



**HAL**  
open science

# Étude du protéome de tumeurs colorectales

Damien Besson

► **To cite this version:**

Damien Besson. Étude du protéome de tumeurs colorectales. Biologie cellulaire. Université d'Angers, 2013. Français. NNT: . tel-00951752

**HAL Id: tel-00951752**

**<https://theses.hal.science/tel-00951752>**

Submitted on 25 Feb 2014

**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.

# ÉTUDE DU PROTÉOME DE TUMEURS COLORECTALES

THÈSE DE DOCTORAT

SPÉCIALITÉ : Biologie Cellulaire  
École Doctorale Biologie Santé Nantes-Angers

Présentée et soutenue publiquement

le 22 octobre 2013

À Angers, par

Damien BESSON

Devant le jury ci-dessous :

Michel SÈVE, rapporteur, Professeur des Universités - Praticien hospitalier, Grenoble  
Jérôme SOLASSOL, rapporteur, Maître de Conférence des Universités - Praticien hospitalier, Montpellier  
Frédéric NOTEL, invité, Fondateur et dirigeant d'Oncothérex, Nantes  
Olivier COQUERET, co-directeur de thèse, Professeur des Universités, Angers  
Catherine GUETTE, directrice de thèse, Chercheur-Biologiste, Angers

Directeurs de thèse :

Dr Catherine GUETTE

Pr Olivier COQUERET

INSERM U892

Institut de cancérologie de l'Ouest Paul Papin

2 rue Moll

49933 Angers Cedex 9



*À mon amour,  
À ma fille,  
À mes parents*

*« La beauté de la science et de l'art est  
consolatrice. » Jean Jaurès*



## Remerciements

MES attentions premières s'adressent sans hésitation à ma directrice de thèse, Madame le Docteur Catherine GUETTE. Qu'elle trouve ici l'expression de ma reconnaissance la plus sincère. Je la remercie pour sa confiance, son enthousiasme, sa rigueur scientifique, et sa grande disponibilité qui ont permis l'aboutissement de ce travail dans les meilleures conditions. Elle a su me faire découvrir et aimer la chimie, par l'intermédiaire de la spectrométrie de masse. Puisse ce travail répondre à ses attentes.

Un grand merci également à Monsieur le Professeur Olivier Coqueret, co-encadrant de ce travail et directeur du laboratoire. Merci à lui de m'avoir accueilli et de m'avoir fait confiance. Son expertise et ses conseils ont très largement alimenté mes réflexions. Merci également de m'avoir permis de côtoyer l'enseignement en m'offrant la possibilité d'encadrer des travaux pratiques. La complémentarité de mes deux directeurs de thèse a permis de nourrir des discussions scientifiques riches et constructives.

Je remercie Monsieur le Professeur Régis Bataille, directeur de l'Institut de Cancérologie de l'Ouest dans lequel j'ai pu réaliser cette étude.

Un grand merci à Messieurs le Docteur Jérôme Solassol et le Professeur Michel Sève d'avoir accepté de consacrer une partie de leur temps pour l'examen de ce travail. Je remercie Monsieur Solassol pour nous avoir permis d'accéder à des échantillons de tumeurs de patients.

Merci à Monsieur le Docteur Frédéric Notel pour sa participation au jury de cette thèse et pour avoir examiné mes travaux.

Je tiens également à remercier Monsieur le Docteur Benjamin Barré (autrement appelé Ben) pour sa bonne humeur, sa vivacité scientifique et son intérêt porté à mon sujet. Je le remercie également pour son aide précieuse lorsque j'ai été livré en pâture aux étudiants de licence.

Un grand merci à Messieurs les Docteurs Fabrice Gouilleux et François Guillonnet pour leur participation à mes comités de suivi de thèse. Leurs conseils avisés auront été

d'une aide précieuse.

J'adresse mes remerciements les plus sincères à tous les membres du laboratoire d'anatomopathologie de l'Institut de Cancérologie de l'Ouest. Plus particulièrement, je souhaite remercier Mesdames les Docteurs Isabelle Valo, Véronique Verrièle et Farida Benallegue pour leur aide précieuse dans l'analyse des lames d'immunohistochimie. Je n'oublie pas non plus Aude-Hélène Pavageau et Caroline Eymerit-Morin, qui étaient toutes deux internes au moment de notre rencontre et qui m'ont aidé à y voir plus clair dans les structures cellulaires tumorales. Un grand merci également à Chantal et Cécile, pour leur bonne humeur matinale. Nous avons l'habitude d'arriver au laboratoire à des horaires similaires.

Merci également à Madame le Docteur Agnès Chassevent, à Madame le Docteur Michelle Boisdrion-Celle, et à Monsieur le Professeur Alain Morel pour ce qu'il m'ont apporté durant cette thèse. Je n'oublie pas non plus Monsieur le Docteur Mario Campone et Monsieur le Professeur Éric Gamelin pour leur aide.

À tous les membres de l'U892 avec qui nous avons eu le plaisir de partager quelques séances scientifiques, un grand merci. Je pense notamment à Monsieur de Docteur Philippe Juin, à Madame le Docteur Sophie Barillé. Je remercie également Fabien, sans qui les échanges avec les labos de Nantes auraient été plus compliqués. Merci aussi à Nourdine pour son dévouement à la cause du logiciel libre et de GNU/Linux. Il est sans doute le plus chevronné de tous ceux que j'ai réussi à convaincre.

Un grand merci à Messieurs les Docteurs Pierre Bigot et Souhil Lebdaï. Ce fût un plaisir de travailler avec vous. Votre soif d'apprendre et votre humilité feront probablement de vous des médecins qu'il fait bon consulter! Un grand merci également à Céline et Caroline, pour leur bonne humeur respective, et joie de vivre. J'en profite ici pour remercier chaleureusement Monsieur le Docteur Sory Thraoré pour nos longues discussions philosophico-politiques dans le bureau, ainsi que pour son aide dans ce domaine sibyllin que constituent les statistiques pour un non initié.

Je présente également mes remerciements aux anciens étudiants qui sont passés par la protéomique et qui m'ont beaucoup appris. Émilie, Anthony, je vous remercie très sincèrement. Merci également à Charlène et Jing pour leur passage au labo. Véronique, merci pour nos discussions d'une grande richesse scientifique, et bon courage dans la protéomique!

Une place toute particulière doit être faite dans ces remerciements pour Alice, Nicolas et leurs enfants, sans qui cette thèse aurait eu une toute autre saveur. Merci à tous deux

pour ces moments de complicité, ces échanges enrichissants, et ces gros moments de délire. Merci également pour votre soutien quotidien. Alice, je n'oublierai pas ces moments passés à discuter de tout, de rien, et parfois de sujets sérieux et profonds. Recevez, à travers ces lignes, toute mon amitié.

Je pense également à Bernard Oudart et à Laurent Postolle pour qui les entrailles des MALDI TOF/TOF 4800 et 5800 n'ont aucun secret. Merci pour leur efficacité et leur réactivité lors des (quelques?) pannes rencontrées.

J'ai également une petite pensée nostalgique pour tous les thésards qui sont partis du labo, et dont les conseils auront toujours été précieux. Un grand merci à Sophie, qui a été la première à partir vers d'autres cieux, et sans qui les fautes d'orthographe et de français ont fait leur retour dans le labo. Ton départ a laissé un grand vide. Merci également à Hélène, qui a su réaliser un parcours sans faute durant sa thèse, et qui a été, et demeure un modèle pour moi. Ton petit brin de folie associé à ta rigueur font de toi une scientifique remarquable. Parmi les « anciens », je n'oublie pas non plus Laurence et Zaza. Ces deux compères vous amènent organisation et franche rigolade dans un laboratoire. J'envie ceux chez qui elles sont parties.

Je présente également mes remerciements à Audrey. J'ai été embarqué en même temps qu'elle dans cette aventure, cela crée forcément des liens. Je n'oublierai pas ces moments à rédiger en même temps, et nos éternels : « alors, tu en es où ? ». Merci aussi pour tous ces moments à rigoler de tout et de rien, pour tes prêts de voiture lorsqu'il fallait aller à la fac pour gérer l'aspect administratif de cette thèse ou aller en cours. Je te souhaite une belle carrière !

Un grand merci à *Chacha*, ta bonne humeur et ton dynamisme quotidien auront souvent été un moteur. Tu as toujours été là pour me soutenir, m'encourager, m'aider. Merci pour ta disponibilité et ton expertise. J'espère que la vie te réserve une belle surprise pour l'après. Je vous remercie également, toi et Thomas, pour les quelques soirées passées dans votre canapée après des soirées poker ou dégustation de spécialités culinaires. D'ailleurs, nos discussions gastronomiques risquent de beaucoup me manquer !

Merci également à David (oui désolé je fais l'impasse sur le : « Monsieur le Docteur David Lam » je trouvais que ça ne s'y prêtait pas) pour ces délires dans le labo à raconter n'importe quoi avec la voix la plus ridicule possible. Merci également pour tout ce que tu nous as apporté, tant au niveau technique (merci au passage pour ton aide sur la mutagénèse dirigée), qu'au niveau de la réflexion scientifique. Merci également de nous avoir fait partager ton amour de l'Oriflamme (non, il ne s'agit pas de l'étendard de

Charlemagne, mais bien de la pâtisserie qui se situe à côté de la gare d'Angers).

Merci à Barbara, souvent de corvée de décongélation de cellules pour tout le monde ! Ton calme et ton sérieux t'aideront certainement pour aller très loin, et apportent au labo un peu de sérénité. J'espère cependant que tu réussiras à soigner ton addiction au tarot. Bon courage pour ta dernière année.

Alexandra, si tu as le temps de lire ces remerciements c'est probablement que tu as oublié ton *timer* quelque part et qu'il est en train de sonner pour te rappeler que tu dois couler huit gels pour ton prochain *western-blot*. Quoiqu'il arrive, si un jour tu as un doute, c'est forcément 8 mL ! Merci pour tous ces moments passés à la paillasse à discuter pendant nos manips.

Je remercie également Julien et son excellent caractère ! Non, attends, ne repars pas tout de suite, je n'ai pas terminé. Il est toujours pressé celui-là. Je tiens à te souhaiter un bon courage pour la suite de ta thèse. Ta curiosité et ta soif de savoir sont des atouts précieux. Profite de chaque instant et ménage toi !

Alexis, Marie, nos chemins en thèse vont se croiser. Je vous souhaite de vous épanouir dans cette aventure ô combien enrichissante. J'aurais aimé apprendre à vous connaître un peu plus. Vous avez l'air d'être de belles personnes... Bon courage !

Une petite place également pour remercier mon frère jumeau, Simon, né le même jour que moi, la même année mais à quelques 800 km d'écart ! Ton passage au labo aura laissé des traces. Ton art du dessin, ta culture, ton sens de l'humour m'auront souvent aidé à garder le sourire durant ces derniers mois. Je te souhaite de trouver un travail qui répondra à tes attentes sur Montpellier. Merci également à ta moitié, qui porte le même prénom que la mienne, et dont la joie de vivre est très contagieuse.

Je remercie les salariés de l'ICO, et plus particulièrement les membres de l'équipe d'oncopharmacologie. Merci à Marie-Claire, Vincent, Adélaïde, Sabrina, Audrey, Marie, Julie, Luc, Sophie, Paula et Catherine pour la bonne humeur qu'ils dégagent et les conditions de travail très agréables qu'ils ont instaurées. Merci pour ces pauses cafés qui s'accompagnaient souvent de délicieux gâteaux confectionnés avec beaucoup de talent. Une attention particulière pour Marie-Claire, toujours prête à prendre la défense des *jeunes du 4<sup>e</sup>* et à soutenir notre cause ! Merci pour les brioches au chocolat blanc de la boulangerie de Cantenay, et la terrine au saumon.

Merci au service informatique du Centre pour leur aide tout au long de ces trois années.

Un grand merci à Jonathan et à Sylvain pour ces midis passés au restaurant asiatique à refaire le monde. Merci également à tous leurs compères de l'avenue Pasteur à Angers.

Le document que vous avez entre les mains constitue la raison principale de mes absences à vos réunions. Chers amis, recevez toute mon amitié!

Chers amis manceaux, ne pensez pas que j'allais vous oublier. Merci à Bouba bien sûr, lui qui a si gentilleusement accepté d'être témoin de mon mariage. Je n'oublie pas que tu dois m'emmener faire un tour de moto. Merci pour ta patience, et ton amitié, et pour ce repas de mariage qui restera inoubliable. Merci Yahnou pour ces soirées Counter Strike (avec Bouba), les débats politiques entre 23h et 3h du matin. Un grand merci à Marie et Mélanie, d'avoir mis au monde notre fille dans des conditions des plus privilégiées. Merci pour les repas à la chevalerie et pour votre bonne humeur. Merci à Aurélie et Luc et leur petite Augustine pour les soirées barbecue, et autres repas vietnamiens et créoles. Luc, rappelle-moi de prendre rendez-vous dans ton cabinet lorsque tout cela sera terminé, il faut vraiment que je fasse quelque chose pour mes dents! Merci également à tous les collègues sage-femme de Charlotte (Véro, Arnault, p'tit kiwi, Floriane, Guen, Céline, et tous les autres) d'être ce qu'ils sont, des gens qui aiment la vie et les autres. Merci de m'avoir fait une petite place dans vos soirées.

Je n'oublierai pas non plus ceux qui sont parti du Mans. Merci Anna pour ta participation à notre mariage en qualité de témoin. Merci pour toutes ces soirées détentes entre le Maine-et-Loire et la Sarthe. C'est marrant de s'être formé dans des villes très différentes l'un et l'autre pour finalement travailler au même endroit! Merci aux Nantais Agathe, Yoann et leur petite Leila. Malgré la distance, c'est toujours un plaisir de se voir. Merci pour votre soutien et votre intérêt constant pour ce que je fais. Merci également, Agathe d'avoir accepté d'être des nôtres le jour de notre mariage en tant que témoin. Tant que nous sommes du côté de Nantes, merci également à Olivier et Tiphanie, toujours enclins à nous recevoir à l'improviste lors de nos virées sur Nantes. Merci pour tout! Faisons maintenant un détour vers Paris. Je remercie très vivement Alban et Virginie. Merci beaucoup Alban, sans toi, cette thèse aurait probablement été écrite depuis Windows, et avec Word! Merci pour tes précieux conseils en informatique, pour nos discussions sur twitter et pour avoir été notre témoin de mariage. Merci Virginie pour ta bonne humeur, ta franchise, et ton talent de chanteuse. Ces concerts sur Paris resteront de très bons souvenirs.

Je ne peux pas ne pas remercier les développeurs de la distribution GNU/Linux Debian, que j'utilise au quotidien. Ils font un travail remarquable et leur réactivité est vraiment exceptionnelle. Merci également aux développeurs de  $\LaTeX$ , qui ont permis l'écriture de cette thèse dans des conditions des plus sereines.

J'adresse mes tendres remerciements à ma famille. Merci à mes parents, à qui je dédie

ce travail, pour leur soutien constant et leur intérêt. Vous avez toujours sû m'encourager et m'accompagner dans ce que je voulais faire. Je vous en suis extrêmement reconnaissant. Merci pour toute l'aide que vous m'avez apportée durant ces longues années d'études.

Merci également à mon frère (et Aurore), et ma sœur pour tout ce que vous représentez pour moi. Désolé d'avoir été un grand frère si occupé.

Une pensée également pour ma belle famille, que je tiens à remercier chaleureusement pour ses encouragements constants. Sophie, Lionel, recevez toute ma considération ! Merci également à Pascal pour tout ce qu'il a fait pour moi lors de ma formation universitaire.

Enfin, un grand merci à mon épouse à qui je dédie ce travail. Tu es certainement la personne qui comprend le mieux ce que j'ai vécu durant ces trois années. Un grand merci pour ta compréhension, pour tes conseils précieux, et ta relecture. Nous avons vécu tellement de choses ensemble durant ces trois années que les mots me manquent pour t'exprimer ma reconnaissance. Nous nous sommes pacsés, puis mariés, nous avons acheté une maison ensemble, et voilà que maintenant nous sommes parents ! Merci d'être ce que tu es ! Un grand merci également à ma petite Jeanne, sans toi et ta petite frimousse cette fin de thèse aurait paradoxalement été beaucoup plus difficile.

# Table des matières

<b>Remerciements</b>	<b>xi</b>
<b>Table des matières</b>	<b>xvi</b>
<b>Liste des figures</b>	<b>xix</b>
<b>Liste des tableaux</b>	<b>xxii</b>
<b>Liste des abréviations</b>	<b>xxiii</b>
<b>I Introduction</b>	<b>1</b>
<b>1 Le cancer colorectal</b>	<b>3</b>
1.1 Données générales . . . . .	3
1.1.1 Données épidémiologiques . . . . .	3
1.1.2 Organisation fonctionnelle du côlon et du rectum . . . . .	4
1.1.2.1 Rappels anatomiques . . . . .	4
1.1.2.2 Histologie . . . . .	4
1.1.2.3 L'unité fonctionnelle : les glandes de Lieberkühn . . . . .	5
1.1.3 Histoire naturelle du cancer colorectal . . . . .	7
1.1.3.1 Le modèle de développement séquentiel . . . . .	7
1.1.3.2 STAT3 et NF $\kappa$ B dans le cancer colorectal . . . . .	9
1.2 Les classifications du cancer colorectal . . . . .	12
1.2.1 Classification histologique . . . . .	12
1.2.2 Classification clinique . . . . .	13
1.2.3 Classification moléculaire . . . . .	16
1.2.3.1 L'instabilité génomique dans le cancer colorectal . . . . .	16
1.2.3.2 L'instabilité épigénétique du cancer colorectal . . . . .	18
1.2.3.3 Les mutations fréquentes . . . . .	18
1.3 Les traitements . . . . .	21

1.3.1	Un traitement principal : la chirurgie . . . . .	21
1.3.2	Présentation des traitements médicamenteux . . . . .	21
1.4	L'apoptose . . . . .	22
1.5	La sénescence . . . . .	23
1.5.1	Définition . . . . .	23
1.5.2	Le SASP . . . . .	24
1.6	SASP et sécrétomes tumoraux . . . . .	27
<b>2</b>	<b>Les sources de biomarqueurs</b>	<b>29</b>
2.1	Définition . . . . .	29
2.2	Les caractéristiques des biomarqueurs . . . . .	30
2.2.1	Rappels sur les vrais/faux négatifs et les vrais/faux positifs . . . . .	30
2.2.2	La sensibilité . . . . .	31
2.2.3	La spécificité . . . . .	31
2.2.4	L'exactitude . . . . .	31
2.3	De l'intérêt des biomarqueurs dans le cancer colorectal . . . . .	32
2.4	Classification . . . . .	32
2.4.1	Les différents types de biomarqueurs . . . . .	33
2.4.1.1	Les biomarqueurs pronostiques . . . . .	33
2.4.1.2	Les biomarqueurs prédictifs . . . . .	34
2.4.1.3	Les biomarqueurs pharmacodynamiques . . . . .	35
2.4.1.4	Les biomarqueurs diagnostiques . . . . .	36
2.4.1.5	Les biomarqueurs de type ADN . . . . .	36
2.4.1.6	Les biomarqueurs de type ARN : les miRNA . . . . .	37
2.4.1.7	Les biomarqueurs de type protéique . . . . .	38
2.4.1.8	Les biomarqueurs de type glucidique . . . . .	38
2.4.1.9	Les biomarqueurs associés à l'imagerie médicale . . . . .	38
2.5	Les sources de biomarqueurs protéiques potentiels . . . . .	39
2.6	Le sécrétome . . . . .	40
2.6.1	Définition . . . . .	40
2.6.2	Les différentes voies de sécrétion . . . . .	41
2.6.2.1	La voie de sécrétion classique . . . . .	42
2.6.2.2	La voie de sécrétion non classique . . . . .	42
2.6.2.3	Le relargage par des exosomes . . . . .	43
2.6.3	Intérêts . . . . .	44
2.7	Biomarqueurs protéiques potentiels et protéomiques . . . . .	44

2.7.1	À partir de modèles <i>in vitro</i> . . . . .	44
2.7.2	À partir de modèles <i>in vivo</i> . . . . .	46
2.7.3	À partir de tissus biologiques . . . . .	48
<b>3</b>	<b>Identification de biomarqueurs par protéomique</b>	<b>51</b>
3.1	Définition de la protéomique . . . . .	51
3.2	Le fractionnement protéique . . . . .	52
3.2.1	L'électrophorèse à une dimension . . . . .	52
3.2.2	L'électrophorèse bidimensionnelle . . . . .	52
3.2.3	La chromatographie . . . . .	54
3.2.4	L'approche MudPIT . . . . .	55
3.2.5	La méthode OFFGEL . . . . .	56
3.3	La spectrométrie de masse en protéomique . . . . .	57
3.3.1	Principe général . . . . .	57
3.3.1.1	L'ionisation MALDI . . . . .	58
3.3.1.2	Les analyseurs de masse . . . . .	60
3.3.1.3	Spectrométrie de masse en tandem . . . . .	63
3.3.1.4	Le détecteur . . . . .	63
3.3.2	Deux grands types d'approche . . . . .	64
3.4	Protéomique quantitative . . . . .	65
3.4.1	La quantification basée sur gels . . . . .	66
3.4.1.1	Comparaison de gels 2D . . . . .	66
3.4.1.2	2D-DIGE . . . . .	66
3.4.2	La quantification basée sur marquage isotopique . . . . .	67
3.4.2.1	Principe général . . . . .	67
3.4.2.2	Le marquage SILAC . . . . .	68
3.4.2.3	Le marquage à l' <sup>18</sup> O . . . . .	70
3.4.2.4	Le marquage ICAT . . . . .	70
3.4.2.5	Le marquage iTRAQ . . . . .	72
3.4.3	La quantification sans marquage . . . . .	79
3.4.4	Quantification par SRM/MRM . . . . .	80
3.4.5	La quantification absolue . . . . .	81
3.4.5.1	Quantification absolue avec marquage isotopique . . . . .	81
3.5	Glycoprotéomique et biomarqueurs . . . . .	82
3.5.1	La glycosylation . . . . .	82
3.5.2	Pourquoi étudier le glycoprotéome . . . . .	85

3.5.3	Glycoprotéomique . . . . .	86
3.5.4	Les modes d'enrichissement . . . . .	87
3.5.4.1	Les lectines . . . . .	87
3.5.4.2	Les résines hydrazines . . . . .	88
3.5.4.3	Une information importante : les sites de glycosylation . . . . .	89
3.5.5	La technique FASP et ses avantages . . . . .	89
 <b>II Matériels et méthodes</b>		 <b>93</b>
<b>1</b>	<b>Détail des techniques utilisées</b>	<b>95</b>
1.1	Les tumeurs . . . . .	95
1.2	Génération des coupes de tumeurs et microdissection laser . . . . .	95
1.3	Extraction des protéines, digestion tryptique et dosage . . . . .	96
1.4	Réduction, Alkylation, marquage iTRAQ . . . . .	97
1.5	Isoélectrofocalisation OFFGEL . . . . .	98
1.6	La séparation par RPLC . . . . .	98
1.7	Analyse par spectrométrie de masse MALDI TOF/TOF . . . . .	99
1.8	Les outils d'analyse en protéomique . . . . .	100
1.9	L'immunohistochimie . . . . .	101
1.10	Western Blot . . . . .	102
1.11	ChIP . . . . .	103
1.12	Les modèles cellulaires . . . . .	103
1.13	Conditions de culture . . . . .	104
1.14	Préparation des extraits protéiques cellulaires . . . . .	105
1.15	ARN interférence . . . . .	105
1.16	Formation des sphères . . . . .	105
1.17	Les tests ELISA de l'OLFM4 . . . . .	106
1.18	Tests de clonogénicité . . . . .	106
1.19	Le FASP . . . . .	106
1.20	L'enrichissement en glycopeptides . . . . .	107
1.20.1	Enrichissement sur lectines . . . . .	107
1.20.2	Enrichissement sur billes hydrazines . . . . .	108
1.21	La mutagénèse dirigée de l'OLFM4 . . . . .	108
1.22	L'étude du stroma tumoral . . . . .	109

<b>III</b>	<b>Résultats</b>	<b>115</b>
<b>1</b>	<b>Analyse protéomique de tumeurs colorectales</b>	<b>117</b>
1.1	Problématique . . . . .	117
1.2	Publication . . . . .	118
1.3	Figure complémentaire . . . . .	133
<b>2</b>	<b>Résultats complémentaires concernant l'OLFM4</b>	<b>135</b>
2.1	L'OLFM4 un marqueur de détection des tumeurs précoces . . . . .	135
2.2	L'OLFM4 : une protéine glycosylée . . . . .	136
2.2.1	Détermination des sites de glycosylation de l'OLFM4 . . . . .	136
2.2.1.1	Approche bioinformatique . . . . .	136
2.2.1.2	Sites de glycosylation par lectines . . . . .	137
2.2.1.3	Identification de sites de glycosylations grâce à l'enrichissement basée sur les billes d'hydrazine . . . . .	138
2.2.2	Validation des sites de glycosylation par mutagenèse dirigée . . . . .	140
2.3	OLFM4 et survie cellulaire . . . . .	141
2.4	OLFM4 et pousse en 3D . . . . .	143
<b>3</b>	<b>Résultats complémentaires</b>	<b>145</b>
3.1	Analyse génétique de tumeurs colorectales . . . . .	145
3.2	OLFM4 et oncogène RAS . . . . .	147
3.3	Confirmation du lien entre l'OLFM4 et KRAS . . . . .	148
3.4	KRAS et OLFM4 par ELISA . . . . .	150
<b>4</b>	<b>Étude du stroma de tumeurs colorectales</b>	<b>153</b>
4.1	Microdissection laser . . . . .	153
4.2	Résultats de l'étude par spectrométrie de masse . . . . .	154
<b>5</b>	<b>Biomarqueurs et lignées cellulaires</b>	<b>157</b>
5.1	Une approche protéomique globales . . . . .	157
5.2	Amélioration de la couverture protéomique . . . . .	158
5.2.1	Publication . . . . .	159
5.3	Analyse du sécrétome des LS180 . . . . .	172
<b>6</b>	<b>Glycoprotéomique</b>	<b>173</b>

<b>IV</b>	<b>Discussion</b>	<b>177</b>
<b>1</b>	<b>Discussion des résultats obtenus</b>	<b>179</b>
1.1	L'analyse protéomique par stade tumoraux . . . . .	179
1.2	Analyse des tumeurs et statut de l'oncogène KRAS . . . . .	183
1.3	Analyse du stroma tumoral . . . . .	183
1.4	Analyse du glycoprotéome tumoral . . . . .	184
1.5	Analyse de lignées cellulaires . . . . .	185
<b>V</b>	<b>Conclusion et perspectives</b>	<b>191</b>
	<b>Références bibliographiques</b>	<b>195</b>
	<b>Annexes</b>	<b>213</b>

## Table des figures

1.1	Organisation histologique de la muqueuse colique . . . . .	6
1.2	Organisation cellulaire des glandes de Lieberkühn . . . . .	7
1.3	Schéma simplifié de la voie Wnt . . . . .	8
1.4	Schéma simplifié de la voie de signalisation médiée par STAT3 . . . . .	11
1.5	Shéma illustrant la classification par stades en fonction de l'invasion de la muqueuse intestinale par la tumeur . . . . .	15
1.6	Représentation schématique de la voie de signalisation médiée par le récepteur à l'EGF et de l'effet du cétuximab sur cette voie. . . . .	20
1.7	Les mécanismes d'activation de la sénescence. . . . .	24
1.8	Schéma des quatre champs d'action du SASP. . . . .	25
2.1	Courbes de survie de patients atteints de cancers colorectaux en fonction du stade tumoral et du taux de CEA en post-opératoire. . . . .	34
2.2	Shéma illustrant la différence entre les biomarqueurs pronostiques et prédictifs . . . . .	35
2.3	Schéma illustrant les différentes voies de sécrétion . . . . .	41
3.1	Principe général de l'électrophorèse bidimensionnelle . . . . .	54
3.2	Schéma simplifié du fonctionnement de la chromatographie liquide haute performance. . . . .	55
3.3	Schéma simplifié représentant la stratégie MudPIT . . . . .	55
3.4	Schéma de fonctionnement du fractionnement OFFGEL . . . . .	57
3.5	Représentation schématique de la structure générale d'un spectromètre de masse – Un spectromètre de masse est constitué de quatre éléments principaux : la source d'ionisation, l'analyseur, le détecteur, et un système informatique permettant l'intégration des données . . . . .	58
3.6	Représentation schématique du mode d'ionisation MALDI . . . . .	59

3.7	<b>Représentation schématique d'un analyseur en temps de vol – TOF –</b>	
	Les ions circulent dans un tube dans lequel un vide poussé est ménagé de sorte que la masse des ions soit proportionnelle au temps qu'ils vont mettre à venir frapper le détecteur. . . . .	61
3.8	<b>Représentation schématique du fonctionnement du mode réflectron . .</b>	62
3.9	<b>Représentation schématique du fonctionnement du mode tandem . . .</b>	63
3.10	<b>Les approches <i>top-down</i> et <i>bottom-up</i> en protéomique . . . . .</b>	64
3.11	<b>Principe générale de la méthode DIGE . . . . .</b>	67
3.12	<b>Niveau de marquage selon les approches utilisées . . . . .</b>	68
3.13	<b>Principe générale de la méthode SILAC . . . . .</b>	69
3.14	<b>Principe générale de la méthode de marquage enzymatique avec utilisation de l'eau lourde . . . . .</b>	71
3.15	<b>Structure chimique du réactif chimique ICAT de première génération .</b>	72
3.16	<b>Principe général de la méthode ICAT . . . . .</b>	73
3.17	<b>Formule chimique du réactif iTRAQ<sup>TM</sup> dans sa version 4-plex . . . . .</b>	73
3.18	<b>Formule chimique du réactif iTRAQ<sup>TM</sup> dans sa version 8-plex . . . . .</b>	74
3.19	<b>Schéma du principe général de marquage iTRAQ<sup>TM</sup> de type 4-plex . . .</b>	74
3.20	<b>Schéma du principe général du fonctionnement du logiciel iQuantitator</b>	75
3.21	<b>Schéma montrant l'évolution de la N-glycosylation au cours de l'avancée des protéines dans le réticulum endoplasmique et dans l'appareil de golgi . . . . .</b>	84
3.22	<b>Réaction de fixation sur la résine hydrazine . . . . .</b>	89
3.23	<b>Réaction de déglycosylation dans l'H<sub>2</sub><sup>18</sup>O . . . . .</b>	90
3.24	<b>Schéma illustrant les étapes nécessaires à la réalisation d'une extraction peptidique par l'approche FASP . . . . .</b>	91
1.1	<b>Comparaison entre le tissu colique sain et le tissu tumoral . . . . .</b>	96
1.2	<b>Schéma de l'approche utilisée . . . . .</b>	113
1.1	<b>Protéines statistiquement dérégulées dans les différents stades du cancer colorectal . . . . .</b>	133
2.1	<b>L'OLFM4 est un marqueur potentiel de tumeurs précoces . . . . .</b>	135
2.2	<b>Séquence protéique de l'OLFM4 . . . . .</b>	137
2.3	<b>Validation des sites de glycosylation par <i>western blot</i> . . . . .</b>	141
2.4	<b>L'OLFM4 améliore la survie des cellules tumorales. . . . .</b>	142
2.5	<b>L'olfactomedine 4 induit une diminution de la taille des sphéroïdes . .</b>	144

3.1	Schéma expérimental . . . . .	146
3.2	Western blot de validation de l'expression de l'OLFM4 dans les lignées LS174t, LS180, SW1463, et SW403 . . . . .	149
3.3	L'OLFM4 est fortement exprimée dans les tumeurs mutées sur l'onco- gène RAS . . . . .	151
3.4	L'OLFM4 est un marqueur potentiel de la mutation de KRAS . . . . .	152
4.1	Microdissection laser du stroma tumoral . . . . .	153
4.2	Analyse du protéome issu du stroma de deux tumeurs coliques . . . . .	155



## Liste des tableaux

1.1	La classification TNM . . . . .	14
1.2	Correspondance TNM, stades cliniques et survie à 5 ans . . . . .	15
2.1	Les notions de vrais/faux négatifs et de vrais/faux positifs . . . . .	31
3.1	Comparaison des principaux colorants protéiques utilisés pour révéler les gels 2D. . . . .	53
3.2	Comparaison des principaux analyseurs de masse . . . . .	60
1.1	Amorces de mutagénèse utilisées pour muter l'OLFM4 . . . . .	109
2.1	Sites de glycosylation potentiels de l'OLFM4 par des outils bioinformatiques . . . . .	137
2.2	Sites de glycosylation potentiels de l'OLFM4 par une approche expérimentale utilisant les lectines . . . . .	138
2.3	Sites de glycosylation potentiels de l'OLFM4 par une approche expérimentale utilisant billes hydrazines . . . . .	138
2.4	Liste et masse des peptides tryptiques de l'OLFM4 détectables entre 1 000 et 4 000 Da . . . . .	139
2.5	Liste et masse des peptides issus de la digestion pas l'enzyme Glu-C de l'OLFM4 . . . . .	140
3.1	Résultats de l'analyse des mutations sur 79 tumeurs du côlon . . . . .	145
3.2	Résultat de l'analyse protéomique de tumeurs mutées sur KRAS et non mutées comparées à un pool de tissu sain . . . . .	147
3.3	Caractéristiques des tumeurs analysées par western blot . . . . .	148
3.4	Niveau d'expression de l'ARNm de l'OLMF4 dans des lignées de cellules colorectales en fonction du statut de RAS. . . . .	149
6.1	Comparaison des résultats de protéomique issus de l'analyse globale du protéome d'un patient et de son glycoprotéome correspondant . . . . .	175

1	Tableau des protéines dont l'expression varie significativement (selon l'analyse par iQuantitator) en fonction des stades tumoraux . . . . .	213
2	Tableau des protéines dont l'expression varie significativement (selon l'analyse par iQuantitator) chez les tumeur présentant une mutation de l'oncogène KRAS . . . . .	221
3	Tableau des protéines sécrétées par les tumeurs et les lignées du cancer colorectal détectées dans la bases de données issue de la littérature (226 protéines) . . . . .	225
4	Tableau des 245 protéines identifiées par l'analyse du stroma tumoral .	229
5	Tableau des 53 protéines glycosylées identifiées dans une tumeur de côlon par enrichissement sur des lectines . . . . .	233

## Liste des abréviations

2D	2 Dimensions
3'UTR	3' UnTranslated Region
5-FU	5-Fluorouracile
ADC	Apparent Diffusion Coefficient
ADN	Acide Désoxyribonucléique
Ago	Argonaute
AGR2	Anterior Gradient homolog 2
APC	Adenomatous Polyposis Coli
Apo B-100	Apolipoprotein B-100
AQUA	Absolute QUAntification
ARN	Acide Ribonucléique
ARNm	ARN messenger
ATP	Adénosine TriPhosphate
Bax	Bcl2-associated X protein
Bcl-xl	B-cell lymphoma gene XL
Bcl2	B-cell lymphoma gene 2
BFA	BreFeldine A
C-terminal	Carboxy terminal
CBP	cAMP response element Binding Protein
CCL	CC chemokine
CD	Cluster of Differentiation
CDH17	Cadherin-17
CEA	CarcinoEmbryonic Antigen
CEACAM5	CarcinoEmbryonic Antigen-Related Cell Adhesion Molecule 5
CHCA	$\alpha$ -Cyano-4-HydroxyCinnamic Acid
CHI3L1	Chitinase-3-like protein 1
CID	Chambre de collision
CIMP	CpG Island Methylator Phenotype
CIMP-H	CIMP-High
CIMP-L	CIMP-Low
CIN	Chromosomal Instability
CLHP	Chromatographie Liquide Haute Performance
CMH	Complexe Majeur d'Histocompatibilité
ConA	Concanavaline A
CpG	Cytosine-phosphate-Guanine
CRLCC	Centre régional du lutte contre le cancer
CRMP-2	Collapsin Reponse Mediator Protein-2

CXCL	C-X-C motif chemokine
Da	Dalton
DHB	2,5-DiHydroxyBenzoic acid
DIGE	Difference In Gel Electrophoresis
DTT	DiThioThreitol
e	Charge élémentaire
E-Cadhérine	Epithéliale Cadhérine
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-Linked ImmunoSorbent Assay
emPAI	Exponentially modified Protein Abundance Index
ESI	ElectroSpray Ionisation
FASP	Filter-Aided Sample Preparation
FDA	Food and Drug Administration
FN	Faux Négatifs
FP	Faux Positifs
FT-ICR	Fourier Transform Ion Cyclotron Resonance
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
GDF15	Growth Differentiation Factor 15
GDP	Guanosine DiPhosphate
GEF	Guanine nucleotide Exchange Factors
GFP	Green Fluorescent Protein
Glc	Glucose
GlcA	Acide Glucoronique
GlcNAc	N-acetylglucosamine
Glod4	Glyoxalase domain-containing protein 4
GLUT1	Glucose transporter 1
gp130	Glycoprotéine 130
GPI	Glycosylphosphatidyl inositol
Grb2	Growth factor receptor-bound protein 2
GTP	Guanosine TriPhosphate
HES	Hématoxiline, Éosine, Safran
HGF	Hepatocyte Growth Factor
HGPI	Human Disease Glycomics/Proteome Initiative
HILIC	Hydrophilic Interaction Liquid Chromatography
HPLC	High-Performance Liquid Chromatography
HPV	Human PapillomaVirus
HSP	Heat Shock Protein
hTERT	Human Telomerase Reverse Transcriptase
HUPO	Human Proteomics Organization
ICAT	Isotope Coded Affinity Tag

IdoA	Acide iduronique
IGF2R	Insulin-like growth factor 2 receptor
IGFBP7	Insulin-like growth factor-binding protein 7
IKK	I $\kappa$ B kinase
IL	Interleukine
INCa	Institut National du Cancer
IRM	Imagerie par Résonance Magnétique
iTRAQ	Isobaric Tag for Relative and Absolute Quantitation
JAK	Janus Kinase
JUP	Junction Plakoglobin
K-Ras ou KRAS	Kirsten Ras
kDa	Kilodalton
LC-MS	Liquid chromatography–mass spectrometry
LCN2	Lipocalin-2
Lef	Lymphoid enhancer-binding factor
Let-7	Lethal-7
Lgr5	Leucin-rich repeat-containing G-protein coupled receptor 5
LRP	Low-density lipoprotein Receptor-related Protein
<i>m</i>	Masse
MALDI	Matrix Assisted Laser Desorption Ionisation
Man	Mannose
MAP kinases	Mitogen Activated Protein kinase
MCM4	Mini-Chromosome Maintenance proteins 4
miRNA	MicroARN
MLH1	MutL homolog 1
MMP	Matrices Métallo-Protéinases
MMR	MisMatch Repair
MMTS	Méthyl Méthane ThioSulfonate
MS	Mass Spectrometry
MSH2	MutS protein homolog 2
MSH6	MutS protein homolog 6
MSI	MicroSatellite Instability
MSI-H	MSI-High
MSI-L	MSI-Low
MSS	Microsatellite Stability
MudPIT	Multidimensional Protein Identification Technology
N-terminal	Amino-terminal
NCI	National Cancer Institute
NF $\kappa$ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NH <sub>2</sub>	Groupement amine
NHS	N-hydroxy-succinimide
NIH	National Institute of Health
Oct4	Octamer-binding transcription factor 4
OLFM4	Olfactoméline 4

OMS	Organisation Mondiale de la Santé
OMS	Organisation Mondiale de la Santé
OST	OligoSaccharylTransferase
PAI	Protein Abundance Index
PDGF	Platelet-Derived Growth Factor
pH	Potentiel hydrogène
pI	point Isoélectrique
PI3K	PhosphoInositol-3-kinase
PMS2	PostMeiotic Segregation increased 2
PNGaseF	Peptide-N4-(N-acetyl-beta-glucosaminy)asparagine amidase
PSA	Prostate-Specific Antigen
PSAQ	Protein Standard Absolute Quantification
PSAQ	Protein Standard Absolute Quantification
Ras	Rat Sarcoma
Rb	Retinoblastome
RCA120	Ricinus Communis Agglutinin
RHD	Rel Homology Domain
RNAses	Ribonuclease
ROS	Reactive Oxygen Species
RPLC	Reversed Phase HPLC
SA	Acide Sialique
SASP	Senescence-Associated Secretory Phenotype
SCX	Strong Cations Exchange
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
SEREX	Serological expressionn analysis
SERPA	Serological proteome analysis
shARN	Small hairpin ARN
SILAC	Stable Isotope Labelling by Amino acids in Cell culture
SM	Spectrométrie de Masse
SMAD4	Mothers against decapentaplegic homolog 4
SN-38	7-éthyl-10-hydroxyl-camptothécine
Src	Sarcoma
SRM	Selected Reaction Monitoring
STATs	Signal Transducer and Activator of Transcription
STOML2	Stomatin-like protein 2
TCEP	Tris-(2-CarboxyÉthyl)Phosphine
Tcf	Transcription Factor
TEAB	TriEthylAmmonium Bicarbonate
TEP	Tomodensitométrie par Émission de Positons
TFA	TriFluoroAcetic acid
TFF3	Trefoil factor 3
TGF $\beta$	Transforming Growth Factor beta

TGM2	Tissue TransGlutaMinase 2
TIMP-1	Tissue inhibitor of metalloproteinases protein 1
TIS	Time Ion Selector
TMT	Tandem Mass Tags
TOF	Time Of Flight
TP53	Tumor protein 53
TRIS	Tris(hydroxymethyl)aminomethane
TYMS	Thymidylate Synthase
v	Vitesse
VEGF	Vascular Endothelial Growth Factor
VN	Vrais Négatifs
VP	Vrais Positifs
WGA	Wheat Germ Agglutinin
Wnt	Wingless iNT
Xyl	Xylose
z	Valence



# Première partie

## Introduction



# Chapitre 1

## Le cancer colorectal

### 1.1 Données générales

#### 1.1.1 Données épidémiologiques

CHACQUE année, à travers le monde, environ un million de personnes développent un cancer et près de 655 000 décèdent des suites de ce dernier.

En France, avec près de 40 500 nouveaux cas par an selon l'INCa<sup>1</sup>, le cancer colorectal est le deuxième en terme de fréquence chez la femme avec 19 000 cas, derrière le cancer du sein qui compte lui 53 000 cas. Il représente le troisième chez l'homme avec 21 500 cas, derrière le cancer du poumon (27 500 cas) et le cancer de la prostate (71 000 cas). En 2011, le cancer colorectal aurait fait 17 500 victimes parmi lesquelles 9 200 hommes et 8 300 femmes. Ces chiffres le placent en deuxième position au sinistre classement des cancers les plus meurtriers en France.

Malgré ces chiffres, d'importants progrès ont été réalisés dans la prise en charge des patients. Ces progrès sont dus à l'amélioration des traitements existants et à la mise en place de nouvelles associations thérapeutiques. Ainsi, malgré une augmentation de l'incidence<sup>2</sup>, la mortalité induite par ce cancer a diminué de 3% pour les femmes et de 4% pour les hommes, à travers le monde, et en vingt ans [Belot et al., 2008].

Pour continuer sur cette voie, il est nettement établi qu'une amélioration du dépistage, en terme quantitatif et qualitatif (précocité) permettrait d'augmenter les chances de survie des patients. En effet, comme indiqué dans le tableau 1.1 p.15, cette dernière est

---

1. Chiffre de l'Institut National du Cancer estimé pour 2011.

2. D'après les données actuelles, il semblerait que cette augmentation d'incidence tende à se stabiliser. Aux États-Unis, une diminution est même observée depuis 1985, mais pour des raisons encore inconnues, cette diminution n'est pas encore observée en France.

fortement augmentée lorsque le diagnostic est posé alors que la maladie n'est encore que débutante. Pour mieux comprendre le développement du cancer colorectal, il nous faut bien comprendre son organisation structurelle et fonctionnelle.

## 1.1.2 Organisation fonctionnelle du côlon et du rectum

### 1.1.2.1 Rappels anatomiques

Le système digestif est constitué du tube digestif et de l'ensemble des glandes qui y déversent leurs sécrétions par des canaux. Le tube digestif est constitué de structures anatomiques distinctes que sont la bouche, le système oropharyngé, l'œsophage, l'estomac, l'intestin grêle (*duodénum*, *jéjunum*, et *iléon*), du gros intestin (autrement appelé côlon) du rectum, et de l'anus. À ce tube digestif sont connectés des organes glandulaires : les glandes salivaires, le foie, la vésicule biliaire, et le pancréas.

Le côlon est un tube d'environ 1,2 m de long et d'environ 6 cm de diamètre. Sa première partie est formée par une poche appelée cæcum à laquelle est appendu l'appendice. Suivent ensuite trois segments que sont le côlon droit (ou ascendant), le côlon transverse, et le côlon gauche ou descendant. Ce dernier se prolonge par une partie terminale ayant une forme de S et constituant le côlon sigmoïde, lui-même se poursuivant par le rectum.

Les fonctions principales du côlon et du rectum sont de stocker, et de concentrer les matières fécales avant la défécation. Ils participent également à la digestion terminale de la cellulose par la flore bactérienne intestinale et à l'évacuation des déchets alimentaires (substances n'ayant pu être digérées). Ils permettent la progression et l'évacuation du bol alimentaire grâce à la présence de cellules caliciformes qui produisent une quantité importante de mucus. Par ailleurs, sous l'effet de l'aldostérone, le côlon absorbe activement le sodium du bol alimentaire entraînant sa déshydratation.

D'après un rapport publié en 2009 par la ligue nationale contre le cancer<sup>3</sup>, 73% des cancers concernent le côlon et la région recto-sigmoïdienne, et 27% touchent le rectum.

Qu'en est-il de la constitution du côlon et du rectum au niveau cellulaire ?

### 1.1.2.2 Histologie

La paroi intestinale est constituée de plusieurs « couches » histologiques dont la composition ne varie que très peu. En effet, comme dans le reste du tube digestif, le

---

3. Le rapport est trouvable à l'adresse suivante : <<http://www.ligue-cancer.net/shared/brochures/cancers-colon-rectum.pdf>> visitée le 4 novembre 2012.

côlon présente une muqueuse, une sous-muqueuse, une musculuse et une séreuse. Cette paroi se distingue cependant de celle de l'intestin grêle par l'absence de certains dispositifs d'amplification de surface, notamment les valvules conniventes ou les villosités. La structure de la muqueuse colique est présentée en figure 1.1 p.6.

**La muqueuse** La muqueuse est constituée de deux couches distinctes. Un épithélium unistratifié, « posé » sur une lame basale, et constitué de nombreuses cellules caliciformes qui sont d'autant plus nombreuses que l'on se rapproche du rectum. Cet épithélium est porté par un chorion – tissu conjonctif de soutien – riche en tissu lymphoïde diffus pouvant former, à certains endroits des follicules lymphoïdes<sup>4</sup>. L'épithélium présente des invaginations en forme de doigts constituant les glandes de Lieberkühn.

**La sous-muqueuse** La muqueuse repose sur une sous-muqueuse constituée de tissu conjonctif et qui contient le plexus nerveux de Meissner, ainsi que des vaisseaux sanguins et lymphatiques. Ce plexus nerveux a pour fonction de contrôler les sécrétions gastro-intestinales et le débit sanguin local.

**La musculuse** La sous-muqueuse repose sur la musculuse constituée de deux couches de cellules musculaires lisses, l'une externe dont les fibres sont disposées longitudinalement par rapport au tube digestif et dont la contraction raccourcit le tube. L'autre, est dite circulaire interne et est beaucoup plus épaisse. Sa contraction diminue le diamètre du tube digestif. Entre ces deux couches de cellules musculaires se trouvent les éléments histologiques formant le plexus myentérique d'Auerbach, principalement responsable du contrôle moteur de l'intestin.

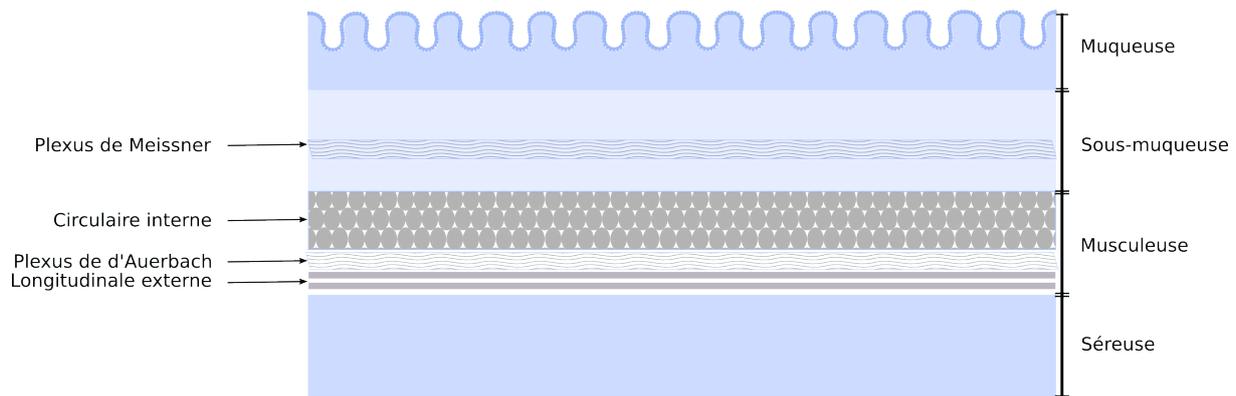
**La séreuse** La dernière couche est un tissu conjonctif tapissé par un mésothélium sur son versant externe et constituant le feuillet viscéral de la séreuse péritonéale.

### 1.1.2.3 L'unité fonctionnelle : les glandes de Lieberkühn

Les glandes de Lieberkühn, que nous avons présentées dans le paragraphe précédent (Cf. paragraphe 1.1.2.2 p. 5), constituent l'unité fonctionnelle du côlon. Elles sont organisées de telle sorte que des cellules souches, capables d'autorenouvellement, soient situées dans le fond des cryptes. En se divisant de façon asymétrique, celles-ci vont permettre d'en

---

4. Il s'agit de zones dans lesquelles les cellules réticulaires du système lymphoïde sont regroupés en amas.



**FIGURE 1.1 – Organisation histologique de la muqueuse colique** – La muqueuse colique est constituée d’une muqueuse, d’une sous-muqueuse, d’une musculeuse et d’une séreuse. La sous-muqueuse est traversée par le plexus de Meissner, et les deux couches de fibres musculaires sont séparées par le plexus d’Auerbach.

produire de nouvelles qui vont remonter le long de la crypte en se différenciant jusqu’à donner des cellules totalement différenciées de type épithéliale au niveau de la lumière intestinale [Gregorieff and Clevers, 2005]. Ces glandes présentent principalement trois types de cellules différenciées. Les plus abondantes dans cette partie de l’épithélium intestinal sont celles dites caliciformes qui ont pour fonction de sécréter du mucus permettant de faciliter le transit du bol alimentaire. Il y a également des entérocytes dont le rôle est d’absorber les minéraux et l’eau contenus dans les fécès, et des cellules entéroendocrines. La figure 1.2 p. 7 adaptée de [Reya and Clevers, 2005] donne une représentation schématique de cette organisation.

