



HAL
open science

Développements méthodologiques en spectrométrie de masse et en analyse protéomique pour la recherche de biomarqueurs de différents types de cancers

Sarah Lennon

► **To cite this version:**

Sarah Lennon. Développements méthodologiques en spectrométrie de masse et en analyse protéomique pour la recherche de biomarqueurs de différents types de cancers. Chimie analytique. Université de Strasbourg, 2014. Français. NNT : 2014STRAF036 . tel-01224067

HAL Id: tel-01224067

<https://theses.hal.science/tel-01224067v1>

Submitted on 4 Nov 2015

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

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

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

UMR7178-IPHC



THÈSE

Présentée par

Sarah LENNON

Soutenue le 25 septembre 2014

pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline / Spécialité : Chimie analytique

Développements méthodologiques en spectrométrie de masse et analyse protéomique pour la recherche de biomarqueurs de différents types de cancers

THÈSE dirigée par :

Dr Alain VAN DORSSELAER
Dr Sarah CIANFERANI

Directeur de recherches, CNRS, Université de Strasbourg
Chargée de recherches, CNRS, Université de Strasbourg

RAPPORTEURS :

Dr Charles PINEAU
Dr Odile SCHILTZ

Directeur de recherches, INSERM, Université de Rennes
Directeur de recherches, CNRS, Université de Toulouse

EXAMINATEURS :

Dr Sylviane MULLER
Dr Hervé CHNEIWEISS

Directeur de recherches, CNRS, Université de Strasbourg
Directeur de recherches, CNRS, Université Pierre et Marie Curie

À ma famille

À ma grand-mère, Jeannine

À Gwenael

« Ce qui est étrange dans l'acquisition de savoir, c'est que plus j'avance, plus je me rends compte que je ne savais même pas que ce que je ne savais pas existait. »

Daniel Keys, extrait de *Des fleurs pour Algernon*



Remerciements

En tout premier lieu, je tiens à remercier Alain Van Dorsselaer de m'avoir permis de bénéficier d'une formation de qualité tout en travaillant dans un environnement scientifique et humain de premier plan.

Je remercie tout particulièrement Sarah Cianféroni de m'avoir encadrée et soutenue tout au long de ma thèse. Merci pour tes précieux conseils et la patience dont tu as fait preuve.

Je remercie vivement Odile Schiltz, Charles Pineau, Sylviane Muller et Hervé Chneiweiss d'avoir accepté d'évaluer mon travail de thèse.

Un grand merci à l'ensemble de mes collaborateurs pour les thématiques variées sur lesquelles j'ai eu l'occasion de travailler : Jacques Haiech et Marie-Claude Kilhoffer – sur les cellules souches cancéreuses –, merci pour votre écoute et vos conseils, Laurent Mauvieux et Laurent Miguet – sur la caractérisation des lymphomes –, merci de m'avoir permis d'accéder au laboratoire d'hématologie de l'hôpital et d'avoir passé du temps à m'en expliquer le fonctionnement, Catherine Thieblemont – sur la prédiction des rechutes du système nerveux central – ainsi que Julie Brunet, Valérie Lamour, Marc Ruff, Arnaud Poterszman et Laura Radu, pour les sujets parallèles non présentés dans cette thèse.

Je remercie l'ensemble du LSMBO pour l'esprit chaleureux, de convivialité et de solidarité qui y règne, sans oublier les échanges scientifiques de qualité. Merci, Jean-Marc, pour ton aide au quotidien et ta gentillesse légendaire ; merci, CC, pour m'avoir montré ton esprit de convivialité ;-) et surtout pour cette attention permanente dont tu fais preuve envers tout le monde, ainsi que pour l'ensemble des conseils avisés que tu as pu me dispenser. Merci, Fabrice, pour ton aide informatique quotidienne et François, pour ta disponibilité et ton aide à réparer des boulettes improbables (je pense à la création d'une piscine intérieure, un cadenas cassé, une voiture sans batterie...). Merci à Fabrice B, Christine S., Agnès, Hélène et Danièle, merci de m'avoir pardonné le carelage bleu. Bien-sûr, je n'oublie pas Kevin et sa ténacité face aux problèmes administratifs.

Mes plus sincères remerciements vont à l'ensemble des étudiants, postdocs et contractuels, pour nos échanges et « disputes » scientifiques, nos blagues plus ou moins drôles, et surtout cette bonne humeur perpétuelle qui fait que l'on ne regrette jamais de se lever pour venir au laboratoire.

Je remercie très chaleureusement Bertaccini et Plumel pour tous ces week-ends et soirées passés à s'énerver devant ces machines « qui ne veulent jamais marcher », pour nos « engueulades » légendaires (les murs du LSMBO en frémissent encore), et surtout pour tous les moments de franches rigolades (*new generation of tips, hit your head here...*). Je n'oublie pas Luc, mon binôme au quotidien : merci pour ta patience et ta gentillesse. Je ne garderai que de bons souvenirs de la semaine passée à préparer les échantillons de Protell et de toutes nos optimisations « gélistiques ». Merci à mon premier binôme, Jean-Michel, pour tous ses précieux conseils. Merci à Seb, pour nous avoir initié au monde de la SRM dans la bonne humeur. Enfin, merci à Gilles, ou plutôt Bullgilles, pour ces bons moments passés à explorer les mystères du G2 en bronzant aux UV.

Je n'oublie pas bien sûr les anciens ou faux anciens – Cyril, Nico, Daniel, Thierry, Stéphanie, Élixa, Quentin, Nicolas, Shaker, Vanessa, Séverine, Cédric, Magali (un réservoir de blagues impressionnant ☺) –, ni les nouveaux – Charlotte, Nina, Gauthier, Margaux –, ni le club VIP des non-covalentistes – Guillaume (now fluent in english), Johann (réfèrent qualité), François (professionnel

boulettes en tout genre) et Julien –, ni le bureau des précieux avec Alisson et Georges. Je remercie le R2 pour son accueil chaleureux : Alex 2, Patrick, Véronique, Laurence, Justine et Benoît (ou plutôt « oa » pour prononcer à la mode Westermann). Un remerciement spécial pour Alex 1 : merci pour ta disponibilité et la grande aide que tu m'as apportée tout au long de ma thèse.

Merci, Leslie, pour ton efficacité dans ton stage, ce fut un plaisir de travailler avec toi. Bonne chance pour la suite.

Je remercie également très chaleureusement Guillaume Béchade, futur collègue, pour tout le soutien qu'il m'a apporté lors de ma candidature chez Waters.

Merci à Amandine, avec qui tout a démarré par un sketch des Inconnus au cours d'une soirée obscure... comme quoi, ça scelle une amitié ☺. Merci pour toute ton aide lors de la rédaction, la préparation de l'oral et même jusqu'au pot !

Un grand merci à mes coloc's d'enfer, Anna et Mathieu, pour ne pas avoir râlé malgré mes horaires improbables et pour toutes leurs petites attentions du quotidien. Je vous attends à Manchester pour la revanche.

Merci à l'ensemble des membres du Bagad Kiz Avel sans qui la vie strasbourgeoise aurait été beaucoup moins fun et beaucoup moins bretonne.

Je tiens à remercier du fond du cœur ma famille, mes parents, mon grand-père, ma grand-mère, ma sœur et mon frère pour leur soutien quotidien et sans failles.

Enfin, j'adresse ma plus profonde gratitude à Gwenael, qui m'a supportée et aidée pendant ces longs mois.

Plan détaillé

Abréviations principales.....	13
Introduction générale	15
Partie I – État de l’art : l’analyse protéomique pour la recherche de biomarqueurs	17
CHAPITRE 1 – L’utilisation de stratégies « omics » pour la recherche de biomarqueurs	19
A. Généralités sur les cancers	20
A.1. Quelques chiffres	20
A.2. Caractéristiques et définitions	22
A.3. Origine de la cancérogenèse	23
A.4. Les traitements actuels	24
A.5. Et maintenant ?	25
B. La recherche de biomarqueurs	26
B.1. Définition.....	26
B.2. Quelles techniques pour quels biomarqueurs ?	27
CHAPITRE 2 – Les outils de la protéomique pour la recherche de biomarqueurs	31
A. La séparation des protéines et des peptides.....	32
A.1. La séparation des protéines.....	32
A.2. La séparation des peptides	33
A.3. Conclusion	35
B. L’analyse par spectrométrie de masse.....	35
B.1. Théorie et instrumentation.....	35
B.2. La spectrométrie de masse en tandem.....	42
B.3. Comparaison des performances des différents couplages utilisés.....	47
C. Identification et validation des spectres de fragmentation.....	48
C.1. Les banques de données.....	48
C.2. Les recherches dans les banques de données	49

C.3. La validation des données	49
D. Quantification.....	50
D.1. Les approches sans marquage	51
D.2. Les approches avec marquage.....	51
D.3. Conclusion.....	52

Partie II – Développement de méthodes d’analyse protéomique adaptées à la recherche de biomarqueurs 53

CHAPITRE 1 Préparation des échantillons et enrichissement en protéines membranaires 57

A. Le choix des protéines membranaires	57
A.1. L’intérêt de travailler sur un sous-protéome.....	57
A.2. Le choix des protéines membranaires	58
B. Les microparticules.....	59
B.1. Les microparticules induites.....	59
B.2. Les microparticules circulantes	60
B.3. Préparation de « <i>ghosts</i> membranaires ».....	66
C. Conclusions.....	67

CHAPITRE 2 – Développement de méthodes nanoLC-MS/MS pour l’analyse de protéines membranaires 69

A. De l’importance d’une optimisation fine des paramètres LC-MS/MS	70
A.1. Choix d’un gradient adapté aux échantillons.....	70
A.2. Paramètres MS et MS/MS.....	73
A.3. Conclusion	79
B. Approfondissement des analyses par la répétition des injections.....	80
B.1. Réplicats d’injections classiques	80
B.2. Réplicats d’injection avec liste d’exclusion	84
B.3. Conclusions.....	85
C. Utilisation d’un nouveau mode d’acquisition : la DIA.....	85
C.1. Description des techniques LC-MS ^E et LC-HDMS ^E	86
C.2. Interprétation des données via PLGS TM	89
C.3. Intérêt de l’utilisation combinée du mode DIA et de la mobilité ionique.....	90
C.4. Conclusions.....	95

D. Conclusion générale	95
CHAPITRE 3 – Recherche dans les banques de données	97
A. Optimisation fine des paramètres du moteur de recherche	98
A.1. Erreur sur la masse du précurseur	98
A.2. Modifications post-traductionnelles.....	101
A.3. Étude de la spécificité de l'enzyme	102
A.4. Évaluation de l'impact des changements de paramètres.....	102
B. Intérêt de l'utilisation de deux moteurs de recherche	103
B.1. Optimisation des résultats	103
B.2. Une meilleure compréhension des paramètres.....	104
C. Conclusions.....	106
CHAPITRE 4 – Quantification sans marquage pour la recherche de biomarqueurs	107
A. Quantification MS2 : le <i>spectral count</i>	108
A.1. Répétabilité des analyses	109
A.2. Quantification de la BSA.....	110
B. Quantification MS1 : l'extraction des courants d'ions.....	112
B.1. Linéarité de réponse de l'appareil.....	115
B.2. Répétabilité de la quantification	117
B.3. Paramètres influençant la quantification.....	118
C. Conclusions.....	121
Partie III – Résultats : application à trois projets de recherche de biomarqueurs.....	123
CHAPITRE 1 – Recherche de biomarqueurs à visée diagnostique pour améliorer la caractérisation des lymphomes B.....	127
A. La nécessité de proposer de nouveaux biomarqueurs candidats des différents types de lymphomes des cellules B.....	127
A.1. Les difficultés du classement de ces pathologies.....	128
A.2. Le cas des lymphomes de la zone marginale (MZL)	128
B. Mise en place d'une stratégie de protéomique globale	129
B.1. Les échantillons et la stratégie mise en œuvre	129

B.2. Les développements méthodologiques	130
B.3. Résultats	131
B.4. Conclusion	134
B.5. Communication des résultats.....	134
C. Développement d'une méthode de protéomique ciblée pour la validation des résultats.....	135
C.1. Quantification protéomique ciblée par SRM	136
C.2. Préparation des échantillons.....	140
C.3. Développement de la méthode LC-SRM	143
C.4. Conclusions.....	149
D. Conclusion générale	150
CHAPITRE 2 – Recherche de biomarqueurs de cellules souches cancéreuses issues de glioblastomes	153
A. Les cellules souches cancéreuses : nouvel axe de la recherche de biomarqueurs	153
B. Développement de la stratégie analytique	155
B.1. Échantillons.....	155
B.2. Stratégie mise en œuvre	156
B.3. Développements méthodologiques.....	157
C. À la recherche des protéines RCPG	162
C.1. Objectifs	162
C.2. Développements méthodologiques.....	163
C.3. Résultats obtenus.....	168
C.4. Conclusion	170
C.5. Communication des résultats.....	170
D. Analyse différentielle des clusters de différenciation	171
D.1. Objectifs	171
D.2. Résultats obtenus.....	171
D.3. Conclusion.....	175
D.4. Communication des résultats	175
E. Conclusions et perspectives	175

CHAPITRE 3 – Recherche de biomarqueurs prédictifs des rechutes du système nerveux central.....	177
A. De la nécessité de proposer des biomarqueurs prédictifs d’une rechute cérébrale dans les cas de lymphomes diffus à grandes cellules B	177
B. Les échantillons	179
C. Stratégie analytique mise en place	180
C.1. Les microparticules circulantes	180
C.2. Évaluation du nouvel instrument Q-TOF	181
C.3. Évaluation du degré de séparation nécessaire pour l’analyse des MPCs.....	181
C.4. Stratégie protéomique mise en place	184
D. Mise en place d’indicateurs de suivi	185
D.1. Contrôle qualité 1 : le digestat de BSA.....	186
D.2. Contrôle qualité 2 : mélange d’échantillons.....	190
D.3. Conclusions	193
E. Résultats préliminaires et analyses statistiques	194
F. Conclusion.....	197
Conclusion générale.....	199
Références bibliographiques.....	203
Partie expérimentale	i
Annexe 1 – Liste des clusters de différenciation	vii
Annexe 2 – Communication par affiche	xi
Annexe 3 – Publication dans le journal <i>Leukemia</i>	xiii
Annexe 4 – Liste des transitions sélectionnées pour le suivi des protéines CD45, CD148, CD5, CD200, CD200_1, CD200_3, CD180 et CD23.....	xv
Annexe 5 – Communication par affiche	xxv
Annexe 6 – Publication dans le journal <i>Plos One</i>	xxvii
Annexe 7 – Communication par affiche	xxix

Abréviations principales

ACN	Acétonitrile	Q	Quadripôle
AQUA	Absolute quantification	QQQ	Triple quadripôle
BPC	Base peak chromatogram	RCPG	Receptor coupled to the protein G
BSA	Albumine de sérum bovin	tr	Temps de rétention
CID	Collision induced dissociation	SILAC	Stable isotope labelling with amino acids in cell culture
CD	Cluster de différenciation	SRM	Selected reaction monitoring
CSC	Cancer stem cell	SSRCalc	Sequence specific retention calculator
DDA	Data dependant acquisition	SDS	Sodium dodecyl sulfate
DIA	Data independant acquisition	SRIG	Stacked ring ion guides
DIGE	Difference gel electrophoresis		
ESI	Ionisation electrospray		
FDR	False discovery rate		
FWMH	Full width at half maximum		
GSC	Glioblastoma stem cell		
HA	Human astrocytes		
HDMS^E	High definition expression mode mass spectrometry		
HPLC	High performance liquid chromatography		
IAA	Iodoacétamide		
ICAT	Isotope-coded affinity tag		
IEF	Isoelectrofocalisation		
IT	Ion trap		
iTRAQ	Isobaric tag for relative and absolute quantitation		
LC	Liquid chromatography		
LDGCB	Lymphome diffus à grandes cellules B		
LNH	Lymphome non hodgkinien		
LPL	Lymphome lymphoplasmocytaire		
MALDI	Matrix assisted laser desorption ionisation		
MCL	Lymphome du manteau		
MPI	Microparticule induite		
MPC	Microparticule circulante		
MS/MS	Spectrométrie de masse en tandem		
m/z	Rapport masse sur charge		
MZL	Lymphome de la zone marginale		
NCBI	National Center for Biotechnology information		
NSC	Neural stem cell		
OMSSA	Open Mass Spectrometry Search Algorithm		
PAGE	Polyacrylamide gel electrophoresis		
PBS	Phosphate buffer saline		



Introduction générale

Le protéome correspond à l'ensemble des protéines présentes à un moment donné dans des conditions données au sein d'un échantillon biologique particulier (tissue, fluide). L'analyse protéomique a pour but de caractériser ce protéome. Depuis le séquençage du génome humain, au début des années 2000, et l'évolution constante de l'instrumentation et des outils de traitement des données, l'analyse protéomique par spectrométrie de masse a connu un essor considérable. Il est, entre autres, désormais possible d'établir le protéome quasi complet d'un échantillon, de caractériser des modifications post-traductionnelles, de quantifier de manière absolue des protéines spécifiques ou encore d'étudier le réseau d'interactions entre les différentes protéines.

Très rapidement, la protéomique clinique a été considérée comme une technique à fort potentiel pour la recherche de biomarqueurs de différentes pathologies, c'est-à-dire « la détection de molécules mesurées et évaluées comme indicatrices d'un processus biologique normal, pathogène ou d'une réponse pharmacologique à une intervention thérapeutique » (définition du National Institute of Health, NIH). De nombreuses études ont été financées et réalisées afin de proposer de nouveaux biomarqueurs de différents types de pathologies par analyse protéomique. Cependant, malgré les financements et les 7840 publications issues de *Pubmed* rassemblant les deux mots clés *proteomics* et *biomarker*, la quête de biomarqueurs reste sur un constat d'échec. Selon E.P. Diamandis, aucun marqueur protéique pertinent du cancer n'a été validé pour une utilisation clinique depuis vingt-cinq ans. Plus grave, plusieurs études protéomiques publiées dans des journaux à *impact factor* élevé ont en fait révélé de « faux » biomarqueurs.

Ces dix dernières années, la communauté protéomique, consciente du problème, a œuvré pour comprendre l'origine des biais pré-analytiques (collecte des échantillons), analytiques (méthode sensible et reproductible) et post-analytiques (interprétation des données), ainsi que pour standardiser et homogénéiser les résultats afin d'accorder plus de confiance dans les identifications et propositions de biomarqueurs. Trois phases sont désormais distinguées :

1. la protéomique globale, qui propose de potentiels biomarqueurs candidats les plus pertinents possible (jusqu'à une centaine) à partir de l'analyse complète du protéome d'une première cohorte ;
2. la protéomique ciblée, qui vérifie la spécificité et l'expression différentielle de cette centaine de marqueurs sur une deuxième cohorte de patients ;
3. une validation par des techniques immunologiques en clinique sur plus de 1000 patients.

Les objectifs de mon travail de thèse s'articulent dans ce contexte autour de la protéomique globale par le développement de nouvelles méthodologies pour s'affranchir au maximum des biais inhérents à l'analyse protéomique dans le cadre de la recherche de potentiels biomarqueurs de différentes pathologies.

La première partie est organisée en deux chapitres :

- Le premier dresse un état des lieux de la recherche contre le cancer et montre la nécessité de la recherche de biomarqueurs. Ce chapitre conclut sur l'intérêt de l'utilisation de l'analyse protéomique tout en mettant en avant ses faiblesses.
- Le second décrit les technologies existantes en analyse protéomique pour la recherche de biomarqueurs.

La deuxième partie décrit l'ensemble des développements méthodologiques réalisés afin d'obtenir le protéome d'un échantillon le plus complet et le plus fiable possible. Pour ce faire, nous avons cherché à répondre à quatre questions principales, organisées en chapitres :

- Comment s'affranchir de la gamme dynamique intrinsèque à chaque échantillon ?
- Quels sont les paramètres importants du couplage chromatographie – spectrométrie de masse à considérer pour développer au mieux les méthodes ?
- Comment utiliser les outils bio-informatiques de manière optimale ?
- Peut-on quantifier les protéines contenues dans un échantillon ?

La troisième partie est consacrée aux résultats de l'application des développements méthodologiques sur trois projets différents :

- La recherche de biomarqueurs à visée diagnostique dans le but d'améliorer la caractérisation des lymphomes des cellules B, et plus particulièrement des lymphomes de la zone marginale.
- L'étude des cellules souches cancéreuses de glioblastomes afin d'en proposer des biomarqueurs potentiels.
- La recherche de biomarqueurs prédictifs des rechutes du système nerveux central dans le cas de lymphomes diffus à grandes cellules B.

Partie I

État de l'art : l'analyse protéomique pour la recherche de biomarqueurs

CHAPITRE 1

L'utilisation de stratégies « omics » pour la recherche de biomarqueurs

En 1971, le président des États-Unis Richard Nixon lançait la première offensive contre le cancer en signant le National Cancer Act (Project Law, 92-218). L'enjeu était de trouver, sous dix ans, des solutions pour guérir les cancers. Quarante ans plus tard, la communauté scientifique reste sur un constat d'échec et cela malgré les extraordinaires avancées de la biologie et l'investissement massif dans la recherche. Depuis 1971, le National Cancer Institute des États-Unis a dépensé plus de 90 milliards de dollars par an dans la recherche, les traitements et les soins hospitaliers. Les résultats en sont mitigés. D'un côté les pessimistes pensent que la guerre est perdue^{1,2} ; d'un autre les optimistes voient de très grandes avancées dans la diminution du taux de mortalité et les progrès de la science^{3,4}. Il n'en reste pas moins que les diagnostics se sont précisés et les thérapies largement améliorées.



Figure 1

R. Nixon signant le premier plan anti-cancer :
"I will also ask for an appropriation of an extra \$100 million to launch an intensive campaign to find a cure for cancer, and I will ask later for whatever additional funds can effectively be used. The time has come in America when the same kind of concentrated effort that split the atom and took man to the moon should be turned toward conquering this dread disease. Let us make a total national commitment to achieve this goal."

Au cours de ce chapitre, nous montrerons la nécessité de la recherche de nouveaux biomarqueurs pour les cancers afin d'améliorer le diagnostic et les traitements actuels. Il s'agira d'exposer l'intérêt des méthodes globales à haut débit, appelées « omics », et d'en discuter les points faibles.

A. Généralités sur les cancers

A.1. Quelques chiffres

Après l'éradication d'une grande partie des maladies infectieuses dans les pays occidentaux, le cancer apparaît aujourd'hui comme un problème majeur de santé publique. Ainsi, en 2012, 14 millions de personnes étaient atteintes d'un cancer et 8,2 millions sont décédées de cette maladie dans le monde⁵.

En France, pour la période 2004-2008, le cancer était la première cause de mortalité chez l'homme (33 % des décès) et la deuxième chez la femme (24 % des décès). Les derniers chiffres publiés par l'Institut National du Cancer (INCa) datent de 2011 et font état de 365 500 nouveaux cas de cancer (207 000 chez l'homme et 158 500 chez la femme) et 147 500 décès (84 500 chez l'homme et 63 000 chez la femme). Comme observé sur les figure 2 et figure 3, tous cancers confondus, la mortalité diminue. Cette tendance positive s'explique par l'optimisation des traitements ainsi que par l'amélioration des diagnostics, qui sont de plus en plus précoces, permettant ainsi une meilleure prise en charge des patients (diminution de la mortalité et augmentation de l'incidence).

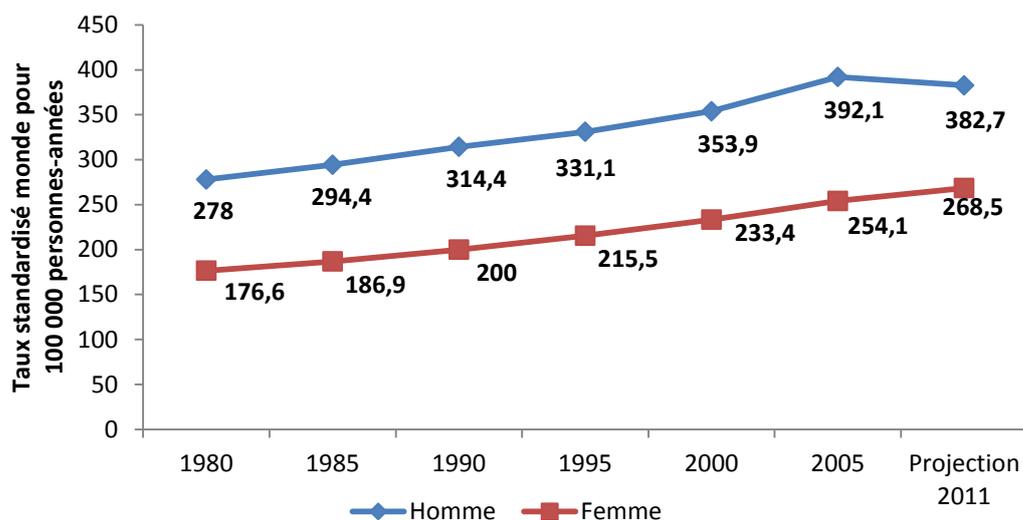


Figure 2

Évolution de l'incidence des cancers de 1980 à 2005 selon le sexe en France (taux standardisé monde estimé, projection pour l'année 2011)

Sources : période 1980 à 1985 [Belot A, 2008 ; période 1990 à 2011 [HCL/InVS/INCa/Francim/Inserm, 2011]

Traitement : INCa 2011

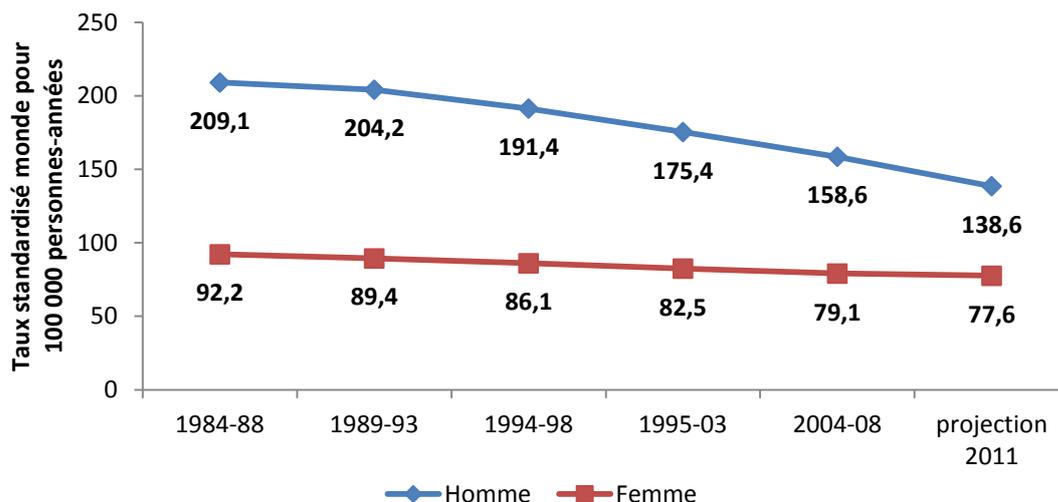


Figure 3
Évolution de la mortalité observée par cancer de 1984-1988 à 2004-2008 selon le sexe en France (taux standardisé monde, projections pour l'année 2011)

Sources : [InVS/Inserm, 2011] [HCL/InVS/INCa/Francim/Inserm, 2011]

Traitement : INCa 2011

Les cancers les plus fréquents sont ceux de la prostate chez l'homme avec 71 000 cas incidents en 2011 et ceux du sein chez la femme avec 53 000 cas incidents (cf. figure 4). Chez l'homme, la majorité des décès sont imputables au cancer du poumon (21 000 décès) et chez la femme, au cancer du sein (11 500 décès).

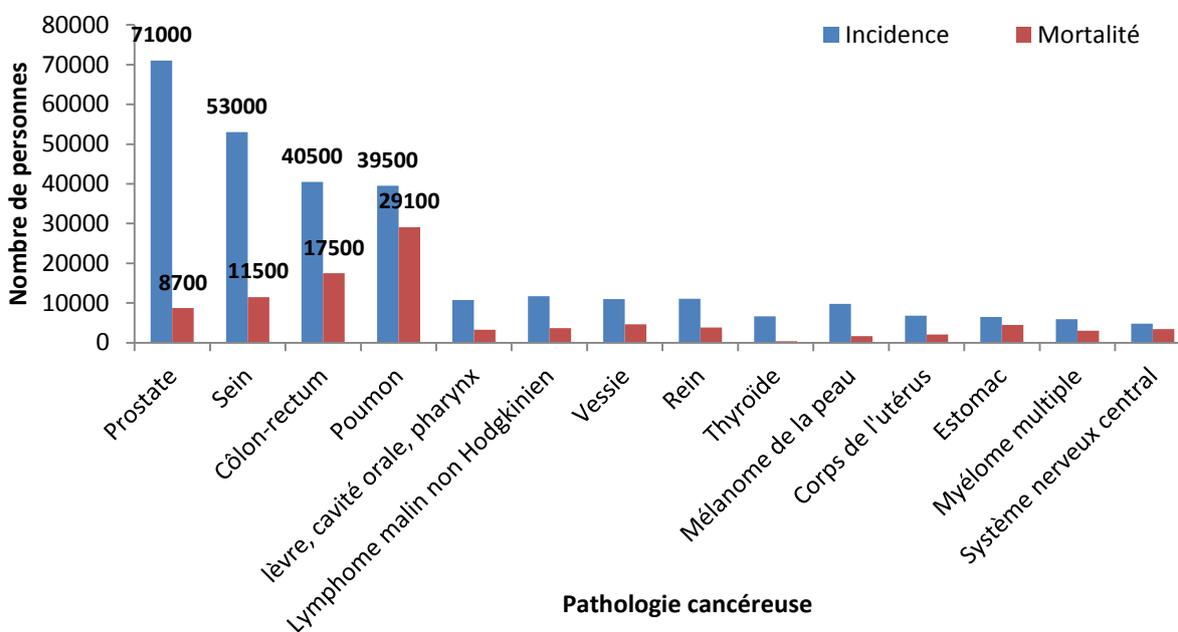


Figure 4
Classement des cancers par incidence et mortalité, tous sexes confondus, en France, 2011

Source : Projection de l'incidence et de la mortalité par cancer en France en 2010 (Hospices civils de Lyon, InVS, INCa, Francim, Inserm).

Traitement : INCa 2011

A.2. Caractéristiques et définitions

Le cancer est une maladie qui a toujours existé, masquée par des fléaux plus grands. Les premières descriptions remontent à 3000 av. J.-C., dans le papyrus d'Edwin Smith. Huit cas de tumeurs du sein y sont présentés. Hippocrate, médecin grec (400-377 av. J.-C.), lui donna le nom de cancer (*Karkinos* en grec, analogie avec la forme des tumeurs lors de leur extension).

La maladie est caractérisée par une dérégulation du fonctionnement de certaines cellules de l'organisme. Celles-ci se multiplient de façon anarchique, d'abord localement, puis se diffusent, formant ainsi des métastases.

Le cancer regroupe un nombre important de maladies — plus d'une centaine — présentant six caractéristiques communes selon un article paru dans *Cells* en 2000, qui faisait un état des recherches effectuées les vingt-cinq années précédentes⁶.

La figure 5, extraite de cette publication, présente les six caractéristiques principales des cancers, qui leur permettent de proliférer et de se disséminer :

- **Autosuffisance en signaux de croissance** caractérisée par une altération des signaux extracellulaires (capacité à autosynthétiser ces signaux auxquels les cellules répondent), par une altération des récepteurs (surexpression des récepteurs aux facteurs de croissance) et par une altération des voies de transduction intracellulaire.
- **Insensibilité aux signaux inhibiteurs de la croissance** : par exemple, la voie de signalisation TGF- β , qui a un effet antiprolifératif, peut dans certains cas être corrompue.
- **Résistance à l'apoptose** : par exemple par surexpression de facteurs antiapoptotique (protéine BCL2), ou encore par perte du régulateur proapoptotique p53.
- **Capacité répllicative infinie** : permet aux cellules cancéreuses de générer des tumeurs macroscopiques par le maintien des télomères, qui protègent les extrémités des chromosomes et évitent ainsi la sénescence des cellules.
- **Induction de l'angiogenèse** : les cellules cancéreuses, comme les cellules saines, ont besoin de nutriments, d'oxygène, ainsi que d'une évacuation des déchets métaboliques et du CO₂. Dans le cas des cellules saines, durant l'embryogenèse, de nouvelles cellules endothéliales naissent et sont assemblées en vaisseaux (vasculogénèse). Puis l'angiogenèse permet de former de nouveaux vaisseaux à partir de vaisseaux préexistants. Le processus de vascularisation devient alors quiescent. Au contraire, dans le cas de cellules cancéreuses, l'angiogenèse reste activée, permettant la croissance des tumeurs.
- **Activation de l'invasion et du processus de métastase**, à l'origine de 90 % des décès par cancer.

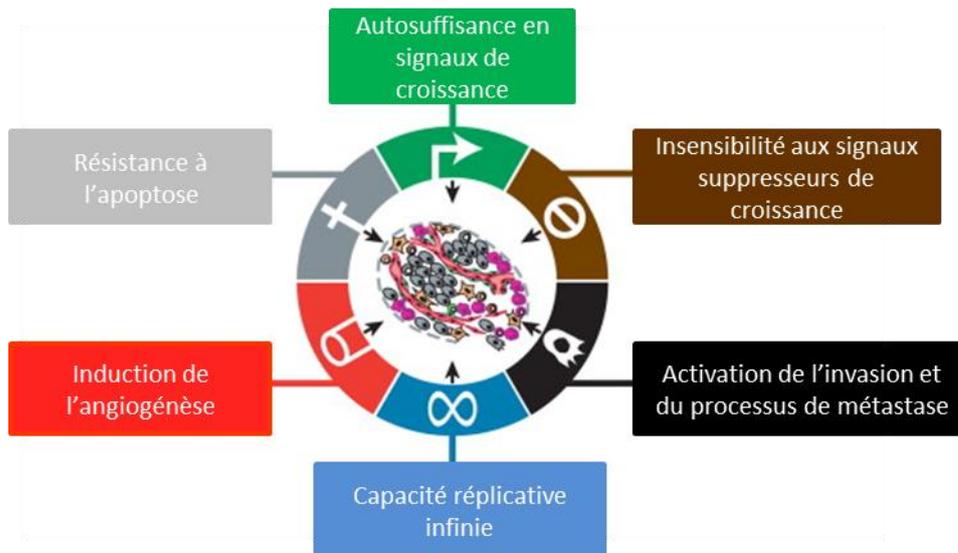


Figure 5

Capacités acquises par l'ensemble des cellules cancéreuses : « Hallmarks of Cancer »

Source : adaptée de Hanahan and al⁶.

En 2011, deux nouvelles caractéristiques sont venues s'ajouter aux six précédentes⁷ :

- La capacité à éviter la destruction par les cellules du système immunitaire.
- Des reprogrammations majeures du métabolisme énergétique des cellules pour leur permettre une croissance et une prolifération continues.

Théoriquement, tous les cancers doivent présenter les huit mêmes caractéristiques. Cependant, l'ordre d'apparition reste variable selon les pathologies.

A.3. Origines de la cancérogenèse

La cancérogenèse a pour origine des mutations qui se produisent à la suite d'erreurs lors de la réparation de l'ADN. Normalement, des complexes enzymatiques sont chargés de réparer ces erreurs (*protein DNA repair*). Cependant, si le nombre d'erreurs est trop important ou si les lésions touchent les deux brins de l'ADN, le risque de mutation est plus élevé.

Pour la plupart des cancers, les causes des mutations restent inconnues. Cependant, certains facteurs de risque ont pu être déterminés⁸⁻¹⁴ :

- les agressions chimiques par des substances telles que le tabac ou l'amiante ;
- les rayonnements solaires ou nucléaires ;
- le vieillissement entraîné par un cumul des agressions extérieures et un moins bon fonctionnement du mécanisme de réparation des cellules ;
- les infections virales, tels le papillomavirus humain, qui peut favoriser l'apparition du cancer du col de l'utérus, ou le virus de l'hépatite C, qui peut entraîner un cancer du foie ;
- la susceptibilité génétique ;

- un problème au niveau de la réplication de l'ADN.

Ces facteurs de risque ont plusieurs types de conséquences :

- des mutations ;
- des altérations des chromosomes telles que des délétions (disparition d'un fragment de chromosome) ;
- des translocations (passage d'un fragment de chromosome sur une autre région de ce même chromosome ou sur un autre chromosome) ;
- des événements épigénétiques qui n'altèrent pas la séquence de base de l'ADN et ne sont pas irréversibles (par exemple, les méthylations de certains dinucléotides).

A.4. Les traitements actuels

Les traitements actuels reposent sur trois stratégies.

La première, la chirurgie, consiste à enlever l'intégralité de la tumeur. L'exérèse est assez large afin d'éviter que la tumeur ne renaisse^{9,15}.

La deuxième, la radiothérapie, constitue 50 % des traitements. Les rayonnements (photons ou électrons de haute énergie, 4 à 25 MeV et rayons gamma du cobalt 60) bloquent la capacité des cellules à se multiplier. La dose doit être bien adaptée afin d'éviter de détruire trop de cellules saines, mais aussi de laisser trop de cellules cancéreuses qui pourront donner naissance à de nouvelles tumeurs¹⁶⁻¹⁹.

Enfin, les traitements médicaux sont basés sur l'utilisation de molécules chimiques. La première chimiothérapie a été développée à partir du gaz moutarde (sulfure de 2,2' -dichlorodiéthyle). En effet, durant la seconde guerre mondiale, il a été observé que les soldats exposés à ce gaz présentaient moins de globules blancs dans le sang ainsi qu'une aplasie (arrêt du développement de la moelle osseuse). L'hypothèse a été émise que ce gaz pourrait empêcher la prolifération de cellules cancéreuses. Les premiers essais cliniques menés en 1943 sur des patients atteints de lymphomes ont montré une rémission brève²⁰. Par la suite, de nombreuses molécules ont été développées et synthétisées. L'inconvénient majeur de cette technique reste que les molécules sont cytotoxiques et ne s'attaquent donc pas spécifiquement aux cellules cancéreuses²¹.

Jusqu'au début des années 2000, le traitement classique des cancers reposait sur une ablation chirurgicale de la tumeur suivie d'une radiothérapie ou de l'administration de divers composés cytotoxiques.

Depuis dix-sept ans émerge un nouveau concept, basé sur l'utilisation de molécules qui visent précisément les cellules cancéreuses^{11,13,22}. Ainsi, des anticorps thérapeutiques ciblent spécifiquement des antigènes tumoraux présents à la surface des cellules pathologiques. En 1997-1998, deux premiers anticorps thérapeutiques sont approuvés par la FDA (Food and Drug Administration). Il s'agit de Trastuzumab, qui cible les récepteurs HER2/neu, et de Rituximab, qui cible le biomarqueur CD20 présent à la surface des lymphocytes B. À ce jour, treize anticorps ont été adoptés par la FDA et sont

actuellement commercialisés (cf. tableau 1)²³. Les modes d'action sont divers. Par exemple, l'anticorps Rituximab induit l'apoptose de la cellule, Bevacizumab se lie à une protéine responsable de la croissance vasculaire (VEGF-A), la séquestre et altère ainsi la vascularisation de la tumeur, ou encore Ipilimumab bloque l'activité d'un régulateur des lymphocytes T (CTLA4) et occasionne ainsi une réponse immunitaire forte du patient contre les cellules cancéreuses²⁴.

Nom de l'anticorps	Cible	Cancers
Cetuximab	EGFR	Colon, cerveau, cou
Panitumumab	EGFR	Colon
Trastuzumab	HER2	Sein, gastrique
Pertuzumab	HER2	Sein
Ado trastuzumabemtansine <i>Anticorps conjugué à un médicament (ADC)</i>	HER2	Sein
Bevacizumab	VEGF	Colon, poumon, glioblastome, reins
Ipilimumab	CTLA4	Mélanome
Rituximab	CD20	Lymphome non Hodgkinien (LNH), Leucémie lymphoïde chronique (LLC)
Ofatumumab	CD20	LLC
⁹⁰ Y-labeled ibritumomabtiuxetan	CD20	Lymphome folliculaire bas grade, NHL
¹³¹ I-labeled tositumomab	CD20	Lymphome folliculaire bas grade, NHL
Alemtuzumab	CD52	LLC
Brentuximabvedotin(ADC)	CD30	Lymphome de Hodgkin

Tableau 1

Liste des anticorps monoclonaux approuvés par la FDA et commercialisés.

Source : Panowski et coll.²⁴

A.5. Et maintenant ?

Malheureusement, si de grandes avancées ont été réalisées depuis le lancement de la guerre contre le cancer par Richard Nixon, beaucoup de mécanismes du cancer restent inconnus. Par ailleurs, les thérapies envisagées ne permettent pas de guérir la majorité des cancers. Force est de constater que le remède miracle n'a pas encore été trouvé même si, à ses débuts, la thérapie ciblée a donné lieu à de grands espoirs. Ainsi, en 2007, des chercheurs de Roche et de Plexikon ont lancé les tests d'un nouveau médicament, PLX4032, sur des patients atteints d'un mélanome. Les résultats étaient incroyables : dans certains cas, la tumeur avait complètement disparu. Cependant, sept mois plus tard, la tumeur avait repris sa croissance chez la plupart des patients. L'une des principales hypothèses expliquant cette résistance est basée sur la grande hétérogénéité des tumeurs²⁵⁻²⁷. Ainsi, la thérapie ciblée éliminerait la majeure partie de la masse tumorale, permettant une rémission du patient.

Cependant, cette réduction laisse de la place pour l'expansion d'un clone minoritaire résistant au traitement initial. Dans cette optique, une nouvelle hypothèse récemment élaborée porte sur l'existence de cellules souches cancéreuses impliquées dans la résurgence des tumeurs²⁸ (ce point sera évoqué de manière plus approfondie dans le chapitre 2 de la Partie III).

Aujourd'hui, trois axes sont développés pour combattre les cancers :

- 1) la prévention, notamment sur l'importance d'une bonne hygiène de vie (nocivité du tabac, alcool, UV, surpoids) ;
- 2) le diagnostic précoce (par la mise en place de dépistages) pour améliorer la prise en charge des patients ;
- 3) l'amélioration des traitements (combinaison de plusieurs thérapies pour cibler différents sous-clones et développement de thérapies personnalisées)^{26, 29, 30}.

Les points 2) et 3) impliquent une caractérisation plus précise des tumeurs et une meilleure compréhension des mécanismes. Une des pistes attractives serait donc de rechercher de nouveaux biomarqueurs des différents types de cellules cancéreuses.

B. La recherche de biomarqueurs

B.1. Définition

D'après les instituts américains de la santé (NIH), un biomarqueur, ou marqueur biologique, est une « caractéristique mesurée objectivement et évaluée comme un indicateur d'un processus biologique normal, pathologique ou d'une réponse pharmacologique à une intervention thérapeutique »³¹.

Les biomarqueurs peuvent être classés en quatre groupes différents. Les marqueurs pronostiques (diagnostiques) sont des marqueurs permettant de séparer une population suivant des caractéristiques évolutives propres. Les marqueurs prédictifs permettent d'estimer le succès du traitement mis en place. La plupart des biomarqueurs vont être à la fois pronostiques et prédictifs. Les biomarqueurs pharmacodynamiques reflètent l'effet d'une molécule sur les cellules hôtes (saines ou tumorales) et permettent de définir les doses de médicaments à utiliser. Les marqueurs de substitution sont des marqueurs permettant d'évaluer l'efficacité d'un traitement en l'absence de données cliniques. Quelques exemples sont présentés dans le tableau 2.

Biomarqueurs	Utilisation	Type de biomarqueur
HER2 (Human epidermal growth factor receptor 2)	Orienter et cibler le traitement	Prédictif
BCR-ABL	Orienter et cibler le traitement	Prédictif
Peptide amyloïde β	Diagnostic du stade prodromal de la maladie d'Alzheimer	Pronostique
Interleukin 6	Marqueur d'inflammation	Pharmacodynamique
Cholestérol LDL (Lipoprotéines de basse densité)	Utilisation dans les maladies coronariennes	Substitution

Tableau 2
Exemples de différents types de biomarqueurs utilisés en clinique³¹

B.2. Quelles techniques pour quels biomarqueurs ?

La carcinogenèse est caractérisée par une expression aberrante de divers gènes qui occasionne des différences d'expression de transcrits d'ARN et de protéines.

Les technologies à haut débit, appelées approches « omics », ont un grand potentiel pour la découverte de nouveaux biomarqueurs de cancers qui pourraient à la fois servir au diagnostic mais aussi orienter la recherche vers de nouvelles cibles thérapeutiques. Ces techniques se basent sur l'étude de chaque étape du processus de transmission de l'information génétique du génome au métabolome, en passant par le transcriptome et le protéome³²⁻³⁴. La figure 6 présente les différents niveaux de biomarqueurs.

La génomique³⁵ consiste à séquencer l'ADN contenu dans diverses cellules par l'analyse du génome entier ou de certaines régions des chromosomes. Ainsi, il est possible d'identifier les mutations somatiques et conductrices et de suivre l'évolution clonale des tumeurs.

La transcriptomique³⁶ consiste à étudier et mesurer les taux d'expression des ARNs messagers et comparer des conditions malades/sains pour identifier les expressions des gènes reliés à la pathologie. Par ailleurs, cette technique permet d'étudier les facteurs qui régulent l'expression des gènes et qui ne sont pas codés dans la séquence ADN (régulation des mRNA, événements épigénétiques).

La protéomique permet d'établir le protéome entier d'un échantillon. Ce terme, proposé pour la première fois par Wilkins en 1996, est une combinaison des noms « **protéines** » et « **génome** », désignant ainsi « l'ensemble des protéines présentes dans un échantillon ou dans un environnement défini ainsi que leur variation dans l'espace et dans le temps »^{37, 38}. Ceci implique aussi l'analyse complète des isoformes, des modifications post-traductionnelles et des interactions protéiques au sein des complexes.

La métabolomique³⁹ aide à comprendre un système biologique à travers l'analyse de ses composés naturels et de ses produits dérivés (substrats, produits, lipides, cofacteurs), générés par l'ensemble des voies métaboliques en réponse à un traitement, ou en comparaison de deux états.

Dans le cadre de la recherche de biomarqueurs, l'ensemble de ces techniques est basé sur une analyse différentielle, qui consiste à comparer différents échantillons représentatifs d'un état donné, par exemple sain/malade ou encore pathologie 1/pathologie 2.

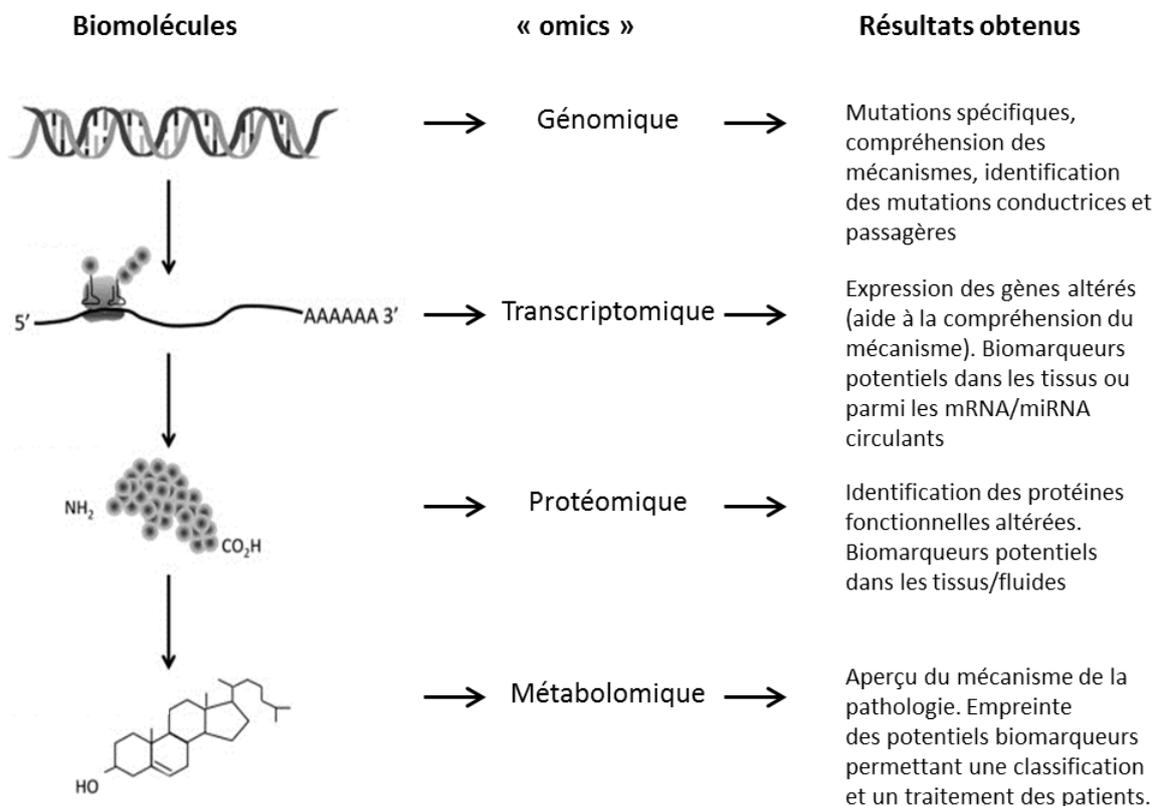


Figure 6
Les stratégies "omics" et leurs applications

Source : adapté de Wood and al⁴⁰

S'il existait un biomarqueur discriminant les sujets sains des sujets malades, les techniques « omics » devaient permettre de l'identifier et la biostatistique d'extraire l'information. La promesse était forte : détecter de nouvelles molécules pour être capable d'établir des diagnostics précoces et prédire les réponses aux thérapies.

Cependant, en dépit du caractère très attractif de ces techniques, les résultats sont décevants et le nombre de biomarqueurs ayant atteint le niveau clinique est faible comparé au nombre proposé par les stratégies « omics » : sur 238 530 publications dans *Pubmed* mentionnant les mots-clés « biomarkers » et « cancer », moins de 1 % des propositions ont pu être validées. Plus grave, plusieurs biomarqueurs proposés se sont avérés non pertinents. Plusieurs types d'erreurs dans le choix des biomarqueurs ont pu être mis en évidence⁴¹⁻⁴⁵ :

- Erreurs A : le marqueur déjà implémenté au niveau clinique se révèle inutile voire nocif. C'est le cas, par exemple, de l'antigène prostatique (PSA) utilisé jusqu'à présent pour dépister les cancers de la prostate. D'après un rapport récent⁴⁶, les cas de diagnostics faux positifs sont nombreux et occasionnent la mise en place de traitements inutiles avec des effets secondaires

importants. Cependant, ce constat est à nuancer : ce marqueur est aujourd'hui suivi de manière différentielle et permet ainsi de suivre l'évolution de la pathologie chez un patient.

- Erreurs B : les biomarqueurs détectés semblent très prometteurs mais s'avèrent finalement faux ou exagérés. Par exemple, en 2002, une étude publiée dans *The Lancet* décrivait une avancée considérable dans le diagnostic des cancers de l'ovaire⁴⁷. La protéomique, ici utilisée sur 50 échantillons malades et 66 échantillons contrôles, avait permis de montrer que la spectrométrie de masse pouvait être utilisée pour différencier les deux cohortes dans le cadre d'un diagnostic précoce, et donc pour améliorer la prise en charge du patient. Cette découverte a levé un enthousiasme considérable avec la commercialisation de ce test par trois grandes firmes. Cependant, une seconde analyse des données par deux biostatisticiens a montré que les différences observées entre les deux types d'échantillons étaient dues à des artefacts expérimentaux⁴⁸.
- Erreurs C : certains biomarqueurs pourtant très prometteurs ne sont jamais validés au niveau clinique.
- Erreurs D : un biomarqueur suscite peu ou pas d'intérêt mais est quand même largement promu pour une validation et une utilisation clinique. Ces marqueurs sont en général issus de la recherche de variants lors du séquençage de génomes de différentes pathologies⁴⁹.

Ces erreurs peuvent avoir plusieurs origines :

- De la découverte du biomarqueur à sa validation et son implémentation en clinique, le chemin est long (cf. figure 7)⁵⁰. En effet, le marqueur potentiel doit être découvert et validé sur une cohorte suffisamment importante puis évalué sur une nouvelle cohorte et, enfin, faire l'objet d'essais cliniques.



Figure 7

L'implémentation d'un nouveau biomarqueur en clinique est un chemin qui peut d'avérer long et périlleux.

Source : adaptée de Mischak and al⁵⁰

- Les techniques « omics » comportent des biais. La figure 8 présente quelques paramètres pouvant influencer le résultat obtenu^{31, 34, 51-57}.
 - Collecte et stockage des échantillons : des protocoles stricts sont aujourd'hui mis en place pour conserver des conditions comparables. Par ailleurs, des biobanques rassemblant plusieurs types de tissus tumoraux ont été mises en place.
 - Variabilité analytique, notamment dans le cas de la protéomique où des résultats différents peuvent être obtenus en fonction des critères de validation appliqués.
 - Mise en évidence de marqueurs non biologiquement pertinents : ainsi, différents types de marqueurs d'inflammation ou de stress ont été considérés comme de potentiels marqueurs de pathologies alors qu'ils sont présents dans tous types de pathologies.
 - Surajustement : le nombre de paramètres étudiés est largement supérieur au nombre d'échantillons. Par exemple, lors des analyses protéomiques, il n'est pas rare de proposer des biomarqueurs à partir de quelques échantillons, tout en étudiant plus de 1000 protéines. Cette pratique augmente le taux de faux positifs et, ainsi, le risque d'identifier un « faux » biomarqueur.

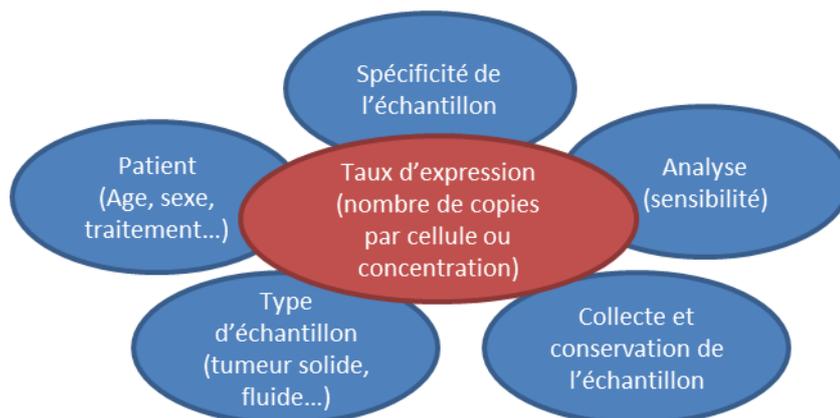


Figure 8

Paramètres pouvant influencer le résultat obtenu lors de l'utilisation d'une stratégie « omics »

Consciente du problème, la communauté « omics » a œuvré ces dix dernières années afin de comprendre l'origine de ces « faux » biomarqueurs^{33, 43, 58-63}. Par ailleurs, des solutions ont été proposées afin d'éviter certains pièges :

- standardisation de la collecte des échantillons (mise en place de biobanques) ;
- procédures standard validées pour la partie analytique ;
- étape de validation sur une deuxième cohorte de patients.

C'est dans ce contexte que s'inscrit la présente thèse : il s'agira de développer de nouvelles méthodologies en analyse protéomique globale afin de proposer des candidats biomarqueurs potentiels les plus pertinents possible.

CHAPITRE 2

Les outils de la protéomique pour la recherche de biomarqueurs

La recherche de biomarqueurs par analyse protéomique implique l'identification, la caractérisation et la quantification des protéines contenues dans un échantillon. Il existe deux types d'approches :

- La stratégie *top down*, qui consiste à chercher des informations sur la séquence à partir de la protéine entière.
- La stratégie *bottom up*, qui consiste à rechercher les informations de séquence à partir de la protéine digérée en peptides.

Au cours de ce chapitre, nous n'évoquerons que la stratégie utilisée dans le cadre de cette thèse, à savoir la méthode *bottom up*.

Nous nous attacherons ici à présenter les points clés de chaque étape de l'analyse protéomique (cf. figure 9).

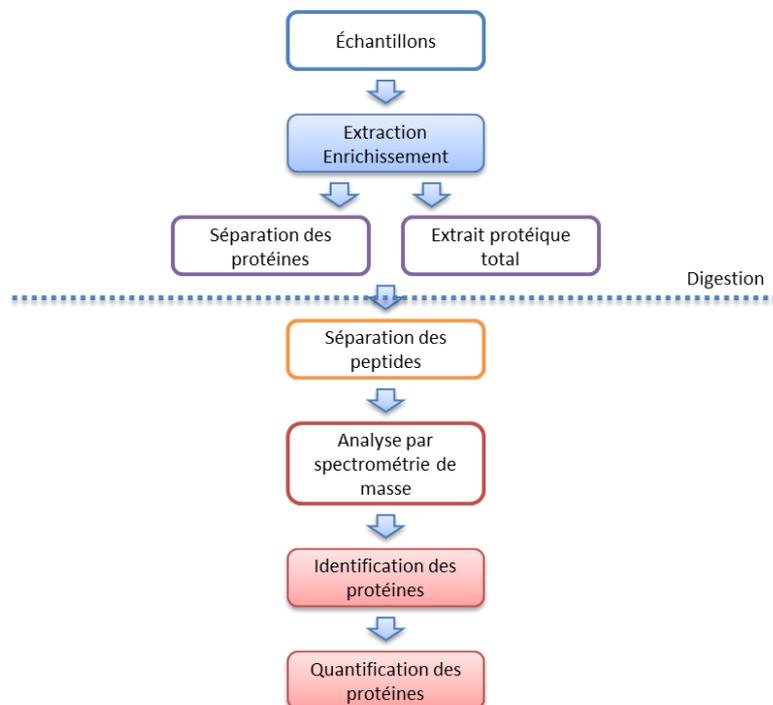


Figure 9

Schéma représentant les différentes étapes de l'analyse protéomique

A. La séparation des protéines et des peptides

Le protéome humain, sujet d'étude de cette thèse, contient environ 21 000 séquences protéiques qui peuvent posséder une ou plusieurs modifications post-traductionnelles, être clivées, issues de produits de dégradation ou d'épissages alternatifs. Les mélanges protéiques étudiés peuvent ainsi s'avérer très complexes et couvrir une large gamme dynamique, rendant l'analyse par spectrométrie de masse difficile. Par ailleurs, dans la majorité des cas, la recherche de biomarqueurs implique l'identification de protéines faiblement abondantes. Pour pallier ce problème, il est possible de simplifier les mélanges en amont par une séparation des protéines et/ou des peptides issus de la digestion trypsique.

A.1. La séparation des protéines

Il existe plusieurs techniques de séparation des protéines basées sur deux critères physico-chimiques majeurs :

- **Le pH** avec des techniques telles que l'IEF (isoélectrofocalisation), où les protéines migrent jusqu'à leur pI (point isoélectrique), et la chromatographie d'échange cationique.
- **La masse moléculaire** avec, par exemple, l'électrophorèse monodimensionnelle, la chromatographie par exclusion de taille (SEC), ou encore la chromatographie par affinité.

Au cours des travaux présentés dans ce manuscrit, la séparation des protéines a été exclusivement réalisée à l'aide d'un gel d'électrophorèse monodimensionnel SDS (sodium dodécylsulfate) PAGE (*polyacrylamide gel electrophoresis*).

L'électrophorèse est une technique analytique qui permet de séparer les molécules suivant leur charge et leur masse moléculaire sous l'effet d'un champ électrique. Dans notre cas, les protéines sont migrées sur un gel en polyacrylamide en condition dénaturante. Elles sont donc séparées uniquement suivant leur masse moléculaire. En effet, le tampon échantillon (Laemmli)⁶⁴ contient un détergent, le SDS, qui se fixe uniformément sur les protéines et leur impose une densité de charge négative. À noter qu'il est possible de réaliser une séparation en conditions natives (*native PAGE*, *Blue native*⁶⁵ ou *high resolution clear native PAGE*⁶⁶).

Le gel de polyacrylamide sert de tamis moléculaire. Un maillage serré permet de séparer les petites molécules et à l'inverse un maillage plus lâche permet de séparer les grandes molécules. La dimension des pores est contrôlée par la concentration en acrylamide et bis-acrylamide. De manière générale, le gel est constitué de deux parties⁶⁷ : un gel de concentration (4 à 5 % en acrylamide/bis-acrylamide) et un gel de séparation (8 à 20 % en acrylamide/bis-acrylamide).

La migration repose sur le principe de l'isotachophorèse⁶⁸ : un ion pilote (Cl⁻), qui migre rapidement et un ion lent (glycinate). Dans le gel de concentration, le pH est proche du pI du glycinate, ce qui conduit les protéines à se concentrer en une bande. Le pH étant plus élevé dans le gel de séparation, la migration est accélérée et les protéines sont séparées suivant leur masse moléculaire. Un exemple de séparation est donné en figure 10.

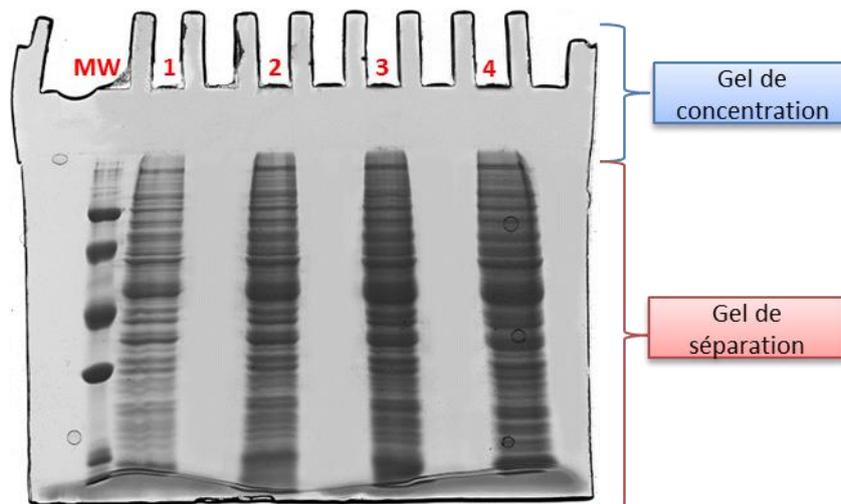


Figure 10

Exemple de séparation réalisée pour un lysat de protéines de levure

MW : marqueur de poids moléculaire ; piste 1 : 5 µg de lysat ; 2 : 10 µg de lysat ; 3 : 20 µg de lysat ; 4 : 50 µg de lysat.

Les protéines sont ensuite révélées par coloration au bleu de Coomassie^{69,70} (détection jusqu'à 100 ng) ou au nitrate d'argent^{71,72} (détection de l'ordre du nanogramme), plus sensible mais moins compatible avec la spectrométrie de masse.

Après décoloration, réduction, alkylation et digestion par une endoprotéase, les peptides sont extraits du gel par un mélange 40/60 v/v ACN/H₂O.

Cette technique de séparation présente un avantage majeur : l'utilisation du SDS facilite grandement la solubilisation des protéines. Cependant, elle reste une technique peu résolutive, qui ne produit qu'une séparation grossière des protéines.

Notons qu'il est possible de combiner deux dimensions et de réaliser un gel 2D SDS PAGE⁷³⁻⁷⁵. La première dimension (IEF) permet de séparer les protéines suivant leur pI puis la deuxième suivant leur masse moléculaire. Cette technique, très résolutive, permet de séparer jusqu'à 5000 protéines et de les quantifier par analyse d'image. Néanmoins, elle est peu adaptée à la recherche de biomarqueurs pour deux raisons principales⁷⁶ :

- elle n'est pas optimale pour la séparation de protéines membranaires du fait du tampon de solubilisation (le SDS est incompatible avec l'IEF) ;
- elle requiert un temps de manipulation important.

A.2. La séparation des peptides

La séparation des peptides joue un rôle important sur deux paramètres principaux de l'analyse protéomique :

- le nombre de peptides, et donc de protéines identifiées ;
- la qualité de la quantification.

La chromatographie HPLC (*high performance liquid chromatography*) est le système le plus répandu de par sa facilité de couplage avec la spectrométrie de masse et ses capacités résolutive⁷⁷.

La séparation est ici basée sur la différence d'affinité des analytes entre la phase stationnaire et la phase mobile. Plusieurs paramètres clés^{78, 79} interviennent :

- Le type de la phase
- La longueur de la colonne
- Le diamètre de la colonne
- La taille des particules

Plusieurs types de phases, basés sur des modes d'interaction différents, peuvent être utilisés : IEC (chromatographie par échange d'ions), SEC (chromatographie d'exclusion stérique), HILIC (chromatographie d'interaction hydrophile) et RP (phase inverse). Cette dernière est la technique la plus courante et aussi celle qui a été utilisée tout au long de ce travail de thèse. Les composés sont séparés suivant leur affinité avec la phase stationnaire apolaire (silice greffée avec des chaînes de carbone, C18) et la phase mobile, plus polaire, constituée d'un mélange d'eau et d'acétonitrile.

Au cours de ces travaux, trois types de systèmes ont été utilisés (cf. tableau 3) :

- UPLCTM : Ultra performance LCTM (Waters Corporation)⁸⁰
- NanoLC-Chip : nanoHPLC-ChipTM (Agilent technologies)^{81, 82}
- MicroLC : Ultimate 3000[®] (Thermo Fisher Scientific)

Système	UPLC TM	nanoLC-Chip	MicroLC
Phase stationnaire	C18	C18	C18
Longueur de la colonne (mm)	200 à 250	150	150
Diamètre de la colonne (µm)	75	75	300
Taille des particules (µm)	1,7	5	3,5
Débit (µL/min)	0,3 à 0,45	0,3	5

Tableau 3

Principales caractéristiques des différents systèmes de chromatographie utilisés

Chacun de ces modes de séparation présente des avantages et des inconvénients. Ainsi, si le système UPLC offre la meilleure qualité de séparation en termes de sensibilité et de résolution, la capacité de charge reste limitée. La miniaturisation du système nanoLC-Chip permet de diminuer les volumes morts, et donc le temps d'analyse. Néanmoins, la capacité de charge est limitée et la séparation reste faiblement résolue. Enfin, la microLC permet d'injecter une quantité de matériel plus importante. Cependant, la résolution chromatographique est largement réduite.

A.3. Conclusion

Aujourd'hui, les performances accrues des spectromètres de masse permettent de réduire le préfractionnement des protéines et d'envisager l'établissement du protéome complet d'un échantillon en une seule analyse. Dans cette optique, les développements de méthodes s'orientent vers l'utilisation de gradients de plus en plus longs, et donc vers l'augmentation de la longueur des colonnes.

B. L'analyse par spectrométrie de masse

En 1913, J.J. Thompson publiait la première étude⁸³ décrivant l'acquisition de spectres de masse de plusieurs composés gazeux (O_2 , CO_2 , CO). Cependant, ce n'est que depuis une vingtaine d'années que l'analyse de biomolécules a été envisagée par spectrométrie de masse grâce à la découverte par J. Fehn et K. Tanaka de méthodes d'ionisations douces : l'ESI (*electrospray ionisation*)⁸⁴ et le MALDI (*matrix assisted laser desorption ionisation*)⁸⁵. Depuis, la spectrométrie de masse s'est progressivement implantée comme la technique de choix pour l'analyse protéomique.

La figure 11 présente les différentes parties d'un spectromètre de masse : dans un premier temps, l'échantillon est ionisé dans la source, puis les ions sont séparés suivant leur rapport masse sur charge (m/z) dans l'analyseur. Enfin le courant ionique est converti en signal électrique au niveau du détecteur. On obtient alors des spectres de masse.

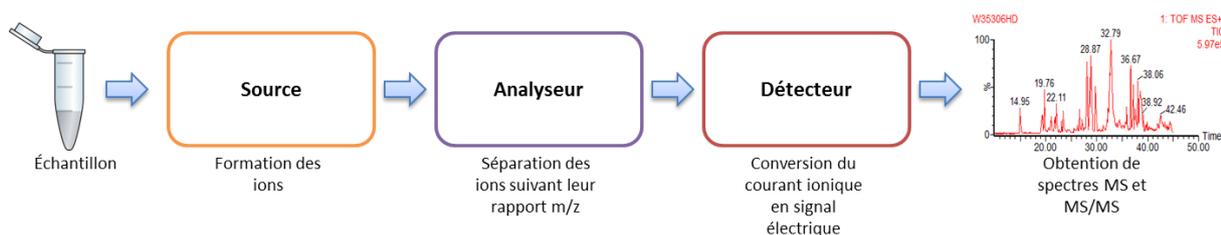


Figure 11
Schéma d'un spectromètre de masse

B.1. Théorie et instrumentation

Au cours de ce paragraphe, les principes des différentes parties du spectromètre de masse seront présentés. Seuls les sources, analyseurs et détecteurs utilisés au cours de ces travaux de thèse seront décrits.

B.1.1. Les sources

L'analyse par spectrométrie de masse de mélanges peptidiques nécessite l'utilisation de sources d'ionisations douces afin d'éviter la fragmentation ou la dégradation des peptides. Deux systèmes sont particulièrement adaptés à l'analyse protéomique : la source ESI⁸⁴ et la source MALDI⁸⁵. Au cours de ce paragraphe, nous ne décrivons que le mode de fonctionnement de la source ESI.

La source ESI est issue d'un long développement démarré par les travaux du groupe de Dole à la fin des années 1960⁸⁶. À la fin des années 1980, le groupe de Fenn réussit à démontrer l'ionisation d'une protéine de 40 kDa sur le couplage ESI-MS⁸⁴.

Les ions sont formés au sein de la source, à pression atmosphérique, par l'intermédiaire d'un capillaire métallique suivant trois étapes^{87, 88} (cf. figure 12) :

- 4) **Production de gouttelettes chargées** à partir de l'électrolyte en solution par l'application d'une différence de potentiel entre le capillaire et la contre-électrode à l'entrée du spectromètre de masse. Les espèces de charges opposées sont séparées et un cône se forme à l'extrémité du capillaire (cône de Taylor)⁸⁹.
- 5) **Diminution de la taille des gouttelettes** par évaporation sous l'influence du chauffage de l'enceinte jusqu'à la limite de Rayleigh^{90, 91}. Les forces de répulsion coulombienne deviennent alors plus importantes que les forces de tension de surface et une série d'explosions coulombiennes engendre la formation de gouttelettes filles.
- 6) **Transfert des ions désolvatés** en phase gazeuse. Deux mécanismes s'affrontent : le mécanisme de la charge résiduelle, où les ions sont progressivement désolvatés⁹², et le mécanisme de l'évaporation des ions, où les ions sont désorbés avant le rayon limite de Rayleigh⁹³.

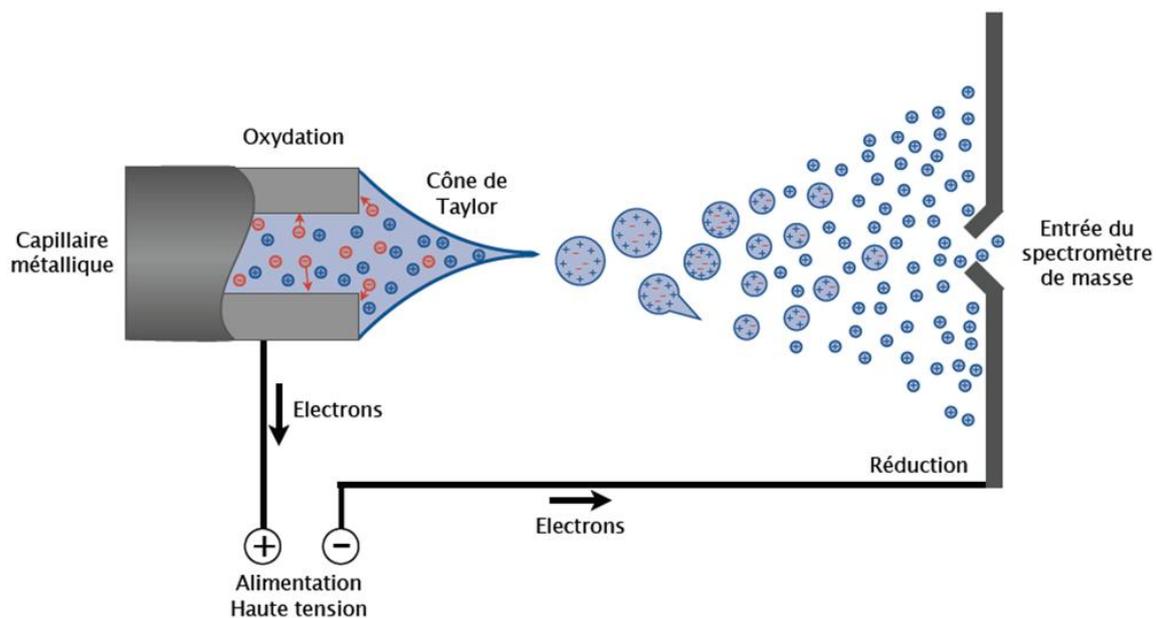


Figure 12
Schéma du mode de fonctionnement de la source électrospray

B.1.2. Les analyseurs

L'analyseur permet de séparer les ions suivant leur rapport masse sur charge (m/z). Il est possible de coupler plusieurs analyseurs afin de réaliser des expériences de spectrométrie de masse en tandem (MS/MS).

En analyse protéomique, quatre types d'analyseurs différents sont couramment utilisés :

- Le tube de vol (TOF)
- Le quadripôle (Q)
- La trappe ionique (IT)
- Les analyseurs à transformée de Fourier (FT-ICR et Orbitrap™)

Les performances de chaque analyseur sont définies par leur résolution, leur limite de détection, leur gamme dynamique, leur précision de masse, leur gamme de rapport masse sur charge analysable et leur capacité à fonctionner en tandem, c'est-à-dire avec un autre analyseur.

Les projets décrits dans ces travaux de thèse ont été réalisés sur plusieurs instruments différents :

- Q-TOF : Synapt G1 (Waters Corporation), Maxis Impact (Bruker Daltonics) et Synapt G2 (Waters Corporation)
- IT : AmaZon (Bruker Daltonics)
- QQQ : TSQ Vantage (Thermo Fisher Scientific)

Seuls les analyseurs utilisés seront décrits dans cette partie.

B.1.2. a. Le tube de vol

La première utilisation d'un tube de vol pour séparer les ions de différents rapports m/z a été publiée en 1946 par W.E. Stephens⁹⁴. Par la suite, dans les années 1950, des TOF ont été conçus et commercialisés⁹⁵⁻⁹⁷. Cependant, les performances de l'époque étaient faibles par rapport à celles d'aujourd'hui et le TOF n'a pas tardé à être supplanté par l'analyseur quadripôle. Ce n'est que dans les années 1980 qu'il a bénéficié d'un regain d'intérêt avec l'invention de la source MALDI⁹⁸ pulsée donc compatible avec des mesures de temps, puis avec le développement de l'injection orthogonale⁹⁹ permettant de passer d'un faisceau d'ions continu à une injection pulsée, ce qui rendait possible le couplage à une source ESI.

Cet analyseur est constitué de deux zones :

- Une zone d'accélération, où les ions sont soumis à une différence de potentiel très élevée de façon à leur conférer une énergie cinétique identique.
- Le tube de vol sous un vide poussé (environ 10^{-7} mbar), où les ions volent librement jusqu'au détecteur.

Grâce à la mesure du temps de vol, il est possible de déterminer le rapport m/z . Ainsi, l'énergie cinétique conférée à chaque ion est définie par l'équation (1) où z est la charge de l'ion, e la charge d'un électron, U la différence de potentiel, m la masse de l'ion et v la vitesse de l'ion.

$$(1) Ec = 2zeU = \frac{1}{2}mv^2$$

L'équation (2) permet de déduire la vitesse d'un ion.

$$(2) v = \sqrt{\frac{2zeU}{m}}$$

Ainsi, on obtient l'équation (3) qui relie le temps de vol à la racine carrée du rapport m/z , où L est la longueur du tube de vol.

$$(3) t = \frac{L}{v} = \frac{L}{\sqrt{2eU}} \sqrt{\frac{m}{z}}$$

Les performances résolutive du TOF ont été améliorées par l'utilisation d'un mode réflecteur permettant de focaliser les ions^{100, 101}. Le réflecteur électrostatique est composé d'une série d'anneaux ou de grilles qui renvoient les ions vers les détecteurs tout en les focalisant. Ainsi, les ions ayant une énergie cinétique plus élevée vont pénétrer plus profondément dans le réflecteur, au contraire des ions ayant une énergie cinétique plus faible.

B.1.2.b. Le quadripôle

L'analyseur quadripolaire a été développé en 1953 par H. Steinwedel et W. Paul¹⁰². Il est composé de quatre barres métalliques parallèles sur lesquelles est appliqué un potentiel ϕ_0 ou $-\phi_0$. Les barres parallèles sont portées au même potentiel et les adjacentes à un potentiel opposé. Ces potentiels possèdent une composante continue U et une composante sinusoïdale ($V\cos\omega t$) avec V amplitude de la radiofréquence et ω la fréquence (cf. équations ci-dessous).

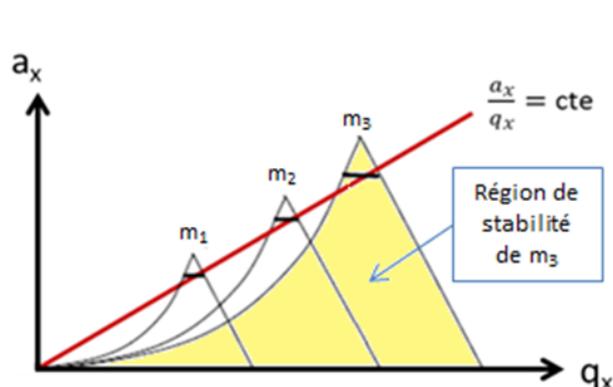
$$\phi_0 = U - V\cos\omega t$$

$$-\phi_0 = -U + V\cos\omega t$$

Les ions entrent dans le quadripôle à une vitesse donnée suivant l'axe z , où ils sont soumis aux potentiels oscillants (ϕ_0 et $-\phi_0$). Ils sont séparés suivant la stabilité de leur trajectoire entre les barres. Les zones de stabilité dans le plan x, y de chacun des ions sont données par la résolution des équations de Mathieu (1866) (cf. figure 13).

Un quadripôle peut ainsi opérer dans plusieurs modes^{103, 104} :

- Full scan : balayage simultané de U et V en maintenant le rapport U/V constant. Les régions de stabilité se succèdent pour chacun des ions selon leurs rapports m/z .
- SIM (*selection ion monitoring*) : une masse particulière est sélectionnée.
- RF only : U est maintenu à 0 afin de transmettre une large bande d'ions.



$$a_x = -\frac{8eU}{m/z \omega^2 r_0^2}$$

$$q_x = \frac{4eV}{m/z \omega^2 r_0^2}$$

e = charge de l'électron = $1,602 \times 10^{-19}$ C

U = tension continue

ω = fréquence

Figure 13

Diagramme de stabilité des ions dans un quadripôle

B.1.2.c. La trappe ionique

La trappe ionique a fait l'objet d'un dépôt de brevet en 1960 par H. Steinwedel et W. Paul¹⁰⁵. Les travaux de Stafford et Finnigan MAT^{106, 107} ont largement amélioré les performances de cet analyseur et en ont permis une première commercialisation en 1983.

La trappe ionique (1 cm^3)^{108, 109} est composée d'une électrode annulaire et de deux électrodes quasi hyperboliques, dites électrodes « chapeaux » (cf. figure 14). Le principe repose sur la création de trajectoires stables pour les ions d'un certain rapport m/z , ou d'une certaine gamme de m/z , tout en défléctant les ions non sélectionnés sur les électrodes. Les ions sont ainsi piégés et éjectés suivant leur rapport m/z . Le vide dans l'enceinte de la trappe est de l'ordre de 10^{-6} mbar. L'hélium présent à une pression résiduelle de 5×10^{-3} mbar, focalise les ions et assure le rôle de gaz de collision. Les ions entrent et sortent de la trappe par les électrodes chapeau.

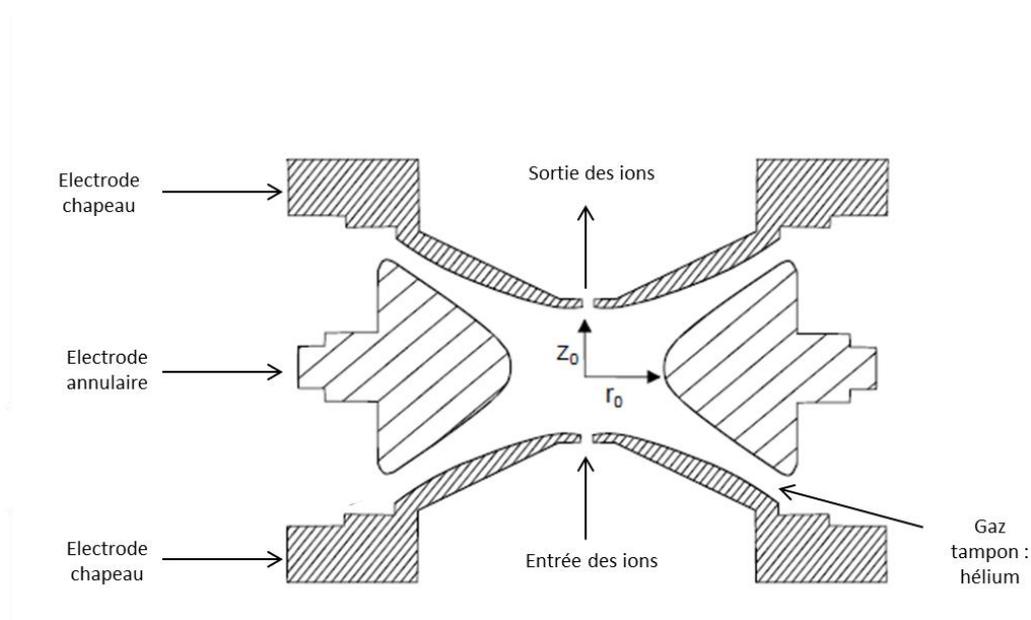


Figure 14

Schéma d'un analyseur à trappe ionique

- **Piégeage des ions dans la trappe**

Une radiofréquence appliquée sur l'électrode annulaire ($V\cos\omega t$) génère un champ quadripolaire afin de piéger les ions dans un puits de potentiel au centre de la trappe ionique. Plus l'amplitude du potentiel est élevée et plus la capacité de piégeage est améliorée. Dans le même temps, ce dernier devient défavorable aux ions de bas m/z .

De la même manière que pour le quadripôle, la stabilité des trajectoires est définie par les équations de Mathieu qui décrivent dans l'espace une courbe de Lissajou (cf. figure 15). Dans le cas présent, la tension continue U est nulle ($a_z = 0$). Ainsi, le diagramme de stabilité est réduit à deux zones paramétrées par le facteur q_z . Pour une valeur inférieure à 0,908, les ions ont une trajectoire stable et sont donc piégés. Pour une valeur supérieure, celle-ci devient instable et les ions sont défectés vers les électrodes.

Ce mode de fonctionnement implique un seuil (*cut off*) de basse masse : il existe des m/z qui seront systématiquement éjectés.

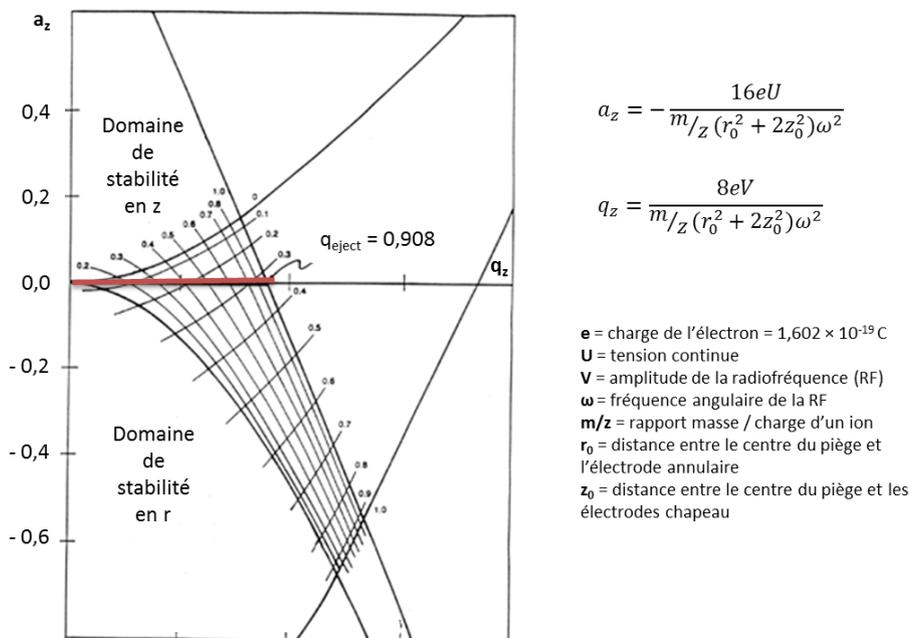


Figure 15

Diagramme de stabilité des ions dans une trappe ionique selon les directions r et z (la droite rouge représente les zones de stabilité pour $a_z = 0$)

- **Éjection des ions de la trappe**

Il existe deux modes d'éjection :

- Éjection simple : un balayage croissant en amplitude permet d'éjecter les ions suivant l'ordre croissant de leur rapport m/z . Cependant, l'amplitude maximale est limitée à 20 kV, et seuls les ions ayant un rapport $m/z < (m/z)_{max}$ seront détectés.

- Éjection résonante : les ions sont mis en résonance par application d'une tension alternative supplémentaire sur les électrodes chapeau de façon à conférer une énergie cinétique aux ions, et donc élargir la gamme de masse analysable.

B.1.3. La détection

Le détecteur va permettre de transformer le courant d'ions en un courant électronique. Le signal est ensuite traité, digitalisé et traduit en valeurs numériques codées sur un certain nombre de bits.

Pour la spectrométrie de masse, il existe deux catégories principales de détecteurs :

- Les détecteurs basés sur un effet inductif par courant image. Ils sont plus couramment employés sur les appareils à transformée de Fourier de types FTICR ou Orbitrap™.
- Les détecteurs nécessitant la conversion des ions en électrons, couramment employés sur la majeure partie des spectromètres de masse utilisés en analyse protéomique (IT, TOF, Q).

De la même manière que précédemment, nous ne décrivons que les détecteurs utilisés au laboratoire, qui sont basés sur l'émission et la multiplication d'électrons secondaires^{110, 111}. Il en existe deux types principaux : les multiplicateurs d'électrode à dynode discrète ou à dynode continue. Parmi ces derniers, on retrouve :

- **Le channeltron**¹¹²

L'impact de l'ion sur la dynode de conversion génère des électrons secondaires qui sont accélérés par des gradients de potentiel au sein du channeltron (électrode). L'impact des ions sur la surface de cette électrode engendre l'émission de nouveaux électrons secondaires (cf. figure 16). Le gain du détecteur va de $10^4 - 10^5$, pour les channeltrons cylindriques, à 10^8 pour les channeltrons en cornet.

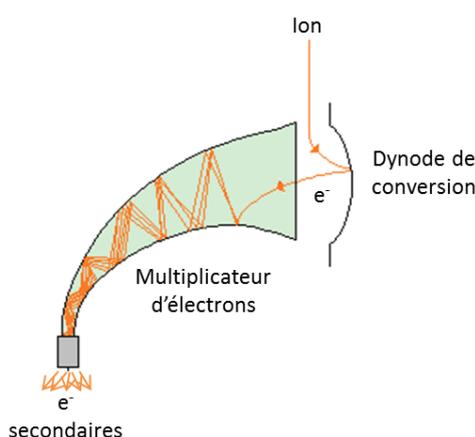


Figure 16

Schéma d'un détecteur channeltron

- **La galette de microcanaux (MCP)**¹¹³

Il s'agit ici d'une galette percée d'un grand nombre de canaux (plusieurs millions) orientés de quelques degrés par rapport à la normale de la surface. Chaque ion pénètre dans un canal et la collision avec la paroi entraîne l'émission d'électrons secondaires (cf. figure 17). Ainsi, chaque

canal peut-être assimilé à un petit multiplicateur d'électrons. Pour un MCP, le gain est de 10^4 . Deux ou trois MCPs sont souvent associés de façon à obtenir un gain allant de 10^6 à 10^7 .

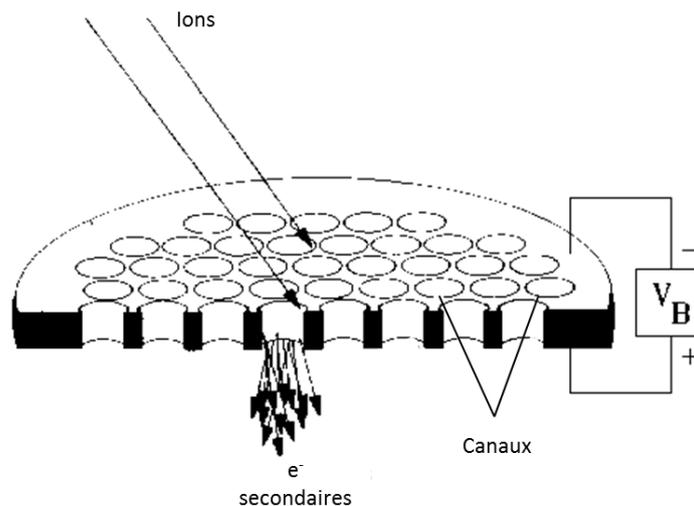


Figure 17
Schéma d'une galette de microcanaux (MCP)

- **Le magneTOF™**

Les électrons secondaires sont ici émis lors de la collision des ions avec une surface plane. L'amplification des électrons est réalisée à l'aide de potentiels électrostatiques déformés par un champ magnétique. Cette technologie récente, proposée par la société ETP, permet d'améliorer le temps de réponse du détecteur.

B.2. La spectrométrie de masse en tandem

L'analyse protéomique à haut débit était à ses débuts réalisée à l'aide d'acquisition MS des peptides issus de la protéine digérée. Par la suite, des techniques de MS/MS, impliquant une fragmentation de l'ion précurseur, ont été implémentées et ont permis une meilleure caractérisation des peptides. Il existe deux façons d'utiliser les analyseurs afin d'effectuer des expériences de spectrométrie de masse en tandem :

- Dans l'espace : utilisation de deux analyseurs (Q-TOF, QQQ)
- Dans le temps : utilisation d'un seul analyseur (IT)

B.2.1. Modes d'acquisition

Le mode d'acquisition principal utilisé au cours de ces travaux est le mode DDA (*data dependent acquisition*). Il consiste en l'alternance de scans MS et MS/MS. Un exemple est montré en figure 18 : les trois ions les plus intenses sont sélectionnés sur le spectre MS pour être fragmentés en MS/MS. Le temps de cycle est alors de 2,2 s.

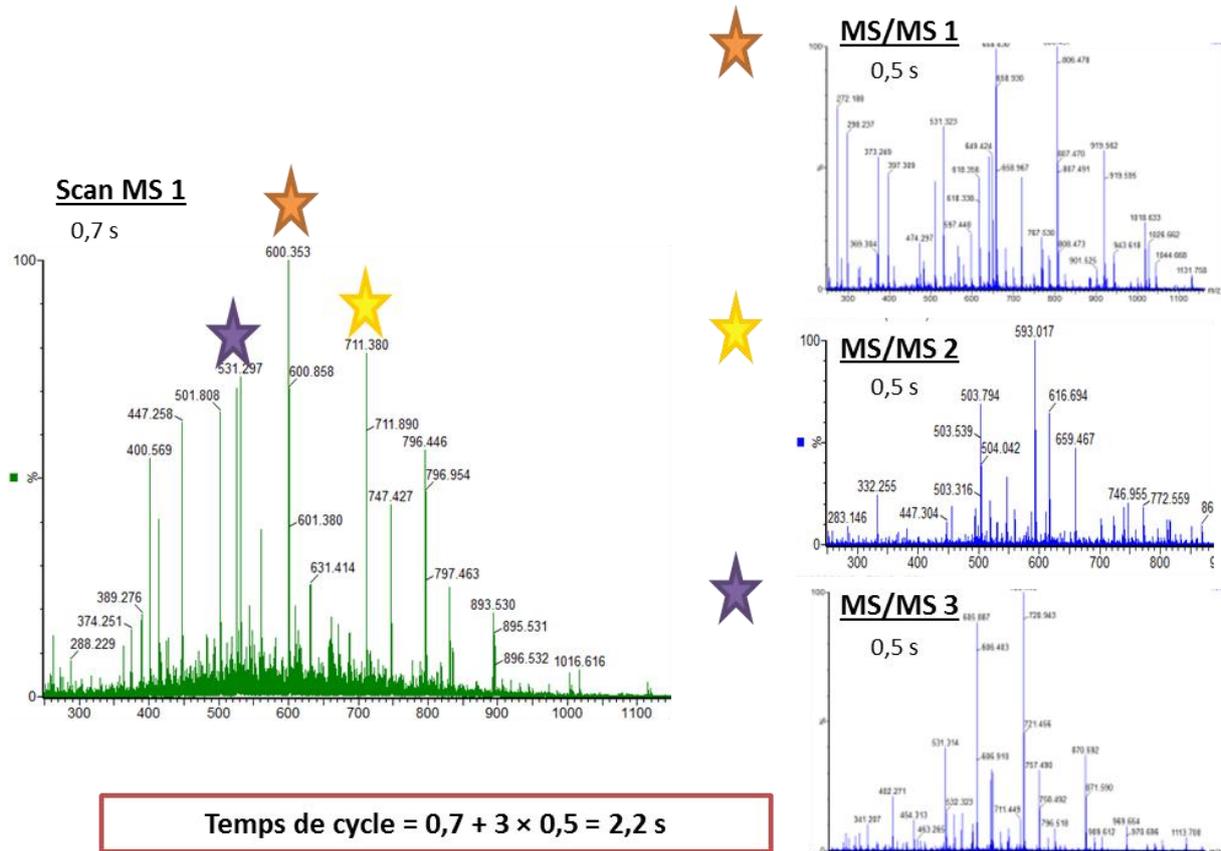


Figure 18
 Représentation schématique du mode d'acquisition DDA (*data dependant acquisition*)
 Dans cet exemple, 3 ions marqués par des étoiles sont sélectionnés sur le spectre MS pour la MS/MS.

Cependant, bien qu'ayant prouvé son efficacité, ce mode d'acquisition présente une limitation majeure : la stochastique de la sélection des ions. Ainsi, tous les ions du spectre MS ne sont pas sélectionnés pour la MS/MS. Par ailleurs, les ions élués par la chromatographie pendant la fragmentation MS/MS ne seront pas sélectionnés.

Afin de s'affranchir de ce problème, des modes d'acquisition appelés DIA (*data independant acquisition*) sont en cours de développement. Dans ce cas, les ions ne sont plus sélectionnés pour la fragmentation. Il existe deux techniques majeures décrites plus précisément dans la Partie III – chapitre 2 :

- Isolation séquentielle et fragmentation des précurseurs sur une certaine fenêtre de masse (SWATH¹¹⁴, PaCIFIC¹¹⁵)
- Alternance de scans basse et haute énergies (MS^E)¹¹⁶

B.2.2. Types de fragmentation

Il existe trois types de fragmentation principaux : CID (*collision induced dissociation*)¹¹⁷, ETD (*electron transfert dissociation*)¹¹⁸ et ECD (*electron collision dissociation*)¹¹⁹. Seul le mode CID a été utilisé au cours de ces travaux et il sera par conséquent le seul à être présenté.

Au sein du mode CID, deux types d'énergies peuvent être appliqués : la basse (quelques eV) et la haute énergie (quelques keV). Les ions précurseurs sont dissociés par collision avec un gaz inerte (He, N₂, Ar).

La fragmentation a lieu sur la chaîne la plus labile à savoir sur la liaison amide (liaison peptidique) par l'apport d'un proton. Des ions b et y, selon la nomenclature de K. Biemann¹²⁰ établie en 1990 (cf. figure 19), sont majoritairement obtenus.

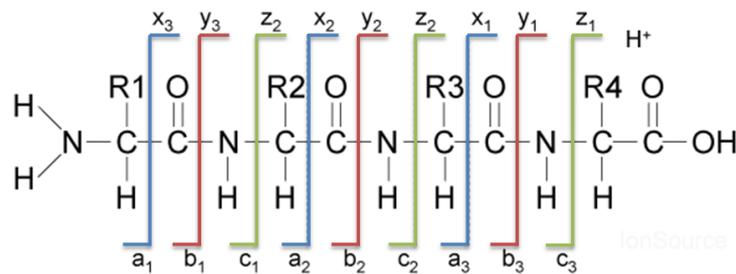


Figure 19

Nomenclature selon Biemann des ions obtenus à la suite d'une fragmentation CID de peptides

La séquence de chaque peptide est ensuite déterminée à partir de la masse des ions fragments successifs. Un exemple de spectre est présenté en figure 20.

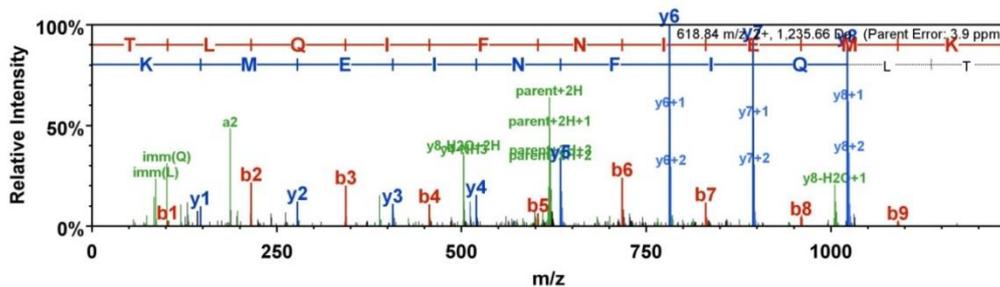


Figure 20

Exemple de spectre MS/MS : peptide KMEINFIQLT 2+ à m/z = 618,84

Du point de vue mécanistique, la protonation a lieu sur des sites basiques avec l'ordre de préférence suivant : groupes basiques de la chaîne latérale > azote N-terminal > oxygène porté par le carbone de la liaison amide > azote de l'amide. Le modèle du proton mobile¹²¹ permet d'expliquer la fragmentation des peptides : l'énergie interne des ions augmente par collision avec les molécules de gaz et entraîne une délocalisation du proton conduisant à la formation d'intermédiaires moins favorables énergétiquement.

Ceci explique la difficulté à fragmenter correctement les peptides tryptiques monochargés. Ainsi, dans ce cas, la charge est « séquestrée » au niveau de l'acide aminé basique en C-terminal (lysine ou arginine). L'énergie nécessaire pour délocaliser cette charge le long du squelette peptidique est élevée. Au contraire, dans le cas de peptides doublement chargés, les charges sont portées par l'acide aminé basique en C-terminal et l'azote N-terminal. Le proton situé en N-terminal est facilement transféré le long de la chaîne peptidique.

B.2.3. Les analyseurs pour la spectrométrie de masse MS/MS

Cette partie présentera plus particulièrement les trois types d'instruments utilisés au cours de ces travaux de thèse :

- La trappe ionique
- Le Q-TOF
- Le QQQ

B.2.3.a. La trappe ionique

La trappe ionique permet de réaliser des expériences de spectrométrie de masse MS^2 , MS^3 , ..., MS^n , grâce à ses capacités de piégeage et d'éjection sélective des ions. L'obtention de spectres MS/MS se déroule en deux étapes :

- **Isolation de l'ion d'intérêt**

Les ions de plus faible m/z que l'ion d'intérêt sont éjectés par un balayage croissant de l'amplitude. Puis les ions de m/z plus élevés sont mis en résonance et éjectés.

- **Excitation et fragmentation des ions précurseurs**

L'ion isolé est fragmenté par excitation à sa fréquence de résonance à l'aide de l'application d'une tension alternative sur l'électrode chapeau. L'amplitude de cette tension est plus faible que celle d'éjection et permet d'augmenter l'énergie cinétique des ions sans pour autant les éjecter de la trappe ionique. L'ion est ensuite fragmenté par conversion de cette énergie cinétique en énergie interne et collision avec les molécules d'hélium.

L'utilisation de la fréquence de résonance permet d'être très spécifique et de ne fragmenter que les ions précurseurs, sans affecter les ions fils.

B.2.3.b. Le Q-TOF

Un exemple de Q-TOF, le Maxis Impact (Bruker Daltonics) est présenté en figure 21. Lors de l'acquisition MS, le quadripôle est paramétré en mode RF only. Une large bande d'ions est transmise et séparée par le TOF avant détection. En mode, MS/MS, un précurseur (ion parent) est sélectionné par le quadripôle avant d'être fragmenté dans la cellule de collision. Les ions fils obtenus sont séparés dans le TOF avant la détection.

