diphosphate glucose; Malonic acid; Ureidosuccinic acid; Beta-Alanine; Methylmalonic acid;	Pyrimidine metabolism	60	13	0,0014	6,5817	0,2756
	Pantothenic acid; Dihydrouracil; Beta-Alanine; Pyruvic acid; Adenosine 3',5'-diphosphate; Uracil;	Pantothenate and CoA biosynthesis	27	6	0,0023	6,0879	0,2736
	L-Phenylalanine; Phenylpyruvic acid; Benzoic acid; Hippuric acid; Pyruvic acid; L-Tyrosine;	Phenylalanine metabolism	45	6	0,0072	4,9364	0,2468
	L-Glutamic acid; L-Glutamine; Oxoglutaric acid	D-Glutamine and D-glutamate metabolism	11	3	0,0124	4,39	0,139
	Isoleucine;Methylmalonic acid	Valine, leucine and isoleucine degradation	40	2	0,015	4,1984	0

	L-Glutamine; Ornithine; Citrulline; L-Arginine; L-Glutamic acid; N-Acetylornithine; L-Proline; Hydroxyproline; Guanidoacetic acid; Creatine; Creatinine; 4-Guanidinobutanoic acid; N2-Succinyl-L-ornithine; Putrescine; Spermidine; N-Acetylputrescine; Pyruvic acid; Glyoxylic acid; Spermine;	Arginine and proline metabolism	77	19	0,0169	4,082	0,6514
	2-Hydroxyethanesulfonate ; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0215	3,8411	0,0324
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,0221	3,8108	0,4122
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0267	3,6235	0,0773
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid	Citrate cycle (TCA cycle)	20	3	0,0302	3,5015	0,176
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,0372	3,2914	0,4699
	Uridine diphosphate glucose; Glycerol 3-phosphate; Glycerol; Glyceric acid; Galactosylglycerol;	Glycerolipid metabolism	32	5	0,0427	3,1546	0,2162
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,0427	3,1536	0,0394

Sparse Kmeans method

Clusters Comparaison	Interaction metabolite		Total Cmpd	Match Status	Raw p	-*log*(p)	Impact
C1 VS C2	L-Methionine;Glutathione	Cysteine and methionine metabolism	56	2	0.007	4.9	0.0454
C1 VS C3							
C1 VS C3	L-Methionine;Glutathione;	Cysteine and methionine metabolism	56	2	0.0020	6.2	0.00454

Spectral clustering method

Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-*log*(p)	Impact
C1 VS C2	Glycine; L-Glutamic acid; L-Threonine;	Porphyrin and chlorophyll metabolism	104	3	0,0240	3,7290	0,0000
	Uridine diphosphate glucose; Glycerol 3-phosphate; Glycerol; Glyceric acid; Galactosylglycerol;	Glycerolipid metabolism	32	5	0,0334	3,4003	0,2162
	Galactitol; Galactosylglycerol; Uridine diphosphate glucose; D-Galactonate; D-Glucose; Glycerol;	Galactose metabolism	41	6	0,0363	3,3154	0,1539
C1 VS C3	Iminoaspartic acid; Quinolinic acid; Niacinamide; Pyruvic acid; Propionic acid;	Nicotinate and nicotinamide metabolism	44	5	0,0024	6,0206	0,0712
	Glyceric acid; Betaine; Guanidoacetic acid; Dimethylglycine; Glycine; Phosphoserine; L-Threonine; O-Phosphohomoserine; L-Aspartyl-4-phosphate; Creatine; Glyoxylic acid; L-Tryptophan	Glycine, serine and threonine metabolism	48	13	0,0040	5,5100	0,4699

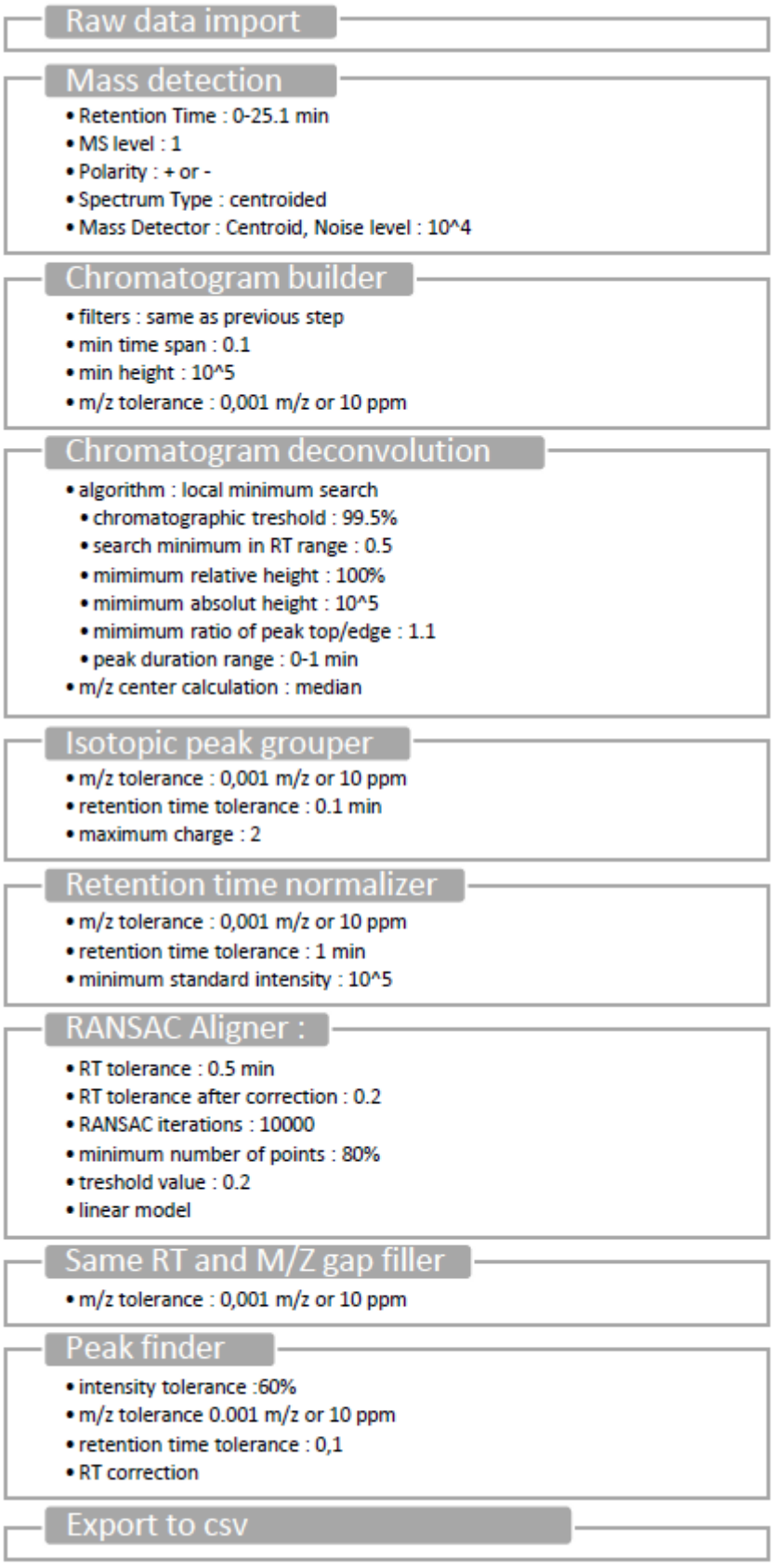
	5'-Methylthioadenosine; N-Formyl-L-methionine; L-Homocysteine; L-Methionine; Glutathione; Phosphoserine; 3-Sulfinoalanine; L-Aspartyl-4-phosphate; Pyruvic acid;	Cysteine and methionine metabolism	56	9	0,0098	4,6232	0,2509
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,0101	4,5961	0,3705
	xoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,0171	4,0710	0,1760
	Pyruvic acid; L-Threonine; L-Isoleucine;	Valine, leucine and isoleucine biosynthesis	27	3	0,0178	4,0277	0,0350
	Pyruvic acid;	Terpenoid backbone biosynthesis	33	1	0,0207	3,8797	0,0000
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate; Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,0210	3,8609	0,0394
	D-Glucose; Glyceric acid; Pyruvic acid;	Pentose phosphate pathway	32	3	0,0232	3,7622	0,0218
	Pyruvic acid; L-Lactic acid; D-Glucose;	Glycolysis or Gluconeogenesis	31	3	0,0249	3,6928	0,0953
	Pyruvic acid; L-Lactic acid;	Pyruvate metabolism	32	2	0,0274	3,5955	0,3201
	L-Glutamic acid; Pyruvic acid; Butyric acid; Oxoglutaric acid;	Butanoate metabolism	40	4	0,0283	3,5644	0,0852
	2-Hydroxyethanesulfonate; Pyruvic acid; 3-Sulfinoalanine;	Taurine and hypotaurine metabolism	20	3	0,0287	3,5525	0,0324
	Glyoxylic acid; Oxoglutaric acid; N-Formyl-L-methionine; Glycolic acid; Glyceric acid; Pyruvic acid;	Glyoxylate and dicarboxylate metabolism	50	6	0,0303	3,4966	0,2680
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0,0330	3,4104	0,1383
	Epinephrine; Dopamine; L-Tyrosine; Homovanillic acid; Pyruvic acid;	Tyrosine metabolism	76	5	0,0385	3,2580	0,1750
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,0390	3,2431	0,4122
	Pyridoxamine; Oxoglutaric acid; 3-Hydroxy-2-methylpyridine-4,5-dicarboxylate; Pyruvic acid;	Vitamin B6 metabolism	32	4	0,0447	3,1074	0,0773
C2 VS C3	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,0209	3,8659	0,4122
	L-Glutamic acid; L-Glutamine; Oxoglutaric acid	D-Glutamine and D-glutamate metabolism	11	3	0,0275	3,5922	0,1390

PCA-Kmeans method

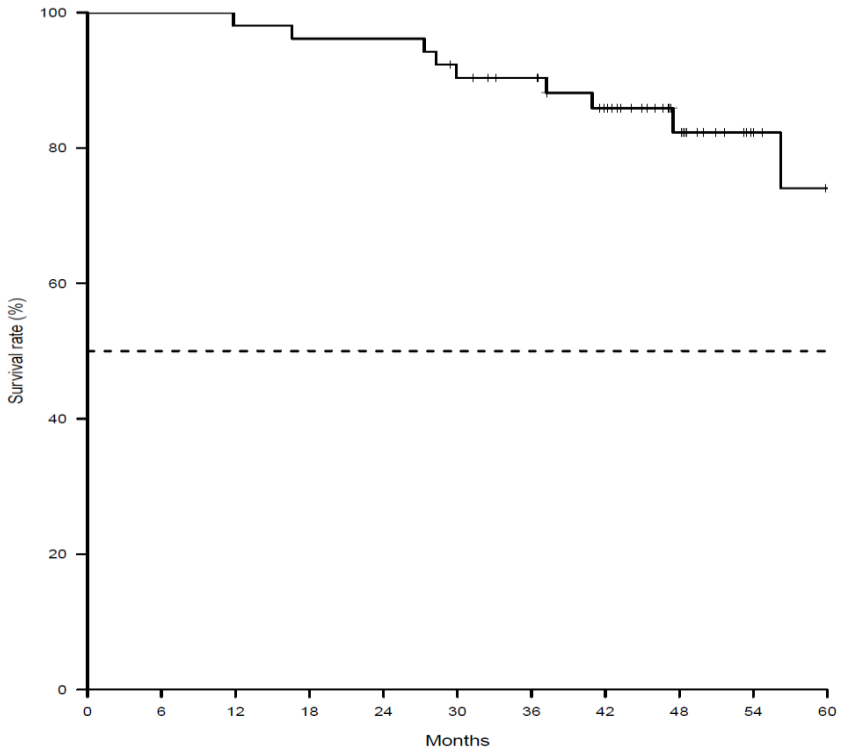
Clusters Comparaison	Interaction metabolite	Pathway Name	Total Cmpd	Match Status	Raw p	-*log*(p)	Impact
C1 vs C2	Iminoaspartic acid; Quinolinic acid; Niacinamide; Pyruvic acid; Propionic acid;	Nicotinate and nicotinamide metabolism	44	5	0,020	3,9278	0,0712
	Galactitol; Galactosylglycerol; Uridine diphosphate glucose; D-Galactonate; D-Glucose; Glycerol;	Galactose metabolism	41	6	0,025	3,6893	0,1539
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0,046	3,0797	0,1383
	Formiminoglutamic acid; L-Glutamic acid; Urocanic acid; L-Histidine; Histamine; D-Erythro-imidazole-glycerol-phosphate; Ergothioneine; Hydantoin-5-propionic acid; Imidazole acetol-phosphate; Oxoglutaric acid;	Histidine metabolism	44	10	0,048	3,0407	0,3705

	L-Glutamic acid; L-Glutamine; Oxoglutaric acid	D-Glutamine and D-glutamate metabolism	11	3	0,049	3,0236	0,1390
C1 vs C3	Iminoaspartic acid; Quinolinic acid; Niacinamide; Pyruvic acid; Propionic acid;	Nicotinate and nicotinamide metabolism	44	5	0,003	5,9412	0,0712
	Oxoglutaric acid; Oxalosuccinic acid; Pyruvic acid;	Citrate cycle (TCA cycle)	20	3	0,011	4,4865	0,1760
	Epinephrine; Dopamine; L-Tyrosine; Homovanillic acid; Pyruvic acid;	Tyrosine metabolism	76	5	0,024	3,7311	0,1750
	Pyruvic acid;	Terpenoid backbone biosynthesis	33	1	0,031	3,4834	0,0000
	Pyruvic acid;L-Lactic acid;	Pyruvate metabolism	32	2	0,043	3,1507	0,3201
	D-Xylose; Uridine diphosphate glucose; D-Glucuronic acid 1-phosphate ;Pyruvic acid;	Pentose and glucuronate interconversions	53	4	0,044	3,1214	0,0394
	Pyruvic acid; L-Threonine; L-Isoleucine;	Valine, leucine and isoleucine biosynthesis	27	3	0,045	3,1107	0,0350
	Ascorbic acid; Uridine diphosphate glucose; Pyruvic acid; D-Glucuronic acid 1-phosphate; Oxoglutaric acid;	Ascorbate and aldarate metabolism	45	5	0,045	3,0926	0,1383
	L-Glutamic acid; Pyruvic acid; Butyric acid; Oxoglutaric acid;	Butanoate metabolism	40	4	0,046	3,0843	0,0852
	D-Glucose; Glyceric acid; Pyruvic acid;	Pentose phosphate pathway	32	3	0,046	3,0769	0,0218
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,048	3,0446	0,4122
C2 vs C3	L-Glutamic acid; L-Glutamine; Oxoglutaric acid	D-Glutamine and D-glutamate metabolism	11	3	0,012	4,4588	0,1390
	N-Acetyl-L-aspartic acid; Pyruvic acid; Ureidosuccinic acid; Oxoglutaric acid; L-Glutamine; L-Glutamic acid; 2-Keto-glutaramic acid;	Alanine, aspartate and glutamate metabolism	24	7	0,046	3,0796	0,4122

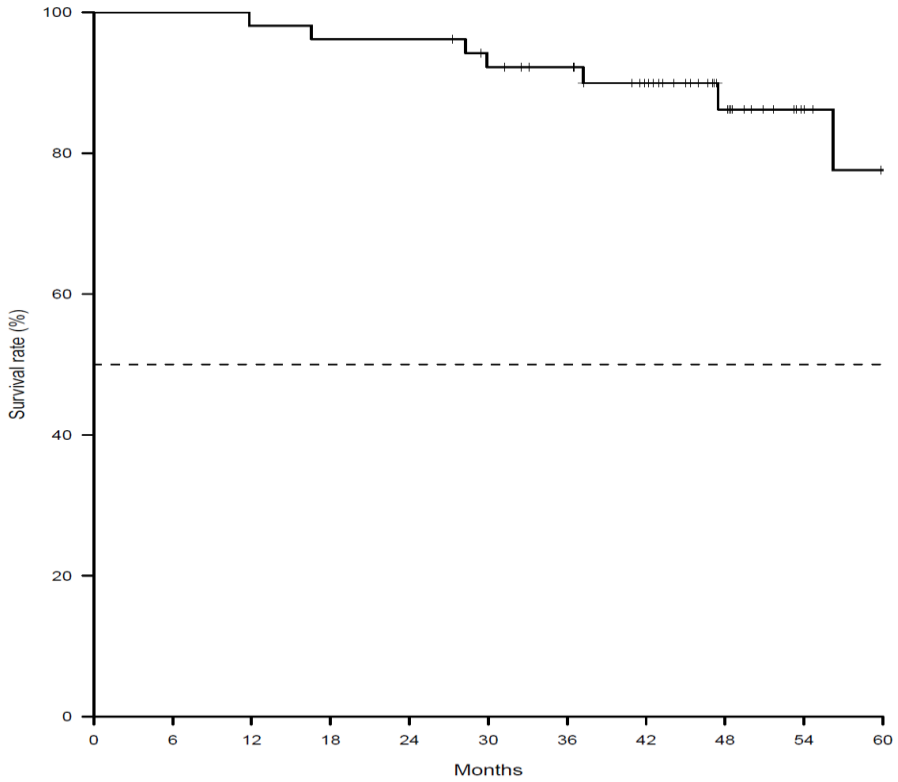
Supplementary Fig. S1: Protocol used in MZmine for the treatment of metabolomic data



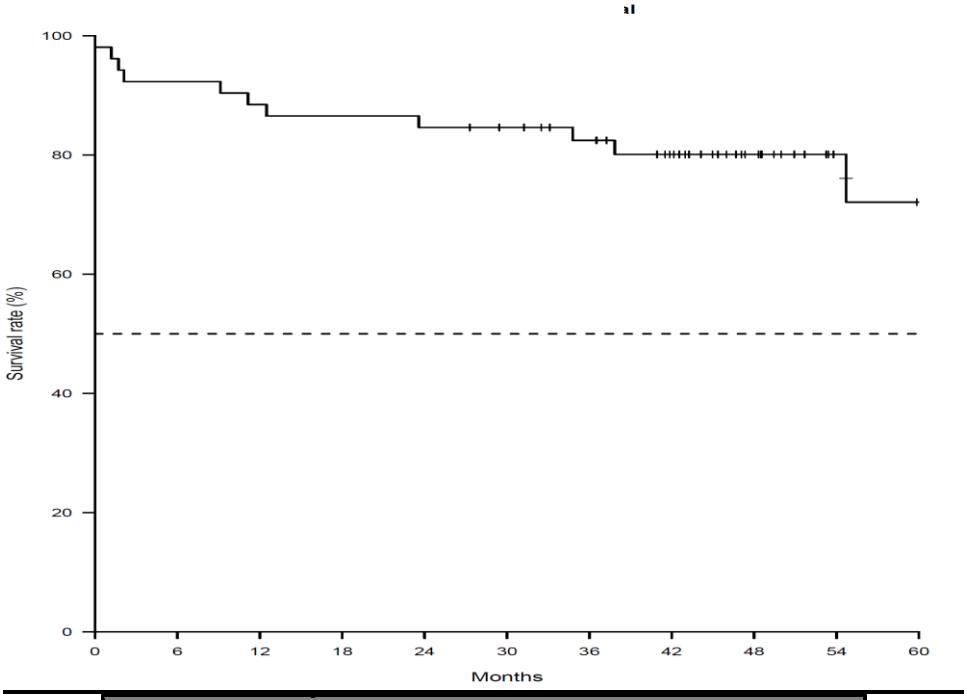
Supplementary Fig. 2A: Overall survival



Supplementary Fig. 2B: Specific survival



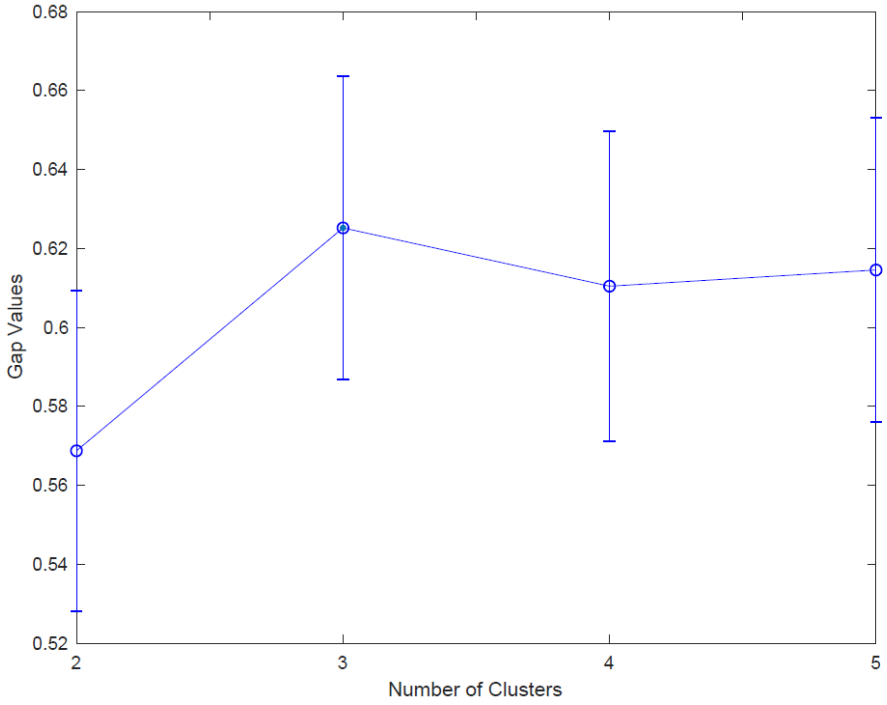
Supplementary Fig. 2C: Recurrence free survival



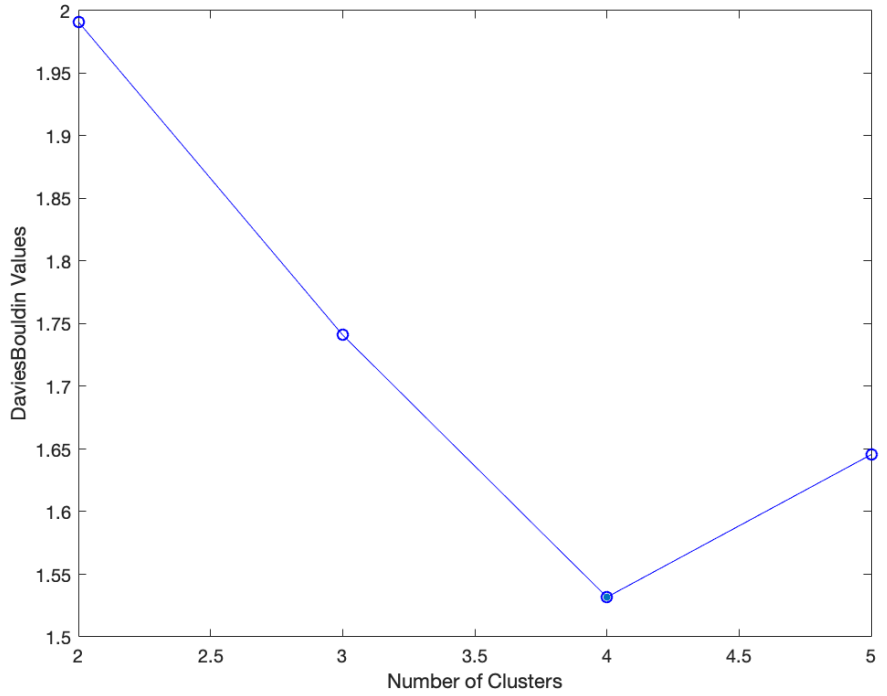
	Months					
time	0	12	24	36	48	60
n.risk	52	51	50	43	23	8
n.event	1	5	2	1	1	1
surv	98.1	85.5	84.6	82.4	80.1	72.1
IC 95% lower	94.4	80.2	75.4	72.6	69.7	56.2
IC 95% upper	100	97.6	95.0	93.6	92.0	92.4

Supplementary Fig. 3: Estimate optimal number of clusters.

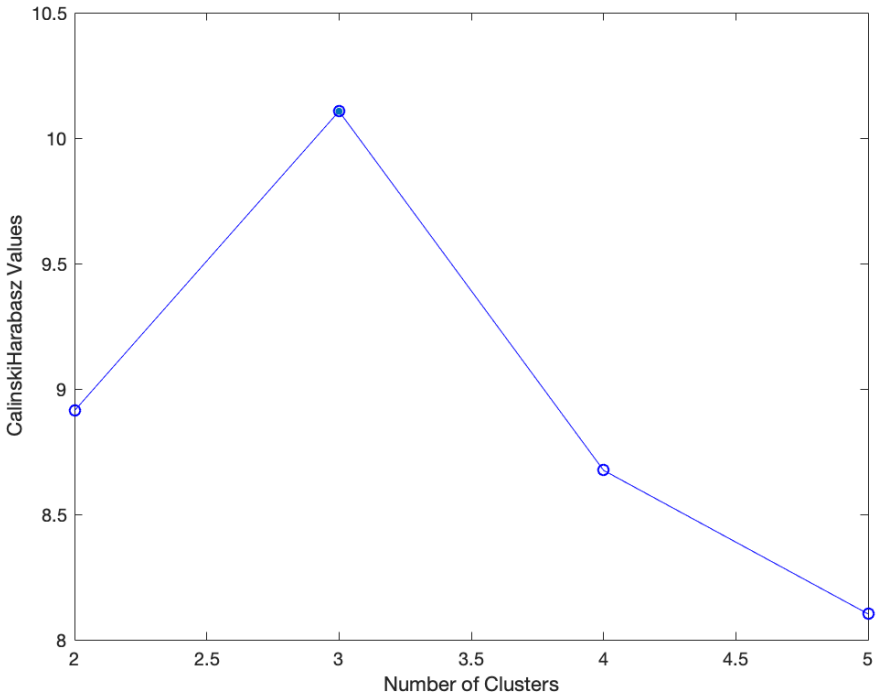
A: Gap statistic criterion



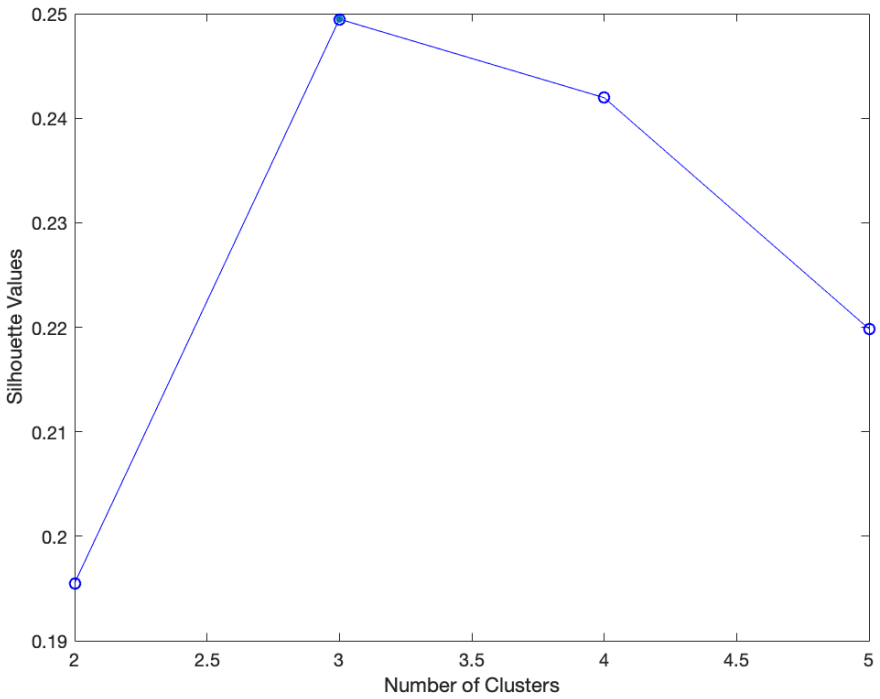
B: Davies-Bouldin criterion



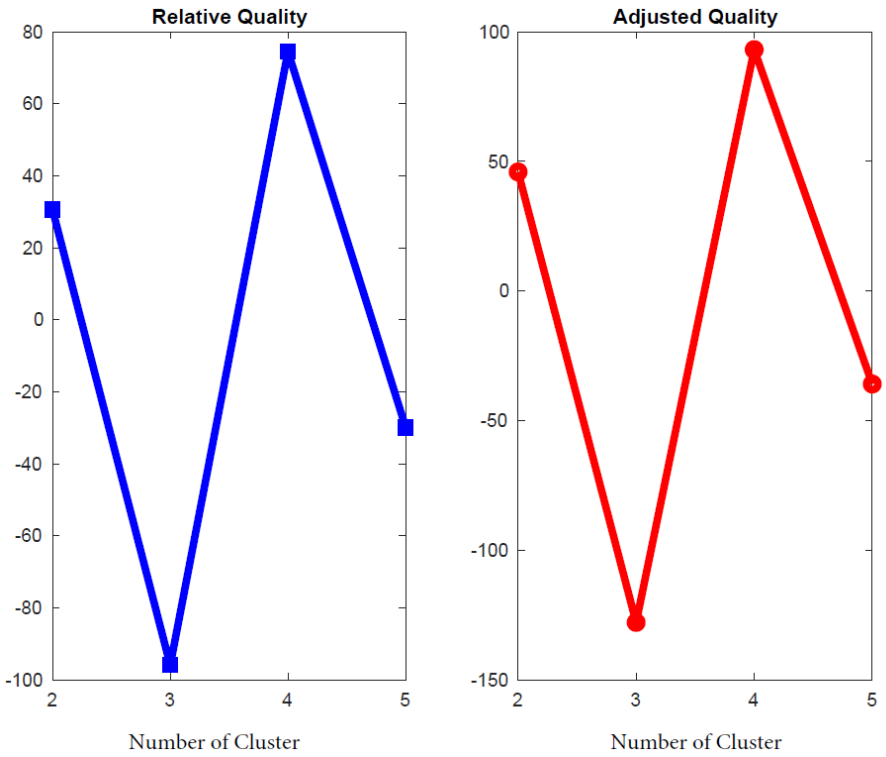
3C: Calinski-Harabasz criterion



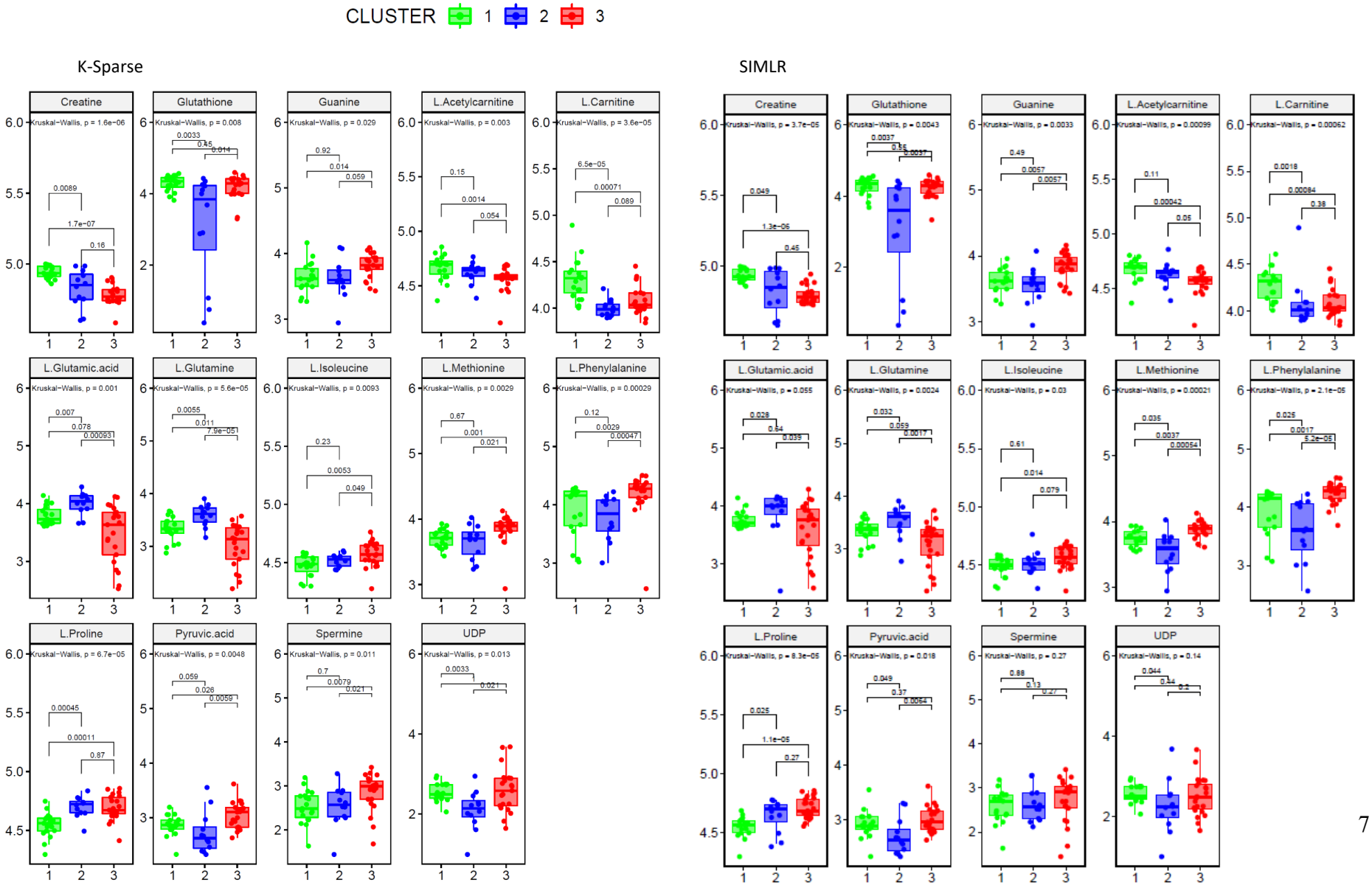
3D: Silhouette criterion



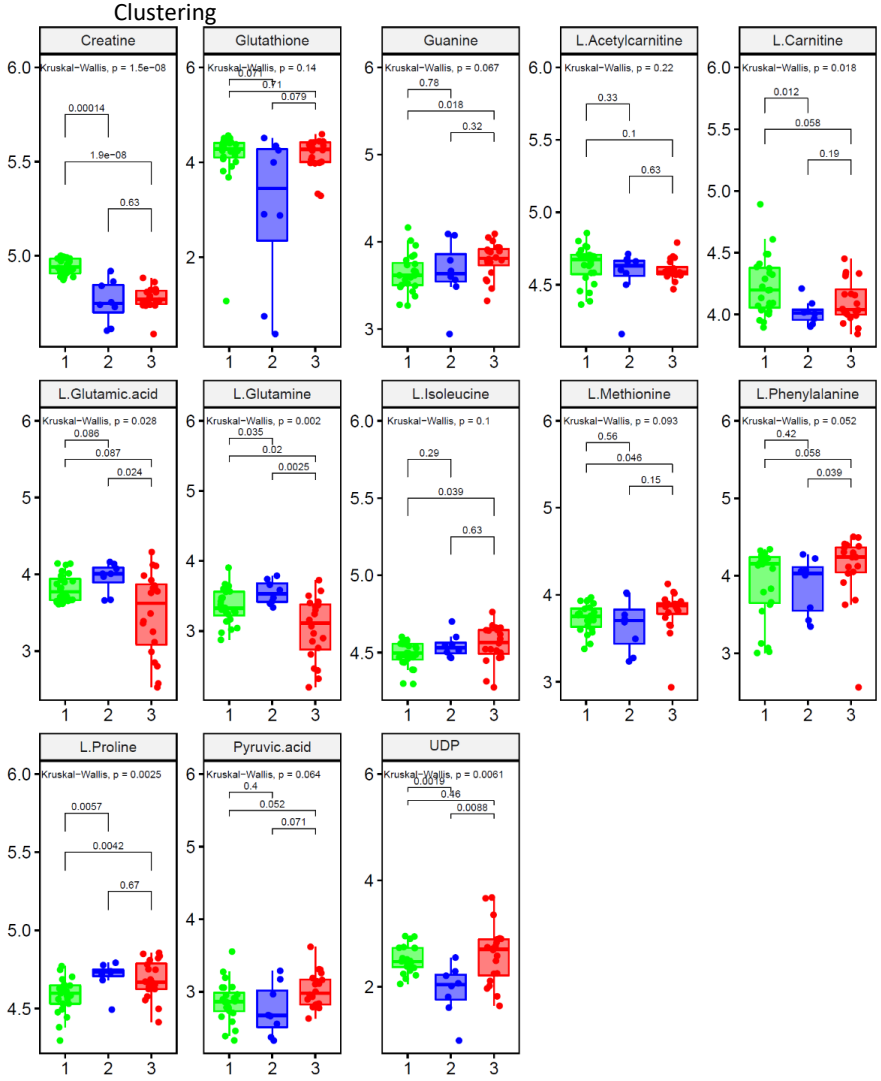
E: SIMLR criterion



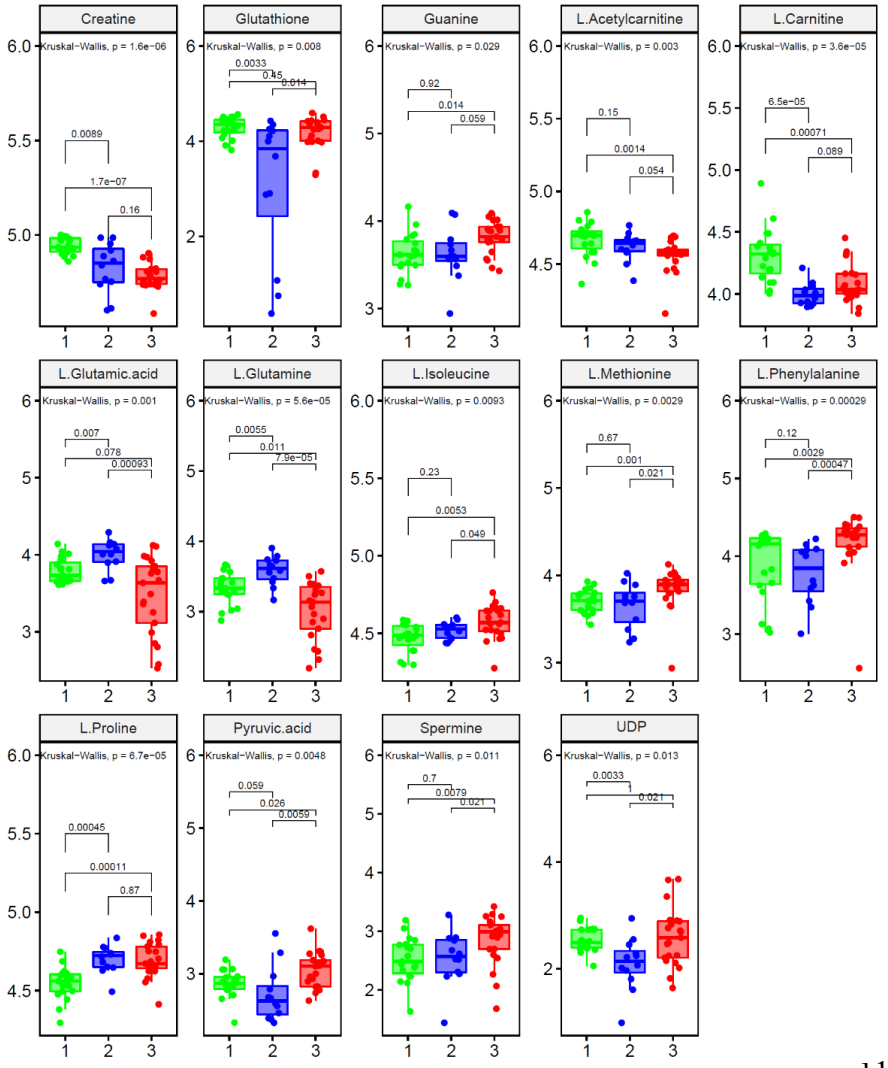
Supplementary Fig. S4: Boxplot of the 14 metabolites extracted from 5 ML methods