Au niveau moléculaire, le phénotype « souche » des cellules du fond des cryptes est permis grâce à l’activation de la voie Wnt. Cette protéine est produite par les myofibroblastes qui se situent sous la membrane basale, et elle va se fixer sur son récepteur Frizzled associé à la protéine LRP (cf figure 1.3 p.8). L’activation de ce récepteur induit la stabilisation de la  $\beta$ -caténine par l’intermédiaire de l’axine et à sa translocation nucléaire qui va pouvoir réguler de nombreux gènes cibles en se fixant aux facteurs de transcription Tcf/Lef [Gregorieff and Clevers, 2005]. Parmi ces gènes notons la régulation de c-Myc, la survivine, ou encore la cycline D [He et al., 1998] [Zhang et al., 2001] [Tetsu and McCormick, 1999]. Des gènes impliqués dans la migration ou l’adhésion cellulaire sont également activés, comme Oct4, Snail, la fibronectine ou encore l’E-cadhérine [Cole et al., 2008] [ten Berge et al., 2008]. Il est également montré que la protéine *Leucin-rich repeat-containing G-protein coupled receptor 5* (Lgr5) est une cible de la voie Wnt [Barker and Clevers, 2010]. Cette dernière est donc un marqueur de cellules souches intestinales, potentiellement régulée par

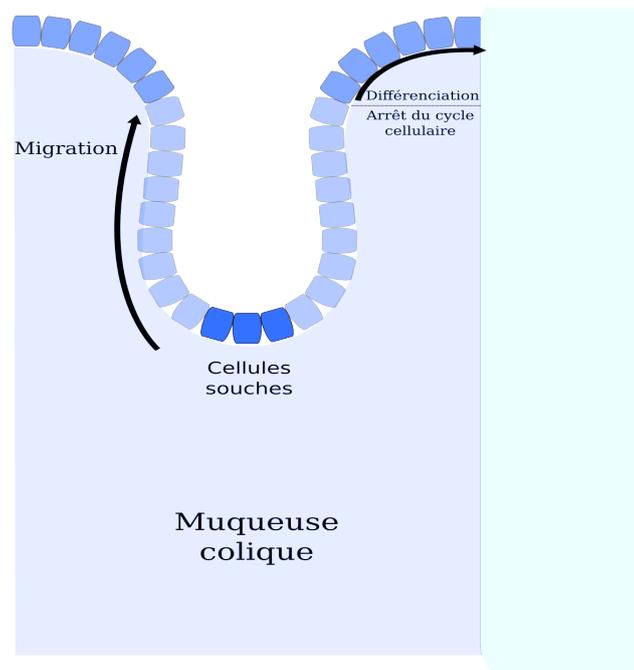


FIGURE 1.2 – **Organisation cellulaire des glandes de Lieberkühn** – Les cellules souches de l'épithélium intestinal colique résident au fond des cryptes (bleu foncé). Les cellules proliférantes occupent environ les deux tiers de la hauteur de la crypte (bleu clair). Les cellules différenciées occupent l'ouverture de la crypte et la surface de la paroi colique (bleu).

la voie Wnt. De plus, il a été montré que les cellules souches étaient très probablement à l'origine du développement des cancers intestinaux chez la souris [Barker et al., 2009].

L'équipe de Clevers a également mis en évidence en 2009 que l'Olfactomédine 4 (OLFM4) était un marqueur robuste des cellules souches intestinales, et a confirmé ces observations en 2010 en démontrant un lien étroit entre cette protéine et la protéine Lgr5 [van der Flier et al., 2009] [Barker and Clevers, 2010]. Il s'agit d'une protéine dont le niveau d'expression est important au niveau de la prostate, de l'intestin grêle et du colon, et dans une moindre mesure au niveau de l'œsophage, de la moelle osseuse et de l'estomac [Zhang et al., 2002a] [Zhang et al., 2004] [Liu et al., 2007].

### 1.1.3 Histoire naturelle du cancer colorectal

#### 1.1.3.1 Le modèle de développement séquentiel

D'un point de vue moléculaire, le modèle décrit par Vogelstein et Fearon en 1990 [Vogelstein et al., 1988] propose une évolution du cancer du côlon selon une succession de mutations. Ils proposent que le phénotype cancéreux apparaît suite à une dérive génétique de cellules dysplasiques, aboutissant à la mutation du gène *Adenomatous Polyposis Coli*

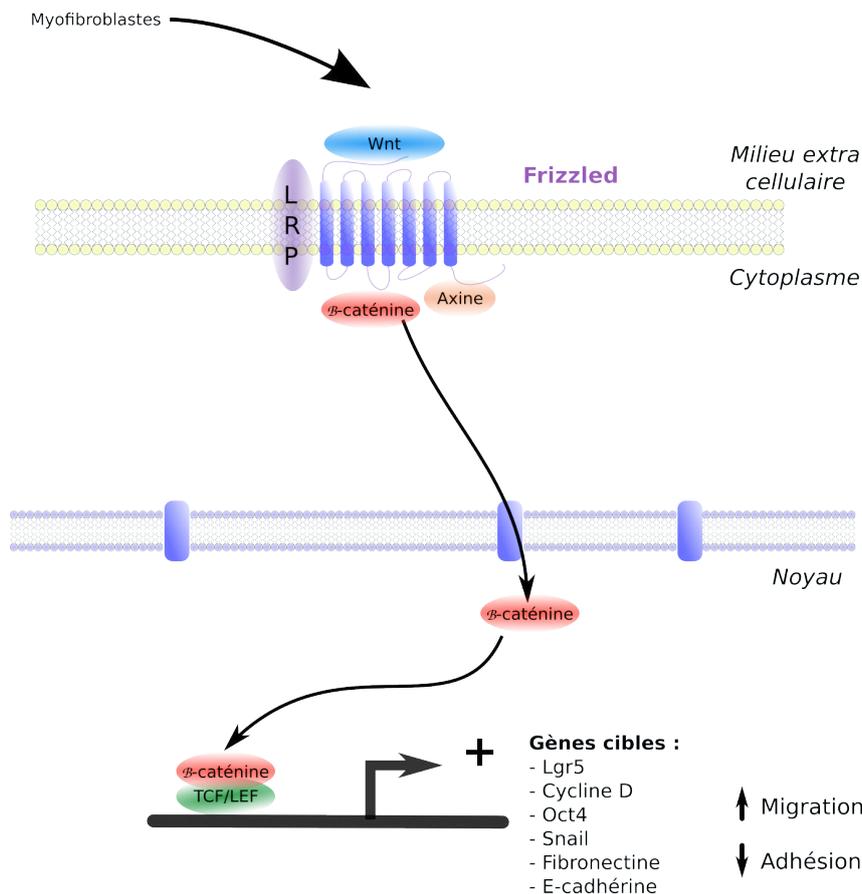


FIGURE 1.3 – **Schéma simplifié de la voie Wnt** – Les myofibroblastes sont responsables de la production du facteur Wnt qui se fixe sur son récepteur Frizzled localisé à la membrane des cellules épithéliales, et qui à son tour permet la translocation de la  $\beta$ -caténine dans le noyau. Elle peut alors activer ses gènes cibles en favorisant notamment la migration et en diminuant l'adhésion cellulaire.

(APC). Les cellules mutées sur ce gène présentent une instabilité génétique qui les conduit à accumuler des mutations aléatoires au sein de leur ADN et notamment sur les gènes K-Ras et p53. Depuis, de nombreux travaux ont montré que ce modèle était imparfait, notamment car ces gènes ne sont pas mutés dans tous les cancers. Pourtant, la cellule tumorale se caractérise par certaines propriétés : une prolifération incontrôlée, la capacité d'envahir les tissus adjacents, et de coloniser des organes, en passant par la circulation. Selon Hanahan et Weinberg [Hanahan and Weinberg, 2011], pour que la cellule évolue vers un phénotype tumoral, elle doit présenter les caractéristiques suivantes :

- une dérégulation de la stabilité du génome ;
- un échappement à la mort cellulaire programmée : l'apoptose ;
- une autosuffisance vis-à-vis des signaux de prolifération conduisant à une croissance indépendante des cellules voisines ;
- une insensibilité aux signaux d'arrêt de croissance ;

- une capacité de division illimitée ;
- une capacité d’induire la production de néovaisseaux ;
- une capacité d’échapper au système immunitaire ;
- une capacité de promouvoir l’inflammation tissulaire ;
- une capacité de produire des métastases, d’envahir les tissus et elle présente des dérèglements d’ordre métabolique.

Ce sont les altérations génétiques successives qui aboutissent à la tumorigenèse et au développement de cancers.

### 1.1.3.2 STAT3 et NF $\kappa$ B dans le cancer colorectal

La plupart des oncogènes sont activés par mutation dans les cancers à cause de l’instabilité génomique décrite précédemment. d’autres facteurs de transcription tels que STAT3 et NF $\kappa$ B ne sont généralement pas activés de cette façon. Pourtant, ils sont impliqués dans un très grand nombre de pathologies, et notamment dans les processus de cancérisation [Bromberg et al., 1999] [Karin, 2006b].

En effet, ces deux facteurs de transcription sont impliqués dans les étapes initiales du développement tumoral du cancer colorectal. Ils sont généralement constitutivement activés dès les premières étapes de la tumorigenèse. Cette activation est généralement due à la stimulation des cellules par les facteurs autocrines et paracrines libérés dans le microenvironnement tumoral [Karin, 2006a]. Ce sont également de puissants activateurs de la prolifération cellulaire, de la survie, de l’angiogenèse et de la réparation tissulaire [Yu et al., 2007] [Karin et al., 2002] [Haura et al., 2005].

Les protéines STATs (*Signal Transducer and Activator of Transcription*) appartiennent à une famille de facteurs de transcription composée de sept membres pour le moment : STAT1, 2, 3, 4, 5A, 5B et 6. En se fixant sur le promoteur de ses gènes cibles, STAT3 régule la prolifération cellulaire et l’entrée en phase S du cycle en activant la transcription de la cycline D1, c-myc et de c-fos [Bromberg, 2002]. Il régule également la survie cellulaire en permettant la transcription de Bcl-x1 et Bcl2 ; et il intervient dans l’angiogénèse, la mobilité cellulaire et l’embryogenèse. STAT3 est aussi impliqué dans le développement tumoral. Il est activé de manière constitutive dans de nombreuses tumeurs et constitue une cible privilégiée d’oncogènes comme Src [Bromberg et al., 1998], un oncogène également fortement activé dans le cancer colorectal [Yeatman, 2004]. Le facteur de transcription STAT3 est présent dans le cytoplasme des cellules dans lequel il est inactif. Son activation

est initiée par la liaison de facteurs de croissance sur leurs récepteurs membranaires comme l'EGF, le PDGF ou encore l'HGF [Boccaccio et al., 1998]. Il peut également être activé par la fixation de cytokines comme l'IL-6, premier activateur de STAT3 décrit dans la littérature [Wegenka et al., 1993]. Ces récepteurs le phosphorylent directement grâce à leur propre activité tyrosine kinase ou de manière indirecte par l'intermédiaire de kinases cytoplasmiques comme Src ou JAK. En effet, la liaison de l'IL-6 sur son récepteur membranaire gp130 entraîne la dimérisation du récepteur, l'activation des kinases JAK et la phosphorylation de STAT3 sur le résidu tyrosine 705 [Yu et al., 2009]. En 1999, l'équipe de Bromberg le décrit comme étant un oncogène [Bromberg et al., 1999]. Ils utilisent une forme dimérique de STAT3, qui est alors capable de se fixer à l'ADN et d'activer des gènes cibles. En transfectant des cellules avec cette forme de STAT3, ils observent que les cellules sont capables de pousser en agar, et de donner des tumeurs lorsqu'elles sont implantées chez des souris « nude »<sup>5</sup>. D'autre part, des études ont montré que la phosphorylation et la translocation du facteur de transcription STAT3 sont corrélées à l'accumulation nucléaire de la  $\beta$ -caténine, un oncogène majeur dans les cancers colorectaux [Kawada et al., 2006]. D'autres publications suggèrent que l'activation de la  $\beta$ -caténine entraîne une augmentation de l'expression de STAT3 dans des cellules souches embryonnaires murines. Cette induction de STAT3 est nécessaire pour le renouvellement des cellules souches et donc pour la prolifération. Dans ces cellules, bien que la voie de wnt ait été activée auparavant, l'absence de STAT3 ou un très faible taux de la protéine, entraîne leur différenciation et elles perdent leur potentiel de cellule souche [Hao et al., 2006]. De plus, des sites consensus TBE sont présents sur le promoteur du gène codant pour STAT3. Ces sites sont les lieux de fixation privilégiés du complexe  $\beta$ -caténine/TCF.

La famille des facteurs de transcription  $\text{NF}\kappa\text{B}$  est constituée de 5 membres : p65 (RelA), RelB, c-Rel, p50/p105 ( $\text{NF}\kappa\text{B1}$ ) et p52/p100 ( $\text{NF}\kappa\text{B2}$ ). Tous les membres de cette famille sont caractérisés par la présence d'un domaine *Rel homology domain* (RHD), de 300 résidus d'acides aminés, localisé en N-terminal [Hayden and Ghosh, 2004]. Ce domaine est responsable de la dimérisation, de la fixation à l'ADN, et du lien avec les régulateurs  $\text{I}\kappa\text{Bs}$ .  $\text{NF}\kappa\text{B}$  est activé de façon constitutive dans beaucoup de tumeurs. Par exemple, la production de cytokines proinflammatoires par le microenvironnement tumoral peut stimuler l'activité d'IKK, conduisant à l'activation constitutive de  $\text{NF}\kappa\text{B}$ . L'expression constitutive est quant à elle à mettre en lien avec une prolifération tumorale excessive,

---

5. Les souris dites « nude » sont des souris présentant une délétion du gène FOXP1 conduisant à une absence de thymus chez ces animaux (ou à un thymus non fonctionnel dans certains cas) conduisant à l'absence de production de lymphocytes T matures. Ces souris présentent donc une immunité très faible.

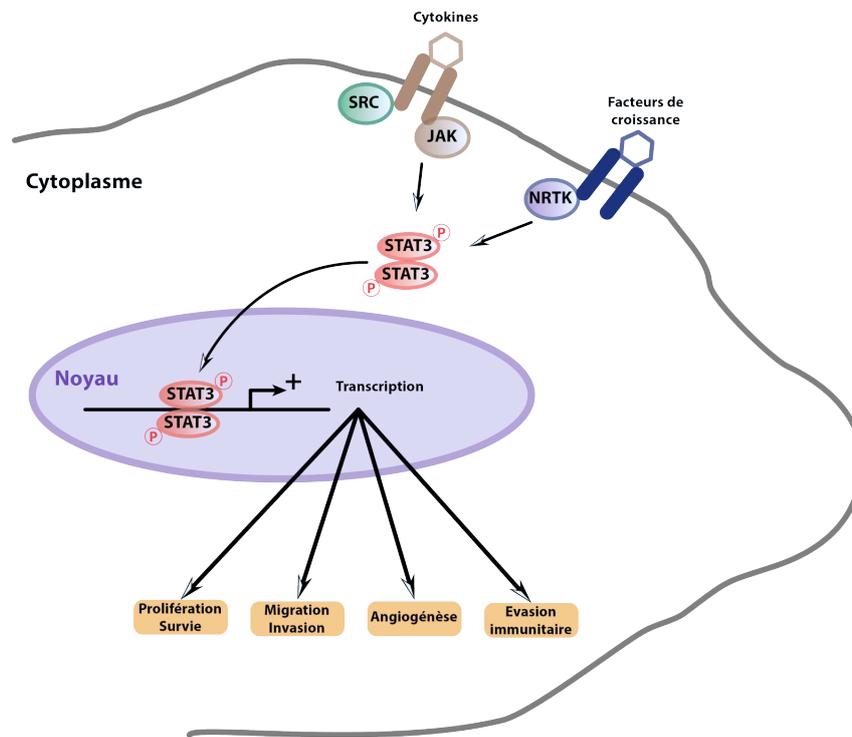


FIGURE 1.4 – Schéma simplifié de la voie de signalisation médiée par STAT3 – La fixation de cytokines ou de facteurs de croissance sur certains récepteurs à activité tyrosine kinase entraîne l’activation des JAK et de SRC qui vont alors pouvoir phosphoryler STAT3. Cette phosphorylation permet la dimérisation du facteur de transcription qui peut alors activer ses gènes cibles.

une inhibition de l’apoptose, et une augmentation de l’angiogénèse ainsi que du potentiel métastatique des tumeurs. Enfin, la plupart des cancers présentant une augmentation de l’activité de  $\text{NF}\kappa\text{B}$  sont des cancers également capable d’une résistance importante aux chimiothérapies et à la radiothérapie [Lee et al., 2007].

Enfin, des travaux récents suggèrent qu’il existe un lien entre les deux facteurs de transcription STAT3 et  $\text{NF}\kappa\text{B}$ . Par exemple, l’activation de STAT3 par le complexe CBP/p300 peut induire le clivage de p100 en p52 pour le rendre actif, en activant  $\text{IKK}\alpha$ . Cette activation conduit à une augmentation de la survie cellulaire et participe à la résistance aux traitements [Nadiminty et al., 2006]. D’autres travaux, menés par l’équipe de K. Struhl montrent que lors de l’activation de Src, il y a une activation de la réponse inflammatoire médiée par  $\text{NF}\kappa\text{B}$  aboutissant à la transcription de Lin28 et à une rapide diminution du micro-ARN Let-7. Ce dernier étant un inhibiteur de l’expression d’Il-6, sa diminution entraîne une augmentation du niveau d’Il-6, qui peut alors activer STAT3 [Iliopoulos et al., 2009].

## 1.2 Les classifications du cancer colorectal

### 1.2.1 Classification histologique

Le cancer colorectal est issu, dans environ 80% des cas, de la transformation d'un polype adénomateux en tumeur. Les 20% restants correspondent à des formes de cancer du colon dites *de novo* et qui correspondent à des cancers qui ne sont pas issus de polypes mais qui se développent à partir de la muqueuse intestinale saine. À l'heure actuelle, on estime qu'environ 5% de l'ensemble des adénomes évoluent en adénocarcinomes [Shinya and Wolff, 1979]. La transformation maligne serait favorisée par certains caractères notamment la taille ( $> 1$  cm), le degré de dysplasie et l'importance de la composante vilieuse.

Parmi les lésions précancéreuses, souvent appelées polypes, il est possible de distinguer les polypes sessiles ou pédiculés selon leur aspect morphologique. Les polypes peuvent être de nature bénigne ou maligne et lorsque leur nombre excède 10, on parle de polypose. Ces polypes colorectaux sont classés en quatre sous-catégories.

Il peut s'agir de polypes adénomateux (autrement appelés adénomes) qui résultent de la prolifération de cellules des glandes de Lieberkhün. Il en existe également trois sous-catégorie selon la classification de l'organisation mondiale de la santé (OMS) :

- les adénomes tubuleux, 75% des cas ;
- les adénomes tubulo-villeux, 20% des cas ;
- les adénomes vilieux, 5% des cas.

Seuls ces polypes adénomateux sont susceptibles de donner des cancers et ils sont caractérisés par leur degré de dysplasie. Ainsi, tout adénome bénin est par définition en dysplasie de bas grade, alors que dysplasie de haut grade correspond au premier stade du cancer, correspondant au cancer intra-épithélial, autrement appelé cancer in-situ.

Il est également possible de distinguer les polypes hyperplasiques, juvéniles et pseudo-inflammatoires, mais nous ne rentrerons pas dans le détail descriptif de ces formes bénignes qui n'évoluent pas en cancer.

Par ailleurs, les polypes adénomateux vont évoluer en carcinomes colorectaux pour lesquels plusieurs types histologiques peuvent être également décrits. L'immense majorité d'entre eux sont des adénocarcinomes dit lieberkühniens (environ 80%). Ces adénocarcinomes se caractérisent par leurs degrés de différenciation. Ainsi, le grade 1 est attribué à des carcinomes bien différenciés, constitués à 95% de glandes coliques. Le grade 2 est attribué à des carcinomes modérément différenciés constitués de 50 à 95% de glandes

coliques. Le grade 3 concerne les carcinomes peu différenciés présentant de 5 à 50% de glandes coliques, et le grade 4 correspond aux carcinomes en ayant moins de 5%. Ces grades constituent des indicateurs pronostiques. Selon une étude publiée en 2004 par Ueno et ses collaborateurs, il est montré que les lésions peu différenciées sont beaucoup plus agressives, et malignes [Ueno et al., 2004]. En pratique courante, la distinction de ces grades est relativement subjective, de sorte qu'une classification en deux grades a été proposée. Ainsi, les adénocarcinomes ayant plus de 50% de glandes sont dits différenciés et ceux en ayant moins de 50% sont dits peu différenciés.

Un autre type d'adénocarcinomes dit mucineux, ou colloïde muqueux représente entre 10% et 20% des tumeurs malignes [Green et al., 1993] qui se développent au niveau colorectal. Depuis leur description en 1923 par Parham [Parham, 1923], quelques éléments clés sont importants dans la prise en charge de ces patients. Ces tumeurs semblent présenter un pronostic moins favorable que les tumeurs non mucineuses [Hussain et al., 1999].

Il existe des formes encore plus rares, représentant autour de 5% des cancers colorectaux. Notamment des cancers issus des cellules nerveuses colorectales constituant des tumeurs carcinoïdes. Mais également des tumeurs touchant les ganglions lymphatiques drainant le côlon et le rectum. Des tumeurs de types sarcomes, touchant les tissus de soutien environnant peuvent également se développer dans ces zones, ou encore des mélanomes issus des mélanocytes présents dans la muqueuse du canal anal notamment.

Enfin, le côlon et le rectum sont des cibles privilégiées pour les métastases issues de cancers de l'ovaire, de la prostate, de l'estomac ou encore du sein.

### 1.2.2 Classification clinique

Les cancers colorectaux répondent à des classifications cliniques prenant en compte les observations anatomopathologiques. Ils vont dans un premier temps ne toucher que l'épithélium intestinal et rester *in situ*, contenus par la membrane basale. Ils peuvent ensuite acquérir la capacité de franchir la membrane basale pour aller envahir les autres couches cellulaires constituant la paroi de l'intestin (sous-muqueuse, musculuse, séreuse). Puis, lorsque toutes les couches cellulaires de la paroi sont franchies, les cellules tumorales peuvent rejoindre la circulation sanguine et donner lieu à des métastases, d'abord ganglionnaires, puis à distance, touchant d'autres organes. Cette évolution histologique du cancer, au sein de la paroi intestinale, est à l'origine de diverses classifications utilisées par les oncologues pour définir la stratégie thérapeutique la plus efficace et la moins invasive

pour le patient.

Parmi celles-ci, la plus utilisée par les médecins est la classification TNM « Tumor, Nodes, Metastasis » [Li and Lai, 2009]. Ainsi, pour chacun de ces critères (tumeur, ganglions, métastases), un chiffre est attribué, reflétant l'importance de l'invasion tumorale dans la paroi intestinale. Pour le critère *Tumor* les valeurs vont de 1 à 4, pour le critère *Nodes* ces valeurs sont comprises entre 0 et 2 et pour le critère *Metastasis* elles sont comprises entre 0 et 1 en fonction de l'importance de l'envahissement, du nombre de ganglions touchés et de la présence ou non de métastases à distance.

Le tableau 1.1 p. 14 présente les critères de cette classification.

Tableau 1.1 – La classification TNM

<b>La tumeur primitive (T)</b>	
Tis	Seule la première couche de la paroi interne du côlon et du rectum est touchée.
T1	La tumeur envahit la sous-muqueuse.
T2	La tumeur envahit la musculuse.
T3	La tumeur a franchi la muqueuse et atteint le tissu péricolique ou périrectal.
T4	La tumeur perforé le péritoine et envahit les organes du voisinage.
<b>Les ganglions régionaux (N)</b>	
N0	Pas d'atteinte glandulaire régionale métastatique.
N1	1 à 3 adénopathies régionales métastatiques.
N2	4 ganglions ou plus sont touchés.
Nx	Les adénopathies régionales ne peuvent pas être évaluées.
<b>Les métastases (M)</b>	
M0	Absence de métastase.
M1	Métastases à distance.
Mx	Le statut des métastases ne peut pas être évalué.

Une classification plus simple a été mise en place par l'*American Joint Committee for Cancer Classification*. Elle propose des grades cliniques permettant d'adapter plus facilement la prise en charge des patients.

Ainsi, les stades 1 et 2 d'une manière générale correspondent à des tumeurs qui ne présentent pas d'invasion des ganglions périphériques. La prise en charge des patients sera donc similaire. Les tumeurs de stade 3 présentent une invasion des ganglions périphériques et nécessitent l'utilisation d'une chimiothérapie adjuvante pour augmenter les chances de réussite du traitement. Enfin, les tumeurs de stade 4 présentent des métastases à distance avec une prise en charge cumulant généralement chirurgie, chimiothérapie et thérapie ciblée.

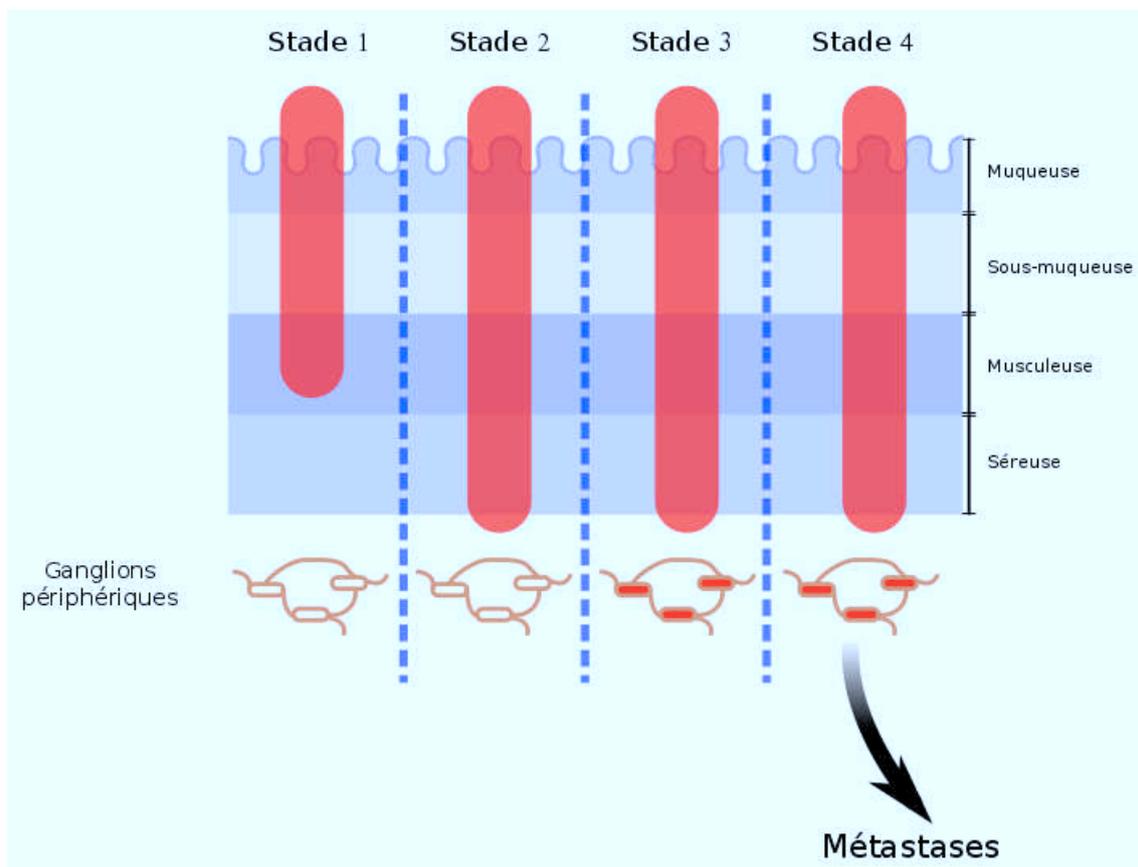


FIGURE 1.5 – Schéma illustrant la classification par stades en fonction de l'invasion de la muqueuse intestinale par la tumeur – Les tumeurs de stade 1 et 2 sont localisées dans la paroi intestinale. Les tumeurs de stade 3 présentent un envahissement des ganglions périphériques. Les tumeurs de stade 4 présentent à la fois un envahissement ganglionnaire, et une invasion d'organes à distance par la tumeur primaire (métastases).

Tableau 1.2 – Correspondance entre la classification TNM, les stades cliniques et survie à 5 ans

Stade	Classification TNM	Survie à 5 ans [O'Connell et al., 2004]
Stade 1A	T1 N0 M0	93%
Stade 1B	T2 N0 M0	
Stade 2A	T3 N0 M0	85%
Stade 2B	T4 N0 M0	72%
Stade 3A	T1-T2 N1 M0	83%
Stade 3B	T3-T4 N1 M0	64%
Stade 3C	T1-T4 N2 M0	44%
Stade 4	T1-T4 N1-N3 M1	8%

### 1.2.3 Classification moléculaire

#### 1.2.3.1 L'instabilité génomique dans le cancer colorectal

**L'instabilité de type CIN** Les tumeurs colorectales présentent deux types d'instabilité génomique ; d'une part celle touchant au caryotype des cellules tumorales, appelée CIN, pour *Chromosomal instability* et concernant 85% des tumeurs sporadiques. Ces dernières présentent généralement des aberrations chromosomiques importantes telles que l'aneuploïdie ou la perte d'hétérozygotie de certaines parties de leur patrimoine génétique. Citons comme exemple la délétion 1p ou 8q, ou encore la perte d'hétérozygotie 5q, 17p et 18q [Soreide et al., 2006]. Ces aberrations apparaissent à la suite d'une dérégulation de l'expression ou d'une mutation d'un gène du point de contrôle du fuseau mitotique, ou de défauts se situant au niveau des microtubules ou des télomères [Grady and Markowitz, 2000]. L'instabilité chromosomique est considérée comme un facteur aggravant la progression tumorale, car la plupart du temps, elle conduit à une perte de gènes suppresseurs de tumeurs ou à un gain du nombre de copies d'oncogènes [Ogino and Goel, 2008]. Le plus souvent, c'est la zone 18q qui est perdue et cela s'accompagne d'un pronostic vital défavorable pour le patient [Weisenberger et al., 2006]. Le gène codant pour la protéine SMAD4 est localisé sur cette portion de chromosome, et la diminution de son taux dans la tumeur semble associée à un mauvais pronostic pour le patient et à une diminution de la réponse au 5-fluorouracile [Alhopuro et al., 2005]. Pour autant, la perte d'hétérozygotie 18q n'est pas systématiquement associée à une diminution du taux de SMAD4 [Alazzouzi et al., 2005].

**L'instabilité de type MSI** Le deuxième type d'instabilité conduit à des mutations sur des régions microsatellitaires, appelée MSI pour *Microsatellite instability*, et concerne 15% des tumeurs sporadiques. Ces dernières présentent des mutations de type insertion ou délétion dans les régions microsatellitaires constituées de séquences de 2 à 10 nucléotides répétées en tandem [Ionov et al., 1993]. Cela s'explique par des défauts importants au sein de leur système de réparation de l'ADN appelé *MisMatch Repair* (MMR). Ce système a pour fonction de repérer les mésappariements au cours de la réplication. Ainsi, lors de la perte de ce système, le taux de mutations ponctuelles augmente considérablement. Il est montré que dans 15% des cancers colorectaux, il est possible de détecter des mutations sur les gènes MLH1, MSH2, MSH6 ou PMS2, et les quatre protéines codées par ces gènes sont impliquées dans le système MMR [Vilar and Gruber, 2010]. De fait, les patients

présentant une instabilité microsatellite, présentent également un défaut du système MMR. Cette instabilité génétique est détectable chez les patients en amplifiant certaines séquences microsatellitaires et en comparant leur taille dans la tumeur et dans le tissu sain périphérique. Des zones de référence sont utilisées pour permettre d'établir le profil du patient, et ces zones constituent 5 à 10 *loci* utilisés pour le diagnostic [Umar et al., 2004]. Pour ces régions, on peut définir trois types de cancers :

- les cancers MSI-H, pour lesquels l'instabilité est supérieure à 30% de la zone de référence ;
- les cancers MSI-L, pour lesquels l'instabilité est comprise entre 10 et 30 % de la zone de référence ;
- les cancers MSS, pour lesquels l'instabilité est stable par rapport à la zone référence.

Bien que ces patients semblent avoir un caryotype diploïde [Lothe et al., 1993], l'absence de mécanisme de réparation entraîne des mutations fréquentes au sein de l'ADN de la tumeur, pouvant toucher des gènes suppresseurs de tumeurs ou des oncogènes [Miquel et al., 2007]. Dans le cancer colorectal, les principales cibles sont  $TGF\beta RII$ ,  $IGF2R$  ou encore  $BAX$  [Jass et al., 1998]. Les tumeurs de phénotype MSI-H sont en revanche rarement mutées sur  $APC$ ,  $KRAS$  ou  $TP53$  et présentent des niveaux standards de  $\beta$ -caténine. À l'inverse, les tumeurs ayant un phénotype plutôt MSS ou MSI-L présentent des fréquences de mutations de ces gènes plus élevées, associées à des pertes d'hétérozygotie concernant les zones 5q, 17q et 18q [Jass et al., 1999]. Pour autant, il existe tout de même des différences entre les MSS et les MSI-L. Il est plus fréquent de détecter des mutations d' $APC$  ou de  $KRAS$  chez les MSI-L que chez les MSS qui présentent, quant à elles des pertes d'hétérozygotie sur le chromosome 5 (5q) plus fréquentes [Konishi et al., 1996].

Dans les études cliniques, les patients ayant un phénotype MSI-H ont un taux de survie supérieur à ceux ayant un phénotype MSS [Popat et al., 2005]. Cependant, le statut de ce marqueur comme indication dans l'utilisation de la chimiothérapie adjuvante pour le traitement des patients est moins clair. Une étude récente suggère que les tumeurs MSI-H répondent bien à une chimiothérapie adjuvante utilisant le 5-Fluorouracile (5-FU) [Elsaleh and Iacopetta, 2001]. Cependant, d'autres études semblent montrer qu'il n'y a pas de différence de sensibilité selon que le patient soit MSI-H ou MSS [Jover et al., 2009].

### 1.2.3.2 L'instabilité épigénétique du cancer colorectal

Les cancers colorectaux présentent des aberrations au niveau de la régulation épigénétique des gènes. C'est ainsi qu'il est possible de définir un *CpG island methylator phenotype* (CIMP), autrement dit, un phénotype de méthylation des îlots CpG. Ce phénotype est défini à partir d'au moins trois *loci* marqueurs [Weisenberger et al., 2006]. Les patients ayant un phénotype MSI-H et CIMP<sup>+</sup>, et présentant la mutation V600E sur le gène BRAF sont dits CIMP-High (CIMP-H) et sont des patients pour lesquels l'issue est plutôt favorable [Ogino et al., 2009]. Il existe un autre groupe de patients CIMP<sup>+</sup> associé à une mutation de KRAS et appelés CIMP-Low (CIMP-L) [Ogino et al., 2006]. Une seule étude semble montrer un lien entre le phénotype MSI et le statut CIMP concernant l'évolution de la pathologie chez les patients [Ogino et al., 2009].

### 1.2.3.3 Les mutations fréquentes

Dès les étapes initiales du développement tumoral, les cellules présentent une forte instabilité génomique. Cette instabilité conduit à l'accumulation de mutations ponctuelles au sein de l'ADN des cellules. Ces mutations, lorsqu'elles touchent des gènes codant pour des protéines de régulation de la prolifération cellulaire, ou des protéines de réparation de l'ADN, peuvent conduire à une augmentation de la multiplication cellulaire, et au phénotype cancéreux.

Il existe également une classification reposant sur l'observation du statut génétique des tumeurs pour certains gènes déterminants dans la prise en charge des patients.

**Les mutations sur KRAS** La détection des mutations de l'oncogène KRAS dans les tumeurs colorectales est dorénavant systématiquement réalisée en routine. Cela permet notamment de savoir s'il sera possible d'utiliser les traitements de chimiothérapie ciblée tels que le cétuximab ou le panitumumab. En effet, ces anticorps bloquent le récepteur à l'EGF, or, dans 35 à 50% des cancers colorectaux il existe une mutation de l'oncogène KRAS qui rend la protéine constitutivement active [Li et al., 2006], [Nandan and Yang, 2011]. Ras est une protéine G de 21kDa associée à la face interne de la membrane plasmique. En temps normal, lorsque l'EGF se fixe sur son récepteur, celui-ci s'autophosphoryle et recrute les protéines *guanine nucleotide exchange factors* GEFs (cf. Fig 1.6 p. 20). Ces protéines ont pour fonction de catalyser l'échange du guanosine diphosphate (GDP) porté par la protéine Ras, par du guanosine triphosphate (GTP). La protéine Ras est alors activée grâce au

chargement de GTP par Grb2/SOS [Cullen and Lockyer, 2002] – protéine appartenant à la famille des GEFs –, et elle va à son tour activer différents facteurs de prolifération tels que Raf, ou la PI3K (PhosphoInositol-3-kinase) [Herbst, 2004]. Lorsque l'oncogène KRAS est muté sur son codon 12, 13 ou 61, la protéine correspondante est constitutivement active de sorte qu'inhiber le récepteur à l'EGF aura peu d'effet sur les tumeurs des patients [Lievre et al., 2008]. En effet, des mutations touchant ces codons vont modifier la poche dans laquelle se loge le guanosine triphosphate (GTP) de sorte qu'il ne peut plus y avoir échange de ce dernier par du GDP. La protéine Ras sous sa forme Ras-GTP va alors s'accumuler dans la cellule, et va pouvoir activer ses gènes cibles [Downward, 2003].

Il a également été proposé récemment que la mutation de l'oncogène KRAS pouvait être responsable de l'accumulation de  $\beta$ -caténine dans le noyau, impliquant l'activation des gènes cibles de cette dernière, notamment c-myc, et la cycline D [Phelps et al., 2009].

**La mutation de BRAF** De la même façon que Ras, la protéine Braf est mutée dans environ 8% de cancers colorectaux. La principale mutation décrite est la substitution d'une valine en position 600 par un acide glutamique. Cette mutation intervient dans 80% des cas [Davies et al., 2002]. Il s'agit, là aussi, d'un élément important dans la prise en charge des patients, car comme pour KRAS, la mutation du gène BRAF rend constitutivement active la protéine correspondante, qui ne répond alors plus aux signaux délivrés par le récepteur à l'EGF (cf. Fig 1.6 p. 20) et médiés par Ras. De la sorte, l'effet des anticorps monoclonaux dirigés contre l'EGF n'auront que peu d'effet [Di Nicolantonio et al., 2008]. Ces données portent donc à environ 60% le nombre de cancers colorectaux dont le traitement par thérapie ciblée sera peu ou pas efficace, car les mutations de KRAS et de BRAF sont exclusives. En effet, les deux mutations n'ont jamais été retrouvées chez le même patient [Li et al., 2006].

**Les mutations de TP53** Le gène TP53 code pour un facteur de transcription activé lors de divers stress, et conduisant à l'arrêt du cycle cellulaire, puis à la réparation de l'ADN, à l'apoptose ou à l'angiogénèse [Vogelstein et al., 2000]. Les mutations entraînant une perte de fonction de ce facteur de transcription constituent des éléments critiques de la progression tumorale. Il a été montré que des tumeurs colorectales de stade III mutées sur le gène suppresseur de tumeurs TP53, et présentant une instabilité microsatellitaire étaient beaucoup moins sensibles à l'action du 5-fluorouracile (5-FU), médicament régulièrement utilisé en thérapie adjuvante [Elsaleh et al., 2001]. La mutation de ce gène constitue donc

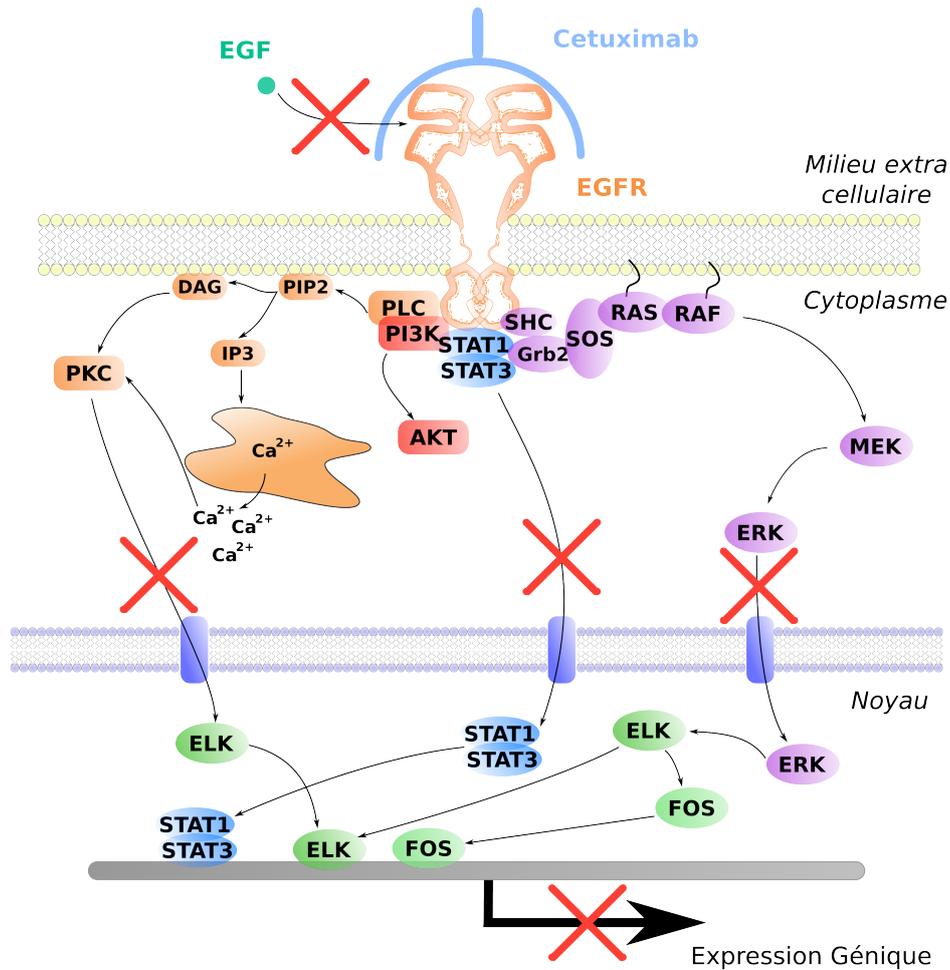


FIGURE 1.6 – Représentation schématique de la voie de signalisation médiée par le récepteur à l'EGF et de l'effet du cétuximab sur cette voie. – La fixation de l'anticorps monoclonal sur l'EGFR empêche sa dimérisation et son autophosphorylation de sorte que les voies de signalisation situées en aval se trouvent également bloquées. En cas de mutation de KRAS, la voie des MAP kinases reste active malgré l'utilisation du médicament.

un facteur de mauvais pronostic car elle conduit à une absence de réponse à un type de chimiothérapie.

**Le niveau d'expression de la thymidylate synthase (TYMS)** La thymidylate synthase (TYMS) est la cible principale du 5-FU, et est responsable de la production de thymidine nécessaire à la réplication de l'ADN. Sous l'effet du 5-FU cette enzyme est inhibée conduisant à une absence de production de thymidine, et la cellule est alors incapable de répliquer son ADN. Deux principaux déterminants génétiques de la régulation de l'expression de la TYMS ont été décrits. L'un correspond à un nombre variable de séquences répétées dans la région promotrice du gène codant pour la TYMS, l'autre correspond à un polymorphisme touchant six paires de bases situées en 3'UTR du gène [Chen et al., 2003].

Ces informations peuvent être utilisées pour avoir une idée du risque de récurrence chez des patients atteints de tumeurs de stade 2 ou 3 [Lurje et al., 2008].

## 1.3 Les traitements

### 1.3.1 *Un traitement principal : la chirurgie*

L'arme principale à disposition des cliniciens dans le traitement du cancer colorectal est la chirurgie. Elle a pour objectif la résection la plus complète possible de la tumeur de même que la résection du plus grand nombre possible de ganglions périphériques colonisés lorsqu'il y en a. Cependant, elle n'est que très rarement suffisante, et elle est souvent complétée par un traitement médicamenteux utilisable en pré-opératoire (néoadjuvant), ou en post-opératoire (adjuvant) [Van Cutsem and Oliveira, 2009]. D'autre part, dans le cas particulier du cancer rectal, une radiothérapie peut être utilisée [David Cunningham et al., 2012].

### 1.3.2 *Présentation des traitements médicamenteux*

Le traitement de référence du cancer du côlon est le 5-FU. Cette molécule a pour fonction d'inhiber la thymidylate synthase qui est une enzyme qui participe à la synthèse de la thymidine. Une fois intégrée dans la cellule, la molécule va donner des métabolites actifs comme la fluorodeoxyuridine monophosphate, la fluorodeoxyuridine triphosphate, et la fluorouridine triphosphate. Ces métabolites vont venir perturber la synthèse des ARN et l'action de la thymidylate synthase, conduisant à des dommages de l'ADN [Longley et al., 2003].

L'oxaliplatine est un sel platine de troisième génération qui est devenu l'un des traitements conventionnels du cancer colorectal. Comme les autres sels de platine, l'oxaliplatine exerce son action anticancéreuse de plusieurs façon, en induisant notamment des dommages de l'ADN, un arrêt de la synthèse d'ADN et/ou d'ARN, conduisant la cellule tumorale à activer les voies d'apoptose. Elle peut également conduire à la mise en place de mécanismes immunologiques conduisant à la destruction des cellules cancéreuses [Alcindor and Beauger, 2011]

L'irinotécan est une prodrogue utilisée dans le traitement du cancer colorectal et qui nécessite d'être métabolisée en SN-38 par la carboxyl-estérase hépatique pour être active [Pinedo and Peters, 1988]. Son action consiste à inhiber l'enzyme topoisomérase I

impliquée dans la gestion du surenroulement de la molécule d'ADN lors de la réplication et de la transcription en générant des cassures réversibles. Pour cela, il se lie au complexe ADN/topoisomérase I et empêche la réparation des cassures générées lors de ces mécanismes.

Le cétuximab et le panitumumab sont des anticorps monoclonaux dirigés contre le récepteur à l'EGF (Cf Figure 1.6 p. 20). Ces deux traitements nécessitent de définir le statut de l'oncogène KRAS avant d'être administrés (Cf paragraphe 1.2.3.3 p. 18) et ont pour fonction de bloquer la voie de prolifération cellulaire médiée par le récepteur à l'EGF.

Ces molécules chimiques utilisées dans le traitement du cancer colorectal ont plusieurs effets. Elles peuvent d'une part aboutir à ce que la cellule mette en place des mécanismes de mort cellulaire (apoptose), auquel cas, on assiste à une régression tumorale. Cependant, les cellules tumorales peuvent également répondre en arrêtant les processus de division cellulaire, et en restant dans un état appelé sénescence.

## 1.4 L'apoptose

Le terme d'apoptose vient d'une locution grecque évocant la « chute des feuilles » à l'automne. Ce terme était déjà utilisé par le célèbre Hippocrate de Kos (460-377 avant Jésus-Christ) qui parlait de ce qui pourrait être traduit par la « chute des os » en évoquant la décomposition des tissus après la mort. L'apoptose cellulaire est un phénomène physiologique par lequel les cellules surnuméraires ou non fonctionnelles de l'organisme sont éliminées. Elle a été mise en évidence en 1972 par Kerr et ses collaborateurs [Kerr et al., 1972] et intervient dans de nombreux processus biologiques et notamment, elle tient un rôle déterminant dans l'embryogenèse. En effet, c'est grâce à l'apoptose que peut avoir lieu la tunnelisation des organes creux (tube digestif), mais aussi la séparation des doigts et des orteils. Elle peut également intervenir dans d'autres processus tels que l'homéostasie cellulaire, ou encore dans la sélection des lymphocytes lors que l'apprentissage du soi.

Ce processus peut être déclenché par des signaux extracellulaires de nature diverse. En effet, la phase d'initiation de l'apoptose peut faire suite à des radiations ionisantes [Farrell et al., 1998], une carence en facteurs de croissance [Martinou et al., 1999], des drogues cytotoxiques [Bose et al., 1995], du glucose [Kaneto et al., 1996], des acides gras [Ishida, 1997], des céramides [Obeid et al., 1993] ou des espèces réactives de l'oxygène (*Reactive Oxygen Species* (ROS)) [Goldkorn et al., 1998]. Suite à ces signaux, la cellule intègre l'information qu'elle reçoit. En fonction de son état physiologique ou du contexte dans

lequel elle se trouve, la cellule va orienter sa réponse vers la mort ou la survie. Pour cela, de nombreux médiateurs entrent en jeu, anti ou proapoptotiques, comme les caspases, les protéines de la famille Bcl-2, la protéine Bax, la mitochondrie, et bien d'autres.

## 1.5 La sénescence

### 1.5.1 Définition

La sénescence cellulaire est un processus initialement décrit comme un arrêt permanent du cycle cellulaire causé par une sorte d'« épuisement réplicatif » chez des fibroblastes diploïdes humains mis en culture [Hayflick, 1965]. Il est à présent admis que la sénescence n'est pas uniquement réplicative, mais qu'elle peut intervenir suite à de nombreux types de stress (cf figure 1.7 p.24). Parmi ceux-ci, nous pouvons citer les dommages de l'ADN, le stress oxydatif lié à l'activité mitochondriale, l'activité de certains oncogènes (*Oncogene induced senescence*), comme KRAS par exemple [de Carne Trecesson et al., 2011], ou encore un défaut de fonctionnement des télomérases dans le cas de la sénescence réplicative [Ben-Porath and Weinberg, 2005]. Elle est à présent décrite comme un arrêt permanent du cycle cellulaire, conduisant la cellule à devenir insensible aux signaux de croissance présents dans son environnement [Stein et al., 1991]. Ce processus cellulaire est principalement médié par deux protéines : p53 et Rb. Les gènes codant pour ces protéines sont régulièrement mutés dans les cellules cancéreuses.

P21, une cible transcriptionnelle de p53, est une protéine particulièrement importante dans l'arrêt du cycle cellulaire en phase G1. Cette protéine est capable d'induire la sénescence dans de nombreuses cellules lorsqu'elle est surexprimée [Chang et al., 2000], alors que sa délétion empêche les cellules de faire de la sénescence [Brown, 1997].

Longtemps considérée comme un mécanisme suppresseur de tumeurs par son action d'arrêt du cycle de cellules tumorales, il est maintenant proposé que la sénescence puisse intervenir dans la prolifération tumorale en modifiant le phénotype sécrétoire des cellules sénescents. En effet, lors de la sénescence, les molécules sécrétées par les cellules varient de sorte que la composition du microenvironnement tumoral est modifiée, perturbant ainsi les interactions entre les cellules [Campisi and d'Adda di Fagagna, 2007]. La sénescence semble donc être à la fois, un mécanisme suppresseur de tumeurs et un mécanisme pouvant favoriser la progression de celle-ci.

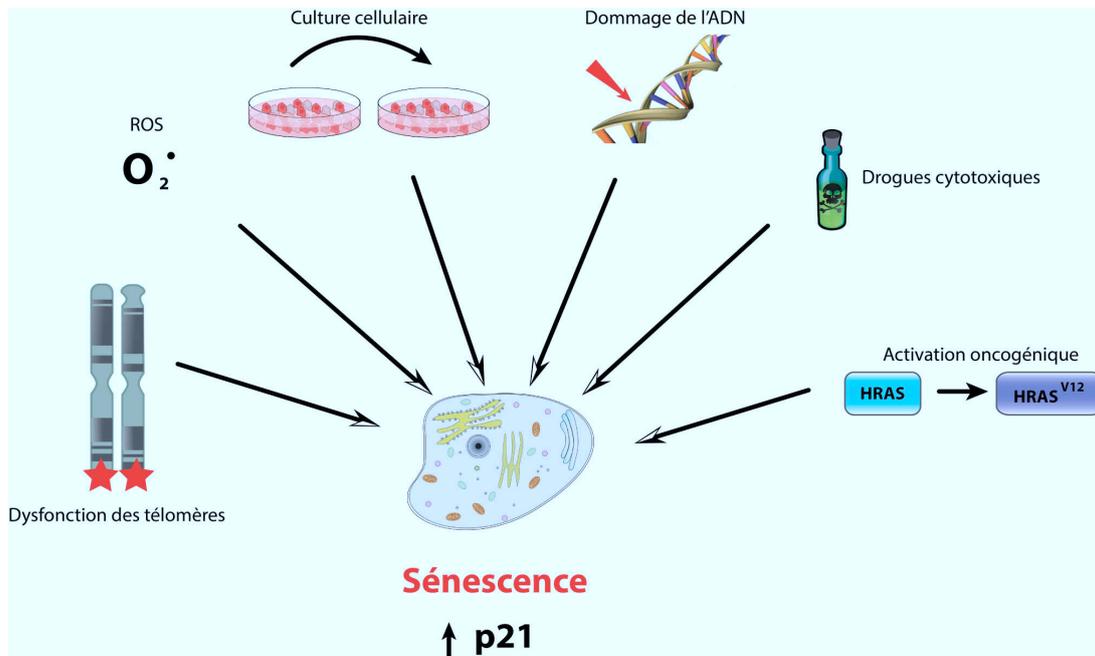


FIGURE 1.7 – **Les mécanismes d’activation de la sénescence.** – La sénescence peut être activée par une dysfonction des télomères (sénescence répllicative), par les espèces réactives de l’oxygène, par le simple fait de cultiver les cellules, par des dommages sur l’ADN, par des drogues cytotoxiques, ou par l’activation de certains oncogènes.