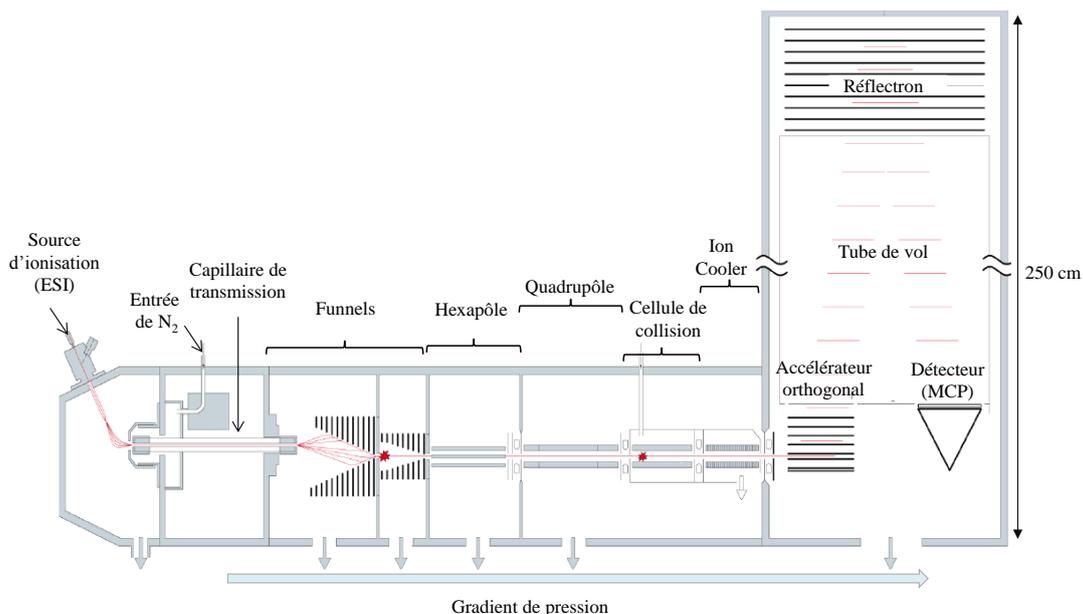


Figure 21
Schéma du spectromètre de masse Q-TOF (Maxis Impact, Bruker Daltonics)

B.2.3.c. Le QQQ

La figure 22 présente le triple quadripôle TSQ Vantage™ (Thermo Fisher Scientific). De par sa géométrie, le QQQ permet d'utiliser des fragmentations caractéristiques pour détecter et quantifier de manière spécifique des composés dans un mélange complexe. Ainsi, il est possible d'envisager plusieurs modes, par exemple¹²² :

- *Product ion scanning* : sélection du précurseur dans Q1, fragmentation dans Q2 et balayage de l'ensemble des fragments dans Q3
- *Precursor ion scanning* : balayage des précurseurs dans Q1, fragmentation dans Q2 et sélection d'un ion fils dans Q3
- *Selected reaction monitoring (SRM)* : sélection d'un ion parent dans Q1, fragmentation dans Q2 et sélection d'un ion fils dans Q3

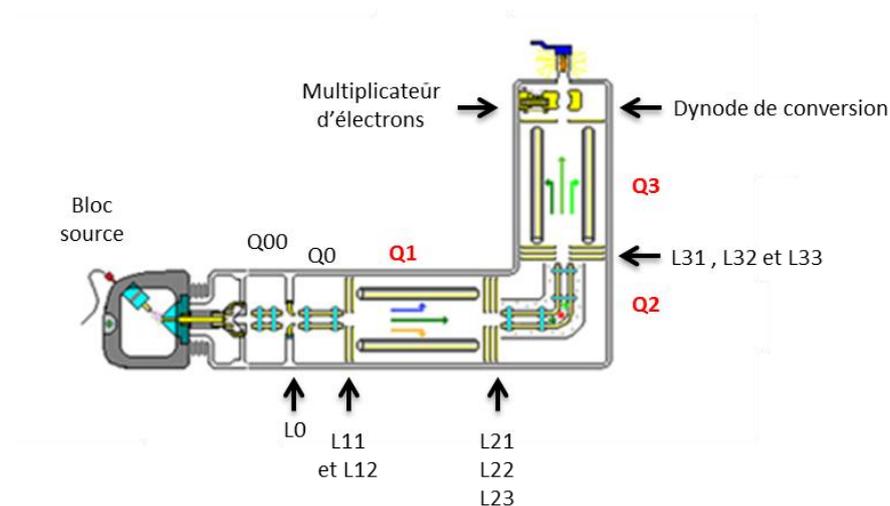


Figure 22

Schéma du triple quadripôle TSQ Vantage (Thermo Fisher Scientific)

Q00 et Q0 sont deux quadripôles utilisés pour la transmission des ions. L0, L11, L12, L21, L22, L23, L31, L32 et L33 sont des lentilles permettant de focaliser les ions. Q1, Q2 et Q3 sont les trois quadripôles. Le détecteur est un channeltron (dynode de conversion et multiplicateur d'électrons).

B.3. Comparaison des performances des différents couplages utilisés

Le tableau 4 présente l'ensemble des couplages utilisés au cours de cette thèse et récapitule leurs performances respectives et leurs domaines d'application dans le cadre de la recherche de biomarqueurs. On notera ici la nette amélioration des performances du Maxis Impact par rapport à un Q-TOF de plus ancienne génération tel que le Synapt G1.

	G1	G2	AmaZon	Maxis Impact	TSQ Vantage
Constructeur	Waters	Waters	Bruker Daltonics	Bruker Daltonics	Thermo Fisher Scientific
Source	ESI	ESI	ESI	ESI	ESI
Analyseur	Q-TOF	Q-TOF	IT	Q-TOF	QQQ
Détecteur	MCP	MCP	Channeltron	MCP	Channeltron
Résolution (m/z = 722,33)	9000	25000	3000	40000	1000
Précision de masse	15 ppm	15 ppm	0,2 Da	10 ppm	0,7 Da
Temps de scan MS/MS (ms)	500	400	250	60	15
Gamme de masse	32 000 m/z	20 000 m/z	3 000 m/z	20 000 m/z	3 000 m/z
Applications	Protéomique globale après fractionnement	Protéomique globale sans fractionnement	Protéomique globale après fractionnement	Protéomique globale sans fractionnement	Protéomique ciblée
Projet	Recherche de biomarqueurs de différents types de lymphomes	Mise en place de la LC-HDMS ^E	Recherche de biomarqueurs de cellules souches de glioblastome	Recherche de biomarqueurs prédictifs des rechutes des lymphomes diffus à larges cellules B	Quantification relative pour la validation de biomarqueurs de différents types de lymphomes

Tableau 4

Comparaison des performances des différents spectromètres de masse utilisés au cours de ces travaux de thèse

C. Identification et validation des spectres de fragmentation

L'interprétation des spectres MS/MS acquis est un point clé de l'analyse protéomique. En effet, lors d'une analyse MS/MS, plus de 100 000 spectres peuvent être acquis et il est crucial de pouvoir les assigner à une séquence peptidique donnée. C'est pourquoi des banques de données, des moteurs de recherche dans ces banques et des algorithmes de validation ont été développés.

C.1. Les banques de données

L'interprétation des spectres MS/MS dépend directement de la disponibilité de la séquence dans la banque. Il existe deux principaux types de banques :

- **Les banques de dépôt**, telles que NCBI's Entrez Protein¹²³, qui regroupent des séquences protéiques issues de la transcription et de la traduction des séquences d'ADN et d'ARN. Ces banques sont très volumineuses et comportent un nombre important de redondances.
- **Les banques corrigées** regroupent des séquences protéiques issues de la littérature avec des annotations de qualité validées manuellement par des experts (UniProtKB Swiss-Prot)¹²⁴ ou automatiquement (UniProtKB/Trembl)¹²⁴. Cette dernière répond à la croissance rapide des

banques de données nucléotidiques et facilite le travail des experts pour le transfert des données vers Swiss-Prot.

Dans le cas de la recherche de biomarqueurs par protéomique globale, la qualité des annotations et la simplicité des interprétations sont primordiales. C'est pourquoi l'interprétation des spectres MS/MS a été réalisée au cours des travaux de thèse à l'aide de la banque humaine Swiss-Prot. Il peut être intéressant de rechercher des variants de séquence lors de la deuxième étape, afin de prévalider les marqueurs.

C.2. Les recherches dans les banques de données

Les protéines sont identifiées par recherche dans les banques de données à l'aide de moteurs de recherche : les spectres MS/MS sont assignés par corrélation avec la digestion *in silico* de l'ensemble des protéines contenues dans une banque donnée. Le regroupement des peptides identifiés permet ensuite de remonter à la protéine.

Différents algorithmes de recherche dans les banques de données ont été développés pour l'identification des spectres MS/MS, parmi lesquels Mascot¹²⁵, Sequest¹²⁶, X!Tandem¹²⁷, Andromeda¹²⁸ ou encore Omssa¹²⁹. Ces algorithmes établissent un score de corrélation basé sur l'adéquation entre le spectre MS/MS expérimental et le spectre MS/MS prédit. Plusieurs paramètres sont pris en compte : le nombre de fragments attribués, le nombre de fragments non attribués, la tolérance sur la masse du précurseur. Le score de corrélation est par la suite converti en une mesure statistique. Au cours de ces travaux, les logiciels Mascot et Omssa ont été évalués.

Mascot est le logiciel le plus couramment utilisé pour l'identification des protéines bien qu'il ne soit pas *open source*. À chaque identification, Mascot attribue un score d'ion, révélateur de la qualité du spectre MS/MS, et un score d'identité permettant de mesurer la taille de l'espace de recherche. Un score d'ion supérieur au seuil d'identité signifie que l'identification a moins de 5 % de probabilité d'être un événement aléatoire.

Depuis 2009, l'équipe de bioinformatique du laboratoire a mis en place une interface permettant d'utiliser Omssa (<https://msda.unistra.fr/>) afin de répondre aux exigences croissantes des éditeurs quant à la confiance dans les identifications. Ce moteur de recherche se base sur l'hypothèse que le nombre d'identifications pour un spectre suit une distribution de Poisson. L'algorithme utilise ensuite ce modèle pour évaluer la significativité d'une identification et établir un score.

C.3. La validation des données

Les recherches dans les banques de données engendrent une possibilité non négligeable d'identifier des faux positifs, c'est-à-dire d'assigner un peptide incorrect à un spectre MS/MS. Aujourd'hui, il est courant d'identifier plus de 30 000 spectres au cours d'une analyse. Il paraît peu envisageable de les vérifier et valider manuellement. C'est pourquoi le laboratoire a décidé d'utiliser la stratégie *Target-decoy*^{130, 131} basée sur l'utilisation d'une banque de données concaténées : c'est-à-dire contenant l'ensemble des séquences de l'organisme considéré ainsi que ces séquences inversées. Le nombre de

séquences assignées dans la banque *decoy* devrait refléter le nombre de faux-positifs (FDR) identifiés dans la banque *target*. L'ensemble des identifications est ensuite filtré de manière à obtenir un taux de FDR < 1 % sur les protéines (cf. équation ci-dessous).

$$\% \text{ FDR} = \frac{\text{Nombre de decoys}}{\text{Nombre de targets} + \text{nombre de decoys}} \times 100$$

D. Quantification

Depuis quelques années, l'analyse protéomique évolue d'un aspect qualitatif, avec l'établissement de listes de protéines, vers un aspect quantitatif. Dans le cadre de la recherche de biomarqueurs, ces évolutions sont particulièrement importantes car elles permettent désormais de proposer un candidat plus fin qu'absent/présent.

Les techniques de quantification pour la recherche de biomarqueurs peuvent être classées en deux catégories (cf. figure 23) :

- La protéomique globale, qui permet de découvrir et de proposer de nouveaux biomarqueurs et donc de quantifier l'intégralité du contenu d'un échantillon.
- La protéomique ciblée, qui permet de quantifier spécifiquement les cibles choisies par des approches SRM (*selected reaction monitoring*).

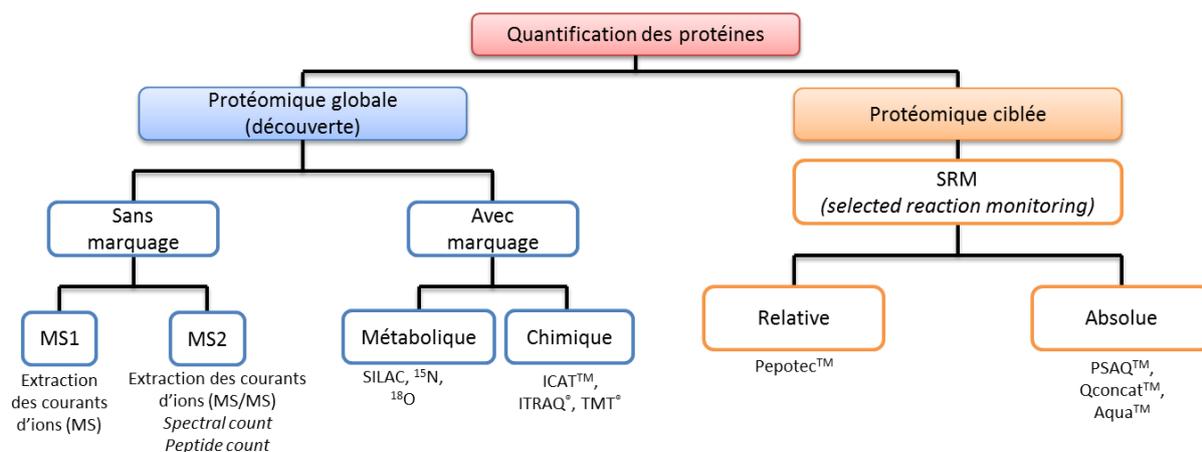


Figure 23

Les différentes options existantes quant au choix d'une méthode de quantification des protéines

Au cours de ce chapitre, nous ne décrivons que les approches concernant la protéomique globale. Les approches de protéomique ciblée seront explicitées en détail dans la Partie III – chapitre 1.

Il existe deux approches principales pour la protéomique globale : les approches sans marquage et les approches avec marquage.

D.1. Les approches sans marquage

Les approches sans marquage, aussi nommées *label free*, sont très populaires et souvent utilisées dans le cadre d'analyses globales. En effet, elles sont peu coûteuses et simples à mettre en œuvre¹³². Il est possible d'utiliser deux techniques basées sur le spectre MS, avec l'extraction des courants d'ions de chaque peptide, ou sur le spectre MS/MS, avec comptage du nombre de spectres (*spectral count*)¹³³ ou extraction des courants d'ions des fragments¹³⁴.

Ces approches sont décrites de manière plus détaillée dans la Partie II – chapitre 4.

D.2. Les approches avec marquage

Les approches avec marquage sont basées sur le fait que les propriétés physicochimiques des protéines et des peptides ne changent pas à l'issue d'un marquage.

D.2.1. Le marquage métabolique

Le SILAC¹³⁵ (*stable isotope labeling with amino acids in cell cultures*) est la méthode la plus courante. Les lignées cellulaires sont cultivées dans des milieux enrichis en résidus d'acides aminés marqués ¹³C et ¹⁵N comme des leucines ou des lysines. Après lyse des cellules, les protéines extraites des milieux marqués et non marqués sont mélangées. Cette technique présente l'immense avantage de s'affranchir de l'ensemble de la variabilité technique. De plus, en utilisant des appareils très haute résolution (Rs = 480 000), du type OrbitrapTM FusionTM TribridTM dernière génération, il est possible de comparer douze échantillons en même temps¹³⁶. Avec un appareil de plus faible résolution, seuls trois échantillons peuvent être comparés.

Une autre technique consiste à marquer de manière métabolique l'ensemble des résidus d'acides aminés contenus dans un échantillon avec de l'azote ¹⁵N¹³⁷ ou du carbone ¹³C¹³⁸. Cependant, cette technique reste difficile à utiliser dans le sens où l'incorporation varie suivant la nature du peptide¹³⁹.

D.2.2. Le marquage chimique

Ce marquage consiste à fixer une fonction chimique sur les peptides des protéines, en général sur les fonctions amines des lysines ou les fonctions thiols des cystéines. On peut en distinguer trois groupes :

- **Le marquage isotopique** : l'ICAT¹⁴⁰ (*Isotope Coded Affinity Tag*), par exemple, comporte deux formes, lourde et légère (deutérée et non-deutérée). Ces réactifs réagissent avec la fonction thiol des cystéines. Il est ensuite possible de réaliser un enrichissement spécifique sur une colonne d'avidine en utilisant la fonction biotine. Cette technique présente trois inconvénients majeurs : seules deux conditions biologiques peuvent être comparées, le nombre de cystéines est variable d'un peptide à l'autre, et la deutération peut occasionner des différences de temps de rétention¹³⁴.
- **Le marquage avec un fluorochrome** : cette technique s'utilise dans le cadre d'application DIGE (*difference in gel electrophoresis*). Les protéines sont marquées sur les cystéines par un fluorochrome¹⁴¹. Elles sont ensuite séparées sur gel 2D SDS PAGE puis révélées. La comparaison

de l'intensité de fluorescence sur différentes images du même gel permet de comparer le taux d'expression des protéines dans plusieurs échantillons. Jusqu'à quatre échantillons peuvent être analysés en simultané sur un gel. Par ailleurs, l'utilisation de standards internes permet d'effectuer une comparaison inter-gels. Cette technique présente cependant l'inconvénient de nécessiter un temps de manipulation important et de ne pas être adaptée à l'analyse de protéines membranaires.

- **Le marquage isobarique** : l'iTRAQ™, par exemple, permet de marquer les peptides avec une fonction chimique de même masse (isobare) mais générant des fragments différents (de 114 à 118 Da)¹⁴². Jusqu'à huit échantillons peuvent être comparés en même temps, voire dix avec la technique similaire TMT (*tandem mass tag*)¹⁴³. De plus, contrairement à la technique SILAC, il n'y a pas d'augmentation de la complexité chromatographique car les peptides co-éluent, ni de la complexité du spectre MS car les peptides sont isobares. Un des inconvénients majeurs est lié aux interférences sur le spectre MS/MS qui peuvent biaiser la quantification.

D.3. Conclusion

La quantification sans marquage est une technique attractive pour la recherche de biomarqueurs dans des échantillons cliniques¹³². En effet, elle peut s'appliquer à tous types d'échantillons (fluides, cellules, lignées cellulaires, tissus) et permet de comparer un nombre important d'échantillons.

Partie II

Développement de méthodes d'analyse protéomique adaptées à la recherche de biomarqueurs

En 2011, Matthias Mann publiait une étude¹⁴⁴ montrant que, dans une analyse nanoLC-MS/MS, plus de 100 000 peptides étaient potentiellement détectables mais que la majorité restait inaccessible à l'analyse DDA (data dependant acquisition) dans le sens où seuls 9797 peptides avaient pu être sélectionnés et identifiés (cf. figure 24).

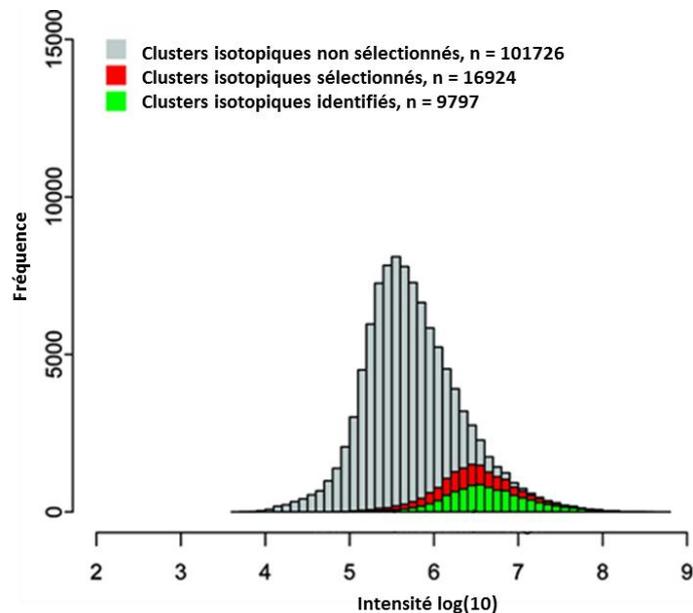


Figure 24

Répartition des clusters isotopiques détectés sur une analyse LC-MS/MS suivant trois catégories : non sélectionnés, sélectionnés et identifiés

Source : adapté de Michalski and al¹⁴⁴

Dans le cadre de l'approche protéomique globale, ayant pour but la proposition de potentiels biomarqueurs candidats, plusieurs défis d'importance restent à relever afin d'identifier la totalité des protéines présentes dans un échantillon :

- Comment s'affranchir de la gamme dynamique intrinsèque à l'échantillon et ne pas identifier uniquement les protéines les plus abondantes ?
- Quelles étapes permettent de réduire la complexité d'un échantillon ?
- Comment utiliser au mieux les outils bio-informatiques pour favoriser le nombre d'identifications ainsi que pour associer une donnée quantitative à chaque protéine ?
- Comment assurer un maximum de confiance dans les résultats obtenus et les biomarqueurs proposés ?

Pour répondre à ces questions, chaque étape de l'analyse protéomique décrite en figure 25 a été finement optimisée. Ainsi, au cours de cette partie, nous démontrerons dans un premier temps l'intérêt de travailler sur la préparation de l'échantillon. Puis nous détaillerons les développements instrumentaux nécessaires. Dans le troisième chapitre, la recherche dans les banques de données sera approfondie. Enfin, nous discuterons de la possibilité de réaliser une quantification de chaque protéine contenue dans un échantillon.

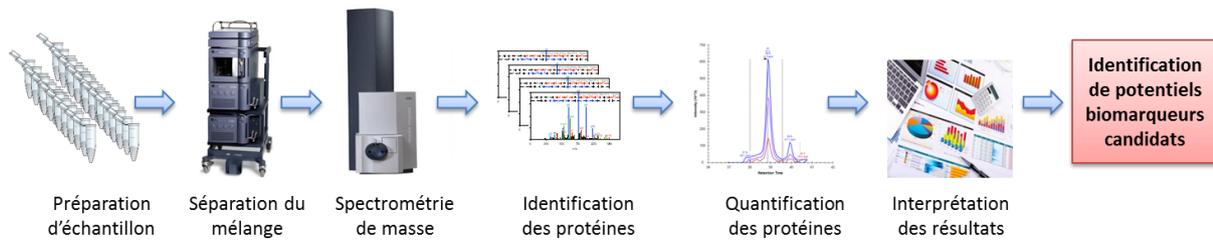


Figure 25

Étapes cruciales de l'analyse protéomique conduisant à l'identification de potentiels biomarqueurs candidats pertinents

Préparation de l'échantillon, séparation du mélange protéique, injection en spectrométrie de masse, identification dans les banques de données, quantification de chaque protéine, interprétation des résultats.

CHAPITRE 1

Préparation des échantillons et enrichissement en protéines membranaires

A. Le choix des protéines membranaires

A.1. L'intérêt de travailler sur un sous-protéome

Idéalement, la recherche de biomarqueurs devrait s'effectuer sur des échantillons directement issus de la zone tumorale (tissus, cellules...). Cependant, le prélèvement de ces échantillons implique un acte médical invasif pour le patient. C'est pourquoi un grand intérêt s'est développé pour les fluides biologiques tels que le sang et les urines (prélèvement rapide et simple, sans conséquence pour le patient)¹⁴⁵⁻¹⁴⁷.

Au cours de mon travail de thèse, j'ai eu l'occasion d'étudier des lignées cellulaires issues de biopsies, des prélèvements sanguins et du plasma. Deux problèmes majeurs se posent lors de l'analyse par protéomique de ces échantillons. D'abord, les mélanges protéiques étudiés (sang, plasma, tissus cérébraux) se montrent souvent très complexes : le protéome d'un mammifère peut en effet compter jusqu'à 50 000 protéines. De plus, la gamme dynamique est très large (6 à 8 ordres de grandeur dans les cellules et 10 à 12 dans le plasma). Par exemple, l'albumine est présente à des concentrations comprises entre 35 et 50 mg/mL alors que l'interleukine 6 est présente à des concentrations comprises entre 0 et 5 pg/mL dans le plasma sanguin¹⁴⁸.

Afin de réduire la gamme dynamique de l'échantillon, plusieurs stratégies sont envisageables :

- **Déplétion des protéines les plus abondantes** par des colonnes d'affinité. Des kits commerciaux permettent par exemple de déléter l'albumine ou les immunoglobulines (Pierce Albumin Depletion Kit)¹⁴⁹. D'autre part, la société Bio-RAD commercialise une technologie nommée Proteominer™ basée sur l'utilisation d'une bibliothèque combinatoire d'hexapeptides (avec 20 de résidus d'acides aminés, il y a 64 millions de combinaisons possibles)¹⁵⁰. Les protéines les plus abondantes vont saturer leurs sites rapidement ce qui permettra d'enrichir l'échantillon en protéines minoritaires.
- **Enrichissement de l'échantillon en protéines possédant des modifications particulières** comme les phosphorylations ou les glycosylations. Les glycosylations sont particulièrement intéressantes car elles sont très abondantes sur les marqueurs de surface cellulaire ainsi que

sur les protéines sécrétées. De plus, plusieurs études ont permis de mettre en évidence une corrélation entre des taux de glycosylations anormaux et certains cas de cancer¹⁵¹⁻¹⁵³. Les protéines glycosylées peuvent être enrichies en utilisant une chromatographie d'affinité possédant une phase stationnaire greffée avec de la lectine¹⁵⁴.

- **Enrichissement en protéines issues d'un compartiment cellulaire particulier**¹⁵⁵⁻¹⁵⁸, comme les mitochondries ou les membranes^{159, 160}, par centrifugation différentielle (composés cellulaires séparés suivant leur taille et leur forme) ou encore, par utilisation de gradients de densité continus ou discontinus. Les gradients de sucrose permettent notamment d'isoler les protéines membranaires.
- **Collecte de microvésicules sécrétées par les cellules**¹⁶¹ : exosomes formés suite à l'exocytose des endosomes ou microparticules issues de la membrane plasmique.

A.2. Le choix des protéines membranaires

Dans le cadre de la recherche de biomarqueurs, nous avons choisi de travailler sur les protéines membranaires. En effet, celles-ci sont situées à une position-clé, enchâssées dans la membrane plasmique, à l'interface entre le milieu extérieur et les compartiments intracellulaires. Cette position leur confère des fonctions biologiques importantes : communication entre les cellules (réception des signaux extracellulaires), transport d'ions et de solutés et transduction du signal¹⁶². Ces protéines sont par conséquent une cible privilégiée pour le développement de nouveaux médicaments. Ainsi, 30 % des médicaments aujourd'hui commercialisés ciblent des protéines membranaires¹⁶³. Par ailleurs, elles sont aussi couramment utilisées pour poser un diagnostic en milieu hospitalier, en utilisant des techniques telles que la cytométrie en flux ou le test Elisa.

Les protéines membranaires apparaissent donc comme des cibles idéales pour de la recherche de biomarqueurs de différents types de cancer. Cependant, l'analyse de cette famille de protéines reste à ce jour un défi analytique.

Les protéines membranaires représentent 20 à 30 % de l'ensemble des cadres de lecture du génome (ORFs)¹⁶⁴. Des domaines transmembranaires composés de 20 à 30 résidus d'acides aminés leur permettent de traverser la bicouche lipidique des cellules. Ces domaines fortement hydrophobes peuvent former des hélices α ou des feuilletts β (majoritairement chez les procaryotes). Les domaines cytosoliques ou extracellulaires sont quant à eux hydrophiles.

Ce caractère amphiphile associé à une faible abondance rend leur détection par protéomique difficile¹⁶⁵. Il est donc nécessaire de mettre en place une stratégie spécifique¹⁶⁶. Trois techniques ont été développées au laboratoire : la génération de microparticules sous l'action d'un stress à partir de cellules sanguines pathologiques, la collecte de microparticules circulantes dans le plasma sanguin et la formation de *ghosts* membranaires.

B. Les microparticules

Il existe trois types de microvésicules générées par la cellule (cf. figure 26A) ¹⁶⁷⁻¹⁶⁹:

- **Les corps apoptotiques** (1000-3000 nm) sont formés exclusivement durant la dernière étape de l'apoptose.
- **Les exosomes** (40-100 nm) sont formés à la suite de l'exocytose des endosomes.
- **Les microparticules** (50 -1000 nm)¹⁷⁰ sont émises à partir de la membrane des cellules sous l'action d'un stress ou en conditions d'apoptose. Ce stimulus occasionne l'apport d'un signal calcique important, ce qui implique une restructuration du squelette avec une externalisation des phosphatidylsérines. La membrane plasmique « bourgeonne », permettant la libération des microparticules (cf. figure 26B). Les microparticules étant issues de la membrane plasmique, elles devraient posséder à leur surface des marqueurs représentatifs issus de cette dernière.

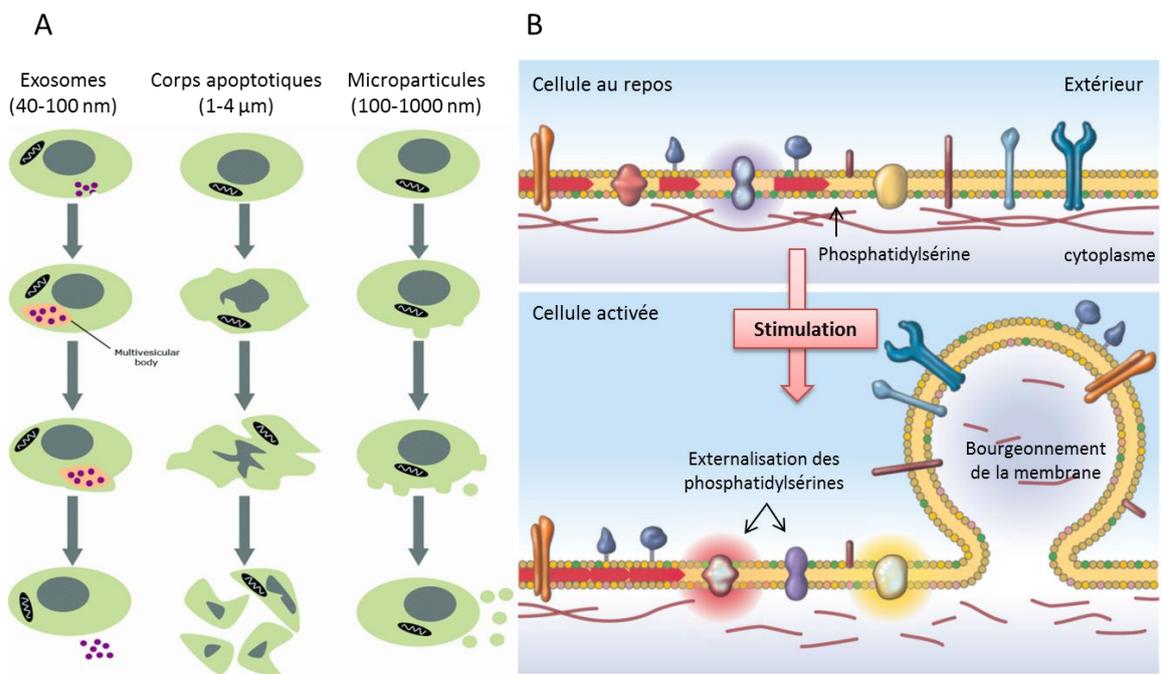


Figure 26

Origines des divers types de microvésicules

A) Les 3 types de microvésicules existantes : les exosomes sont formés par un bourgeonnement interne de la membrane formant des microvésicules qui, après accumulation, sont libérées dans le milieu extracellulaire. Les corps apoptotiques sont formés au cours de la dernière étape de l'apoptose. Les microparticules sont issues d'un bourgeonnement de la membrane vers le milieu extérieur.

B) La réponse de la membrane plasmique suite à une stimulation de la cellule : externalisation des phosphatidylsérines, bourgeonnement de la membrane et libération de microparticules vers le milieu extracellulaire.

Source : adapté de Burger and al and Hugel and al ^{169, 170}

B.1. Les microparticules induites

L'intérêt de l'utilisation des microparticules induites (MPIs) pour la recherche de biomarqueurs a été démontré ^{171, 172} lors de précédentes études menées au laboratoire. Les MPIs sont générées en utilisant

un antibiotique, l'actinomycine D, en tant qu'agent de stress de la membrane plasmique. L'enrichissement en protéines membranaires a ainsi été confirmé : 34 % de protéines issues de la membrane plasmique lors de la collecte des microparticules contre 24 % si utilisation d'un gradient de sucrose. Par ailleurs, la possibilité d'utiliser ce type de préparation membranaire dans le cadre de la recherche de biomarqueurs par protéomique a été prouvée par la mise en évidence de la surexpression du CD148 dans les lymphomes du manteau (MCL).

Forte de son succès, cette technique d'enrichissement a été une nouvelle fois utilisée pour la comparaison d'échantillons de patients atteints de lymphome (cf. Partie III – chapitre 1). De la même manière que précédemment, un enrichissement en protéines membranaires s'élevant de 25 à 30 % a été observé (cf. figure 27). Par ailleurs, un certain nombre de clusters de différenciation (CDs) ont pu être identifiés. Ces protéines, en tant que molécules de surface caractérisant la cellule, sont particulièrement importantes dans le cadre de la recherche de biomarqueurs. Depuis le Human leucocyte differentiation workshop (HLDA) de 1984¹⁷³, un anticorps a été produit pour l'ensemble de ces protéines, permettant ainsi une validation rapide par des techniques de routine utilisées en milieu hospitalier (cytométrie en flux, tests Elisa).

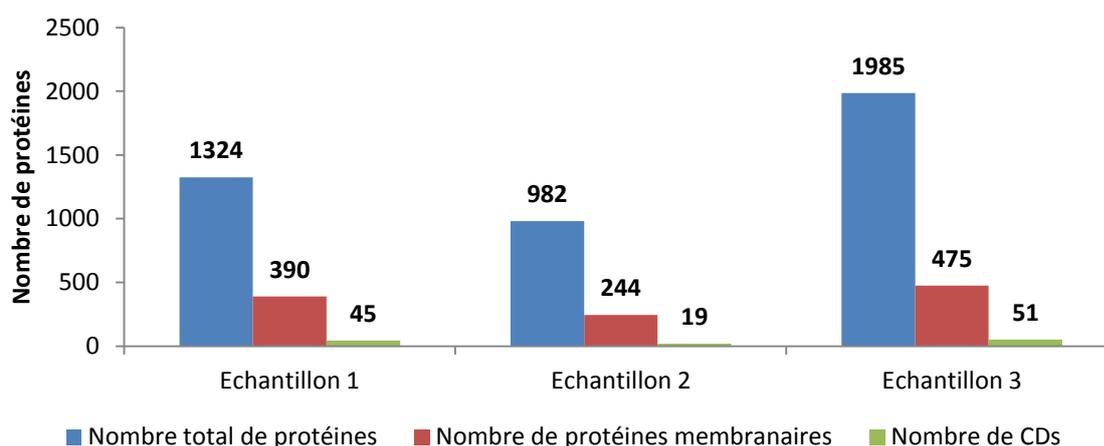


Figure 27

Nombre total de protéines, nombre de protéines membranaires et nombre de CDs identifiés dans les trois échantillons issus de patients atteints de trois types de lymphomes différents

Cependant, cette technique n'est pas aisée à mettre en place, notamment pour un nombre important d'échantillons, car il faut réaliser des cultures cellulaires ainsi que de multiples centrifugations. Par ailleurs, l'impossibilité de préparer des microparticules induites à partir d'échantillons témoins reste un inconvénient majeur. La préparation nécessite en effet un nombre important de lymphocytes B (200×10^6) que ne possède pas un échantillon témoin.

B.2. Les microparticules circulantes

Afin de s'affranchir de ces problèmes, un travail sur les microparticules circulantes (MPCs) a été entrepris. Ces MPCs générées par diverses cellules (principalement les plaquettes, les cellules endothéliales et les cellules sanguines) circulent dans tous les plasmas humains et peuvent être très

facilement collectées¹⁷⁴. Par ailleurs, plusieurs études récemment parues ont décrit le potentiel des MPCs en tant que marqueurs de différentes pathologies, en particulier dans le système vasculaire (inflammation, coagulation, maladies cardiaques...) ¹⁷⁵⁻¹⁷⁹. Un lien a même été suggéré entre un nombre élevé de microparticules et la sévérité de la maladie. Par ailleurs, des microparticules générées à partir de cellules tumorales ont été décrites et il a été possible de montrer qu'elles portaient des marqueurs spécifiques de leur cellule d'origine^{167, 180-184}.

Les microparticules circulantes sont obtenues à partir d'un plasma pauvre en plaquettes par centrifugations et lavages successifs au PBS (cf. figure 28).

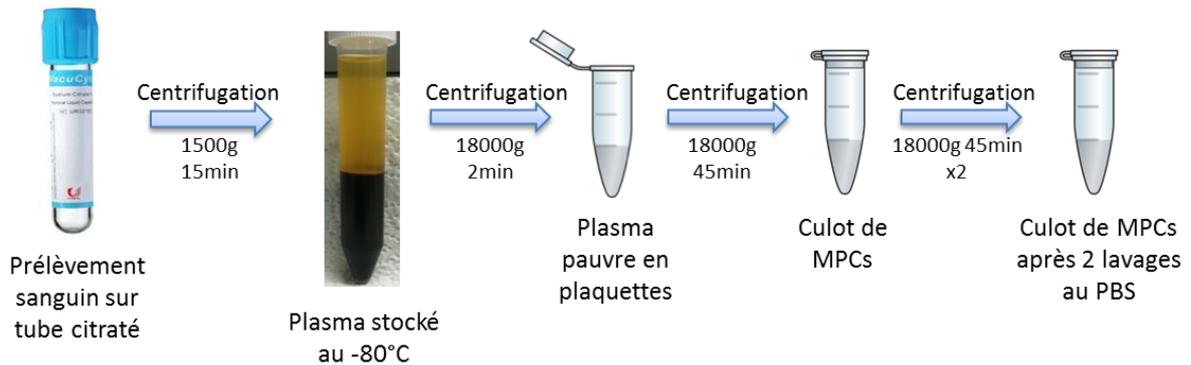


Figure 28

Schéma des différentes étapes permettant d'obtenir un culot de microparticules à partir du prélèvement sanguin

Cette préparation a été appliquée à des échantillons provenant de patients atteints de lymphome. La difficulté majeure était de détecter des MPCs de cellules sanguines. En effet, la plupart des cellules circulantes viennent des plaquettes et une petite partie a pour origine les cellules endothéliales, les lymphocytes ou les érythrocytes.

- **Établissement du protocole : détermination du nombre de centrifugations à réaliser**

Dans un premier temps, le nombre de lavages nécessaires afin d'éliminer le maximum de contaminants plasmatiques a été évalué. À partir d'un échantillon de plasma pauvre en plaquettes, trois centrifugations et lavages successifs des microparticules ont été effectués. La figure 29 présente un gel réalisé en déposant les lavages successifs (piste 2 à 4) ainsi que l'échantillon de microparticules (piste 1).

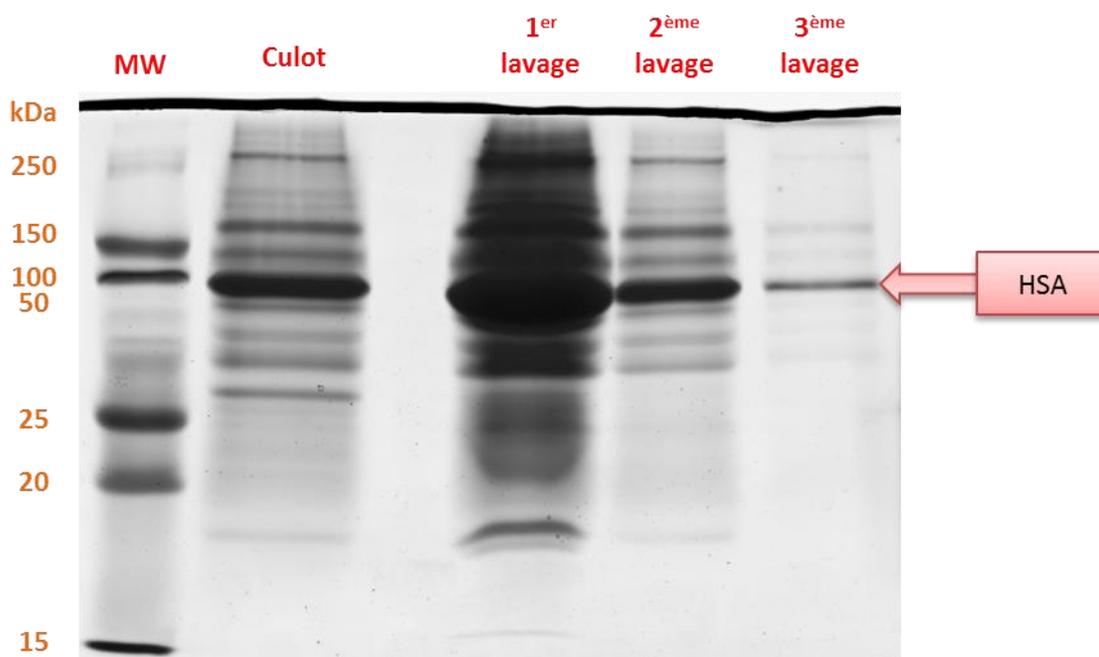


Figure 29
Gel 1D SDS PAGE de MPCs

Piste 1 : culot de microparticules après trois lavages ; piste 2 : surnageant issu du lavage 1 ; piste 3 : surnageant issu du lavage 2 ; piste 4 : surnageant issu du lavage 3.

Au vu de la coloration du gel, les lavages 1 et 2 s'avèrent indispensables alors que le troisième ne l'est pas. En effet, le taux d'HSA (composant majeur du plasma) ne diminue que très légèrement. Par ailleurs, une étude¹⁸⁵ récemment publiée décrit la réalisation de dix lavages au PBS des MPCs. Au dixième lavage, une bande correspondant à l'HSA est toujours visible sur le gel. L'étude conclut sur une possibilité d'interaction entre l'HSA et les MPCs. Pour la suite, deux lavages au PBS seront effectués. La figure 3 décrit les différentes étapes du processus.

- **Répétabilité de la collecte de microparticules plasmatiques**

Dans un deuxième temps, la répétabilité de la collecte des MPCs a été évaluée par deux techniques :

- Un comptage des MPCs par cytométrie en flux en utilisant un marquage annexin V qui cible les phosphatidylsérines exprimées à la surface des cellules.
- Un dosage de Bradford pour évaluer la quantité de protéines présentes dans l'échantillon.

L'ensemble du processus a été répété cinq fois sur le même échantillon issu d'un témoin (du prélèvement sanguin en tube citraté au culot de MPCs). Trois des culots de microparticules ont été mélangés puis redivisés en trois pour évaluer en parallèle la répétabilité du marquage des microparticules et de la détection par cytométrie en flux (cf. figure 30).

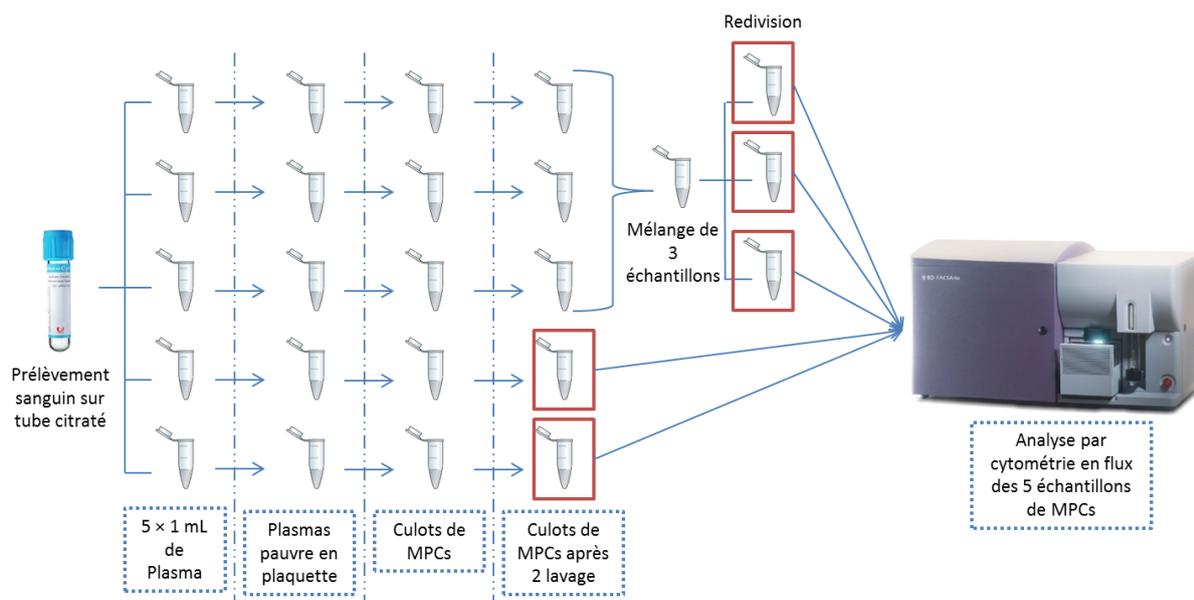


Figure 30

Schéma de l'étude de répétabilité menée sur la collecte des microparticules à partir d'un prélèvement sanguin réalisé sur un échantillon témoin

Les résultats sont présentés dans la figure 31. Un coefficient de variation (CV) de 4 % a été déterminé pour la répétabilité du marquage (3 marquages annexins V différents) et de 9 % pour la répétabilité des centrifugations (exécution du processus complet 3 fois du tube sanguin à la collecte des MPCs). Ces résultats attestent la répétabilité de la collecte des MPCs.

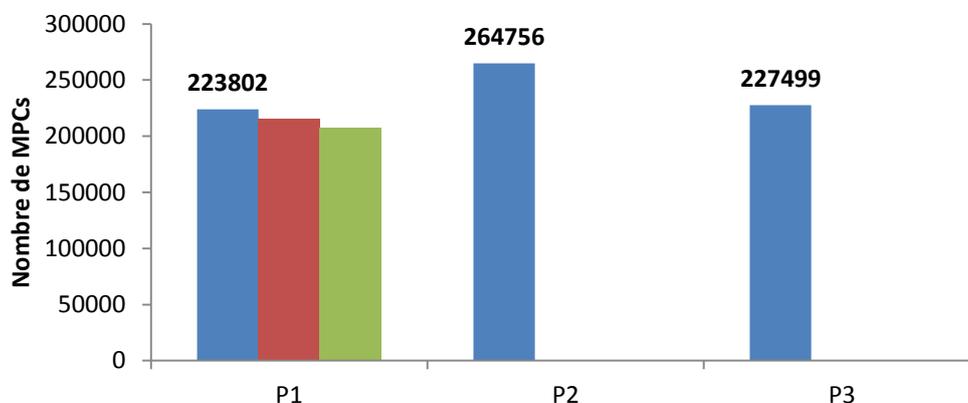


Figure 31

Résultats présentant le nombre de MPCs détectées par cytométrie en flux à la suite de réplicats de marquage et de réplicats de procédure

Un dosage Bradford a été effectué en parallèle sur chaque échantillon (P1, P2 et P3) avec une dilution au cinquième et au dixième. La quantité de protéines présentes a été estimée grâce à une droite d'étalonnage réalisée à partir de quantités croissantes de la protéine BSA (sérum d'albumine bovine). Les résultats, présentés dans le tableau 5, montrent une variabilité importante (coefficient de variation > 31 %). Ceci peut s'expliquer par la présence de protéines majeures (HSA, IGG...) couvrant l'intégralité du signal UV visible. La méthode Bradford dose en

effet la quantité totale de protéines et non spécifiquement la quantité de protéines contenues dans les microparticules.

	P1		P2		P3	
Dilution	dil 1/10	dil 1/5	dil 1/10	dil 1/5	dil 1/10	dil 1/5
Quantité de protéines (µg)	13	12	18	18	24	27

Tableau 5

Quantité de protéines déterminées pour chaque échantillon (P1, P2 et P3) par méthode Bradford

Deux dilutions différentes ont été effectuées : 1/10 et 1/5.

- **Variabilité biologique**

Une étude a été menée afin d'établir la variabilité biologique d'un échantillon à un autre au niveau du nombre de microparticules circulantes détectées dans le plasma. De la même manière que précédemment, un marquage à l'annexin V a été réalisé puis les intensités de fluorescence ont été évaluées par cytométrie en flux. Cinq échantillons de plasmas issus de témoins différents ont été comparés. La figure 32 montre une importante variabilité d'un témoin à l'autre.

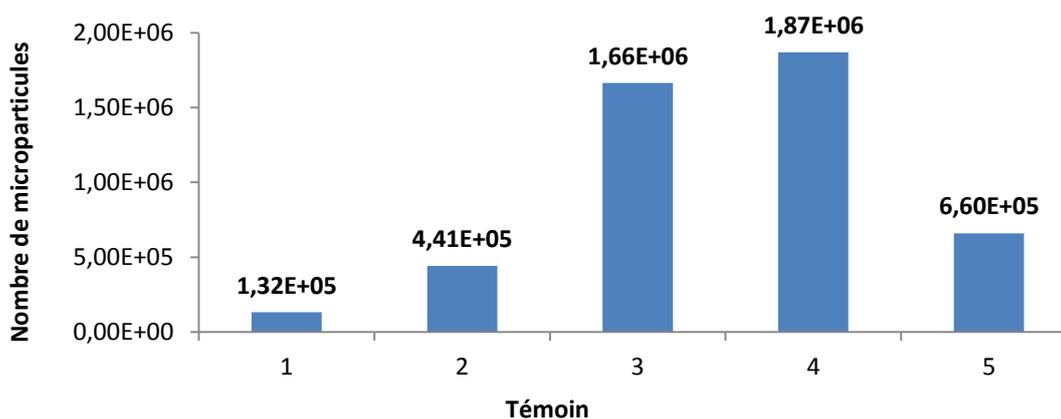


Figure 32

Nombre de microparticules circulantes identifiées dans chaque échantillon de plasma témoin par cytométrie en flux

L'ensemble des résultats montre la répétabilité de la collecte des microparticules à partir du même échantillon, mais la non-répétabilité de la quantité de protéines contenues dans chaque échantillon et une variabilité biologique importante. Au vu de ces conclusions, nous avons estimé qu'il était plus judicieux de partir de 1 mL de plasma pour chaque échantillon et ainsi s'affranchir du problème des protéines contaminantes et de la variabilité biologique. En parallèle, un comptage des MPCs par cytométrie en flux sera effectué sur chaque échantillon.

- **Étude de la composition des microparticules circulantes**

Des microparticules circulantes ont été collectées à partir du plasma de cinq témoins. Les échantillons ont ensuite été injectés en spectrométrie de masse afin d'identifier les protéines et les familles de protéines contenues dans chaque échantillon à l'aide des annotations Gene Ontology. Pour les cinq témoins, la répartition des annotations de localisation est similaire (cf. figure 33). Il est intéressant de noter la forte proportion de protéines membranaires, de 34 à

50 %. Dans un deuxième temps, les clusters de différenciation ont été étudiés et classés suivant leurs cellules d'origine (cf. figure 34). Malgré une forte proportion d'entités provenant des plaquettes, il a été possible de détecter des clusters de différenciation provenant de diverses cellules et notamment des leucocytes. Plus particulièrement, le CD45, marqueur spécifique des lymphocytes B, a pu être identifié dans 4 échantillons témoin sur 5.

L'ensemble de ces résultats atteste la présence de microparticules dans l'échantillon, et plus particulièrement de microparticules d'origine leucocytaire. L'étude de ces entités apparaît donc intéressante dans le cadre de la recherche de biomarqueurs de lymphomes.

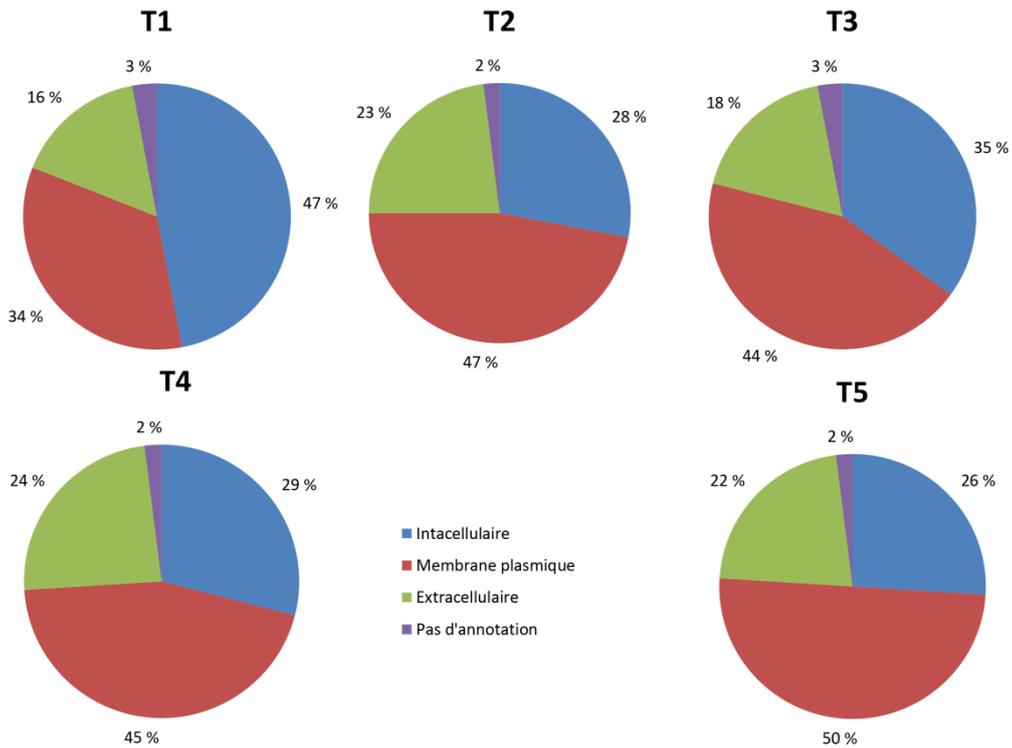


Figure 33 Répartition des annotations de localisation issues de Gene Ontology pour cinq témoins différents

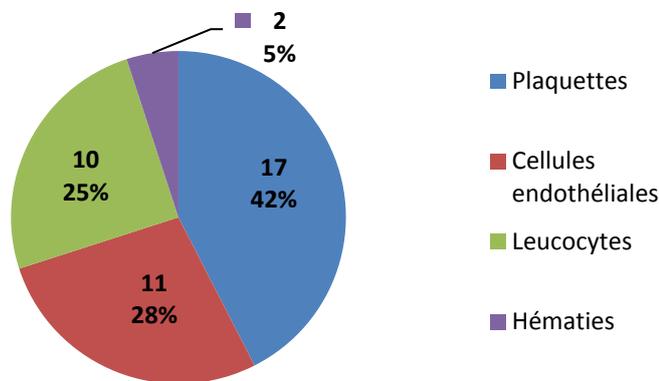


Figure 34 Répartition des cellules d'origine des clusters de différenciation

B.3. Préparation de « *ghosts* membranaires »

Au cours d'un projet d'étude du protéome membranaire de lignées de cellules souches cancéreuses (CSC) issues de glioblastomes, un autre type de préparation membranaire a été évalué. En effet, comme il ne nous a pas été possible de générer des microparticules induites à partir des CSCs, nous avons formé des *ghosts* membranaires (fantômes membranaires) par lyse mécanique des cellules. Les débris cellulaires et les protéines cytoplasmiques ont ensuite été éliminés par centrifugation. Cette technique, bien que ne permettant pas l'élimination de la majeure partie des organelles, permet la conservation des complexes associés à la membrane (cf. figure 35).

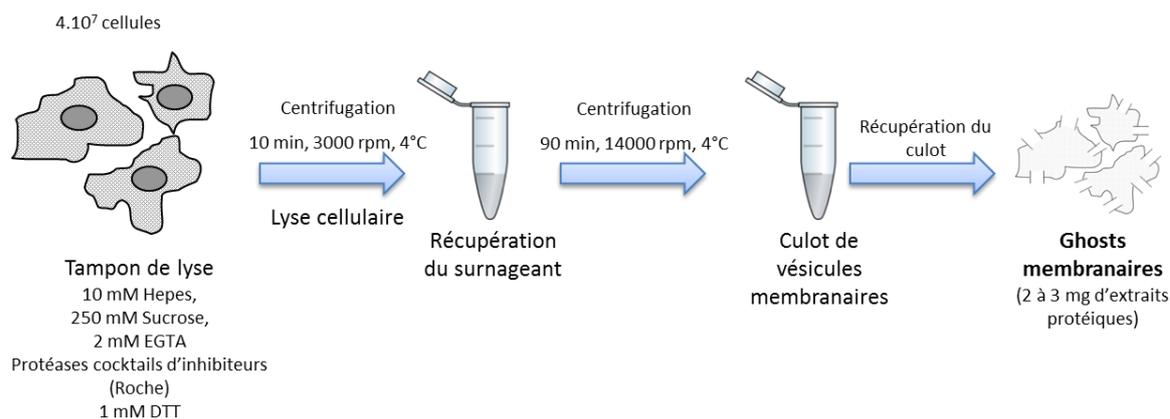


Figure 35

Schéma de l'obtention des *ghosts* membranaires

4.10⁷ cellules sont lysées avec un tampon de lyse puis centrifugées, le surnageant est récupéré pour une nouvelle centrifugation. Un culot de vésicules membranaires est alors obtenu, contenant 2 à 3 mg d'extraits protéiques.

Source : adapté de la thèse du Dr Émilie Audran, Université de Strasbourg¹⁸⁶

Cette préparation a été utilisée sur sept lignées cellulaires différentes : NSC (cellules souches neurales), HA (astrocytes humains), U87 (cellules cancéreuses issues de glioblastomes), TG01, TG10, OB1 et TG16 (cellules souches cancéreuses issues de cultures de biopsies de patients atteints de glioblastomes).

Dans ces préparations, des protéines appartenant aux complexes endomembranaires (noyau, réticulum endoplasmique, ribosomes, golgi, lysozymes, vésicules de sécrétion et membranes plasmiques) ainsi que des complexes associés à la membrane plasmique ont pu être identifiés. La figure 36 montre la répartition des protéines identifiées. Pour l'ensemble des lignées cellulaires, la répartition est similaire. En moyenne, 24 % de protéines membranaires ont été détectées, 20 % issues des mitochondries et 22 % du noyau. Seulement 12 % de protéines annotées « cytoplasme » ont été identifiées.

La proportion de protéines membranaires obtenue est globalement plus faible que celle obtenue suite à la génération de microparticules induites. Cependant, ces résultats concordent avec les enrichissements membranaires moyens observés dans la littérature¹⁷².

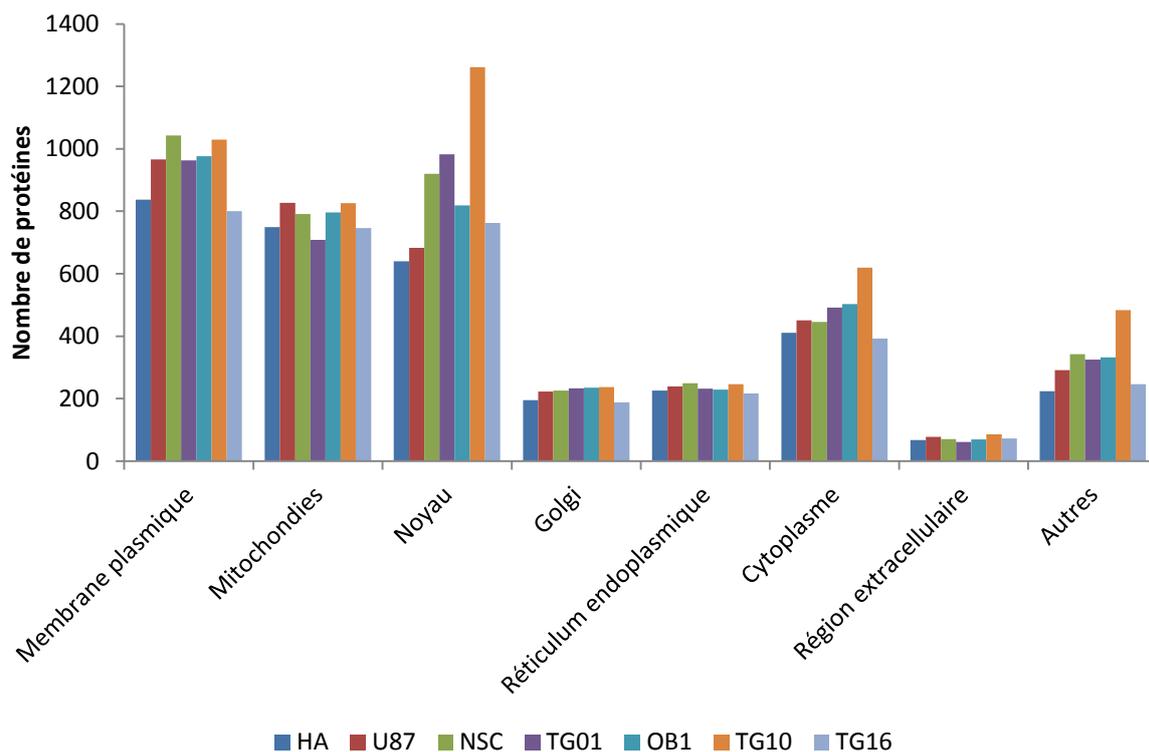


Figure 36

Répartition des protéines en fonction de leurs annotations Gene Ontology pour les lignées cellulaires étudiées

C. Conclusions

Un enrichissement en protéines membranaires a été observé dans l'ensemble des préparations testées, avec une plus forte proportion lors de la génération ou la collecte de microparticules. Cependant, chaque technique présente ses avantages et ses inconvénients. Avant de choisir une méthode, il est important de prendre en compte plusieurs paramètres tels que le type et le nombre d'échantillons. Par exemple, il n'est pas envisageable de générer des microparticules induites pour les 125 échantillons du projet décrit dans la Partie III – chapitre 3.

CHAPITRE 2

Développement de méthodes nanoLC-MS/MS pour l'analyse de protéines membranaires

L'échec de la recherche de biomarqueurs par spectrométrie de masse a grandement affecté la perception de l'analyse protéomique dans ce domaine. Ainsi, en 2002, une étude majeure publiée dans *The Lancet* a présenté des résultats qui n'ont jamais pu être validés¹⁸⁷⁻¹⁹⁰. Une étude inter-laboratoires a été réalisée¹⁹¹ plus récemment, en 2009 : un échantillon contenant 20 protéines a été remis à 27 laboratoires différents dont seuls 7 ont pu identifier correctement les protéines.

Depuis, la communauté protéomique a œuvré pour comprendre les raisons liées au manque d'homogénéité des résultats. Plusieurs points concernant l'étape d'analyse par nanoLC-MS/MS ont pu être mis en évidence¹⁹² :

- De petites différences au niveau de la séparation des peptides par chromatographie liquide peuvent changer l'ordre d'éluion de ces derniers ou altérer leur sélection pour la fragmentation MS/MS.
- Le sous-échantillonnage lié au mode d'acquisition DDA (*data dependant acquisition*) privilégie la sélection des peptides les plus intenses pour la fragmentation, tandis que les autres ne sont pas sélectionnés. Par ailleurs, les ions élués pendant la durée de l'acquisition MS/MS ne sont pas analysés (cf. figure 37).

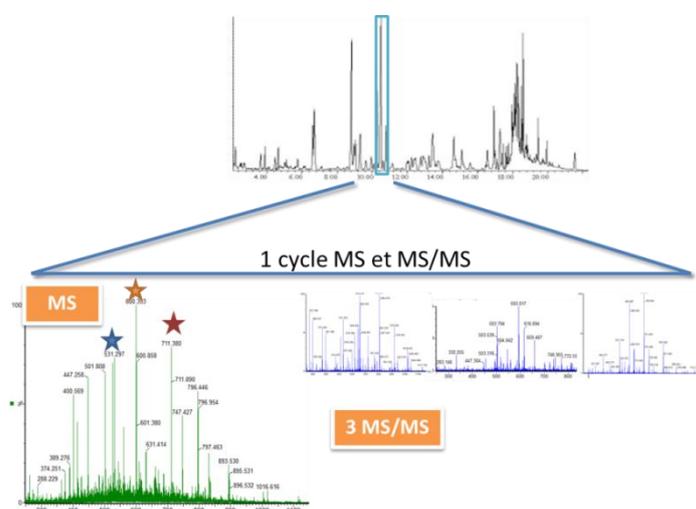


Figure 37
Origine du sous-échantillonnage lors d'acquisitions en mode DDA

L'objectif, lors de la recherche de biomarqueurs, est d'obtenir le protéome le plus complet de chaque échantillon, car les protéines d'intérêt sont souvent celles qui sont exprimées à un niveau plus faible. Des développements méthodologiques se sont avérés nécessaires :

- Optimisation fine des paramètres d'acquisition du spectromètre de masse utilisé. Au cours de mon travail de thèse, j'ai eu l'occasion de travailler sur divers couplages nanoLC-MS/MS : nanoAcquity – Synapt G1 (Chromatographie et Q-TOF Waters), nanoChip – AmaZon (système chromatographique Agilent couplé à la trappe ionique de Bruker) et nanoAcquity – Maxis Impact (système chromatographique Waters couplé à un Q-TOF Bruker). J'ai été chargée de développer l'analyse haut débit de mélanges complexes séparés sur bandes de gels 1D sur la trappe ionique AmaZon qui était jusqu'à présent utilisée pour des analyses de mélanges peu complexes tels que des spots de gel 2D.
- Évaluation de la répétabilité des analyses et mise en œuvre de solutions pour identifier le maximum de protéines d'un échantillon.
- Évaluation d'un nouveau mode d'acquisition, la DIA (*data independent acquisition*), associée à une séparation en mobilité ionique, la LC-HDMSE (nanoAcquity-Synapt G2, Waters).

A. De l'importance d'une optimisation fine des paramètres LC-MS/MS

Les travaux présentés dans ce paragraphe ont été réalisés sur le couplage nanoHPLC-Chip/MS (Agilent) – trappe ionique AmaZon (Bruker Daltonics).

Plusieurs paramètres clés de la chromatographie et du spectromètre de masse ont été finement optimisés afin d'obtenir le protéome le plus complet possible pour l'analyse de bandes de gel 1D SDS PAGE en utilisant des acquisitions en mode DDA :

- le gradient de chromatographie ;
- le nombre et type d'ions sélectionnés pour la MS/MS ;
- la vitesse de balayage ;
- le nombre d'ions accumulés dans la trappe pour la fragmentation ;
- le seuil de sélection du précurseur pour la fragmentation.

Les tests présentés dans cette partie ont été réalisés dans le cadre du travail sur les cellules souches cancéreuses et sont à ce titre issus de lysats cellulaires enrichis en protéines membranaires par la méthode des « ghosts membranaires » et déposés sur un gel 1D-SDS PAGE (cf. Partie III – chapitre 2).

A.1. Choix d'un gradient adapté aux échantillons

La trappe AmaZon est couplée à un système nanoHPLC-Chip/MS (Chip) commercialisé par Agilent Technologies et composé d'une puce microfluidique comportant la précolonne, la colonne et l'aiguille

nanospray ainsi qu'un cube permettant d'interfacer la chromatographie liquide au spectromètre de masse (cf. figure 38).

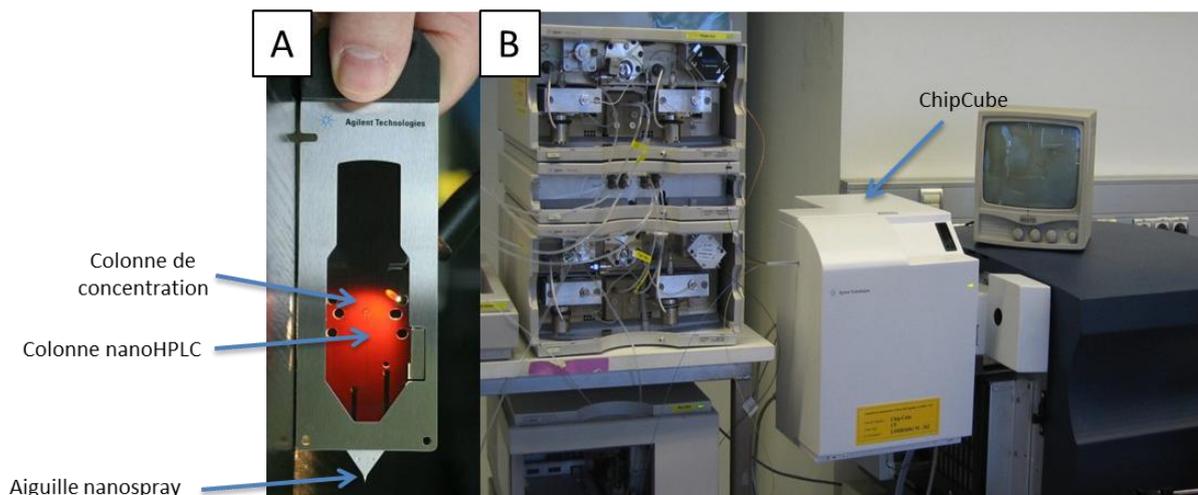


Figure 38

Configuration de l'interface nanoHPLC-Chip/MS

La puce microfluidique contenant la précolonne et la colonne nanoHPLC ainsi que l'aiguille nanospray (A) est reliée à l'HPLC et au spectromètre de masse via le ChipCube (B).

La puce microfluidique présente un avantage majeur : les volumes sont largement réduits et permettent ainsi d'appliquer plus rapidement le gradient (2 min au lieu de 12 min sur une nanoHPLC classique) et de réduire le temps de rééquilibrage (5 min au lieu de 20 min sur une nanoHPLC classique)^{193, 194}.

Les chips utilisées sont commercialisées par la société Agilent et contiennent des colonnes analytiques de 150 mm de long et 75 μm de diamètre avec une phase stationnaire de type ZORBAX 300SB-C18 (particules de 5 μm) et une précolonne de 40 nl. Après chargement sur la colonne, les composés sont élués de la précolonne et de la colonne avec un gradient eau/acétonitrile.

De manière générale, les peptides tryptiques sont élués entre 2 et 40 % en acétonitrile. Plusieurs longueurs de gradient ont été testées : 30, 40 et 60 min dans un premier temps, puis 60, 75 et 90 min dans un deuxième temps. Les résultats présentés en figure 39 montrent le nombre de peptides identifiés dans 3 bandes différentes issues du gel. Un filtre a été appliqué sur le score Mascot ($s > 20$) pour attester la qualité de l'identification. Ce nombre augmente avec le temps de gradient. Le gradient de 90 min semble donc être le plus approprié pour les analyses à réaliser.

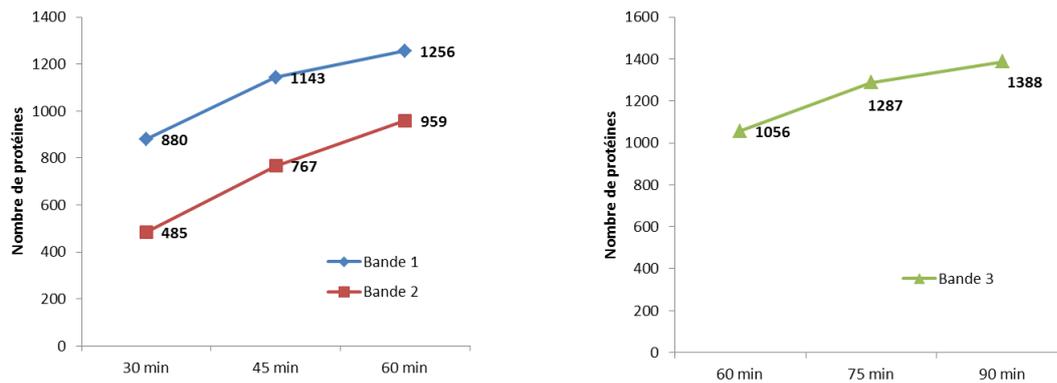


Figure 39

Nombre de peptides avec un score Mascot supérieur à 20 identifiés lors du test de plusieurs gradients sur 3 bandes provenant d'un gel 1D SDS PAGE

Cependant, dans le contexte de ce projet, un gel de 20 cm avec découpe systématique des bandes tous les 2 mm a été réalisé, ce qui représente 90 bandes à analyser avec un gradient de 90 min soit une semaine complète d'analyse pour un échantillon. Il a donc été décidé de travailler sur un gradient à deux paliers afin de réduire le temps d'analyse tout en conservant une longue séparation sur les zones désirées. La figure 40 montre le nombre de peptides avec des scores Mascot supérieurs à 20 élués de la colonne à chaque instant pour un gradient de 90 min. Ainsi, si beaucoup de peptides sont identifiés en début de gradient, peu le sont à la fin. Pour réduire le temps d'analyse et optimiser l'efficacité de la séparation, un gradient de 61 min à deux paliers a été utilisé. La figure 41 montre la séparation obtenue en utilisant le nouveau gradient par rapport au précédent. Un gain en sensibilité est observé en fin de gradient, avec des intensités plus importantes. Par ailleurs, aucune différence majeure n'est observée entre les deux gradients au niveau du nombre total de peptides identifiés (1348 peptides avec le gradient de 90 min et 1467 avec le gradient de 61 min). Pour la suite, le gradient de 61 min à deux paliers sera donc conservé.

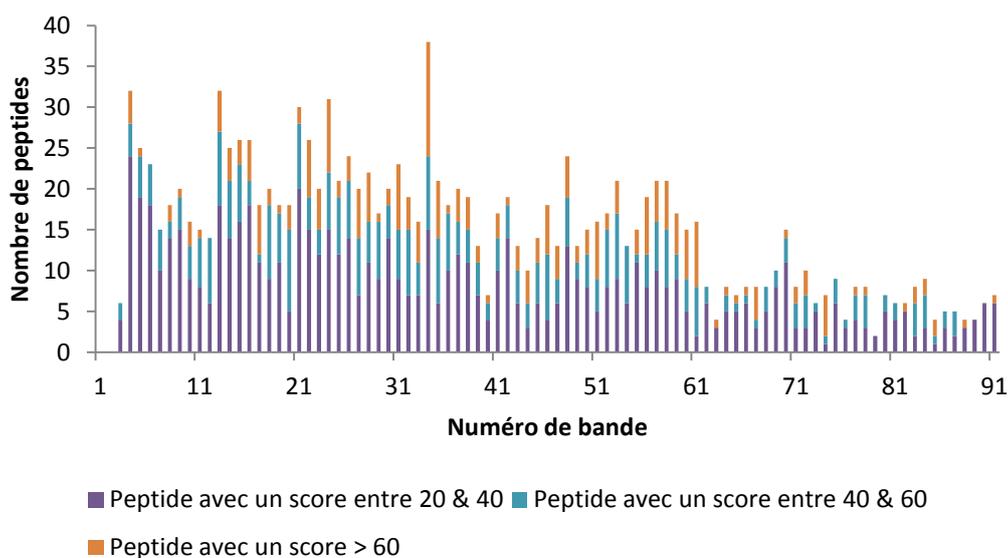


Figure 40

Nombre de peptides identifiés suivant le temps de gradient et le score Mascot

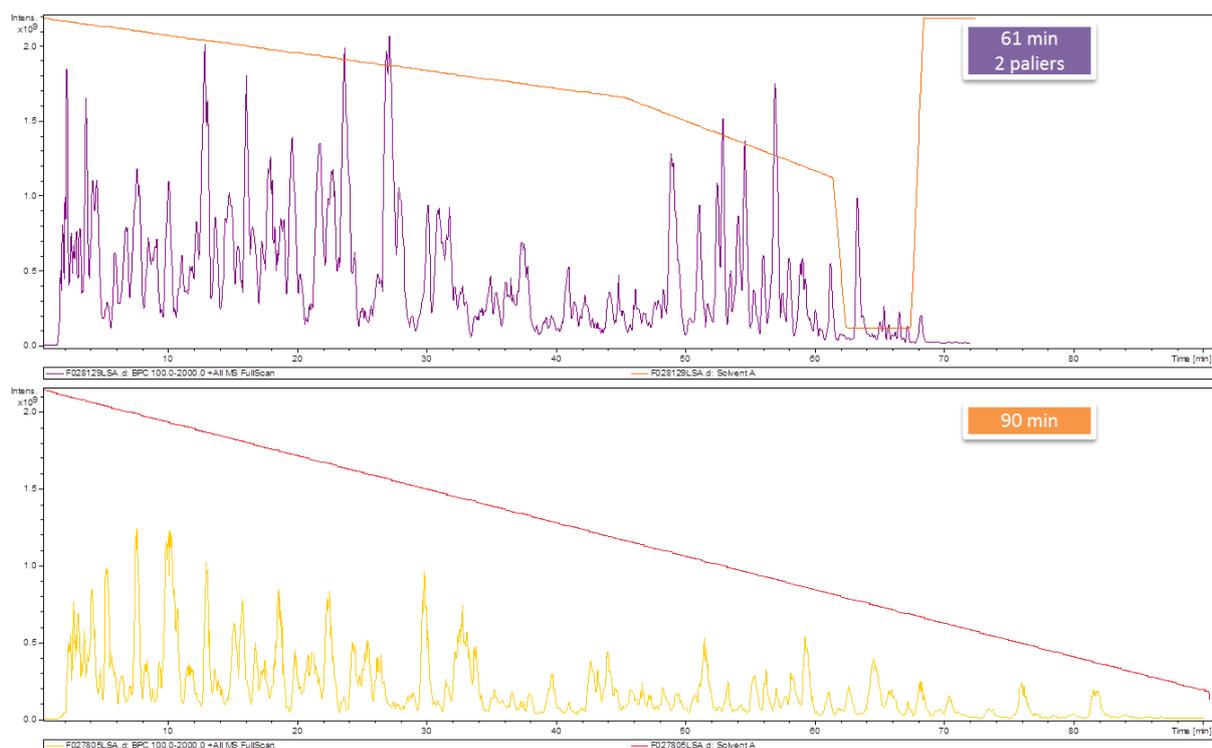


Figure 41

Trace chromatographique obtenue suivant le gradient utilisé

61 min avec deux paliers ou 90 min un gradient linéaire. La trace rouge montre la variation du solvant A (% H₂O).

A.2. Paramètres MS et MS/MS

Avec le gradient choisi, la largeur des pics chromatographiques est de 0,5 à 1 min à la base. Pour définir correctement un pic chromatographique, il faut environ 10 points. Le temps de cycle MS-MS/MS sera donc idéalement compris entre 3 et 6 s.

La figure 42 montre l'organisation des cycles MS et MS/MS sur la trappe ionique. Plusieurs paramètres sont à prendre en compte pour obtenir une durée de cycle optimale, c'est-à-dire suffisamment longue pour que la MS et la MS/MS soient informatives et suffisamment courte pour ne pas perdre d'informations.

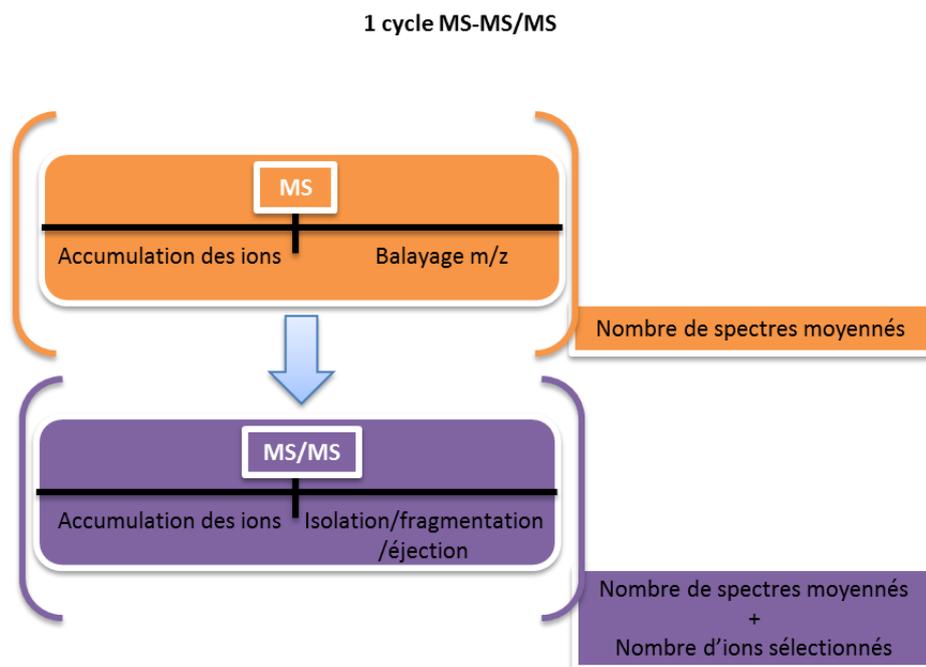


Figure 42

Représentation d'un cycle MS-MS/MS sur la trappe ionique

Pendant la séquence MS, les ions sont accumulés dans la trappe, puis balayés en m/z par l'analyseur. Il est possible de moyenniser plusieurs spectres pour augmenter le ratio signal/bruit. Pendant la séquence MS/MS, les ions sont accumulés dans la trappe puis isolés, fragmentés et éjectés. De la même manière, plusieurs spectres peuvent être moyennés. En général, plusieurs ions sont sélectionnés pour la MS/MS sur un même scan MS.

Les paramètres suivants ont été étudiés dans l'optique de réaliser un maximum d'identifications :

- Nombre et type d'ions sélectionnés
- Accumulation des ions dans la trappe
- Mode de balayage
- Seuils de sélection

A.2.1. Nombre d'ions sélectionnés

Le même échantillon a été injecté suivant trois conditions : sélection de 6, 8 ou 10 ions. Trois réplicats d'injection ont été réalisés pour chaque condition.

Comme le montre la figure 43, l'augmentation du nombre d'ions sélectionnés pour la MS/MS augmente le nombre de spectres réalisés mais n'a pas de réelle influence sur le nombre de spectres MS/MS assignés. De manière arbitraire, le paramètre 8 ions sera conservé pour la suite.

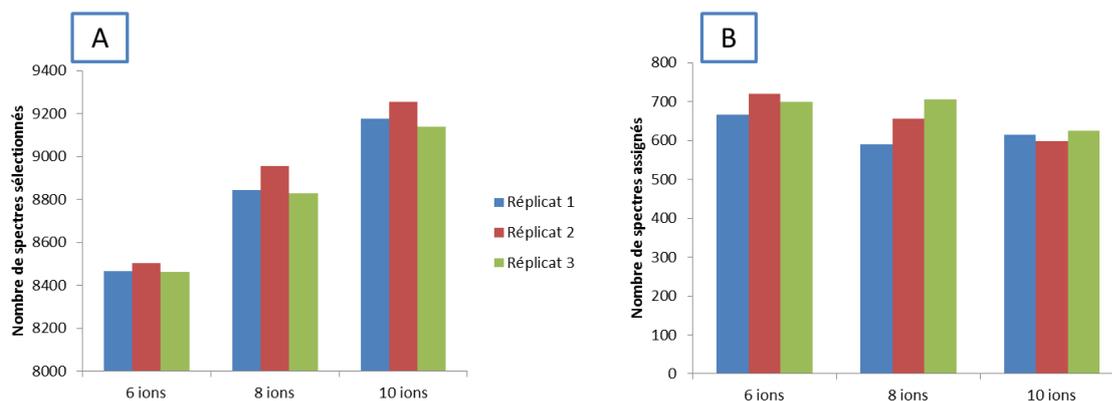


Figure 43

Résultats obtenus lors de l'injection en triplicat d'une même bande de gel suivant trois paramètres

Sélection de 6, 8 et 10 ions pour la MS/MS. Nombre de spectres réalisés (A) et nombre de spectres assignés avec un score Mascot supérieur à 20 (B).

En outre, il est possible de choisir un état de charge préférentiel pour la sélection des ions dans les paramètres des méthodes : 2+, > 2+ ou pas de sélection préférentielle. Une rapide étude (cf. figure 44) a montré que les ions monochargés sélectionnés ne permettent pas de réaliser d'identifications et que les ions deux fois chargés ont plus de chance de donner une identification que les ions trois fois chargés. La méthode utilisera donc la sélection préférentielle des ions 2+ ainsi que l'exclusion des ions 1+ dans le cas de l'analyse de peptides tryptiques.

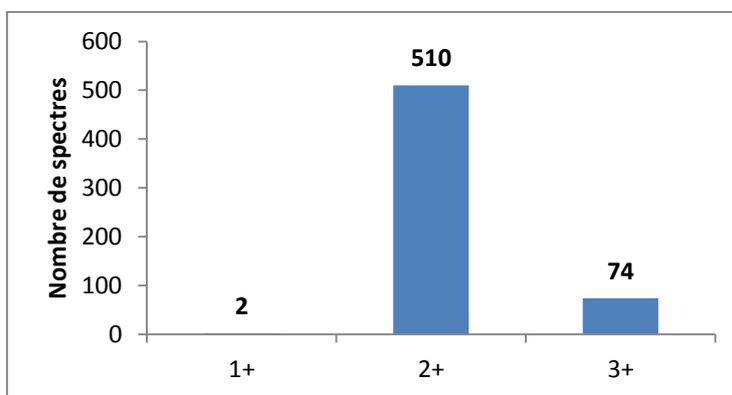


Figure 44

Nombre de spectres assignés avec un score Mascot supérieur à 20 suivant la charge de l'ion sélectionné pour la fragmentation

Une exclusion dynamique sur une fenêtre de 2 à 3 Da a été paramétrée afin d'éviter une multiple sélection d'ions très abondants. Celle-ci a été fixée à 0,4 min, permettant deux sélections de l'ion sur un même pic chromatographique ou la sélection d'un ion qui aurait la même masse et le même temps de rétention.

A.2.2. Balayage des ions

Sur la trappe ionique AmaZon, il est possible d'utiliser quatre modes de balayage :

- XtremeScan : 52000 (m/z)/s (mode non utilisable en auto MS/MS)
- UltraScan : 32500 (m/z)/s
- Enhanced Resolution : 8100 (m/z)/s
- Maximum Resolution : 4650 (m/z)/s

La vitesse de balayage influe sur la sensibilité et sur la résolution. Une vitesse de scan plus élevée occasionne une perte de résolution due à une éjection plus rapide et donc moins spécifique des ions et un gain en sensibilité dû à une diminution des phénomènes d'espace charge¹⁹⁵⁻¹⁹⁷.

Lors d'une précédente thèse¹⁹⁸, les différents niveaux de résolution spectrale ont été évalués en suivant deux peptides issus d'un mélange, l'un doublement chargé et l'autre triplement chargé. La figure 45 montre une perte de résolution avec une augmentation de la vitesse de balayage ainsi qu'un gain en sensibilité.

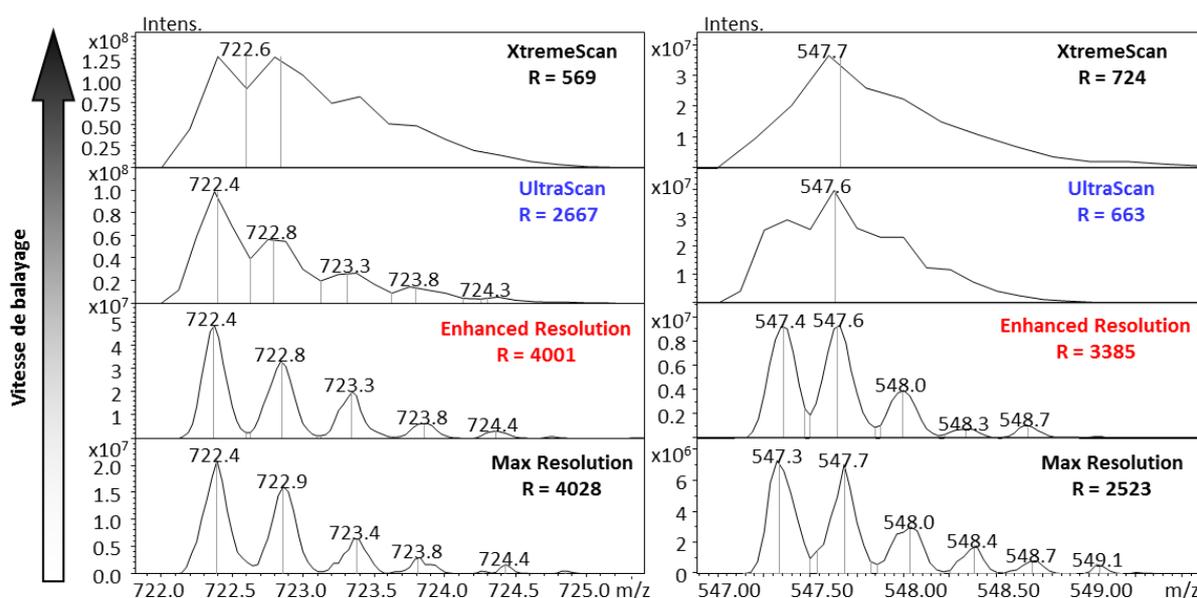


Figure 45

Suivi de la résolution et de l'intensité de deux peptides, l'un doublement chargé et l'autre triplement chargé, suivant le mode de balayage utilisé

Source : adaptée de la thèse du Dr Agnès Hovasse, Université de Strasbourg¹⁹⁸

Lors du scan MS, un maximum de résolution est requis afin de pouvoir déconvoluer l'ion de façon optimale. Le mode le plus adapté est donc le mode *Enhanced Resolution*. À noter que le mode *Max Resolution*, à cause d'une vitesse de balayage plus faible, ne permet pas une augmentation de résolution significative. La MS/MS était à l'origine paramétrée en *UltraScan*. Un test a été effectué afin de vérifier que ce mode est bien le plus pertinent. Un même échantillon a été injecté plusieurs fois en mode *UltraScan* et en mode *Enhanced Resolution*. La figure 46 montre un gain en ions assignés avec un score supérieur à 20 en *UltraScan*. Ce mode sera conservé par la suite.

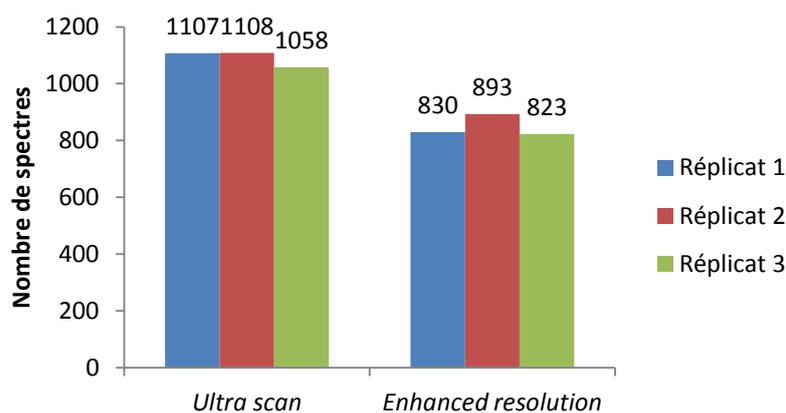


Figure 46
Nombre de spectres identifiés avec un score Mascot supérieur à 20 suivant le mode de balayage utilisé pour la MS/MS (UltraScan ou Enhanced Resolution)

Par ailleurs, la gamme de masse permettant une sélection optimale des ions pour l'identification a été évaluée. En effet, la durée d'un scan MS dépend de la vitesse de balayage et donc directement de la gamme de masse scannée. Dans un premier temps, des acquisitions ont été réalisées sur une gamme de m/z de 250 à 2500. Les résultats présentés dans la figure 47 montrent qu'au-delà d'un m/z de 1200, aucun ion n'est identifié. La gamme de sélection m/z sera donc comprise entre 250 et 1200 en MS. En MS/MS, elle sera comprise entre 100 et 1500.

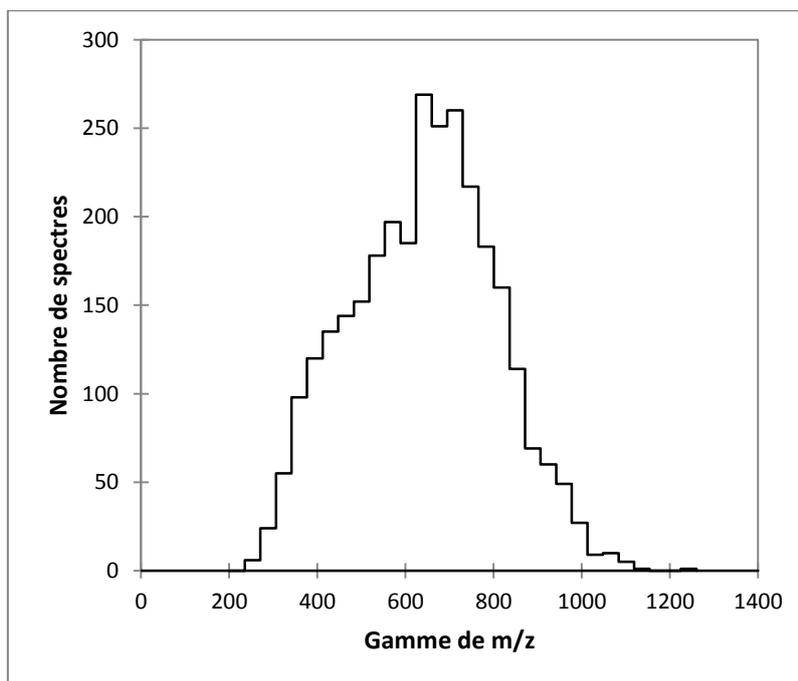


Figure 47
Nombre de spectres identifiés suivant la partie de gamme de m/z scannée

A.2.3. Accumulation des ions dans la trappe

Le temps d'accumulation des ions dans la trappe est automatiquement ajusté suivant l'ICC (*ion charge control*) paramétré, c'est-à-dire suivant le nombre d'ions accumulés dans la trappe. Notons que le temps maximum d'accumulation est fixé à 200 ms. Un ICC élevé permettra d'augmenter la sensibilité mais aussi l'effet espace charge : une accumulation trop importante des ions dans la trappe entraîne une augmentation des répulsions coulombiennes. Le piégeage est moins efficace et la précision sur la masse mesurée est altérée¹⁹⁹.

La valeur d'ICC a été maintenue à 200 000 pour la MS et testée pour plusieurs valeurs en MS/MS : 200 000, 300 000 et 400 000. La figure 48 montre les résultats obtenus. L'augmentation de la valeur d'ICC en MS/MS diminue le nombre de spectres réalisés mais augmente le nombre de spectres assignés. La valeur de 400 000 sera conservée dans la méthode.

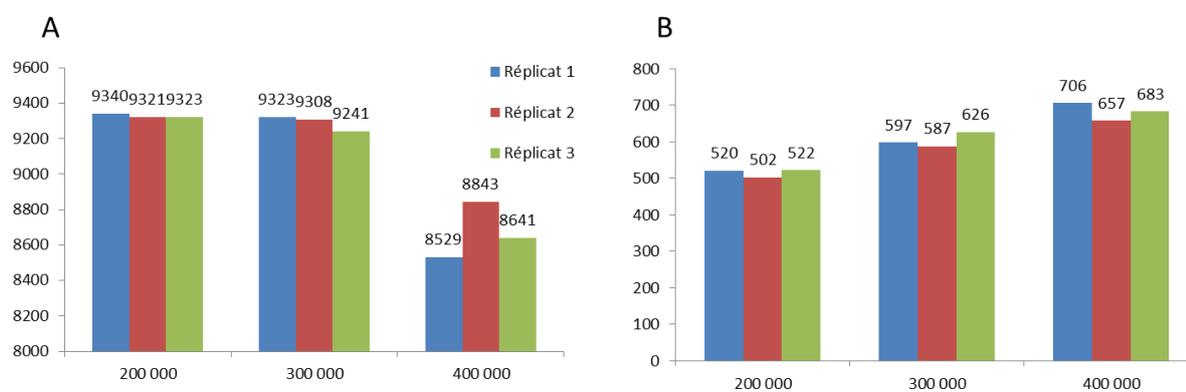


Figure 48

Tests sur différentes valeurs pour l'accumulation des ions dans la trappe (200 000, 300 000 et 400 000)

(A) Nombre de spectres totaux réalisés. (B) Nombre de spectres assignés avec un score supérieur à 20.

A.2.4. Seuil de sélection des ions

Le seuil de sélection des ions est un paramètre clé. En effet, s'il est trop bas, les ions risquent d'être sélectionnés trop tôt, c'est-à-dire dans le bruit de fond, et donc d'avoir une MS/MS de basse qualité. Au contraire s'il est trop haut, il y a un risque de ne pas sélectionner les ions peu abondants. La méthode standard à partir de laquelle a été développé le nouveau protocole utilisait des paramètres non adaptés pour les échantillons complexes. Les ions étaient sélectionnés à partir d'un seuil relatif de 5 % par rapport à l'ion le plus intense. Si des peptides très abondants sont présents dans l'échantillon, cela n'entraîne la sélection que d'un seul type d'ions au cours d'un pic chromatographique. Ce concept est illustré par la figure 49, qui présente le nombre d'ions sélectionnés à chaque minute du gradient. Ainsi, on peut noter que la présence d'un ion abondant occasionne une rupture dans les effectifs de sélection. Afin de pallier ce problème, un seuil de sélection absolu a été fixé en évaluant l'intensité du bruit de fond. Sa valeur est de 600 000.

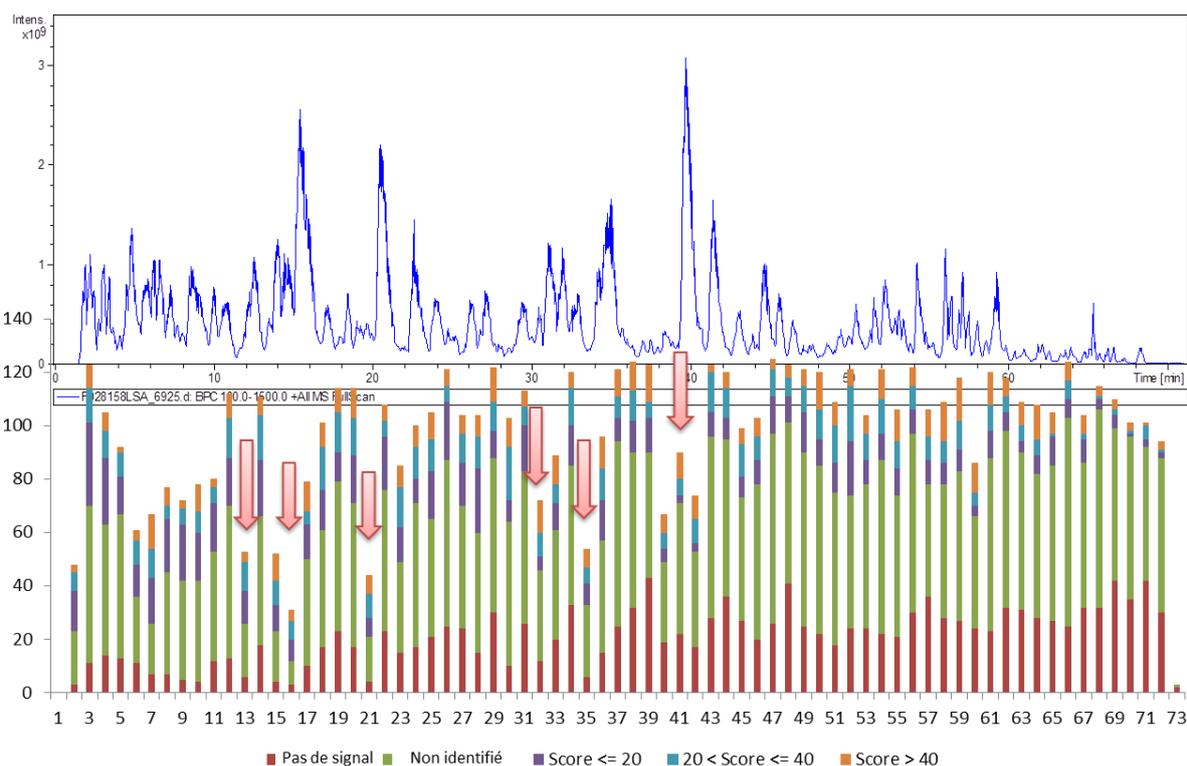


Figure 49

Présentation du nombre d'ions sélectionnés à chaque temps du gradient (pas de 1 min) en fonction de la qualité du spectre MS/MS (évaluée par le score Mascot)

Les flèches rouges indiquent les ruptures dans le nombre d'ions sélectionnés.

A.3. Conclusion

Dans cette partie, plusieurs paramètres ont été testés afin d'évaluer leur influence sur la qualité et le nombre de spectres obtenus. Une méthode optimisée a été développée pour l'analyse haut débit de bandes de gel 1D SDS PAGE d'échantillons complexes. La durée d'un cycle peut désormais être évaluée (cf. figure 50).

$$\begin{aligned} \text{Temps de cycle MS} &= (\text{accumulation} + \text{temps de scan}) \times \text{nombre de spectres moyennés} \\ &= (0,2 + (1200 - 250)/8100) \times 2 = 0,6 \text{ s} \end{aligned}$$

$$\begin{aligned} \text{Temps de cycle MS/MS} &= (\text{accumulation} + \text{isolation} + \text{fragmentation} + \text{éjection}) \times \text{nombre de} \\ &\quad \text{spectres moyennés} \times \text{nombre d'ions sélectionnés} \\ &= (0,2 + 0,04 + (1500 - 100)/26000) \times 2 \times 8 = 4,64 \text{ s} \end{aligned}$$

Figure 50

Détermination de la durée d'un cycle MS-MS/MS suite à l'optimisation fine de l'ensemble des paramètres clés

Au maximum, le temps de cycle sera de 5,54 s, ce qui correspond aux critères fixés en début de chapitre (temps de cycle compris entre 3 et 6 s). Il est important de noter ici que le temps d'accumulation a été largement surestimé. En effet, dans la majeure partie des cas, il ne sera pas nécessaire d'accumuler jusqu'au temps maximum.

Cette partie illustre l'importance d'optimiser et d'adapter les méthodes de chromatographie, MS et MS/MS à chaque échantillon afin d'obtenir le maximum d'informations possible.

B. Approfondissement des analyses par la répétition des injections

Il a été montré qu'une optimisation fine des méthodes d'acquisition permet d'augmenter la profondeur des analyses. Cependant, la non-répétabilité reste une caractéristique inhérente à l'analyse protéomique en mode d'acquisition DDA. En effet, ce dernier implique la sélection de x ions pour la fragmentation à l'issue d'un spectre MS. Tous les ions ne sont pas sélectionnés, ce qui entraîne une perte d'informations. Réaliser plusieurs injections du même échantillon semble donc être une étape incontournable pour avoir une vision la plus significative de son protéome.

Afin de confirmer cette hypothèse, la répétabilité des identifications peptidiques et protéiques a été évaluée sur l'ensemble des couplages nanoLC-MS/MS utilisés dans le cadre de la recherche de biomarqueurs : nanoAcquity – Synapt G1 (Chromatographie et Q-TOF Waters), nanoHPLC – Chip AmaZon (système Agilent couplé à la trappe ionique de Bruker) et nanoAcquity – Maxis Impact (système Waters couplé à un Q-TOF Bruker).

Dans un deuxième temps, nous avons montré qu'une répétition avec liste d'exclusion était plus performante qu'une répétition simple.

B.1. Réplicats d'injections classiques

B.1.1. Couplage nanoAcquity – Synapt G1

Le projet de recherche de biomarqueurs du lymphome de la zone marginale, détaillé en Partie III – chapitre 1, a été réalisé sur le couplage nanoAcquity – Synapt G1. Les échantillons tests sont des microparticules générées à partir de lymphocytes B de patients atteints de différents types de lymphomes. Ces microparticules ont été séparées par un gel 1D SDS PAGE (90 bandes) et injectées trois fois chacune. Dans ce projet ont été évalués la capacité de l'instrument à répéter les injections ainsi que ses résultats en termes d'identifications.