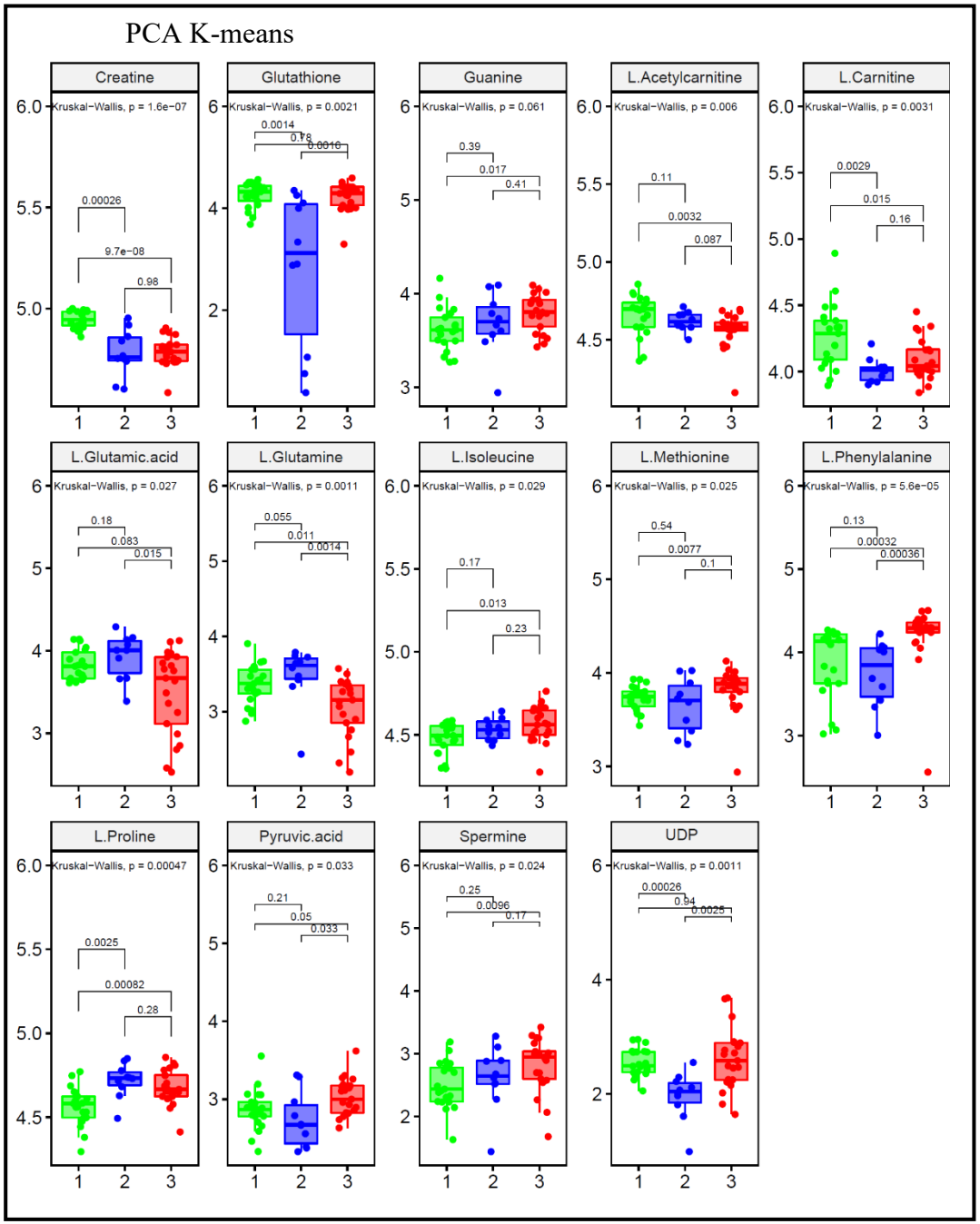


Sparse K-means Clustering



Spectral




PCA K-means
Creatine
Kruskal−Wallis, p = 1.6e−07
0.00026
9.7e−08
0.98
Glutathione
Kruskal−Wallis, p = 0.0021
0.0014
0.18
0.0016
Guanine
Kruskal−Wallis, p = 0.061
0.39
0.017
0.41
L.Acetylcarnitine
Kruskal−Wallis, p = 0.006
0.11
0.0032
0.087
L.Carnitine
Kruskal−Wallis, p = 0.0031
0.0029
0.015
0.16
L.Glutamic.acid
Kruskal−Wallis, p = 0.027
0.18
0.083
0.015
L.Glutamine
Kruskal−Wallis, p = 0.0011
0.055
0.011
0.0014
L.Isoleucine
Kruskal−Wallis, p = 0.029
0.17
0.013
0.23
L.Methionine
Kruskal−Wallis, p = 0.025
0.54
0.0077
0.1
L.Phenylalanine
Kruskal−Wallis, p = 5.6e−05
0.13
0.00032
0.00036
L.Proline
Kruskal−Wallis, p = 0.00047
0.0025
0.00082
0.28
Pyruvic.acid
Kruskal−Wallis, p = 0.033
0.21
0.05
0.033
Spermine
Kruskal−Wallis, p = 0.024
0.25
0.0096
0.17
UDP
Kruskal−Wallis, p = 0.0011
0.00026
0.94
0.0025

c) DISCUSSION

Nos résultats fournissent une preuve de concept de l'utilisation de la métabolomique avec des méthodes de ML non supervisée pour la stratification, et à terme, la gestion personnalisée des patients atteints de cancer du sein.

Cependant, deux limitations majeures ressortent de ce premier travail. Tout d'abord, même si le métabolisme des acides aminés, du glucose et de la glutaminolyse sont les voies les plus fréquemment discriminants pour le clustering, aucune voie métabolique universelle n'a été identifiée. Par conséquent, le choix de la méthode de ML va influencer grandement les voies d'activation métabolique sélectionnées par la suite pour établir une signature métabolomique.

La cohérence des clusters établis a été vérifiée grâce à l'analyse clinique et histopronostique des 3 groupes de patients et à l'interprétation des données de survie simulée par PREDICT tool qui s'appuie sur les critères cliniques et histopronostiques également. Le suivi était encore insuffisant et ne permettait pas une analyse de survie réelle.

Cependant, il existait une tendance nette en défaveur du cluster 3 en survie sans évènement à 18 mois, ce qui renforçait de nouveau le potentiel de la métabolomique en analyse non supervisée.

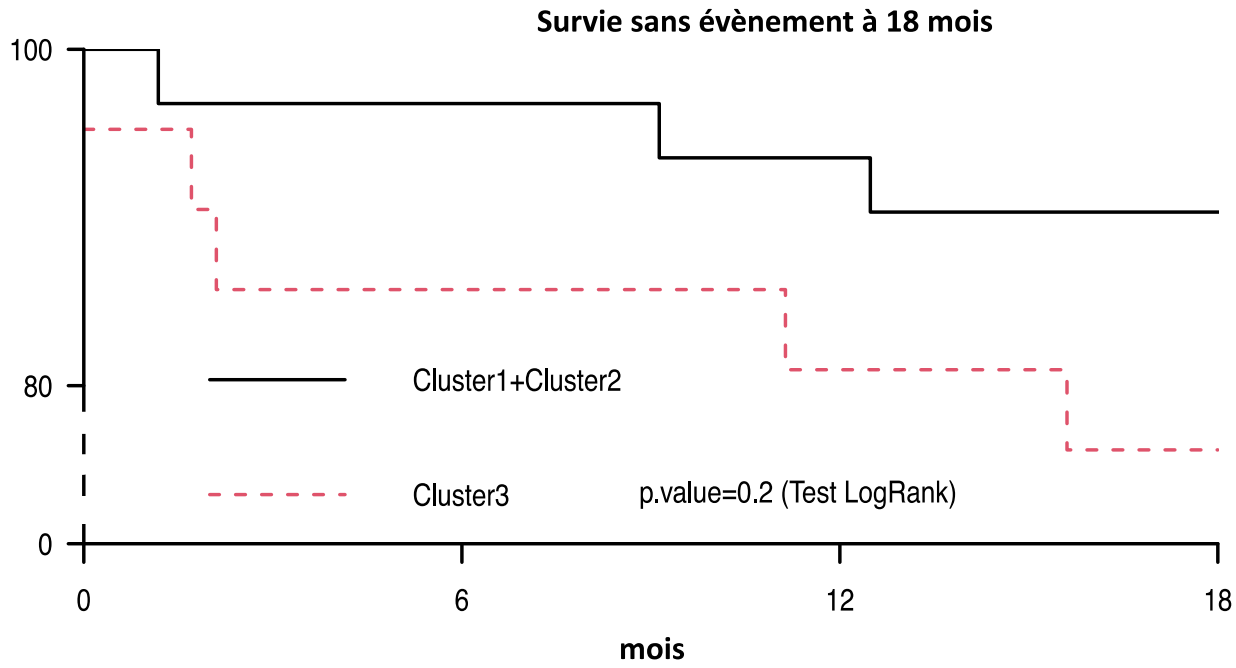


Figure 29. Premières estimations des données de survie de l'étude EMMEA.

La deuxième limitation est la taille de l'échantillon limitée et l'absence de cohorte de validation. Des études de validation sont donc nécessaires pour confirmer les premiers résultats obtenus.

2. Article EMMEA – Analyse de survie

a) INTRODUCTION

Après avoir mis en évidence une capacité de la métabolomique à différencier des clusters de patients avec des critères histopronostiques et des caractéristiques cliniques distinctes, nous voulions évaluer la validité clinique par des analyses de survie. Les analyses de survie simulées réalisées par PREDICT tool dans le premier article montraient déjà une tendance intéressante. Cependant, ces analyses simulées ne permettaient pas de conclure en une validité clinique de la métabolomique. Après 3 années de suivi supplémentaires, nous avons actualisé les données de survie. Le but était d'analyser la survie de des groupes de patients définis par les méthodes de ML non supervisées.

b) RESUME DES PRINCIPAUX RESULTATS & ARTICLE

Dans cette analyse de survie, 49 patientes, atteintes d'un cancer du sein localisé avec une indication de chimiothérapie adjuvante entre 2013 et 2016, ont été inclues. Trois patientes oligométastatiques d'emblée ou diagnostiquées métastatiques dans les 3 mois ont été retirées de l'analyse. Les groupes 1 et 2 ont été regroupés et comparés au groupe 3 au vu des premiers résultats de survie obtenus précédemment (figure 29). Le suivi médian a été étendu à 85,8 mois (IC95% [83,6-97,9]). Les analyses de survie (survie globale, survie spécifique au cancer et survie sans progression) obtenues avec les mêmes 5 méthodes ML non supervisées ont été analysées et une optimisation Bootstrap a été appliquée. Les méthodes PCA k-means, K-sparse et Spectral clustering ont été les plus performantes pour prédire la survie sans progression à 2 ans avec optimisation bootstrap (SSPb) ; comme exemple bootstrap, avec la méthode PCA k-means, la SSPb était de 94% (95%CI [90-98]) pour les clusters 1&2 contre 82% (95%CI [75-91]) pour le cluster 3 (p=0.01). La méthode PCA k-means a eu la meilleure performance avec une plus grande reproductibilité (HR moyen 2 (95%CI [1,4-2,7]); probabilité de p≤0,05 85%). Les analyses CSS et OS ont trouvé une discordance entre les 5 méthodes de ML non supervisées.

Survival analysis of patient groups defined by unsupervised machine learning clustering methods based on patient metabolomic data.

Caroline Bailleux[1,2] , David Chardin[3,2], Jean-Marie Guigonis[2], Jean-Marc Ferrero[1], Yann Chateau[4], Olivier Humbert[3,2], Thierry Pourcher[2], Jocelyn Gal[4].*

1 University Côte d'Azur, Centre Antoine Lacassagne, Medical Oncology Department, Nice, F-06189, France
2 University Côte d'Azur, Commissariat à l'Energie Atomique et aux énergies alternatives, Institut Frédéric Joliot, Service Hospitalier Frédéric Joliot, laboratory Transporters in Oncology and Radiotherapy in Oncology (TIRO), School of medicine, Nice, F-06100, France
3 University Côte d'Azur, Centre Antoine Lacassagne, Nuclear medicine Department, Nice, F-06189, France
4 University Côte d'Azur, Centre Antoine Lacassagne, Epidemiology and Biostatistics Department, Nice, F-06189, France

*Corresponding author:
Caroline Bailleux
Medical Oncology Department, Centre Antoine Lacassagne, University Côte d'Azur
33 avenue de Valombrose
06189 Nice, France
Phone : +33-4-92-03-11-34
E-mail : caroline.bailleux@nice.unicancer.fr

Key words: Unsupervised machine learning, clustering, breast cancer, survival.

Abstract

Purpose: We previously published the comparison of 5 different unsupervised machine learning methods (PCA k-means, Sparse k-means, Spectral clustering, SIMLR and k-sparse) to establish a metabolomic signatures of breast cancer (BC). Are reported here the survival analysis with extended follow-up.

Experimental design: Forty-nine consecutive patients, with non-metastatic BC and an indication of adjuvant chemotherapy between 2013 and 2016, were included retrospectively. Median follow-up was extended to 85.8 months (95%CI [83.6-97.9]). As previously reported, tumor resection samples were analyzed, and 449 metabolites were extracted by combined LC-MS. Survival analysis (overall survival, cancer-specific survival, and progression-free survival) obtained with the same 5 unsupervised ML methods were reported. Cluster 1 and 2 were regrouped and compared versus cluster 3. Bootstrap optimization was applied.

Results: PCA k-means, K-sparse and Spectral clustering were the most performing methods to predict 2-year progression-free survival with bootstrap optimization (PFSb); as bootstrap example, with PCA k-means method, PFSb were 94% (95%CI [90-98]) for cluster 1&2 versus 82% (95%CI [75-91] for cluster 3 (p=0.01). PCA k-means method had the best performance with higher reproducibility (mean HR 2 (95%CI [1.4-2.7]); probability of p≤0.05 85%). CSS and OS analyses found discordance between the 5 ML unsupervised methods.

Conclusion: Our study is a proof-of-principle that it is possible to use unsupervised ML methods on metabolomic data to predict PFS survival outcomes with the best performance for PCA k-means. A larger population study is needed to conclude on CSS and OS analyses.

Introduction

Worldwide, breast cancer(BC) is the most common cancer in women and the second leading cause of cancer deaths(93). Metabolic pathway alterations associated with BC tumors and disease progression have been widely explored at the genomic level(188,278,279). Proteomics studies have also revealed alterations in metabolism-associated protein expression in BC tumors with a correlation with overall and recurrence-free survival(280). Metabolomics is a new and rapidly developing field of investigation dedicated to the study of metabolism in tissues and fluids. There are two distinct approaches to metabolomics: a targeted approach aiming to precisely quantifying a limited number of predefined metabolites of interest(234) and a non-targeted approach aiming to measure, without a priori, the largest possible number of metabolites in a sample(235,281). Only few studies have reported associations between metabolic alterations and early BC patient survival outcomes based on serum analyses(282,283). To our knowledge, no study has been performed on tumor tissue.

Metabolomics can generate a large amount of data, which can make their analysis difficult, hence the interest of machine learning (ML) methods to extract useful information. In the case of metabolomics, ML involve supervised or unsupervised methods. Supervised method can be used to predict metabolites or biomarkers associated with a particular disease from labeled metabolomic data. Unsupervised learning can be used to identify patterns or groups of patients and metabolites that may be associated with specific diseases or phenotypes from unlabeled metabolomic data. The unsupervised algorithm takes a dataset and attempts to find a structure in the data by grouping or clustering the data points(284,285).

We previously published a comparison of 5 different unsupervised machine learning methods (PCA k-means, Sparse k-means, Spectral clustering, SimLR and K-sparse) to establish a metabolomic signature of breast cancer (BC)(286). In-silico survival analysis based on survival data simulated by predict tool (https://breast.predict.nhs.uk/tool)(114,287) revealed a significant difference for 5-year predicted overall survival (OS) and cancer-specific survival (CSS) between the 3 clusters(286). As event occurs many years after initial diagnosis, data was initially not mature enough to allow real data survival analyses. However, these simulated data may also be biased. Therefore, with extended follow-up of additional three years, we analyzed the real survival data. The objective of this study was to compare 5 different methods of unsupervised machine learning (PCA k-means, Sparse k-means, Spectral clustering, SimLR and K-sparse) to predict progression free survival (PFS), CSS and OS.

Material and methods

Selection and data collection of patients

A cohort of patients treated in our institution between March 2013 and September 2016 for a clinical stage I to III_B biopsy-proven BC, with an indication for adjuvant therapy after surgery, was included retrospectively in the study. Compared to the first publication(286), 3 metastatic patients were excluded from the survival analysis. A patient was considered de novo metastatic if metastatic diagnosis occurred within the first two months of treatment. Patients were treated according to national guidelines. Clinical, histological, radiological, and therapeutical data were retrospectively extracted from our facility's digital records or collected by a clinical data monitor. Follow up data was either extracted from our facility's digital records or retrieved through telephone communications if patients had changed facilities during surveillance. The date of the latest news was updated at the time of the final survival analysis, on December 2022. Written informed consent was obtained from all study participants. All procedures performed in studies involving tissue collection and analyses were in accordance with the ethical standards of the institutional and/or national research committee (French National Commission for Informatics and Liberties N°17003 and National Institute Health data N° 1515251018).

Statistical analysis

Relevance of the discovered clusters were assessed by comparing the clinical and survival characteristics between clusters by using χ^2 or Fisher's exact tests for categorical data, analysis of variance or Mann-Whitney's test for continuous variables and log-rank test for censored data. *P*-values inferior to 0.05 (two-sided) were considered statistically significant. Overall survival (OS) was defined by the time between diagnosis and death due to any cause. Cancer-specific survival (CSS) was defined by the time between diagnosis and death due to breast cancer. Progression-Free Survival (PFS) was defined by the time between diagnosis and the first progression (local, regional and metastasis). Patients showing no event (death or recurrence) or lost to follow-up were censored at the date of their last contact. OS, CSS and PFS were estimated using the Kaplan-Meier method. Median follow-up with a 95% confidence interval (95%CI) was calculated by reverse Kaplan–Meier method. 2-year outcomes were detailed, and bootstrap optimization were applied on these results to simulate the effect in a larger study population and highlight first trends : 200 subjects were randomly sampled to assess effectiveness without overpowering; 1000 subjects were randomly sampled to provide reproducibility and performance criteria (*P*-values, 95%CI) in a sample size comparable to similar studies dealing with genomic signature. For each sampling, the survival was compared between each cluster of all 5

methods. The relationship between clusters and the OS, CSS and PFS was analyzed by hazard ratio (95% confidence interval). A total of 1,500 replicates were performed.

Metabolomic analysis

Sample collection and preparation, details of LC-MS analysis, data preprocessing and metabolite identification have been already reported previously(286). Final metabolomic analysis was performed on 449 metabolites.

Results

Clinical and tumor characteristics

Forty-nine consecutive patients with non-metastatic breast cancer were analyzed. Tumor and treatment characteristics are described in **Table 1**. Median age was 65 years (range: 37-8). Main histological type and tumor stage were invasive ductal carcinoma (91.8%), T1 (40.8%) and T2 (44.9%) respectively. Twenty-one patients (42.9%) presented axillary lymph node invasion. Five patients (10.2%) had histological grade I tumors, 20 patients (40.8%) had histological grade II tumors and 23 patients (46.9%) had grade III tumors. Half of the patient's tumors had negative hormone receptor status (46.9%) and 18.4% had a Her-2 overexpression. To study the survival behavior of the supposedly aggressive cancers grouped in cluster 3 and to deal with small population size in each cluster, cluster 1 and 2 (cluster 1&2) were regrouped to be compared to cluster 3. As previously described, patients in cluster 3 were more often those with unfavorable prognostic factors: grade III, non-luminal with negative hormone receptor or triple negative phenotype. On the contrary, patients in cluster 1&2 had more often favorable prognosis factors: tumour stage T1, N0, histological grade I/II, and luminal phenotype. Details of patient's characteristics for the five unsupervised machine learning methods cluster 1&2 and cluster 3 are shown in **Table 1.**

Survival outcomes for the entire cohort

Median follow–up was extended to 85.8 months (95%CI, [83.6-97.9]). In the entire cohort, 2-year PFS and 5 year PFS were 98% (95%CI [94%-100%]) and 80% (95%CI [69%-92%]) respectively; 2-year CSS and 5 year CSS were 98% (95%CI [94%-100%]) and 85% (95%CI [76%-96%]) respectively; 2-year OS and 5 year OS were 88% (95%CI [79%-97%]) and 79% (95%CI [69%-92%]) respectively (**FigureS1**).

Survival Analysis of 2-year PFS with 5 unsupervised ML methods

As shown on table 2, the survival analysis with the previous clustering (k=3) did not show a statistical difference on the PFS data. The survival analysis with the new clustering regrouping clusters 1&2 showed a clinical trend, increased using the censored PFS at 2 years (**Figure 1** and **Figure S2** a). However, the result was still not statistically significant. We presented in **Figure 1** and **Figure S2**, an example of progression-free survival with bootstrap optimization and censored data at 2 years with the 5 unsupervised machine learning methods (n=200). With n=200 bootstrap optimization, PCA k-means, k-sparse and spectral clustering were the most performing methods to predict 2-year progression-free survival with bootstrap optimization (PFSb); PCA k-means 2-year PFSb: 94% (95%CI [90%-98%]) for cluster 1&2 versus 82% (95%CI [75%-91%] for cluster 3 (p=0.01). K-sparse 2-year PFSb: 94% (95%CI [90%-98%]) for cluster 1&2 versus 82% (95%CI [74%-91%] for cluster 3 (p=0.01). Spectral clustering also demonstrated significant efficiency for PFSb (p=0.02) (**Figure 1, Table 3**). To evaluate bootstrap reproducibility and performances, we applied a n=1000 bootstrap optimization. PCA k-means obtained the best performances (mean HR =2 (95%CI [1.4-2.7]); probability of p≤0.05; 85%) followed by k-sparse (mean HR = 1.6 (95%CI [1.1-2.4]); probability of p≤0.05; 83%) and spectral clustering (mean HR = 1.48 (95%CI [1.05-2.1]); probability of p≤0.05; 84%). The results of other methods were less statistically significant (**Figure S2, Table 2-3**).

Survival Analysis of 5-year survival outcomes with 5 unsupervised ML methods

Progression free survival curves were consistent between 4 ML methods and different with SimLR clustering. In the first 2 years, cluster 3 showed lower PFS than clusters 1&2. After 2 years, events in cluster 3 became rarer while events in cluster 1&2 were consistent over time becoming progressively numerically higher. At the end of 5 years, PFS was lower in clusters 1&2 than in cluster 3 (**Figure 2A, Table 3**). With SimLR clustering, the switch occurred earlier, at 1 year (**Figure S2A**). Concerning OS and CSS, results were homogenous between 4 ML methods and different with Sparse K-means. Cluster 3 had better survival outcomes, except for Sparse K-means clustering, where the trend was in disfavor of cluster 3, but only for OS (no significance reached for CSS) (**Figure 2B-C, Table 3**). Only Sparse K-means OS results were consistent with *in silico* survival analysis previously performed with PREDICT Tool(286), although the difference found was not statistically significant. With a n=1000 bootstrap optimization, Sparse K-means obtained a mean HR 1.6 (95%CI [1.2-2]) and a probability of p≤0.05 81% for OS prediction.

Discussion

To the best of our knowledge, this proof-of-concept study is the first to compare different unsupervised methods to identify metabolomics-based prognostics signatures in BC with survival analysis. We demonstrated that K-sparse, Spectral clustering and PCA k-means methods has a higher performance to predict 2-year PFS after bootstrap optimization than the other two ML methods. In contrast, for CSS and OS analyses, results were not consistent with *in silico* survival analysis previously performed with PREDICT Tool, except for Sparse K-means method, and only for OS.

From a clinical point of view, the ML methods were able to identify a distinct group of patients with a poor prognosis and a high risk for early recurrence (cohort 3). The PFS behavior switch at 2-years between cluster 3 and cluster 1&2 could be explained by the heterogenicity of the entire population. Patients in cluster 3 more often had triple-negative or HER-2 overexpressed tumors, which are known to be aggressive and relapse mainly in the first 2 years. In contrast, patients in cluster 1&2 were more likely to have HR+ tumors, which are less aggressive, but with a consistent risk of relapse over time. Indeed, for patients with aggressive tumors, PFS is lower the first 2 years, but for patients without relapse at 2-years, the risk of late relapse decrease compared to the risk for patients with HR+ tumors(288). With SimLR clustering, the switch occurred earlier, at 1 year, and some late relapses were observed. This could traduce a less strict selection in cluster 3 for aggressiveness, but better performance in clustering patients with relapse overall. However, even with a median follow-up of only 85.8 months, the analyses failed to find a significant difference in term of OS and CSS, contrasting with previously published *in silico* analysis(286). Only Sparse K-means method yielded the expected trend, but only for OS. This result is not sufficiently consistent to recommend the use of Sparse K-means method for survival analyses. The failure of the analysis is probably due to the limited sample size and the rarity of reported events. In addition, the retrospective nature of our study may interfere with long-term follow-up and survival analyses. To finish, survival outcomes are largely dependent on histological subtype and received treatment. Therefore, future studies should analyze a specific subtype of breast cancer with homogenous clinical setting and treatment to be able to study long-term outcomes.

From a methodological perspective, new clustering and bootstrap optimization may be a suitable option when the sample size is too small for significant statistical analysis. The latest genomic signature trials have examined several thousand patients to show a difference of a few percent(144,147,149). For example, the RxPonder trial dealing with Oncotype DX signature, randomized a total of 5083 women and 5018 participated in the trial. Among postmenopausal women, invasive disease–free survival at 5 years was 91.9% in the endocrine-only group and 91.3% in the chemoendocrine group,

with no chemotherapy benefit. Among premenopausal women, invasive disease–free survival at 5 years was 89.0% with endocrine-only therapy and 93.9% with chemoendocrine therapy (hazard ratio, 0.60; 95% CI, 0.43 to 0.83; P=0.002)(149). It is therefore acceptable that our study size is too small to show a significant difference. Here, bootstrap optimization was applied to simulate a larger study (n=200 and n=1000) and see if such a study would be worthwhile to conduct, as a proof-of-principle.

From a biological point of view, only few studies have reported associations between metabolic alterations and early BC patient survival. To our knowledge, no study has been performed on tumor tissue, but only on serum. Fahrmann et al. reported serum analyses of Diacetylspermine in patients with triple negative breast cancer (TNBC)(214). Diacetylspermine levels were higher in serum samples from patients with triple-negative breast cancer than in samples from patients without triple-negative breast cancer and from healthy volunteers. In a prospective cohort, the authors observed that serum Diacetylspermine levels were significantly increased in patients with early recurrence (<1 year). Higher serum Diacetylspermine levels were also associated with lower 5-year distant metastasis-free survival and 5-year overall survival. Asiago et al. published very interesting results on early detection of recurrent breast cancer using metabolite profiling with 7 metabolite markers. More than a half of the patients were predicted to have recurrence 13 months (on average) before the recurrence clinical diagnosis. However, this metabolomic signature allows an early detection and not a prediction of relapse. Oakman et al. calculated individual early patient 'metabolomic risk' derived from forty-four early breast cancer patients compared with fifty-one metastatic patients served as control. Metabolomic risk was compared with the Adjuvantionline 10-year mortality estimate. The comparison with Adjuvantionline revealed discordance like in our study. Of 21 patients assessed as high-risk by Adjuvantionline, 10 (48%) and 6 (29%) were at high metabolomic risk pre- and postoperatively, respectively. Of the 23 low-risk patients evaluated by Adjuvantionline, 11 (48%) preoperatively and 20 (87%) postoperatively were at low metabolomic risk. However, these simulated data may also be biased, hence the interest of our study and future studies on real survival data to distinguish limitations due to metabolomics from those due to simulated survival data.

Conclusion

The objective of our study was to compare different unsupervised machine learning algorithms on untargeted metabolomics data and to evaluate the performance of these methods to predict survival outcomes. Our results showed that it is possible to use unsupervised machine learning methods on metabolomic unlabeled data to identify clusters of patients with worse 2-year PFS. Among the 5 unsupervised ML methods reported here, PCA k-means, K-sparse and spectral clustering outperformed

the other two unsupervised methods. However, because of the retrospective study design and the small number of patients, no conclusion could be drawn for the prediction of CSS and OS. Future studies are needed with a larger population on specific histologic subtypes.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Author‘s contributions

Conception and design: C.B., J.G., D.C., T.P.; development of methodology: J.G. ; acquisition of data: O.H., C.B., J-M.G, Y.C. ; analysis and interpretation of data: J.G, C.B., T.P. ; writing, review, and/or revision of the manuscript: C.B., J.G., T.P. and all the authors; study supervision: T.P., O.H., J-M.F.

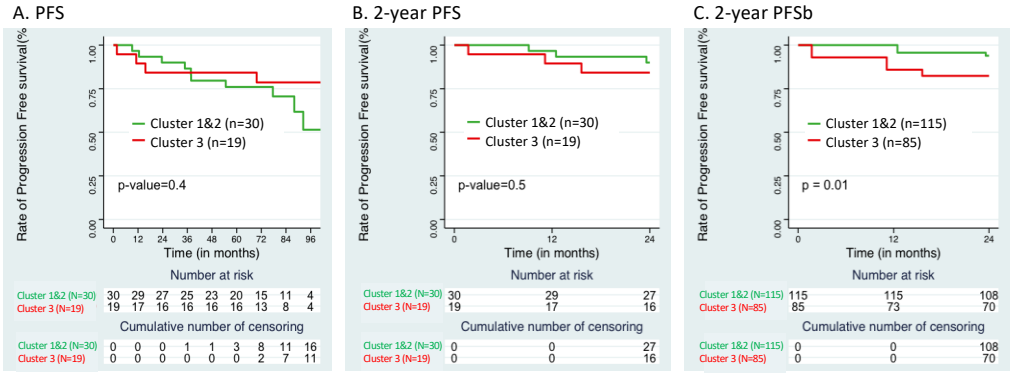
Grand Support

The authors declare no competing financial interests.

Acknowledgements

The authors acknowledge support from Centre Antoine Lacassagne and TIRO Unit, University Côte d’Azur, France.

PCA k-means

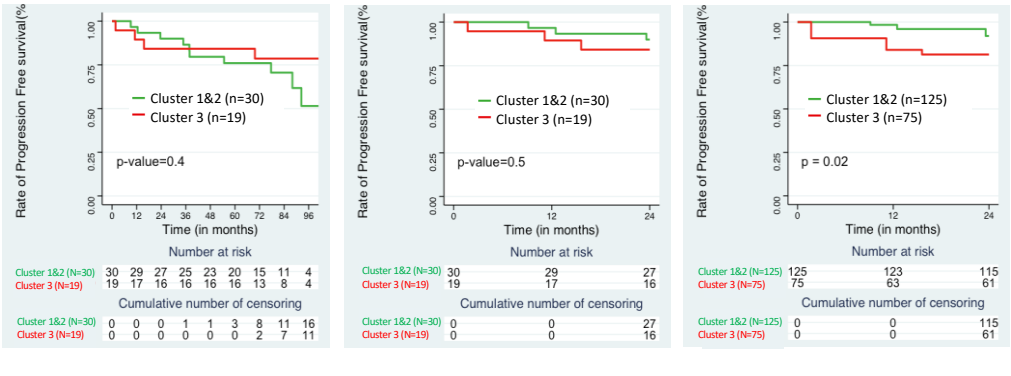


name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	30	20	19	16
number of events	0	7	0	3
censored patients	0	3	0	0
rate of survival (%)	1	0.76	1	0.84
lower limit IC95%	1	0.62	1	0.69
upper limit IC95%	1	0.93	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	30	27	19	16
number of events	0	3	0	3
censored patients	0	27	0	16
rate of survival (%)	1	0.9	1	0.84
lower limit IC95%	1	0.8	1	0.69
upper limit IC95%	1	1	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	115	108	85	70
number of events	0	7	0	15
censored patients	0	108	0	70
rate of survival (%)	1	0.94	1	0.82
lower limit IC95%	1	0.9	1	0.75
upper limit IC95%	1	0.98	1	0.91

Spectral Clustering

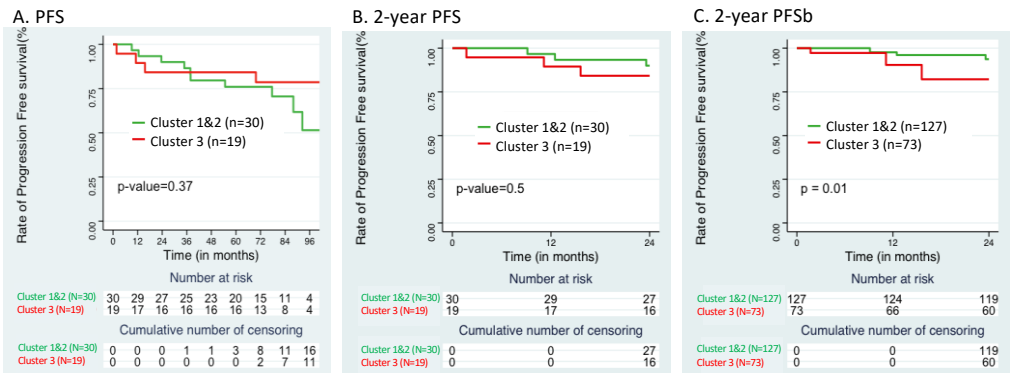


name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	30	20	19	16
number of events	0	7	0	3
censored patients	0	3	0	0
rate of survival (%)	1	0.76	1	0.84
lower limit IC95%	1	0.62	1	0.69
upper limit IC95%	1	0.93	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	30	27	19	16
number of events	0	3	0	3
censored patients	0	27	0	16
rate of survival (%)	1	0.9	1	0.84
lower limit IC95%	1	0.8	1	0.69
upper limit IC95%	1	1	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	30	27	19	16
number of events	0	3	0	3
censored patients	0	27	0	16
rate of survival (%)	1	0.9	1	0.84
lower limit IC95%	1	0.8	1	0.69
upper limit IC95%	1	1	1	1

K-sparse



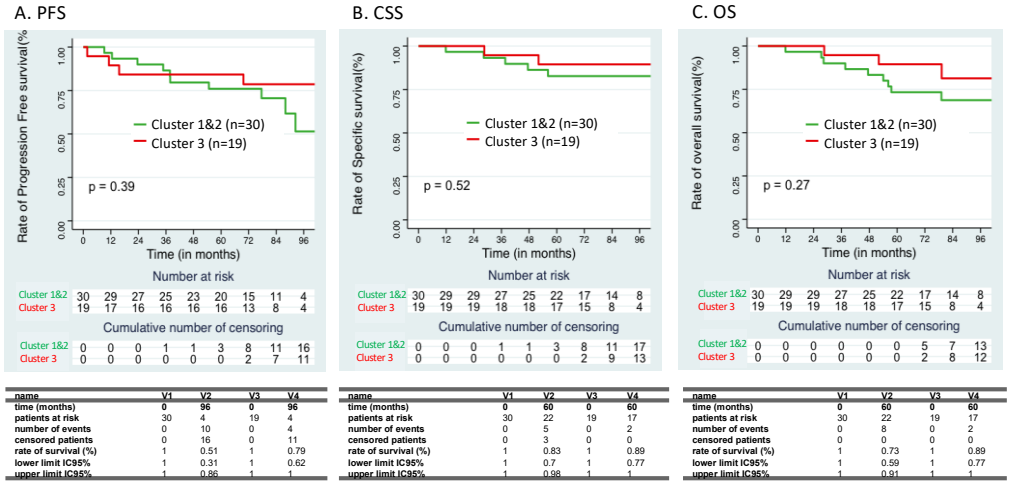
name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	30	20	19	16
number of events	0	7	0	3
censored patients	0	3	0	0
rate of survival (%)	1	0.76	1	0.84
lower limit IC95%	1	0.62	1	0.69
upper limit IC95%	1	0.93	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	30	27	19	16
number of events	0	3	0	3
censored patients	0	27	0	16
rate of survival (%)	1	0.9	1	0.84
lower limit IC95%	1	0.8	1	0.69
upper limit IC95%	1	1	1	1

name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	127	119	73	60
number of events	0	8	0	13
censored patients	0	119	0	60
rate of survival (%)	1	0.94	1	0.82
lower limit IC95%	1	0.9	1	0.74
upper limit IC95%	1	0.98	1	0.91

Figure 1. Survival Analysis of PFS with PCA k-means, Spectral clustering and K-sparse unsupervised machine learning methods. Cluster 1 and cluster 2 were regrouped in Cluster1&2 and compared to cluster 3. (A) PFS : progression-free survival; (B) 2-year PFS : censored data at 2-year; (C) 2-year PFSb : example of progression-free survival with bootstrap optimization and censored data at 2-year. Bootstrap optimization performance details are exposed in Table 3.