### 1.5.2 Le SASP

Lors de la mise en place de la sénescence en réponse à un stress génotoxique, les cellules se mettent à sécréter une grande quantité de facteurs associés à l’inflammation ou à la progression tumorale constituant le phénotype sécrétoire associé à la sénescence (*Senescence-Associated Secretory Phenotype* ou SASP) [Coppé et al., 2008]. Ces facteurs sont donc sécrétés et peuvent agir de manière autocrine ou paracrine. L’effet paracrine du SASP semble dépendre du contexte cellulaire (cf. figure 1.8 p.25). En effet, les cellules normales ou les cellules précancéreuses de bas grade semblent répondre au SASP en induisant elles aussi la sénescence alors que les cellules précancéreuses de haut grade ou les cellules cancéreuses vont plutôt progresser dans leur développement tumoral [Gorgoulis and Halazonetis, 2010]. Ces facteurs sécrétés peuvent donc s’avérer être de potentiels marqueurs du développement tumoral, ou pourraient permettre de suivre l’évolution de la tumeur en réponse aux traitements notamment génotoxiques. De plus, il semble que la perte du gène suppresseur de tumeur TP53 et l’activation de l’oncogène KRAS – qui induit un stress génotoxique et une entrée en sénescence dans des cellules normales [Di Micco et al., 2006] – exacerbent le versant paracrine de l’action du SASP, favorisant ainsi le développement tumoral. Ces molécules produites par les cellules sénescents ont également

un rôle de régulation du système immunitaire.

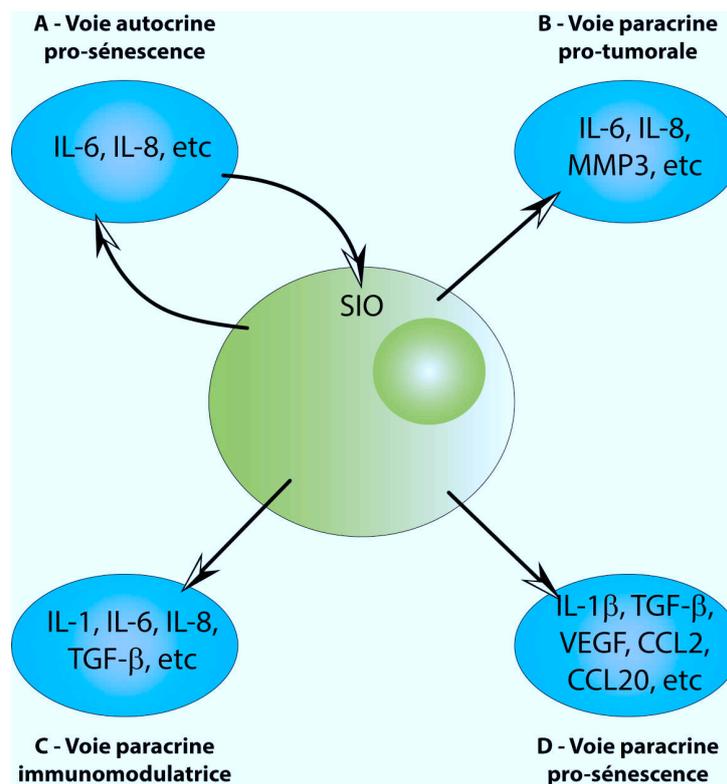


FIGURE 1.8 – Schéma des quatre champs d'action du SASP [Hoare and Narita, 2013]. – (A) La voie autocrine du SASP favorise la sénescence induite par les oncogènes. (B) La voie paracrine peut favoriser le développement tumoral lorsqu'elle agit sur des cellules précancéreuses de haut grade ou sur des cellules cancéreuses. (C) La voie paracrine du SASP a une fonction immunorégulatrice. (D) La voie paracrine du SASP peut favoriser l'apparition ou le maintien de la sénescence lorsqu'elle agit sur des cellules normales ou des cellules cancéreuses de bas grade.

La principale cytokine associée au SASP est l'interleukine-6 (Il-6), un facteur pléiotropique proinflammatoire. Cette protéine a été montrée comme étant surexprimée suite à des dommages sur l'ADN, suite à la sénescence induite par l'oncogène chez la souris, et dans des kératinocytes humains, des mélanocytes, des monocytes, des fibroblastes, et des cellules épithéliales [Lu, 2006] [Sarkar, 2004] [Kuilman et al., 2008]. Rappelons également qu'il s'agit d'un puissant activateur de l'oncogène STAT3 précédemment décrit (cf paragraphe 1.1.3.2 p. 9).

Une autre cytokine importante dans le SASP est l'interleukine-1 (Il-1) [Garfinkel et al., 1994]. Elle est produite par les cellules endothéliales, les fibroblastes, ou encore par les cellules épithéliales sous l'effet de traitements de chimiothérapie [Chang et al., 2002]. Les deux formes de cette cytokine ( $\alpha$  et  $\beta$ ) peuvent agir sur les cellules environnantes si celles-ci expriment à leur surface des récepteurs à l'Il-1 d'une part, mais aussi si elles présentent

des récepteurs de type toll-like. Cette activation est médiée par facteur de transcription  $\text{NF}\kappa\text{B}$  [Mantovani et al., 2001].

Bien d'autres molécules voient leur expression augmenter lors de la mise en place du processus de sénescence. On peut notamment citer certaines chémokines comme CXCL-1, 2, 4, 5, 6, 8, 12 pour celles de type CXCL. Ces molécules sont classées selon l'organisation des cystéines permettant leur conformation tridimensionnelle. On peut également citer les CCL-1, 2, 3, 4, 6, 7, 8, 13, 16, 20, 26 pour celles de type CCL. Il y a également des matrices métallo-protéinases (MMP) dont l'expression est augmentée suite à la mise en place de la sénescence, de même que des protéines de la matrice extracellulaire telles que la fibronectine, ou encore certaines intégrines [Coppé et al., 2010].

Ces protéines, sécrétées par les cellules sénescents, ont été étudiées dans différents contextes et notamment ceux des cancers du sein [Krtolica et al., 2001], de la peau [Sun et al., 2007], de la prostate [Choi et al., 2000], du pancréas [Ohuchida et al., 2004], et des tissus oropharyngés [Coppe et al., 2008]. D'autre part, de nombreux facteurs de ce sécrétome semblent être des protéines particulièrement conservées entre les différents types cellulaires [Coppé et al., 2008] conduisant certains auteurs à penser que ce phénomène touche tous les types de tissu.

En ce qui concerne le cancer colorectal, peu de données sont disponibles dans la littérature. Il est malgré tout démontré que les tumeurs associées à l'oncogène RAS contiennent souvent un nombre important de cellules sénescents [Collado et al., 2005]. Or cet oncogène est particulièrement important de le développement des cancers colorectaux [Nandan and Yang, 2011]. Il est donc probable que l'action du SASP dans ce type de cancer soit particulièrement important. De plus, les tumeurs mutées sur RAS sont hautement vascularisées et cette vascularisation pourrait être la résultante de l'action proangiogénique du SASP associées aux protéines *vascular endothelial growth factor* (VEGF), IL-8 et eotaxine [Salcedo et al., 2001] [Coppé et al., 2006]. Cet effet du SASP a été démontré dans des tumeurs du sein, et la quantité de vaisseaux sanguins est supérieure dans les tumeurs qui se sont développées en présence de fibroblastes en état de sénescence par rapport à celles développées en présence de fibroblastes non sénescents [Coppé et al., 2006].

Tous ces exemples illustrent le fait que, lors du développement tumoral, une sous-population cellulaire va entrer en sénescence pour lutter contre une prolifération non contrôlée entraînant une modification du micro-environnement tumoral par sécrétion d'une quantité importante de facteurs constituant le SASP. Ces molécules, souvent protéiques, peuvent potentiellement diffuser dans la circulation sanguine à proximité de la tumeur,

ce d'autant plus que l'un des effe et constituer une source potentielle de biomarqueurs tumoraux.

## 1.6 Le SASP et des autres types de sécrétomes tumoraux

Par l'intermédiaire des protéines qui le constitue, le SASP intervient dans différents phénomènes physiologiques importants dans le développement tumoral, avec notamment l'inflammation, la croissance cellulaire, le remodelage de la matrice. Ces fonctions régulées par les protéines du SASP montrent la capacité des cellules sénescents à intervenir dans des modifications du microenvironnement de la tumeur, dans la régulation de l'inflammation au niveau de la tumeur et leur capacité à rendre les cellules environnantes plus agressives. Ses fonctions sont contexte physiologique dans lequel se trouve la cellule, et il est généralement associé à un stress génotoxique touchant la cellule sénescents. Finalement, les molécules présentes dans le SASP correspondent à tout ce dont la tumeur a besoin pour progresser, exception faite de l'activation du phénomène inflammatoire. Cela ne constitue pas un problème pour la tumeur, car dans la plupart des cas, celle-ci présente une diminution des antigènes de surface et notamment des molécules du CMHI lui permettant d'échapper au contrôle du système immunitaire [Seliger, 2005].

Il semble à présent clairement établit qu'il existe une relation de compétition entre les cellules tumorales et les cellules non tumorales [Huang et al., 2008]. Il semble cependant que cela ne suffise pas à expliquer que toutes les tumeurs échappent au système immunitaire. Il a été démontré que les sera de patients atteints de tumeurs contenaient une quantité impressionnante de molécules immunosuppressives produites soit par la tumeurs par la le tissu de l'hôte [Ochsenbein, 2002]. Au final,



# Chapitre 2

## Les sources de biomarqueurs

### 2.1 Définition

L'ÉTUDE des biomarqueurs est loin d'être nouvelle. En effet, en 1847, le médecin et chimiste anglais Henry Bence-Jones met en évidence qu'une chaîne légère d'immunoglobuline est retrouvée dans les urines de patients présentant des gammopathies monoclonales encore appelées maladies de Valdenström. Ainsi, la présence de cette protéine dans les urines constitue un élément diagnostique et son taux permet de suivre l'efficacité du traitement [Jones, 1848].

Selon le *National Institute of Health* (NIH), et plus précisément selon le groupe travaillant sur le thème des biomarqueurs, il s'agit d'une caractéristique qui peut être objectivement mesurée et qui reflète l'état physiologique, pathologique ou la réponse pharmacologique à un traitement donné.

Dans son dictionnaire des mots relatifs au cancer, le *National Cancer Institute* (NCI) du NIH donne la définition suivante :

*Biomarqueur* : molécule biologique trouvée dans le sang ou dans tout autre fluide biologique ou tissu, qui est le reflet d'un processus normal ou pathologique, ou d'une condition, ou d'une maladie. Un biomarqueur peut être utilisé pour suivre l'évolution de la réponse à un traitement pour une pathologie donnée ou un état. Ils sont également nommés marqueurs moléculaires ou molécules signatures.

L'organisation mondiale de la santé (OMS) les définit comme toute substance, structure ou processus pouvant être mesuré dans le corps, ou ses produits ou ses influences, et qui permet de prédire l'évolution d'une maladie ou simplement sa présence.

Roy Perlis, un médecin psychiatre travaillant depuis longtemps sur la question des

applications des biomarqueurs au domaine clinique, définit un biomarqueur comme une caractéristique individuelle mesurable qui reflète un facteur de risque pour une pathologie donnée ou un moyen de suivre son évolution. Cette caractéristique peut également permettre de suivre l'éventuelle toxicité d'un médicament. Dans tous les cas, elle doit faire état d'une grande spécificité (dans l'idéal, un marqueur pour une pathologie) et d'une grande sensibilité (la mesure doit être facile à réaliser).

Toutes ces définitions, émanant de différents organismes internationaux, ou de spécialistes de la question des biomarqueurs, tournent autour des mêmes problématiques et se complètent.

Dans une revue publiée récemment dans le journal *Nature* [Perlis, 2011], l'auteur passe en revue les différentes caractéristiques des biomarqueurs que nous allons présenter dans les lignes qui suivent.

## 2.2 Les caractéristiques des biomarqueurs

### 2.2.1 Rappels sur les vrais/faux négatifs et les vrais/faux positifs

Les biomarqueurs sont caractérisés par des données objectives donnant à voir leur intérêt et leur efficacité au premier coup d'œil. Parmi ces données, il est possible de citer les deux principales que sont la sensibilité et la spécificité d'un biomarqueur. Dans les lignes qui suivent nous allons tenter de donner une définition simple de ces données, mais avant cela il nous faut revenir sur quelques notions.

Tout d'abord, un biomarqueur peut donner deux types de réponses :

- Le patient présente un risque d'être malade.
- Le patient ne présente pas de risque d'être malade.

Ensuite, la réalité nous amène à considérer deux conditions également :

- Le patient est malade.
- Le patient n'est pas malade.

Et toutes les combinaisons sont possibles. Cela nous amène à définir les quatre notions suivantes :

- les vrais négatifs (VN) ;
- les faux négatifs (FN) ;
- les vrais positifs (VP) ;
- les faux positifs (FP).

Tableau 2.1 – Les notions de vrais/faux négatifs et de vrais/faux positifs

Statut du biomarqueur	Réalité	
	Positif	Négatif
Positif	vrai positif (VP)	faux positif (FP)
Négatif	faux négatif (FN)	vrai négatif (VN)

### 2.2.2 La sensibilité

La sensibilité d'un test ou d'un biomarqueur correspond au pourcentage de cas positifs correctement identifiés. C'est ainsi que :

$$Sensibilite = \frac{VP}{(VP + FN)}$$

De la sorte, plus le nombre de faux négatifs est faible plus le pourcentage de vrais positifs sera proche de 100% et plus la sensibilité du test sera élevée.

### 2.2.3 La spécificité

La spécificité d'un test ou d'un biomarqueur correspond au pourcentage de cas négatifs correctement identifiés. C'est ainsi que :

$$Specificite = \frac{VN}{(VN + FP)}$$

De la sorte, plus le nombre de faux positifs est faible plus le pourcentage de vrais négatifs sera proche de 100% et plus la spécificité du test sera élevée.

### 2.2.4 L'exactitude

Dans certains cas, il peut être intéressant d'accéder à une autre donnée qui correspond à la capacité d'un test ou d'un biomarqueur à identifier correctement les cas négatifs et les cas positifs. On appelle cette donnée exactitude et elle est définie de la manière suivante :

$$Exactitude = \frac{VN + VP}{(VN + VP + FP + FN)}$$

## 2.3 De l'intérêt des biomarqueurs dans le cancer colorectal

Comme nous l'avons vu, le cancer colorectal constitue un véritable enjeu de santé publique car à l'heure actuelle, il continue de tuer 17 500 personnes par an en France. De plus, comme nous l'avons également déjà mentionné, les chances de survie des patients sont très fortement augmentées lors d'un dépistage précoce de la maladie. Cependant, environ un patient sur cinq est diagnostiqué au stade métastatique (IV). L'un des enjeux actuels pour permettre d'améliorer la précocité du diagnostic réside dans la mise en évidence de biomarqueurs, produits par la tumeur et intervenant rapidement dans le développement tumoral.

En cancérologie, les biomarqueurs sont d'une importance capitale dans le diagnostic mais ils peuvent également s'avérer très utiles dans le suivi thérapeutique, en donnant des informations sur la réponse de la tumeur à la stratégie thérapeutique utilisée.

## 2.4 Classification

Face à l'augmentation importante du nombre de candidats biomarqueurs apparus dans la littérature ces dernières années, il a été tentant d'essayer de les classer pour tenter d'y voir plus clair. Malgré des tentatives nombreuses, il n'existe, à l'heure actuelle, aucun consensus international permettant de proposer une classification claire [Mishra and Verma, 2010]. En effet, le large champ disciplinaire intervenant dans la mise en évidence de biomarqueurs conduit à des difficultés d'homogénéisation des critères de classement. De nombreuses spécialités interviennent, de la biologie cellulaire à la médecine, en passant par les approches globales, rendant l'harmonisation de la classification complexe. Malgré tout, une tentative généraliste a vu le jour dernièrement divisant les biomarqueurs en trois grandes catégories contenant chacune différentes sous catégories.

Tout d'abord les biomarqueurs relatifs à l'état d'une pathologie donnée. Ces derniers peuvent être :

- des biomarqueurs pronostiques ;
- des biomarqueurs prédictifs ;
- des biomarqueurs pharmacodynamiques ;
- des biomarqueurs diagnostiques.

Il est également possible de classer les biomarqueurs selon leur nature moléculaire. Il existe notamment :

- des biomarqueurs ADN ;
- des biomarqueurs ARN ;
- des biomarqueurs protéiques.

### 2.4.1 Les différents types de biomarqueurs

#### 2.4.1.1 Les biomarqueurs pronostiques

Les cancers sont particulièrement hétérogènes, aussi bien d'un patient à l'autre, qu'au sein même de la tumeur [Marusyk et al., 2012]. D'autre part, la plupart des traitements disponibles présentent d'importants effets secondaires chez les patients. Aussi, il est nécessaire d'adapter au mieux le traitement afin d'éviter d'infliger inutilement de lourdes complications à un patient. C'est dans ce contexte qu'interviennent les biomarqueurs pronostiques, qui permettent de classer les tumeurs en fonction de leur agressivité. Concernant le cancer du côlon, le *carcinoembryonic antigen* (CEA) a été montré comme étant un bon « prédicateur » de rechute.

Dans une étude publiée dès 1978 portant sur quelques 358 patients, il est démontré que ceux ayant une tumeur de stade 2 et 3, et présentant un taux de CEA supérieur à 5 ng/ml en pré-opératoire ont un plus fort risque de rechute [Wanebo et al., 1978]. De plus, les patients ayant des tumeurs de stade 3 présentent un temps moyen de rechute de 13 mois lorsque le taux de CEA était supérieur à 5 ng/ml avant l'intervention chirurgicale, alors qu'il est en moyenne de 28 mois lorsque ce taux était inférieur à cette valeur seuil en pré-opératoire [Wanebo et al., 1978].

Une étude publiée plus récemment s'est attachée à faire un bilan de l'utilisation de ce biomarqueur sur 1 361 patients entre les années 2000 et 2004 [Lin et al., 2011]. De cette analyse ils démontrent que le CEA est un facteur pronostique pertinent du cancer colorectal. Les patients ayant un taux de CEA élevé en post-opératoire présentent un risque de récurrence plus important et doivent faire l'objet d'une surveillance régulière (cf figure 2.1 p.34). Ils présentent également une propension à développer des métastases hépatiques importantes.

Pour le cancer du sein, il existe un kit qui permet au clinicien d'avoir une idée du type de tumeur auquel ils sont confrontés. Il s'agit donc de biomarqueurs permettant d'appréhender le comportement de la tumeur et la façon dont elle est susceptible d'évoluer.

Cette figure ne peut être présentée dans ce document en raison d'une restriction des droits de reproduction.

**FIGURE 2.1 – Courbes de survie de patients atteints de cancers colorectaux en fonction du stade tumoral et du taux de CEA en post-opératoire.** Figure reproduite à partir de la publication [Lin et al., 2011] avec l'aimable autorisation de l'éditeur Springer [Numéro de licence : 3198231488336]. – La ligne verte représente les patients ayant un taux « normal » de CEA en post-opératoire. La ligne rouge représente ceux ayant un taux élevé en post-opératoire. (A) Patients atteints de tumeurs de stade 1 : le taux de CEA n'affecte pas la survie des patients. (B) Patients atteints de tumeurs de stade 2 : la survie sans progression à 5 ans des patients ayant un taux de CEA « normal » en post-opératoire est de 87% alors que pour ceux présentant des taux élevés elle est de 62% (C) Patients atteints de tumeurs de stade 3 : la survie sans progression à 5 ans des patients ayant un taux de CEA « normal » en post-opératoire est de 62% alors que pour ceux présentant des taux élevés elle est de 44%.

#### 2.4.1.2 Les biomarqueurs prédictifs

Dans de nombreux cas, lorsque les médecins traitent un patient pour un cancer touchant un organe spécifique, il y a deux types de patients. Ceux qui répondent au traitement, et ceux qui ne répondent pas ou peu. L'objectif des biomarqueurs prédictifs est de discriminer ces deux groupes de patients afin de savoir quel traitement utiliser. Dans le cas du cancer colorectal, la mutation de l'oncogène KRAS peut être utilisée comme prédictive. Cette mutation est une contre indication à l'utilisation du traitement de chimiothérapie ciblée touchant le récepteur à l'EGF [Lievre et al., 2006]. En effet, en cas de mutation, l'effet sur la tumeur est négligeable comparé à la lourdeur des effets secondaires liés à l'utilisation de ces

molécules (violentes réactions cutanées, déshydratation suite à des diarrhées importantes, hypocalcémies, ou encore anorexie pouvant entraîner de fortes pertes de poids, maux de tête fréquents, etc...) [Van Cutsem et al., 2010].

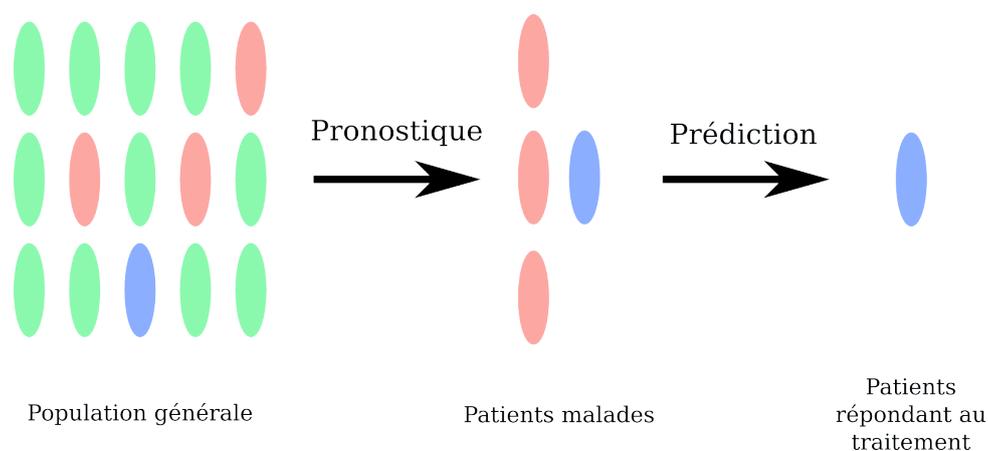


FIGURE 2.2 – **Shéma illustrant la différence entre les biomarqueurs pronostiques et prédictifs** – En vert sont représentés des patients ne présentant pas de cancer. En rouge, des patients ayant un cancers ne répondant pas au traitement. En bleu, des patients ayant un cancer répondant au traitement.

Comme le montre la schéma en figure 2.2 p. 35 la différence entre les biomarqueurs diagnostics et prédictifs est assez subtile. Les premiers permettent de déterminer le type de tumeur que présente le patient, et ses caractéristiques d'agressivité. Les seconds permettent de discriminer quels sont les patients « répondeurs » parmi les malades.

### 2.4.1.3 Les biomarqueurs pharmacodynamiques

Dans le cadre de certaines pathologies, il existe des biomarqueurs permettant de suivre l'évolution de la réponse au traitement. Depuis les années 60, le CEA est décrit comme un biomarqueur et utilisé par les cliniciens. Il s'agit d'une grosse protéine caractérisée par des domaines immunoglobulines présents sur le feuillet extracellulaire de la membrane plasmique des cellules épithéliales, et ancrés à cette membrane par une molécule de glycosylphosphatidyl inositol (GPI). Cette protéine est retrouvée lors du développement embryonnaire, et également au cours de la progression du cancer colorectal. Sa demi-vie dans les fluides biologiques est d'environ deux semaines ce qui lui permet d'être dosée dans le sang des patients au cours de leur traitement. Des taux élevés sont généralement à mettre en lien avec une progression importante de la maladie, et ce taux à tendance à diminuer après l'intervention chirurgicale de résection. En 1986, l'équipe de Fletsher montre qu'en utilisant une valeur seuil de  $2,5\mu\text{g/l}$ , on obtient une sensibilité de l'ordre de

40% [Fletcher, 1986].

Plus récemment, l'équipe de Hundt a réalisé une méta-analyse des études portant sur l'utilisation de ce biomarqueur et ils ont ainsi montré une sensibilité allant de 43 à 69%. Ces résultats sont peu satisfaisants d'autant que dans le détail, la spécificité de ce marqueur tombe à 8% pour les tumeurs de stade 1 [Hundt et al., 2007]. Sa pertinence quant au diagnostic précoce du cancer colorectal est de la sorte particulièrement faible. Bien que spécifique à plus de 90%, ce marqueur est également augmenté dans d'autres types de cancers que ceux touchant le côlon et le rectum, et notamment les cancers de l'estomac, du sein ou du pancréas. Toutes ces données aboutissent à ce que le CEA soit utilisé pour suivre l'évolution de la pathologie cancéreuse, mais son efficacité en terme de diagnostic précoce reste faible.

#### 2.4.1.4 Les biomarqueurs diagnostiques

Ces biomarqueurs sont sans doute les plus recherchés. En effet, il s'agit de biomarqueurs qui, par leur présence ou leur absence, permettent de poser un diagnostic précis d'une pathologie particulière. Il en existe très peu. Les tumeurs de la thyroïde de type médullaires expriment des quantités importantes de calcitonine dès les premières étapes du développement de la tumeur [Niccoli et al., 1997]. De la même façon, certains considèrent la présence du virus *Human Papilloma Virus* (HPV) comme un marqueur diagnostique du cancer du col de l'utérus dans la mesure où celui-ci est présent dans plus de 90% des lésions cancéreuses. Un autre exemple, récemment approuvé par la *Food and Drug Administration* (FDA), le test diagnostique OVA1, basé sur la détection d'un panel de cinq protéines découvertes par protéomique et qui renseigne sur la malignité des tumeurs de l'ovaire [Ueland et al., 2011]<sup>1</sup>.

#### 2.4.1.5 Les biomarqueurs de type ADN

Cette classe de biomarqueurs est caractérisée par une atteinte du matériel génétique. Il peut donc s'agir de modifications importantes, telles que des translocations, des aneuploïdies ou des pertes d'hétérozygoties. Mais il peut également s'agir de modifications beaucoup plus discrètes, telles que des mutations ponctuelles touchant des gènes spécifiques, induisant une perte ou un gain de fonction pour la protéine correspondante. Des exemples ont déjà

---

1. La validation par la FDA est trouvable ici : FDA Clears a Test for Ovarian Cancer Test can help identify potential malignancies, guide surgical decisions : <<http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2009/ucm182057.htm>> (Page consultée le 31 janvier 2013)

été décrits dans ce tapuscrit avec notamment les délétions 1p ou 8q détectées dans le cancer colorectal, de même que certaines pertes d'hétérozygotie touchant les régions 5q, 17p ou encore 18q [Soreide et al., 2006]. On peut également mentionner l'instabilité microsatellitaire et les différentes mutations touchant les gènes importants de régulation de la prolifération tumorale. Il peut s'agir d'oncogène – KRAS muté code pour une protéine qui est constitutivement active – ou de gènes suppresseurs de tumeurs – TP53 muté code pour une protéines qui ne joue plus son rôle de « gardien du génome » – ce qui conduit à des phénotypes différents [Tejpar et al., 2010].

#### 2.4.1.6 Les biomarqueurs de type ARN : les miRNA

Les microARN (miRNA) sont des séquences d'ARN non codantes de 19 à 22 nucléotides impliquées dans la régulation du développement cellulaire, de la différenciation, de la prolifération, de l'apoptose et du métabolisme [Bartel, 2004]. Environ 30% des gènes codant pour des protéines seraient régulés par des miRNA [Filipowicz et al., 2008]. À l'heure actuelle, il est assez clairement établi que ces miRNAs interviennent dans la mise en place de processus pathologiques tels que les cancers [Calin and Croce, 2006]. Ainsi, une expression aberrante de miRNA constitue un marqueur potentiellement utilisable pour le suivi de certains cancers.

Une méta-analyse publiée en 2011 résume les données obtenues dans le cadre de l'étude du cancer colorectal [Luo et al., 2011]. Elle prend en compte vingt études qui ont cherché à déterminer les modifications intervenant dans le niveau d'expression des miRNA au niveau tissulaire ; et trois les ont regardées au niveau plasmatique. En effet, un article récent a montré que les miRNA pouvaient se retrouver dans le milieu extracellulaire [Turchinovich et al., 2011] [Valadi et al., 2007]. On les retrouve dans divers fluides biologiques tels que les urines, la salive, le sérum ou le plasma [Chen et al., 2008]. Dans ce cas, ils sont pris en charge par des protéines Ago, ce qui permet de les protéger de l'attaque des RNAses – enzymes ayant pour fonction de dégrader les ARN – circulantes [Turchinovich et al., 2011]. Au total, c'est 164 miRNA qui semblent être dérégulés dans le cancer colorectal sur les quelques 700 étudiés sur 1 525 adénomes et adénocarcinomes et 1 102 contrôles. Dans les tissus, les miR-20a et miR-31 sont détectés dans plusieurs études comme étant sur-exprimés de façon significative, et les miR-143 et miR-145 sont statistiquement sous-exprimés. En conclusion de leur article, ils indiquent que les miRNA présents dans le plasma des patients sont des sources intéressantes de biomarqueurs mais des étapes de validation à large échelle

restent nécessaires. D'autre part, des études complémentaires devront être menées sur des tumeurs plus précoces, et des adénomes.

#### **2.4.1.7 Les biomarqueurs de type protéique**

Nous ne développeront pas ici cette catégorie de biomarqueurs car elle fait l'objet d'un développement très approfondi dans les lignes qui suivent (Cf. chapitre 2.5 p. 39).

#### **2.4.1.8 Les biomarqueurs de type glucidique**

L'aspect ubiquitaire de la glycosylation, et son implication dans divers processus biologiques tels que les interactions cellule-cellule, le contrôle structural des protéines, ou encore la fixation sur des récepteurs, en font une source potentielle de biomarqueurs [Ruhaak et al., 2013]. Il est à présent clairement établi que les N-glycosylations sont affectées lors de la mise en place des processus de cancérogénèse [Dube and Bertozzi, 2005]. De ce constat est née une spécialité appelée *glycomique* dont l'objectif est d'analyser les structures glucidiques complexes fixées sur les protéines, et notamment les N-glycosylations. L'un des intérêts de cette approche est également de donner un moyen d'affronter un déficit technologique de taille. En effet, nous y reviendrons, mais la complexité du sérum est un facteur limitant son analyse par des approches globales. Plusieurs études ont ainsi réussi à montrer des modifications de glycosylation entre un état pathologique, et un état contrôle. Une étude publiée par l'équipe de Miyamoto en 2008 montre des différences entre les motifs glucidiques de protéines issues de milieux conditionnés de cellules du cancer de l'ovaire, mais surtout entre le sérum de patientes malades et celui de patientes non malades [Li et al., 2008]. Ils montrent ainsi des changements du type de glycosylation affectant notamment la protéine Apo B-100. Une autre étude a par exemple cherché des modifications ciblant les protéines durant une phase inflammatoire aigüe. Ils expliquent ainsi que ces glycoprotéines, une fois modifiées, peuvent activer des récepteurs différents, et mettre en jeu de nouvelles voies de l'inflammation [Dempsey and Rudd, 2012].

#### **2.4.1.9 Les biomarqueurs associés à l'imagerie médicale**

Ces dernières années, l'imagerie médicale a connu un formidable essor, permettant d'adapter au mieux les thérapeutiques au patient. Il peut alors s'agir d'imagerie anatomique, fonctionnelle, ou encore moléculaire. Deux techniques sont principalement utilisées dans ce cadre, la tomographie par émission de positons (TEP) et l'imagerie par résonance

magnétique (IRM). Il est ainsi possible d'utiliser des marqueurs de prolifération comme par exemple, la thymidine couplée à un atome de  $^{18}\text{F}$  qui vont pouvoir marquer les cellules présentant une forte activité proliférative. La mort cellulaire peut également être suivie par IRM en déterminant l'ADC, un coefficient de diffusion (Apparent Diffusion Coefficient). Il en existe d'autres mais il ne sera pas fait état de ces approches plus en détail dans ce tapuscrit.

## 2.5 Les sources de biomarqueurs protéiques potentiels

Dans la littérature, les protéines sécrétées sont étudiées à partir de différentes sources. L'hypothèse consiste à penser que, du fait de leur rôle dans les contacts cellules-cellules, dans la communication intercellulaire ou encore dans la croissance, ces protéines reflètent le niveau d'évolution d'une pathologie donnée [Doroudgar and Glembotski, 2011]. Elles sont analysées par des techniques de spectrométrie de masse sur lesquelles nous reviendrons dans la suite de l'exposé.

En première intention, les efforts se sont d'abord portés sur l'analyse directe des fluides biologiques, tels que le sang, le plasma, les urines [Stastna and Van Eyk, 2012] pour les plus courantes, mais on peut également citer le liquide amniotique, le lait maternelle, le sperme, ou encore la salive.

Plus récemment, de nouvelles sources d'études sont apparues. Des équipes se sont par exemple intéressées à étudier directement les exosomes et les membranes vésiculaires par protéomique après les avoir isolés [Raimondo et al., 2011]. Certains ont également étudié les liquides interstitiels tumoraux que ce soit chez l'homme [Ambartsumian et al., 2004] ou chez la souris [Xu et al., 2011]. Pour ce faire, il est possible de mettre en culture des fragments de tumeurs et de récupérer les protéines sécrétées. Une technique de microdialyse a également été développée pour permettre une étude intratumorale [Xu et al., 2011]. Ces études semblent majoritairement réalisées pour les tumeurs du sein. Beaucoup d'auteurs s'intéressent également à des milieux conditionnés de cellules en culture. Une revue, publiée en 2011 par Dowling et Clynes présente ces approches dans le détail [Dowling and Clynes, 2011]. Dans la majorité des études, les cellules sont cultivées quelques temps en conditions normales, c'est à dire avec du sérum de veau foetal, puis elles sont lavées plusieurs fois et déplétées en sérum de veau pour la culture pendant généralement 24h. Au bout de ce temps de carence, le surnageant de culture des cellules est analysé.

Connus depuis les années 60, grâce notamment à l'équipe de R. Baldwin, les autoanticorps

sont produits par le système immunitaire en réponse au développement d'une tumeur [Baldwin, 1966] [Baldwin, 1971]. En effet, dès les premières étapes du développement de la tumeur, les cellules acquièrent à leur surface des autoantigènes tumoraux fortement immunogènes capables de déclencher une réponse humorale. Ces anticorps constituent une source de biomarqueur intéressante dans la mesure où ce sont des molécules qui apparaissent très précocément dans le développement de la tumeur. Leur étude est rendue possible grâce aux avancées technologiques qui ont eu lieu ces dernières années en protéomique. Les principales approches sont le SEREX (approche basée sur l'utilisation de banques d'ADN complémentaires), le SERPA (approche basée sur l'utilisation de l'électrophorèse bidimensionnelle), et les puces à protéines [Desmetz et al., 2011].

## 2.6 Le sécrétome

### 2.6.1 Définition

Le sécrétome est constitué par l'ensemble des protéines sécrétées par une cellule, un tissu ou un organisme à un instant  $t$  donné et soumis à des conditions environnementales précises. Il s'agit d'un sous-protéome correspondant à l'ensemble des protéines retrouvées dans le milieu extracellulaire. Il peut s'agir d'un milieu conditionné de cellules en culture, ou des protéines présentes dans la matrice extracellulaire.

Depuis les années 2000, date à laquelle le terme de « sécrétome » a été employé pour la première fois par Tjalsma et son équipe [Tjalsma et al., 2000], le nombre de publications ne cesse de croître chaque année sur le sujet. Dans leur étude, ils se sont intéressés aux protéines sécrétées par la bactérie *Bacillus Subtilis* et ils ont appelé ces protéines, le sécrétome de la bactérie. Dix ans plus tard, en 2010, 349 articles étaient publiés sur le thème du sécrétome [Makridakis and Vlahou, 2010]. À l'heure actuelle<sup>2</sup> ce ne sont pas moins de 907 disponibles sur le sujet.

Chez les organismes supérieurs, la sécrétion de protéines est un processus extrêmement contrôlé, qui peut emprunter plusieurs voies, et de la sorte être soumis à différentes régulations. Dans la plupart des cas, les molécules sécrétées sont des molécules actives, qui vont avoir des fonctions physiologiques indispensables à la communication des cellules avec leur environnement. L'environnement peut également conditionner le sécrétome des cellules. Une perturbation de ces mécanismes de sécrétion peut être le reflet d'une perturbation

---

2. le 14 août 2013

cellulaire importante ou sa cause [Chenau et al., 2008].

L'ensemble de ces caractéristiques font du sécrétome une source intéressante de biomarqueurs, ou de cibles thérapeutiques dans de nombreuses pathologies et notamment dans les cancers. Nous reviendrons plus tard dans l'exposé sur cet aspect. Pour le moment, arrêtons nous quelques instants sur les différentes possibilités qu'a la cellule pour « adresser » une protéine vers l'espace extracellulaire.

### 2.6.2 Les différentes voies de sécrétion

Il y a plusieurs voies qui permettent à la cellule de sécréter des protéines :

- la voie de sécrétion classique ;
- la voie de sécrétion non-classique ;
- la voie de sécrétion utilisant les exosomes.

Ces voies de sécrétion sont utilisées par la cellule pour adresser des protéines spécifiques à l'espace extracellulaire. La figure 2.3 p. 41 illustre ces différentes voies.

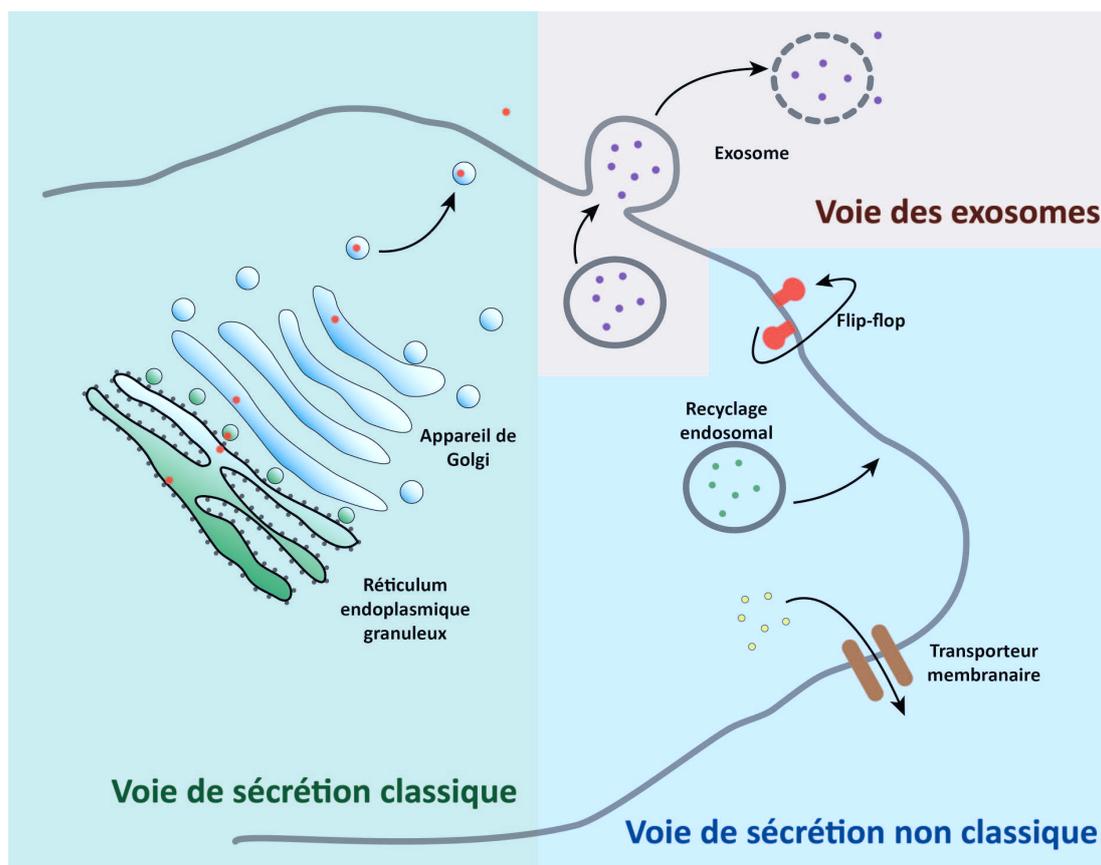


FIGURE 2.3 – Schéma illustrant les différentes voies de sécrétion – En vert, la voie de sécrétion classique passant par le réticulum endoplasmique et l'appareil de Golgi dans lequel est clivé le peptide signal. En orange, la voie des exosomes. En bleu, les mécanismes faisant partie de la voie de sécrétion non classique.

### 2.6.2.1 La voie de sécrétion classique

Les protéines sécrétées par la voie classique présentent un peptide signal de 15 à 30 acides aminés à leur extrémité N-terminale. Ces protéines sont traduites par un ribosome « chargé » sur le réticulum endoplasmique cellulaire. À mesure de la traduction, la protéine croît dans la lumière, et est prise en charge par des protéines chaperonnes qui vont lui permettre d'acquérir la structure tridimensionnelle compatible avec sa fonction biologique. Dans le cas contraire, la protéine sera orientée vers le mécanisme d'épuration des protéines mal conformées : le protéasome.

Le peptide signal dirige la protéine vers le réticulum endoplasmique, dans lequel il sera clivé pour libérer la protéine dans la lumière de ce dernier et permettre la maturation de la protéine, et notamment les glycosylations. Une fois cette étape réalisée, la protéine est orientée vers l'appareil de Golgi pour terminer sa maturation. En effet, certaines glycosylations ne peuvent avoir lieu que dans l'appareil de Golgi. D'une manière générale, les protéines majoritairement O-glycosylées sont adressées au compartiment intracellulaire, alors que les protéines N-glycosylées sont surtout adressées à la face externe de la membrane plasmique et au milieu extracellulaire (sécrétion) [Wisniewski, 2010a] [Roth, 2002]. Enfin, les protéines ayant vocation à être sécrétées vont se retrouver dans des vésicules de Golgi qui vont les transporter jusqu'à la membrane plasmique le long du cytosquelette cellulaire. Ces vésicules vont alors fusionner avec la membrane plasmique au niveau de structures appelées les porosomes, permettant le relargage dans le milieu extracellulaire de la protéine sous une forme soluble.

### 2.6.2.2 La voie de sécrétion non classique

Les protéines sécrétées par la voie non classique ne possèdent pas de peptide signal et elles sont sécrétées par des mécanismes ne faisant intervenir ni le réticulum ni l'appareil de Golgi. C'est il y a une vingtaine d'années que cette voie a été mise au jour, en s'intéressant à deux protéines particulières que sont l'interleukine-1 $\beta$  [Rubartelli et al., 1990] et la galectine-1 [Cooper and Barondes, 1990]. Bien qu'il ait été cru un temps que ces protéines faisaient office d'exceptions, il s'avéra plus tard qu'il s'agissait en réalité d'une voie de sécrétion certes non classique, mais malgré tout, beaucoup utilisée par la cellule. À la fin des années 80, il était supposé que ces voies de sécrétion étaient mise en jeu lors de stress, ou de blessures mécaniques telles que celles qui peuvent être mises en jeu lors de l'angiogénèse ou l'inflammation [McNeil et al., 1989]. Or, des études ultérieures ont montré

qu'il s'agissait bel et bien d'un phénomène parallèle à la sécrétion classique, mais pour lesquels les protéines ne présentent pas de peptides signaux dans leur séquence primaire. D'autre part, il existe une drogue, la brefeldine A (BFA), qui est capable de bloquer la voie de sécrétion classique, mais cette drogue ne bloque pas la sécrétion non classique, preuve que ces deux voies sont indépendantes [Cleves, 1997].

Les protéines utilisant cette voie ne sont pas des protéines d'importance moindre par rapport à celles qui empruntent la voie classique. En effet, on y retrouve des facteurs de croissance, des cytokines et d'autres molécules intervenant dans les phénomènes d'inflammation, d'angiogénèse, de différenciation ou encore de prolifération cellulaire.

Il existe à l'heure actuelle, quatre processus décrits par Nickel [Nickel, 2005] expliquant ce que subissent les protéines sécrétées par la voie non classique.

1. Les protéines peuvent être sécrétées en passant par des lysosomes particuliers capables de fusionner avec la membrane plasmique, de la même manière que les vésicules de l'appareil de Golgi dans la voie classique.
2. Les protéines peuvent également être sécrétées par l'action d'exosomes particuliers issus de corps multivésiculaires présents dans le cytoplasme cellulaire et recyclés pour ce mode de sécrétion.
3. Les protéines peuvent être sécrétées en formant des exovésicules constituées par la membrane plasmique qui accompagne la protéine sécrétée et qui sera dégradée par la suite pour permettre sa libération dans le milieu extracellulaire.
4. Les protéines peuvent être à l'état soluble dans le cytoplasme et passer à travers la membrane plasmique, directement dans le milieu extracellulaire, grâce à l'action de récepteurs spécifiques, ou par un mécanisme de *flip-flop*. Certaines protéines peuvent se fixer à la face interne de la membrane plasmique et se retrouver à la face externe par une inversion des phospholipides membranaires.

Enfin, certaines protéines sont également sécrétées à l'aide de transporteurs membranaires spécifiques permettant à ces molécules de franchir la bicouche lipidique de la membrane plasmique.

### 2.6.2.3 Le relargage par des exosomes

La troisième voie de sécrétion correspond au relargage de protéines contenues dans les exosomes. Il s'agit donc de l'exocytose de vésicules membranaires contenant des protéines ubiquitaires telles que des enzymes du métabolisme, des protéines du cytosquelette ou

encore des protéines chaperonnes. Ce mode de sécrétion est connu depuis le début des années 80. Cependant, il a connu un regain d'intérêt durant les dernières années, car des équipes ont pu montrer que les exosomes pouvaient contenir, en plus des protéines sécrétées, des ARN messagers (ARNm) et des microARN [Valadi et al., 2007].

### 2.6.3 Intérêts

Comme nous l'avons vu, les cellules tumorales, parmi lesquelles certaines sont en sénescence présentent une modification de leur phénotype sécrétoire. On peut donc émettre l'hypothèse selon laquelle, lors de son développement, la tumeur augmente le niveau d'expression de certaines protéines sécrétées qui pourraient ainsi se retrouver libres dans la circulation sanguine. C'est la raison pour laquelle de nombreuses publications s'intéressent maintenant à étudier plus spécifiquement le sécrétome de lignées de cancers ou de tumeurs [Surinova et al., 2011]. À l'heure actuelle, tous les biomarqueurs connues sont des protéines sécrétées, ou membranaires, ce qui conduit à considérer le sécrétome des tumeurs comme une source importante de potentiels biomarqueurs. Son étude par des approches globales est de ce fait un enjeu primordial. C'est ce que nous allons tenter de voir dans les prochains paragraphes.

## 2.7 Des biomarqueurs protéiques potentiels obtenus par protéomique

Dans la littérature, il existe d'ores et déjà pléthore de biomarqueurs potentiels obtenus par protéomique. Ces derniers ont été obtenus à partir de différents types de modèles d'étude. C'est ce que nous allons tenter de voir dans les lignes qui suivent.

### 2.7.1 À partir de modèles in vitro

L'équipe de Xue et ses collaborateurs s'est intéressée au sécrétome des lignées SW480, et de la lignée métastatique correspondante, les SW620. Ils ont pu mettre en évidence 145 protéines différentiellement exprimées dans le sécrétome de la lignée SW620. Parmi les 145 protéines qu'ils ont identifiées comme différentiellement exprimées, ils en ont validées 6 par *western blot*, à savoir les protéines TFF3, GDF15, AGR2, TGM2, LCN2, et IGFBP7 [Xue et al., 2010]. D'autre part, TFF3 et GDF15 ont été dosés par ELISA dans une large cohorte d'échantillons de sérums (114 *séra* de patients ayant un cancer colorectal et 156 sujet non

malades). Le taux de ces deux protéines est significativement supérieur pour les patients présentant un cancer colorectal avec envahissement ganglionnaire.

Parmi les études du sécrétome, une autre étude paraît importante à mentionner, celle de Wu et de ses collaborateurs. Ils ont réalisé une étude portant sur 21 lignées cellulaires correspondant à 12 localisations différentes de cancer parmi lesquelles deux lignées de cancer colorectal [Wu et al., 2008]. De la sorte, ils ont pu mettre en évidence la protéine *collapsin response mediator protein-2* (CRMP-2) comme une protéine sur-exprimée dans les cancers colorectaux précoces et ceux présentant des métastases ganglionnaires.

Une autre approche a été également proposée par l'équipe de Klein-Scory. Elle est basée sur le fait qu'une tumeur génère une réponse immunitaire de la part du patient conduisant à la production d'antigènes spécifiques de cette dernière [Klein-Scory et al., 2010]. Ils ont donc pris 5 sécrétomes de lignées colorectales qu'ils ont groupés et fait migrer sur un gel 2D pour avoir une référence. Ils ont ensuite hybridé différents sérums de patients afin de détecter s'ils contenaient des anticorps capables de reconnaître des antigènes tumoraux. Ensuite ils ont révélé leurs gels avec des anticorps secondaires couplés à des fluorochromes et ils ont identifié les antigènes présents sous les anticorps de plasma de patients. Grâce à cette approche, ils ont mis au jour la protéine Glod4 et un fragment C-terminal de l'agrine comme de potentiels biomarqueurs du cancer colorectal.

Par ailleurs, les protéines fixées à la membrane sont particulièrement intéressantes, surtout lorsqu'elles ont un domaine extracellulaire pouvant servir de récepteur. Plusieurs études montrent comment extraire ces protéines par une technique basée sur l'utilisation de la biotinylation de protéines de surface. L'une d'entre elles a comparé le surfaceome d'une lignée colorectale ayant un faible pouvoir métastatique (KM12C), à celui d'une autre lignée ayant un fort pouvoir métastatique (KM12SM). Ils montrent alors que les protéines sur-exprimées dans les KM12SM sont des biomarqueurs métastatiques potentiels. Ils ont ainsi validé CDH17, JUP et l'EGFR par immunohistochimie sur des tumeurs [Luque-García et al., 2010].

Une autre de ces études s'est penchée sur l'identification de biomarqueurs de la transition adénome à adénocarcinome. Ils ont pu mettre en évidence la protéine PrP<sup>c</sup> et GLUT1 comme étant sur-exprimées à la surface des adénocarcinomes par rapport aux adénomes, et ils les ont validées par des marquages immunohistochimiques réalisés sur puces de tissus (*tissue microarrays*) [de Wit et al., 2012]. Par ailleurs, les protéines identifiées dans ces deux publications sur des lignées différentes présentent 80% de recouvrement, indiquant une relative constance dans l'expression des antigènes de surface des lignées de cancer

colorectal.

Un autre moyen de faire de la protéomique est de s'intéresser à des sous-compartiments cellulaires. Cette approche a été utilisée par l'équipe de Boisvert, sur des lignées colorectales HCT116. Ils ont ainsi généré des fractions cytosoliques, nucléaires et nucléolaires qu'ils ont analysés par spectrométrie de masse de façon indépendante [Boisvert et al., 2010]. Ils ont également appliqué ce protocole pour analyser les mêmes cellules traitées à l'étoposide – une molécule responsable de dommages à l'ADN – et leurs données révèlent notamment une relocalisation des protéines du protéasome du cytoplasme vers le noyau. Cette approche paraît donc particulièrement puissante pour mettre en évidence, par approche globale, des relocalisations de protéines au sein des différents compartiments cellulaires.