La figure 51 présente le bilan des injections. Il apparaît clairement que les analyses ne sont pas répétables. Ainsi, sur les 1515 protéines et 8175 peptides identifiés, seuls 69 % des protéines et 61 % des peptides sont identifiés dans les trois analyses. 16 % des protéines et 21 % des peptides identifiés ne le sont qu'une fois. L'analyse nanoLC-MS/MS possède donc un caractère aléatoire. Par ailleurs, il est important de noter que les peptides très abondants vont être systématiquement identifiés alors que les moins abondants ne seront pas forcément sélectionnés pour une analyse MS/MS en mode DDA.

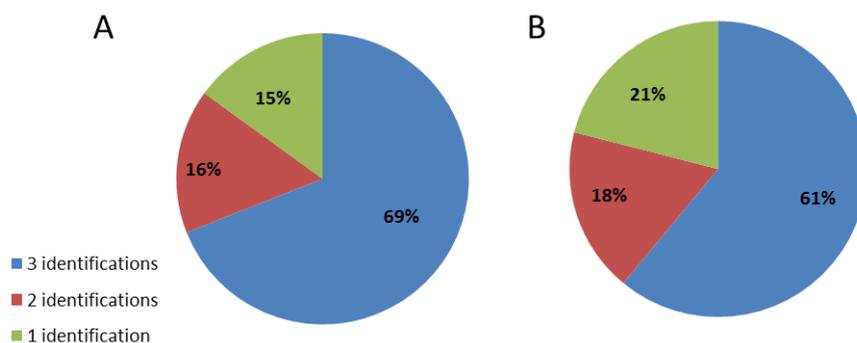


Figure 51
Répartition des protéines (A) et des peptides (B) suivant le nombre d'analyses où ils sont identifiés

Dans un deuxième temps, le nombre d'injections à réaliser afin d'obtenir un nombre d'identifications représentatif de l'échantillon a été déterminé. Les résultats obtenus sont illustrés par la figure 52.

L'effet d'ordre, dû au fait que les analyses successives n'ont pas été effectuées de manière aléatoire, a été pris en compte. Ainsi, une moyenne a été réalisée sur les six permutations possibles. Lors de la première injection, 100 % de protéines et peptides sont nouvellement identifiés. La deuxième injection permet d'identifier 12 % de protéines et 16 % de peptides en plus. À la troisième injection, 6 % de protéines et 8 % de peptides sont nouvellement identifiés par rapport aux deux premières analyses.

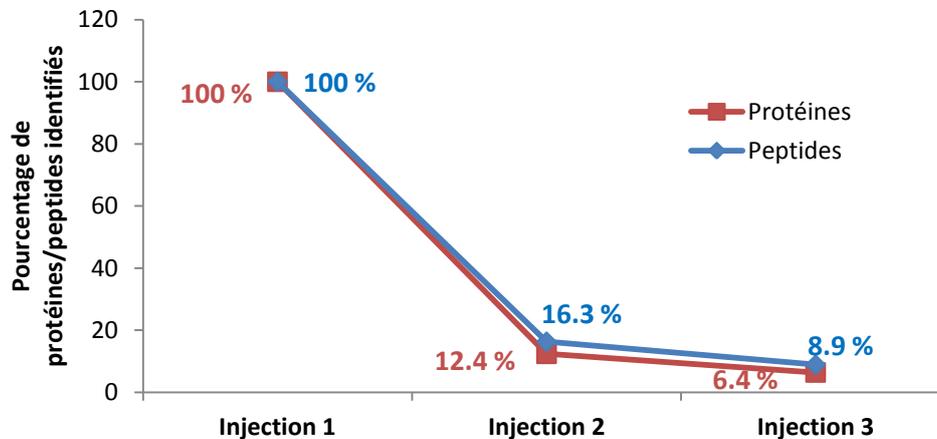


Figure 52
Nombre de protéines ou de peptides nouvellement identifiés à chaque injection

Enfin, si l'on pose l'hypothèse que 100 % des protéines et des peptides sont identifiés après la troisième injection, la figure 53 montre que plus de 80 % des protéines et des peptides sont identifiés dès la première analyse, et plus de 90 % après la deuxième analyse. Le nombre d'analyses à réaliser résulte quant à lui d'un compromis entre la durée totale de l'analyse d'un échantillon, la qualité des identifications souhaitées et la quantité de matériel biologique disponible. Il semble ici judicieux de considérer que deux analyses suffisent pour obtenir une vision significative de l'échantillon puisqu'elles permettent d'identifier 95 % des protéines et 93 % des peptides.

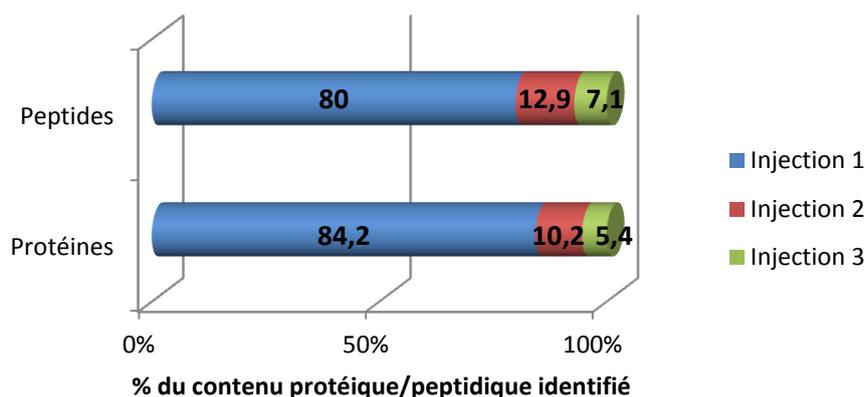


Figure 53

Pourcentage du contenu protéique (A) et peptidique (B) identifié à chaque analyse

B.1.2. Couplage nanoHPLC – Chip AmaZon

Le projet portant sur l'étude de cellules souches cancéreuses, détaillé dans la Partie III – chapitre 2, a été réalisé sur le couplage nanoHPLC – Chip AmaZon. Un lysat de cellules souches cancéreuses issues d'un patient atteint de glioblastome a été déposé deux fois sur un gel 1D SDS PAGE (2 × 90 bandes) et chaque bande a été injectée une fois.

La répétabilité en termes d'identifications implique ici la répétabilité de la séparation sur gel 1D, de l'extraction et de la digestion des peptides ainsi que la répétabilité de l'injection. Les résultats présentés en figure 54 montrent une augmentation de 12 % du nombre de protéines et de 16 % du nombre de peptides identifiés à la suite d'un duplicat d'injection. Il est intéressant de noter que les résultats sont similaires à ceux obtenus lors de l'étude de répétabilité menée sur le couplage nanoAcquity – Synapt G1.

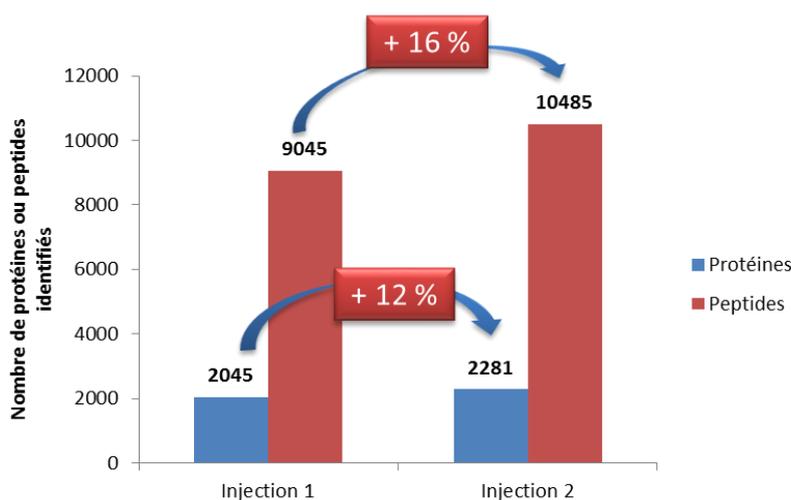


Figure 54

Apport d'une deuxième injection d'un même échantillon à l'identification de protéines et peptides.

B.1.3. Couplage nanoAcquity – Maxis Impact

L'étude sur la prédiction de rechute du système nerveux central dans le cadre de lymphome diffus à grandes cellules B (cf. Partie III – chapitre 3, étude Protell) a été réalisée sur le couplage nanoAcquity – Maxis Impact. Ce Q-TOF récemment acquis par le laboratoire est un instrument dernière génération caractérisé par une vitesse de scan plus élevée (possibilité de travailler jusqu'à 25 Hz) et une gamme dynamique plus étendue.

Un même échantillon de microparticules plasmatiques issues d'un pool de patients a été déposé 12 fois sur un gel 1D SDS PAGE (1 bande). Chaque bande a été injectée une fois. De la même manière que précédemment, la répétabilité des identifications implique ici la répétabilité de la séparation sur gel, de l'extraction et de la digestion des peptides ainsi que la répétabilité de l'injection. La figure 55 présente le nombre de protéines et de peptides identifiés dans les différentes injections. Ce nombre est quasi identique d'une injection à une autre pour ce qui est des peptides comme des protéines.

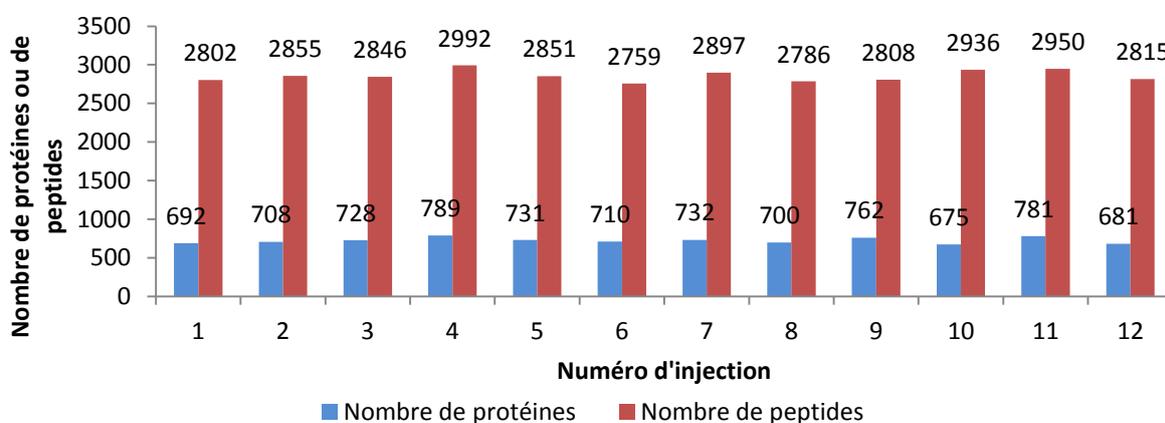


Figure 55
Nombre de protéines et de peptides identifiés à chaque injection

La figure 56 présente le pourcentage de protéines nouvellement identifiées à chaque injection. À la deuxième injection, 11 % et 15 % de protéines et de peptides sont nouvellement identifiés, et 5 et 9 % à la troisième injection. Ces pourcentages sont identiques à ceux obtenus sur le G1 et sur l'AmaZon.

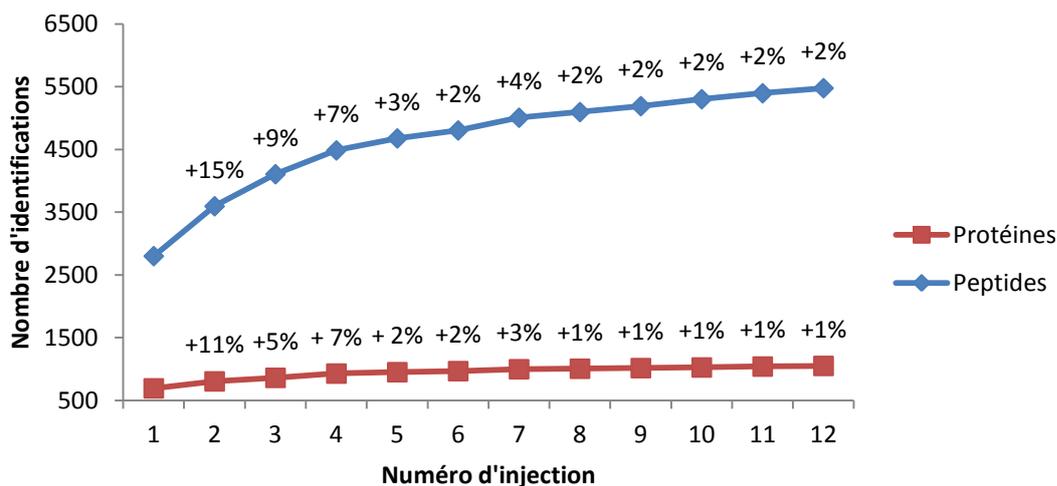


Figure 56

Pourcentage de protéines et de peptides nouvellement identifiés à chaque injection

Contrairement à ce que l'on pouvait supposer, l'augmentation de la vitesse de scan n'améliore pas la répétabilité. Cet aspect reste inhérent à l'analyse par spectrométrie de masse en mode DDA. Afin d'analyser un échantillon en profondeur, il apparaît ici judicieux de réaliser deux répétitions pour chaque échantillon, quel que soit l'instrument utilisé.

B.2. Réplicats d'injection avec liste d'exclusion

Sur les instruments Bruker Daltonics, il est possible de paramétrer facilement une liste d'exclusion. L'analyse est réalisée une première fois en mode classique et les peptides sont identifiés via Mascot. Un script permet ensuite d'extraire l'ensemble des peptides identifiés avec un score supérieur à une valeur donnée (ici, 30). Ces ions sont ensuite inclus dans une liste d'exclusion qui est ajoutée à la méthode MS et MS/MS. Cette technique permet de ne pas sélectionner une deuxième fois les ions déjà bien identifiés lors de la première analyse. Quelques études ont permis de mettre en évidence l'intérêt de cette liste d'exclusion en termes d'identifications par rapport à une répétition simple²⁰⁰⁻²⁰⁴.

La liste d'exclusion a été testée sur la trappe ionique AmaZon. Une exclusion sur 60 s autour du temps de rétention détecté dans la première analyse et une exclusion en masse de 0.3 Da à gauche et 3 Da à droite ont été établies. Les résultats présentés en figure 57 sont à comparer à ceux obtenus en figure 54 dans le point 2.1.2. Une augmentation légèrement supérieure est observée pour le nombre de protéines identifiées en utilisant une liste d'exclusion. La différence reste cependant faible. Par contre, le nombre de peptides identifiés augmente de 35 % contre 16 % sans liste d'exclusion.

L'utilisation de cette liste d'exclusion pour le réplicat permet ainsi d'augmenter :

- la qualité des identifications (67 % des identifications à un peptide passent à des identifications à plusieurs peptides) ;
- la couverture de séquence des protéines détectées.

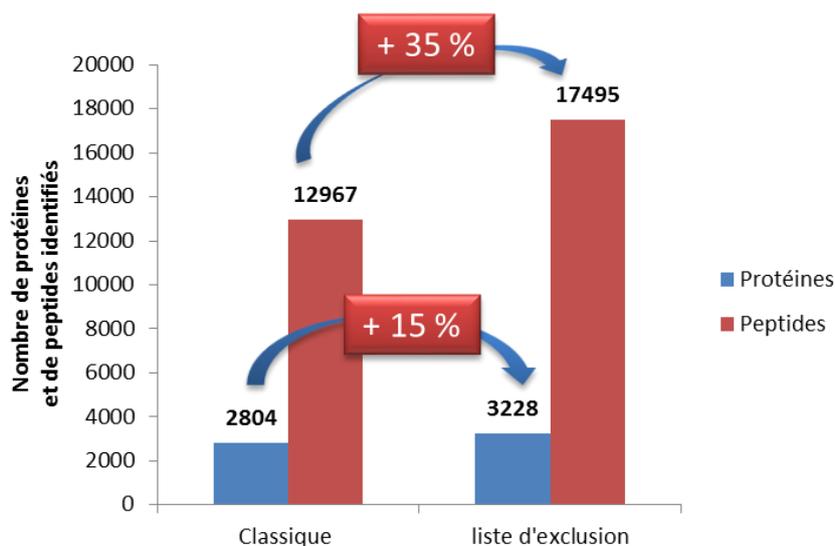


Figure 57

Apport d'une deuxième injection d'un même échantillon en utilisant une liste d'exclusion

B.3. Conclusions

La répétition simple d'une analyse permet d'augmenter le nombre de protéines et de peptides identifiés d'environ 12 % et de peptides de 16 %, quels que soient le couplage et l'instrument utilisé. Cet état de fait est inhérent au mode d'acquisition utilisé (DDA) qui dépend du scan MS.

La répétition en utilisant une liste d'exclusion des peptides identifiés avec de bons scores Mascot lors de la première injection permet un léger gain concernant le nombre de protéines identifiées. En revanche, le nombre de peptides identifiés est largement amélioré par rapport à une répétition simple.

C. Utilisation d'un nouveau mode d'acquisition : la DIA

Un nouveau type d'acquisition appelé DIA (*data independant acquisition*) a été récemment développé afin de s'affranchir des limitations de la DDA. Ce mode permet de réaliser une fragmentation simultanée de plusieurs précurseurs sans tenir compte de l'intensité ou de la charge des ions parents^{205, 206}. Ainsi, des spectres de fragmentation de l'ensemble des ions contenus dans un échantillon sont obtenus, permettant théoriquement de s'affranchir de l'aspect stochastique des analyses tout en augmentant la sensibilité (fragmentation des ions peu abondants).

Deux techniques principales ont été mises au point :

- **L'isolation séquentielle et la fragmentation des précurseurs sur une certaine fenêtre de masse** : les ions sont sélectionnés sur des fenêtres de 10 à 25 Da pour la fragmentation (SWATH)^{114, 206}, voire 2,5 Da (PACIFIC)¹¹⁵. Dans le cas d'un cycle SWATH, technologie développée sur des instruments de type TripleTOF™ (ABSciex, Concord, Canada), 32 fenêtres de 25 Da seront réalisées au cours d'un cycle d'acquisition pour un temps de scan total de 3 s. Cette technique a été proposée en 2004 mais ce n'est que récemment qu'elle a pu être réellement

appliquée grâce notamment à l'augmentation de la vitesse d'acquisition des spectromètres de masse et au développement de logiciels permettant de traiter les données. Pour le moment, il n'y a cependant pas de logiciel disponible pour interpréter les spectres MS/MS. Les identifications sont réalisées grâce à des bibliothèques de spectres^{207, 208} précédemment obtenues lors d'analyses en mode DDA. En effet, il n'est pas aisé de lier les fragments au bon précurseur. Le SWATH est principalement utilisé à des fins quantitatives (quantification sans marquage).

- **L'alternance entre des scans basse et haute énergie : la MS^E.** Cette technologie, publiée pour la première fois en 2003²⁰⁵ et actuellement commercialisée par Waters, propose de réaliser en alternance des scans basse et haute énergie. Les spectres sont identifiés grâce au logiciel ProteinLynx Global SERVER™ (PLGS™, Waters). Il est ainsi théoriquement possible d'identifier, voire de quantifier l'ensemble du protéome accessible d'un échantillon en une seule analyse.

Au cours de ma thèse au laboratoire, j'ai été chargée d'implémenter les technologies LC-MS^E et LC-HDMS^E (mobilité ionique-MS^E) sur le couplage nanoAcquity– Synapt G2 (Waters) et d'évaluer l'intérêt de cette méthode pour l'analyse d'échantillons complexes.

C.1. Description des techniques LC-MS^E et LC-HDMS^E

La MS^E consiste à réaliser en alternance des scans basse et haute énergie (cf. figure 58) afin de fragmenter l'ensemble des peptides contenus dans un échantillon. Le scan basse énergie permet d'obtenir une liste d'ions parents et le scan haute énergie, de fragmenter en une seule fois l'ensemble des précurseurs. Le scan haute énergie est réalisé en appliquant une rampe d'énergie de collision, en général de 15 à 45 eV²⁰⁹.

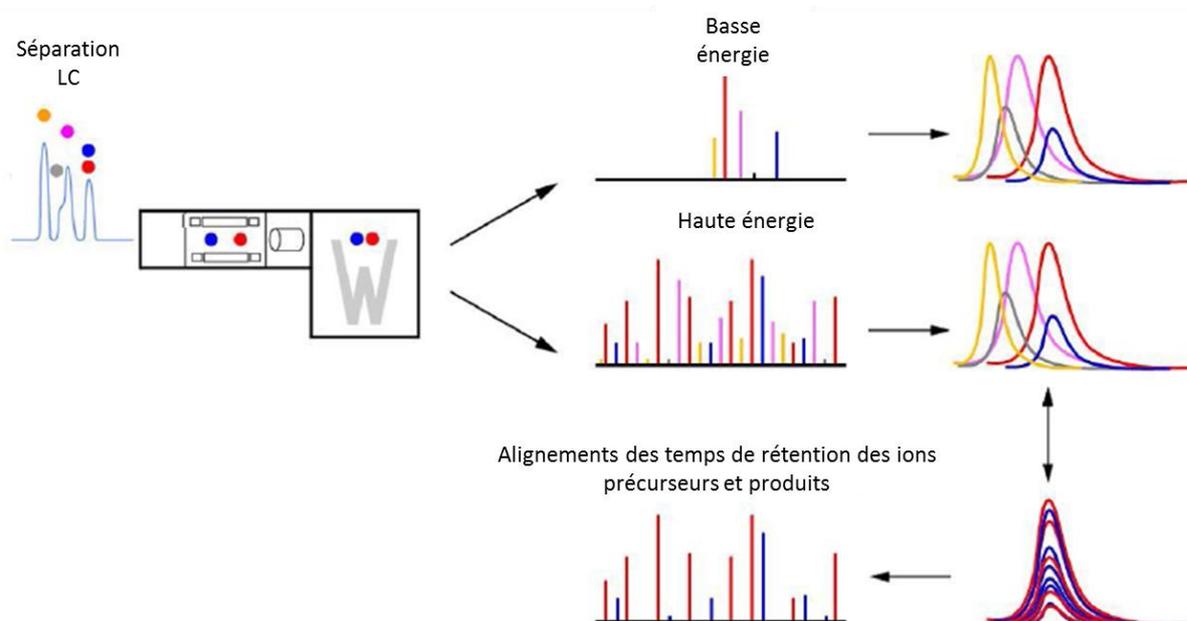


Figure 58

Principe de l'acquisition en mode MS^E

Les ions sont scannés à haute et basse énergie. Précurseurs et fragments sont ensuite réassignés grâce aux temps de rétention.

La difficulté consiste ici à assigner correctement les fragments à leurs ions précurseurs. Afin de simplifier et d'améliorer les assignations, Waters a commercialisé la LC-HDMS^E, une technologie développée par Hoaglung-Hizer et coll. en 2001²¹⁰. Le mode d'acquisition DIA est associé à une dimension de séparation supplémentaire : la mobilité ionique. Cette dernière permet de séparer des analytes en phase gazeuse sous l'effet d'un champ électrique.

Il existe plusieurs types de cellules de mobilité ionique²¹¹⁻²¹³. Nous ne décrivons ici que la technologie développée par Waters : la cellule TWIMS (*travelling wave ion mobility*). Bien que peu résolutive, cette cellule est assez sensible et permet de déterminer des sections efficaces. Le champ des applications possibles est donc plus large. La cellule est composée d'un guide d'ions de type SRIG (*stacked ring ion guides*). Des impulsions de potentiel sont appliquées successivement sur ces guides afin de générer un champ électrique qui va drainer les ions vers la sortie de la cellule de mobilité. En parallèle, des potentiels de radiofréquences sont appliqués afin d'empêcher la diffusion radiale des ions (cf. figure 59)²¹⁴⁻²¹⁶.

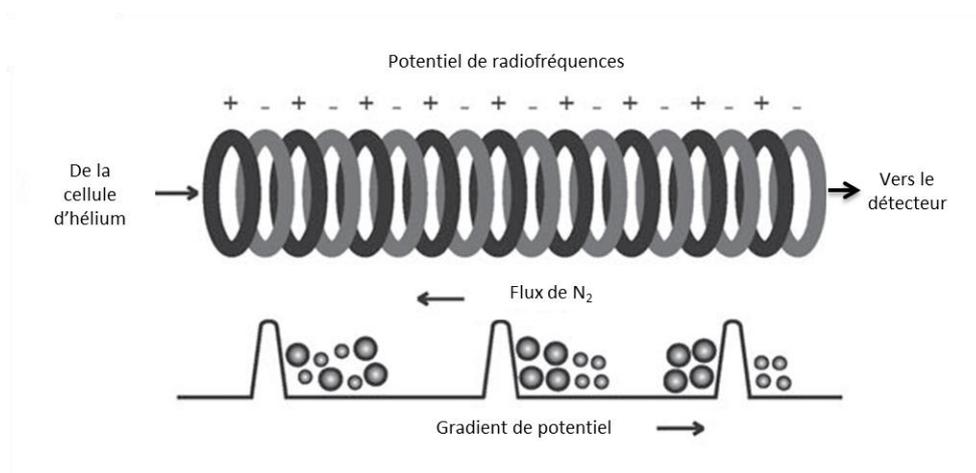


Figure 59

Schéma d'un guide d'ions de type SRIG constituant la cellule de mobilité du Synapt G2

Sur chaque lentille, des potentiels de radiofréquence sont appliqués afin d'éviter la diffusion radiale des ions. Des impulsions de potentiel sont superposées à ces radiofréquences afin de drainer les ions vers la sortie de la cellule de mobilité.

Source : d'après Utrecht et coll.²¹⁷

La cellule de mobilité est insérée dans une cellule Tri-Wave (cf. figure 60). Les ions arrivant du quadripôle sont échantillonnés dans la cellule Trap, puis confinés dans la cellule d'hélium afin d'éviter les pertes entre chaque cycle IMS. Ils atteignent ensuite la cellule de mobilité ionique soumise à une pression en azote de l'ordre de 0,5 mbar. Suivant les collisions avec les molécules de gaz, les ions seront entraînés vers la sortie sous l'effet des vagues de potentiel avec un temps de passage dans la cellule plus ou moins long. Les ions ne pouvant suivre la vague n subissent un mouvement de retour en arrière et sont drainés vers la sortie par la vague $n+1$ ^{216, 217}. La séparation dépend de :

- La charge des ions : plus un ion sera chargé et plus son passage dans la cellule de mobilité sera court.

- Leur conformation : plus un ion est compact et plus son passage dans la cellule de mobilité est court.

Les ions sont ensuite transférés via le guide d'ion Transfer au TOF, où ils peuvent être fragmentés.

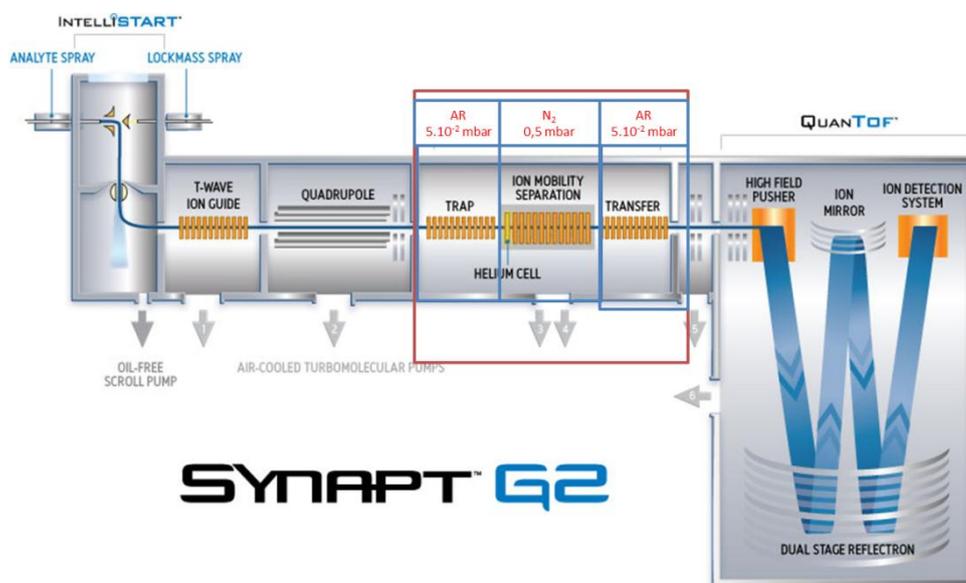


Figure 60

Schéma du Synapt G2

La cellule Tri-Wave est composée des cellules Trap et Transfer avec une pression en Argon de 5.10^{-2} mbar, d'une cellule de mobilité avec une pression de 0,5 mbar en azote. La cellule d'hélium permet de thermaliser les ions.

Source : site internet de la société Waters

La cellule IMS et le TOF fonctionnent de manière synchronisée. Pour un cycle IMS, 200 cycles TOF sont réalisés. Cette valeur est fixe sur les Q-TOF de type Synapt. Un cycle IMS dure de 5 à 40 ms et un cycle TOF de 25 à 200 μ s. Le signal est accumulé pendant 0,4 à 2 s suivant les paramètres définis par l'utilisateur.

Au cours d'une analyse protéomique, les ions à séparer sont de natures et de masses différentes. L'optimisation des paramètres de séparation dans la cellule de mobilité ionique est moins cruciale que lors d'analyses de grands édifices telles que des protéines entières ou des complexes protéiques. La figure 61 montre le nombre de peptides identifiés à chaque cycle TOF au cours de l'analyse d'un digestat d'E.Coli (MassPREP™ digestion standard mix 1, Waters) avec utilisation des paramètres de mobilité ionique indiqués par le constructeur. La plupart des peptides sont élués entre le cycle 40 et le cycle 120, signe d'une séparation optimale.

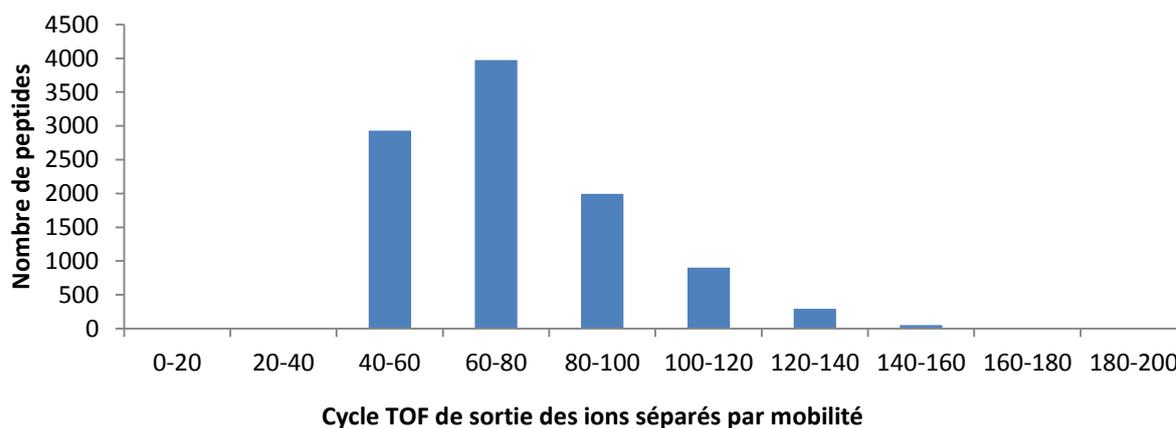


Figure 61
Nombre de peptides identifiés à chaque cycle TOF

C.2. Interprétation des données via PLGS™

La société Waters a développé le logiciel PLGS™, qui permet de réaliser les assignations entre les ions précurseurs et les ions fragments. Les différentes étapes de traitement sont explicitées dans la figure 62^{218, 219}.

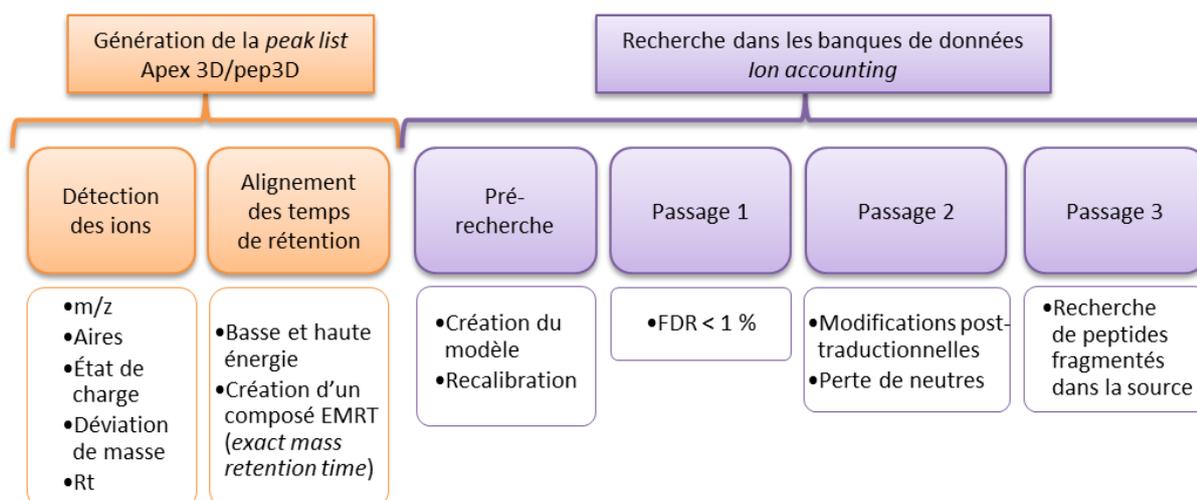


Figure 62
Différentes étapes menées par PLGS pour identifier correctement chaque spectre réalisé lors de l'analyse nanoLC-MS^E

L'algorithme Apex 3D permet de soustraire le bruit de fond et d'intégrer les signaux par rapport au temps de rétention chromatographique. Ensuite, pep3D va rassembler les isotopes de chaque ion. Chaque peptide est alors caractérisé par une masse, un temps de rétention (t_r), une intensité et éventuellement un temps de passage dans la cellule de mobilité ionique (t_d). Les fragments sont ensuite associés aux précurseurs suivant les similitudes d'éluion (t_r et t_d). Un même fragment peut-être assigné à plusieurs précurseurs.

Enfin, l'algorithme *ion accounting* identifie les spectres par recherche dans les banques de données en trois étapes. Dans un premier temps, les peptides et les protéines sont identifiés de manière itérative. À chaque cycle, les ions et fragments identifiés précédemment sont retirés de la liste des spectres à

assigner. Cette étape est arrêtée dès que le taux de FDR désiré est atteint. Ensuite, des modifications sur les peptides appartenant à des protéines précédemment identifiées dans le passage 1 sont recherchées. Enfin, lors du dernier passage, les fragments ayant une intensité plus élevée que le précurseur sont autorisés pour identifier des peptides labiles qui auraient été fragmentés dans la source.

Les identifications peptidiques sont ensuite classées et un score (pepscore) leur est attribué suivant l'adéquation entre l'identification et le modèle physico-chimique selon 15 critères :

- Précision de masse sur la mesure du parent
- Précision de masse sur la mesure des fragments
- Somme de l'intensité des fragments
- Nombre d'ions *b* et *y* consécutifs
- Complémentarité des ions *b* et *y*
- Fragmentation observée au niveau des sites de clivage préférentiels
- Rapport entre l'intensité totale des ions *y* et celles des ions *b*
- Adéquation entre le temps de rétention théorique et le temps de rétention expérimental
- Adéquation entre la mobilité ionique théorique et celle qui a été obtenue expérimentalement
- Rapport entre l'intensité du précurseur et l'intensité totale des fragments
- Adéquation entre les pertes de neutre et la composition en résidus d'acides aminés (H₂O et NH₃)
- Multiplicité des états de charge conforme au modèle théorique établi suivant la composition en résidus d'acides aminés
- Nombre de peptides identifiés conformes au modèle
- Nombre de fragments identifiés conformes au modèle
- Rapport de la somme des intensités des ions fragments sur la somme des intensités des précurseurs.

C.3. Intérêt de l'utilisation combinée du mode DIA et de la mobilité ionique

Afin d'évaluer l'intérêt de l'utilisation du mode d'acquisition DIA ainsi que celui de l'association DIA-mobilité ionique, appelée HD-MS^E, un digestat d'E-Coli (MassPREPTM digestion standard mix 1, Waters) a été injecté suivant les trois modes d'acquisition possibles : DDA, MS^E et LC-HDMS^E. La DDA a été réalisée sur le Synapt G1 car les méthodes d'acquisition *y* sont déjà opérationnelles et optimisées.

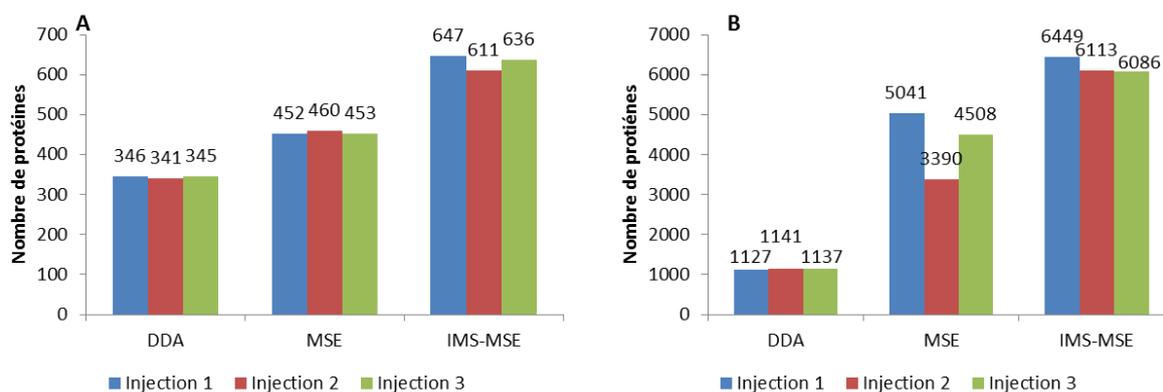


Figure 63

Nombre d'identifications protéiques (A) et peptidiques (B) obtenues lors d'acquisition en mode DDA, MS^E et HDMS^E

La figure 63 présente un gain en identifications au niveau protéique et peptidique avec la MS^E et, surtout, la LC-HDMS^E par rapport à une acquisition DDA classique. Cette expérience montre l'efficacité de la MS^E et de la LC-HDMS^E pour obtenir la couverture du protéome d'un échantillon plus large.

Deux raisons principales expliquent ce résultat :

- La fragmentation de l'ensemble des précurseurs en mode MS^E n'occasionne pas de perte d'informations.
- La mobilité ionique en HDMS^E facilite l'association des fragments aux précurseurs en créant une dimension de séparation supplémentaire.

Par ailleurs, la répétabilité des injections a été évaluée aux niveaux peptidique et protéique (cf. tableau 6). Les résultats ne montrent pas une amélioration de la répétabilité des analyses par rapport à la DDA. La MS^E montre même des performances largement inférieures, notamment au niveau peptidique. Cela peut-être imputé au logiciel PLGSTM. En effet, le traitement des données est réalisé de manière itérative de façon à assigner le maximum de spectres, indépendamment de la qualité de ces derniers.

Technique utilisée	% Protéines	% Peptides
DDA	73	54
MS ^E	61	28
LC-HDMS ^E	74	61

Tableau 6

Pourcentage d'identifications protéiques et peptidiques communes pour trois injections du digestat d'E.Coli

Ainsi, si le mode d'acquisition DIA ne permet pas d'améliorer la répétabilité des analyses, il permet néanmoins d'en augmenter la profondeur. Cependant, ces bons résultats sont à pondérer par la qualité des spectres. En effet, le logiciel PLGSTM autorise l'assignation de spectres MS/MS de qualité douteuse sous couvert du respect des 14 critères (ou 15 si utilisation de la mobilité ionique) et du taux de FDR (< 1 %).

La figure 64 montre quelques spectres obtenus et validés par PLGS lors d'une analyse LC-HDMS^E avec les pepscores assignés. Une analyse qualitative des spectres a montré qu'en dessous d'un score de 8, la

qualité est discutable. Or, d'après la figure 65, la majorité des spectres validés sont identifiés avec un pepscore entre 6 et 7. Pour les peptides *decoy (random)*, les pepscores maxima sont compris entre 5 et 6. Ainsi, si la qualité des spectres reste discutable, très peu de *decoys* sont identifiés avec un score supérieur à 6, ce qui justifie la pertinence des 15 critères.

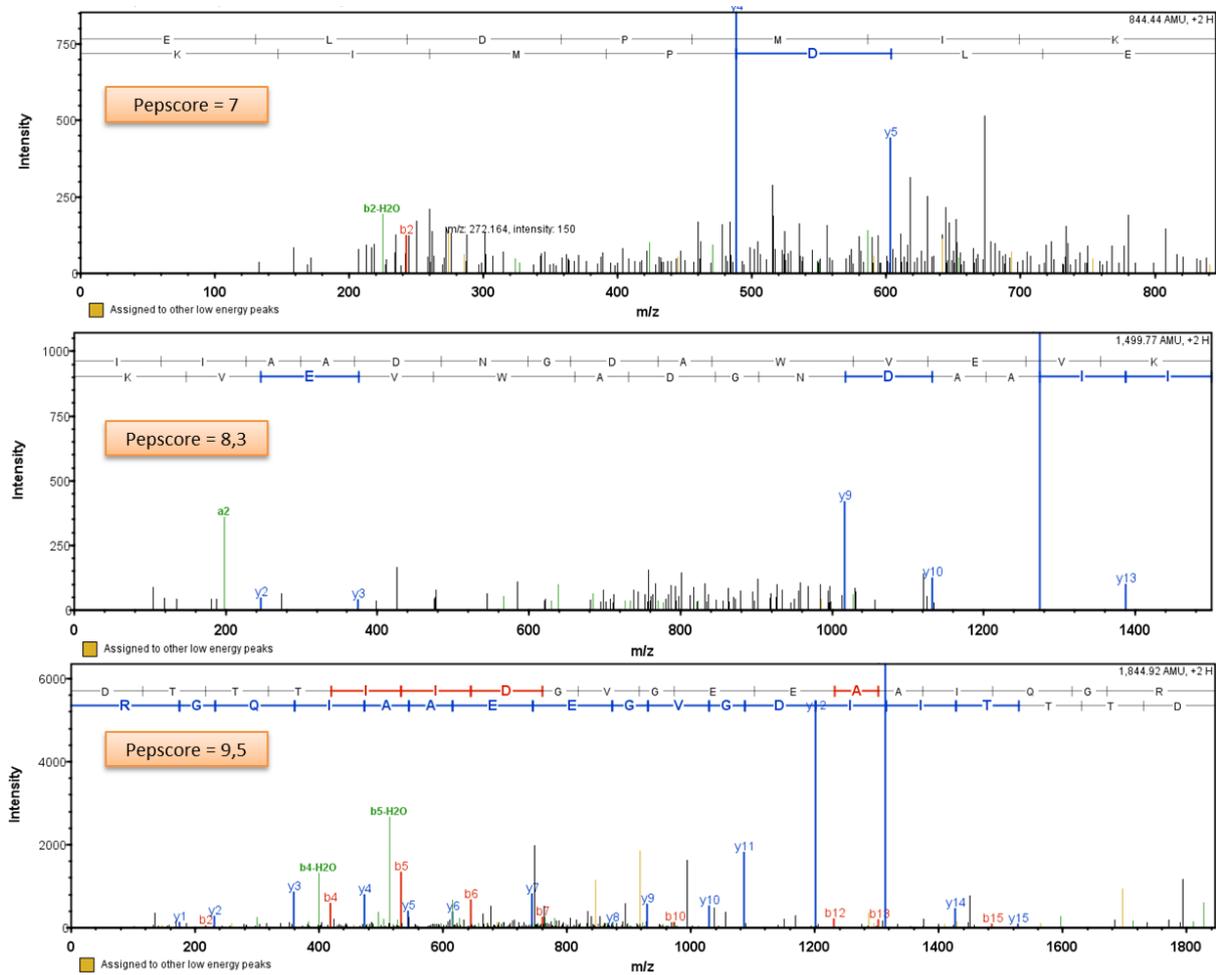


Figure 64

Exemple de spectres assignés par PLGS avec des scores de 9,5, de 8,3 et de 7

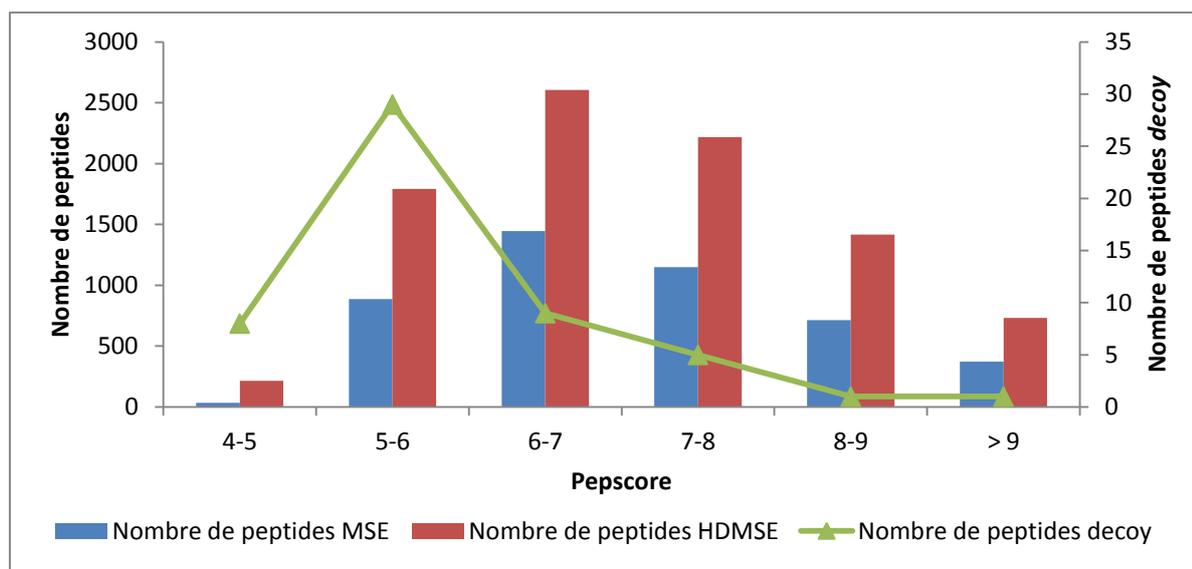


Figure 65

Nombre de peptides identifiés suivant les plages de score en MS^E et LC-HDMS^E et nombre de peptides *decoy* (*random*) identifiés suivant les mêmes plages de scores

Une question importante est ici soulevée : quelle confiance peut-on accorder à PLGSTM pour les identifications ainsi qu'à des critères issus de modèles basés sur les propriétés physico-chimiques des peptides, et non plus uniquement sur la qualité des spectres ?

D'autre part, si la LC-HDMS^E apporte une dimension de séparation supplémentaire, elle occasionne aussi une réduction de la gamme dynamique pour deux raisons^{220,221} :

- **La perte par diffusion lors de la séparation des ions dans la cellule de mobilité** : en effet, les ions passent de la cellule Trap à basse pression ($5 \cdot 10^{-2}$ mbar) à la cellule d'IMS où règne une haute pression (0,5 mbar) et à la cellule Transfer où la pression redevient basse ($5 \cdot 10^{-2}$ mbar).
- **La saturation du digitaliseur** : les ions ne sont plus distribués de manière équitable entre les 200 cycles TOF mais sont enregistrés sur un faible nombre de cycles.

Ces deux effets sont décrits dans la figure 66. L'intensité du BPI est largement moins élevée dans le cadre d'une acquisition utilisant la mobilité ionique et un effet de saturation est observé sur les ions les plus intenses. Le massif isotopique a la « tête coupée », ce qui occasionne un changement des rapports isotopiques. Par ailleurs, on observe un phénomène de résonance du digitaliseur.

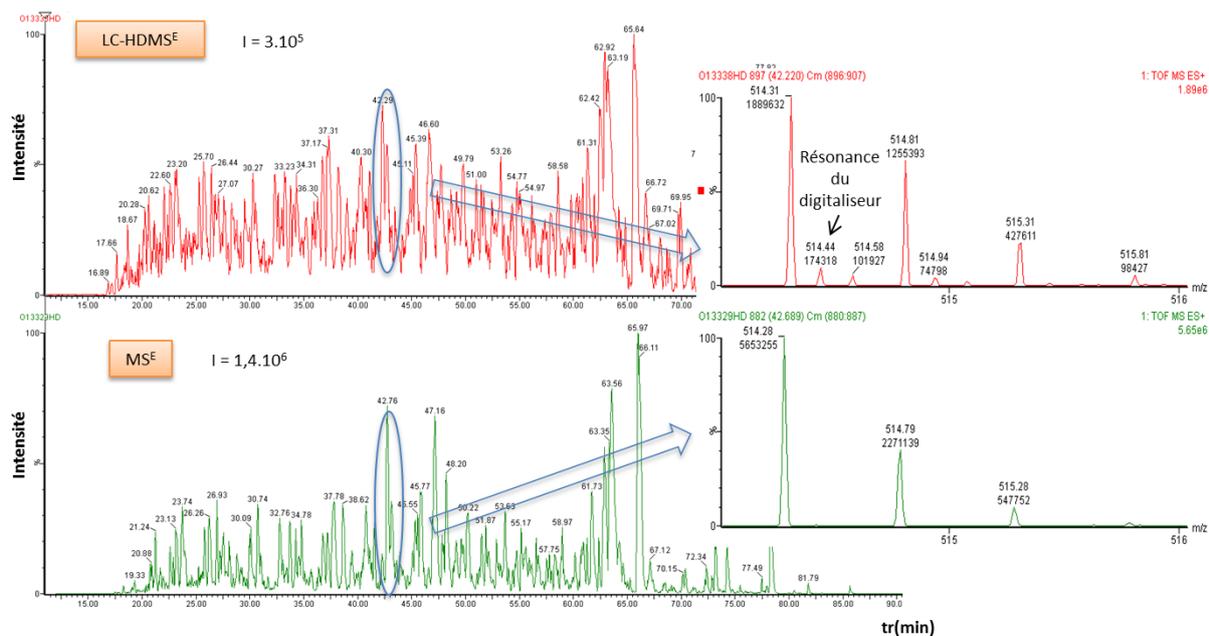


Figure 66

BPI obtenu lors d'acquisitions utilisant la LC-HDMS^E et la MS^E

L'intensité globale du BPI est moins importante dans le cas d'une acquisition utilisant la mobilité. Un phénomène de saturation est observé sur les ions les plus intenses (résonance du digitaliseur et modification des rapports isotopiques).

C.4. Conclusions

La MS^E permet d'augmenter la profondeur de l'analyse des échantillons. Cependant, la difficulté reste l'identification et la validation des spectres. Pour le moment, il n'existe qu'un seul logiciel commercial interprétant ces données et l'assignation de certains spectres reste discutable. Si la mobilité ionique permet, quant à elle, d'améliorer la qualité des assignations, elle réduit aussi la gamme dynamique.

Des développements méthodologiques restent à apporter à cette technique, tant au niveau instrumental que logiciel. Il y a fort à parier que ce mode d'acquisition remplacera un jour la DDA, comme en témoignent les publications récentes parues dans des journaux à facteur d'impact élevé²²⁰⁻²²⁴.

D. Conclusion générale

Ce chapitre illustre les limitations du mode d'acquisition DDA : le protéome complet d'un échantillon ne peut être identifié à l'issue d'une seule analyse. C'est pourquoi il est important de réaliser une optimisation fine de la séparation chromatographique et des paramètres d'acquisition du spectromètre de masse afin d'obtenir la vision la plus significative possible d'un échantillon. De plus, il a été démontré qu'une deuxième injection permettait, quel que soit l'instrument utilisé, d'apporter de nouvelles identifications protéiques et peptidiques. Dans ce contexte, la répétition avec liste d'exclusion s'est avérée très efficace.

L'avenir de l'analyse haut débit semble ici se diriger vers le développement de nouveaux modes d'acquisition DIA. Cependant, si le concept de la LC-HDMS^E est prometteur (possibilité d'identifier et de quantifier le protéome complet d'un échantillon), quelques points techniques restent à améliorer, notamment au niveau du logiciel permettant de valider les spectres identifiés.

CHAPITRE 3

Recherche dans les banques de données

La recherche dans les banques de données est un paramètre clé de l'analyse protéomique. Le choix du moteur de recherche et des paramètres spécifiés ainsi que la manière de filtrer les identifications peuvent influencer les résultats.

Ainsi, lors de ma thèse, j'ai eu l'occasion de participer à une démonstration d'un constructeur sur un nouvel appareil. Un digestat de levure a été analysé et les données ont été exploitées par le constructeur et deux laboratoires publics (dont le LSMBO). La figure 67 montre les disparités obtenues quant au nombre d'identifications : de 1200 à 2000 protéines détectées suivant le lieu où a été effectuée la recherche.

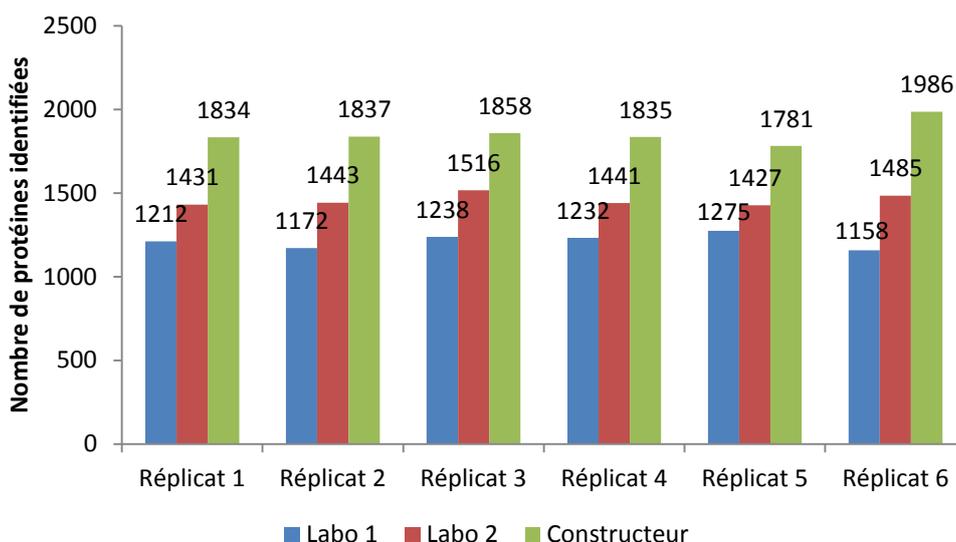


Figure 67

Nombre d'identifications protéiques obtenues sur 6 injections d'un digestat de levure suivant les applications de filtres et le lieu où ont été effectuées les recherches dans les banques de données

Les disparités constatées peuvent s'expliquer par trois raisons principales :

- le paramétrage des recherches dans les banques de données ;
- le moteur de recherche utilisé ;
- les filtres appliqués pour valider les résultats.

L'objectif de ce chapitre est de montrer qu'une optimisation fine des paramètres de recherche dans les banques de données permet d'améliorer les résultats obtenus concernant le nombre et la qualité des

identifications. Par ailleurs, l'intérêt de combiner plusieurs moteurs de recherche (Mascot¹²⁵ et Omssa¹²⁹) sera mis en évidence.

A. Optimisation fine des paramètres du moteur de recherche

En moyenne, quel que soit l'instrument du laboratoire utilisé, environ 75 % des spectres ne sont pas assignés après recherche dans les banques de données. Afin de réduire ce nombre, les paramètres de recherche du moteur Mascot ont été évalués en utilisant des données issues d'un gel 90 bandes injecté sur la trappe ionique Amazon précédemment décrite dans la Partie II – chapitre 2. L'échantillon compte environ 2000 protéines et 10 000 peptides.

Le moteur de recherche Mascot, présenté de manière plus détaillée dans la Partie I – chapitre 2, se base sur l'utilisation de deux scores : le score de corrélation, qui évalue la qualité du spectre, et le score d'identité qui prend en compte l'espace de recherche généré pour l'identification de chaque peptide.

Plusieurs paramètres, réglés par l'utilisateur, peuvent influencer le résultat final :

- la taille de la banque protéique ;
- la tolérance définie sur le précurseur et les fragments ;
- l'enzyme utilisée ;
- les modifications recherchées ;
- le nombre de coupures protéolytiques manquées.

A.1. Erreur sur la masse du précurseur

Dans un premier temps, l'erreur optimale à utiliser pour effectuer les recherches dans les banques de données a été évaluée. Il s'agit ici de réduire au maximum le nombre d'identifications possibles afin de limiter les faux positifs. Une recherche a été réalisée en tolérant une erreur sur la détermination du précurseur de $\pm 0,3$ Da. La figure 68 présente, pour chaque spectre assigné, l'erreur sur le précurseur en fonction du score de corrélation attribué par Mascot. Les résultats montrent qu'il est possible de restreindre la tolérance sur la masse du précurseur à $\pm 0,2$ Da.

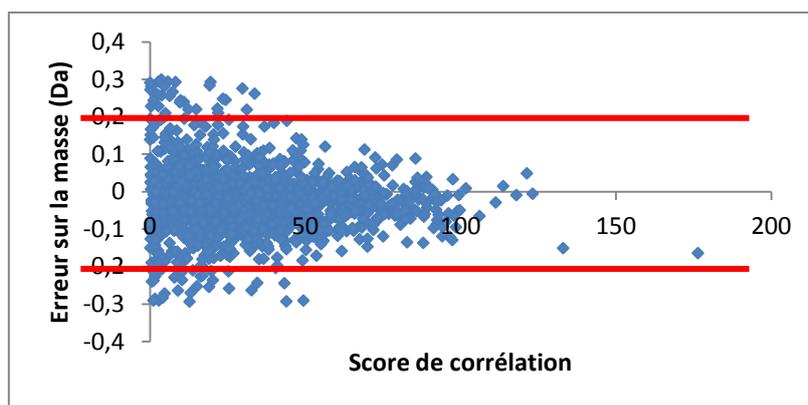


Figure 68

Erreur sur la masse obtenue pour chaque spectre identifié sur la trappe ionique Amazon en fonction du score de corrélation (erreur de $\pm 0,3$ Da sur le parent)

Dans un deuxième temps, la fenêtre de recherche a été ouverte à 3 Da. La figure 69 montre l'erreur obtenue sur l'ensemble des peptides identifiés en fonction du score de corrélation. De manière surprenante, des massifs sont observés autour de 0,1 et 2 Da et non autour de -1 et -2 Da. Deux hypothèses peuvent expliquer cet état de fait :

- Des modifications post-traductionnelles non prises en compte affectent la masse des peptides, par exemple des amidations ou des déamidations.
- Les massifs isotopiques sont mal interprétés et c'est l'isotope p+1 qui est pris en compte comme masse du précurseur au lieu du pic mono-isotopique p.

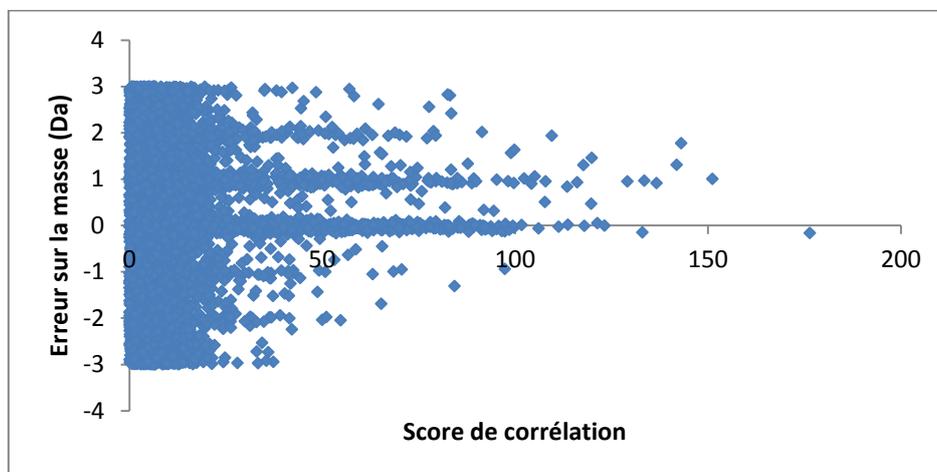


Figure 69

Erreur sur la masse obtenue pour chaque spectre assigné sur la trappe ionique Amazon en fonction du score de corrélation (erreur de ± 3 Da sur le parent)

Afin de valider l'une ou l'autre des hypothèses, la même expérience a été réalisée sur les mêmes données acquises sur un appareil haute résolution ($R_s = 60\,000$) : le Q-TOF Maxis de Bruker Daltonics. Contrairement aux résultats précédents, un massif principal est observé, centré autour de 0 Da, et des identifications ponctuelles sont réalisées à ± 1 , ± 2 et ± 3 Da (cf. figure 70). Cela montre que les modifications post-traductionnelles induisant ces différences sont des événements rares par rapport à

l'ensemble des identifications, et donc que les massifs observés dans les résultats obtenus avec un instrument basse résolution sont dus à des erreurs d'attribution du pic mono-isotopique.

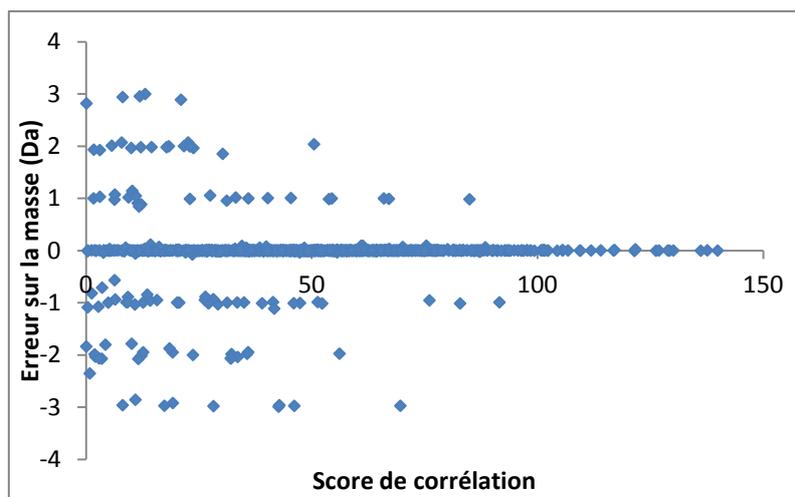


Figure 70

Erreur sur la masse obtenue pour chaque spectre assigné sur le Q-TOF Maxis en fonction du score de corrélation (erreur de ± 3 Da sur le parent)

Afin de pallier ce problème, nous avons mis en œuvre l'option « 2^{13}C », qui permet de tenir compte des erreurs d'attribution du pic mono-isotopique. Notons que l'option « multi-isotopes », présente dans le moteur de recherche Omssa, joue le même rôle. La figure 71 montre les massifs d'identifications obtenus à 0,1 et 2 Da avec une erreur de $\pm 0,2$ Da. Ce paramètre permet d'identifier les ions dont le pic mono-isotopique est mal assigné tout en gardant une erreur sur la masse du précurseur réaliste.

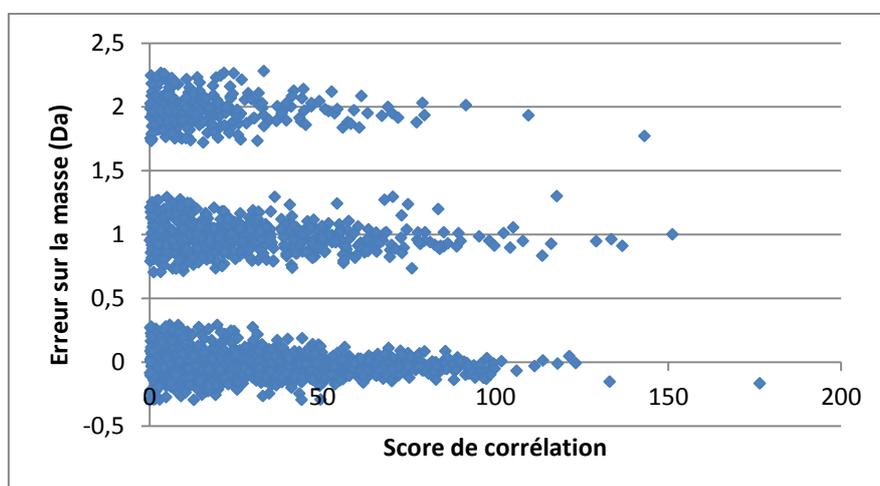


Figure 71

Erreur sur la masse obtenue sur la trappe ionique Amazon pour chaque spectre assigné en fonction du score de corrélation (erreur de $\pm 0,2$ Da sur le parent et utilisation de l'option 2^{13}C)

A.2. Modifications post-traductionnelles

Afin de détecter la présence de modifications post-traductionnelles non attendues, le mode « *Error tolerant* » a été utilisé²²⁵. Une première recherche est effectuée dans la banque de données suivant les paramètres indiqués par l'utilisateur. Lors de la deuxième recherche, les paramètres sont considérablement élargis pour tenter d'identifier les spectres non assignés :

- Des coupures semi-spécifiques sont autorisées
- Le nombre de coupures manquées est augmenté de 1
- La liste des modifications est testée en série
- 150 types de substitutions de résidus d'acides aminés sont testés en série

Une seule déviation par peptide est autorisée. La figure 72 montre les résultats d'identification obtenus sur l'ensemble des 10 000 peptides. Comme attendu, la carbamidométhylation des cystéines et l'oxydation des méthionines font partie des modifications majeures. Par contre, il est intéressant d'observer une forte proportion de modifications propionamides et déamidations. Afin de vérifier qu'il ne s'agit pas d'un effet dû à la basse résolution de la trappe ionique AmaZon (Bruker Daltonics), la même recherche a été effectuée sur un échantillon semblable injecté sur un appareil haute résolution, le Q-TOF Maxis (Bruker Daltonics). La modification propionamide apparaît aussi sur un grand nombre de peptides, mais peu de déamidations sont observées.

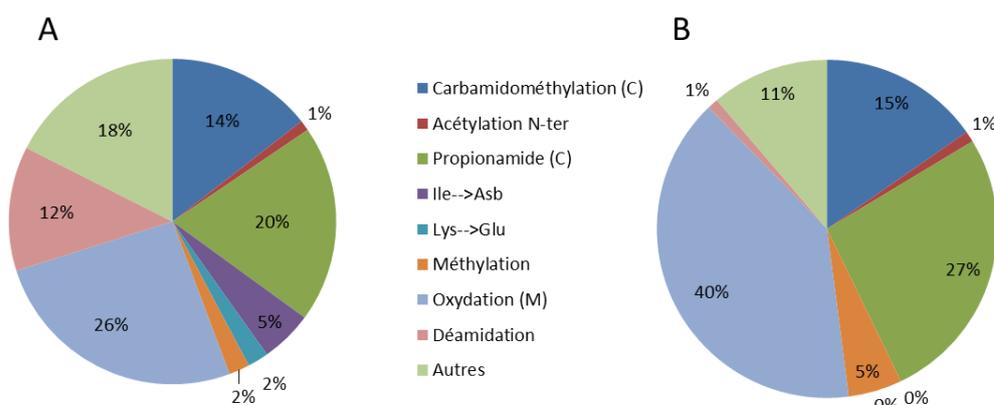


Figure 72

Répartition des modifications identifiées sur chaque spectre pour des analyses réalisées sur un appareil basse résolution, la trappe ionique AmaZon (A), et un appareil haute résolution, le Q-TOF Maxis (B)

La modification propionamide est causée par le gel de polyacrylamide. En effet, la double liaison de l'acrylamide libre dans un gel peut réagir avec des groupements sulfhydryles (–SH) libres pour former un adduit de cystényl S-propionamide. Grâce à l'ajout de cette modification dans les paramètres de recherche, 5 % de peptides supplémentaires sont identifiés (non identifiés si la modification n'est pas implémentée) et la proportion de spectres non assignés est réduite de 4 %.

Finalement, les modifications oxydation, carbamidométhylation et propionamide seront conservées pour effectuer les recherches. L'ajout des modifications méthylation et acétylation N-ter pourrait éventuellement être envisagé mais la partie 1.4. montrera que cette option n'est pas judicieuse.

A.3. Étude de la spécificité de l'enzyme

La spécificité de l'enzyme a été étudiée de la même manière que précédemment, c'est-à-dire en utilisant l'option « *Error tolerant* ». La figure 73 présente le pourcentage de coupures manquées identifiées ainsi que le nombre de peptides semi-tryptiques. Les résultats montrent qu'un nombre de coupures manquées de 1 est largement suffisant. Par ailleurs, 4 % des spectres correspondent à des identifications de peptides semi-tryptiques. Malgré le gain en identifications, cette option n'a pas été choisie car elle augmente considérablement l'espace de recherche et donc le temps d'analyse ainsi que la monopolisation des ressources informatiques du laboratoire.

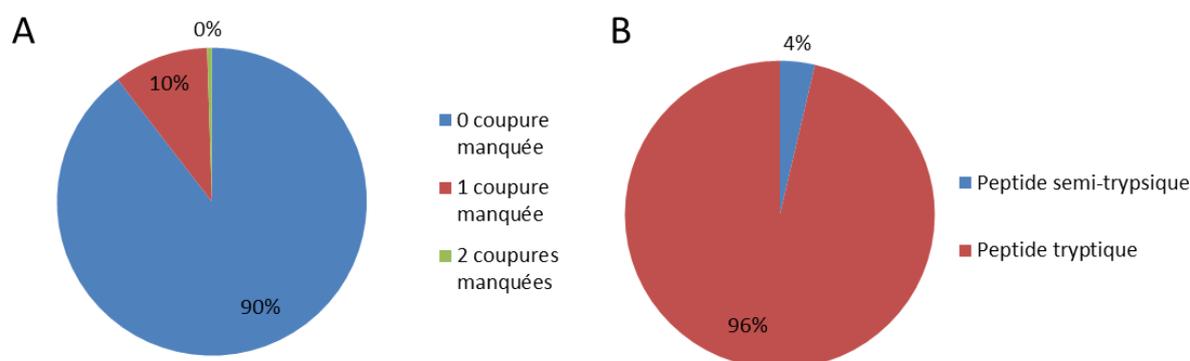


Figure 73

Spécificité de l'enzyme évaluée grâce à la recherche Mascot en mode "Error tolerant"

(A) Pourcentage de coupures manquées de l'enzyme observée. (B) Pourcentage de peptides semi-tryptiques identifiés.

Pour la suite, une coupure manquée sera autorisée et l'enzyme sera considérée comme spécifique.

A.4. Évaluation de l'impact des changements de paramètres

L'impact de divers changements des paramètres de recherche a été évalué sur le score « identité ». En effet, ce score est révélateur de la taille de l'espace de recherche et une augmentation importante de celui-ci va conduire à une augmentation de la probabilité d'identifier des faux positifs. Une première méthode a été établie avec sept modifications, deux carbones 13 et 0,2 Da d'erreur. Les paramètres ont été ensuite modifiés séquentiellement à partir de cette méthode. La figure 74 montre l'évolution du score d'identité moyen suivant les paramètres choisis. Le passage de la méthode 1 à la méthode 2 (tolérance sur le précurseur augmentée à 0,3 Da) n'a que peu d'influence. En revanche, le passage de la méthode 1 à la méthode 3 (trois modifications au lieu de sept) permet de diminuer de 10 points le score identité. Enfin, le passage à la méthode 4 (pas d'utilisation de l'option ^{13}C) a une influence légère. Ainsi, il est possible de conclure que le nombre de modifications recherchées a un impact élevé sur le score d'identité. Par contre, l'erreur sur la masse du précurseur ou l'ajout de carbones 13 est moins

critique. Dans cette optique, les modifications précédemment suggérées comme les méthylations et acétylations ne seront pas utilisées pour les recherches.

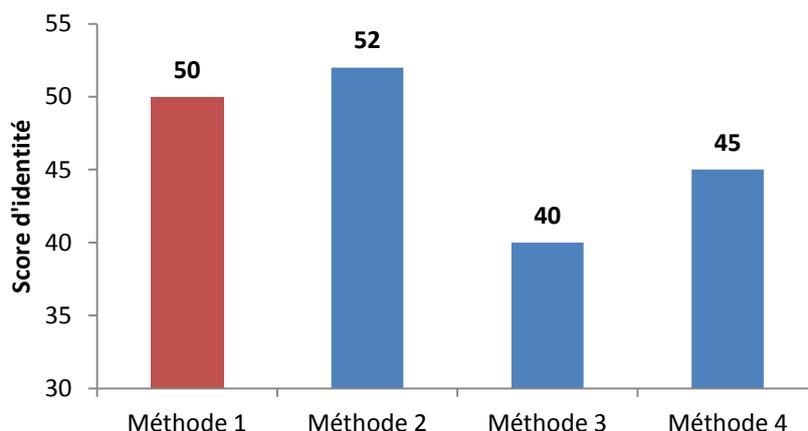


Figure 74

Évaluation de la variation du score d'identité moyen suivant la méthode utilisée

Méthode 1 : 7 modifications, 2 13C et $\pm 0,2$ Da d'erreur sur le précurseur

Méthode 2 : 7 modifications, 2 13C et $\pm 0,3$ Da d'erreur sur le précurseur

Méthode 3 : 3 modifications, 2 13C et $\pm 0,2$ Da d'erreur sur le précurseur

Méthode 4 : 7 modifications, 0 13C et $\pm 0,2$ Da d'erreur sur le précurseur

B. Intérêt de l'utilisation de deux moteurs de recherche

Différents algorithmes de recherche dans les banques de données ont été développés pour l'identification des protéines à partir des spectres MS/MS : Mascot¹²⁵, Sequest¹²⁶, X!Tandem¹²⁷, Andromeda¹²⁸ ou encore Omssa¹²⁹. Ces algorithmes sont basés sur des techniques d'attribution de scores de corrélation différents. Afin de répondre à deux problématiques, identification exhaustive du contenu d'un échantillon et assurance de la confiance dans les résultats²²⁶⁻²²⁹, le laboratoire s'est doté depuis 2009 d'un moteur de recherche non licencié complémentaire, Omssa.

Dans cette partie, nous discuterons de l'intérêt d'utiliser deux moteurs de recherche, Mascot et Omssa.

B.1. Optimisation des résultats

Trois analyses issues de bandes de gel 1D et injectées sur le Q-TOF Synapt G1 (Waters) ont été traitées par les deux moteurs de recherche. La figure 75 présente les résultats obtenus en termes d'identifications protéiques et peptidiques. Les résultats montrent un nombre d'identifications légèrement plus élevé pour Omssa que pour Mascot. Par ailleurs, 87 % et 80 % d'identifications protéiques et peptidiques communes ont été constatées. Les différences minimales sont dues à une attribution différente des scores.

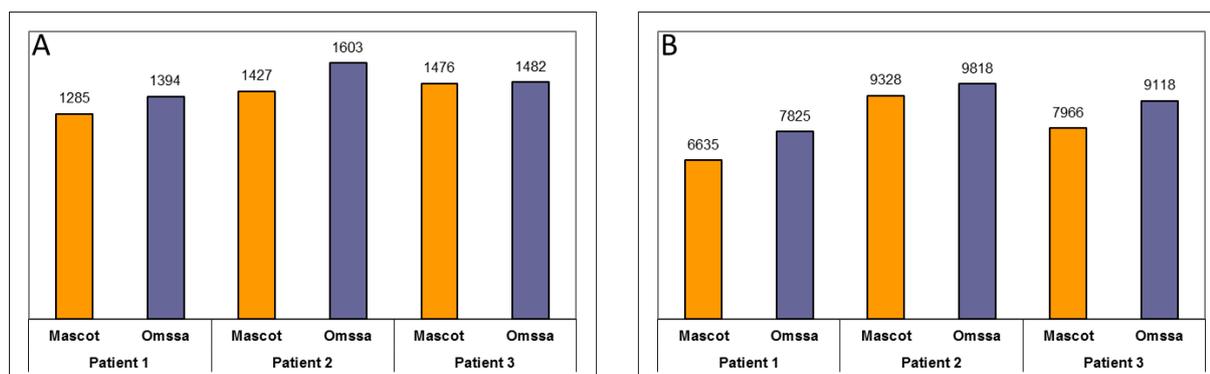


Figure 75

Nombre d'identifications protéiques (A) et peptidiques (B) réalisées par chaque moteur de recherche pour trois échantillons issus de trois patients différents

À ce stade, deux stratégies peuvent être envisagées :

- Conserver l'ensemble des assignations afin de maximiser le nombre d'identifications.
- Conserver uniquement les identifications communes aux deux moteurs de recherche afin d'assurer le maximum de confiance dans les résultats.

La stratégie employée lors des projets menés au cours de ce travail a été adaptée suivant l'objectif à atteindre.

B.2. Une meilleure compréhension des paramètres

Lors des expériences menées sur la trappe ionique AmaZon (Bruker Daltonics), l'utilisation de deux moteurs de recherche a permis de souligner l'importance d'un paramètre lors du traitement de données issues d'appareils à basse résolution. De grandes disparités ont été mises en évidence entre les résultats émanant de chacun des deux moteurs de recherche pour une série d'analyses issues d'un gel 90 bandes. Ainsi, Omssa a permis d'identifier 647 protéines supplémentaires, c'est-à-dire 7 379 peptides et 35 847 spectres en plus (cf. figure 76).

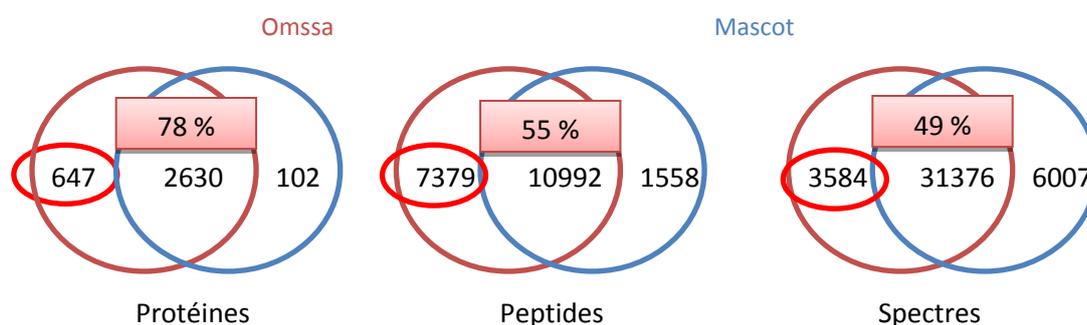


Figure 76

Comparaison des identifications obtenues par Mascot (en bleu) et Omssa (en rouge) au niveau des protéines, peptides et spectres

Les pourcentages d'identifications communes sont indiqués dans l'encadré rouge.

Une analyse plus fine des résultats a montré que la proportion d'ions +3 identifiés à l'aide d'Omssa était beaucoup plus importante (cf. figure 77).

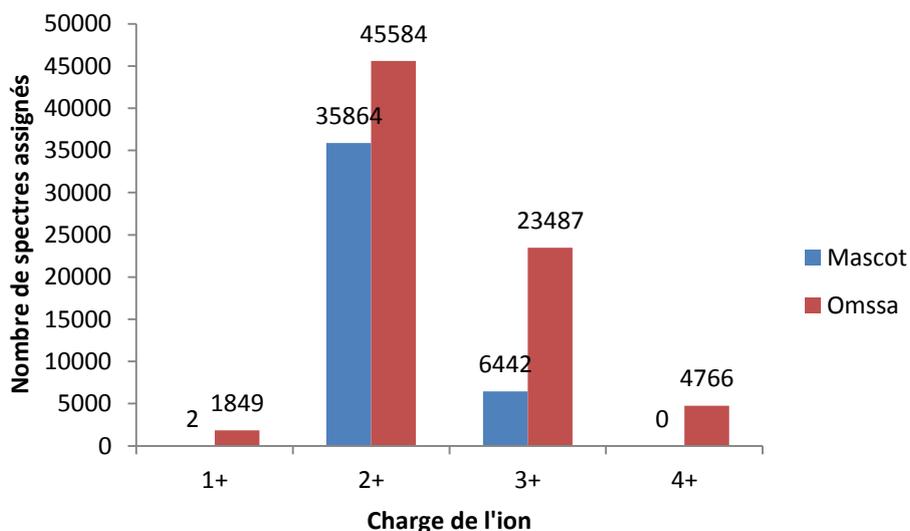


Figure 77

Nombre de spectres assignés à un peptide suivant la charge de l'ion et le moteur de recherche utilisé

Cet effet est occasionné par une option choisie dans les paramètres avancés du moteur de recherche Omssa. Ainsi, pour la détermination de l'état de charge du précurseur, il y a deux possibilités :

- « *Use the given range* » : si le spectre n'est pas assigné, Omssa va tester tous les autres états de charge possibles (indiqués par l'utilisateur), pour voir si ce spectre peut correspondre à un peptide avec un état de charge différent de celui indiqué par le logiciel de déconvolution (ici Data Analysis, Bruker Daltonics).
- « *Read input file* » : Omssa ne va considérer que les états de charge indiqués par le logiciel de déconvolution et ne va pas chercher à attribuer un état de charge différent. Cela implique qu'en cas de non-déconvolution par Data Analysis, le spectre concerné ne sera pas inclus dans les recherches dans les banques de données.

Pour valider ces observations, une comparaison a été effectuée entre les recherches Mascot et Omssa (en utilisant les deux options) avec deux types de *peak list* : une liste classique et cette même liste sans les états de charge indiqués par Data Analysis. Le tableau 7 montre que, lorsque l'option « *Input reading charge* » est cochée dans Omssa, aucune identification n'est réalisée si le logiciel de déconvolution n'a attribué aucune charge. Cette option va donc obliger le moteur à considérer uniquement les états de charge attribués par le logiciel de déconvolution. Par ailleurs, plus d'identifications sont réalisées à la fois par Mascot et Omssa (option « *Use the given range* »), si une *peak list* sans états de charge est utilisée.

	Options	<i>Peak list</i> classique	<i>Peak list</i> sans charges
Mascot		1635	2561
Omsa	Read input file	1869	0
	Use the given range	2119	3009

Tableau 7

Nombre d'identifications protéiques réalisées par les deux moteurs de recherche Mascot et Omsa

Avec une *peak list* classique et cette *peak list* sans les états de charge donnés par le logiciel de déconvolution. Deux paramètres ont été testés pour Omsa : « Read input file » et « Use the given range ».

L'utilisation d'une *peak list* sans indication des états de charge issus de la déconvolution par Data Analysis semble ici être une perspective attrayante. Cependant, il ne faut pas oublier que les expériences ont été réalisées sur un appareil basse résolution. Nous devons donc mettre en parallèle deux considérations :

- Le logiciel a réalisé une mauvaise déconvolution et attribué un mauvais état de charge. Ne jamais considérer l'état de charge indiqué dans la *peak list* permet de s'affranchir de ce problème et donc de réaliser plus d'identifications.
- La tolérance sur le précurseur et les fragments étant élevée, le nombre d'identifications possibles augmente. Si on ne considère pas les états de charge indiqués par le logiciel, la probabilité de réaliser de fausses identifications augmente aussi.

Dans le cadre du projet concerné par cette analyse (cf. Partie III – chapitre 2), il a été estimé préférable de placer une plus grande confiance dans les résultats de déconvolution donnés par Data Analysis et de ne pas chercher à modifier les états de charge lors de la recherche dans les banques de données.

Cet exemple illustre très bien l'importance de l'utilisation de deux moteurs de recherche, qui permet de limiter les erreurs et de mieux appréhender l'influence de chaque paramètre sur les résultats d'identifications.

C. Conclusions

Cette partie a permis :

- de mettre en évidence l'importance de l'évaluation des paramètres de recherche dans les banques de données pour optimiser le nombre d'identifications obtenues ;
- de montrer l'intérêt de combiner plusieurs moteurs de recherche, à la fois pour augmenter la confiance accordée aux identifications et pour maximiser le nombre d'identifications obtenues.

CHAPITRE 4

Quantification sans marquage pour la recherche de biomarqueurs

La quantification sans marquage est une technique attractive pour la recherche de biomarqueurs dans des échantillons cliniques¹³². En effet, elle peut s'appliquer à tous types d'échantillons (fluides, cellules, lignées cellulaires, tissus) et permet d'en analyser un nombre important. À titre de comparaison, le marquage métabolique SILAC nécessite une mise en culture des lignées dans des milieux enrichis en réidus d'acides aminés marqués. De plus, il ne permet d'analyser que douze conditions différentes avec un spectromètre de très haute résolution ($R_s = 480\,000$)¹³⁶. Par ailleurs, les technologies sans marquage sont moins coûteuses et plus simples à mettre en place²³⁰⁻²³² (pas de pré-traitement de l'échantillon nécessaire, développement d'une méthode d'acquisition par spectrométrie de masse simplifiée).

Deux approches sont envisageables (cf. figure 78)²³³ :

- **MS1** : L'extraction des courants d'ions des scans MS
- **MS2** : Le comptage du nombre de spectres MS/MS ou l'extraction des courants d'ions des scans MS/MS

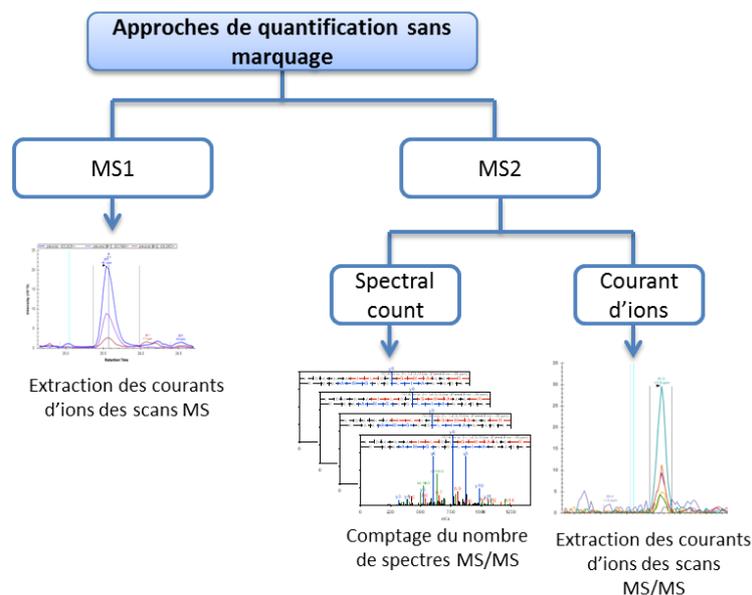


Figure 78

Schéma représentant les deux approches envisageables lors de la quantification sans marquage

Pour chaque projet présenté dans ce travail de thèse, la méthode de quantification a été adaptée à l'instrumentation et au protocole mis en place, ainsi qu'aux résultats attendus.

MS1 et MS2 sont les deux méthodes principales qui ont été mises en œuvre.

A. Quantification MS2 : le *spectral count*

En 2004, Liu et coll.²³⁴ montraient une corrélation entre le nombre de spectres MS/MS et la quantité de protéines contenue dans un échantillon. Depuis, le *spectral count* n'a cessé de gagner en popularité, comme en témoignent les nombreuses publications y faisant référence²³⁵⁻²³⁹. Cette technique sans marquage présente un intérêt majeur : le traitement des données y est largement simplifié. Par ailleurs, elle est très bien adaptée pour les instruments de basse résolution qui ne permettent pas d'extraire des courants d'ions avec une précision importante¹³³.

Cependant, cette technique est limitée par le mode d'acquisition DDA qui engendre une non-répétabilité de la sélection des ions, et notamment des ions les moins abondants. En outre, l'exclusion dynamique paramétrée afin de ne pas sélectionner de manière trop récurrente les ions les plus abondants permet certes d'augmenter la couverture d'identification mais empêche dans le même temps une estimation précise de la quantité de ces mêmes ions¹³³. Ces deux paramètres inhérents à l'analyse protéomique engendrent des valeurs manquantes difficiles à pallier lors d'analyses statistiques. Enfin, reste la question des peptides partagés²⁴⁰ : comment attribuer de la manière la plus juste possible les spectres partagés entre plusieurs protéines ? Plusieurs solutions ont été proposées, chacune comportant des biais : ne pas les prendre en compte et réaliser le calcul uniquement sur les peptides protéotypiques, c'est-à-dire uniques à la protéine, les attribuer plusieurs fois ou encore effectuer une pondération, par exemple suivant la quantité de protéines présente dans l'échantillon.

Ces limitations engendrent des variations dans le nombre de spectres comptés pour chaque protéine. C'est pourquoi des techniques de normalisation ont été mises en œuvre :

- Normalisation par le nombre total de spectres assignés au cours de l'analyse
- Normalisation par une protéine connue pour ne pas être régulée dans l'échantillon
- Normalisation par des facteurs physico-chimiques : nombre de peptides tryptiques de la protéine (*emPAI*, *exponentially Modified Protein Abundance Index*)^{241, 242}, probabilité de détecter les peptides dans l'échantillon suivant leur longueur et leur composition en résidus d'acides aminés (*APEX*, *Absolute Protein Expression*)^{243, 244}, longueur de la protéine (*NSAF*, *Normalized Spectral Abundance Factor*)²⁴⁵
- Prise en compte de l'intensité totale des fragments en MS/MS (*SI*, *Spectral Index*)²⁴⁶

Afin d'évaluer la pertinence des résultats, des tests statistiques peuvent être utilisés^{233, 247-250} :

- Le *t*-test (test de Student) permet de mesurer la distance entre la moyenne des réplicats d'une catégorie et la moyenne des réplicats d'une autre catégorie. Le résultat est donné sous forme de probabilité : la valeur *p*. Cette valeur donne la probabilité que la distance entre les deux catégories soit due au hasard. La méthode nécessite au moins six réplicats et l'hypothèse d'une distribution normale.
- Le LPE (*local pooled error*) est une variante du *t*-test qui ne nécessite que deux réplicats.
- Le G-test et le test exact de Fisher sont basés sur des distributions multinomiales et hypergéométriques et ne nécessitent qu'un seul réplicat. En cas de multiples injections du même échantillon, les réplicats sont groupés^{251, 252}.
- Le beta-binomial test^{253, 254} : ce test tenant compte de la variabilité inter et intra échantillons a été récemment implémenté au laboratoire, au sein de la plateforme MSDA.

Dans cette partie, la qualité du *spectral count* sur la trappe ionique AmaZon sera discutée.

A.1. Répétabilité des analyses

Afin d'évaluer la qualité du *spectral count* de l'AmaZon, un digestat de levure dopé avec des quantités croissantes de BSA (10, 20, 50, 100, 500 et 1000 fmol) a été injecté en triplicat.

Dans l'échantillon analysé, les protéines de levure ont un niveau d'expression constant. Dans un premier temps, la répétabilité du *spectral count* a été évaluée en utilisant diverses techniques de normalisation proposées par Scaffold™ (Proteome Software Inc., Portland, OR)²⁵⁵ : pas de normalisation, normalisation par le nombre total de spectres, techniques NSAF et emPAI. La figure 79 présente le coefficient de variation médian obtenu sur les 18 injections du digestat de levure suivant les quatre types de normalisation utilisés et les filtres imposant un nombre minimum de spectres pour chaque protéine (aucun, nombre de spectres > 2 dans au moins une analyse et nombre de spectres > 3 dans au moins une analyse).

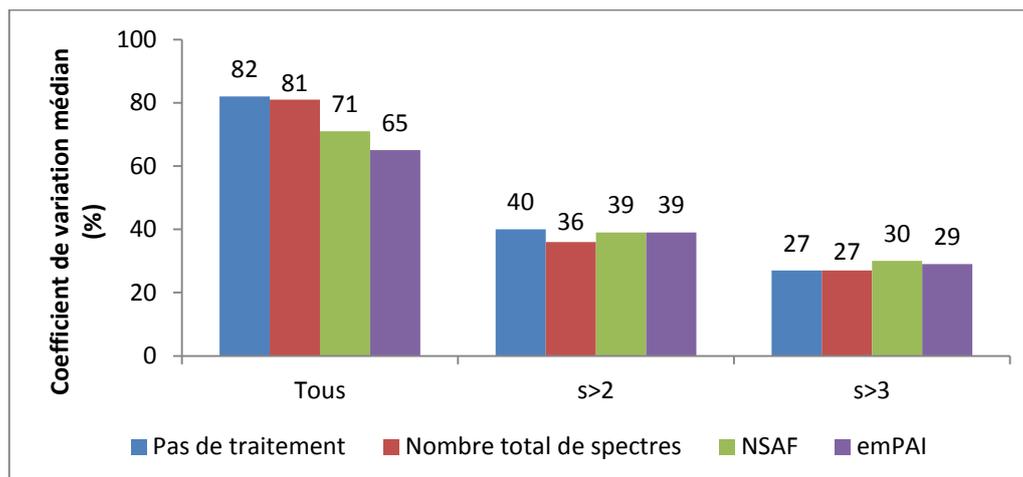


Figure 79

Coefficient de variation médian obtenu sur 18 injections du même échantillon de levure sur la trappe ionique Amazon suivant le type de normalisation utilisé

Quatre types de normalisation sont ici présentés (aucune, nombre total de spectres, NSAF et emPAI). Des filtres ont été appliqués sur les protéines : (1) Toutes les protéines ont été considérées (Tous), (2) uniquement les protéines avec un nombre de spectres supérieur à deux dans au moins une analyse ($s > 2$) et les protéines avec un nombre de spectres supérieur à 3 dans au une analyse ($s > 3$).

Les coefficients de variation médians obtenus sont très élevés si aucun filtre n'est appliqué sur le nombre de spectres. Cette observation confirme la difficulté d'évaluer la quantité des protéines peu abondantes. Si des filtres autorisant uniquement les protéines avec un nombre minimum de spectres de 2 et de 3 sont appliqués, les coefficients de variation médians atteignent un niveau correct. Il est important de noter ici que l'application des filtres sur le nombre de spectres entraîne une diminution importante du nombre de protéines quantifiables : 377 protéines identifiées au total, 205 si le nombre de spectres minimum est supérieur à 2 et 131 si supérieur à 3. Ces données sont corrélées avec les valeurs rapportées dans la littérature (% CV médian de 16 à 38 %) ^{233, 256}.

Par ailleurs, aucune différence notable ni amélioration des résultats n'est observée à la suite d'application de différents types de normalisation. Cela ne semble pas aberrant dans la mesure où les analyses ont été réalisées sur une seule journée, réduisant ainsi au maximum les possibilités de variation.

Par la suite, afin d'obtenir des résultats les plus fiables possibles, aucune normalisation ne sera appliquée et les protéines quantifiées et étudiées devront présenter un minimum de trois spectres dans au moins l'un des échantillons.

A.2. Quantification de la BSA

Dans un deuxième temps, la gamme de BSA a été considérée. La figure 80 présente le nombre de spectres obtenus suivant la quantité de BSA dopée. Une corrélation est bien observée entre les deux paramètres ($R^2=0,98$). Cependant, notons que pour les ajouts de 10 et 20 fmol, aucun peptide de BSA

n'est détecté. Le *spectral count* est ici limité par le mode d'acquisition DDA qui défavorise les peptides les moins abondants.

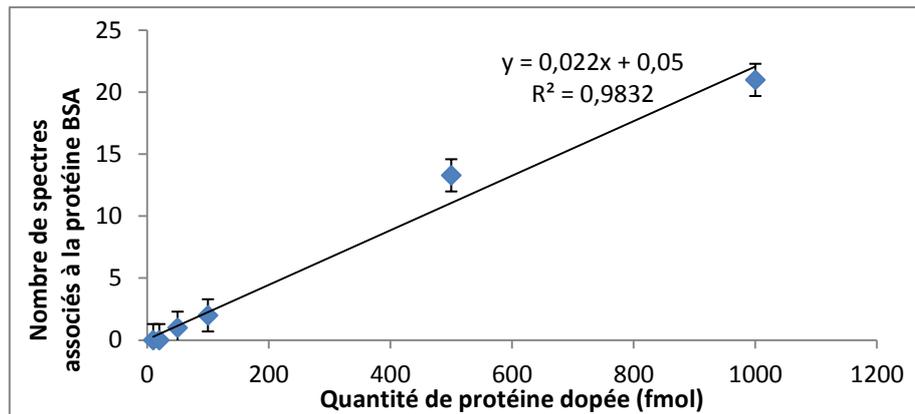


Figure 80

Corrélation entre la quantité de protéine BSA dopée dans le digestat de levure et le nombre de spectres associés à la protéine BSA

Au laboratoire, le test beta-binomial a été implémenté pour effectuer des traitements statistiques des données. La variabilité est modélisée suivant deux directions : la variation intra-échantillon est modélisée par une distribution binomiale (bien adaptée pour les valeurs discrètes de *spectral count*) et la variation inter-échantillons est modélisée en traitant les paramètres de la distribution β -binomiale comme des valeurs aléatoires issues d'une distribution β .

Afin d'évaluer la pertinence de ce test statistique, certains échantillons ont été comparés deux à deux : échantillon 100-1000 fmol (ratio de 10), échantillon 100-500 fmol (ratio de 5) et échantillon 500-1000 fmol (ratio de 2). Les « graphiques volcano » (*volcano plot*) obtenus pour les différents ratios considérés sont présentés en figure 4. La BSA est détectée sans ambiguïté, significativement différente pour des ratios de 5 et 10 (cf. figure 81B et C). Pour un ratio de 2, la différence est significative en termes de valeur p , par contre, au niveau du ratio, elle ne se distingue pas des protéines non régulées (cf. figure 81A). Par conséquent, la validation des ratios de quantité 2 apparaît ici délicate.

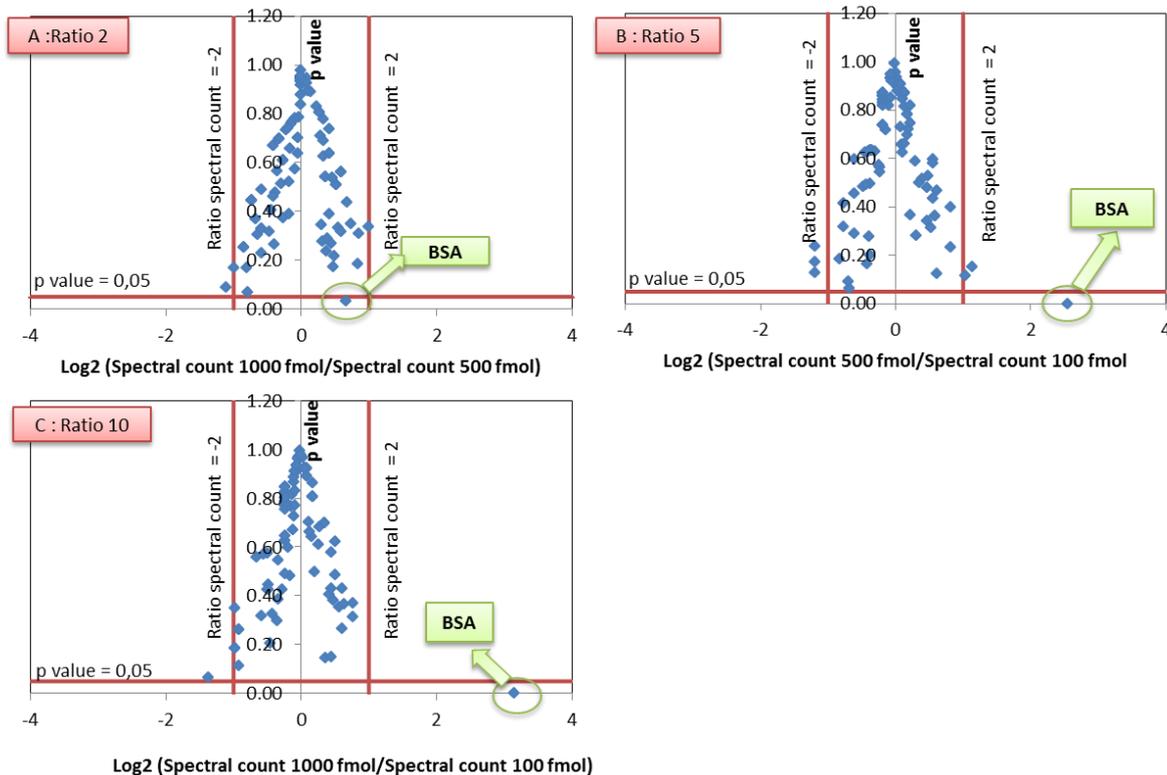


Figure 81

Graphiques "volcano" issus de la comparaison deux à deux du *spectral count* d'échantillons d'un digestat de levure dopé avec différentes quantités de BSA

(A) ratio 2 : *spectral count* 1000 fmol/500 fmol; (B) ratio 5 : *spectral count* 500 fmol/100 fmol; (C) ratio 10 : *spectral count* 1000 fmol/100 fmol.

Par ailleurs, un second écueil majeur est à considérer. En effet, le test ne permet pas de traiter les valeurs manquantes. Deux solutions peuvent être envisagées : ne pas tenir compte des protéines possédant des valeurs manquantes ou attribuer par défaut une valeur de 1. La première solution n'est pas idéale dans le cadre de la recherche de biomarqueurs. En effet, les détections absent/présent restent les plus simples à mettre en évidence. La deuxième solution va quant à elle introduire des biais : une protéine non détectée ou non présente sera considérée comme identifiée.

B. Quantification MS1 : l'extraction des courants d'ions

Depuis la preuve de principe en 2002, par Chelsius et coll.¹³⁴, la détermination de la concentration relative d'une protéine par extraction des courants d'ions est une technique de quantification très souvent employée dans l'étape découverte de la recherche de potentiels biomarqueurs candidats²⁵⁷⁻²⁶⁴. Cette dernière repose sur l'intégration de l'aire sous le pic chromatographique d'un peptide donné pour une masse et un temps de rétention spécifiques²³¹. Ainsi, même si le peptide a été sélectionné et identifié par recherche dans les banques de données lors d'une seule analyse, les courants d'ions peuvent être extraits dans l'ensemble des analyses. La concentration relative de la protéine dans chaque échantillon est ensuite déduite de la somme de celles des peptides.