PCA k-means

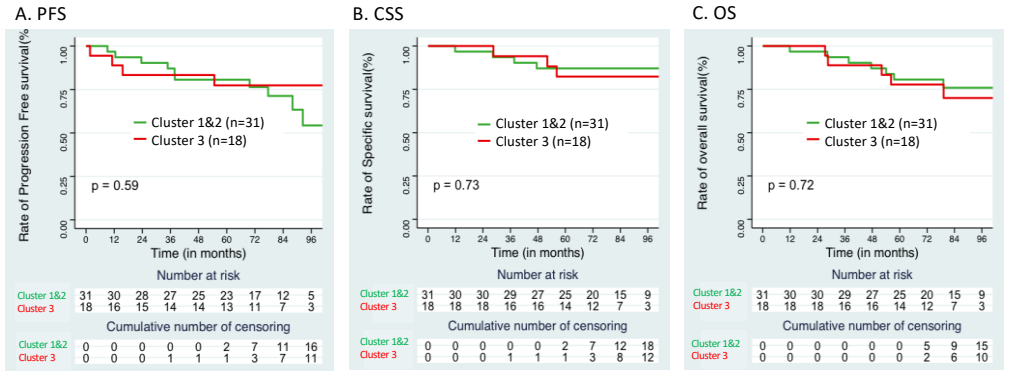


name	V1	V2	V3	V4
time (months)	0	96	0	96
patients at risk	30	4	19	4
number of events	0	10	0	4
censored patients	0	16	0	11
rate of survival (%)	1	0.51	1	0.79
lower limit IC95%	1	0.31	1	0.62
upper limit IC95%	1	0.86	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	30	22	19	17
number of events	0	5	0	2
censored patients	0	3	0	0
rate of survival (%)	1	0.83	1	0.89
lower limit IC95%	1	0.7	1	0.77
upper limit IC95%	1	0.98	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	30	22	19	17
number of events	0	8	0	2
censored patients	0	0	0	0
rate of survival (%)	1	0.73	1	0.89
lower limit IC95%	1	0.59	1	0.77
upper limit IC95%	1	0.91	1	1

Sparse K-means

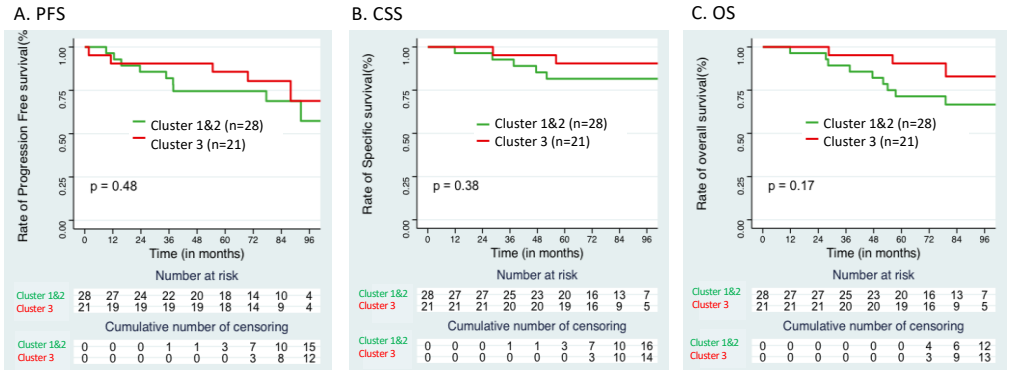


name	V1	V2	V3	V4
time (months)	0	96	0	96
patients at risk	31	5	18	3
number of events	0	10	0	4
censored patients	0	16	0	11
rate of survival (%)	1	0.54	1	0.77
lower limit IC95%	1	0.35	1	0.6
upper limit IC95%	1	0.85	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	31	25	18	14
number of events	0	4	0	3
censored patients	0	2	0	1
rate of survival (%)	1	0.87	1	0.82
lower limit IC95%	1	0.76	1	0.66
upper limit IC95%	1	1	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	31	25	18	14
number of events	0	6	0	4
censored patients	0	0	0	0
rate of survival (%)	1	0.81	1	0.78
lower limit IC95%	1	0.68	1	0.61
upper limit IC95%	1	0.96	1	1

SimLR



name	V1	V2	V3	V4
time (months)	0	96	0	96
patients at risk	28	4	21	4
number of events	0	9	0	5
censored patients	0	15	0	12
rate of survival (%)	1	0.57	1	0.69
lower limit IC95%	1	0.37	1	0.48
upper limit IC95%	1	0.9	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	28	20	21	19
number of events	0	5	0	2
censored patients	0	3	0	0
rate of survival (%)	1	0.82	1	0.9
lower limit IC95%	1	0.68	1	0.79
upper limit IC95%	1	0.98	1	1

name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	28	20	21	19
number of events	0	8	0	2
censored patients	0	0	0	0
rate of survival (%)	1	0.71	1	0.9
lower limit IC95%	1	0.57	1	0.79
upper limit IC95%	1	0.9	1	1

Figure 2. Survival Analysis with extended follow-up with PCA k-means, Sparse K-means et SimLR. Cluster 1 and cluster 2 were regrouped in Cluster1&2 and compared to cluster 3. (A) PFS : progression free survival; (B) CSS : cancer-specific survival ; (C) OS : overall survival. Bootstrap optimization performance details are exposed in Table 3.

Table 1. Clinical comparison of 49 patients between clusters

		K-sparse			Spectral Clustering			PCA-K-means			SIMLR			Sparse K-means		
Characteristics	All (n=49)	Cluster 1&2 (n=30)	Cluster 3 (n=19)	p-value	Cluster 1&2 (n=30)	Cluster 3 (n=19)	p-value	Cluster 1&2 (n=30)	Cluster 3 (n=19)	p-value	Cluster 1&2 (n=28)	Cluster 3 (n=21)	p-value	Cluster 1&2 (n=31)	Cluster 3 (n=18)	p-value
	Nb of patients (%)	Nb of patients (%)	Nb of patients (%)		Nb of patients (%)	Nb of patients (%)		Nb of patients (%)	Nb of patients (%)		Nb of patients (%)	Nb of patients (%)		Nb of patients (%)	Nb of patients (%)	
Age (median min-max)*	65 (37-88)	65 (38-88)	63 (37-84)	0.525	65 (38-88)	63 (37-84)	0.525	65 (38-88)	65 (37-84)	0.547	65 (38-88)	63 (37-84)	0.424	65 (38-88)	65.5 (37-82)	0.881
Histology type£				0.144			0.144			0.727			0.756			0.378
Invasive ductal carcinoma	45 (91.8)	27 (90.0)	18 (94.7)		27 (90.0)	18 (94.7)		28 (93.3)	17 (89.5)		26 (92.9)	19 (90.5)		27 (87.1)	18 (100.0)	
Invasive lobular carcinoma	3 (6.1)	3 (10.0)	0 (0.0)		3 (10.0)	0 (0.0)		2 (6.7)	1 (5.3)		2 (7.1)	1 (4.8)		3 (9.7)	0 (0.0)	
Other	1 (2.0)	0 (0.0)	1 (5.3)		0 (0.0)	1 (5.3)		0 (0.0)	1 (5.3)		0 (0.0)	1 (4.8)		1 (3.2)	0 (0.0)	
Tumor stage£				0.073			0.073			0.074			0.080			0.237
T1	20 (40.8)	16 (53.3)	4 (21.1)		16 (53.3)	4 (21.1)		16 (53.3)	4 (21.1)		15 (53.6)	5 (23.8)		15 (48.4)	5 (27.8)	
T2	22 (44.9)	10 (33.3)	12 (63.2)		10 (33.3)	12 (63.2)		10 (33.3)	12 (63.2)		9 (32.1)	13 (61.9)		11 (35.5)	11 (61.1)	
T3	7 (14.3)	4 (13.3)	3 (15.8)		4 (13.3)	3 (15.8)		4 (13.3)	3 (15.8)		4 (14.3)	3 (14.3)		5 (16.1)	2 (11.1)	
Axillary lymph node status#				0.090			0.090			0.090			0.243			0.441
N0	28 (57.1)	20 (66.7)	8 (42.1)		20 (66.7)	8 (42.1)		20 (66.7)	8 (42.1)		18 (64.3)	10 (47.6)		19 (61.3)	9 (50.0)	
N+	21 (42.9)	10 (33.3)	11 (57.9)		10 (33.3)	11 (57.9)		10 (33.3)	11 (57.9)		10 (35.7)	11 (52.4)		12 (38.7)	9 (50.0)	
Histological grade£				**0.023**			**0.023**			**0.049**			**0.018**			**0.049**
I	5 (10.2)	5 (16.7)	0 (0.0)		5 (16.7)	0 (0.0)		5 (16.7)	0 (0.0)		5 (17.9)	0 (0.0)		5 (16.1)	0 0	
II	20 (40.8)	15 (50.0)	5 (26.3)		15 (50.0)	5 (26.3)		14 (46.7)	6 (31.6)		14 (50.0)	6 (28.6)		14 (45.2)	6 (33.3)	
III	23 (46.9)	10 (33.3)	13 (68.4)		10 (33.3)	13 (68.4)		11 (36.7)	12 (63.2)		9 (32.1)	14 (66.7)		11 (35.5)	12 (66.7)	
Hormonal status#				**0.016**			**0.016**			0.070			**0.017**			**0.035**
Negatif	23 (46.9)	10 (33.3)	13 (68.4)		10 (33.3)	13 (68.4)		11 (36.7)	12 (63.2)		9 (32.1)	14 (66.7)		11 (35.5)	12 (66.7)	
Positif	26 (53.1)	20 (66.7)	6 (31.6)		20 (66.7)	6 (31.6)		19 (63.3)	7 (36.8)		19 (67.9)	7 (33.3)		20 (64.5)	6 (33.3)	
Her-2 status£				0.282			0.282			0.282			0.470			0.708
Non-over-expressed	40 (81.6)	26 (86.7)	14 (73.7)		26 (86.7)	14 (73.7)		26 (86.7)	14 (73.7)		24 (85.7)	16 (76.2)		26 (83.9)	14 (77.8)	
Over-expressed	9 (18.4)	4 (13.3)	5 (26.3)		4 (13.3)	5 (26.3)		4 (13.3)	5 (26.3)		4 (14.3)	5 (23.8)		5 (16.1)	4 (22.2)	
Triple-negatif status#				**0.043**			**0.043**			0.165			**0.025**			**0.025**
No	34 (69.4)	24 (80.0)	10 (52.6)		24 (80.0)	10 (52.6)		23 (76.7)	11 (57.9)		23 (82.1)	11 (52.4)		25 (80.6)	9 (50.0)	
Yes	15 (30.6)	6 (20.0)	9 (47.4)		6 (20.0)	9 (47.4)		7 (23.3)	8 (42.1)		5 (17.9)	10 (47.6)		6 (19.4)	9 (50.0)	
Luminal#				**0.006**			**0.006**			**0.030**			**0.006**			**0.013**
No	24 (49.0)	10 (33.3)	14 (73.7)		10 (33.3)	14 (73.7)		11 (36.7)	13 (68.4)		9 (32.1)	15 (71.4)		11 (35.5)	13 (72.2)	
Yes	25 (51.0)	20 (66.7)	5 (26.3)		20 (66.7)	5 (26.3)		19 (63.3)	6 (31.6)		19 (67.9)	6 (28.6)		20 (64.5)	5 (27.8)	
Adjuvant chemotherapy£				0.323			0.323			0.323			0.192			1
No	12 (24.5)	9 (30.0)	3 (15.8)		9 (30.0)	3 (15.8)		9 (30.0)	3 (15.8)		9 (32.1)	3 (14.3)		8 (25.8)	4 (22.2)	
Yes	37 (75.5)	21 (70.0)	16 (84.2)		21 (70.0)	16 (84.2)		21 (70.0)	16 (84.2)		19 (67.9)	18 (85.7)		23 (74.2)	14 (77.8)	
Adjuvant radiotherapy£				1			1			1			1			1
No	5 (10.2)	3 (10.0)	2 (10.5)		3 (10.0)	2 (10.5)		3 (10.0)	2 (10.5)		3 (10.7)	2 (9.5)		3 (9.7)	2 (11.1)	
Yes	44 (89.8)	27 (90.0)	17 (89.5)		27 (90.0)	17 (89.5)		27 (90.0)	17 (89.5)		25 (89.3)	19 (90.5)		28 (90.3)	16 (88.9)	
Adjuvant hormone therapy#				0.181			0.181			0.181			0.154			0.319
No	20 (40.8)	10 (33.3)	10 (52.6)		10 (33.3)	10 (52.6)		10 (33.3)	10 (52.6)		9 (32.1)	11 (52.4)		11 (35.5)	9 (50.0)	
Yes	29 (59.2)	20 (66.7)	9 (47.4)		20 (66.7)	9 (47.4)		20 (66.7)	9 (47.4)		19 (67.9)	10 (47.6)		20 (64.5)	9 (50.0)	

Table 1. Clinical comparison of 49 patients between clusters.

£: Fisher's exact test; #: Chi^2-test; *: student t-test

Table 2. Survival outcomes with 5 different methods of unsupervised machine learning

ML Method		Previous clustering (k=3)	New clustering (k=2)	2-year survival outcome (k=2)
k-sparse	OS	p=0.5	p=0.3	p=0.3
	CSS	p=0.8	p=0.5	p=0.3
	PFS	p=0.7	p=0.4	p=0.5
PCA k-means	OS	p=0.5	p=0.3	p=0.5
	CSS	p=0.7	p=0.5	p=0.5
	PFS	p=0.5	p=0.4	p=0.5
Spectral clustering	OS	p=0.5	p=0.3	p=0.3
	CSS	p=0.8	p=0.5	p=0.4
	PFS	p=0.7	p=0.4	p=0.5
Sparse k-means	OS	p=0.8	p=0.7	p=1
	CSS	p=0.9	p=0.7	p=1
	PFS	p=0.9	p=0.6	p=0.5
SimLR	OS	p=0.3	p=0.2	p=0.3
	CSS	p=0.7	p=0.4	p=0.3
	PFS	p=0.8	p=0.5	p=0.7

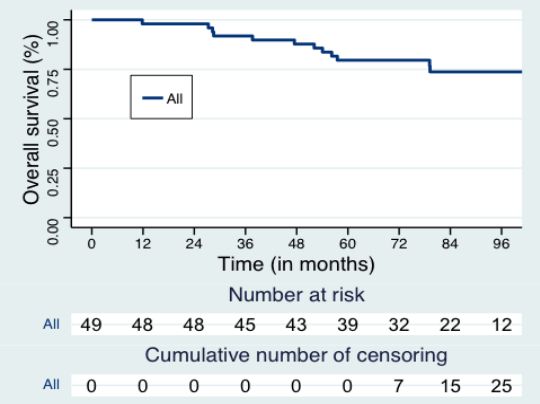
ML: machine learning; k: number of clusters; OS: overall survival; CSS: cancer-specific survival; PFS: progression-free survival; k: number of clusters.

Table 3. Survival outcomes with 5 different methods of unsupervised machine learning and Bootstrap optimization (k=2)

ML Method		mean HR [95%CI]	probability of p≤0.05
k-sparse	OS	0.53 [0.4-0.7]*	82%
	CSS	0.6 [0.4-0.9]*	83%
	2-years PFS	1.6 [1.1-2.4]*	83%
PCA k-means	OS	0.5 [0.3-0.6]*	85%
	CSS	0.6 [0.4-0.9]*	85%
	2-years PFS	2 [1.4-2.7]*	85%
Spectral clustering	OS	0.5 [0.35-0.65]*	85%
	CSS	0.6 [0.4-0.8]*	85%
	2-years PFS	1.48 [1.05-2.1]*	84%
Sparse k-means	OS	1.6 [1.2-2.0]*	81%
	CSS	1.1 [0.8-1.5]	80%
	2-years PFS	1.3 [0.9-1.9]	82%
SimLR	OS	0.35 [0.25-0.45]*	83%
	CSS	0.5 [0.35-0.7]*	84%
	2-years PFS	0.65 [0.4-0.9]*	83%

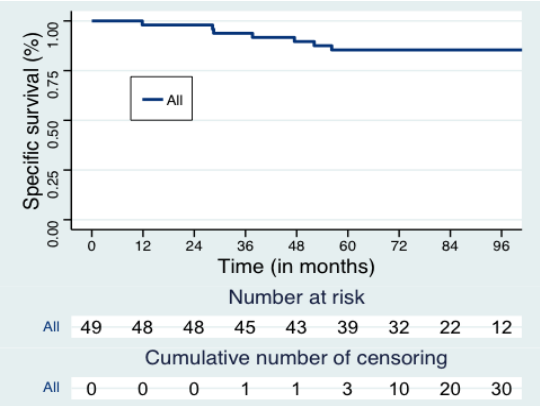
ML: machine learning; k: number of clusters; OS: overall survival; CSS: cancer-specific survival; PFS: progression-free survival; HR: hazard ratio; 95CI%: 95% confidence interval; k: number of clusters; *: statistically significant.

A.



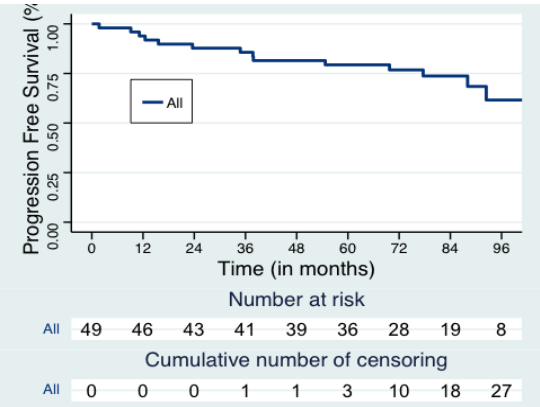
name	V1	V2	V3	V4	V5	V6	V7	V8	V9
time (months)	0	12	24	36	48	60	72	84	96
patients at risk	49	48	48	45	43	39	32	22	12
number of events	0	1	0	3	2	4	0	2	0
censored patients	0	0	0	0	0	0	7	8	10
rate of survival (%)	1	0.98	0.98	0.92	0.88	0.8	0.8	0.74	0.74
lower limit IC95%	1	0.94	0.94	0.84	0.79	0.69	0.69	0.62	0.62
upper limit IC95%	1	1	1	1	0.97	0.92	0.92	0.88	0.88

B.



name	V1	V2	V3	V4	V5	V6	V7	V8	V9
time (months)	0	12	24	36	48	60	72	84	96
patients at risk	49	48	48	45	43	39	32	22	12
number of events	0	1	0	2	2	2	0	0	0
censored patients	0	0	0	1	0	2	7	10	10
rate of survival (%)	1	0.98	0.98	0.94	0.9	0.85	0.85	0.85	0.85
lower limit IC95%	1	0.94	0.94	0.87	0.81	0.76	0.76	0.76	0.76
upper limit IC95%	1	1	1	1	0.99	0.96	0.96	0.96	0.96

C.

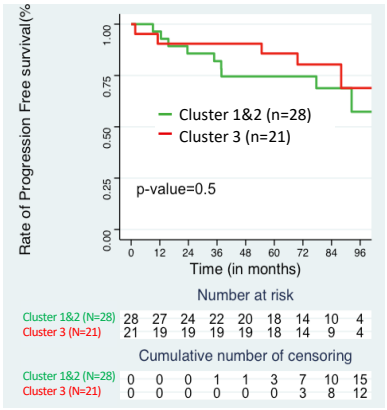


name	V1	V2	V3	V4	V5	V6	V7	V8	V9
time (months)	0	12	24	36	48	60	72	84	96
patients at risk	49	46	43	41	39	36	28	19	8
number of events	0	3	3	1	2	1	1	1	2
censored patients	0	0	0	1	0	2	7	8	9
rate of survival (%)	1	0.94	0.88	0.86	0.81	0.79	0.77	0.74	0.62
lower limit IC95%	1	0.87	0.79	0.76	0.71	0.69	0.66	0.62	0.45
upper limit IC95%	1	1	0.97	0.96	0.93	0.92	0.9	0.88	0.84

Figure S1. Survival outcomes for the entire population (n=49). (A) Overall survival; (B) Specific survival; (C) Progression Free Survival.

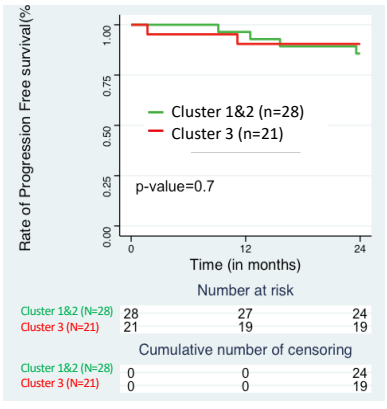
SimLR

A. PFS



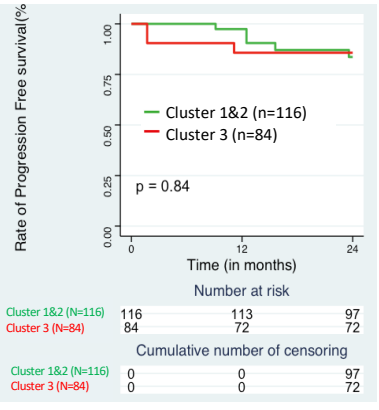
name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	28	18	21	18
number of events	0	7	0	3
censored patients	0	3	0	0
rate of survival (%)	1	0.75	1	0.86
lower limit IC95%	1	0.6	1	0.72
upper limit IC95%	1	0.93	1	1

B. 2-year PFS



name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	28	24	21	19
number of events	0	4	0	2
censored patients	0	24	0	19
rate of survival (%)	1	0.86	1	0.9
lower limit IC95%	1	0.74	1	0.79
upper limit IC95%	1	1	1	1

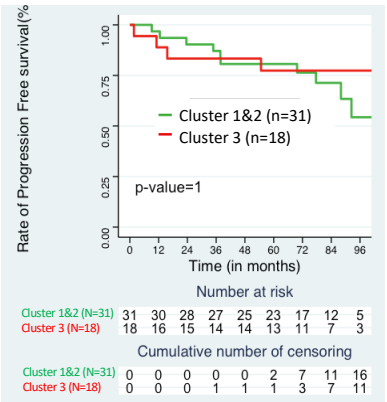
C. 2-year PFSb



name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	116	97	84	72
number of events	0	19	0	12
censored patients	0	97	0	72
rate of survival (%)	1	0.84	1	0.86
lower limit IC95%	1	0.77	1	0.79
upper limit IC95%	1	0.91	1	0.94

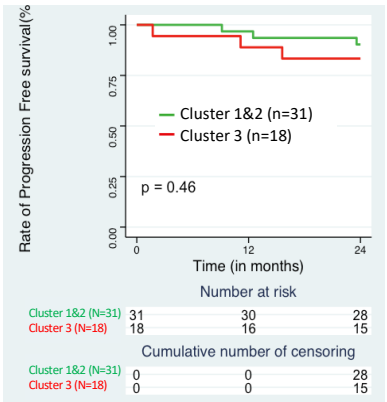
Sparse-K-means

A. PFS



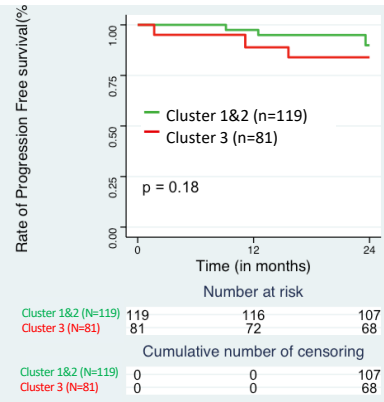
name	V1	V2	V3	V4
time (months)	0	60	0	60
patients at risk	31	23	18	13
number of events	0	6	0	4
censored patients	0	2	0	1
rate of survival (%)	1	0.81	1	0.77
lower limit IC95%	1	0.68	1	0.6
upper limit IC95%	1	0.96	1	1

B. 2-year PFS



name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	31	28	18	15
number of events	0	3	0	3
censored patients	0	28	0	15
rate of survival (%)	1	0.9	1	0.83
lower limit IC95%	1	0.8	1	0.68
upper limit IC95%	1	1	1	1

C. 2-year PFSb



name	V1	V2	V3	V4
time (months)	0	24	0	24
patients at risk	119	107	81	68
number of events	0	12	0	13
censored patients	0	107	0	68
rate of survival (%)	1	0.9	1	0.84
lower limit IC95%	1	0.85	1	0.76
upper limit IC95%	1	0.95	1	0.92

Figure S2. Survival Analysis of PFS with SimLR and Sparse-K-means unsupervised machine learning methods. Cluster 1 and cluster 2 were regrouped in Cluster1&2 and compared to cluster 3. (A) PFS : progression-free survival; (B) 2-year PFS : censored data at 2-year; (C) 2-year PFSb : example of progression-free survival with bootstrap optimization and censored data at 2-year. Bootstrap optimization performance details are exposed in Table 3.

c) DISCUSSION

Notre étude est une preuve de principe qu'il est possible d'utiliser des méthodes ML non supervisées sur des données métabolomiques pour prédire les résultats de la SSP à 2 ans avec la meilleure performance pour PCA k-means.

En revanche, pour l'analyse de survie à 5 ans (survie globale et survie spécifique), les résultats ne sont pas cohérents avec les analyses de survie in silico réalisées précédemment avec l'outil PREDICT uk (article 1), à l'exception de la méthode Sparse K-means, et uniquement pour la survie globale. Ce résultat n'est pas suffisamment consistant pour recommander l'utilisation de la méthode Sparse K-means pour les analyses de survie. L'échec de l'analyse est probablement dû à la taille limitée de l'échantillon et à la rareté des événements rapportés. De plus, la nature rétrospective de notre étude peut interférer avec le suivi à long terme et les analyses de survie. Enfin, la survie dépend largement du sous-type histologique et des traitements reçus. Par conséquent, les études futures devraient analyser un sous-type spécifique de cancer du sein avec un contexte clinique homogène et un traitement comparable pour pouvoir étudier les résultats à long terme sans artefacts.

Afin de poursuivre ce travail, les principaux métabolites d'intérêt prédits vont être validés et une analyse supervisée va être réalisée en censurant les données de survie sans progression à 2 ans. L'objectif sera alors d'obtenir un modèle prédictif plus puissant et de déterminer les métabolites et les voies métaboliques en jeu dans la récidive précoce spécifiquement. Ce travail sera réalisé en post-thèse (cf. Partie Conclusion et Perspectives, projet EMMEA-S).

3. Article Grade – Analyse supervisée

a) INTRODUCTION

Une analyse a été réalisée pour définir des biomarqueurs d'agressivité tumorale. La première question était le choix du critère clinico-histologique représentatif de cette agressivité tumorale. Plusieurs marqueurs auraient pu être analysés comme le Ki67 ou la surexpression de HER2. Cependant, la difficulté de détermination exacte de ces facteurs dans la pratique courante pour le Ki67, et les changements d'annotations récentes pour HER2 ont rendu l'analyse compliquée dès les premières étapes de monitoring des données. Nous avons donc décidé d'étudier le grade. Les échantillons étaient donc labélisés de haut grade (Grade III) ou de grade bas ou intermédiaire (Grade I&II).

Cette analyse supervisée a été réalisée sur deux cohortes. La cohorte niçoise, qui a servi de cohorte d'entrainement (n=51, 1 patiente exclue : grade non disponible), et la cohorte dijonnaise, qui a servi de cohorte test (de validation) (n=49). Au niveau métabolomique, cela nous a obligés à considérer des étapes supplémentaires de filtrage pour obtenir une base de données harmonisée : l'utilisation des métabolites communs par fusion des 2 bases, la suppression des doublons et la fusion des lignes dédoublées.

L'analyse pré-analytique a été améliorée lors d'un travail de méthodologie (cf. Paragraphe Méthodologie pré-analytique spécifique) et a permis d'extraire 602 métabolites d'intérêt communs et identifiés.

b) RESUME DES PRINCIPAUX RESULTATS & ARTICLE

Toutes les tumeurs de cancer du sein étaient classées comme étant de haut grade (grade III) ou de grade bas ou intermédiaire (grade I-II). Une signature métabolomique composée des 12 premiers métabolites a été identifiée à partir de la base de données des 602 métabolites communs et identifiés. Les analyses de PLS-DA ont montré une bonne performance pour cette signature. Pour la cohorte d'apprentissage et la cohorte de

validation, les accuracy étaient respectivement de 0,81 et 0,82, les scores R2 de 0,57 et 0,55, et les scores Q2 de 0,44431 et 0,40147; les AUC des courbes ROC étaient de 0,882 et 0,886. Le modèle pouvait donc distinguer les tumeurs de grade élevé et de grade faible/intermédiaire avec une probabilité de près de 90 %. Bien que non utile en routine clinique, cette capacité prédictive a permis de mettre en évidence la fiabilité de la signature et son interprétation au niveau biologique.

Le métabolite le plus pertinent était la N1,N12-diacétylspermine. Les analyses d'enrichissement et les analyses des voies métaboliques ont mis en évidence la voie du tryptophane avec une forte variabilité de concentration entre les différents échantillons analysés. La voie de la sérotonine était également modifiée.

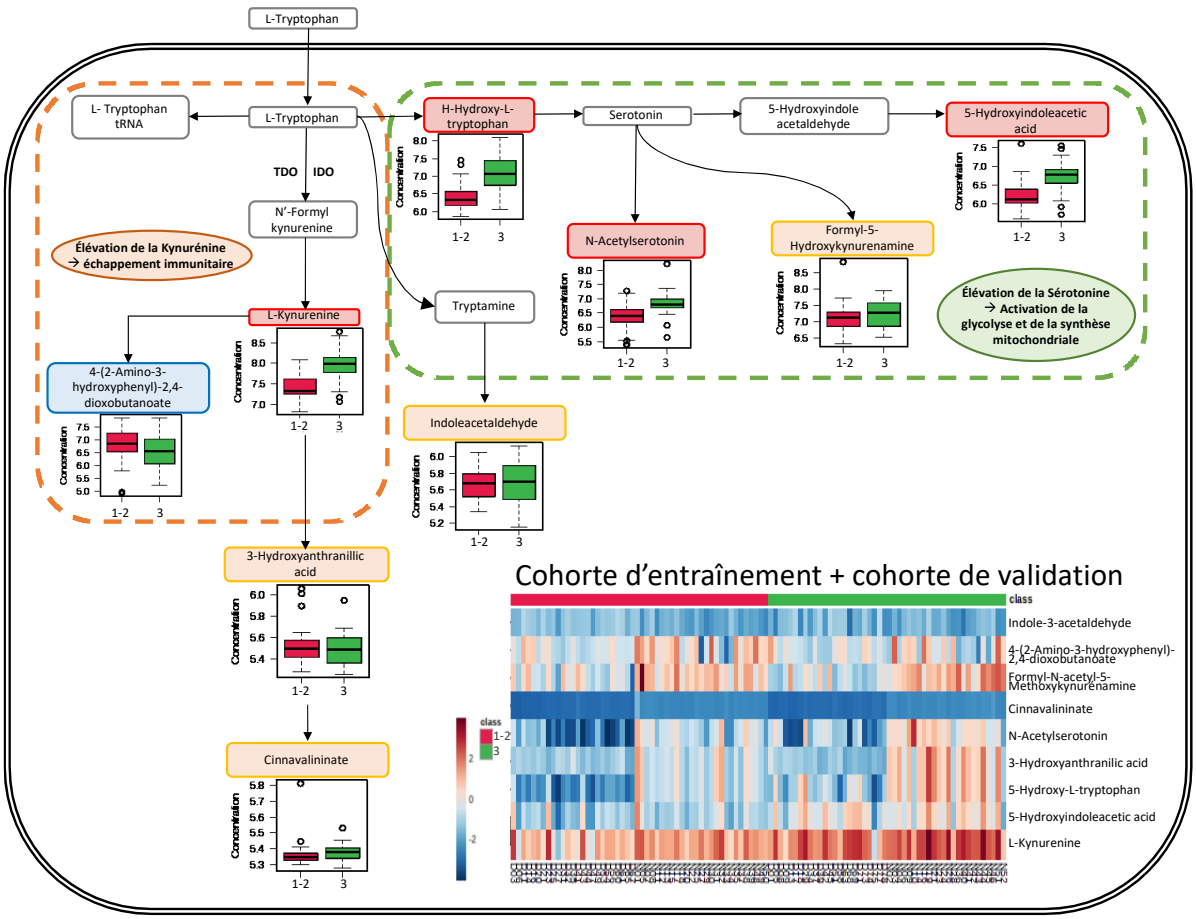


Figure 30. Principaux résultats de l'étude Grade (Article 3). *TDO: Tryptophan-2,3-dioxygenase; IDO: indoléamine 2, 3-dioxygénase.*

cancers

MDPI

Article

Metabolomic Signatures of Scarff–Bloom–Richardson (SBR) Grade in Non-Metastatic Breast Cancer

Caroline Bailleux [1,2], David Chardin [1,3], Jocelyn Gal [4], Jean-Marie Guigonis [1], Sabine Lindenthal [1], Fanny Graslin [1,3], Laurent Arnould [5,6], Alexandre Cagnard [1], Jean-Marc Ferrero [2], Olivier Humbert [1,3] and Thierry Pourcher [1,*]

1 Laboratory Transporter in Imaging and Radiotherapy in Oncology (TIRO), Direction de la Recherche Fondamentale (DRF), Institut des Sciences du Vivant Fréderic Joliot, Commissariat à l'Energie Atomique et aux Énergies Alternatives (CEA), Université Côte d'Azur (UCA), 06100 Nice, France
2 Medical Oncology Department, Centre Antoine Lacassagne, University Côte d'Azur, 06189 Nice, France
3 Department of Nuclear Medicine, Antoine Lacassagne Centre, 06189 Nice, France
4 Department of Epidemiology and Biostatistics, Antoine Lacassagne Centre, University of Côte d'Azur, 06189 Nice, France
5 Department of Tumour Biology and Pathology, Georges-François Leclerc Centre, 21079 Dijon, France
6 Cenre de Ressources Biologiques (CRB) Ferdinand Cabanne, 21000 Dijon, France
* Correspondence: thierry.pourcher@univ-cotedazur.fr; Tel.: +33-4-89-15-35-12

Simple Summary: Breast cancer is a heterogeneous disease with multiple biological, molecular, and histological subtypes. Several metabolomics studies have been performed on breast cancer cells highlighting their metabolic heterogeneity with a potential impact on the efficiency of personalized therapies. In our study, we performed an untargeted metabolomic analysis of breast cancer tumors and identified a metabolic signature for high-grade invasive tumors. AUCs for both the training set and validation set were above 0.88. This result indicates that the model can distinguish high-grade and low-grade tumors with a probability of almost 90%. We also identified several biomarkers of tumor aggressiveness, such as N1,N12-diacetylspermine and tryptophan catabolites, both of which are involved in the inhibition of the immune response. Our study thus provides new insights into the biological mechanisms underlying tumor aggressiveness. Furthermore, the identified biomarkers will enable the development of new strategies for better selection of patients in different immune therapy clinical trials, and thus, for better patient management. All these findings are discussed in relation to the latest publications in the field.

Abstract: Purpose: Identification of metabolomic biomarkers of high SBR grade in non-metastatic breast cancer. Methods: This retrospective bicentric metabolomic analysis included a training set (n = 51) and a validation set (n = 49) of breast cancer tumors, all classified as high-grade (grade III) or low-grade (grade I–II). Metabolomes of tissue samples were studied by liquid chromatography coupled with mass spectrometry. Results: A molecular signature of the top 12 metabolites was identified from a database of 602 frequently predicted metabolites. Partial least squares discriminant analyses showed that accuracies were 0.81 and 0.82, the R2 scores were 0.57 and 0.55, and the Q2 scores were 0.44431 and 0.40147 for the training set and validation set, respectively; areas under the curve for the Receiver Operating Characteristic Curve were 0.882 and 0.886. The most relevant metabolite was diacetylspermine. Metabolite set enrichment analyses and metabolic pathway analyses highlighted the tryptophan metabolism pathway, but the concentration of individual metabolites varied between tumor samples. Conclusions: This study indicates that high-grade invasive tumors are related to diacetylspermine and tryptophan metabolism, both involved in the inhibition of the immune response. Targeting these pathways could restore anti-tumor immunity and have a synergistic effect with immunotherapy. Recent studies could not demonstrate the effectiveness of this strategy, but the use of theragnostic metabolomic signatures should allow better selection of patients.