Les lignées cellulaires peuvent également être utilisées pour des approches protéomiques en modifiant uniquement une voie de signalisation particulière. L'équipe de Flavin s'est par exemple intéressée à la régulation de hTERT [Flavin et al., 2011]. L'effet de la diminution de l'expression de c-Myc par shARN a également été étudiée par Zhang et ses collaborateurs [Zhang et al., 2011]. À l'inverse, l'équipe de Pocsfalvi a pu étudier l'effet d'une sur-expression de c-Myc dans des lignées par approche protéomique [Pocsfalvi et al., 2011]. D'autres, ont comparé une lignée mutée sur APC, et cette même lignée dans laquelle ils avaient restauré l'expression de cette protéine [Halvey et al., 2011]. Ces études montrent qu'il est techniquement possible d'utiliser des modèles cellulaires afin d'établir des protéomes signatures de régulation de voies de signalisation.

### 2.7.2 À partir de modèles *in vivo*

Les modèles animaux sont souvent des modèles de choix pour les études *in vivo* et notamment les souris. En protéomique, seule une poignée de publications propose des approches protéomiques réalisées sur des modèles murins. Parmi ces études, une équipe s'est intéressée aux cellules souches de tumeurs colorectales, menée par Di Palma et ses collaborateurs [Di Palma et al., 2011]. Pour cela, ils ont récupéré des souris qu'ils ont sacrifiées pour récupérer les cellules de l'épithélium intestinal. Ils ont ensuite trié les cellules par cytométrie en flux et ont alors pu sélectionner les cellules ayant un phénotype souche. Afin de permettre une bonne isolation des cellules souches, les souris ont été modifiées génétiquement de sorte que la protéine LGR5 – marqueur de cellules souches – (Cf. paragraphe 1.1.2.3 p. 5) soit couplée à la *Green Fluorescent Protein* (GFP). Ils ont ensuite réalisé une étude protéomique de ces cellules à partir de laquelle ils ont pu identifier

3 775 protéines.

Une autre étude, celle de Ang, s'est intéressée aux selles de souris [Ang et al., 2010]. Ils ont alors pu comparer des selles issues de souris mutées sur le gène APC, donc susceptibles de développer des tumeurs colorectales plus aisément, à des souris contrôles. Ils ont ainsi mis en évidence 336 protéines parmi lesquelles 115 protéines d'origine murines, 201 protéines issues d'organismes de la flore intestinale essentiellement bactériens, et 18 protéines issues de résidus du bol alimentaire. Par cette étude, ils font la preuve de concept que l'on peut faire de la protéomique sur des matières fécales pour obtenir l'identification de potentiels biomarqueurs concernant des pathologies digestives. Cinq de ces protéines ont retenu leur attention : l'haptoglobine, l'hémoglobine, l'hémopexine, l'alpha-2-macroglobuline et la cadhérine 17.

L'équipe de Fijneman s'est également penchée sur l'utilisation d'un modèle de souris pour mettre en évidence des biomarqueurs précoces du cancer colorectal [Fijneman et al., 2012]. Pour cela, ils ont utilisé des souris présentant une propension à développer des cancers, et ils ont réalisé des extraits de ces tumeurs à 202 jours de vie. Ils ont ainsi pu identifier 2 172 protéines parmi lesquelles 192 varient de façon significative, et s'avèrent être cinq fois plus sécrétées dans les tumeurs. Une analyse statistique de leurs données les a conduit à définir 52 biomarqueurs potentiels tels que, S100A9 et MCM4 validés par immunohistochimie, et la protéine CHI3L1 validée par dosage *enzyme-linked immunosorbent assay* (ELISA) sur des *séras* de patients. Ces derniers étaient soit bien portant, soit ils présentaient un adénome, soit ils présentaient un cancer colorectal.

Naba et ses collaborateurs, se sont quant à eux intéressés à ce qu'ils ont appelé le « matrisome ». Ces travaux publiés en 2012, avaient pour objectif de définir les protéomes de la matrice extracellulaire des tissus de cancer de poumon ou du côlon. Ils ont ainsi mis en évidence une liste de 100 protéines potentiellement utilisables pour définir une signature de ces cancers. Ils ont également remarqué que, lorsqu'ils injectent des cellules tumorales humaines dans leurs souris, ces cellules et les cellules stromales environnantes participent au modelage de la matrice extracellulaire. Ils ont pu montrer cela en faisant des xénogreffes sur les souris et en analysant leurs échantillons en comparant leurs séquences protéiques aux bases de données humaines, et murines. De la sorte, ils prouvent la contribution de la tumeur elle-même, mais aussi de son stroma dans le remodelage de la matrice extracellulaire [Naba et al., 2012]. De plus, selon leurs résultats, les protéines de la matrices sécrétées par la tumeur sont très différentes en fonction du potentiel métastatique de la tumeur.

### 2.7.3 À partir de tissus biologiques

Les sources de biomarqueurs précédemment décrites que sont les urines, le sang, ou encore les modèles *in vitro* sont des sources qui, la plupart du temps sont assez faciles d'accès. Cependant, la plupart des études qui les utilisent sont confrontées à des difficultés pour analyser ces fluides. Ces difficultés résident dans la complexité même de l'échantillon qui rend son analyse complexe, en particulier pour le protéome du plasma. Il s'agit donc d'un véritable défi technologique. De plus, ces analyses nécessitent d'avoir un nombre important de patients analysés pour avoir une forte reproductibilité statistique. En effet, les variations inter-individus sont particulièrement importantes dans le cas des fluides biologiques. Il est donc nécessaire d'avoir un échantillonnage important pour discriminer les variations dues à la tumeur de celles dues à la variabilité entre les sujets analysés [Surinova et al., 2011]. L'autre difficulté de cette approche, ce sont les grandes variations de concentrations entre les différentes protéines constituant les fluides biologiques. En effet, il est montré qu'il y a un facteur pouvant aller jusqu'à  $10^{12}$  entre les protéines les plus abondantes – dont les concentrations se situent autour du mg/mL – et les plus faiblement concentrées – dont les concentrations se situent autour du pg/mL – [Anderson and Anderson, 2002].

Quelques équipes se sont donc proposées de travailler directement sur des tissus biologiques avec pour objectif de comparer les tissus tumoraux aux tissus sains correspondants. L'objectif consiste à détecter d'éventuelles variations de niveau d'expression protéique entre la tumeur et l'épithélium sain correspondant et nécessite d'utiliser des approches de protéomique quantitative. Suite à une analyse globale utilisant l'approche de 2D-DIGE – nous reviendrons plus en détails sur cette technique ci-après (cf. paragraphe 3.4.1.2 p. 66) – une équipe française a pu mettre en évidence la protéine *Heat Shock Protein 60* (HSP60) comme marqueur tardif du cancer colorectal. Pour cela, ils ont analysé le protéome global de 23 tumeurs de patients et leurs tissus sains associés. L'étape de validation a été réalisée par des tests ELISA sur quelques 112 patients ayant un cancer et 90 patients sains. Ils démontrent ainsi que HSP60 présente une spécificité proche de celle du CEA [Hamelin et al., 2011].

Par une autre approche, l'équipe de Yang a montré que l'association entre la cathepsine D et l' $\alpha_1$ -antitrypsine permet de discriminer les tissus sains des tissus tumoraux à plus de 96%. Une analyse des protéomes de 37 tumeurs comparés aux tissus sains associés et une validation par immunomarquage sur puces de tissus les a conduits à proposer ces deux protéines comme biomarqueurs précoces potentiels du cancer colorectal [Xie et al., 2010].

Deux équipes ont, quant à elles, axé leur étude sur une validation robuste de leur analyse globale. Ainsi, l'équipe de Tweedle et celle d'O'Dwyer ont respectivement mis en évidence qu'HSP27 et la protéine 14-3-3 $\beta$  sont associées à un mauvais pronostic, en réalisant leur validation sur des cohortes de plus de 400 tissus de patients par immunohistochimie sur puces de tissus [Tweedle et al., 2010] [O'Dwyer et al., 2011]. Kim et ses collègues ont souhaité proposer une approche pour tenter de discriminer les patients à fort risque de rechute de ceux à faible risque. Pour cela ils ont fait une double approche protéomique quantitative, l'une utilisant la technique DIGE et l'autre utilisant une forme clivable de l'ICAT (cf. paragraphe 3.4.2.4 p. 70). Ils ont alors appliqué ces approches en poolant les lysats de tumeurs de patients ayant un bon pronostic pour les comparer à un pool de lysats de tumeurs de patients ayant un mauvais pronostic. Ainsi, ils proposent un panel de cinq protéines permettant de différencier les deux groupes [Kim et al., 2011].

Une approche particulièrement importante a été proposée en 2011 par Wisniewsky de l'équipe de Matthias Mann. Cette approche consiste à utiliser du tissu fixé en paraffine pour réaliser une approche de protéomique quantitative. Basée sur leur technique de FASP (*Filter-Aided Sample Preparation*) ils ont réussi à extraire les protéines de ces tissus fixés, et ils ont obtenu un profil protéomique de plus de 6 000 protéines [Wisniewski et al., 2011a].

Il est également possible de se limiter à l'étude de cellules souches colorectales. Ce sont des approches en cours de développement mais qui semblent malgré tout prometteuses. Ainsi, l'équipe de DiPalma a réalisé une étude sur les cellules Lgr5 positives, autrement dit, sur des cellules considérées comme les cellules souches colorectales [Di Palma et al., 2011].

L'étude de Fang et ses collaborateurs a utilisé le modèle de culture en trois dimensions (sphéroïdes) pour étudier le protéome de cellules souches, exprimant le marqueur CD133. Ces sphéroïdes, formées à partir de cellules de tumeurs primaires de patients sont particulièrement résistantes à l'irinotécan. Ils montrent que des cellules maintenues dans ces conditions de culture gardent leurs propriétés de cellules souches en restant CD133<sup>+</sup>. D'autre part, une analyse protéomique montre que ces cellules CD133<sup>+</sup> expriment aussi fortement les marqueurs CD44, CD166, CD29, CEACAM5, CADH17, et biglycan, qui sont des marqueurs de cellules souches [Fang et al., 2010].

Comme nous l'avons déjà mentionné, la complexité de l'échantillon est une limite importante pour une bonne identification des protéines d'un échantillon. Aussi, certaines équipes ont proposé des techniques permettant de diminuer cette complexité, y compris pour les extraits tumoraux. L'équipe de Yang s'est intéressée à la fraction non soluble des protéines de 13 couples tumeur/tissu sain. Ils ont donc récupéré les protéines non solubles

qu'ils ont analysées par une technique de quantification sans marquage et ils ont mis en évidence 56 protéines variant significativement entre le tissu sain et la tumeur [Yang et al., 2012]. Les protéines membranaires ont particulièrement intéressé l'équipe de Han. Ils ont effectué une étude comparative des protéines membranaires de 28 tumeurs et de leur tissu sain associé, toujours par protéomique quantitative sans marquage (cf. paragraphe 3.4.3 p. 79). De cette analyse, ils proposent la protéine STOML2 comme biomarqueur potentiel, après l'avoir validé par immunohistochimie sur 205 patients. De plus, un test ELISA qu'ils ont développé montre que la détection de cette protéine présente une meilleure sensibilité que le test basé sur le dosage du CEA pour la détection du cancer colorectal [Han et al., 2011].

# Chapitre 3

## Identification de biomarqueurs par protéomique

### 3.1 Définition de la protéomique

**F**AISANT le pendant de la génomique, le terme de protéomique est apparu en 1997 et a été utilisé pour la première fois par James P lors de son article intitulé : *Protein identification in the post-genome era : the rapid rise of proteomics* [James, 1997]. Il s'agissait au départ de définir l'ensemble des protéines exprimées par un génome. Quelques années plus tôt, Marc Wilkins parlait déjà de la notion de protéome [Wilkins et al., 1996]. Depuis la définition s'est précisée et s'applique à l'étude de l'ensemble d'un compartiment cellulaire, d'un tissu, ou d'un organe. Contrairement au génome, dont l'une des caractéristiques principale est sa stabilité dans le temps (à l'exception de quelques mutations ponctuelles), le protéome est extrêmement dynamique [Shen and Smith, 2002]. Cette dynamique résulte de différences d'expression régulées par des mécanismes cellulaires sensibles. Une autre dimension peut également être prise en compte lors de l'étude de la dynamique du protéome, à savoir les modifications post-traductionnelles. C'est ainsi que l'on considère actuellement que pour un génome donné, il existe pléthore de protéomes correspondants, définis par un moment  $t$  donné, dans un état physiologique donné. D'autre part, il est assez clair maintenant, que la présence d'un ARNm n'est pas nécessairement synonyme de la présence de la protéine pour laquelle il code [Rogers et al., 2008]. Cela a conduit les scientifiques à s'intéresser plus particulièrement aux protéines, par l'intermédiaire des approches de protéomique, car ce sont elles, les actrices des fonctions biologiques.

Bien que l'étude du protéome ait commencé dans le milieu des années 70, avec la technique d'électrophorèse bidimensionnelle [O'Farrell, 1975] développée par O'Farrell en 1975, la protéomique a véritablement pris son envol dans le courant des années 90, avec l'amélioration importante des techniques de spectrométrie de masse et des techniques de

séparation. C'est en effet au cours de cette période que se sont vus décerner deux Prix Nobel, l'un pour Kiochi Tanaka et l'autre pour John Fenn, récompensant des travaux proposant des techniques d'ionisation douce permettant de faire de l'analyse de biomolécules par spectrométrie de masse (SM). Ces deux modes d'ionisation sont les deux principaux utilisés à l'heure actuelle pour les analyses de biomolécules. Nous reviendrons plus en détails sur ces questions au paragraphe 3.3.1.1 p. 58.

## 3.2 Le fractionnement protéique

Les fluides biologiques et les mélanges moléculaires complexes ne peuvent être analysés en une seule fois, et ne le sont même pas de manière complète en l'état actuel des connaissances et des techniques disponibles. En effet, des milliers de protéines sont exprimées à un instant  $t$  donné [Anderson and Anderson, 2002], chacune portant des modifications post-traductionnelles pouvant conduire à des fonctions biologiques complètement différentes. Pour diminuer la complexité du mélange, des techniques de séparation protéiques ou peptidiques ont été optimisées.

### 3.2.1 L'électrophorèse à une dimension

Sans doute la méthode de séparation la plus ancienne, il s'agit de réaliser une gel d'électrophorèse de type *sodium dodecyl sulfate polyacrylamide gel electrophoresis* (SDS-PAGE) afin de séparer les protéines en fonction de leur poids moléculaire. On obtient alors une trainée de protéines qu'il est possible de découper en plusieurs bandes qui seront analysées séparément. C'est une méthode de séparation rapide et simple à mettre en place. Chacune des bandes est ensuite digérée à la trypsine générant les peptides correspondant aux protéines présentes dans ces bandes. En numérotant les bandes, il est également possible d'obtenir une validation approximative de l'identification des protéines par la taille de la protéine identifiée [Jafari et al., 2012].

### 3.2.2 L'électrophorèse bidimensionnelle

L'électrophorèse bidimensionnelle est apparue au début des années 1970 et fait appelle à une double séparation des protéines [Kenrick and Margolis, 1970]. Tout d'abord, les protéines sont séparées selon leur point isoélectrique (pI) sur un gel d'acrylamide dans lequel est ménagé un gradient de pH allant généralement de 3 à 10. Un champ électrique est

appliqué et les protéines vont s'immobiliser dans la zone de pH où leur charge nette est nulle, c'est à dire où le pH est égal au point isoélectrique de l'analyte. C'est ce que l'on appelle l'isoélectrofocalisation. Dans une seconde étape, les protéines sont à nouveau séparées dans un gel d'acrylamide, cette fois selon le poids moléculaire grâce à une électrophorèse de type SDS-PAGE (Cf. figure 3.1 p. 54).

Dans une seconde étape, les protéines sont révélées dans le gel d'acrylamide grâce à des colorants spécifiques des protéines [Harris et al., 2007]. On peut mentionner ici les plus utilisés. Par exemple, malgré sa faible sensibilité (100 à 500ng de protéines par bande [Ren et al., 2012]), le bleu de Coomassie est totalement adapté aux études comparative car il s'agit d'un colorant homogène, de sorte que la coloration est proportionnelle à la quantité de protéine présente. D'autre part, ce colorant est tout à fait adapté à la spectrométrie de masse [Gauci et al., 2013]. Le nitrate d'argent est un autre colorant couramment utilisé. Ce dernier est le plus sensible (5ng de protéines par bande [Ren et al., 2012]) mais il présente un marquage peu linéaire car dépendant des propriétés physico-chimiques des protéines, et il est également difficilement compatible avec la spectrométrie de masse. Des techniques spécifiques permettent malgré tout son utilisation [Chevallet et al., 2006]. Un bon compromis entre les deux précédents existe et fait intervenir un marquage de type fluorescent avec notamment les différents SYPRO. Leur sensibilité est bonne (10 à 50ng de protéine par bande [Ren et al., 2012]), la coloration est linéaire et homogène, et totalement compatible avec la spectrométrie de masse.

Tableau 3.1 – **Comparaison des principaux colorants protéiques utilisés pour révéler les gels 2D.**

Colorant	Sensibilité	Linéarité
Bleu de Coomassie R250	100 à 500 ng	Bonne
Nitrate d'argent	5 ng	Dépend des propriétés physico-chimiques des protéines
Sypro Ruby	10 à 50 ng	Bonne

Bien que très utilisée, l'électrophorèse présente des limites importantes. Les protéines faiblement concentrées sont difficiles à étudier. D'autre part, les petites protéines inférieures à 10 kDa sont perdues, de même que les protéines ayant un pH extrême, inférieur à 3 ou supérieur à 10 [Gygi et al., 2000].

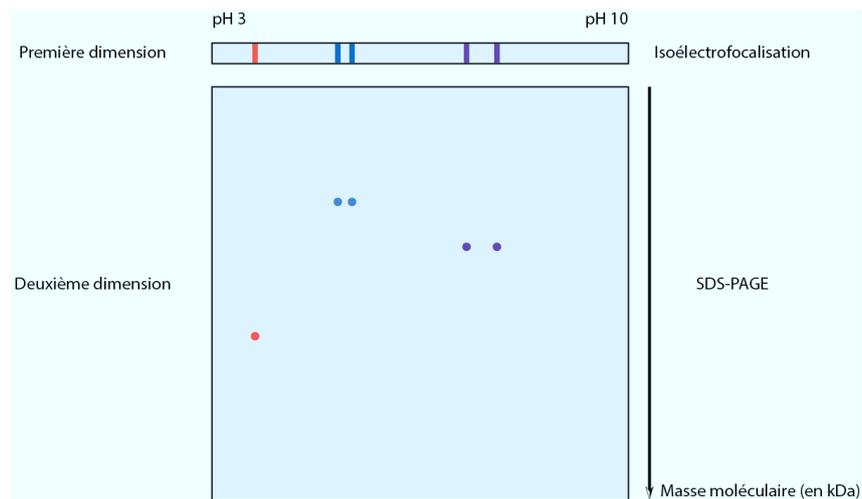


FIGURE 3.1 – **Principe général de l'électrophorèse bidimensionnelle** – Les protéines sont séparées dans une première étape en fonction de leur pI dans un gel d'acrylamide dans lequel est ménagé un gradient de pH allant généralement de 3 à 10. Dans une seconde étape, les protéines de ce gel sont séparées dans un gel SDS-PAGE en fonction de leur masse moléculaire.

### 3.2.3 La chromatographie

La chromatographie est une technique de séparation de molécules chimiques en fonction de leurs caractéristiques physico-chimiques. Elle met en jeu deux phases, l'une dite mobile (liquide ou gaz) et l'autre dite stationnaire (papier, gélatine, silice, polymère, silice greffée...). C'est dans les années 1970 que cette technique a connu un véritable essor avec l'apparition de la chromatographie liquide haute performance (CLHP) qui a permis de diminuer considérablement les temps de séparation en travaillant à des pressions élevées. Elle est alors devenue une technique de choix dans la séparation des mélanges peptidiques et/ou protéiques [Mitulovic and Mechtler, 2006].

Dans le cas particulier de l'étude des peptides, plusieurs types de séparation sont possibles. La plupart des études utilisent un mode de séparation par chromatographie liquide en phase inverse (RPLC) reposant sur l'interaction hydrophobe entre les peptides et la phase stationnaire constituée de silice greffée par des chaînes de 18 atomes de carbone. Les peptides sont déposés sur la colonne avec laquelle ils vont interagir puis ils sont élués à mesure que la concentration en solvant organique de la phase mobile augmente (cf figure 3.2 p.55).

Au début des années 1970, il est apparu l'idée de coupler plusieurs analyses chromatographiques afin d'arriver à une séparation des plus optimales [Edelman and Gall, 1971]. Cette idée a, par la suite, été reprise dans le cadre de la protéomique et c'est ainsi que sont nées les analyses multidimensionnelles couplant notamment des chromatographies

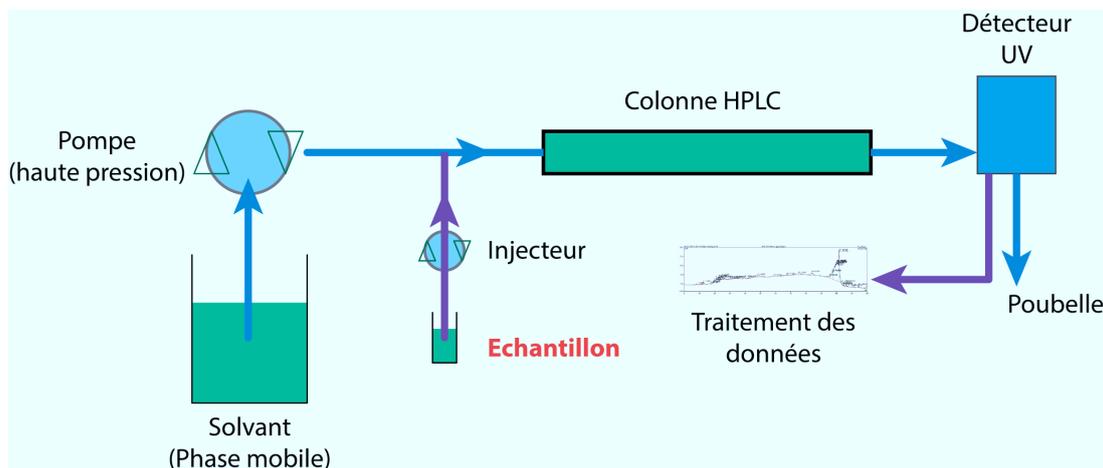


FIGURE 3.2 – Schéma simplifié du fonctionnement de la chromatographie liquide haute performance.

échangeuses de cations (SCX), une séparation sur gel SDS-PAGE à une dimension, ou encore la technique d'isoélectrofocalisation en phase liquide.

### 3.2.4 L'approche MudPIT

Apparue vers les années 2000, et développée par l'équipe de Yates, la technique MudPIT (*Multidimensional Protein Identification Technology*) est une technique d'analyse de protéomes globaux réalisée à la suite de séparation par chromatographie multidimensionnelle [Link et al., 1999]. La première étude à avoir utilisé cette technique a étudié le protéome global d'une levure (*Saccharomyces cerevisiae* BJ5460) [Washburn et al., 2001]. Le mélange complexe de peptides est acidifié et injecté sur une colonne SCX (Strong Cations Exchange) – une colonne échangeuse de cations – et à mesure de l'augmentation de la force ionique, les peptides sont directement envoyés sur une colonne en phase reverse sur laquelle ils seront à nouveau séparés (cf. figure 3.3 p.55), et analysés par spectrométrie de masse [Link et al., 1999]. Cette approche a donc permis de diminuer la complexité de l'échantillon de façon séquentielle, en augmentant parallèlement la couverture protéomique.

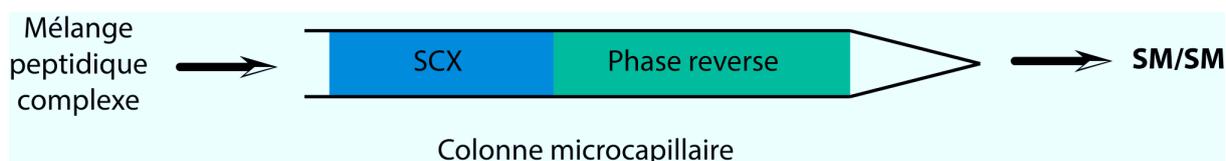


FIGURE 3.3 – Schéma simplifié représentant la stratégie MudPIT

### 3.2.5 La méthode OFFGEL

Il est possible de séparer les peptides, ou les protéines, en fonction de leur point isoélectrique (pI), et d'appliquer cette séparation à des études protéomiques globales. L'isoélectrofocalisation peut s'effectuer de différentes manières. D'une part, comme nous l'avons mentionné au paragraphe 3.2.2 p. 52, elle peut s'effectuer dans un gel d'acrylamide dans lequel est ménagé un gradient de pH. Bien qu'utilisée pour des analyses de mélanges complexes [Essader et al., 2005], cette approche demande un traitement assez contraignant de l'échantillon après l'isoélectrofocalisation, car il faut découper le gel en petites bandes desquelles seront extraits les peptides.

Une nouvelle approche a été développée en 2002 par Ros et ses collaborateurs. Cette méthode est basée sur une séparation de molécules amphotères, en fonction de leur point isoélectrique, dans un gradient de pH soumis à un champ électrique [Ros et al., 2002]. À la différence d'une isoélectrofocalisation classique, cette méthode de séparation permet de garder les analytes en solution dans un milieu liquide (Cf. figure 3.4 p. 57). Cela facilite grandement la récupération des échantillons après la séparation, et ces derniers peuvent alors directement être utilisés pour une deuxième étape de séparation lors d'une analyse multidimensionnelle. L'intérêt de cette technique dans les approches protéomique n'est plus à démontrer tant le nombre de laboratoires l'utilisant a considérablement augmenté ces dernières années, et tant les publications utilisant cette technique ont vu le jour [Ernoul et al., 2008] [Ernoul et al., 2009] [Besson et al., 2011] [Michel et al., 2006] [Keidel et al., 2011] [Krishnan et al., 2011]... Notre laboratoire a par exemple démontré la supériorité de cette approche par rapport à la technique de séparation classiquement utilisée jusqu'alors, la SCX [Ernoul et al., 2008].

D'autre part, cette séparation permet, dans le cadre d'une identification par spectrométrie de masse, d'introduire un paramètre de validation supplémentaire [Cargile et al., 2004]. En effet, si un peptide ayant un pI de 8 est identifié dans une fraction contenant théoriquement des analytes ayant un pI autour de 3, il est probable que l'identification de ce peptide soit aberrante.

La méthode permet d'utiliser des quantités importantes de matériel, pouvant aller jusqu'à 5 mg, ce qui permet d'envisager toutes sortes d'expériences complémentaires à cette séparation.

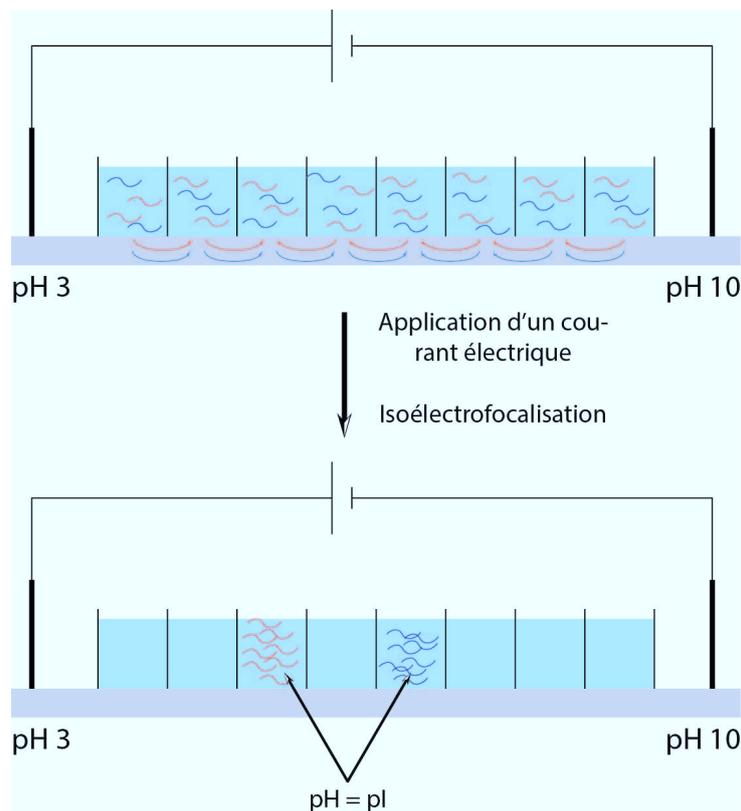


FIGURE 3.4 – **Schéma de fonctionnement du fractionnement OFFGEL** – Les protéines ou les peptides sont séparées en fonction de leur  $pI$  grâce à un gel d'acrylamide dans lequel est ménagé un gradient de pH allant généralement de 3 à 10. Les peptides isoélectrofocalisés restent en solution dans le tampon.

### 3.3 La spectrométrie de masse en protéomique

#### 3.3.1 Principe général

Un spectromètre de masse est un appareil capable de mesurer le rapport  $m/z$  d'une molécule chargée en phase gazeuse. Autrement dit, le rapport qu'il existe entre la masse d'une molécule et sa charge électrique. Quatre parties, schématisées en figure 3.5 p. 58, le constitue :

- La source d'ionisation qui permet de générer des ions chargés en phase gazeuse.
- L'analyseur qui permet de séparer les ions en fonctions de leur rapport  $m/z$ .
- Le détecteur qui permet de détecter et de compter les ions en leur associant leur rapport  $m/z$ .
- Un système informatique permettant l'analyse du signal généré.

Dans le contexte de la protéomique, les deux types d'ionisation utilisés sont principalement l'ionisation de type MALDI pour *Matrix Assisted Laser Desorption Ionisation* et

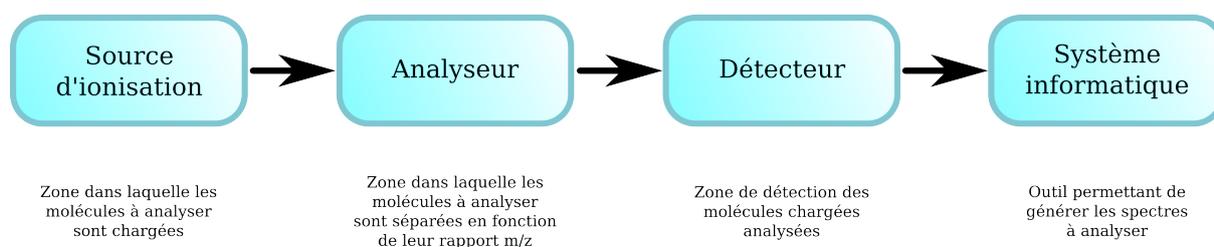


FIGURE 3.5 – **Représentation schématique de la structure générale d'un spectromètre de masse** – Un spectromètre de masse est constitué de quatre éléments principaux : la source d'ionisation, l'analyseur, le détecteur, et un système informatique permettant l'intégration des données

l'ionisation de type ESI pour *ElectroSpray Ionisation*. On doit l'ionisation de type MALDI à Kiochi Tanaka, un scientifique japonais, dont les travaux ont permis de mettre au point cette technique d'ionisation [Tanaka et al., 1988]. Ses travaux lui ont valu l'attribution d'un Prix Nobel de chimie en 2002. Mentionnons également ici les travaux réalisés par Karas et Hillenkamp qui ont mis au point l'application de cette technique, un an plus tard [Karas and Hillenkamp, 1988]. L'ionisation ESI a quant à elle été mise au point par John Fenn, un professeur américain en chimie analytique [Whitehouse et al., 1985]. Il s'est vu récompensé du Prix Nobel de chimie, la même année que Tanaka. Dans notre exposé, nous ne rentrerons pas dans le détail descriptif de ce mode d'ionisation, car nous n'avons pas utilisé cette technique dans le cadre de nos analyses et il ne nous paraît pas indispensable d'en faire état dans ces lignes. Comme nous l'avons mentionné précédemment, ces techniques permettent de réaliser des ionisations douces, permettant de garder les molécules intactes<sup>1</sup>.

### 3.3.1.1 L'ionisation MALDI

Dans le cadre de ce travail, seule la technique MALDI-TOF/TOF sera exposée car c'est cette approche qui a été utilisée dans notre étude. Il existe cependant d'autres modes d'analyse par spectrométrie de masse dont il ne sera pas fait état ici.

L'ionisation MALDI passe par la co-cristallisation des molécules à analyser avec une matrice présente en large excès. Le co-cristal ainsi formé peut être irradié avec un laser présentant une énergie pouvant être absorbée par la matrice, et non par les molécules à analyser (cf. Figure 3.6 p. 59). Il existe plusieurs types de matrices utilisables, mais elles ont toutes la particularité d'être de petites molécules organiques acides ayant un spectre

1. À noter qu'ils ont partagé ce Prix Nobel de chimie 2002 avec une autre personnalité importante du domaine, à savoir le physicien, mathématicien et chimiste Kurt Wüthrich qui est à l'origine de la résonance magnétique nucléaire [Wüthrich and Shulman, 1969].

d'absorption proche de celui du laser utilisé par le spectromètre. La désorption de la matrice conduit au transfert d'un proton de la matrice aux molécules à analyser qui se retrouvent alors chargées et qui peuvent être accélérées dans un champ électrique [El-aneid et al., 2009]. Les ions générés sont généralement monochargés, mais le processus qui permet d'obtenir ces ions est assez peu connu [Hillenkamp and Karas, 2007]. Deux modèles existent pour rendre compte du phénomène d'ionisation. Le modèle le plus ancien propose que la matrice soit dans un premier temps ionisée sous l'effet des photons du laser, et il s'en suit un transfert de charge de la matrice vers l'analyte dans le nuage de désorption lors de collisions. Le second modèle, plus récent, avance que les charges sont transférées à l'intérieur du cristal formé par les analytes et la matrice [Karas et al., 2000]. Ce débat n'est à l'heure actuelle toujours pas clos.

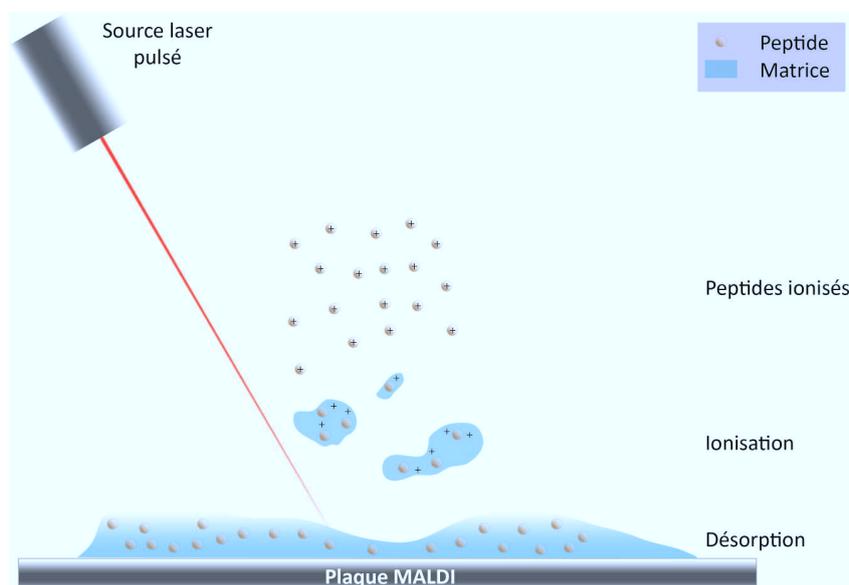


FIGURE 3.6 – **Représentation schématique du mode d'ionisation MALDI** – Le laser frappe la matrice qui absorbe l'énergie et désorbe de la plaque MALDI. Se faisant, elle transfère une charge aux molécules à analyser qui sont alors accélérables dans un champ électrique.

En protéomique, trois matrices sont principalement utilisées :

- le CHCA ( *$\alpha$ -cyano-4-hydroxycinnamic acid*), une molécule de 189 Da, généralement utilisée pour étudier les peptides et les protéines [Beavis et al., 1992] ;
- l'acide sinapinic (*3,5-dimethoxy-4-hydroxy cinnamic acid*), une molécule de 224 Da utilisée pour l'analyse des peptides ou des protéines, mais aussi pour les glycoprotéines [Beavis et al., 1989] ;
- le DHB (*2,5-dihydroxybenzoic acid*), une molécule de 154 Da pouvant être utilisée pour ioniser des peptides, des protéines, des glycoprotéines, des polymères synthétiques,

des lipides ou des hydrates de carbones [Strupat et al., 1991].

Le choix de la matrice en fonction de l'analyte à étudier est primordial car lorsqu'une molécule n'est pas ionisée, elle n'est pas du tout détectée par le spectromètre de masse [Aebersold et al., 2003a]. Pour autant, le mode d'ionisation MALDI permet, comme précédemment indiqué, d'analyser de l'ADN [Gut, 2004], des lipides [Balazy, 2004] ou encore des molécules glycoconjuguées [Harvey, 2003]. La matrice CHCA a un spectre d'absorption dans l'ultra violet situé entre 337 nm et 353 nm. Le mode d'ionisation MALDI associé à l'utilisation d'une matrice de CHCA constitue la méthode de choix pour l'étude des peptides [Mann et al., 2001].

### 3.3.1.2 Les analyseurs de masse

Concernant l'analyseur, il en existe de plusieurs types. Ce peut être un temps de vol (TOF : Time Of Flight), un quadrupôle, une orbitrappe, un analyseur à résonance cyclotronique d'ions, ou encore une trappe ionique. Ces différents analyseurs se distinguent par leur sensibilité, leur résolution et leur précision. Ces paramètres sont d'une importance cruciale car ce sont eux qui vont conduire à une identification la plus fiable possible.

Tableau 3.2 – Comparaison des principaux analyseurs de masse

Analyseur de masse	Limite de masse	Résolution ( $m/z = 1\ 000$ )
Quadrupole	8 000	2 000
Trappe ionique	6 000	5 000
TOF	En théorie illimité : 100 000	25 000 en mode réflectron
Analyseur à résonance cyclotronique d'ions (FT-ICR)	4 000	>100 000
Orbitrappe	4 000	100 000

L'ionisation de type MALDI est généralement associée à un analyseur en temps de vol, ce dernier étant particulièrement adapté à la nature pulsée de la désorption. Cette analyseur est constitué d'un tube dans lequel est ménagé un vide poussé, et possédant la source d'ions à une extrémité et le détecteur à l'autre extrémité (cf figure 3.7 p. 61). Nous devons le principe général de cet analyseur au physicien Stephens [Stephens, 1946].

D'après le théorème de l'énergie cinétique :

$$E_c = \frac{mv^2}{2}$$

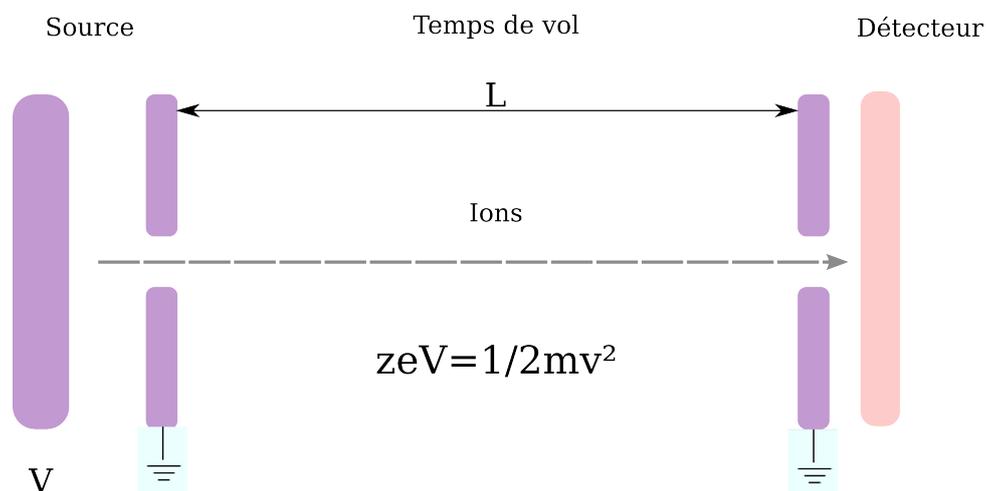


FIGURE 3.7 – **Représentation schématique d'un analyseur en temps de vol – TOF** – Les ions circulent dans un tubel dans lequel un vide poussé est ménagé de sorte que la masse des ions soit proportionnelle au temps qu'ils vont mettre à venir frapper le détecteur.

or

$$Ec = zeV$$

donc

$$Ec = zeV = \frac{mv^2}{2}$$

d'où

$$\frac{m}{z} = \frac{2.e.V}{v^2}$$

de plus,

$$v = \frac{L}{t}$$

donc

$$\frac{m}{z} = \frac{2.e.V.t^2}{L^2}$$

$z$  étant la charge de l'ion,  $e$  la charge élémentaire ( $\approx 1,6.10^{-19}$  C),  $V$  étant la tension d'accélération,  $m$  étant la masse de l'ion,  $v$  étant la vitesse,  $t$  le temps et  $L$  la longueur du tube.

D'après ces lois physiques, valables dans le vide, on remarque que la vitesse des ions dans le tube est directement proportionnelle à leur masse. Ainsi, connaissant la longueur du tube  $L$ , le temps  $t$  mis par un ion pour percuter le détecteur suffit à retrouver la masse de l'ion. Ainsi plus un ion a une masse élevée, plus il va mettre du temps à arriver au détecteur.

Le temps de vol est un analyseur qui présente un nombre important d'avantages. Tout d'abord, d'un point de vue théorique, la gamme de masse analysable est infinie, même si, dans la pratique, il est nécessaire de réussir à ioniser les molécules d'intérêt. Il est idéal pour les modes d'ionisation pulsée ce qui en fait un bon candidat à associer avec une source de type MALDI. Il permet d'obtenir un spectre complet pour chaque événement d'ionisation. La transmission des ions produits dans la source est très élevée de sorte que peu d'informations sont perdues [Medzihradszky et al., 2000].

Une approche permet d'augmenter la possibilité de séparer les signaux dont les masses sont proches et donc la résolution de l'appareil (cf figure 3.8 p.62). C'est en 1973 que le russe Boris Aleksandrovich Mamyryn a mis au point une technique permettant d'affiner les propriétés de l'analyseur TOF [Mamyryn and Shmikk, 1979]. Il a proposé une technique permettant à des ions de même rapport  $m/z$  mais possédant des énergies cinétiques différentes, d'arriver au même instant au détecteur, à condition que ceux-ci soient placés au plan de « focalisation en temps » en utilisant un miroir électrostatique générant un champ électrique de direction opposée à celui généré par la source. Ce mode d'utilisation du TOF est appelé mode réflectron, par opposition au mode précédemment décrit, qui constituait le mode linéaire.

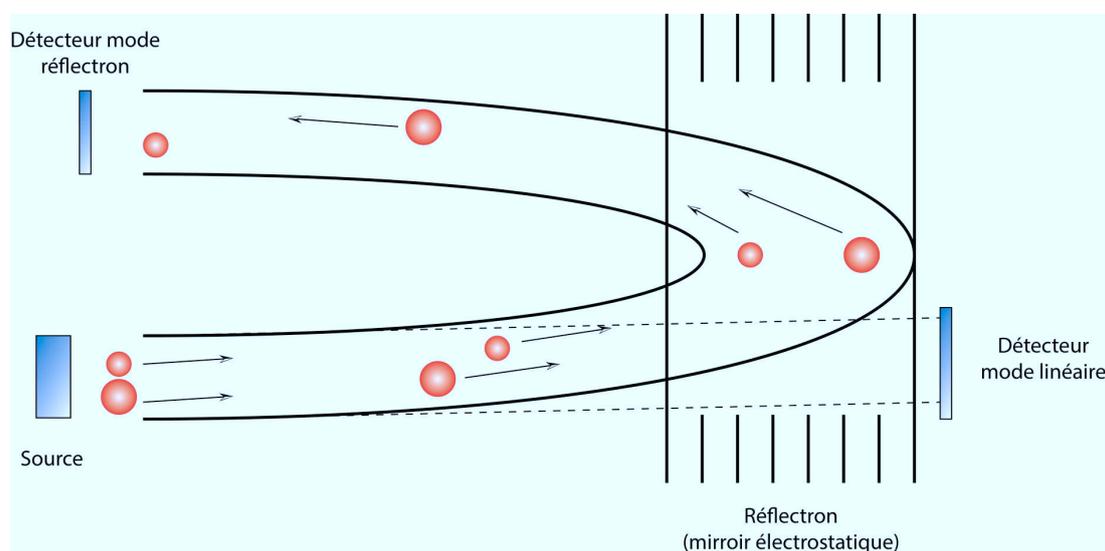


FIGURE 3.8 – **Représentation schématique du fonctionnement du mode réflectron** – Les ions issus de la source voient leur trajectoire modifiée grâce à un champ électromagnétique inversé de sorte que leur temps de parcours dans le vide est augmenté et la séparation des ions de masses proches se fait de façon plus nette. L'appareil gagne donc en sensibilité grâce à ce mode.

### 3.3.1.3 Spectrométrie de masse en tandem

Il est également possible de coupler deux analyseurs de masse séparés l'un de l'autre par une chambre de collision, permettant la fragmentation des molécules à analyser (libération d'ions fils) [de Hoffmann, 1996]. Cette chambre de collision contient soit des molécules de gaz inerte soit des molécules d'air (ne pouvant pas réagir chimiquement avec les ions parents). Les molécules à analyser vont entrer en contact avec les différents gaz présents dans la chambre et se casser (cf figure 3.9 p.63). Dans le cadre de la protéomique, cette fragmentation a lieu sur les peptides constituant les ions parents. Cette fragmentation libère des ions fils qui seront séparés dans le deuxième temps de vol afin de déterminer leur masse et d'en déduire la séquence en acides aminés du peptide. En effet, la zone de fragilité de ces molécules se situe au niveau de la liaison peptidique associant deux acides aminés [Roepstorff, 1984]. Le TIS (*Time Ion Selector*) est situé à l'entrée de la chambre de collision, et son rôle consiste à ne laisser entrer dans la chambre de collision que l'ion parent à analyser.

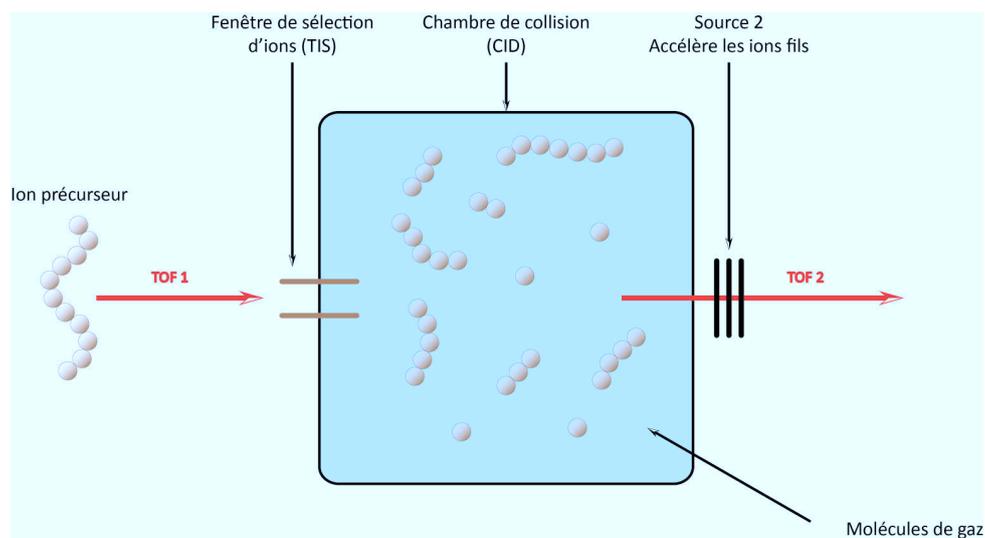


FIGURE 3.9 – Représentation schématique du fonctionnement du mode tandem – Les ions parents qui entrent dans la chambre de collision vont entrer en contact avec les molécules de gaz présentes dans cette dernière et se casser pour générer des ions fils. Ces ions peuvent ensuite être analysés dans un deuxième temps de vol.

### 3.3.1.4 Le détecteur

Les ions générés par la source, et séparés dans l'analyseur, doivent être détectés et transformés en un signal électrique utilisable par un système informatique. Il existe différents types de détecteurs, basés sur des principes physiques pouvant être différents,

mais leur rôle final reste le même, compter les ions selon le principe que plus le nombre d'ions est élevé plus le courant est important. Il doit donc permettre de mesurer un courant électrique très faible ( $10^{-13}$  à  $10^{-17}$  A) après une amplification basée sur le processus d'amplification d'ions.

### 3.3.2 Deux grands types d'approche

En protéomique, l'identification des protéines se fait à partir des peptides qui la constituent. Dans la plupart des cas, la protéase utilisée pour obtenir ces peptides est la trypsine car elle possède une grande spécificité. En effet, la trypsine hydrolyse de façon efficace et spécifique les protéines après les acides aminés basiques que sont la lysine et l'arginine. D'autres protéases sont utilisables (notamment la chymotrypsine), comme dans cette étude de Fischer dans laquelle ils testent les meilleures conditions de digestion pour analyser un grand nombre de protéines membranaires [Fischer and Poetsch, 2006] – ou encore la Glu-C qui est souvent utilisée dans le cadre d'études du glycoprotéome [Zielinska et al., 2012] [Kolarich et al., 2012].

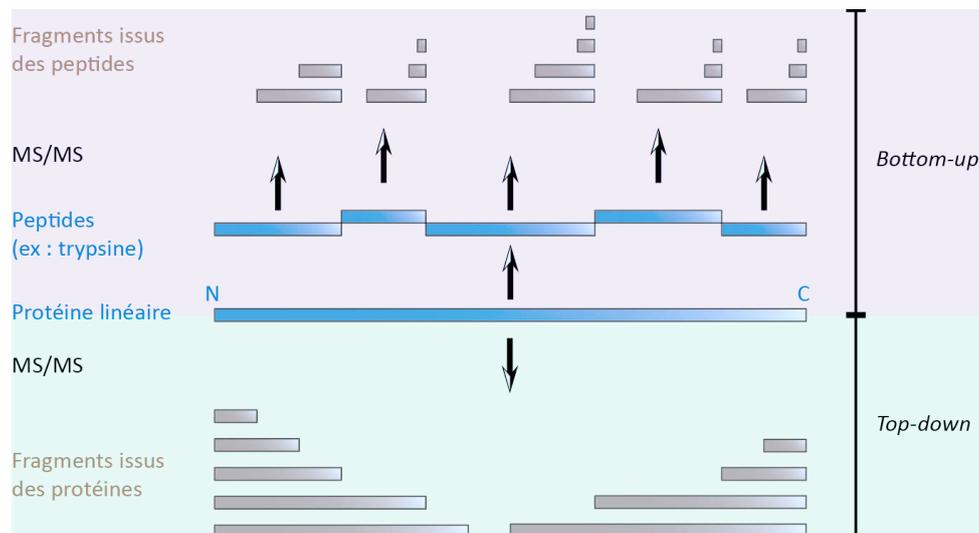


FIGURE 3.10 – Les différentes approches *top-down* et *bottom-up* en protéomique – Dans les analyses *bottom-up*, les protéines sont protéolysées directement après leur extraction de façon à obtenir des peptides qui pourront être séparés et analysés ultérieurement (exemple : les analyses multidimensionnelles). Dans les analyses de type *top-down*, les protéines sont séparées dans leur état initial, généralement sur des gels SDS-PAGE avant d'être récupérées et identifiées.