Plusieurs points clés influencent la qualité de la quantification^{132, 230} :

- **La robustesse du système chromatographique** : en cas de non identification du peptide, la stabilité des temps de rétention assure un meilleur alignement entre les analyses et facilite le choix du pic à intégrer. De plus, une résolution chromatographique élevée permet d'augmenter la gamme dynamique et de quantifier des peptides peu abondants. Enfin, l'efficacité de l'ionisation est influencée par la coélution des peptides et peut ainsi différer d'une analyse à l'autre en cas de non-répétabilité.
- **La résolution de l'appareil de spectrométrie de masse** : une résolution élevée facilite l'extraction des courants d'ions et limite les interférences causées par la coélution de peptides de mêmes masses. Dans le logiciel Skyline, il est possible de définir la résolution de l'instrument. Afin d'évaluer de manière détournée l'importance de ce paramètre, quatre répliquats d'analyse d'un échantillon de microparticules plasmatiques (5 400 peptides, soit 1 000 protéines) ont été traités avec une résolution à 10 000, 20 000, 30 000 et 40 000. Les figure 82 et figure 83 montrent l'importance de la résolution, qui facilite l'intégration par une augmentation de la sensibilité et permet ainsi de diminuer la variabilité entre les analyses.

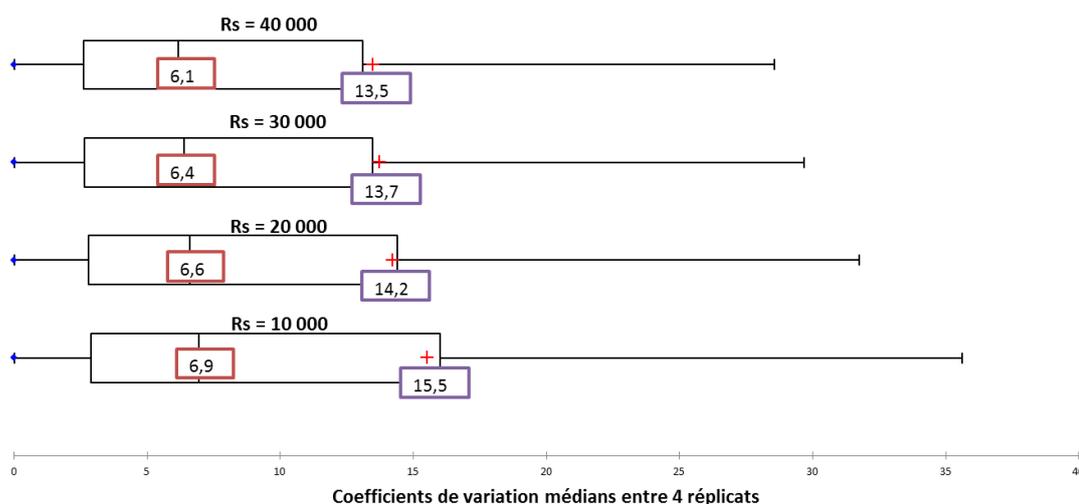


Figure 82

Boîtes à moustaches montrant la répartition des valeurs des coefficients de variation issus de quatre répliquats d'analyse d'un échantillon de microparticules plasmatiques (environ 800 protéines) suivant la résolution instrumentale (10 000, 20 000, 30 000 et 40 000)

Les encadrés rouges donnent la valeur de la médiane, les mauves, la moyenne.

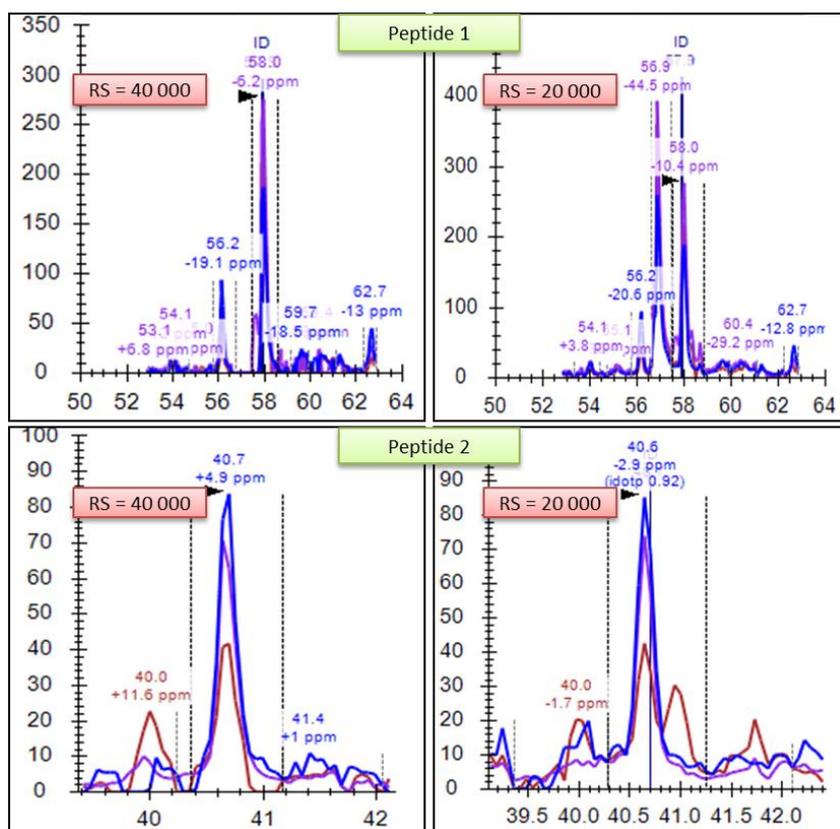


Figure 83

Exemples d'intégrations facilitées par l'utilisation de la haute résolution

Comparaison pour deux peptides issus de l'analyse d'un échantillon de microparticules plasmatiques suivant la valeur de la résolution instrumentale (20 000 et 40 000).

- **La précision de masse** : une précision de masse élevée permet d'assurer un maximum de confiance dans la séquence peptidique identifiée.

Par ailleurs, les spécificités du mode DDA impliquent un compromis entre le temps de scan MS et le nombre de scans MS/MS. En effet, d'un côté il est primordial d'effectuer un nombre élevé de scans MS afin de définir au mieux les pics chromatographiques. D'un autre côté, il est nécessaire de réaliser un nombre de scans MS/MS suffisant pour assurer un grand nombre d'identifications.

Récemment, l'arrivée au laboratoire d'un spectromètre de masse Q-TOF de dernière génération (Maxis Impact, Bruker Daltonics) a permis d'envisager le développement de méthodes de quantification basées sur l'extraction des courants d'ions. En effet, cet appareil est caractérisé par une vitesse de scan accrue (jusqu'à 25 Hz), une meilleure sensibilité (augmentation de la gamme dynamique jusqu'à 3 log) ainsi qu'une haute résolution ($R_s = 40\,000$) et une haute précision de masse (5 ppm).

Contrairement au *spectral count*, l'extraction des courants d'ions nécessite un traitement informatique important des données. Plusieurs étapes sont cruciales^{265, 266} :

- l'élimination du bruit de fond ;
- la détection des ions et leur intégration sur le massif isotopique ;
- l'alignement des temps de rétention entre les différentes analyses ;

- la normalisation²⁶⁷ des données pour tenir compte de la variabilité de l'intensité des signaux due à des erreurs systématiques dans les expériences ou encore dans la préparation des échantillons. Il est important ici d'établir une liste systématique des biais. En effet, une normalisation non appropriée des données peut conduire à l'identification de faux positifs significatifs²⁶⁸.

Pour répondre à ces besoins, plusieurs logiciels commerciaux et *open source* ont été développés : SIEVE (<http://www.thermo.com>), Progenesis (<http://www.nonlinear.com>), GE Decyder MS (<http://www.gelifesciences.com>), Max Quant²⁶⁹, Skyline²⁷⁰, OpenMS²⁷¹... Le laboratoire s'est orienté vers l'utilisation de Skyline, développé par le laboratoire de Michael J. MacCoss. Ce logiciel présente en effet plusieurs avantages majeurs :

- Il est *open source* et compatible avec des données issues de n'importe quel constructeur.
- Il permet de visualiser très facilement les données d'intégration. L'utilisateur a ainsi la possibilité de vérifier manuellement ce qui a été intégré (ce n'est pas le cas pour Max Quant par exemple).
- L'interaction avec les développeurs est très facile.

Une fois les résultats de quantification obtenus, la question épineuse des statistiques se pose : comment identifier les différences significatives entre les nombreux échantillons analysés et parmi l'ensemble des protéines identifiées ? Plusieurs tests statistiques ont été adaptés pour l'analyse protéomique²⁷² : le test de Student, le test exact de Fisher ou encore le modèle linéaire mix²⁷³. Ce dernier, particulièrement adapté à la spécificité des analyses protéomiques, permet de compenser les variations techniques et biologiques et d'adapter le traitement des valeurs manquantes.

Au cours de ce chapitre, nous montrerons les résultats produits lors d'une évaluation de la répétabilité de la quantification sur le couplage nanoLC-MS/MS Maxis Impact ainsi que la linéarité qu'il est possible d'obtenir. Puis nous nous attacherons à présenter le logiciel Skyline ainsi que les divers paramètres qui influencent la qualité de la quantification.

B.1. Linéarité de réponse de l'appareil

Dans un premier temps, nous avons vérifié la linéarité du Q-TOF. L'objectif était de déterminer la gamme de linéarité de l'aire totale d'un peptide (somme des aires des trois premiers isotopes de chaque ion) sur laquelle il est possible de travailler en assurant une détection suffisante des ions et en évitant la saturation du détecteur.

Ce test a été réalisé directement sur un échantillon d'intérêt, à savoir les microparticules plasmatiques utilisées dans le cadre du projet Protell (cf. Partie III – chapitre 3). Un même échantillon a été déposé 12 fois sur un gel 1D SDS PAGE (cf. figure 84). La bande contenant les protéines a été découpée puis les échantillons ont été réduits, alkylés, digérés, extraits et dopés avec différentes quantités de peptides iRT avant lyophilisation. Les échantillons ont ensuite été repris dans un même volume puis injectés sur le Maxis Impact.

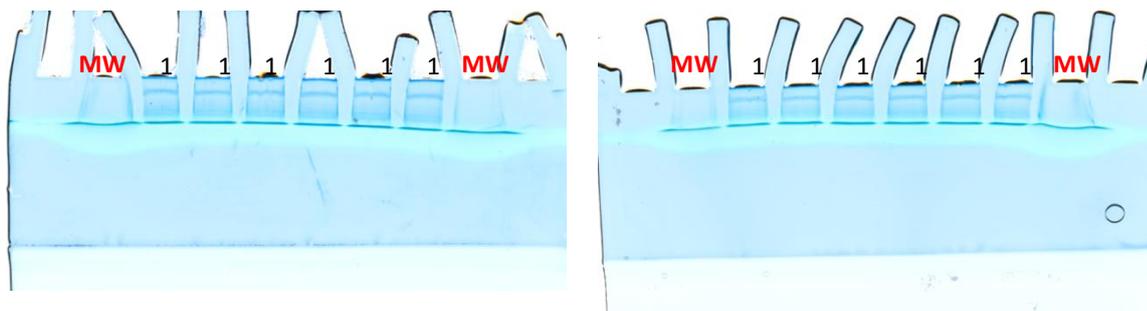


Figure 84

Photo des deux gels sur lesquels ont été déposés 12 fois un même échantillon de microparticules plasmatiques

Le standard iRT, commercialisé par la société Biognosys, a ici une utilisation détournée de celle prévue initialement. Il est constitué de 11 peptides synthétiques non identiques à une séquence connue, c'est-à-dire non rapportés dans la banque de recherche NCBI et répartis sur l'ensemble du gradient chromatographique. Ces peptides permettent de normaliser les temps de rétention et donc, par exemple, d'adapter une méthode à un nouveau gradient.

Dans le cas présent, une gamme a été réalisée à partir des iRT²⁷⁴, qui ont été dopés dans chaque échantillon en duplicat aux quantités suivantes : 0,2 - 0,5 - 0,75 - 1 - 3 - 5.

Les résultats obtenus sont présentés en figure 85. Parmi les 9 peptides qui ont pu être détectés et quantifiés, 8 présentent un excellent coefficient de corrélation sur deux log d'aire totale, de 1×10^6 à 1×10^8 . La non-linéarité du peptide FLLQFGAQGSPLFK s'explique par son hydrophobicité. En effet, les peptides très hydrophobes sont caractérisés par une grande variabilité car ils présentent une solubilité plus faible et peuvent s'agréger.

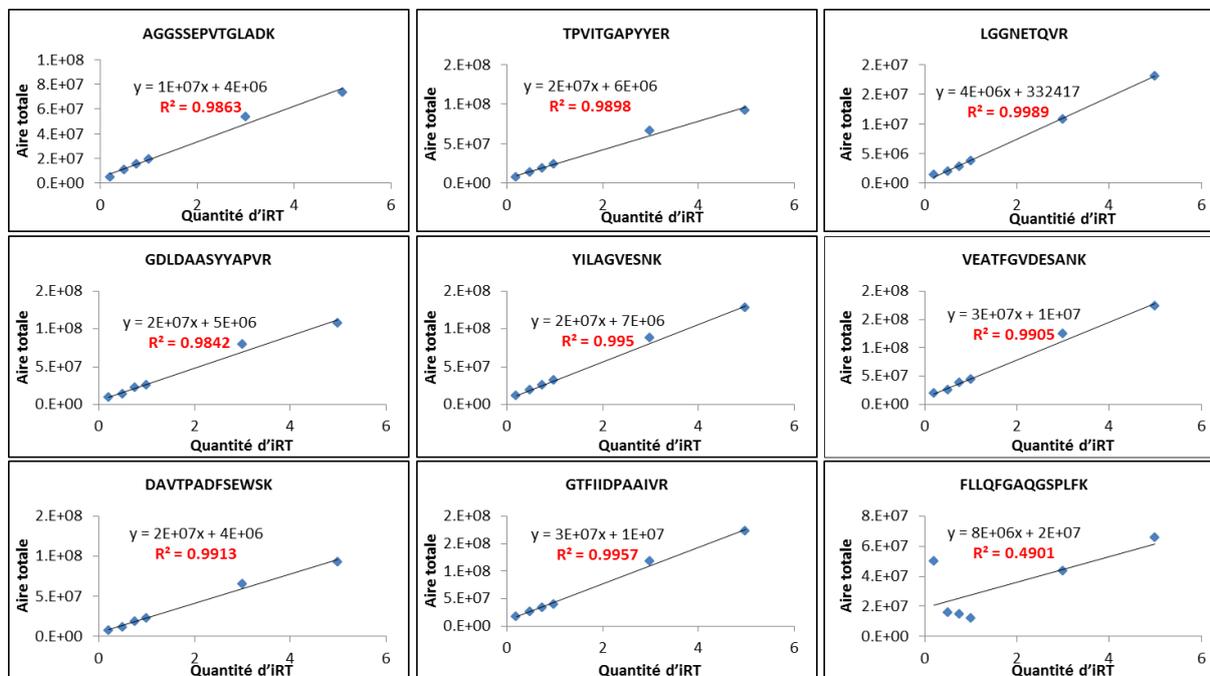


Figure 85

Régression linéaire de l'aire totale (somme des aires des trois premiers isotopes) pour chaque peptide en fonction de la quantité d'iRT dopée dans l'échantillon

L'équation de la droite ainsi que le coefficient de corrélation sont indiqués.

B.2. Répétabilité de la quantification

La répétabilité de la quantification a été évaluée sur l'ensemble des peptides identifiés (sauf les IRT) dans l'échantillon (environ 5 400 peptides identifiés soit 1 000 protéines).

Dans ce cas, il n'a pas été nécessaire d'effectuer de normalisation. En effet, les analyses n'ont duré que deux jours. La figure 86 le confirme dans un graphique MA. Un graphique MA est une représentation permettant d'évaluer la nécessité de normaliser les données. Ce graphique a été tracé pour des analyses injectées à des temps éloignés : ajout 0,5 - injection 1 et ajout - injection 2. Il montre bien que les données sont centrées sur 0 et qu'aucune normalisation n'est nécessaire entre les deux conditions étudiées.

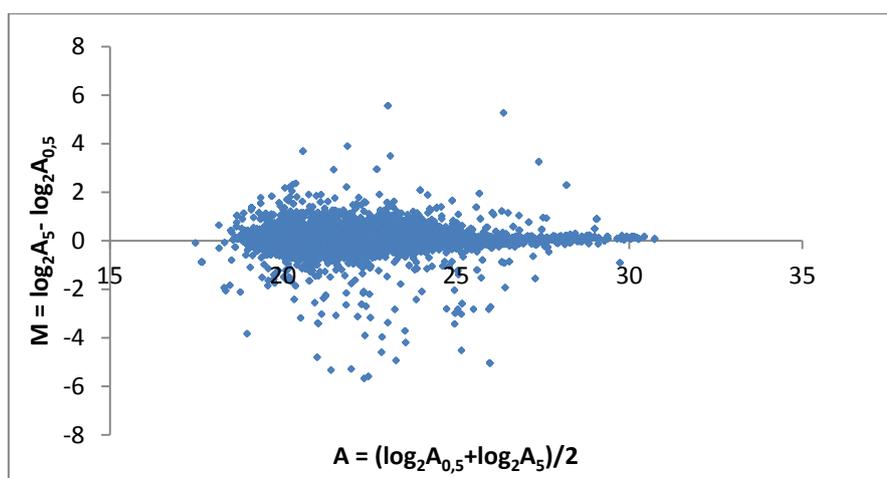


Figure 86

Graphique MA plot pour l'ensemble des peptides identifiés dans les conditions 0,5 et 5

En ordonnée, ratio des aires totales des deux conditions, et en abscisse, moyenne des aires des deux conditions.

Les coefficients de variation de l'aire totale de chaque peptide entre les 12 injections ont été calculés et représentés sous la forme d'une boîte à moustaches (cf. figure 87). Leur valeur médiane est de 14 % et leur moyenne, de 18 %. Ces résultats montrent une excellente répétabilité de la quantification.

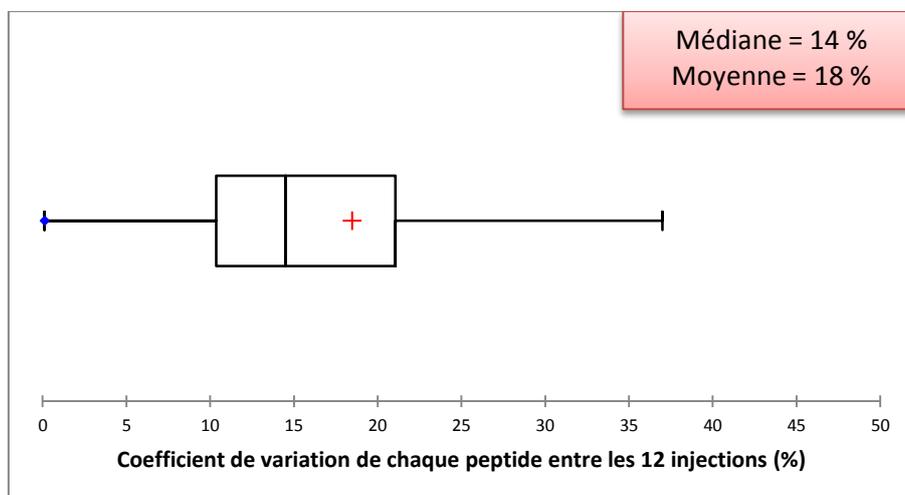


Figure 87

Boîte à moustaches des coefficients de variation observés pour chacun des peptides entre les 12 réplicats

B.3. Paramètres influençant la quantification

Comme décrit précédemment, le choix du laboratoire s'est porté sur le logiciel Skyline pour la quantification de protéines sans marquage.

Skyline extrait l'intensité des x premiers isotopes (définis par l'utilisateur) au cours du temps. La fenêtre d'extraction est calculée selon la résolution de l'instrument et correspond à deux fois la largeur à mi-hauteur (cf. figure 88). L'utilisateur peut ensuite visualiser une trace chromatographique pour l'ensemble des isotopes de chaque peptide obtenu. Un *isotopic dot product (idotp)* est calculé, permettant ainsi de comparer la distribution observée à la distribution attendue (score de 0 à 1, 1 représentant la plus forte corrélation).

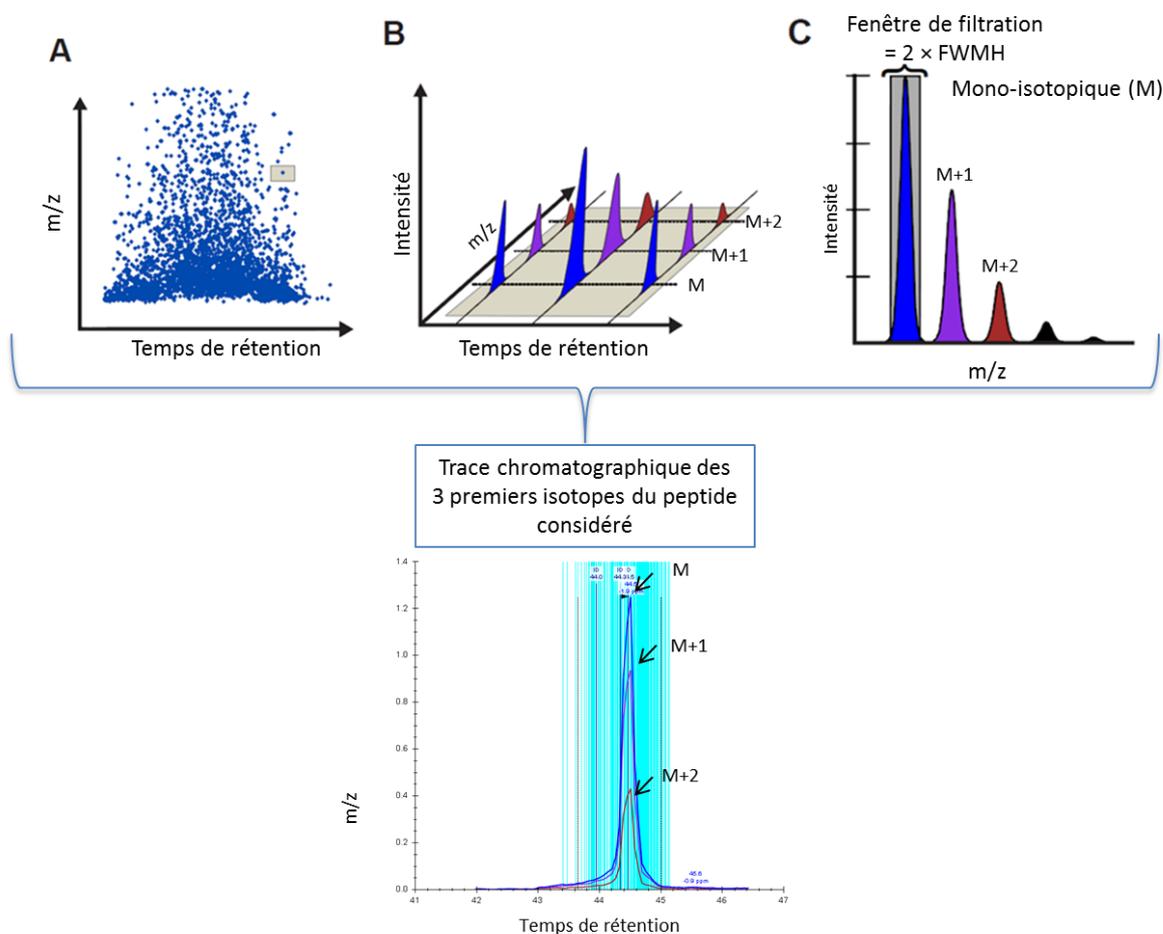


Figure 88

Schéma représentant le mode de fonctionnement de Skyline pour l'extraction des courants d'ions

Skyline repère un composé d'intérêt à un certain temps de rétention (A), extrait les trois premiers isotopes sur une certaine fenêtre de temps de rétention (B), puis calcule la somme des intensités de chaque extraction pour chaque isotope sur le temps de rétention défini (C) afin d'obtenir une trace chromatographique des trois premiers isotopes du peptide considéré.

La particularité de l'extraction de Skyline réside dans l'utilisation de bibliothèques spectrales. Ces dernières sont obtenues lors de l'assignation des spectres par les moteurs de recherche, par exemple Mascot. L'ensemble des spectres des ions identifiés est rassemblé dans une bibliothèque, et les redondances sont conservées. Ainsi, pour chaque analyse, une fenêtre de temps est établie selon des critères définis par

l'utilisateur, par exemple ± 3 min, autour du temps de rétention auquel le peptide a été sélectionné pour la MS/MS. Cette fenêtre est ensuite associée au rapport masse/charge (m/z) du peptide considéré afin de détecter le pic à intégrer. Cette technique permet par ailleurs de rechercher le pic à intégrer dans les analyses où le peptide n'a pas été identifié (cf. figure 89). Il est important de noter ici que Skyline ne réalise pas de réalignement en temps de rétention, contrairement à d'autres logiciels tels que MaxQuant.

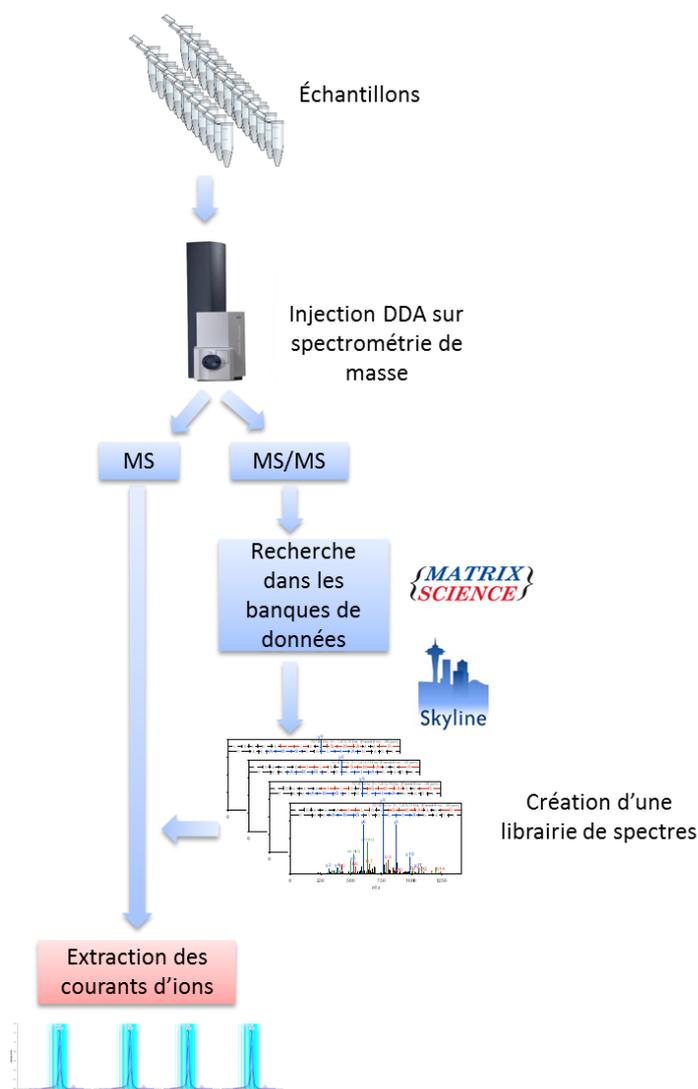


Figure 89

Étapes consécutives de la quantification sans marquage lors de l'utilisation de Skyline

Une fois analysés, les spectres MS/MS sont identifiés et une bibliothèque spectrale est créée. Cette dernière est ensuite utilisée pour cibler l'extraction des courants d'ions de la MS.

Afin d'étudier l'influence de différents paramètres sur les résultats obtenus, et notamment sur les moyennes et médianes des coefficients de variation, 4 réplicats issus des 12 réplicats réalisés précédemment ont été comparés.

Dans un premier temps, la variation de la quantification entre les quatre réplicats a été évaluée sur trois niveaux : isotopique, peptidique et protéique. La figure 90 montre les résultats obtenus. Les valeurs de médiane et de moyenne sont sensiblement améliorées lorsque la quantification est réalisée au niveau

protéique. Ce constat peut s'expliquer par le fait que les peptides les moins abondants et les plus faibles en intensité sont ceux qui présentent en général une variation plus élevée. Par contre, du fait de leur faible abondance, leur impact sur la quantification de la protéine est faible.

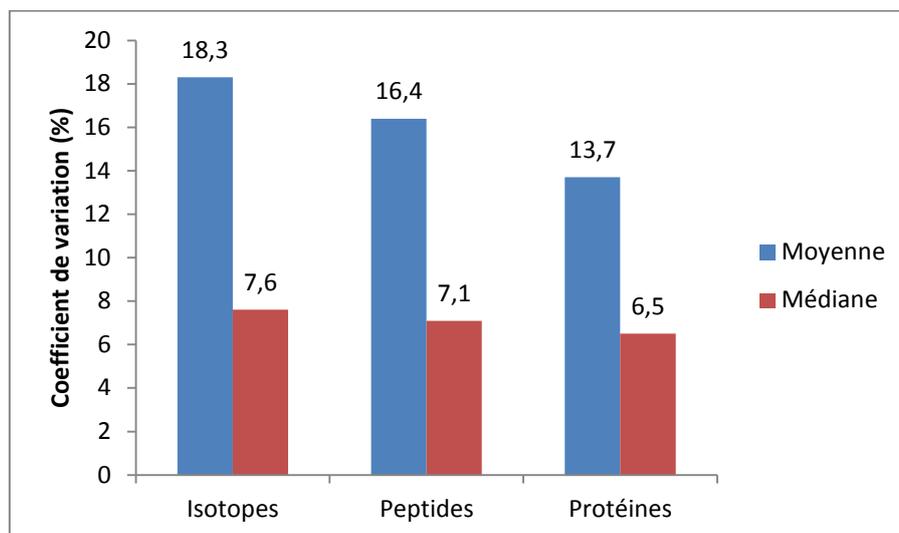


Figure 90

Coefficients de variation moyen et médian obtenus pour chaque niveau de quantification, isotopique, peptidique et protéique

Dans le cas présent, la quantification au niveau protéique a été réalisée en sommant l'intensité de l'ensemble des peptides identifiés pour les protéines considérées. Plusieurs facteurs peuvent expliquer un coefficient de variation élevé pour un peptide donné :

- erreur du logiciel d'extraction (extraction du mauvais pic ou mauvaise intégration) ;
- présence d'une interférence au même temps de rétention et à la même masse ;
- instabilité du peptide du fait de modifications post-traductionnelles, de la nature hydrophile ou hydrophobe de celui-ci.

Afin d'améliorer les résultats obtenus, plusieurs filtres peptidiques permettant de limiter le nombre de peptides à forte variation ont été testés :

- suppression des peptides avec un *isotope dot product* < 0,9 afin de s'assurer une qualité d'identification et d'intégration : le pic intégré est correct et il n'y a pas d'interférence ;
- suppression des peptides à méthionine pour s'affranchir des problèmes liés à l'oxydation ;
- suppression des peptides à cystéine (carbamidométhylation).

Les résultats obtenus sont présentés sous forme de boîtes à moustaches dans la figure 91. La médiane et la moyenne des coefficients de variation sont très faiblement améliorées si les peptides à méthionine ne sont pas considérés. De plus, cela occasionne la suppression de 30 % des peptides quantifiés. Lors de la suppression des peptides à méthionine et à cystéine, les résultats sont en revanche nettement améliorés. Mais, une diminution de 62 % de peptides quantifiés est observée. Dans le cas de ces

analyses, il apparaît donc peu judicieux de supprimer les peptides modifiés. En effet, les résultats sont très faiblement améliorés alors que beaucoup de peptides sont supprimés.

Le filtre sur *idotp* permet quant à lui d'améliorer de façon non négligeable les résultats, tout en ne diminuant pas de manière excessive le nombre de peptides identifiés (20 % en moins).

Idotp permet ici d'associer un score à la qualité de l'intégration et est donc particulièrement bien adapté pour filtrer les grandes séries de données. Cependant, un bémol est à évoquer. En effet, dans le cadre de la recherche de biomarqueurs, les protéines les plus intéressantes restent celles qui sont absentes dans une condition et présentes dans l'autre. Le filtre sur les *idotp* éliminera l'ensemble des peptides de ces protéines car il y a peu de chance que l'intégration du bruit de fond corrèle avec les rapports isotopiques théoriques.

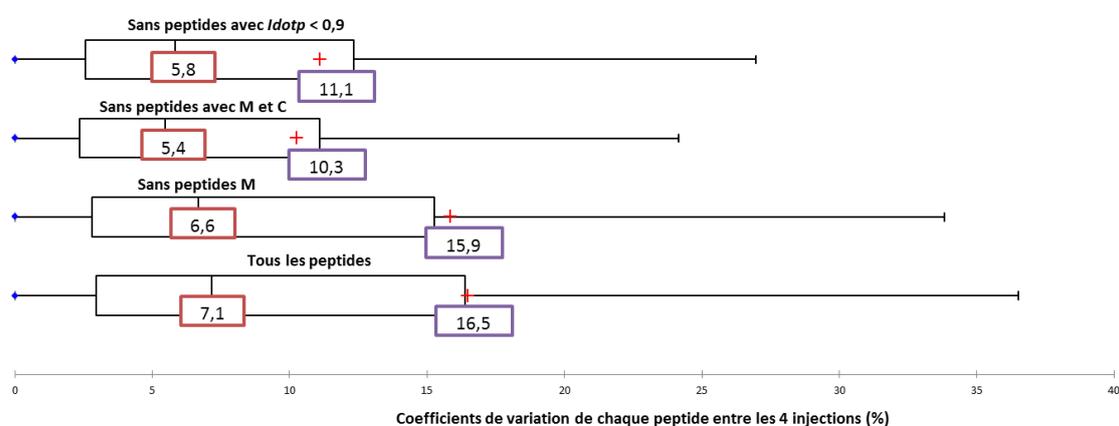


Figure 91

Boîtes à moustaches montrant la variation de quantification sur les peptides entre les 4 injections d'un échantillon de microparticules plasmatiques suivant différents filtres : sans les peptides à méthionine, sans les peptides à cystéine, sans les peptides avec un *idotp* < 0,9

Les valeurs médianes sont indiquées dans l'encadré rouge et les valeurs moyennes, dans l'encadré mauve.

C. Conclusions

Les techniques de quantification des protéines sans marquage connaissent aujourd'hui un succès croissant, de par leur simplicité de mise en œuvre et leur faible coût. Cependant, chacune des deux techniques présentées au cours de ce chapitre possède des inconvénients qui nécessitent une étape de validation des différences de quantité mises en évidence entre chaque échantillon. Des techniques de quantification basées sur le mode d'acquisition DIA (*data independent analysis*) émergent aujourd'hui. Elles présentent le grand avantage de s'affranchir des problèmes relatifs à la DDA.

Dans le cadre de la recherche de biomarqueurs, la quantification sans marquage doit être considérée comme une étape préliminaire permettant d'émettre des hypothèses et doit impérativement être suivie d'une étape de validation par des méthodes plus ciblées telles que la SRM ou encore la cytométrie en flux.

Dans le cadre de la recherche de biomarqueurs, l'identification et la quantification la plus exhaustive possible des protéines contenues dans un échantillon sont capitales. Les protéines les plus intéressantes sont souvent celles qui sont faiblement exprimées et les plus difficiles à détecter par spectrométrie de masse.

Cette partie souligne l'importance de l'évaluation et du développement de chaque étape lors de la mise en place d'une stratégie de protéomique globale. De la préparation d'échantillons, en passant par l'analyse nanoLC-MS/MS et la recherche dans les banques de données, à la quantification des protéines, chaque étape influe sur la qualité des résultats obtenus et sur la confiance que l'on peut leur accorder.

À l'heure actuelle, la spectrométrie de masse subit une véritable révolution. Les appareils sont de plus en plus rapides et de plus en plus sensibles. La multiplicité des opérations se développe. Ainsi, sur le nouvel appareil de Thermo Scientific, l'Orbitrap™ Fusion, il est possible d'utiliser trois modes de fragmentation différents (HCD, ETD et CID), de réaliser en parallèle les MS et MS/MS, de faire de la MSⁿ... En une heure, il a été possible d'obtenir la quasi-totalité du protéome de la levure²⁷⁵. Cette évolution est particulièrement intéressante pour la recherche de biomarqueurs dans la mesure où les médecins s'orientent de plus en plus vers des thérapies personnalisées et adaptées à chaque patient. L'apport de l'identification du protéome complet de chaque patient serait ici une avancée majeure.

Cependant, si les instruments deviennent de plus en plus performants, il reste encore un champ peu exploré : la statistique. Comment identifier les différences significatives parmi les milliers de protéines issues de l'analyse protéomique ? Lors d'analyses de larges cohortes, il est absolument nécessaire d'associer un statisticien à la fois à l'étape de construction des groupes ainsi qu'au traitement des données.

Partie III

Résultats : application à trois projets de recherche de biomarqueurs

L'ensemble des développements décrits dans la Partie II ont été appliqués à trois projets différents concernant la recherche de biomarqueurs par analyse protéomique :

- Le premier concerne la recherche de biomarqueurs à visée diagnostique dans le but d'améliorer la caractérisation des lymphomes des cellules B.
- Le second a pour but d'étudier les cellules souches cancéreuses des glioblastomes afin d'en proposer des biomarqueurs potentiels.
- Le troisième traite de la recherche de biomarqueurs prédictifs des rechutes du système nerveux central dans le cas de lymphomes diffus à grandes cellules B.

Les deux premiers projets ont requis une analyse protéomique en profondeur du fait du faible nombre d'échantillons ainsi qu'une validation stricte des résultats pour proposer les biomarqueurs candidats les plus pertinents possible.

Le dernier projet a différé par le grand nombre d'échantillons à analyser (125). Il a été nécessaire de développer une méthode d'analyse rapide associée à des contrôles qualité stricts tout au long des injections.

CHAPITRE 1

Recherche de biomarqueurs à visée diagnostique pour améliorer la caractérisation des lymphomes B

Cette étude a été réalisée en collaboration avec l'équipe du professeur Laurent Mauvieux, du pôle d'onco-hématologie des hôpitaux universitaires de Strasbourg et fait suite à un projet initié en 2004 durant la thèse du Dr Laurent Miguet (2002-2006) et poursuivi durant la thèse du Dr Guillaume Béchade (2006-2009).

Laurent Miguet a étudié différentes méthodes d'enrichissement en protéines membranaires et a proposé l'utilisation de microparticules induites à partir de lymphocytes B¹⁷². Guillaume Béchade a pu grâce à une analyse différentielle, proposer un biomarqueur candidat potentiel du lymphome du manteau : le CD148¹⁷¹. Ce marqueur a par la suite été validé par cytométrie en flux.

Le Dr Luc Fornecker, hématologue à l'hôpital de Strasbourg et doctorant au laboratoire depuis octobre 2012 et Sebastian Vaca, également doctorant au laboratoire depuis octobre 2012, ont été impliqués dans le projet à partir de 2013, notamment en ce qui concerne le développement de méthodes de spectrométrie de masse ciblées pour la validation des biomarqueurs.

A. De la nécessité de proposer de nouveaux biomarqueurs candidats des différents types de lymphomes des cellules B

Les lymphomes sont des maladies du sang caractérisées par une prolifération maligne de lymphocytes dans les organes lymphoïdes secondaires (rate, ganglions lymphatiques...). L'ensemble de ce chapitre porte sur l'étude des lymphomes non hodgkiniens à cellules B.

Les hémopathies lymphoïdes B chroniques représentent un ensemble de syndromes lymphoprolifératifs développés à partir de la lignée lymphoïde. Avec une incidence annuelle en augmentation de 5 à 10 % dans les pays occidentaux, ces pathologies représentent chaque année 5 % des nouveaux cas de cancer, soit environ 10 000 personnes par an en France (source : association France Lymphome Espoir).

A.1. Les difficultés du classement de ces pathologies

L'histoire naturelle et la survie globale des patients sont extrêmement hétérogènes selon les hémopathies. L'Organisation mondiale de la santé (OMS) en a établi une classification selon des caractéristiques morphologiques, génotypiques et immunophénotypiques²⁷⁶.

Pour un certain nombre de cas atypiques, la classification s'avère délicate du fait de caractéristiques cellulaires proches ou d'une absence de biomarqueurs. Cependant, malgré leurs similitudes, ces lymphomes peuvent avoir des prises en charge différentes, d'où la nécessité d'un diagnostic le plus précis possible^{277, 278}.

A.2. Le cas des lymphomes de la zone marginale (MZL)

Les lymphomes de la zone marginale (MZL)²⁷⁹ sont rares. Ils représentent 5 à 17 % de l'ensemble des lymphomes non hodgkiniens. Dans la plupart des cas, ils concernent des adultes d'un âge médian de 60 ans. Cette pathologie est indolente, dans le sens où son évolution est lente : il peut s'écouler plusieurs années entre le début de la maladie et l'apparition des premiers symptômes. Ces lymphomes sont originaires des lymphocytes B mémoires, normalement présents dans un compartiment cellulaire appelé la zone marginale des follicules lymphoïdes. Il en existe trois formes :

- le lymphome du tissu lymphoïde associé aux muqueuses extra-ganglionnaires (MALT) ;
- le lymphome de la zone marginale ganglionnaire (NMZL) ;
- le lymphome de la zone marginale splénique (SMZL).

Il est à noter que la première pathologie s'accompagne très rarement de cellules circulant dans le sang au contraire des deux suivantes.

À partir des lymphocytes circulants, le diagnostic des MZL peut s'avérer délicat²⁸⁰. En effet, il n'existe pas de marqueur permettant de les différencier des lymphomes du manteau (MCL) et des lymphomes lymphoplasmocytiques (LPL). Le tableau 8 décrit les caractéristiques²⁷⁶ de chaque pathologie. La translocation indiquée pour le lymphome de la zone marginale impliquant MALT1 (t(11 ;18)) concerne plus particulièrement les lymphomes du MALT et n'est donc pas toujours présente sur l'ensemble des lymphomes de la zone marginale.

Pathologie	Éléments phénotypiques discriminants	Autre information pour le diagnostic
Lymphome du manteau (MCL)	CD5 ⁺ , CD10 ⁻ , CD20 ^m , slg ^m , CD23 ^{+/-} , FMC-7 ^{+/-}	Translocation t(11,14) entraînant une surexpression du gène de la cycline D1
Lymphome de la zone marginale (MZL)	CD5 ⁺ , CD10 ⁻ , CD23 ⁻ , souvent CD11c ^{+/-} , CD103 ^{+/-} , parfois clg uniquement	Croissance autour et dans les follicules, translocation t(11;18) impliquant le gène MALT1
Lymphome lymphoplasmocytaire (LPL)	CD5 ⁺ , CD10 ⁻ , CD23 ^{-/f} , parfois slg ⁻ mais clg ⁺	Petites cellules, différenciation plasmocytaire

Tableau 8

Caractéristiques des hémopathies lymphoïdes B chroniques étudiées permettant le diagnostic et la classification

+ : habituellement positif ; - : habituellement négatif ; +/- : peut être positif ou négatif ; f : intensité faible ; m : intensité moyenne ; e : intensité élevée ; slg : immunoglobuline de surface ; clg : immunoglobuline cytoplasmique.

Source : d'après l'Organisation mondiale de la santé²⁷⁶

Le but du projet était ici double :

- proposer de nouveaux biomarqueurs candidats potentiels des lymphomes de la zone marginale afin d'aider au diagnostic de routine ;
- développer une méthode de protéomique quantitative permettant la validation de ces marqueurs.

Il est important de noter qu'à l'hôpital, la caractérisation des hyper-lymphocytoses est effectuée par cytométrie en flux²⁸¹. Cette technique permet d'évaluer l'absence ou la présence ainsi que le niveau d'expression d'antigènes spécifiques (phénotypes) sur des cellules en suspension à l'aide d'anticorps couplés avec des fluorochromes.

Dans cette perspective, il est crucial de s'orienter vers le développement de stratégies permettant l'analyse de protéines membranaires.

Le projet s'est déroulé en deux parties :

- i) Recherche et proposition d'un potentiel biomarqueur du lymphome de la zone marginale et validation par cytométrie en flux.
- ii) Mise en place d'un protocole de collecte dans le but de constituer une cohorte d'analyse et développement préliminaire d'une stratégie protéomique quantitative permettant de valider plusieurs candidats biomarqueurs.

B. Mise en place d'une stratégie de protéomique globale

B.1. Les échantillons et la stratégie mise en œuvre

Trois échantillons issus de patients atteints de différents lymphomes ont été sélectionnés : 1 MZL, 1 LPL et 1 MCL.

Des microparticules induites ont été générées à partir de 200×10^6 lymphocytes B issus de prélèvements sanguins (tube EDTA, 5 mL par patient) par traitement à l'actinomycine D pendant 18 heures dans un milieu sans sérum. Elles ont ensuite été récupérées par centrifugation et déposées sur un gel 1D SDS PAGE de 20 cm découpé tous les 2 mm (90 bandes). Les protéines ont été réduites, alkylées, digérées et le mélange peptidique a été injecté en nanoLC-MS/MS (nanoACQUITY UltraPerformance LC® Waters couplé au spectromètre de masse de type Q-TOF Synapt G1 Waters). Les recherches ont été effectuées dans la banque de données humaine Swiss-Prot à l'aide des moteurs Mascot et Omssa. Puis, les résultats issus de ces deux algorithmes ont été combinés dans Scaffold et des filtres de sélection ont été appliqués afin d'obtenir un taux de faux positifs inférieur à 1 % (FDR < 1 %). Une liste de protéines a ainsi été validée pour chaque échantillon. Par ailleurs, à chaque identification est associée une donnée de quantification basée sur le comptage du nombre de peptides et de spectres MS/MS. (cf. figure 92).

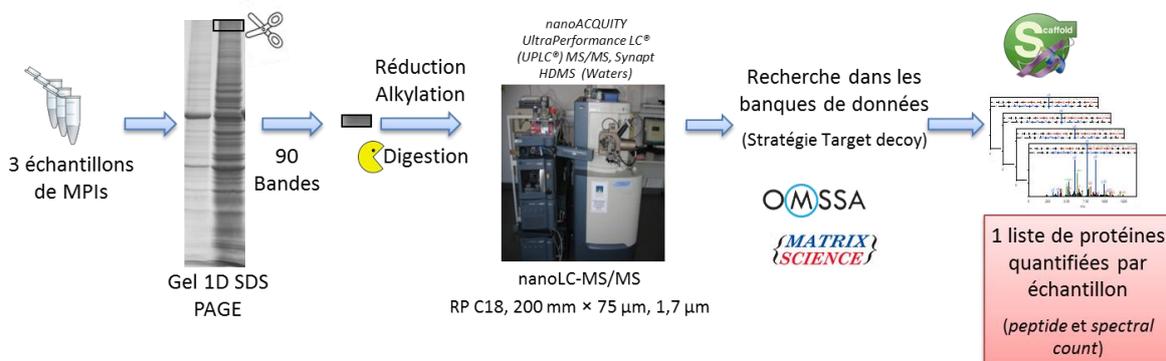


Figure 92

Schéma présentant les différentes étapes de la stratégie de protéomique globale mise en place pour la recherche de biomarqueurs de lymphopathies B

B.2. Les développements méthodologiques

Du fait du faible nombre de patients (3 personnes), l'objectif de ces analyses était d'identifier la liste la plus exhaustive possible du protéome membranaire de chaque échantillon afin de proposer des biomarqueurs candidats les plus fiables possible. À cet effet, plusieurs étapes de la stratégie mise en œuvre ont été évaluées et optimisées. Les apports des développements méthodologiques seront ici décrits de manière succincte, pour plus de détails se reporter à la Partie II.

B.2.1. Enrichissement de l'échantillon en protéines membranaires par génération de microparticules induites

Nous avons orienté notre développement analytique vers la recherche de protéines membranaires différentiellement exprimées car celles-ci sont couramment utilisées pour poser un diagnostic dans le domaine de l'hématologie, en utilisant des techniques telles que la cytométrie en flux. Cependant, cette famille de protéines caractérisée par une faible abondance reste un défi pour l'analyse protéomique. Afin de faciliter leur identification, nous avons choisi de travailler sur un matériel original : les microparticules induites (MPI). Ces dernières sont des vésicules issues du bourgeonnement de la membrane plasmique des cellules sous l'action d'un stress ou en condition d'apoptose. L'hypothèse a été émise que ces MPIs, directement issues des cellules tumorales, sont enrichies en protéines

membranaires et sont par conséquent un matériel intéressant pour l'étude du protéome des lymphocytes B malins. Nous avons pu constater sur nos trois patients un enrichissement allant de 25 à 30 % et démontrer ainsi l'intérêt de leur utilisation.

B.2.2. Répétabilité des analyses et détermination du nombre minimum d'injections à réaliser pour obtenir la vision la plus significative possible du contenu protéique d'un échantillon

Le mode d'acquisition utilisé (DDA – *Data Dependant Acquisition*), caractérisé par une alternance d'acquisition de spectre MS et MS/MS ne permet pas l'identification de l'ensemble des peptides élués dans une fraction chromatographique. En effet, au cours de l'analyse MS, seuls les cinq ions les plus intenses seront sélectionnés pour la fragmentation MS/MS. Par ailleurs, les peptides élués par la chromatographie pendant l'analyse MS/MS ne seront pas sélectionnés.

Conscients de ce problème, nous avons cherché à évaluer la répétabilité de nos analyses afin de déterminer la meilleure stratégie à mettre en œuvre. Dans ce but, chaque bande d'un échantillon de LPL déposé sur un gel 1 D SDS PAGE (90 bandes) a été injectée trois fois. Ainsi, sur les trois injections, 70 % des protéines et 61 % des peptides sont identifiés trois fois. Après deux injections, nous avons pu montrer que 95 % des protéines de l'échantillon et 92 % des peptides étaient identifiés.

Il a été considéré que deux analyses du même échantillon permettent d'obtenir une vision significative de son protéome.

B.2.3. Intérêt de l'utilisation de deux moteurs de recherche

Afin d'assurer un maximum de confiance dans les résultats, les recherches dans les banques de données ont été effectuées à l'aide de deux algorithmes : Mascot et Omssa. 87 % de protéines et 80 % de peptides ont été identifiés par les deux logiciels.

Nous avons choisi ici de ne conserver que les identifications communes aux deux moteurs de recherche afin de proposer la liste d'identifications la plus fiable possible.

B.3. Résultats

Pour un FDR < 1 %, entre 1000 et 2000 protéines ont été identifiées dans chaque pathologie : 2016 dans MZL, 1476 dans LPL et 1066 dans MCL, soit une liste de 1977 protéines uniques. 45 % des protéines sont communes aux trois échantillons. La figure 93 décrit le processus de sélection des protéines d'intérêt. Un premier filtre a été appliqué sur les annotations Gene Ontology (« plasma membrane ») afin de conserver uniquement les protéines membranaires : 489 pour MZL, 371 pour LPL et 224 pour MCL.

Puis, dans le but d'assurer des identifications fiables, seules les protéines identifiées avec plus de deux peptides ont été choisies, soit 388 protéines pour les trois pathologies. Ensuite, nous nous sommes focalisés sur les clusters de différenciation. En effet, ces molécules possédant chacune un anticorps commercial, la validation par cytométrie en flux sera simplifiée. La liste des 59 protéines résultantes est présentée en annexe 1. Enfin, une étude biologique associée à des données de semi-quantification

(comptage du nombre de peptides et de spectres) nous a permis de proposer le CD180 en tant que candidat biomarqueur pertinent surexprimé dans les lymphomes de la zone marginale.

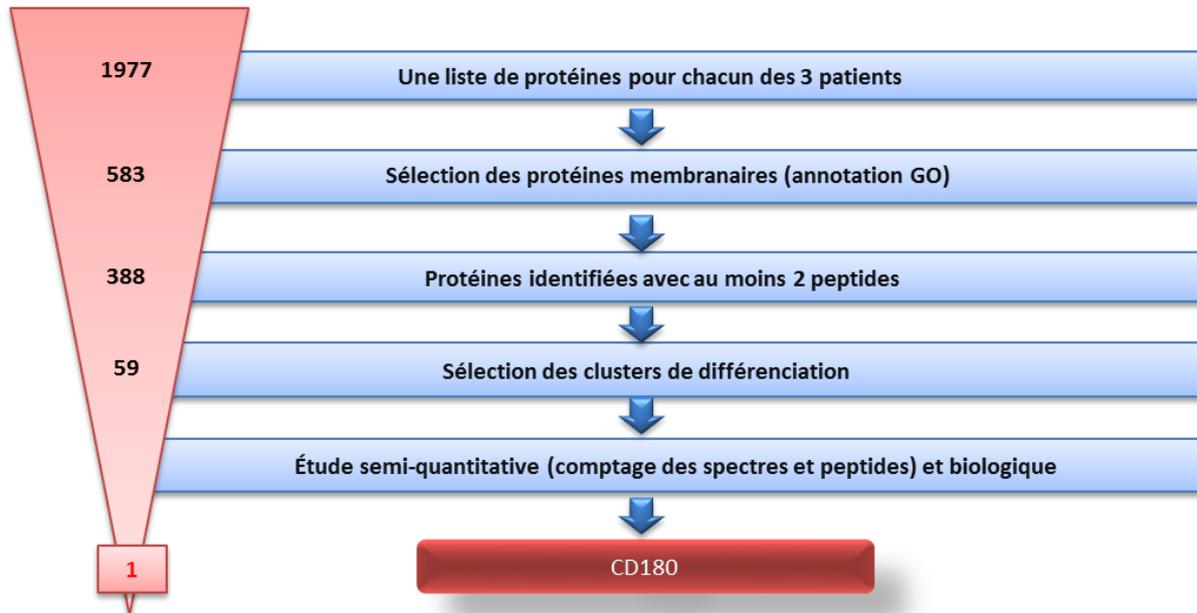


Figure 93
Filtres appliqués pour la mise en évidence du CD180 comme potentiel biomarqueur des lymphomes de la zone marginale

Le tableau 9 montre les résultats obtenus s'agissant des nombres de peptides et de spectres MS/MS. Une légère surexpression est observée pour le CD180 dans les lymphomes de la zone marginale.

	MZL	LPL	MCL
Nombre de peptides	9	0	2
Nombre de spectres	29	0	14

Tableau 9
Nombre de peptides et nombre de spectres identifiés dans chaque pathologie pour le CD180

Répartis sur l'ensemble de la séquence, 29 spectres et 9 peptides ont pu être identifiés, conduisant à un pourcentage de recouvrement de 30 % (cf. figure 94). La qualité des spectres a été validée manuellement afin d'assurer un maximum de confiance dans le résultat proposé. Un exemple est présenté en figure 95 avec le peptide GIYLNLAANSINIISPR.

CD180 – Homo Sapiens (Human)

(ou lymphocyte antigen 64 ou radioprotective 105 kDa protein)

MAFDVSCFFWVVLFSAGCKVITSWDQMCIEKEANKTYNCENLGLSEIPDTLPNTTEFLEFSFNFLPTIHNRTFSRLMN
 LTFDLTRCQINWIHEDTFQSHHQLSTLVLTGNPLIFMAETSLNGPKSLK**HFLFIQTGISNLEFIPVHNLENLESYLGS**
NHISSIKFPKDFPARNLK**VLDFQNNAIHYISR**EDMRSLEQAINLSLNFNGNNVKGIELGAFDSTIFQSLNFGGTPNLSV
 IFNGLQNSTTQSLWLGTFFEDIDDEDISSAMLKGLCEMSVESLNLQEH**RFSDISSTTFQCFTQLQELDLTATHLKG**LPS
 GMKGLNLLKK**LVLSVNHFDQLCQISAANFPSLTHLYIR**GNVKKLH**LGVGCLEKLGNLQTLDSLHNDIEASDCCSLQL**
KNLSHLQTLNLSHNEPLGLQSQAFKECPQLELLDLAFTRLHINAPQSPFQNLHFLQVLNLTFCFLDTSNQHLLAGLPV
 LRHLNLKGNHFQDGTITKTNLLQTVGSLEVLILSSCGLLSIDQQAFHSLGKMSHVDSLHNSLTCDSDLSHLK**GIYLNL**
AANSINIISPRLLPILSQQSTINLSHNPLDCTCSNIHFLTWYKENLHK**LEGSEETCANPPSLR**GVKLSVDVKLSCGITAIGI
 FFLIVFLLLLAILFFAVKYLLRWKYQHI

Figure 94

Séquence de la protéine CD180

Les peptides identifiés par nanoLC-MS/MS pour l'échantillon MZL sont indiqués en rouge. Le pourcentage de recouvrement s'élève à 30 %.

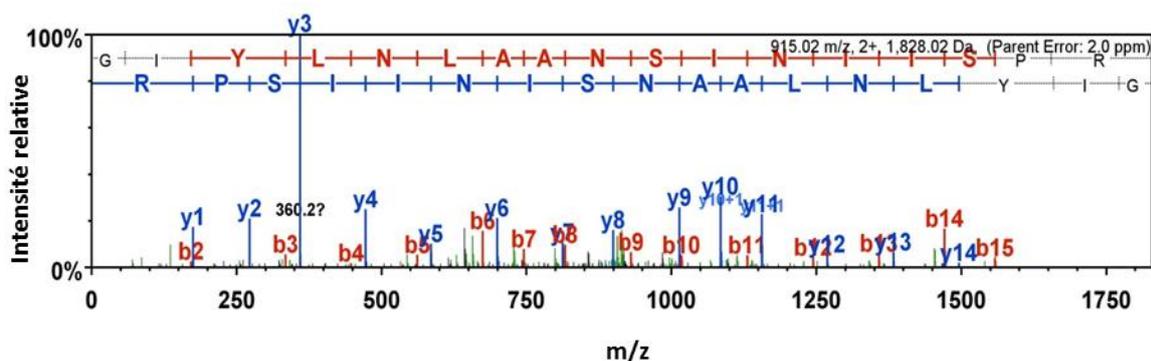


Figure 95

Exemple d'un spectre de fragmentation de l'ion $[M+2H]^{2+}$ à m/z 915,02, correspondant au peptide GIYLNLAANSINIISPR extrait de la protéine CD180

Même si la surexpression de CD180 peut paraître discutable en termes de statistiques, il ne faut pas oublier que ce marqueur a aussi été sélectionné en fonction de sa pertinence biologique. Ainsi, le CD180 est une protéine membranaire appartenant la famille des récepteurs de type Toll (TLR). Il a été précédemment montré que cette molécule est majoritairement exprimée à la surface des cellules porteuses d'antigènes (cellules B, monocytes, macrophages et cellules dendritiques) et que son expression est différente dans les cas de CLL (leucémie lymphoïde chronique) comparé à des témoins²⁸²⁻²⁸⁴. Par ailleurs, ces résultats ont fait l'objet d'une validation par cytométrie en flux dans deux services d'hématologie différents, à Strasbourg et à Lyon, sur 214 cas dont 60 atteints d'un lymphome de la zone marginale. La figure 96 présente les résultats obtenus pour l'analyse de l'ensemble des patients sur chacun des sites. Les MZL présentent une forte surexpression de CD180, avec une intensité de fluorescence médiane élevée, alors que dans le cas des LPL et des MCL, l'intensité médiane est faible.

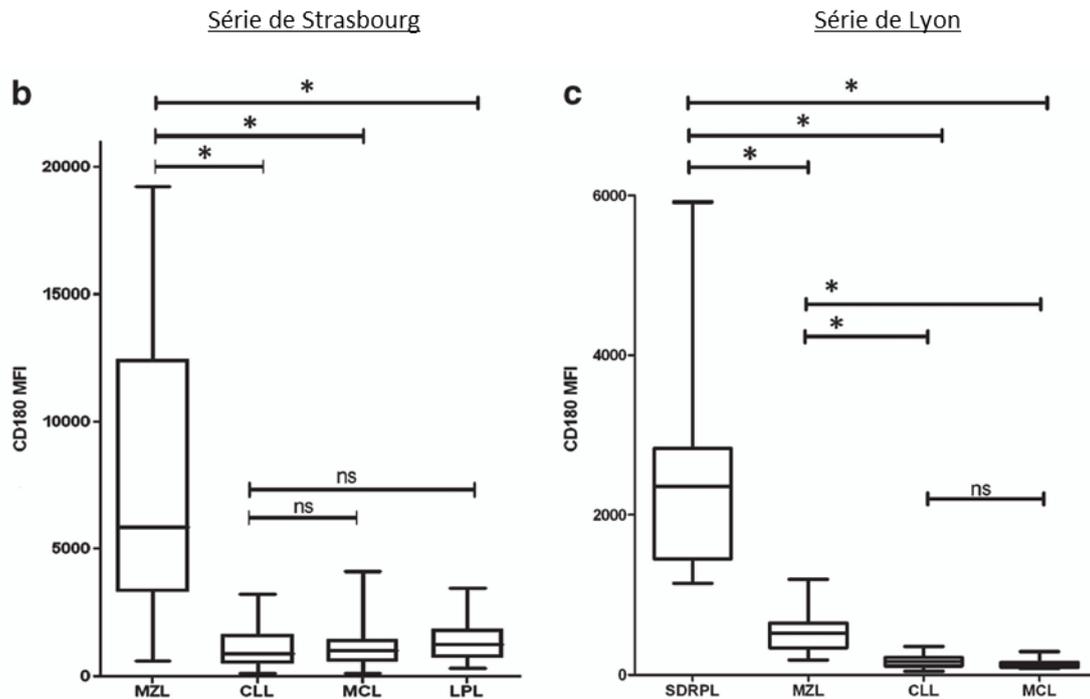


Figure 96

Comparaison de l'intensité de fluorescence moyenne pour CD180 mesurée par cytométrie en flux, pour les cas de MZL, CLL (leucémie lymphoïde chronique B), MCL et LPL dans la série de Strasbourg, et les cas de SDRPL (lymphome de la zone marginale splénique), MZL, CLL et MCL dans la série de Lyon

B.4. Conclusion

La mise en évidence du CD180 et sa pré-validation ont permis de montrer l'efficacité de la stratégie protéomique mise en œuvre malgré le faible nombre d'échantillons étudiés. Ce succès s'explique en grande partie par le développement d'une stratégie adaptée à la problématique et à la question posée.

Actuellement, le marqueur est utilisé en routine dans 8 hôpitaux de France (Lyon, Dijon, Nantes, Saint-Étienne, Angers, Poitiers, Strasbourg et Nancy) au sein d'une tripléte : CD148, surexprimé dans les lymphomes du manteau, CD180, surexprimé dans les lymphomes de la zone marginale et CD200 surexprimé dans les leucémies lymphoïdes chroniques. Associés, ces trois marqueurs aident à la caractérisation des entités lymphoprolifératives et sont une aide précieuse pour le diagnostic de cas difficiles.

B.5. Communication des résultats

Ces résultats ont fait l'objet :

- D'une **communication par affiche** au 61^e congrès ASMS (American society of mass spectrometry) de 2013 à Minneapolis (cf. annexe 2) :

Lennon S., Miguet L., Carapito C., Fornecker L., Van Dorsselaer A., Mauvieux L., Cianféroni S.

CD180: a new Marginal Zone B-cell Lymphoma Biomarker discovered by Plasma Membrane Microparticle Proteomic Analysis

- De **deux communications orales**, l'une au 29^e congrès de la SFEAP (Société française d'électrophorèse et d'analyse protéomique) à Rouen en 2012, et l'autre aux 29^{es} journées de la SFSM (Société française de spectrométrie de masse) à Orléans en 2012 :

Lennon S., Miguet L., Carapito C., Van Dorsselaer A., Mauvieux L., Cianférani S.

Development of a Proteomic Approach to Search for Biomarkers of B-lymphoma from Microparticles

- D'une **publication** dans le journal *Leukemia* (cf. annexe 3)

Miguet L.*, Lennon S.*, Baseggio L., Traverse-Glehen A., Berger F., Perrusson N., Chenard M. P., Galois A. C., Eischen A., Mayeur-Rousse C., Maar A., Fornecker L., Herbrecht R., Felman P., Van Dorsselaer A., Carapito C., Cianferani S., Mauvieux L.

Cell-surface Expression of the TLR homolog CD180 in Circulating Cells from Splenic and Nodal Marginal Zone Lymphomas

Leukemia, 2013

*Ces auteurs ont contribué de manière égale à ce travail.

C. Développement d'une méthode de protéomique ciblée pour la validation des résultats

La mise en œuvre de stratégies de protéomiques globales au laboratoire a permis de mettre en évidence deux marqueurs d'intérêt facilitant le diagnostic des lymphomes : CD148 et CD180. Ces biomarqueurs ont fait l'objet d'une validation par cytométrie en flux.

Cependant, les techniques de validation immunologiques (cytométrie en flux, ELISA) présentent un certain nombre de limitations :

- nécessité de disposer d'un anticorps afin de ne pas engendrer des coûts de production élevés ;
- un seul épitope de la protéine est ciblé par l'utilisation d'anticorps.

Dans ce contexte, l'objectif de la deuxième partie de ce projet a été de proposer une alternative aux méthodes immunologiques pour la validation de potentiels candidats biomarqueurs : la protéomique quantitative ciblée basée sur la technique SRM (selected reaction monitoring). Cette initiative s'inscrit dans un effort d'amélioration de notre pipeline actuel de découverte et de validation de biomarqueurs (cf. tableau 10), afin de leur faciliter l'accès à la validation clinique.

Étapes	Découverte	Vérification	Validation clinique
Stratégie	Protéomique globale <i>Identification de l'ensemble des protéines</i>	Protéomique ciblée <i>Quantification des candidats pré-sélectionnés</i>	Immunologie <i>Quantification des candidats finaux</i>
Echantillons	≈10	>100	>1000
Protéines	>1000	>10	4-10

Tableau 10

Politique du laboratoire quant à la recherche de biomarqueurs

L'étape de découverte est associée à une étape de vérification effectuée au laboratoire en utilisant une technique de protéomique ciblée.

Afin de réaliser la quantification la plus juste et la plus précise possible des candidats biomarqueurs, il est nécessaire de limiter au maximum le nombre d'étapes de préparation de l'échantillon. Le défi est grand dans le sens où il s'agit de détecter des protéines membranaires caractérisées par une faible abondance et une forte hydrophobicité dans un mélange complexe de protéines à large gamme dynamique.

Dans un premier temps, la faisabilité de la mise en œuvre de cette technique a été démontrée sur 6 protéines d'intérêt :

- le CD148, surexprimé dans les lymphomes du manteau¹⁷¹ ;
- le CD180, surexprimé dans les lymphomes de la zone marginale²⁸⁵ ;
- le CD200, surexprimé dans les leucémies lymphoïdes chroniques (données issues de la littérature)²⁸⁶ ;
- le CD23, surexprimé dans les leucémies lymphoïdes chroniques²⁸⁷ ;
- le CD45, exprimé uniformément à la surface de l'ensemble des lymphocytes B (marqueur témoin) ;
- le CD5, surexprimé dans les leucémies lymphoïdes chroniques et les lymphomes du manteau²⁸⁷.

Pour la bonne réalisation de ces développements méthodologiques, un protocole de collecte des échantillons a été établi. Il s'agit ici de mettre au point une méthode de traitement simple qui n'ajoute pas un travail supplémentaire trop important pour les laboratoires d'hématologie dans la gestion des prélèvements.

C.1. Quantification protéomique ciblée par SRM

L'approche ciblée SRM (selected reaction monitoring), méthode de référence depuis de nombreuses années pour l'analyse quantitative de petites molécules en chimie²⁸⁸⁻²⁹⁰, a récemment été adaptée à l'analyse quantitative de peptides et de protéines²⁹¹, atteignant des seuils de sensibilité identiques à ceux atteints par les techniques immunologiques dans le cadre de procédures diagnostiques,

pronostiques et/ou de suivis thérapeutiques²⁹²⁻²⁹⁴. Un des premiers avantages de cette approche ciblée réside dans le fait que celle-ci ne dépend pas du développement d'un anticorps de haute affinité. Un autre avantage de l'approche SRM est sa capacité à suivre plusieurs peptides²⁹⁵ pour une protéine donnée. De manière schématique, un anticorps utilisé pour une validation immunologique ne cible qu'un seul épitope, tandis qu'une approche SRM permettrait de cibler simultanément plusieurs peptides (épitopes) de la même protéine. À cela s'ajoute la capacité de multiplexage de la SRM²⁹⁶, c'est-à-dire de quantifier parallèlement et simultanément plusieurs protéines (une cinquantaine en routine)^{293, 294} dans des échantillons biologiques complexes. Enfin, il est possible de quantifier de manière fiable des isoformes d'une même protéine ainsi que des modifications post-traductionnelles.

Ces approches SRM sont en majeure partie mises en œuvre sur un spectromètre de masse de type « triple quadripôles » (QQQ)²⁹⁷. Le principe de fonctionnement repose sur une double sélection des ions ce qui rend la technique sensible et spécifique (cf. figure 97) : le premier quadripôle sélectionne un peptide d'un certain m/z (ion parent), le deuxième quadripôle permet la fragmentation du peptide et le troisième quadripôle sélectionne les fragments spécifiques (ions fils).

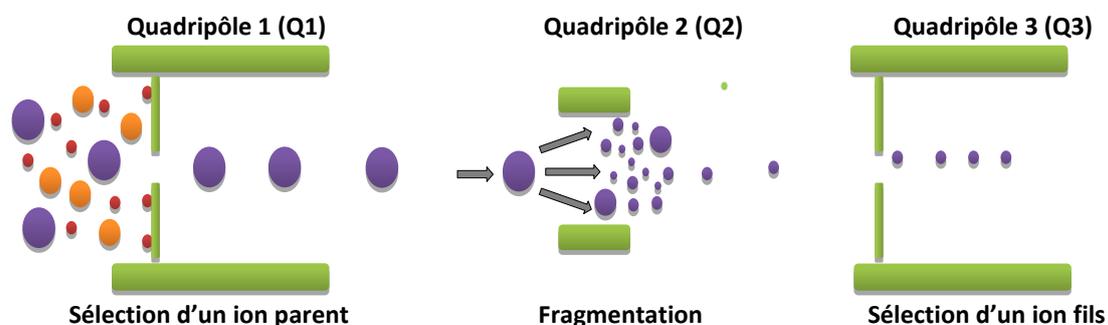


Figure 97

Sélection d'une transition par un triple quadripôle

Sélection d'un ion parent dans le quadripôle Q1, puis fragmentation de cet ion dans le quadripôle Q2 et sélection d'un ion fils dans le quadripôle Q3.

La sélection d'un ion parent et d'un ion fils est appelée transition et atteste de la spécificité de la technique. Afin d'assurer un maximum de confiance, chaque peptide est quantifié par la mesure des aires totales d'au moins 3 transitions et les résultats d'au moins 3 peptides sont pris en compte²⁹⁶ pour chaque protéine. En effet, la quantification d'un seul peptide peut induire des biais d'interprétation de la concentration de la protéine, qui peut être modifiée, avoir subi des clivages protéolytiques ou encore résulter d'épissages alternatifs différents. Ce biais a notamment été constaté au laboratoire lors de la quantification de la vitellogénine dans du plasma de tortue²⁹⁸. Un peptide issu de cette protéine présentait une évolution de concentration différente par rapport aux trois autres de la même protéine. Les auteurs expliquent que ceci pourrait être dû à des épissages alternatifs différents suivant les conditions. Par ailleurs, suivre 3 transitions permet de passer outre les possibles biais rencontrés à la suite d'interférences. En effet, les échantillons étudiés en analyse protéomique sont souvent très complexes et la probabilité d'observer des peptides isobariques (de même rapport m/z) est élevée, ce qui entraîne une possibilité non négligeable de rencontrer des interférences.

Afin de s'affranchir des effets d'ionisation ou de matrice qui peuvent engendrer une suppression ou une augmentation du signal, l'utilisation de standards marqués isotopiquement est donc recommandée²⁹⁶.

La quantification peut être absolue ou relative et s'effectuer de trois manières différentes :

- **SID** (stable isotope dilution) : utilisation de peptides de référence marqués isotopiquement (¹³C, ¹⁵N) et chimiquement identiques aux peptides endogènes, ajoutés en quantités connues à l'échantillon avant l'injection sur LC-SRM²⁹⁹. Un ratio entre les formes lourdes et légères est mesuré et la quantité de peptides est déterminée grâce à une courbe de calibration. Il est possible de réaliser une quantification absolue en utilisant des peptides quantifiés de manière précise (AQUA™, Sigma). Cependant, même si cette technique est dite absolue, elle ne donnera qu'une estimation relative de la concentration d'une protéine dans l'échantillon dans la mesure où le peptide marqué ne subira pas toutes les étapes de préparation³⁰⁰.
- **QconCAT™** (Quantification concatamers)^{301, 302} : deux peptides ou plus représentant les protéines ciblées sont concaténés dans une seule protéine chimérique, exprimée dans E.Coli, dans un milieu enrichi en arginine et en lysine marquées. La protéine chimérique est ajoutée dans l'échantillon au début du processus de traitement.
- **PSAQ™** (Protein Standard Absolute Quantification)³⁰³ : utilisation des protéines entières marquées, ajoutées dès le début du processus de traitement. La technique PSAQ est la plus précise pour deux raisons :
 - La séquence de la protéine marquée est identique à la séquence de la protéine endogène. Les effets de digestion seront ainsi limités par rapport à la stratégie QconCAT™.
 - La protéine marquée est ajoutée à l'échantillon dès la première étape de la préparation.

Cependant, les approches QconCAT et PSAQ restent des techniques très coûteuses et difficiles à mettre en place pour la validation de biomarqueurs où l'intérêt est de suivre plusieurs protéines dans le même échantillon (jusqu'à 50).

La stratégie développée dans ce chapitre est basée sur l'utilisation de peptides standards marqués et non quantifiés (PEPotec™, Thermo Fisher Scientific). La mise au point de cette méthode de quantification relative a nécessité plusieurs développements analytiques décrits dans la figure 98.



Figure 98

Liste des étapes à développer lors de la mise en place d'une méthode LC-SRM

La stratégie mise en place lors de notre étude est décrite dans la figure 99. Tous les développements méthodologiques ont été réalisés sur des lymphocytes issus de témoins sains. Les cellules sont comptées avec un automate de numération puis 1,5 million en sont déposées sur un gel 1D SDS PAGE 1

bande, nommé gel « stacking ». Les protéines sont ensuite réduites, alkylées et digérées. Après extraction, un mélange de peptides endogènes est obtenu, auquel sont ajoutés les peptides marqués (PEPotec™, Thermo Fisher Scientific). L'ensemble est injecté sur le couplage SRM (Ultimate 3000 couplé au triple quadripôle TSQ Vantage, Thermo Fisher Scientific). La quantification est réalisée à l'aide du logiciel Skyline qui extrait pour chaque transition les courants d'ions correspondants, et ceci pour l'ensemble des peptides endogènes et marqués qui sont suivis.

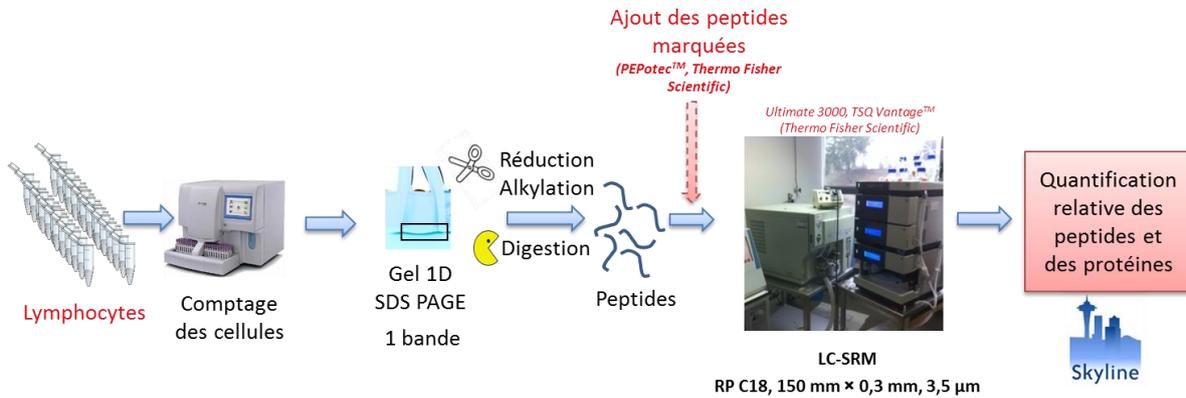


Figure 99

Schéma représentant les différentes étapes de la stratégie développée afin de quantifier de manière relative les peptides issus d'échantillons de lymphocytes par une méthode LC-SRM

La figure 100 décrit le mode de quantification d'un peptide dans un échantillon. Les aires des trois transitions de chaque peptide, endogène et marqué, sont extraites et sommées. Un ratio entre les formes lourdes et légères est ensuite calculé donnant une quantification relative pour chaque peptide. Une comparaison des ratios peut alors être effectuée entre les différents échantillons. Dans le cas présent, ce ratio est de 0,5.

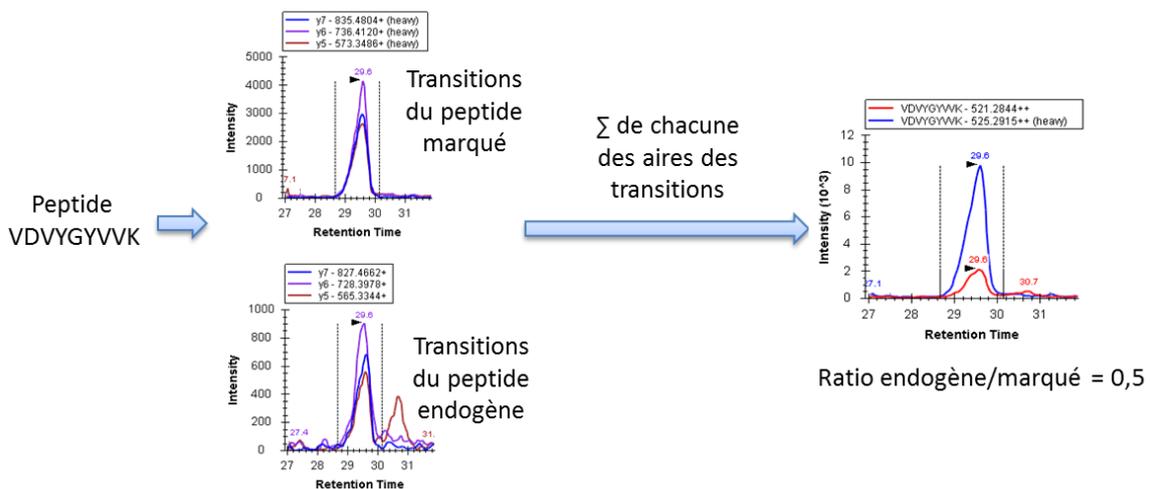


Figure 100

Quantification relative du peptide VDVYGYVVK dans un échantillon donné

C.2. Préparation des échantillons

C.2.1. Particularité des échantillons

Le sang total a été collecté sur des témoins sains (non atteints de lymphomes). Une centrifugation par gradient de densité sur un Ficoll a permis de récupérer les lymphocytes T et B. Il est important de noter ici que les échantillons issus de témoins sains sont plus riches en lymphocytes T qu'en lymphocytes B : ces derniers représentent 10 à 15 % du nombre total de lymphocytes. Dans les échantillons de la future cohorte, les patients sont caractérisés par une hyperlymphocytose des cellules B, c'est-à-dire un nombre de lymphocytes B très important pouvant parfois atteindre 90 % de l'ensemble des lymphocytes. Ceci aura une influence sur les résultats attendus et les marqueurs potentiellement détectables.

Par ailleurs, lors de la collecte des lymphocytes B et T, une contamination plaquettaire a été observée. Celle-ci est « patient et opérateur dépendant » (cf. figure 101). Il sera important par la suite d'évaluer ce taux de plaquettes dans les échantillons de patients atteints de lymphomes. En effet, certains marqueurs d'intérêt tels que le CD148 et le CD45 sont aussi exprimés à la surface des plaquettes et peuvent ainsi fausser la quantification.

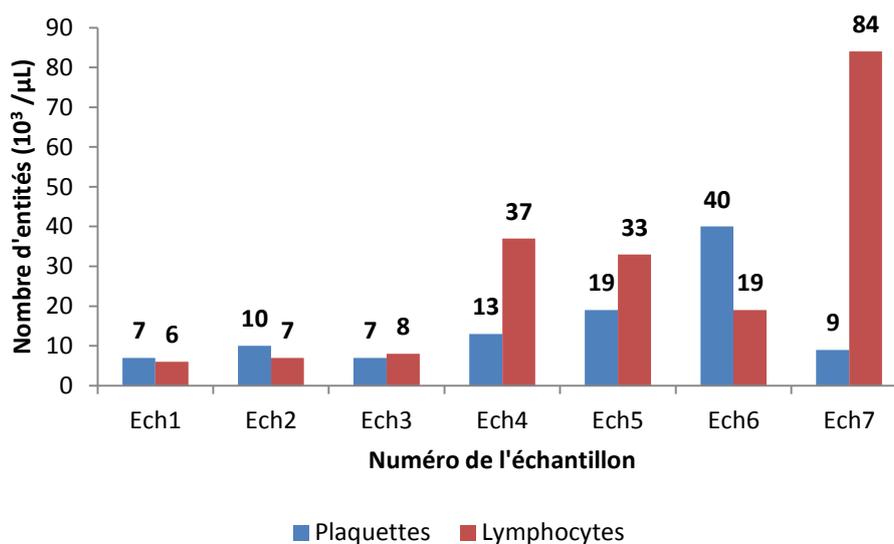


Figure 101

Nombre de plaquettes et de lymphocytes détectés dans les sept échantillons témoins ayant servi au développement de la méthode LC-SRM

C.2.2. Dosage protéique des échantillons

Pour chaque échantillon, un dosage RC DC (réducteur compatible, détergent compatible)³⁰⁴ a été effectué afin de déterminer la concentration protéique. La figure 102 montre que la quantité de protéines détectées en μg par million de cellules varie d'un échantillon à l'autre (de 26 à 73 μg/million de cellules). En cytométrie en flux, les intensités de fluorescence sont évaluées sur des cellules en suspension. Afin d'obtenir des résultats comparables, nous avons choisi de partir sur 1,5 million de cellules, soit environ 50 μg de protéines.

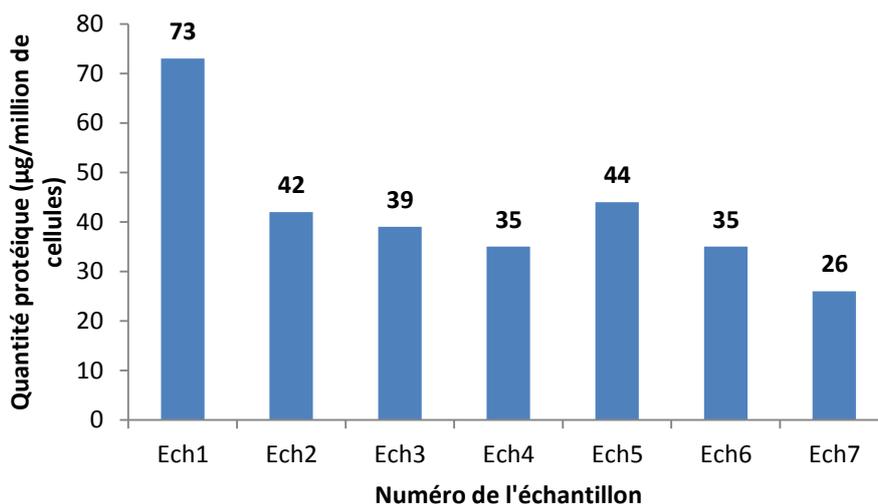


Figure 102
Quantité de protéines (µg/million de cellules) déterminée par dosage RC DC dans les sept échantillons témoins utilisés pour le développement de la méthode

C.2.3. Choix du mode de préparation des échantillons

Afin de choisir la meilleure solubilisation possible, des tests ont été effectués sur un même échantillon, en utilisant quatre détergents différents (cf. figure 103) : pas de détergent (classique), 2 % sodium dodécyl sulfate (SDS), 1 % deoxycolate de sodium (DOC) ou 0,1 % Rapidigest™ SF (Waters). Chaque préparation a été réalisée en duplicat puis injectée deux fois en nanoLC-MS/MS, couplage nanoAcquity – Maxis Impact, Waters-Bruker (Q-TOF).

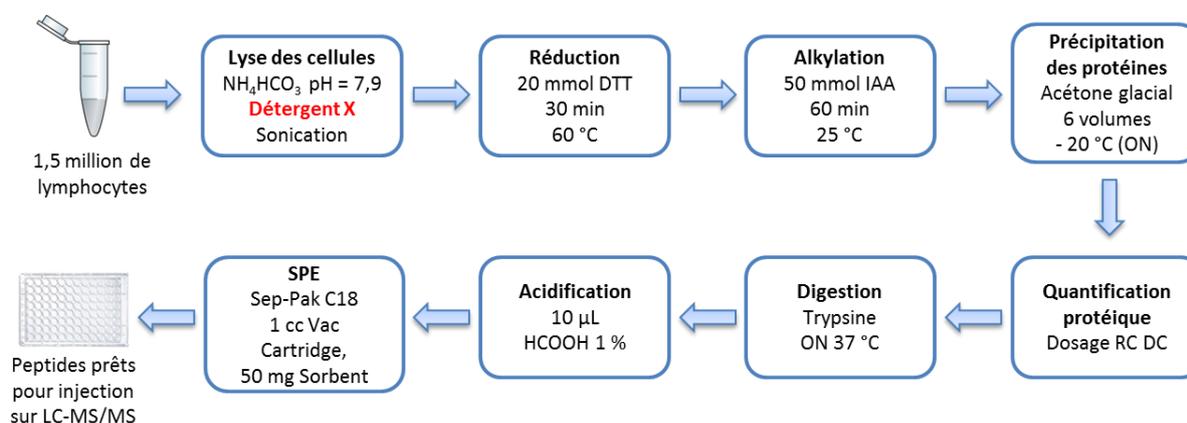


Figure 103
Processus de préparation des échantillons à partir de 1,5 million de lymphocytes
Plusieurs détergents ont été testés : aucun, DOC (déoxycolate de sodium), Rapidigest™ SF, SDS (dodécylsulfate de sodium).

Les résultats obtenus, en termes de nombre de protéines identifiées (FDR < 1 %), sont présentés en Figure 104. Comme attendu, l'utilisation du SDS favorise nettement la solubilisation des protéines. Par ailleurs, le DOC et Rapidigest™ SF présentent un nombre d'identifications protéiques comparables, bien qu'inférieur de 25 % à celui obtenu avec le SDS. Enfin, l'absence de détergent a un effet très négatif sur le nombre de protéines identifiées : 65 % en moins par rapport au SDS.

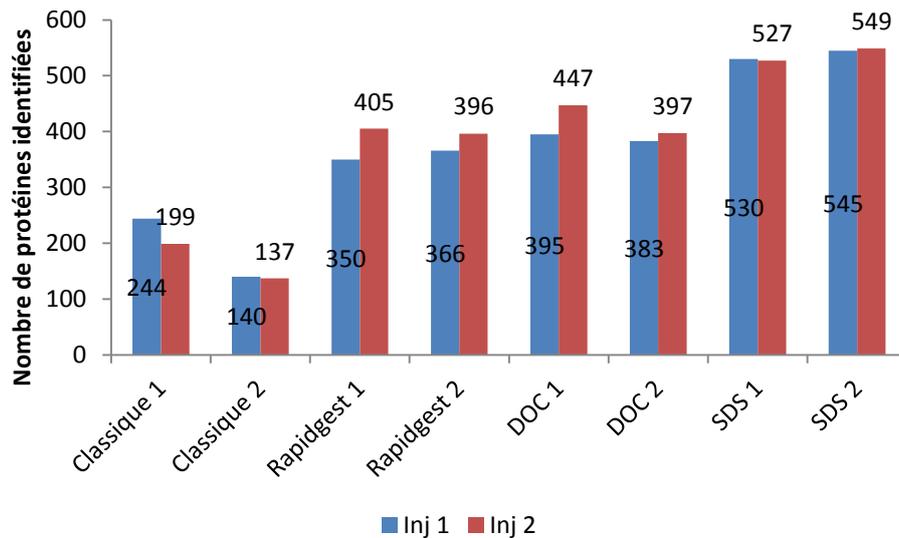


Figure 104

Nombre de protéines identifiées (FDR < 1%) suivant la technique de solubilisation employée

Classique = sans détergent, Rapidigest™SF, DOC (Déoxycholate de sodium), SDS (dodécyl sulfate de sodium).

Pour la suite, le protocole SDS a été retenu. Cependant, lors d'expériences ultérieures, nous avons observé de fortes suppressions de signal causées par du SDS résiduel ainsi que par la présence de contaminants dans plusieurs échantillons.

Afin de contourner le problème, il a été décidé d'employer un gel 1D SDS PAGE stacking. Ainsi, les cellules sont lysées dans le tampon de Laemmli (10 mM Tris pH8, 1 mM EDTA, 5 % β -mercaptoéthanol, 5 % SDS, 10 % glycérol et 0,1 % bleu de bromophénol), puis déposées sur un gel. La migration est effectuée uniquement dans le gel de concentration, ce qui permet d'obtenir une seule bande contenant l'ensemble des protéines. Le gel est ensuite découpé selon le modèle présenté en figure 105. Les protéines sont réduites, alkylées, digérées et extraites avant analyse par spectrométrie de masse.

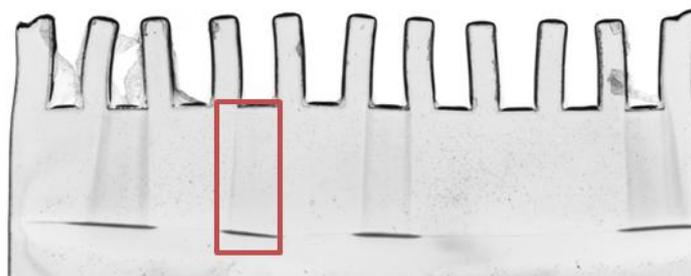


Figure 105

Photo d'un gel de concentration

Les protéines migrent dans le gel pendant 45 min à 50 V. La bande contenant les protéines est ensuite découpée.

Remarque complémentaire : dans le cadre de ce projet, il a été décidé de découper la bande concentrée ainsi que la partie supérieure. En effet, lors de précédentes expériences, plusieurs protéines d'intérêt (CD45 et CD148) y ont été identifiées.

C.3. Développement de la méthode LC-SRM

C.3.1. Choix des peptides

Pour chaque protéine, une première série de critères a été appliquée afin de sélectionner les dix peptides les plus fiables, c'est-à-dire ayant le plus de chances d'être détectés par spectrométrie de masse :

- **Peptides protéotypiques**, c'est-à-dire uniques à la protéine dans la banque considérée³⁰⁵. Dans le cas présent, il a été vérifié que chaque peptide sélectionné était unique dans la banque humaine Swiss-Prot ainsi que dans la banque humaine contenant les variants d'épissage. Les protéines d'intérêt (CD148, CD180, CD45, CD200, CD5 et CD23) possèdent toutes des peptides uniques sauf le CD200 qui partage un peptide commun avec ses trois isoformes. Il existe deux peptides discriminants pour l'isoforme 1 et deux pour l'isoforme 3. L'isoforme 2 n'en possède pas.
- **Peptides identifiés lors de précédentes analyses nanoLC-MS/MS**. Le tableau 11 montre le nombre de peptides uniques déjà identifiés pour chaque marqueur au cours d'analyses nanoLC-MS/MS réalisées sur divers projets du laboratoire concernant l'étude de protéomes humains. Pour les CD148 et CD45, des filtres seront nécessaires afin de sélectionner les peptides les plus pertinents. Pour les autres, en plus de la sélection de peptides théoriques, le SRM Atlas³⁰⁶ sera utilisé. Cette banque de données accessible à l'ensemble de la communauté scientifique indique pour chaque protéine humaine au minimum trois peptides ayant été synthétisés et analysés par SRM.

Projet	1	2	3	4	5	6	7
CD148	15	7	23	nd	13	34	3
CD180	1	4	4	2	2	8	1
CD45	37	46	53	34	37	56	43
CD200	1	nd	nd	nd	nd	nd	2
CD5	7	3	nd	nd	nd	11	2
CD23	nd						

Tableau 11

Nombre de peptides identifiés pour chaque protéine ciblée dans des analyses nanoLC-MS/MS précédemment réalisées au laboratoire dans le cadre de sept projets différents portant sur l'étude du protéome humain (nd : non détectée)

- **Nombre de résidus d'acides aminés** compris entre 8 et 25.
- **Peptides sans coupures manquées.**
- **Peptides non situés dans un domaine transmembranaire** : ces domaines sont en effet caractérisés par des peptides longs et très hydrophobes plus difficiles à détecter en spectrométrie de masse. Il existe plusieurs logiciels permettant de prédire les domaines transmembranaires d'une protéine donnée à partir de sa séquence. Afin d'être le plus juste possible, nous avons combiné l'utilisation de quatre logiciels : SOSUI³⁰⁷ (basé sur une détermination d'hydrophobicité des séquences), TMHMM³⁰⁸, PHOBIUS³⁰⁹ (tous deux basés sur une approche statistique) et TOP CONS³¹⁰ (basé sur une approche consensus, c'est-à-dire

utilisant à la fois l'hydrophobicité de la séquence et l'approche statistique). Un exemple est donné pour le CD180 en figure 106. En combinant les quatre logiciels, trois domaines transmembranaires ont été identifiés. Les peptides uniques issus de ces domaines seront exclus de la sélection pour l'analyse SRM.

CD180 – Homo Sapiens (Human)

(ou lymphocyte antigen 64 ou radioprotective 105 kDa protein)

MAFDVSCFFWVVLFSAGCKVITSWDQMCIEKEANKTYNCENLGLSEIPDTPNTEFLEFSFNFLPTIHNRTFSRLM
 NLFLDLTRCQINWIHEDTFQSHHQLSTLVLTGNPLIFMAETSLNGPKSLKHLFLIQTGISNLEFIPVHNLENLESYLG
 NHISSIKFPKDFPARNLKVLDFQNNAIHYISREDMRSLEQAINLSLNFNGNNVKGIELGAFDSTIFQSLNFGGTPNLSVI
 FNLQNSTTQSLWLGTFFEDIDDEISSAMLKGLCEMSVESLNLQEHFSDISSTTFQCFTQLQELDLTATHLKGPLSG
 MKGLNLLKKLVLSVNHFDQLCQISAANFPSLTHLYIRGNVKKLHLGVGCKLEKLGNLQTLDSLHNDIEASDCCSLQKLN
 LSHLQTLNLSHNEPLGLQSQAFKECPQLELLDLAFTRLHINAPQSPFQNLHFLQVLNLTQCFLDTSNQHLLAGLPVLRH
 LNLKGNHFQDGTITKTN**LLQTVGSLEVLILSSCGLLSI**DQQAFHSLGKMSHVDSLHNSLTCDSDLSLHKGIYLNLA
 NSINIISPRLLPILSQSTINLSHNPLDCTCSNIHFLTWYKENLHKLEGSEETTCANPPSLRGVKLSVKSCG**ITAIGIFFL**
IVFLLLAILLFFAVKYLLRWKYQH I

Figure 106

Séquence de la protéine CD180

Les domaines transmembranaires détectés par l'utilisation des 4 logiciels (SOSUI, PHOBIUS, THMM et TOP CONS) sont indiqués en vert gras.

- **Pas de sélection de peptides à cystéine ou méthionine** afin d'éviter, dans la mesure du possible, les complications dues aux modifications pouvant survenir durant l'étape de préparation des échantillons (ici carbamidométhylation et oxydation)

Le tableau 12 récapitule le nombre de peptides choisis et commandés (PEPotec™) pour chaque protéine suivant les critères établis précédemment. Pour certains marqueurs, il n'a pas été possible de sélectionner dix peptides car la séquence était trop courte ou possédait peu de sites de coupures tryptiques. Par ailleurs, pour certaines protéines (CD200, CD180, CD5 et CD23), des peptides modifiés ont dû être choisis. Enfin, les isoformes CD200_1 et CD200_3 présentent deux peptides spécifiques (non partagés avec les autres isoformes) et un peptide partagé. L'isoforme CD200_2 présente un seul peptide partagé avec CD200_1 et CD200_3. Il ne sera donc pas possible de suivre CD200_2.

Protéine	CD148	CD180	CD45	CD200_1	CD200_3	CD5	CD23
Nombre de peptides	10	7	10	2 (1)	2 (1)	10	7

Tableau 12

Nombre de peptides sélectionnés pour chaque protéine d'intérêt

Les PEPotec™ correspondants ont été commandés.

Remarque : les isoformes CD200_1 et CD200_3 possèdent deux peptides spécifiques et un peptide partagé indiqué en parenthèses.

Dans une deuxième étape, les transitions optimales ont été choisies (ion parent – ion fils) et les peptides sélectionnés ont été validés.

Les transitions optimales ont été sélectionnées sur la base de la qualité des spectres MS/MS obtenus lors de l'injection du mélange de peptides marqués en nanoLC-MS/MS. Une validation manuelle de chaque peptide et de chaque transition a ensuite été effectuée sur la base de l'injection du même mélange en LC-SRM. Deux exemples de peptides retenus et non retenus sont montrés en figure 107 : le peptide 1 présente un profil chromatographique idéal et trois transitions aux intensités correctes. Au contraire, le peptide 2 a une largeur à la base de 8 min avec un spray très instable.

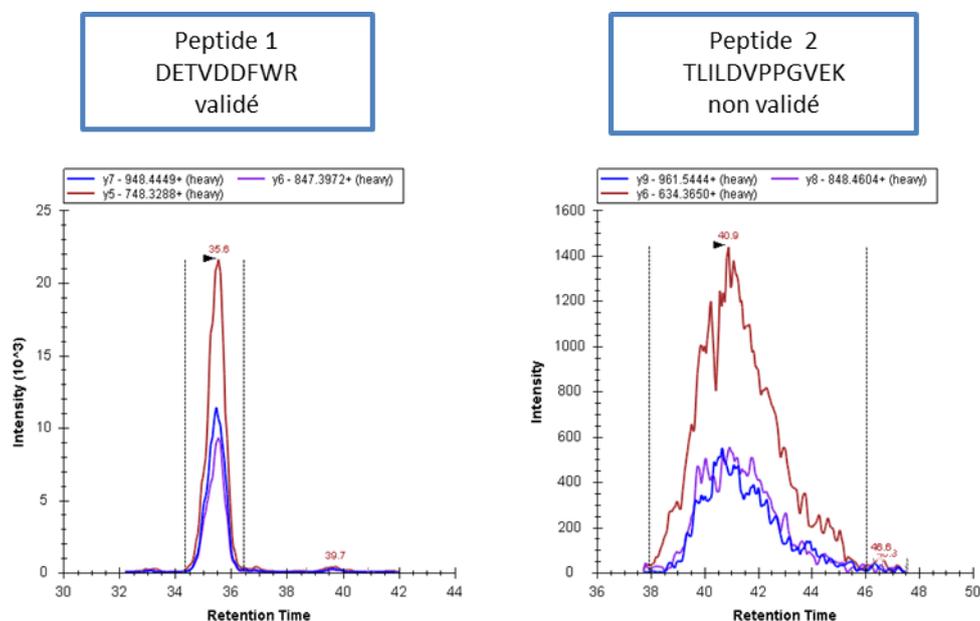


Figure 107 Exemples d'un peptide retenu et d'un peptide non retenu sur la base de leur profil chromatographique

Le tableau 13 récapitule le nombre de peptides validés et donc suivis pour chaque protéine. La liste complète des transitions suivies est présentée en annexe 4. Pour la plupart des protéines, le nombre de peptides suivis est largement suffisant et correspond aux critères définis en début de chapitre. Par contre, il ne sera pas possible de suivre plus de peptides pour les isoformes du CD200. En effet, la séquence est courte (31 kDa) et compte un domaine transmembranaire.

Protéine	CD148	CD180	CD45	CD200_1	CD200_3	CD5	CD23
Nombre de peptides validés	8	4	10	1 (1)	1 (1)	8	5

Tableau 13 Nombre final de peptides validés pour chaque protéine

Ces peptides seront suivis lors des analyses LC-SRM. Remarque : le CD200 possède un peptide spécifique à chaque isoforme et un peptide partagé entre les deux.

À titre d'exemple, les peptides sélectionnés pour le CD180 sont présentés dans sa séquence en figure 108.

CD180 – Homo Sapiens (Human)

(ou lymphocyte antigen 64 ou radioprotective 105 kDa protein)

MAFDVSCFFWVVLFSAGCKVITSWDQMCIEKEANKTYNCENLGLSEIPDTLPNTTEFLEFSFNFLPTIHNRTFSRLMN
 LTFDLTRCQINWIHEDTFQSHHQLSTLVLTGNPLIFMAETSLNGPKSLKHLFLIQTGISNLEFIPVHNLENLESYLGSN
 HISSIKFPKDFPARNLKVLD**FQNNAIHYISRE**DMRS**LEQAINLSLNFNGNNVK**GIELGAFDSTIFQSLNFGGTPNLSVI
 FNLQNSTTQSLWLGTFEDIDDEDISSAMLKGLCEMSVESLNLQEHFSDISSTTFQCFTQLQELDLTATHLKGLPSG
 MKGLNLLKLVLSVNHFDQLCQISAANFPSLTHLYIRGNVKKLHLGVGCLEKLGNLQTLDSLHNDIEASDCCSLQL**KN**
LSHLQTLNLSHNEPLGLQSQAFKECPQLELLDLAFTRLHINAPQSPFQNLHFLQVLNLTFCFLDTSNQHLLAGLPVLR
 HLNLK**GNHFQDGTITK**TNLLQTVGSLEVLILSSCGLLSIDQQAFHSLGKMSHVDLSHNSLTCDSIDSLSHLKGIYLNLA
 ANSINIISPRLLPILSQQSTINLSHNPLDCTCSNIHFLTWYKENLHKLEGSEETTCANPPSLRQVGLSDVKLSGCGITAIGIFF
 LIVFLLLLAILFFAVKYLLRWKYQH I

Figure 108

Séquence de la protéine CD180

Les peptides sélectionnés pour la méthode SRM sont indiqués en gras bleu.

C.3.2. Développement de la méthode LC-SRM

Une fois les transitions choisies et validées, une méthode LC-SRM a été développée. Cette dernière est basée sur des cycles répétés d'acquisition des transitions choisies. Le temps de cycle de chacune des mesures est défini par le nombre de transitions suivies et le temps de scan de chacune de ces transitions (*dwell time*) selon la relation présentée ci-dessous :

$$\text{Temps de cycle} = \text{nombre de transitions} \times \text{temps de scan}$$

Il s'agit ici de trouver un compromis entre le temps de scan et le temps de cycle. Un temps de cycle long permettra d'augmenter le temps passé à acquérir chaque transition et donc d'améliorer le rapport signal/bruit. Un temps de cycle court permettra de réaliser plus de points pour chaque transition et donc de mieux définir la trace chromatographique. Le temps de scan minimum auquel peut travailler le QQQ (TSQ Vantage, Thermo Fisher Scientific) est de 10 ms et 10 points sont requis pour bien définir un pic chromatographique. Ces données sont basées sur des tests réalisés à l'installation et décrits dans la thèse du Dr Amandine Bœuf³¹¹.