Keywords: metabolomic signature; breast cancer; SBR grade; immunosuppression

Citation: Bailleux, C.; Chardin, D.; Gal, J.; Guigonis, J.-M.; Lindenthal, S.; Graslin, F.; Arnould, L.; Cagnard, A.; Ferrero, J.-M.; Humbert, O.; et al. Metabolomic Signatures of Scarff–Bloom–Richardson (SBR) Grade in Non-Metastatic Breast Cancer. *Cancers* **2023**, *15*, 1941. https://doi.org/10.3390/cancers15071941

Academic Editors: Maurizio Di Bonito, Michelino De Laurentiis and Monica Cantile

Received: 6 February 2023
Revised: 19 March 2023
Accepted: 21 March 2023
Published: 23 March 2023

1. Introduction

Breast cancer (BC) is a heterogeneous disease that includes several biological, molecular, and histological subtypes. Targeted and non-targeted metabolomics are promising approaches in the field of personalized medicine because they relate to the patient's phenotype as closely as possible [1]. The targeted approach aims to identify a pathway or metabolite of interest based on a previously identified relationship. The untargeted approach seeks to identify and quantify as many metabolites as possible in a sample. Appropriate statistical analyses are then performed to determine which metabolites differ between the sample groups. Metabolite production changes when healthy cells turn into tumor cells with altered metabolism. This leads to metabolomic signatures that can reveal the presence of cancer cells with a specific cell behavior [2].

The study of metabolites in cancer can provide insights into how impaired metabolism can trigger proliferation, angiogenesis, and epithelial-mesenchymal transition (EMT) [3,4]. Because cancer cells have a sustained rate of growth and proliferation that requires a constant supply of metabolic precursors, significant changes in cell metabolism occur [5]. Metabolic reprogramming of cells and adjacent stroma is a key step in cancer development. The current biological model of carcinogenesis highlights various pathways for this process, such as escape from mechanisms involved in cell growth suppression, resistance to cell death, genomic instability and mutations, replication of immortalized cells, induction of metastasis capacity, tumor-induced inflammation, and immune system escape [6,7]. Several metabolomics studies have been performed with breast cancer cells [8–10]. For example, Gong et al. investigated metabolic dysregulation in Triple Negative Breast Cancers (TNBCs) using a multi-omics database. They classified TNBC samples into three heterogeneous metabolic-pathway-based subtypes (lipogenic, glycolytic or mixed) with distinct prognoses, molecular subtype distributions, genomic alterations, and distinct responses to personalized therapies targeting specific metabolic profiles [11]. To our knowledge, there is no publication reporting on studies that have specifically focused on the metabolomics of high-grade tumors.

Alterations in the metabolome can also be used as a potential indicator of breast cancer aggressiveness [12]. For example, metabolites of energy-generating metabolic pathways, such as glycolysis, TCA cycle, and beta-oxidation are present at higher levels in non-hormone-dependent breast cancer and triple-negative breast cancer than in hormone-dependent breast cancer, which correlates with breast cancer aggressiveness [13]. Metabolites of secondary bile acid metabolism, amino acid degradation, short-chain fatty acid production, and deconjugated hormones have also been shown to predict cancer aggressiveness [14–16].

The aim of our study was to identify metabolomic biomarkers specific to high-grade SBR in early-stage breast cancer. After identifying a reliable metabolomic signature, metabolic pathway analyses were performed.

2. Materials and Methods

2.1. Population

The training population consisted of 51 patients treated at our institution (Centre Antoine Lacassagne, Cancer Center of Nice) between March 2013 and September 2016 for a clinical stage I to III_B biopsy-proven breast cancer with an indication for adjuvant therapy after surgery. The validation population consisted of 49 patients treated in another institution (Centre Georges-François Leclerc, Cancer Center of Dijon) between February 2007 and July 2012 for a clinical stage II_A to IV biopsy-proven BC, with an indication for neoadjuvant therapy before surgery. All patients were included retrospectively in the study. The biopsy and tumor resection samples were quick-frozen and stored in the tumor biobanks of our respective facilities. All patients were treated according to current guidelines, with sequential chemotherapy including anthracyclines (epirubicin and cyclophosphamide) and taxanes before or after surgery and radiotherapy. HER2-positive status was defined as IHC3+ or IHC2+/FISH+. Patients with HER2-positive tumors were

treated with trastuzumab and taxanes simultaneously for one year (total duration). Patients with luminal BC were then treated by endocrine therapy with Tamoxifen or an aromatase inhibitor, based on menopausal status.

2.2. Patient Data Collection and Statistical Analysis

Clinical, histological, radiological, and therapeutic data were retrospectively extracted from our facility's digital records or collected by a clinical data monitor, including the SBR (Scarff–Bloom–Richardson) grade used to stratify breast cancer into low, intermediate, and high grades based on the nuclear grade, tubule formation, and mitotic rate [17,18]. Since the two study populations (training set and validation set) were different and to be able to extrapolate our results to real-life study populations, we analyzed and compared the clinical and tumor characteristics between the training set and the validation set using the *t*-student and Fisher's exact test.

2.3. Sample Collection

Samples for the training set were collected during breast surgery. Samples for the validation set were collected during the diagnostic biopsy prior to neoadjuvant chemotherapy. All the samples were quickly deep-frozen and transferred to our facilities' respective biobanks where they were stored at −80 °C until analysis. Samples from Dijon were transported to Nice at −80 °C prior to the metabolomic analysis. All samples were prepared and analyzed in the same facility.

2.4. Sample Preparation

Samples (50–100 mg tumor tissue or 20–40 mg biopsy sample) were placed in 1.5 mL Eppendorf tubes containing 1 mL of methanol, grinded manually with a piston and stored at −20 °C overnight. Samples were then centrifuged at 13,000 rpm for 15 min at 0 °C. Supernatants were transferred into new tubes and placed in a Speed Vac until complete liquid evaporation occurred. Samples were then stored at −80 °C until LC-MS analyses. They were resuspended in 100 µL of a 50% acetonitrile and 50% water mix before LC-MS analysis [19].

2.5. LC-MS Analysis

Liquid chromatography analysis was performed using a DIONEX Ultimate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA, USA). From each sample, 10 µL was injected onto a Synergi 4 µm Hydro-RP 80 Å, 250 × 3.0 mm column (Phenomenex, Le Pecq, France). The mobile phases were composed of 0.1% formic acid (Thermo Fisher Scientific) in water (A) and 0.1% formic acid in acetonitrile (B). The gradient was set as follows with a flow rate of 0.9 mL/min: 0% phase B from 0 to 5 min, 0–95% B from 5 to 21 min, holding at 95% B until 21.5 min, 95–0% B from 21.5 to 22 min, holding at 0% B until 25 min for column equilibration. Mass spectrometry analysis was carried out on a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) with a heated electrospray ionization source, HESI II, operating in both positive and negative mode. High-resolution accurate-mass full-scan MS and the top 5 MS2 spectra were collected in a data-dependent fashion at a resolving power of 70,000 and 35,000 at m/z 400, respectively. This standard procedure has been described in more detail in the cited publications [20–25]. The analyses were performed separately on each of the two groups: the first group consisted of the 51 tumors of the training set and the second of the 49 tumors of the validation set.

2.6. Data Preprocessing and Metabolite Identification

The raw data obtained for the two groups in positive and negative ionization modes were analyzed separately with MzMine (Version 2.38) [26,27]. Individual chromatograms were built for each mass with a noise threshold of 10^5. A local minimum search algorithm was used to select the validated peaks. Peaks were then aligned by RANSAC (random

sample consensus) algorithm with a tolerance of 10 ppm in m/z and 1 min retention time. Missing values were filled, as far as possible, with the same m/z and RT range as observed for detected samples, using the gap-filling tool. Peaks were then predicted using the Human Metabolome DataBase (HMDB, version 3.0) by searching for M + H^+ and M − H^+ ion forms in positive and negative modes, respectively, with a mass tolerance of 15 ppm. Only predicted peaks were included in the final analysis. A linear normalization was performed using the average intensity of each sample as a normalization factor. Only metabolites with no null values after pre-processing were selected for final analysis. If a metabolite was detected in both positive and negative modes, only the mode with the highest average intensity was considered. Finally, a filtering function was applied before statistical analysis selecting only the metabolites with the highest average intensity. This step allowed us to eliminate metabolites that could be considered as background signals or for which quantification was not robust enough.

2.7. Metabolite Selection

The metabolite selection methodology was established as follows to ensure the reproducibility of the analyses. Since the two raw databases (the training set and the validation set) had been merged, only common predicted metabolites were kept. Data were filtered for correlated metabolites, signal intensity, isotope, duplicates, artifacts, and drugs. Metabolite validations were performed with MS2 (from MZmine and/or using Compound Discoverer analysis). MS2 matches of the first 25 metabolites of interest (top list of the statistical analysis) are available in the Supplementary Materials (ms2.xls). The final table with all metabolites is available in the "HMDBval_PLSNice" sheet of the "MS2" Excel file (ms2.xls Supplementary Materials).

2.8. Statistical and Pathway Analyses

All statistical analyses were performed online using MetaboAnalyst (https://www.metaboanalyst.ca/, accessed on 21 December 2022) version 5.0 [28]. The only sample normalization, data transformation, and data scaling method used was the log transformation. Sum or median sample normalizations did not improve the performance of the chemometrics analysis (Principal Component Analysis or PCA; Partial Least Squares Discriminant Analysis or PLS-DA). PLS-DA analysis was used to establish score plots, loading plots, and cross validations (performance accuracy, R2, Q2). Receiver Operating Characteristic (ROC) curves, heatmap graphs, exploration of metabolite set enrichment, and metabolic pathway analyses were generated online using MetaboAnalyst (https://www.metaboanalyst.ca/, accessed on 21 December 2022). The tryptophan pathway was interpreted using data from the SMP and Kegg pathway.

3. Results

3.1. Clinical and Tumor Characteristics

Fifty-one patients were analyzed in the training set and 49 patients in the validation set. Clinical and tumor characteristics are described in Table 1. Median ages were statistically different ($p < 0.00001$) with 65 years (range: 37–88) for the training set and 51 years (range: 26–70) for the validation set. Tumor size, T stage, and N stage also differed statistically with more unfavorable tumor characteristics in the validation set compared to the training set: median tumor size 40 mm, 10.2% of T4, 71.4% of axillary lymph node invasion vs. median tumor size 30 mm, 1.9% of T4, 47.1% of axillary lymph node invasion. These differences could be explained by locally advanced and localized settings. However, the cellular characteristics of the two groups were comparable: the main histological feature was invasive ductal carcinoma (82.5% and 91.8%), almost half of the patients had SBR grade 3 tumors in both populations, and no statistical differences were observed for Ki67, estrogen-receptor, progesterone-receptor, and HER2-receptor status. Despite clinically different study populations, these two groups could therefore be used to analyze intra-tumor cellular aggressiveness.

Table 1. Clinical and tumor characteristics (training set and validation set).

	Training Set		**Validation Set**		
	(n = 51)		(n = 49)		
	N/med	(%/SD)	N/med	(%/SD)	p
Age					$p < 0.00001$ (£)
median	65		51		
min-max	37–88		26–70		
Histology					NS ($)
DIC	48	(82.5%)	45	(91.8%)	
LIC	3	(12.5%)	3	(6.1%)	
other	0	(0.0%)	1	(2.0%)	
Tumor size (mm)	30 *	(21.9)	40 **	(22.4)	$p < 0.00001$ (£)
T					$p = 0.001$ ($)
T1	13	(25.5%)	3	(6.1%)	
T2	26	(51.0%)	37	(75.5%)	
T3	11	(21.6%)	3	(6.1%)	
T4	1	(1.9%)	5	(10.2%)	
unknown	0	(0.0%)	1	(2.0%)	
N					$p = 0.002$ ($)
N0	26	(51.0%)	14	(28.6%)	
N1	18	(35.3%)	34	(69.4%)	
N2	3	(5.9%)	1	(2.0%)	
N3	3	(5.9%)	0	(0.0%)	
unknown	1	(1.9%)	0	(0.0%)	
SBR grading					NS ($)
I	5	(9.8%)	5	(10.2%)	
II	22	(43.1%)	20	(40.8%)	
III	24	(47.1%)	24	(50.0%)	
Ki67%					NS (£)
median	35	(29.3)	60	(23.0)	
≤10%	4	(7.8%)	1	(2.0%)	
Estrogen-receptor					NS (£/$)
Mean	50.2	(47.9)	65.4	(43.6)	
≥10% of cells	29	(56.9%)	28	(57.1%)	
Progesteron-receptor					NS (£/$)
Mean	40.3	(42.5)	43.4	(38.3)	
≥10% of cells	28	(54.9%)	31	(63.3%)	
HER2-positive receptor					NS ($)
HER2 not amplified	40	(78.4%)	41	(83.7%)	
HER2 amplified	11	(21.6%)	8	(16.3%)	

Data retrospectively extracted from digital records or collected by a clinical data monitor. DIC: Ductal Invasive Carcinoma; pT: primary tumor (TNM); pN: regional lymph nodes (TNM); SBR: Scarff–Bloom and Richardson; med: median; SD: standard deviation. * Size assessed on excisional specimen (n = 52). ** Size assessed on ultrasound mammography (n = 48). (£) t-student test. ($) Fisher's exact test. NS: not statistically significant.

3.2. SBR Grade Metabolomic Signature Discriminated between High-Grade (Grade III) and Low-Grade (Grade I–II) Groups

The metabolome from samples collected during breast surgery (training set) and those collected during diagnostic biopsy (validation set) were analyzed by liquid chromatography coupled with mass spectrometry (LC-MS) according to our standard procedures [20,22,29]. Posttreatment of the obtained data generated a database of 602 predicted metabolites. Peak intensities of these predicted metabolites in the 100 tumor samples are included in the Supplementary Materials (training_set.csv and validation_set.csv). Patients in both groups were classified as high-grade (grade III) or low-grade (grade I-II) according to their clinical characteristics. Principal Component Analyses (PCA) performed with MetaboAnalyst showed that the two groups could not be distinguished with this unsupervised method (score plots illustrated in Figure S1A,B). Supervised analyses were subsequently performed on the two cohorts independently. For the training set, the best PLS-DA model was obtained

for three components with an accuracy value of 0.79, R2 = 0.84 and Q2 = 0.38 (Figure 1A,B, values illustrated in Figure S1C–E). For the validation set, the best model was obtained for two components with an accuracy value of 0.78, R2 = 0.69 and Q2 = 0.38 (Figure 1C,D, values illustrated in Figure S1D–F). Multivariate Receiver Operating Characteristic (ROC) curve analyses were also performed using MetaboAnalyst. Areas Under the Curves (AUCs) reached 0.884 (CI95% 0.778–0.995) for the training set and 0.84 (CI95% 0.668–0.969) for the validation set.

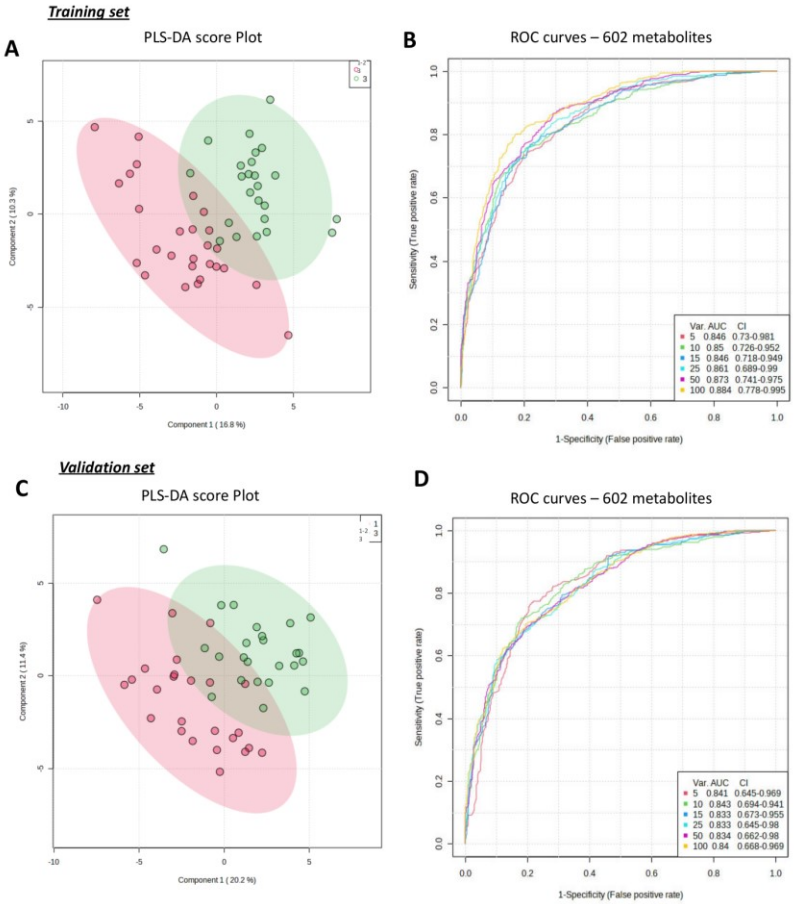


Figure 1. Metabolomic fingerprinting allowed accurate discrimination of SBR grades using 602 predicted metabolites found in both the training set and the validation set. (**A**) shows the score plot of the PLS-DA on the training set, which can accurately discriminate between high-grade (grade III—green dots) and low-grade (grade I–II—red dots) groups. (**B**) shows the AUCs of the ROC of different metabolomic signatures for the training set, which included an increasing number of metabolites (var.), with their respective 95% confidence interval values (95%CI). The score plot of the PLS-DA and ROC curves for the validation set are shown in (**C**,**D**), respectively.

Score plots of PLS-DA analyses using the top 12 metabolites are illustrated in Figure 2A–C. The best models were obtained with two components. After cross-validation, the accuracy values were 0.81 and 0.82, R2 scores were 0.57 and 0.55, and Q2 scores were 0.44431 and 0.40147 for the training set and the validation set, respectively (Figure S2). AUC or ROC curves were 0.882 (CI95% 0.727–0.977) for the training set and 0.886 (CI95% 0.742–0.997) for the validation set (Figure 2B–D). The performance of the grade SBR metabolomic signature could not be improved by either sample normalization (Figure S3) or by increasing the number of metabolites included up to 25 (Figure S4). The top

12 metabolites as well as the top 25 were validated using MS2 matches (for details see Supplementary Materials).

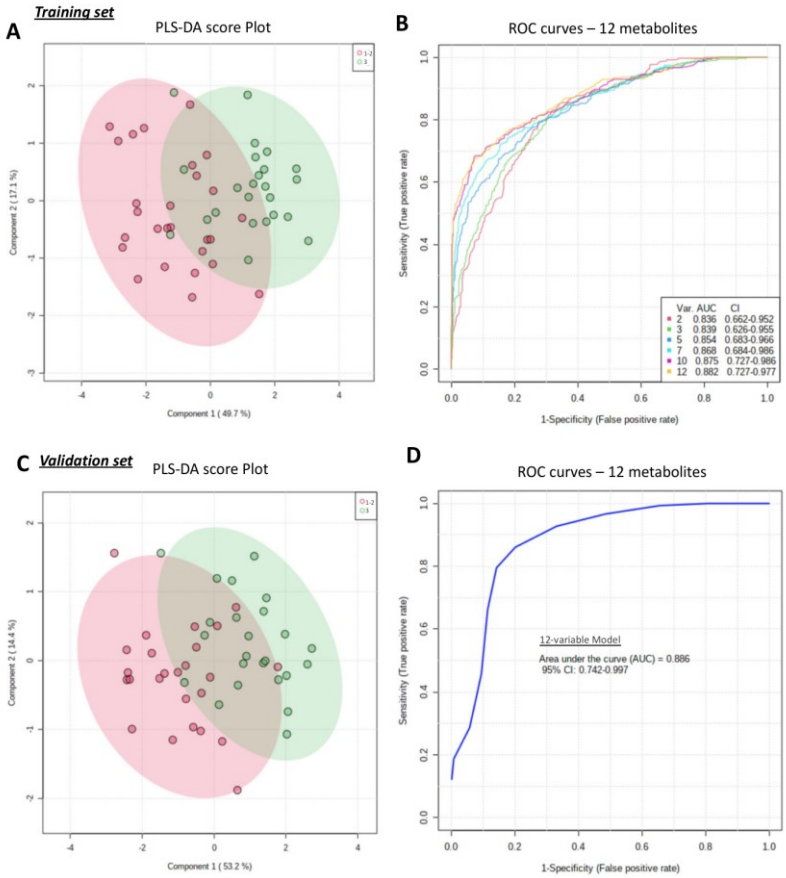


Figure 2. Metabolomic fingerprinting allowed accurate discrimination of SBR grades using the top 12 most important metabolites. The top 12 metabolites (N1,N12-Diacetylspermine, N'Formylkynurenine, N-(1-Deoxy-1-fructosyl)phenyalanine, fructoseglycine, malonylcarnitine, L-L-Homoglutathione, 5-Hydroxy-L-tryptophan, 8-Methpxykynurenate, L-Dopa, L-Kynurenine, N-Acetylproline and 5-Hydroxyindoleacetic acid) were determined from previous Partial Least Squares Discriminant Analyses (PLS-DA—see Figure 1A). (**A**) shows the score plot of the PLS-DA for the training set, which accurately discriminates between high-grade (grade III—green dots) and low-grade (grade I–II—red dots) groups. (**B**) shows the AUCs of the ROC of different metabolomic signatures in the training set, which included an increasing number of metabolites (var.), with their respective 95% confidence interval values (95%CI). The score plot of the PLS-DA and ROC curves for the validation set are shown in (**C**,**D**), respectively.

3.3. PLS-DA Models Identified a Discriminatory Signature with the Top 12 Metabolites

The top 12 metabolites that provide a putative discriminatory signature are shown according to their coefficient scores in Figure 3. These 12 most relevant metabolites were N1,N12-Diacetylspermine (coefficient score = 100), N'Formylkynurenine (coefficient 65.7), N-(1-Deoxy-1-fructosyl)phenyalanine (coefficient 57.3), fructoseglycine (coefficient 53.8), malonylcarnitine (coefficient 49.1), L-L-Homoglutathione (coefficient 48.9), 5-Hydroxy-L-tryptophan (coefficient 46.9), 8-Methpxykynurenate (coefficient 46.5), L-Dopa (coefficient 44.0), L-Kynurenine (coefficient 43.0), N-Acetylproline (coefficient 39.6), and 5-Hydroxyindoleacetic acid (coefficient 39.0).

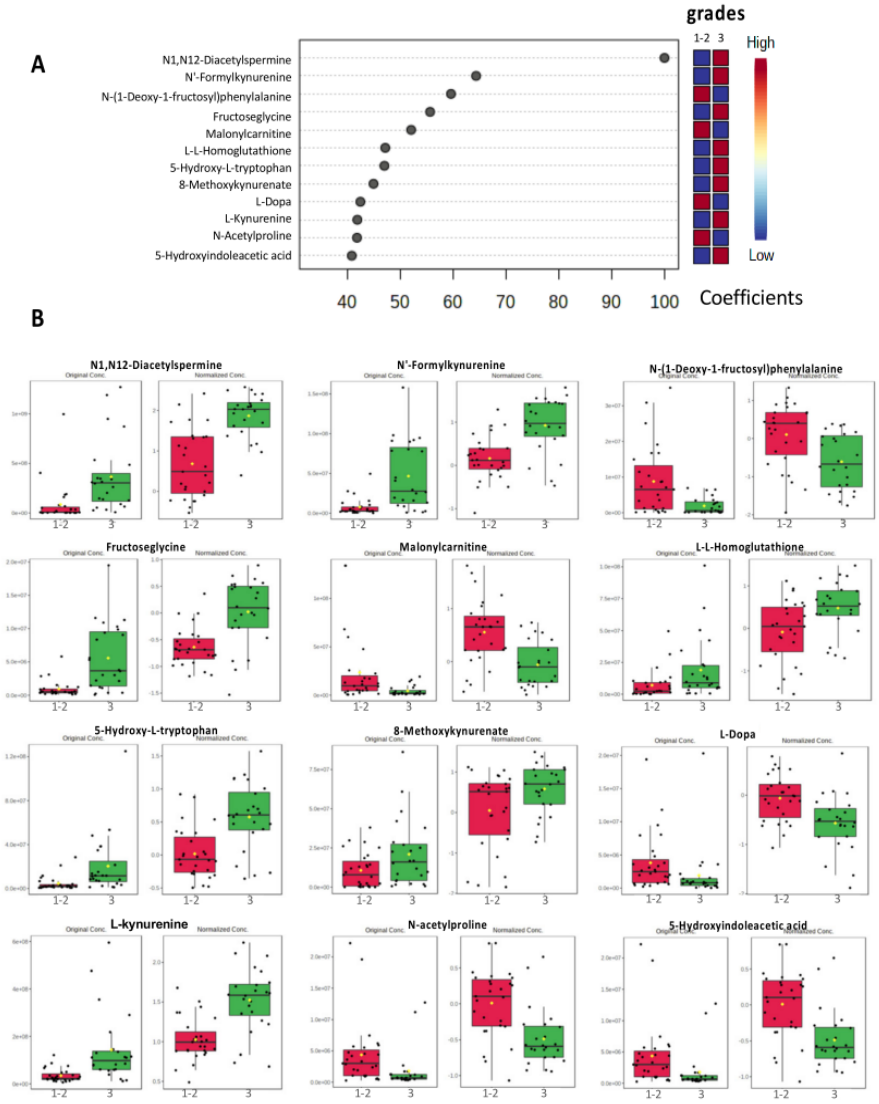
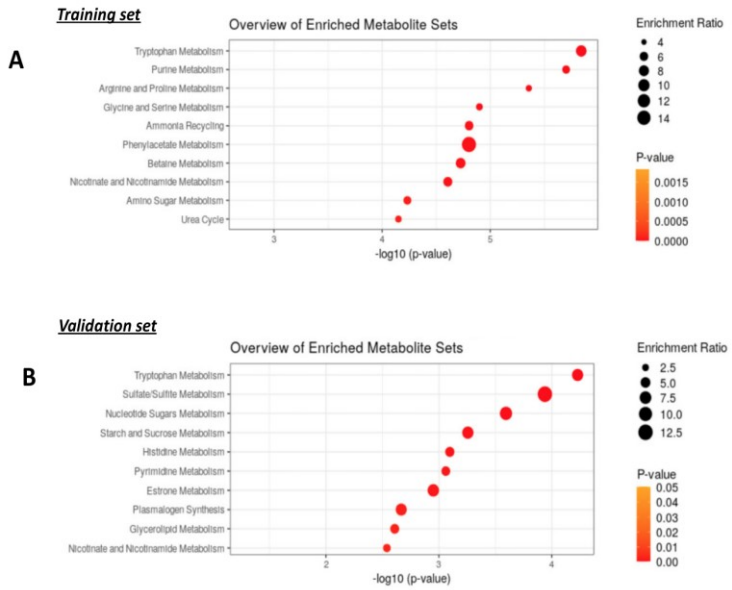


Figure 3. Importance and variation of the top 12 most important metabolites. (**A**) shows the coefficient score plot for the top 12 most important metabolite features identified by PLS-DA. In the right column, the relative concentration of the metabolite is represented in blue when reduced or in red when increased. (**B**) Box plots illustrate the relative concentration of the top 12 most important metabolite features identified by PLS-DA in high-grade (grade III—green boxes) and low-grade (grade I-II—red boxes) groups. The exact names of the metabolites were verified by matching experimental MS2 results with MS2 databases (HMDB).

3.4. Metabolic Pathway Analysis

Metabolite set enrichment analyses were performed separately on the training set and the validation set. For both sets, the most significant metabolic pathway (p-value < 0.0005) with an enrichment ratio of eight was the tryptophan pathway (the top 10 enrichments are shown in Figure 4A,B for both the training and the validation set). A similar result was obtained with metabolic pathway analyses (the top seven common pathways are shown in Figure 4C, more details are shown in Table S1). The most relevant common pathway between the training set and the validation set was the tryptophan metabolism pathway with 9 hits and p-values < 0.00005 (training set: $p = 1.09 \times 10^{-5}$,

validation set: $p = 3.13 \times 10^{-5}$ Figure 4C). The matched metabolites of the tryptophan metabolism pathway were N-Acetylserotonin, 5-Hydroxyindoleacetate, 5-Hydroxy-L-tryptophan, 3-Hydroxyanthranilate, L-Kynurenine, Indole-3-acetaldehyde, Formyl-N-acetyl-5-methoxykynurenamine, Cinnavalininate, and 4-(2-Amino-3-hydroxyphenyl)-2, 4-dioxobutanoate.



Common pathways

C

Pathway	Total Cmpd	Hits	Training Set Raw p	Validation Set Raw p
Tryptophan metabolism	41	9	1.09E-05	3.13E-05
Glyoxylate and dicarboxylate metabolism	32	9	9.12E-05	3.01E-02
Histidine metabolism	16	5	9.43E-04	8.31E-03
Glycine, serine and threonine metabolism	33	12	1.58E-03	3.07E-02
Nicotinate and nicotinamide metabolism	15	5	1.90E-03	3.48E-02
Glutathione metabolism	28	9	2.52E-03	3.17E-02
Pyrimidine metabolism	39	10	5.28E-03	5.41E-04
Glycerolipid metabolism	16	2	8.38E-03	2.08E-02

Figure 4. Metabolite set enrichment analyses and metabolic pathway analyses highlighted the tryptophan metabolism. The top 10 enriched metabolite sets in the analyses performed on the training set and the validation set are shown in (**A**,**B**), respectively. Metabolic pathway analyses were also performed on both sets and the top 7 common significant metabolic pathways are illustrated in (**C**). More details are provided in Table S1.

The analysis of the tryptophan pathway using the KEGG pathway database (Kyoto Encyclopedia of Genes and Genomes (https://www.genome.jp/kegg/pathway.html, accessed on 21 December 2022) (Figure S5) and the SMP database (Small Molecule Pathway) (https://www.smpdb.ca/, accessed on 21 December 2022) revealed an activation of the aromatic amino acid metabolism and serotonin metabolism pathways with a noticeable increase of L-Kynurenine, 5-Hydroxy-L-tryptophan, N-acetylserotonin, and 5-Hydroxyindolacetate in high-grade tumors (results are shown in Figure 5, see also Figure S6). The relative metabolite levels are also presented in a heatmap revealing the considerable variation in metabolite levels between the different samples (included in Figure 5, see also Figure S7).

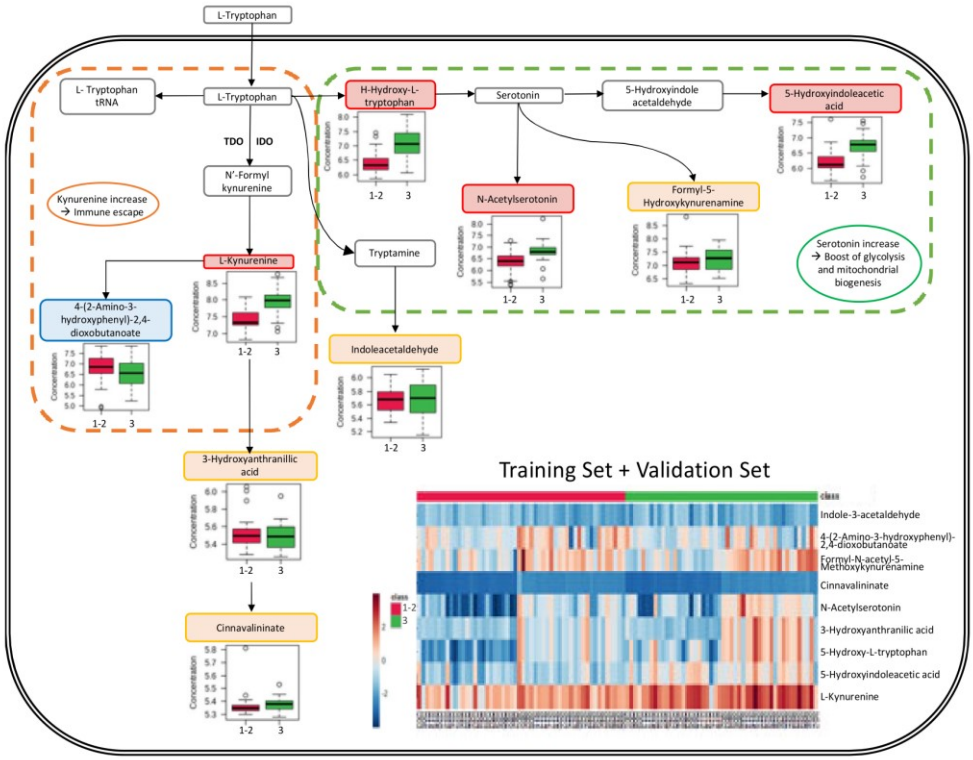


Figure 5. Schematic representation of metabolic pathway changes. The tryptophan pathway illustrations have been adapted from the Small Molecule Pathway database (https://www.smpdb.ca/view/SMP0000063, accessed on 21 December 2022). Box plots illustrate the relative concentration of the main tryptophan catabolites in high-grade (grade III—green boxes) and low-grade (grade I–II—red boxes) groups. Metabolite names are shown in colored boxes: red boxes relate to higher concentrations in high-grade samples; orange boxes to equivalent concentrations in high-grade and low-grade samples; green boxes to lower concentrations in high-grade samples. Heatmap representations of relative concentrations of tryptophan catabolites are shown for all samples (together for the training set and the validation set). Results of the low-grade (grade I–II—red labels on the top line) group are positioned in the left part of the heatmap and those of the high-grade (grade III—green labels) group in the right part.

4. Discussion

This study is the first to analyze the metabolomic profiles of high-grade tumors regardless of their histologic subtype. We identified a metabolic signature for high-grade tumors and obtained AUCs for the training and the validation set above 0.88, showing that our model discriminates high-grade from low-grade tumors with a probability of almost 90%. This signature is not intended to replace the classification system currently used in clinical practice, but it does provide a better analysis of the underlying cellular signaling pathways.

To date, only a few studies have been published on the metabolomic signatures of high-grade SBR. In a study of 139 serum samples from grades I, II, III breast cancer patients and 155 healthy volunteers, Hadi NI et al. [30] showed that the increased levels of glucopyranoside, tetradecane, mannose, and benzene 1,2-dicarboxylic acid allow a differentiation between the various grades. Despite their encouraging results, the authors concluded that a larger sample was needed to further support their findings and to define the metabolic differences between tumor grades more precisely [30]. However, a comparison with our results is not possible because Hadi's group analyzed serum samples while we worked

on tissue samples. In addition, Hadi and her colleagues performed gas chromatography analyses coupled with a mass spectrometer (GC-MS), while we performed LC-MS analyses, which may lead to the identification of different metabolites.

4.1. Strengths and Weaknesses of the Study

We have already performed several studies using similar experimental procedures and have shown that it reliably identifies many metabolites. Despite the use of only one method (LC-MS), this study allowed us to identify and evaluate a large number of metabolites in only small amounts of tumor tissue using biopsy samples (validation set) and comparing them with larger samples from breast surgery (training set). One of the main strengths of our study is that it was conducted on two different sample cohorts from two different patient groups (i.e., biopsies of locally advanced tumors collected from patients in the Dijon aera for the validation set and breast surgery samples of localized tumors on from patients in the Nice aera for the training set). Furthermore, the samples were analyzed in two separate and independent runs (first, the 51 tumors from the training set and second, the 49 tumors from the validation set). This could have led to the statistically significant differences observed in clinical and tumor characteristics, but it also allowed to detect only large differences and identify only robust signatures.

In the present study, metabolic analyses were performed on breast tumor tissue only. No analysis was performed on peripheral blood samples. Since the metabolite signature identified in breast tumor tissue cannot be extrapolated to the signature expected in peripheral blood, it is not suitable for the early detection of tumors in clinical routine. However, in the case of primary surgical treatment, metabolomic analysis of tumor resection samples allows, for example, the prediction of the occurrence of immunosuppression (and thus provides information about the efficiency of a potential immunotherapy).

4.2. N1,N12-Diacetylspermine Metabolite (DiAcSpm)

In our SBR signature, the most relevant metabolite was N1,N12-Diacetylspermine, an alkylamine with multiple amino groups (polyamine). In both sample sets, higher levels of N1,N12-Diacetylspermine were found in high-grade tumor samples than in samples from low-grade tumors (Figure 3). Polyamines are produced during cell division. They are then acetylated in the liver and finally excreted in the urine [31]. MYC is an oncogenic driver of tumor development, progression, and immune-suppression in triple-negative breast cancer (TNBC) [32–35]. A downstream target of MYC is ornithine decarboxylase (ODC), a rate-limiting enzyme of the polyamine metabolism [36,37]. Polyamines have been described to play a functional role in promoting neoplastic transformation and growth [38,39]. Among polyamine derivatives, N1,N12-diacetylspermines have recently attracted much attention in oncology, and urinary diacetylspermines have been described as highly sensitive tumor markers in many cancers, including breast cancer [31,40–42]. Previous studies have shown that high levels of acetylated polyamines are found in breast cancer in association with a simultaneous increase in spermidine and spermine N1 acetyltransferase (SAT1) activity and decreased polyamine oxidase activity [43]. A functional study investigated the effects of spermine on the estrogen receptor (ER) [44]. The obtained results suggest that spermine plays an important role in the regulation of ER ligand-binding and gene activation and thus also in hormone resistance. DiAcSpm was studied by Fahrmann et al. in triple negative breast cancer (TNBC) patients [45]. Serum samples from TNBC patients showed a higher DiAcSpm level than samples from non-TNBC patients and healthy volunteers. In addition, Fahrmann et al. provided evidence that elevated plasma DiAcSpm levels are associated with low immune infiltrate, reduced immune-related gene signatures, early recurrence (<1 year), worse 5-year distant metastasis-free survival and 5-year overall survival. Here, we report the increase of DiAcSpm in breast tissue from low- and high-grade tumors, regardless of their histological subtype.