Il existe deux grands types d'approches, différenciés en fonction du moment auquel a lieu la digestion trypsique. En effet, les analyses protéomiques nécessitent de fractionner les échantillons que l'on souhaite traiter afin d'améliorer le nombre de protéines identifiées et de permettre une identification des protéines faiblement concentrées. Pourtant, il est

possible de choisir de fractionner le mélange de protéines issues de l'échantillon d'intérêt – dans ce cas on parle d'une approche « *Top-Down* » – ou on peut choisir de travailler sur un mélange encore plus complexe, et de fractionner l'échantillon peptidique obtenu après la digestion préalable de l'échantillon – on parle alors d'approche « *Bottom-up* » – [Han et al., 2008] (cf. figure 3.10 p.64). Chacune de ces approches présente des avantages et des inconvénients. L'approche « *Top-Down* » permet notamment d'avoir à traiter un échantillon moins complexe ce qui peut parfois simplifier les étapes de fractionnement. Cependant, en digérant les protéines des échantillons rapidement après l'extraction, on obtient un mélange extrêmement complexe de peptides, qu'il faudra ensuite être capable de séparer selon des critères physico-chimiques. L'intérêt de cette deuxième approche réside dans le fait que, lorsqu'une protéine est réduite à l'état de peptides, sa stabilité est nettement augmentée. Cela simplifie alors les traitements ultérieurs de l'échantillon.

### 3.4 Protéomique quantitative

Bien qu'il soit intéressant de pouvoir réaliser l'identification des protéines présentes dans un échantillon à un instant  $t$  donné, il est encore plus informatif de savoir comment varie le niveau d'expression d'une protéine donnée entre au moins deux conditions. Par nature, la spectrométrie de masse ne permet pas, d'emblée, de faire de la quantification, mais il a été développé un nombre important d'approches permettant de mesurer, souvent de façon relative, les variations de niveau d'expression des protéines dans des échantillons [Aebersold et al., 2003b]. En spectrométrie de masse, l'intensité d'un pic correspondant à un peptide sur le spectre n'est pas le reflet de son abondance relative, par rapport aux autres pics. En effet, tous les peptides d'un échantillon n'ont pas les mêmes propriétés physico-chimiques de sorte que leur capacité à être ionisé varie en fonction de leur nature. Un pic de forte intensité peut donc représenter un peptide très faiblement concentré mais ayant une forte capacité à donner ou à recevoir des charges. D'autre part, le nombre de protons transférables depuis la matrice vers les analytes est limité, de sorte que seuls les peptides de forte affinité protonique et de concentration élevée seront détectés, entraînant un effet de suppression d'ions [Karas et al., 2000]. Il a donc fallu développer des stratégies d'analyse permettant de quantifier les protéines présentes dans les échantillons biologiques.

Nous pouvons considérer qu'il existe quatre grands types de méthode de quantification. Il y a d'une part les techniques basées sur l'utilisation de gels. D'autre part, il y a les techniques utilisant les marquages isotopiques. Il existe également des approches sans

marquage [Aebersold et al., 2003b]. Et enfin, des approches récentes permettent de faire de la quantification absolue par spectrométrie de masse [Nanavati et al., 2008].

### 3.4.1 La quantification basée sur gels

#### 3.4.1.1 Comparaison de gels 2D

Par cette approche, les protéines de chacune des conditions analysées sont séparées selon une électrophorèse bidimensionnelle (séparation en fonction du point isoelectrique suivie d'une séparation en fonction du poids moléculaire : cf figure 3.1 p.54) puis les gels sont colorés à l'aide d'un des colorants décrits dans le paragraphe 3.2.2 p.52. Ils sont ensuite scannés, superposés et alignés pour pouvoir être analysés grâce à des logiciels dédiés tels que PDQuest ou ImageMaster, qui vont comparer l'intensité de coloration des spots et leur taille. Ces variations correspondent à des variations de quantité de la protéine marquée. Il est alors possible d'établir un ratio d'une même protéine entre différentes conditions. La digestion de la protéine à l'intérieur du gel permet ensuite de réaliser une identification de celle-ci. Cette approche « *Top-Down* » présente l'avantage de visualiser les différents états d'une même protéine en fonction de ses modifications post-traductionnelles. Elle présente également quelques inconvénients majeurs avec notamment un besoin en réplicats techniques important puisqu'on estime qu'il faut trois réplicats par condition pour permettre de conclure de façon statistiquement significative quant aux variations de niveau d'expression des protéines étudiées.

#### 3.4.1.2 2D-DIGE

Pour palier aux inconvénients liés aux réplicats techniques de la comparaison de gels 2D, il a été mis au point en 1997 une technique assez proche mais utilisant des fluorochromes, les cyanines [M et al., 1997]. Cette technique appelée DIGE (cf figure 3.11 p.67) pour *Difference In Gel Electrophoresis* permet de limiter le nombre de gels utilisés pour réaliser les analyses. Les protéines des deux échantillons à comparer sont marquées sur les lysines de façon covalente avec les cyanines 3 et 5. Un mélange 1 : 1 des deux conditions (ou plus) est également marqué avec la cyanine 2 et va servir de standard interne pour normaliser les variations d'intensité de fluorescence. Les différentes cyanines ont des caractéristiques physico-chimiques identiques de sorte que la migration des protéines marquées se fera de la même façon quelque soit la cyanine fixée. Cependant, les deux échantillons et le standard interne sont déposés sur le même gel, et les cyanines ont des longueurs d'ondes d'excitation

différentes. Il va alors être possible d'obtenir trois images parfaitement superposables et les niveaux d'intensité de fluorescence vont être fonction de la quantité de protéine présente et marquée. L'abondance de chaque protéine peut être exprimée comparativement au standard interne [Marouga et al., 2005]. Grâce à cette technique, on s'affranchit de la reproductibilité intra expérimentation, et la présence du standard interne permet d'aligner et de normaliser les gels d'une même expérience dans le cas de l'utilisation de plusieurs gels.

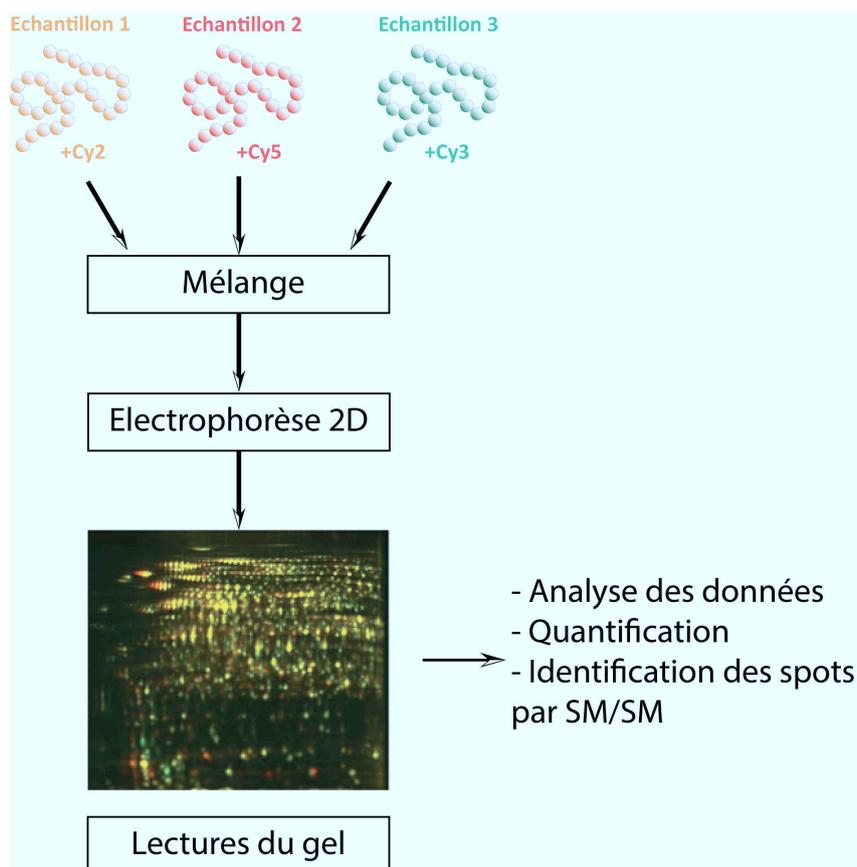


FIGURE 3.11 – **Principe générale de la méthode DIGE** – Les protéines des deux échantillons à comparer sont marquées avec les cyanines 3 et 5. Un mélange 1 : 1 des deux conditions avec la cyanine 2 va servir de standard interne pour normaliser les variations d'intensité de fluorescence. La lecture des gels avec un lecteur à fluorescence permet ensuite de déterminer les variations du niveau d'expression des différentes protéines.

### 3.4.2 La quantification basée sur marquage isotopique

#### 3.4.2.1 Principe général

D'une manière générale, les peptides marqués avec des isotopes ont, pour chaque condition étudiée, le même comportement physico-chimique. En effet, l'incorporation d'un

isotope stable d'un atome au sein du molécule ne modifie pas ses propriétés intrinsèques. Cela permet, une fois le marquage réalisé, de séparer les molécules marquées différemment de façon similaire. Autrement dit, ces molécules différentes par leur composition isotopique auront le même temps de rétention en HPLC, et auront le même point isoélectrique. En revanche, les isotopes ont des masses différentes qui peuvent être détectées spécifiquement par le spectromètre de masse. Il est ainsi possible de comparer l'aire sous la courbe des pics correspondant aux isotopes et de déterminer une information quantitative relative. Selon le type de marquage, les isotopes peuvent être introduits au niveau protéique ou peptidique dans la séquence en acides aminés par une réaction métabolique ou enzymatique, ou par fixation chimique [Aebersold et al., 2003b].

La réaction de couplage du réactif de quantification peut avoir lieu à différents niveaux du processus de traitement des échantillons en fonction des approches (cf figure 3.12 p.68).

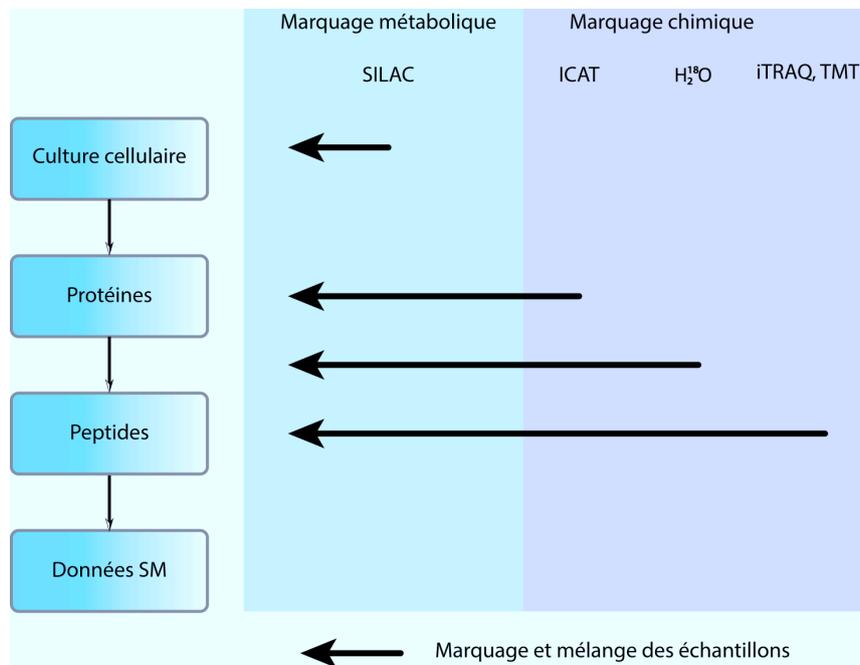


FIGURE 3.12 – Niveau de marquage selon les approches utilisées – Le SILAC permet un marquage au niveau de la culture cellulaire, l'ICAT permet un marquage au niveau protéique, le marquage à l'eau lourde permet un marquage au moment de la digestion protéolytique, et les marquages iTRAQ<sup>TM</sup> et TMT permettent de marquer les peptides.

### 3.4.2.2 Le marquage SILAC

Apparu au début des années 2000, le marquage SILAC (*Stable Isotope Labelling by Amino acids in Cell culture*) est un marquage métabolique permettant de comparer des cellules cultivables (cf figure 3.13 p.69). Il a été popularisé par l'équipe de Matthias Mann,

avec notamment une étude parue en 2002 dans la revue *Molecular and cellular proteomics* qui justifie l'utilisation de ce marquage en protéomique pour les études *in vitro* [Ong et al., 2002]. Le principe consiste à remplacer un acide aminé présent dans le milieu de culture par son équivalent portant une modification isotopique dans l'une des deux conditions de culture. À mesure de la croissance cellulaire, les cellules vont utiliser les acides aminés modifiés (généralement  $^{13}\text{C}$ ,  $^{15}\text{N}$ -lysine et  $^{13}\text{C}$ ,  $^{15}\text{N}$ -arginine) pour synthétiser leurs protéines et vont les incorporer pour finalement n'exprimer plus que des protéines avec les acides aminés modifiés. L'intérêt d'utiliser la lysine ou l'arginine réside dans le fait que la trypsine hydrolyse après chacun ces acides aminés de sorte que la totalité des peptides générés par la digestion possèdera au moins un acide aminé modifié. Cette méthode conduit à des peptides tryptiques marqués sur un seul acide aminé (sauf en cas de clivage manqué par la trypsine : *miscleavage*), ce qui permet une comparaison directe.

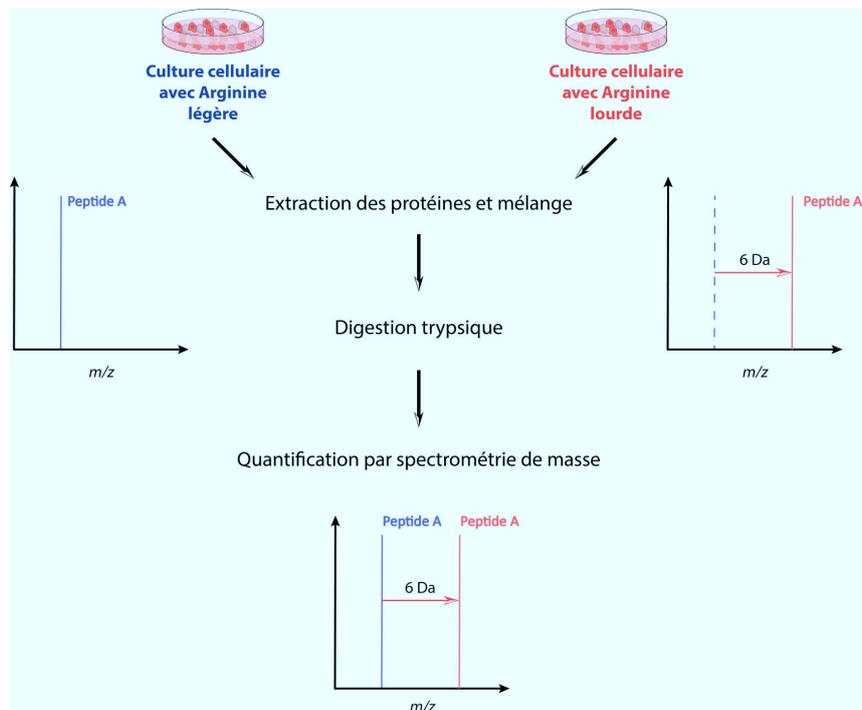


FIGURE 3.13 – **Principe générale de la méthode SILAC** – Les cellules sont mises en culture en présence des différents acides aminés modifiés ou non, puis les protéines sont extraites, mélangées et digérées avant d'être identifiées et quantifiées par spectrométrie de masse.

Il existe cependant une donnée à laquelle il faut être particulièrement attentif. En effet, durant les processus métaboliques mis en jeu par les cellules, l'arginine peut être convertie en proline ce qui conduit à complexifier l'analyse finale et surtout la quantification. Il est possible de limiter cet effet en diminuant le taux d'arginine dans le milieu jusqu'à une concentration suffisamment faible pour rendre la réaction de conversion moins favorable [Bla-

goev and Mann, 2006]. Malgré tout, bien que cela permette de diminuer la réaction de conversion, il n'est pas possible de l'empêcher complètement [Hwang et al., 2006]. Plus récemment, l'équipe de Lajoie a montré qu'il était possible d'améliorer la quantification en adaptant la quantité d'arginine disponible dans le milieu, et en le supplémentant en proline [Bendall et al., 2008].

Il existe également des différences entre les cellules en terme d'incorporation des différents acides aminés. Par exemple, il est très facile d'utiliser la méthode SILAC sur des lignées de mammifères car elles ne sont pas capables de synthétiser tous les acides aminés. En revanche, les cellules végétales étant autotrophes, elles n'ont pas besoin de puiser leurs acides aminés dans leur environnement, donc il est difficile d'utiliser le SILAC. De plus, certaines cellules ayant des difficultés de croissance ne pourront pas être utilisées car elles ne supporteront pas les carences en facteurs de croissance nécessaires à la mise en œuvre de cette approche [Elliott et al., 2009].

### 3.4.2.3 Le marquage à l' $^{18}\text{O}$

Ce type de marquage repose sur la réaction d'hydrolyse médiée par la trypsine lors de son action. En effet, elle incorpore deux atomes d'oxygène dans le peptide hydrolysé [Yao, ]. Fort de ce constat, il a suffi d'imaginer utiliser de l'eau modifiée pour réaliser la réaction d'hydrolyse, ce qui permet d'introduire ces  $^{18}\text{O}$  en C-terminal des peptides [Mirgorodskaya et al., 2000]. Ainsi, il existe une variation de quatre daltons entre le peptide marqué et le peptide non marqué (cf figure 3.14 p.71). D'autre part, cette approche permet de faciliter l'identification car les ions sont marqués en C-terminal ce qui permet d'attribuer facilement l'ion « y ». D'autre part, afin d'éviter les contaminations par de l' $^{16}\text{O}$  les échantillons doivent être complètement déshydratés avant le marquage.

Avec cette approche, les deux échantillons à comparer sont donc digérés séparément, ce qui peut affecter la précision de la quantification de même que la reproductibilité expérimentale. De plus, tous les peptides n'incorporent pas le marquage de la même façon à cause de leurs propriétés physico-chimiques. Par exemple, il semble que le marquage soit plus lent et donc souvent incomplet pour les gros peptides acides [Gevaert et al., 2008].

### 3.4.2.4 Le marquage ICAT

Le marquage ICAT (*Isotope Coded Affinity Tag*) est le premier réactif permettant de réaliser un marquage chimique sur des peptides à des fins de quantification qui ait été

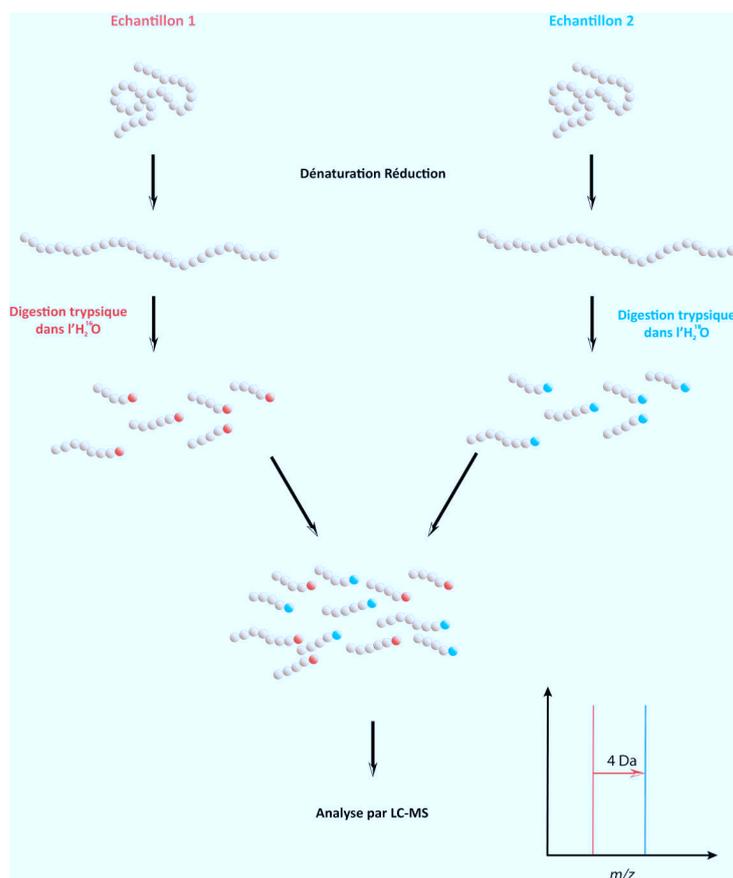


FIGURE 3.14 – **Principe générale de la méthode de marquage enzymatique avec utilisation de l'eau lourde** – Les protéines issues des échantillons sont dénaturées et réduites, puis elles sont digérées par des protéases soit en condition normale soit en présence d'eau lourde. Les échantillons sont ensuite mélangés puis les protéines sont identifiées et quantifiées par spectrométrie de masse.

développé. C'est ainsi, qu'en 1999, Gygi et Aebersold ont publié un article dans la revue *Nature Biotechnology* dans lequel ils exposent le principe sur lequel repose ce mode de quantification relative. Le réactif original comprend trois principaux groupements (cf figure 3.15 p.72) ; d'une part, un groupement ayant la possibilité de réagir sur les groupements thiols (iodoacétamide) des cystéines, un groupement biotine permettant de réaliser un couplage avec l'avidine, et un groupement marqueur contenant soit 0 soit 8 atomes de deutérium. Ainsi, les deux formes du réactif présentent une différence de masse de 8 daltons. En 2003, une nouvelle génération de réactif ICAT est apparue, possédant cette fois 9 <sup>13</sup>C au lieu de 8 atomes de deutérium [Hansen et al., 2003]. Il avait été remarqué que l'utilisation du deutérium modifiait le temps de rétention des peptides sur les colonnes en phase reverse, de sorte que les peptides marqués avec la forme « lourde » du réactif soient élués légèrement plus tôt que ceux marqués avec la forme légère [Zhang et al., 2002b]. L'utilisation du <sup>13</sup>C permet de limiter cet effet.

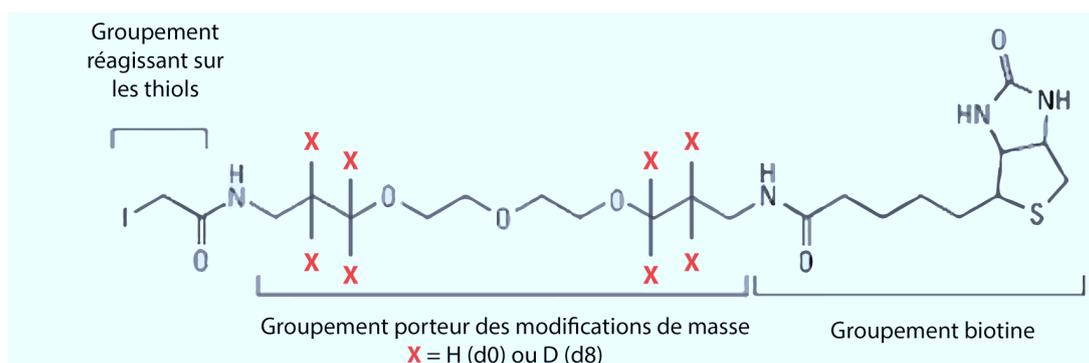


FIGURE 3.15 – **Structure chimique du réactif chimique ICAT de première génération** – Trois éléments principaux constituent le réactif ICAT, un groupement chimique capable de réagir sur les groupements thiols des protéines, un groupement porteur des modifications isotopiques permettant la discrimination de la contribution de chacun des échantillons, et un groupement biotine.

L'un des avantages de l'ICAT réside dans sa capacité à permettre une diminution de la complexité d'un échantillon car il se fixe sur les cystéines (cf figure 3.16 p.73), un acide aminé particulièrement rare dans les protéines puisque représentant environ 1,5% de tous les acides aminés [Gevaert et al., 2008]. Cependant, cet aspect comporte aussi des inconvénients majeurs, puisque la quantification des protéines va être effectuée sur un nombre limité de peptides par protéine. De plus, cela conduit à ce que des protéines ne contenant pas de cystéines ne puissent être pas analysées, or ces dernière représentent environ 10 à 13% des protéines [Vaughn et al., 2006].

### 3.4.2.5 Le marquage iTRAQ

Le réactif en lui-même est constitué de trois parties [Ross et al., 2004] (cf. figure 3.17 p.73 et figure 3.18 p.74) :

1. **Le groupement reporter (N-méthyl piperazine)** : il diffère entre les quatre ou huit formes du réactif ( $m/z = 114, 115, 116, 117$ ) ou ( $m/z = 113, 114, 115, 116, 117, 118, 119, 121$ ).
2. **Le groupement « balance »** : il permet de compenser les différences du groupement reporter afin que les peptides marqués aient tous la même masse d'étiquette ( $m/z = 145$  Da pour le 4-plex et 305 Da pour le 8-plex).
3. **Le groupement fonctionnel** : ce groupement N-hydroxy-succinimide (NHS) lie de manière covalente le réactif iTRAQ<sup>TM</sup> de masse 145 Da pour le 4-plex et 305 Da pour le 8-plex, aux amines primaires libres des peptides<sup>2</sup>.

2. Le NH<sub>2</sub> terminal ou les fonctions amines primaires des lysines

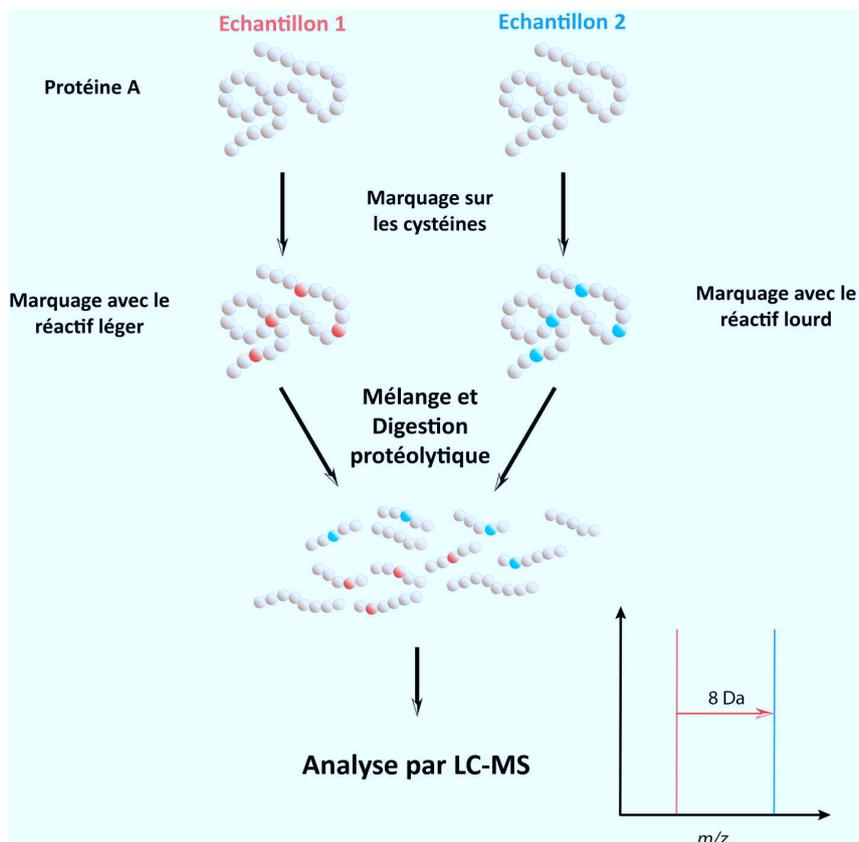


FIGURE 3.16 – Principe général de la méthode ICAT – Les protéines issues de l'échantillon sont marquées soit avec le réactif léger soit avec le réactif lourd, puis ces protéines sont mélangées et digérées avant l'analyse par spectrométrie de masse.

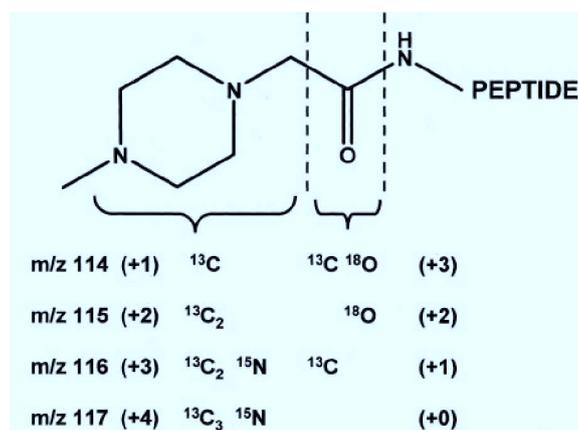


FIGURE 3.17 – Formule chimique du réactif iTRAQ<sup>TM</sup> dans sa version 4-plex – À gauche, le groupement reporter, au centre, le groupement balance, à droite, le groupement fonctionnel NHS [Ross et al., 2004]

Le réactif réagit avec les amines primaires situées en N-terminal des chaînes peptidiques, et avec celles des chaînes latérales des lysines. Il existe également une variante de ce réactif qui permet de marquer huit échantillons en même temps, ce qui offre la possibilité de comparer plus d'échantillons au cours de la même analyse.

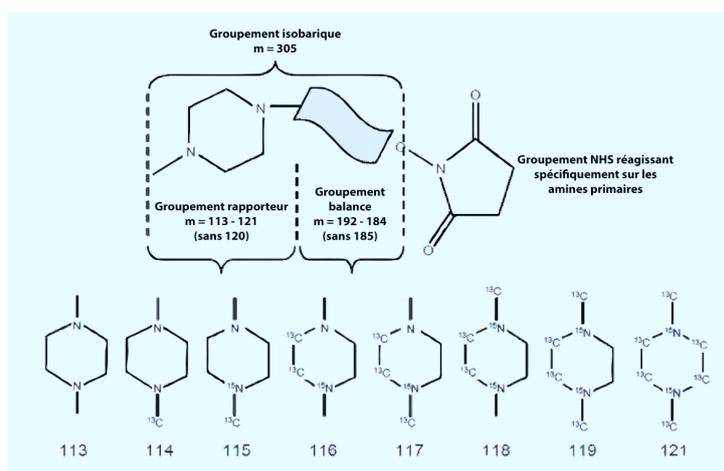


FIGURE 3.18 – Formule chimique du réactif iTRAQ<sup>TM</sup> dans sa version 8-plex – À gauche, le groupement reporter, au centre, le groupement balance, à droite, le groupement fonctionnel NHS

Ainsi, quatre ou huit échantillons peuvent être marqués simultanément, et mélangés selon un rapport 1 : 1 : 1 : 1 ou 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1. Puis les peptides peuvent être séparés selon les méthodes de séparation classique (OFFGEL, SCX, Nano-LC sur colonne de C18 en phase reverse).

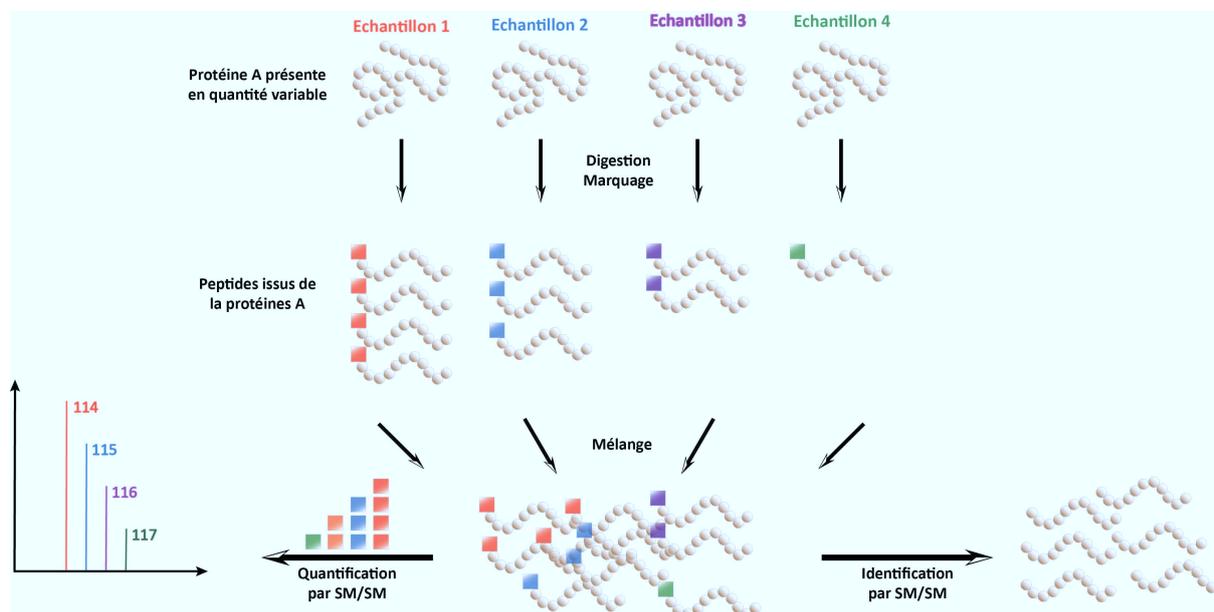


FIGURE 3.19 – Schéma du principe général de marquage iTRAQ<sup>TM</sup> dans sa version 4-plex – La protéine A présente dans les quatre échantillons en quantité variable est digérée et marquée avec le réactif iTRAQ<sup>TM</sup>. Les peptides des quatre échantillons sont ensuite fragmentés libérant ainsi les étiquettes. L'aire sous la courbe du pic des étiquettes représente la quantité relative du peptide présent dans l'échantillon correspondant à l'étiquette. Ces variations sont extrapolables pour permettre de définir la variation relative de la protéine dans les échantillons analysés.

Lors de l'analyse par spectrométrie de masse en tandem, les groupements balances sont clivés, de façon à libérer les étiquettes sous la forme d'ions rapporteurs. Comme le montre la figure 3.19 p. 74, dans le cas du 4-plex il y aura quatre pics (114, 115, 116, 117), et dans le cas du 8-plex, il y en aura huit (113, 114, 115, 116, 117, 118, 119, 121) [Choe et al., 2007]. L'aire sous la courbe de ces pics correspond à la quantité relative du peptide.

D'autre part, le spectre SM/SM permet de dégager l'information concernant la séquence du peptide ce qui permet de remonter à la protéine par l'intermédiaire de l'interrogation des bases de données. L'un des avantages du marquage iTRAQ<sup>TM</sup>, réside dans le fait qu'il permette de prendre en compte plusieurs peptides pour consolider la variation d'une même protéine. Ainsi, si deux peptides issus d'une même protéine ont des variations similaires cela rend la quantification plus robuste. Alors que si deux peptides issus d'une protéine varient de façon opposés, cela conduit à s'interroger sur la recevabilité des données. Cela peut arriver notamment lorsqu'une protéine est mal identifiée. Des logiciels, comme *Protein Pilot* ou encore *iQuantitator* [Schwacke et al., 2009a] ont été développés pour permettre de regrouper toutes les informations peptidiques relatives à une protéine pour ainsi globaliser les informations concernant la quantification. Cela permet également de réaliser des tests statistiques inférentiels<sup>3</sup> pour déterminer si la variation observée est due au hasard ou si elle est statistiquement significative.

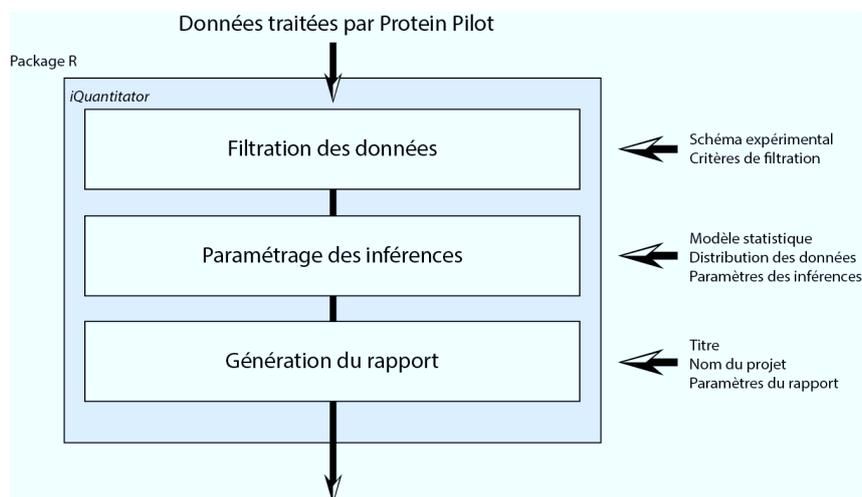


FIGURE 3.20 – **Schéma du principe général du fonctionnement du logiciel iQuantitator** – Les données issues du logiciel *Protein Pilot* sont traitées statistiquement de sorte que les variations significatives puissent être mise en lumière.

Quelques biais techniques sont à prendre en considération lors de l'utilisation du marquage iTRAQ<sup>TM</sup>. Par exemple, le réactif iTRAQ<sup>TM</sup> 8-plex a été pensé pour ne pas

3. Cela signifie qu'on extrapole la variation de la protéine en fonction de celle des peptides la constituant.

posséder d'étiquette ayant une masse de 120 Da. La raison réside dans le fait que les ions immoniums, issus de la phénylalanine des peptides, possèdent une masse de 120,08 Da. Ainsi, ces ions pourraient interférer avec la quantification des peptides marqués par cette étiquette [Evans et al., 2012]. Malgré ces précautions, il semble également qu'il faille interpréter avec prudence le marquage avec l'étiquette 121 car il peut y avoir une contamination croisée de ce pic avec un isotope de cet ion immonium de la phénylalanine ayant une masse de 121.08 [Ow et al., 2009]. Selon cette étude, cette contamination peut représenter une surestimation pouvant aller jusqu'à 40% de la référence sur des appareils de basse résolution. Cette approximation est moins vraie sur des appareils haute résolution car l'étiquette possède une masse de 121.12 Da, et l'ion immonium +1 une masse de 121.08 Da. Sur ce type d'appareil, ces masses pourront être discriminées.

D'autre part, le groupe de Ow a également montré qu'il existait une sous-estimation de l'amplitude réelle des variations d'intensité. Ils mettent en évidence que cela est dû au fait que les peptides avec les mêmes  $m/z$  sont co-élués, créant ainsi un mélange d'étiquettes dans les échantillons complexes. Des stratégies ont été développées afin de palier à cette sous-estimation et elles résident principalement dans l'amélioration de la séparation en première dimension. Par exemple, l'utilisation de chromatographie d'interaction hydrophile (HILIC) permet d'améliorer le nombre de protéines identifiées, mais permet également d'amoindrir le phénomène de sous-estimation [Ow et al., 2011]. Cependant, selon Evans [Evans et al., 2012], cette sous-estimation est surtout à prendre en compte dans le cadre du développement d'une approche méthodologique, car elle reste négligeable dans le cas de l'interprétation biologique des données. De plus, le logiciel commercial *Protein Pilot* propose un algorithme capable de modéliser le bruit de fond d'une expérience, et d'appliquer une « correction » aux données obtenues. Ils ont montré que l'application de cette correction sur leurs données aboutissait à des rapport de quantification supérieurs de 34%, compensant ainsi les 40% de sur-estimation démontré par Ow et son équipe, et ils expliquent ainsi qu'il est possible d'améliorer les sous-estimations par l'utilisation de traitements post-acquisitions. À noter tout même qu'il existe ainsi un risque de sur-estimation et d'amplification du bruit de fond.

D'autres équipes se sont intéressées à la reproductibilité des marquages, et donc leur précision. Elles ont ainsi démontré que les coefficients de variation étaient plus important pour les pics présentant une faible intensité [Karp et al., 2010] [Wang et al., 2011]. Les mesures iTRAQ<sup>TM</sup> issues de chaque peptide peuvent être combinées pour obtenir une idée de la variation de la protéine correspondante. Il existe plusieurs modes de calcul pour obtenir

cette information. Certaines études excluent de leur calcul les peptides identifiés à partir de spectres dont les pics correspondant aux étiquettes sont de trop faible intensité [Hu et al., 2006]. D'autres appliquent une pondération de l'importance d'un peptide par rapport à un autre, en fonction de l'intensité de ses ions rapporteurs [Gan et al., 2007]. Une autre étude a quant à elle utilisé un modèle de stabilisation de la variance par une transformation mathématique [Karp et al., 2010].

Apparu dans un second temps, le marquage 8-plex est nécessairement moins présent dans la littérature. Il semble cependant que son apparition tardive ne soit pas la seule explication. Trois études, complémentaires à la notre [Besson et al., 2011], ont en effet montré une supériorité du marquage 4-plex en terme d'identification par rapport au 8-plex [Pichler et al., 2010] [Thingholm et al., 2010] [Mahoney et al., 2011]. Dans cette dernière analyse les auteurs décrivent notamment que les expériences réalisées avec le réactif 8-plex aboutissent à une diminution d'environ 40% de protéines identifiées et quantifiées. L'équipe de Thingholm propose un modèle pouvant rendre compte de ces différences. Ils ont tout d'abord observé que la charge moyenne des peptides marqués était supérieure pour les peptides marqués avec le réactif 8-plex probablement à cause de la taille du groupement qui est supérieure. Cet état de charge réduirait l'intensité des peptides marqués conduisant à ce que les peptides les moins présents se retrouvent sous la limite du seuil de détection.

Il existe une difficulté importante dans le traitement des données qui consiste à déterminer à partir de quelle amplitude une variation peut avoir une signification biologique. Le logiciel *iQuantitator* nous permet de nous affranchir de cette question puisqu'il regroupe toutes les données relatives à la quantification d'une protéine, obtenues pour chaque peptide ayant permis l'identification, et il réalise un test statistique, définissant un intervalle de confiance. Lorsque les deux bornes de cet intervalle de confiance sont strictement supérieures à 1 la protéine est statistiquement sur-exprimée, lorsque ces bornes sont strictement inférieures à 1, la protéine est statistiquement sous-exprimée, et lorsque les bornes sont de part et d'autre de 1, la variation de la protéine n'est pas significative [Schwacke et al., 2009a]. Ce logiciel permet également de regrouper plusieurs expériences iTRAQ<sup>TM</sup> facilitant ainsi les analyses de grande ampleur et donc de plus forte significativité statistique.

Le marquage iTRAQ<sup>TM</sup> semble être une bonne méthode pour l'identification de nouveaux biomarqueurs selon Elliott et ses collaborateurs [Elliott et al., 2009]. De plus, cette technique permet de réaliser un multiplexage des échantillons aboutissant à une amélioration de la détection des protéines faiblement concentrées [Chaerkady and Pandey, 2007].

Plusieurs études ont utilisé ce marquage sur des échantillons tumoraux dans le but de mettre en évidence des protéines différentiellement exprimées par la tumeur. L'une des premières est celle publiée par Keshamouni et ses collègues au cours de laquelle ils identifient que les protéines sur-exprimées lors de la transition épithélio-mésenchymateuse médiée par le TGF $\beta$  sont majoritairement des protéines d'adhésion cellulaire, de migration, et impliquées dans les phénomènes de migration cellulaire [Keshamouni et al., 2006]. Depuis, l'association des mots-clés `iTRAQ`, `cancer`, et `biomarker` donne plus de cent dix résultats<sup>4</sup>. Cette technique est donc de plus en plus utilisée, aussi bien sur des tissus ou des fluides biologiques que sur des lignées. Ainsi, il a été utilisé pour mettre en évidence des biomarqueurs impliqués dans le rejet de greffe [Freue et al., 2010]. Il a été utilisé à de nombreuses reprises pour étudier des lignées de cancer, colorectal par exemple, pour mettre en évidence des biomarqueurs potentiels. Une étude publiée par Ghosh et ses collaborateurs a permis d'étudier les différences entre deux lignées, l'une étant le phénotype métastatique de l'autre (SW480 et SW620). Ils ont pu mettre en évidence 1 140 protéines dont 94 étaient sous-exprimées et 53 étaient sur-exprimées [Ghosh et al., 2011]. Nous pouvons également citer une autre étude récente au cours de laquelle l'équipe de Chen et ses collaborateurs a travaillé sur une comparaison entre le tissu sain et le tissu tumoral de 8 patients atteints de cancers colorectaux. Ils ont identifié 438 protéines parmi lesquelles 34 sur-exprimées et 8 sous-exprimées [Chen et al., 2010]. Mentionnons également ici que l'iTRAQ<sup>TM</sup> est utilisé dans plusieurs études dans le cas d'analyses de modifications post-traductionnelles telles que des phosphorylations [Hoffert et al., 2012], des méthylations, des acétylations ou encore des fucosylations [Rosenzweig et al., 2008].

Au laboratoire, il a également été montré que l'iTRAQ<sup>TM</sup> permettait d'améliorer la couverture protéomique lors de l'identification de protéines par approches globales [Ernault et al., 2008]. Il semble que la fixation de l'iTRAQ<sup>TM</sup> améliore l'ionisation des peptides, et plus particulièrement de ceux contenant une lysine. Ces observations nous amènent à penser que l'utilisation de ce réactif, en améliorant la couverture protéomique, augmente également les chances d'identification des protéines faiblement exprimées.

À la vue du nombre, sans cesse croissant de publications utilisant l'iTRAQ<sup>TM</sup>, on comprend que la position d'Elliott décrite plus tôt concernant son utilisation dans la recherche de biomarqueurs semble particulièrement pertinente.

---

4. Recherche réalisée sur le site Pubmed : <<http://www.ncbi.nlm.nih.gov/pubmed>> le 20 août 2013

### 3.4.3 La quantification sans marquage

Malgré leur efficacité démontrée, les techniques de quantification avec marquage présentent quelques limites. Parmi les plus décrites, on peut noter une préparation des échantillons complexe, un prix particulièrement élevé des réactifs, un risque de marquage incomplet, et la nécessité d'avoir à disposition des logiciels adaptés au traitement des données [Zhu et al., 2010]. De plus, à l'exception de l'iTRAQ<sup>TM</sup> et du marquage TMT, il n'est généralement possible de comparer que 2 à 3 échantillons entre eux. Des techniques de quantification sans marquage ont été imaginées pour compléter les approches utilisant un marquage. Ces approches ont été développées dans l'objectif de s'affranchir des étapes de ce marquage ; elles sont dites « *label-free* ».

La quantification relative basée sur l'intensité des pics en LC-MS est une approche basée sur le présupposé qu'un ion est détecté avec une intensité précise à un temps précis sur le chromatogramme et l'intensité du signal d'un ion obtenu par électrospray présente une corrélation avec la concentration de cet ion [Voyksner and Lee, 1999]. Cette approche présente des limites importantes d'un point de vue expérimental. En effet, lors d'une répétition d'expérience, il est très difficile d'obtenir des chromatogrammes strictement comparables à cause des variations expérimentales. Il peut exister d'infimes variations de temps de rétention des peptides d'une expérience à l'autre. Des normalisations sont donc nécessaires pour pouvoir interpréter convenablement ces expériences. Enfin, la quantité importante de données accumulées nécessite des puissances de calcul informatique très importantes, et l'utilisation d'algorithmes d'automatisation.

Le comptage des spectres est une des premières méthodes développées pour faire de la spectrométrie de masse quantitative sans marquage. Elle est basée sur l'idée qu'il existe une proportionnalité entre le nombre de spectres SM/SM obtenus et la quantité de peptide, et donc de protéine présente dans l'échantillon. Cette approche est rendue possible par le fait qu'une protéine présente en quantité importante dans un échantillon va générer un grand nombre de peptides protéolytiques. Cette augmentation du nombre de peptides conduit généralement à une augmentation de la couverture de la séquence de la protéine correspondante. L'équipe de Liu a analysé les corrélations entre l'abondance relative d'une protéine et le recouvrement de la séquence d'une part, le nombre de peptides d'autre part, et enfin avec le nombre de spectres. Ils ont ainsi démontré que seul le paramètre nombre de spectres présente une forte corrélation linéaire avec l'abondance relative de la protéine ( $r^2 = 0.9997$ ). En réalité cette approximation ne semble vraie que jusqu'à un facteur deux de

variation [Liu et al., 2004]. D'autre part, ce mode d'analyse ne semble pas adapté à l'étude des molécules faiblement concentrées [Schulze and Usadel, 2010]. Une étude parue en 2005 utilise la comparaison entre ce type de quantification et une quantification avec marquage de type SILAC. Ils démontrent ainsi que les variations observées sont comparables entre les deux approches [Zybailov et al., 2005].

La méthode du *Protein Abundance Index* (PAI) est une technique fiable basée sur le calcul du nombre de peptides observés pour une protéine par rapport au nombre de peptides théoriquement observables pour la protéine en question [Rappsilber et al., 2002]. Ainsi, le nombre de peptides identifiant une protéine augmente selon l'abondance de cette protéine. Plus tard, cet index a été modifié et passé en base exponentielle pour devenir le emPAI [Ishihama et al., 2005]. Dans l'étude d'Ishihama, les auteurs démontrent la possibilité de réaliser une quantification absolue sur 46 protéines issues d'un extrait total de cellules de souris en utilisant des peptides synthétiques.

#### 3.4.4 Quantification par SRM/MRM

Les méthodes de quantification de protéines par SRM/MRM sont utilisées dans une toute autre optique. Elles permettent de doser une protéine cible dans un mélange complexe. Ces méthodes présentent l'avantage de ne pas dépendre de l'utilisation d'anticorps de haute affinité, souvent difficiles à développer comme pour les tests ELISA. Par ailleurs, elles permettent d'accéder à des informations complémentaires telles que des mutations ponctuelles ou des modifications post-traductionnelles.

Les méthodes dites de « Selected Reaction Monitoring » (SRM) offrent la possibilité de quantifier une ou plusieurs molécules d'intérêt dans un mélange complexe [Lange et al., 2008]. Elles sont en place depuis de nombreuses années et elles représentent une approche de choix dans le suivi des petites molécules, notamment les médicaments et leurs métabolites. Les progrès réalisés dans le développement de spectromètres de masse plus sensibles rendent possible l'utilisation de ces approches pour des molécules de plus haut poids moléculaire, parmi lesquelles les peptides. Ces analyses sont classiquement réalisées à l'aide de spectromètres de masse possédant des technologies de type « triple quadrupôle » ou « Quadrupôle-Trap ». La première étape de l'analyse consiste à sélectionner un ion précurseur possédant un rapport  $m/z$  prédéfini dans le premier quadrupôle. Cet ion est ensuite fragmenté dans un deuxième quadrupôle dans lequel il va entrer en contact avec des molécules de gaz. Un ion fils, issu de ce précurseur est ensuite sélectionné dans un

troisième quadropôle, et c'est l'intensité du signal généré par cet ion qui est utilisé pour la quantification de la molécule à doser. Les rapports  $m/z$  de l'ion précurseur et de l'ion fils constituent une transition. Des logiciels informatiques permettent de sélectionner ces transitions au préalable de l'expérimentation, avec par exemple le logiciel ATAQS [Brusniak et al., 2011].

Par ailleurs, l'utilisation d'un standard interne marqué avec un isotope lourd et présent à une concentration connue dans l'échantillon permet de réaliser une quantification absolue de la protéine d'intérêt [Anderson and Hunter, 2006].

### 3.4.5 La quantification absolue

La quantification absolue par spectrométrie de masse fait partie des défis actuels. Pour le moment, la technique de référence pour faire de la quantification absolue est en effet, le test ELISA. Cependant, cette technique nécessite de connaître la protéine d'intérêt à doser, et de posséder un ou deux anticorps capables de la reconnaître.

#### 3.4.5.1 Quantification absolue avec marquage isotopique

Pour faire de la quantification absolue en spectrométrie, il est possible d'utiliser un standard interne de concentration connue. Un certain nombre ont été développés, et nous pouvons notamment citer la méthode AQUA (*Absolute QUAntification*) décrite en 2003 par l'équipe de Gygi [Gerber et al., 2003]. Deux approches similaires sont utilisables, l'une avec un peptide comme standard (AQUA), l'autre avec la protéine entière (méthode PSAQ).

La quantification absolue par peptide AQUA nécessite de faire synthétiser un peptide synthétique de la protéine à doser, modifié isotopiquement ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) sur certains acides aminés. Ce peptide est ajouté en quantité connue dans l'échantillon protéique dans lequel on veut doser la protéine d'intérêt. On effectue ensuite une digestion protéolytique à la trypsine. Le spectre SM met alors en évidence deux pics pour le peptide d'intérêt et la quantité du peptide natif peut être déterminée à partir du niveau d'expression du peptide marqué par rapport au peptide natif.

Une autre approche utilise la protéine entière et est appelée PSAQ (*Protein Standard Absolute Quantification*). Le principe consiste à utiliser une protéine recombinante entière alourdie par incorporation de lysine ou d'arginine modifiées, comme étalon interne [Brun et al., 2009]. Ces protéines sont biochimiquement identiques à la protéine naturelle à doser mais leur masse est différente. Cela permet d'ajouter la protéine modifiée préalablement à

tout traitement, limitant ainsi les biais liés au processus analytique [Dupuis et al., 2008].