Les échantillons étant très complexes, un gradient de 70 min a été choisi (de 5 à 35 % en acétonitrile). À partir de ce dernier et des temps de rétention des peptides marqués, une méthode *scheduled* a été paramétrée. Ce type de méthode permet de suivre les transitions uniquement sur une certaine gamme de temps de rétention et ainsi d'augmenter le nombre de transitions suivies par analyse (cf. figure 109).

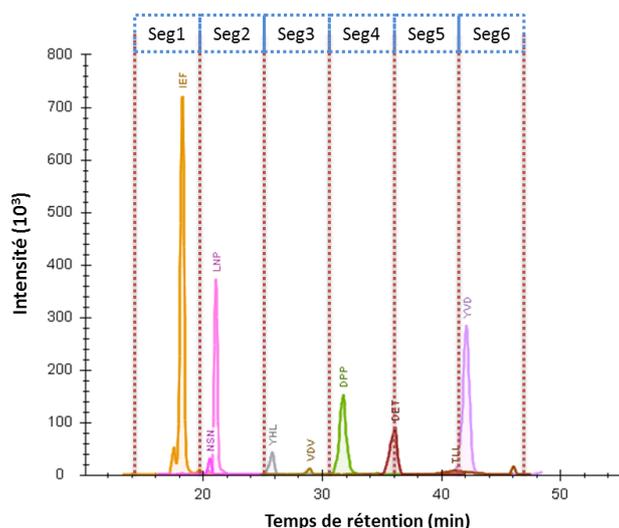


Figure 109

Méthode SRM *scheduled* : les transitions suivies sont différentes selon le segment considéré

Dans le cas présenté ci-dessus, il y a 6 segments différents. Ce type de méthode permet de suivre plus de transitions mais nécessite aussi des temps de rétention stables.

Dans un premier temps, afin d'éviter divers problèmes pouvant survenir suite à des changements de connexions et de solvants, nécessaires durant l'étape d'optimisation, une fenêtre de recherche du peptide a été fixée à 10 min. La figure 110 présente une estimation du nombre de peptides suivis dans chaque fenêtre de temps : 2, 5 et 10 minutes. Ainsi, si l'on suit chaque transition pendant 10 min, 120 transitions au maximum seront suivies à la fois.

La largeur des pics chromatographiques étant de 30 s, le temps de cycle a été fixé à 2,4 s, ce qui correspond à un minimum de 20 ms par transition lorsque 120 transitions sont suivies en même temps. Ces paramètres remplissent bien les critères définis précédemment.

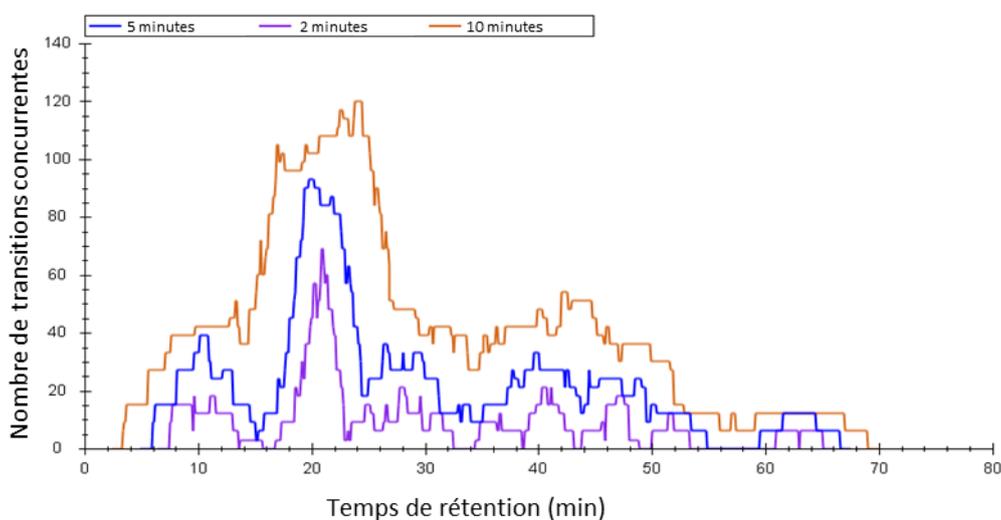


Figure 110

Nombre de transitions suivies à chaque instant du gradient selon les fenêtres de temps utilisées

C.3.3. Étude de répétabilité et résultats

Une étude préliminaire a été menée afin de déterminer la reproductibilité de l'ensemble du processus expérimental (cf. figure 111). Un même échantillon a été déposé trois fois sur trois gels différents et injecté trois fois, soit un total de 27 injections.

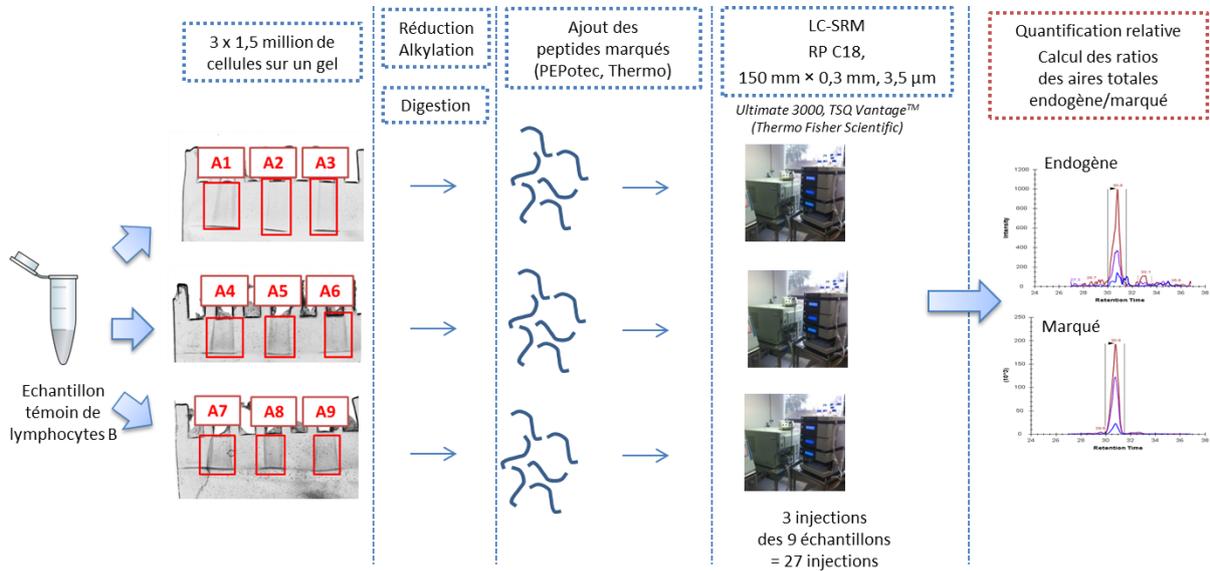


Figure 111

Schéma du processus suivi pour l'évaluation de la reproductibilité de la stratégie analytique

Les résultats obtenus sont présentés sous forme de boîte à moustaches (cf. figure 112) regroupant l'ensemble des ratios des aires totales pour les peptides détectés. Un coefficient de variation inférieur à 20 % est observé pour l'ensemble des peptides. Par ailleurs, la médiane des valeurs se situe à 12 %. Ces résultats, couramment rencontrés dans la littérature³¹², attestent une bonne reproductibilité de l'ensemble du processus.

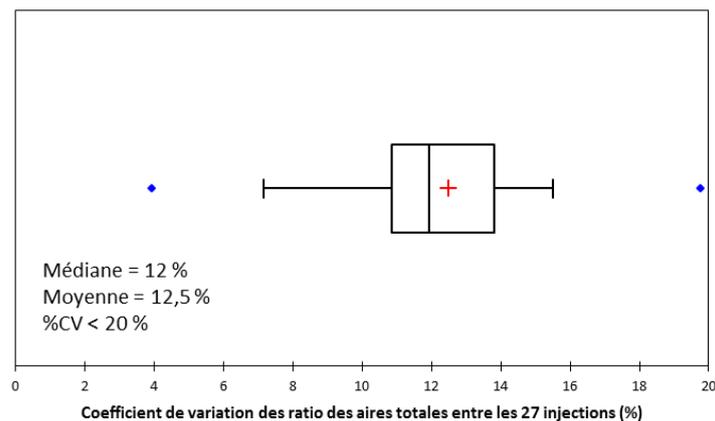


Figure 112

Boîte à moustaches des coefficients de variation des ratios des aires totales des peptides détectés par LC-SRM

Les peptides détectés sont présentés dans le tableau 14. Les protéines CD148, CD5 et CD45 sont identifiées, ce qui est conforme à leur expression sur les lymphocytes T (composants majoritaires des globules blancs chez les témoins). Pour la même raison, il n'est pas aberrant de ne pas détecter les protéines CD200, CD180 et CD23 qui ne sont pas exprimées sur les lymphocytes T.

Protéine	CD148	CD180	CD45	CD200_1	CD200_3	CD5	CD23
Nombre de peptides suivis	8	4	10	2	2	8	5
Nombre de peptides détectés	5	0	7	0	0	4	0

Tableau 14

Nombres de peptides suivis et de peptides détectés au cours de l'expérience de reproductibilité pour chaque protéine d'intérêt

C.4. Conclusions

Un protocole de préparation des échantillons de lymphocytes permettant une analyse par LC-SRM a été mis en place (cf. figure 113). Ce protocole permet la détection des protéines membranaires.

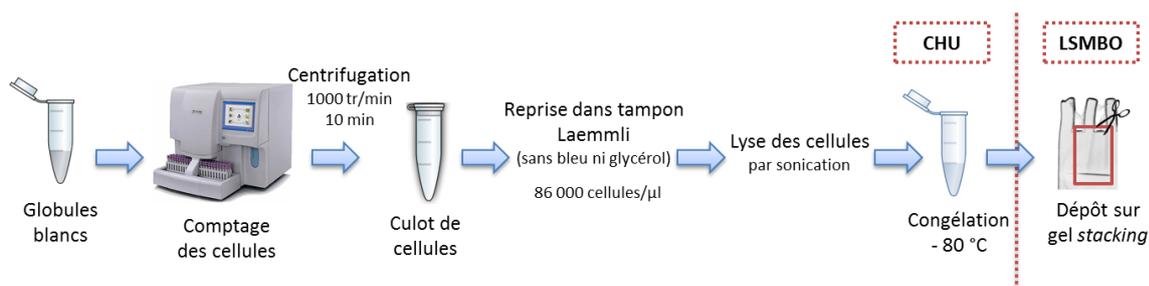


Figure 113

Protocole de préparation des échantillons mis en place

Les premières étapes – comptage des cellules, centrifugation, reprise dans le tampon et stockage à -80°C – seront effectuées au CHU. Le dépôt sur gel se fera au LSMBO.

Quelques apports restent nécessaires à cette étude avant de démarrer la constitution d'une cohorte à grande échelle. Il faudrait notamment travailler sur des échantillons issus de patients hyperlymphocytaires, afin d'évaluer la proportion de lymphocytes B ainsi que l'influence de la contamination plaquettaire, et rechercher les trois protéines non identifiées.

Par ailleurs, dans un deuxième temps, de nouveaux marqueurs candidats issus d'études bibliographiques seront inclus dans notre méthode LC-SRM (environ 50) afin d'exploiter au maximum les capacités de la technique. Chaque nouvelle protéine incluse nécessitera un travail de développement. La dernière étape consistera à valider la méthode en déterminant des limites de détection et de quantification et en évaluant la linéarité.

Cette étude est désormais incluse dans un projet nommé Malyzoma et soutenu par le Cancéropôle Grand-Est. Ce projet implique la collecte d'une large cohorte d'échantillons au sein des CHU de Strasbourg et Nancy afin, notamment, de valider plusieurs marqueurs d'intérêt par SRM.

D. Conclusion générale

Pour la deuxième fois, nous avons relevé le défi de la recherche de biomarqueurs candidats d'hémopathies lymphoïdes B afin d'en faciliter le diagnostic. Le cluster de différenciation CD180 a été révélé par l'analyse protéomique globale comme candidat biomarqueur potentiel des lymphomes de la zone marginale à partir d'un nombre restreint d'échantillons. Son expression a été vérifiée cliniquement sur plus de 200 patients et il est en cours de validation à l'échelle nationale.

La littérature compte, selon l'expression de Steven Carr³⁰⁰, « legions of biomarkers » de diverses pathologies identifiés par analyse protéomique, non validés et restés au stade de candidats. Cependant, notre deuxième succès valide l'intérêt de la recherche de biomarqueurs par protéomique et, surtout, met en évidence l'importance de quatre points :

- la définition claire et précise de la question posée ;
- le choix du matériel biologique ;
- la réflexion autour de la confiance que l'on peut accorder aux identifications réalisées par protéomique afin de développer une stratégie analytique adaptée ;
- la collaboration étroite avec les cliniciens qui sont le plus à même de choisir des candidats pertinents et d'entamer le processus de vérification (indispensable pour tout biomarqueur issu d'une stratégie de protéomique globale) et de validation clinique.

Les approches de vérification de l'expression différentielle des protéines candidates par protéomique quantitative connaissent aujourd'hui un engouement croissant et sont considérées comme une passerelle entre la découverte du biomarqueur et son utilisation en clinique^{313, 314}. En effet, la cytométrie en flux n'est un outil efficace que pour les candidats possédant déjà un anticorps commercial. Par ailleurs, dans notre cas, le choix des candidats potentiels a été restreint par la capacité limitée du laboratoire d'hématologie à effectuer ces analyses en plus des analyses de routine. Afin d'affermir et d'augmenter l'efficacité de notre « pipeline » de validation de biomarqueurs, nous avons développé une méthode SRM permettant de cibler 3 des 5 marqueurs candidats précédemment mis en évidence. La figure 114 conclut sur le « pipeline » mis en place au laboratoire pour la recherche et validation de candidats.

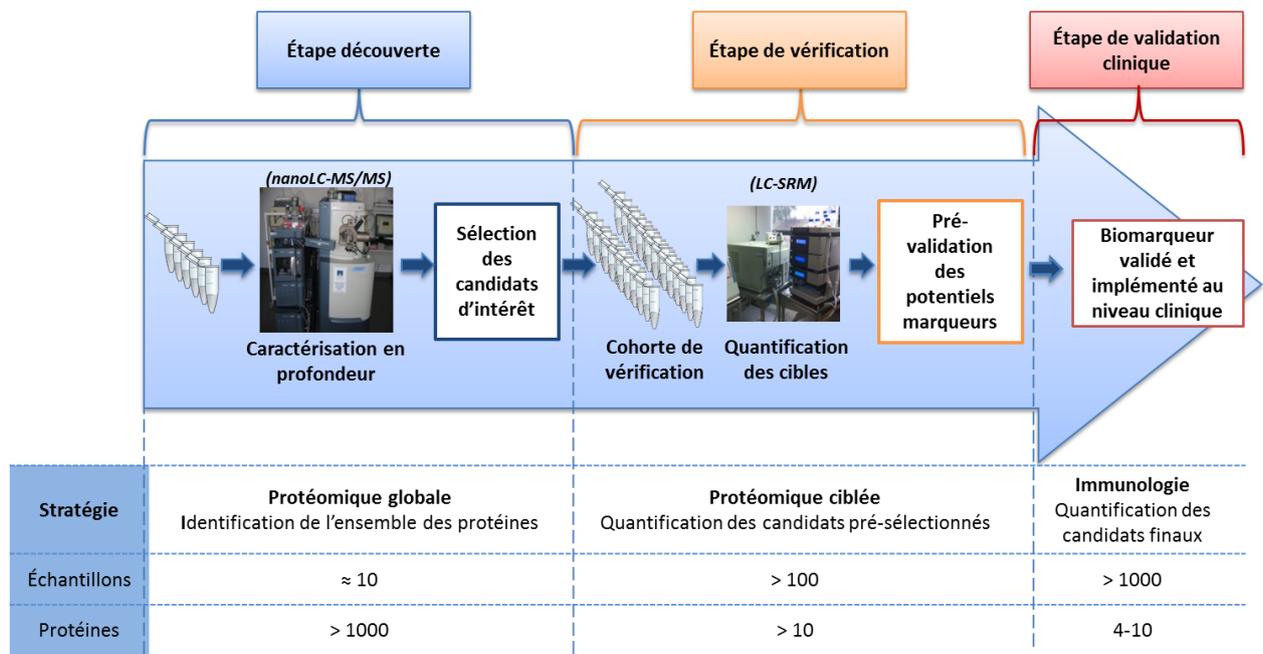


Figure 114

Pipeline de validation des biomarqueurs candidats

Les étapes de découverte et de vérification pourront désormais être effectuées au laboratoire.

CHAPITRE 2

Recherche de biomarqueurs de cellules souches cancéreuses issues de glioblastomes

Cette étude a été réalisée en collaboration avec les professeurs Jacques Haiech et Marie-Claude Kilhoffer du laboratoire d'innovation thérapeutique de la faculté de pharmacie d'Illkirch et avec le Dr Hervé Chneiweiss, neurobiologiste et neurologue au CNRS, à Paris.

J. Haiech et M.-C. Kilhoffer travaillent sur l'étude de potentielles cibles thérapeutiques, principalement les protéines RCPG (récepteurs couplés aux protéines G) et les protéines liées à la signalisation calcique, comme marqueurs de différenciation des cellules souches.

H. Chneiweiss est spécialisé dans l'étude des astrocytes et s'intéresse plus particulièrement aux mécanismes liant les astrocytes à la formation de tumeurs cérébrales.

A. Les cellules souches cancéreuses : nouvel axe de la recherche de biomarqueurs

Les gliomes sont des tumeurs cérébrales développées aux dépens des cellules gliales. Le glioblastome, gliome de grade IV suivant la classification de l'Organisation Mondiale pour la Santé, en est la forme la plus commune et concerne 50 % des cas³¹⁵. Ce cancer affecte cinq à dix personnes sur 100 000 par an³¹⁶. Il est très agressif et la survie médiane des patients est inférieure à seize mois. Les traitements classiques associent des techniques de radiothérapie, chimiothérapie (témozolomide) et chirurgie³¹⁷⁻³²⁰. Cependant, même après une ablation chirurgicale complète et un traitement agressif, la tumeur récidive³²¹.

Une théorie est aujourd'hui émise : des cellules souches cancéreuses (CSC)^{28, 322, 323} seraient à l'origine de l'initiation de la tumeur, de son expansion et des rechutes, du fait des caractéristiques propres aux cellules souches :

- **Capacité d'auto-renouvellement** : Cette propriété est nécessaire au développement de la tumeur. Ainsi, la cellule souche subit une division symétrique qui génère deux nouvelles cellules souches³²⁴.
- **Capacité à se différencier** afin de générer des cellules matures. La division est ici asymétrique³²⁴.

- **Quiescence** (capacité à sommeiller) pendant une longue période dans des endroits zones (niches)³²¹.
- **Résistance aux traitements**³²⁵.

L'origine de ces CSCs est encore inconnue à ce jour. Deux théories sont explorées³²⁶ (cf. figure 115) :

- Des cellules souches neurales (NSC) donnent naissance à des CSCs suite à un événement oncogénique³²⁷ (mutation).
- Des cellules différenciées deviennent malignes suite à un premier événement oncogénique (mutation) puis CSCs suite à un deuxième événement^{328, 329}.

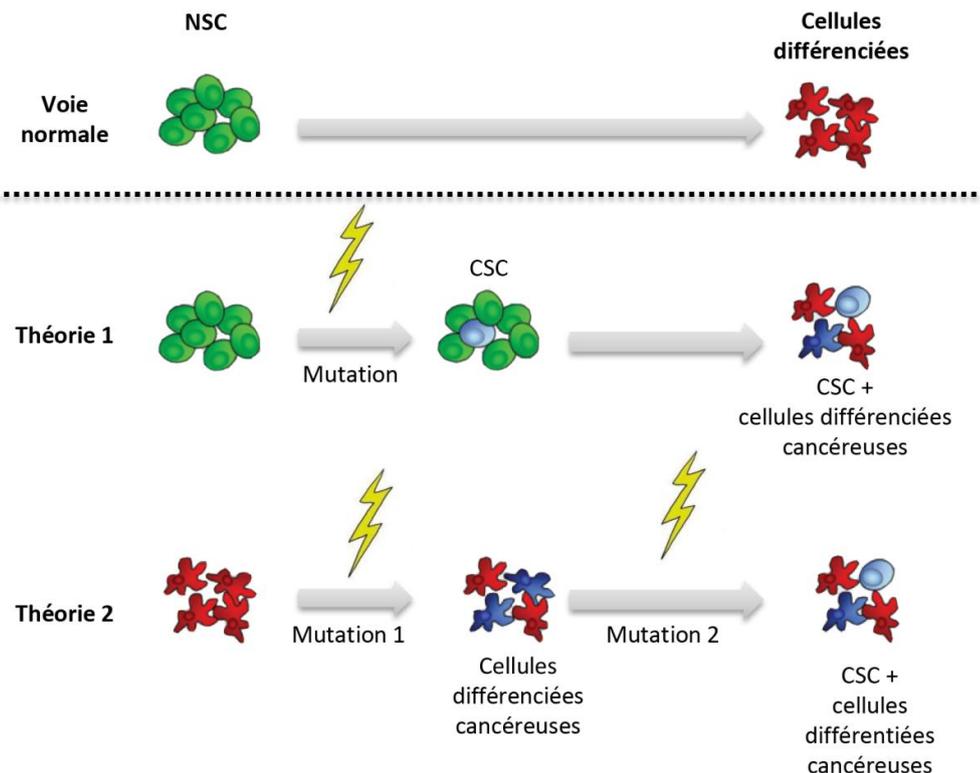


Figure 115

Schéma présentant les deux théories de l'origine des cellules souches cancéreuses

Dans ce contexte, l'idée a été émise de combiner les thérapies actuelles, qui permettent de réduire la masse tumorale, à de nouvelles thérapies ciblées sur l'élimination des cellules souches cancéreuses^{324, 330}.

Cependant, à ce jour, aucun marqueur spécifique des cellules souches cancéreuses issues de glioblastomes (GSC, *glioblastoma stem cell*) n'a été mis en évidence. En effet, les marqueurs actuels sont soit non spécifiques, c'est-à-dire exprimés de la même manière sur les cellules souches normales, soit non-ubiquitaires sur l'ensemble des GSC étudiées³³¹⁻³³⁹.

Le projet a pour but de proposer de nouvelles cibles qui permettraient de caractériser et/ou cibler thérapeutiquement les GSCs par analyse protéomique différentielle. Sept lignées cellulaires ont été utilisées :

- U87-MG (cellules cancéreuses issues de glioblastomes)

- HA (astrocytes humains)
- f-NSC (cellules souches neurales)
- TG01 et OB1 (GSCs p53 non-mutées), TG10 et TG16 (GSCs p53 mutées)

La recherche a été orientée dans un premier temps vers l'identification des récepteurs couplés à la protéine G (RCPG), dont les transcrits ont été précédemment analysés. Notons que l'étude différentielle a ici été effectuée sur U87-MG, HA, f-NSC, TG01 et OB1. En effet, les lignées TG10 et TG16 n'étaient pas encore disponibles. Dans un deuxième temps, l'étude s'est portée sur la détection et la quantification (*spectral count*) des clusters de différenciation (CD) sur l'ensemble des échantillons. L'augmentation maximale de la profondeur d'analyse a été ici de la première importance. En effet, les protéines RCPG et les CD sont des protéines membranaires faiblement abondantes.

B. Développement de la stratégie analytique

B.1. Échantillons

Les analyses ont été réalisées sur sept échantillons différents :

- HA : culture primaire d'astrocytes humains (ScienCell, Research laboratories, Carlsbad, Etats-Unis)
- U87-MG : lignées cellulaires de glioblastome (American Type Culture, collection ATCC, Molsheim, France)
- f-NSC : cellules souches neurales de fœtus, préparées par le laboratoire du Dr Hervé Chneiweiss, Inserm U752, Hôpital Sainte-Anne, Paris³³⁶
- TG01, OB1, TG10B, TG16 : cellules souches cancéreuses issues de patients atteints de glioblastomes, cultures réalisées au laboratoire du Dr Hervé Chneiweiss³³⁶

Les échantillons HA et U87-MG ont été cultivés selon les instructions du fournisseur.

Pour les GSCs, des biopsies de 1 mm³ de tumeurs de patients atteints d'un glioblastome ont été réalisées et ont ensuite été placées en culture dans un milieu sans sérum avec des facteurs de croissance tels que EGF (Epidermal Growth Factor). Une fois par semaine, les neurosphères formées sont dissociées mécaniquement et remises en culture dans les mêmes conditions. Environ cinquante passages sont nécessaires pour obtenir une population homogène de cellules souches cancéreuses. Les cellules f-NSC sont cultivées exactement de la même manière que les cellules souches cancéreuses. La seule différence réside dans le prélèvement des biopsies qui est effectué dans une zone potentiellement neurogénique.

Il est important de noter que le caractère souche des cellules cancéreuses a été vérifié à l'aide de plusieurs tests (cf. figure 116) : vérification de la capacité à s'auto-renouveler, de la clonogénicité, de la capacité à générer une nouvelle tumeur si ces cellules sont greffées dans le cerveau d'une souris immunodéficente. Par ailleurs, l'expression des marqueurs de cellules souches neurales et des cellules différenciées a été vérifiée.

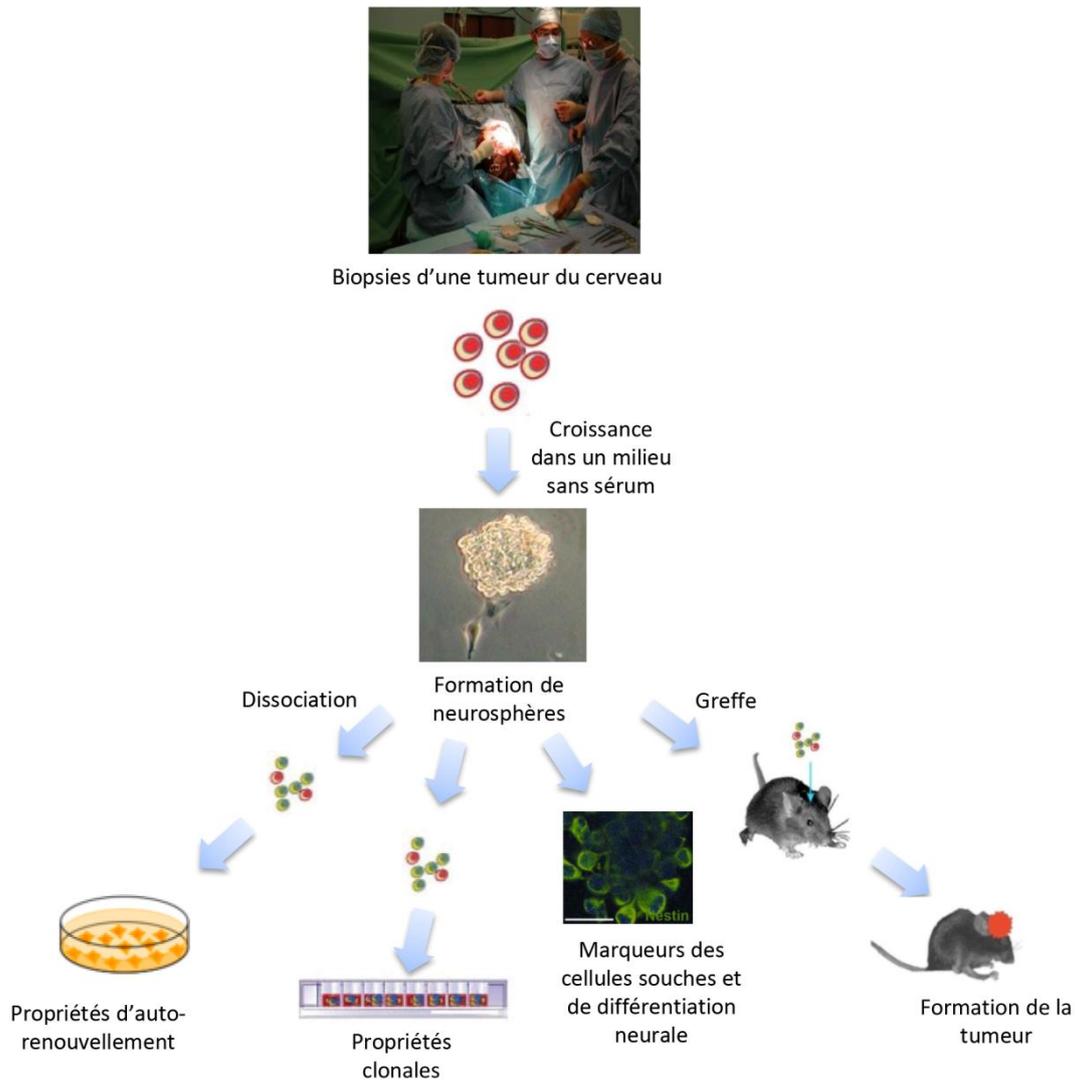


Figure 116

Les différentes étapes de la préparation des cellules souches cancéreuses

Une biopsie (1 mm³) de tumeur du cerveau est effectuée sur un patient. La croissance cellulaire est réalisée dans un milieu sans sérum avec des facteurs de croissance (EGF, Epidermal Growth Factor) jusqu'à l'obtention de neurosphères. Ces dernières sont ensuite contrôlées au niveau de leur capacité à s'auto-renouveler, de leurs propriétés clonales, des marqueurs présents à la surface des cellules et de la formation d'une tumeur à l'issue de la greffe de ces cellules chez une souris.

Source : adapté d'une présentation du Dr Hervé Chneiweiss

B.2. Stratégie mise en œuvre

Les différentes étapes de la stratégie mise en œuvre sont décrites dans la figure 117. Les 7 échantillons ont été enrichis en protéines membranaires par lyse mécanique de $4 \cdot 10^7$ cellules conduisant à la formation de *ghosts*. Ces derniers sont récupérés par centrifugation, repris dans du PBS (*phosphate buffered saline*) et déposés sur un gel 1D SDS PAGE de 20 cm découpé tous les 2 mm (90 bandes). Les protéines ont été réduites, alkylées, digérées et le mélange peptidique a été injecté sur le couplage nanoLC-MS/MS (nanoHPLC-Chip, Agilent, couplée à la trappe ionique AmaZon, Bruker Daltonics). Les recherches ont été effectuées dans la banque de données Swiss-Prot à l'aide des moteurs Mascot et Omssa. Puis, les résultats de ces algorithmes ont été combinés dans le logiciel Scaffold (Proteome

Software) et des filtres de sélection ont été appliqués afin d'obtenir un taux de faux positifs inférieur à 1 % (FDR < 1 %). Une liste de protéines a ainsi été validée pour chaque échantillon. Par ailleurs, à chaque identification est associée une donnée de quantification basée sur le comptage du nombre de peptides et de spectres MS/MS.

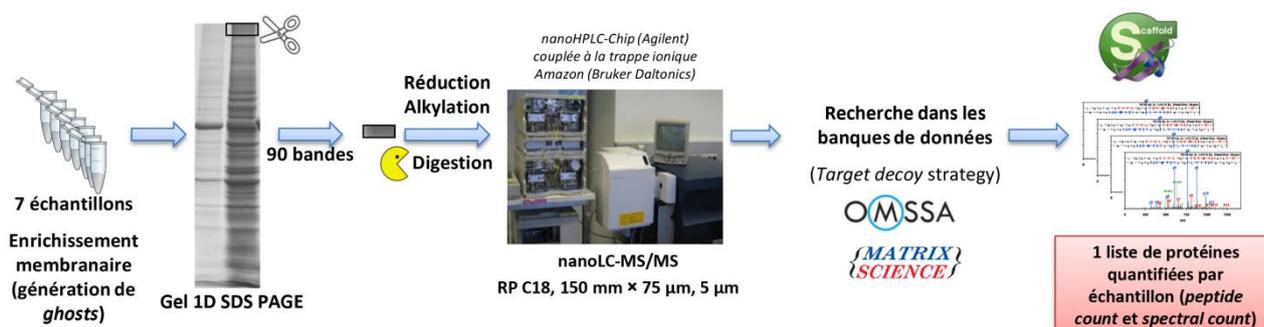


Figure 117

Stratégie protéomique mise en place pour l'analyse différentielle des sept lignées cellulaires

B.3. Développements méthodologiques

La culture de cellules souches cancéreuses est un travail long et difficile. Le nombre d'échantillons issus de patients différents à disposition est par conséquent peu élevé (4 GSCs). En outre, les protéines membranaires (RCPGs et CDs) sont faiblement abondantes. Dans ce contexte, la stratégie mise en œuvre a été optimisée de manière à obtenir une vision du protéome membranaire de chaque échantillon la plus complète et la plus fiable possible.

Les apports des développements méthodologiques sont ici décrits de manière succincte. Pour plus de détails se reporter à la Partie II.

B.3.1. Enrichissement des échantillons en protéines membranaires par la génération de *ghosts*

De la même manière que pour le projet « lymphome » précédemment décrit, nous avons choisi de travailler sur les protéines membranaires. En effet, celles-ci sont des cibles thérapeutiques privilégiées dans le cadre des cancers.

Comme il n'a pas été possible de générer des microparticules induites à partir des GCSs, nous nous sommes orientés vers une lyse mécanique des cellules conduisant à la formation de *ghosts* membranaires. En moyenne, un enrichissement de 24 % a été observé dans chaque échantillon, valeur couramment relevée dans la littérature.

B.3.2. Optimisation fine des paramètres de chromatographie liquide et d'acquisition du couplage

Les analyses ont été réalisées sur le couplage nanoLC-Chip (Agilent) – trappe ionique Amazon (Bruker Daltonics). Les paramètres clés de ce dernier ont été finement étudiés et optimisés afin d'extraire d'un échantillon le maximum d'informations possibles :

- **Temps de gradient** : il s'agissait ici de déterminer le temps de gradient optimal de façon à séparer au maximum les peptides pour faciliter leur sélection, tout en conservant un temps

d'analyse correct. Finalement, un gradient de 61 min de 2 à 40 % en ACN comportant deux pentes a été choisi.

- **Nombre d'ions sélectionnés pour la MS/MS** : des sélections de 6, 8 et 10 ions pour la MS/MS ont été testées. Aucun changement dans les résultats obtenus n'a été observé. De manière arbitraire, le paramètre 8 ions a été choisi.
- **Balayage des ions** : sur la trappe ionique, il est possible de paramétrer la vitesse de balayage des ions suivant quatre modes. Cependant, si une vitesse élevée permet de gagner en sensibilité et en nombre de spectres réalisés, elle entraîne une perte de résolution. Les tests nous ont permis de montrer que le mode UltraScan, 32 500 (m/z)/s, était le plus pertinent à utiliser pour l'acquisition MS/MS. En effet, le nombre d'identifications était augmenté. À noter qu'il s'agit du mode de balayage le plus rapide lors de l'acquisition en DDA.
- **Accumulation des ions dans la trappe ionique** : il a été observé qu'une augmentation du nombre d'ions accumulés (ICC = 400 000) dans la trappe ionique lors de l'acquisition MS/MS permettait d'augmenter le nombre d'identifications.
- **Le seuil de sélection absolu des ions** a été évalué en fonction du bruit de fond.

L'optimisation fine de chacun de ces paramètres a permis d'augmenter de manière significative les performances instrumentales quant au nombre de protéines identifiées.

B.3.3. Recherche dans les banques de données

Les paramètres du moteur de recherche Mascot ont été étudiés afin d'être adaptés à notre analyse, c'est-à-dire aux types d'échantillons et à l'appareil utilisé.

- **Erreur sur la masse du précurseur** : ce paramètre est important dans la mesure où il influe sur l'espace de recherche. Or plus l'espace de recherche est grand, plus l'identification de faux positifs sera favorisée. Il a été montré qu'au maximum, l'erreur était de 0,2 Da.
- **Modifications post-traductionnelles** : dans notre cas, les modifications post-traductionnelles ont été apportées par le processus de traitement des échantillons. Il est connu que les cystéines sont couramment carbamidométhylées et les méthionines, oxydées. Notre étude a permis de montrer que la propionamidation des cystéines issues de leur réaction avec l'acrylamide non polymérisée concernait un nombre important de peptides de notre échantillon.
- **Spécificité de l'enzyme** : il a été démontré que très peu de peptides sont semi-tryptiques (4 %) et que la quasi-totalité de ces derniers comporte au maximum une coupure manquée.

Dans un deuxième temps, les résultats obtenus à l'aide du moteur de recherche Omssa ont été comparés à ceux obtenus en utilisant Mascot. De manière surprenante, Omssa identifiait 24 % de protéines supplémentaires. Cette observation s'explique par la présence d'un paramètre dans Omssa qui permet de s'affranchir des problèmes de déconvolution, courant sur les instruments de basse résolution. Afin de maximiser le nombre d'identifications, nous avons choisi de conserver celles des deux moteurs de recherche.

Comme précédemment, l'étude fine de l'ensemble de ces paramètres de recherche dans les banques de données nous a permis d'augmenter la qualité ainsi que le nombre de protéines identifiées.

B.3.4. Évaluation de la qualité du *spectral count*

Le *spectral count* est une technique basée sur le comptage du nombre de spectres afin d'évaluer la quantité d'une protéine présente dans un échantillon donné. Elle est couramment utilisée lors de l'analyse différentielle de divers échantillons. Notre étude a permis de montrer qu'il était pertinent d'utiliser le *spectral count* sur des résultats issus d'une trappe ionique.

Cependant, deux biais importants ont nécessité une adaptation des étapes développées lors des tests réalisés sur la levure :

- Des réplicats d'analyse n'ont pu être réalisés. En effet, le projet nécessitait la comparaison de sept échantillons déposés sur un gel 1D (90 bandes) soit une semaine d'analyse par échantillon.
- Des échantillons supplémentaires ont été ajoutés par le collaborateur à la suite de la première série de manipulations (comparaison de résultats issus de deux gels différents).

Ainsi, dans la première série de manipulations, cinq échantillons ont été analysés. Dans la deuxième série, deux échantillons supplémentaires ont été ajoutés. Afin d'évaluer la reproductibilité de nos analyses, et notamment du dépôt sur gel, deux échantillons déposés sur gel lors de la première manipulation ont été redéposés sur le deuxième gel. La figure 118 présente le schéma d'analyse.

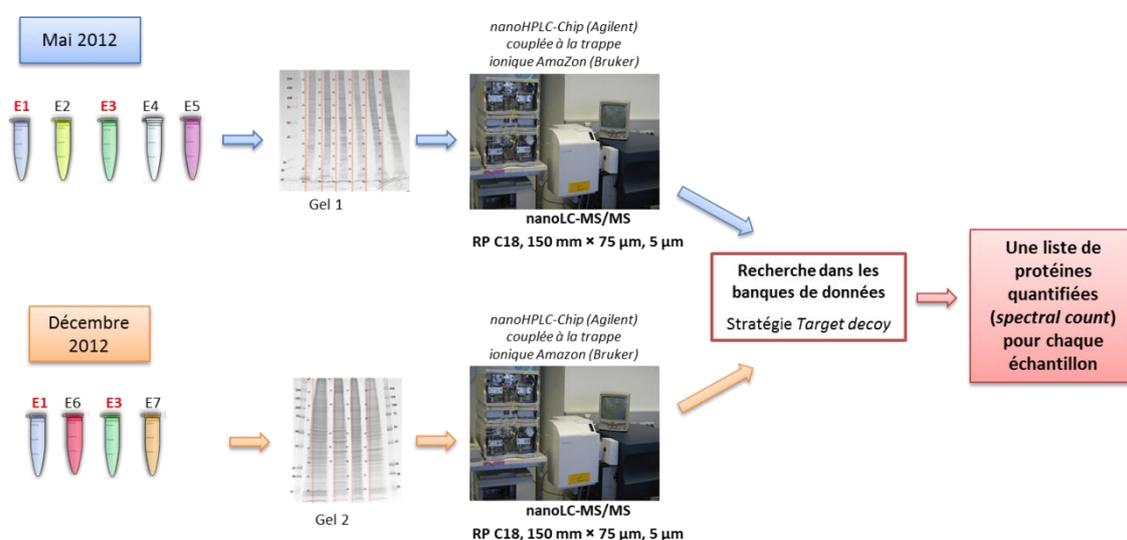


Figure 118
Schéma de l'analyse réalisée

La première série d'échantillons a été déposée sur le gel 1 en mai 2012 puis analysée sur la trappe Amazon. La deuxième série a été déposée en décembre 2012 sur le gel 2 puis analysée sur la trappe Amazon.

À noter que les échantillons 1 et 3 ont été déposés et analysés deux fois. Les recherches dans les banques de données ont été effectuées en même temps pour les deux analyses. Une liste de protéines quantifiées par *spectral count* a pu être proposée pour chaque échantillon.

La figure 119 montre les résultats obtenus en termes de nombre d'identifications protéiques. Dans l'analyse 1, environ 30 % de protéines supplémentaires ont été identifiées.

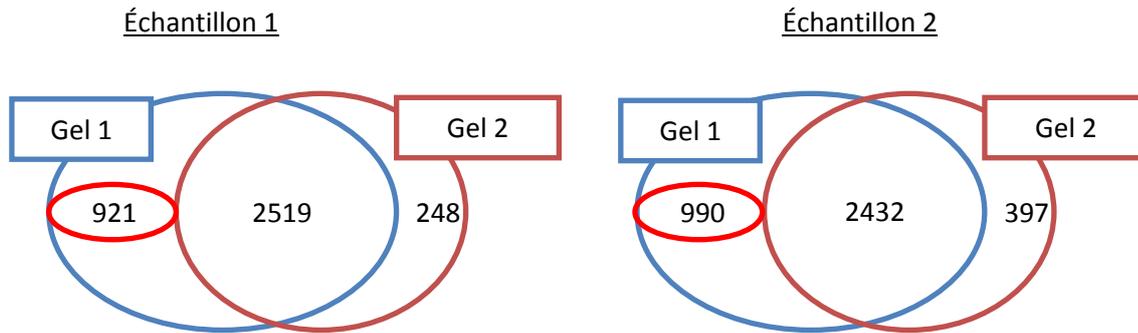


Figure 119

Comparaison du nombre de protéines identifiées dans chacun des deux échantillons déposés sur chaque gel

Une analyse du nombre de protéines identifiées par bande montre des différences qui semblent résulter d'une migration différente au sein du gel (cf. figure 120). Ainsi, pour chaque échantillon, le nombre de protéines identifiées diffère d'une bande à l'autre.

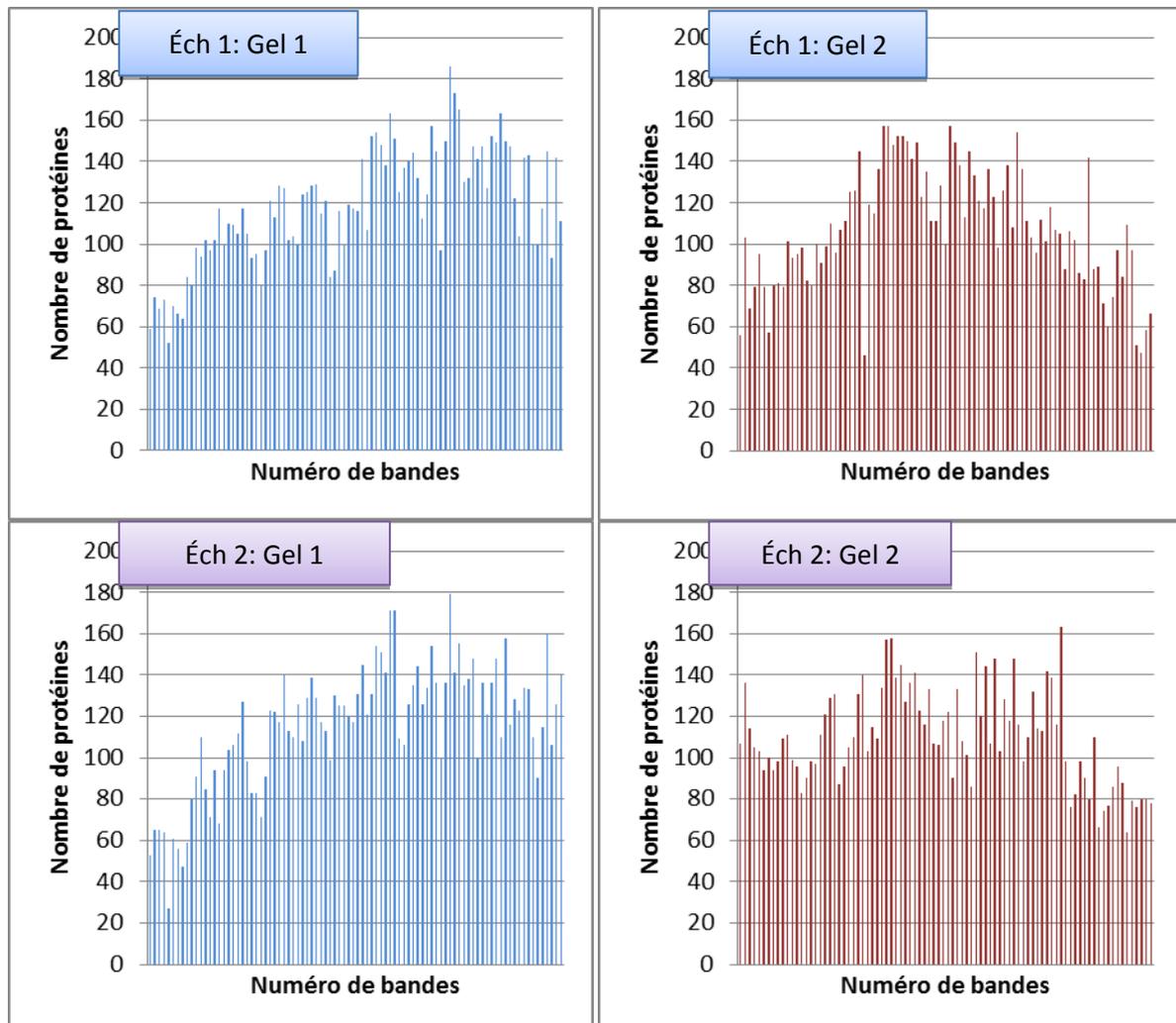


Figure 120

Nombre de protéines identifiées par bande dans chaque échantillon et sur chaque gel

Afin de s'affranchir de ce problème, une normalisation des données a été envisagée. En effet, les variations étant ici plus fortes que celles observées sur le digestat de levure, l'hypothèse a été formulée que la normalisation permettrait de corriger ces dernières. Plusieurs méthodes ont été testées :

- Pas de normalisation
- Normalisation par le nombre total de spectres totaux assignés
- Normalisation par le nombre total de spectres assignés dans chaque groupe de 10 bandes
- Normalisation par le nombre total de spectres assignés par groupe de masses (9 groupes de 20 000 Da)
- Normalisation par le nombre total de spectres assignés à la trypsine par groupe de 10 bandes

Les résultats obtenus, présentés dans la figure 121, montrent le coefficient de variation médian issu de la comparaison de l'échantillon 1 (gels 1 et 2) suivant différents types de normalisation et en appliquant des filtres sur le nombre de spectres minimum exigé par protéine (aucun, nombre de spectres supérieur à 1 et nombre de spectres supérieur à 5). La normalisation n'apporte pas de réelle amélioration des coefficients de variation médians. Elle empire même parfois les résultats (cf. normalisation par rapport à la masse de la protéine ou normalisation par la quantité de trypsine). Par contre, si l'on considère uniquement les protéines avec un nombre de spectres supérieur à 5, les coefficients de variation médians deviennent corrects (environ 21 % de variation entre les deux analyses). Ici, on notera que les valeurs obtenues sont comparables à celles obtenues pour la levure malgré l'ajout d'une importante source de variabilité technique.

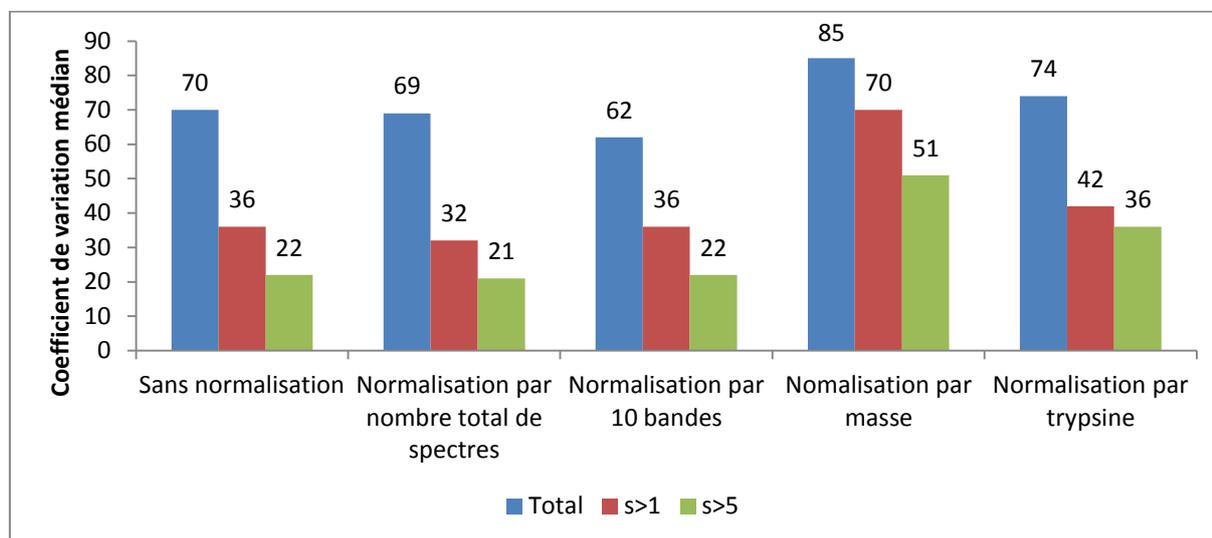


Figure 121
Coefficient de variation médian calculé pour les analyses de l'échantillon 1, entre le gel 1 et le gel 2, en fonction du type de normalisation utilisé et du nombre minimum de spectres exigés

Ainsi, si l'on considère uniquement les protéines ayant un nombre de spectres minimal de 5, plus de 50 % des protéines ont un coefficient de variation médian inférieur à 30 % (cf. figure 122).

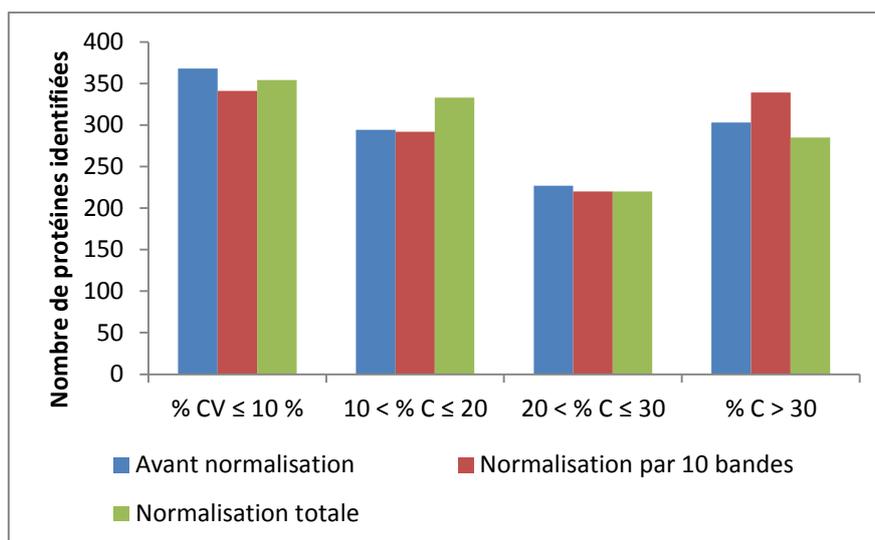


Figure 122

Nombre de protéines identifiées suivant le coefficient de variation médian et le type de normalisation appliqué

Pour conclure, une comparaison des échantillons analysés au cours du projet « glioblastome » semble pertinente car les coefficients de variation médians restent similaires à ceux qui ont été obtenus lors de l'évaluation de la reproductibilité des analyses du digestat de levure. Cependant, l'utilisation de statistiques apparaît ici risquée : au vu des biais, le nombre de faux positifs risque d'être élevé. C'est pourquoi, dans le cadre de ce projet, une interprétation manuelle sur des jeux de protéines potentiellement intéressantes a été réalisée.

C. À la recherche des protéines RCPG

C.1. Objectifs

L'objectif de cette première partie est de corréliser l'expression des transcrits RCPG (récepteurs couplés à la protéine G), déterminée par transcriptomique par l'équipe de Jacques Haiech et Marie-Claude Kilhoffer, à celle des protéines obtenues par analyse protéomique. Les RCPGs sont particulièrement intéressantes dans la recherche de biomarqueurs car elles jouent un rôle crucial dans la plupart des fonctions biologiques clés telles que la communication, la signalisation et la migration cellulaire, la prolifération et l'apoptose³⁴⁰. Elles sont de ce fait des cibles idéales pour les thérapies^{341, 342}. Ainsi, 30 % des médicaments actuellement commercialisés sont dirigés contre des RCPGs.

Afin de proposer de nouvelles cibles thérapeutiques, une étude transcriptomique différentielle a été menée sur les échantillons HA, U87, f-NSC, TG01 et OB1. L'expression de 356 RCPGs a été déterminée par RT-QPCR (*real time quantitative polymerase chain reaction*) sur une puce à ADN (TaqMan® GPCR low density microarrays, Applied Biosystems). 138 protéines ont été retenues pour une étude descriptive.

La détection de l'expression de ces gènes au niveau protéique reste un challenge analytique du fait de leur faible abondance et de leur structure. Ces cellules possèdent en effet (cf. figure 123) :

- 7 domaines transmembranaires aux séquences très hydrophobes de 20 résidus d'acides aminés enchâssés dans la double couche lipidique ;
- des domaines accessibles (milieu intra ou extracellulaire) avec des séquences courtes.

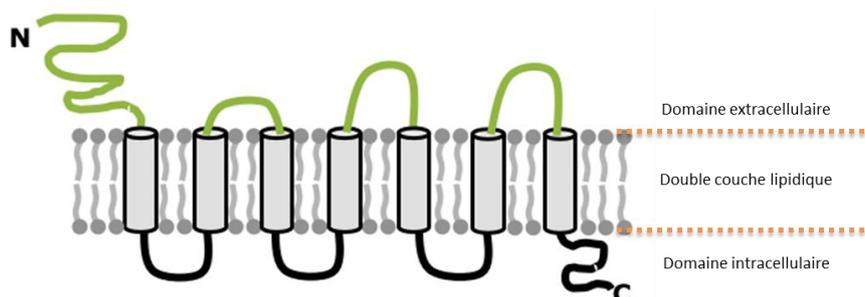


Figure 123

Représentation schématique d'une protéine RCPG

Cette famille de récepteurs membranaires possède sept domaines transmembranaires reliés entre eux par trois boucles extracellulaires et trois boucles intracellulaires.

Par ailleurs, il en existe un nombre important d'isoformes. Enfin, ces protéines sont très peu abondantes au sein de l'échantillon.

Du fait de ces limitations, une stratégie de validation par utilisation d'anticorps reste peu envisageable. En effet, la production de ces molécules a un prix dissuasif et la technique reste peu spécifique pour cibler des RCPGs.

Dans cette optique, une stratégie protéomique d'analyse haut débit a été mise en place. Le but était ici, dans un premier temps, de cartographier le protéome membranaire des cinq échantillons considérés, puis de comparer l'expression des RCPGs détectées par spectrométrie de masse à celle obtenue par transcriptomique.

C.2. Développements méthodologiques

Conscients des difficultés liées à la détection des RCPGs par analyse protéomique, nous avons œuvré pour augmenter la profondeur de nos analyses :

- **Répétition simple** : une deuxième injection du même échantillon permet d'apporter 12 % et 16 % d'identifications protéiques et peptidiques supplémentaires.
- **Répétition avec liste d'exclusion** : une deuxième injection du même échantillon avec une liste d'exclusion des peptides validés par des spectres MS/MS de bonne qualité lors de la première injection permet d'apporter 15 % et 35 % d'identifications protéiques et peptidiques supplémentaires.

Malgré l'augmentation du nombre de protéines et de peptides identifiés grâce à ces deux types de techniques, aucune RCPG supplémentaire n'a pu être détectée. Nous nous sommes alors orientés vers l'utilisation de liste d'inclusion. Il s'agit ici de cibler spécifiquement 55 RCPGs sélectionnés à partir de leur taux d'expression en transcriptomique.

La liste d'inclusion permet d'imposer une liste d'ions à sélectionner de manière continue et obligatoire sur l'ensemble des analyses. Chaque ion est sélectionné sur une certaine gamme de masse et pour un temps de rétention défini. Cette sélection forcée peut être utilisée afin d'obtenir des spectres de fragmentation de meilleure qualité, d'augmenter le nombre de peptides identifiés pour une protéine donnée et de rechercher des protéines supposées présentes. Il s'agit ici de s'approcher des techniques de protéomique ciblée et de forcer l'identification de protéines d'intérêt.

Une liste d'inclusion contenant les peptides d'intérêt a été créée pour chercher à identifier ces protéines. Un temps de rétention théorique ainsi qu'une erreur sur la détermination de la masse ont été assignés à chaque peptide.

C.2.1. Le choix des peptides

Les peptides ont été choisis suivant plusieurs critères :

- peptides protéotypiques, c'est-à-dire tryptiques et uniques à la protéine ;
- peptides présents dans le peptide Atlas³⁴³⁻³⁴⁶ (cette banque de données regroupe l'ensemble des peptides identifiés par nanoLC-MS/MS pour les protéines de la banque humaine) ;
- peptides de longueur de 8 à 25 résidus d'acides aminés.

Au total, pour les 55 RCPGs sélectionnées, 277 peptides ont été suivis, soit environ 5 peptides par protéine.

C.2.2. Application de critères de suivi des peptides en cours d'analyse

L'analyseur trappe ionique ayant une précision de mesure faible ($\pm 0,25$ Da), les peptides ciblés ont été recherchés avec une erreur de 0,7 Da à droite et 0,5 Da à gauche.

Par ailleurs, deux critères ont été appliqués afin de réduire la taille des listes d'inclusion :

- **Masse de la protéine**

Une liste d'inclusion différente a été paramétrée suivant la bande dans laquelle la protéine est supposée se trouver (cf. figure 124).

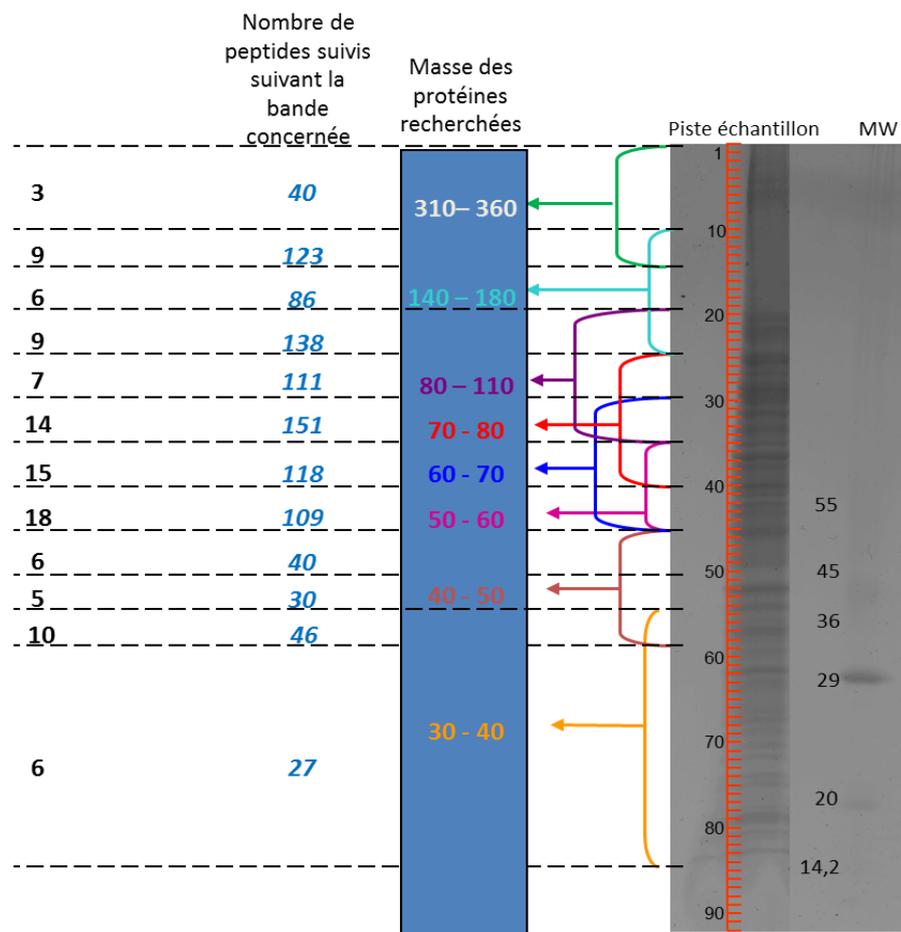


Figure 124

Division des listes d'inclusion suivant la zone de gel concernée

Les protéines ont été réparties en 8 groupes suivant leur masse. Chaque groupe est exclusivement sélectionné sur une partie du gel.

- **Temps de rétention théorique**

L'algorithme SSRCalc (*sequence specific retention calculator*)³⁴⁷⁻³⁵² a été utilisé afin de déterminer des temps de rétention théoriques pour chaque peptide suivi. SSRCalc permet de calculer l'hydrophobicité relative des peptides suivant plusieurs critères :

- la composition en résidus d'acides aminés et leur position (en position N- ou C-terminale, les résidus d'acides aminés n'ont plus de groupements amines ou carboxylates) ;
- la longueur du peptide ;
- l'hydrophobicité globale ;
- le pI (point isoélectrique) ;
- l'effet de charge des proches voisins (pour les résidus d'acides aminés adjacents à des lysines, arginines ou histidine) ;
- l'effet du voisinage d'une proline liée de manière covalente avec l'amine dans la chaîne peptidique ;

- la capacité à former des structures hélicoïdales (hélices α , structures β).

Le processus employé est détaillé dans la figure 125. Dans un premier temps, une droite de régression linéaire est tracée pour l'ensemble des peptides identifiés (11 228 peptides) dans la première analyse, représentant l'hydrophobicité en fonction du temps de rétention expérimental (cf. figure 126). Ensuite, afin de déterminer une erreur sur l'estimation du temps théorique, le temps de rétention expérimental a été tracé en fonction du temps de rétention théorique, déterminé à partir de la première droite (cf. figure 127). Une erreur de ± 20 min a ainsi été estimée sur la détermination théorique des temps de rétention. L'hydrophobicité relative de chaque peptide a été précisée puis, en utilisant la première droite tracée, un temps de rétention théorique a été calculé.

Pour chaque peptide, un temps de rétention théorique associé à une tolérance a ainsi pu être établi.

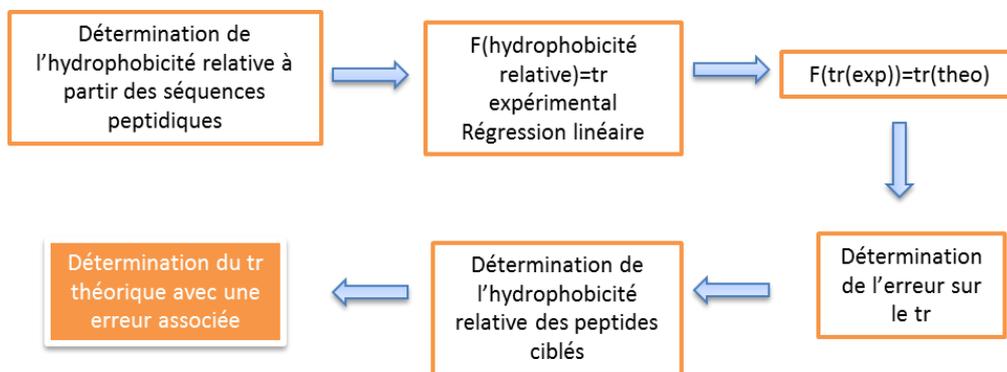


Figure 125

Processus permettant de déterminer le temps de rétention théorique de chaque peptide ciblé, associé à une erreur sur cette détermination

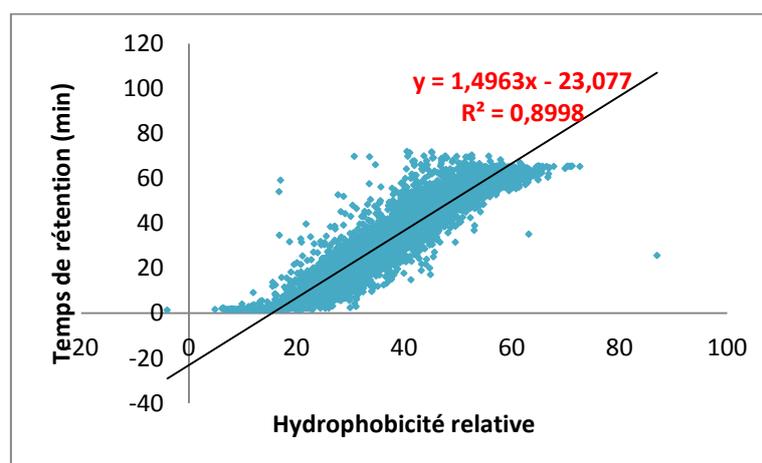


Figure 126

Régression linéaire : temps de rétention expérimental pour 11 228 peptides en fonction de l'hydrophobicité relative déterminée avec l'algorithme SSRCalc

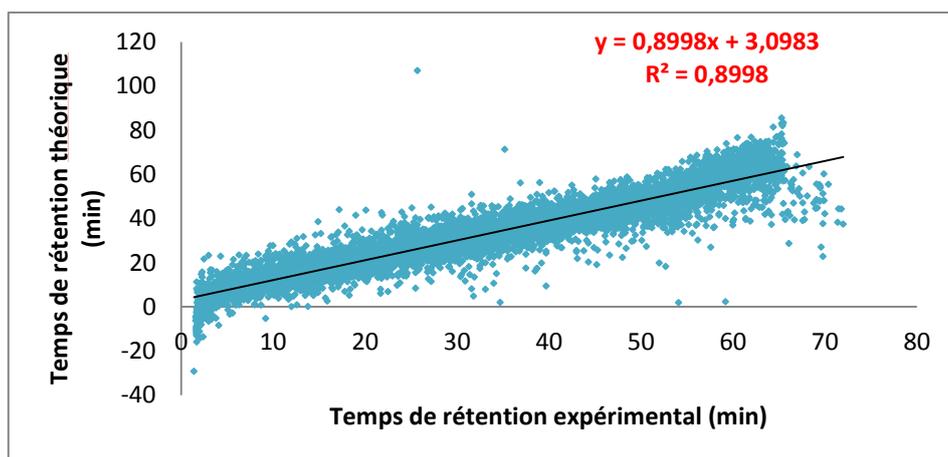


Figure 127

Régression linéaire du temps de rétention théorique en fonction du temps de rétention expérimental

C.2.3. Résultats

Les résultats obtenus sont décrits dans le tableau 15. Au total, parmi les 55 protéines recherchées, 6 ont pu être identifiées dont une jamais vue dans les précédentes analyses. Et parmi les 277 peptides ciblés, 11 ont pu être identifiés.

Protéine identifiée	Nombre de peptides issus de l'analyse classique	Nombre de peptides ciblés	Nombre de peptides ciblés déjà identifiés	Nombre de peptides supplémentaires identifiés grâce à la liste d'inclusion
CD97	2	8	1	2
GPR56	2	8	1	3
LPHN2	8	7	1	2
GPR5B	0	5	0	2
GPR107	1	8	1	1
GPR180	1	8	0	1

Tableau 15

Ensemble des protéines identifiées par les protéines ciblées avec le nombre de peptides préalablement identifiés, le nombre de peptides ciblés, le nombre de peptides ciblés déjà identifiés et le nombre de peptides identifiés grâce à la liste d'inclusion

Même si une nouvelle protéine a pu être identifiée (GPR5B), les résultats sont plutôt décevants. Finalement, la liste d'inclusion a permis de ré-identifier des peptides pour la plupart déjà détectés précédemment avec de meilleurs spectres de fragmentation. Cela montre que la liste d'inclusion ne permet pas d'identifier de nouvelles protéines ou de nouveaux peptides mais plutôt de confirmer les identifications. Par ailleurs, avec le recul, cette expérience a été conduite sur un spectromètre peu adapté, sa résolution étant faible et sa gamme dynamique peu étendue.

C.3. Résultats obtenus

Grâce à notre stratégie protéomique, nous avons détecté 7 protéines RCPGs. Le tableau 16 présente une comparaison du taux d'expression du transcrit avec celui obtenu au niveau protéique par le *peptide count* et le *spectral count* pour les 15 transcrits RCPGs les plus exprimés. Un nombre de peptides tryptiques potentiellement détectables par spectrométrie de masse a été déterminé suivant 4 critères :

- Pas de coupures manquées
- Masse comprise entre 1 000 et 3 000 Da
- Pas de modifications post-traductionnelles
- Pas de peptides issus du domaine transmembranaire

Les résultats montrent une corrélation pour certaines protéines, par exemple GPR56 (cf. figure 128). Par contre, d'autres protéines ne sont pas détectées par spectrométrie de masse malgré un taux élevé d'expression du transcrit. Deux explications peuvent être avancées :

- Pas de corrélation entre l'expression du transcrit et celle de la protéine correspondante. Cet état de fait est discuté dans une publication récente de C. Vogel et coll. : seuls 40 % des concentrations protéiques peuvent être reliées à l'abondance des transcrits. Ceci montre l'importance des phénomènes de régulation qui ont lieu pendant ou après la traduction^{353, 354}.
- Peu de chances de détecter la protéine car il y a peu de peptides tryptiques potentiellement détectables.

Au vu des résultats, nous pouvons considérer que les transcrits sont détectés en protéomique sur le couplage nanoLC-Chip – trappe ionique AmaZon, si les protéines comprennent au moins 5 peptides détectables et un niveau d'expression de 20 000 en transcriptomique, à la condition qu'ils soient traduits.

Nom du RCPG	Numéro d'accension	Poids moléculaire (kDa)	Nombre de peptides attendus	OB1		TG01		f-NSC		HA		U87	
				T Expression du transcrit ($2^{-Ct} \times 10^{12}$)	P Expression de la protéine : nb de peptides uniques (Nb de spectres)	T	P	T	P	T	P	T	P
GPR56	Q9Y653	78	14	42820	4 (8)	52347	4 (10)	155121	8 (37)	625	nd	32369	3 (6)
LPHN2	Q95490	163	50	37915	9 (15)	62685	12 (24)	2338	nd	5921	nd	3495	nd
F2R	P25116	44	5	48045	2 (6)	26751	nd	18912	nd	7701	nd	1259	nd
BDKRB2	P30411	44	nd	78	nd	0	nd	0	nd	1655	nd	42664	nd
FZD7	O75084	61	3	39613	nd	10182	nd	14788	nd	30943	nd	4011	nd
GPRC5B	Q9NZH0	41	3	24853	1 (3)	24464	1 (4)	33650	3 (7)	2888	nd	248	nd
TM7SF1	O60478	45	3	15303	nd	3322	nd	29968	nd	2855	nd	1743	nd
CXCR4	P61073	40	2	29644	nd	14194	nd	11354	nd	120	nd	19	nd
BAI2	60241	173	40	19770	nd	15863	nd	29243	1 (1)	5050	nd	5478	nd
BDKRB1	P46663	40	2	5	nd	0	nd	9	nd	65	nd	27717	nd
GPR153	Q6NV754	65	1	8111	nd	26080	nd	7762	nd	2438	nd	342	nd
CD97	P48960	90	19	14593	10 (40)	24959	9 (31)	405	nd	1079	nd	5233	6 (21)
LGR4	Q8N537	104	8	22726	nd	4136	nd	19526	nd	1804	nd	507	nd
CNR1	P31554	53	3	20381	nd	3161	nd	14232	nd	190	nd	0	nd
CELSR2	Q9HCU4	317	79	7543	2 (2)	1769	2 (2)	3501	8 (12)	60	nd	118	nd

Tableau 16

Présentation des 15 gènes possédant le taux d'expression le plus fort en transcriptomique

Pour chaque protéine, le nombre de peptides attendus a été déterminé suivant 4 critères : pas de coupures manquées, masse comprise entre 1 000 et 3 000 Da, pas de modifications post-traductionnelles, pas de peptides transmembranaires. Pour chaque lignée étudiée, la colonne T indique le taux d'expression du transcrit $2^{-Ct} \times 10^{12}$ et la colonne P le taux d'expression de la protéine (nombre de spectres MS/MS). À noter que Ct se réfère au cycle seuil (*cycle threshold*) à partir duquel l'intensité de fluorescence est plus importante.

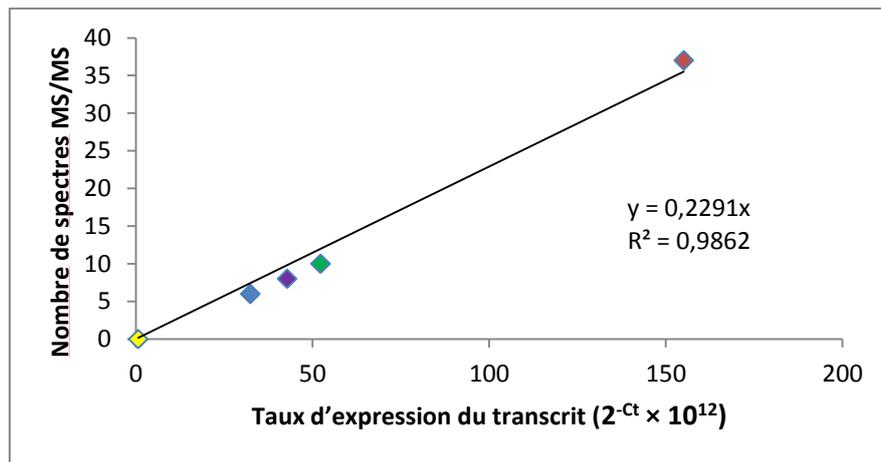


Figure 128

Corrélation linéaire entre le nombre de spectres MS/MS et le taux d'expression du transcrite pour la protéine GPR56

C.4. Conclusion

Il apparaît ici difficile de conclure quant à la comparaison des expressions des RCPGs au niveau du transcrite et de la protéine. En effet, deux points clés sont à considérer :

- les limites de l'analyse protéomique à haut débit, et notamment la difficulté d'identifier les protéines membranaires peu abondantes et très hydrophobes ;
- la possibilité de non-corrélation entre l'expression du transcrite et l'expression de la protéine.

Il serait ici judicieux d'envisager une méthode d'enrichissement supplémentaire à l'image de celle proposée par Eisen et coll.³⁵⁵, basée sur l'utilisation d'anticorps ciblant des épitopes spécifiques de plusieurs RCPG.

C.5. Communication des résultats

Ces résultats ont fait l'objet :

- d'une **communication par affiche** au second meeting annuel du GDR RCPG-PhysioMed à Strasbourg en octobre 2013 (annexe 5) :

Lennon S., Saliou JM., Fève M., Zeniou M., Carapito C., Dong J., Van Dorsselaer A., Junier MP., Chneiweiss H., Cianféran S., Haiech J., Kilhoffer MC.

Proteomic Analysis of the GPCR Repertoire in Glioblastoma Stem Cells. Comparison to Non-Cancer Cells

- d'une **publication dans le journal Plos One** (annexe 6) :

Fève M.*, Saliou JM.*, Zeniou M.*, Lennon S., Carapito C., Dong J., Van Dorsselaer A., Junier MP., Chneiweiss H., Cianféran S., Haiech J., Kilhoffer MC.

Comparative Expression Study of the Endo-G Protein Coupled Receptor (GPCR) Repertoire in Human Glioblastoma Cancer Stem-like Cells, U87-MG Cells and Non-Malignant Cells of Neural Origin Unveils New Potential Therapeutic Targets

Plos One, 2014, **9**, Issue 3

*Ces auteurs ont contribué de manière égale à ce travail.

D. Analyse différentielle des clusters de différenciation

D.1. Objectifs

Malgré les développements spécifiques réalisés, le bilan de la détection des RCPGs dans les échantillons et l'analyse différentielle de cette famille de molécules restait décevant. Nous avons alors décidé de réorienter le projet vers une étude différentielle du protéome membranaire global.

L'objectif de cette deuxième partie était d'établir le protéome membranaire des échantillons HA, f-NSC, U87-MG, TG01, OB1, TG16 et TG10B afin de comparer le protéome des cellules souches cancéreuses à celui des astrocytes. f-NSC et U87-MG sont utilisées en tant qu'échantillons standards. Il est important de noter ici que la lignée U87-MG est issue d'un patient atteint d'un glioblastome de grade IV décédé en 1966. Depuis ce temps, les cellules ont subi des mutations et leur phénotype s'est peu à peu éloigné de celui de la pathologie. Cependant, il s'agit d'une lignée couramment utilisée pour l'étude des glioblastomes, elle a donc été incluse dans notre analyse à titre indicatif.

Par ailleurs, il sera important de tenir compte du fait que TG01 et OB1 sont p53 non mutées alors que TG16 et TG10 sont p53 mutées. Le gène p53^{356, 357} est le plus communément muté dans les cancers. Il a été élu molécule de l'année en tant que gardien du génome par le journal *Science* en 1993³⁵⁸. Ce facteur de transcription permet en effet la réparation de l'ADN, ainsi que le contrôle du cycle cellulaire et de l'apoptose. Il est muté dans la moitié des tumeurs.

D.2. Résultats obtenus

TG16 et TG10B ont été injectées six mois après la première analyse pour des raisons techniques. Malgré cela, il a été montré que les différentes lignées cellulaires restaient comparables.

Entre 3000 et 4000 protéines ont été identifiées dans chaque échantillon (FDR < 1 %), dont 800 à 1000 protéines membranaires (cf. figure 129).

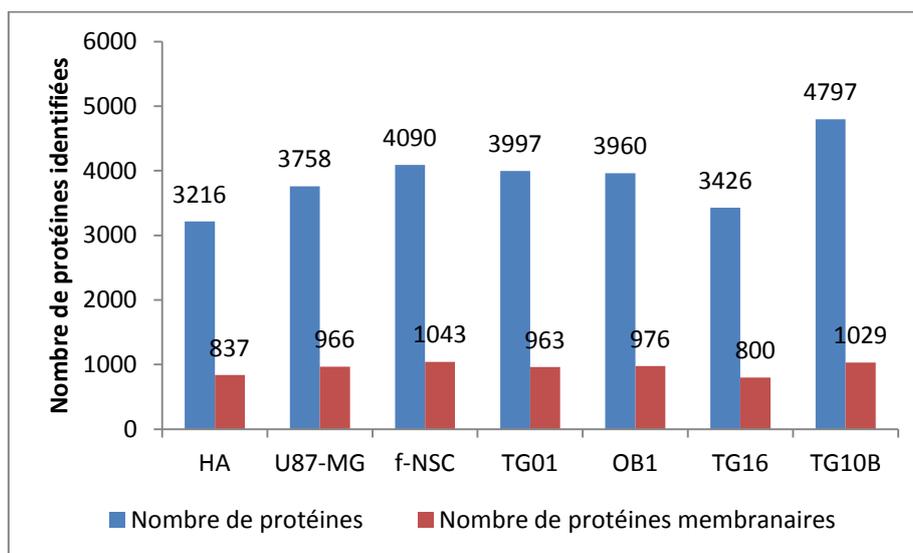


Figure 129

Nombre de protéines totales et de protéines membranaires identifiées dans chaque type de cellules

Dans un premier temps, les protéines membranaires et en particulier les clusters de différenciation (CD) ont été étudiés. 117 CD ont ainsi été identifiés. Afin de déterminer les cibles d'intérêt, une *heat map*, présentée en figure 130, a été réalisée sur la base du *spectral count* obtenu pour chaque CD dans chaque lignée. La *heat map* permet de regrouper les échantillons similaires. Il est donc possible de noter que TG01/OB1, d'une part, et TG16/TG10, d'autre part, présentent des profils similaires. Il est intéressant de relever que cette classification concorde avec la mutation ou non du gène p53.

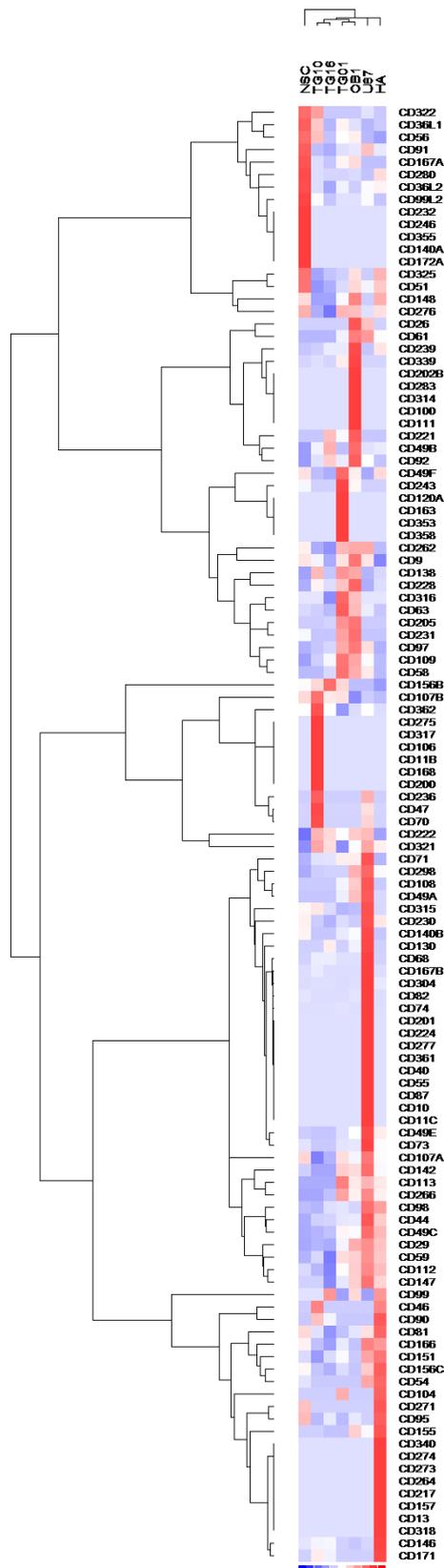


Figure 130
 Heat map représentant l'abondance de l'ensemble des clusters de différenciation des 7 échantillons (NSC, TG10, TG16, TG01, OB1, U87, HA)

Les similitudes entre les profils sont visualisées par un regroupement hiérarchique. Les couleurs bleu et rouge indiquent le niveau d'expression de chaque protéine (bleu = sous-exprimé, rouge = surexprimé).

Arbitrairement, seules les protéines ayant un nombre de spectres supérieur à 5 dans au moins un des échantillons ont été conservées. Puis HA a été comparée à TG01/OB1 d'une part et TG16/TG10B d'autre part par le calcul de ratio. À cette étape, deux critères ont été établis pour la sélection :

- Ratio des valeurs de *spectral count* > 5 dans au moins une des comparaisons
- Pertinence biologique

Ainsi, 12 CDs, dont les valeurs de *spectral count* et de *peptide count* sont présentés dans le tableau 17 ont été retenus.

	Nombre de peptides (nombre de spectres)						
	NSC	TG10	TG16	TG01	OB1	U87	HA
CD108	nd	nd	nd	2 (4)	5 (10)	9 (21)	nd
CD109	nd	16 (28)	22 (52)	32 (140)	29 (111)	24 (55)	9 (17)
CD167A	16 (47)	4 (6)	1 (1)	8 (15)	11 (20)	nd	nd
CD205	nd	nd	2 (2)	21 (51)	27 (60)	nd	nd
CD221	nd	nd	2 (5)	1 (2)	6 (9)	nd	nd
CD231	1 (2)	nd	nd	1 (6)	1 (8)	nd	nd
CD243	2 (4)	nd	nd	13 (25)	4 (6)	nd	nd
CD339	nd	1 (1)	nd	4 (6)	9 (22)	nd	nd
CD56	18 (110)	10 (73)	5 (18)	7 (45)	9 (52)	6 (14)	2 (3)
CD63	2 (25)	2 (20)	2 (11)	6 (113)	4 (71)	4 (33)	3 (17)
CD97	nd	4 (4)	2 (7)	9 (31)	10 (40)	6 (21)	nd
CD9	4 (8)	4 (6)	1 (2)	4 (8)	4 (14)	4 (8)	nd

p53 mutée

p53 non mutée

Tableau 17

Nombre de peptides avec, entre parenthèses, le nombre de spectres MS/MS identifiés dans chaque échantillon pour les 12 clusters de différenciation potentiellement intéressants

Les résultats sont actuellement en validation par immunohistochimie au laboratoire du Pr Jacques Haiech. Des résultats préliminaires ont permis de confirmer l'expression du CD205 d'une part (surexpression sur les GSCs non mutées et sous-expression sur les GSC mutées et HA) et CD109 d'autre part (surexpression forte sur les GSCs non mutées et surexpression plus légère sur les GSCs mutées par rapport à HA).

D.3. Conclusion

Les résultats sont très encourageants dans le sens où deux nouveaux potentiels biomarqueurs des cellules souches cancéreuses issues de glioblastomes ont pu être proposés grâce à une cartographie membranaire la plus exhaustive possible.

D.4. Communication des résultats

Les résultats obtenus ont fait l'objet :

- d'une **communication par affiche** au congrès SMAP de 2011 à Avignon (Spectrométrie de masse et analyse protéomique (cf. annexe 7) :

Lennon S., Saliou JM., Audran E., Chneiweiss H., Carapito C., Van Dorselaer A., Junier MP., Haiech J., Cianféroni S.

Développements méthodologiques en analyse protéomique pour l'établissement du protéome membranaire de cellules souches cancéreuses issues de gliomes

- d'une **communication orale** au 6^e workshop bio-informatique du Cancéropôle de 2013 à Strasbourg :

Lennon S., Saliou JM., Dong J., Chneiweiss H., Van Dorselaer A., Carapito C., Kilhoffer MC., Haiech J., Cianféroni S.

Membrane Proteomics to Search for Specific Glioblastoma Cancer Stem Cells Biomarkers

Une publication est en cours de rédaction.

E. Conclusions et perspectives

L'analyse protéomique mise en place a permis de :

- confirmer l'expression protéique de 7 RCPGs identifiés comme étant de potentiels candidats d'intérêt pour cibler les cellules souches cancéreuses issues de glioblastomes ;
- proposer une liste de 12 nouveaux candidats biomarqueurs potentiels.

Nous pouvons cependant regretter à l'issue de ce projet que, malgré l'ensemble des solutions proposées, les protéines RCPGs soient restées réfractaires à l'identification par analyse protéomique. Pour ce type de protéines, il serait peut-être judicieux d'envisager des enrichissements très spécifiques.

Les résultats issus de l'étude des clusters de différenciation sont toutefois très encourageants et prometteurs. Si leur expression différentielle est confirmée, ils ouvriront de belles perspectives pour le développement de nouveaux médicaments permettant de cibler spécifiquement les GSCs.

Preuve est faite que la stratégie protéomique mise en œuvre est efficace pour un faible nombre d'échantillons. L'idée était ici de cartographier de la manière la plus complète possible le protéome membranaire en se basant sur un enrichissement, et sur une optimisation fine de l'ensemble des paramètres de chromatographie, d'acquisition et de recherche dans les banques de données. La collaboration étroite avec les biologistes qui ont une connaissance approfondie du sujet a été, à mon

sens, le point le plus important de cette recherche. En effet, ces derniers sont le plus à même de poser les questions pertinentes et d'identifier les cibles potentielles.

Pour la suite du projet, il serait particulièrement intéressant d'envisager une validation à plus grande échelle sur des lignées cellulaires issues de différents patients. L'arrivée au laboratoire de nouveaux appareils, plus rapides, plus résolutifs et avec une meilleure gamme dynamique, permet d'envisager deux types de stratégies :

- **La protéomique globale**

En effet, la performance accrue des nouveaux instruments permet de ne réaliser qu'une seule analyse par échantillon ainsi qu'une quantification par extraction des courants d'ions un peu plus précise que le *spectral count*.

- **La protéomique ciblée**

Validation des candidats précédemment mis en évidence par le développement de méthodes de spectrométrie de masse ciblées (SRM) sur l'ensemble des lignées cellulaires disponibles.

CHAPITRE 3

Recherche de biomarqueurs prédictifs des rechutes du système nerveux central

Cette étude d'envergure nationale et financée par l'INCA (Institut National du Cancer) implique :

- 6 services cliniques dirigés par le Pr Catherine Thieblemont (Hôpital Saint-Louis, Paris), le Pr Guy Laurent (Toulouse), le Dr Rémy Gressin (Grenoble), le Pr Raoul Herbrecht (Strasbourg), le Dr Carole Soussain (Centre René Huguenin, Paris) et le Dr Hervé Ghesquières (Centre Léon Bernard, Lyon).
- 3 équipes de protéomique : équipe des Drs Bernard Monsarrat et Odile Schiltz (IPBS, Toulouse), du Pr François Berger (Groupe Neuro-oncologie, Grenoble) et notre équipe du LSMBO (Strasbourg) associée à celle du Pr Laurent Mauvieux (Hôpitaux Universitaires de Strasbourg).
- 1 équipe de mathématiques appliquées à la statistique encadrée par le Pr Philippe Besse (Institut de Mathématiques, Toulouse).
- 1 équipe de transcriptomique dirigée par le Dr Rémy Houlgatte (Plateforme « Puces à ADN », Nantes)

L'ensemble est coordonné par le Pr Catherine Thieblemont.

A. De la nécessité de proposer des biomarqueurs prédictifs d'une rechute cérébrale dans les cas de lymphomes diffus à grandes cellules B

Les lymphomes diffus à grandes cellules B (LDGCB) concernent 30 à 40 % des cas de lymphomes, ce qui représente en France environ 15 à 20 cas pour 100 000 habitants par an (source : INCA). L'âge médian des patients est situé entre 60 et 70 ans. Cette pathologie très hétérogène est caractérisée par la prolifération de grandes cellules B. Elle peut apparaître dans les ganglions lymphatiques, mais aussi au sein d'organes extraganglionnaires comme l'os, le système nerveux central (SNC), les seins, le tube digestif, les sinus, les testicules, les amygdales ou la peau. Par ailleurs, certains cas résultent de l'évolution d'un lymphome indolent, tel que le lymphome folliculaire, le lymphome de la zone marginale ou le lymphome lymphoplasmocytaire³⁵⁹.

Le traitement combine une chimiothérapie à l'utilisation d'un anticorps monoclonal anti-CD20 (Rituximab). Une rémission, dont la durée est variable, est observée chez 50 à 60 % des patients³⁶⁰.

Environ 5 % des patients peuvent être concernés par une rechute au niveau du SNC. Cette complication rare est fatale et extrêmement rapide : les rechutes interviennent dans un délai médian de 6 mois et la durée médiane de survie après la rechute est de 5 mois^{361, 362}.

Certains facteurs permettent d'estimer le risque de rechute au niveau du SNC^{363, 364} :

- Un IPI³⁶⁵ (index pronostique international) élevé et la présence de symptômes B (fièvre, sueurs nocturnes, perte de poids). L'IPI est un outil permettant aux cliniciens de prévoir l'évolution des patients avec un lymphome non hodgkinien (LNH) agressif. Il existe plusieurs critères ayant un impact négatif :
 - Âge > 60 ans
 - Stade III ou IV, c'est-à-dire caractérisé par une maladie disséminée sus- et sous-diaphragmatique ou une atteinte extraganglionnaire
 - Taux élevé de LDH (lactate déshydrogénase)
 - Nombre de sites extraganglionnaires atteints > 1
 - Performance statut ≥ 2 (ce test permet d'évaluer l'état général d'un patient)
- Localisation à certains organes :
 - Testicules, ovaires, seins
 - Peau
 - Tissus mous
 - Moelle osseuse
 - Glandes surrénales
- Localisation tumorale dans des régions proches de la base du crâne :
 - Base de la langue
 - Glandes salivaires
 - Orbites
 - Sinus
 - Cavum

En cas de risque élevé de rechute, la prophylaxie consiste à mettre en place une thérapie sur la base d'injections intrathécales de méthotrexate. Cependant, ces mesures préventives ne sont pas optimales et ne permettent pas d'éliminer la possibilité de rechute. Le risque est ici de sur- ou sous-traiter les patients.

Dans ce contexte, l'objectif du projet Protell (PROtéines prédictives de la dissémination cérébrale des LDGCB) est d'identifier des biomarqueurs prédictifs des rechutes du SNC à la fois dans le LCR (liquide céphalo-rachidien), dans les microparticules plasmatiques circulantes issues de cellules tumorales

leucocytaires, et au sein de la tumeur primaire. Les deux premiers matériaux seront étudiés par protéomique et le dernier par transcriptomique.

Notre rôle dans le projet a été de rechercher des biomarqueurs prédictifs de rechute à partir de microparticules circulant dans le plasma.

B. Les échantillons

Les échantillons ont été collectés pendant trois ans sur des patients issus des six centres impliqués dans le projet en fonction de critères stricts afin de former deux cohortes :

- Cohorte 1 : patients à haut risque de rechute
 - Groupe 1 – présence de 3 des 4 critères suivants : taux élevé de LDH, nombre de sites extraganglionnaires ≥ 2 , PS ≥ 2 et symptômes B
 - Groupe 2 – Atteinte d'organes à risque
 - Groupe 3 – Localisation dans des régions proches de la base du crâne
- Cohorte 2 : patients présentant une tumeur cérébrale à l'inclusion
 - Groupe 1 – Lymphome LDGCB primitif du SNC
 - Groupe 2 – Rechute cérébrale

La création des deux cohortes a nécessité deux ans de collecte. Pour chaque patient, des échantillons de sang et de LCR ont été prélevés au moment de l'inclusion dans le protocole. Dans les deux heures suivant ce prélèvement, les échantillons sanguins ont été centrifugés pendant 10 minutes à 1300 g et 4 °C, de manière à collecter le plasma. Ce dernier a ensuite été réparti en aliquots de 1 mL dans des tubes en polypropylène et stocké à - 80 °C.

Au total, 128 patients ont été inclus dans le protocole, dont 125 sur lesquels un prélèvement plasmatique a été effectué. Parmi les 125 analyses réalisées par spectrométrie de masse, 102 seront conservées pour l'analyse statistique. En effet, 23 patients ont été jugés non pertinents au niveau clinique du fait de non-conformités lors de la collecte. Le tableau 18 présente la répartition des échantillons dans les différents groupes décrits précédemment.

Cohorte	1	1	2	2
Caractéristique	Non-rechute	Rechute au niveau du SNC	Atteinte primaire du SNC	Rechute du SNC à l'inclusion
Nombre d'échantillons	69	9	21	3

Tableau 18

Nombre d'échantillons suivant la cohorte concernée

C. Stratégie analytique mise en place

Dans un premier temps, nous nous sommes basés sur la stratégie mise en place pour le projet d'identification de potentiels biomarqueurs candidats des lymphomes à cellules B (cf. figure 131).

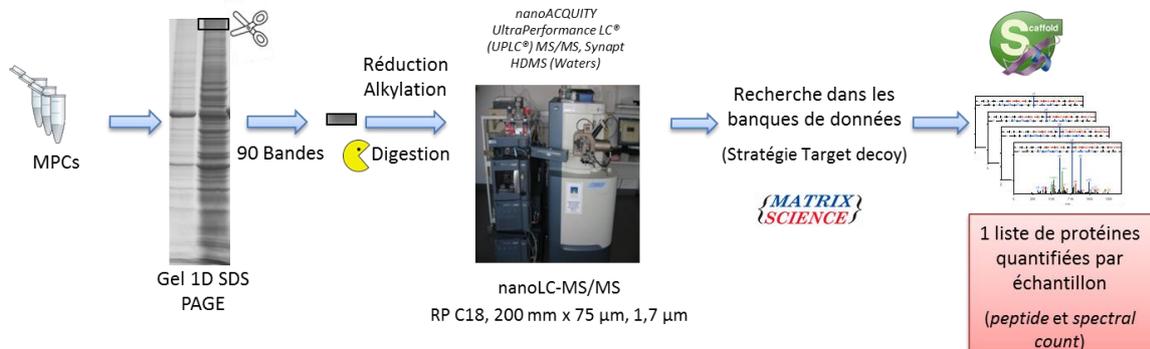


Figure 131

Stratégie mise en place pour la recherche de biomarqueurs candidats à visée diagnostique dans le cas des lymphomes B

Cependant, cette stratégie a nécessité quelques adaptations pour différentes raisons :

- L'étude a porté ici sur des microparticules plasmatiques circulantes (MPCs) et non plus des microparticules induites (MPIs) directement à partir de lymphocytes malins. En effet, les MPIs nécessitent une mise en culture des cellules malignes, stratégie peu envisageable pour l'analyse de 120 patients.
- Une semaine était nécessaire à l'analyse d'un échantillon. Cette durée est incompatible avec un grand nombre d'échantillons.
- Un nouvel instrument de type Q-TOF (Maxis Impact, Bruker Daltonics), caractérisé par une vitesse d'acquisition accrue, une meilleure sensibilité, une haute résolution ainsi qu'une précision de masse élevée, a été acquis par le laboratoire.

C.1. Les microparticules circulantes

Les MPCs ont été étudiées dans le chapitre 1 de la Partie II. Ainsi, nous avons montré :

- la possibilité de détecter des MPCs d'origine leucocytaire dans les échantillons malgré le nombre important de MPCs d'origine plaquettaire ;
- la répétabilité de la collecte en termes de nombre de MPCs obtenues ;
- l'intérêt d'utiliser 1 mL de plasma pour chaque patient afin de collecter les MPCs.

C.2. Évaluation du nouvel instrument Q-TOF

Les performances du nouveau Q-TOF ont été évaluées dans le chapitre 4 de la Partie II. Ainsi, nous avons démontré :

- la linéarité de réponse de l'appareil ;
- la répétabilité de la quantification des protéines sur 12 injections du même échantillon avec extraction des courants d'ions par le logiciel Skyline (coefficient de variation médian de 14 % et moyen de 18 %).

C.3. Évaluation du degré de séparation nécessaire pour l'analyse des MPCs

Plusieurs types de séparation de l'échantillon ont été testés. La question sous-jacente était de déterminer s'il était possible de réduire le fractionnement en 90 bandes sur gel de l'échantillon grâce à l'efficacité du nouvel appareil de spectrométrie de masse.

Comme discuté précédemment dans le chapitre 1 de la Partie III, le SDS reste le détergent le plus efficace pour la solubilisation des protéines. Dans cette optique, nous avons testé trois types de séparation sur gel :

- Gel 2 bandes
- Gel 6 bandes
- Gel 15 bandes

Le protocole de test est illustré par la figure 132. 6 échantillons de MPCs ont été mélangés puis redivisés en 6. Chaque aliquot a été déposé deux fois sur des gels différents pour chacune des trois séparations. Après réduction, alkylation et digestion des protéines, les mélanges peptidiques obtenus ont été injectés deux fois pour chaque préparation sur le Maxis Impact.

Le gradient a été adapté à chaque séparation dans le but de réaliser un maximum d'identifications dans un temps d'analyse raisonnable :

- 60 min de gradient pour chacune des bandes du gel 15 bandes (soit 15 h d'analyse par échantillon)
- 90 min pour le gel 6 bandes (soit 9 h d'analyse par échantillon)
- 120 min pour le gel 2 bandes (soit 4 h d'analyse par échantillon)

Pour chacune des 12 injections réalisées, nous avons obtenu une liste de protéines identifiées via le moteur de recherche Mascot et quantifiées via l'extraction des courants d'ions par Skyline.

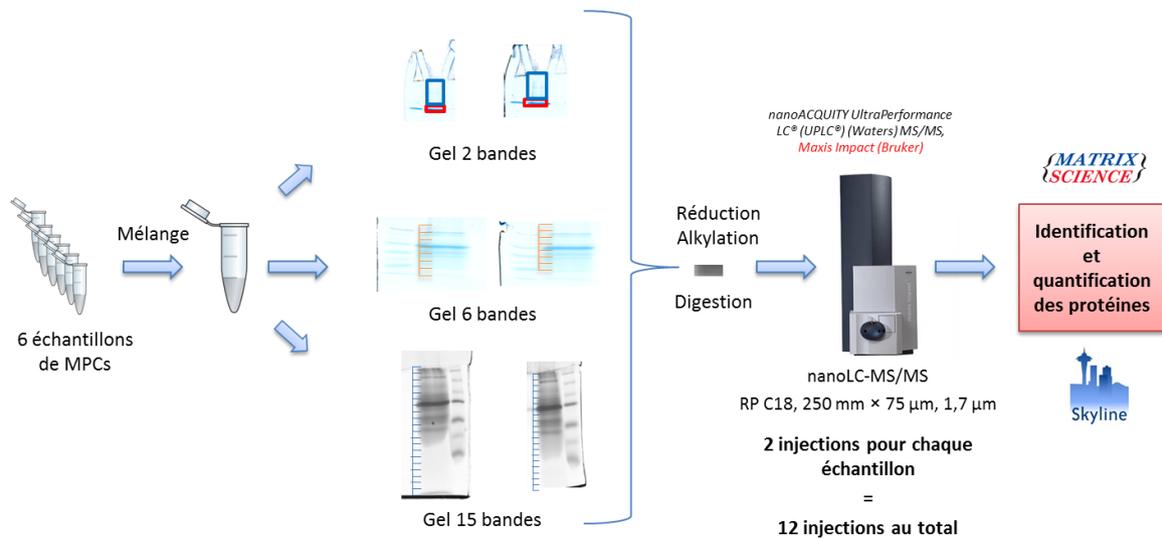


Figure 132

Test de différents types de séparation sur gel d'un même échantillon de MPCs

Pour chaque technique, les MPCs ont été déposées deux fois et chaque réplicat technique a été injecté deux fois sur le Maxis Impact.