4.3. Kynurenine Synthesis via the Tryptophan Pathway

The kynurenine to tryptophan catabolism is a known mechanism involved in the modulation of the immune system and has been extensively studied in cancer (33). Tryptophan is converted to kynurenine by indoleamine 2,3-dioxygenase 1 (IDO1), its splice variant IDO2 and tryptophan 2,3-dioxygenase(TDO) [46]. IDO1 is a key factor in maintaining immune tolerance [47]. Its expression increases in response to several inflammatory cytokines, such as interferon-γ, which acts as an endogenous mechanism to prevent an excessive immune response [48]. IDO1 is expressed in multiple tumor types and is associated with reduced activation of cytotoxic cells, increased infiltration of tumor-regulating T-cells, poorer survival rates [49–55], and increased drug resistance [56–59]. Wei et al. [60] measured IDO1 expression in paraffin-embedded breast cancer tissue samples. The group found that IDO is expressed in 64% of the samples. D'Amato et al. [61] suggested an important role for TDO in aggressive breast cancer subtypes, as high TDO levels were found in primary breast tumors associated with shorter overall survival.

Several molecular mechanisms have been described to explain how IDO contributes to tumor-induced tolerance [62–66]. IDO promotes, for example, the formation of immunosuppressive antigen presenting cells (APCs). Furthermore, overexpression of IDO1 in APCs activates the kynurenine pathway, which facilitates kynurenine release and tryptophan consumption. Tryptophan catabolites (kynurenine and its downstream metabolites) operate by activating the aryl hydrocarbon receptor involved in the immune response. Consumption of tryptophan leads to the activation of GCN2 and inhibition of mTOR, which in turn is responsible for Treg differentiation, MDSCS activation and inhibition of T-lymphocytes and natural killer cells [62–66]. In our study, both metabolite-set enrichment analyses and metabolic pathway analyses showed that the tryptophan pathway is more strongly activated in high-grade tumors than in low-grade tumors. The SBR signature of the 12 major metabolites revealed increased levels for 4 tryptophan catabolites (N'-formylkynurenine, 5-hydroxy-L-tryptophan, 8-methoxykynurenate, and L-kynurenine) (Figure 3) in high-grade tumors compared with low-grade tumors. Using breast cancer tissue provided by Duke University Medical Center, Tang et al. showed that kynurenine levels are significantly higher in ER-negative tumors than in ER-positive tumors [67]. Here, we report the first results that associate high-grade tumors with the tryptophan pathway in breast cancer regardless of the histological subtype.

4.4. Serotonin Implications

Our results also suggest a greater activation of the serotonin pathway in high-grade tumors. Although serotonin is mainly known as a neurotransmitter, it is also synthesized by epithelial cells in the mammary gland by tryptophan hydroxylase 1 (TPH1) and plays a role in regulating epithelial homeostasis in breast cancers. Serotonin may produce multiple effects through interaction with a variety of receptors involved in different signaling pathways [68]. The alteration of serotonin and serotonin receptor expression patterns leads to dysregulation of epithelial homeostasis, which has been associated with the initial events of breast cancer development, tumorigenesis and tumor progression [69–72]. Tumors can down-regulate enzymes of serotonin synthesis, decreasing the consumption of tryptophan by the serotonin pathway to increase the consumption of tryptophan by the tryptophan/kynurenine pathway [73]. However, in our study, levels of metabolites of the serotonin pathway were also higher with an increase of N-Acetylserotonin and 5-Hydroxyindoleacetic acid. The serotonin pathway could therefore be involved in tumor aggressiveness in breast cancer independently of the tryptophan/kynurenine pathway. Serotonin has already been shown to affect the proliferation and metabolism of breast cancer cells by triggering two distinct signaling pathways: Jak1/STAT3 which boosts glycolysis by upregulating PKM2, and adenylyl cyclase/PKA which promotes mitochondrial biogenesis [74]. In addition, several studies have suggested that the expression of serotonin and its receptors in immune cells can modulate the immune response, especially in the case of inflammation [75,76]. Other studies have indicated that the immune effects of serotonin

include the suppression of IL-1β and TNF-α release in peripheral blood cells and the activation of T-cells [77]. Here we identified high levels of metabolites of the serotonin pathway in high-grade patients. This finding suggests that the serotonin pathway is involved in the aggressiveness and immunosuppression of high-grade breast cancer.

4.5. Grade and Immune Response

Our study showed that high-grade tumors are related to higher levels of DiAcSp and tryptophan-derived metabolites, both of which are involved in the immune response through Treg differentiation, T cell and natural killer cell inhibition. These findings raise the question of whether the aggressiveness of high-grade tumors could depend on immune escape. Other studies have already indicated that T cells play an essential role in limiting tumor development and that in breast cancer, CD4+ and CD8+ infiltrating T cells are abundant in high-grade ductal carcinoma in situ as well as in invasive carcinoma [78–80]. Higher T_{reg}infiltration is associated with high grade but not with tumor subtype, size of the invasive tumor, lymph node status, or disease stage [81].

Considering these immune escape mechanisms, targeting spermine and tryptophan metabolism could decrease Treg differentiation and reactivate T cells and natural killer cells, thereby reducing immune escape and restoring anti-tumoral immunity. Moreover, targeting both spermine and tryptophan metabolism could create a synergistic effect. Several strategies have been outlined by Peyraud et al. including three different strategies that target the IDO/TDO-Kyn-AhR signaling circuit in cancer treatment: (i) pharmacological inhibition of IDO/TDO by IDO inhibitors, (ii), systemic depletion of Kyn by engineered kynureninase, and (iii) blockade of AhR activation by synthetic AhR modulators [82] (Table 2). To date, no study has been able to demonstrate the benefit of these targeted therapies.

Table 2. Clinical trials targeting the IDO/TDO-Kyn-AhR signaling. Past and recruiting trials, adapted from Peyraud et al. [82] IDO: indoleamine 2,3-dioxygenase; TDO: Tryptophan 2,3-dioxygenase; TNBC: triple-negative breast cancer; BID: twice daily; Q3W: every 3 weeks; ORR: objective response rate; DCR: disease control rate; PR: partial response; SD:s table disease; QD: daily; PD1: programmed cell death protein 1; PD-L1: programmed death-ligand 1; Kyn: kynurenine; AhR: aryl hydrocarbon receptor.

NCT Number	Phase	Number of Patients	Trial Title	Intervention	Main Results
Pharmacological Inhibition of IDO-TDO/IDO Inhibitor					
NCT02178722	I/II	3 TNBC	Study to explore the safety, tolerability and efficacy of MK-3475 combined with INCB024360 in participants with selected cancers	Epacadostat 1 BID combined with pembrolizumab Q3W	Acceptable safety profile TNBC: ORR 10%; DCR 36%
NCT02471846	I	25 (17 TNBC)	A study of GDC-0919 and atezolizumab combination treatment in participants with locally advanced or metastatic solid tumors	Navoximod BID combined with atezolizumab Q3W	Advanced cancer: PR 9%; ORR 10%, SD 24%; Decreasing plasma Kyn with increasing doses
NCT02658890	I/II	627 advanced cancer	An investigational immuno-therapy study of BMS-986205 given combined with nivolumab and combined with both nivolumab and ipilimumab in cancers that are advanced or have spread	Linrodostat combined with immunotherapy (nivolumab or nivolumab+ipilimumab)	Acceptable safety profile No efficicacy results yet
NCT03343613	I	90 advanced cancer	A study of LY3381916 alone or combined with LY3300054 in participants with solid tumors	LY3381916 QD combined with LY3300054 (anti-PD-L1) Q2W	Best response: SD
NCT03328026	I/II	60 breast cancer	Study of SV-BR-1-GM combined with retifanlimab	Epacadostat + Retifanlimab (anti-PD1) + SV-BR-1-GM (vaccine)	Recruiting
Systemic depletion of Kyn/Kynureninase					

Table 2. *Cont.*

NCT Number	Phase	Number of Patients	Trial Title	Intervention	Main Results
-	-	-	-	-	-
Blockade of AhR activation / synthetic AhR modulator					
NCT04200963	I	93 advanced cancer	A phase 1a/b study of IK-175 as a single agent and combined with nivolumab in patients with locally advanced or metastatic solid tumors and urothelial carcinoma	IK-175 combined with nivolumab	Recruiting

Interestingly, our study showed that the activation of the tryptophan pathway was not homogenous among all high-grade patients. Indeed, the L-Kynurenine levels were not high in the analyzed samples from high-grade patients (Figure 5), which may have an impact on the efficacy of the targeted therapies tested. Better selection of targeted therapies for each candidate using previous metabolomic assays may improve efficacy. One should note that, after determination of discriminant biomarkers, accessibility of the targeted metabolomic technique is a major element of applicability in routine care. Such putative personalized medicine will be analyzed in further studies. Finally, with the advent of immunotherapy in neo-adjuvant [83] and first-line metastatic [84] triple-negative breast cancer, the theragnostic value of the activation of these metabolic pathways may be analyzed in the future.s

5. Conclusions

Here, we report the identification of a metabolic signature for high-grade invasive tumors with AUCs greater than 0.88 on both the training set and the validation set, suggesting that the model has a nearly 90% chance of being able to distinguish high-grade from low-grade tumors. This may be of interest in cases of heterogeneity but essentially confirms the performance of the metabolomic analysis. Our results showed that high-grade invasive tumors are related to the metabolism of DiAcSp and tryptophan, both involved in the inhibition of the immune response. Targeting these pathways could restore anti-tumor immunity or activate immunogenicity and create a synergistic effect with immunotherapy. Although the efficacy of this strategy has not been demonstrated in recent studies, metabolic analysis may allow better selection of the most appropriate therapy for each patient. Personalized immunotherapy using theragnostic metabolomic signatures needs to be evaluated in further studies.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers15071941/s1, Figure S1: Untargeted metabolomic statistical analyses of SBR grades of the training set (n = 51) and the validation set (n = 49) using 602 metabolites; Figure S2: Untargeted metabolomic statistical analyses of SBR grades in the training set (n = 51) and the validation set (n = 49) using the top 12 metabolites; Figure S3: Metabolomic fingerprinting allowed accurate discrimination of SBR grades using the top 25 most important metabolites; Figure S4: Untargeted metabolomic statistical analyses of SBR grades of the training set and validation set using the top 25 metabolites; Figure S5: Schematic representation of metabolic pathway changes; Figure S6: Schematic representation of metabolic pathway changes; Figure S7: Heatmap representation of relative concentrations of tryptophan catabolites of the training set and the validation set; Table S1: Significant pathways of the SBR grade analysis.

Author Contributions: Conceptualization, C.B., D.C., L.A., O.H. and T.P.; methodology, J.G.; software, D.C., C.B. and T.P.; validation, A.C., C.B., J.-M.G. and T.P.; formal analysis, J.-M.G., C.B., F.G. and T.P.; investigation, C.B, J.-M.F. and O.H.; writing—original draft preparation, C.B. and T.P.; writing—review and editing, C.B., T.P., S.L., D.C., L.A., O.H., J.-M.F. and J.G.; supervision, T.P., O.H. and J.-M.F. All authors have read and agreed to the published version of the manuscript.

Funding: Equipment for this study was purchased through grants from the Recherche en Matières de Sûreté Nucléaire et Radioprotection program from the French National Research Agency and the Conseil Départemental 06.

Institutional Review Board Statement: Tissue collection and analyses were approved by French ethics committees (French National Commission for Informatics and Liberties N°17003 and National Institute Health data N°1515251018).

Informed Consent Statement: Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: Peak intensities of theses predicted metabolites in the 100 tumor samples are provided as Supplementary Materials (training_set.csv and validation_set.csv). MS2 matches of the first 25 metabolites of interest (top list of the statistical analysis) are available in the Supplementary Materials (ms2.xls).

Acknowledgments: Our thanks go to all TIRO Team members and to the Antoine Lacassagne Center. The authors thank the Cancer Center of Dijon for providing the samples for the validation set. We thank Yvonne van der Does for editorial correction of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

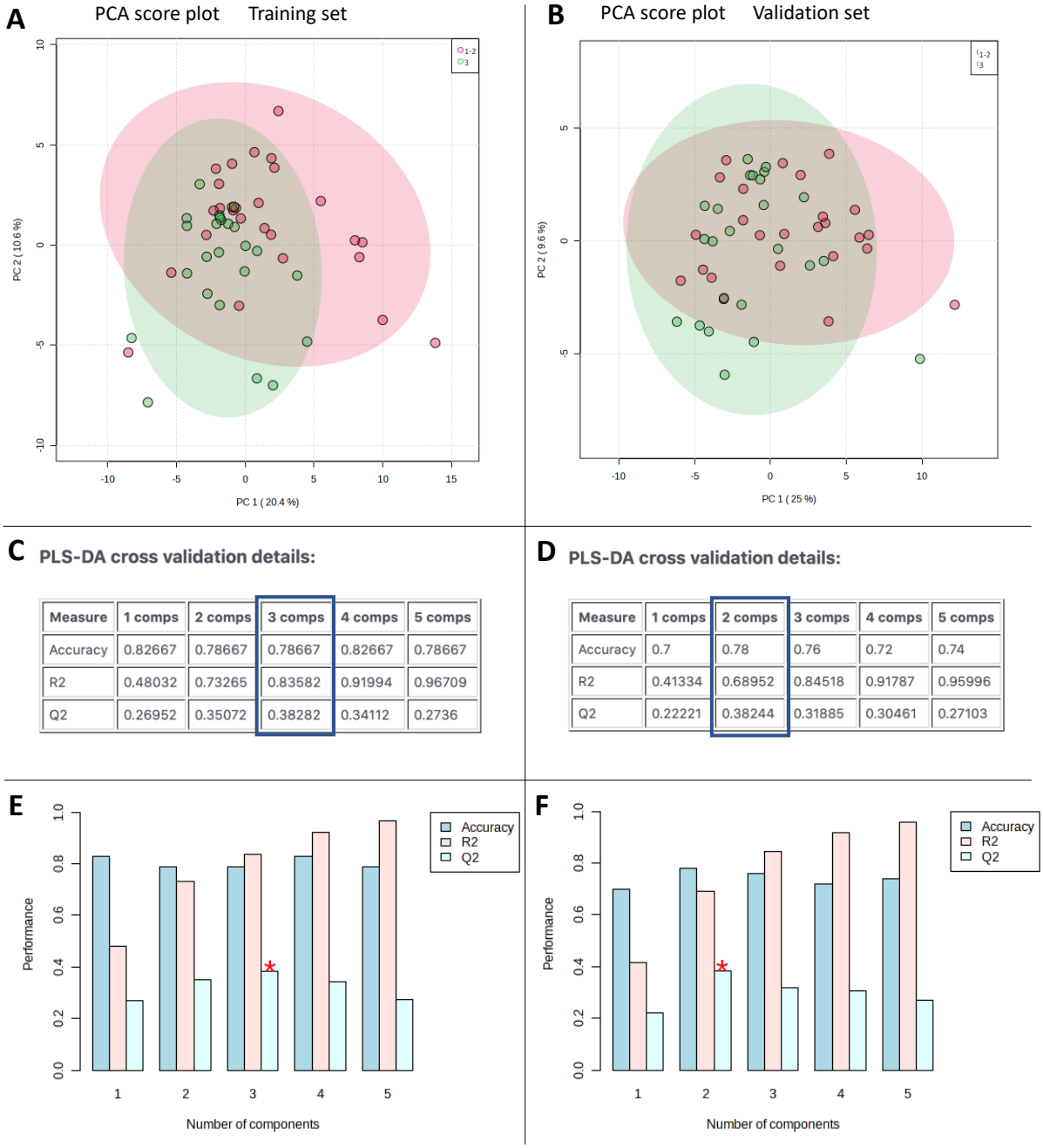
References

1. Aboud, O.A.; Weiss, R.H. New Opportunities from the Cancer Metabolome. *Clin. Chem.* **2013**, *59*, 138–146. [CrossRef] [PubMed]
2. Cardoso, M.; Santos, J.; Ribeiro, M.; Talarico, M.; Viana, L.; Derchain, S. A Metabolomic Approach to Predict Breast Cancer Behavior and Chemotherapy Response. *Int. J. Mol. Sci.* **2018**, *19*, 617. [CrossRef] [PubMed]
3. Mikó, E.; Kovács, T.; Sebő, É.; Tóth, J.; Csonka, T.; Ujlaki, G.; Sipos, A.; Szabó, J.; Méhes, G.; Bai, P. Microbiome—Microbial Metabolome—Cancer Cell Interactions in Breast Cancer—Familiar, but Unexplored. *Cells* **2019**, *8*, 293. [CrossRef] [PubMed]
4. Wu, J.; Yang, R.; Zhang, L.; Li, Y.; Liu, B.; Kang, H.; Fan, Z.; Tian, Y.; Liu, S.; Li, T. Metabolomics research on potential role for 9-cis-retinoic acid in breast cancer progression. *Cancer Sci.* **2018**, *109*, 2315–2326. [CrossRef] [PubMed]
5. Kapoore, R.V.; Coyle, R.; Staton, C.A.; Brown, N.J.; Vaidyanathan, S. Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: A case study with the metastatic breast cancer cell line MDA-MB-231. *Analyst* **2017**, *142*, 2038–2049. [CrossRef] [PubMed]
6. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef]
7. Hanahan, D. Rethinking the war on cancer. *Lancet* **2014**, *383*, 558–563. [CrossRef]
8. Subramani, R.; Poudel, S.; Smith, K.D.; Estrada, A.; Lakshmanaswamy, R. Metabolomics of Breast Cancer: A Review. *Metabolites* **2022**, *12*, 643. [CrossRef] [PubMed]
9. Pal, A.K.; Sharma, P.; Zia, A.; Siwan, D.; Nandave, D.; Nandave, M.; Gautam, R.K. Metabolomics and EMT Markers of Breast Cancer: A Crosstalk and Future Perspective. *Pathophysiology* **2022**, *29*, 17. [CrossRef]
10. Chen, Z.; Li, Z.; Li, H.; Jiang, Y. Metabolomics: A promising diagnostic and therapeutic implement for breast cancer. *OncoTargets Ther.* **2019**, *12*, 6797–6811. [CrossRef]
11. Gong, Y.; Ji, P.; Yang, Y.-S.; Xie, S.; Yu, T.-J.; Xiao, Y.; Jin, M.-L.; Ma, D.; Guo, L.-W.; Pei, Y.-C.; et al. Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. *Cell Metab.* **2021**, *33*, 51–64.e9. [CrossRef] [PubMed]
12. Dougan, M.M.; Li, Y.; Chu, L.W.; Haile, R.W.; Whittemore, A.S.; Han, S.S.; Moore, S.C.; Sampson, J.N.; Andrulis, I.L.; John, E.M.; et al. Metabolomic profiles in breast cancer: A pilot case-control study in the breast cancer family registry. *BMC Cancer* **2018**, *18*, 532. [CrossRef]
13. Kanaan, Y.M.; Sampey, B.P.; Beyene, D.; Esnakula, A.K.; Naab, T.J.; Ricks-Santi, L.J.; Dasi, S.; Day, A.; Blackman, K.W.; Frederick, W.; et al. Metabolic profile of triple-negative breast cancer in African-American women reveals potential biomarkers of aggressive disease. *Cancer Genom. Proteom.* **2014**, *11*, 279–294.
14. Kisanga, E.R.; Mellgren, G.; Lien, E.A. Excretion of hydroxylated metabolites of tamoxifen in human bile and urine. *Anticancer Res.* **2005**, *25*, 4487–4492.
15. Visekruna, A.; Luu, M. The Role of Short-Chain Fatty Acids and Bile Acids in Intestinal and Liver Function, Inflammation, and Carcinogenesis. *Front. Cell Dev. Biol.* **2021**, *9*, 703218. [CrossRef] [PubMed]
16. Arnone, A.A.; Cline, J.M.; Soto-Pantoja, D.R.; Cook, K.L. Investigating the role of endogenous estrogens, hormone replacement therapy, and blockade of estrogen receptor-α activity on breast metabolic signaling. *Breast Cancer Res. Treat.* **2021**, *190*, 53–67. [CrossRef]
17. Scarff, R.; Torloni, H. *Histological Typing of Breast Tumors*; International Histological Classification of Tumours; World Health Organization: Geneva, Switerland, 1968; Volume 2, pp. 13–20.
18. Bloom, H.J.G.; Richardson, W.W. Histological Grading and Prognosis in Breast Cancer: A Study of 1409 Cases of which 359 have been Followed for 15 Years. *Br. J. Cancer* **1957**, *11*, 359–377. [CrossRef]

19. Prasad Maharjan, R.; Ferenci, T. Global metabolite analysis: The influence of extraction methodology on metabolome profiles of Escherichia coli. *Anal. Biochem.* **2003**, *313*, 145–154. [CrossRef] [PubMed]
20. Jing, L.; Guigonis, J.-M.; Borchiellini, D.; Durand, M.; Pourcher, T.; Ambrosetti, D. LC-MS based metabolomic profiling for renal cell carcinoma histologic subtypes. *Sci. Rep.* **2019**, *9*, 15635. [CrossRef]
21. Hichri, M.; Vassaux, G.; Guigonis, J.-M.; Juhel, T.; Graslin, F.; Guglielmi, J.; Pourcher, T.; Cambien, B. Proteomic Analysis of Iodinated Contrast Agent-Induced Perturbation of Thyroid Iodide Uptake. *J. Cell. Mol.* **2020**, *9*, 329. [CrossRef] [PubMed]
22. Suissa, L.; Guigonis, J.-M.; Graslin, F.; Doche, E.; Osman, O.; Chau, Y.; Sedat, J.; Lindenthal, S.; Pourcher, T. Metabolome of Cerebral Thrombi Reveals an Association between High Glycemia at Stroke Onset and Good Clinical Outcome. *Metabolites* **2020**, *10*, 483. [CrossRef] [PubMed]
23. Suissa, L.; Flachon, V.; Guigonis, J.-M.; Olivieri, C.-V.; Burel-Vandenbos, F.; Guglielmi, J.; Ambrosetti, D.; Gérard, M.; Franken, P.; Darcourt, J.; et al. Urinary ketone body loss leads to degeneration of brain white matter in elderly SLC5A8-deficient mice. *J. Cereb. Blood Flow Metab.* **2020**, *40*, 1709–1723. [CrossRef]
24. Suissa, L.; Guigonis, J.-M.; Graslin, F.; Robinet-Borgomano, E.; Chau, Y.; Sedat, J.; Lindenthal, S.; Pourcher, T. Combined Omic Analyzes of Cerebral Thrombi: A New Molecular Approach to Identify Cardioembolic Stroke Origin. *Stroke* **2021**, *52*, 2892–2901. [CrossRef]
25. Castillo-Rivera, F.; Ondo-Méndez, A.; Guglielmi, J.; Guigonis, J.-M.; Jing, L.; Lindenthal, S.; Gonzalez, A.; López, D.; Cambien, B.; Pourcher, T. Tumor microenvironment affects exogenous sodium/iodide symporter expression. *Transl. Oncol.* **2021**, *14*, 100937. [CrossRef] [PubMed]
26. Katajamaa, M.; Orešič, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinform.* **2005**, *6*, 179. [CrossRef]
27. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform.* **2010**, *11*, 395. [CrossRef] [PubMed]
28. Pang, Z.; Zhou, G.; Ewald, J.; Chang, L.; Hacariz, O.; Basu, N.; Xia, J. Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat. Protoc.* **2022**, *17*, 1735–1761. [CrossRef] [PubMed]
29. Occelli, C.; Guigonis, J.-M.; Lindenthal, S.; Cagnard, A.; Graslin, F.; Brglez, V.; Seitz-Polski, B.; Dellamonica, J.; Levraut, J.; Pourcher, T. Untargeted plasma metabolomic fingerprinting highlights several biomarkers for the diagnosis and prognosis of coronavirus disease 19. *Front. Med.* **2022**, *9*, 995069. [CrossRef] [PubMed]
30. Hadi, N.I.; Jamal, Q.; Iqbal, A.; Shaikh, F.; Somroo, S.; Musharraf, S.G. Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry. *Sci. Rep.* **2017**, *7*, 1715. [CrossRef]
31. Kato, M.; Onishi, H.; Matsumoto, K.; Motoshita, J.; Tsuruta, N.; Higuchi, K.; Katano, M. Prognostic significance of urine N1, N12-diacetylspermine in patients with non-small cell lung cancer. *Anticancer Res.* **2014**, *34*, 3053–3059. [PubMed]
32. Xu, J.; Chen, Y.; Olopade, O.I. MYC and Breast Cancer. *Genes Cancer* **2010**, *1*, 629–640. [CrossRef]
33. Fallah, Y.; Brundage, J.; Allegakoen, P.; Shajahan-Haq, A.N. MYC-Driven Pathways in Breast Cancer Subtypes. *Biomolecules* **2017**, *7*, 53. [CrossRef]
34. Gatza, M.L.; Lucas, J.E.; Barry, W.T.; Kim, J.W.; Wang, Q.; Crawford, M.D.; Datto, M.B.; Kelley, M.; Mathey-Prevot, B.; Potti, A.; et al. A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6994–6999. [CrossRef] [PubMed]
35. Zimmerli, D.; Brambillasca, C.S.; Talens, F.; Bhin, J.; Linstra, R.; Romanens, L.; Bhattacharya, A.; Joosten, S.E.P.; Da Silva, A.M.; Padrao, N.; et al. MYC promotes immune-suppression in triple-negative breast cancer via inhibition of interferon signaling. *Nat. Commun.* **2022**, *13*, 6579. [CrossRef] [PubMed]
36. Bello-Fernandez, C.; Packham, G.; Cleveland, J.L. The ornithine decarboxylase gene is a transcriptional target of c-Myc. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7804–7808. [CrossRef]
37. Bachmann, A.S.; Geerts, D. Polyamine synthesis as a target of MYC oncogenes. *J. Biol. Chem.* **2018**, *293*, 18757–18769. [CrossRef]
38. Funakoshi-Tago, M.; Sumi, K.; Kasahara, T.; Tago, K. Critical Roles of Myc-ODC Axis in the Cellular Transformation Induced by Myeloproliferative Neoplasm-Associated JAK2 V617F Mutant. *PLoS ONE* **2013**, *8*, e52844. [CrossRef] [PubMed]
39. Casero, R.A.; Marton, L.J. Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. *Nat. Rev. Drug Discov.* **2007**, *6*, 373–390. [CrossRef]
40. Sugimoto, M.; Hiramatsu, K.; Kamei, S.; Kinoshita, K.; Hoshino, M.; Iwasaki, K.; Kawakita, M. Significance of urinaryN 1,N 8-diacetylspermidine andN 1,N 12-diacetylspermine as indicators of neoplastic diseases. *J. Cancer Res. Clin. Oncol.* **1995**, *121*, 317–319. [CrossRef]
41. Hiramatsu, K.; Sugimoto, M.; Kamei, S.; Hoshino, M.; Kinoshita, K.; Iwasaki, K.; Kawakita, M. Diagnostic and prognostic usefulness of N 1, N 8 -diacetylspermidine and N 1, N 12 -diacetylspermine in urine as novel markers of malignancy. *J. Cancer Res. Clin. Oncol.* **1997**, *123*, 539–545. [CrossRef]
42. Hiramatsu, K.; Takahashi, K.; Yamaguchi, T.; Matsumoto, H.; Miyamoto, H.; Tanaka, S.; Tanaka, C.; Tamamori, Y.; Imajo, M.; Kawaguchi, M.; et al. *N* 1, *N* 12-Diacetylspermine as a Sensitive and Specific Novel Marker for Early- and Late-Stage Colorectal and Breast Cancers. *Clin. Cancer Res.* **2005**, *11*, 2986–2990. [CrossRef]

43. Cervelli, M.; Bellavia, G.; Fratini, E.; Amendola, R.; Polticelli, F.; Barba, M.; Federico, R.; Signore, F.; Gucciardo, G.; Grillo, R.; et al. Spermine oxidase (SMO) activity in breast tumor tissues and biochemical analysis of the anticancer spermine analogues BENSpm and CPENSpm. *BMC Cancer* **2010**, *10*, 555. [CrossRef] [PubMed]
44. Lu, B.; Liang, X.; Scott, G.K.; Chang, C.-H.; Baldwin, M.A.; Thomas, T.; Benz, C.C.; Weinstein, I.B. Polyamine inhibition of estrogen receptor (ER) DNA-binding and ligand-binding functions. *Breast Cancer Res. Treat.* **1998**, *48*, 243–257. [CrossRef]
45. Fahrmann, J.F.; Vykoukal, J.; Fleury, A.; Tripathi, S.; Dennison, J.B.; Murage, E.; Wang, P.; Yu, C.-Y.; Capello, M.; Creighton, C.J.; et al. Association between Plasma Diacetylspermine and Tumor Spermine Synthase with Outcome in Triple-Negative Breast Cancer. *J. Natl. Cancer Inst.* **2020**, *112*, 607–616. [CrossRef]
46. Platten, M.; Wick, W.; Van den Eynde, B.J. Tryptophan Catabolism in Cancer: Beyond IDO and Tryptophan Depletion. *Cancer Res.* **2012**, *72*, 5435–5440. [CrossRef] [PubMed]
47. Wu, H.; Gong, J.; Liu, Y. Indoleamine 2, 3-dioxygenase regulation of immune response (Review). *Mol. Med. Rep.* **2018**, *17*, 4867–4873. [CrossRef]
48. Muller, A.J.; Sharma, M.D.; Chandler, P.R.; DuHadaway, J.B.; Everhart, M.E.; Johnson, B.A.; Kahler, D.J.; Pihkala, J.; Soler, A.P.; Munn, D.H.; et al. Chronic inflammation that facilitates tumor progression creates local immune suppression by inducing indoleamine 2,3 dioxygenase. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17073–17078. [CrossRef]
49. Brandacher, G.; Perathoner, A.; Ladurner, R.; Schneeberger, S.; Obrist, P.; Winkler, C.; Werner, E.R.; Werner-Felmayer, G.; Weiss, H.G.; G√∂bel, G.; et al. Prognostic value of indoleamine 2,3-dioxygenase expression in colorectal cancer: Effect on tumor-infiltrating T cells. *Clin. Cancer Res.* **2006**, *12*, 1144–1151. [CrossRef]
50. Ino, K.; Yamamoto, E.; Shibata, K.; Kajiyama, H.; Yoshida, N.; Terauchi, M.; Nawa, A.; Nagasaka, T.; Takikawa, O.; Kikkawa, F. Inverse Correlation between Tumoral Indoleamine 2,3-Dioxygenase Expression and Tumor-Infiltrating Lymphocytes in Endometrial Cancer: Its Association with Disease Progression and Survival. *Clin. Cancer Res.* **2008**, *14*, 2310–2317. [CrossRef]
51. Ino, K.; Yoshida, N.; Kajiyama, H.; Shibata, K.; Yamamoto, E.; Kidokoro, K.; Takahashi, N.; Terauchi, M.; Nawa, A.; Nomura, S.; et al. Indoleamine 2,3-dioxygenase is a novel prognostic indicator for endometrial cancer. *Br. J. Cancer* **2006**, *95*, 1555–1561. [CrossRef] [PubMed]
52. Okamoto, A.; Nikaido, T.; Ochiai, K.; Takakura, S.; Saito, M.; Aoki, Y.; Ishii, N.; Yanaihara, N.; Yamada, K.; Takikawa, O.; et al. Indoleamine 2,3-Dioxygenase Serves as a Marker of Poor Prognosis in Gene Expression Profiles of Serous Ovarian Cancer Cells. *Clin. Cancer Res.* **2005**, *11*, 6030–6039. [CrossRef]
53. Nakamura, T.; Shima, T.; Saeki, A.; Hidaka, T.; Nakashima, A.; Takikawa, O.; Saito, S. Expression of indoleamine 2,3-dioxygenase and the recruitment of Foxp3-expressing regulatory T cells in the development and progression of uterine cervical cancer. *Cancer Sci.* **2007**, *98*, 874–881. [CrossRef] [PubMed]
54. Witkiewicz, A.; Williams, T.K.; Cozzitorto, J.; Durkan, B.; Showalter, S.L.; Yeo, C.J.; Brody, J.R. Expression of Indoleamine 2,3-Dioxygenase in Metastatic Pancreatic Ductal Adenocarcinoma Recruits Regulatory T Cells to Avoid Immune Detection. *J. Am. Coll. Surg.* **2008**, *206*, 849–854. [CrossRef]
55. Brody, J.R.; Costantino, C.L.; Berger, A.C.; Sato, T.; Lisanti, M.P.; Yeo, C.J.; Emmons, R.V.; Witkiewicz, A.K. Expression of indoleamine 2,3-dioxygenase in metastatic malignant melanoma recruits regulatory T cells to avoid immune detection and affects survival. *Cell Cycle* **2009**, *8*, 1930–1934. [CrossRef] [PubMed]
56. Seeber, A.; Klinglmair, G.; Fritz, J.; Steinkohl, F.; Zimmer, K.; Aigner, F.; Horninger, W.; Gastl, G.; Zelger, B.; Brunner, A.; et al. High IDO -1 expression in tumor endothelial cells is associated with response to immunotherapy in metastatic renal cell carcinoma. *Cancer Sci.* **2018**, *109*, 1583–1591. [CrossRef]
57. Creelan, B.C.; Antonia, S.J.; Bepler, G.; Garrett, T.J.; Simon, G.R.; Soliman, H.H. Indoleamine 2,3-dioxygenase activity and clinical outcome following induction chemotherapy and concurrent chemoradiation in Stage III non-small cell lung cancer. *OncoImmunology* **2013**, *2*, e23428. [CrossRef] [PubMed]
58. Wang, W.; Huang, L.; Jin, J.-Y.; Jolly, S.; Zang, Y.; Wu, H.; Yan, L.; Pi, W.; Li, L.; Mellor, A.L.; et al. IDO Immune Status after Chemoradiation May Predict Survival in Lung Cancer Patients. *Cancer Res.* **2018**, *78*, 809–816. [CrossRef] [PubMed]
59. Botticelli, A.; Cerbelli, B.; Lionetto, L.; Zizzari, I.; Salati, M.; Pisano, A.; Federica, M.; Simmaco, M.; Nuti, M.; Marchetti, P. Can IDO activity predict primary resistance to anti-PD-1 treatment in NSCLC? *J. Transl. Med.* **2018**, *16*, 219. [CrossRef] [PubMed]
60. Wei, L.; Zhu, S.; Li, M.; Li, F.; Wei, F.; Liu, J.; Ren, X. High Indoleamine 2,3-Dioxygenase Is Correlated with Microvessel Density and Worse Prognosis in Breast Cancer. *Front. Immunol.* **2018**, *9*, 724. [CrossRef]
61. D'Amato, N.C.; Rogers, T.J.; Gordon, M.A.; Greene, L.I.; Cochrane, D.R.; Spoelstra, N.S.; Nemkov, T.G.; D'Alessandro, A.; Hansen, K.C.; Richer, J.K. A TDO2-AhR Signaling Axis Facilitates Anoikis Resistance and Metastasis in Triple-Negative Breast Cancer. *Cancer Res.* **2015**, *75*, 4651–4664. [CrossRef] [PubMed]
62. Terness, P.; Bauer, T.M.; Röse, L.; Dufter, C.; Watzlik, A.; Simon, H.; Opelz, G. Inhibition of Allogeneic T Cell Proliferation by Indoleamine 2,3-Dioxygenase–expressing Dendritic Cells. *J. Exp. Med.* **2002**, *196*, 447–457. [CrossRef]
63. Frumento, G.; Rotondo, R.; Tonetti, M.; Damonte, G.; Benatti, U.; Ferrara, G.B. Tryptophan-derived Catabolites Are Responsible for Inhibition of T and Natural Killer Cell Proliferation Induced by Indoleamine 2,3-Dioxygenase. *J. Exp. Med.* **2002**, *196*, 459–468. [CrossRef] [PubMed]
64. Prendergast, G.C.; Malachowski, W.J.; Mondal, A.; Scherle, P.; Muller, A.J. Indoleamine 2,3-Dioxygenase and Its Therapeutic Inhibition in Cancer. *Int. Rev. Cell Mol. Biol.* **2018**, *336*, 175–203. [PubMed]