Une équipe, menée par Alain Dupuis a démontré l'intérêt de cette stratégie en la comparant avec l'AQUA. Ils ont, pour cela, dosé une toxine bactérienne dans l'eau, l'urine, ou des échantillons alimentaires [Dupuis et al., 2008]. Leurs résultats semblent indiquer une meilleure précision de la méthode PSAQ par rapport à la méthode AQUA.

## 3.5 L'étude du glycoprotéome pour améliorer l'identification de biomarqueurs

### 3.5.1 La glycosylation

La glycosylation est une réaction chimique consistant en l'ajout d'un résidu glucidique sur un groupement accepteur et il s'agit d'une modification post-traductionnelle qui concerne environ 50% [Apweiler et al., 1999] des protéines cellulaires. Cette réaction est médiée par des enzymes<sup>5</sup> (principalement des glycosyltransférases) et donne lieu à une liaison covalente entre le résidu et le groupement accepteur dont les principaux sont les groupements hydroxyles (O-glycosylation) et les groupements amines (N-glycosylation) portés par les peptides et les protéines. Il existe environ 200 à 250 enzymes différentes intervenant dans le processus de glycosylation des protéines. Elles seraient codées par 1 à 2% du génome total chez l'homme [Schachter and Freeze, 2009]. Les glycosyltransférases sont métallo-dépendantes, envers le magnésium, ou le manganèse, qui jouent le rôle d'acides de Lewis en acceptant les électrons du groupe activateur.

Dans le cas des N-glycosylations, elles catalysent le transfert de sucres activés par leur liaison pyrophosphate à un lipide spécifique (le dolichol), sur des accepteurs (Cf. figure 3.21 p.84). Ce lipide est une longue chaîne carbonée constituée de dizaines de groupements isopréniques et pouvant traverser plusieurs fois la membrane du réticulum. Dans un premier temps, elles interviennent sur le précurseur de façon à ajouter progressivement les monosaccharides qui vont le constituer. La première réaction consiste à fixer un N-acétylglucosamine-1-phosphate sur le groupement phosphate porté par le dolichol. Par la suite, un N-acétylglucosamine et 5 mannoses sont ajoutés, et le précurseur passe ensuite à

---

5. Ne pas confondre avec la glycation qui consiste en l'ajout d'un groupement glucidique de façon spontanée. Cette réaction appelée réaction de Maillard a lieu dans de nombreux systèmes biologiques et consiste en la fixation du groupement carbonyle d'un sucre sur le groupement amine d'une protéine. Ce processus est impliqué dans de nombreuses pathologies, notamment dans le diabète. Dans cette pathologie, l'hémoglobine glyquée est un marqueur permettant d'évaluer l'efficacité du traitement sur un long terme, et de l'adapter si besoin.

la face interne du réticulum endoplasmique. Quatre mannoses et trois glucoses sont alors ajoutés. Ce complexe est ensuite transféré d'un bloc sur l'asparagine d'une protéine en cours de synthèse par un complexe enzymatique appelé OST pour oligosaccharyltransférase. Des modifications vont alors avoir lieu sur la structure glucidique. Ainsi, trois résidus glucoses et un résidu mannose vont être hydrolysés par des enzymes spécifiques (des glycosides hydrolases). Ces hydrolyses sont une sorte de signal aboutissant au transfert de la protéine vers l'appareil de golgi pour que la maturation glucidique puisse continuer. L'arbre glucidique va ensuite être modifié en fonction de son cheminement dans l'appareil de golgi pour être mûré. Il va ainsi subir des hydrolyses de mannoses ou des fixations d'autres monosaccharides. Cette maturation complexe fait intervenir plusieurs autres enzymes – hydrolases et glycosyltransférases – pour aboutir à une structure complexe, spécifique, donnant à la protéine sa structure tridimensionnelle définitive [Sears and Wong, 1998].

La O-glycosylation est nettement moins bien connue. En effet, cette modification semble être beaucoup plus hétérogène que la précédente, pouvant être constituée d'un monosaccharide seul (N-acétylglucosamine ou glucose) ou d'une chaîne oligosaccharidique extrêmement complexe. Il semble que ces modifications aient lieu sur des protéines déjà repliées (contrairement à la N-glycosylation) mais les signaux qui déterminent quel acide aminé sera modifié sont encore peu connus. La conformation tridimensionnelle des protéines semble donc être particulièrement importante pour la détermination des sites de glycosylation sur les groupement hydroxyles. Il s'agit d'une modification post-traductionnelle qui se déroule dans l'appareil de Golgi, mais la localisation exacte des enzymes impliquées est encore inconnue.

Pour permettre l'édification de ces « arbres glucidiques », la cellule utilise environ dix monomères glucidiques qui sont : le fucose (Fuc), le galactose (Gal), le glucose (Glc), le N-acétylgalactosamine (GalNAc), le N-acétylglucosamine (GlcNAc), l'acide glucuronique (GlcA), l'acide iduronique (IdoA), le mannose (Man), l'acide sialique (SA) et le xylose (Xyl). D'autres dérivés peuvent également être retrouvés mais en quantité moindre. Ces structures peuvent être fixées à environ 8 groupements accepteurs différents comme l'asparagine, la sérine, et la théonine pour les principaux. Mais il en existe d'autres comme par exemple, l'arginine, l'hydroxylysine, l'hydroxyproline, la tyrosine ou encore le tryptophane [Spiro, 2002].

Bien que très complexe, la glycosylation intervient dans de nombreux processus physiologiques. Ainsi, elle participe à la configuration tridimensionnelle des protéines, et elle

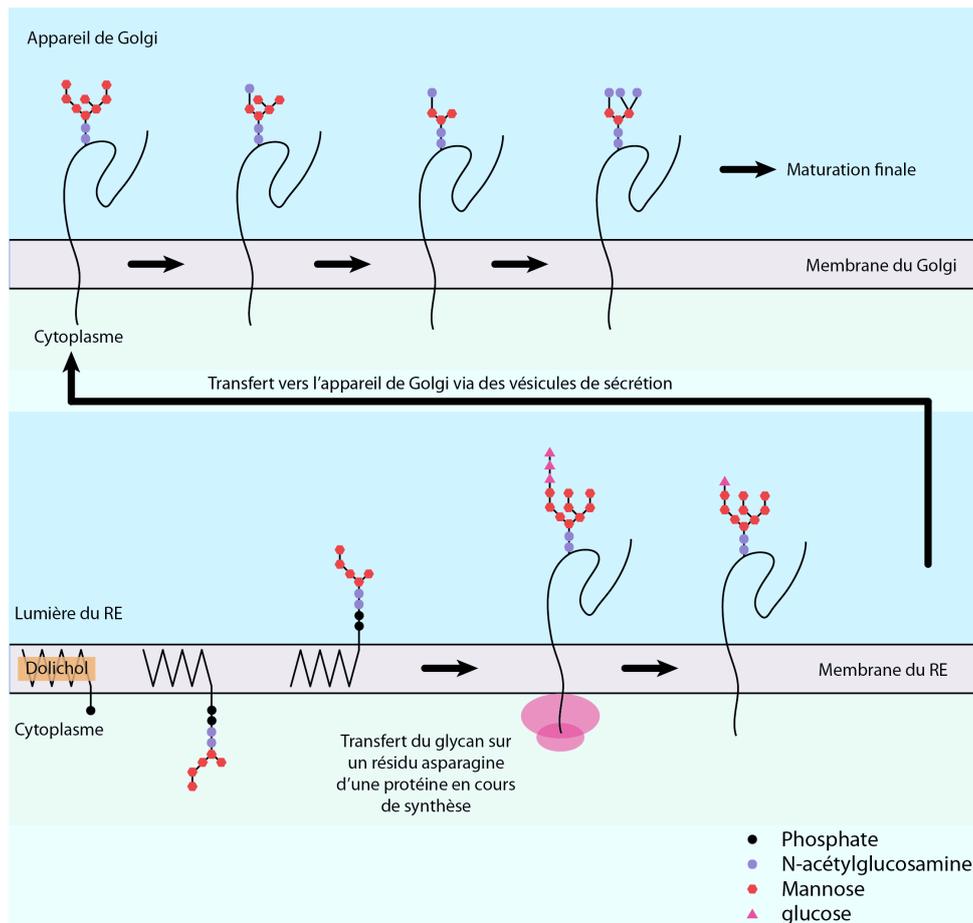


FIGURE 3.21 – Schéma montrant l'évolution de la N-glycosylation au cours de l'avancée des protéines dans le réticulum endoplasmique et dans l'appareil de Golgi. Un N-acétylglucosamine-1-phosphate est fixé sur le groupement phosphate porté par le dolichol. Un N-acétylglucosamine et 5 mannoses sont ajoutés, et le précurseur passe ensuite à la face interne du réticulum endoplasmique. Quatre mannoses et trois glucoses sont alors ajoutés. Ce complexe est ensuite transféré d'un bloc sur l'asparagine d'une protéine en cours de synthèse par un complexe enzymatique appelé OST. Trois résidus glucoses et un résidu mannose vont être hydrolysés par des glycosides hydrolases. La protéine est transférée vers l'appareil de Golgi pour que la maturation glucidique puisse continuer. L'arbre glucidique est modifié en fonction de son cheminement dans l'appareil de Golgi pour être mûri. Il va ainsi subir des hydrolyses de mannoses ou des fixations d'autres monosaccharides. Cette maturation complexe fait intervenir plusieurs autres enzymes – hydrolases et glycosyltransférases – pour aboutir à une structure complexe, spécifique, donnant à la protéine sa structure tridimensionnelle définitive.

intervient dans de nombreux mécanismes de reconnaissance moléculaire [Varki, 1993]. Elle participe aussi aux interactions cellules-cellules, aux interactions ligand-récepteur, à la réponse immunitaire, et à l'apoptose. Cela conduit à ce que des perturbations de la glycosylation soient aussi impliquées dans des pathologies, et notamment dans des défaillances du système immunitaire, des maladies neurodégénératives, ou encore dans l'apparition de cancers [Dennis et al., 2009].

Elle possède également une fonction d'adressage, selon la nature des sucres fixés,

les protéines vont être orientées soit vers le compartiment intracellulaire, soit vers une voie de sécrétion à la sortie de l'appareil de golgi. D'une manière générale, les protéines majoritairement O-glycosylées sont adressées au compartiment intracellulaire, alors que les protéines N-glycosylées sont surtout adressées à la face externe de la membrane plasmique et au milieu extracellulaire (sécrétion) [Wisniewski, 2010a] [Roth, 2002]. En effet, il est supposé que la O-glycosylation des résidus sérine ou thréonine des protéines permette d'inhiber transitoirement leur phosphorylation de sorte que les voies de signalisation en aval peuvent être modifiées [Wells, 2001]. Par ailleurs, la N-glycosylation qui apparaît sur les asparagines, répond à une séquence consensus. En effet, lorsque cet acide aminé est située dans une séquence N-X-S/T (où X peut être n'importe lequel des acides aminés excepté la proline), celui-ci pourra être glycosylé.

### 3.5.2 Pourquoi étudier le glycoprotéome

À l'heure actuelle, il est techniquement impossible d'avoir une vision globale du glycoprotéome. En effet, les méthodes d'analyse ne permettent pas de regrouper l'analyse de la structure des sucres fixés sur des protéines tout en ayant accès à une autre information toute aussi importante, le site sur lequel est fixé le groupement glucidique.

Cependant, il y a quelques points importants à noter ; tout d'abord, la plupart des protéines sécrétées par les cellules sont glycosylées. Ainsi, ces protéines constituent à elles seules une réserve extrêmement importante de biomarqueurs tumoraux [Janković, 2011]. De plus, la glycosylation, en ajoutant d'importants résidus glucidiques pouvant représenter près de 40% de la masse de la protéine, constitue un moyen efficace pour protéger les protéines sécrétées des attaques de potentielles protéases extracellulaires [Russell et al., 2009]. Ainsi, la demi-vie dans les fluides biologiques de ces protéines est augmentée, ce qui permet une détection plus aisée chez les patients. Ensuite, l'ajout de sucres permet également d'augmenter la solubilité des protéines dans les milieux aqueux.

Signe de l'importance croissante de l'étude du glycoprotéome, l'organisation internationale HUPO (*Human Proteomics Organization*) a ouvert en 2002 un axe de recherche spécialement consacré à la recherche de biomarqueurs du sang et des urines. Cet axe, appelé *Human Disease Glycomics/Proteome Initiative* (HGPI)<sup>6</sup>, a pour objectif de mettre en évidence, par des moyens d'analyse haut débit, des modifications du glyprotéome du sang ou des urines survenant suite à une pathologie afin de mettre en évidence des biomarqueurs

---

6. Source : <<http://www.hupo.org/research/hgpi/>> consultée le 20 août 2013

diagnostiques et/ou pronostiques de ces dernières.

La technique de choix pour l'étude du protéome, et par extension du glycoprotéome est la spectrométrie de masse. Cependant, concernant l'étude du protéome, il est totalement impossible à l'heure actuelle d'être certain d'identifier toutes les protéines exprimées par les cellules à un instant  $t$  donné. Ainsi, enrichir en protéines glycosylées l'échantillon à analyser permet de diminuer sa complexité (et donc d'améliorer l'identification des protéines) et permet également d'améliorer les chances d'identifier un biomarqueur pertinent dans la mesure où la plupart des protéines sécrétées sont glycosylées.

D'autre part, les protéines N-glycosylées sont adressées principalement à la membrane plasmique et au compartiment extracellulaire, cela en fait une réserve importante de biomarqueurs potentiels. Une illustration de ceci réside dans le fait que la plupart des biomarqueurs existants, décrits par des méthodes d'analyse globale classique, s'avèrent être des protéines glycosylées [Pan et al., 2011]. Ainsi, la protéine TIMP-1, biomarqueur du cancer colorectal, est une protéine glycosylée, tout comme le CEA. De la même façon, le CA125 serait également une protéine potentiellement glycosylée. Il y a également l'exemple de l'antigène PSA utilisé dans le cancer de la prostate ou encore l' $\alpha$ -foetoprotéine, utilisée dans le cancer des testicules.

### 3.5.3 Glycoprotéomique

Tout d'abord, il est nécessaire de distinguer dès maintenant deux disciplines différentes mais complémentaires. Dans la littérature, les auteurs parlent de glycomique pour parler des études qui traitent de l'analyse structurale des sucres portés par les protéines. La glycomique est à distinguer de la glycoprotéomique qui consiste à analyser les protéines porteuses de groupements glucidiques. Cette dernière est donc basée sur l'enrichissement d'un échantillon peptidique complexe en peptides glycosylés afin d'analyser principalement ces dernières par spectrométrie de masse. En effet, la caractérisation du glycoprotéome d'un tissu ou d'un sérum nécessite de passer par une étape d'enrichissement des protéines glycosylées qui présentent des écarts de concentration importants. Pour mener à bien cet enrichissement, deux approches complémentaires ont été développées. L'une est basée sur l'utilisation de lectines et l'autre est basée sur la propriété des sucres oxydés à réagir sur des groupements hydrazines. Ces deux approches ne sont pas exclusives et certaines équipes ont fait le choix d'utiliser les deux pour leur complémentarité [Pan et al., 2006]. D'autres outils plus secondaires ont également été développés comme l'utilisation de l'acide boronique

à la place de l'hydrazine, mais les résultats semblent moins probants [Sparbier et al., 2006]. La technique de gel filtration [Alvarez-Manilla et al., 2006] ou la chromatographie d'interactions hydrophiles (HILIC) [Hägglund et al., 2004], de même que des colonnes de poudre de graphite [Larsen et al., 2005] sont également utilisées dans quelques études, mais ces outils restent marginaux.

### 3.5.4 Les modes d'enrichissement

#### 3.5.4.1 Les lectines

Le mot *lectine* vient du latin *legere* qui signifie entre autres, sélectionner, ramener à soi. Il s'agit en effet de protéines ayant une affinité particulière pour les groupements glucidiques fixés sur des glycoprotéines ou sur des glycopeptides. On les retrouve chez les plantes, les champignons, les bactéries et les animaux. Certains virus les utilisent pour se fixer sur les membranes des cellules à infecter, ce qui induit une endocytose de ce dernier et lui permet de poursuivre son cycle répliatif. Il existe une variété importante de lectines, chacune étant capable de reconnaître spécifiquement une structure glucidique spécifique. Ainsi, la concanavaleine A (ConA) est capable de reconnaître les résidus glucose et mannose fixés aux protéines ou aux peptides [Becker et al., 1975]. La *wheat germ agglutinin* (WGA) est capable de reconnaître spécifiquement les résidus N-acétylglucosamines et l'acide sialique [Han et al., 1987]. La *ricinus communis agglutinin* (RCA<sub>120</sub>) est également utilisée et elle peut reconnaître spécifiquement le galactose en position terminale de l'arbre glucidique ou lorsqu'il est modifié en position 0 ou 3 [Zielinska et al., 2010]. Afin d'améliorer les capacités d'enrichissement, plusieurs lectines (souvent ConA et WGA) sont associées lors des étapes de décomplexification de l'échantillon. Ainsi, les chances de retenir le plus grand nombre possible de protéines glycosylées se trouvent augmentées [Yang, 2006]. Une étude récente associe la technique *filter aided sample preparation* (FASP) (Cf. paragraphe 3.5.5 p.89) à ces étapes d'enrichissement multilectines dans le cadre de l'étude du N-glycoprotéome du plasma ou de tissus de souris [Zielinska et al., 2010]. Il s'agit de l'une des études couvrant le glycoprotéome le plus large avec une identification de plus de 74% des sites de glycosylation connus chez la souris et de 5753 nouveaux sites. Le principe général de ces approches basées sur l'utilisation des lectines est de mettre en contact les protéines ou les peptides avec les lectines immobilisées sur un support. Les protéines ou les peptides non glycosylés sont évacués par des étapes de lavage, puis les peptides sont libérés des lectines par l'action d'une enzyme capable de couper la liaison entre le sucre et l'acide aminé. Ils

peuvent également être libérés en utilisant des réactifs chaotropiques ou en utilisant des sucres de façon compétitive [Wisniewski, 2010b]. Un autre moyen d'augmenter l'efficacité de l'enrichissement consiste à réaliser l'étape d'enrichissement sur les protéines et sur les peptides. C'est ce que propose par exemple l'équipe de Wisniewski [Bunkenborg et al., 2004].

Il existe cependant une limite avec cette approche. En effet, les lectines sont spécifiques de certains motifs glucidiques de sorte qu'il est peu probable que l'utilisation de deux ou trois lectines différentes soit suffisante pour enrichir la totalité des motifs existants. Il se peut que certaines protéines ou certains peptides, porteurs de motifs glucidiques atypiques, soient évacués lors des étapes de lavage des lectines. De plus, la constante d'association entre les lectines et les sucres est relativement faible ; elle est comprise entre  $10^6$  et  $5.10^7$   $M^{-1}$ . Pour comparaison, celle de la liaison antigène/anticorps est entre 100 et 10 000 fois plus forte, avec des valeurs comprises entre  $10^8$  et  $10^{12}$   $M^{-1}$  [Meany et al., 2009].

### 3.5.4.2 Les résines hydrazines

Pour remédier au problème de spécificité des lectines, une autre approche a été développée en utilisant les propriétés physico-chimiques des sucres. Cette approche décrite par Zhang en 2003 utilise la capacité des glucides oxydés à réagir sur des billes saturées d'hydrazine [Zhang et al., 2003a].

Au cours de la première étape, les glycopeptides sont oxydés grâce à du méta-périodate de sodium sur les résidus sucrés. Les fonctions *cis-diols* deviennent alors des fonctions aldéhydes. Les fonctions aldéhydes libres sont extrêmement réactives et elles vont réagir sur les fonctions amines ( $NH_2$ ) des hydrazines pour former une liaison covalente (Cf. figure 3.22 p.89). Les protéines ou peptides non liés peuvent alors être élués en réalisant des lavages, ce qui permet de ne garder que ce qui est fixé sur les billes. L'action d'une enzyme de déglycosylation comme la PNGaseF permet ensuite de libérer les peptides fixés pour pouvoir ensuite les analyser par spectrométrie de masse.

Cette méthode d'enrichissement, bien que très efficace, comporte également une limite essentielle. En effet, il n'existe pas d'enzyme spécifique capable de cliver les O-Glycosylations, ce qui conduit à ce que seule les N-glycosylations soient détectées par cette technique. Cependant cela peut être compensé par l'utilisation des lectines et dans tous les cas, dans le cadre d'une approche orientée vers la mise en évidence de biomarqueurs, nous avons déjà mentionné que les N-glycosylations étaient les plus importantes [McDonald et al., 2009].

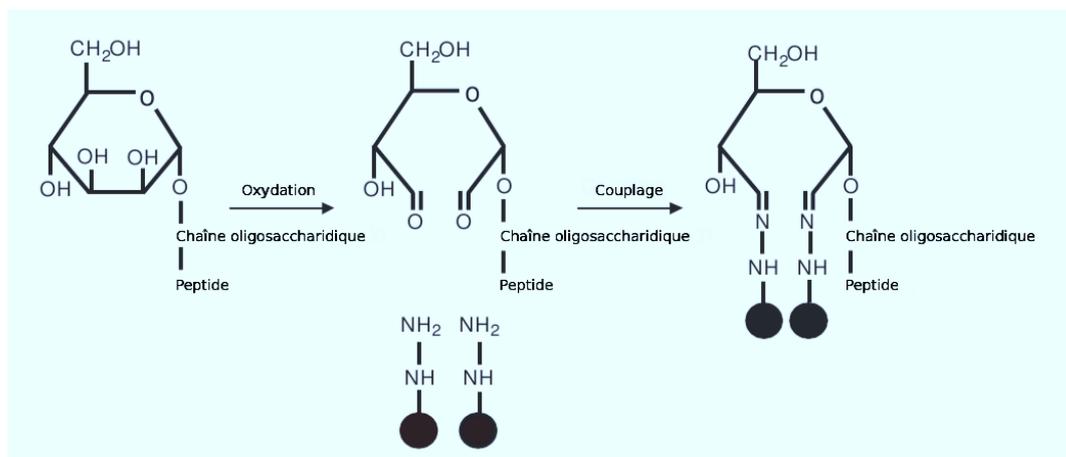


FIGURE 3.22 – Réaction de fixation sur la résine hydrazine [Zhang et al., 2003a] – Les sucres fixés sur les peptides sont oxydés transformant les fonctions diols en aldéhydes qui vont ensuite pouvoir réagir sur les fonctions hydrazines fixées sur les billes et se fixer de façon covalente à ces dernières. Une déglycosylation permet alors de récupérer les peptides d'intérêt.

Ce sont en effet celles qui sont le plus souvent retrouvées sur les protéines extracellulaires.

### 3.5.4.3 Une information importante : les sites de glycosylation

Il est également possible, grâce à ces deux approches, d'obtenir avec précision les sites sur lesquels étaient fixés les sucres qui ont servi à l'enrichissement. En effet, il suffit pour cela de réaliser la déglycosylation dans l'eau lourde [Kaji et al., 2003]. Autrement dit, de l'eau sur laquelle l'atome d'oxygène est remplacé par son isotope stable ayant une masse de 18. Ainsi, lors de l'action de la PNGaseF, l'hydrolyse du sucre va permettre d'incorporer un  $^{18}\text{O}$  dans la chaîne latérale de l'asparagine qui devient un acide aspartique, dont la masse passe alors d'environ 132 à 135 par addition d'un proton et d'un atome d'oxygène lourd. L'analyse MS/MS permet ensuite de retrouver ces acides aminés qui ont été modifiés par l'action de l'enzyme et de déterminer leur position dans la protéine.

### 3.5.5 La technique FASP et ses avantages

La technique FASP (Filter aided sample preparation) a été développée en 2009 par l'équipe de Matthias Mann [Wisniewski et al., 2009]. Cette technique a pour objectif de créer un réacteur dans lequel va se dérouler l'ensemble des étapes aboutissant à la digestion des peptides.

L'échantillon est lysé dans un tampon de lyse contenant du DTT (dithiothreitol) à 100mM, du TRIS (tris(hydroxyméthyl)aminométhane) et 4% de SDS (Sodium dodecyl sulfate). La présence de ce dernier dans le tampon de lyse permet de réaliser la réduction

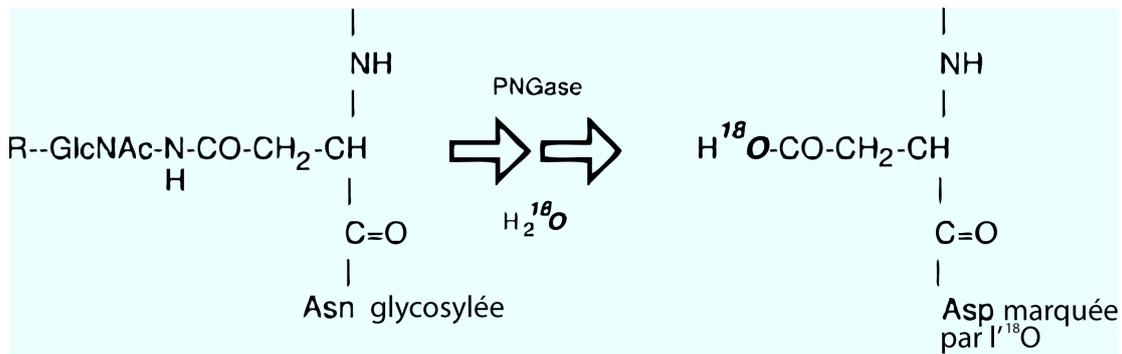


FIGURE 3.23 – Réaction de déglycosylation dans l' $\text{H}_2^{18}\text{O}$  [Kaji et al., 2003] – L'action de l'enzyme sur les asparagines induit une déamidation de cette dernière en acide aspartique.

des ponts disulfures dès les premières étapes de lyse. L'échantillon est ensuite soniqué pour être clarifié, puis le lysat est placé dans un centricon contenant une membrane ayant la capacité de laisser passer les molécules ayant un poids moléculaire inférieur à une valeur consigne choisie. Dans notre cas, nous avons choisi une valeur consigne de 30kDa car l'équipe de Wisniewsky a montré qu'il s'agissait de la meilleure solution pour l'étude des glycopeptides [Wisniewski et al., 2011b]. Des lavages avec un tampon contenant 8M d'urée et ayant donc un pouvoir chaotropique important vont permettre d'évacuer le SDS. C'est également dans ce tampon que l'alkylation des groupements thiols libres va avoir lieu avec l'ajout de 50mM de iodoacétamide ou de 10mM de méthylméthane thiosulfonate (MMTS). Puis des lavages sont à nouveau réalisés pour évacuer le réactif chimique alkylant à l'aide du tampon contenant 8M d'urée. Ce tampon est alors remplacé progressivement par un tampon compatible avec l'action de la trypsine tel que le bicarbonate d'ammonium ou le TEAB (TriEthylAmmonium Bicarbonate). Puis les protéines sont digérées par la trypsine. Les peptides générés sont alors capables de passer à travers le filtre et peuvent ainsi être récupérés par une simple centrifugation dans un nouveau tube collecteur.

L'un des avantages majeurs de cette technique est une limitation de la perte de matériel par les différentes étapes de traitement de l'échantillon et notamment l'étape de précipitation à l'acétone. Un deuxième avantage réside dans le tampon de lyse utilisé. En effet, ce dernier contient 4% de SDS ce qui permet de solubiliser une partie des protéines membranaires des cellules. Les cellules ou les tissus sont donc lysés dans ce tampon qui contient également du DTT (DiThioThreitol) pour permettre la réduction des ponts disulfures des protéines.

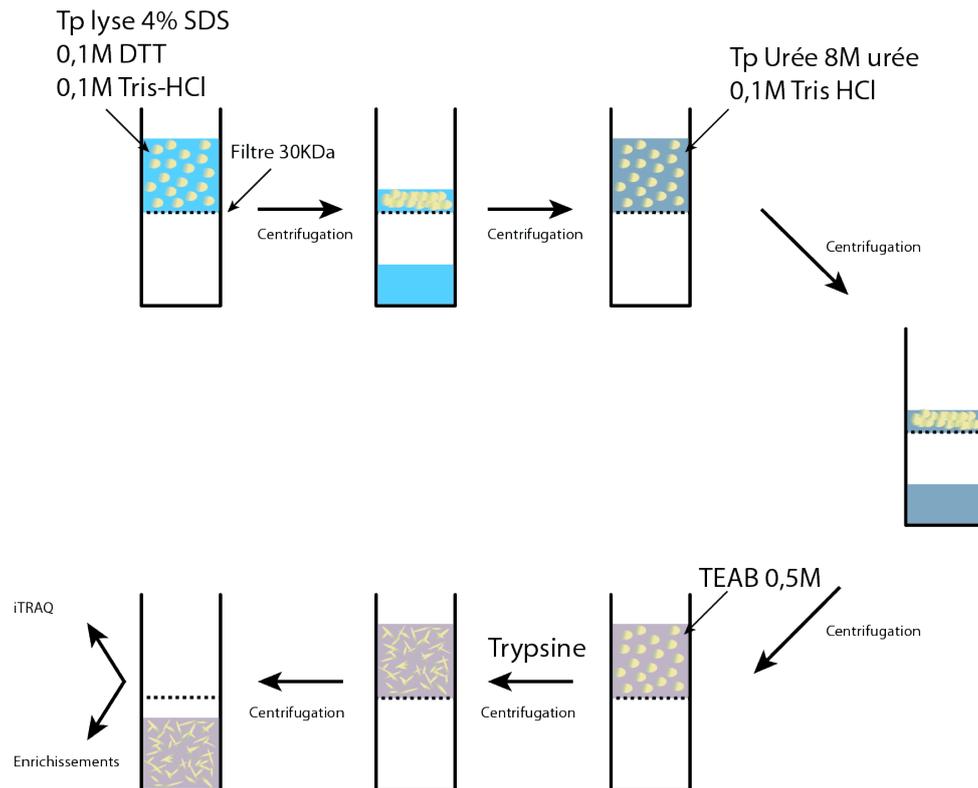


FIGURE 3.24 – Schéma illustrant les étapes nécessaires à la réalisation d'une extraction peptidique par l'approche FASP – Les protéines solubilisées dans le tampon de lyse sont placées au dessus du centricon, puis des lavages à l'urée vont permettre de chasser le SDS puis l'urée est ensuite échangée avec un tampon compatible avec l'action de protéases. Les peptides générés peuvent être récupérés par une simple centrifugation.



## Deuxième partie

### Matériels et méthodes



# Chapitre 1

## Détail des techniques utilisées

### 1.1 Les tumeurs

P OUR mener à bien cette étude, des pièces de tissus présents dans la tumorothèque du Centre Régional de Lutte Contre le Cancer (CRLCC) *Paul Papin* ont été utilisées. Le protocole d'étude, ainsi que la forme du consentement des patients ont été validés par le comité d'éthique de l'hôpital d'Angers et de l'institut de cancérologie de l'Ouest Paul Papin. Ces tissus, collectés entre 1999 et 2004, sont conservés dans l'azote liquide à environ  $-170^{\circ}\text{C}$  peu de temps après leur prélèvement par les chirurgiens et constituent une tumorothèque propre à l'institut. Le tissu sain est obtenu par dissection d'un morceau de muqueuse saine distante d'au moins 10 cm de la tumeur. Le statut de ces tumeurs a été évalué par au moins deux médecins anatomopathologistes suite à une coloration hématoxyline, éosine, safran (HES) de coupes paraffinées issues de ces prélèvements.

### 1.2 Génération des coupes de tumeurs et microdissection laser

À l'aide d'un cryostat (Bright Instrument Co. Ltd., Huntingdon, UK), des coupes de  $12\ \mu\text{m}$  de ces tissus ont été réalisées, déposées sur lame, puis colorées au bleu de Toluidine pour avoir une idée de la structure générale du tissu (cf. figure 1.1 page 96). À cette étape, une lecture par un médecin anatomopathologiste a été réalisée afin de s'assurer du caractère tumoral du prélèvement. Dans un deuxième temps, d'autres coupes ont subi une déshydratation grâce à des bains successifs d'éthanol de concentrations croissantes. Un bain de 30 secondes dans de l'éthanol à 70%, un bain de 30 secondes dans de l'éthanol à 90% et deux bains de 2 minutes dans de l'éthanol absolu et deux bains de xylène de cinq minutes ont été réalisés sur les coupes. Une fois déshydratées, les coupes ont subi une

étape de microdissection laser, permettant de récupérer uniquement les cellules tumorales. L'appareil utilisé est un PixCell II (Arcturus Engineering, Mountain View, CA) associé au logiciel *PixCell II archiving image*. Ce microdissecteur est équipé d'un laser ayant une longueur d'onde  $\lambda = 810$  nm. Les paramètres de microdissection utilisés sont les suivants :

- diamètre du spot :  $7,5 \mu\text{m}$  ;
- durée du tir : 70 ms ;
- puissance du laser : 70 mW.

Les films plastiques des capsules de microdissection (CapSure LCM caps -Molecular Devices Corporation- généralement au nombre de 5) contenant les cellules microdisséquées (environ 30 000) ont été placés dans des tubes dans lesquels a été ajouté le tampon de lyse protéique.

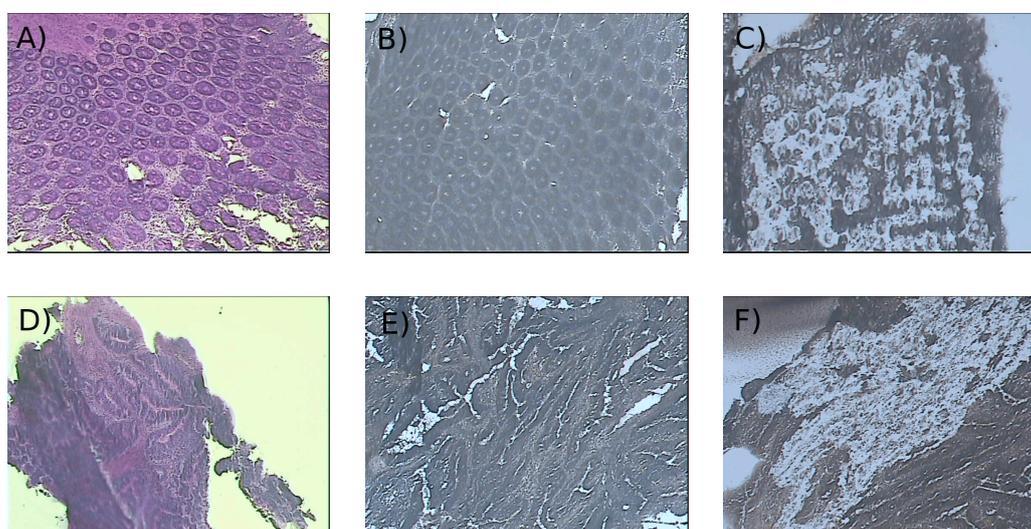


FIGURE 1.1 – **Comparaison entre le tissu colique sain et le tissu tumoral** – A) Coupe de tissu sain colorée au bleu de toluidine, B) Coupe de tissu sain déshydratée, C) Coupe de tissu sain après déshydratation, D) Coupe de tissu tumoral colorée au bleu de toluidine, E) Coupe de tissu tumoral déshydraté, F) Coupe de tissu tumoral après microdissection.

### 1.3 Extraction des protéines, digestion trypsique et dosage

L'extraction des protéines a été réalisée à l'aide d'un kit d'extraction (*Expression Pathology Inc., Gaithersburg, MD*). Ce kit appelé *Liquid Tissue MS protein preparation*, permet une extraction optimale des protéines issues d'échantillons tumoraux.

Après que les cinq films plastiques des capsules aient été déposés dans des tubes à faible pouvoir adsorbant, 20 à 30  $\mu\text{L}$  de tampon d'extraction du kit ont été ajoutés et le tout chauffé à  $95^\circ\text{C}$  pendant 90 minutes. Les tubes ont ensuite été incubés à température ambiante pendant environ 2 minutes, le temps que la température de l'échantillon diminue.

Ensuite, 5 à 10  $\mu\text{g}$  de trypsine – traitée au L-(tosylamido-2-phényl) éthyl chlorométhyl kétone (TPCK, inhibiteur de l'activité chymotrypsique contaminante) en présence de  $\text{CaCl}_2$  (*AB SCIEX, Foster City, CA, USA*) – ont été ajoutés dans les tubes et ces derniers ont été placés à 37 ° C. Une agitation régulière et vive a été ménagée pendant la première heure d'agitation, puis une agitation légère a été maintenue durant toute la durée de la digestion qui s'est déroulée pendant une nuit complète. Une centrifugation a été réalisée afin d'éliminer les débris de capsule restants, et le surnageant a été prélevé et transféré dans un nouveau tube. Ce dernier a ensuite été complètement déshydraté en utilisant un *speedvac*.

Le culot sec ainsi obtenu a été repris dans 100  $\mu\text{L}$  d'une solution contenant 0,5% d'acide trifluoroacétique (TFA), et 5% d'acétonitrile (ACN). Les peptides ont ensuite été dessalés sur des cartouches Sep-Pack C18 (*Waters Corporation, Milford, MA, USA*). Pour cela, la colonne a été activée à l'aide de 3 mL de méthanol 100%, puis rincée avec 5 mL d'eau déionisée. L'échantillon a ensuite été déposé sur la colonne, les peptides se sont alors fixés sur les chaînes carbonées. Ils ont été rincés avec 5 mL d'eau afin d'éliminer les sels. L'élution a ensuite été réalisée grâce à l'utilisation d'un mL de méthanol 100% puis ce méthanol a été évaporé au *speedvac*.

Les peptides ont ensuite été resuspendus dans 30 $\mu\text{L}$  de triéthylammonium bicarbonate (TEAB) (*Sigma Aldrich Corporation*) à pH 8,5, et dosés en utilisant le kit de dosage microBCA (*Pierce*). L'échantillon a ensuite été dilué de façon à obtenir une concentration de 33,3  $\mu\text{g}/\mu\text{L}$ , autrement dit, qui les 100  $\mu\text{g}$  de peptides soient dans 30  $\mu\text{L}$  de TEAB.

## 1.4 Réduction, Alkylation, marquage iTRAQ

Les 100  $\mu\text{g}$  de peptides issus de chacune des tumeurs ont ensuite été réduits en ajoutant 5 mM final de tris-(2-carboxyethyl)phosphine (TCEP) à 60 ° C pendant une heure. Puis les groupements thiols ainsi réduits ont été alkylés en ajoutant 10 mM final de méthyl méthane thiosulfonate (MMTS) pendant 10 à 15 minutes à température ambiante. Ensuite, chaque étiquette iTRAQ<sup>TM</sup> a été reprise dans 70  $\mu\text{L}$  d'éthanol pour le réactif 4-plex, ou dans 50  $\mu\text{L}$  d'isopropanol pour le réactif 8-plex. Ces étiquettes ont ensuite été mélangées aux 100  $\mu\text{g}$  de peptides de chaque échantillon et le tout a été incubé deux heures à température ambiante. Cette réaction de marquage a été stoppée par l'ajout de 200  $\mu\text{L}$  d'eau. 5  $\mu\text{L}$  de chaque échantillon ont ensuite été prélevés et mélangés pour être analysés en SM/SM afin de vérifier l'égalité des quantités de peptides issues des échantillons. Pendant ce temps, le

reste des mélanges peptides/étiquettes a été évaporé au *speedac*. Les culots ont ensuite été repris dans 100  $\mu\text{L}$  de TFA 0,1% puis dessalés sur une cartouche Sep-Pack de C18 (Waters), et le méthanol servant à l'élution a été évaporé au *speedvac*. C'est à cette étape que les quatre ou les huit échantillons marqués avec des étiquettes différentes, ont été mélangés.

## 1.5 Isoélectrofocalisation OFFGEL

La première dimension de séparation des peptides a été effectuée selon leur point isoélectrique grâce à la technique OFFGEL. Pour cela, nous avons utilisé un *3100 OFFGEL Fractionator* (Agilent) en configuration 24 puits. Vingt quatre fractions ont ainsi été générées. Les culots de peptides sont ressolubilisés dans 3,6 mL d'une solution à 2,4% de glycérol, et 1% d'*IPG buffer* pH 3-10 (*GE Healthcare*). La bande contenant un gel d'acrylamide dans lequel est ménagé un gradient de pH allant de 3 à 10, a été déposée au fond des cupules et réhydratée avec cette même solution, pendant 15 minutes, afin de permettre l'étanchéité des cupules. Ensuite, 150  $\mu\text{L}$  de la solution contenant les échantillons ont été déposés dans les cupules. L'électrofocalisation a ensuite été réalisée à 20 ° C et 50  $\mu\text{A}$ . Après l'isoélectrofocalisation, les 24 fractions ont été prélevées, puis les cupules ont été lavées avec 200  $\mu\text{L}$  d'une solution composée de 50% de méthanol et d'1% d'acide formique. Après 30 minutes, la solution de lavage des cupules a été ajoutée à l'échantillon correspondant, et le tout a été évaporé au *speedvac*. Les culots, ne contenant plus que le glycérol et les peptides marqués ont été conservés à -20 ° C jusqu'à la réalisation de l'étape de séparation sur nano-RPLC.

## 1.6 La séparation par RPLC

Les peptides de chacune des fractions ont été ressolubilisés dans 20  $\mu\text{L}$  d'une solution de TFA 0,1% avant d'être injectés sur la colonne de C18 de l'*Ultimate 3000 nano-LC system* (*Dionex, Sunnyvale, CA*). Nous avons utilisé une colonne C18 de type PepMap100, 3  $\mu\text{m}$ , 100 Å, 75  $\mu\text{m}$  de diamètre interne par 15 cm de long (*Dionex*) avec un flux de 300 nL/min. Le solvant A était composé de 2% d'ACN dans de l'eau contenant 0,05% de TFA et le solvant B était constitué de 80% d'ACN dans de l'eau contenant 0,04% de TFA. Les peptides ont été dessalés pendant trois minutes grâce au solvant A sur la précolonne, puis une étape de séparation de 105 min a été réalisée. Le gradient utilisé était le suivant :

- de 0 à 20% de B : 10 min ;
- de 20 à 45% de B : 85 min ;
- de 45 à 100% de B : 10 min.

À la deuxième minute de ce programme, un signal de démarrage était envoyé au Probot (*Dionex*) permettant la collecte et le dépôt de l'échantillon directement sur les plaques compatibles avec le spectromètre de masse (1664 spots/plaque; *Applied Biosystems, Foster City, CA*). Cet appareil permet de co-déposer l'échantillon avec une matrice choisie au préalable. Ici, les échantillons ont été co-déposés avec de l'acide  $\alpha$ -cyano-4-hydroxycinnamique CHCA (*LaserBioLabs, Sophia-Antipolis, France*) à une concentration de 2 mg/mL dans un mélange composé de 70% d'ACN et de 0,1% de TFA. Cette dernière est continuellement injectée dans le système de T du Probot à un débit de 1,2  $\mu$ L/min, ce qui permet d'obtenir des spots présentant un rapport 1 d'échantillon pour 4 de matrice. Ces spots ont été collectés durant une durée six secondes.

## 1.7 Analyse par spectrométrie de masse MALDI TOF/TOF

Les analyses des plaques MALDI ont pu être réalisées sur deux spectromètres de masse, le MALDI TOF/TOF 4800 accompagné de la version 3.5 du logiciel *Protein Pilot* et le 5800 accompagné de la version 4.0 de ce même logiciel (*Applied biosystem / ABSciex*).

Ils étaient utilisés en mode positif et les plaques étaient calibrées à l'extérieur grâce à un kit de calibration (*ABSciex*). La puissance du laser était comprise entre 2800 et 3400 pour la SM et entre 3600 et 4200 pour l'acquisition SM/SM.

Après avoir réalisé l'acquisition des spectres SM à raison de 1500 tirs laser par spectre, les 12 ions les plus intenses (ayant un rapport signal/bruit supérieur à 12) étaient automatiquement sélectionnés et fragmentés en SM/SM dans la chambre de collision en utilisant l'air comme gaz de fragmentation. L'énergie de la chambre de collision maintenue à 1 kV et le vide y était d'environ  $2 \times 10^{-6}$  Torr. Chacun des spectres SM/SM résultait de l'acquisition de 2000 tirs laser.

Une calibration interne a également été utilisée en utilisant l'ion Glu-1 du fibrinopeptide ayant une masse de 1570,677 Da, présent dans la matrice à une concentration de 30 fmol/spot.

L'identification a été réalisée grâce à l'algorithme Paragon du logiciel *Protein Pilot*. Les paramètres de cette identification, sauf mention contraire dans les résultats, a été réalisée avec au moins deux peptides, une précision en masse de 50 uma et un score de

confiance supérieur ou égal à 95%. Un indice de confiance appelé « *unused score* » a été défini comme devant être supérieur à 1,3 pour atteindre les 95% de confiance. Chaque spectre SM/SM a été comparé à la base de données Uniprot/Swissprot en sélectionnant l'espèce *Homo Sapiens*. Les modifications telles que la fixation du groupement MMTS sur les cystéines, la spécificité des clivages tryptiques ont été recherchées par l'analyse avec le logiciel. Les protéines identifiées ont été groupées avec l'algorithme *Progroup* du logiciel afin de limiter la redondance. La correction des biais a également été exécutée. Le taux de faux positif résultant de ces critères se situe autour de 2%.

Une question récurrente se pose dans le cadre des analyses protéomiques quantitatives. À partir de quel seuil de variation considérer qu'une protéine varie de façon statistiquement significative? Pour s'affranchir de ce problème, nous avons eu recours à une version personnalisée du logiciel iQuantitator publié par John Schwacke en 2009 [Schwacke et al., 2009b]. Ce logiciel a pour objectif de normaliser l'ensemble des rapports iTRAQ<sup>TM</sup> des protéines d'une expérience, et de définir un intervalle de confiance à l'intérieur duquel il est possible de considérer que la variation du niveau d'expression de la protéine n'est pas dû au hasard, mais qu'il s'agit bien d'une différence significative (cf figure 3.20 p.75). Ainsi, lorsque les deux bornes de l'intervalle de confiance sont supérieures à 1, la protéine est significativement sur-exprimée, et inversement, lorsque les deux bornes sont inférieures à 1, la protéine est significativement sous-exprimée. D'autre part, des intervalles de confiance sont également appliqués à chaque peptide ayant permis d'identifier et de quantifier une protéine. L'intervalle de confiance pour la protéine tient donc compte des intervalles de confiance de chaque peptide ayant permis d'identifier la protéine. La valeur statistique ainsi obtenue est donc particulièrement robuste, elle permet également de mettre en évidence des protéines dérégulées dont la variation serait faible mais significative.

## 1.8 Les outils d'analyse en protéomique

L'ontologie des protéines d'intérêt a été analysée à l'aide du logiciel GoMiner. L'ensemble des protéines qui avaient été sélectionnées pour l'analyse iQuantitator a été utilisé comme liste de comparaison. Chacun des termes ontologiques regroupant plus de 5 protéines a été utilisé pour le calcul statistique.

Une valeur de  $p$  a pu être calculée pour chaque terme en utilisant un test exact de Fisher. Le taux de faux positifs a été déterminé en utilisant des pools de 1 000 protéines choisies aléatoirement comme liste de comparaison.

Par ailleurs, une analyse des réseaux dans lesquels interviennent les protéines candidates a été réalisée à l'aide de la suite logiciel MetaCore<sup>TM</sup> v4.7 (GeneGo, Inc., St. Joseph, MI) et leur variation d'expression ont été comparées en utilisant les valeurs de  $p$  inférieures à 0,01 comme indicateurs statistiques. Les identifiants de gènes des fichiers téléchargés ont été appariés avec ceux des ontologies fonctionnelles dans MetaCore<sup>TM</sup>.

## 1.9 L'immunohistochimie

Cent vingt-six patients atteints d'adénomes ou d'adénocarcinomes colorectaux ont été étudiées par immunohistochimie. Toutes les tumeurs ont été obtenues à partir du laboratoire d'anatomopathologie du centre Paul Papin et du CHU d'Angers et collectées entre 2000 et 2005. Certains échantillons ont été exclus notamment les jeunes patients (moins de 40 ans) ou les tumeurs ayant reçu une chimiothérapie. Ces tumeurs étaient presque toutes localisées au niveau du côlon ou du haut rectum. Il y avait 30 adénomes avec 15 adénomes de bas grade et 15 adénomes de haut grade, dont la dysplasie a été classée selon les caractéristiques architecturales et cytologiques. Il y avait également 72 adénocarcinomes colorectaux. Selon le septième système de classification TNM de l'*American Joint Committee on Cancer*, la profondeur de l'invasion tumorale dans chacun des carcinomes a été classée en cinq groupes : Tis, le carcinome in situ ou limité à la muqueuse ; T1, envahissant la sous-muqueuse ; T2, envahissant la musculature ; T3, envahissant la sous-séreuse et T4, envahissant au delà de la séreuse et les organes contigus. Le statut des métastases ganglionnaires était donc classé de la sorte : N0, absence de métastases régionales dans les ganglions lymphatiques ; N1, une à trois métastases régionales présentes dans les ganglions lymphatiques ; N2, quatre métastases ganglionnaires ou plus présentes dans les ganglions lymphatiques. La présence de métastases à distance a été classée ainsi : M0, absence de métastases à distance ; M1, présence de métastases à distance.

Une coupe représentative d'une section transversale de la tumeur et de la muqueuse saine a été sélectionnée pour chaque échantillon. L'immunohistochimie a été réalisée sur des coupes paraffinées de 4  $\mu\text{m}$  d'épaisseur fixés au formol en utilisant un anticorps dirigé contre l'OLFM4 (numéro de catalogue ab78496 ; Abcam, Cambridge, MA ; 1 :25). La technique d'immuno-marquage a été réalisée par un système faisant référence en terme de coloration des tissus automatisée (Ventana).

Pour l'analyse des images, nous avons attribué un score en fonction du pourcentage de cellules marquées –  $< 10\% \Rightarrow 0$  ; entre 10 et 30%  $\Rightarrow 1$  ; entre 30 et 50%  $\Rightarrow 2$  ; entre

50 et 70%  $\Rightarrow$  3;  $> 70\%$   $\Rightarrow$  4 -. Ensuite, l'intensité de marquage des cellules marquées a été évaluée - aucun  $\Rightarrow$  0; faible  $\Rightarrow$  1; intermédiaire  $\Rightarrow$  2; et fort  $\Rightarrow$  3 - puis un score a été établi à partir de ces données en multipliant ces valeurs - négatif  $\Rightarrow$  0; faiblement positif  $\Rightarrow$  1-3; positif  $\Rightarrow$  4-7; fortement positif  $\Rightarrow$  8-12 - et les résultats ont été analysés statistiquement.

Pour exclure la subjectivité, toutes les coupes ont été évaluées par deux pathologistes qui n'avaient aucune connaissance de l'identité des patients ou du statut clinique. Dans les cas discordants, les deux médecins ont examiné les coupes ensemble et sont parvenus à un consensus.

## 1.10 Western Blot

Les extraits totaux tumoraux ont été préparés à partir de tissus normaux et des tissus tumoraux. Des échantillons de tissus congelés ont été homogénéisés et lysés dans un tampon contenant 7 M d'urée, 2 M de thiourée, et 4% (p / v) de CHAPS à 4 ° C pendant 1 h sur la roue. La clarification de l'échantillon a été réalisée par sonication sur glace (trois impulsions de 5 s), et les lysats ont été centrifugés à 14 000 g à 4 ° C pendant 15 min. Les concentrations de protéines ont été déterminées à l'aide du kit de quantification des protéines FluoroProfile (Sigma), en utilisant de la *bovin serum albumin* BSA pour la gamme, et des quantités égales de protéines (50  $\mu$ g par piste) ont été séparés sur un gel SDS-PAGE contenant 10% d'acrylamide. Les protéines ont ensuite été transférées sur des membranes de PVDF. Après blocage avec 5% de BSA dans du TBS (0,1 M, pH 7,4), les membranes ont été incubées avec les anticorps primaires (dilution 1 :500) à 4 ° C pendant la nuit. L'abondance de la  $\beta$ -tubuline a été utilisée comme contrôle de charge et a été déterminée avec un anticorps polyclonal de lapin anti- $\beta$ -tubuline (H-235) (sc-9104, Santa Cruz Biotechnology Inc.). Les membranes ont été incubées avec l'anticorps secondaire correspondant à l'anticorps primaire, et dilué avec 1% de BSA à température ambiante, pendant une heure. Après cette étape, les membranes ont été lavées trois fois avec 0,05% de Tween, dans du TBS. La membrane a été sondée avec les anticorps indiqués et révélée avec l'ECL.