Les résultats ont été évalués sur la base de deux critères :

- le nombre de protéines identifiées (FDR < 1 %) ;
- la répétabilité de la quantification.

La figure 133 montre le nombre de protéines identifiées pour chaque séparation. En moyenne, le gel 6 bandes permet de réaliser 34 % d'identifications supplémentaires par rapport au gel 15 bandes et 50 % par rapport au gel 2 bandes. L'augmentation de la séparation de 2 à 6 bandes a donc un fort impact sur le nombre de protéines détectées. Par contre, la séparation à 15 bandes est moins efficace. Ceci peut s'expliquer par la réduction du temps de gradient.

La séparation 6 bandes est ici la plus efficace en termes d'identifications. L'exploitation du gel 15 bandes ne sera pas poursuivie. En effet, il apparaît peu judicieux d'utiliser cette séparation dans le sens où le nombre d'identifications n'augmente pas alors que le temps d'analyse est supérieur à celui du gel 6 bandes.

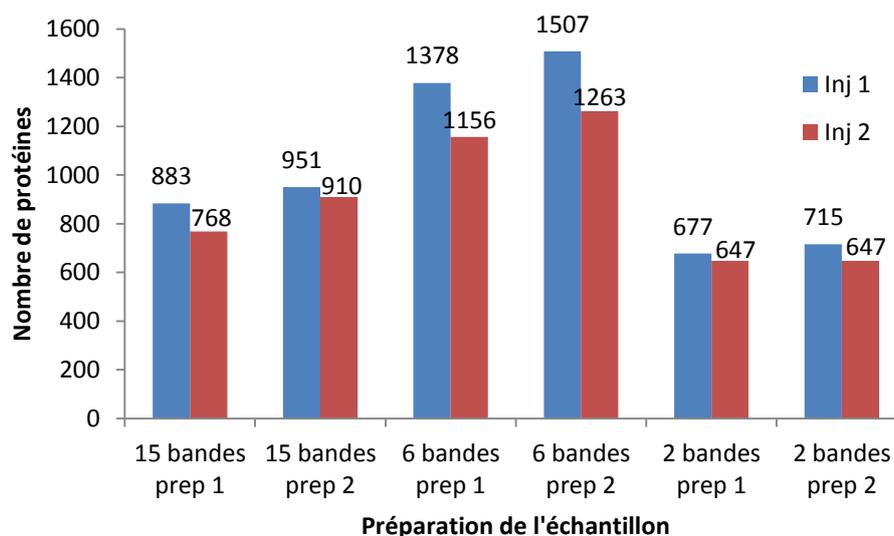


Figure 133
Nombre de protéines identifiées suivant la séparation sur le gel 1D SDS PAGE (15, 6 et 2 bandes), le réplicat de préparation et le réplicat d'injection

Dans un deuxième temps, nous avons évalué la reproductibilité de la quantification. La figure 134 présente la répartition des coefficients de variation de l'intensité de quantification obtenus pour chaque protéine sous forme de boîtes à moustaches. La médiane et la moyenne des coefficients de variation pour le gel 2 bandes sont quasiment deux fois plus faibles que celles du gel 6 bandes. Ceci est dû à l'amélioration de la séparation qui permet l'identification de protéines moins abondantes et donc plus difficilement quantifiables.

La séparation sur gel 6 bandes induit des questions et des difficultés supplémentaires : sera-t-elle reproductible pour une centaine d'échantillons ? Ces derniers seront-ils comparables ? Comment combiner les résultats des extractions d'ions pour une protéine identifiée dans plusieurs bandes ? Sans oublier que le temps d'analyse est multiplié par deux.

Finalement, au vu de toutes ces considérations, afin de simplifier l'analyse et l'interprétation des résultats et de limiter au maximum le temps passé à analyser chaque échantillon, nous avons opté pour le gel 2 bandes.

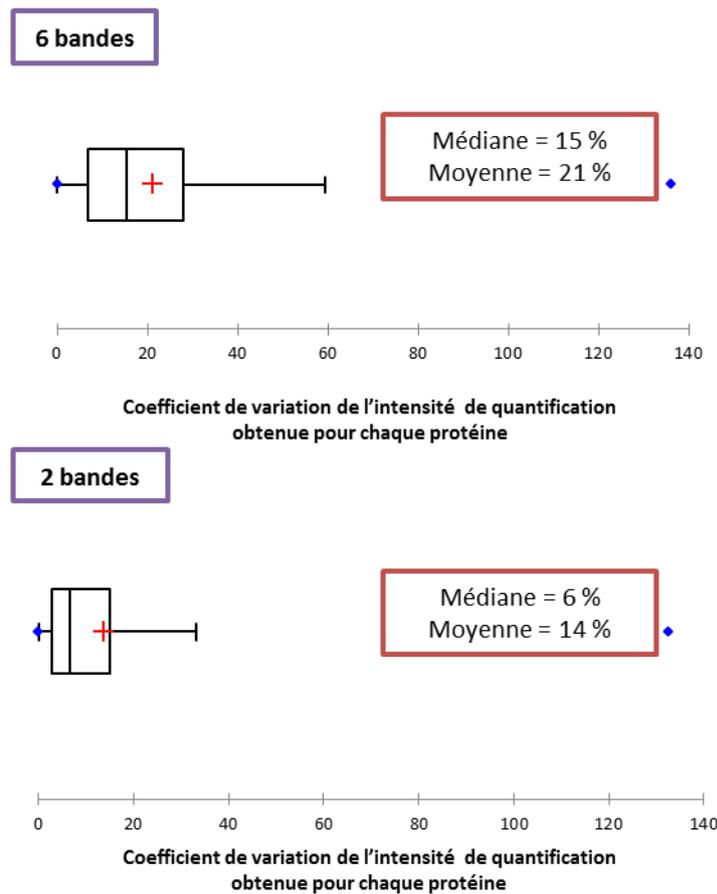


Figure 134

Boîtes à moustaches des coefficients de variation de la quantité de chaque protéine obtenue à la suite d'une séparation sur gel 6 bandes et gel 2 bandes

Remarque complémentaire : il a été envisagé de ne découper qu'une bande sur le gel, à l'image de ce qui a été réalisé pour le projet de quantification par SRM des marqueurs de lymphomes B (cf. Partie III – chapitre 1). Cependant, les MPCs sont contaminées par des protéines plasmatiques en forte abondance et notamment par des protéines de haut poids moléculaire, telles que l'apolipoprotéine B100 (515 kDa), ou la filamine A (280 kDa), qui occasionnent une forte diminution de signal et donc du nombre de protéines identifiées.

C.4. Stratégie protéomique mise en place

La stratégie protéomique mise en place pour l'analyse des 125 échantillons de MPCs est présentée en figure 135. Une fois décongelés, les échantillons de plasma sont centrifugés 2 min à 18 000 g afin d'obtenir des plasmas pauvres en plaquettes. Une centrifugation à 18 000 g pendant 45 min permet l'obtention de culots de MPCs qui sont ensuite lavés deux fois au PBS. Les microparticules sont ensuite reprises dans du tampon Laemmli et migrées pendant 45 min à 50 V dans un gel de concentration à 4 %. Deux bandes sont découpées puis les protéines sont réduites et alkylées. À ce stade, les échantillons sont stockés à - 20 °C. 26 gels 1D SDS PAGE ont ainsi été réalisés.

Les protéines ont été digérées et après extraction, des iRTs (mélange de 11 peptides standard commercialisés par la société Biognosys) ont été ajoutés dans le mélange de peptides avant injection sur le couplage nanoLC-MS/MS (Maxis Impact, Bruker Daltonics). Puis, les identifications ont été réalisées par recherche dans les banques de données avec un FDR < 1 %. Les courants d'ions ont été extraits via Skyline afin d'associer une donnée de quantification à chaque protéine.

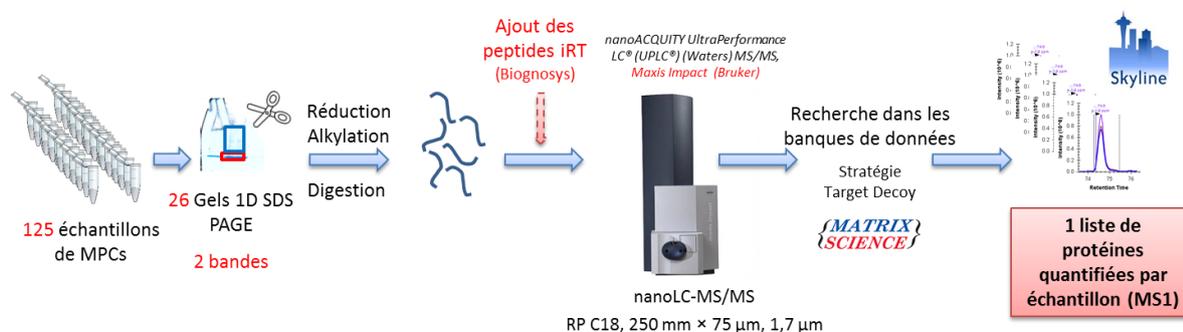


Figure 135

Schéma d'injection finale des 125 échantillons de MPCs collectés sur 125 patients atteints d'un LDGCB

Remarque 1 : les 125 échantillons ont nécessité quatre semaines d'analyse sur le couplage. Afin d'éviter une dégradation des mélanges peptidiques, 32 échantillons ont été digérés par semaine.

Remarque 2 : au vu du nombre d'échantillons, il n'a pas été possible d'envisager de réplicats techniques. Dans cette manipulation, seules les différences entre réplicats biologiques seront évaluées.

D. Mise en place d'indicateurs de suivi

Quatre semaines d'injection ont été nécessaires afin d'analyser l'ensemble des 125 échantillons. La mise en place de critères stricts de suivi de la qualité des analyses apparaissait ici obligatoire.

À cet effet, deux types d'échantillons contrôles ont été utilisés :

- analyse d'un digestat de BSA (albumine de sérum bovin) avant chaque échantillon afin d'évaluer par un contrôle rapide (gradient 9 min) l'état du couplage nanoLC-MS/MS ;
- analyse d'un mélange contrôle préparé avec 26 échantillons de MPCs, issus de patients atteints de diverses lymphopathies, afin d'évaluer la stabilité instrumentale dans les conditions d'injection des échantillons ainsi que la stabilité de nos préparations.

La figure 136 présente le schéma d'analyse mis en place, avec l'alternance des contrôles qualités et des échantillons. Les régénérations consistent en des lavages à 50/50 H₂O/ACN de la colonne pendant 20 min. Les échantillons ont été analysés dans un ordre aléatoire.

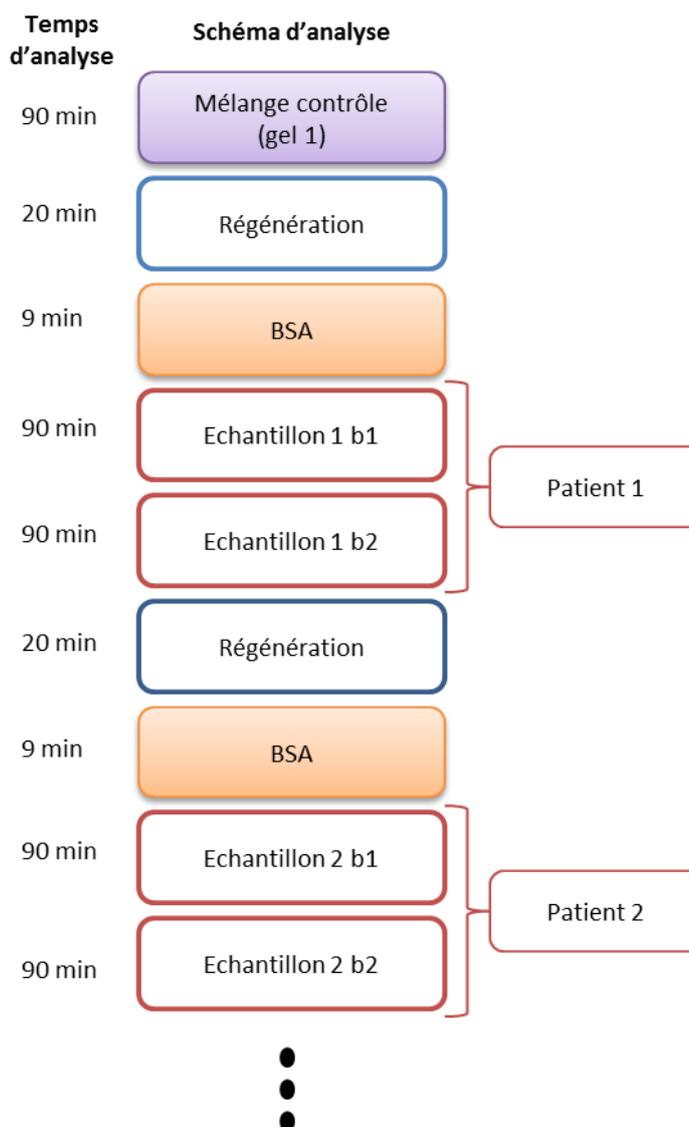


Figure 136

Schéma des injections réalisées lors de l'analyse des 125 échantillons du projet

Chaque gel comprend 1 mélange contrôle et 5 échantillons. Ces derniers sont découpés en deux bandes (b1 et b2).

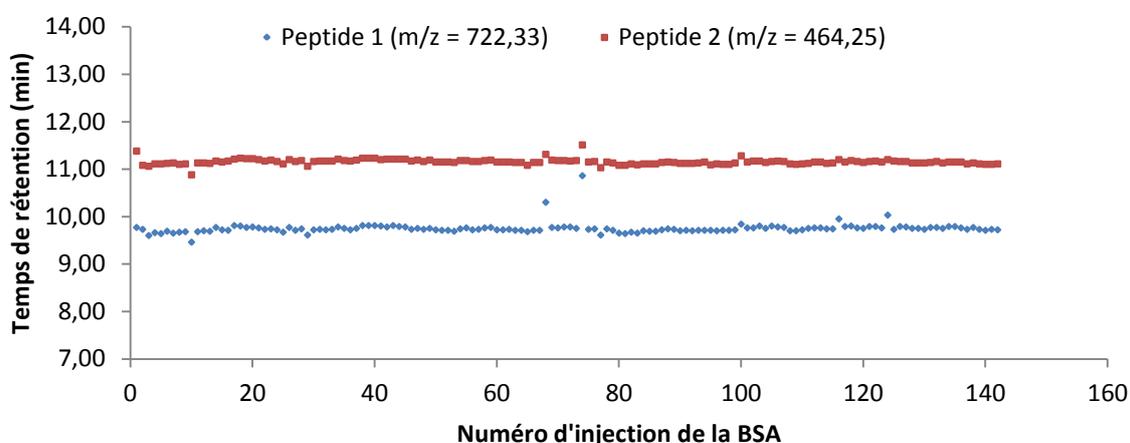
D.1. Contrôle qualité 1 : le digestat de BSA

Le contrôle qualité 1 est basé sur le suivi de deux peptides de masses différentes ($m/z = 722,33$ et $m/z = 464,25$) extraits du digestat de BSA. L'objectif est ici de vérifier de manière simple et rapide le bon fonctionnement du couplage nanoLC-MS/MS avant l'injection de chaque nouvel échantillon correspondant à un patient différent. Au total, sur les 4 semaines d'analyse, 143 digestats de BSA ont été injectés. À noter que la *lock mass* ($m/z = 922,0098$), masse de référence ionisée en permanence dans la source et permettant de mesurer à chaque instant la dérive de l'étalonnage en masse de l'appareil, a aussi été utilisée. Ce composé sert notamment à étalonner les spectres MS et MS/MS après acquisition. Il permet de s'affranchir des effets de dérive de masse dus à la dilation du tube de vol lors de variations de température.

D.1.1. Contrôles de la chromatographie

La reproductibilité chromatographique est un paramètre crucial pour assurer une extraction optimale des courants d'ions de chacun des peptides et donc une comparaison fiable des échantillons. Il s'agit ici de vérifier la stabilité de la colonne, l'absence de volumes morts, le bon fonctionnement du système et notamment des pompes. Deux paramètres ont été suivis sur chacun des deux peptides : reproductibilité des temps de rétention et largeur à mi-hauteur du pic chromatographique.

La stabilité des temps de rétention a été contrôlée tout au long des injections pour chacun des deux peptides extraits du digestat de BSA (cf. figure 137). Les coefficients de variation du peptide 1 et du peptide 2 sont de 1,2 % et 0,6 % attestant ainsi la stabilité de ce paramètre au cours des 143 injections réalisées. Les quelques valeurs déviantes ont été observées à la suite d'injections d'échantillons particulièrement concentrés en protéines.



[Figure 137](#)
Suivi des temps de rétention de deux peptides extraits du digestat de BSA (m/z = 722,33 et m/z = 465,25) au cours des 143 injections

La largeur à mi-hauteur (FWMH) de chacun des deux peptides a été suivie pour les 143 injections de BSA (cf. figure 138). Hormis pour les quinze premières injections, les valeurs sont largement inférieures au seuil critique d'intervention fixé à 0,10 min. La variation observée pour les quinze premiers échantillons est due à l'ordre d'injection. En effet, le digestat de BSA était analysé directement après l'échantillon. Ceci nous a permis de constater qu'un rééquilibrage de la colonne était nécessaire après l'analyse de chaque échantillon. Ce protocole a été mis en place pour les injections 15 à 143.

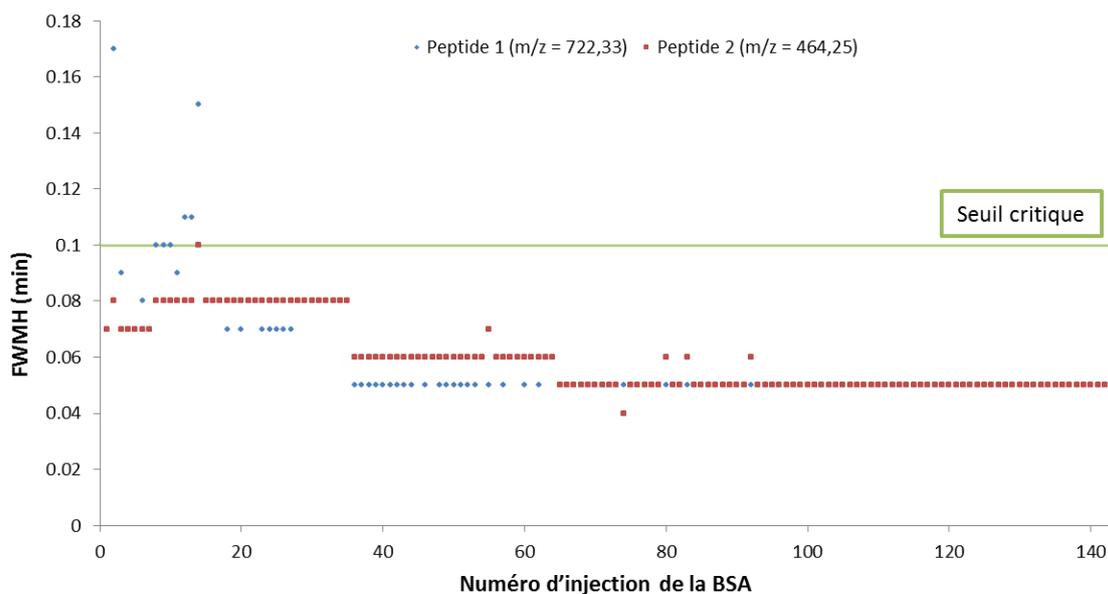


Figure 138

Suivi de la largeur à mi-hauteur (FWMH) pour chacun des deux peptides extraits du digestat de BSA ($m/z = 722,33$ et $m/z = 465,25$) tout au long des 143 injections

D.1.2. Contrôle des paramètres de spectrométrie de masse

Le suivi des paramètres de spectrométrie de masse a pour but d'évaluer la nécessité d'un nettoyage ou d'un réétalonnage de l'appareil. Trois indicateurs ont été suivis : l'intensité et la résolution spectrale des deux peptides extraits du digestat de BSA ainsi que l'erreur constatée sur la *lock mass* ($m/z = 922,0098$). Des valeurs définissant un seuil d'intervention ont été établies pour chacun des paramètres.

La figure 139 montre la variation des intensités des deux peptides extraits du digestat de BSA. Il s'agissait ici de vérifier que l'intensité était supérieure au seuil critique d'intervention ($I > 10^6$) déterminé lors d'expériences précédentes. Il est important de noter que ce paramètre ne permet pas de vérifier la reproductibilité des injections dans le sens où le digestat de BSA n'est pas stable au cours du temps.

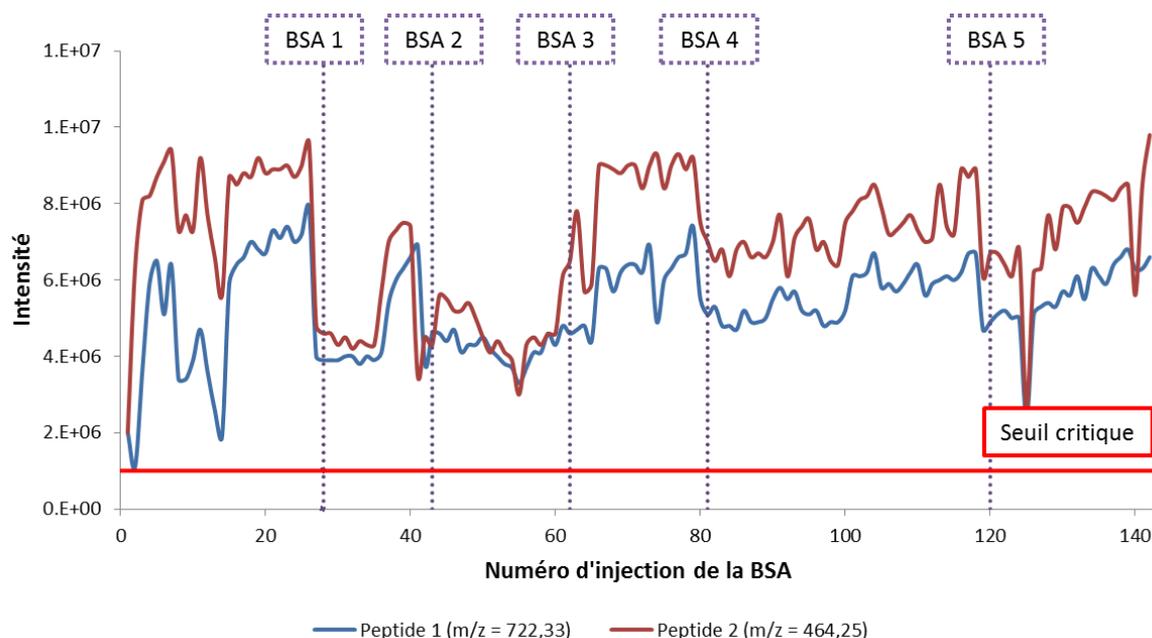


Figure 139
 Suivi de l'intensité des deux peptides extraits du digestat de BSA ($m/z = 722,33$ et $m/z = 465,25$) tout au long des 143 analyses
 La ligne rouge définit le seuil critique d'intervention ($I = 10^6$). Les traits verticaux indiquent le numéro de préparation de la BSA.

La figure 140 montre l'évolution de la résolution spectrale tout au long des analyses. Le seuil critique d'intervention a été fixé pour une résolution inférieure à 30 000. Toutes les valeurs sont bien supérieures à la valeur seuil.

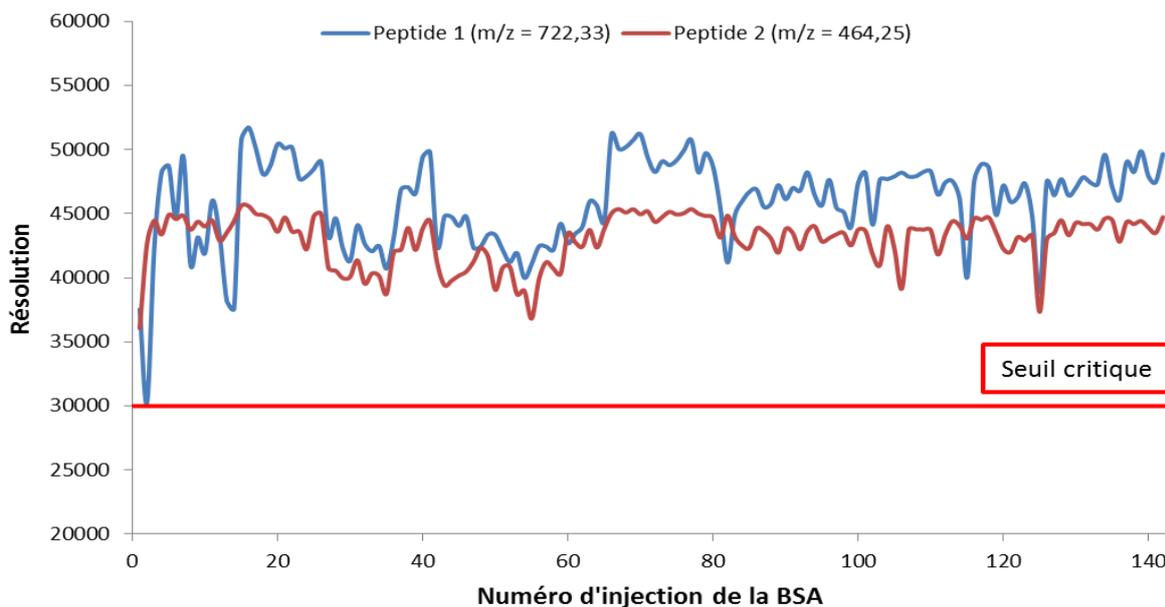


Figure 140
 Variation de la résolution spectrale pour les deux peptides extraits du digestat de BSA ($m/z = 722,33$ et $m/z = 465,25$) tout au long des 143 analyses
 La ligne rouge définit un seuil critique d'intervention ($R_s = 30\ 000$).

La figure 141 montre la variation de l'erreur sur la *lock mass* ($m/z = 922,0098$) tout au long des 143 analyses de la BSA. Il s'agissait ici d'évaluer directement le moment où un étalonnage de l'appareil était absolument nécessaire, c'est-à-dire lorsque l'erreur dépassait 20 ppm. Ce suivi a été crucial pour deux raisons :

- Un nouvel appareil de type Q-TOF était en cours d'installation dans la pièce où ont été réalisées les analyses. Les démarrages et arrêts successifs de cet instrument ont entraîné des variations de température importantes et donc une stabilité de l'étalonnage amoindrie.
- Tout nouvel étalonnage entraîne des manipulations de la source et comporte ainsi le risque non négligeable de détériorer le *sprayer*.

Au cours des quatre semaines, quatre étalonnages ont été effectués.

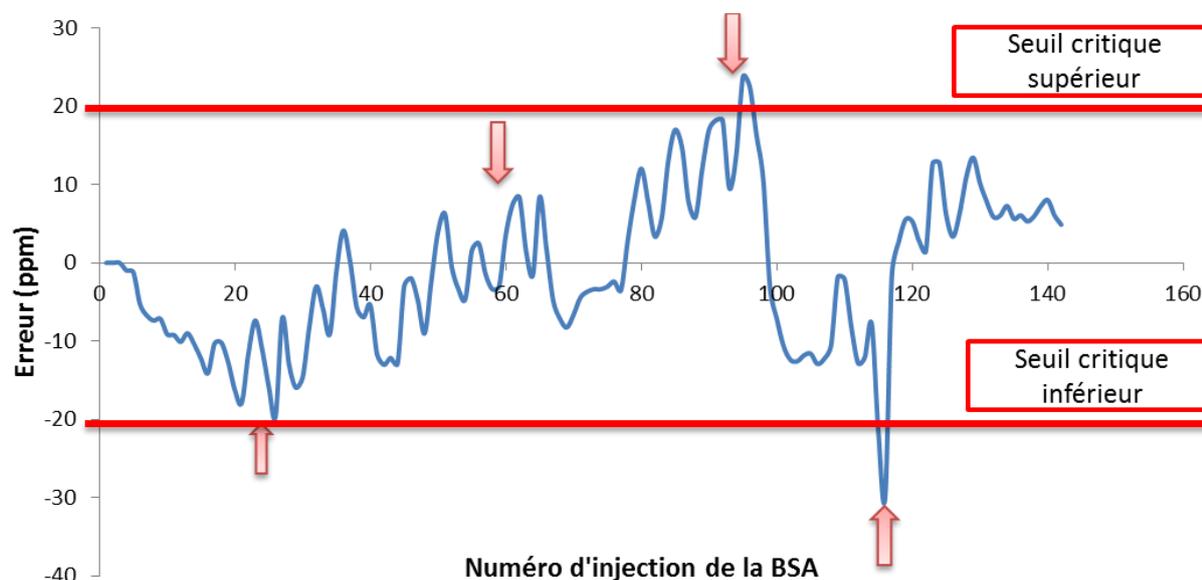


Figure 141

Variation de l'erreur sur la détermination de la masse de la *lock mass* ($m/z = 922,09$)

Le seuil critique d'étalonnage indiqué par les lignes rouges a été fixé pour une erreur de 20 ppm. Les flèches rouges indiquent les moments où un réétalonnage de l'appareil a été réalisé.

D.2. Contrôle qualité 2 : mélange d'échantillons

26 échantillons de plasma collectés à l'hôpital de Strasbourg par le Dr Luc Fornecker sur des patients atteints de diverses lymphopathies ont été mélangés puis redivisés en 26. Les MPCs ont été collectées et déposées sur chacun des gels à partir des 26 nouveaux échantillons. Les bandes de gel issues de ces dépôts ont été traitées suivant la procédure décrite précédemment en figure 5. De la même manière, des peptides iRT ont été ajoutés avant injection sur le couplage nanoLC-MS/MS.

L'objectif est ici double : il s'agit d'évaluer la stabilité instrumentale dans les conditions d'injection des échantillons de la cohorte et la stabilité des échantillons dans le temps.

D.2.1. Stabilité de l'instrument

L'ajout des iRTs dans le mélange nous a permis d'évaluer la variation des temps de rétention au cours des injections successives dans les conditions de gradients utilisées pour les échantillons Protell. Ainsi, les coefficients de variation des temps de rétention sont inférieurs à 1 % pour l'ensemble des 9 peptides détectés et la variation entre les points extrêmes est de 1 min, excepté pour le peptide le plus hydrophile (cf. tableau 19). Ces résultats confirment l'excellente reproductibilité de notre système chromatographique avec un gradient de 90 min.

Séquence du peptide	% CV	Δt (minimum et maximum)
LGGNETQVR	2.2	1.42
AGGSSEPVTGLADK	1.0	1.09
VEATFGVDESANK	0.9	0.98
YILAGVESNK	0.8	1.01
TPVITGAPYYER	0.7	1.09
GDLDAASYAPVR	0.7	1.12
DAVTPADFSEWSK	0.6	1.12
GTFIIDPAAIVR	0.3	0.84
FLLQFGAQGSPLFK	0.3	0.87

Tableau 19

Variation des temps de rétention des peptides iRT au cours des quatre semaines d'analyse

La colonne % CV donne la valeur du coefficient de variation sur les 26 analyses. La colonne Δt indique l'écart entre les valeurs extrêmes.

Un coefficient de variation d'intensité a été calculé pour chacun des peptides iRTs détectés suivant la semaine pendant laquelle ils ont été ajoutés dans les mélanges contrôles (cf. tableau 20). Pour chacune des semaines 2,3 et 4, une valeur aberrante a été enlevée.

Les résultats montrent d'excellents coefficients de variation, inférieurs à 20 % pour la majorité des peptides. Notons que, comme attendu, le peptide le plus hydrophile et le plus hydrophobe ne sont pas très stables. En effet, les peptides hydrophiles sont susceptibles de ne pas être retenus sur la colonne C18 et d'être élués vers la poubelle lors du chargement sur la précolonne. Par ailleurs, il a été observé que le spray était moins stable en début de gradient. Les peptides hydrophobes sont, quant à eux, susceptibles de ne pas être élués de la colonne, ou même de se coller aux parois des différents contenants.

Séquence du peptide	% CV semaine 1	% CV semaine 2	% CV semaine 3	% CV semaine 4
LGGNETQVR	48	16	20	49
AGGSSEPVTGLADK	10	12	11	9
VEATFGVDESANK	10	10	12	14
YILAGVESNK	9	14	14	15
TPVITGAPYYER	10	9	11	4
GDLDAASYAPVR	11	9	11	4
DAVTPADFSEWSK	10	13	18	15
GTFIIDPAAIVR	12	9	23	8
FLLQFGAQGSPLFK	64	50	72	18

Tableau 20

Coefficients de variation d'intensité observés pour chacun des peptides suivant la semaine pendant laquelle ils ont été ajoutés au mélange contrôle

D.2.2. Stabilité de l'échantillon

Afin d'évaluer la stabilité des échantillons, 32 peptides du mélange, avec des temps de rétention répartis sur l'ensemble du gradient, ont été suivis. Les coefficients de variation d'intensité de chacun de ces peptides sont présentés dans le tableau 21. La majorité des coefficients de variation sont inférieurs à 20 %, ce qui montre la stabilité des échantillons au cours des quatre semaines d'analyse, et cela malgré des digestions à des temps différents. Par ailleurs, les résultats obtenus concordent avec ceux qui ont été observés dans la littérature²³³.

Séquence du peptide	Nom de la protéine	% CV
LLIYDASNLETGVPSR	sp P01613 KV121_HUMAN	10
FDSVGEFR	sp P01911 2B1F_HUMAN	10
GHSLQPTGFLS	sp P01911 2B1F_HUMAN	17
YFYNQEEVR	sp P01911 2B1F_HUMAN	14
IAQSDYIPTQQDVLRL	sp P04899 GNAI2_HUMAN	12
LWADHGVQACFGR	sp P04899 GNAI2_HUMAN	17
FQYYEDSSGK	sp P05106 ITB3_HUMAN	18
LAPLAEDVR	sp P06727 APOA4_HUMAN	20
SALSGHLETVILGLLK	sp P07355 ANXA2_HUMAN	21
LTSPLGALR	sp P07359 GP1BA_HUMAN	21
TLPPGLLTPTPK	sp P07359 GP1BA_HUMAN	9

EENFYVDETTVVK	sp P08185 CBG_HUMAN	18
WSAGLTSSQVDLYIPK	sp P08185 CBG_HUMAN	20
DGEDQTQDTELVETRPAGDGTFFQK	sp P10314 1A32_HUMAN	19
FIAVGYVDDTQFVR	sp P10314 1A32_HUMAN	17
VEIANDQGNR	sp P11021 GRP78_HUMAN	11
DIHSLGIGLLICQK	sp P12259 FA5_HUMAN	17
ITAIITQGCK	sp P12259 FA5_HUMAN	19
QWLEIDLLK	sp P12259 FA5_HUMAN	20
AFASPVENPDNYCFCTEK	sp P16671 CD36_HUMAN	17
LQVNLVKPSEK	sp P16671 CD36_HUMAN	19
QVVLEEGTIAFK	sp P16671 CD36_HUMAN	15
SIYAVFESDVNLK	sp P16671 CD36_HUMAN	17
VAIIDTYK	sp P16671 CD36_HUMAN	11
DGEDQTQDTELVETRPAGDR	sp P30460 1B08_HUMAN	19
FLEQELETITIPDLR	sp P55058 PLTP_HUMAN	16
TFHETLDCCGSSTLTALTTSVLK	sp P60033 CD81_HUMAN	24
IIAPPER	sp P68133 ACTS_HUMAN	19
SYELPDGQVITIGNER	sp P68133 ACTS_HUMAN	20
FISVGYVDDTQFVR	sp Q95365 1B38_HUMAN	16
SLITFVNK	sp Q9HBI1 PARVB_HUMAN	13
DITYFIQQLR	sp Q9P1U1 ARP3B_HUMAN	18

Tableau 21

Coefficient de variation de l'aire totale des 32 peptides suivis durant les 4 semaines d'analyse

D.3. Conclusions

Les échantillons de contrôle qualité ont ici rempli deux rôles majeurs :

- vérifier à chaque instant la stabilité du couplage nanoLC-MS/MS et de nos échantillons ;
- attester la qualité de nos acquisitions nanoLC-MS/MS et la reproductibilité de notre processus.

Ainsi, pour chaque échantillon issu de la cohorte Protell, nous sommes en mesure d'affirmer que toute variation observée n'est pas due à la stratégie analytique.

La mise en place de ces contrôles qualité à intervalles réguliers lors de l'analyse de larges cohortes s'inscrit au cœur des préoccupations actuelles de la communauté des protéomistes^{266, 366-369}. Cette dernière cherche en effet à standardiser l'analyse protéomique pour obtenir les résultats les plus fiables possible.

E. Résultats préliminaires et analyses statistiques

L'objectif du travail est ici de trouver des marqueurs prédictifs des rechutes du SNC. Afin d'identifier les protéines potentiellement intéressantes exprimées de manière différentielle, nous avons, dans un premier temps, comparé la cohorte 1 « rechute » à la cohorte 1 « non-rechute » (comparaison initiale). Puis nous avons comparé la cohorte 1 « non-rechute » à la cohorte 2 (comparaison contrôle). En effet, la cohorte 2 est constituée de patients souffrant d'une atteinte primitive du SNC.

La figure 142 montre que le nombre de protéines identifiées dans chaque échantillon est compris entre 200 et 500, hormis pour une analyse qui en compte plus de 800 (FDR < 1 %). Afin d'éviter de futurs biais, cet échantillon a été considéré comme aberrant. Au total, 1450 protéines uniques ont été identifiées.

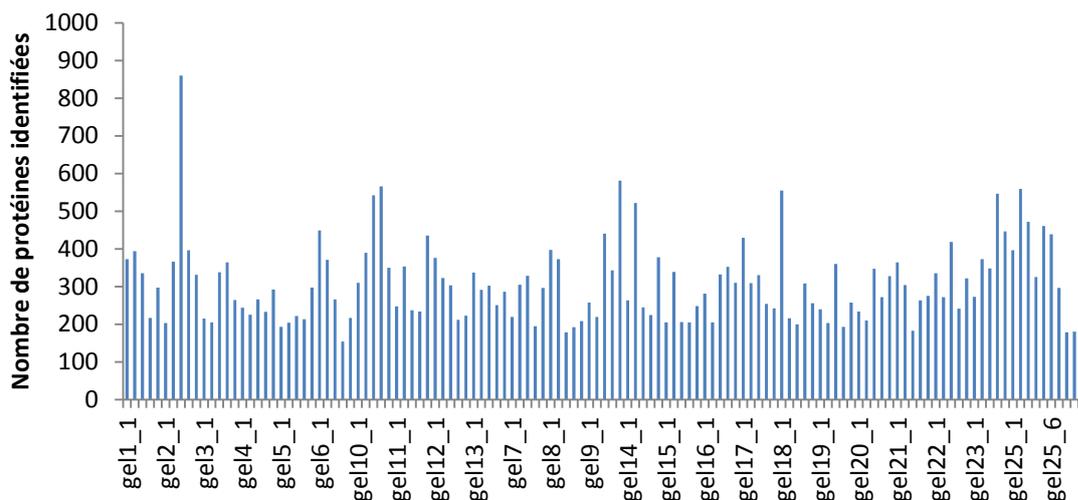


Figure 142

Nombre de protéines identifiées suivant l'échantillon injecté

Au vu du nombre d'échantillons, la mise en évidence de potentielles protéines d'intérêt passe par une analyse statistique. Dans ce but, nous avons voulu évaluer un outil développé par le laboratoire du Dr Olga Vitek (Purdue University) : MSstats (<http://www.msstats.org/>). Ce script disponible dans R et directement associé au logiciel de quantification Skyline, est une suite disposant :

- d'un outil de visualisation des variations potentielles au cours des analyses (détection de problèmes tels que les valeurs aberrantes, les analyses de mauvaise qualité et les erreurs d'intégration) ;

- d'un outil statistique permettant de détecter les protéines significativement exprimées, d'associer une valeur de quantification à chaque entité et, éventuellement, de prédire un modèle d'analyse pour des expériences futures.

L'analyse statistique²⁷³ est ici basée sur le modèle linéaire mixte, qui permet d'exprimer la quantité de chaque protéine via une combinaison linéaire de plusieurs paramètres. Ainsi, ce modèle va permettre de différencier les réplicats biologiques des réplicats techniques, mais aussi de tenir compte des spécificités propres à l'analyse protéomique (différence d'ionisation et interférences). Ensuite, une procédure standard de test statistique est appliquée : le ratio des différences estimées sur l'erreur standard est calculé pour chaque protéine et la valeur obtenue est comparée à une distribution de Student pour obtenir une valeur p .

À noter qu'il est possible de réaliser une comparaison multigroupes en utilisant une valeur p ajustée.

La normalisation des analyses est réalisée à partir de l'hypothèse que la majeure partie des protéines ne doit pas présenter de variations entre les différents échantillons. À ce titre, MSstats propose deux types de normalisation :

- Un ajustement global : ce mode force la distribution des aires totales de chaque peptide autour d'une valeur constante, dans notre cas la médiane.
- Un ajustement par quantile : ce mode non paramétrique force une répartition égale des abondances peptidiques entre les différents quantiles. Il s'agit en fait de faire coïncider les boîtes à moustaches de chaque analyse.

De façon arbitraire, il a été décidé de normaliser par un ajustement global. Les résultats de la comparaison initiale (cohorte 1 « rechute » vs. « non-rechute ») sont présentés en figure 143, sous forme de *volcano plot*, et les résultats de la validation (cohorte 1 « non-rechute » vs. cohorte 2), en figure 144. Les protéines significatives ont été sélectionnées sur la base d'une valeur $p < 0,05$. Il est courant de se baser sur un ratio des quantités dans les deux conditions > 2 . Cependant, dans notre cas, il y a très peu de protéines répondant à ce critère. Le ratio le plus important est de 1,5 et pour les deux comparaisons, il concerne une protéine glycogène phosphorylase impliquée dans la transformation du glycogène en glucose dans le foie et donc a priori peu pertinente, les échantillons étant ici très proches.

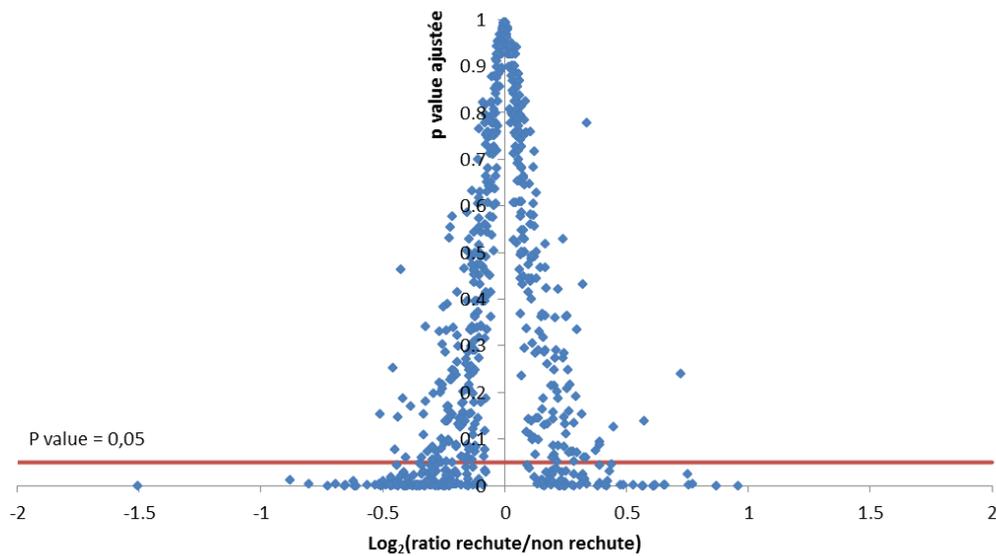


Figure 143

Volcano plot représentant le log en base 2 du ratio de chaque protéine dans la comparaison « rechute » vs. « non-rechute » pour les échantillons de la cohorte 1 en fonction de la valeur p ajustée

Le seuil de significativité des identifications ($p < 0,05$) est indiqué par la droite rouge.

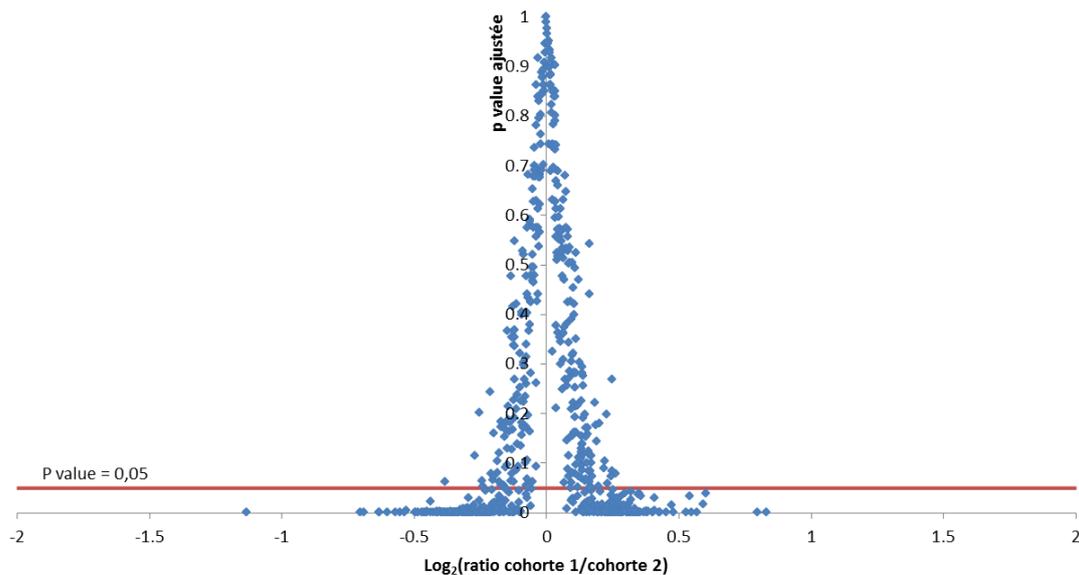


Figure 144

Volcano plot représentant le log en base 2 du ratio de chaque protéine dans la comparaison cohorte 1 « non-rechute » vs. cohorte 2 en fonction de la valeur p ajustée

Le seuil de significativité des identifications ($p < 0,05$) est indiqué par la droite rouge.

Suite à la comparaison initiale, 257 protéines sortent différemment exprimées, et 256 suite à la comparaison contrôle. 168 protéines sont communes aux deux échantillons. Parmi ces protéines différemment exprimées de manière significative, deux éléments importants sont à noter :

- La présence importante de protéines non pertinentes dans le cadre de notre recherche, telles que des protéines plasmatiques (apolipoprotéines, HSA) et des immunoglobulines. Celles-ci ont été publiées maintes fois comme différemment exprimées dans le cadre de la recherche de

biomarqueurs et sont désormais considérées comme « déjà vues »³⁷⁰. Par ailleurs, il ne faut pas oublier que l'objet de l'analyse est le protéome des microparticules et non celui du plasma.

- La présence de quelques protéines potentiellement intéressantes telles que le CD44 ou le CD14, clusters de différenciation non exprimés à la surface des plaquettes.

Il pourrait être instructif de chercher à relier l'ensemble de ces protéines à des voies de régulation pour essayer de détecter des tendances.

Cependant, avant de démarrer une analyse biologique plus poussée, il est plus judicieux d'attendre le traitement de nos données par Caroline Le Gall, statisticienne à l'institut de Mathématiques de Toulouse. En effet, cette dernière a développé un scénario directement adapté à nos échantillons en tenant compte notamment du déséquilibre des cohortes. La cohorte 1 « rechute » contient uniquement 9 échantillons. Par ailleurs, des outils et des tests différents seront utilisés et, plus particulièrement, des analyses de type PLSDA (*partial least square data analyse*) seront mises en place dans le but de chercher une combinaison de protéines permettant de discriminer au mieux les différentes cohortes.

Par la suite, il sera intéressant de confronter les résultats obtenus en utilisant Skyline et MSstats à ceux obtenus par la statisticienne par application d'un scénario spécifiquement adapté à notre analyse.

L'étude finale des protéines mises en évidence se fera en collaboration avec des biologistes et des cliniciens afin de proposer des marqueurs pertinents. En outre, l'étude protéomique menée sur le LCR sera mise en perspective avec notre analyse. Enfin, il sera nécessaire de mettre en place une méthode de validation de potentielles protéines d'intérêt.

F. Conclusion

Ce projet visant à rechercher des biomarqueurs prédictifs des rechutes du système nerveux central dans les cas de lymphome à grandes cellules B a été un défi analytique pour deux raisons principales :

- Les microparticules circulantes sont peu abondantes dans le plasma et majoritairement d'origine plaquettaire.
- Le nombre d'échantillons était très élevé et c'était la première fois qu'une telle cohorte était analysée au laboratoire.

Dans le cadre de ce projet, il a été particulièrement important de développer une méthode d'analyse robuste et d'implémenter des contrôles qualité tout au long des injections afin de surveiller l'évolution de l'instrument et des échantillons. Les résultats obtenus montrent une grande stabilité des échantillons et de l'instrument tout au long des quatre semaines d'injections.

Le nombre d'échantillons est tel qu'une analyse statistique apparaît incontournable. Les résultats préliminaires semblent prometteurs, mais il est important à ce stade de demander conseil à un statisticien afin de mettre en évidence des différences significatives et pertinentes. Par ailleurs, il sera indispensable de travailler en étroite collaboration avec les biologistes et les cliniciens afin d'éviter de

sélectionner des protéines « déjà vues »³⁷⁰. Par la suite, les marqueurs feront l'objet d'une validation par spectrométrie de masse quantitative (SRM) avant d'être étudiés en milieu clinique.

Conclusion générale

Ce travail de thèse s'est articulé autour de deux objectifs principaux :

- le développement de nouvelles méthodologies en analyse protéomique pour la recherche et la proposition de candidats biomarqueurs les plus pertinents possible ;
- l'adaptation de ces méthodologies à trois projets relatifs à la recherche de biomarqueurs.

Développement de nouvelles méthodologies en analyse protéomique pour la recherche de candidats biomarqueurs

L'objectif de cette partie était d'étudier et d'optimiser finement l'ensemble des étapes de la stratégie protéomique globale. L'importance majeure de plusieurs points a pu être démontrée :

- **La réduction de la complexité de l'échantillon par l'enrichissement en protéines membranaires.** Ces dernières sont des cibles privilégiées dans le cadre de la recherche de biomarqueurs par leur position clé à l'interface entre le milieu extracellulaire et intracellulaire ainsi que par leur rôle important dans la communication entre les cellules. Nous nous sommes orientés vers le développement de trois stratégies : la génération de microparticules induites par l'application d'un stress sur la cellule, la formation de *ghosts* membranaires par lyse mécanique et la collecte de microparticules circulant dans le plasma. Ces trois stratégies ont démontré leur efficacité quant à l'enrichissement obtenu. Il est ici important de relever que le choix de la stratégie dépend en premier lieu du type et du nombre d'échantillons.
- **L'optimisation fine des paramètres de chromatographie et de spectrométrie de masse** afin d'identifier le maximum de protéines contenues dans un échantillon. L'importance de l'étude fine des paramètres des systèmes nanoLC-MS/MS a pu être démontrée. D'autre part, la répétabilité de l'identification de chaque système a été évaluée et il a été montré qu'en moyenne, une deuxième injection apporte 12 % d'identifications supplémentaires quel que soit le couplage utilisé. En parallèle, l'apport d'une liste d'exclusion a pu être déterminé. Enfin, un nouveau mode d'acquisition a été étudié : le mode *data independent acquisition* (DIA) de type HDMS^F (Waters). Ce mode ouvre de nouvelles perspectives quant à l'analyse protéomique, telle l'identification de l'ensemble des protéines contenues dans un échantillon. Cependant, si cette technique est attractive, quelques améliorations restent à apporter, notamment en ce qui concerne la qualité des spectres MS/MS.
- **La réflexion sur l'utilisation des moteurs de recherche.** Une étude fine des paramètres de chacun des moteurs de recherche s'est avérée indispensable. De plus, il a été montré que la

combinaison de deux moteurs de recherche augmente la confiance dans les résultats obtenus.

- **Le choix de la méthode de quantification.** Les **méthodes** de quantification **sans marquage** (*label free*) basées soit sur le comptage du nombre de spectres ou sur une extraction des courants d'ions ont été évaluées. Ces deux méthodes se sont avérées justes et particulièrement utiles pour la recherche de biomarqueurs permettant ainsi de détecter des changements plus fins que des protéines « absent/présent ».

Adaptation de ces stratégies à la recherche de biomarqueurs de différents cancers

Les stratégies développées ont pu être adaptées à trois projets de recherche différents. Il est important de noter que les deux premiers projets ont requis une analyse protéomique en profondeur du fait du faible nombre d'échantillons ainsi qu'une validation stricte des résultats pour proposer les biomarqueurs candidats les plus pertinents possible.

Le dernier projet a différé par le grand nombre d'échantillons à analyser (125). Il a été nécessaire de développer une méthode d'analyse rapide associée à des contrôles qualité stricts tout au long des injections.

- **La recherche de biomarqueurs à visées diagnostiques de différents types de lymphomes B.** Le CD180 a ainsi été déterminé comme étant surexprimé dans les lymphomes de la zone marginale à la suite d'une analyse protéomique globale. Ces résultats ont été validés par cytométrie en flux sur plus de 200 patients et le marqueur est actuellement utilisé en routine dans huit hôpitaux de France, au sein d'une triplette avec le CD200 et le CD148 (précédemment mis en évidence au laboratoire). Par ailleurs, la possibilité de mettre en place une méthode SRM pour la quantification et la validation de ces marqueurs a été démontrée. Dans la suite de ce projet, un panel de marqueurs potentiellement intéressants, issus de recherche en cytométrie en flux, en transcriptomique et de la littérature, sera validé par cette méthode, ce qui constituera un premier pas vers la validation clinique.
- **La recherche de biomarqueurs de cellules souches cancéreuses issues de glioblastomes.** Ce projet est un sujet phare de la recherche actuelle contre le cancer dans le sens où ces cellules sont supposées être à l'origine de la croissance et de la résurgence des tumeurs. Cette étude a tout d'abord été orientée vers la recherche de RCPGs (récepteurs couplés à la protéine G), protéines cibles de la plupart des médicaments. L'analyse protéomique a été confrontée à un défi important : les RCPGs sont des protéines très peu abondantes, possédant sept domaines transmembranaires, et donc très peu détectables par spectrométrie de masse. Les résultats sont ici modestes par rapport à ceux obtenus en transcriptomique. La deuxième partie du projet, orientée vers une analyse différentielle de plusieurs types de cellules : cellules souches cancéreuses de glioblastome, U87 (cellules cancéreuses de glioblastomes),

astrocytes humains et cellules souches neurales. Elle a donné l'occasion de mettre en évidence douze marqueurs d'intérêt cours de validation par immunohistochimie.

- **La recherche de biomarqueurs prédictifs des rechutes du système nerveux central dans le cadre de lymphomes diffus à grandes cellules B.** L'objectif était ici de taille dans le sens où il s'agissait d'analyser une cohorte de patients comptant 125 échantillons. Le développement d'une stratégie adaptée ainsi que l'établissement d'un contrôle qualité se sont avérés cruciaux. Nous avons pu démontrer la reproductibilité de l'ensemble du processus analytique ainsi que la stabilité de l'instrument et des échantillons tout au long des quatre semaines d'analyse. Des résultats préliminaires sont ressorties de potentielles protéines d'intérêt qui restent cependant à valider, notamment par la mise en place d'une analyse statistique développée par une spécialiste du domaine.

Perspectives

Mon travail de thèse s'inscrit directement au cœur d'une des préoccupations principales de la communauté protéomiste : comment rendre l'analyse protéomique fiable pour la recherche de biomarqueurs et regagner la confiance perdue ? Cet enjeu devient chaque jour de plus en plus accessible, notamment grâce au développement :

- de nouveaux instruments aux performances sans cesse améliorées et de nouveaux modes d'acquisition permettant d'envisager l'analyse et la quantification complètes d'un protéome donné ;
- de contrôles qualités permettant d'attester de la justesse des analyses protéomiques ;
- de méthodes statistiques adaptées à l'analyse protéomique et à ses spécificités afin de dégager les protéines d'intérêt dans les cohortes d'analyses ;
- de stratégies de validation de potentiels candidats biomarqueurs par des techniques de spectrométrie de masse quantitative telles que la SRM.

L'ensemble de ces points devrait permettre de rassurer les cliniciens quant à la qualité des résultats obtenus par analyse protéomique. L'objectif est désormais de créer des passerelles entre les deux mondes pour mener les biomarqueurs potentiels jusqu'à la validation clinique, et ceci passe par la mise en place de collaborations rassemblant des scientifiques issus de divers domaines : médecine, biologie, transcriptomique, génétique, protéomique, bio-informatique et statistique.

Néanmoins, il est important de ne pas négliger l'aspect le plus important : à quelle question biologique répond notre analyse protéomique ? Orienter ainsi la recherche de biomarqueurs devrait permettre d'éviter des écueils majeurs, notamment sur la sélection des patients et la pertinence des échantillons biologiques.

Enfin, l'étape ultime serait l'implémentation de la spectrométrie de masse à l'hôpital. Pour le moment, cette idée apparaît peu envisageable dans le cadre de la protéomique découverte. Par contre, les méthodes ciblées de type SRM sont aujourd'hui de plus en plus robustes et font l'objet de

mise en place de *guidelines* par la FDA (*Food and Drug Administration*) et le NIH (*National Institutes of Health*). Il est fort probable que, dans quelques années, la validation à grande échelle des biomarqueurs issus de la protéomique découverte pourra être réalisée directement à l'hôpital en routine, ce qui facilitera l'accès aux échantillons. Cette technique pourra également être envisagée pour établir des diagnostics basés sur des protéines ne possédant pas d'anticorps.

Références bibliographiques

1. Bailar, J.C. 3rd Rethinking the war on cancer. *Issues Sci Technol* **4**, 16-21 (1987).
2. El-Deiry, W.S. Are we losing the war on cancer? *Cancer Biol Ther* **14** (2013).
3. Kiberstis, P. & Marshall, E. Cancer crusade at 40. Celebrating an anniversary. Introduction. *Science* **331**, 1539 (2011).
4. Alberts, B. The challenge of cancer. *Science* **331**, 1491 (2011).
5. IARC Globocan. (2012).
6. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
7. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
8. Calapre, L., Gray, E.S. & Ziman, M. Heat stress: a risk factor for skin carcinogenesis. *Cancer Lett* **337**, 35-40 (2013).
9. Lackey, A. & Donington, J.S. Surgical Management of Lung Cancer. *Semin Intervent Radiol* **30**, 133-140 (2013).
10. Motavaf, M., Safari, S., Saffari Jourshari, M. & Alavian, S.M. Hepatitis B virus-induced hepatocellular carcinoma: the role of the virus x protein. *Acta Virol* **57**, 389-396 (2013).
11. Sanna, V., Pala, N. & Sechi, M. Targeted therapy using nanotechnology: focus on cancer. *Int J Nanomed* **9**, 467-483 (2014).
12. Schmid, D., Behrens, G., Jochem, C., Keimling, M. & Leitzmann, M. Physical activity, diabetes, and risk of thyroid cancer: a systematic review and meta-analysis. *Eur J Epidemiol* **28**, 945-958 (2013).
13. Wormann, S.M. & Algul, H. Risk Factors and Therapeutic Targets in Pancreatic Cancer. *Front Oncol* **3**, 282 (2013).
14. Yang, I.A., Holloway, J.W. & Fong, K.M. Genetic susceptibility to lung cancer and co-morbidities. *J Thorac Dis* **5**, S454-S462 (2013).
15. Kranzfelder, M., Buchler, P. & Friess, H. Surgery within multimodal therapy concepts for esophageal squamous cell carcinoma (ESCC): the MRI approach and review of the literature. *Adv Med Sci-Poland* **54**, 158-169 (2009).
16. Dunne-Daly, C.F. Principles of radiotherapy and radiobiology. *Semin Oncol Nurs* **15**, 250-259 (1999).
17. Hawley, L. Principles of radiotherapy. *Br J Hosp Med (Lond)* **74**, C166-169 (2013).
18. Hellevik, T. & Martinez-Zubiaurre, I. Radiotherapy and the Tumor Stroma: The Importance of Dose and Fractionation. *Front Oncol* **4**, 1 (2014).
19. Roopashri, G. & Baig, M. Current advances in radiotherapy of head and neck malignancies. *J Int Oral Health* **5**, 119-123 (2013).
20. Hirsch, J. An anniversary for cancer chemotherapy. *JAMA* **296**, 1518-1520 (2006).
21. Chabner, B.A. & Roberts, T.G., Jr. Timeline: Chemotherapy and the war on cancer. *Nat Rev Cancer* **5**, 65-72 (2005).
22. Paulmurugan, R. et al. Real time dynamic imaging and current targeted therapies in the war on cancer: a new paradigm. *Theranostics* **3**, 437-447 (2013).
23. Sliwkowski, M.X. & Mellman, I. Antibody therapeutics in cancer. *Science* **341**, 1192-1198 (2013).
24. Panowski, S., Bhakta, S., Raab, H., Polakis, P. & Junutula, J.R. Site-specific antibody drug conjugates for cancer therapy. *MAbs* **6**, 34-45 (2014).
25. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108**, 479-485 (2013).
26. Hanahan, D. Rethinking the war on cancer. *Lancet* **383**, 558-563 (2014).
27. Yap, T.A., Gerlinger, M., Futreal, P.A., Pusztai, L. & Swanton, C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med* **4**, 127ps110 (2012).
28. Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111 (2001).
29. Brakenhoff, R.H. Another NOTCH for Cancer. *Science* **333**, 1102-1103 (2011).
30. Erren, T.C., Gross, J.V., Koch, M.S., Erren, M. & Meyer-Rochow, V.B. What Do We Know 40 Years After Nixon Declared the 'War on Cancer'? On the Origin, Prevention and Treatment of Cancer. *J Cancer Educ* **27**, 597-600 (2012).
31. Gosho, M., Nagashima, K. & Sato, Y. Study designs and statistical analyses for biomarker research. *Sensors (Basel)* **12**, 8966-8986 (2012).
32. Hanash, S. & Taguchi, A. TIMELINE The grand challenge to decipher the cancer proteome. *Nature Reviews Cancer* **10**, 652-660 (2010).
33. Pesch, B. et al. Biomarker research with prospective study designs for the early detection of cancer. *Biochim Biophys Acta* (2013).
34. Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C. & Liotta, L.A. Clinical proteomics: Translating bedside promise into bedside reality. *Nat Rev Drug Discov* **1**, 683-695 (2002).
35. Sanger, F., Nicklen, S. & Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *P Natl Acad Sci USA* **74**, 5463-5467 (1977).
36. Morin, R.D. et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* **45**, 81-+ (2008).
37. Wilkins, M.R. et al. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65 (1996).
38. Yadav, S.P. The wholeness in suffix -omics, -omes, and the word om. *J Biomol Tech* **18**, 277 (2007).
39. Novotny, M.V., Soini, H.A. & Mechref, Y. Biochemical individuality reflected in chromatographic, electrophoretic and mass-spectrometric profiles. *J Chromatogr B* **866**, 26-47 (2008).

40. Wood, S.L., Westbrook, J.A. & Brown, J.E. Omic-profiling in breast cancer metastasis to bone: implications for mechanisms, biomarkers and treatment. *Cancer Treat Rev* **40**, 139-152 (2014).
41. Buchen, L. Missing the Mark. *Nature* **471**, 428-432 (2011).
42. Check, E. Proteomics and cancer - Running before we can walk? *Nature* **429**, 496-497 (2004).
43. Diamandis, E.P. Cancer Biomarkers: Can We Turn Recent Failures into Success? *J Natl Cancer I* **102**, 1462-1467 (2010).
44. Ioannidis, J.P. Biomarker failures. *Clin Chem* **59**, 202-204 (2013).
45. Kurzrock, R., Kantarjian, H. & Stewart, D.J. A cancer trial scandal and its regulatory backlash. *Nat Biotechnol* **32**, 27-31 (2014).
46. Moyer, V.A. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **157**, 120-134 (2012).
47. Petricoin, E.F. et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572-577 (2002).
48. Sorace, J.M. & Zhan, M. A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* **4**, 24 (2003).
49. Roberts, N.J. et al. The predictive capacity of personal genome sequencing. *Sci Transl Med* **4**, 133ra158 (2012).
50. Mischak, H. et al. Implementation of proteomic biomarkers: making it work. *European Journal of Clinical Investigation* **42**, 1027-1036 (2012).
51. Ransohoff, D.F. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* **4**, 309-314 (2004).
52. Mitchell, P. Proteomics retrenches. *Nature Biotechnology* **28**, 665-670 (2010).
53. McShane, L.M. et al. Criteria for the use of omics-based predictors in clinical trials. *Nature* **502**, 317-320 (2013).
54. Matafora, V., Bachi, A. & Capasso, G. Genomics and Proteomics: How Long Do We Need to Reach Clinical Results? *Blood Purificat* **36**, 7-11 (2013).
55. Mallick, P. & Kuster, B. Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
56. Leichle, A.B., Dufour, J.F. & Fiedler, G.M. Potentials and pitfalls of clinical peptidomics and metabolomics. *Swiss Med Wkly* **143**, w13801 (2013).
57. Boja, E. et al. Evolution of clinical proteomics and its role in medicine. *J Proteome Res* **10**, 66-84 (2011).
58. Rundle, A., Ahsan, H. & Vineis, P. Better cancer biomarker discovery through better study design. *European Journal of Clinical Investigation* **42**, 1350-1359 (2012).
59. Poste, G. Bring on the biomarkers. *Nature* **469**, 156-157 (2011).
60. Ioannidis, J.P.A. & Khoury, M.J. Improving Validation Practices in "Omics" Research. *Science* **334**, 1230-1232 (2011).
61. Carr, S. et al. The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* **3**, 531-533 (2004).
62. Pepe, M.S., Feng, Z.D., Janes, H., Bossuyt, P.M. & Potter, J.D. Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design. *J Natl Cancer I* **100**, 1432-1438 (2008).
63. Picotti, P., Bodenmiller, B. & Aebersold, R. Proteomics meets the scientific method. *Nat Methods* **10**, 24-27 (2013).
64. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
65. Schagger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* **199**, 223-231 (1991).
66. Wittig, I., Karas, M. & Schagger, H. High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol Cell Proteomics* **6**, 1215-1225 (2007).
67. Righetti, P.G. The Monkey King: A personal view of the long journey towards a proteomic Nirvana. *J Proteomics* (2013).
68. Rabilloud, T. et al. Power and limitations of electrophoretic separations in proteomics strategies. *Mass Spectrom Rev* **28**, 816-843 (2009).
69. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255-262 (1988).
70. Candiano, G. et al. Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327-1333 (2004).
71. Rabilloud, T. Silver staining of 2D electrophoresis gels. *Methods Mol Biol* **893**, 61-73 (2012).
72. Switzer, R.C., 3rd, Merril, C.R. & Shifrin, S. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal Biochem* **98**, 231-237 (1979).
73. Rogowska-Wrzesinska, A., Le Bihan, M.C., Thaysen-Andersen, M. & Roepstorff, P. 2D gels still have a niche in proteomics. *Journal of Proteomics* **88**, 4-13 (2013).
74. Rabilloud, T., Chevallet, M., Luche, S. & Lelong, C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J Proteomics* **73**, 2064-2077 (2010).
75. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021 (1975).
76. Rabilloud, T. The whereabouts of 2D gels in quantitative proteomics. *Methods Mol Biol* **893**, 25-35 (2012).

77. Neverova, I. & Van Eyk, J.E. Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **815**, 51-63 (2005).
78. Xie, F., Smith, R.D. & Shen, Y. Advanced proteomic liquid chromatography. *J Chromatogr A* **1261**, 78-90 (2012).
79. Lavrik, N.V., Taylor, L.T. & Sepaniak, M.J. Nanotechnology and chip level systems for pressure driven liquid chromatography and emerging analytical separation techniques: a review. *Anal Chim Acta* **694**, 6-20 (2011).
80. Kay, R.G., Gregory, B., Grace, P.B. & Pleasance, S. The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum. *Rapid Commun Mass Spectrom* **21**, 2585-2593 (2007).
81. Yin, H. et al. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* **77**, 527-533 (2005).
82. Yin, H. & Killeen, K. The fundamental aspects and applications of Agilent HPLC-Chip. *J Sep Sci* **30**, 1427-1434 (2007).
83. Thomson, J.J. Rays of positive electricity and their application to chemical analysis. (London, Longmans, 1913).
84. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
85. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **2**, 151-153 (1988).
86. Covey, T.R., Thomson, B.A. & Schneider, B.B. Atmospheric pressure ion sources. *Mass Spectrom Rev* **28**, 870-897 (2009).
87. Kebarle, P. & Verkerk, U.H. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev* **28**, 898-917 (2009).
88. Wilm, M.S. & Mann, M. Electrospray and Taylor-Cone Theory, Doles Beam of Macromolecules at Last. *Int J Mass Spectrom* **136**, 167-180 (1994).
89. Taylor, G.I. Disintegration of Water Drops in an Electric Field. *Proc. Royal. Soc. London A* **280**, 383-397 (1964).
90. Rayleigh On the equilibrium of Liquid Conducting Mass Charged with Electricity. *London, Edinburgh, Dublin Phil. Mag. J. Sci.* **14**, 184-186 (1882).
91. Duft, D., Achtzehn, T., Muller, R., Huber, B.A. & Leisner, T. Coulomb fission - Rayleigh jets from levitated microdroplets. *Nature* **421**, 128-128 (2003).
92. dole, M. et al. Molecular Beams of Macroions. *The Journal Of Chemical Physics* **49**, 2240-2249 (1968).
93. iribarne, J.V. & thomson, B.A. On the evaporation of small ions from charged droplets. *The Journal Of Chemical Physics* **64**, 2287-2294 (1976).
94. Stephens, W.E. A Pulsed Mass Spectrometer With Time Dispersion. *Phys. Rev.* **69**, 691 (1946).
95. Cameron, A.E. & Eggers, D.F. An ion "Velocitron". *Rev. Sci. Instrum.* **19**, 605-607 (1948).
96. Wolff, M.M. & Stephens, W.E. A Pulsed Mass Spectrometer with Time Dispersion. *Rev. Sci. Instrum.* **24**, 616-617 (1953).
97. Wiley, W.C. & McLaren, I.H. Time-of-flight Mass Spectrometer with Improved Resolution. *Rev. Sci. Instrum.* **26**, 1150-1157 (1955).
98. Mamyrin, B.A. Laser-Assisted Reflectron Time-of-Flight Mass-Spectrometry. *International Journal of Mass Spectrometry and Ion Processes* **131**, 1-19 (1994).
99. Dodonov, A.F., Chernushevich, I.V. & Laiko, V.V. Electrospray-Ionization on a Reflecting Time-of-Flight Mass-Spectrometer. *Acs Sym Ser* **549**, 108-123 (1994).
100. Karas, M. Time-of-flight mass spectrometer with improved resolution. *J Mass Spectrom* **32**, 1-3 (1997).
101. Wiley, W.C. & McLaren, I.H. Time-of-flight mass spectrometer with improved resolution (Reprinted from Review of Scientific Instruments vol 26, pg 1150, 1995). *J Mass Spectrom* **32**, 4-11 (1997).
102. Paul, W. & Steinwedel, H. Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforschung A* **8**, 448-450 (1953).
103. Campana, J.E. & Jurs, P.C. Computer-Simulation of the Quadrupole Mass Filter. *International Journal of Mass Spectrometry and Ion Processes* **33**, 119-137 (1980).
104. Campana, J.E. Elementary Theory of the Quadrupole Mass Filter. *International Journal of Mass Spectrometry and Ion Processes* **33**, 101-117 (1980).
105. Steinwedel, H. & Paul, H. (Google Patents, 1960).
106. Stafford, G.C., Kelley, P.E., Syka, J.E.P., Reynolds, W.E. & Todd, J.F.J. Recent Improvements in and Analytical Applications of Advanced Ion Trap Technology. *International Journal of Mass Spectrometry and Ion Processes* **60**, 85-98 (1984).
107. Stafford, G.C., Kelley, P.E. & Stephens, D.R. (Google Patents, 1985).
108. March, R.E. Quadrupole Ion Traps. *Mass Spectrom Rev* **28**, 961-989 (2009).
109. Kaiser, R.E., Cooks, R.G., Stafford, G.C., Syka, J.E.P. & Hemberger, P.H. Operation of a Quadrupole Ion Trap Mass-Spectrometer to Achieve High Mass Charge Ratios. *International Journal of Mass Spectrometry and Ion Processes* **106**, 79-115 (1991).
110. Stanton, H.E., Chupka, K., Walder, A.J. & Inghram, M.G. Electron Multipliers in Mass Spectrometry ; Effect of Molecular Structure. *Rev. Sci. Instrum.* **27**, 109 (1956).
111. Busch, K.L. The electron multiplier. *Spectroscopy* **15**, 28-+ (2000).
112. Kurz, E.A. Channel Electron Multipliers. *Am Lab* **11**, 67-& (1979).

113. Laprade, B.N. & Labich, R.J. Microchannel Plate-Based Detectors in Mass-Spectrometry. *Spectroscopy* **9**, 26-& (1994).
114. Gillet, L.C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717 (2012).
115. Panchaud, A. et al. Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal Chem* **81**, 6481-6488 (2009).
116. Plumb, R.S. et al. UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation (vol 20, pg 1989, 2006). *Rapid Commun Mass Sp* **20**, 2234-2234 (2006).
117. McLafferty, F.W. & Bryce, T.A. Metastable-ion characteristics: characterization of isomeric molecules. *Chem. Commun.*, 1215-1217 (1967).
118. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J. & Hunt, D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533 (2004).
119. Zubarev, R.A. et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem* **72**, 563-573 (2000).
120. Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol* **193**, 886-887 (1990).
121. Dongre, A.R., Jones, J.L., Somogyi, A. & Wysocki, V.H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *Journal of the American Chemical Society* **118**, 8365-8374 (1996).
122. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212-217 (2006).
123. Maglott, D., Ostell, J., Pruitt, K.D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research* **39**, D52-D57 (2011).
124. Apweiler, R. et al. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research* **42**, D191-D198 (2014).
125. Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
126. Eng, J.K., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).
127. Craig, R., Cortens, J.P. & Beavis, R.C. Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* **3**, 1234-1242 (2004).
128. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).
129. Geer, L.Y. et al. Open mass spectrometry search algorithm. *J Proteome Res* **3**, 958-964 (2004).
130. Huttlin, E.L., Hegeman, A.D., Harms, A.C. & Sussman, M.R. Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. *J Proteome Res* **6**, 392-398 (2007).
131. Elias, J.E. & Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
132. Megger, D.A., Bracht, T., Meyer, H.E. & Sitek, B. Label-free quantification in clinical proteomics. *Biochim Biophys Acta* **1834**, 1581-1590 (2013).
133. Lundgren, D.H., Hwang, S.I., Wu, L. & Han, D.K. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics* **7**, 39-53 (2010).
134. Chelius, D. & Bondarenko, P.V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research* **1**, 317-323 (2002).
135. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1**, 376-386 (2002).
136. Hebert, A.S. et al. Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat Methods* **10**, 332-334 (2013).
137. Gouw, J.W., Tops, B.B., Mortensen, P., Heck, A.J. & Krijgsveld, J. Optimizing identification and quantitation of 15N-labeled proteins in comparative proteomics. *Anal Chem* **80**, 7796-7803 (2008).
138. Kline, K.G. & Sussman, M.R. Protein Quantitation Using Isotope-Assisted Mass Spectrometry. *Annu Rev Biophys* **39**, 291-308 (2010).
139. Zhang, Y.Y. et al. A MS data search method for improved N-15-labeled protein identification. *Proteomics* **9**, 4265-4270 (2009).
140. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994-999 (1999).
141. Unlu, M., Morgan, M.E. & Minden, J.S. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-2077 (1997).
142. Ross, P.L. et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* **3**, 1154-1169 (2004).
143. Thompson, A. et al. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).

144. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* **10**, 1785-1793 (2011).
145. Christensen, E., Evans, K.R., Menard, C., Pintilie, M. & Bristow, R.G. Practical approaches to proteomic biomarkers within prostate cancer radiotherapy trials. *Cancer Metast Rev* **27**, 375-385 (2008).
146. Gravett, M.G. et al. Proteomic analysis of cervical-vaginal fluid: Identification of novel biomarkers for detection of intra-amniotic infection. *Journal of Proteome Research* **6**, 89-96 (2007).
147. Theodorescu, D. et al. Discovery and validation of new protein biomarkers for 4 urothelial cancer: a prospective analysis. *Lancet Oncology* **7**, 230-240 (2006).
148. Anderson, N.L. & Anderson, N.G. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-867 (2002).
149. Ramstrom, M. et al. Depletion of high-abundant proteins in body fluids prior to liquid chromatography fourier transform ion cyclotron resonance mass spectrometry. *J Proteome Res* **4**, 410-416 (2005).
150. Boschetti, E. & Righetti, P.G. The art of observing rare protein species in proteomes with peptide ligand libraries. *Proteomics* **9**, 1492-1510 (2009).
151. Blomme, B., Van Steenkiste, C., Callewaert, N. & Van Vlierberghe, H. Alteration of protein glycosylation in liver diseases. *J Hepatol* **50**, 592-603 (2009).
152. Dube, D.H. & Bertozzi, C.R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **4**, 477-488 (2005).
153. Dwek, R.A. Glycobiology: Toward Understanding the Function of Sugars. *Chem Rev* **96**, 683-720 (1996).
154. Nawarak, J., Phutrakul, S. & Chen, S.T. Analysis of lectin-bound glycoproteins in snake venom from the Elapidae and Viperidae families. *J Proteome Res* **3**, 383-392 (2004).
155. Dreger, M. Subcellular proteomics. *Mass Spectrom Rev* **22**, 27-56 (2003).
156. Drissi, R., Dubois, M.L. & Boisvert, F.M. Proteomics methods for subcellular proteome analysis. *FEBS J* **280**, 5626-5634 (2013).
157. Falcon-Perez, J.M., Lu, S.C. & Mato, J.M. Sub-proteome approach to the knowledge of liver. *Proteomics Clin Appl* **4**, 407-415 (2010).
158. Lescuyer, P., Chevallet, M., Luche, S. & Rabilloud, T. Organelle proteomics. *Curr Protoc Protein Sci* **Chapter 24**, Unit 24 22 (2006).
159. Fischer, F., Wolters, D., Rogner, M. & Poetsch, A. Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection. *Mol Cell Proteomics* **5**, 444-453 (2006).
160. Speers, A.E. & Wu, C.C. Proteomics of integral membrane proteins--theory and application. *Chem Rev* **107**, 3687-3714 (2007).
161. Gyorgy, B. et al. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. *Cellular and Molecular Life Sciences* **68**, 2667-2688 (2011).
162. Jang, J.H. & Hanash, S. Profiling of the cell surface proteome. *Proteomics* **3**, 1947-1954 (2003).
163. Vuckovic, D., Dagley, L.F., Purcell, A.W. & Emili, A. Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. *Proteomics* **13**, 404-423 (2013).
164. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7**, 1029-1038 (1998).
165. Macher, B.A. & Yen, T.Y. Proteins at membrane surfaces - a review of approaches. *Molecular Biosystems* **3**, 705-713 (2007).
166. Josic, D. & Clifton, J.G. Mammalian plasma membrane proteomics. *Proteomics* **7**, 3010-3029 (2007).
167. Giusti, I., D'Ascenzo, S. & Dolo, V. Microvesicles as potential ovarian cancer biomarkers. *Biomed Res Int* **2013**, 703048 (2013).
168. Barteneva, N.S. et al. Circulating microparticles: square the circle. *BMC Cell Biol* **14**, 23 (2013).
169. Burger, D. et al. Microparticles: biomarkers and beyond. *Clin Sci (Lond)* **124**, 423-441 (2013).
170. Hugel, B., Martinez, M.C., Kunzelmann, C. & Freyssinet, J.M. Membrane microparticles: two sides of the coin. *Physiology (Bethesda)* **20**, 22-27 (2005).
171. Miguet, L. et al. Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* **8**, 3346-3354 (2009).
172. Miguet, L. et al. Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* **6**, 153-171 (2006).
173. Zola, H. et al. CD molecules 2006 - Human cell differentiation molecules. *Journal of Immunological Methods* **319**, 1-5 (2007).
174. Mause, S.F. & Weber, C. Microparticles: protagonists of a novel communication network for intercellular information exchange. *Circ Res* **107**, 1047-1057 (2010).
175. Diamant, M., Tushuizen, M.E., Sturk, A. & Nieuwland, R. Cellular microparticles: new players in the field of vascular disease? *Eur J Clin Invest* **34**, 392-401 (2004).
176. Amabile, N. et al. Circulating endothelial microparticles are associated with vascular dysfunction in patients with end-stage renal failure. *J Am Soc Nephrol* **16**, 3381-3388 (2005).

177. Boulanger, C.M., Amabile, N. & Tedgui, A. Circulating microparticles: a potential prognostic marker for atherosclerotic vascular disease. *Hypertension* **48**, 180-186 (2006).
178. VanWijk, M.J., VanBavel, E., Sturk, A. & Nieuwland, R. Microparticles in cardiovascular diseases. *Cardiovasc Res* **59**, 277-287 (2003).
179. Budaj, M. et al. Microparticles: a component of various diseases. *Pol Arch Med Wewn* **122 Suppl 1**, 24-29 (2012).
180. Jaiswal, R., Luk, F., Dalla, P.V., Grau, G.E. & Bebawy, M. Breast cancer-derived microparticles display tissue selectivity in the transfer of resistance proteins to cells. *PLoS One* **8**, e61515 (2013).
181. Park, J.O. et al. Identification and characterization of proteins isolated from microvesicles derived from human lung cancer pleural effusions. *Proteomics* **13**, 2125-2134 (2013).
182. Shao, H. et al. Protein typing of circulating microvesicles allows real-time monitoring of glioblastoma therapy. *Nat Med* **18**, 1835-1840 (2012).
183. Martins, V.R., Dias, M.S. & Hainaut, P. Tumor-cell-derived microvesicles as carriers of molecular information in cancer. *Curr Opin Oncol* **25**, 66-75 (2013).
184. D'Souza-Schorey, C. & Clancy, J.W. Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. *Genes Dev* **26**, 1287-1299 (2012).
185. Ostergaard, O. et al. Quantitative proteome profiling of normal human circulating microparticles. *J Proteome Res* **11**, 2154-2163 (2012).
186. Audran, E., PhD 54-58 (Université de Strasbourg, 2012).
187. Diamandis, E.P. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **360**, 170; author reply 170-171 (2002).
188. Elwood, M. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **360**, 170; author reply 170-171 (2002).
189. Pearl, D.C. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **360**, 169-170; author reply 170-161 (2002).
190. Rockhill, B. Proteomic patterns in serum and identification of ovarian cancer. *Lancet* **360**, 169; author reply 170-161 (2002).
191. Bell, A.W. et al. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Methods* **6**, 423-430 (2009).
192. Tabb, D.L. et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **9**, 761-776 (2010).
193. Yin, H.F. & Killeen, K. The fundamental aspects and applications of Agilent HPLC-Chip. *Journal of Separation Science* **30**, 1427-1434 (2007).
194. Yin, N.F. et al. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. *Anal Chem* **77**, 527-533 (2005).
195. Guan, S.H. & Marshall, A.G. Equilibrium Space-Charge Distribution in a Quadrupole Ion-Trap. *J Am Soc Mass Spectr* **5**, 64-71 (1994).
196. Li, G.Z., Guan, S.H. & Marshall, A.G. Comparison of equilibrium ion density distribution and trapping force in Penning, Paul, and combined ion traps. *J Am Soc Mass Spectr* **9**, 473-481 (1998).
197. Yang, C.G. & Bier, M.E. Investigation of the rapid scan on an electrospray ion trap mass spectrometer. *Anal Chem* **77**, 1663-1671 (2005).
198. Hovasse, A. PhD 65-70 (Université de Strasbourg, 2010).
199. Cox, K.A., Clevon, C.D. & Cooks, R.G. Mass Shifts and Local Space-Charge Effects Observed in the Quadrupole Ion-Trap at Higher Resolution. *International Journal of Mass Spectrometry and Ion Processes* **144**, 47-65 (1995).
200. Hodge, K., Ten Have, S., Hutton, L. & Lamond, A.I. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics* **88**, 92-103 (2013).
201. McQueen, P. et al. Information-dependent LC-MS/MS acquisition with exclusion lists potentially generated on-the-fly: Case study using a whole cell digest of *Clostridium thermocellum*. *Proteomics* **12**, 1160-1169 (2012).
202. Muntel, J., Hecker, M. & Becher, D. An exclusion list based label-free proteome quantification approach using an LTQ Orbitrap. *Rapid Commun Mass Sp* **26**, 701-709 (2012).
203. Voisin, S.N. et al. Identification of Novel Molecular Targets for Endometrial Cancer Using a Drill-Down LC-MS/MS Approach with iTRAQ. *PLoS One* **6** (2011).
204. Zhang, Z. Automated precursor ion exclusion during LC-MS/MS data acquisition for optimal ion identification. *J Am Soc Mass Spectrom* **23**, 1400-1407 (2012).
205. Purvine, S., Eppel, J.T., Yi, E.C. & Goodlett, D.R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics* **3**, 847-850 (2003).
206. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods* **1**, 39-45 (2004).
207. Frank, A.M. et al. Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. *Nat Methods* **8**, 587-591 (2011).
208. Yen, C.Y., Houel, S., Ahn, N.G. & Old, W.M. Spectrum-to-Spectrum Searching Using a Proteome-wide Spectral Library. *Molecular & Cellular Proteomics* **10** (2011).
209. Geromanos, S.J. et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683-1695 (2009).

210. Hoaglund-Hyzer, C.S. & Clemmer, D.E. Ion trap/ion mobility/quadrupole/time of flight mass spectrometry for peptide mixture analysis. *Anal Chem* **73**, 177-184 (2001).
211. Guevremont, R. High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. *J Chromatogr A* **1058**, 3-19 (2004).
212. Kolakowski, B.M. & Mester, Z. Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **132**, 842-864 (2007).
213. Zimmermann, S., Abel, N., Baether, W. & Barth, S. An ion-focusing aspiration condenser as an ion mobility spectrometer. *Sensor Actuat B-Chem* **125**, 428-434 (2007).
214. Shvartsburg, A.A. & Smith, R.D. Fundamentals of traveling wave ion mobility spectrometry. *Anal Chem* **80**, 9689-9699 (2008).
215. Michaelevski, I., Kirshenbaum, N. & Sharon, M. T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J Vis Exp* (2010).
216. Giles, K., Williams, J.P. & Campuzano, I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun Mass Spectrom* **25**, 1559-1566 (2011).
217. Uetrecht, C., Rose, R.J., van Duijn, E., Lorenzen, K. & Heck, A.J. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* **39**, 1633-1655 (2010).
218. Li, G.Z. et al. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696-1719 (2009).
219. Geromanos, S.J. et al. Simulating and validating proteomics data and search results. *Proteomics* **11**, 1189-1211 (2011).
220. Bond, N.J., Shliha, P.V., Lilley, K.S. & Gatto, L. Improving Qualitative and Quantitative Performance for MSE-based Label-free Proteomics. *Journal of Proteome Research* **12**, 2340-2353 (2013).
221. Shliha, P.V., Bond, N.J., Gatto, L. & Lilley, K.S. Effects of Traveling Wave Ion Mobility Separation on Data Independent Acquisition in Proteomics Studies. *Journal of Proteome Research* **12**, 2323-2339 (2013).
222. Distler, U. et al. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nature Methods* **11**, 167-+ (2014).
223. Benjamin, A.M. et al. A flexible statistical model for alignment of label-free proteomics data - incorporating ion mobility and product ion information. *BMC Bioinformatics* **14** (2013).
224. Muntel, J. et al. Comprehensive absolute quantification of the cytosolic proteome of *Bacillus subtilis* by multiplexed LC/MS (LC/MSE). *Mol Cell Proteomics* (2014).
225. Creasy, D.M. & Cottrell, J.S. Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2**, 1426-1434 (2002).
226. Akhtar, M.N., Southey, B.R., Andren, P.E., Sweedler, J.V. & Rodriguez-Zas, S.L. Evaluation of Database Search Programs for Accurate Detection of Neuropeptides in Tandem Mass Spectrometry Experiments. *J Proteome Res* (2012).
227. Alves, G., Wu, W.W., Wang, G., Shen, R.F. & Yu, Y.K. Enhancing peptide identification confidence by combining search methods. *J Proteome Res* **7**, 3102-3113 (2008).
228. Chang, K.Y. & Muddiman, D.C. Identification of alternative splice variants in *Aspergillus flavus* through comparison of multiple tandem MS search algorithms. *BMC Genomics* **12**, 358 (2011).
229. Lietzen, N., Natri, L., Nevalainen, O.S., Salmi, J. & Nyman, T.A. Compid: a new software tool to integrate and compare MS/MS based protein identification results from Mascot and Paragon. *J Proteome Res* **9**, 6795-6800 (2010).
230. Bantscheff, M., Lemeer, S., Savitski, M.M. & Kuster, B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404**, 939-965 (2012).
231. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031 (2007).
232. Neilson, K.A. et al. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11**, 535-553 (2011).
233. Tu, C., Li, J., Sheng, Q., Zhang, M. & Qu, J. Systematic Assessment of Survey Scan and MS2-based Abundance Strategies for Label-Free Quantitative Proteomics Using High-resolution MS Data. *J Proteome Res* (2014).
234. Liu, H., Sadygov, R.G. & Yates, J.R., 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201 (2004).
235. Choi, H., Glatter, T., Gstaiger, M. & Nesvizhskii, A.I. SAINT-MS1: Protein-Protein Interaction Scoring Using Label-free Intensity Data in Affinity Purification-Mass Spectrometry Experiments. *Journal of Proteome Research* **11**, 2619-2624 (2012).
236. Gokce, E., Shuford, C.M., Franck, W.L., Dean, R.A. & Muddiman, D.C. Evaluation of Normalization Methods on GeLC-MS/MS Label-Free Spectral Counting Data to Correct for Variation during Proteomic Workflows. *J Am Soc Mass Spectr* **22**, 2199-2208 (2011).
237. Schmidt, C. et al. Mass spectrometry-based relative quantification of proteins in precatalytic and catalytically active spliceosomes by metabolic labeling (SILAC), chemical labeling (iTRAQ), and label-free spectral count. *RNA* **20**, 406-420 (2014).

238. Uen, Y.H. et al. Comparative proteomics, network analysis and post-translational modification identification reveal differential profiles of plasma Con A-bound glycoprotein biomarkers in gastric cancer. *J Proteomics* **83**, 197-213 (2013).
239. Wu, Q. et al. Improved accuracy for label-free absolute quantification of proteome by combining the Absolute Protein EXpression profiling algorithm and summed tandem mass spectrometric total ion current. *Analyst* **139**, 138-146 (2014).
240. Chen, Y.Y. et al. Refining comparative proteomics by spectral counting to account for shared peptides and multiple search engines. *Anal Bioanal Chem* **404**, 1115-1125 (2012).
241. Ishihama, Y. et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**, 1265-1272 (2005).
242. Shinoda, K., Tomita, M. & Ishihama, Y. emPAI Calc--for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* **26**, 576-577 (2010).
243. Braisted, J.C. et al. The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **9**, 529 (2008).
244. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**, 117-124 (2007).
245. Paoletti, A.C. et al. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *P Natl Acad Sci USA* **103**, 18928-18933 (2006).
246. Griffin, N.M. et al. Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis. *Nat Biotechnol* **28**, 83-89 (2010).
247. Zhang, B. et al. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* **5**, 2909-2918 (2006).
248. Zhou, J.Y. et al. Improved LC-MS/MS spectral counting statistics by recovering low-scoring spectra matched to confidently identified peptide sequences. *J Proteome Res* **9**, 5698-5704 (2010).
249. Kannaste, O. et al. Cross-correlation of spectral count ranking to validate quantitative proteome measurements. *J Proteome Res* (2014).
250. Choi, H., Fermin, D. & Nesvizhskii, A.I. Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics* **7**, 2373-2385 (2008).
251. Sokal, R. & Rohlf, F., Edn. Third. (New York; 1995).
252. Colinge, J., Chiappe, D., Lagache, S., Moniatte, M. & Bougueleret, L. Differential proteomics via probabilistic peptide identification scores. *Anal Chem* **77**, 596-606 (2005).
253. Pham, T.V. & Jimenez, C.R. An accurate paired sample test for count data. *Bioinformatics* **28**, 1596-1602 (2012).
254. Pham, T.V., Piersma, S.R., Warmoes, M. & Jimenez, C.R. On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* **26**, 363-369 (2010).
255. Searle, B.C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265-1269 (2010).
256. Hoehenwarter, W. & Wienkoop, S. Spectral counting robust on high mass accuracy mass spectrometers. *Rapid Commun Mass Spectrom* **24**, 3609-3614 (2010).
257. Bleijerveld, O.B. et al. Deep Proteome Profiling of Circulating Granulocytes Reveals Bactericidal/Permeability-Increasing Protein as a Biomarker for Severe Atherosclerotic Coronary Stenosis. *Journal of Proteome Research* **11**, 5235-5244 (2012).
258. Cole, L.M. et al. MALDI-MSI and label-free LC-ESI-MS/MS shotgun proteomics to investigate protein induction in a murine fibrosarcoma model following treatment with a vascular disrupting agent. *Proteomics* **14**, 890-903 (2014).
259. Han, N.Y. et al. Label-free quantification for discovering novel biomarkers in the diagnosis and assessment of disease activity in inflammatory bowel disease. *J Dig Dis* **14**, 166-174 (2013).
260. Higgs, R.E., Knierman, M.D., Gelfanova, V., Butler, J.P. & Hale, J.E. Label-free LC-MS method for the identification of biomarkers. *Methods Mol Biol* **428**, 209-230 (2008).
261. Kikuchi, T. et al. In-depth proteomic analysis of nonsmall cell lung cancer to discover molecular targets and candidate biomarkers. *Mol Cell Proteomics* **11**, 916-932 (2012).
262. Mutsaers, C.A., Lamont, D.J., Hunter, G., Wishart, T.M. & Gillingwater, T.H. Label-free proteomics identifies Calreticulin and GRP75/Mortalin as peripherally accessible protein biomarkers for spinal muscular atrophy. *Genome Med* **5** (2013).
263. Nie, S. et al. Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. *J Proteome Res* **13**, 1873-1884 (2014).
264. Perrin, R.J. et al. Quantitative label-free proteomics for discovery of biomarkers in cerebrospinal fluid: assessment of technical and inter-individual variation. *PLoS One* **8**, e64314 (2013).
265. Nahnsen, S., Bielow, C., Reinert, K. & Kohlbacher, O. Tools for label-free peptide quantification. *Mol Cell Proteomics* **12**, 549-556 (2013).
266. Sandin, M., Teleman, J., Malmstrom, J. & Levander, F. Data processing methods and quality control strategies for label-free LC-MS protein quantification. *Biochim Biophys Acta* **1844**, 29-41 (2014).
267. Rudnick, P.A., Wang, X., Yan, X., Sedransk, N. & Stein, S.E. Improved Normalization of Systematic Biases Affecting Ion Current Measurements in Label-free Proteomics Data. *Mol Cell Proteomics* (2014).

268. Callister, S.J. et al. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res* **5**, 277-286 (2006).
269. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
270. Schilling, B. et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol Cell Proteomics* **11**, 202-214 (2012).
271. Sturm, M. et al. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**, 163 (2008).
272. Matzke, M.M. et al. A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics* **13**, 493-503 (2013).
273. Clough, T., Thaminy, S., Ragg, S., Aebersold, R. & Vitek, O. Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. *BMC Bioinformatics* **13 Suppl 16**, S6 (2012).
274. Escher, C. et al. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111-1121 (2012).
275. Hebert, A.S. et al. The one hour yeast proteome. *Mol Cell Proteomics* **13**, 339-347 (2014).
276. Swerdlow, S.H., Campo, E., Harris, N.L., Jaffe, E.S., Pileri, S.A., Stein, H., Thiele, J., Vardiman, J.W. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition, Vol. 2. (IARC, 2008).
277. Flowers, C.R. & Armitage, J.O. A Decade of Progress in Lymphoma: Advances and Continuing Challenges. *Cl Lymph Myelom Leuk* **10**, 414-423 (2010).
278. Harris, M. Monoclonal antibodies as therapeutic agents for cancer. *Lancet Oncology* **5**, 292-302 (2004).
279. Zinzani, P.L. The many faces of marginal zone lymphoma. *Hematol-Am Soc Hemat*, 426-432 (2012).
280. Van den Brand, M. & van Krieken, J.H.J.M. Recognizing nodal marginal zone lymphoma: recent advances and pitfalls. A systematic review. *Haematologica* **98**, 1003-1013 (2013).
281. Craig, F.E. & Foon, K.A. Flow cytometric immunophenotyping for hematologic neoplasms. *Blood* **111**, 3941-3967 (2008).
282. Porakishvili, N. et al. Differential expression of CD180 and IgM by B-cell chronic lymphocytic leukaemia cells using mutated and unmutated immunoglobulin VH genes. *Brit J Haematol* **131**, 313-319 (2005).
283. Miura, Y. et al. RP105 is associated with MD-1 and transmits an activation signal in human B cells. *Blood* **92**, 2815-2822 (1998).
284. Chaplin, J.W., Kasahara, S., Clark, E.A. & Ledbetter, J.A. Anti-CD180 (RP105) Activates B Cells To Rapidly Produce Polyclonal Ig via a T Cell and MyD88-Independent Pathway. *Journal of Immunology* **187**, 4199-4209 (2011).
285. Miguet, L. et al. Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas. *Leukemia* (2013).
286. Sandes, A.F. et al. CD200 has an important role in the differential diagnosis of mature B-cell neoplasms by multiparameter flow cytometry. *Cytometry B Clin Cytom* **86**, 98-105 (2014).
287. Matutes, E. Contribution of immunophenotype in the diagnosis and classification of haemopoietic malignancies. *J Clin Pathol* **48**, 194-197 (1995).
288. Baty, J.D. & Robinson, P.R. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. *Biomed Mass Spectrom* **4**, 36-41 (1977).
289. Zakett, D., Flynn, R.G.A. & Cooks, R.G. Chlorine Isotope Effects in Mass-Spectrometry by Multiple Reaction Monitoring. *J Phys Chem-Us* **82**, 2359-2362 (1978).
290. Yost, R.A. & Enke, C.G. Triple Quadrupole Mass-Spectrometry for Direct Mixture Analysis and Structure Elucidation. *Anal Chem* **51**, 1251-& (1979).
291. Barr, J.R. et al. Isotope dilution mass spectrometric quantification of specific proteins: Model application with apolipoprotein A-I. *Clinical Chemistry* **42**, 1676-1682 (1996).
292. Percy, A.J., Chambers, A.G., Yang, J. & Borchers, C.H. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. *Proteomics* **13**, 2202-2215 (2013).
293. Pan, S. et al. Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. *J Proteome Res* **11**, 1937-1948 (2012).
294. Keshishian, H. et al. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics* **8**, 2339-2349 (2009).
295. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555-566 (2012).
296. Gallien, S., Duriez, E. & Domon, B. Selected reaction monitoring applied to proteomics. *J Mass Spectrom* **46**, 298-312 (2011).
297. Angel, T.E. et al. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews* **41**, 3912-3928 (2012).
298. Plumel, M.I. et al. Mass spectrometry-based sequencing and SRM-based quantitation of two novel vitellogenin isoforms in the leatherback sea turtle (*Dermochelys coriacea*). *J Proteome Res* **12**, 4122-4135 (2013).
299. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **100**, 6940-6945 (2003).

300. Gillette, M.A. & Carr, S.A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat Methods* **10**, 28-34 (2013).
301. Beynon, R.J., Doherty, M.K., Pratt, J.M. & Gaskell, S.J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* **2**, 587-589 (2005).
302. Simpson, D.M. & Beynon, R.J. QconCATs: design and expression of concatenated protein standards for multiplexed protein quantification. *Anal Bioanal Chem* **404**, 977-989 (2012).
303. Brun, V. et al. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics* **6**, 2139-2149 (2007).
304. Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**, 265-275 (1951).
305. Mallick, P. et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* **25**, 125-131 (2007).
306. Picotti, P. et al. A database of mass spectrometric assays for the yeast proteome. *Nat Methods* **5**, 913-914 (2008).
307. Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378-379 (1998).
308. Kall, L., Krogh, A. & Sonnhammer, E.L. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027-1036 (2004).
309. Sonnhammer, E.L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182 (1998).
310. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* **37**, W465-468 (2009).
311. Boeuf, A., PhD 140-150 (Université de Strasbourg, 2013).
312. Percy, A.J. et al. Method and platform standardization in MRM-based quantitative plasma proteomics. *J Proteomics* **95**, 66-76 (2013).
313. Whiteaker, J.R. et al. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat Biotechnol* **29**, 625-634 (2011).
314. Smith, R.D. Mass spectrometry in biomarker applications: from untargeted discovery to targeted verification, and implications for platform convergence and clinical application. *Clin Chem* **58**, 528-530 (2012).
315. Dumas-Duport, C., Beuvon, F., Varlet, P. & Fallet-Bianco, C. Gliomas : WHO and Sainte-Anne Hospital classification. *Ann Pathol* **20**, 413-428 (2000).
316. Dolecek, T.A., Propp, J.M., Stroup, N.E. & Kruchko, C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. *Neuro Oncol* **14 Suppl 5**, v1-49 (2012).
317. Yin, A.A., Cheng, J.X., Zhang, X. & Liu, B.L. The treatment of glioblastomas: a systematic update on clinical Phase III trials. *Crit Rev Oncol Hematol* **87**, 265-282 (2013).
318. Sarkaria, J.N. et al. Combination of temsirolimus (CCI-779) with chemoradiation in newly diagnosed glioblastoma multiforme (GBM) (NCCTG trial N027D) is associated with increased infectious risks. *Clin Cancer Res* **16**, 5573-5580 (2010).
319. Stupp, R., Hegi, M. & Weller, M. Neuro-oncology, a decade of temozolomide and beyond. *Expert Rev Anticancer Ther* **10**, 1675-1677 (2010).
320. Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* **352**, 987-996 (2005).
321. Chen, J. et al. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488**, 522-+ (2012).
322. Clevers, H. The cancer stem cell: premises, promises and challenges. *Nat Med* **17**, 313-319 (2011).
323. Dalerba, P., Cho, R.W. & Clarke, M.F. Cancer stem cells: Models and concepts. *Annu Rev Med* **58**, 267-284 (2007).
324. Persano, L., Rampazzo, E., Basso, G. & Viola, G. Glioblastoma cancer stem cells: Role of the microenvironment and therapeutic targeting. *Biochemical Pharmacology* **85**, 612-622 (2013).
325. Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nature Reviews Cancer* **5**, 275-284 (2005).
326. Visvader, J.E. Cells of origin in cancer. *Nature* **469**, 314-322 (2011).
327. Holland, E.C. et al. Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice. *Nat Genet* **25**, 55-57 (2000).
328. Bruggeman, S.W. et al. Bmi1 controls tumor development in an Ink4a/Arf-independent manner in a mouse model for glioma. *Cancer Cell* **12**, 328-341 (2007).
329. Bachoo, R.M. et al. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer Cell* **1**, 269-277 (2002).
330. Scatena, R., Bottoni, P., Pontoglio, A. & Giordina, B. The proteomics of cancer stem cells: potential clinical applications for innovative research in oncology. *Proteomics Clin Appl* **5**, 590-602 (2011).
331. Chinnaiyan, P. et al. The prognostic value of nestin expression in newly diagnosed glioblastoma: Report from the Radiation Therapy Oncology Group. *Radiat Oncol* **3** (2008).
332. Christensen, K., Schroder, H.D. & Kristensen, B.W. CD133 identifies perivascular niches in grade II-IV astrocytomas. *J Neuro-Oncol* **90**, 157-170 (2008).
333. Dahlrot, R.H., Hermansen, S.K., Hansen, S. & Kristensen, B.W. What is the clinical value of cancer stem cell markers in gliomas? *Int J Clin Exp Pathol* **6**, 334-348 (2013).