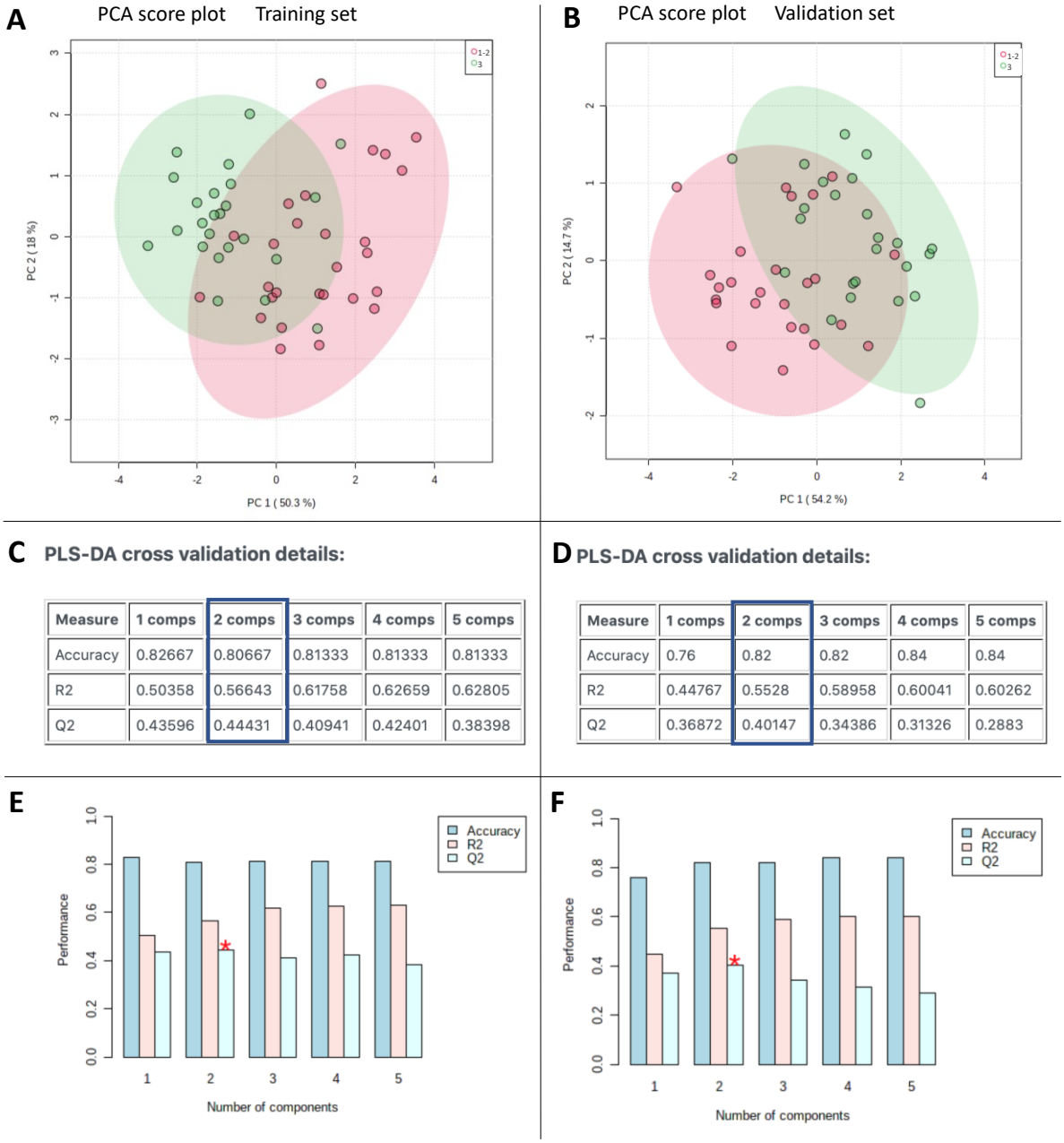
65. Ye, Z.; Yue, L.; Shi, J.; Shao, M.; Wu, T. Role of IDO and TDO in Cancers and Related Diseases and the Therapeutic Implications. *J. Cancer* **2019**, *10*, 2771–2782. [CrossRef] [PubMed]
66. Ramapriyan, R.; Caetano, M.S.; Barsoumian, H.B.; Mafra, A.C.P.; Zambalde, E.P.; Menon, H.; Tsouko, E.; Welsh, J.W.; Cortez, M.A. Altered cancer metabolism in mechanisms of immunotherapy resistance. *Pharmacol. Ther.* **2019**, *195*, 162–171. [CrossRef]
67. Tang, X.; Lin, C.-C.; Spasojevic, I.; Iversen, E.S.; Chi, J.-T.; Marks, J.R. A joint analysis of metabolomics and genetics of breast cancer. *Breast Cancer Res.* **2014**, *16*, 415. [CrossRef]
68. Olfati, Z.; Rigi, G.; Vaseghi, H.; Zamanzadeh, Z.; Sohrabi, M.; Hejazi, S.H. Evaluation of serotonin receptors (5HTR2A and 5HTR3A) mRNA expression changes in tumor of breast cancer patients. *Med. J. Islam. Repub. Iran* **2020**, *34*, 99. [CrossRef]
69. Ballou, Y.; Rivas, A.; Belmont, A.; Patel, L.; Amaya, C.; Lipson, S.; Khayou, T.; Dickerson, E.; Nahleh, Z.; Bryan, B. 5-HT serotonin receptors modulate mitogenic signaling and impact tumor cell viability. *Mol. Clin. Oncol.* **2018**, *9*, 243–254. [CrossRef] [PubMed]
70. Gautam, J.; Banskota, S.; Regmi, S.C.; Ahn, S.; Jeon, Y.H.; Jeong, H.; Kim, S.J.; Nam, T.; Jeong, B.-S.; Kim, J.-A. Tryptophan hydroxylase 1 and 5-HT7 receptor preferentially expressed in triple-negative breast cancer promote cancer progression through autocrine serotonin signaling. *Mol. Cancer* **2016**, *15*, 75. [CrossRef]
71. Balakrishna, P.; George, S.; Hatoum, H.; Mukherjee, S. Serotonin Pathway in Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 1268. [CrossRef]
72. Jose, J.; Tavares, C.D.J.; Ebelt, N.D.; Lodi, A.; Edupuganti, R.; Xie, X.; Devkota, A.K.; Kaoud, T.S.; Van Den Berg, C.L.; Anslyn, E.V.; et al. Serotonin Analogues as Inhibitors of Breast Cancer Cell Growth. *ACS Med. Chem. Lett.* **2017**, *8*, 1072–1076. [CrossRef]
73. Badawy, A.A.-B. Tryptophan metabolism and disposition in cancer biology and immunotherapy. *Biosci. Rep.* **2022**, *42*, BSR20221682. [CrossRef] [PubMed]
74. Sola-Penna, M.; Paixão, L.P.; Branco, J.R.; Ochioni, A.C.; Albanese, J.M.; Mundim, D.M.; Baptista-de-Souza, D.; Figueiredo, C.P.; Coelho, W.S.; Marcondes, M.C.; et al. Serotonin activates glycolysis and mitochondria biogenesis in human breast cancer cells through activation of the Jak1/STAT3/ERK1/2 and adenylate cyclase/PKA, respectively. *Br. J. Cancer* **2020**, *122*, 194–208. [CrossRef] [PubMed]
75. Agus, A.; Planchais, J.; Sokol, H. Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host Microbe* **2018**, *23*, 716–724. [CrossRef] [PubMed]
76. Baganz, N.L.; Blakely, R.D. A Dialogue between the Immune System and Brain, Spoken in the Language of Serotonin. *ACS Chem. Neurosci.* **2013**, *4*, 48–63. [CrossRef]
77. Herr, N.; Bode, C.; Duerschmied, D. The Effects of Serotonin in Immune Cells. *Front. Cardiovasc. Med.* **2017**, *4*, 48. [CrossRef] [PubMed]
78. Koretzky, G.A. Multiple Roles of CD4 and CD8 in T Cell Activation. *J. Immunol.* **2010**, *185*, 2643–2644. [CrossRef] [PubMed]
79. Baxevanis, C.N.; Fortis, S.P.; Perez, S.A. The balance between breast cancer and the immune system: Challenges for prognosis and clinical benefit from immunotherapies. *Semin. Cancer Biol.* **2021**, *72*, 76–89. [CrossRef] [PubMed]
80. Vazquez, M.I.; Catalan-Dibene, J.; Zlotnik, A. B cells responses and cytokine production are regulated by their immune microenvironment. *Cytokine* **2015**, *74*, 318–326. [CrossRef] [PubMed]
81. Lal, A.; Chan, L.; DeVries, S.; Chin, K.; Scott, G.K.; Benz, C.C.; Chen, Y.-Y.; Waldman, F.M.; Hwang, E.S. FOXP3-positive regulatory T lymphocytes and epithelial FOXP3 expression in synchronous normal, ductal carcinoma in situ, and invasive cancer of the breast. *Breast Cancer Res. Treat.* **2013**, *139*, 381–390. [CrossRef] [PubMed]
82. Peyraud, F.; Guegan, J.-P.; Bodet, D.; Cousin, S.; Bessede, A.; Italiano, A. Targeting Tryptophan Catabolism in Cancer Immunotherapy Era: Challenges and Perspectives. *Front. Immunol.* **2022**, *13*, 807271. [CrossRef] [PubMed]
83. Schmid, P.; Cortes, J.; Pusztai, L.; McArthur, H.; Kümmel, S.; Bergh, J.; Denkert, C.; Park, Y.H.; Hui, R.; Harbeck, N.; et al. Pembrolizumab for Early Triple-Negative Breast Cancer. *N. Engl. J. Med.* **2020**, *382*, 810–821. [CrossRef] [PubMed]
84. Cortes, J.; Cescon, D.W.; Rugo, H.S.; Nowecki, Z.; Im, S.-A.; Yusof, M.M.; Gallardo, C.; Lipatov, O.; Barrios, C.H.; Holgado, E.; et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): A randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet* **2020**, *396*, 1817–1828. [CrossRef] [PubMed]



Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.82667	0.78667	0.78667	0.82667	0.78667
R2	0.48032	0.73265	0.83582	0.91994	0.96709
Q2	0.26952	0.35072	0.38282	0.34112	0.2736

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.7	0.78	0.76	0.72	0.74
R2	0.41334	0.68952	0.84518	0.91787	0.95996
Q2	0.22221	0.38244	0.31885	0.30461	0.27103

Figure S1. Untargeted metabolomic statistical analyses of **SBR grades of the training set (n=51) (Figures S1 A, C** and **D) and the validation set (n=49) (Figures S1 B, D** and **F) using 602 metabolites.** Principal component analysis (PCA) score plots which shows no significant separation trends between high-grade (grade III – green dots) and low-grade (grade I-II –red dots) groups of the training and the validation set are shown in **Figure S1A** and **B**, respectively. Values of the accuracy, R2 and Q2 with 1 to 5 components of PLS-DA cross validations are shown for the training (**Figure S1C** and **E**) and of the validation (**Figure S1D** and **F**) set analyses.



Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.82667	0.80667	0.81333	0.81333	0.81333
R2	0.50358	0.56643	0.61758	0.62659	0.62805
Q2	0.43596	0.44431	0.40941	0.42401	0.38398

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.76	0.82	0.82	0.84	0.84
R2	0.44767	0.5528	0.58958	0.60041	0.60262
Q2	0.36872	0.40147	0.34386	0.31326	0.2883

Figure S2. Untargeted metabolomic statistical analyses of **SBR grades in** the **training set (n=51) (Figures S2 A, C** and **D) and the validation set (n=49) (Figures S2 B, D** and **F) using the top-12 metabolites.** Principal component analysis (PCA) score plots which show no significant separation trends between high-grade (grade III – green dots) and low-grade (grade I-II –red dots) groups of the training and the validation set are shown in **Figure S2A** and **B**, respectively. Values for accuracy, R2 and Q2 according 1 to 5 components of the PLS-DA cross validations are shown for the training (**Figure S2C** and **2E**) and the validation (**Figure S2D** and **F**) set analyses.

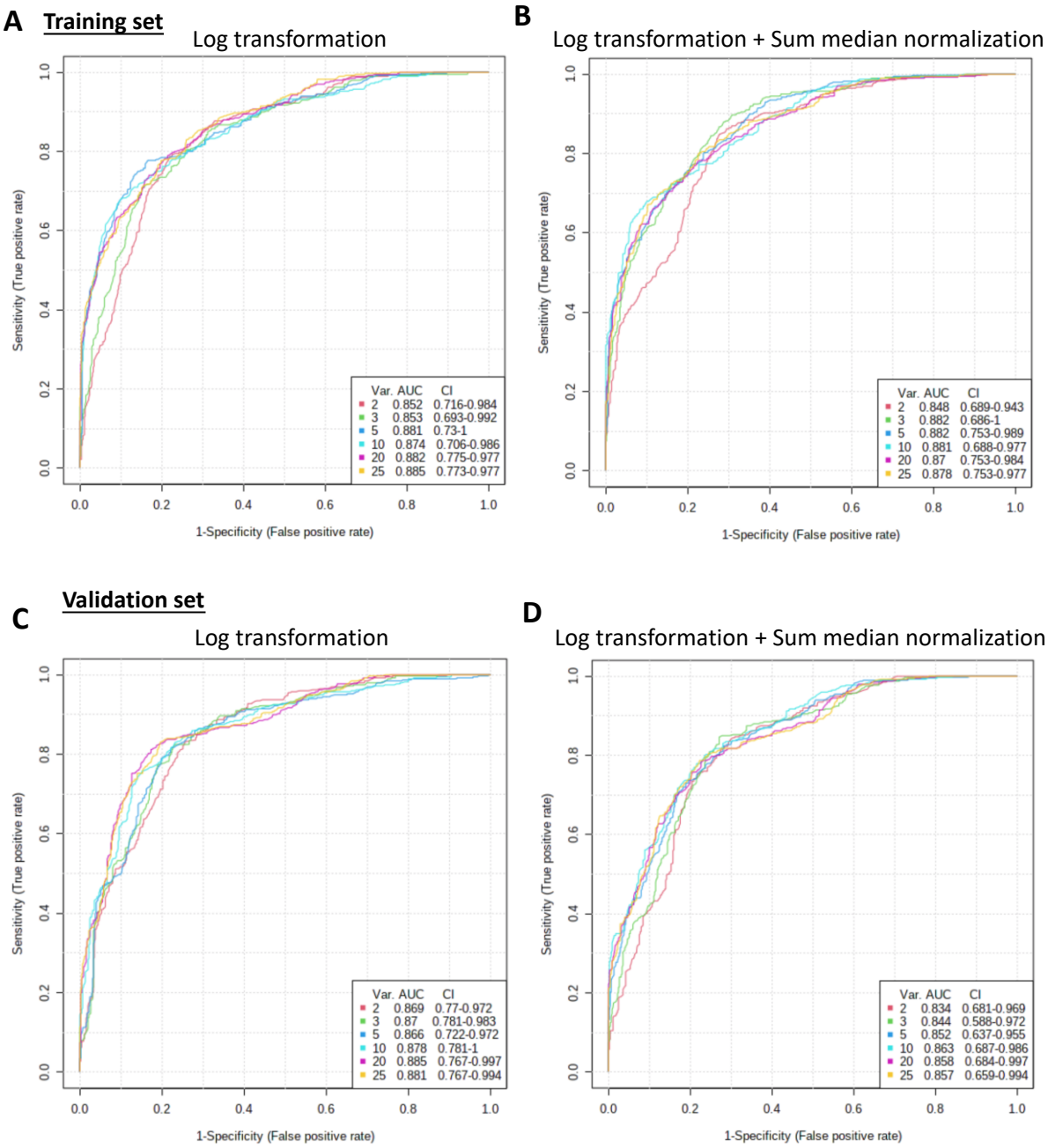
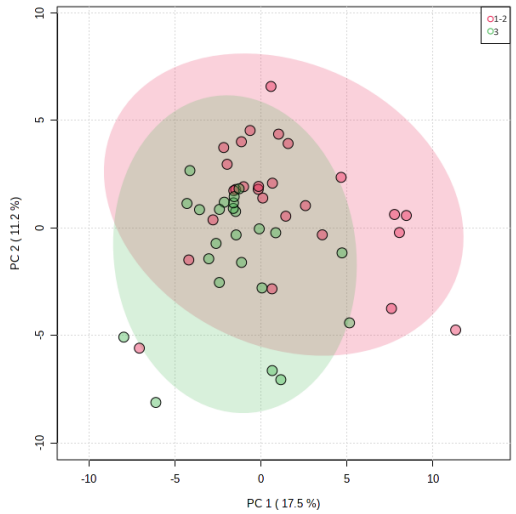
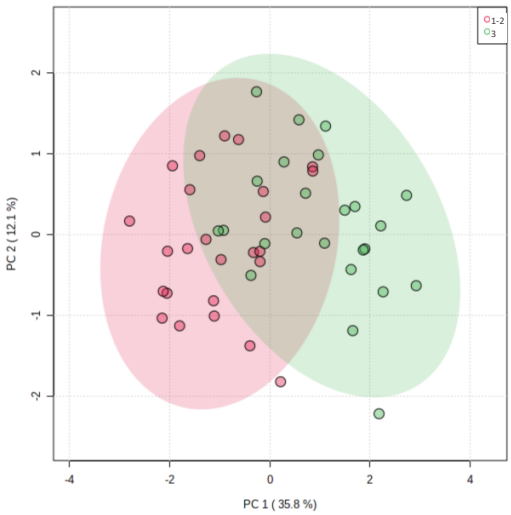


Figure S3. Metabolomic fingerprinting allowed accurate discrimination of SBR grades using the top 25 most important metabolites. The areas under the curve (AUCs) of the receiver operating characteristic (ROC) of different metabolomic signatures are illustrated for the training set (**Figures S3 A, B**) and the validation set (**Figure S3 C, D**), which included an increasing number of metabolites (var.), with their respective 95% confidence interval values (95%CI). Data were normalized using Log transformation. Sample normalization by sum was also performed and shown in **Figures S3 B** and **D**.

A PCA score plot Training set



B PCA score plot Validation set



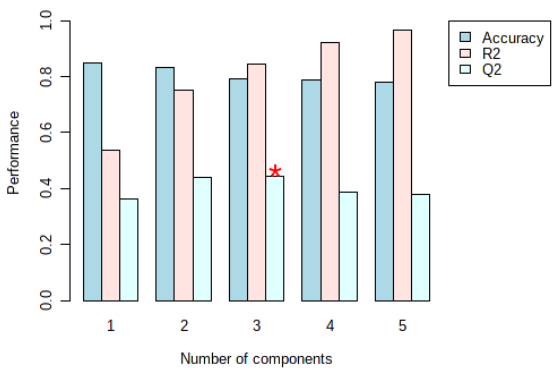
C **PLS-DA cross validation details:**

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.84667	0.83333	0.79333	0.78667	0.78
R2	0.53487	0.75237	0.84559	0.92016	0.96737
Q2	0.36443	0.43975	0.44491	0.38535	0.3783

D **PLS-DA cross validation details:**

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.8	0.82	0.74	0.72	0.72
R2	0.5258	0.58199	0.61884	0.63367	0.64166
Q2	0.44717	0.33581	0.21578	0.14141	0.03647

E



F

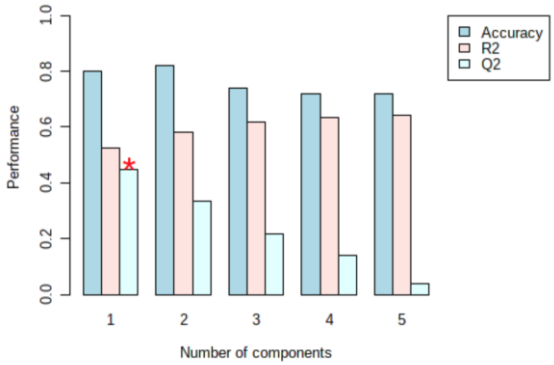


Figure S4. Untargeted metabolomic statistical analyses of **SBR grades of the training set (Figures S4 A, C** and **D) and validation set (Figures S4 B, D** and **F) using the top-25 metabolites.** Principal component analysis (PCA) score plots showing no significant separation trend between high-grade (grade III – green dots) and low-grade (grade I-II –red dots) groups of the training set and the validation set are shown in **Figure S4A** and **B**, respectively. Values for accuracy, R2 and Q2 with 1 to 5 components of the PLS-DA cross validations are shown for the training (**Figure S4C** and **E**) and validation (**Figure S4D** and **F**) set analyses.

Figure S5. Tryptophan Metabolism – Kegg database

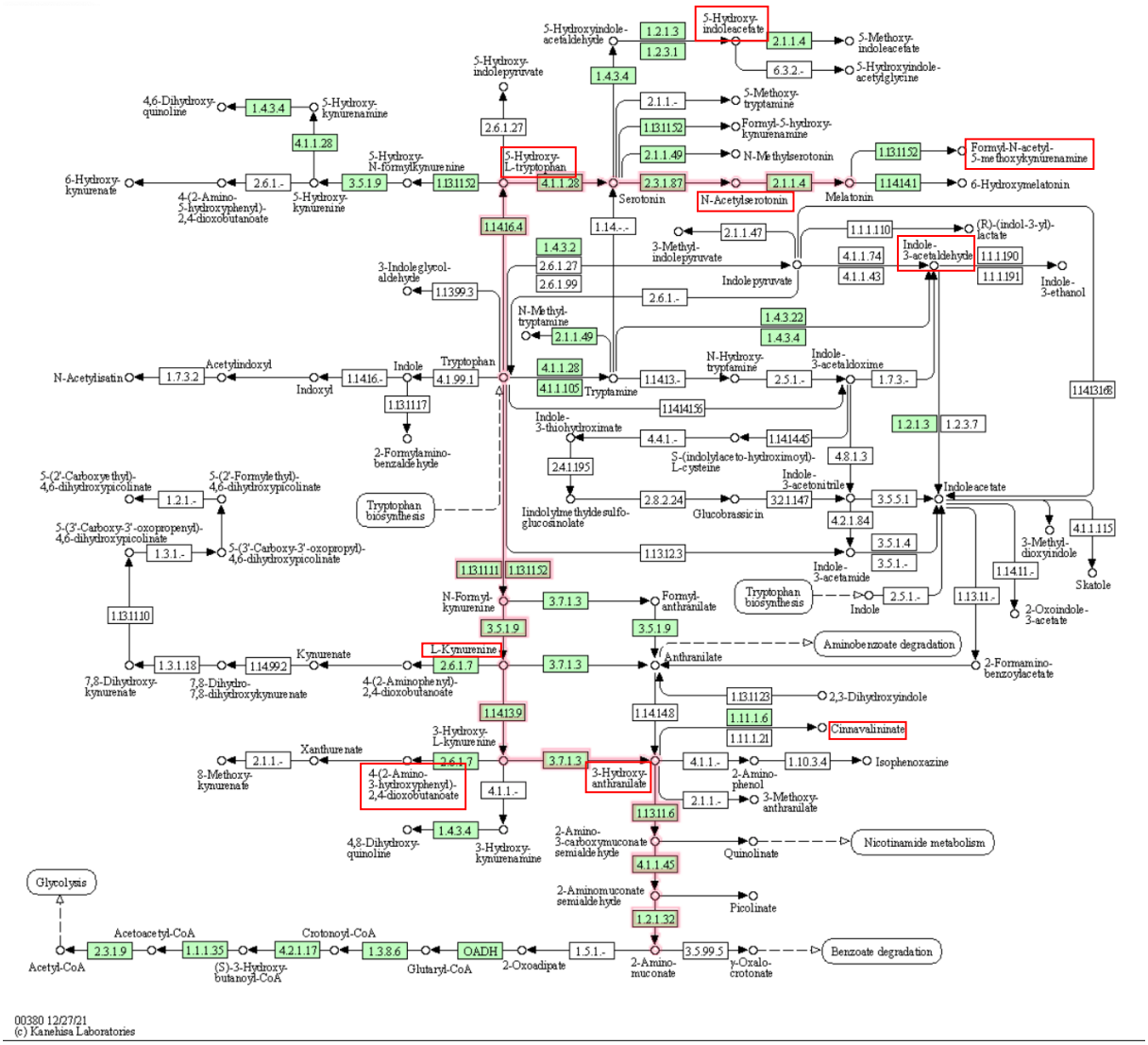


Figure S5. Schematic representation of metabolic pathway changes. The tryptophan pathway is illustrated according to the Kegg database (https://www.genome.jp/kegg-bin/show_pathway?hsa00380). Metabolite names are shown in colored boxes: Red boxes relate to higher concentration in high-grade samples.

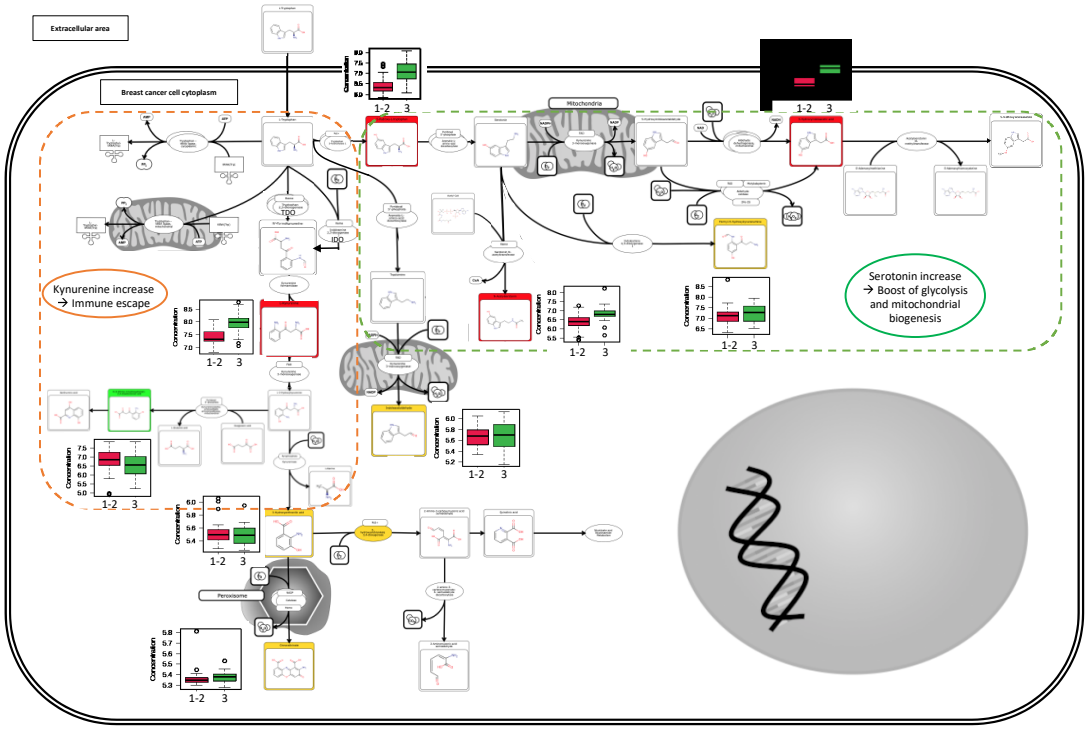
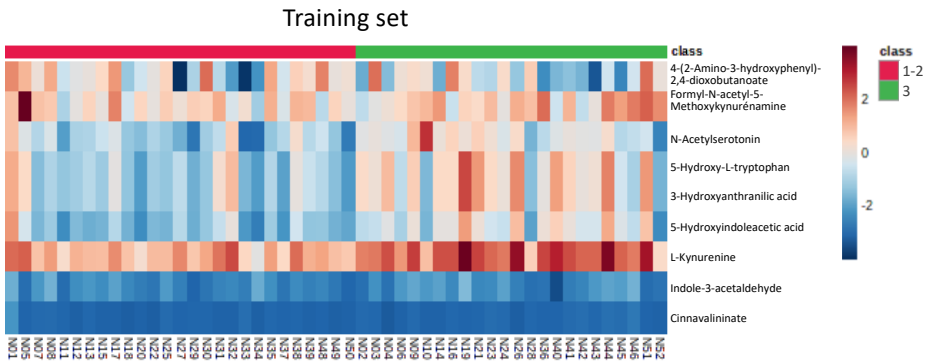


Figure S6. Schematic representation of metabolic pathway changes. The tryptophan pathway have been adapted from the Small Molecule Pathway database (https://www.smpdb.ca/view/SMP0000063). Box plots illustrate the relative concentration of the main tryptophan catabolites in high-grade (grade III – green boxes) and low-grade (grade I-II –red boxes) groups. Metabolite names are shown in colored boxes: Red boxes relate to higher concentration in high-grade samples; yellow boxes to equivalent concentrations in high-grade and low-grade samples; green boxes to lower concentrations in high-grade samples.

A



B

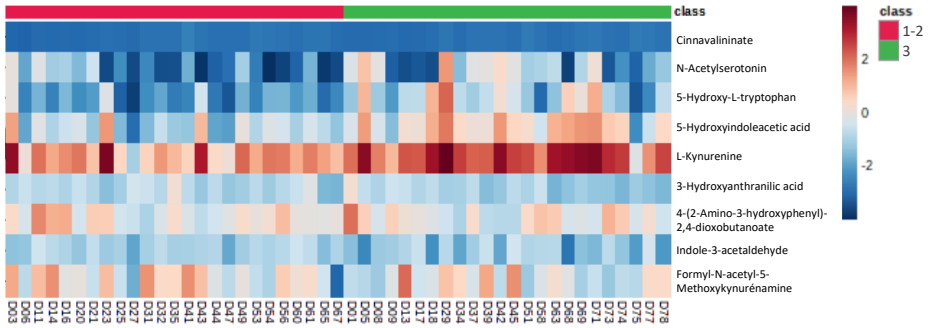


Figure S7. Heatmap representation of relative concentrations of tryptophan catabolites of the training set (Figures S7 A) and the validation set (Figures S7). Results of the low-grade (grade I-II –red labels of the top line) group are positioned in the left part of the heatmap and those of the high-grade (grade III – green labels) group in the right part.

Table S1. Exploration of Pathway Analysis

Training set

	Total Cmpd	Hits	Raw p	-log10(p)	Holm adjust	FDR	Impact
Purine metabolism	65	12	8.77E-06	5.06E+00	5.00E-04	3.11E-04	0.14
Tryptophan metabolism	41	9	1.09E-05	4.96E+00	6.11E-04	3.11E-04	0.34
Nitrogen metabolism	6	2	3.72E-05	4.43E+00	2.05E-03	6.42E-04	0.00
Aminoacyl-tRNA biosynthesis	48	11	4.51E-05	4.35E+00	2.43E-03	6.42E-04	0.17
Glyoxylate and dicarboxylate metabolism	32	9	9.12E-05	4.04E+00	4.83E-03	9.49E-04	0.58
Arginine biosynthesis	14	9	9.99E-05	4.00E+00	5.19E-03	9.49E-04	0.60
Neomycin, kanamycin and gentamicin biosynthesis	2	1	1.50E-04	3.82E+00	7.63E-03	1.22E-03	0.00
Valine, leucine and isoleucine biosynthesis	8	1	2.32E-04	3.63E+00	1.16E-02	1.31E-03	0.00
D-Glutamine and D-glutamate metabolism	6	3	2.35E-04	3.63E+00	1.16E-02	1.31E-03	0.50
Terpenoid backbone biosynthesis	18	2	2.37E-04	3.63E+00	1.16E-02	1.31E-03	0.11
Arginine and proline metabolism	38	13	2.54E-04	3.60E+00	1.19E-02	1.31E-03	0.53
Histidine metabolism	16	5	9.43E-04	3.03E+00	4.34E-02	4.48E-03	0.41
Cysteine and methionine metabolism	33	7	1.38E-03	2.86E+00	6.20E-02	5.89E-03	0.21
Pentose phosphate pathway	22	6	1.45E-03	2.84E+00	6.37E-02	5.89E-03	0.37
Glycine, serine and threonine metabolism	33	12	1.58E-03	2.80E+00	6.78E-02	5.99E-03	0.71
Nicotinate and nicotinamide metabolism	15	5	1.90E-03	2.72E+00	7.97E-02	6.76E-03	0.23
Glutathione metabolism	28	9	2.52E-03	2.60E+00	1.03E-01	8.45E-03	0.15
beta-Alanine metabolism	21	7	2.77E-03	2.56E+00	1.11E-01	8.76E-03	0.56
Alanine, aspartate and glutamate metabolism	28	9	5.19E-03	2.28E+00	2.02E-01	1.50E-02	0.61
Pyrimidine metabolism	39	10	5.28E-03	2.28E+00	2.02E-01	1.50E-02	0.36
Tyrosine metabolism	42	5	7.33E-03	2.13E+00	2.71E-01	1.94E-02	0.14
Amino sugar and nucleotide sugar metabolism	37	4	7.49E-03	2.13E+00	2.71E-01	1.94E-02	0.15
Glycerolipid metabolism	16	2	8.38E-03	2.08E+00	2.93E-01	2.08E-02	0.33
Butanoate metabolism	15	2	9.63E-03	2.02E+00	3.27E-01	2.29E-02	0.00
Inositol phosphate metabolism	30	3	1.11E-02	1.96E+00	3.66E-01	2.51E-02	0.00
Sphingolipid metabolism	21	3	1.15E-02	1.94E+00	3.66E-01	2.51E-02	0.02
Riboflavin metabolism	4	3	2.01E-02	1.70E+00	6.24E-01	4.25E-02	1.00
Glycerophospholipid metabolism	36	8	2.24E-02	1.65E+00	6.71E-01	4.55E-02	0.38
Fructose and mannose metabolism	20	2	2.44E-02	1.61E+00	7.08E-01	4.80E-02	0.07
Pantothenate and CoA biosynthesis	19	7	3.08E-02	1.51E+00	8.62E-01	5.85E-02	0.13
Selenocompound metabolism	20	1	4.56E-02	1.34E+00	1.00E+00	8.39E-02	0.00

Validation set

	Total Cmpd	Hits	Raw p	-log10(p)	Holm adjust	FDR	Impact
Tryptophan metabolism	41	9	3.13E-05	4.50E+00	1.79E-03	1.79E-03	0.34
Ascorbate and aldarate metabolism	8	2	2.26E-04	3.65E+00	1.27E-02	4.13E-03	0.50
Starch and sucrose metabolism	18	2	2.55E-04	3.59E+00	1.40E-02	4.13E-03	0.15
Pentose and glucuronate interconversions	18	3	2.90E-04	3.54E+00	1.56E-02	4.13E-03	0.20
Pyrimidine metabolism	39	10	5.41E-04	3.27E+00	2.86E-02	6.16E-03	0.36
Galactose metabolism	27	3	7.31E-04	3.14E+00	3.80E-02	6.94E-03	0.00
Lysine degradation	25	3	9.24E-04	3.03E+00	4.71E-02	7.52E-03	0.14
Vitamin B6 metabolism	9	1	1.29E-03	2.89E+00	6.44E-02	9.18E-03	0.05
Fatty acid degradation	39	1	3.45E-03	2.46E+00	1.69E-01	2.18E-02	0.00
Glycerophospholipid metabolism	36	8	6.24E-03	2.20E+00	3.00E-01	3.56E-02	0.38
Histidine metabolism	16	5	8.31E-03	2.08E+00	3.90E-01	4.30E-02	0.41
Pyruvate metabolism	22	5	1.44E-02	1.84E+00	6.65E-01	6.86E-02	0.24
Glycerolipid metabolism	16	2	2.08E-02	1.68E+00	9.38E-01	9.14E-02	0.33
Ether lipid metabolism	20	2	2.77E-02	1.56E+00	1.00E+00	9.19E-02	0.00
Citrate cycle (TCA cycle)	20	5	2.88E-02	1.54E+00	1.00E+00	9.19E-02	0.18
Glyoxylate and dicarboxylate metabolism	32	9	3.01E-02	1.52E+00	1.00E+00	9.19E-02	0.58
Glycine, serine and threonine metabolism	33	12	3.07E-02	1.51E+00	1.00E+00	9.19E-02	0.71
Glutathione metabolism	28	9	3.17E-02	1.50E+00	1.00E+00	9.19E-02	0.15
Pantothenate and CoA biosynthesis	19	7	3.20E-02	1.49E+00	1.00E+00	9.19E-02	0.13
Nicotinate and nicotinamide metabolism	15	5	3.48E-02	1.46E+00	1.00E+00	9.19E-02	0.23
Phenylalanine, tyrosine and tryptophan biosynthesis	4	1	3.63E-02	1.44E+00	1.00E+00	9.19E-02	0.50
Phenylalanine metabolism	10	1	3.63E-02	1.44E+00	1.00E+00	9.19E-02	0.36
Fructose and mannose metabolism	20	2	3.71E-02	1.43E+00	1.00E+00	9.19E-02	0.07

Training set + validation set

			Training Set	Validation Set
Pathway	Total Cmpd	Hits	Raw p	Raw p
Tryptophan metabolism	41	9	1.09E-05	3.13E-05
Glyoxylate and dicarboxylate metabolism	32	9	9.12E-05	3.01E-02
Histidine metabolism	16	5	9.43E-04	8.31E-03
Glycine, serine and threonine metabolism	33	12	1.58E-03	3.07E-02
Nicotinate and nicotinamide metabolism	15	5	1.90E-03	3.48E-02
Glutathione metabolism	28	9	2.52E-03	3.17E-02
Pyrimidine metabolism	39	10	5.28E-03	5.41E-04
Glycerolipid metabolism	16	2	8.38E-03	2.08E-02

Table S1. Significant Pathways of the SBR grade Analysis. Only pathways with p-values < 0.05 are shown.

c) DISCUSSION

Le métabolite le plus pertinent était la N1,N12-diacétylspermine. La diacétylspermine intervient dans la transformation, la prolifération cellulaire et l'invasion métastatique par le biais de l'ornithine décarboxylase (ODC) et c-MYC(289,290). Une étude fonctionnelle a examiné les effets de la spermine sur le récepteur des œstrogènes. Les résultats obtenus suggèrent que la spermine joue un rôle important dans la régulation de la liaison du ligand au RE, dans l'activation des gènes, et dans la résistance aux hormones(213). Fahrmann *et al.* ont démontré que des niveaux élevés de DiAcSpm dans le plasma étaient associés à un faible infiltrat immunitaire, à des signatures génétiques liées à l'immunité réduites, à une récidive précoce (< 1 an), à une survie sans maladie métastatique à 5 ans plus faible et à une survie globale à 5 ans plus faible(214).

Les analyses d'enrichissement et des voies métaboliques ont montré que la voie du tryptophane (kynurénine et sérotonine) était plus fortement activée dans les tumeurs de haut grade que dans les tumeurs de grade bas ou intermédiaire. La signature métabolomique des 12 principaux métabolites a révélé des niveaux accrus pour 4 catabolites du tryptophane (N'-formylkynurénine, 5-hydroxy-L-tryptophane, 8-méthoxykynurénate et L-kynurénine) et 2 métabolites de la voie de la sérotonine (N-acétylsérotonine et acide 5-hydroxyindoleacétique). Le tryptophane est converti en kynurénine par l'indoleamine 2,3-dioxygénase 1 (IDO1), sa variante d'épissage IDO2 et la tryptophane 2,3-dioxygénase (TDO)(291). L'IDO1 est un facteur clé dans le maintien de la tolérance immunitaire(292). IDO1 est exprimée dans plusieurs types de tumeurs et est associée à une activation réduite des cellules cytotoxiques, à une infiltration accrue des lymphocytes T régulateurs, à des taux de survie plus faibles(293,294) et à une résistance accrue aux médicaments(295,296). L'élévation des concentrations de kynurénine pourrait être impliquée dans l'inhibition de la réponse immunitaire(297,298). Par ailleurs, l'altération des profils d'expression de la sérotonine et de ses récepteurs a été associée aux événements initiaux du développement du cancer du sein et à la progression

tumorale(216,217). Plusieurs études ont suggéré que l'expression de la sérotonine et de ses récepteurs dans les cellules immunitaires pouvait moduler la réponse immunitaire, en particulier en cas d'inflammation(299). D'autres études ont indiqué que les effets immunitaires de la sérotonine comprenaient la suppression de la libération d'IL-1β et de TNF-α dans les cellules du sang périphérique, et l'activation des lymphocytes T(300). La voie du tryptophane, par le biais de la kynurénine et de la sérotonine pourrait donc influencer l'immunité antitumorale.