## 1.11 CHIP

Les cellules HT29 ont été fixées avec 1% de formaldéhyde. Après 10 minutes, les cellules ont été lavées avec du PBS froid et lysées avec 500  $\mu$ l de tampon CHIP (50 mM de Tris-HCl, pH 8,1, 1% de SDS, 10 mM EDTA, 1 mM PMSF, 10  $\mu$ g/mL d'aprotinine, 10  $\mu$ g/mL de leupeptine, 10  $\mu$ g/mL pepstatine, 1 mM  $\text{Na}_3\text{VO}_4$ , 50 mM NaF). La chromatine a été fragmentée par sonication en fragments ayant une taille moyenne de 500 pb. La solution contenant la chromatine est diluée avec 1 volume de tampon de dilution (2 mM d'EDTA, 20 mM de Tris-HCl, pH 8,1, 1% de Triton X-100, 0,1% de NP-40, 150 mM de NaCl, 1 mM PMSF, 10  $\mu$ g/ml aprotinine, 10  $\mu$ g/ml de leupeptine, 10  $\mu$ g/ml pepstatine, 1 mM  $\text{Na}_3\text{VO}_4$ , 50mM NaF) et incubée pendant une heure à 4 ° C sur la roue avec de la protéine A-agarose et de la protéine G-Sépharose prétraitée avec de l'ADN de sperme de saumon. La chromatine a ensuite été incubée pendant une nuit à 4 ° C sur la roue avec 1  $\mu$ g des anticorps spécifiques ou d'anticorps dirigés contre la protéine bactérienne GAL4. Après précipitation avec la protéine A-agarose et protéine G-Sépharose (prétraitées avec l'ADN de sperme de saumon), la chromatine a été éluée avec un tampon d'éluion (1% SDS, 100 mM  $\text{NaHCO}_3$ ) pendant 5 heures à 65 ° C. L'ADN a ensuite été extrait avec du phénol chloroforme, puis précipité avec de l'éthanol. Les culots ont alors été mis à sécher à l'air, puis dissous dans 100  $\mu$ l d'eau stérile. Ce sont ensuite 4  $\mu$ l qui ont été utilisés pour réaliser les réactions de PCR.

## 1.12 Les modèles cellulaires

Afin d'étudier les rôles de l'OLFM4, nous avons utilisé plusieurs modèles cellulaires. Dans un premier temps, nous avons créé un modèle de cellules d'adénocarcinome colorectal HCT116 (ATCC : CCL-247) dont l'expression de l'OLFM4 était sous la dépendance d'un promoteur inductible à la doxycycline. Ces cellules n'expriment pas d'OLFM4 au niveau basal.

À l'inverse, nous avons pu réaliser des expériences d'ARN interférence dans une autre lignée d'adénocarcinome colorectal, les LS180 (ATCC : CL-187) qui elles expriment fortement l'OLFM4, à la fois au niveau transcriptionnel et traductionnel.

L'étude du lien entre l'OLFM4 et Ras a pu être réalisée grâce à un modèle présent au laboratoire de cellules d'adénocarcinome colique HT29 (ATCC : HTB-38) dont l'expression de l'oncogène H-RAS<sup>V12</sup> était sous la dépendance d'un promoteur inductible à la

doxycycline.

Afin de produire des quantités importantes d'OLFM4 nous avons également utilisé des cellules COS-7 (ATCC : CRL-1651). Elles sont issues du rein de *Cercopithecus aethiops* et permettent de produire des quantités importantes de protéines. Ces cellules ont pu être transfectées avec un vecteur d'expression codant pour l'OLFM4. La sécrétion de la protéine nous a permis d'en récupérer d'importantes quantités dans les surnageants de culture.

Nous avons également utilisé d'autres lignées que nous avons au laboratoire afin d'étudier leur protéome. Nous avons de la sorte analysé le protéome des cellules LS174T (ATCC : CL-188). Il s'agit de cellules d'adénome colorectal issues d'une patiente de 58 ans.

Nous avons étudié les cellules RKO (ATCC : CRL-2577) issues d'un carcinome colique.

Les cellules SW48 (ATCC : CCL-231) ont été analysées. Ce sont des cellules issues de l'adénocarcinome d'une patiente de 82 ans.

Des cellules SW1417 (ATCC : CCL-238) ont été étudiées. Elles sont issues d'une patiente de 53 ans.

Les cellules SW1463 (ATCC : CCL-234) sont issues d'un adénocarcinome d'une patiente de 66 ans.

Les cellules SW480 (ATCC : CCL-228) et SW620 (ATCC : CCL-227) sont issues du même patient. La lignée SW620 a été générée à partir d'un site métastatique de ce patient.

### 1.13 Conditions de culture

Les cellules présentées au paragraphe précédent étaient maintenues en milieu RPMI-1640 (Lonza) sans antibiotique, et complété avec 10% de sérum de veau fœtal (SVF). Les cellules étaient passées lorsque la confluence atteignait environ 80% et elles étaient diluées au 1/6. Chaque début de semaine l'absence de contamination par des mycoplasmes était contrôlée à l'aide d'un kit MycoAlert (Lonza). Les cellules étaient cultivées dans une atmosphère contenant 5% de CO<sub>2</sub> et à une température de 37 ° C.

Dans le cas des analyses de surnageants, les cellules étaient généralement maintenues 24 heures en condition de culture normale. Puis le milieu RPMI-1640 contenant 10% de SVF était éliminé. Deux lavages étaient alors réalisés avec du PBS, puis le milieu était remplacé par du RPMI-1640 ne contenant pas de SVF. Les cellules ont été ainsi maintenues dans l'incubateur pendant 24 heures et le milieu a alors pu être récupéré.

## 1.14 Préparation des extraits protéiques cellulaires

Dans des boîtes de culture de 15 cm de diamètre, 20 millions de cellules ont été ensemencées. Après 48h de culture en condition normale, les boîtes ont été récupérées, et le milieu aspiré. Un lavage PBS a été réalisé puis 10 mL ont alors été ajouté dans lesquels les cellules ont été grattées mécaniquement. Elles ont ensuite été placées dans un tube de 15 mL puis culotées par centrifugation et transférées dans un tube eppendorf de 1,5 mL.

Environ 300  $\mu$ L de tampon de lyse FASP (0,1 M Tris-HCl, 0,1 M DTT, 4% SDS pH 7,3) ont alors été ajoutés afin de lyser les cellules. Le tube a été chauffé 5 minutes à 95 ° C puis soniqué trois fois 20 secondes toutes les 20 secondes. Une centrifugation à 14 000 g a ensuite été réalisée afin de culoter les débris potentiels. Le surnageant était alors prêt pour le FASP.

## 1.15 ARN interférence

Afin d'inhiber l'expression de l'OLFM4 dans les cellules LS180, un pool de deux siRNA (20 pM chacun) dirigé contre l'OLFM4 a été transfecté avec l'agent de transfection Dharmafect-4 (Dharmacon) selon les instructions du fabricant.

Les séquences de ces siRNA sont les suivantes :

- 1 : ACCUAACUGUCCGAAUUGA (19pb)
- 2 : AUAUGACUCUCUUGGUAGA (19pb)

Une suspension contenant 800 000 cellules par mL de RPMI-1640 a été préparée et 800  $\mu$ L de cette dernière ont été déposés au fond du puits. Pendant ce temps, une solution contenant 4  $\mu$ L de Dharmafect-4 et 2  $\mu$ L de chacun des deux siRNA ont été incubés pendant 20 minutes à température ambiante dans 800  $\mu$ L d'OptiMEM. Ces 800  $\mu$ L ont ensuite été déposés goutte à goutte sur la suspension cellulaire. Au bout de 24 heures de culture, le milieu était changé par du RPMI-1640 neuf avec 10% de SVF.

## 1.16 Formation des sphères

Dans une plaque 96 puits, 70  $\mu$ L de RPMI-1640 2% agar chaud étaient déposés au fond des puits. Ces puits étaient remplis de 100  $\mu$ L de RPMI-1640. Des gouttes de 10  $\mu$ L contenant 2 000 cellules étaient déposées dans le couvercle de la plaque, bien au centre de chaque délimitation des puits. Le couvercle était ensuite posé délicatement sur la plaque pendant 48h. Les cellules sédimentaient dans la goutte s'agglomérant les unes aux autres.

Au bout de 48h, les puits étaient vidés de leur contenu et la plaque tapée contre la paille afin de faire descendre la structure trimentionnelle produite. Les puits étaient alors remplis avec 100  $\mu$ L de RPMI-1640 contenant 10% de SVF. Les cellules étaient ainsi mises en culture pendant 7 jours.

Des photos de chaque puits ont ensuite été réalisées et analysées grâce au logiciel ImageJ. Il s'agit d'un logiciel disponible en open source à l'adresse suivante : <http://rsbweb.nih.gov/ij/><sup>1</sup>.

## 1.17 Les tests ELISA de l'OLFM4

Les tests ELISA de l'OLFM4 ont été réalisés sur des sérums de patients. Ils ont été réalisés à l'aide d'un kit commercial selon les recommandations du fabricant (SEA162Hu, USCN Life Science Inc. Wuhan, Chine).

Nous avons pu doser l'OLFM4 chez 60 patients sains et chez 14 patients atteints de polypes pour tester notre hypothèse de candidat biomarqueur précoce.

Nous avons également réalisé ces dosages chez 45 patients différents afin de tester l'hypothèse de l'OLFM4 comme protéine dérégulée suite à la mutation de KRAS dans les tumeurs.

Les séra ont été conservés à -80 ° C directement après leur prélèvement, et aliquotés afin de n'être décongelés qu'une seule fois pour les tests.

Une fois décongelé, le sérum était dilué au 1/20<sup>e</sup> dans du PBS selon les préconisations du fabricant.

## 1.18 Tests de clonogénicité

Les cellules ont étéensemencées dans des plaques 6 puits à raison de 300 cellules par puit. Après adhésion des cellules au fond de la plaque (environ 6h) l'induction de l'expression de l'OLFM4 a été permise par l'ajout de doxycycline à 200 ng/ml final. Le traitement au Sn38 était réalisé de façon concomitante à une concentration de 0,5 ng/ml.

## 1.19 Le FASP

Les échantillon ont été lysés dans un tampon de lyse contenant du DTT (dithiothreitol) à 100mM, du TRIS (tris(hydroxyméthyl)aminométhane) et 4% de SDS (Sodium dodecyl

---

1. Page consultée le 10 juin 2013

sulfate). La présence du DTT dans le tampon de lyse permet de réaliser la réduction des ponts disulfures dès les premières étapes de lyse.

Lorsque l'échantillon était un lysat de lignées cellulaires, il était chauffé pendant 5 minutes à 95 ° C pour permettre une lyse efficace. Dans le cas des échantillons de tumeurs obtenus à partir de tissus congelés, l'échantillon lysé était chauffé pendant 90 minutes pour dénaturer complètement la structure du tissu.

L'échantillon était ensuite soniqué pour être clarifié, puis le lysat était placé dans un centricon contenant une membrane ayant la capacité de laisser passer les molécules ayant un poids moléculaire inférieur à une valeur consigne choisie. Dans notre cas, nous avons choisi une valeur consigne de 30kDa (Millipore) car l'équipe de Wisniewsky a montré qu'il s'agissait de la meilleure solution pour l'étude des glycopeptides [Wisniewski et al., 2011b].

Des lavages (6 fois 200  $\mu$ L) avec un tampon contenant 8M d'urée, 0,1 M de Tris-HCl pH 8,5 et donc un pouvoir chaotropique important permettaient d'évacuer le SDS. Ces lavages ont été permis par des centrifugations de 6 minutes à 14 000 g.

C'est également dans le tampon urée que l'alkylation des groupements thiols libres eu lieu avec l'ajout de 50mM de iodoacétamide (GE Healthcare) ou de 10mM de méthylméthane thiosulfonate (MMTS) (Sigma).

Puis des lavages (6 fois 200 $\mu$ L) ont à nouveau été réalisés pour évacuer le réactif chimique alkylant à l'aide du tampon contenant l'urée. Ce tampon était alors remplacé progressivement (6 lavages de 200  $\mu$ L) par un tampon compatible avec l'action de la trypsine tel que le bicarbonate d'ammonium 50 mM à pH 8 ou le TEAB (Sigma) 0,5 M pH 8,5 (Triethylammonium bicarbonate). Puis les protéines étaient digérées avec la trypsine (10  $\mu$ g pour 100  $\mu$ g de peptides). Les peptides générés étaient alors capables de passer à travers le filtre et pouvaient ainsi être récupérés par une simple centrifugation dans un nouveau tube collecteur.

Les peptides ont ensuite été dosés à l'aide du kit microBCA (Pierce) en comparaison à une gamme de BSA.

## 1.20 L'enrichissement en glycopeptides

### 1.20.1 Enrichissement sur lectines

Les peptides issus du FASP (800  $\mu$ g) ont été dessalés sur Sep-Pack sur une colonne de C18 (Waters) puis les culots secs ont été ressolubilisés dans 400  $\mu$ L d'un tampon de

fixation aux lectines (1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MnCl}_2$ , 0,5 M NaCl in 20 mM Tris-HCl pH 7,3). Ensuite, 800  $\mu\text{g}$  de concanavaleine A et 800  $\mu\text{g}$  de WGA ont été ajoutés dans le tube, qui a ensuite été placé à 4 ° C pour la nuit. Cette étape permet la fixation des peptides glycosylés sur les lectines.

L'ensemble du contenu du tube est ensuite placé sur un centricon de 30 kDa afin de permettre les rinçages. Deux lavages au tampon de fixation et deux lavages au bicarbonate d'ammonium 50 mM pH 8 repris dans l'eau lourde ont été réalisés. Puis 3 unités de PNGase F ont été ajoutés et l'ensemble a été incubé 4h à 37 ° C pour permettre la déglycosylation et le relargage des peptides glycosylés qui ont ensuite pu être récupérés par centrifugation.

Les peptides obtenus ont été séparés par la technique OFFGEL en douze fractions allant d'un pH 3 à un pH 10.

### 1.20.2 Enrichissement sur billes hydrazines

Au cours de la première étape, les glycopeptides sont oxydés grâce à du méta-périodate de sodium à 15 mM final sur les résidus sucrés. Les fonctions *cis-diols* deviennent alors des fonctions aldéhydes. Les fonctions aldéhydes libres sont extrêmement réactives et elles vont réagir sur les fonctions amines ( $\text{NH}_2$ ) des hydrazines (Bio-Rad) pour former une liaison covalente. Les protéines ou peptides non liés peuvent alors être élués en réalisant des lavages, ce qui permet de ne garder que ce qui est fixé sur les billes. L'action d'une enzyme de déglycosylation comme la PNGaseF (P7367, Sigma-Aldrich) permet ensuite de libérer les peptides fixés pour pouvoir ensuite les analyser par spectrométrie de masse. Dans nos tests, nous avons utilisé 3 unités de l'enzyme pour 100  $\mu\text{g}$  de peptides au départ.

## 1.21 La mutagenèse dirigée de l'OLFM4

Nous avons à notre disposition un plasmide d'expression pCMV6-XL4 codant pour l'OLFM4 (SC108914, Origene). Nous avons donc utilisé un kit de mutagenèse dirigée afin de réaliser l'insertion des mutations dans ces plasmides (Applied biosystem).

Les asparagines ont toutes été substituées par des alanines. Les séquences des amorces sont les suivantes :

Tableau 1.1 – Amorces de mutagénèse utilisées pour muter l'OLFM4

Site	Amorce sens	Amorce antisens
72	5'-tgtcccagttgtttccgctttcaccggtccgtgg-3'	5'-ccacggagccggtgaaagcggaaaacaactgggaca-3'
253	5'-gtcatggtggtgtggtggccatcagcaaaccgtctg-3'	5'-cagacggtttgctgatggccaccacaccatgac-3'
352	5'-cgggaatattgccagagttgccctgaccaccaacacga-3'	5'-tcgtgttggtggtcagggcaactctggcaatattcccg-3'
356	5'-cagagttaacctgaccaccgcccagattgctgtgactcaa-3'	5'-ttgagtcacagcaatcgtggcggtggtcaggttaactctg-3'

## 1.22 L'étude du stroma tumoral

Pour cette étude, nous avons analysé deux tumeurs colorectales. Les coupes ont été déshydratées selon le même protocole que pour l'analyse des tumeurs. Nous avons généré 50 capsules par tumeurs sur lesquelles nous avons microdisséqué uniquement les zones correspondant au stroma tumoral. Cette microdissection a été réalisée sous le contrôle d'un médecin anatomopathologiste.

Les cellules fixées sur les capsules ont alors été lysées dans 500  $\mu$ L de tampon de FASP, chauffées 90 minutes à 95 ° C et centrifugées pour récupérer le surnageant contenant les protéines. Le protocole FASP a alors pû être appliqué.



## Objectifs de l'étude et résumé de l'approche utilisée

LA compréhension de plus en plus détaillée des mécanismes aboutissant aux cancers conduit à mettre en exergue le fait que chaque tumeur possède ses caractéristiques propres, et ce y compris pour une même localisation tissulaire. C'est dans ce contexte que le concept de médecine personnalisée a fait son apparition, et a connu un véritable essor ces dernières années. Aujourd'hui, cette approche commence à devenir une réalité quotidienne. De plus, une meilleure compréhension des processus physiopathologiques permet d'optimiser le choix des médicaments ainsi que le dosage. Cette optimisation n'est pas sans conséquences, et elle est bénéfique pour l'ensemble de la société. En effet, l'administration d'un médicament a un coût non négligeable, de sorte qu'utiliser un médicament dont on aurait pu prévoir qu'il serait inefficace chez un patient constitue un gaspillage qu'il peut être intéressant de limiter. L'apport de biomarqueurs, permettant de réaliser des diagnostics moléculaires précis est donc un enjeu de taille. C'est dans ce cadre que s'inscrivent nos travaux, dont les résultats sont présentés dans les lignes qui suivent.

Pour cela, nous avons souhaité réaliser une étude protéomique globale de la tumeur du côlon avec l'objectif de mettre en évidence des biomarqueurs précoces potentiels mais également des biomarqueurs en lien avec l'activation de l'oncogène KRAS. Afin d'étendre cette étude, nous avons également analysé des lignées cellulaires, ainsi que, pour certaines d'entre elles, le sécrétome associé. Nous avons également souhaité étudier le sécrétome de tumeurs en analysant non pas la tumeur, mais le stroma tumoral. Enfin, nous avons voulu développer une approche d'analyse du glycoprotéome afin de mettre en évidence des candidats biomarqueurs susceptibles d'être dosés dans le sang des patients. À la suite de cette étude, nous avons proposé une liste de protéines sécrétées par les tumeurs ou les lignées colorectales, pouvant constituer une base de données de recherche de biomarqueurs potentiels.

La première partie de ce travail nous a conduit à étudier le protéome global de tumeurs colorectales en fonction de leur stade histologique, afin de rechercher des biomarqueurs

potentiels précoces. Cela nous a conduit à identifier plusieurs marqueurs potentiels dont un, l'OLFM4 a fait l'objet d'une publication. Nous avons également recalculé nos expériences iTRAQ<sup>TM</sup> en fonction du statut KRAS des patients pour nous apercevoir que notre protéine d'intérêt était effectivement augmentée dans les tumeurs mutées sur KRAS. Nous avons alors souhaité savoir si cette protéine était détectable dans le sérum des patients avant toute analyse complémentaire.

La deuxième partie de ces travaux nous a conduit à nous intéresser plus en profondeur aux protéines sécrétées dans un contexte de biomarqueurs. Pour cela, nous avons analysé l'environnement cellulaire immédiat de la tumeur, le stroma tumoral, dans lequel se situe probablement une quantité importante des protéines sécrétées par la tumeur.

Au cours de la troisième partie, nous avons analysé en détail le protéome de lignées colorectales, et également leur sécrétome. L'objectif était de permettre de détecter des protéines sécrétées susceptibles d'être des biomarqueurs potentiels et qui auraient échappé aux analyses des tumeurs et de leur environnement.

Dans la dernière partie de ce travail, nous nous sommes intéressés aux protéines glycosylées présentes dans les tumeurs colorectales. Ces protéines présentent des avantages importants en terme de stabilité dans les milieux biologiques leur assurant une demie-vie suffisante pour être accumulées dans le sang des patients, et détectées par des méthodes de routine telle que l'ELISA [Russell et al., 2009]. Nous avons donc souhaité mettre en place une méthode d'analyse globale permettant de les étudier plus spécifiquement afin d'améliorer nos chances de détection lors des expériences d'analyse globale.

Par ailleurs, toujours dans l'objectif d'identifier le protéome le plus large possible, nous avons réalisé une mise au point technologique basée sur une diminution de la complexité de l'échantillon par OFFGEL séquentiels. Cette approche technologique a fait l'objet d'une publication.

La figure 1.2 p.113 présente un schéma récapitulatif de notre approche.

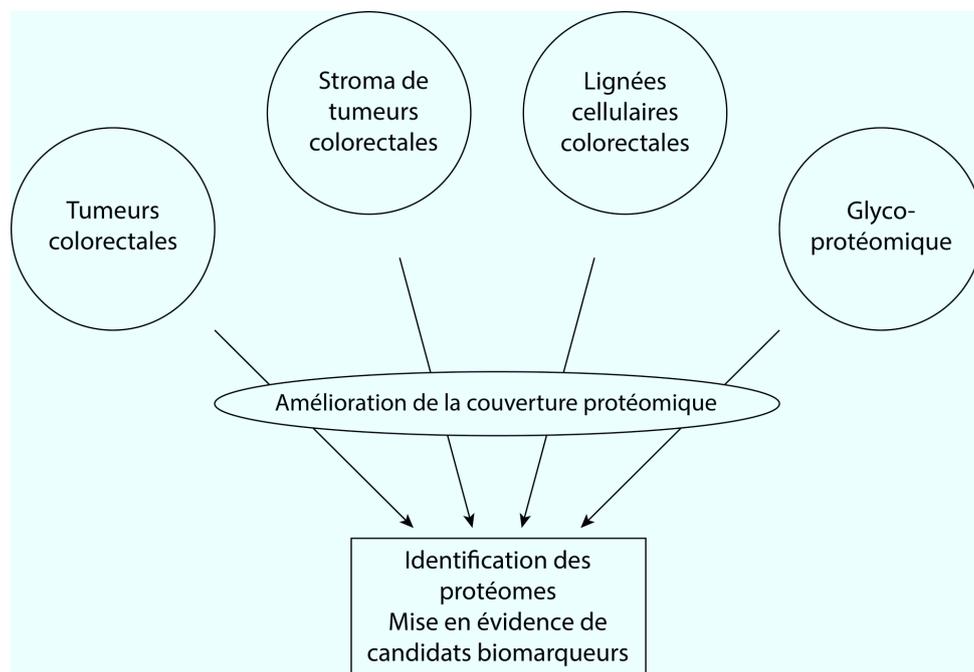


FIGURE 1.2 – Schéma de l'approche utilisée pour tenter de mettre en évidence des candidats biomarqueurs.



## Troisième partie

### Résultats



# Chapitre 1

## Analyse protéomique globale de tumeurs colorectales en fonction de leur stade histologique

LES résultats de notre étude ont fait l'objet d'une publication dans la revue *Molecular and cellular proteomics*. La référence est la suivante :

*A Quantitative Proteomic Approach of the Different Stages of Colorectal Cancer Establishes OLFM4 as a New Nonmetastatic Tumor Marker.*

Damien Besson, Aude-Hélène Pavageau, Isabelle Valo, Anthony Bourreau, Audrey Bélanger, Caroline Eymerit-Morin, Alice Moulière, Agnès Chassevent, Michelle Boisdron-Celle, Alain Morel, Jérôme Solassol, Mario Campone, Erick Gamelin, Benjamin Barré, Olivier Coqueret, and Catherine Guette.

*Molecular and Cellular Proteomics* 10 : 10.1074/mcp.M111.009712, 1–14, 2011.

### 1.1 Problématique

Bien qu'utiles, les différentes classifications histologiques du cancer colorectal montrent des limites dans leur utilisation comme élément de choix du traitement. Elles ne permettent pas d'établir un profil moléculaire de la tumeur et ne donnent pas d'information concernant d'éventuelles résistances aux traitements. En outre, parmi les patients présentant un cancer de stade 1 ou 2 traités par chirurgie, 15 à 20% vont développer une récurrence de la pathologie [Ho et al., 2004]. Cela montre bien les limites de ces modes de traitements basés sur l'utilisation de la classification histologique. Nous avons donc souhaité mettre en place une approche nous permettant de caractériser les tumeurs colorectales au niveau moléculaire et ce, en fonction des différents stades cliniques afin de tenter de dégager un ou plusieurs biomarqueurs potentiels qui pourraient être recherchés chez les patients pour

déterminer la présence d'un cancer précoce.

Pour répondre à notre problématique, nous avons développé une approche protéomique basée sur l'utilisation du réactif iTRAQ<sup>TM</sup> dont la structure chimique est présentée en figure 3.17 p. 73 [Ross et al., 2004]. Les échantillons à notre disposition n'étaient pas tous associés à leur tissu sain. Nous avons donc fait le choix de prendre comme référence un pool de trois tissus sains, toujours les mêmes pour chaque expérience. Nous avons alors pu mettre en évidence l'OLFM4 comme biomarqueur diagnostic potentiel du cancer colorectal. Les résultats complet de cet analyse par stade clinique sont présentés dans le tableau 1 joint en annexe de ce document p.213. Par ailleurs, représentation graphique de ces résultats sous la forme d'une carte de variation est présentée en figure 1.1 p.133.

## 1.2 Publication

## Research

© 2011 by The American Society for Biochemistry and Molecular Biology, Inc.  
This paper is available on line at <http://www.mcponline.org>

# A Quantitative Proteomic Approach of the Different Stages of Colorectal Cancer Establishes OLFM4 as a New Nonmetastatic Tumor Marker\*

Damien Besson‡, Aude-Hélène Pavageau‡, Isabelle Valo§, Anthony Bourreau‡, Audrey Bélanger‡, Caroline Eymerit-Morin‡, Alice Moulière‡, Agnès Chassevent§, Michelle Boisdron-Celle‡, Alain Morel‡, Jérôme Solassol¶, Mario Campone||, Erick Gamelin‡, Benjamin Barré‡, Olivier Coqueret‡, and Catherine Guette‡\*\*

Expression profiles represent new molecular tools that are useful to characterize the successive steps of tumor progression and the prediction of recurrence or chemotherapy response. In this study, we have used quantitative proteomic analysis to compare different stages of colorectal cancer. A combination of laser microdissection, OFFGEL separation, iTRAQ labeling, and MALDI-TOF/TOF MS was used to explore the proteome of 28 colorectal cancer tissues. Two software packages were used for identification and quantification of differentially expressed proteins: Protein Pilot and iQuantator. Based on ~1,190,702 MS/MS spectra, a total of 3138 proteins were identified, which represents the largest database of colorectal cancer realized to date and demonstrates the value of our quantitative proteomic approach. In this way, individual protein expression and variation have been identified for each patient and for each colorectal dysplasia and cancer stage (stages I–IV). A total of 555 proteins presenting a significant fold change were quantified in the different stages, and this differential expression correlated with immunohistochemistry results reported in the Human Protein Atlas database. To identify a candidate biomarker of the early stages of colorectal cancer, we focused our study on secreted proteins. In this way, we identified olfactomedin-4, which was overexpressed in adenomas and in early stages of colorectal tumors. This early stage overexpression was confirmed by immunohistochemistry in 126 paraffin-embedded tissues. Our results also indicate that OLFM4 is regulated by the Ras-NF- $\kappa$ B2 pathway, one of the main oncogenic pathways deregulated

in colorectal tumors. *Molecular & Cellular Proteomics* 10: 10.1074/mcp.M111.009712, 1–14, 2011.

Every year, more than one million individuals around the world are diagnosed with colorectal cancer (CRC),<sup>1</sup> and with a death rate ~33% (1), this disease is an important cause of mortality. CRC diagnosis and prognosis rely on the tumor-node-metastasis and clinical staging systems, which illustrate local lymph node and distal organ invasion. These clinical stages are important prognostic factors because survival rates of 5 years or more are observed for more than 90% of patients diagnosed with Stage I CCR, whereas survival rates drop to only ~10% for CRC that have metastasized to distant organs (stage IV) (2). As a consequence, early stage detection has the most impact on cancer incidence and mortality in this disease (3, 4). As initially described by Vogelstein *et al.* (5), colorectal transformation is explained by the sequential accumulation of genetic alterations that generate malignant cells (6). Mutations of the adenomatous polyposis coli gene and the subsequent activation of  $\beta$ -catenin is probably the most common initiating event of CRC, leading to the transformation of normal colonic epithelium into adenomas (7–10). This stage represents an intermediate lesion where cells exhibit autonomous growth and probably genetic instability but are incapable of invasive growth and metastasis. It is estimated that only a small proportion of ~5% of adenomas will progress to the next CRC stages, implying that the transition from normal cells to adenoma differs from the progression from adenomas to adenocarcinomas. Following the loss of adenomatous pol-

From the ‡Institut de Cancérologie de l'Ouest, Paul Papin Cancer Center; INSERM U892, Angers, France, the §Institut de Cancérologie de l'Ouest, Paul Papin Cancer Center Paul Papin; Angers, France, the ¶Department of Cellular Biology, CHU Arnaud de Villeneuve, Montpellier, France, and the ||Institut de Cancérologie de l'Ouest, René Gauducheau Cancer Center, Nantes, France

Received March 24, 2011, and in revised form, September 29, 2011  
Published, MCP Papers in Press, October 10, 2011, DOI 10.1074/mcp.M111.009712

<sup>1</sup> The abbreviations used are: CRC, colorectal cancer; IHC, immunohistochemistry; HPA, Human Protein Atlas; LCM, laser capture microdissection; FDR, false discovery rate; GO, gene ontology; CHAPS, 3-[[3-(cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; ChIP, chromatin immunoprecipitation; STAT, signal transducers and activators of transcription; OIS, oncogene-induced senescence.

### Quantitative Proteomic Analysis of Colorectal Tumors

yposis coli, it has been proposed recently that KRAS mutations are essential to allow the nuclear accumulation of  $\beta$ -catenin and the subsequent progression to the adenocarcinoma step (11). Although Ras mutations have probably no prognosis value, this modification is associated with resistance to anti-epidermal growth factor receptor-targeted therapies (12), indicating that this transformation pathway is associated with an intrinsic drug resistance program.

Despite their utility, tumor-node-metastasis and clinical staging remain relatively imprecise and are not well characterized at the molecular level. For this reason, the development of new prognostic tools would be useful to characterize the successive steps of the disease and predict the risk or recurrence of chemotherapy escape. Elegant results have recently identified gene expression profiles that associate with specific oncogenic pathways and can eventually predict chemotherapy sensitivity (13, 14). This approach has also been successfully used in CRC to identify a 50-gene signature that distinguished patients with low or high risks or recurrence at the early stage of the disease (15). These results also led to the identification of a therapeutic approach that would be specific of a particular stage of CRC. Besides these genomic data, proteomics analysis is also a powerful tool for the global evaluation of protein expression and the identification of prognosis or predictive signatures. However, although recent in-depth proteomics analyses have generated large protein data sets, only a few proteins such as carcinoembryonic antigen, CA19.9, and CA125 have been described as potential prognosis or prognostic biomarkers, and none of them are recommended for clinical screening (16–20). These analyses essentially used two-dimensional gel strategies combined with image analysis, thereby limiting the analyses to the more abundant CRC proteins. A few recent studies have combined more targeted approaches with two-dimensional electrophoresis, including studies focusing on membrane proteins (21), basic proteins (22), heparin-affinity isolated proteins (23), or proteasome (24). Some of these studies have identified novel candidate CRC serum biomarkers with comparable or better sensitivity than carcinoembryonic antigen, such as nicotinamide *N*-methyltransferase (25), proteasome activator complex subunit, PSME3 (26), S100A9, S100A8 (27), and Desmin (28). However, these results are often largely limited to abundant proteins that are commonly overexpressed in cancers (structural proteins, glycolytic enzymes, annexins, cathepsins, and heat shock protein). Because these proteins are probably not specific to CRC, the benefits in clinical staging or in predicting the success of targeted therapies remain to be determined. Tumor analysis is challenging, given the heterogeneity of the colorectal cancer tissue and the limited number of tumor cells generally available. To increase the specificity of the analysis, tumor cells should be ideally isolated from heterogeneous samples by laser capture microdissection (LCM) (29). Although this approach is widely used in the genomic field, proteomic analyses of tumor tissue are limited,

TABLE I  
Clinical features of human tissue samples

Clinical features	iTRAQ	IHC
Gender (male/female)	16/12	70/56
Mean age (years)	68.7 ± 8.9	69.39 ± 10.56
Age range (years)	48–87	41–94
Union Internationale contre le Cancer stage		
Adenoma	4 (14.3%)	30 (23.8%)
Intramucosal carcinoma	0	12 (9.5%)
<b>Adenocarcinoma</b>		
Stage I	6 (21.5%)	26 (20.6%)
Stage II	5 (17.8%)	14 (11.1%)
Stage III	9 (32.1%)	25 (19.8%)
Stage IV	4 (14.3%)	19 (15.2%)
<b>Location</b>		
Colon	17 (58.6%)	92 (73.0)
Rectum	1 (3.4%)	10 (7.9%)
Rectosigmoid hinge	6 (20.7%)	9 (7.2%)
Sigmoid	3 (10.4%)	10 (7.9%)
Caecum	2 (6.9%)	5 (4.0%)

and to our knowledge only one study has been performed in CRC, using adenoma tissue (29).

In this study, we have performed comparative proteomic profiling of laser capture microdissected adenoma, stage I–IV adenocarcinoma tissues, and normal colorectal tissues from 28 different human specimens. Using only the limited amount of material collected by LCM, a proteomic profile for each sample has been separately acquired by iTRAQ labeling using our previously developed method (30, 31). 555 proteins were quantified in the different stages. Among these proteins, we focused on OLFM4 (olfactomedin-4), because this protein has recently been shown to be expressed in colorectal stem cells in association with Lgr5 (32, 33). We found that OLFM4 is a secreted protein that is regulated by Ras-NF- $\kappa$ B2, one of the main oncogenic pathway in colorectal cells. These results highlight the power of quantitative proteomic approaches to allow the identification of stage-specific markers in colorectal cancer. To our knowledge, this is the most comprehensive global study that compares proteomic results of phenotypically normal, adenoma, early stage tumor, and metastatic adenocarcinoma in multiple individual samples.

#### MATERIALS AND METHODS

**Sample Collection**—The study protocol and patient consent forms were approved by the Angers Hospital and CRLCC Paul Papin Ethic Committee. Twenty-eight colorectal frozen tissue samples, collected between 1999 and 2004, were obtained from the Cancer Center tumor bank. All of the tissue specimens were obtained from surgical resection. Normal colonic tissue was obtained from the distal edge of the resection at least 10 cm from the tumor. After hematoxylin and eosin staining, the paraffin-embedded tissue sections from all of the specimens were evaluated by two experienced pathologists independently according to Union Internationale contre le Cancer staging. Clinical features of tissue candidates are summarized in Table I.

**Quantitative Proteomic Analysis of Colorectal Tumors**

**Laser Capture Microdissection**—Frozen sections (12  $\mu\text{m}$  thick) of either colon cancer or normal colonic mucosa were cut on a cryostat (Bright Instrument Co. Ltd., Huntingdon, UK). Specific sections were stained with toluidine blue for visual reference. Tissue sections were incubated in 70% ethanol for 1 min, 95% ethanol for 2 min, 95% ethanol for 2 min, and finally twice in 100% xylene for 5 min. Xylene was evaporated, and sections were microdissected using a PixCell II laser capture microdissection system (Arcturus Engineering, Mountain View, CA) equipped with the PixCell II image archiving software (Arcturus Engineering). Laser settings were as follows:  $\lambda = 810$  nm, spot diameter set at 7.5  $\mu\text{m}$ , pulse duration = 70 ms, and power = 70 mW. After microdissection, the plastic film containing the microdissected cells was removed, the film containing the tumor cells was placed in a microcentrifuge tube, and the protein lysis solution was added. Approximately 30,000 cells were captured from either a single or consecutive tissue sections using up to five CapSure LCM caps (Molecular Devices Corporation).

**Protein Extraction and Digestion**—Protein extraction was carried out using the Liquid Tissue MS protein preparation kit according to manufacturer's protocol (Expression Pathology Inc., Gaithersburg, MD). Briefly, the films from the underside of the caps for all samples were removed, transferred to low binding reaction tubes, incubated with 20  $\mu\text{l}$  of Liquid Tissue extraction, and heated at 95  $^{\circ}\text{C}$  for 90 min. After cooling for 2 min on ice, 5  $\mu\text{l}$  of trypsin reagent was added and incubated at 37  $^{\circ}\text{C}$  for 1 h with vigorous shaking for 30 s at 20 min intervals. The samples were further incubated overnight at 37  $^{\circ}\text{C}$  followed by heating at 95  $^{\circ}\text{C}$  for 5 min. The samples were then harvested via centrifugation at 10,000  $\times g$ , dried completely using a SpeedVac, resuspended in 100  $\mu\text{l}$  of 0.5% TFA in 5% acetonitrile, desalted via PepClean C-18 spin columns (Pierce), and dried for iTRAQ<sup>TM</sup> processing.

**Peptide Labeling with iTRAQ Reagents**—Peptides samples were resuspended with 30  $\mu\text{l}$  of iTRAQ dissolution buffer (AB Sciex). They were reduced with 5 mM tris-(2-carboxyethyl)phosphine at 60  $^{\circ}\text{C}$  for 1 h, and the cysteine-groups were blocked using a 10 mM methyl methanethiosulfonate solution at room temperature for 10 min. Each peptide solution was labeled at room temperature for 2 h with one iTRAQ reagent vial previously reconstituted with 70  $\mu\text{l}$  of ethanol for 4-plex iTRAQ reagent and reconstituted with 50  $\mu\text{l}$  of isopropanol for 8-plex iTRAQ reagent. A mixture containing small aliquots from each labeled sample was analyzed by MS/MS to determine a proper mixing ratio to correct for unevenness in peptide yield from Liquid Tissues procedures. Labeled peptides were then mixed in a 1:1:1:1 (or 1:1:1:1:1:1:1:1) ratio. The peptide mixture was then dried completely using a SpeedVac.

**Peptide OFFGEL Fractionation**—For pl-based peptide separation, we used the 3100 OFFGEL Fractionator (Agilent Technologies, Böblingen, Germany) with a 24-well set-up. Prior to electrofocusing, the samples were desalted onto a Sep-Pak C18 cartridge (Waters). For a 24-well set-up, peptide samples were diluted to a final volume of respectively 3.6 ml using OFFGEL peptide sample solution. To start, the 24-cm-long IPG gel strip (GE Healthcare) with a 3–10 linear pH range was rehydrated with the peptide IPG strip rehydration solution according to the protocol of the manufacturer for 15 min. Then 150  $\mu\text{l}$  of sample was loaded in each well. Electrofocusing of the peptides was performed at 20  $^{\circ}\text{C}$  and 50  $\mu\text{A}$  until the 50-kVh level was reached. After focusing, the 24 peptide fractions were withdrawn, and the wells were washed with 200  $\mu\text{l}$  of a solution of water/methanol/formic acid (49:50:1). After 15 min, the washing solutions were pooled with their corresponding peptide fraction. All of the fractions were evaporated by centrifugation under vacuum and maintained at  $-20$   $^{\circ}\text{C}$ . Just prior nano-LC, the fractions were resuspended in 20  $\mu\text{l}$  of  $\text{H}_2\text{O}$  with 0.1% (v/v) TFA.

**Capillary LC Separation**—The samples were separated on an Ultimate 3,000 nano-LC system (Dionex, Sunnyvale, CA) using a C18

column (PepMap100, 3  $\mu\text{m}$ , 100 A, 75  $\mu\text{m}$  inner diameter  $\times$  15 cm; Dionex) at 300 nl/min a flow rate. Buffer A was 2% ACN in water with 0.05% TFA, and buffer B was 80% ACN in water with 0.04% TFA. The peptides were desalted for 3 min using only buffer A on the precolumn, followed by a separation for 105 min using the following gradient: 0 to 20% B in 10 min, 20% to 45% B in 85 min, and 45% to 100% B in 10 min. Chromatograms were recorded at the wavelength of 214 nm. Peptide fractions were collected using a Probot microfraction collector (Dionex). We used CHCA (LaserBioLabs, Sophia-Antipolis, France) as MALDI matrix. The matrix (concentration of 2 mg/ml in 70% ACN in water with 0.1% TFA) was continuously added to the column effluent via a micro T mixing piece at 1.2  $\mu\text{l}/\text{min}$  flow rate. After 12-min run, a start signal was sent to the Probot to initiate fractionation. Fractions were collected for 10 s and spotted on a MALDI sample plate (1,664 spots/plate; Applied Biosystems, Foster City, CA).

**MALDI-MS/MS**—MS and MS/MS analyses of off-line spotted peptide samples were performed using the 4800 or 5800 MALDI-TOF/TOF Analyzers (Applied Biosystems/ABSciex) and 4000 Series Explorer software, version 3.5 (with MALDI 4800) and version 4.0 with MALDI 5800). The instrument was operated in a positive ion mode and externally calibrated using a mass calibration standard kit (ABSciex). The laser power was set between 2800 and 3400 for MS and between 3600 and 4200 for MS/MS acquisition. After screening all LC-MALDI sample positions in MS-positive reflector mode using 1500 laser shots, the fragmentation of automatically selected precursors was performed at a collision energy of 1 kV using air as collision gas (pressure of  $\sim 2 \times 10^{-6}$  Torr) with an accumulation of 2000 shots for each spectrum. MS spectra were acquired between  $m/z$  800 and 4000. For internal calibration, we used the parent ion of Glu-1 fibrinopeptide at  $m/z$  1570.677 diluted in the matrix (30 fmol/spot). Up to 12 of the most intense ion signals per spot position having a S/N of  $>12$  were selected as precursors for MS/MS acquisition. Peptide and protein identification were performed by the ProteinPilot<sup>TM</sup> software version 3.0 (AB Sciex) using the Paragon algorithm as the search engine (34). Each MS/MS spectrum was searched for *Homo sapiens* species against the Uniprot/Swissprot database (UniProtKB/Sprot 20090414 release 15.0, with 525,997 sequence entries). The searches were run using with the fixed modification of methylmethanethiosulfate-labeled cysteine parameter enabled. Other parameters such as tryptic cleavage specificity, precursor ion mass accuracy and fragment ion mass accuracy are MALDI 4800 or 5800 built-in functions of ProteinPilot software. The detected protein threshold (unused protscore (confidence)) in the software was set to 2 to achieve 99% confidence, and identified proteins were grouped by the ProGroup algorithm (ABSciex) to minimize redundancy. The bias correction option was executed.

A decoy database (based on a reverse sequence database concatenated with the forward sequence database) search strategy was also used to estimate the false discovery rate (FDR = number of validated decoy hits/(number of validated target hits + number of validated decoy hits)  $\times$  100). The FDR was calculated by searching the spectral against the Uniprot *H. sapiens* decoy database. The FDR for each iTRAQ experiment is indicated in Table II.

**Quantification of Relative Protein Expression**—We employed a customized software package, iQuantitor (35, 36), to infer the magnitude of change in protein expression. The software infers treatment-dependent changes in expression using Bayesian statistical and Markov Chain Monte Carlo methods. Basically, this approach was used to generate means, medians, and 95% confidence intervals (upper and lower) for each treatment-dependent change in protein expression by using peptide level data for each component peptide and integrating data across the two experiments. For proteins whose iTRAQ ratios were down-regulated in tissues, the extent of down-

### Quantitative Proteomic Analysis of Colorectal Tumors

regulation was considered further if the higher limit of the confidence interval had a value lower than 1. Conversely, for proteins whose iTRAQ ratios were up-regulated in tumors, the extent of up-regulation was considered further if the lower limit of the confidence interval had a value greater than 1. The width of these credible intervals depends on the data available for a given protein. Because the number of peptides observed and the number of spectra used to quantify the change in expression for a given protein are taken into consideration, it is possible to detect small but significant changes in up- or down-regulation when many peptides are available. For each protein and each peptide associated with a given protein, the mean, median, and 95% credible intervals were computed for each of the protein and peptide level treatment effects.

The peptide selection criteria for relative quantification were performed as follows. Only peptides unique for a given protein were considered for relative quantification, excluding those common to other isoforms or proteins of the same family. Proteins were identified on the basis of having at least two peptides with an ion score above 95% confidence. The protein sequence coverage (95%) was estimated for specific proteins by the percentage of matching amino acids from the identified peptides having confidence greater than or equal to 95% divided by the total number of amino acids in the sequence.

**Functional Annotation and Network Analysis**—Gene ontology (GO) terms for identified proteins were extracted, and overrepresented functional categories for differentially abundant proteins were determined by the high throughput GOMiner tool (National Cancer Institute, <http://discover.nci.nih.gov.gate2.inist.fr/gominer/>) (37). All proteins that were subjected to iQuantitor analysis served as the background list, and GO terms with at least five proteins were used for statistical calculations. A *p* value for each term was calculated via the one-sided Fisher's exact test, and FDR was estimated by permutation analysis using 1000 randomly selected sets of proteins sampled from the background list. Statistically significant (FDR < 25%) GO terms were clustered based on the correlation of associated proteins to minimize potential redundancy in significant GO terms.

Network analyses of protein candidates and the ratios of their expression in tumor and nontumor tissues (obtained from eight independent experiments) were performed using the MetaCore™ analytical suite version 4.7 (GeneGo, Inc., St. Joseph, MI) and compared using *p* values of <0.01 as statistical metrics. For enrichment analysis, gene identifiers of the uploaded files were matched with gene identifiers in functional ontologies in MetaCore™ (38), which included canonical pathway maps (GeneGo maps), GeneGo cellular processes, GO cellular process, and diseases categories.

**Western Blot Analysis of Proteins in Normal Tissues and CRC**—Whole cell lysates were prepared from normal tissues and tumoral tissues. Frozen tissue samples were homogenized and lysed in a buffer containing 7 M urea, 2 M thiourea, and 4% (w/v) CHAPS at 4 °C for 1 h using a rotary shaker. Lysis was achieved by sonication on ice (three 5-s pulses), and the lysates were clarified by centrifugation at 14,000 × *g* at 4 °C for 15 min. Protein concentrations were determined using the FluoroProfile protein quantification kit (Sigma-Aldrich), with BSA as the standard, and equal amounts of proteins (80 μg/lane) from the samples tissues were resolved on a 10% SDS-polyacrylamide gel. The proteins were then electrotransferred onto PVDF membranes. After blocking with 3% BSA in TBS (0.1 M, pH 7.4), blots were incubated with the respective primary antibodies (1:200 dilution) at 4 °C overnight. The protein abundance of β-tubulin was used as a control for protein loading and was determined with rabbit polyclonal anti-β-tubulin: (H-235) antibody (sc-9104; Santa Cruz Biotechnology Inc.). The membranes were incubated with the respective secondary antibody, horseradish peroxidase-conjugated rabbit anti-IgG (goat anti-rabbit IgG, 1:5000; Santa Cruz Biotechnology Inc.), and diluted

with 1% bovine serum albumin for 1 h at room temperature. After each step, blots were washed three times with 0.05% Tween, TBS. The membrane was probed with the indicated antibodies and developed with the ECL.

**Immunohistochemistry**—One hundred twenty-six patients with adenoma or colorectal adenocarcinoma were studied by immunohistochemistry. All of the tumors were obtained from the Departments of Pathology at the Paul Papin Cancer Center and at the University Hospital of Angers and from the Center of Pathology of Angers between 2000 and 2005. Some samples were excluded: young patients (under 40 years old) and tumors having received chemotherapy, metastatic tumors. The location was almost colic or from high rectum. There were 30 adenomas with 15 low grade adenomas and 15 high grade adenomas; the dysplasia was classified according to the established criteria of architectural features and cytological atypia. There were 72 colorectal adenocarcinomas. According to the seventh tumor-node-metastasis staging system (39) of the American Joint Committee on Cancer, the depth of tumor invasion in each of the carcinomas was classified into five groups, as follows: Tis, carcinoma *in situ* or limited to mucosa; T1, invading the submucosa; T2, invading the muscularis propria; T3, invading either the subserosa; and T4, invading through the serosa or invading contiguous organs. The status of lymph node metastasis was therefore stratified as follows: N0, absence of regional lymph node metastasis; N1, one to three regional lymph node metastasis; N2, four or more regional lymph node metastasis. The presence of distant metastasis was noted as follows: M0, absence of distant metastasis; M1, presence of distant metastasis. The clinicopathological parameters are summarized in Table 1. One representative slide with a transversal section of the tumor and with safe mucosa for each sample was selected. The immunohistochemistry was carried out on 4-μm-thick paraffin-embedded sections of formalin-fixed tumor samples using an antibody directed against olfactomedin-4 (catalog number ab78496; Abcam, Cambridge, MA; 1:25). The immunolabeling technique was performed by a Benchmark automated tissue staining system (Ventana).

The immunohistochemistry was evaluated semi-quantitatively by the percentage of cytoplasmic staining cells, the intensity, and the presence or lack of secretory granules. To exclude subjectivity, all of the slides were evaluated by two pathologists who had no knowledge of the patients' identities or clinical status. In discrepant cases, the two pathologists reviewed the slides together and reached a consensus.

The percentage of immunopositive stained cells (A) was divided into five grades as: <10% (score 0); 10–30% (score 1); 30–50% (score 2); 50–70% (score 3); and >70% (score 4). Second, the intensity of staining was scored by evaluating the average staining intensity (B) of the positive cells (0, none; 1, weak; 2, intermediate; and 3, strong). The score for each section was measured as A × B, and the result was defined as negative (–, 0), weakly positive (+, 1–3), positive (++, 4–7), and strongly positive (+++, 8–12). The immunohistochemical data were subjected to statistical analysis. All of the quantitative data were recorded as the means ± S.D. Comparison between multiple groups were performed by one-way analysis of variance and Wilcoxon rank tests (*p* < 0.05).

**Specificity of the OLFM4 Antibody**—The specificity of the OLFM4 antibody was determined by Western blot analyses using protein extracts from four different cell lines (supplemental Fig. 1, A and B). RNA interference data were obtained from colorectal cell lines transfected with OLFM4 specific or control siRNA oligonucleotides (supplemental Fig. 1, C and D). The validation of the OLFM4 antibody was also determined by Western blot analysis using cell extract from COS7 cells overexpressing OLFM4 (supplemental Fig. 1E). Finally, *in situ* hybridization for OLFM4 in the human colon realized in Clevers's group (9) and our immunohistochemistry (IHC) images were compared (supplemental Fig. 1F).

### Quantitative Proteomic Analysis of Colorectal Tumors

**Cell Lines and Transfection**—Human cell lines (American Type Culture Collection) were maintained in antibiotic-free RPMI 1640 medium (Lonza). The cultures were supplemented with 10% fetal bovine serum. The cell lines were maintained at 37 °C in 5% carbon dioxide and were tested to rule out mycoplasma contamination. For transfection experiments, the cells were seeded into 60-mm culture dishes and grown until 80% confluence. The empty plasmid pcDN4/T0 and the pcDNA4/OLFM4 plasmids were stably cotransfected with the pcDNA6/TR using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. The cells were selected with 100 µg/ml blasticidin (Sigma-Aldrich) and 500 µg/ml zeocin (Invitrogen) for 2 weeks and maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum containing 100 µg/ml zeocin and 2.5 µg/ml blasticidin.