334. Ehrmann, J., Kolar, Z. & Mokry, J. Nestin as a diagnostic and prognostic marker: immunohistochemical analysis of its expression in different tumours. *Journal of Clinical Pathology* **58**, 222-223 (2005).
335. Griguer, C.E. et al. CD133 is a marker of bioenergetic stress in human glioma. *PLoS One* **3**, e3655 (2008).
336. Patru, C. et al. CD133, CD15/SSEA-1, CD34 or side populations do not resume tumor-initiating properties of long-term cultured cancer stem cells from human malignant glioma-neuronal tumors. *BMC Cancer* **10**, 66 (2010).
337. Phi, J.H. et al. Sox2 expression in brain tumors: A reflection of the neuroglial differentiation pathway. *American Journal of Surgical Pathology* **32**, 103-112 (2008).
338. Son, M.J., Woolard, K., Nam, D.H., Lee, J. & Fine, H.A. SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. *Cell Stem Cell* **4**, 440-452 (2009).
339. Thon, N. et al. Presence of pluripotent CD133(+) cells correlates with malignancy of gliomas. *Mol Cell Neurosci* **43**, 51-59 (2010).
340. Jacoby, E., Bouhelal, R., Gerspacher, M. & Seuwen, K. The 7 TM G-protein-coupled receptor target family. *ChemMedChem* **1**, 761-782 (2006).
341. Audigier, Y., Picault, F.X., Chaves-Almagro, C. & Masri, B. G Protein-Coupled Receptors in cancer: biochemical interactions and drug design. *Prog Mol Biol Transl Sci* **115**, 143-173 (2013).
342. Garland, S.L. Are GPCRs still a source of new targets? *J Biomol Screen* **18**, 947-966 (2013).
343. Desiere, F. et al. The PeptideAtlas project. *Nucleic Acids Res* **34**, D655-658 (2006).
344. Desiere, F. et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biology* **6** (2005).
345. Deutsch, E.W. The PeptideAtlas Project. *Methods Mol Biol* **604**, 285-296 (2010).
346. Farrah, T. et al. State of the human proteome in 2013 as viewed through PeptideAtlas: comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven Human Proteome Project. *J Proteome Res* **13**, 60-75 (2014).
347. Krokhin, O.V. & Spicer, V. Predicting peptide retention times for proteomics. *Curr Protoc Bioinformatics* **Chapter 13**, Unit 13 14 (2010).
348. Spicer, V., Grigoryan, M., Gotfrid, A., Standing, K.G. & Krokhin, O.V. Predicting retention time shifts associated with variation of the gradient slope in peptide RP-HPLC. *Anal Chem* **82**, 9678-9685 (2010).
349. Tripet, B. et al. Requirements for prediction of peptide retention time in reversed-phase high-performance liquid chromatography: hydrophilicity/hydrophobicity of side-chains at the N- and C-termini of peptides are dramatically affected by the end-groups and location. *J Chromatogr A* **1141**, 212-225 (2007).
350. Spicer, V. et al. Sequence-specific retention calculator. A family of peptide retention time prediction algorithms in reversed-phase HPLC: applicability to various chromatographic conditions and columns. *Anal Chem* **79**, 8762-8768 (2007).
351. Krokhin, O.V. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal Chem* **78**, 7785-7795 (2006).
352. Krokhin, O.V. et al. Use of peptide retention time prediction for protein identification by off-line reversed-phase HPLC-MALDI MS/MS. *Anal Chem* **78**, 6265-6269 (2006).
353. Vogel, C. & Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* **13**, 227-232 (2012).
354. Ginestier, C. et al. Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *American Journal of Pathology* **161**, 1223-1233 (2002).
355. Eisen, D. et al. G protein-coupled receptor quantification using peptide group-specific enrichment combined with internal peptide standard reporter calibration. *Journal of Proteomics* **90**, 85-95 (2013).
356. Aloni-Grinstein, R., Shetzer, Y., Kaufman, T. & Rotter, V. p53: The barrier to cancer stem cell formation. *FEBS Lett* (2014).
357. Muller, P.A. & Vousden, K.H. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. *Cancer Cell* **25**, 304-317 (2014).
358. Harris, C.C. P53 - at the Crossroads of Molecular Carcinogenesis and Risk Assessment. *Science* **262**, 1980-1981 (1993).
359. Martelli, M. et al. Diffuse large B-cell lymphoma. *Crit Rev Oncol Hematol* **87**, 146-171 (2013).
360. Cai, Q. et al. Accelerated therapeutic progress in diffuse large B cell lymphoma. *Ann Hematol* **93**, 541-556 (2014).
361. Feugier, P. et al. Incidence and risk factors for central nervous system occurrence in elderly patients with diffuse large-B-cell lymphoma: influence of rituximab. *Ann Oncol* **15**, 129-133 (2004).
362. Haioun, C. et al. Incidence and risk factors of central nervous system relapse in histologically aggressive non-Hodgkin's lymphoma uniformly treated and receiving intrathecal central nervous system prophylaxis: a GELA study on 974 patients. Groupe d'Etudes des Lymphomes de l'Adulte. *Ann Oncol* **11**, 685-690 (2000).
363. Laskin, J.J., Savage, K.J., Voss, N., Gascoyne, R.D. & Connors, J.M. Primary paranasal sinus lymphoma: natural history and improved outcome with central nervous system chemoprophylaxis. *Leuk Lymphoma* **46**, 1721-1727 (2005).
364. Van Besien, K. et al. Risk factors, treatment, and outcome of central nervous system recurrence in adults with intermediate-grade and immunoblastic lymphoma. *Blood* **91**, 1178-1184 (1998).
365. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med* **329**, 987-994 (1993).

366. Bramwell, D. An introduction to statistical process control in research proteomics. *J Proteomics* **95**, 3-21 (2013).
367. Gallien, S., Bourmaud, A. & Domon, B. A Simple Protocol To Routinely Assess the Uniformity of Proteomics Analyses. *J Proteome Res* (2014).
368. Gu, Q. & Yu, L.R. Proteomics quality and standard: from a regulatory perspective. *J Proteomics* **96**, 353-359 (2014).
369. Tabb, D.L. Quality assessment for clinical proteomics. *Clin Biochem* **46**, 411-420 (2013).
370. Petrak, J. et al. Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins. *Proteomics* **8**, 1744-1749 (2008).

Partie expérimentale

Préparation des échantillons et enrichissement en protéines membranaires

A. Préparation des microparticules induites

Projet : recherche de biomarqueurs à visée diagnostique pour la caractérisation des lymphomes B

200 × 10⁶ lymphocytes B sont isolés à partir de prélèvements sanguins (5 mL par patient dans un tube EDTA) issus de trois patients atteints de trois pathologies différentes : maladie de Waldenström, lymphome du manteau et lymphome de la zone marginale. Ils sont traités avec de l'actinomycine D (0,5 µg/mL, Sigma) pendant 18 h dans un milieu sans sérum X-VIVO 15 medium (Cambrex). Une centrifugation à 750 g pendant 15 min puis à 1500 g pendant 5 min est réalisée afin d'éliminer les débris cellulaires importants. Le surnageant contenant les microparticules est récupéré et centrifugé pendant 45 min à 15 000 g. Les microparticules contenues dans le culot sont lavées dans 15 mL d'une solution de tampon salin (HBSS) et centrifugé 45 min à 15 000 g. Les cellules sont alors reprises dans 1 mL de PBS et conservées à - 80 °C jusqu'à l'analyse protéomique.

B. Collecte des microparticules circulantes

Projet : recherche de biomarqueurs prédictifs des rechutes du système nerveux central

5 mL de sang sont prélevés sur tube citraté. Une première centrifugation à 1 500 g pendant 15 min est effectuée afin de récupérer le plasma. Puis, une deuxième centrifugation à 18 000 g pendant 2 min permet d'obtenir un plasma pauvre en plaquettes. Le culot de microparticules est obtenu par une troisième centrifugation à 18 000 g pendant 45 min. Le culot est ensuite lavé deux fois au PBS à 18 000 g pendant 45 min, repris dans 15 µL de tampon de Laemmli (10 mM Tris pH 8, 1 mM EDTA, 5 % β-mercaptoéthanol, 5 % SDS, 10 % glycérol et 0,1 % bleu de bromophénol) et stocké à - 20 °C.

C. Préparation des *ghosts* membranaires

Projet : recherche de biomarqueurs de cellules souches cancéreuses issues de glioblastomes

Les cellules sont lavées avec du PBS et du tampon de lyse (10 mM HEPES pH 7.4, 2 mM EGTA, cocktail d'inhibiteur de protéases). La lyse des cellules s'effectue de façon mécanique dans 10 mL de tampon par des allers-retours de la suspension à travers une aiguille 22G montée sur une seringue de 25 mL toutes les 5-10 min jusqu'à la lyse de plus de 90 % des cellules. Les débris cellulaires sont éliminés par centrifugation (10 min, 3 000 rpm) et le surnageant est centrifugé 1 h 30 à 14 000 rpm. Les *ghosts* membranaires sont re-suspendus dans 300 µL de tampon de lyse, dosés par Bradford et conservés à - 20 °C.

Extraction des protéines et dépôt sur gel 1D SDS PAGE

Pour chacun des projets présentés, un gel aux caractéristiques différentes a été réalisé.

A. Recherche de biomarqueurs diagnostiques des lymphomes des cellules B

A.1. Fractionnement des microparticules induites

Pour chaque échantillon, 200 µg de microparticules induites sont repris dans 80 µL de tampon de Laemmli (5 % SDS, 5 % β-mercaptoéthanol, 1 mM EDTA, 10 % glycérol, 10 mM Tris pH 8, 0,01 % bleu de bromophénol) et chauffés à 100 °C pendant 3 min. Les solutions sont ensuite déposées sur un gel polyacrylamide 8-15 % de dimension 20 × 20 cm. Les gels migrent pendant 5 h à 30 mA. La coloration des gels est réalisée au bleu de Coomassie. Des bandes sont coupées tous les 2 mm à l'emporte-pièce.

A.2. Réalisation d'un gel 1 bande (analyses ciblées)

Les lymphocytes sont séparés du sang total par centrifugation différentielle sur un Ficoll à 700 g pendant 25 min à température ambiante. Les cellules sont ensuite centrifugées à 1000 tr/min avant reprise dans du tampon de Laemmli sans bleu de façon à obtenir une concentration de 86 000 cellules/µL. La quantité de protéines totale est évaluée par dosage RCDC à titre indicatif. 17,5 µL, soit 1,5 million de cellules, sont déposés sur petit gel 1D SDS PAGE (6 cm). Il est ici important de noter que le gel de concentration de 4 % occupe la moitié du volume par rapport au gel de séparation de 10 %. Les protéines migrent pendant 45 min à 50 V dans le gel de concentration. La coloration est réalisée au bleu de Coomassie pendant 20 min. La fine bande contenant des protéines est découpée ainsi que la bande supérieure (chemin de migration des protéines).

B. Recherche de biomarqueurs de cellules souches cancéreuses

Pour chaque échantillon, 150 µg de *ghosts* membranaires sont repris dans 80 µL de tampon de Laemmli et chauffés à 100 °C pendant 3 min. Les solutions sont ensuite déposées sur un gel polyacrylamide 10-20 % de dimension 20 × 20 cm. Les cellules migrent pendant 1 h avec un ampérage de 10 mA, puis 14 h à 12 mA et 1 h 30 à 15 mA. La coloration des gels est réalisée au bleu de Coomassie. Des bandes sont coupées tous les 2 mm à l'emporte-pièce.

C. Recherche de biomarqueurs prédictifs des rechutes cérébrales dans le cadre de lymphomes diffus à larges cellules B

Les microparticules circulantes sont reprises dans 15 µL de tampon de Laemmli. Un protocole identique à la partie A.2. est appliqué.

Réduction/alkylation et digestion des protéines

Les bandes de gel sont décolorées trois fois par ajout de 100 µL d'une solution 25/75 NH₄HCO₃ 25 mM/ACN. Puis, une déshydratation est réalisée par ajout de 50 µL d'acétonitrile. Les protéines sont ensuite réduites avec 50 µL de DTT à 10 mM et laissées 30 min à 60 °C et 30 min à température ambiante. 50 µL d'iodoacétamide à 55 mM sont ajoutés pour l'alkylation des cystéines.

La réaction est réalisée pendant 20 min à température ambiante dans le noir. Ensuite, 100 µL d'acétonitrile sont ajoutés et l'ensemble des solutions est retiré. Les bouts de gels sont lavés trois fois avec un mélange 50/50 NH₄HCO₃ 25 mM/ACN. Enfin, les bandes de gels sont déshydratées trois fois avec 50 µL d'acétonitrile.

À ce stade, les plaques peuvent être stockées à - 20 °C.

Les protéines sont ensuite digérées par de la trypsine à 12,5 ng/µL, selon le rapport enzyme/protéine 1/50, pendant la nuit à 37 °C.

Les peptides sont extraits par 40 µL d'une solution à 40/60/0,1 H₂O/ACN/HCOOH pendant 1 h sous agitation avant transfert dans une nouvelle plaque. Une deuxième extraction est réalisée avec 40 µL d'une solution d'acétonitrile. L'excès d'acétonitrile est réduit par évaporation avec une centrifugeuse sous vide (Speedvac® de Thermo Savant SPD111V).

Analyse nanoLC-MS/MS et identification dans les banques de données

A. Couplage nanoAcquity – Synapt G1 (système chromatographique Waters couplé à un Q-TOF Waters)

Projet : recherche de biomarqueurs à visée diagnostique pour la caractérisation des lymphomes B

Le système de chromatographie nanoAcquity est couplé à un spectromètre de masse hybride quadripôle – temps de vol (Q-TOF). La précolonne est une phase inverse C18 de dimension 180 µm × 20 mm nanoAcquity™ UPLC™ de porosité 5 µm (Waters). La colonne analytique est une phase inverse C18 de dimension 75 µm × 200 mm nanoAcquity™ UPLC™, de porosité 1,7 µm (Waters). Le gradient est réalisé par un mélange de deux solvants A et B. Le solvant A est une solution à 0,1 % en acide formique dans de l'eau et le solvant B est une solution à 0,1 % en acide formique dans de l'acétonitrile. Le gradient va de 1 à 40 % de B en 30 min. Le débit est fixé à 0,300 µL/min.

L'étalonnage interne est réalisé grâce à un système de *lock mass*. Celui-ci consiste en une infusion continue d'un composé de référence, ici une solution d'acide formique à 0,1 % et de GFP (glucofibrinopeptide) à 200 fmol à un débit de 0,200 µL/min. La tension du capillaire de l'électrospray est de 3 kV, celle du cône d'échantillonnage, à 35 V et celle du cône d'extraction, à 40 V. La température de la source est de 90 °C.

L'acquisition est réalisée à l'aide du logiciel Masslynx 4.1 (Waters). Chaque balayage MS dure 1 s pour une gamme de masse de m/z de 250-2000 Da et est espacé du suivant par 0,1 s. Un spectre MS/MS est réalisé pour les trois ions les plus intenses sur une gamme de masse de 50 à 2000 Da. L'acquisition dure 0,7 s par spectre MS/MS. Un ion sélectionné est exclu de la sélection pour la fragmentation pendant les 10 secondes qui suivent. Le gaz de collision utilisé pour la fragmentation est de l'argon.

Les fichiers obtenus après les analyses sont convertis en *peak list* (fichier *.pkl) à l'aide du logiciel ProteinLynx Browser 2.3 (Waters). Ces fichiers sont ensuite chargés dans les logiciels de recherche peptidique – Mascot, Matrix Science et Omssa, (*Open Mass Spectrometry Search Algorithm*) – avec les critères suivants : banque humaine *Target decoy* Swiss-Prot (28/01/2010, 40 556 séquences), deux modifications variables (carbamidométhylation des cystéines et oxydation des méthionines), 1 coupure manquée autorisée, 25 ppm d'erreur sur la MS et 0,07 Da sur la MS/MS.

Les fichiers *.mgf de Mascot et les fichiers *.omx de Omssa sont chargés dans Scaffold 3.00.03. (Proteome Software Inc.). Des filtres de sélection sont appliqués afin d'obtenir un taux de faux positifs (FDR) inférieur à 1 %.

B. Couplage Ultimate 3000 – TSQ Vantage (système chromatographique Thermo Fisher Scientific couplé à un QQQ Waters)

Projet : recherche de biomarqueurs à visée diagnostique pour la caractérisation des lymphomes B

Le mélange peptidique est chargé pendant 3 min à 50 µL/min sur une précolonne d'enrichissement Agilent Zorbax 300SB-C18 (40 nL, 5 µm) et séparé sur une colonne analytique Agilent Zorbax 300SB-C18 (150 × 0,3 mm, 3,5 µm) à 5 µL/min. Le solvant A contient 98 % d'eau, 2 % d'acétonitrile et 0,1 % d'acide formique et le B, 98 % d'acétonitrile, 2 % d'eau et 0,1 % d'acide formique. Le gradient comporte trois pentes : 8 à 15 % de B de 3 à 15 min, puis 15 à 22 % B de 15 à 65 min et de 22 à 40 % de B de 60 à 90 min. Le temps de cycle est de 2,4 s. 273 transitions sont suivies en mode *scheduled*. La tension appliquée sur le capillaire est de 3000 V et la température de la source de 300 °C. La résolution sur Q1 est de 0,7 Da. L'acquisition des données est réalisée avec le logiciel Xcalibur 2.1 (Thermo Fisher Scientific).

C. Couplage nanoChip – AmaZon (système chromatographique Agilent couplé à la trappe ionique de Bruker)

Projet : recherche de biomarqueurs de cellules souches cancéreuses issues de glioblastomes

L'ensemble des conditions de fonctionnement de ce couplage est décrit dans la publication des résultats présentée dans la Partie III – chapitre 2.

D. Couplage nanoAcquity – Impact (système chromatographique Waters couplé au Q-TOF de Bruker)

Projet : recherche de biomarqueurs prédictifs des rechutes du système nerveux central

Le système chromatographique utilisé est identique à celui de la partie A. excepté pour la longueur de la colonne qui est de 25 cm. Le gradient va de 1 à 33 % de B en 90 min. Le débit est fixé à 0,450 µL/min.

L'étalonnage interne est réalisé grâce à un système de *lock mass*. Celui-ci consiste en une infusion continue d'un composé de référence de masse $m/z = 922,09$. La tension du capillaire de l'électrospray est de 1,3 kV. La température de la source est de 150 °C. Le temps de cycle est de 3 s.

Un scan MS dure 0,5 s. Le nombre et la durée des scans MS/MS sont établis suivant l'intensité des peptides en MS. L'appareil fonctionne à 4 Hz si l'intensité est de 3000 cps et à 25 Hz si l'intensité est de 25 000 cps. Les ions monochargés sont exclus de la sélection. L'exclusion dynamique est de 1 min. L'acquisition est réalisée à l'aide du logiciel HyStar v.3.2.

Les *peak lists* sont générées à l'aide du logiciel Compass Data Analysis v.4.2 (Bruker Daltonics). Ces fichiers sont ensuite chargés dans les logiciels de recherche peptidique (Mascot, Matrix Science) avec les critères suivants : banque humaine *Target decoy* Swiss-Prot (31/12/2013, 40 744 séquences), trois modifications variables (carbamidométhylation des cystéines, oxydation des méthionines et propionamidation des cystéines), 1 coupure manquée autorisée, 15 ppm d'erreur sur la MS et 0,05 Da sur la MS/MS.

Les fichiers *.mgf de Mascot sont chargés dans Scaffold 3.00.03. (Proteome Software Inc.) Des filtres de sélection sont appliqués afin d'obtenir un taux de FDR < 1 %.

E. Couplage nanoAcquity – Synapt G2 (système chromatographique Waters couplé à un Q-TOF Waters)

Projet : mise en place du couplage LC-HDMS^F au laboratoire

Le système chromatographique utilisé est identique à celui de la partie A. excepté pour la longueur de la colonne qui est de 25 cm. Le gradient est adapté suivant la complexité du mélange. Le débit est fixé à 0,450 µL/min.

L'étalonnage interne est réalisé grâce à un système de *lock mass*. Celui-ci consiste en une infusion continue d'un composé de référence, ici une solution d'acide formique à 0,1 % et de GFP (glucofibrinopeptide) à 200 fmol, à un débit de 0,200 µL/min. La tension du capillaire de l'électrospray est de 3 kV, celle du cône d'échantillonnage à 30 V et celle du cône d'extraction à 40 V. La température de la source est de 80 °C.

L'acquisition est réalisée à l'aide du logiciel Masslynx 4.1 (Waters). Une alternance est réalisée entre les scans basse et haute énergie. Chaque spectre est acquis pendant 1 s sur une gamme de masse de 200 à 1700 m/z. L'azote pour la cellule de mobilité et l'hélium de la cellule d'hélium sont délivrés à un débit de 90 mL/min.

Les protéines sont identifiées par recherche dans les banques de données E.Coli, fournies par la société Waters, et dans le logiciel ProteinLynx Browser 2.5.2 with Expression 2 avec les critères suivants : une modification fixe (carbamidométhylation des cystéines), une modification variable (oxydation des méthionines), 1 fragment MS/MS minimum par peptide et 3 fragments MS/MS minimum pour une protéine. Les filtres sont appliqués automatiquement par le logiciel de manière à obtenir un taux de FDR < 1 %.

Annexe 1

Liste des clusters de différenciation

Clusters identifiés dans les échantillons de MZL, LLC et LPL
avec au moins deux peptides dans la Partie III – chapitre 1.

CDs	Masse moléculaire (Da)	Nombre total de peptides			Nombre de spectres MS/MS			Couverture de séquence		
		LPL	MCL	MZL	LPL	MCL	MZL	LPL	MCL	MZL
CD98	67978		2	9		7	67		4.29%	20.5%
CD233	101778	10	2		80	2		17.0%	3.29%	
CD79A	25021			3			31			31.0%
CD79B	26031			5			98			31.0%
CD72	40203			5			14			25.9%
CD22	95329		12	13		25	56		13.1%	20.4%
CD19	61112	5		7	18		62	15.5%		19.6%
CD20	33061	6	7	9	40	84	215	22.2%	29.0%	29.3%
CD277	57660	1		3	2		15	2.53%		14.8%
CD222	274355	4		13	7		22	1.77%		7.19%
CD196	42478		1	1		2	16		3.21%	3.21%
CD197	42857	1		1	13		7	2.91%		2.91%
CD180	74165		2	9		6	29		4.54%	30.3%
CD44	81520	7	4	8	37	22	110	9.43%	6.87%	10.6%
CD48	27665	2		5	10		67	9.47%		24.7%
CD81	25792	4	4	4	6	15	21	28.4%	28.4%	28.4%
CD82	29608	2	2	6	8	2	108	9.36%	9.36%	28.1%
CD9	25399	2	1	3	4	4	13	15.4%	4.39%	20.6%
CD99	18830	1	1		3	4		7.57%	7.57%	
CD55	41382			6			18			13.1%
CD35	223642	1	19	28	4	49	103	0.589%	12.2%	18.0%
CD21	112897	4		3	13		3	5.13%		3.87%
CD184	39729	2			4			8.52%		
CD39	57948	2		8	3		30	4.71%		20.2%
CD74	33498	1	1	1	4	2	5	4.73%	4.73%	4.73%

CD49B	129280	13			37			16.9%	
CD49D	114884		12			40		16.2%	
CD49F	126618	13			27			17.0%	
CD41	113359	32	12		335	20		40.6%	15.1%
CD11A	128754	9	23		16	157		12.0%	26.4%
CD11B	127161	11	28		26	139		12.2%	32.6%
CD29	88397	13	16		71	70		19.5%	23.6%
CD18	84764	9	6	31	20	10	106	11.6%	8.06% 53.2%
CD161	87040	30	1		203	2		58.2%	1.90%
CD54	57807	3	2	4	9	4	7	7.52%	5.64% 9.59%
CD102	30636	2	3		8	11		9.09%	9.09%
CD50	59521	6	10		87	93		15.7%	19.0%
CD225	13947	1			3			12.8%	
CD321	32565	6			16			19.4%	
CD37	31687	4	4	6	28	8	90	19.2%	19.2% 29.2%
CD85J	70802		2			7			4.62%
CD32	34026	2	3		9	19		9.68%	8.71%
CD107A	44865	3	1	1	12	3	8	8.63%	3.84% 3.84%
CD31	82518	16	4	19	92	10	132	30.1%	7.99% 34.7%
CD36	53036	8	1		37	5		21.2%	3.60%
CD42C	21701	3			14			15.5%	
CD42A	19028	4			5			27.1%	
CD232	175727		12			41			8.74%
CD361	48650	2	2		2	9		5.80%	5.80%
CD62P	90815	6			19			7.83%	
CD45	147238	45	42	60	525	182	1735	35.4%	29.8% 45.2%
CD148	145924	17	32		57	202		19.0%	33.8%
CD352	37328	1	1		1	2		2.41%	6.93%
CD298	31496	1	1	4	8	4	16	5.02%	5.02% 12.5%

CD6	71782	3	5	5	19	6.29%	12.0%
CD1C	37636		2		9		5.41%
CD4	51093	1	1	1	2	2.84%	2.84%
CD5	54559	6	11	24	75	16.6%	29.9%
CD71	84856		9		20		16.2%

Annexe 2

Communication par affiche

61^e congrès ASMS (American Society of Mass Spectrometry)
2013, Minneapolis

Lennon S., Miguet L., Carapito C., Fornecker L., Van Dorsselaer A., Mauvieux L., Cianférani S.

CD180 : a new Marginal Zone B-cell Lymphoma biomarker discovered by plasma membrane microparticle proteomic analysis

Sarah Lennon ⁽¹⁾, Laurent Miguet ⁽²⁾, Christine Carapito ⁽¹⁾, Luc Fornecker ⁽¹⁾, Alain Van Dorsselaer ⁽¹⁾, Laurent Mauvieux ⁽²⁾, Sarah Cianférani-Sanglier ⁽¹⁾

(1) Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, DSA, CNRS UMR7178, Uds, 67087 Strasbourg, France
 (2) Institut d'Hématologie et d'Immunologie, Faculté de Médecine, Université de Strasbourg Cedex, France



Overview

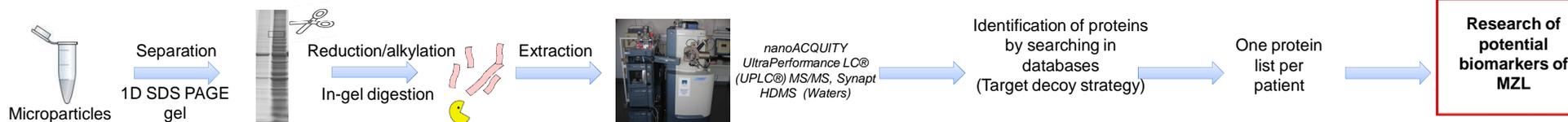
Aim: Find new plasma membrane biomarker by proteomics
Method: Semi quantitative (spectral counting) approach based on nanoLC-MS/MS
Results: Optimisation of the proteomic workflow allows to suggest CD180 as candidate biomarker

Introduction

Marginal zone B-cell lymphomas (MZL) diagnosis can be of great difficulty, particularly because there is no immunological-positive marker to distinguish MZL from chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL) and lymphoplasmocytic lymphoma (LPL). To search for new potential MZL biomarkers, we used large scale proteomics. In this strategy, the major difficulties concern the sample complexity and the large dynamic range of protein abundances. Working on subproteomes is of utmost interest. In our case, membrane proteins appeared relevant as they can easily be used for immunological validation during routine diagnostic in hospitals. Therefore, **we developed a 1D-gel proteomic-based strategy on plasma membrane microparticles (MPs) to search for new potential MZL biomarkers.**

Samples were collected from the university hospitals of Strasbourg : 1 LPL, 1 MZL and 1 MCL **Method**

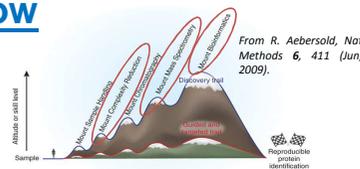
L. Miguet et al., *J Proteome Res* 8, 3346 (Jul, 2009)



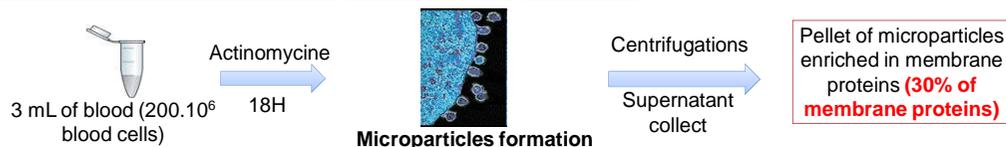
Optimisation of the proteomic workflow

Main drawback of proteomic analysis

Non-repeatability and non-reproducibility of the analysis

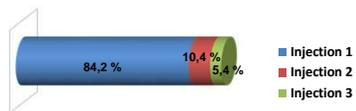


1- Enrichment in membrane proteins: MPs generation



2-Repetability studies: 3 injections of sample 1 (LPL)

1515 proteins : 70% of the proteins are identified 3 times



Identification of almost 95 % of proteins at the end of the second injection

2 injections = significant evaluation of the proteome

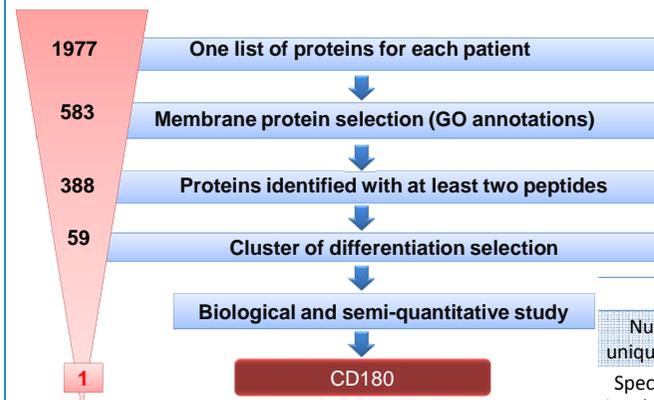
3-Evaluation of two search engines: Mascot and Omssa ⁽⁴⁾ (comparison shown for sample 1, LPL)



- More identifications with Omssa, differences rely on scoring
- Only common identification will be considered

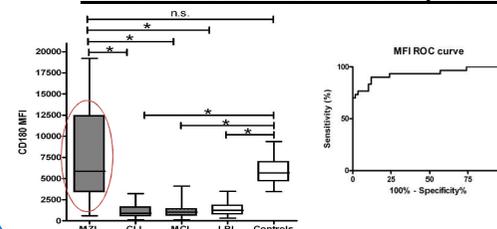
Results

Selection of the most promising biomarker candidate



	LPL	MCL	MZL
Number of unique peptides	0	2	9
Spectral count (MS/MS spectra)	0	14	29

Validation of the CD180 on 120 patients



→ Normal expression of CD180 in MZL (compare to other lymphoma)
 → Sensitivity of 87% and specificity of 88% for MFI>2122

L. Miguet et al., *Leukemia*, (Jan 10, 2013).

Conclusion

- The proteomics based nanoLC-MS/MS strategy was optimized to minimize the non-repeatability and non reproducibility of the analysis.
- Our optimisation allowed to identify around 2200 proteins. After a biological and semi-quantitative study, we were able to propose the CD180 as a potential biomarker of MZL.
- The CD180 has currently been pre-validated by flow cytometry on 120 patients.

Annexe 3

Publication dans le journal *Leukemia*

Miguet L.*, Lennon S.*, Baseggio L., Traverse-Glehen A., Berger F., Perrusson N., Chenard M. P., Galois A. C., Eischen A., Mayeur-Rousse C., Maar A., Fornecker L., Herbrecht R., Felman P., Van Dorselaer A., Carapito C., Cianferani S., Mauvieux L.

Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas

Leukemia, 2013

*Ces auteurs ont contribué de manière égale à ce travail.

REFERENCES

- Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J *et al.* A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood* 1995; **85**: 1435–1441.
- Chaplin T, Bernard O, Beverloo HB, Saha V, Hagemeyer A, Berger R *et al.* The t(10;11) translocation in acute myeloid leukemia (M5) consistently fuses the leucine zipper motif of AF10 onto the HRX gene. *Blood* 1995; **86**: 2073–2076.
- Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J *et al.* New insights to the MLL recombinome of acute leukemias. *Leukemia* 2009; **23**: 1490–1499.
- Morerio C, Rapella A, Tassano E, Rosanda C, Panarello C. MLL-MLLT10 fusion gene in pediatric acute megakaryoblastic leukemia. *Leuk Res* 2005; **29**: 1223–1226.
- Beverloo HB, Le Coniat M, Wijsman J, Lillington DM, Bernard O, de Klein A *et al.* Breakpoint heterogeneity in t(10;11) translocation in AML-M4/M5 resulting in fusion of AF10 and MLL is resolved by fluorescent *in situ* hybridization analysis. *Cancer Res* 1995; **55**: 4220–4224.
- DiMartino JF, Ayton PM, Chen EH, Naftzger CC, Young BD, Cleary ML. The AF10 leucine zipper is required for leukemic transformation of myeloid progenitors by MLL-AF10. *Blood* 2002; **99**: 3780–3785.
- Van Limbergen H, Poppe B, Janssens A, De Bock R, De Paepe A, Noens L *et al.* Molecular cytogenetic analysis of 10;11 rearrangements in acute myeloid leukemia. *Leukemia* 2002; **16**: 344–351.
- Chen J, Feng W, Jiang J, Deng Y, Huen MS. Ring finger protein RNF169 antagonises the ubiquitin-dependent signaling cascade at sites of dna damage. *J Biol Chem* 2012; **287**: 27715–27722.
- Poulsen M, Lukas C, Lukas J, Bekker-Jensen S, Mailand N. Human RNF169 is a negative regulator of the ubiquitin-dependent response to DNA double-strand breaks. *J Cell Biol* 2012; **197**: 189–199.

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Cell-surface expression of the TLR homolog CD180 in circulating cells from splenic and nodal marginal zone lymphomas

Leukemia (2013) **27**, 1748–1750; doi:10.1038/leu.2013.3

Marginal zone B-cell lymphomas (MZL) are classically characterized by an indolent proliferation of CD5[−], CD23[−], CD43[−], CCND1[−] and CD10[−] B cells that strongly express CD19, CD20, CD22, IgM and IgD, even if CD5⁺ cases have been characterized.¹ The World Health Organization (WHO)² classifies marginal B-cell lymphomas into splenic marginal zone lymphoma (SMZL), nodal MZL and extranodal MZL of mucosa-associated lymphoid tissue (MALT).² Among the lymphomas developed from the spleen, the splenic diffuse red pulp lymphoma with villous lymphocytes (SDRPLs) has been described as provisional entity close to but distinct from SMZL.^{2,3}

For small to intermediate-sized circulating abnormal lymphocytes, MZL diagnosis can be of great difficulty, particularly because there is no immunological-positive marker to distinguish MZL from atypical chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL) and lymphoplasmacytic lymphoma (LPL). Moreover, if translocations involving MALT1 (t(11;18), t(14;18)), BCL10 (t(1;14)) or FOXP1 (t(3;14)) are primarily associated with MALT, and del7q is frequent in SMZL, these cytogenetic markers are not always present. Therefore, there is no specific immunologic marker for splenic or nodal MZL.

To improve the diagnosis of chronic B-cell lymphoproliferative disease (B-LPD) with circulating cells, we have previously developed an innovative mass spectrometry-based proteomic workflow on plasma membrane microparticles derived from B-LPD⁴ and proposed CD148 as a potential marker for MCL diagnosis panel.⁵ Using again the proteomic toolbox, a differential proteomic analysis was applied here to compare the membrane proteomes originating from MPs derived from the following three different B-LPD: LPL, MCL and MZL.

A list of 388 unique membrane proteins was obtained including 59 CD annotated cell-surface proteins (Supplementary Table). Within this list of characterized cell-surface proteins, we focused on CD180, also called Radio-Protective 105 (RP105) or lymphocyte antigen 64, which was considered overexpressed in the MZL sample by our semi-quantitative mass spectrometry analysis (peptide and spectral counting). CD180 is a leucine-rich repeat type 1 membrane protein belonging to the Toll-like receptor (TLR)

family, with significant extracellular homology to the lipopolysaccharide receptor.⁶ Its expression is restricted to antigen presenting cells, such as B cells, monocytes, macrophages and dendritic cells, and anti-CD180 antibodies activate the majority of human and mouse B cells *in vitro*.^{7,8} The literature describes opposite functions for CD180; however, it has been shown that CD180 ligation induces significant B-cell proliferation and Ig production, in both humans and mice.^{7,8}

CD180 expression in B-CLL has been previously described to be different from that in normal CD19⁺ blood cells.⁹ In this paper, we evaluated CD180 expression using flow cytometry in two independent series of B-LPD patients from two French hematology departments, including 60 cases of MZL that were centrally reviewed.

Ninety-six blood samples from patients referred for investigation or follow-up of B-LPD in the Oncology and Hematology Department of the Strasbourg University Hospital were tested: 30 patients with CLL, 16 patients with MCL, 15 patients with LPL and 35 patients with MZL (25 splenic, 10 nodal without splenomegaly), in addition to 25 controls. Blood cells samples were collected after signed informed consent was obtained (Comité de protection des personnes EST-IV, # AC-2008-438) in accordance with the Declaration of Helsinki, and the Ethic Committee of the Faculty of Medicine of Strasbourg approved the study in 2011. Over 93% of the patients were included in the study at the time of diagnosis, and all patients presented a low count of circulating residual normal B cells. All cases of MCL displayed a t(11;14) and/or cyclin D1 hyperexpression. All cases of CLL had a Matutes score of 4 or 5. Only cases of MZL without plasma differentiation were selected in this study. Nine patients with SMZL displayed circulating villous lymphocytes.

As previously described, normal B cells displayed strong staining for CD180. CD27⁺ and CD27[−] cells were equally stained and only a small percentage of CD180[−] B cells were observed in the controls (mean 5.3%, range 0.4–14%, data not shown). In agreement with previous studies⁹ malignant B cells of CLL displayed weaker expression of CD180 ($P < 0.0001$ Mann-Whitney *U*-test) (Figure 1b).

We observed that CD180 staining was also weak in MCL ($P < 0.001$) and LPL ($P < 0.001$) B cells (Figures 1a and b). CD180 mean fluorescence intensities (MFIs) of CLL, MCL and LPL B cells

did not differ significantly ($P > 0.05$, one-way analysis of variance (ANOVA)) and were all significantly lower than the controls ($P < 0.001$, one-way ANOVA and Figure 1b).

By contrast, as hypothesized from the mass spectrometry studies, circulating MZL cells had a high CD180 MFI, which was not different from the controls ($P > 0.05$) but was significantly higher than the CLL, MCL and LPL B cells ($P < 0.001$, non-parametric one-way ANOVA). CD180 staining appeared more heterogeneous in the MZL cells than the control B cells (mean 5865, s.d. = 5643 vs mean 5671, s.d. = 1708). Within the MZL group, CD180 staining of the SMZL and the nodal MZL were not different from the controls ($P = 0.094$ and 0.162 , respectively, Figure 1d), and the SMZL displayed a higher MFI than nodal MZL (4941 vs 8708, $P = 0.0327$ Mann-Whitney U -test).

Similar results were obtained using a second anti-CD180 clone that was used to stain approximately one-third of the patients and control cells (clone MHR73-11, eBiosciences, San Diego, CA, USA; data not shown).

In the same way, the CD180 expression was performed in an independent series of 123 consecutive and prospective peripheral blood samples from patients with B-LPD diagnosed in the Hematology Department of Lyon-Sud University Hospital, according to 2008 WHO recommendations and after integration of all data (cytology, histology, flow cytometry immunophenotyping, conventional and molecular cytogenetics). These cases were classified as CLL ($n = 80$), MZL ($n = 25$), MCL ($n = 10$) and SDRPL ($n = 8$).

The MZL cases presented a strong CD180 expression (median = 537), whereas the MCL and CLL cases showed a very dim or negative CD180 expression (median = 126 and median = 171, respectively) (Table 1, Figures 1a and c). CD180 expression was statistically higher in the eight cases of SDRPL (median = 2 356) than in MZL (median = 537) (Table 1, Figure 1c). The difference of CD180 MFI values between Strasbourg and Lyon could be explained by lack

of standardization between both cytometers. In both centers CLL and MCL B cells displayed a low CD180 staining, always weaker than in those of MZL, whatever their mutational status.

To confirm the flow cytometric protein expression data, total RNA was extracted from highly purified B cells. CD180 RQ-PCR was performed on CD19+ and CD19- fractions of separated blood lymphocytes of 10 CLL, 6 MZL and 6 MCL patient samples, and 5 normal control individuals. As expected, the CD180 transcript was

Table 1. Median of CD180 expression in chronic B-cell lymphoproliferative diseases

	MFI	RFI
<i>Strasbourg series</i>		
CLL ($n = 30$)	906 (111–3226) \pm 794.4	8.429 (2.8–35.5) \pm 7.6
MCL ($n = 16$)	1028 (128–4124) \pm 1084	8.610 (2.18–28.32) \pm 8.8
LPL ($n = 15$)	1247 (316–3475) \pm 905.5	7.188 (3.764–31.03) \pm 10.1
MZL ($n = 30$)	5865 (608–19221) \pm 5374	60.79 (7.753–220.7) \pm 61.24
<i>Lyon series</i>		
CLL ($n = 80$)	171 (44–352) \pm 117	3.4 (1.0–9.1) \pm 2.2
MCL ($n = 10$)	126 (78–286) \pm 80	2.4 (1.5–3.5) \pm 0.7
MZL ($n = 25$)	537 (295–1197) \pm 303	8.3 (4.2–18.7) \pm 4.9
SDRPL ($n = 8$)	2 356 (1 145–5 920) \pm 1 509	35.9 (16.7–103.9) \pm 27.2

Abbreviations: CLL, chronic lymphocytic leukemia; LPL, lymphoplasmacytic lymphoma; MCL, mantle cell lymphoma; MFI, mean fluorescence intensity; MZL, marginal zone lymphoma; RFI, relative fluorescence intensity (over negative controls); SMZL, splenic MZL; SDRPL, splenic diffuse red pulp B-cell lymphoma with villous lymphocytes. Median, (range) and s.d. of CD180 expression. CD180 immunolabelling was performed with the phycoerythrin-conjugated RP105 (clone G28.8, BD Biosciences, Franklin Lakes, NJ, USA) and CD19 (clone 5J25C1, Per-CP, BD Biosciences) using standard procedures. An IgG1 isotype control was also tested for each patient, and the whole blood samples were analyzed within 24 h.

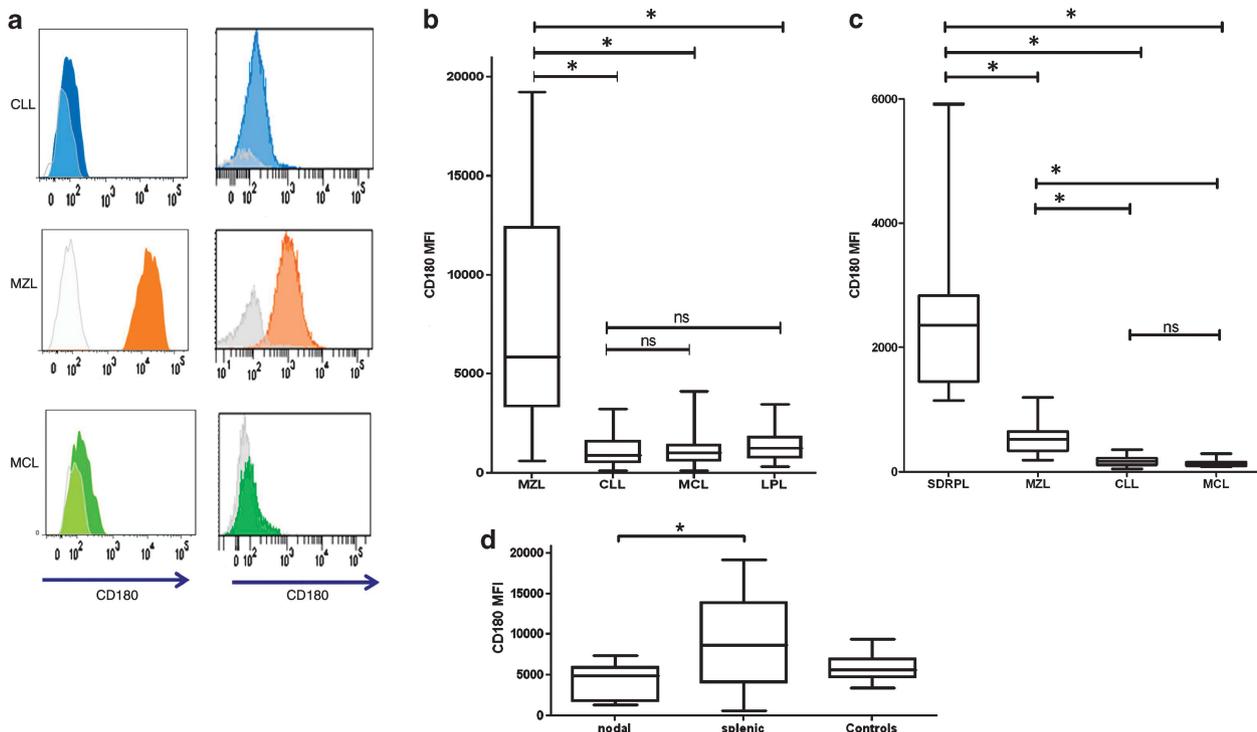


Figure 1. Fluorescence intensity of CD180 in normal and malignant B cells. (a) Representative histograms of CD180 staining of malignant lymphocytes. Right panel: Strasbourg series; left panel: Lyon series. (b–c) Fluorescence intensity of CD180 in malignant B cells. The results were calculated by the FACSDiva software (Becton-Dickinson, Franklin Lakes, NJ, USA) and are expressed as the MFI. The box plot represents the median, 25/75 percentiles and extreme values. $*P < 0.001$; ns, nonsignificant. (b) Strasbourg series, (c) Lyon series. (d) MFI of splenic and nodal MZL of the Strasbourg series.

detected in normal B cells. It was very weakly detected in the CD19⁺ lymphocyte fraction, which primarily contains T cells that do not express CD180 at their surface. The CD180 transcript was also detected in all B-LPD tested, and was expressed at variable but comparable levels to the CD19⁺ lymphocytes of normal controls. Accordingly, CD180 expression was also observed in 56 samples of normal CD19⁺ B cells and B-CLP using microarray analyses (public NCI GEO dataset GSE6691¹⁰).

To test whether variant transcripts may explain the lower expression of CD180 at the cell surface, RT-PCR products covering the entire transcript were sequenced in four CLL and three MZL-purified CD19⁺ cells. According to whole genome sequencing studies of B-LPD,^{11,12} no somatic mutations to the coding sequence that was published previously (Genbank NM_005582.2) were observed. Only the rs169714 polymorphism (667A>G, Ile223Val) was detected in both CLL and MZL samples (data not shown). These results suggest that the lower expression of CD180 in CLL is not associated with the presence of an abnormal transcript.

To test the intracytoplasmic localization of the CD180 protein in B-LPD, FITC- and a PE-conjugated anti-CD180 antibodies were used alternatively for membrane and intracellular staining in CLL ($n=5$), MCL ($n=5$), MZL ($n=5$) and control blood samples ($n=5$). The two combinations revealed comparable intracytoplasmic staining of CD180 in all samples tested, which was not observed in control T cells, whereas the surface staining remained unchanged (data not shown). These findings suggest that the CD180 transcript and protein are present in normal and malignant B cells at similar levels. The area under the curve and receiver operating characteristic curve analysis were used to verify the accuracy of CD180 expression in MZL vs the other B-LPD (Supplementary Figure 1). It appeared that the CD180 MFI or RFI could positively discriminate MZL from CLL, LPL and MCL with more than 86% sensitivity and specificity. These results were obtained in two independent sets of patients, tested in two different laboratories.

To the best of our knowledge, no other available positive marker can be used for this type of characterization. Also, a high CD180 staining may be in favor of a splenic origin of the lymphoma, as SMZL and SDRPL display particularly high levels of CD180 expression. These results require however further confirmation in a larger cooperative study to assess its diagnostic potential.

In conclusion, this study shows the differential expression of CD180 in circulating cells of B-LPD, with high expression restricted to MZL and SDRPL B cells.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by research grants from the Institut National du Cancer (INCa PAIR) Lymphome 2008–2011, the Association pour la Recherche en Hématologie Biologique (ARHEB), the Ligue Régionale Contre le Cancer (Comité du Bas-Rhin), the Fondation pour la Recherche Médicale (FRM), the CNRS, the University of Strasbourg Uds, the Région Alsace and the GIS IBiSA. SL was supported by a studentship from the French Ministère de la Recherche. This work is dedicated to the memory of Dr Estelle Zink.

L Miguet^{1,2,8}, S Lennon^{3,8}, L Baseggio⁴, A Traverse-Glehen⁵, F Berger⁵, N Perrusson², M-P Chenard⁶, A-C Galois¹, A Eischen¹, C Mayeur-Rousse¹, A Maar¹, L Fornecker^{3,8}, R Herbrecht⁷, P Felman⁴, A Van Dorssele³, C Carapito³, S Cianféroni³ and L Mauvieux^{1,2}

¹Laboratoire d'Hématologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France;

²Laboratoire d'Hématologie cellulaire, EA 3430, Institut d'Hématologie et d'Immunologie, Faculté de Médecine de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg, Université de Strasbourg, Strasbourg, France;

³Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), Université de Strasbourg, IPHC, Strasbourg, France. CNRS, UMR7178, Strasbourg, France;

⁴Laboratoire d'Hématologie Cellulaire, Centre Hospitalier Lyon-Sud/Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Strasbourg, France;

⁵Service d'Anatomie-pathologique, Centre Hospitalier Lyon-Sud/Hospices Civils de Lyon, Université Claude Bernard Lyon 1, Pierre-Bénite, France;

⁶Service d'Anatomie Pathologique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France and

⁷Département d'Oncologie et d'Hématologie, Hôpitaux Universitaires de Strasbourg, Strasbourg, France
E-mail: laurent.mauvieux@chru-strasbourg.fr

⁸These authors contributed equally to this work.

REFERENCES

- Baseggio L, Traverse-Glehen A, Petinataud F, Callet-Bauchu E, Berger F, Ffrench M *et al.* CD5 expression identifies a subset of splenic marginal zone lymphomas with higher lymphocytosis: a clinico-pathological, cytogenetic and molecular study of 24 cases. *Haematologica* 2010; **95**: 604–612.
- Swerdlow SH, Campo E, Harris N, Jaffe ES, Pileri SA, Stein H *et al.* *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues*. IARC, 2008.
- Traverse-Glehen A, Baseggio L, Bauchu EC, Morel D, Gazzo S, Ffrench M *et al.* Splenic red pulp lymphoma with numerous basophilic villous lymphocytes: a distinct clinicopathologic and molecular entity? *Blood* 2008; **111**: 2253–2260.
- Miguet L, Pacaud K, Felden C, Hugel B, Martinez MC, Freyssinet J-M *et al.* Proteomic analysis of malignant lymphocyte membrane microparticles using double ionization coverage optimization. *Proteomics* 2006; **6**: 153–171.
- Miguet L, Béchade G, Fornecker L, Zink E, Felden C, Gervais C *et al.* Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 2009; **8**: 3346–3354.
- Miura Y, Miyake K, Yamashita Y, Shimazu R, Copeland NG, Gilbert DJ *et al.* Molecular cloning of a human RP105 homologue and chromosomal localization of the mouse and human RP105 genes (Ly64 and LY64). *Genomics* 1996; **38**: 299–304.
- Miura Y, Shimazu R, Miyake K, Akashi S, Ogata H, Yamashita Y *et al.* RP105 is associated with MD-1 and transmits an activation signal in human B cells. *Blood* 1998; **92**: 2815–2822.
- Chaplin JW, Kasahara S, Clark EA, Ledbetter JA. Anti-CD180 (RP105) activates B cells to rapidly produce polyclonal Ig via a T cell and MyD88-independent pathway. *J Immunol* 2011; **187**: 4199–4209.
- Porakishvili N, Kulikova N, Jewell AP, Youinou PY, Yong K, Nathwani A *et al.* Differential expression of CD 180 and IgM by B-cell chronic lymphocytic leukaemia cells using mutated and unmutated immunoglobulin VH genes. *Br J Haematol* 2005; **131**: 313–319.
- Gutiérrez NC, Ocio EM, de Las Rivas J, Maiso P, Delgado M, Fermiñán E *et al.* Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals. *Leukemia* 2007; **21**: 541–549.
- Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigel HG, Chiang MY *et al.* Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med* 2012; **209**: 1553–1565.
- Rossi D, Trifonov V, Fangazio M, Brusca A, Rasi S, Spina V *et al.* The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med* 2012; **209**: 1537–1551.

Supplementary Information accompanies the paper on the Leukemia website (<http://www.nature.com/leu>)

Annexe 4

Liste des transitions sélectionnées pour le suivi des protéines CD45, CD148, CD5, CD200, CD200_1, CD200_3, CD180 et CD23.

Au total, 37 peptides sont suivis soit 228 transitions.

Les séquences annotées avec une étoile indiquent qu'il s'agit du peptide marqué.

Protéine	Séquence peptidique	m/z précurseur	Temps de rétention (min)	m/z fragments
cd45	DETVDDFWR	591.757	37.16	938.4
				738.3
				837.4
	DETVDDFWR*	596.761	37.16	748.3
				847.4
				948.4
	DPPSESPLEAEFQR	849.902	32.59	175.1
				1173.6
				1486.7
	DPPSESPLEAEFQR*	854.906	32.59	185.1
				1183.6
				1496.7
	IEFDNEVDK	554.761	17.19	719.3
				995.4
				866.4
	IEFDNEVDK*	558.769	17.19	727.3
				874.4
				1003.4
	LFLAEFQSIPR	660.869	65	272.2
				747.4
				1060.6
LFLAEFQSIPR*	665.873	65	282.2	
			1070.6	
			757.4	
LNPTPGSNAISDVPGER	862.434	19.86	458.2	
			1298.6	

			557.3
LNPTPGSNAISDVPGER*	867.438	19.86	468.2
			567.3
			1308.6
NSNVIPYDYNR	677.823	19.4	452.2
			940.5
			827.4
NSNVIPYDYNR*	682.827	19.4	837.4
			950.5
			462.2
VDVYGYVVK	521.284	29.46	565.3
			827.5
			573.3
VDVYGYVVK*	525.291	29.46	736.4
			835.5
			728.4
YHLEVEAGNTLVR	500.932	25.92	602.4
			730.4
			659.4
YHLEVEAGNTLVR*	504.268	25.92	740.4
			612.4
			669.4
TLILDVPPGVEK	640.877	42.87	840.4
			626.4
			953.5
TLILDVPPGVEK*	644.884	42.87	961.5
			634.4

				848.5	
	YVDILPYDYNR	715.851	44.06	452.2	
				940.5	
				827.4	
	YVDILPYDYNR*	720.855	44.06	462.2	
				837.4	
				950.5	
	cd148	DTEVLLVGLEPGTR	749.909	47.05	729.4
					828.5
					430.2
		DTEVLLVGLEPGTR*	754.913	47.05	440.2
					739.4
838.5					
FSGFEASHGPIK		638.82	19.21	1042.5	
				1129.6	
				838.4	
FSGFEASHGPIK*		642.827	19.21	846.5	
				1050.5	
				1137.6	
GLIDGAESYVSFSR		750.87	42	974.5	
				496.3	
				845.4	
GLIDGAESYVSFSR*		755.874	42	855.4	
				984.5	
				506.3	
SQLSEVLK	495.777	20.66	147.1		
			575.3		

			775.5
SQSLSEVLK*	499.784	20.66	155.1
			583.4
			783.5
TGTFIAIDR	497.272	24.36	474.3
			734.4
			587.4
TGTFIAIDR*	502.276	24.36	744.4
			484.3
			597.4
VITEPIPVSDLR	674.889	39.58	1136.6
			696.4
			906.5
VITEPIPVSDLR*	669.885	39.58	686.4
			896.5
			1126.6
YAAELAENR	518.757	11.29	602.3
			802.4
			489.2
YAAELAENR*	523.761	11.29	499.3
			612.3
			812.4
YNNVLPYDISR	677.344	31.67	1076.6
			750.4
			863.5
YNNVLPYDISR*	682.348	31.67	760.4
			1086.6

				873.5
cd5	HLPETEAGR	505.257	6.31	872.4
				533.3
				759.4
	HLPETEAGR*	510.261	6.31	543.3
				769.4
				882.5
	EHQPLPIQWK	638.346	29.5	1009.6
				881.5
				671.4
	EHQPLPIQWK*	642.353	29.5	889.5
				679.4
				1017.6
	LSWYDPDFQAR	699.328	47.08	733.4
				848.4
				1197.5
	LSWYDPDFQAR*	704.332	47.08	743.4
				858.4
				1207.5
	LVGGSSIC[+57.0]EGTVEVR	781.896	20.5	660.4
				949.4
1350.6				
LVGGSSIC[+57.0]EGTVEVR*	786.9	20.5	1360.6	
			959.4	
			670.4	
NSHLSAYPALEGALHR	868.447	28.41	1284.7	
			1197.6	

				963.5	
	NSHLSAYPALEGALHR*	873.451	28.41	973.5	
				1294.7	
				1207.6	
				732.4	
	QWEDPSQASK	588.27	10.08	861.4	
				617.3	
				869.4	
	QWEDPSQASK*	592.277	10.08	740.4	
				625.3	
				823.4	
	VLALLC[+57.0]SGFQPK	666.871	50	936.5	
				1120.6	
				831.4	
	VLALLC[+57.0]SGFQPK*	670.878	50	944.5	
				1128.6	
				460.3	
	VLDAGDPTSR	515.762	9.2	703.3	
				818.4	
				470.3	
	VLDAGDPTSR*	520.766	9.2	713.3	
				828.4	
				515.3	
CD200	EQLYTPASK	575.311	20.64	616.4	
					779.4
					787.4
	EQLYTPASK*	579.318	20.64	523.3	

				624.4
cd200_1	GELSQGVQK	473.254	6.48	646.4
				759.4
				431.3
	GELSQGVQK*	477.261	6.48	439.3
				654.4
				767.5
cd200_3	TTINDYQVIR	611.825	20.28	678.4
				793.4
				907.5
	TTINDYQVIR*	616.829	20.28	688.4
				803.4
				917.5
cd180	NLSHLQTLNLSHNEPLGLSQAFK	1345.206	49	580.3
				878.5
				1088.6
	NLSHLQTLNLSHNEPLGLSQAFK*	1349.214	49	588.3
				886.5
				1096.6
	GNHFQDGTITK	609.299	9.45	1046.5
				909.5
				248.2
	GNHFQDGTITK*	613.306	9.45	256.2
				917.5
				1054.5
	SLEQAINLSLNFNGNNVK	988.016	60	1106.6
				906.4

				1019.5	
	SLEQAINLSLNFNGNNVK*	992.023	60	914.5	
				1027.5	
				1114.6	
	VLDFQNNAIHYISR	845.439	42.27	675.4	
				538.3	
				1087.6	
	VLDFQNNAIHYISR*	850.443	42.27	685.4	
				548.3	
				1097.6	
	cd23	HASHTGSWIGLR	661.342	19.38	1113.6
					1184.6
788.4					
HASHTGSWIGLR*		666.346	19.38	798.5	
				1123.6	
				1194.6	
LPTPSAPLHS		510.28	17.51	611.3	
				708.4	
				809.4	
LPTPSAPLHS*		512.283	17.51	712.4	
				813.4	
				615.3	
NEASDLLER	523.759	18.41	803.4		
			645.4		
			732.4		
NEASDLLER*	528.763	18.41	655.4		
			742.4		

				813.4
				788.4
	SQSTQISQELEELR	824.413	42.47	546.3
				1003.5
	SQSTQISQELEELR*	829.417	42.47	798.4
				556.3
				1013.5
	WNDAFC[+57.0]DR	542.219	17.99	450.2
				668.3
				783.3
	WNDAFC[+57.0]DR*	547.223	17.99	460.2
				607.3
				793.3

Annexe 5

Communication par affiche

2nd meeting annuel du GDR RCPG-PhysioMed à Strasbourg en octobre 2013

Lennon S., Saliou JM., Fève M., Zeniou M., Carapito C., Dong J., Van Dorselaer A., Junier MP., Chneiweiss H., Cianférani S., Haiech J., Kilhoffer MC.

Proteomic analysis of the GPCR repertoire in glioblastoma stem cells, comparison to non-cancer cells

Proteomic analysis of the GPCR repertoire in glioblastoma stem cells. Comparison to non-cancer cells.

Sarah Lennon ⁽¹⁾, Jean-Michel Saliou ⁽¹⁾, Maria Zenou ⁽²⁾, Marie Fève ⁽²⁾, Christine Carapito ⁽¹⁾, Jihu Dong ⁽²⁾, Alain Van Dorselaer ⁽¹⁾, Hervé Chneiweiss ⁽³⁾, Jacques Haiech ⁽²⁾, Sarah Cianféroni ⁽¹⁾, Marie-Claude Killhofer ⁽²⁾

(1) Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, DSA, CNRS UMR7178, UoS, 67087 Strasbourg, France

(2) Intitut Gilbert Laustriat, CNRS, UMR 7175, UoS, Illkirch, France

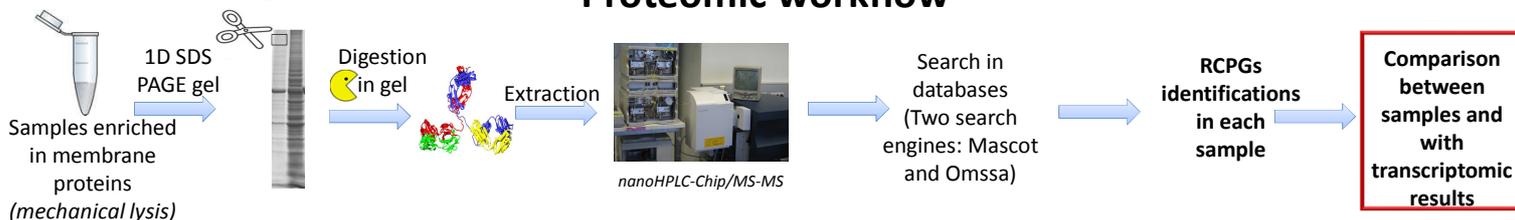
(3) Plasticité gliale et tumeurs au cerveau, Inserm UMR 894, Université Paris Descartes, Paris, France

Introduction

Glioblastomas (GBMs) are highly aggressive and invasive brain tumors with bad prognosis. As other solid tumors, they are heterogeneous being composed by different cell types. Among these cells figure cancer stem cells (GCSs) considered as the culprits of tumor maintenance and relapse after treatment. CSCs are malignant cells with stem properties, able to propagate the tumor in animal models. They are particularly resistant to chemo- and radiotherapies. GCSs appear nowadays as the prominent targets for new chemotherapies development. Understanding their pathophysiology is thus essential. We focused our work on G-proteins coupled receptors (GPCRs), a family of membrane receptors playing a prominent role in cell signaling, cell communication and crosstalk with the microenvironment. The role of GPCRs in cancer has been highlighted but remains largely unexplored.

We report here a comparative expression study of the GPCR's proteome using mass spectrometry in glioblastoma cancer stem cells (GBM CSCs), glioblastoma cancer cells (U87 MG), human astrocytes (HA) and fetal neural stem cells (f-NSC). Proteomic data are presented along with transcriptomic data.

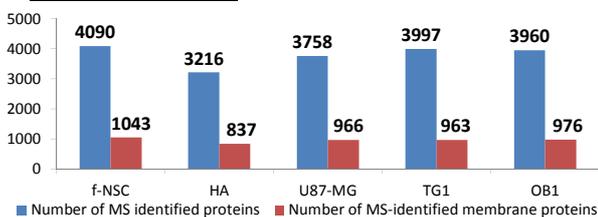
Proteomic workflow



Results

Proteomics results

Global Identifications



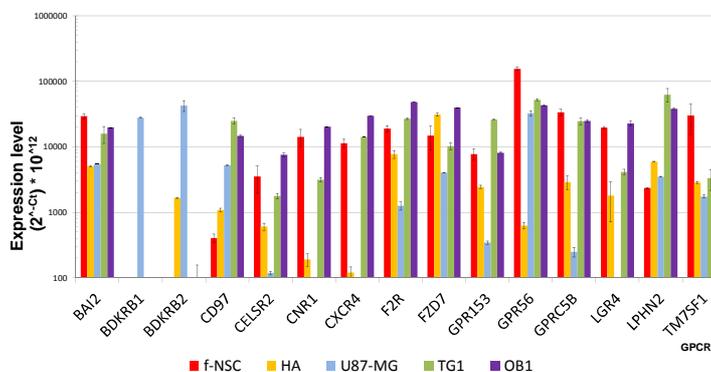
GPCRs expression level : peptide counts and in bracket spectral count

GPCR Name	Accession number	Molecular weight (kDa)	Expected peptides	f-NSC protein level	HA protein level	U87-MG protein level	OB1 protein level	TG1 protein level
GPR55	Q8Y653	78	14	8 (37)	0	3 (6)	4 (8)	4 (10)
LPHN2	Q95490	163	50	0	0	0	9 (15)	12 (24)
F2R	P25116	44	5	0	0	0	2 (6)	0
BDKRB2	F30411	44	0	0	0	0	0	0
FZD7	O75084	61	3	0	0	0	0	0
GPRC5B	Q9NZ00	41	3	3 (7)	0	0	1 (3)	1 (4)
TM7SF1	Q90478	45	3	0	0	0	0	0
CXCR4	P61073	40	2	0	0	0	0	0
BAI2	60241	173	40	1 (1)	0	0	0	0
BDKRB1	P46663	40	2	0	0	0	0	0
GPR153	Q8NV754	65	1	0	0	0	0	0
CD97	P48960	90	19	0	0	6 (21)	10 (40)	9 (31)
LGR4	Q9N637	104	8	0	0	0	0	0
CNR1	P31554	53	3	0	0	0	0	0
CELSR2	Q9HCU4	317	79	8 (12)	0	0	2 (2)	2 (2)

Transcriptomic results

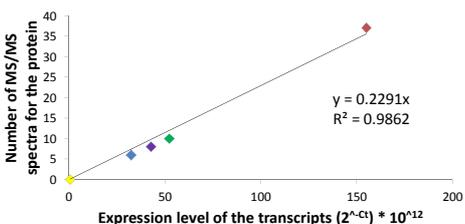
Expression level of the transcripts

GPCRs' expression was analyzed by RT-QPCR using TaqMan assays. The most highly expressed GPCRs are represented.



Correlation between transcriptomics and proteomics

1st case Results in agreement: example with GPR56



2nd case No correlation between technics: example with GPR153

High transcripts expression but not detected by mass spectrometry

Discrepancies can be explained by the number of MS detectable peptides in each protein (purple = tryptic peptides from transmembrane domain, red = MS potentially detectable tryptic peptides, underlined when detected in the experiment)

GPR56
 MTFQSLQTTFLFLSLLFLVQAGHRRGRHDFRFSQRRNQTRHRSLSLYKPPFDLRIESTN
 SERALVYRPFPAABASPSFPDPRGLVHCLVWNRHAGRLHLLVGRDFLSDKASLLL
 CFQDQESLAQDFLLATSYVSWRSRQNSLSLSASSTFFSRFSRFAAHSNVDKCEKLE
 RDLLLSQFLHFQASRRFAASASQQLSLSKLTYSVYMGDMVSEFDRIINATVNRLL
 QPTAGLDLHFSRQEPSPSINMYSVLLRFLRFTFORRSGEAEKRLLVDPSSQALPQ
 NNSQSLGSEVFLGKLVYKRYVALYKTVKLVVQKQALQPPFTLCCYFVDEKPLASRGR
 NSSAGCETVRETQSCFNHLLYFVLMVSYVEVDVRRKHLVLSLVYVQVSAALCLV
 TIANVLSNVLPCRFRFRDITVRRHLLLVFLDPSFLISEVFLTQSAQCPASAT
 FLRFLSLTCSNGLGSLHLLVQKLVYKTVKLVVQKQALQPPFTLCCYFVDEKPLASRGR
 NYGFIILAVRHTFPGVYFSKWCIRLSVSYITNLGSLVFLFNMGMLATMVOQLRLR
 PHTQWRSHVLLDGLSLVGLDVALFLWFLPFSQTFQVLVFLSFTTFSQGLFIWVWS
 HLLQARQSPFLKRWSDARLPSRSGTSSSR

GPR153
 MSEEARLPGSAVQNVVCSGLSLLADAMNLLSYVQAGQRFKFLVLLCTLAATMVMVAVP
 LAIYVQVLRQRQPEFNEGLCKVYVSTFTLLIATCFVLSVYSHHQQVCFVNRFL
 SNARKGAVTVYVGMVYFLLSALFVQWVOTSEPTTHQCFIVRAIQLGDFVCLLY
 GSEFVMQVTCALPFLYVQVQKQASRAFFWYFVETDAQDQSRSSIDSEPAKSL
 QTSVTVTVTVVQVDFVYVQVFSLSAASAPQMLVLCVSVAGLLEFVFLKAC
 IIRFALLAVRCKMLNARSESQDSTLQSFQVLSLIVYRQSFVLDQNRM
 YEISALGQLPQLYRFLRLEQDMVYVQVFTFRSHDDADVRAVPLPAPLFRWSSGSD
 LAALNLYLPAQERRRALLAPADAFPPRARSRASLILRFSALDSDRGRARDSPF
 CSPRRRQSPKSAKALLRFLALTLKPFQKALRFPFPKAPKAPQSDQSCAFTF
 FSSAQRSPRRSAHSHASRFPOLSAHQEPPQLRAAGQGGSTSSFLSPSSSESOYATL
 HSDLSGAS

- To be MS detectable, peptide should have
- 1) 1000 < MW < 2000 Da
 - 2) Cleavage sites
 - 3) No PTMs

GPCRs with more than 5 potentially MS-detectable peptides and at least 20 000 units expression in transcriptomics will be detected by mass spectrometry

Conclusions

- First comparative study on GPCRs issued from GBMs cells in non cancer neural cells to combine transcriptomics and proteomics
- Among the 7000 identified proteins, 7 GPCRs were detected by mass spectrometry
- Level of expression were comparable to transcriptomic studies if more than 5 potentially MS-detectable peptides are expected and if the transcriptomics expression is high.

Annexe 6

Publication dans le journal *Plos One*

Fève M.*, Saliou JM.*, Zeniou M.*, Lennon S., Carapito C., Dong J., Van Dorsselaer A., Junier MP., Chneiweiss H., Cianférani S., Haiech J., Kilhoffer MC.

Comparative Expression Study of the Endo-G Protein Coupled Receptor (GPCR) Repertoire in Human Glioblastoma Cancer Stem-like Cells, U87-MG Cells and Non Malignant Cells of Neural Origin Unveils New Potential Therapeutic Targets

Plos One, 2014, **9**, Issue 3

*Ces auteurs ont contribué de manière égale à ce travail.

Comparative Expression Study of the Endo-G Protein Coupled Receptor (GPCR) Repertoire in Human Glioblastoma Cancer Stem-like Cells, U87-MG Cells and Non Malignant Cells of Neural Origin Unveils New Potential Therapeutic Targets

Marie Fève¹✉, Jean-Michel Saliou²✉, Maria Zeniou¹✉, Sarah Lennon², Christine Carapito², Jihu Dong¹, Alain Van Dorsselaer², Marie-Pierre Junier³, Hervé Chneiweiss³, Sarah Cianférani², Jacques Haiech¹, Marie-Claude Kihloffer¹*

1 Laboratoire d'Innovation Thérapeutique, UMR7200, Laboratoire d'Excellence Medalis, CNRS, Université de Strasbourg, Faculté de Pharmacie, Illkirch, France,

2 Laboratoire de Spectrométrie de Masse BioOrganique, UMR7178, CNRS, Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, Strasbourg, France,

3 Neuroscience Paris Seine, UMR8246, Inserm U1130, Institut de Biologie Paris Seine, CNRS, Université Pierre et Marie Curie, Paris, France

Abstract

Glioblastomas (GBMs) are highly aggressive, invasive brain tumors with bad prognosis and unmet medical need. These tumors are heterogeneous being constituted by a variety of cells in different states of differentiation. Among these, cells endowed with stem properties, tumor initiating/propagating properties and particularly resistant to chemo- and radiotherapies are designed as the real culprits for tumor maintenance and relapse after treatment. These cells, termed cancer stem-like cells, have been designed as prominent targets for new and more efficient cancer therapies. G-protein coupled receptors (GPCRs), a family of membrane receptors, play a prominent role in cell signaling, cell communication and crosstalk with the microenvironment. Their role in cancer has been highlighted but remains largely unexplored. Here, we report a descriptive study of the differential expression of the endo-GPCR repertoire in human glioblastoma cancer stem-like cells (GSCs), U-87 MG cells, human astrocytes and fetal neural stem cells (f-NSCs). The endo-GPCR transcriptome has been studied using Taqman Low Density Arrays. Of the 356 GPCRs investigated, 138 were retained for comparative studies between the different cell types. At the transcriptomic level, eight GPCRs were specifically expressed/overexpressed in GSCs. Seventeen GPCRs appeared specifically expressed in cells with stem properties (GSCs and f-NSCs). Results of GPCR expression at the protein level using mass spectrometry and proteomic analysis are also presented. The comparative GPCR expression study presented here gives clues for new pathways specifically used by GSCs and unveils novel potential therapeutic targets.

Citation: Fève M, Saliou J-M, Zeniou M, Lennon S, Carapito C, et al. (2014) Comparative Expression Study of the Endo-G Protein Coupled Receptor (GPCR) Repertoire in Human Glioblastoma Cancer Stem-like Cells, U87-MG Cells and Non Malignant Cells of Neural Origin Unveils New Potential Therapeutic Targets. *PLoS ONE* 9(3): e91519. doi:10.1371/journal.pone.0091519

Editor: Roland Seifert, Medical School of Hannover, Germany

Received: December 28, 2013; **Accepted:** February 10, 2014; **Published:** March 24, 2014

Copyright: © 2014 Fève et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by La Ligue contre le Cancer, La Fondation pour la Recherche Médicale, the Agence Nationale pour la Recherche (ANR), the Proteomic French Infrastructure (ANR-10-INSB-08-03), the Région Alsace, Université de Strasbourg, CNRS, the French Ministère de l'enseignement supérieur et de la recherche, the Fondation ARC (SFI20121205939, HC/MPJ) and has been preformed within the LABEX ANR-10-LABX-0034_Medalis and received a financial support from French government managed by "Agence Nationale de la Recherche" under "Programme d'investissement d'avenir". Marie Fève is indebted to French Ministère de l'enseignement supérieur et de la recherche for doctoral fellowship and Sarah Lennon to the Fondation ARC pour la Recherche sur le Cancer. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mckihloffer@unistra.fr

✉ These authors contributed equally to this work.

Introduction

Glioblastomas (World Health Organization (WHO) grade IV astrocytomas) are highly aggressive, angiogenic and infiltrating brain tumors representing more than 50% of all gliomas [1]. Their outcome is poor, most treatments currently in use remaining inefficient on long term survival [2,3,4] and less than 5% of patients survive 5 years post diagnosis [1]. Numerous clinical trials, mainly in phase I/II, are ongoing worldwide to find treatments with increased efficacy (<http://clinicaltrials.gov/>, [<http://apps.who.int/trialsearch/> and for recent review see \[5\]\). Improvement of the standard Stupp protocol \[3,6\], association to topoisomerase inhibitors, alkylating agents, tyrosine kinase inhibitors, intercalating agents or antibodies targeting VEGF or the EGF receptor, development of new combinations of chemotherapeutic agents and testing new delivery drug devices have been used in different protocols \(for exhaustive review see \[5\]\). However no therapy appears as the panacea. Seeking for new molecules acting on targets different from those explored so far and improvement of](http://apps.</p></div><div data-bbox=)

multi-targeted protocols are necessary to decrease recurrence and morbidity of this brain cancer that afflicts humans of all ages.

A new paradigm in cancer therapies arose a decade ago with the identification, within tumors, of cells endowed with stem cell properties and able to propagate tumors in immunodeficient mice with high efficacy. Such cells were first identified in hematological cancers [7]. In 2001, Reya postulated that they may be an integral part of most if not all tumors [8]. Cells with stem properties, long term survival, able to self-renew, differentiate into various cells types and lead to tumor formation after serial transplantation of few cells to immuno-deficient mice were found in many solid tumors including glioblastomas (GBM) [9,10], breast [11], colon cancer [12,13], melanoma [14,15] and pancreatic cancer [16]. Recent experiments, using genetic techniques, allowing the tracing of cancer cells within tumors strongly argue in favor of the presence of tumor stem cells within GBM [17], skin [18] and adenoma [19] tumors *in vivo*. These cancer stem cells (CSCs), also termed tumor stem cells, tumor initiating cells, tumor propagating cells or stem-like cancer propagating cells are considered as the cells responsible for tumor maintenance and relapse after treatment. Classical radiotherapies and chemotherapies which often induce tumor regression are thought to act preferentially on the non-stem cancer cells sparing CSCs and thus causing cancer recurrence. In addition, CSCs are thought to be able to enter a quasi-quiescent state compared to the rest of the tumor cells and thus escape treatments often designed to act on highly proliferating cells. CSCs have thus become the cells to target in order to increase cancer therapy efficacy. Numerous reviews are highlighting this point (for the most recent ones see [20,21,22,23,24,25,26]).

In the present study, we focused our attention on G-protein coupled receptors' (GPCRs) expression in cancer cells with stem and tumor propagating properties isolated from human GBM biopsies [27] and called herein glioblastoma stem-like cells (GSCs). The approach relied on the pivotal role played by this large family of membrane receptors in most key physiological functions including cell communication, cell signalling, cell migration, cell proliferation and apoptosis. These seven trans-membrane domain receptors convey information from outside the cell to inside by binding a variety of physiological compounds of wide chemical diversity. Their paramount importance in cell activity places GPCRs among the most appealing pharmacological targets. This has been extensively exploited to develop major drugs of therapeutic interest for a variety of pathologies including cardiovascular diseases, psychological disorders, allergies, so that nowadays 30–50% of current drugs target GPCRs [28,29,30,31]. Recently, GPCRs also gained attention in anti-tumor drug discovery due to the prominent role they can play, directly or indirectly through transactivation of other receptors, in cancer cell pathophysiology including proliferation, migration/metastasis, angiogenesis and metabolism [31,32,33,34,35,36]. The SMO antagonist vismodegib has recently been FDA-approved in basal cell carcinoma [37] and several molecules targeting GPCRs are in clinical trials as adjuvants of chemotherapies (namely Bosentan, an Endothelin A/B subtype receptor antagonist for melanoma or BKT140, a CXCR4 antagonist for multiple myeloma). In addition, the possible use in cancer therapy of monoclonal antibodies targeting GPCRs or their ligands has been reviewed in the literature [33,38].

In order to get a deeper insight into the pathophysiology of GSCs and find new potential therapeutic targets, we undertook a descriptive study of the set of GPCRs expressed in CSCs isolated from patient primary GBMs in comparison with the GPCR expression repertoires of the GBM cell line U87-MG, non-cancerous human fetal neural stem cells (f-NSCs) and primary

human astrocytes (HA). This set of cells was chosen as a first step to compare cancerous and non-cancerous neural cells with stem properties or in a more differentiated state. The study was performed both at the transcriptomic and proteomic levels. The expression profile of this family of proteins appears finely regulated despite karyotype alterations and changes in DNA ploidy [27]. The exquisitely tuned GPCR expression changes observed in the different cell types tested (consisting of cancerous and non-cancerous cells, stem and non-stem cells) lead to a molecular signature evidenced by hierarchical clustering. Finally, the study discloses several GPCRs with high expression specificity in GSCs compared to non-cancerous stem cells or the U-87 GBM cell line. In addition, GPCRs related to stemness have been highlighted. Those GPCRs may give a hook to modify the fate of cancer stem-like cells and more differentiated cancer cells.

Results

Global analysis of endo-GPCR gene expression in human GSCs, fetal NSCs, astrocytes in primary culture and GBM U-87 MG cells

Expression of GPCRs was studied in five different human cell types comprising TG1 and OB1 GSCs isolated from GBM biopsies of two patients, the GBM cancer cell line U-87 MG, fetal neural stem cells (f-NSCs) and human astrocytes (HA) in primary cultures. Expression studies were performed by RT-QPCR using the GPCR low density microarrays (Applied Biosystems) in the 384-well plate format. These microarrays allow testing 356 different GPCRs, 2 members of the LANCL family (LANCL1 and LANCL2), which are not considered as GPCRs, 14 different housekeeping genes (with 18 s ribosomal RNA present in quadruplicate) and 11 other genes not related to GPCRs. The list of genes present on the microarray is presented in Table S1. To analyze GPCR expression in the different cell types, data were normalized taking RPLP0, one of the most stable housekeeping genes in our experiments, as reference (Figure S1). The use of RPLP0 alone gave results similar to those using geometric averaging of multiple internal control genes [39]. The results were also similar to those obtained after normalization against TBP previously reported as being an adequate housekeeping gene for expression studies in GBM [40]. In the present experiments, the cycle threshold (Ct) value of RPLP0 was 21.1 ± 0.28 . RPLP0 is expressed at a ~ 6000 fold lower level compared to the ribosomal RNA 18 s. The most highly expressed GPCRs in each cell type are given in Table 1. GPR56 is highly expressed in all tumor cells but also and even at a higher level, in f-NSCs, the non-tumor stem cells of fetal origin. Its expression in astrocytes is reduced (Ct = 31.2; ratio over RPLP0 = 0.0009). For comparison, low levels of GPR56 were also observed in HEK293 cells, another non-stem cell line (Ct = 30.6 (after normalization); ratio over RPLP0 = 0.001) (data not shown).

Considering the whole set of 356 GPCRs (list in Table S1), we focused our attention on those expressed at Ct values ≤ 31.5 (i.e. expressed down to 1.33×10^{-3} fold the level of RPLP0) in at least one of the five cell types tested. This stringent condition gives the best signal to noise ratio, although it may leave aside some interesting genes expressed at lower levels. Out of the 356 different GPCRs present on the array, 138 satisfy this criterion (see list and expression values in Table S1). Expression of these 138 GPCRs in the different cell types is represented in Figure 1, where GPCRs are listed, in alphabetical order, as a function of their expression level (expressed as $2^{-Ct} * 10^{12}$ (arbitrary units)). Among these 138 GPCRs, 90 genes were expressed above the cutoff level in OB1 cells, 83 genes in TG1 cells, 77 genes in f-NSCs, 61 genes in

Table 1. GPCRs most highly expressed in the five cell types.

Cell type	GPCR	Ct	Delta Ct (Ct _{GPCR} - Ct _{RPLP0})	GPCR expression compared to RPLP0 (fold)
TG1	LPHN2	24±0.3	2.88±0.58	7.36±1.49
	GPR56	24.2±0.04	3.08±0.32	8.46±1.25
OB1	F2R	24.3±0.01	3.18±0.28	9.06±1.22
	GPR56	24.4±0.01	3.28±0.29	9.71±1.22
	FZD7	24.5±0.03	3.38±0.31	10.41±1.24
f-NSC	LPHN2	24.6±0.04	3.48±0.32	11.16±1.25
	GPR56	22.6±0.08	1.48±0.36	2.79±1.28
	FZD3	24.7±0.18	3.58±0.46	12.38±1.38
HA	GPRCSB	24.8±0.15	3.68±0.43	13.09±1.35
	FZD1	25±0.08	3.88±0.36	14.72±1.28
U87	FZD7	25±0.1	3.88±0.38	14.72±1.3
	BDKRB2	24.5±0.3	3.38±0.58	10.41±1.49
	GPR56	24.9±0.13	3.78±0.41	13.74±1.33

GPCRs with Ct values ≤ 25 are indicated. TG1 and OB1 correspond to the GBM cancer stem-like cells. f-NSC, HA and U-87 stand for fetal neural stem cells, human astrocytes and the U-87 GBM cell line, respectively. Ct corresponds to GPCR cycle threshold normalized to the Ct of RPLP0 (Ct = 21.12). GPCR fold expression corresponds to the expression of a given GPCR in a given cell type compared to the expression of RPLP0 calculated as: fold = $2^{\Delta(\text{Ct}_{\text{RPLP0}} - \text{Ct}_{\text{GPCR}})}$. doi:10.1371/journal.pone.0091519.t001

human astrocytes and 57 genes in the U-87 MG GBM cell line (see expression data in Table S1).

Hierarchical clustering of GSCs, f-NSCs, human astrocytes and GBM U-87 MG cells using endo-GPCR gene expression

The 138 GPCRs significantly expressed (with Ct ≤ 31.5 or expression level >329 units) in at least one of the five cell types studied were used to perform unsupervised hierarchical clustering. Three main clusters could be observed (Figure 2). A first cluster gathers the GSCs TG1 and OB1, the second assembles human astrocytes (HA) and GBM U-87 MG cells and the third cluster corresponds to f-NSCs, which appeared closer to the tumor stem cells than to the astrocytes and GBM U-87 MG cells. The same three clusters are obtained when using all GPCRs present on the microarray without filtering (data not shown). The clustering clearly points to sets of genes specifically expressed or overexpressed either in a given cell type or common to two or more cell types. GPCR expression pattern thus constitutes a signature specific to a given cell type and allows clustering cells as a function of their physiological or pathophysiological states. Of note, the GBM U-87 MG cells cluster apart from the GSCs TG1 and OB1.

Comparative analysis of the differences in GPCR expression among the five cell types, with emphasis on GPCRs specifically expressed in the GBM CSCs will be performed in the following paragraphs.

Analysis of the core of GPCRs expressed in all of the five cell types

Among the 138 GPCRs retained in our study, 26 are present at levels above the cutoff value (expression level ≥ 329 units, i.e. Ct ≤ 31.5) in the five cell types (Figure 3A and 3B). Receptors of the family of adhesion GPCRs, including BAI2, CD97, CELSR1, GPR125, LPHN1, LPHN2 and GPR56, are the most highly represented (7/26), followed by members of the Frizzled receptor family (5/26) comprising FZD1, FZD2, FZD5, FZD7 and FZD8. The vomeronasal GPCR VN1R1, is also a member of this group of “common GPCRs”. Although expressed in all five cell types,

most of the genes present a differential expression level between the five cell types.

As aforementioned, GPR56 is one of the most highly expressed genes in the three cancer cell types tested (TG1, OB1 and U-87 MG). Figure 3 shows that this gene is also highly expressed in f-NSCs. Its expression level in this cell type is respectively 3, 3.6 and 4.8 fold higher compared to that in TG1, OB1 and U-87 MG cells and largely exceeds its expression level in human astrocytes, which is close to the limit fixed in our study. Besides GPR56, the thrombin receptor F2R/PAR1 and to a lesser extent BAI2, GABBR1 and the orphan receptor GPR153 show significantly higher expression in the cells with stem cell properties (f-NSCs, TG1 and OB1), compared to the two other cell types. Only one gene in this group, LPHN2, appears to be overexpressed in the GSCs TG1 and OB1, but not in f-NSCs, nor in the GBM U-87 MG cells or in human astrocytes. Its expression in these two cell types exceeds the mean expression in the three other cell types HA, f-NSC and U-87 MG by 13 fold. On the overall, astrocytes and U-87 MG cells express the “commonly expressed” GPCR genes at the lowest level (with the exception of FZD1 and FZD7 in astrocytes).

GPCRs with expression specificity in TG1 and OB1 GSCs

Expression data analyses unveil nine GPCRs with expression specificity in TG1 and OB1 cells (Figure 4). LPHN2, one of these GPCRs, belongs to the group of “commonly expressed” genes (Figure 3 and previous section). However, LPHN2 appears significantly overexpressed in the two GSC cell types compared to the GBM U-87 MG cells or to the non-cancerous cells (Figure 4).

Seven other genes are expressed with specificity in TG1 and OB1 cells. These include CALCRL, GPR37, GPR73, HRH2, GPR103 and GPR128, the levels of the latter two being close to the limit set in our study (Figure 4). The calcitonin receptor-like receptor CALCRL (or CRLR or CLR) is expressed above, but close to the cutoff level in astrocytes and f-NSCs, whereas it is downregulated in U-87 MG cells. Six genes, GPR37, HRH2, GPR73, EDG8, GPR128 and GPR103 are expressed above the cutoff level only in TG1 and OB1 cells. GPR37 is expressed well

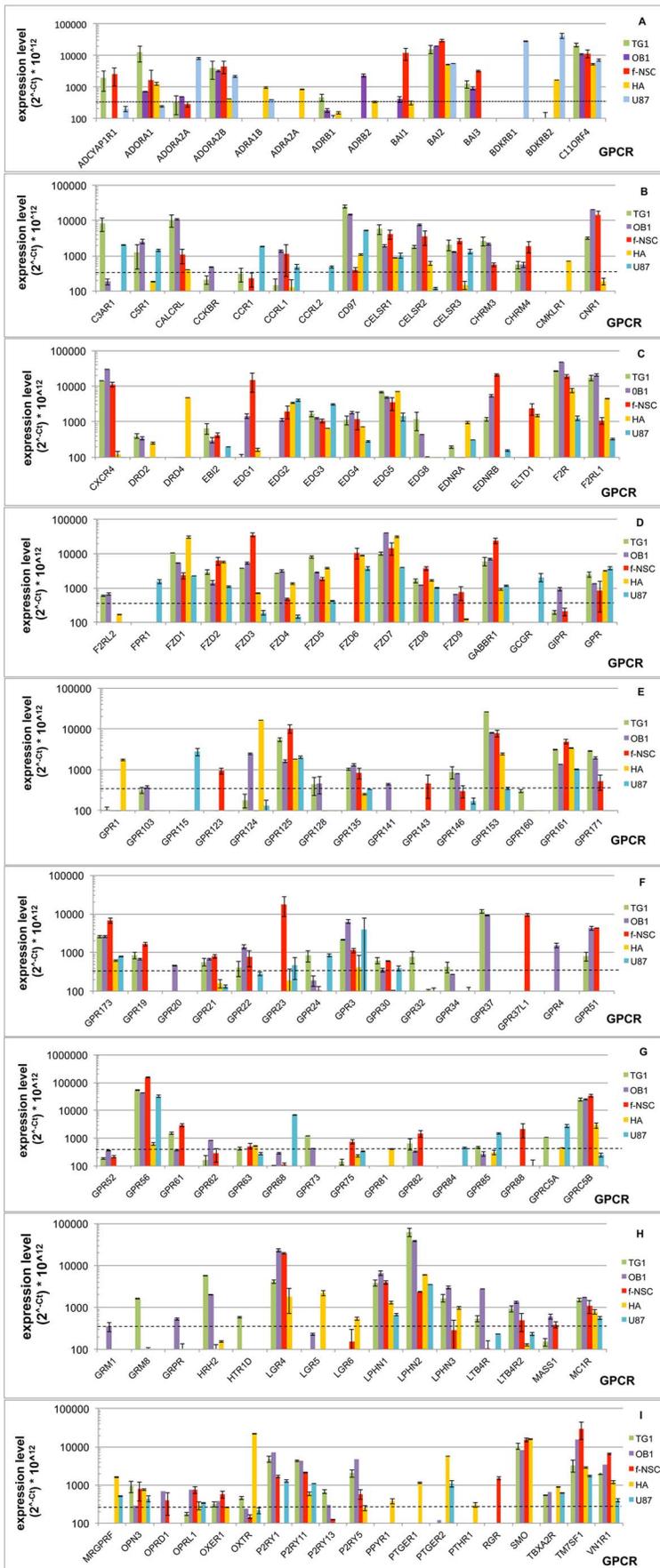


Figure 1. GPCR expression in TG1, OB1, f-NSC, HA and U-87 -MG cells. GPCR expression was determined by RT-QPCR. GPCRs are presented in alphabetical order. Expression levels are related to the cycle threshold value and expressed as $2^{-Ct} * 10^{-12}$, where Ct corresponds to the cycle threshold of a given GPCR. A Ct of 31.5 corresponds to an expression level of 329 (arbitrary units). Dotted lines indicate the cutoff level (Ct=31.5). Expression levels are presented on a logarithmic scale. All data were normalized using RPLP0 as housekeeping gene. doi:10.1371/journal.pone.0091519.g001

above the cutoff value in TG1 and OB1 cells and is strongly upregulated in these cells in comparison to f-NSCs, HA and U-87 MG cells. GPR73 expression is increased in both TG1 and OB1 cells, with highest amplitude in TG1 cells. The last three receptors of the group, EDG8, GPR128 and GPR103 are specifically expressed in TG1 and OB1 cells, but at levels close to the cutoff value fixed in the present study. GPR103, although expressed in the lower range level, shows total specificity for TG1 and OB1 cells.

Among the 138 selected GPCRs, only one, FZD6, appears downregulated in TG1 and OB1 cells. FZD6 shows a rather high expression in HA cells and f-NSCs. It is downregulated in U-87 MG cells and absent in the two GSC types. FZD6 belongs to the frizzled family of GPCRs that counts a total of ten members (FZD1-10). Nine FZD genes (FZD1-9) have been analyzed in the present study (Figure 1). FZD10 mRNA could not be detected at a significant level in any of the five cell types tested. Among the nine remaining FZD genes, two may be noticed. FZD3 which shows the highest expression in the cells with stem properties and FZD6, the expression of which is almost completely turned down in the GSCs tested.

GPCRs showing expression specificity in cells with stem cell properties

Besides the nine genes with specific changes in TG1 and OB1 cells compared to f-NSCs, astrocytes and U-87 MG cells (Figure 4), other genes show more specific expression in cells with stem properties. Ten genes show expression higher or in the same order of magnitude in f-NSC compared to TG1 and OB1 cells. These include BAI2, CXCR4, F2R, GABBR1 and GPRC5B (Figure 5A) and EDNRB (Figure 5B), BAI3, CHRM4, GPR19, and GPR82 (Figure 5C). Among these genes, CXCR4 appears to be the most specifically expressed in cells with stem properties. Its expression is similar in f-NSCs and TG1 cells, increased in OB1 cells, lower in HA cells and largely downregulated in U-87 MG cells (Figure 5A). Similarly to CXCR4, F2R/PAR1 shows the highest expression in OB1 cells. F2R belongs to the Proteinase activated receptors (PAR) also called Thrombin receptors or Coagulation factor II receptors. Two out of the four PARs, namely FR2 (PAR-1) and F2RL1 (PAR-2) are significantly expressed in four (PAR-2) and all (PAR-1) of the five cell types tested, respectively (Figure 1C). Both F2R/PAR-1 and F2RL1/PAR-2 show overexpression in TG1 and OB1 cells, but F2R/PAR-1 has also increased expression in f-NSCs, whereas F2RL1 (PAR-2) has higher expression in HA cells.

Three genes, CHRM3, GPR171 and P2Y5, although expressed in f-NSCs, show higher expression in TG1 and OB1 cells (Figure 5C).

Three more genes LGR4, CNR1 and GPR51 (GABBR2), although expressed in TG1 cells, appear more specific to f-NSCs and OB1 cells (Figure 5B and C).

Finally, CELSR2 expression is also upregulated in these cells, with the highest expression in OB1 cells (Figure 5B).

GPCRs with higher expression in the cancer cells TG1, OB1 and U-87 MG compared to HA and f-NSCs

We previously showed that GPR56 was overexpressed in TG1, OB1, U-87 MG and f-NSC cells (Figure 2). Further analysis of the GPCR transcriptome indicates that the expression of one GPCR,

CD97, appears more specifically related to the cancer phenotype. Among these cancer cells, CD97 expression appears upregulated in cancer cells with stem cell properties compared to U-87 MG cells.

GPCRs characteristic of a given cell type

Analysis of Figure 1 also points to GPCRs that appear highly expressed in only one of the five cell types (Figure 6). Thus, HA cells are characterized by the specific expression of OXTR and DRD4, f-NSCs by BAI1, GPR37L1 and U-87 MG cells by BDKRB1 and to some extent BDKRB2 (although this GPCR is also expressed at low levels in HA cells).

Proteomic analysis of GPCR expression in human GBM CSCs, f-NSCs, HA and GBM U-87 MG cells

One step further in our investigation was to study GPCR expression at the protein level. Antibodies are often used as tools to investigate protein expression and localization. However the use of antibodies is cumbersome if a great number of proteins have to be analyzed. Moreover, the use of antibodies and especially those targeting GPCRs is often hampered by their lack of specificity. This has been reported for several receptors [41] including $\alpha 1$ adrenergic [42], dopamine [43], beta adrenergic [44], muscarinic receptors [45] and histamine H4 receptors [46]. To overcome the tedious steps needed for generation and validation of reliable antibodies, other approaches are needed. In recent years, high throughput protein profiling, even in complex biological systems, has gained considerable interest thanks to the development of mass spectrometry techniques (MS) coupled to liquid chromatography and bioinformatics.

Thus, in parallel to transcriptomic analysis, we developed a mass spectrometry-based proteomic approach to investigate the expression of GPCRs at the protein level. This task appeared challenging as GPCRs are membrane proteins mostly expressed at low levels. Therefore, a dedicated membrane proteomic approach based on 1D gel protein separation, followed by reduction, alkylation and enzymatic digestion prior to nanoLC-MS/MS analysis of peptide mixtures was developed (see material and methods). With a 1% false discovery rate (FDR), between 3000 and 5000 unique proteins were identified for each cell type (more precisely 4090 for f-NSCs, 3997 for TG1, 3960 for OB1, 3758 for U-87 MG and 3216 for HA cells). This led to more than 7000 non-redundant protein identifications when the results from the five different cell types were combined. Among those proteins, about 20% were annotated as plasma membrane proteins according to their GO annotation: 1043 for f-NSCs, 976 for OB1, 966 for U-87 MG, 963 for TG1 and 837 for HA cells. A final list of seven GPCRs was obtained, among which five (GPR56, LPHN2, F2R/PAR1, GPRC5B and CD97) belong to the fifteen most expressed transcripts (expression level >20000 units) (Table 2). Interestingly, CELSR2 which is expressed at a lower mRNA level (7542 units maximum in OB1 cells) was identified with at least 2 peptides in f-NSCs, TG1 and OB1 cells. A detailed analysis of proteomic results was performed in order to explain why some highly abundant gene transcripts were not detected at the protein level, whereas less abundant ones were. Based on in depth GPCR amino acid sequence analysis, the number of MS detectable peptides was estimated (Table 2). To be



Figure 2. Hierarchical clustering of TG1, OB1, f-NSC, HA and U-87 MG cells based on GPCR expression. The set of 138 GPCRs with Ct values ≤ 31.5 (or expression levels expressed as $2^{-Ct} * 10^{-12} \geq 329$ units) was used for the clustering. For a given cell type, mean expression values (m) from at least two experiments were used for the clustering. Clustering was performed using the DChip software [180]. doi:10.1371/journal.pone.0091519.g002

best detected in MS, peptides should have: i) molecular weights comprised between 1000 and 3000 Da; ii) extracellular or cytoplasmic regions exposed to trypsin digestion and characterized by lower hydrophobicity than trans-membrane domains and iii) no posttranslational modifications. Considering those criteria, our analyses suggest that, grossly, GPCRs with more than 5 potentially detectable peptides and with gene expression levels over 20000 units were observed in the present proteomic study.

To obtain semi-quantitative information on proteins identified by mass spectrometry, we performed spectral and peptide counting analysis. For GPR56, one of the best identified GPCRs by proteomic analysis, spectral counting led to an expression pattern similar to that observed at the transcript level, i.e. expression of GPR56 being higher in f-NSCs than in TG1, OB1 and U-87 MG cells and not detected in HA (Table 2). LPHN2 which shows high transcript levels specifically in OB1 and TG1 cells, exhibits similar results at the protein level (LPHN2 is detected with a high number of peptides and spectra in both TG1 and OB1 cells and is not detected in f-NSCs, U-87 MG cells or in astrocytes) (Table 2). Interestingly, all the MS/MS selected peptides are located in the extracellular domain of LPHN2 even if 34 peptides could have been expected from the cytoplasmic domain. This observation could be consistent with cytoplasmic domain processing which eliminates part of the C-terminal domain of the receptor. CD97 is another GPCR for which proteomics follows the trend of transcriptomics. CD97 transcript was found only in the cancer cells with overexpression in GSCs. At the protein level CD97 appears highly overexpressed in TG1 (9 peptides), OB1 (10 peptides) and U-87 cells (6 peptides) while it was not identified in f-NSCs and HA cells (Table 2). Also, CELSR2 which showed a transcript level in $OB1 > f-NSC > TG1 > HA$ was detected with the same expression hierarchy at the proteomic level with $OB1 > f-NSC > TG1 > HA$ (Table 2). F2R (PAR-1), GPRC5B and BAI2 belong to the group of GPCRs overexpressed at the mRNA level in cells with stem properties. With only five potentially detectable peptides, F2R was identified by proteomic analysis only in OB1 cells that present the highest mRNA level for this GPCR (Table 2). Despite a lower level of transcript and few peptides detectable by MS/MS, GPCR5B was identified in f-NSCs, OB1 and TG1 cells in agreement with the RT-QPCR analysis (Table 2). Surprisingly, BAI2, a 173 kDa protein for which 40 unique peptides could be expected by MS analysis and with a mRNA level similar to that of CD97 (Table 2), was identified by a single peptide in f-NSCs (the cell type which showed the highest transcript level) (Table 2).