Le ciblage de la voie du tryptophane, par le biais d'un ciblage de la kynurénine et/ou de la sérotonine, pourrait restaurer l'immunité antitumorale et avoir un effet synergique avec l'immunothérapie. Des études récentes n'ont pas pu démontrer l'efficacité de cette stratégie(301) dans une population non sélectionnée par la métabolomique. De manière intéressante, notre étude a montré que l'activation de la voie du tryptophane n'était pas homogène chez tous les patients de haut grade. En effet, les niveaux de L-Kynurénine n'étaient pas élevés dans tous les échantillons analysés de patients de haut grade, ce qui pourrait avoir un impact sur l'efficacité des combinaisons ciblant l'immunité antitumorale. L'utilisation de la métabolomique et du profil de concentration des métabolites de la voie du tryptophane pourrait permettre une meilleure sélection des patients éligibles à l'immunothérapie. Avec l'avènement de l'immunothérapie dans le cancer du sein triple négatif néoadjuvant(82) et métastatique de première ligne(302), la recherche de biomarqueurs prédictifs devient primordiale, ce d'autant plus que pour le moment aucun marqueur n'a été identifié en contexte néoadjuvant. La première étape va être de montrer que le profil de concentration des métabolites de la voie du tryptophane est prédictif d'une réponse à l'immunothérapie en analysant la réponse clinico-biologique après chimio-immunothérapie néo-adjuvante dans le cancer du sein triple négatif stade II-III (cf. Perspective Projet EMMENEO-TN).

IX. Conclusion et Perspectives

1. Conclusion

Les résultats rapportés dans les 3 articles de thèse montrent une applicabilité de la métabolomique dans le cancer du sein. Lors de ce travail, la validité analytique (reproductibilité, robustesse) et la validité clinique (capacité d'un facteur à séparer une population d'intérêt en deux sous-groupes) de la métabolomique ont été prouvées. Il reste cependant à confirmer ces résultats sur des cohortes de plus grande ampleur (cf. chapitre Perspectives ci-dessous).

Les prochaines études devront s'appliquer à démontrer une utilité clinique. L'utilité clinique implique qu'un facteur soit utile dans la prise en charge des patients. Pour cela, il faudra apporter la preuve d'un bénéfice sur les données de survie lorsqu'il est comparé à une pratique clinique sans lui dans un contexte spécifique où le facteur est pertinent. Un facteur pronostique utile dans le cancer du sein devra présenter une valeur pronostique ou prédictive significative et indépendante, validée lors d'un essai thérapeutique.

De plus, sa détermination devra être faisable, reproductible et largement disponible, avec un contrôle de qualité. Sa mesure ne devra pas consommer de tissus nécessaires à d'autres tests, en particulier à l'évaluation histopathologique réalisée en routine clinique et les résultats devront être facilement interprétables par le clinicien.

2. Perspectives

Cinq projets ont émergé de ce travail de thèse et vont être menés en post-thèse pour renforcer la validité analytique et clinique de la métabolomique. De projets sont en continuité directe : EMMEA-Survie, EMMEA-Validation. Les trois autres sont des projets émergents nécessitant l'élaboration de nouvelles cohortes d'analyse : METABOPREDICT, TISSUBLOC, et EMMENEO-TN.

a) PROJET EMMEA-S

Le projet EMMEA-S a pour objectif principal la réalisation d'une analyse supervisée en fonction de la survie sans progression à 2 ans. Vu qu'un signal est ressorti des analyses de survie réalisées après clustering par les méthodes non supervisées (Article 1&2), nous espérons mettre en évidence des métabolites d'intérêt significatif en lien avec une récidive précoce à 2 ans. La signature établie pourra alors être validée sur la cohorte METABOPREDICT.

b) PROJET EMMEA-VALIDATION

Le projet EMMEA-Validation a pour objectif principal de confirmer les résultats méthodologiques établis par l'étude EMMEA (Article 1) en comparant les performances des 5 méthodes de machine learning sur la cohorte dijonnaise. Si la méthode K-sparse se révèle de nouveau être la meilleure méthode d'analyse non-supervisée malgré une population cliniquement différente (contexte localement avancé versus contexte localisé), alors cette méthode pourra être recommandée pour la suite des analyses non-supervisées de métabolomique dans le cancer du sein non-métastatique. Par contre, l'analyse de survie ne pourra pas être réalisée car les 2 populations d'étude sont trop différentes en termes de pronostic et de traitements administrés, ce qui rendrait difficile l'interprétation des données de survie au long cours.

c) PROJET METABOPREDICT

Le Projet METABOPREDICT a pour but de réaliser une analyse multiOMIQUE associant des données de transcriptomique, métabolomique et données de survie simulées par PREDICT uk. Dans un premier temps, nous allons comparer la valeur prédictive et pronostique de la métabolomique et des tests de prédiction comme ENDOPREDICT et PREDICT uk. Dans un second temps, nous allons évaluer la valeur prédictive et pronostique d'un test multi-OMIQUE. Une cinquantaine de tumeurs congelées ont déjà été collectées et traitées en métabolomique. Les données brutes de métabolomique sont disponibles. Le recueil de données cliniques a été réalisé ainsi que la centralisation des données ENDOPREDICT. Deux appels à projet VALO-DATA ont été soumis, acceptés en lettre d'intention mais non retenus par la suite, ce qui a ralenti l'avancée du projet. Cependant, le projet METABOPREDICT a obtenu un financement dons et legs Octobre rose en 2022, ce qui va permettre de réaliser l'analyse statistique multiOMIQUE en association avec le département de statistiques (DEBDS) du Centre Antoine Lacassagne. De plus, une nouvelle soumission a été réalisée cette année pour obtenir un financement VALO-DATA complémentaire.

Si METABOPREDICT met en évidence un intérêt à l'adjonction de la métabolomique à une signature génomique, ici ENDOPREDICT, un projet similaire pourra être réalisé avec la signature ONCOTYPE DX désormais majoritairement recommandé en pratique clinique.

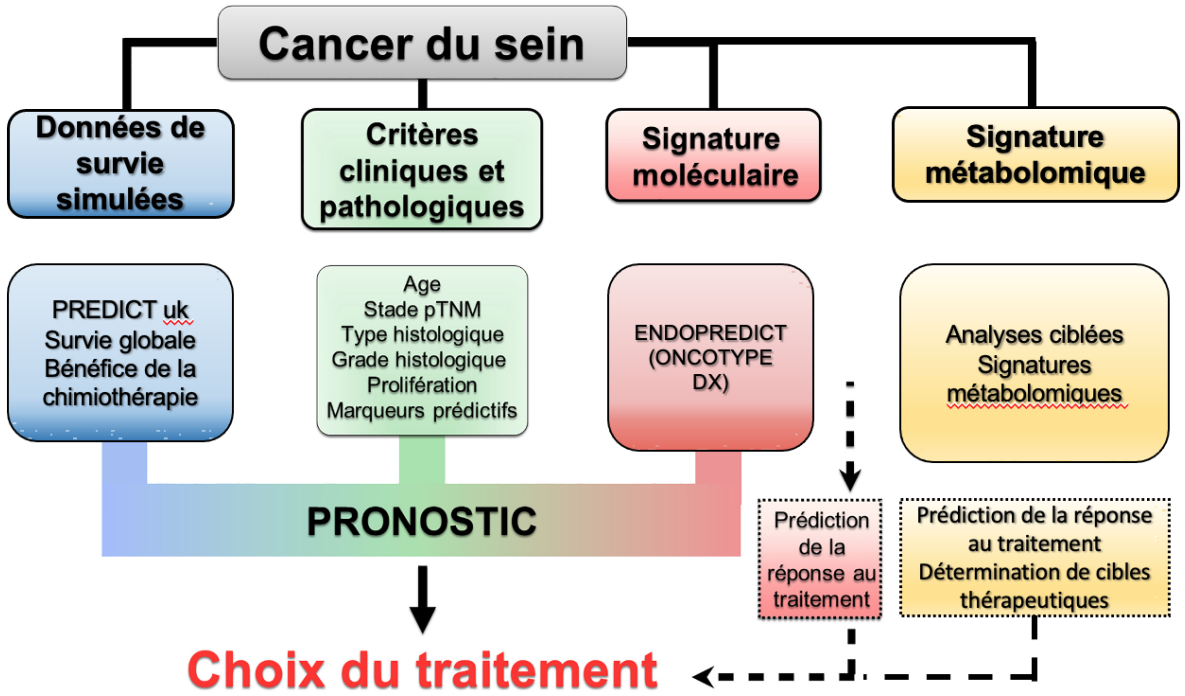


Figure 31. Modèle d'intégration des paramètres clinicopathologiques, moléculaires et métabolomiques dans la prise en charge du cancers du sein non-métastatique *(d'après Franchet et al.(303) et Joyon et al.(304)). Le profil des cancers du sein est évalué en premier lieu sur les paramètres clinicopathologiques « traditionnels ». Les outils moléculaires et métabolomiques peuvent venir nuancer l'évaluation pronostique des patientes et donner une dimension prédictive pour guider la stratégie thérapeutique.*

d) PROJET TISSUBLOC

Le Projet TISSUBLOC est un projet méthodologique qui va comparer des signatures obtenues sur tissus congelés et en paraffine. Pour cela, en parallèle des tumeurs congelées recueillies pour le projet METABOPREDICT, des copeaux de paraffine vont être extraits pour une analyse parallèle de métabolomique sur paraffine. Si les résultats obtenus sont comparables, les analyses métabolomiques pourront être étendues à l'ensemble des tissus fixés en paraffine et non plus limitées aux prélèvements congelés. Ce projet a pour bénéfice attendu la généralisation possible de la métabolomique et la simplification de la gestion des prélèvements.

e) PROJET EMMENEO-TN

Le projet EMMENEO-TN a pour objectif de montrer que le profil de concentration des métabolites de la voie du tryptophane est prédictif d'une réponse à l'immunothérapie en analysant la réponse clinico-biologique après chimio-immunothérapie néo-adjuvante dans le cancer du sein triple négatif stade II-III. Pour cela, des biopsies diagnostiques vont être prospectivement collectées (une cinquantaine). Les tumeurs seront labélisées en fonction de la réponse clinico-biologique (réponse complète RCB 0, réponse partielle RCB I-II, mauvaise réponse RCB III) et une analyse ciblée sera réalisée sur les métabolites de la voie du tryptophane. Si un lien est établi entre les métabolites de la voie du tryptophane et la réponse clinico-biologique, une étude de phase I associant un inhibiteur du tryptophane et l'immunochimiothérapie pourra être proposée dans la population tryptophane-activée.

X. Annexes

Annexe 1. Classification TNM 8e édition (305)

Tumeur primitive (T)

TX Tumeur primaire non connue ou tumeur prouvée par la présence de cellules malignes dans les sécrétions broncho-pulmonaires mais non visible aux examens radiologiques et endoscopiques.

T0 Absence de tumeur identifiable.

Tis Carcinome in situ.

T1 Tumeur de 3 cm ou moins dans ses plus grandes dimensions, entourée par du poumon ou de la plèvre viscérale, sans évidence d'invasion plus proximale que les bronches lobaires à la bronchoscopie (c'est-à-dire pas dans les bronches souches).

T1a(mi) : Adénocarcinome minimalement-invasif

T1a : Tumeur de 1 cm ou moins dans sa plus grande dimension.

T1b : Tumeur de plus de 1 cm sans dépasser 2 cm dans sa plus grande dimension.

T1c : Tumeur de plus de 2 cm sans dépasser 3 cm dans sa plus grande dimension.

T2 Tumeur de plus de 3 cm, mais de 5 cm ou moins, avec quelconque des éléments suivants:

-envahissement d'une bronche souche quelle que soit sa distance par rapport à la carène mais sans envahissement de la carène,

-envahissement de la plèvre viscérale,

-existence d'une atélectasie ou pneumonie obstructive s'étendant à la région hilaire ((sub)lobaire ou pulmonaire).

T2a : Tumeur de plus de 3 cm sans dépasser 4 cm dans sa plus grande dimension.

T2b : Tumeur de plus de 4 cm sans dépasser 5 cm dans sa plus grande dimension.

T3 Tumeur de plus de 5 cm et de 7 cm ou moins, ou associée à un(des) nodule(s) tumoral(aux) distinct(s) et dans le même lobe, ou ayant au moins l'un des

caractères invasifs suivants:

-atteinte de la paroi thoracique (incluant les tumeurs du sommet),

-atteinte du nerf phrénique,

-atteinte de la plèvre pariétale ou du péricarde.

T4 Tumeur de plus de 7 cm ou associée à un(des) nodule(s) pulmonaire(s) distinct(s) comportant un envahissement quelconque parmi les suivants : médiastin, cœur ou gros vaisseaux, trachée, diaphragme, nerf récurrent, œsophage, corps vertébraux, carène, nodules tumoraux séparés dans deux lobes différents du même poumon.

Ganglions lymphatiques régionaux (N)

NX Envahissement locorégional inconnu.

N0 Absence de métastase dans les ganglions lymphatiques régionaux.

N1 Métastases ganglionnaires péri-bronchiques homolatérales et/ou hilaires homolatérales incluant une extension directe.

N2 Métastases dans les ganglions médiastinaux homolatéraux ou dans les ganglions sous-carénaires.

N3 Métastases ganglionnaires médiastinales controlatérales ou hilaires controlatérales ou scaléniques, sus-claviculaires homo- ou controlatérales.

Métastases à distance (M)

M0 Pas de métastase à distance.

M1 Existence de métastases:

M1a : Nodules tumoraux séparés dans un lobe controlatéral, ou nodules pleuraux ou pleurésie maligne ou péricardite maligne.

M1b : Une seule métastase dans un seul site métastatique.

M1c : Plusieurs métastases dans un seul site ou plusieurs sites atteints.

Annexe 2. Stadification cancer du sien (305)

0	Tis	N0	M0
IA	T1[1]	N0	M0
IB	T0, T1[1]	N1mi	M0
IIA	T0, T1[1]	N1	M0
	T2	N0	M0
IIB	T2	N1	M0
	T3	N0	M0
IIIA	T0, T1[1], T2	N2	M0
	T3	N1, N2	M0
IIIB	T4	N0, N1, N2	M0
IIIC	tous T	N3	M0
IV	tous T	tous N	M1

[1] : IA : T1a ; IB : T1b-c ; IIIA : tout T1.

Annexe 3. Grade du cancer du sien (306)

1. Différenciation tubulo-glandulaire : proportion de tubes ou glandes dans la tumeur (en % de surface tumorale)	**Score**
>75 % : tumeur bien différenciée	1
10 à 75 % : tumeur moyennement différenciée	2
<10 % : tumeur peu différenciée	3
2. Pléomorphisme nucléaire : degré d'atypie apprécié sur la population tumorale prédominante	
Noyaux petits, réguliers, uniformes	1
Pléomorphisme modéré	2
Variations marquées de taille, de forme, avec nucléoles proéminents	3
Nombre de mitoses (à compter sur 10 champs au grossissement x400 ; valeurs définies pour un champ de 0,48 mm de diamètre ; calibrage du microscope nécessaire pour des champs différents)	
0 à 6 mitoses	1
7 à 12 mitoses	2
>12 mitoses	3
AU TOTAL	
Grade I	3 ou 4 ou 5
Grade II	6 ou 7
Grade III	8 ou 9

XI. Bibliographie

1. Oliver S. Systematic functional analysis of the yeast genome. Trends Biotechnol. 1 sept 1998;16(9):373-8.
2. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. janv 2002;48(1-2):155-71.
3. Gika HG, Theodoridis GA, Wilson ID. Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry fingerprinting of human urine. J Chromatogr A. mai 2008;1189(1-2):314-22.
4. Townsend MK, Clish CB, Kraft P, Wu C, Souza AL, Deik AA, et al. Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. Clin Chem. nov 2013;59(11):1657-67.
5. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical Aspects and Sample Quality Assessment in Metabolomics Studies of Human Blood. Clin Chem. 1 mai 2013;59(5):833-45.
6. Courant F, Antignac JP, Dervilly-Pinel G, Le Bizec B. Basics of mass spectrometry based metabolomics. PROTEOMICS. nov 2014;14(21-22):2369-88.
7. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin. juill 2021;71(4):333-58.
8. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. Sci Adv. 6 mai 2016;2(5):e1600200.
9. Hsu PP, Sabatini DM. Cancer Cell Metabolism: Warburg and Beyond. Cell. sept 2008;134(5):703-7.
10. Dang CV, Semenza GL. Oncogenic alterations of metabolism. Trends Biochem Sci. févr 1999;24(2):68-72.
11. Shaw RJ. LKB1 and AMP-activated protein kinase control of mTOR signalling and growth. Acta Physiol. mai 2009;196(1):65-80.
12. Worby CA, Dixon JE. PTEN. Annu Rev Biochem. 2 juin 2014;83(1):641-69.
13. Dang CV. MYC, Metabolism, Cell Growth, and Tumorigenesis. Cold Spring Harb Perspect Med. 1 août 2013;3(8):a014217-a014217.
14. Kimmelman AC. Metabolic Dependencies in *RAS* -Driven Cancers. Clin Cancer Res. 15 avr 2015;21(8):1828-34.
15. Vousden KH, Ryan KM. p53 and metabolism. Nat Rev Cancer. oct 2009;9(10):691-700.
16. Jordan KW, He W, Halpern EF, Wu CL, Cheng LL. Evaluation of Tissue Metabolites with High Resolution Magic Angle Spinning MR Spectroscopy Human Prostate Samples After Three-Year Storage at -80 degrees C. Biomark Insights. 18 avr 2007;2:147-54.
17. Jordan KW, Cheng LL. NMR-based metabolomics approach to target biomarkers for human prostate cancer. Expert Rev Proteomics. juin 2007;4(3):389-400.
18. Cheng LL, Anthony DC, Comite AR, Black PM, Tzika AA, Gonzalez RG.

Quantification of microheterogeneity in glioblastoma multiforme with exvivo high-resolution magic-angle spinning (HRMAS) proton magnetic resonancespectroscopy. Neuro-Oncol. 1 avr 2000;2(2):87-95.
19. Cheng LL, Burns MA, Taylor JL, He W, Halpern EF, McDougal WS, et al. Metabolic Characterization of Human Prostate Cancer with Tissue Magnetic Resonance Spectroscopy. Cancer Res. 15 avr 2005;65(8):3030-4.
20. Cheng LL, Chang IW, Louis DN, Gonzalez RG. Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens. Cancer Res. 1 mai 1998;58(9):1825-32.
21. Cheng LL, Chang IW, Smith BL, Gonzalez RG. Evaluating Human Breast Ductal Carcinomas with High-Resolution Magic-Angle Spinning Proton Magnetic Resonance Spectroscopy. J Magn Reson. nov 1998;135(1):194-202.
22. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. Nat Rev Cancer. 1 juill 2004;4(7):551-61.
23. Griffiths JR, Tate AR, Howe FA, Stubbs M. Magnetic Resonance Spectroscopy of cancer—practicalities of multi-centre trials and early results in non-Hodgkin's lymphoma. Eur J Cancer. nov 2002;38(16):2085-93.
24. Millis K, Weybright P, Campbell N, Fletcher JA, Fletcher CD, Cory DG, et al. Classification of human liposarcoma and lipoma using ex vivo proton NMR spectroscopy. Magn Reson Med. févr 1999;41(2):257-67.
25. Morvan D, Demidem A, Papon J, De Latour M, Madelmont JC. Melanoma tumors acquire a new phospholipid metabolism phenotype under cystemustine as revealed by high-resolution magic angle spinning proton nuclear magnetic resonance spectroscopy of intact tumor samples. Cancer Res. 15 mars 2002;62(6):1890-7.
26. Morvan D, Demidem A, Papon J, Madelmont JC. Quantitative HRMAS proton total correlation spectroscopy applied to cultured melanoma cells treated by chloroethyl nitrosourea: Demonstration of phospholipid metabolism alterations. Magn Reson Med. févr 2003;49(2):241-8.
27. Tate AR, Foxall PJ, Holmes E, Moka D, Spraul M, Nicholson JK, et al. Distinction between normal and renal cell carcinoma kidney cortical biopsy samples using pattern recognition of (1)H magic angle spinning (MAS) NMR spectra. NMR Biomed. avr 2000;13(2):64-71.
28. Kapoore RV, Coyle R, Staton CA, Brown NJ, Vaidyanathan S. Influence of washing and quenching in profiling the metabolome of adherent mammalian cells: a case study with the metastatic breast cancer cell line MDA-MB-231. The Analyst. 2017;142(11):2038-49.
29. Warburg O. On the Origin of Cancer Cells. Science. 24 févr 1956;123(3191):309-14.
30. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science. 22 mai 2009;324(5930):1029-33.
31. Koppenol WH, Bounds PL, Dang CV. Otto Warburg's contributions to current concepts of cancer metabolism. Nat Rev Cancer. mai 2011;11(5):325-37.
32. Vander Heiden MG, DeBerardinis RJ. Understanding the Intersections between

Metabolism and Cancer Biology. Cell. févr 2017;168(4):657-69.
33. Luengo A, Li Z, Gui DY, Sullivan LB, Zagorulya M, Do BT, et al. Increased demand for NAD+ relative to ATP drives aerobic glycolysis. Mol Cell. févr 2021;81(4):691-707.e6.
34. Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, Pfeiffer HK, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc Natl Acad Sci. 2 déc 2008;105(48):18782-7.
35. Reynolds MR, Lane AN, Robertson B, Kemp S, Liu Y, Hill BG, et al. Control of glutamine metabolism by the tumor suppressor Rb. Oncogene. 30 janv 2014;33(5):556-66.
36. Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt Stimulates Aerobic Glycolysis in Cancer Cells. Cancer Res. 1 juin 2004;64(11):3892-9.
37. Roberts DJ, Tan-Sah VP, Smith JM, Miyamoto S. Akt Phosphorylates HK-II at Thr-473 and Increases Mitochondrial HK-II Association to Protect Cardiomyocytes. J Biol Chem. août 2013;288(33):23798-806.
38. Sano H, Kane S, Sano E, Mîinea CP, Asara JM, Lane WS, et al. Insulin-stimulated Phosphorylation of a Rab GTPase-activating Protein Regulates GLUT4 Translocation. J Biol Chem. avr 2003;278(17):14599-602.
39. Waldhart AN, Dykstra H, Peck AS, Boguslawski EA, Madaj ZB, Wen J, et al. Phosphorylation of TXNIP by AKT Mediates Acute Influx of Glucose in Response to Insulin. Cell Rep. juin 2017;19(10):2005-13.
40. Düvel K, Yecies JL, Menon S, Raman P, Lipovsky AI, Souza AL, et al. Activation of a Metabolic Gene Regulatory Network Downstream of mTOR Complex 1. Mol Cell. juill 2010;39(2):171-83.
41. Hudson CC, Liu M, Chiang GG, Otterness DM, Loomis DC, Kaper F, et al. Regulation of Hypoxia-Inducible Factor 1α Expression and Function by the Mammalian Target of Rapamycin. Mol Cell Biol. 1 oct 2002;22(20):7004-14.
42. Zhong H, Chiles K, Feldser D, Laughner E, Hanrahan C, Georgescu MM, et al. Modulation of hypoxia-inducible factor 1alpha expression by the epidermal growth factor/phosphatidylinositol 3-kinase/PTEN/AKT/FRAP pathway in human prostate cancer cells: implications for tumor angiogenesis and therapeutics. Cancer Res. 15 mars 2000;60(6):1541-5.
43. Majumder PK, Febbo PG, Bikoff R, Berger R, Xue Q, McMahon LM, et al. mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. Nat Med. juin 2004;10(6):594-601.
44. Majmundar AJ, Wong WJ, Simon MC. Hypoxia-Inducible Factors and the Response to Hypoxic Stress. Mol Cell. oct 2010;40(2):294-309.
45. Denko NC. Hypoxia, HIF1 and glucose metabolism in the solid tumour. Nat Rev Cancer. sept 2008;8(9):705-13.
46. Kim J whan, Tchernyshyov I, Semenza GL, Dang CV. HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. Cell Metab. mars 2006;3(3):177-85.
47. Mikó E, Kovács T, Sebő É, Tóth J, Csonka T, Ujlaki G, et al. Microbiome—Microbial Metabolome—Cancer Cell Interactions in Breast Cancer—Familiar, but Unexplored. Cells.

29 mars 2019;8(4):293.
48. Wu J, Yang R, Zhang L, Li Y, Liu B, Kang H, et al. Metabolomics research on potential role for 9-cis-retinoic acid in breast cancer progression. Cancer Sci. juill 2018;109(7):2315-26.
49. Elia I, Haigis MC. Metabolites and the tumour microenvironment: from cellular mechanisms to systemic metabolism. Nat Metab. 4 janv 2021;3(1):21-32.
50. Lyssiotis CA, Kimmelman AC. Metabolic Interactions in the Tumor Microenvironment. Trends Cell Biol. nov 2017;27(11):863-75.
51. Hanahan D. Rethinking the war on cancer. The Lancet. févr 2014;383(9916):558-63.
52. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. mars 2011;144(5):646-74.
53. O'Neill LAJ, Kishton RJ, Rathmell J. A guide to immunometabolism for immunologists. Nat Rev Immunol. sept 2016;16(9):553-65.
54. Klein Geltink RI, Kyle RL, Pearce EL. Unraveling the Complex Interplay Between T Cell Metabolism and Function. Annu Rev Immunol. 26 avr 2018;36(1):461-88.
55. Michalek RD, Gerriets VA, Jacobs SR, Macintyre AN, MacIver NJ, Mason EF, et al. Cutting Edge: Distinct Glycolytic and Lipid Oxidative Metabolic Programs Are Essential for Effector and Regulatory CD4+ T Cell Subsets. J Immunol. 15 mars 2011;186(6):3299-303.
56. Wang R, Dillon CP, Shi LZ, Milasta S, Carter R, Finkelstein D, et al. The Transcription Factor Myc Controls Metabolic Reprogramming upon T Lymphocyte Activation. Immunity. déc 2011;35(6):871-82.
57. Phan AT, Goldrath AW. Hypoxia-inducible factors regulate T cell metabolism and function. Mol Immunol. déc 2015;68(2):527-35.
58. Su W, Chapman NM, Wei J, Zeng H, Dhungana Y, Shi H, et al. Protein Prenylation Drives Discrete Signaling Programs for the Differentiation and Maintenance of Effector Treg Cells. Cell Metab. déc 2020;32(6):996-1011.e7.
59. Viola A, Munari F, Sánchez-Rodríguez R, Scolaro T, Castegna A. The Metabolic Signature of Macrophage Responses. Front Immunol. 3 juill 2019;10:1462.
60. Nebert DW, Vesell ES. Can personalized drug therapy be achieved? A closer look at pharmaco-metabonomics. Trends Pharmacol Sci. nov 2006;27(11):580-6.
61. Badawy AAB. Tryptophan metabolism and disposition in cancer biology and immunotherapy. Biosci Rep. 30 nov 2022;42(11):BSR20221682.
62. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. mai 2021;71(3):209-49.
63. Gautier Defossez, Sandra Le Guyader-Peyrou, Zoé Uhry, Pascale Grosclaude, Marc Colonna, Emmanuelle Dantony, et al. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Étude à partir des registres des cancers du réseau Francim. In 2019.
64. Santé publique, france. Cancer du sein. 2018;
65. Institut National du Cancer. Panorama des cancers en France. In: 2022e éd.

66. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging: The Eighth Edition AJCC Cancer Staging Manual. CA Cancer J Clin. mars 2017;67(2):93-9.
67. Haute autorité de santé (HAS) Dépistage et prévention du cancer du col de l'utérus Actualisation du référentiel de pratiques de l'examen périodique de santé (EPS) HAS, Saint-Denis La Plaine (2013) Available online at: http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-08/referentieleps_format2clic_kc_col_uterus_2013-30-08_vf_mel.pdf [accesed on 29 April 2014].
68. Haute Autorité de Santé (HAS). Actualisation du référentiel de pratiques de l'examen périodique de santé - Dépistage et prévention du cancer du sein. 2015.
69. Li CI, Uribe DJ, Daling JR. Clinical characteristics of different histologic types of breast cancer. Br J Cancer. oct 2005;93(9):1046-52.
70. Rosenberg PS, Barker KA, Anderson WF. Estrogen Receptor Status and the Future Burden of Invasive and In Situ Breast Cancers in the United States. JNCI J Natl Cancer Inst [Internet]. sept 2015 [cité 9 avr 2023];107(9). Disponible sur: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djv159
71. Howlader N, Altekruse SF, Li CI, Chen VW, Clarke CA, Ries LAG, et al. US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status. JNCI J Natl Cancer Inst [Internet]. mai 2014 [cité 9 avr 2023];106(5). Disponible sur: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/dju055
72. Boyle P. Triple-negative breast cancer: epidemiological considerations and recommendations. Ann Oncol. août 2012;23:vi7-12.
73. Kohler BA, Sherman RL, Howlader N, Jemal A, Ryerson AB, Henry KA, et al. Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. JNCI J Natl Cancer Inst [Internet]. juin 2015 [cité 25 mars 2023];107(6). Disponible sur: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djv048
74. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. The Lancet. févr 2012;379(9814):432-44.
75. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. The Lancet. août 2011;378(9793):771-84.
76. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10 801 women in 17 randomised trials. The Lancet. nov 2011;378(9804):1707-16.
77. Munoz D, Near AM, van Ravesteyn NT, Lee SJ, Schechter CB, Alagoz O, et al. Effects of Screening and Systemic Adjuvant Therapy on ER-Specific US Breast Cancer

Mortality. JNCI J Natl Cancer Inst [Internet]. nov 2014 [cité 25 mars 2023];106(11). Disponible sur: https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/dju289
78. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. The Lancet. févr 2012;379(9814):432-44.
79. Wagner LI, Gray RJ, Sparano JA, Whelan TJ, Garcia SF, Yanez B, et al. Patient-Reported Cognitive Impairment Among Women With Early Breast Cancer Randomly Assigned to Endocrine Therapy Alone Versus Chemoendocrine Therapy: Results From TAILORx. J Clin Oncol. 10 juin 2020;38(17):1875-86.
80. Vaz-Luis I, Cottu P, Mesleard C, Martin AL, Dumas A, Dauchy S, et al. UNICANCER: French prospective cohort study of treatment-related chronic toxicity in women with localised breast cancer (CANTO). ESMO Open. 2019;4(5):e000562.
81. Rosenstock AS, Niu J, Giordano SH, Zhao H, Wolff AC, Chavez-MacGregor M. Acute myeloid leukemia and myelodysplastic syndrome after adjuvant chemotherapy: A population-based study among older breast cancer patients: AML/MDS After Adjuvant Chemotherapy. Cancer. 1 mars 2018;124(5):899-906.
82. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, et al. Pembrolizumab for Early Triple-Negative Breast Cancer. N Engl J Med. 27 févr 2020;382(9):810-21.
83. Adami HO, Malker B, Holmberg L, Persson I, Stone B. The Relation between Survival and Age at Diagnosis in Breast Cancer. N Engl J Med. 28 août 1986;315(9):559-63.
84. Billena C, Wilgucki M, Flynn J, Modlin L, Tadros A, Razavi P, et al. 10-Year Breast Cancer Outcomes in Women ≤35 Years of Age. Int J Radiat Oncol. mars 2021;109(4):1007-18.
85. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast Cancer in Young Women: Poor Survival Despite Intensive Treatment. Aziz SA, éditeur. PLoS ONE. 11 nov 2009;4(11):e7695.
86. Bastiaannet E, Liefers GJ, de Craen AJM, Kuppen PJK, van de Water W, Portielje JEA, et al. Breast cancer in elderly compared to younger patients in the Netherlands: stage at diagnosis, treatment and survival in 127,805 unselected patients. Breast Cancer Res Treat. déc 2010;124(3):801-7.
87. Eaker S, Dickman PW, Bergkvist L, Holmberg L, The Uppsala/Örebro Breast Cancer Group. Differences in Management of Older Women Influence Breast Cancer Survival: Results from a Population-Based Database in Sweden. Franco E, éditeur. PLoS Med. 17 janv 2006;3(3):e25.
88. van de Water W, Markopoulos C, van de Velde CJH, Seynaeve C, Hasenburg A, Rea D, et al. Association Between Age at Diagnosis and Disease-Specific Mortality Among Postmenopausal Women With Hormone Receptor–Positive Breast Cancer. JAMA [Internet]. 8 févr 2012 [cité 25 mars 2023];307(6). Disponible sur: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2012.84
89. Swain SM, Jeong JH, Geyer CE, Costantino JP, Pajon ER, Fehrenbacher L, et al.

Longer Therapy, Iatrogenic Amenorrhea, and Survival in Early Breast Cancer. N Engl J Med. 3 juin 2010;362(22):2053-65.
90. Fisher B, Slack NH, Bross IDF, Cooperating Investigators. Cancer of the breast: Size of neoplasm and prognosis. Cancer. nov 1969;24(5):1071-80.
91. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1 janv 1989;63(1):181-7.
92. Koscielny S, Tubiana M, Lê MG, Valleron AJ, Mouriesse H, Contesso G, et al. Breast cancer: Relationship between the size of the primary tumour and the probability of metastatic dissemination. Br J Cancer. juin 1984;49(6):709-15.
93. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. janv 2017;67(1):7-30.
94. Elston CW, Ellis IO. pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. nov 1991;19(5):403-10.
95. Pinder SE, Ellis IO, Galea M, O'Rouke S, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. III. Vascular invasion: relationship with recurrence and survival in a large study with long-term follow-up. Histopathology. janv 1994;24(1):41-7.
96. Luporsi E, André F, Spyratos F, Martin PM, Jacquemier J, Penault-Llorca F, et al. Ki-67: level of evidence and methodological considerations for its role in the clinical management of breast cancer: analytical and critical review. Breast Cancer Res Treat. avr 2012;132(3):895-915.
97. de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. Br J Cancer. mai 2007;96(10):1504-13.
98. Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients. The Breast. août 2008;17(4):323-34.
99. Harvey JM, Clark GM, Osborne CK, Allred DC. Estrogen Receptor Status by Immunohistochemistry Is Superior to the Ligand-Binding Assay for Predicting Response to Adjuvant Endocrine Therapy in Breast Cancer. J Clin Oncol. mai 1999;17(5):1474-1474.
100. Bartlett JMS, Brookes CL, Robson T, van de Velde CJH, Billingham LJ, Campbell FM, et al. Estrogen Receptor and Progesterone Receptor As Predictive Biomarkers of Response to Endocrine Therapy: A Prospectively Powered Pathology Study in the Tamoxifen and Exemestane Adjuvant Multinational Trial. J Clin Oncol. 20 avr 2011;29(12):1531-8.
101. Pertschuk LP, Kim DS, Nayer K, Feldman JG, Eisenberg KB, Carter AC, et al. Immunocytochemical estrogen and progestin receptor assays in breast cancer with monoclonal antibodies. Histopathologic, demographic, and biochemical correlations and relationship to endocrine response and survival. Cancer. 15 oct 1990;66(8):1663-70.
102. Colzani E, Liljegren A, Johansson ALV, Adolfsson J, Hellborg H, Hall PFL, et al. Prognosis of Patients With Breast Cancer: Causes of Death and Effects of Time Since Diagnosis, Age, and Tumor Characteristics. J Clin Oncol. 20 oct 2011;29(30):4014-21.
103. Colleoni M, Sun Z, Price KN, Karlsson P, Forbes JF, Thürlimann B, et al. Annual

Hazard Rates of Recurrence for Breast Cancer During 24 Years of Follow-Up: Results From the International Breast Cancer Study Group Trials I to V. J Clin Oncol. 20 mars 2016;34(9):927-35.
104. Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA, et al. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. Br J Cancer. févr 2014;110(3):565-72.
105. Thakkar JP, Mehta DG. A Review of an Unfavorable Subset of Breast Cancer: Estrogen Receptor Positive Progesterone Receptor Negative. The Oncologist. 1 mars 2011;16(3):276-85.
106. Gusterson BA, Gelber RD, Goldhirsch A, Price KN, Säve-Söderborgh J, Anbazhagan R, et al. Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. J Clin Oncol. juill 1992;10(7):1049-56.
107. Chia S, Norris B, Speers C, Cheang M, Gilks B, Gown AM, et al. Human Epidermal Growth Factor Receptor 2 Overexpression As a Prognostic Factor in a Large Tissue Microarray Series of Node-Negative Breast Cancers. J Clin Oncol. 10 déc 2008;26(35):5697-704.
108. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. janv 2018;19(1):40-50.
109. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol. févr 2015;26(2):259-71.
110. Leone JP, Graham N, Tolaney SM, Leone BA, Freedman RA, Hassett MJ, et al. Estimating long-term mortality in women with hormone receptor-positive breast cancer: The 'ESTIMATE' tool. Eur J Cancer. sept 2022;173:20-9.
111. ESTIMATE tool. Available online at: estimatetool.org (Accessed on March 25, 2023). In.
112. Dowsett M, Sestak I, Regan MM, Dodson A, Viale G, Thürlimann B, et al. Integration of Clinical Variables for the Prediction of Late Distant Recurrence in Patients With Estrogen Receptor–Positive Breast Cancer Treated With 5 Years of Endocrine Therapy: CTS5. J Clin Oncol. 1 juill 2018;36(19):1941-8.
113. Noordhoek I, Blok EJ, Meershoek-Klein Kranenbarg E, Putter H, Duijm-de Carpentier M, Rutgers EJT, et al. Overestimation of Late Distant Recurrences in High-Risk Patients With ER-Positive Breast Cancer: Validity and Accuracy of the CTS5 Risk Score in the TEAM and IDEAL Trials. J Clin Oncol. 1 oct 2020;38(28):3273-81.
114. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. févr 2010;12(1):R1.
115. POSH Steering Group, Maishman T, Copson E, Stanton L, Gerty S, Dicks E, et al. An evaluation of the prognostic model PREDICT using the POSH cohort of women aged ⩽40

years at breast cancer diagnosis. Br J Cancer. mars 2015;112(6):983-91.
116. Wishart GC, Bajdik CD, Azzato EM, Dicks E, Greenberg DC, Rashbass J, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. Eur J Surg Oncol EJSO. mai 2011;37(5):411-7.
117. de Glas NA, Bastiaannet E, Engels CC, de Craen AJM, Putter H, van de Velde CJH, et al. Validity of the online PREDICT tool in older patients with breast cancer: a population-based study. Br J Cancer. févr 2016;114(4):395-400.
118. Engelhardt EG, van den Broek AJ, Linn SC, Wishart GC, Rutgers EJTh, van de Velde AO, et al. Accuracy of the online prognostication tools PREDICT and Adjuvant! for early-stage breast cancer patients younger than 50 years. Eur J Cancer. juin 2017;78:37-44.
119. Wong HS, Subramaniam S, Alias Z, Taib NA, Ho GF, Ng CH, et al. The Predictive Accuracy of PREDICT: A Personalized Decision-Making Tool for Southeast Asian Women With Breast Cancer. Medicine (Baltimore). févr 2015;94(8):e593.
120. Wishart GC, Rakha E, Green A, Ellis I, Ali HR, Provenzano E, et al. Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. BMC Cancer. déc 2014;14(1):908.
121. Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. Br J Cancer. août 2012;107(5):800-7.
122. Cao L, Stabellini N, Towe CW, Miller ME, Shenk R, Amin AL, et al. BPI22-014: Independent Validation of the PREDICT Prognostication Tool in U.S. Breast Cancer Patients Using the National Cancer Database (NCDB). J Natl Compr Canc Netw. 31 mars 2022;20(3.5):BPI22-014.
123. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. oct 2012;490(7418):61-70.
124. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. oct 2010;12(5):R68.
125. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey R, Robertson JF, et al. Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol. juin 2004;203(2):661-71.
126. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 17 août 2000;406(6797):747-52.
127. Perreard L, Fan C, Quackenbush JF, Mullins M, Gauthier NP, Nelson E, et al. Classification and risk stratification of invasive breast carcinomas using a real-time quantitative RT-PCR assay. Breast Cancer Res. avr 2006;8(2):R23.
128. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci. 11 sept 2001;98(19):10869-74.
129. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci. 8 juill 2003;100(14):8418-23.

130. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci. 2 sept 2003;100(18):10393-8.
131. Yu K, Lee CH, Tan PH, Tan P. Conservation of Breast Cancer Molecular Subtypes and Transcriptional Patterns of Tumor Progression Across Distinct Ethnic Populations. Clin Cancer Res. 15 août 2004;10(16):5508-17.
132. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DSA, Nobel AB, et al. Concordance among Gene-Expression–Based Predictors for Breast Cancer. N Engl J Med. 10 août 2006;355(6):560-9.
133. Hu Z, Fan C, Oh DS, Marron J, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics. déc 2006;7(1):96.
134. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, et al. Definition of Clinically Distinct Molecular Subtypes in Estrogen Receptor–Positive Breast Carcinomas Through Genomic Grade. J Clin Oncol. 1 avr 2007;25(10):1239-46.
135. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast Cancer Subtypes and the Risk of Local and Regional Relapse. J Clin Oncol. 1 avr 2010;28(10):1684-91.
136. Prat A, Chaudhury A, Solovieff N, Paré L, Martinez D, Chic N, et al. Correlative Biomarker Analysis of Intrinsic Subtypes and Efficacy Across the MONALEESA Phase III Studies. J Clin Oncol. 1 mai 2021;39(13):1458-67.
137. Carey LA, Berry DA, Cirrincione CT, Barry WT, Pitcher BN, Harris LN, et al. Molecular Heterogeneity and Response to Neoadjuvant Human Epidermal Growth Factor Receptor 2 Targeting in CALGB 40601, a Randomized Phase III Trial of Paclitaxel Plus Trastuzumab With or Without Lapatinib. J Clin Oncol. 20 févr 2016;34(6):542-9.
138. Llombart-Cussac A, Cortés J, Paré L, Galván P, Bermejo B, Martínez N, et al. HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. Lancet Oncol. avr 2017;18(4):545-54.
139. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol. 2007;8(8):R157.
140. Lehmann BD, Jovanović B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. Sapino A, éditeur. PLOS ONE. 16 juin 2016;11(6):e0157368.
141. Weigelt B, Horlings H, Kreike B, Hayes M, Hauptmann M, Wessels L, et al. Refinement of breast cancer classification by molecular characterization of histological special types. J Pathol. oct 2008;216(2):141-50.
142. Untch* M, Gerber B, Harbeck N, Jackisch C, Marschner N, Möbus V, et al. 13th St. Gallen International Breast Cancer Conference 2013: Primary Therapy of Early Breast Cancer Evidence, Controversies, Consensus - Opinion of a German Team of Experts (Zurich 2013).

Breast Care. 2013;8(3):221-9.
143. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A Comparison of PAM50 Intrinsic Subtyping with Immunohistochemistry and Clinical Prognostic Factors in Tamoxifen-Treated Estrogen Receptor–Positive Breast Cancer. Clin Cancer Res. 1 nov 2010;16(21):5222-32.
144. Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N Engl J Med. 25 août 2016;375(8):717-29.
145. Tian S, Roepman P, van't Veer LJ, Bernards R, De Snoo F, Glas AM. Biological Functions of the Genes in the Mammaprint Breast Cancer Profile Reflect the Hallmarks of Cancer. Biomark Insights. janv 2010;5:BMI.S6184.
146. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, et al. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. JNCI J Natl Cancer Inst. 15 févr 2006;98(4):262-72.
147. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 12 juill 2018;379(2):111-21.
148. on behalf of the Austrian Breast and Colorectal Cancer Study group (ABCSG), Fitzal F, Filipits M, Rudas M, Greil R, Dietze O, et al. The genomic expression test EndoPredict is a prognostic tool for identifying risk of local recurrence in postmenopausal endocrine receptor-positive, her2neu-negative breast cancer patients randomised within the prospective ABCSG 8 trial. Br J Cancer. avr 2015;112(8):1405-10.
149. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, et al. 21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer. N Engl J Med. 16 déc 2021;385(25):2336-47.
150. Delaloge S, Saghatchian M, Ghouadni A, Fekih M, André F. Les signatures moléculaires commerciales : quelle utilité clinique ? Bull Cancer (Paris). juin 2015;102(6):S102-5.
151. Andre F, Ismaila N, Allison KH, Barlow WE, Collyar DE, Damodaran S, et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. J Clin Oncol. 1 juin 2022;40(16):1816-37.
152. Beckonert O, Monnerjahn J, Bonk U, Leibfritz D. Visualizing metabolic changes in breast-cancer tissue using1H-NMR spectroscopy and self-organizing maps. NMR Biomed. févr 2003;16(1):1-11.
153. Mountford CE, Somorjai RL, Malycha P, Gluch L, Lean C, Russell P, et al. Diagnosis and prognosis of breast cancer by magnetic resonance spectroscopy of fine-needle aspirates analysed using a statistical classification strategy. Br J Surg. 29 nov 2002;88(9):1234-40.
154. Sitter B, Lundgren S, Bathen TF, Halgunset J, Fjosne HE, Gribbestad IS. Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. NMR Biomed. févr 2006;19(1):30-40.
155. Sitter B, Sonnewald U, Spraul M, Fjösne HE, Gribbestad IS. High-resolution magic angle spinning MRS of breast cancer tissue: HR MAS OF BREAST CANCER. NMR

Biomed. août 2002;15(5):327-37.
156. Sharma U, Mehta A, Seenu V, Jagannathan NR. Biochemical characterization of metastatic lymph nodes of breast cancer patients by in vitro 1H magnetic resonance spectroscopy: a pilot study. Magn Reson Imaging. juin 2004;22(5):697-706.
157. Aboud OA, Weiss RH. New Opportunities from the Cancer Metabolome. Clin Chem. 1 janv 2013;59(1):138-46.
158. Da Cunha PA, Nitusca D, Canto LMD, Varghese RS, Ressom HW, Willey S, et al. Metabolomic Analysis of Plasma from Breast Cancer Patients Using Ultra-High-Performance Liquid Chromatography Coupled with Mass Spectrometry: An Untargeted Study. Metabolites. 17 mai 2022;12(5):447.
159. Yuan B, Schafferer S, Tang Q, Scheffler M, Nees J, Heil J, et al. A plasma metabolite panel as biomarkers for early primary breast cancer detection. Int J Cancer. juin 2019;144(11):2833-42.
160. Jové M, Collado R, Quiles JL, Ramírez-Tortosa MC, Sol J, Ruiz-Sanjuan M, et al. A plasma metabolomic signature discloses human breast cancer. Oncotarget. 21 mars 2017;8(12):19522-33.
161. Slupsky CM, Steed H, Wells TH, Dabbs K, Schepansky A, Capstick V, et al. Urine Metabolite Analysis Offers Potential Early Diagnosis of Ovarian and Breast Cancers. Clin Cancer Res. 1 déc 2010;16(23):5835-41.
162. Omran MM, Rashed RE, Darwish H, Belal AA, Mohamed FZ. Development of a gas chromatography–mass spectrometry method for breast cancer diagnosis based on nucleoside metabolomes 1-methyl adenosine, 1-methylguanosine and 8-hydroxy-2′-deoxyguanosine. Biomed Chromatogr [Internet]. janv 2020 [cité 24 avr 2023];34(1). Disponible sur: https://onlinelibrary.wiley.com/doi/10.1002/bmc.4713
163. Murata T, Yanagisawa T, Kurihara T, Kaneko M, Ota S, Enomoto A, et al. Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination. Breast Cancer Res Treat. oct 2019;177(3):591-601.
164. Cheng F, Wang Z, Huang Y, Duan Y, Wang X. Investigation of salivary free amino acid profile for early diagnosis of breast cancer with ultra performance liquid chromatography-mass spectrometry. Clin Chim Acta. juill 2015;447:23-31.
165. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A, et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. déc 2019;17(1):178.
166. Meadows AL, Kong B, Berdichevsky M, Roy S, Rosiva R, Blanch HW, et al. Metabolic and Morphological Differences between Rapidly Proliferating Cancerous and Normal Breast Epithelial Cells. Biotechnol Prog. 4 avr 2008;24(2):334-41.
167. Bathen TF, Geurts B, Sitter B, Fjøsne HE, Lundgren S, Buydens LM, et al. Feasibility of MR Metabolomics for Immediate Analysis of Resection Margins during Breast Cancer Surgery. Han A, éditeur. PLoS ONE. 17 avr 2013;8(4):e61578.
168. Maria RM, Altei WF, Andricopulo AD, Becceneri AB, Cominetti MR, Venâncio T, et al. Characterization of metabolic profile of intact non-tumor and tumor breast cells by high-resolution magic angle spinning nuclear magnetic resonance spectroscopy. Anal Biochem.

nov 2015;488:14-8.
169. Cala MP, Aldana J, Medina J, Sánchez J, Guio J, Wist J, et al. Multiplatform plasma metabolic and lipid fingerprinting of breast cancer: A pilot control-case study in Colombian Hispanic women. Bathen TF, éditeur. PLOS ONE. 13 févr 2018;13(2):e0190958.
170. Suman S, Sharma RK, Kumar V, Sinha N, Shukla Y. Metabolic fingerprinting in breast cancer stages through 1H NMR spectroscopy-based metabolomic analysis of plasma. J Pharm Biomed Anal. oct 2018;160:38-45.
171. Kanaan YM, Sampey BP, Beyene D, Esnakula AK, Naab TJ, Ricks-Santi LJ, et al. Metabolic profile of triple-negative breast cancer in African-American women reveals potential biomarkers of aggressive disease. Cancer Genomics Proteomics. 2014;11(6):279-94.
172. Tayyari F, Gowda GAN, Olopade OF, Berg R, Yang HH, Lee MP, et al. Metabolic profiles of triple-negative and luminal A breast cancer subtypes in African-American identify key metabolic differences. Oncotarget. 20 févr 2018;9(14):11677-90.
173. Cha Y, Kim ES, Koo J. Amino Acid Transporters and Glutamine Metabolism in Breast Cancer. Int J Mol Sci. 19 mars 2018;19(3):907.
174. Blücher C, Stadler SC. Obesity and Breast Cancer: Current Insights on the Role of Fatty Acids and Lipid Metabolism in Promoting Breast Cancer Growth and Progression. Front Endocrinol. 30 oct 2017;8:293.
175. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature. août 2011;476(7360):346-50.
176. Du T, Zhu L, Levine KM, Tasdemir N, Lee AV, Vignali DAA, et al. Invasive lobular and ductal breast carcinoma differ in immune response, protein translation efficiency and metabolism. Sci Rep. 8 mai 2018;8(1):7205.
177. Kulkoyluoglu-Cotul E, Arca A, Madak-Erdogan Z. Crosstalk between Estrogen Signaling and Breast Cancer Metabolism. Trends Endocrinol Metab. janv 2019;30(1):25-38.
178. Nie C, Lv H, Bie L, Hou H, Chen X. Hypoxia-inducible factor 1-alpha expression correlates with response to neoadjuvant chemotherapy in women with breast cancer. Medicine (Baltimore). déc 2018;97(51):e13551.
179. Schito L, Rey S. Hypoxic pathobiology of breast cancer metastasis. Biochim Biophys Acta BBA - Rev Cancer. août 2017;1868(1):239-45.
180. Craze ML, Cheung H, Jewa N, Coimbra NDM, Soria D, El-Ansari R, et al. MYC regulation of glutamine–proline regulatory axis is key in luminal B breast cancer. Br J Cancer. janv 2018;118(2):258-65.
181. Kim S, Kim DH, Jung WH, Koo JS. Expression of glutamine metabolism-related proteins according to molecular subtype of breast cancer. Endocr Relat Cancer. juin 2013;20(3):339-48.
182. El Ansari R, McIntyre A, Craze ML, Ellis IO, Rakha EA, Green AR. Altered glutamine metabolism in breast cancer; subtype dependencies and alternative adaptations. Histopathology. janv 2018;72(2):183-90.
183. Gandhi N, Das G. Metabolic Reprogramming in Breast Cancer and Its Therapeutic Implications. Cells. 26 janv 2019;8(2):89.

184. Raphael J, Desautels D, Pritchard KI, Petkova E, Shah PS. Phosphoinositide 3-kinase inhibitors in advanced breast cancer: A systematic review and meta-analysis. Eur J Cancer. mars 2018;91:38-46.
185. Keegan NM, Gleeson JP, Hennessy BT, Morris PG. PI3K inhibition to overcome endocrine resistance in breast cancer. Expert Opin Investig Drugs. 2 janv 2018;27(1):1-15.
186. Hadad SM, Baker L, Quinlan PR, Robertson KE, Bray SE, Thomson G, et al. Histological evaluation of AMPK signalling in primary breast cancer. BMC Cancer. déc 2009;9(1):307.
187. Laderoute KR, Calaoagan JM, Chao W ru, Dinh D, Denko N, Duellman S, et al. 5′-AMP-activated Protein Kinase (AMPK) Supports the Growth of Aggressive Experimental Human Breast Cancer Tumors. J Biol Chem. août 2014;289(33):22850-64.
188. Wang L, Zhang S, Wang X. The Metabolic Mechanisms of Breast Cancer Metastasis. Front Oncol. 7 janv 2021;10:602416.
189. Alakwaa FM, Chaudhary K, Garmire LX. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. J Proteome Res. 5 janv 2018;17(1):337-47.
190. Dougan MM, Li Y, Chu LW, Haile RW, Whittemore AS, Han SS, et al. Metabolomic profiles in breast cancer:a pilot case-control study in the breast cancer family registry. BMC Cancer. déc 2018;18(1):532.
191. Fan Y, Zhou X, Xia TS, Chen Z, Li J, Liu Q, et al. Human plasma metabolomics for identifying differential metabolites and predicting molecular subtypes of breast cancer. Oncotarget. 1 mars 2016;7(9):9925-38.
192. Roig B, Rodríguez-Balada M, Samino S, Lam EWF, Guaita-Esteruelas S, Gomes AR, et al. Metabolomics reveals novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of human breast cancer. Sci Rep. 19 déc 2017;7(1):17831.
193. Gong Y, Ji P, Yang YS, Xie S, Yu TJ, Xiao Y, et al. Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. Cell Metab. janv 2021;33(1):51-64.e9.
194. Wu Q, ba-alawi W, Deblois G, Cruickshank J, Duan S, Lima-Fernandes E, et al. GLUT1 inhibition blocks growth of RB1-positive triple negative breast cancer. Nat Commun. 21 août 2020;11(1):4205.
195. Kisanga ER, Mellgren G, Lien EA. Excretion of hydroxylated metabolites of tamoxifen in human bile and urine. Anticancer Res. 2005;25(6C):4487-92.
196. Visekruna A, Luu M. The Role of Short-Chain Fatty Acids and Bile Acids in Intestinal and Liver Function, Inflammation, and Carcinogenesis. Front Cell Dev Biol. 26 juill 2021;9:703218.
197. Arnone AA, Cline JM, Soto-Pantoja DR, Cook KL. Investigating the role of endogenous estrogens, hormone replacement therapy, and blockade of estrogen receptor-α activity on breast metabolic signaling. Breast Cancer Res Treat. nov 2021;190(1):53-67.
198. Giskeødegård GF, Lundgren S, Sitter B, Fjøsne HE, Postma G, Buydens LMC, et al. Lactate and glycine-potential MR biomarkers of prognosis in estrogen receptor-positive breast

cancers: METABOLIC BIOMARKERS OF BREAST CANCER PROGNOSIS. NMR Biomed. nov 2012;25(11):1271-9.
199. Özer Ö, Nemutlu E, Reçber T, Eylem CC, Aktas BY, Kır S, et al. Liquid biopsy markers for early diagnosis of brain metastasis patients with breast cancer by metabolomics. Eur J Mass Spectrom. avr 2022;28(1-2):56-64.
200. Zhu W, Qian W, Liao W, Huang X, Xu J, Qu W, et al. Non-Invasive and Real-Time Monitoring of the Breast Cancer Metastasis Degree via Metabolomics. Cancers. 14 nov 2022;14(22):5589.
201. Nees J, Schafferer S, Yuan B, Tang Q, Scheffler M, Hartkopf A, et al. How previous treatment changes the metabolomic profile in patients with metastatic breast cancer. Arch Gynecol Obstet. 25 avr 2022;306(6):2115-22.
202. Lin J, Lee IM, Song Y, Cook NR, Selhub J, Manson JE, et al. Plasma Homocysteine and Cysteine and Risk of Breast Cancer in Women. Cancer Res. 15 mars 2010;70(6):2397-405.
203. Nittoli AC, Costantini S, Sorice A, Capone F, Ciarcia R, Marzocco S, et al. Effects of α-zearalenol on the metabolome of two breast cancer cell lines by 1H-NMR approach. Metabolomics. mars 2018;14(3):33.
204. Abdelmagid SA, Rickard JA, McDonald WJ, Thomas LN, Too CKL. CAT-1-mediated arginine uptake and regulation of nitric oxide synthases for the survival of human breast cancer cell lines. J Cell Biochem. avr 2011;112(4):1084-92.
205. Cao Y, Feng Y, Zhang Y, Zhu X, Jin F. L-Arginine supplementation inhibits the growth of breast cancer by enhancing innate and adaptive immune responses mediated by suppression of MDSCs in vivo. BMC Cancer. déc 2016;16(1):343.
206. Nagata C, Wada K, Tsuji M, Hayashi M, Takeda N, Yasuda K. Plasma amino acid profiles are associated with biomarkers of breast cancer risk in premenopausal Japanese women. Cancer Causes Control. févr 2014;25(2):143-9.
207. Knott SRV, Wagenblast E, Khan S, Kim SY, Soto M, Wagner M, et al. Asparagine bioavailability governs metastasis in a model of breast cancer. Nature. 15 févr 2018;554(7692):378-81.
208. Kato M, Onishi H, Matsumoto K, Motoshita J, Tsuruta N, Higuchi K, et al. Prognostic significance of urine N1, N12-diacetylspermine in patients with non-small cell lung cancer. Anticancer Res. juin 2014;34(6):3053-9.
209. Sugimoto M, Hiramatsu K, Kamei S, Kinoshita K, Hoshino M, Iwasaki K, et al. Significance of urinaryN 1,N 8-diacetylspermidine andN 1,N 12-diacetylspermine as indicators of neoplastic diseases. J Cancer Res Clin Oncol. mai 1995;121(5):317-9.
210. Hiramatsu K, Sugimoto M, Kamei S, Hoshino M, Kinoshita K, Iwasaki K, et al. Diagnostic and prognostic usefulness of N 1 , N 8 -diacetylspermidine and N 1 , N 12 -diacetylspermine in urine as novel markers of malignancy. J Cancer Res Clin Oncol. 27 oct 1997;123(10):539-45.
211. Hiramatsu K, Takahashi K, Yamaguchi T, Matsumoto H, Miyamoto H, Tanaka S, et al. *N* 1, *N* 12-Diacetylspermine as a Sensitive and Specific Novel Marker for Early- and Late-Stage Colorectal and Breast Cancers. Clin Cancer Res. 15 avr 2005;11(8):2986-90.

212. Cervelli M, Bellavia G, Fratini E, Amendola R, Polticelli F, Barba M, et al. Spermine oxidase (SMO) activity in breast tumor tissues and biochemical analysis of the anticancer spermine analogues BENSpm and CPENSpm. BMC Cancer. déc 2010;10(1):555.
213. Lu B, Liang X, Scott GK, Chang CH, Baldwin MA, Thomas T, et al. Polyamine inhibition of estrogen receptor (ER) DNA-binding and ligand-binding functions. Breast Cancer Res Treat. avr 1998;48(3):243-57.
214. Fahrmann JF, Vykoukal J, Fleury A, Tripathi S, Dennison JB, Murage E, et al. Association Between Plasma Diacetylspermine and Tumor Spermine Synthase With Outcome in Triple-Negative Breast Cancer. JNCI J Natl Cancer Inst. 1 juin 2020;112(6):607-16.
215. Tang X, Lin CC, Spasojevic I, Iversen ES, Chi JT, Marks JR. A joint analysis of metabolomics and genetics of breast cancer. Breast Cancer Res. août 2014;16(4):415.
216. Ballou Y, Rivas A, Belmont A, Patel L, Amaya C, Lipson S, et al. 5-HT serotonin receptors modulate mitogenic signaling and impact tumor cell viability. Mol Clin Oncol [Internet]. 19 juill 2018 [cité 23 janv 2023]; Disponible sur: http://www.spandidos-publications.com/10.3892/mco.2018.1681
217. Gautam J, Banskota S, Regmi SC, Ahn S, Jeon YH, Jeong H, et al. Tryptophan hydroxylase 1 and 5-HT7 receptor preferentially expressed in triple-negative breast cancer promote cancer progression through autocrine serotonin signaling. Mol Cancer. déc 2016;15(1):75.
218. Balakrishna P, George S, Hatoum H, Mukherjee S. Serotonin Pathway in Cancer. Int J Mol Sci. 28 janv 2021;22(3):1268.
219. Jose J, Tavares CDJ, Ebelt ND, Lodi A, Edupuganti R, Xie X, et al. Serotonin Analogues as Inhibitors of Breast Cancer Cell Growth. ACS Med Chem Lett. 12 oct 2017;8(10):1072-6.
220. Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF, et al. NCCN Task Force Report: Evaluating the Clinical Utility of Tumor Markers in Oncology. J Natl Compr Canc Netw. nov 2011;9(Suppl_5):S-1-S-32.
221. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. Genet Med. janv 2009;11(1):3-14.
222. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Anal Bioanal Chem. janv 2007;387(2):525-7.
223. Riekeberg E, Powers R. New frontiers in metabolomics: from measurement to insight. F1000Research. 2017;6:1148.
224. McLafferty FW. Tandem mass spectrometry. Science. 16 oct 1981;214(4518):280-7.
225. Pavia DL, éditeur. Introduction to organic laboratory techniques: a small scale approach. 2nd ed. Belmont, CA: Thomson Brooks/Cole; 2005. 1028 p. (Brooks/Cole laboratory series for organic chemistry).
226. Gwenola, Jean-Louis. Methodes instrumentales d'analyse chimique et applications methodes chromatographiques, electrophoreses, methodes spectrales et thermiques. Tec & Doc; 2011.

227. Annesley TM. Ion suppression in mass spectrometry. Clin Chem. juill 2003;49(7):1041-4.
228. Antignac JP, de Wasch K, Monteau F, De Brabander H, Andre F, Le Bizec B. The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. Anal Chim Acta. janv 2005;529(1-2):129-36.
229. Wishart DS. Advances in metabolite identification. Bioanalysis. août 2011;3(15):1769-82.
230. Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. The Analyst. 2005;130(5):606.
231. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. The Analyst. 2012;137(2):293-300.
232. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. PloS One. 16 févr 2011;6(2):e16957.
233. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The human urine metabolome. PloS One. 2013;8(9):e73076.
234. Roberts LD, Souza AL, Gerszten RE, Clish CB. Targeted Metabolomics. Curr Protoc Mol Biol [Internet]. avr 2012 [cité 22 févr 2023];98(1). Disponible sur: https://onlinelibrary.wiley.com/doi/10.1002/0471142727.mb3002s98
235. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. J Am Soc Mass Spectrom. 1 déc 2016;27(12):1897-905.
236. Vinayavekhin N, Saghatelian A. Untargeted metabolomics. Curr Protoc Mol Biol. avr 2010;Chapter 30:Unit 30.1.1-24.
237. Zubarev RA, Makarov A. Orbitrap Mass Spectrometry. Anal Chem. 4 juin 2013;85(11):5288-96.
238. Sugimoto M, Kawakami M, Robert M, Soga T, Tomita M. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. Curr Bioinforma. 1 mars 2012;7(1):96-108.
239. Pedrioli PGA, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, et al. A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol. 1 nov 2004;22(11):1459-66.
240. Gorrochategui E, Jaumot J, Lacorte S, Tauler R. Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow. TrAC Trends Anal Chem. sept 2016;82:425-42.
241. Katajamaa M, Orešič M. Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics. déc 2005;6(1):179.
242. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. déc 2010;11(1):395.
243. Wu Y, Li L. Sample normalization methods in quantitative metabolomics. J Chromatogr A. janv 2016;1430:80-95.
244. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al.

HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 4 janv 2018;46(D1):D608-17.
245. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, Hermann G, et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal Chem. 6 mars 2018;90(5):3156-64.
246. Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. J Chem Educ. 1 nov 2010;87(11):1123-4.
247. Kind T, Fiehn O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics. déc 2006;7(1):234.
248. Wiener MC, Sachs JR, Deyanova EG, Yates NA. Differential Mass Spectrometry: A Label-Free LC−MS Method for Finding Significant Differences in Complex Peptide and Protein Mixtures. Anal Chem. 1 oct 2004;76(20):6085-96.
249. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics. 12 janv 2007;2(4):171-96.
250. Shaffer JP. Multiple Hypothesis Testing. Annu Rev Psychol. janv 1995;46(1):561-84.
251. Boccard J, Veuthey JL, Rudaz S. Knowledge discovery in metabolomics: An overview of MS data handling. J Sep Sci. févr 2010;33(3):290-304.
252. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. sept 1933;24(6):417-41.
253. Nyamundanda G, Brennan L, Gormley IC. Probabilistic principal component analysis for metabolomic data. BMC Bioinformatics. déc 2010;11(1):571.
254. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci. 8 déc 1998;95(25):14863-8.
255. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst. oct 2001;58(2):109-30.
256. Kettaneh-Wold N. Analysis of mixture data with partial least squares. Chemom Intell Lab Syst. avr 1992;14(1-3):57-69.
257. Wold S, Ruhe A, Wold H, Dunn, III WJ. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. SIAM J Sci Stat Comput. sept 1984;5(3):735-43.
258. Lindgren F, Hansen B, Karcher W, Sjöström M, Eriksson L. Model validation by permutation tests: Applications to variable selection. J Chemom. sept 1996;10(5-6):521-32.
259. Berry KJ, Johnston JE, Mielke PW. Permutation methods: Permutation methods. Wiley Interdiscip Rev Comput Stat. nov 2011;3(6):527-42.
260. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. Bioinformatics. 1 août 2005;21(15):3301-7.
261. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. J R Stat Soc Ser B Methodol. janv 1974;36(2):111-33.
262. Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, et al. Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of

global metabolomics data. Nat Protoc. août 2022;17(8):1735-61.
263. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr. mars 2008;17(2):145-51.
264. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform. 1 mars 2018;19(2):286-302.
265. Rosato A, Tenori L, Cascante M, De Atauri Carulla PR, Martins dos Santos VAP, Saccenti E. From correlation to causation: analysis of metabolomics data using systems biology approaches. Metabolomics. avr 2018;14(4):37.
266. Xia T, Jiang H, Li C, Tian M, Zhang H. Molecular imaging in tracking tumor stem-like cells. BioMed Res Int [Internet]. 2012 [cité 18 janv 2015];2012. Disponible sur: http://downloads.hindawi.com/journals/biomed/2012/420364.pdf
267. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. mars 1982;28(2):129-37.
268. Steinhaus, H. Sur la division des corps matériels en parties. In 1957. p. 801-4. (Bulletin L'Académie Polonaise des Science; vol. 4).
269. Mercer J. XVI. Functions of positive and negative type, and their connection the theory of integral equations. Philos Trans R Soc Lond Ser Contain Pap Math Phys Character. janv 1909;209(441-458):415-46.
270. Canu S, Mary X, Rakotomamonjy A. Functional learning through kernels. 2009 [cité 8 mars 2023]; Disponible sur: https://arxiv.org/abs/0910.1013
271. Gönen M, Kandemir M, Kaski S. Multitask Learning Using Regularized Multiple Kernel Learning. In: Lu BL, Zhang L, Kwok J, éditeurs. Neural Information Processing [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011 [cité 8 mars 2023]. p. 500-9. (Lecture Notes in Computer Science; vol. 7063). Disponible sur: http://link.springer.com/10.1007/978-3-642-24958-7_58
272. Bach FR, Lanckriet GRG, Jordan MI. Multiple kernel learning, conic duality, and the SMO algorithm. In: Twenty-first international conference on Machine learning - ICML '04 [Internet]. Banff, Alberta, Canada: ACM Press; 2004 [cité 8 mars 2023]. p. 6. Disponible sur: http://portal.acm.org/citation.cfm?doid=1015330.1015424
273. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nat Methods. avr 2017;14(4):414-6.
274. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J R Stat Soc Ser B Methodol. janv 1996;58(1):267-88.
275. Witten DM, Tibshirani R. A Framework for Feature Selection in Clustering. J Am Stat Assoc. juin 2010;105(490):713-26.
276. Gilet C, Deprez M, Caillau J-B et al. Clustering with feature selection using alternating minimization, Application to computational biology. 2017;
277. Condat L. Fast projection onto the simplex and the $$\pmb {l}_\mathbf {1}$$ l 1 ball. Math Program. juill 2016;158(1-2):575-85.
278. Gaude E, Frezza C. Tissue-specific and convergent metabolic transformation of cancer

correlates with metastatic potential and patient survival. Nat Commun. 10 oct 2016;7(1):13041.
279. Rosario SR, Long MD, Affronti HC, Rowsam AM, Eng KH, Smiraglia DJ. Pan-cancer analysis of transcriptional metabolic dysregulation using The Cancer Genome Atlas. Nat Commun. 14 déc 2018;9(1):5330.
280. Bernhardt S, Bayerlová M, Vetter M, Wachter A, Mitra D, Hanf V, et al. Proteomic profiling of breast cancer metabolism identifies SHMT2 and ASCT2 as prognostic factors. Breast Cancer Res. déc 2017;19(1):112.
281. Vinayavekhin N, Saghatelian A. Untargeted Metabolomics. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al., éditeurs. Current Protocols in Molecular Biology [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010 [cité 22 févr 2023]. p. mb3001s90. Disponible sur: https://onlinelibrary.wiley.com/doi/10.1002/0471142727.mb3001s90
282. Asiago VM, Alvarado LZ, Shanaiah N, Gowda GAN, Owusu-Sarfo K, Ballas RA, et al. Early Detection of Recurrent Breast Cancer Using Metabolite Profiling. Cancer Res. 1 nov 2010;70(21):8309-18.
283. Oakman C, Tenori L, Claudino WM, Cappadona S, Nepi S, Battaglia A, et al. Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods. Ann Oncol. juin 2011;22(6):1295-301.
284. Galal A, Talal M, Moustafa A. Applications of machine learning in metabolomics: Disease modeling and classification. Front Genet. 24 nov 2022;13:1017340.
285. Dhall D, Kaur R, Juneja M. Machine Learning: A Review of the Algorithms and Its Applications. In: Singh PK, Kar AK, Singh Y, Kolekar MH, Tanwar S, éditeurs. Proceedings of ICRIC 2019 [Internet]. Cham: Springer International Publishing; 2020 [cité 8 avr 2023]. p. 47-63. (Lecture Notes in Electrical Engineering; vol. 597). Disponible sur: http://link.springer.com/10.1007/978-3-030-29407-6_5
286. Gal J, Bailleux C, Chardin D, Pourcher T, Gilhodes J, Jing L, et al. Comparison of unsupervised machine-learning methods to identify metabolomic signatures in patients with localized breast cancer. Comput Struct Biotechnol J. 2020;18:1509-24.
287. Candido dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. Breast Cancer Res. déc 2017;19(1):58.
288. Darlix A, Louvel G, Fraisse J, Jacot W, Brain E, Debled M, et al. Impact of breast cancer molecular subtypes on the incidence, kinetics and prognosis of central nervous system metastases in a large multicentre real-life cohort. Br J Cancer. 10 déc 2019;121(12):991-1000.
289. Fallah Y, Brundage J, Allegakoen P, Shajahan-Haq AN. MYC-Driven Pathways in Breast Cancer Subtypes. Biomolecules. 11 juill 2017;7(4):53.
290. Bello-Fernandez C, Packham G, Cleveland JL. The ornithine decarboxylase gene is a transcriptional target of c-Myc. Proc Natl Acad Sci. 15 août 1993;90(16):7804-8.
291. Platten M, Wick W, Van den Eynde BJ. Tryptophan Catabolism in Cancer: Beyond

IDO and Tryptophan Depletion. Cancer Res. 1 nov 2012;72(21):5435-40.
292. Wu H, Gong J, Liu Y. Indoleamine 2, 3-dioxygenase regulation of immune response (Review). Mol Med Rep [Internet]. 1 févr 2018 [cité 23 janv 2023]; Disponible sur: http://www.spandidos-publications.com/10.3892/mmr.2018.8537
293. Ino K, Yamamoto E, Shibata K, Kajiyama H, Yoshida N, Terauchi M, et al. Inverse Correlation between Tumoral Indoleamine 2,3-Dioxygenase Expression and Tumor-Infiltrating Lymphocytes in Endometrial Cancer: Its Association with Disease Progression and Survival. Clin Cancer Res. 15 avr 2008;14(8):2310-7.
294. Brody JR, Costantino CL, Berger AC, Sato T, Lisanti MP, Yeo CJ, et al. Expression of indoleamine 2,3-dioxygenase in metastatic malignant melanoma recruits regulatory T cells to avoid immune detection and affects survival. Cell Cycle. 15 juin 2009;8(12):1930-4.
295. Seeber A, Klinglmair G, Fritz J, Steinkohl F, Zimmer K, Aigner F, et al. High IDO -1 expression in tumor endothelial cells is associated with response to immunotherapy in metastatic renal cell carcinoma. Cancer Sci. mai 2018;109(5):1583-91.
296. Botticelli A, Cerbelli B, Lionetto L, Zizzari I, Salati M, Pisano A, et al. Can IDO activity predict primary resistance to anti-PD-1 treatment in NSCLC? J Transl Med. déc 2018;16(1):219.
297. Frumento G, Rotondo R, Tonetti M, Damonte G, Benatti U, Ferrara GB. Tryptophan-derived Catabolites Are Responsible for Inhibition of T and Natural Killer Cell Proliferation Induced by Indoleamine 2,3-Dioxygenase. J Exp Med. 19 août 2002;196(4):459-68.
298. Ramapriyan R, Caetano MS, Barsoumian HB, Mafra ACP, Zambalde EP, Menon H, et al. Altered cancer metabolism in mechanisms of immunotherapy resistance. Pharmacol Ther. mars 2019;195:162-71.
299. Baganz NL, Blakely RD. A Dialogue between the Immune System and Brain, Spoken in the Language of Serotonin. ACS Chem Neurosci. 16 janv 2013;4(1):48-63.
300. Herr N, Bode C, Duerschmied D. The Effects of Serotonin in Immune Cells. Front Cardiovasc Med. 20 juill 2017;4:48.
301. Peyraud F, Guegan JP, Bodet D, Cousin S, Bessede A, Italiano A. Targeting Tryptophan Catabolism in Cancer Immunotherapy Era: Challenges and Perspectives. Front Immunol. 31 janv 2022;13:807271.
302. Cortes J, Cescon DW, Rugo HS, Nowecki Z, Im SA, Yusof MM, et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. The Lancet. déc 2020;396(10265):1817-28.
303. Franchet C, Duprez-Paumier R, Lacroix-Triki M. Cancer du sein luminal et apport des classifications intrinsèques moléculaires : comment identifier les tumeurs luminales A et B en 2015 ? Bull Cancer (Paris). juin 2015;102(6):S34-46.
304. Joyon N, Penault-Llorca F, Lacroix-Triki M. Classification et signatures moléculaires des cancers du sein en 2017. Oncologie. avr 2017;19(3-4):64-70.
305. Giuliano AE, Edge SB, Hortobagyi GN. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. Ann Surg Oncol. juill 2018;25(7):1783-5.

306. Bloom HJG, Richardson WW. Histological Grading and Prognosis in Breast Cancer: A Study of 1409 Cases of which 359 have been Followed for 15 Years. Br J Cancer. sept 1957;11(3):359-77.