Discussion

This paper presents the first exhaustive study of GPCR expression at both the transcriptomic and proteomic levels in GSCs. Analysis of the difference in GPCR expression between several cell types of neural origin highlights potential functioning specificities. Investigating the GPCR proteome was challenging as most of these membrane receptors are expressed at low levels. Besides the expression level, receptor accessibility to proteolytic enzymes, which can be related to the length of the extra- and intracellular domains and post-transcriptional modifications, must

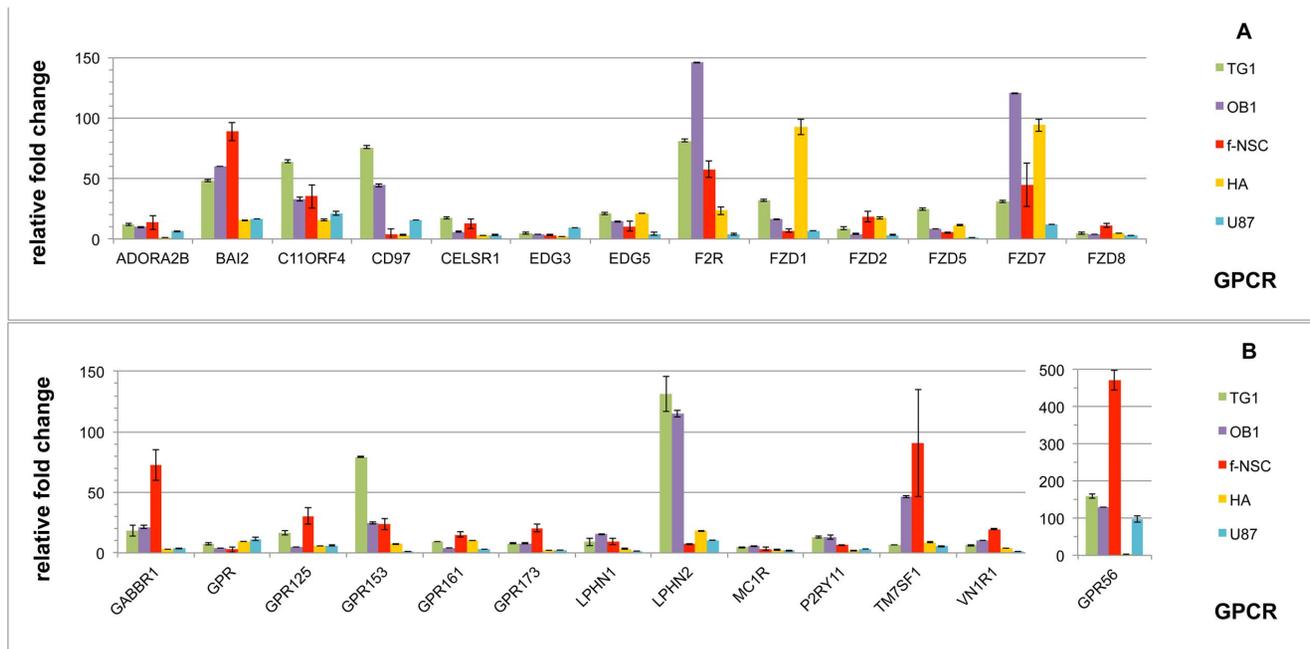


Figure 3. Ensemble of GPCRs expressed in all of the five cell types tested. GPCRs expressed with a cycle threshold (Ct) \leq to 31.5 in GSCs (TG1 and OB1) human fetal neural stem cells (f-NSCs), human astrocytes (HA) and U-87 MG GBM cells (U-87) are represented. Ct = 31.5 (or expression level expressed as $2^{-Ct} * 10^{12} = 329$ units) was taken as the lowest limit for expression. All data were normalized using RPLP0 as housekeeping gene. Ordinate indicates the fold change in expression of a given GPCR relative to the Ct value of 31.5. Relative fold change is calculated as $2^{-(31.5 - Ct_{GPCR})}$, where Ct_{GPCR} corresponds to the cycle threshold of a given GPCR in a given cell type. Values correspond to the mean of a least 2 experiments (from 2 to 4). GPCRs are presented in alphabetical order. Note the ordinate scale change for GPR 56. doi:10.1371/journal.pone.0091519.g003

be considered. Seven GPCRs were detected by MS - based proteomics. For all GPCRs, but BAI2, detection by mass spectrometry correlates with their transcript level and number of potential proteolytic peptides. These results thus demonstrate that MS - based proteomic approaches nowadays allow both identification of proteins in complex mixtures of thousands of proteins and quantification with good confidence. It also reinforces transcriptomic data by showing (for the receptors that could be detected) a good correlation between levels evaluated by RT-Q-PCR and mass spectrometry, i.e. the highest protein levels being observed for the highest mRNA levels. Due to the lack of selectivity of most commercial GPCR - directed antibodies, MS appears as an attractive methodology for GPCR expression profiling.

The different GPCRs presented in the result section, with expression specificity in either GBM cancer cells, GBM cancer cells and f-NSCs, GSCs and GSCs/f-NSC are presented in Table 3 along with their known ligand(s), class, function and chromosome localization. In the discussion we will focus our attention on the most highly expressed genes detected by proteomics and on some GPCRs with specific expression in TG1 and OB1 cells. The relevant GPCRs will be discussed in order to highlight their main known functions and, if documented, their implications in glioma or, more generally, in cancer.

Genes with specific overexpression in GBM cancer cells

One gene, CD97, emerges for its specificity in GBM cancer cells. This GPCR shows significant expression change in the brain cancer cells *vs* non-cancer cells and its expression is highest in cancer cells with stem cell properties (TG1 and OB1). This expression specificity is observed both at the mRNA and protein levels. CD97 is a group II adhesion GPCR, member of the

epidermal growth factor-seven transmembrane (EGF-TM7) family. Aberrant expression of this gene has been reported in various cancers, namely thyroid, gastric, pancreatic, and esophageal carcinomas [47] and expression seems to be related to tumor aggressiveness. Recently, CD97 was shown to be upregulated in three different GBM cell lines and to be one of the targets of the transcription factor WT-1 (Wilms tumor protein), the expression of which is also upregulated in gliomas [48]. CD97 was found to be involved in invasiveness and metastasis of prostate cancer, (where it forms heterodimers with the lysophosphatidic acid receptor LPAR1 [49,50]) and more recently of GBMs [48,51]. Its overexpression is correlated with decreased survival in patients [51]. The present study clearly shows the expression specificity of this receptor in GBM cancer cells compared to non-cancerous neural cells. Further studies need to be performed to establish which of the possible splice variants [52] are upregulated. So far there are no reports on the physiological role of CD97 in brain. The low expression of this receptor in astrocytes and f-NSCs compared to cancer cells places it as a good candidate for further investigation of its role in cancer in general and GSC cells in particular.

Genes expressed in GBM cancer cells with and without stem properties and in f-NSCs

GPR56, another adhesion GPCR belonging to class VIII is expressed in GBM cells, but its expression is even higher in f-NSCs. GPR56 is known for its expression in cancer cells, including gliomas [53]. In several tumors (melanoma, prostate cancer) [54,55,56], the level of GPR56 expression appears to be inversely correlated to malignancy and metastatic potential. Physiologically, GPR56 plays a key role in brain development, being involved in cortical development and radial migration of neural progenitors

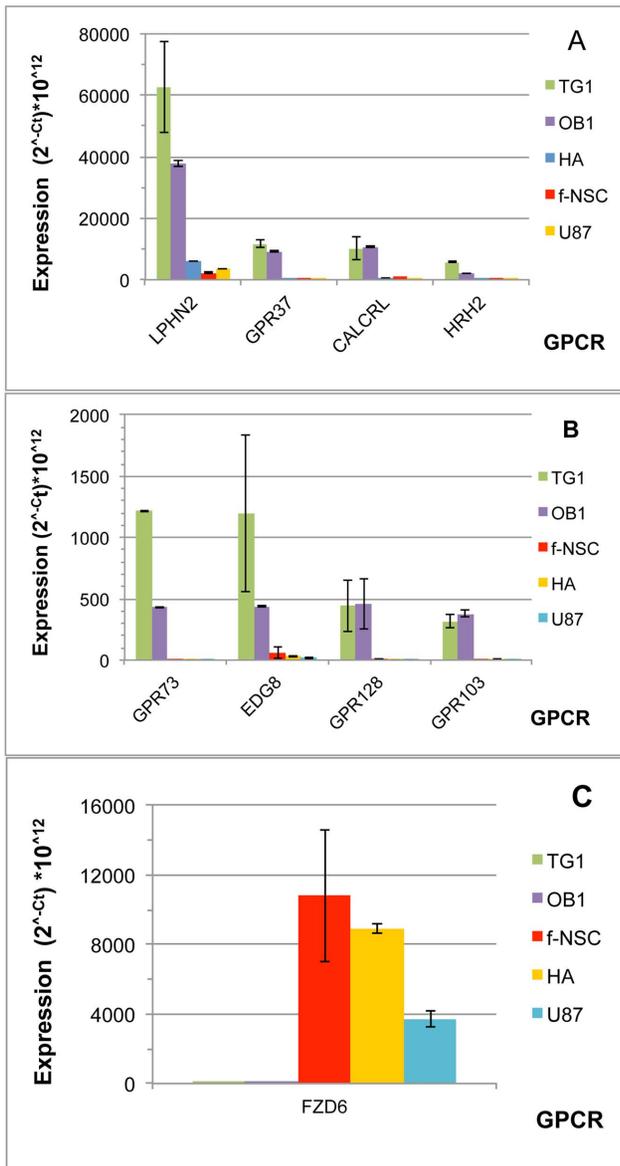


Figure 4. GPCRs preferentially expressed in TG1 and OB1 GSCs. GPCR expression was determined by RT-QPCR. Expression levels are presented as $2^{-Ct} * 10^{12}$, where Ct corresponds to the cycle threshold of a given GPCR. A) and B) present GPCRs overexpressed in TG1 and OB1 cells and GPCRs are presented as a function of decreasing expression level in TG1 cells; C) presents the GPCR specifically underexpressed in TG1 and OB1 GSCs. Data correspond to the mean of 4 experiments. doi:10.1371/journal.pone.0091519.g004

(NPC/NSC) from the lateral ventricles to the cortical plate. Collagen has been reported to be its major ligand for this function [57]. GPR56 has been shown to be highly expressed in nestin-positive neural stem/progenitor cells in the ventricular/subventricular zone of human and mouse fetal brains as well as in cultured neurospheres derived from both human and mouse fetal brains [58] [59]. Its expression is downregulated during differentiation [59].

The high expression of GPR56 in f-NSCs and its low expression in astrocytes are in agreement with the pre-cited studies performed on neural stem/progenitor cells. However the present study which compares the expression of GPR56 in GBM cells and f-NSCs

shows for the first time that GPR56 can be expressed at the same and even higher level in normal (non-cancerous) stem cells than in cancer cells.

GPCRs showing expression specificity in cells with stem cells properties TG1, OB1 and f-NSC cells

At the transcriptomic level, 17 GPCRs show specific expression in cells with stem properties (Figure 5A–C; Table 3). Four of them, BAI2, F2R, GPCR5B and CELSR2 are detected at the protein level. Information relative to the role of these receptors in stem cells is limited and their role in CSCs remains to be explored.

BAI-2 is a member of the brain angiogenesis inhibitor family (BAIs) which belongs to group VII of adhesion GPCRs and counts 3 members, BAI1 to 3, located on chromosomes 8, 1 and 6 respectively. BAIs are mainly expressed in brain. This group exhibits an N-terminal CUB (for complement C1r/C1s, Uegf, Bmp1) domain, 3 to 5 thrombospondin type 1 repeats (TSR) discovered to regulate the anti-angiogenic activity of thrombospondin-1, one hormone-binding domain, the evolutionary conserved GPCR proteolytic site (GPS) and multiple glycosylation sites. The three receptors are involved in neo-vascularization and have been reported to be involved in apoptosis and regulation of tumor progression (for review see [60]). The antiangiogenic activity of BAI1 in normal and pathological states is well documented [61,62,63]. It is triggered by vasculostatin-40, a 40 kDa fragment induced by matrix metalloprotease cleavage [63]. Tumor-suppressor activity has thus been attributed to BAI1 and its use in cancer gene therapy has been proposed [62]. The angiostatic effect of BAI2 (and BAI3) is less documented. Activation through N-terminal cleavage has been reported for BAI2 [64]. Concerning a role of this receptor family in high grade gliomas, data are scarce and concern global levels of expression measured on tissue samples composed of a mixture of cells [65]. The present study which gives a comparative overview of the transcription expression pattern in five different cell types shows that f-NSCs express BAI1-3 at the highest level (Figure 1). BAI2-3 are overexpressed in cells with stem properties compared to more differentiated cells. Such observations have not yet been reported in the literature. Only the lack of expression of BAI1 in various GBM cell lines, including the U-87 MG cells has been documented [61], which is in agreement with the results of the present study (Figure 1). The different members of the family and namely BAI2 and BAI3 may be involved in functions different from neovascularization, related to stemness maintenance or regulation of apoptosis. Further experiments are necessary to uncover the role of these receptors in GBM and in cancer, more generally. Of the three BAIs, the transcription level of BAI2 is highest. Despite its high molecular weight and high level of mRNA in f-NSCs, TG1 and OB1 cells, BAI2 is only scarcely detected (one peptide) at the proteomic level in f-NSCs (Table 2). This raises further questions concerning the expression of the gene at the protein level or receptor processing in the stem-like cancer cells.

F2R is a second GPCR detected by mass spectrometry with expression specificity in cells with stem properties. F2R belongs to the Proteinase activated receptors (PAR) also called Thrombin receptors or Coagulation factor II receptors which count four members. Two PAR receptors, namely FR2 (PAR-1) and F2RL1 (PAR-2) are significantly expressed in, respectively, four (PAR-2) and all (PAR-1) of the five cell types tested (Figure 1). Both F2R/PAR-1 and F2RL1/PAR2 show overexpression in TG1 and OB1 cells. F2R, but not F2RL1 is also overexpressed in f-NSCs. PAR receptors have plural effects on neural cells. PAR receptors and namely F2R/PAR-1 and F2RL1/PAR-2 are expressed in most tumors (for review see [66,67]) and are known for their

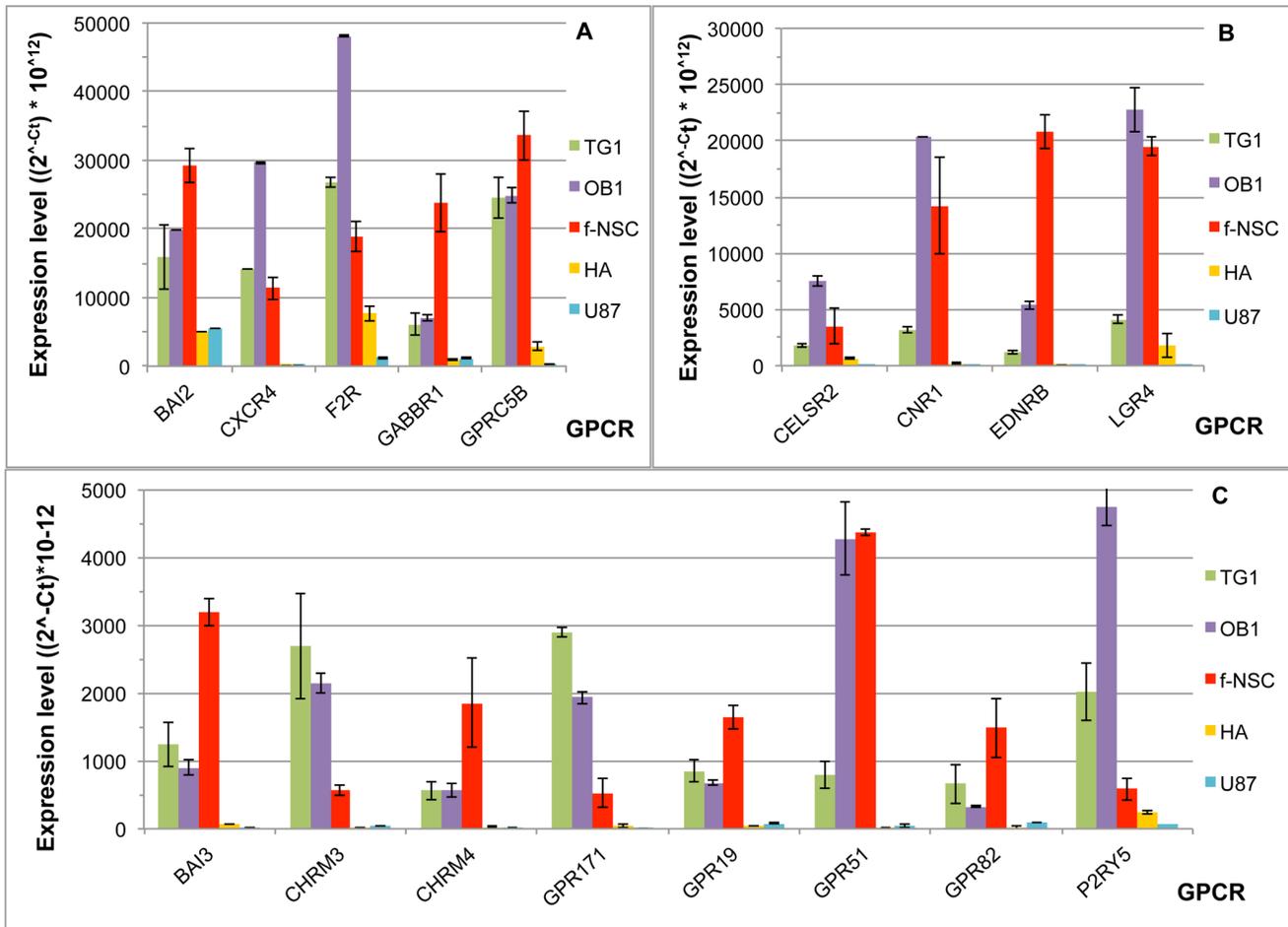


Figure 5. GPCRs overexpressed in the GSCs TG1 and OB1 and the fetal neural stem cells (f-NSCs). The ordinate represents the expression level of the different GPCRs shown as $(2^{-Ct} * 10^{12})$ where Ct corresponds to the cycle threshold of a given GPCR. Note the difference in scale between A, B and C. HA corresponds to human astrocytes and U-87 to the GBM cell line U-87 MG. doi:10.1371/journal.pone.0091519.g005

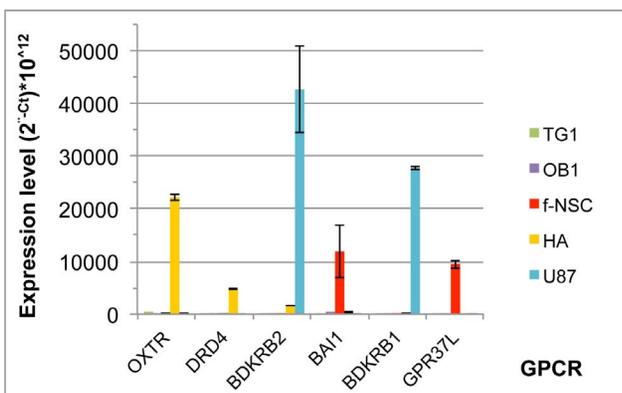


Figure 6. GPCRs specifically expressed in one of the five cell types. Expression of the given GPCR is compared between TG1 and OB1 GSCs, fetal neural stem cells (f-NSCs), human astrocytes (HA) and the GBM cell line U-87 MG. The ordinate represents the expression level of the different GPCRs shown as $(2^{-Ct} * 10^{12})$ where Ct corresponds to the cycle threshold of a given GPCR. Note the difference in scale between A) and B). TG1 corresponds to GSCs, HA to human astrocytes and U-87 to the GBM cell line U-87 MG. doi:10.1371/journal.pone.0091519.g006

cooperation with EGFR signaling [68]. F2RL1/PAR-2 triggers proliferation, migration, invasion of GBM cells and activation of downstream MAPKs including ERKs after stimulation by the HIF-inducible tissue factor (TF) [69,70]. F2R/PAR-1, the principal thrombin-activated receptor involved in platelet aggregation and endothelial cell proliferation, has been shown to be overexpressed in various invasive and metastatic tumors and its expression levels directly correlate with the degree of invasiveness of the cancer. A link between angiogenesis and EGFR (via the ligand independent EGFRvIII mutant), tissue factor (TF) and PAR receptors has recently been established in GBM cell lines [71]. In melanoma, cell signaling through PAR releases several factors, such as cytokines, proteases, adhesion molecules and growth factors causing changes in the tumor microenvironment [72]. This has not been shown so far for GBMs. PAR receptors probably play a major role in GBM, and even more generally in cancer, linking the tumor microenvironment, hypoxia and angiogenesis to tumor cell proliferation, migration and tissue invasion. However, PARs also play a major role in normal cell functioning and are expressed, as shown in the present study, in astrocytes and fetal neural stem cells. The overexpression of these receptors in cancer stem-like cells has not been reported so far. F2R is detected by mass spectrometry in OB1 cells that have the highest mRNA level for this receptor (Table 2).

Table 2. GPCRs detected by proteomic analysis and comparison to those most highly expressed at the transcriptomic level.

GPCR Name	Access number	Molecular weight (kDa)	Nb of expected unique peptides	OB1			TG1			f-NSC			HA			U87		
				Transcript level ($2^{-Ct} * 10^{12}$)	Nb of unique peptide (nb of spectrum)	Nb of unique Transcript peptide (nb level of spectrum) ($2^{-Ct} * 10^{12}$)	Transcript level ($2^{-Ct} * 10^{12}$)	Nb of unique peptide (nb of spectrum)	Nb of unique Transcript peptide (nb level of spectrum) ($2^{-Ct} * 10^{12}$)	Transcript level ($2^{-Ct} * 10^{12}$)	Nb of unique peptide (nb of spectrum)	Nb of unique Transcript peptide (nb level of spectrum) ($2^{-Ct} * 10^{12}$)	Transcript level ($2^{-Ct} * 10^{12}$)	Nb of unique peptide (nb of spectrum)	Nb of unique Transcript peptide (nb level of spectrum) ($2^{-Ct} * 10^{12}$)			
<i>GRP56</i>	Q91653	78	14	42820	4 (8)	52347	4 (10)	155121	8 (37)	625	32369	3 (6)						
<i>LPHN2</i>	O95490	163	50	37915	9 (15)	62684	12 (24)	2338	nd	5921	3495	nd						
<i>F2R</i>	P25116	44	5	48045	2 (6)	26751	nd	18912	nd	7701	1259	nd						
<i>BDKRB2</i>	P30411	44	0	78	nd	0	nd	1	nd	1655	42664	nd						
<i>FZD7</i>	O75084	61	3	39613	nd	10182	nd	14788	nd	30943	4011	nd						
<i>GPRC5B</i>	Q9NZH0	41	3	24853	1 (3)	24463	1 (4)	33650	3 (7)	2888	248	nd						
<i>TIM7SF1</i>	O60478	45	3	15303	nd	3322	nd	29968	nd	2855	1743	nd						
<i>CXCR4</i>	P61073	40	2	29644	nd	14193	nd	11354	nd	120	19	nd						
<i>BAL2</i>	O60241	173	40	19770	nd	15863	nd	29243	1 (1)	5050	5478	nd						
<i>BDKRB1</i>	P46663	40	2	5	nd	0	nd	9	nd	65	27717	nd						
<i>GPR153</i>	Q6NV75	65	1	8111	nd	26080	nd	7762	nd	2438	342	nd						
<i>CD97</i>	P48960	90	19	14593	10 (40)	24958	9 (31)	405	nd	1079	5233	6 (21)						
<i>LGR4</i>	Q8N537	104	8	22726	nd	4135	nd	19526	nd	1804	51	nd						
<i>CNR1</i>	P21554	53	3	20381	nd	3161	nd	14232	nd	190	1	nd						
<i>CELSR2</i>	Q9HCU4	317	79	7543	2 (2)	1769	2 (2)	3501	8 (12)	601	118	nd						

List of the GPCRs with the highest transcription level in at least one of the five cell types (expressed as $2^{-Ct} * 10^{12}$ values >20000) and their identification by proteomic analysis. GPCRs are listed from top to bottom as a function of their mRNA level (the most highly expressed in one or several cell types appears at the top). Number of expected unique peptides corresponds to the unique peptides potentially detectable by MS/MS analysis. Expected peptides were calculated after *in silico* trypsin digestion of a given GPCR taking into account different criteria: i) no miscleavage, ii) molecular weights comprised between 1000 and 3000 Da, iii) presence in extracellular or cytoplasmic regions and iv) no posttranslational modifications. GPCRs identified by proteomic analysis are indicated in italics. OB1 and TG1 correspond to the GSCs. f-NSC, HA and U-87 stand for fetal neural stem cells, human astrocytes and the U-87 MG GBM cell line, respectively. Ct = cycle threshold; nd = not detected.
doi:10.1371/journal.pone.0091519.t002

Table 3. Summary of the GPCRs expressed with the highest specificity in different subgroups of cells comprising at least the GSCs TG1 and OB1.

Subgroup	Function	Ligand	Class	Chromosomal localisation
TOU				
CD97	tumor angiogenesis, cell migration, EMT	CD55	B Adhesion/EGF-like family	19p13
NTOU				
GPR56	brain development; tumor metastasis; cell plasticity	collagen III	B Adhesion	16q13
TO				
CALCRL/CRLR	angiogenesis, cell proliferation, antiapoptotic, resistance to hypoxia	adrenomedullin, CGRP	B Calcitonin group	2q32.1
EDG8/S1P5	cell migration/proliferation	lysophosphatidic acid, sphingosine 1-phosphate	A	19p13.2
FZD6	tumor suppressor	Wnt	B Frizzled	8q22.3-q23.2
GPR103	pain. RF-amide Receptor	QRFP family peptides	A	4q27
GPR128	orphan no report	orphan	B Adhesion	3q12.2
GPR37	macroautophagy, glio and neuronal protection	prosaposin/Prosapeptide?	A	7q31
GPR73/PKR1	angiogenesis, cell proliferation, involved in pain such as GPR103	prokinectin-1 and 2	A	2p13.1
HRH2	angiogenesis, cell proliferation	histamine	A	5q35.2
LPHN2	EMT in heart development	lectomedin	B Adhesion	1p31.1
NTO				
BAI2	unclear (neovascularization, apoptosis)	orphan	B Adhesion	1p25.2
BAI3	synapse formation, dendritic morphogenesis	C1q-like proteins (C1q1-4) ?	B Adhesion	6q12
CELSR2	planar cell polarity	Cadherin domains - homophilic binding	B Adhesion	11p13.3
CHRM3	Cell proliferation, differentiation, survival. Link with EGFR activation. Implicated in cancer	acetylcholine	A	1q43
CHRM4	Cell proliferation, differentiation, survival. Implicated in cancer	acetylcholine	A	11p11.2
CNR1	energy metabolism; glucose oxidation, ketogenesis, apoptosis, autophagy; stem gene expression	anandamide, 2-arachidonyl-glycerol	A	6q14-q15
CXCR4	cell migration, angiogenesis, invasiveness hypoxia controlled, member of the CRG (cooperation regulatory genes).	CXCL12	A	2q22.1
EDNRB	self renewal, migration, proapoptotic	endothelins 1 to 3	A	13q22.3
F2R/PAR-1	Various pathophysiological processes. Participates in EGFR signalling and thus involved in carcinoma. HIF inducible.	proteolytic cleavage of N-terminus	A	5q13.3
GABBR1	cell proliferation	GABA	C Metabotropic GABA(B)R	6p22.1
GABBR2 (GPR51)	cell proliferation	GABA	C Metabotropic GABA(B)R	9q22.1-22.3
GPR19	links with stemness	orphan	A	12p12.3
GPR82	orphan, food intake, not essential	orphan	A	Xp11.4
GPR171		orphan	A	3q25.1
GPRC5B/Retinoic acid-induced gene 2 protein2	Unknown; retinoic acid induced	orphan	C Glutamate metabotropic	16p12.3
LGR4	stem cell proliferation/differentiation; Invasiveness metastasis	R-spondin	A	11p14-p13
P2RY5/LPAR2	Spars information. Involved in cancer	oleoyl-L-alpha-lysophosphatidic acid	A	13q14.2

GPCRs with expression specificity in the different subgroups of cells are listed together with i) their reported function in cancer or their physiological implications (if known), ii) their ligand(s), iii) the GPCR class they belong to and iv) their chromosome localization. T, O, U, N, stands for TG1, OB1, U-87 -MG and f-NSC, respectively. GPCRs in TO thus stands for the receptors specifically expressed in TG1 and OB1 GSCs compared to their expression in the three other cell types; TOU for those expressed in TG1, OB1 and U-87 MG and so on.

doi:10.1371/journal.pone.0091519.t003

GPRC5B (or RAIG-2), a retinoic acid-inducible orphan GPCR, is a member of the group C of metabotropic glutamate receptor family proteins with short N-termini. It is expressed in human brain [35,73,74]. Studies performed on *Xenopus* suggest a signaling through the non-canonical Wnt/frizzled pathway and a possible cross-talk between the two pathways [75]. This receptor may play an important role in brain maturation and neural circuitry building during brain development. Expression variants, due to alternative splicing have been reported in mouse brain [76]. GPRC5B-v1 is ubiquitously expressed, whereas GPRC5B-v2 is more specific to brain. Nevertheless, besides its implication in behavioral abnormalities, information concerning this receptor family is limited. GPRC5B is almost absent from U-87 MG cells. Its mRNA level is highest in f-NSCs, TG1 and OB1 cells with f-NSC>OB1~TG1 (Figure 1). Interestingly, mass spectrometry confirms its presence at the protein level with the same relative abundance in the three cell types.

CELSR2 (Cadherin EFG LAG seven-pass G-type receptor 2) is a non-classical cadherin adhesion receptor (group IV of adhesion GPCRs), member of the flamingo subfamily (fmi) characterized by the presence of a seven transmembrane domain [77]. The three members of the family, CELSR1-3, have multiple functions in epithelial and nervous systems and are among the core planar cell polarity (PCP) receptors [78,79,80]. A role in maintaining quiescence of long term hematopoietic stem cells has been reported [81]. They play a pivotal role in contact mediated communication. PCP regulation appears to take place through cooperation with the Frizzled receptors and the non-canonical Wnt pathway [81,82]. Aberrant regulation of these signaling pathways has been suggested to be involved in cancer invasion and metastasis [83,84]. Although expressed in HA cells, CELSR2 mRNA is overexpressed in cells with stem properties with levels in OB1>f-NSC>TG1. The same order of expression is also found at the protein level. The overexpression of this receptor in CSCs deserves further investigation.

Several other highly expressed GPCRs, namely CXCR4, GABA(B)Rs, LGR4 and CNR1, although not detected by mass spectroscopy, will be discussed.

CXCR4 presents the same relative expression as F2R/PAR1 at the mRNA level (Figure 4A) and similarities in functional properties. CXCR4 is a member of the chemokine receptor family and its endogenous ligand CXCL12 has been implicated in tumor development, including gliomas, facilitating angiogenesis, infiltration and metastasis through modulation of VEGF signaling, PAI-1 (plasminogen activator inhibitor-1) production and membrane type-2 matrix metalloproteinase expression (reviewed in [85,86]). In agreement with the results of the present study, overexpression of CXCR4 in GBMs has been recognized as an attribute of cells with progenitor/stem properties and CXCL12 promotes specifically the proliferation of these cells compared to more differentiated cancer cells [87,88]. Of note, the micro-RNA cluster miR 302–367 has recently been shown to inhibit the GSCs'self-renewal and infiltrative properties through CXCR4 repression [89]. Recent studies also show a role of the CXCR4/CXCL12 axis in triggering interaction of cancer cells with microvascular endothelial cells and GBM cancer stem cell transdifferentiation into pericytes [90,91]. The prominent role of this axis in tumor growth and cancer stem-like cells' interaction with their microenvironment and the perivascular space, designed CXCR4 as a potential target for adjuvant chemotherapies. So far the CXCR4 antagonist BKT140 is in phase I/IIA of a clinical trial for multiple myeloma (<http://www.clinicaltrials.gov> NCT01010880; for review see [86]). Expression of CXCR4 in neural stem cells has also recently been reported [92]. The present

study shows that the expression level of CXCR4 in f-NSCs can reach that observed in GSCs. Using antibodies, CXCR4 was found to be expressed in U-87 GBM cells. Our study indicates low expression of the receptor at the mRNA level in U-87 MG cells (below the cutoff level used in the study (Table S1)). Despite its high transcription level in f-NSCs, TG1 and OB1 cells, the protein could not be detected by mass spectrometry.

GABBR1 and GABBR2 (GPR51) which belong to the metabotropic GABA(B) receptors (class C GPCRs) present a relatively high expression in f-NSCs (Figures 5A and 5C). These two receptors are largely distributed in the central nervous system (CNS) and are particularly present on neurons where they play a role in neurotransmission both at the presynaptic and postsynaptic level. The two receptors function by forming heteromeric complexes of GABA(B) R1/R2 subunits [93,94,95,96], but homodimers [97], higher order oligomers [98] and complexes with other GPCRs have also been reported (e.g. the closely related extracellular calcium sensing receptor CaR [99], the muscarinic receptor M2 [100], as well as with the non GPCR GABA A receptors [Balasubramanian, 2004 #242]. GABA(A)R and GABA(B)R have been shown to exert opposite effects on rodent NSCs. GABA(A)R activation restricts NSC proliferation [101,102], whereas GABA(B)R activation promotes in vitro neurospheres' development [103]. Besides neurons, GABA(B) R1 and R2 are also present in astrocytes [104] and cells outside the CNS [105,106,107]. GABA(B) receptors have been involved in various cancers and the GABA(B) receptor agonist baclofen was shown to reduce cancer development in various animal models [108,109,110,111]. However, GABA signaling remains extremely complex and the multiple possible interaction of GABA(B) subunits with other partners, as well as the regulatory role played by the GABA(B)R2 in receptor trafficking [112] renders assessing a clear cut role of these receptors in cancer difficult. The high expression of the two GABA(B) receptors in f-NSCs may suggest a role in stem cells.

Similarly to GABBR2 (GPR51), LGR4 and CNR1 present an increased mRNA level in f-NSCs and OB1 cells.

LGR4/GPR48, a class A receptor, belongs to the family of leucine-rich repeat (LRR)-containing G protein-coupled receptors (or LGRs) composed of six members. LGR1-3 are hormone receptors. LGR4-6 were orphanized in 2011 when De Lau and Carmon [113,114], working on intestinal crypt cells, showed that these receptors bind R-spondin, a molecule secreted by the crypt cells and known as a positive regulator of Wnt signaling. In the study of Carmon, LGR4 and 5 were also shown to interact with the Wnt-binding LRP/Frizzled receptor complex. LGR4-6 have been found in various tissues [115,116,117,118], where they regulate stem cell proliferation and differentiation. Inactivation of the LGR4 coding gene decreases cell proliferation and strongly reduces terminal differentiation of Paneth cells in postnatal mouse intestinal crypts [119]. Physiologically, LGR4 appears to play a key role in organism and tissue development [100,120,121,122]. Expression of LGR4 was also associated with invasiveness and metastatic activity of colon and cervical cancer cells [123,124]. So far no information is available on LGR4 expression in gliomas. Of note, in our study, among the three R-spondin binding receptors, only LGR4 is expressed at a significant level in TG1 and OB1 cells.

CNR1, which encodes the CB1 cannabinoid receptor (CB1), exhibits an expression profile comparable to LGR4 (Figure 5B). CB1 has a widespread distribution in various brain regions [125] and has two known endogenous ligands (anandamide and 2-arachidonoyl-glycerol). Several genetic studies have been performed to link CNR1 mutations to psychiatric but also metabolic disorders

[126,127,128]. In rat astrocytes CB1 plays pivotal roles in energy metabolism and activation of this receptor increases the rate of glucose oxidation and ketogenesis, two mechanisms involved in energy supply of the CNS (for review see [125]). CNR1 expression has been reported in various glioma cell lines and receptor agonists trigger apoptosis and autophagy of these cells [129,130]. However cannabinoids' effect is complex and relies on both ligand concentrations and receptor levels [131]. In glioma stem like cells, CB1 agonists have been reported to modulate the expression of stem genes [132]. In human glioma tissues, increased expression of cannabinoid receptors (CB1 and CB2) has been reported to correlate with higher tumor grades [133]; but this remains controversial [134,135,136]. In other tumors, a role of CNR1 in invasion and metastasis has been reported [137,138] and the endocannabinoid system has been reported as a promising tool to improve the efficacy of steroids in colon cancer [139]. Globally, the cannabinoid system appears to present anti-proliferative, anti-angiogenic and pro-apoptotic properties. Nevertheless, its potential use for treating glioma and may be cancer more generally needs further analysis and is hindered by side effects such as anxiety and depression.

Several less abundant GPCRs of the subgroup of genes expressed in cells with stem properties (Table 3), show specific upregulation in GSCs. These include CHRM3, GPR171 and P2RY5. As for F2R/PAR-1 and CXCR4, CHRM3 activation may synergistically act on the EGFR signaling pathway. Much less is known for P2RY5 and GPR171 and their relation to cancer.

GPCRs specifically expressed/overexpressed in TG1 and OB1 GSCs

At the transcriptomic level, eight GPCRs appeared specifically upregulated and one downregulated in TG1 and OB1 GSCs (Figure 4 and Table 3) compared to their expression in HA or fNSCs.

LPHN2 shows the highest expression among the different GPCRs of this group. The specific upregulation of LPHN2 at the mRNA level is corroborated at the protein level and both transcriptomic and proteomic data are in agreement with a level in TG1 > OB1. LPHN2 is an adhesion GPCR (group I) belonging to the family of latrophilins composed of three members (LPHN1-3) (for review see [140]). Expression of latrophilin 1 and 3 is mainly found in the central nervous system, whereas LPHN2 is more ubiquitous [140,141,142,143]. LPHN2 shares with the two other members a large N-terminus composed of lectin, olfactomedin and hormone binding domains. The receptors bind α -latrotoxin, the black widow spider venom and this binding triggers strong exocytosis at neuron synapses. LPHN2 however shows lower binding affinities than the LPHN1 and LPHN3 isoforms. Not much is known about the functional role of these receptors in vertebrates. Through their large N-terminus, latrophilins may establish interactions with the extracellular matrix. Such interactions have been reported recently for LPHN1 [144,145] and LPHN3 [146]. Information on the function of LPHN2 is scarce. It has been reported to play a role in the epithelial mesenchymal transition (EMT) necessary for heart valve development during embryogenesis [143]. In addition, one report indicates a higher expression of the gene in breast tumors [147]. More recently, in a genome wide association study approach aimed at identifying genes involved in the response of tumor cells to paclitaxel, a microtubule cytotoxic used in chemotherapies, a SNP located in the intronic part of the LPHN2 coding gene was identified as being associated with the cellular response to this compound [148]. However there is no further information as to how an intronic SNP in LPHN2 may interfere with the cellular effect of

paclitaxel. In the current state of knowledge, one may speculate that the overexpression of LPHN2 in TG1 and OB1 GSCs may be related to EMT, the developmental program often activated during cancer metastasis and regulating the equilibrium between CSCs and the non-stem cancer cells [149]. A recent study using in silico analysis and aiming at finding genes under p53 control, designed LPHN2 as a novel target gene of p53 [150]. LPHN1 and LPHN3 are also upregulated in TG1 and OB1 cells, but both receptors appear less specific than LPHN2 and we did not detect them by mass spectrometry.

Other receptors, namely the parkin-associated endothelin B-like receptor (GPR37 or PAEL-R), the calcitonin receptor-like receptor (CALCRL or CRL), the histamine receptor HRH2, the prokinectin receptor (GPR73 or PKR1), the endothelial differentiation G-protein coupled receptor 8 (EDG8 or S1P5), GPR128 and GPR103, show specific increase in their mRNA levels in TG1 and OB1 GSCs (Figure 4 and Table 3). However their expression is too low to be detected by mass spectrometry.

Concerning the histamine receptors, only changes in HRH2 could be observed in the present study. HRH1, H3 and H4 were expressed below the threshold level used in the study. HRH2 has a large tissue distribution and is involved in a plethora of physiological functions as reviewed in [151,152]. As other members of the histamine receptor family [153] it has been reported to play a role in cell proliferation, differentiation energy metabolism and thus to be potentially involved in cancer. HRH2 are found in human mammary and gastric carcinoma cells and several human melanoma cell lines and the HRH2 antagonist cimetidine appears to be endowed with anticancer properties, although its action is not exclusively mediated through HRH2 [154,155]. Clinical trials using HRH2 antagonists as adjuvant treatment in resected colorectal cancer have also been reported ([156]; for review see [157]), but their real benefits need to be assessed. One additional report points to an antiproliferative activity of histamine mediated through HRH2 in pancreatic cells (PANC-1), with a G0/G1-phase arrest and a modulation of the Bcl-2 family proteins [158]. In our hands, two HRH2 agonists (anthamine and dimaprit) and five antagonists (cimetidine, zolandine, ranitidine, thiotidine and famotidine) tested at 5 and 50 μ M for 24 hours did not show a significant effect on TG1 GSCs' proliferation and survival. The specific overexpression of this histamine receptor subtype in TG1 and OB1 GSCs suggests its potential involvement in the cells' pathophysiology that needs further and deeper investigation.

The function of the calcitonin receptor-like receptor **CALCRL** depends on the sub-type of RAMP (receptor-activity modulating protein) to which it is associated. It functions either as a receptor for calcitonin-gene-related peptide (CGRP) when associated to RAMP1 or for adrenomedullin (AM) when associated to RAMP2 or RAMP3 [159]. AM has vasorelaxation properties and its expression is controlled by the hypoxia inducible factor (HIF). Cells overexpressing AM exhibit higher resistance to hypoxia induced apoptosis [160]. AM induces overexpression of anti-apoptotic factors and decreases expression of pro-apoptotic factors [160,161]. The location of AM close to necrotic zones in tumors suggests a major role of this peptide in the resistance of tumor cells to hypoxia. Numerous data are available on AM and its expression in tumors. However data on its receptor are scarce. A recent work shows that CALCRL and its partners RAMP2 and RAMP3 are present at the mRNA and protein levels in GBM tumors isolated from several patients [162]. Moreover, systemic injections of AM receptor antibodies in animals were shown to inhibit both tumor growth and neo-vascularization of U87 xenografted cells in a dose dependent manner [163]. Nevertheless, information on the

distribution, regulation and function of endogenous AM receptors remains poor. Overall, data from the literature suggest a role of AM and its receptor in the survival of tumor cells within the hypoxic regions of solid tumors. AM may induce neovascularization and promote tumor growth. However, in the absence of functional data, it remains difficult to assess the exact relationship between AM, hypoxia, cell proliferation, cell survival and CALCR.

Concerning EDG8/S1P5, the role of the ligand S1P (sphingosine 1-phosphate) in cancer has recently been reviewed [164]. For the other receptors of this subgroup, information relative to their implication in cancer is rare (one report for GPR37 [165]) or lacking.

Concluding Remarks

GPCRs, which mediate extracellular to intracellular signaling, are key receptors in cell functioning in both physiological and disease conditions. Over the years, a great number of GPCRs have successfully been developed as pharmacological targets in various diseases. However, the complexity of GPCR signaling offered by the possible formation of homo- or hetero oligomers, interactions with a wealth of other proteins, including growth factor receptors [166,167,168,169] and downstream effectors, and its regulatory role in several pathways including the Hippo pathway which coordinates tissue homeostasis [170], unveils bewildering opportunities to target new diseases and fuel drug discovery [31]. Oncology is one of the sectors that may benefit from these new developments as deregulation of GPCR-dependent cell signaling sustains and may even drive tumorigenesis [32,35,36]. Exploration of the comparative expression of 356 GPCRs (among which 138 were retained in the present study) in GSCs and in non-cancer neural cells with stem cell properties (human fetal neural stem cells) or without stem cell properties (human astrocytes) and to the established GBM U-87 MG cell line, unveils differences in GPCR expression between the five cell types of neural origin. Several genes, including GPR37, GPR73 (PKR1), GPR103, GPR128, EDG8 (S1P5) and HRH2 appear as the most specifically enriched in TG1 and OB1 cells. LPHN2, which is expressed in the five cell types, is predominantly upregulated in TG1 and OB1 cells. LPHN2 is the only GPCR of the group detected by mass spectrometry. The specific under-expression of FZD6 in TG1 and OB1 cells is noteworthy as FZD6 is the only GPCR showing such a profile. The GPCRs emerging from our analysis are related to metabolic functions, pro-angiogenic activities, cell proliferation, resistance to hypoxia and interactions with the microenvironment (Table 3), an ensemble of characteristics which endow these cancer cells with properties allowing them to be maintained and develop in the specific microenvironment created by the tumor and which may induce death of the normal neural cells. GSCs may originate from differentiated cells that acquired stem properties or from stem/immature progenitor cells that lost their normal regulatory capacities to become cancerous. Whatever the origin, the GSCs may have acquired or retained gene expression found in normal stem/progenitor cells. In this GPCR transcriptome study, several GPCRs show a differential expression pattern in f-NSCs, TG1 and OB1 cells compared to HA or U-87 MG cells. Some GPCRs are already known for their involvement in GBMs. Others have been reported for their involvement in different cancers, but not in brain cancers. To our knowledge, this is the first exhaustive report presenting comparable expression profiles of the whole set of GPCRs in cells from GBM tumors and non-cancerous tissues at the transcriptomic and proteomic levels. Not many GPCRs have so far been reported for their role in GSCs. For those genes that are already documented, our findings mostly correlate

with the reported studies obtained using other techniques such as use of agonists and antagonists, or knockout/knockdown experiments. GPCRs specific to GSCs or U-87 MG cells constitute the most appealing ones for cancer therapy. In addition to changes in the level of expression, genomic modifications with mutations that can alter the integrity of the receptor, namely its structure, interaction with partner proteins, affinity for its ligand(s) and downstream signaling have to be considered [171]. Protein altering mutations affecting several GPCRs as well as G protein subunits, have been reported in a study of 441 breast, lung, ovarian, pancreatic, prostate cell types and subtypes [172]. Targeting the finely tuned-GPCR dependent signaling with biased agonists or antagonists or allosteric effectors [173] selected to act on the identified deregulated pathways opens new avenues for more efficient and personalized multi-target oriented chemotherapies. Getting deeper insight into tumor pathophysiology, identifying cancer subtypes and developing system biology approaches to simulate cell intra- and intercellular communication networks will pave the route for disruptive innovation in oncology. The present descriptive presentation of GPCRs exhibiting specific expression changes in GSCs is a first step that calls for future studies at the functional level in order to decipher the precise role of these receptors in sustaining the cancerous phenotype.

Materials and Methods

Ethics statement

The biomedical research was conducted according to the declaration of Helsinki, to the French laws and was approved by the institutional review board of Ste Anne Hospital, Paris. Patients have given written informed consent. Isolation and characterization of neural stem cells from human fetal brain at embryonic day 50–55 (Carnegie stage 19–22) were performed under ethical approval from the University Paris-Descartes internal review board using tissue donated with written informed consent after elective termination of pregnancy.

Cell Cultures

Patients' derived cells were isolated from surgical resections of human primary GBMs provided by the Neurosurgery department of Sainte Anne Hospital in Paris. Patients' tumors were diagnosed as GBMs according to the World Health Organization (WHO) classification system [174] and as malignant glio-neuronal tumor cells according to the Sainte Anne's classification and grading system [175]. Malignant glio-neuronal tumor cells (TG1 and OB1 cells) with stem cell properties were isolated from those biopsies as described in [27]. Their characterization was previously reported [27,176]. *TP53* and *IDH1* were not mutated in these cells and *SHH* and *NOTCH* pathways were active. TG1 and OB1 cells were able to differentiate and to induce tumors when orthotopically grafted (<500 cells) in immune-compromised mice. Cells were expanded in Dulbecco's modified Eagle's: F-12 medium (1:1) supplemented with N2, G5 (containing bFGF and EGF) and B27 (all from Invitrogen, France) at a density of 3.10^4 cells/cm². Human fetal neural stem cells (f-NSCs/NSC XXIV) were derived from human fetal brain and characterized as previously described [177]. Cells were cultured in "Neurocult"[®] NSC Proliferation Medium supplemented with basic fibroblast growth factor (10 ng/ml, bFGF, Peprotech, Rocky Hill, NJ) and with epidermal growth factor (20 ng/ml, EGF, Peprotech, Rocky Hill, NJ). Under the culture conditions used, TG1, OB1 and f-NSC cells grew as neurospheres. For cell expansion, once a week, neurospheres were mechanically dissociated into a single-cell suspension and 90% of their culture medium was renewed. Human Astrocyte primary

cells (HA cells, ScienCell Research Laboratories) were expanded in Astrocyte Medium (ScienCell Research Laboratories) according to the manufacturer's instructions. Human brain tumor cells U-87 MG (American Type Culture Collection, ATCC) were expanded in ATCC complete growth medium according to the manufacturer's instructions. For all cell lines, Master Cell Banks were established and Working Cell Banks were derived from these Master Cell Banks for further experiments.

G-protein-coupled receptor transcriptome

Total RNA was isolated from 10^7 cells of each cell type with TRI ReagentTM (Euromedex, France) according to the manufacturer's instructions and further purified with RNeasy mini kit columns (Qiagen). RNA purity was assessed by absorption spectra analysis using a NanoDrop ND-1000 apparatus (Labtech). Absorption ratios A260/A280 and A260/A230 were comprised between 1.8 and 2.1. RNA integrity and concentration were checked using an Agilent 2100 Bioanalyzer and the RNA 6000 LabChip kit. Only RNA with a RNA Integrity Number (RIN) higher than 8.5 was processed (2100 expert software, Agilent Technologies). 10 μ g of total RNA were reverse transcribed to single-stranded cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). Real-time PCR analysis was performed using TaqMan[®] Low Density Arrays (Applied Biosystems). TaqMan[®] GPCR Gene Expression Arrays correspond to quantitative PCR (Q-PCR) assays spotted onto a 384-well card. cDNA (850 ng) was carefully mixed with 850 μ L of TaqMan Universal PCR Master Mix (PE Applied Biosystems). Real-time RT-PCR was performed using an ABI Prism 7900HT apparatus (Applied Biosystems, Foster City, CA). Thermal cycling was carried out for 2 min at 50°C (to activate uracil-DNA glycosylase), 10 min at 94.5°C (activation) and 40 cycles of 30 s at 97°C and 60 s at 59.7°C. Each reaction well contained cDNA derived from about 1.5 ng of total RNA. Experiments were performed in triplicates for f-NSCs and duplicates for the four other cell types.

Cell membrane preparation, protein extraction and separation

Cells grown as neurospheres were dissociated, washed in PBS and suspended in buffer A consisting of 10 mM Hepes buffer pH 7.4, 2 mM EDTA and protease inhibitors (complete proteases inhibitors cocktail from Roche). Cells were mechanically lysed by pipetting the solution up and down with a P1000 pipette and pressing the solution through a 22G needle mounted on a syringe several times (~15) until 90% of the cells were lysed. Cell debris were removed by a 10 min. centrifugation at 1560 g. The supernatant was collected and centrifuged for 90 min at 20 800 g. Cell ghosts were resuspended in buffer A. After denaturation at 100°C in 5% SDS, 5% β mercaptoethanol, 1 mM EDTA, 10% glycerol, 10 mM Tris pH 8 buffer for 3 min, protein samples were fractionated on 1D SDS-PAGE gradient gel (8–15%). Gels were stained with colloidal Coomassie Blue. 90 slices were cut and stored at -20°C prior to mass spectrometry analysis.

Mass spectrometry proteomic analysis

In gel digestion of gel slices was performed as previously described [178]. The resulting peptide extracts were analysed by C18 reversed phase nanoHPLC on a nanoHPLC-Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to an ion trap amaZon (Bruker Daltonics, Bremen, Germany) mass spectrometer. Peptide mixtures were loaded on a 40 nl enrichment column and separated on a 75 μ m \times 150 mm column packed with Zorbax 300SB-C18 5 μ m material (Agilent Technologies, Palo

Alto, USA). Trapping was performed during 3 min, at 3.75 μ L/min, with solvent A (98% H₂O, 2% ACN and 0.1% FA). Elution was performed using two solvents, A (98% H₂O, 2% ACN and 0.1% FA) and B (98% ACN, 2% H₂O and 0.1% FA) at a flow rate of 300 nL/min, using a 2–24% B gradient over 45 min and a 24–40% B gradient over 16 min, followed by 70% B over 5 min.

For tandem MS experiments, the system was operated with automatic switching between MS and MS/MS modes. The 8 most abundant peptides, preferably doubly and triply charged ions, were selected on each MS spectrum for further isolation and CID fragmentation. Then, selected ions were excluded during 0.6 min. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 32 500 m/z per second. The complete system was fully controlled by Hystar 3.2 (Bruker Daltonics).

Proteomic Data Analysis

Raw data collected during nanoLC-MS/MS analyses were processed and converted into *.mgf peak list format with DataAnalysis 4.0 (Bruker Daltonics). MS/MS data were interpreted using two search engines: Mascot (version 2.3.02, Matrix Science, London, UK) installed on a local server and OMSSA (Open Mass Spectrometry Search Algorithm version 2.1.9; [179]) interfaced for grid computing on an in-house developed and freely available software suite (<https://msda.unistra.fr>). Searches were performed against a composite target-decoy UniProtKB-SwissProt human entries (TaxID = 9606, 10 July 2012, 40500 total entries). Searches were performed with a tolerance on mass measurement of 0.2 Da for precursor and 0.3 Da for fragment ions. Carbamidomethylation of cysteine residues, oxidation of methionine residues and propionamidation of cysteine residues were searched as variable modifications. Up to one trypsin missed cleavage was allowed. Scaffold software (version 3_6_5 Proteome Science, Portland, USA) was used for validation and false positive rate estimation of protein identifications. For each sample and each search engine, peptides were filtered out according to the cutoff set for protein hits with 1 or more peptides and a false positive identification rate less than 1%. Further data manipulations were performed with Microsoft Excel. Semi-quantification was performed by peptide and spectral counting.

Supporting Information

Figure S1 Expression of 14 housekeeping genes present on the TaqMan GPCR Array Card (Applied Biosystems).

Expression distribution for each housekeeping gene, in the different experiments performed, is given as box plots. Ordinates represent the cycle threshold (Ct) of the quantitative PCR experiment for each housekeeping gene indicated. The middle line in the boxes corresponds to the median. The end of the whiskers corresponds to the lowest and highest values within 1.5 IQR (interquartile range) of the lower and upper quartile. Outliers are represented by stars. Abbreviations: 18 s, 18 s ribosomal RNA; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; POLR2A, DNA dependent RNA polymerase 2A subunit; ACTB, actin β ; PPIA, peptidylpropyl isomerase A; PGK1, phosphoglycerate kinase 1; B2M, β 2-microglobulin; GUSB, glucuronidase β ; HPRT1, hypoxanthine phosphoribosyltransferase 1; TBP, TATA-box binding protein; TFRC, transferrin receptor; HMBS, hydroxymethylbilane synthase; IPO8, importin 8; PHGDH, phosphoglycerate dehydrogenase. (DOCX)

Table S1 Gene assays present on the GPCR microarray plates and their expression values in five cell types.

381 different assays are present on the GPCR microarray, allowing the analysis of the expression of 356 GPCRs, 14 different housekeeping genes (rRNA 18 s is in quadruplicate) and 11 other genes

(among which two genes of the LANCL family). Expression data are presented in four groups. Group I corresponds to GPCRs expressed with Ct values ≤ 31.5 (138 genes) in at least one of the five cell types, group II to GPCRs expressed at $Ct > 31.5$ in all five cell types, group III to the housekeeping genes (14 genes) and group IV to eleven genes not related to housekeeping genes or GPCRs. In each group, genes are listed in alphabetical order. Expression values are indicated as cycle threshold values (Ct) and as $2^{-Ct} * 10^{12}$ (arbitrary units). The cutoff level was set at $Ct \leq 31.5$ (i.e. expression level ≥ 329 units). For GPCRs which did not show amplification, Ct values were arbitrarily set at 40. (XLSX)

References

- Dolecek TA, Propp JM, Stroup NE, Kruchko C (2012) CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009. *Neuro Oncol* 14 Suppl 5: v1–49.
- Schechter CB (1999) Re: Brain and other central nervous system cancers: recent trends in incidence and mortality. *J Natl Cancer Inst* 91: 2050–2051.
- Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, et al. (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 352: 987–996.
- Holdhoff M, Grossman SA (2011) Controversies in the adjuvant therapy of high-grade gliomas. *Oncologist* 16: 351–358.
- Yin AA, Cheng JX, Zhang X, Liu BL (2013) The treatment of glioblastomas: a systematic update on clinical Phase III trials. *Crit Rev Oncol Hematol* 87: 265–282.
- Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, et al. (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 10: 459–466.
- Bonnet D, Dick JE (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 3: 730–737.
- Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414: 105–111.
- Hemmati HD, Nakano I, Lazareff JA, Masterman-Smith M, Geschwind DH, et al. (2003) Cancerous stem cells can arise from pediatric brain tumors. *Proc Natl Acad Sci U S A* 100: 15178–15183.
- Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, et al. (2003) Identification of a cancer stem cell in human brain tumors. *Cancer Res* 63: 5821–5828.
- Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* 100: 3983–3988.
- Dalerba P, Cho RW, Clarke MF (2007) Cancer stem cells: models and concepts. *Annu Rev Med* 58: 267–284.
- O'Brien CA, Pollett A, Gallinger S, Dick JE (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445: 106–110.
- Schatton T, Frank MH (2008) Cancer stem cells and human malignant melanoma. *Pigment Cell Melanoma Res* 21: 39–55.
- Schatton T, Murphy GF, Frank NY, Yamaura K, Waaga-Gasser AM, et al. (2008) Identification of cells initiating human melanomas. *Nature* 451: 345–349.
- Hermann PC, Huber SL, Herrler T, Aicher A, Ellwart JW, et al. (2007) Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell* 1: 313–323.
- Chen J, Li Y, Yu TS, McKay RM, Burns DK, et al. (2012) A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*.
- Driessens G, Beck B, Caauwe A, Simons BD, Blanpain C (2012) Defining the mode of tumour growth by clonal analysis. *Nature*.
- Schepers AG, Snippet HJ, Stange DE, van den Born M, van Es JH, et al. (2012) Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science* 337: 730–735.
- Frank NY, Schatton T, Frank MH (2010) The therapeutic promise of the cancer stem cell concept. *J Clin Invest* 120: 41–50.
- Alison MR, Lin WR, Lim SM, Nicholson IJ (2012) Cancer stem cells: in the line of fire. *Cancer Treat Rev* 38: 589–598.
- Sharkis SJ, Jones RJ, Civin C, Jang YY (2012) Pluripotent stem cell-based cancer therapy: promise and challenges. *Sci Transl Med* 4: 127ps129.
- Lee C, Dunn SE, Yip S (2012) Stem cells in brain tumour development and therapy—two-sides of the same coin. *Can J Neurol Sci* 39: 145–156.
- Nduom EK, Hadjipanayis CG, Van Meir EG (2012) Glioblastoma cancer stem-like cells: implications for pathogenesis and treatment. *Cancer J* 18: 100–106.
- Malik B, Nie D (2012) Cancer stem cells and resistance to chemo and radio therapy. *Front Biosci (Elite Ed)* 4: 2142–2149.
- Persano L, Rampazzo E, Basso G, Viola G (2013) Glioblastoma cancer stem cells: role of the microenvironment and therapeutic targeting. *Biochem Pharmacol* 85: 612–622.
- Patru C, Romao L, Varlet P, Coulombel L, Raponi E, et al. (2010) CD133, CD15/SSEA-1, CD34 or side populations do not resume tumor-initiating properties of long-term cultured cancer stem cells from human malignant glioblastoma tumors. *BMC Cancer* 10: 66.
- Drews J (2000) Drug discovery: a historical perspective. *Science* 287: 1960–1964.
- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1: 727–730.
- Jacoby E, Bouhelal R, Gerspacher M, Seuwen K (2006) The 7 TM G-protein-coupled receptor target family. *ChemMedChem* 1: 761–782.
- Garland SL (2013) Are GPCRs Still a Source of New Targets? *J Biomol Screen*.
- Lappano R, Maggiolini M (2011) G protein-coupled receptors: novel targets for drug discovery in cancer. *Nat Rev Drug Discov* 10: 47–60.
- Herr DR (2012) Potential use of G protein-coupled receptor-blocking monoclonal antibodies as therapeutic agents for cancers. *Int Rev Cell Mol Biol* 297: 45–81.
- Dorsam RT, Gutkind JS (2007) G-protein-coupled receptors and cancer. *Nat Rev Cancer* 7: 79–94.
- Audigier Y, Picault FX, Chaves-Almagro C, Masri B (2013) G Protein-Coupled Receptors in cancer: biochemical interactions and drug design. *Prog Mol Biol Transl Sci* 115: 143–173.
- Feigin ME (2013) Harnessing the genome for characterization of GPCRs in cancer pathogenesis. *FEBS J*.
- Axelsson M, Liu K, Jiang X, He K, Wang J, et al. (2013) U.S. Food and Drug Administration approval: vismodegib for recurrent, locally advanced, or metastatic basal cell carcinoma. *Clin Cancer Res* 19: 2289–2293.
- Webb DR, Handel TM, Kretz-Rommel A, Stevens RC (2013) Opportunities for functional selectivity in GPCR antibodies. *Biochem Pharmacol* 85: 147–152.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.
- Valente V, Teixeira SA, Neder L, Okamoto OK, Oba-Shinjo SM, et al. (2009) Selection of suitable housekeeping genes for expression analysis in glioblastoma using quantitative RT-PCR. *BMC Mol Biol* 10: 17.
- Mackrill JJ (2004) Generation, use, and validation of receptor-selective antibodies. *Methods Mol Biol* 259: 47–65.
- Jensen BC, Swigart PM, Simpson PC (2009) Ten commercial antibodies for alpha-1-adrenergic receptor subtypes are nonspecific. *Naunyn Schmiedebergs Arch Pharmacol* 379: 409–412.
- Bodei S, Arrighi N, Spano P, Sigala S (2009) Should we be cautious on the use of commercially available antibodies to dopamine receptors? *Naunyn Schmiedebergs Arch Pharmacol* 379: 413–415.
- Hamdani N, van der Velden J (2009) Lack of specificity of antibodies directed against human beta-adrenergic receptors. *Naunyn Schmiedebergs Arch Pharmacol* 379: 403–407.
- Pradicharceep W, Stallen J, Labruyere WT, Dabhoiwala NF, Michel MC, et al. (2009) Lack of specificity of commercially available antisera against muscarinic and adrenergic receptors. *Naunyn Schmiedebergs Arch Pharmacol* 379: 397–402.
- Beermann S, Seifert R, Neumann D (2012) Commercially available antibodies against human and murine histamine H(4)-receptor lack specificity. *Naunyn Schmiedebergs Arch Pharmacol* 385: 125–135.
- Aust G, Eichler W, Laue S, Lehmann I, Heldin NE, et al. (1997) CD97: a dedifferentiation marker in human thyroid carcinomas. *Cancer Res* 57: 1798–1806.
- Chidambaram A, Fillmore HL, Van Meter TE, Dumur CI, Broaddus WC (2012) Novel report of expression and function of CD97 in malignant gliomas: correlation with Wilms tumor 1 expression and glioma cell invasiveness. *J Neurosurg* 116: 843–853.

Acknowledgments

We thank Dr René Frydman (Hopital Antoine Bécélère, Clamart, France) and Dr Hervé Coffiny for providing fetal brain tissues and the Sainte Anne neurosurgery team for providing glioblastoma tissues.

Author Contributions

Conceived and designed the experiments: JMS CC AVD HC SC JH MCK. Performed the experiments: MF JMS MZ SL JD. Analyzed the data: MF JMS MZ SL CC JD JH MCK. Contributed reagents/materials/analysis tools: MPJ HC. Wrote the paper: JMS SL CC SC JH MCK. Revised the paper critically: HC AVD MPJ HC.

49. Ward Y, Lake R, Yin JJ, Heger CD, Raffeld M, et al. (2011) LPA receptor heterodimerizes with CD97 to amplify LPA-initiated RHO-dependent signaling and invasion in prostate cancer cells. *Cancer Res* 71: 7301–7311.
50. Ward Y, Lake R, Martin PL, Killian K, Salerno P, et al. (2013) CD97 amplifies LPA receptor signaling and promotes thyroid cancer progression in a mouse model. *Oncogene* 32: 2726–2738.
51. Safaei M, Clark AJ, Oh MC, Ivan ME, Bloch O, et al. (2013) Overexpression of CD97 confers an invasive phenotype in glioblastoma cells and is associated with decreased survival of glioblastoma patients. *PLoS One* 8: e62765.
52. Liu D, Trojanowicz B, Radestock Y, Fu T, Hammje K, et al. (2010) Role of CD97 isoforms in gastric carcinoma. *Int J Oncol* 36: 1401–1408.
53. Shashidhar S, Lorente G, Nagavarapu U, Nelson A, Kuo J, et al. (2005) GPR56 is a GPCR that is overexpressed in gliomas and functions in tumor cell adhesion. *Oncogene* 24: 1673–1682.
54. Xu L, Begum S, Hearn JD, Hynes RO (2006) GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. *Proc Natl Acad Sci U S A* 103: 9023–9028.
55. Xu L, Begum S, Barry M, Crowley D, Yang L, et al. (2010) GPR56 plays varying roles in endogenous cancer progression. *Clin Exp Metastasis* 27: 241–249.
56. Yang L, Chen G, Mohanty S, Scott G, Fazal F, et al. (2011) GPR56 Regulates VEGF production and angiogenesis during melanoma progression. *Cancer Res* 71: 5558–5568.
57. Luo R, Jeong SJ, Jin Z, Strokes N, Li S, et al. (2011) G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc Natl Acad Sci U S A* 108: 12925–12930.
58. Terskikh AV, Easterday MC, Li L, Hood L, Kornblum HI, et al. (2001) From hematopoiesis to neurogenesis: evidence of overlapping genetic programs. *Proc Natl Acad Sci U S A* 98: 7934–7939.
59. Bai Y, Du L, Shen L, Zhang Y, Zhang L (2009) GPR56 is highly expressed in neural stem cells but downregulated during differentiation. *Neuroreport* 20: 918–922.
60. Cork SM, Van Meir EG (2011) Emerging roles for the BAI1 protein family in the regulation of phagocytosis, synaptogenesis, neurovasculature, and tumor development. *J Mol Med (Berl)* 89: 743–752.
61. Kaur B, Brat DJ, Calkins CC, Van Meir EG (2003) Brain angiogenesis inhibitor 1 is differentially expressed in normal brain and glioblastoma independently of p53 expression. *Am J Pathol* 162: 19–27.
62. Kang X, Xiao X, Harata M, Bai Y, Nakazaki Y, et al. (2006) Antiangiogenic activity of BAI1 in vivo: implications for gene therapy of human glioblastomas. *Cancer Gene Ther* 13: 385–392.
63. Cork SM, Kaur B, Devi NS, Cooper L, Saltz JH, et al. (2012) A proprotein convertase/MMP-14 proteolytic cascade releases a novel 40 kDa vasculostatin from tumor suppressor BAI1. *Oncogene*.
64. Okajima D, Kudo G, Yokota H (2010) Brain-specific angiogenesis inhibitor 2 (BAI2) may be activated by proteolytic processing. *J Recept Signal Transduct Res* 30: 143–153.
65. Kee HJ, Ahn KY, Choi KC, Won Song J, Heo T, et al. (2004) Expression of brain-specific angiogenesis inhibitor 3 (BAI3) in normal brain and implications for BAI3 in ischemia-induced brain angiogenesis and malignant glioma. *FEBS Lett* 569: 307–316.
66. Garcia-Lopez MT, Gutierrez-Rodriguez M, Herranz R (2010) Thrombin-activated receptors: promising targets for cancer therapy? *Curr Med Chem* 17: 109–128.
67. Elste AP, Petersen I (2010) Expression of proteinase-activated receptor 1–4 (PAR 1–4) in human cancer. *J Mol Histol* 41: 89–99.
68. Arora P, Cuevas BD, Russo A, Johnson GL, Trejo J (2008) Persistent transactivation of EGFR and ErbB2/HER2 by protease-activated receptor-1 promotes breast carcinoma cell invasion. *Oncogene* 27: 4434–4445.
69. Gessler F, Voss V, Dutzmann S, Seifert V, Gerlach R, et al. (2010) Inhibition of tissue factor/protease-activated receptor-2 signaling limits proliferation, migration and invasion of malignant glioma cells. *Neuroscience* 165: 1312–1322.
70. Dutzmann S, Gessler F, Harter PN, Gerlach R, Mittelbronn M, et al. (2010) The pro-migratory and pro-invasive role of the procoagulant tissue factor in malignant gliomas. *Cell Adh Migr* 4: 515–522.
71. Magnus N, Garnier D, Rak J (2010) Oncogenic epidermal growth factor receptor up-regulates multiple elements of the tissue factor signaling pathway in human glioma cells. *Blood* 116: 815–818.
72. Zigler M, Kamiya T, Brandley EC, Villares GJ, Bar-Eli M (2011) PAR-1 and thrombin: the ties that bind the microenvironment to melanoma metastasis. *Cancer Res* 71: 6561–6566.
73. Robbins MJ, Michalovich D, Hill J, Calver AR, Medhurst AD, et al. (2000) Molecular cloning and characterization of two novel retinoic acid-inducible orphan G-protein-coupled receptors (GPC5B and GPC5C). *Genomics* 67: 8–18.
74. Brauner-Osborne H, Krosgaard-Larsen P (2000) Sequence and expression pattern of a novel human orphan G-protein-coupled receptor, GPC5B, a family C receptor with a short amino-terminal domain. *Genomics* 65: 121–128.
75. Harada Y, Yokota C, Habas R, Shusarski DC, He X (2007) Retinoic acid-inducible G protein-coupled receptors bind to frizzled receptors and may activate non-canonical Wnt signaling. *Biochem Biophys Res Commun* 358: 968–975.
76. Cool BH, Chan GC, Lee L, Oshima J, Martin GM, et al. (2010) A flanking gene problem leads to the discovery of a Gpr5b splice variant predominantly expressed in C57Bl/6j mouse brain and in maturing neurons. *PLoS One* 5: e10351.
77. Usui T, Shima Y, Shimada Y, Hirano S, Burgess RW, et al. (1999) Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled. *Cell* 98: 585–595.
78. Boutin C, Goffinet AM, Tissir F (2012) Celsr1-3 cadherins in PCP and brain development. *Curr Top Dev Biol* 101: 161–183.
79. Tissir F, Goffinet AM (2006) Expression of planar cell polarity genes during development of the mouse CNS. *Eur J Neurosci* 23: 597–607.
80. Tissir F, Qu Y, Montcouquiol M, Zhou L, Komatsu K, et al. (2010) Lack of cadherins Celsr2 and Celsr3 impairs ependymal ciliogenesis, leading to fatal hydrocephalus. *Nat Neurosci* 13: 700–707.
81. Sugimura R, He XC, Venkatraman A, Arai F, Box A, et al. (2012) Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* 150: 351–365.
82. Wada H, Tanaka H, Nakayama S, Iwasaki M, Okamoto H (2006) Frizzled3a and Celsr2 function in the neuroepithelium to regulate migration of facial motor neurons in the developing zebrafish hindbrain. *Development* 133: 4749–4759.
83. Katoh M (2005) WNT/PCP signaling pathway and human cancer (review). *Oncol Rep* 14: 1583–1588.
84. Huang H, Groth J, Sossey-Alaoui K, Hawthorn L, Beall S, et al. (2005) Aberrant expression of novel and previously described cell membrane markers in human breast cancer cell lines and tumors. *Clin Cancer Res* 11: 4357–4364.
85. Terasaki M, Sugita Y, Arakawa F, Okada Y, Ohshima K, et al. (2011) CXCL12/CXCR4 signaling in malignant brain tumors: a potential pharmacological therapeutic target. *Brain Tumor Pathol* 28: 89–97.
86. Domanska UM, Kruijzinga RC, Nagengast WB, Timmer-Boscha H, Huls G, et al. (2013) A review on CXCR4/CXCL12 axis in oncology: no place to hide. *Eur J Cancer* 49: 219–230.
87. Ehteshami M, Mapara KY, Stevenson CB, Thompson RC (2009) CXCR4 mediates the proliferation of glioblastoma progenitor cells. *Cancer Lett* 274: 305–312.
88. Schulte A, Gunther HS, Phillips HS, Kemming D, Martens T, et al. (2011) A distinct subset of glioma cell lines with stem cell-like properties reflects the transcriptional phenotype of glioblastomas and overexpresses CXCR4 as therapeutic target. *Glia* 59: 590–602.
89. Fareh M, Turchi L, Virolle V, Debruyne D, Almairac F, et al. (2012) The miR 302–367 cluster drastically affects self-renewal and infiltration properties of glioma-initiating cells through CXCR4 repression and consequent disruption of the SHH-GLI-NANOG network. *Cell Death Differ* 19: 232–244.
90. Cheng L, Huang Z, Zhou W, Wu Q, Donnola S, et al. (2013) Glioblastoma stem cells generate vascular pericytes to support vessel function and tumor growth. *Cell* 153: 139–152.
91. Rao S, Sengupta R, Choe EJ, Woerner BM, Jackson E, et al. (2012) CXCL12 mediates trophic interactions between endothelial and tumor cells in glioblastoma. *PLoS One* 7: e33005.
92. Li M, Chang CJ, Lathia JD, Wang L, Pacentia HL, et al. (2011) Chemokine receptor CXCR4 signaling modulates the growth factor-induced cell cycle of self-renewing and multipotent neural progenitor cells. *Glia* 59: 108–118.
93. Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, et al. (1998) GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* 396: 674–679.
94. White JH, Wise A, Main MJ, Green A, Fraser NJ, et al. (1998) Heterodimerization is required for the formation of a functional GABA(B) receptor. *Nature* 396: 679–682.
95. Kaupmann K, Malitschek B, Schuler V, Heid J, Froestl W, et al. (1998) GABA(B)-receptor subtypes assemble into functional heteromeric complexes. *Nature* 396: 683–687.
96. Billinton A, Ige AO, Wise A, White JH, Disney GH, et al. (2000) GABA(B) receptor heterodimer-component localisation in human brain. *Brain Res Mol Brain Res* 77: 111–124.
97. Villemure JF, Adam L, Bevan NJ, Gearing K, Chenier S, et al. (2005) Subcellular distribution of GABA(B) receptor homo- and hetero-dimers. *Biochem J* 388: 47–55.
98. Maurel D, Comps-Agrar L, Brock C, Rives ML, Bourrier E, et al. (2008) Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nat Methods* 5: 561–567.
99. Chang W, Tu C, Cheng Z, Rodriguez L, Chen TH, et al. (2007) Complex formation with the Type B gamma-aminobutyric acid receptor affects the expression and signal transduction of the extracellular calcium-sensing receptor. Studies with HEK-293 cells and neurons. *J Biol Chem* 282: 25030–25040.
100. Boyer SB, Clancy SM, Terunuma M, Revilla-Sanchez R, Thomas SM, et al. (2009) Direct interaction of GABAB receptors with M2 muscarinic receptors enhances muscarinic signaling. *J Neurosci* 29: 15796–15809.
101. Fernando RN, Eleuteri B, Abdelhady S, Nussenzweig A, Andang M, et al. (2011) Cell cycle restriction by histone H2AX limits proliferation of adult neural stem cells. *Proc Natl Acad Sci U S A* 108: 5837–5842.
102. Kriegstein AR (2005) GABA puts the brake on stem cells. *Nat Neurosci* 8: 1132–1133.

103. Fukui M, Nakamichi N, Yoneyama M, Ozawa S, Fujimori S, et al. (2008) Modulation of cellular proliferation and differentiation through GABA(B) receptors expressed by undifferentiated neural progenitor cells isolated from fetal mouse brain. *J Cell Physiol* 216: 507–519.
104. Oka M, Wada M, Wu Q, Yamamoto A, Fujita T (2006) Functional expression of metabotropic GABAB receptors in primary cultures of astrocytes from rat cerebral cortex. *Biochem Biophys Res Commun* 341: 874–881.
105. Mizuta K, Osawa Y, Mizuta F, Xu D, Emala CW (2008) Functional expression of GABAB receptors in airway epithelium. *Am J Respir Cell Mol Biol* 39: 296–304.
106. Tamura S, Watanabe M, Kanbara K, Yanagawa T, Watanabe K, et al. (2009) Expression and distribution of GABAergic system in rat knee joint synovial membrane. *Histol Histopathol* 24: 1009–1019.
107. Zhou Z, Sun H, Li X, Li Y, Zhao S, et al. (2010) A local GABAergic system is functionally expressed in human fallopian tube. *Biochem Biophys Res Commun* 398: 237–241.
108. Jiang X, Su L, Zhang Q, He C, Zhang Z, et al. (2012) GABAB Receptor Complex as a Potential Target for Tumor Therapy. *J Histochem Cytochem* 60: 269–279.
109. Roberts SS, Mendonca-Torres MC, Jensen K, Francis GL, Vasko V (2009) GABA receptor expression in benign and malignant thyroid tumors. *Pathol Oncol Res* 15: 645–650.
110. Wang T, Huang W, Chen F (2008) Baclofen, a GABAB receptor agonist, inhibits human hepatocellular carcinoma cell growth in vitro and in vivo. *Life Sci* 82: 536–541.
111. Tatsuta M, Iishi H, Baba M, Taniguchi H (1992) Attenuation by the GABA receptor agonist baclofen of experimental carcinogenesis in rat colon by azoxymethane. *Oncology* 49: 241–245.
112. Hannan S, Wilkins ME, Dehghani-Tafii E, Thomas P, Baddeley SM, et al. (2011) GABAB receptor internalisation is regulated by the R2 subunit. *J Biol Chem*.
113. Carmon KS, Gong X, Lin Q, Thomas A, Liu Q (2011) R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/beta-catenin signaling. *Proc Natl Acad Sci U S A* 108: 11452–11457.
114. de Lau W, Barker N, Low TY, Koo BK, Li VS, et al. (2011) Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* 476: 293–297.
115. Birchmeier W (2011) Stem cells: Orphan receptors find a home. *Nature* 476: 287–288.
116. Sato T, van Es JH, Snippert HJ, Stange DE, Vries RG, et al. (2011) Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* 469: 415–418.
117. Snippert HJ, Clevers H (2011) Tracking adult stem cells. *EMBO Rep* 12: 113–122.
118. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, et al. (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449: 1003–1007.
119. Mustata RC, Van Loy T, Lefort A, Libert F, Strollo S, et al. (2011) Lgr4 is required for Paneth cell differentiation and maintenance of intestinal stem cells ex vivo. *EMBO Rep* 12: 558–564.
120. Kato S, Matsubara M, Matsuo T, Mohri Y, Kazama I, et al. (2006) Leucine-rich repeat-containing G protein-coupled receptor-4 (LGR4, Gpr48) is essential for renal development in mice. *Nephron Exp Nephrol* 104: e63–75.
121. Oyama K, Mohri Y, Sone M, Nawa A, Nishimori K (2011) Conditional knockout of Lgr4 leads to impaired ductal elongation and branching morphogenesis in mouse mammary glands. *Sex Dev* 5: 205–212.
122. Wang Z, Jin C, Li H, Li C, Hou Q, et al. (2010) GPR48-Induced keratinocyte proliferation occurs through HB-EGF mediated EGFR transactivation. *FEBS Lett* 584: 4057–4062.
123. Gao Y, Kitagawa K, Hiramatsu Y, Kikuchi H, Isobe T, et al. (2006) Up-regulation of GPR48 induced by down-regulation of p27Kip1 enhances carcinoma cell invasiveness and metastasis. *Cancer Res* 66: 11623–11631.
124. Gao Y, Shan ZY, Wang H, Zhang HM, Teng WP (2009) Inhibitory effect of shRNA targeting GPR48 on invasion and metastasis of human cervical carcinoma cell line HeLa. *Ai Zheng* 28: 104–107.
125. Stella N (2010) Cannabinoid and cannabinoid-like receptors in microglia, astrocytes, and astrocytomas. *Glia* 58: 1017–1030.
126. Zhuang M, Yang Y, Cao F, Lu M, Wang X, et al. (2012) Associations of variants of CNR1 with obesity and obesity-related traits in Chinese women. *Gene* 495: 194–198.
127. Verde Z, Santiago C, Rodriguez Gonzalez-Moro JM, de Lucas Ramos P, Lopez Martin S, et al. (2011) ‘Smoking genes’: a genetic association study. *PLoS One* 6: e26668.
128. Smith SB, Maixner DW, Fillingim RB, Slade G, Gracely RH, et al. (2012) Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. *Arthritis Rheum* 64: 584–593.
129. Carracedo A, Lorente M, Egia A, Blazquez C, Garcia S, et al. (2006) The stress-regulated protein p8 mediates cannabinoid-induced apoptosis of tumor cells. *Cancer Cell* 9: 301–312.
130. Salazar M, Carracedo A, Salanueva JJ, Hernandez-Tiedra S, Lorente M, et al. (2009) Cannabinoid action induces autophagy-mediated cell death through stimulation of ER stress in human glioma cells. *J Clin Invest* 119: 1359–1372.
131. Cudaback E, Marrs W, Moeller T, Stella N (2010) The expression level of CB1 and CB2 receptors determines their efficacy at inducing apoptosis in astrocytomas. *PLoS One* 5: e8702.
132. Aguado T, Carracedo A, Julien B, Velasco G, Milman G, et al. (2007) Cannabinoids induce glioma stem-like cell differentiation and inhibit gliomagenesis. *J Biol Chem* 282: 6854–6862.
133. Wu X, Han L, Zhang X, Li L, Jiang C, et al. (2012) Alteration of endocannabinoid system in human gliomas. *J Neurochem* 120: 842–849.
134. Held-Feindt J, Dorner L, Sahan G, Mehdorn HM, Mentlein R (2006) Cannabinoid receptors in human astroglial tumors. *J Neurochem* 98: 886–893.
135. Schley M, Stander S, Kerner J, Vajkoczy P, Schupfer G, et al. (2009) Predominant CB2 receptor expression in endothelial cells of glioblastoma in humans. *Brain Res Bull* 79: 333–337.
136. De Jesus ML, Hostalot C, Garibi JM, Salles J, Meana JJ, et al. (2010) Opposite changes in cannabinoid CB1 and CB2 receptor expression in human gliomas. *Neurochem Int* 56: 829–833.
137. Marshall AD, Lagutina I, Grosveld GC (2011) PAX3-FOXO1 induces cannabinoid receptor 1 to enhance cell invasion and metastasis. *Cancer Res* 71: 7471–7480.
138. Kenessey I, Banki B, Mark A, Varga N, Tovari J, et al. (2012) Revisiting CB1 Receptor as Drug Target in Human Melanoma. *Pathol Oncol Res*.
139. Proto MC, Gazzzerro P, Di Croce L, Santoro A, Malfitano AM, et al. (2012) Interaction of endocannabinoid system and steroid hormones in the control of colon cancer cell growth. *J Cell Physiol* 227: 250–258.
140. Silva JP, Ushkaryov YA (2010) The latrophilins, “split-personality” receptors. *Adv Exp Med Biol* 706: 59–75.
141. Matsushita H, Lelianaova VG, Ushkaryov YA (1999) The latrophilin family: multiply spliced G protein-coupled receptors with differential tissue distribution. *FEBS Lett* 443: 348–352.
142. Ichtschenko K, Bittner MA, Krasnoperov V, Little AR, Chepurny O, et al. (1999) A novel ubiquitously expressed alpha-latrotoxin receptor is a member of the CIRL family of G-protein-coupled receptors. *J Biol Chem* 274: 5491–5498.
143. Doyle SE, Scholz MJ, Greer KA, Hubbard AD, Darnell DK, et al. (2006) Latrophilin-2 is a novel component of the epithelial-mesenchymal transition within the atrioventricular canal of the embryonic chicken heart. *Dev Dyn* 235: 3213–3221.
144. Boucard AA, Ko J, Sudhof TC (2012) High Affinity Neuroxin Binding to Cell Adhesion G-protein-coupled Receptor CIRL1/Latrophilin-1 Produces an Intercellular Adhesion Complex. *J Biol Chem* 287: 9399–9413.
145. Silva JP, Lelianaova VG, Ermolyuk YS, Vysokov N, Hitchen PG, et al. (2011) Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. *Proc Natl Acad Sci U S A* 108: 12113–12118.
146. O’Sullivan ML, de Wit J, Savas JN, Comoletti D, Otto-Hitt S, et al. (2012) FLRT Proteins Are Endogenous Latrophilin Ligands and Regulate Excitatory Synapse Development. *Neuron* 73: 903–910.
147. White GR, Varley JM, Heighway J (1998) Isolation and characterization of a human homologue of the latrophilin gene from a region of 1p31.1 implicated in breast cancer. *Oncogene* 17: 3513–3519.
148. Eng L, Ibrahim-zada I, Jarjanazi H, Savas S, Meschian M, et al. (2011) Bioinformatic analyses identifies novel protein-coding pharmacogenomic markers associated with paclitaxel sensitivity in NCI60 cancer cell lines. *BMC Med Genomics* 4: 18.
149. Yang G, Quan Y, Wang W, Fu Q, Wu J, et al. (2012) Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. *Br J Cancer*.
150. Garritano S, Inga A, Gemignani F, Landi S (2013) More targets, more pathways and more clues for mutant p53. *Oncogenesis* 2: e54.
151. Del Valle J, Gantz I (1997) Novel insights into histamine H2 receptor biology. *Am J Physiol* 273: G987–996.
152. Hill SJ, Ganellin CR, Timmerman H, Schwartz JC, Shankley NP, et al. (1997) International Union of Pharmacology. XIII. Classification of histamine receptors. *Pharmacol Rev* 49: 253–278.
153. Medina VA, Rivera ES (2010) Histamine receptors and cancer pharmacology. *Br J Pharmacol* 161: 755–767.
154. Kubecova M, Kolostova K, Pinterova D, Kacprzak G, Bobek V (2011) Cimetidine: an anticancer drug? *Eur J Pharm Sci* 42: 439–444.
155. Zheng Y, Xu M, Li X, Jia J, Fan K, et al. (2013) Cimetidine suppresses lung tumor growth in mice through proapoptosis of myeloid-derived suppressor cells. *Mol Immunol* 54: 74–83.
156. Lisy K (2013) Histamine type 2 receptor antagonists as adjuvant treatment for resected colorectal cancer. *Clin J Oncol Nurs* 17: 211–212.
157. Deva S, Jameson M (2012) Histamine type 2 receptor antagonists as adjuvant treatment for resected colorectal cancer. *Cochrane Database Syst Rev* 8: CD007814.
158. Crizzo G, Martin G, Medina V, Nunez M, Mohamad N, et al. (2006) Histamine inhibits cell proliferation and modulates the expression of Bcl-2 family proteins via the H2 receptor in human pancreatic cancer cells. *Anticancer Res* 26: 4443–4450.
159. McLatchie LM, Fraser NJ, Main MJ, Wise A, Brown J, et al. (1998) RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* 393: 333–339.
160. Oehler MK, Norbury C, Hague S, Rees MC, Bicknell R (2001) Adrenomedullin inhibits hypoxic cell death by upregulation of Bcl-2 in endometrial

- cancer cells: a possible promotion mechanism for tumour growth. *Oncogene* 20: 2937–2945.
161. Martinez A, Vos M, Guedez L, Kaur G, Chen Z, et al. (2002) The effects of adrenomedullin overexpression in breast tumor cells. *J Natl Cancer Inst* 94: 1226–1237.
 162. Metellus P, Voutsinos-Porche B, Nanni-Metellus I, Colin C, Fina F, et al. (2011) Adrenomedullin expression and regulation in human glioblastoma, cultured human glioblastoma cell lines and pilocytic astrocytoma. *Eur J Cancer* 47: 1727–1735.
 163. Kaafarani I, Fernandez-Sauze S, Berenguer C, Chinot O, Delfino C, et al. (2009) Targeting adrenomedullin receptors with systemic delivery of neutralizing antibodies inhibits tumor angiogenesis and suppresses growth of human tumor xenografts in mice. *FASEB J* 23: 3424–3435.
 164. Pyne NJ, Tonelli F, Lim KG, Long JS, Edwards J, et al. (2012) Sphingosine 1-phosphate signalling in cancer. *Biochem Soc Trans* 40: 94–100.
 165. Huang X, Wang Y, Nan X, He S, Xu X, et al. (2013) The role of the orphan G protein-coupled receptor 37 (GPR37) in multiple myeloma cells. *Leuk Res*.
 166. Bholra NE, Grandis JR (2008) Crosstalk between G-protein-coupled receptors and epidermal growth factor receptor in cancer. *Front Biosci* 13: 1857–1865.
 167. Daub H, Weiss FU, Wallasch C, Ullrich A (1996) Role of transactivation of the EGF receptor in signalling by G-protein-coupled receptors. *Nature* 379: 557–560.
 168. Delcourt N, Bockaert J, Marin P (2007) GPCR-jacking: from a new route in RTK signalling to a new concept in GPCR activation. *Trends Pharmacol Sci* 28: 602–607.
 169. Lappano R, De Marco P, De Francesco EM, Chimento A, Pezzi V, et al. (2013) Cross-talk between GPER and growth factor signaling. *J Steroid Biochem Mol Biol*.
 170. Yu FX, Guan KL (2013) The Hippo pathway: regulators and regulations. *Genes Dev* 27: 355–371.
 171. O'Hayre M, Vazquez-Prado J, Kufareva I, Stawiski EW, Handel TM, et al. (2013) The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. *Nat Rev Cancer* 13: 412–424.
 172. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, et al. (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 466: 869–873.
 173. Wootten D, Christopoulos A, Sexton PM (2013) Emerging paradigms in GPCR allostery: implications for drug discovery. *Nat Rev Drug Discov* 12: 630–644.
 174. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, et al. (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114: 97–109.
 175. Varlet P, Soni D, Miquel C, Roux FX, Meder JF, et al. (2004) New variants of malignant glioneuronal tumors: a clinicopathological study of 40 cases. *Neurosurgery* 55: 1377–1391; discussion 1391–1372.
 176. Galan-Moya EM, Le Guelte A, Lima Fernandes E, Thirant C, Dwyer J, et al. (2011) Secreted factors from brain endothelial cells maintain glioblastoma stem-like cell expansion through the mTOR pathway. *EMBO Rep* 12: 470–476.
 177. Thirant C, Besette B, Varlet P, Puget S, Cadusseau J, et al. (2011) Clinical relevance of tumor cells with stem-like properties in pediatric brain tumors. *PLoS One* 6: e16375.
 178. Miguet L, Bechade G, Fornecker L, Zink E, Felden C, et al. (2009) Proteomic analysis of malignant B-cell derived microparticles reveals CD148 as a potentially useful antigenic biomarker for mantle cell lymphoma diagnosis. *J Proteome Res* 8: 3346–3354.
 179. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, et al. (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3: 958–964.
 180. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98: 31–36.

Annexe 7

Communication par affiche

Congrès SMAP de 2011 à Avignon (Spectrométrie de masse et analyse protéomique)

Lennon S., Saliou JM., Audran E., Chneiweiss H., Carapito C., Van Dorsselaer A., Junier MP., Haiech J., Cianférani S.

Développements méthodologiques en analyse protéomique pour l'établissement du protéome membranaire de cellules souches cancéreuses issues de gliome

Développements méthodologiques en analyse protéomique pour l'établissement du protéome membranaire de cellules souches cancéreuses issues de gliomes

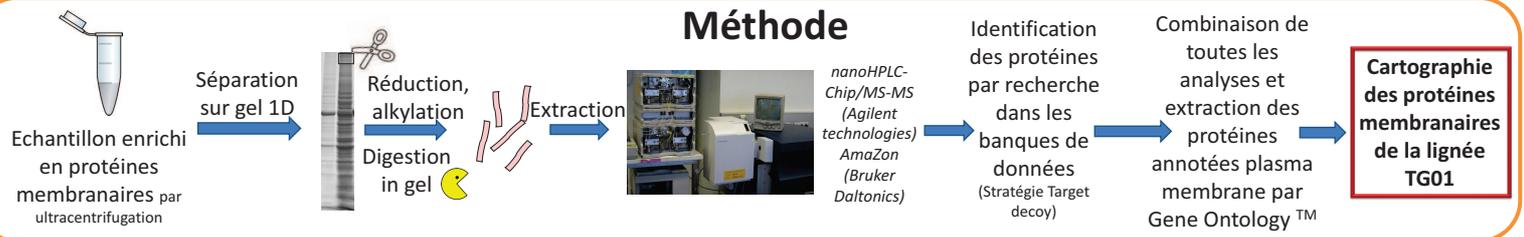
Sarah Lennon ⁽¹⁾, Jean-Michel Saliou ⁽¹⁾, Emilie Audran ⁽²⁾, Hervé Chneiweiss ⁽³⁾, Christine Carapito ⁽¹⁾, Alain Van Dorsselaer ⁽¹⁾, Jacques Haiech ⁽²⁾, Sarah Cianféroni-Sanglier ⁽¹⁾

(1) Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, DSA, CNRS UMR7178, Uds, 67087 Strasbourg, France
 (2) Intitut Gilbert Laustriat, CNRS, UMR 7175, Uds, Illkirch, France
 (3) Plasticité gliale et tumeurs au cerveau, Inserm UMR 894, Université Paris Descartes, Paris, France

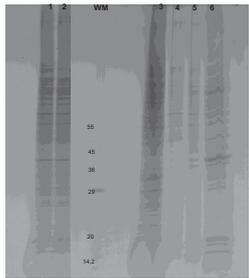
Introduction

Les gliomes concernent l'ensemble des tumeurs du système nerveux central. Malgré les avancées en chirurgie médicale et dans les thérapies médicamenteuses, ces tumeurs restent difficiles à traiter. Très récemment, la présence de cellules souches cancéreuses (CSC) au sein de ces tumeurs a été mise en évidence ¹. Du fait de leur résistance aux traitements de chimio/radiothérapie actuels et de leur capacité à sommeiller pendant une longue période de temps, les CSC pourraient être à l'origine des rechutes et sont donc des cibles intéressantes pour la recherche de nouveaux biomarqueurs thérapeutiques. L'objectif global de ce projet est la recherche de biomarqueurs de CSC du gliome.

BUT de cette étude: réaliser une cartographie la plus complète possible du protéome membranaire d'une lignée de CSC du gliome (TG01)

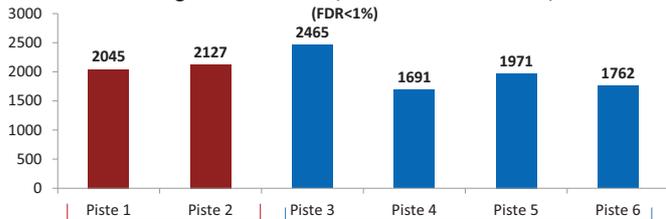


Tests réalisés



Piste 1: Témoin TG01 réplikat 1
 Piste 2: Témoin TG01 réplikat 2
 Piste 3: Culot insoluble
 Piste 4: Culot 30 % sulfate d'ammonium
 Piste 5: Culot 40 % sulfate d'ammonium
 Piste 6: Culot 80 % sulfate d'ammonium

Résultats globaux: Nombre de protéines identifiées dans chaque test (FDR<1%)

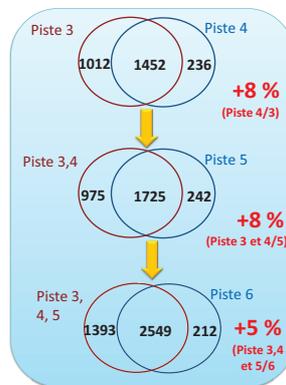


Réplicats techniques 2281 protéines + 21% Fractionnement de l'échantillon 2904 protéines

Au total: en combinant toutes les analyses
 3080 protéines -> Annotation « plasma membrane » -> 599 protéines (19%) -> 61 clusters de différenciation

Apport du fractionnement par sulfate d'ammonium

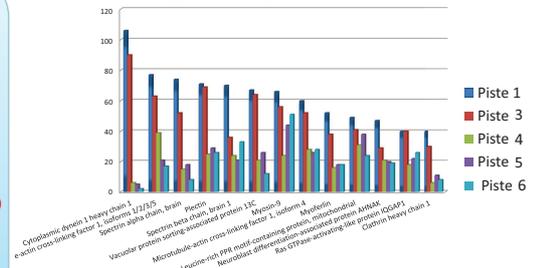
1) Augmentation du nombre de protéines identifiées



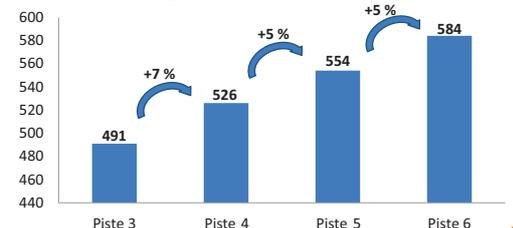
3) Identification de protéines membranaires supplémentaires

2) Décomplexification de l'échantillon

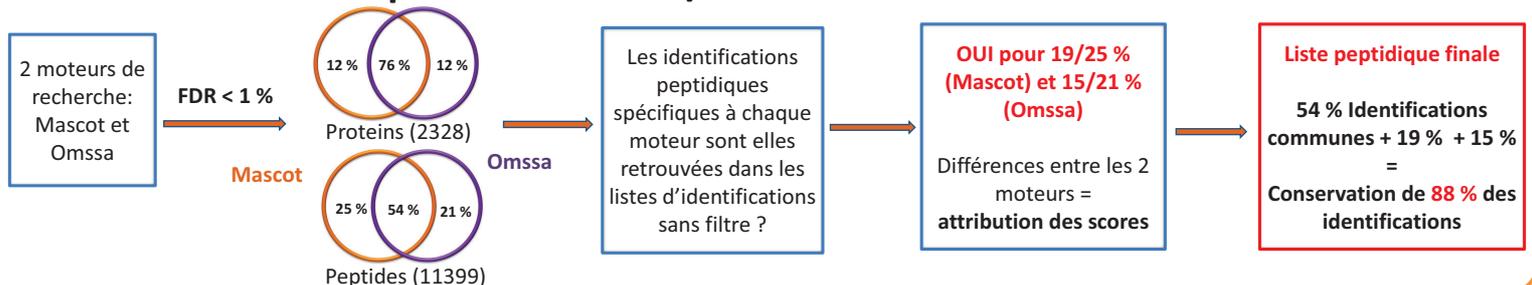
Nombre de peptides identifiés dans chaque piste pour les 13 protéines les plus abondantes



Nombre supplémentaire d'identification de protéines annotées « plasma membrane » dans chaque piste



Comparaison Mascot/Omsa (Comparaison réalisée pour la piste 1)



Conclusion

- Identification de plus de 3000 protéines, dont 599 annotées « plasma membrane » (soit 20 % de l'ensemble des protéines)
- Augmentation de la sensibilité de détection: identification de protéines telles que EGFR et CD221, jamais vues en protéomique auparavant ²
- Utilisation de deux moteurs de recherche pour augmenter le nombre d'identifications et la confiance que l'on peut leur accorder

1. Clevers, H. The cancer stem cell: premises, promises and challenges. *Nat Med* 17, 313-319 (2011).

2. Deighton, R.F., McGregor, R., Kemp, J., McCulloch, J. & Whittle, I.R. Glioma pathophysiology: insights emerging from proteomics. *Brain Pathol* 20, 691-703 (2010).

Développements méthodologiques en spectrométrie de masse et en analyse protéomique pour la recherche de biomarqueurs de différents types de cancers

Résumé

La protéomique clinique est, depuis ses débuts, considérée comme une technique à fort potentiel pour la recherche de biomarqueurs des cancers. Cependant, cette quête reste à ce jour sur un constat d'échec. Ces dix dernières années, la communauté protéomiste a œuvré pour comprendre l'origine des biais et pour homogénéiser les résultats. Les objectifs de mon travail de thèse s'articulent, dans ce contexte, en deux parties :

- La première partie décrit l'ensemble des développements méthodologiques réalisés en analyse protéomique pour la recherche de biomarqueurs.

- La seconde partie est consacrée à l'application de ces développements sur trois projets : (i) La recherche de biomarqueurs à visée diagnostique dans le but d'améliorer la caractérisation des lymphomes des cellules B. (ii) L'étude des cellules souches cancéreuses des glioblastomes. (iii) La recherche de biomarqueurs prédictifs des rechutes du système nerveux central dans le cas de lymphomes diffus à grandes cellules B.

Mots clés : analyse protéomique, biomarqueurs, spectrométrie de masse, chromatographie liquide, lymphomes, cellules souches cancéreuses de glioblastomes

Abstract

At first, clinical proteomics received great enthusiasm for the search of biomarkers from different pathologies. However, up to date, the biomarker quest dwells on a failure. These last ten years, the proteomics community has worked to understand the origins of the bias and and to homogenize the results in order to improve the confidence in the identifications and the proposals of biomarkers. In this context my PhD thesis is divided in two parts :

- The first part describes the methodological developments done in order to obtain the most complete and confident proteome of one sample.

- The second part is devoted to the results obtained on three different projects: (i) Diagnostic biomarkers search in order to better characterize B cells lymphoma and particularly marginal zone lymphoma. (ii) Glioblastoma stem cell study to propose new therapeutic targets. (iii) The search for predictive biomarkers of the central nervous system in the case of diffuse large B cell lymphoma.

Keywords : proteomics, biomarkers, mass spectrometry, liquid chromatography, lymphoma, glioblastoma stem cells