**Chromatin Immunoprecipitation Assays**—ChIP experiments were performed as previously described (40–43). Briefly, HT29 cells were fixed with 1% formaldehyde. After 10 min, the cells were washed with ice-cold Tris-buffered saline and lysed with 500 µl of ChIP buffer (50 mM Tris-HCl, pH 8.1, 1% SDS, 10 mM EDTA, 1 mM PMSF, 10 µg/ml aprotinin, 10 µg/ml leupeptin, 10 µg/ml pepstatin, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 50 mM NaF). Chromatin was sheared by sonication to an average size of 500 bp. The chromatin solution was diluted with 1 volume of dilution buffer (2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 1% Triton X-100, 0.1% Nonidet P-40, 150 mM NaCl, 1 mM PMSF, 10 µg/ml aprotinin, 10 µg/ml leupeptin, 10 µg/ml pepstatin, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 50 mM NaF) and incubated 1 h at 4 °C on a rotating platform with protein A-agarose and protein G-Sepharose that was pretreated with sheared DNA salmon sperm. Chromatin was then incubated overnight at 4 °C on a rotating platform with 1 µg of the indicated antibodies or anti-GAL4 antibodies. Following precipitation with protein A-agarose and protein G-Sepharose (pretreated with sheared DNA salmon sperm), chromatin was eluted with elution buffer (1% SDS, 100 mM NaHCO<sub>3</sub>) for 5 h at 65 °C. DNA was extracted with phenol chloroform, precipitated with ethanol, allowed to air dry, and then dissolved in 100 µl of sterile H<sub>2</sub>O. Four µl of the DNA samples were then subjected to PCR amplification.

### RESULTS

**iTRAQ Analysis of Colorectal Tumors**—Given the cellular heterogeneity of colorectal cancer, LCM was first applied on tissue to obtain a highly purified population of tumor cells. Representative images of pre- and postmicrodissected tissue images as well as purified cells are presented Fig. 1A. Note that UV laser capture induced cell damage and reduced protein yield during microdissection so that all experiments were performed using infrared laser capture. Approximately 100,000 cells were collected from multiple consecutive tissue sections, and quantitative expression profiles were obtained using iTRAQ labeling coupled with OFFGEL fractionation and off-line nanoLC/MS/MS as we previously described (30). To obtain proteomic maps of the successive steps of colorectal cancer, four adenomas and 24 adenocarcinomas representing the four clinical stages of the disease were subjected to eight different iTRAQ experiments (six 4-plex and two 8-plex; see Fig. 1B). To compare the different results, a pool of three different normal tissues was included in each experiment, labeled with a 114 tag. Finally, among these eight experiments, five were performed using 100 µg of tissue, and three were conducted with less than 70 µg, for each sample.

**Proteomic Analyses of Colorectal Cancer**—We used the ProteinPilot algorithm (34) to characterize the number of identified proteins (see the identification criteria under “Materials and Methods”). In a first three 4-plex analyses (iTRAQ-1, -2, and -7; supplemental Tables 1–3), 1672 unique proteins were identified (more than 1100 unique proteins by experiment; Table II). Among these, 734 (43.9%) were detected in each of the three iTRAQ experiments, whereas 366 (21.9%) were common to at least two analysis (Fig. 2A). This indicates that approximately two-thirds of the identified proteins can be detected in at least two of the three iTRAQ experiments. To extend these results, we then compared iTRAQ results, using either a 4800 (Fig. 2B) or 5800 (Fig. 2C) MALDI TOF/TOF. Using the 4800 approach, we were able to identify 1457 unique proteins with two 4-plex and one 8-plex experiments. 598 (41.0%) common proteins were detected in all three experiments, and 371 (25.4%) were shared by at least two experiments, indicating again that two-third of the identified proteins can be detected in at least two of the three iTRAQ experiments (Fig. 2B). Using the 5800 MALDI TOF/TOF and one 4-plex and one 8-plex experiments, we were able to identify 1443 unique proteins; among which 821 (56.9%) were common to the two experiments (Fig. 2C).

To define a colorectal cancer proteome, we then determined the total number of unique proteins that can be identified from the 28 tumor samples. As a first attempt, according to the criteria, two peptides/protein with score >95% for each iTRAQ experiment, 2141 unique proteins were identified (Table II and supplemental Table 4). We then repeated this identification but combined the eight iTRAQ experiments into a single group. In this case, 3138 unique proteins were identified with at least two peptides (confidence score > 99%) (supplemental Table 5). We then classified this colorectal cancer proteome using GOMiner software. 17 GO cellular components terms were defined among the 3138 proteins (supplemental Fig. 2A), and the majority (37%) was attributed to membrane proteins from cytoplasm, nucleus, mitochondrion, endoplasmic reticulum, and Golgi apparatus. This demonstrated the effectiveness of the reported approach for the identification of hydrophobic species. Proteins were also grouped based on their biological functions (supplemental Fig. 2B): 68% were involved in metabolism, 39% were involved in the regulation of biological process, and 23% were transport proteins. In addition, 21 and 19% were involved in development and cell communication, respectively. GO molecular functions annotation terms indicated that 64% were associated with protein binding, 47% had a catalytic activity, 23% were involved in nucleotide, and 27% were involved in nucleic acid binding (supplemental Fig. 2C).

**Technical and Experimental Reproducibility of iTRAQ Experiments**—To estimate the analytical reproducibility of our results, two experimental replicates (identical sample in two different iTRAQ sets) of a technical duplicate (two identical

## Quantitative Proteomic Analysis of Colorectal Tumors

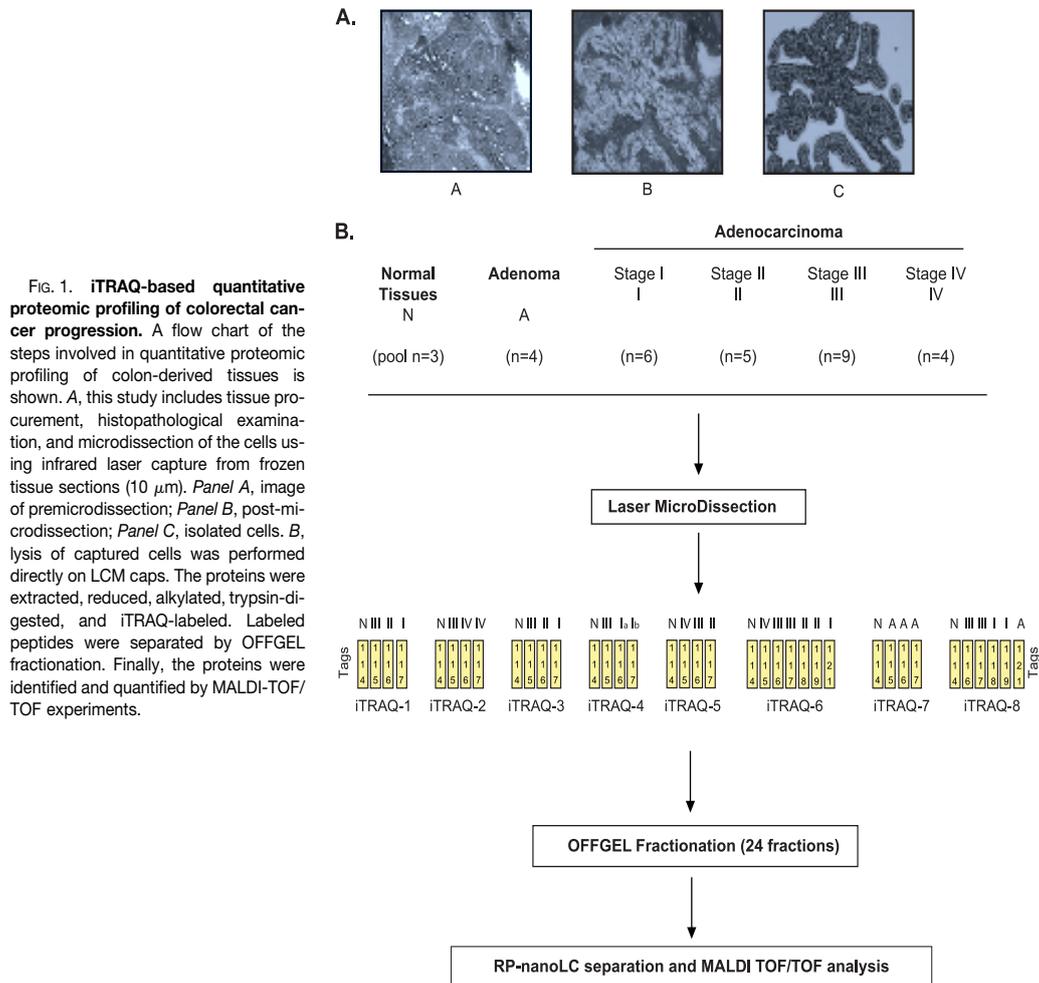


FIG. 1. **iTRAQ-based quantitative proteomic profiling of colorectal cancer progression.** A flow chart of the steps involved in quantitative proteomic profiling of colon-derived tissues is shown. *A*, this study includes tissue procurement, histopathological examination, and microdissection of the cells using infrared laser capture from frozen tissue sections (10  $\mu$ m). *Panel A*, image of pre-microdissection; *Panel B*, post-microdissection; *Panel C*, isolated cells. *B*, lysis of captured cells was performed directly on LCM caps. The proteins were extracted, reduced, alkylated, trypsin-digested, and iTRAQ-labeled. Labeled peptides were separated by OFFGEL fractionation. Finally, the proteins were identified and quantified by MALDI-TOF/TOF experiments.

TABLE II  
Spectra, peptides, and proteins identified and quantified of eight independent iTRAQ experiments

Experiment	iTRAQ	Mass spectrometer	Quantity	Total spectra	Identified peptides (confidence = 95%)	Identified peptides (confidence = 95%, contribution = 2)	Identified proteins (after grouping)		Quantified proteins (unused >4)
							Unused >2	Unused >4	
			$\mu$ g						
iTRAQ-1	4-plex	4800	100	146,063	51,399	17,287	2175	1210	1180
iTRAQ-2	4-plex	4800	100	105,013	52,462	6638	2017	1127	1102
iTRAQ-7	4-plex	5800	100	229,545	88,106	8215	2251	1253	1224
iTRAQ-8	8-plex	5800	100	338,262	67,606	6573	1937	1041	1012
iTRAQ-6	8-plex	4800	100	131,143	42,005	12,997	1406	742	740
iTRAQ-3	4-plex	4800	70	86,380	23,768	8299	1214	604	603
iTRAQ-4	4-plex	4800	70	97,110	46,234	6844	1075	557	557
iTRAQ-5	4-plex	4800	70	57,186	26,247	3945	1242	431	431

samples in an iTRAQ set) were run (supplemental Fig. 3A). In total, 1282 proteins were identified (with at least two peptides) across both iTRAQ experiment replicates using Protein Pilot

(1222 in replicate A1 and 1175 in replicate A2) (supplemental Table 6). The Venn diagram provided in supplemental Fig. 3B shows that 1115 (87%) of these proteins were

## Quantitative Proteomic Analysis of Colorectal Tumors

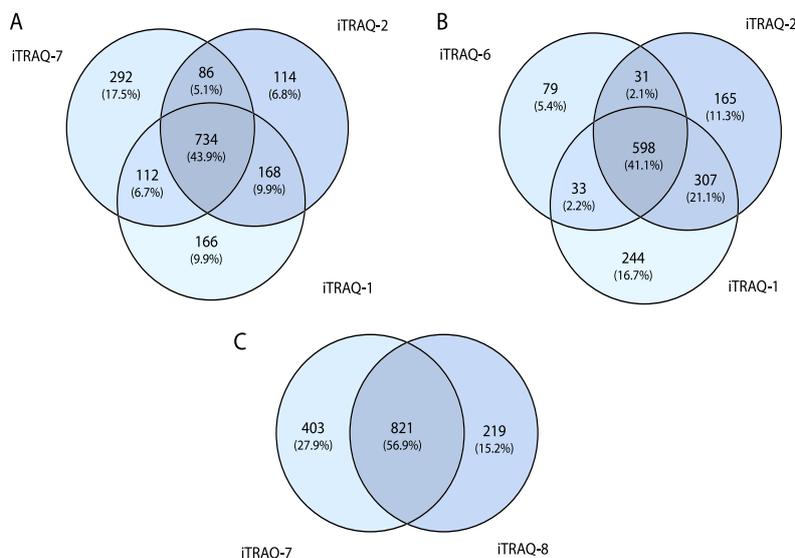


FIG. 2. Venn diagrams for the distribution of proteins between the different iTRAQ experiments. A, comparison between 4-plex experiments (two MALDI 4800 and one MALDI 5800 experiments). B, comparison between MALDI 4800 experiments (two 4-plex and one 8-plex experiments). C, comparison between MALDI 5800 experiments (one 4-plex and one 8-plex experiments).

common to both sets. Linear regression analyses were performed on ratios obtained from the duplicate analyses. Pearson correlation coefficients between both technical samples were 0.94 and 0.95, and those between the technical duplicate were 0.91 and 0.92 ( $p < 0.0001$ ) (supplemental Fig. 3C). Thus, the duplicate ratios were significantly positively correlated, indicating a good technical sample preparation and a good analytical reproducibility of the OFFGEL-LC/MS/MS approach.

**Proteomic Analysis of the Different Stages of Colorectal Cancer**—We then used the iQuantator software to quantify protein expression between the different stages of colorectal cancer. This approach generates variation means and 95% credible intervals for each expression change. For proteins whose iTRAQ ratios were down-regulated, down-regulation was considered to be significant if the upper limit of the credible interval was below 1. Conversely, for proteins whose iTRAQ ratios were up-regulated, up-regulation was considered to be significant if the lower limit of the credible interval had a value greater than 1. By taking into consideration both the peptide and spectra numbers, this approach allowed us to detect small but significant expression changes, provided that several peptides are detected. Using this analysis, we were able to obtain a list of quantified proteins from the eight iTRAQ experiments (supplemental Table 7). This allowed us to determine the individual protein expression of each patient (supplemental Table 8) and consequently to define the variation of protein expression between the different stages of colorectal cancer (adenoma, stages I–IV). 555 proteins were

identified that vary significantly between these different steps of the disease (supplemental Table 9). At the adenoma stage, 164 proteins were expressed differently as compared with normal tissue, and significant variations were also observed for the other stages (stage 1, 183 proteins; stage 2, 304 proteins; stage 3, 194 proteins; and stage 4, 69 proteins; in each case the variation is expressed as compared with normal tissue). Note that in each condition, both up- or down-regulated proteins can be detected, indicating that this approach is suitable to identify proteins that are inactivated during the transformation process and not only oncogenes that are overexpressed (supplemental Table 9). Using the Metacore data mining tool (<http://www.genego.com/metacore.php>), we then analyzed the signaling pathways represented at the different stages. The cell adhesion-cell matrix pathway was identified as the most significant network associated with adenoma and stage I ( $p = 4.20348E-10$  and  $1.81119E-12$ , respectively), the cytoskeleton-actin filament pathway was significantly associated with stage II ( $p = 1.5857E-19$ ), the integrin-mediated cell adhesion and migration pathway was correlated with stage III ( $p = 1.41513E-07$ ), and the cytoskeleton-intermediate filament pathway was significantly associated with stage IV ( $p = 1.642E-09$ ). Using this analysis, we also found as expected that the adenoma signature was associated with intestinal diseases ( $p = 3.4659E-21$ ); stage 1 was associated with gastrointestinal neoplasms ( $p = 2.0057E-16$ ); stage II was associated with digestive system neoplasms ( $p = 7.7799E-24$ ); stage III was associated with pathologic processes ( $p = 2.0623E-28$ ); and stage IV was

### Quantitative Proteomic Analysis of Colorectal Tumors

associated with intestinal diseases ( $p = 7,1607E-19$ ) (supplemental Tables 10–14).

*Validation of the Proteomic Analysis Using the Human Atlas Protein*—The verification of proteomic results involves IHC analysis on tumor tissue where only a few proteins are generally examined. Rather than performing IHC analysis on a limited number of proteins, we took advantage of IHC data available in the Human Protein Atlas (<http://www.proteinatlas.org>, 8832 antibodies and 7,334,244 images). As an unbiased approach, we selected all of proteins for which expression differed significantly as compared with normal tissue and assessed HPA IHC data. Because all of these proteins were not represented in the HPA, most of the time as a consequence of missing data, we further selected within this list 83 proteins that met the following criteria: 1) IHC expression reported in more than one normal colorectal tissue; 2) IHC expression reported in more than eight colorectal cancer samples; and 3) the HPA antibody verification score was moderate or high (see <http://www.proteinatlas.org>). HPA images were then manually inspected to confirm that the protein of interest was overexpressed, unchanged, or underexpressed in tumor cells as compared with normal tissue. Of the 83 selected proteins, 44 (53%) presented consistent expression ratios between iTRAQ and IHC results (supplemental Table 15). Twenty-seven proteins (33%) were determined to be unchanged by IHC, whereas they were down- or up-regulated in tumor samples, and 12 proteins (14%) in IHC did not fit with the expression in our study. Examination of the literature allowed the confirmation of our iTRAQ results for 21 of 39 proteins (unchanged or no fit by IHC); no information was found for the 16 remaining proteins. Only the tenascin result does not seem to match the information in the literature and the IHC results. Note that the different colon cancer stages are not specified in the HPA database. This prevents the detection of variation that would be stage-specific and might explain these discrepancies.

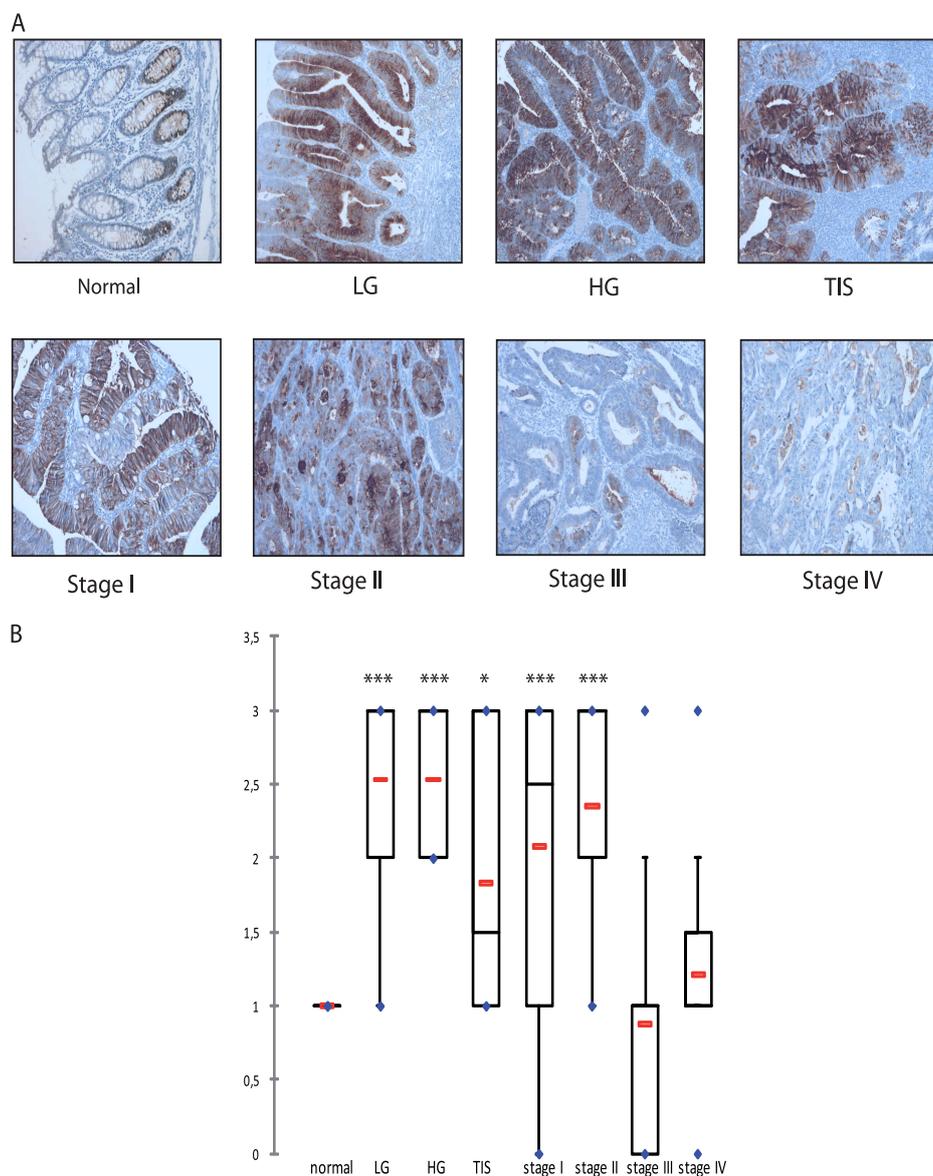
*OLFM4 Is Expressed at the Early Stages of Colorectal Cancer*—We then asked whether we could identify some proteins that are specific to the early stages of colorectal cancer. To this end, proteins were classified with the following criteria: 1) significant up-regulation in adenoma and stages I/II according to iQuantitor analysis and 2) not significantly expressed or underexpressed in stages III/IV. With this approach, only four proteins were identified (supplemental Table 9), the aldehyde dehydrogenase (ALDH1A1), the heat shock protein 1 (HSPE1), the sorbitol dehydrogenase (SORD), and OLFM4. Interestingly, OLFM4 encodes a protein that has been recently described as a specific marker of colorectal stem cells in association with Lgr5 (32, 33). In our experimental conditions, OLFM4 was detected with the highest statistical confidence, and given its importance in colorectal cancer, we focused on this protein for the following part of the study. To confirm its dysregulation in adenoma and in the early stage of CCR, its expression was first analyzed by immunohistochem-

istry using paraffin-embedded tissues isolated from 126 patients. Representative pictures of OLFM4 staining in adenomas, early or metastatic CRC cases are shown Fig. 3A. Whereas normal intestinal crypts showed moderate nuclear staining, results showed that the cytoplasmic and nuclear staining increased significantly in dysplasia tissue and in noninvasive tumors. OLFM4 was found to be significantly up-regulated in low grade adenoma, high grade adenoma, *in situ* adenoma, and stages I and II as compared with normal crypts (one-way analysis of variance test,  $p < 0.05$ ). By contrast, OLFM4 expression was not significantly different between invasive tissues (stage III/IV) and normal tissues (Fig. 3B).

*OLFM4 Expression Is Activated by the Ras-NF- $\kappa$ B2 Pathway*—These observations suggested to us that OLFM4 is expressed at the early stages of colorectal cancer, probably in response to oncogenes that are involved in the initial step of cell transformation. Although it is well known that the adenomatous polyposis coli/ $\beta$ -catenin pathway plays an important role in the initial transformation of intestinal crypts, we and others have also shown that the STAT3 and NF- $\kappa$ B transcription factor plays an essential role in this disease (43–45). Because STAT3 and NF- $\kappa$ B2 subunits are known to interact in tumor cells, we investigated the role of these two transcription factors in the regulation of OLFM4 expression. Transcription factor recognition site analysis of the OLFM4 promoter revealed the presence of several potential binding sites for these proteins. To determine whether STAT3 and NF- $\kappa$ B2 can be found associated with the OLFM4 gene, ChIP experiments were performed in growing HT29 cells using pair of primers corresponding to the proximal promoter. Although we were not able to detect any association of STAT3 with this region, ChIP results showed that NF- $\kappa$ B2 and its cofactor BCL3 are associated with the OLFM4 proximal promoter and that this was correlated with the binding of the RNA polymerase II (Fig. 4A). Note that this effect was noticed on endogenous proteins and not following overexpression. In addition, using RNA interference, we also noticed that the down-regulation of NF- $\kappa$ B2 inhibits OLFM4 expression at the protein and mRNA levels (Fig. 4B).

It has been shown recently that the Ras oncogene plays an important role in the initial stages of colorectal cancer and that this signaling pathway can deregulate the NF- $\kappa$ B transcription factor to allow abnormal cell cycle progression and survival (44, 46, 47). To determine whether Ras regulates OLFM4, we used stable HT29 cells expressing the RasV12 oncogene under the control of a doxycycline-inducible promoter (48). As expected, Ras was up-regulated in response to doxycycline, and a significant activation of the NF- $\kappa$ B2 transcription factor was detected (Fig. 4C). Interestingly, the induction of Ras was correlated with an up-regulation of OLFM4 at the protein level (Fig. 4D). In addition, semi-quantitative PCR experiments showed that this effect was regulated at the transcriptional level (Fig. 4D).

## Quantitative Proteomic Analysis of Colorectal Tumors



**FIG. 3. The validation of OLFM4 in human CRC tissue.** *A*, the immunohistochemistry staining of the different colorectal tumor stages (100 $\times$ ). *LG*, low grade adenoma; *HG*, high grade adenoma; *TIS*, carcinoma *in situ*. *Stages I and II*, early stage carcinomas; *stages III and IV*, metastatic carcinomas. *B*, different distribution of OLFM4 IHC scores (staining intensity  $\times$  staining percentage) between normal crypts, adenomas, carcinomas *in situ*, and different colorectal tumor stages ( $p < 0.05$ , significant for normal *versus* low grade adenoma, high grade adenoma, carcinoma *in situ*, stage I, and stage II, not significant for normal *versus* stages III and IV). The *red lines* represent the mean, and the *blue points* represent the minimum and maximum values. Dunnett's multiple comparison tests were performed;  $q = 3.20$  (\*) for normal *versus* carcinoma *in situ*;  $q = 6.35$  (\*\*\*) for normal *versus* low and high grade adenoma;  $q = 5.21$  and  $5.49$  for normal *versus* stages I and II, respectively.

*OLFM4 Is Overexpressed in Ras-mutated Tumors*—These observations suggested to us that the expression of OLFM4 might be enhanced in tumor samples expressing the Ras-NF-

$\kappa$ B2 pathway. Because NF- $\kappa$ B2 is generated as a cleavage product of its p100 precursor, the detection of the active form of this transcription factor is difficult in tumor samples. How-

## Quantitative Proteomic Analysis of Colorectal Tumors

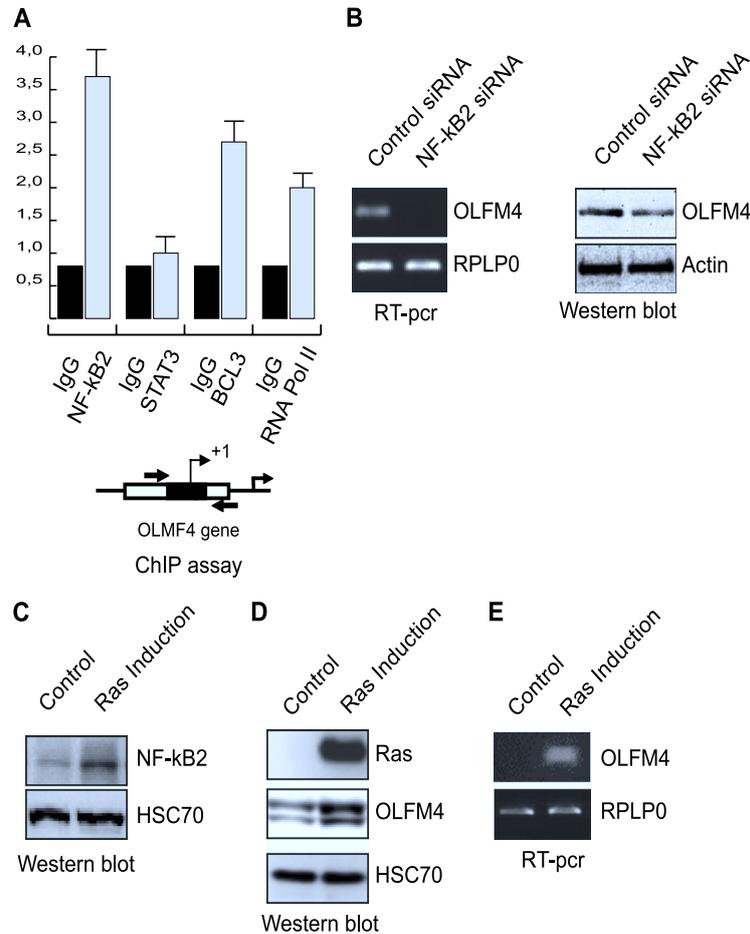


FIG. 4. **NF-κB2 regulates OLFM4 expression following Ras induction.** A, soluble chromatin from growing HCT116 cells was prepared and immunoprecipitated with antibodies targeted against NF-κB2, Bcl3, and the RNA polymerase II. DNA was amplified using pair of primers that cover the NF-κB proximal binding site of the OLFM4 promoter. ChIP assays were then quantified by real time RT-PCR as compared with the signal obtained on a control sequence with a control IgG. The schematic representation of the potential NF-κB binding sites of the OLFM4 promoter is presented. The reading frame of the binding site is indicated by + for 5'-3' and - for 3'-5'. B, HCT116 colorectal cell lines were either transfected with NF-κB2 specific or control siRNA oligonucleotides as indicated. The mRNA extracts (*left panels*) and whole cell extracts (*right panels*) were processed, and OLFM4 expression was analyzed 48 h after siRNA transfection. C, HT29 colorectal cell lines expressing an incible vector of the RasV12 oncogene were treated or not with doxycyclin for 48 h, and whole cell extracts were prepared and analyzed using antibodies directed against NF-κB2 and HSC70 as a loading control. D and E, cells were treated as above, and the expression of OLFM4 or Ras was evaluated by Western blot analysis (D) or RT-PCR (E).

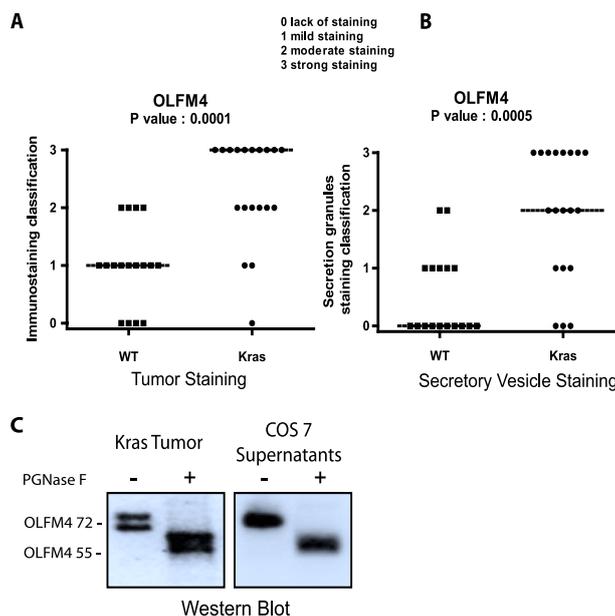
ever, the presence of the RasV12 oncogene can be determined by DNA sequencing. For this reason, we then analyzed OLFM4 expression in tumors expressing or not a mutated form of this oncogene. Interestingly, results presented Fig. 5A indicate that OLFM4 expression was significantly enhanced in Ras-mutated tumors ( $p < 0.0001$ ) as compared with wild type tumor samples.

*OLF4 Is Present in the Cytoplasm and as a Secreted Protein*—During the course of the IHC experiments, we no-

ticed that OLFM4 was expressed in the cytoplasm, and there was significant expression in the secretory vesicles (Fig. 5B). This result suggested to us that this protein might be secreted, and OLFM4 was effectively detected in the cell supernatants (data not shown and see Fig. 5C). In addition, we also noticed that two bands can be detected by Western blot, one at the expected molecular mass of 55 kDa and another band ~72 kDa. This observation suggested to us that OLFM4 was modified by glycosylation. To verify this hypothesis, extracts

## Quantitative Proteomic Analysis of Colorectal Tumors

FIG. 5. **OLFM4 is glycosylated and secreted in KRAS tumors.** A and B, OLFM4 IHC scores (staining intensity  $\times$  staining percentage) between WT and KRAS tissue in tumor tissue (A) and in secretory vesicles (B). C, deglycosylation analysis of OLFM4. OLFM4 was significantly secreted from the COS-7 cells; the secreted form at 72 kDa was treated with (+) or (-) peptide:*N*-glycosidase F (PGNase F). The size of deglycosylated OLFM4 is consistent with the molecular mass of calculated OLFM4 gene product. Treatment of the 72-kDa band in KRAS tumor leads to the same band at 55 kDa.



obtained from tumor samples were incubated with the peptide:*N*-glycosidase F deglycosylating enzyme. Following incubation, a shift in the molecular mass from 72 to 55 kDa was observed, indicating that the protein is effectively modified by *N*-glycosylation (Fig. 5C). To confirm this observation, COS cells were transfected with a vector allowing OLFM4 expression, and its potential secretion was analyzed by Western blot. Interestingly, results showed that only the 72-kDa band was detected in the supernatant, suggesting that OLFM4 is secreted as a glycosylated protein. Following peptide:*N*-glycosidase F treatment, the same migration shift was observed, and OLFM4 migrated as a 55-kDa protein (Fig. 5C, right panel). Importantly, in Ras-mutated tumors, OLFM4 was essentially detected at 72 kDa, suggesting that this protein is effectively secreted *in vivo*.

## DISCUSSION

Following the initial description of the genetic modifications occurring during colorectal cancer transformation, several studies have clearly shown that CRC results from multiple mutations that induce the deregulation of cell cycle and cell death pathways. This led to the important conclusion that CRC is a heterogeneous disease, which certainly explains why patients suffering from the same apparent disease have distinct outcome and different responses to the same anti-cancer treatment. Molecular clustering has therefore become an essential goal of cancer treatment, not only to establish tumor prognosis but also to identify the specific additive oncogenic pathways that should be targeted (49–51). Recent results have shown the interest of using genomic signatures

to identify these deregulated pathways and characterize prognosis markers. These gene signatures can also be used to characterize predictive markers that reflect the response to a particular treatment, but in this case, the predicting value of this approach remains to be fully validated (13, 52). In addition to these genomic experiments, quantitative proteomics also appears as a powerful tool to define cancer signatures that would identify disease subtypes, predict tumor escape, or characterize new molecular targets.

In this study, we provide what is to our knowledge the most extensive proteome database established so far for colorectal cancer and illustrate the value of using the combination of OFFGEL-iTRAQ labeling and MALDI-TOF/TOF approaches to explore the deep proteome of frozen tissues. It should be noted that the use of LCM favored the identification of many low concentrated proteins by removing abundant stromal proteins. This approach enables the identification and quantification of ~1100 proteins by patient, allowing the identification of a proteomic map for each tumor, which could be used in the future for individual clinical monitoring.

From a technical point of view, this study allowed us to compare the two MALDI-TOF/TOF 4800 and 5800 from AB sciex. With the iTRAQ technology, the 5800 MALDI seems to be slightly more sensitive than the 4800 MALDI (increase in identification of ~6% of proteins). However, in quite a surprising way, with the iTRAQ 8-plex, we can identify 35% more proteins with the MALDI 5800 than with the MALDI 4800. Although it is well known that the number of identified proteins and peptides is larger when using iTRAQ 4-plex than with iTRAQ 8-plex (53), it seems that the use of the MALDI

### Quantitative Proteomic Analysis of Colorectal Tumors

5800 lessens this difference. The difference in proteins identified changes from 35% in favor of the 4-plex when using the MALDI 4800 as compared with 18% with the 5800, thereby reviving interest in this reactive agent that can compare eight samples instead of four.

Starting from a cohort of 28 frozen tumors, we compared the protein profiles of adenomas or adenocarcinomas at the early stages (I and II) or metastatic stages (III and IV). The results indicate that the expression of a total of 555 proteins was significantly over- or underexpressed in colorectal cancer as compared with normal tissue, representing 16% of the total identified proteome. This approach allowed us to characterize the proteins that expression varied significantly between the different stages (adenoma, stages I-IV) and to establish what is to our knowledge the first proteomic analysis of these different steps. As compared with normal tissue, it is interesting to note that the most important variation was observed between stages II and III, which corresponds to the transition between a nonmetastatic and metastatic tumor. In this case, 304 proteins were found to vary significantly, whereas 180 were differently regulated in adenomas and stages I and III, and only 68 were modified in stage IV. Although this remains to be demonstrated, it is tempting to speculate that some proteins expressed in the stage II are necessary for invasive migration. In line with this hypothesis, we observed using gene ontology analysis that the expression of extracellular matrix proteins varies significantly. Among these proteins, we have focused on secreted proteins present at the early stage of the disease because they can be easily detected by ELISA, and they can be useful to distinguish these aggressive cancers from early stage cases. It is important to consider that patients with stage I/II cancer are believed to be cured after surgery but that ~20% of them will relapse. The distinction between these two cases is actually difficult, and for this reason, the identification of specific biomarkers of stage II colorectal tumors is an important goal that would allow the prediction of recurrence events. A 50-gene signature has been recently described that can separate early tumors depending on their relapse probability (15). Therefore, it will be interesting to determine whether the protein list identified in this study can be used to improve this early stage stratification to predict tumor relapse.

Among these proteins, we focused on OLFM4 because its expression was significantly up-regulated in adenomas, further increased in stage I, and maximum in stage II before dropping considerably in stages III and IV. Importantly, OLFM4 has recently been shown to be expressed in colorectal stem cells in association with Lgr5 (32, 33), further confirming its expression at the early stages of the disease. This protein plays probably an important role in cancer because it has been recently shown in myeloid cells that its promoter is probably methylated. Its re-expression induced cell cycle arrest and cell death in myeloid cells (54). Interestingly, the same effect has also been observed in prostate cancer cells

where OLFM4 levels are down-regulated during cancer progression, most significantly in tumors with high Gleason scores (55). Importantly, restoring the OLFM4 level through overexpression led to reduced proliferation and invasiveness. Because this protein is known to interact with lectins as well as cadherin (56), this effect might be explained by a better adhesion to the extracellular matrix or to the surrounding cells. However, this study also proposed that this effect is related to an inhibition of cathepsin D expression and an enhanced autophagic activity of prostate cancer cells. Because autophagy plays an important role in tumor suppression (57), it will be interesting to determine whether this effect of OLFM4 can be extended to other experimental models. Moreover, it should be noted that OLFM4 is a member of the olfactomedin domain-containing protein. This family includes Noelin (OLFM1), which prolongs neural crest production. The involvement of nervous system proteins, such as neurotrophins, in carcinogenesis has been reported for prostate (58) and breast (59) cancers, and it will be interesting to test whether OLFM4 behaves in the same way.

In our experimental conditions, it is striking to note that OLFM4 level was very significantly down-regulated in stages III and IV compared with stages I/II tumors and that a reduced expression of this protein has been recently correlated with poor prognosis. In light of these observations, it is tempting to speculate that the inactivation of OLFM4 is a necessary event to prevent cell death and allow tumor progression and metastasis in colorectal cancer. In line with this hypothesis, it has been recently reported that OLFM4 expression is reduced in poorly differentiated colon cancers, as well as at the late tumor-node-metastasis stage (60). If feasible, the detection of OLFM4 variations in the serum of the patients might therefore be an interesting tool to follow the evolution of stage II tumors. Interestingly, we also noticed that OLFM4 was regulated by the Ras-NF- $\kappa$ B2 pathway and that the expression of this protein was significantly enhanced in Ras-mutated tumors. We have recently shown that senescence is induced in response to the Ras oncogene in colorectal cell lines (48). Oncogene-induced senescence (OIS) is a powerful antitumor mechanism that induces permanent cell cycle arrest in response to abnormal proliferative signals (61). Originally described in cell culture, OIS has been recently shown to occur also *in vivo* as an early protection against carcinogenesis. In light of these observations, one can speculate that OLFM4 overexpression is an early event occurring in response to OIS in Ras-expressing cells. Further experiments are therefore necessary to characterize the effect of OLFM4 on cell cycle and cell death pathways in colorectal cell lines and to determine whether this effect is deregulated by the Ras oncogene to allow OIS escape and tumor progression.

Through the characterization of the molecular aberrations present in cancer cells, it is now widely accepted that the identification of new biomarkers will improve the outcome prognosis or the prediction of therapy response. In addition, it

## Quantitative Proteomic Analysis of Colorectal Tumors

is also expected that molecular clustering will help to separate apparently similar tumors to provide rationale treatments. Besides genomic approaches, our results indicate that proteomic analysis can be used on tumor samples to provide not only a better understanding of cell transformation in colorectal cancer but also to identify new biomarkers of the different tumor stages such as OLFM4.

**Acknowledgments**—We thank John Schwacke for help with iQuantator software.

\* This work was supported by grants from the Ligue Contre le Cancer (Comité du Maine et Loire), the Institut du Cancer, the Région Pays de Loire, and Amgen France. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ This article contains [supplemental Tables 1–15 and Figs. 1–3](#).

\*\* To whom correspondence should be addressed: Paul Papin Cancer Center, INSERM U892 2 rue Moll, 49033 Angers, France. Tel.: 33-2-41-35-29-14, E-mail: catherine.guette@ico.unicancer.fr.

## REFERENCES

- Cunningham, D., Atkin, W., Lenz, H. J., Lynch, H. T., Minsky, B., Nordlinger, B., and Starling, N. (2010) Colorectal cancer. *Lancet* **375**, 1030–1047
- Jemal, A., Siegel, R., Xu, J., and Ward, E. (2010) Cancer statistics, 2010. *CA Cancer J. Clin.* **60**, 277–300
- Huang, C. S., Lai, S. K., and Farraye, F. A. (2005) Colorectal cancer screening in average risk individuals. *Cancer Causes Control* **16**, 171–188
- van Rossum, L. G., van Rijn, A. F., Laheij, R. J., van Oijen, M. G., Fockens, P., van Krieken, H. H., Verbeek, A. L., Jansen, J. B., and Dekker, E. (2008) Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology* **135**, 82–90
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988) Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319**, 525–532
- Pino, M. S., and Chung, D. C. (2010) The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059–2072
- Clevers, H. (2006) Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469–480
- Reya, T., and Clevers, H. (2005) Wnt signalling in stem cells and cancer. *Nature* **434**, 843–850
- Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J., and Clevers, H. (2009) Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611
- van der Flier, L. G., and Clevers, H. (2009) Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–260
- Phelps, R. A., Chidester, S., Dehghanizadeh, S., Phelps, J., Sandoval, I. T., Rai, K., Broadbent, T., Sarkar, S., Burt, R. W., and Jones, D. A. (2009) A two-step model for colon adenoma initiation and progression caused by APC loss. *Cell* **137**, 623–634
- Lièvre, A., Bachet, J. B., Boige, V., Cayre, A., Le Corre, D., Buc, E., Ychou, M., Bouché, O., Landi, B., Louvet, C., André, T., Bibeau, F., Diebold, M. D., Rougier, P., Ducreux, M., Tomasic, G., Emile, J. F., Penault-Llorca, F., and Laurent-Puig, P. (2008) KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J. Clin. Oncol.* **26**, 374–379
- Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., Joshi, M. B., Harpole, D., Lancaster, J. M., Berchuck, A., Olson, J. A., Jr., Marks, J. R., Dressman, H. K., West, M., and Nevins, J. R. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357
- Chang, J. T., Carvalho, C., Mori, S., Bild, A. H., Gatzka, M. L., Wang, Q., Lucas, J. E., Potti, A., Febbo, P. G., West, M., and Nevins, J. R. (2009) A genomic strategy to elucidate modules of oncogenic pathway signaling networks. *Mol. Cell* **34**, 104–114
- Garman, K. S., Acharya, C. R., Edelman, E., Grade, M., Gaedcke, J., Sud, S., Barry, W., Diehl, A. M., Provenzale, D., Ginsburg, G. S., Ghadimi, B. M., Ried, T., Nevins, J. R., Mukherjee, S., Hsu, D., and Potti, A. (2008) A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19432–19437
- Crawford, N. P., Colliver, D. W., and Galandiuk, S. (2003) Tumor markers and colorectal cancer: Utility in management. *J. Surg. Oncol.* **84**, 239–248
- Sidransky, D. (2002) Emerging molecular markers of cancer. *Nat. Rev. Cancer* **2**, 210–219
- Duffy, M. J., van Dalen, A., Haglund, C., Hansson, L., Holinski-Feder, E., Klapdor, R., Lamerz, R., Peltomaki, P., Sturgeon, C., and Topolcan, O. (2007) Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *Eur. J. Cancer* **43**, 1348–1360
- Locker, G. Y., Hamilton, S., Harris, J., Jessup, J. M., Kemeny, N., Macdonald, J. S., Somerfield, M. R., Hayes, D. F., and Bast, R. C., Jr. (2006) ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–5327
- Jimenez, C. R., Knol, J. C., Meijer, G. A., and Fijneman, R. J. (2010) Proteomics of colorectal cancer: Overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. *J. Proteomics* **73**, 1873–1895
- Alfonso, P., Cañamero, M., Fernández-Carboni, F., Núñez, A., and Casal, J. I. (2008) Proteome analysis of membrane fractions in colorectal carcinomas by using 2D-DIGE saturation labeling. *J. Proteome Res.* **7**, 4247–4255
- Bi, X., Lin, Q., Foo, T. W., Joshi, S., You, T., Shen, H. M., Ong, C. N., Cheah, P. Y., Eu, K. W., and Hew, C. L. (2006) Proteomic analysis of colorectal cancer reveals alterations in metabolic pathways: Mechanism of tumorigenesis. *Mol. Cell. Proteomics* **5**, 1119–1130
- Rho, J. H., Qin, S., Wang, J. Y., and Roehrl, M. H. (2008) Proteomic expression analysis of surgical human colorectal cancer tissues: Up-regulation of PSB7, PRDX1, and SRP9 and hypoxic adaptation in cancer. *J. Proteome Res.* **7**, 2959–2972
- Ducoux-Petit, M., Uttenweiler-Joseph, S., Brichory, F., Bousquet-Dubouch, M. P., Bulet-Schiltz, O., Haeuw, J. F., and Monsarrat, B. (2008) Scaled-down purification protocol to access proteomic analysis of 20S proteasome from human tissue samples: Comparison of normal and tumor colorectal cells. *J. Proteome Res.* **7**, 2852–2859
- Roessler, M., Rollinger, W., Palme, S., Hagmann, M. L., Berndt, P., Engel, A. M., Schneider, B., Pfeffer, M., Andres, H., Karl, J., Bodenmüller, H., Rüschoff, J., Henkel, T., Rohr, G., Rossol, S., Rösch, W., Langen, H., Zolg, W., and Tacke, M. (2005) Identification of nicotinamide N-methyltransferase as a novel serum tumor marker for colorectal cancer. *Clin. Cancer Res.* **11**, 6550–6557
- Roessler, M., Rollinger, W., Mantovani-Endl, L., Hagmann, M. L., Palme, S., Berndt, P., Engel, A. M., Pfeffer, M., Karl, J., Bodenmüller, H., Rüschoff, J., Henkel, T., Rohr, G., Rossol, S., Rösch, W., Langen, H., Zolg, W., and Tacke, M. (2006) Identification of PSME3 as a novel serum tumor marker for colorectal cancer by combining two-dimensional polyacrylamide gel electrophoresis with a strictly mass spectrometry-based approach for data analysis. *Mol. Cell. Proteomics* **5**, 2092–2101
- Kim, H. J., Kang, H. J., Lee, H., Lee, S. T., Yu, M. H., Kim, H., and Lee, C. (2009) Identification of S100A8 and S100A9 as serological markers for colorectal cancer. *J. Proteome Res.* **8**, 1368–1379
- Ma, Y., Peng, J., Liu, W., Zhang, P., Huang, L., Gao, B., Shen, T., Zhou, Y., Chen, H., Chu, Z., Zhang, M., and Qin, H. (2009) Proteomics identification of desmin as a potential oncofetal diagnostic and prognostic biomarker in colorectal cancer. *Mol. Cell. Proteomics* **8**, 1878–1890
- Xu, B. J., Li, J., Beauchamp, R. D., Shyr, Y., Li, M., Washington, M. K., Yeatman, T. J., Whitehead, R. H., Coffey, R. J., and Caprioli, R. M. (2009) Identification of early intestinal neoplasia protein biomarkers using laser capture microdissection and MALDI MS. *Mol. Cell. Proteomics* **8**, 936–945
- Ernout, E., Gamelin, E., and Guette, C. (2008) Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation. *Proteome Sci.* **6**, 27

## Quantitative Proteomic Analysis of Colorectal Tumors

31. Ernout, E., Bourreau, A., Gamelin, E., and Guette, C. (2010) A proteomic approach for plasma biomarker discovery with iTRAQ labelling and OFF-GEL fractionation. *J. Biomed. Biotechnol.* **2010**, 927917
32. Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J., and Clevers, H. (2007) Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007
33. van der Flier, L. G., Haegebarth, A., Stange, D. E., van de Wetering, M., and Clevers, H. (2009) OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137**, 15–17
34. Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., and Schaeffer, D. A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* **6**, 1638–1655
35. Schwacke, J. H., Hill, E. G., Comte-Walters, S., and Schey, K. L. (2009) iQuantitor: A tool for protein expression inference using iTRAQ. *BMC Bioinformatics* **10**, 342
36. Grant, J. E., Bradshaw, A. D., Schwacke, J. H., Baicu, C. F., Zile, M. R., and Schey, K. L. (2009) Quantification of protein expression changes in the aging left ventricle of *Rattus norvegicus*. *J. Proteome Res.* **8**, 4252–4263
37. Zeeberg, B. R., Qin, H., Narasimhan, S., Sunshine, M., Cao, H., Kane, D. W., Reimers, M., Stephens, R. M., Bryant, D., Burt, S. K., Elnekave, E., Hari, D. M., Wynn, T. A., Cunningham-Rundles, C., Stewart, D. M., Nelson, D., and Weinstein, J. N. (2005) High-Throughput GoMiner, an “industrial-strength” integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics* **6**, 168
38. Ekins, S., Nikolsky, Y., Bugrim, A., Kirillov, E., and Nikolskaya, T. (2007) Pathway mapping tools for analysis of high content data. *Methods Mol. Biol.* **356**, 319–350
39. Sobin, L. H., Gospodarowicz, M. K., and C., W. (eds.) (2009) *TNM Classification of Malignant Tumours*, 7th Ed., Wiley-Blackwell, Chichester, UK
40. Bienvendu, F., Barré, B., Giraud, S., Avril, S., and Coqueret, O. (2005) Transcriptional regulation by a DNA-associated form of cyclin D1. *Mol. Biol. Cell* **16**, 1850–1858
41. Courapied, S., Sellier, H., de Carné Trécesson, S., Vigneron, A., Bernard, A. C., Gamelin, E., Barré, B., and Coqueret, O. (2010) The *cdk5* kinase regulates the STAT3 transcription factor to prevent DNA damage upon topoisomerase I inhibition. *J. Biol. Chem.* **285**, 26765–26778
42. Courapied, S., Chérier, J., Vigneron, A., Troadec, M. B., Giraud, S., Valo, I., Prigent, C., Gamelin, E., Coqueret, O., and Barré, B. (2010) Regulation of the *Aurora-A* gene following topoisomerase I inhibition: Implication of the Myc transcription factor. *Mol. Cancer* **9**, 205
43. Vigneron, A., Gamelin, E., and Coqueret, O. (2008) The EGFR-STAT3 oncogenic pathway up-regulates the Eme1 endonuclease to reduce DNA damage after topoisomerase I inhibition. *Cancer Res.* **68**, 815–825
44. Greten, F. R., Eckmann, L., Greten, T. F., Park, J. M., Li, Z. W., Egan, L. J., Kagnoff, M. F., and Karin, M. (2004) IKK $\beta$  links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell* **118**, 285–296
45. Bollrath, J., Phesse, T. J., von Burstin, V. A., Putoczki, T., Bennecke, M., Bateman, T., Nebelsiek, T., Lundgren-May, T., Canli, O., Schwitalla, S., Matthews, V., Schmid, R. M., Kirchner, T., Arkan, M. C., Ernst, M., and Greten, F. R. (2009) gp130-mediated Stat3 activation in enterocytes regulates cell survival and cell-cycle progression during colitis-associated tumorigenesis. *Cancer Cell* **15**, 91–102
46. Mathew, R., Karp, C. M., Beaudoin, B., Vuong, N., Chen, G., Chen, H. Y., Bray, K., Reddy, A., Bhanot, G., Gelinas, C., Dipaola, R. S., Karantza-Wadsworth, V., and White, E. (2009) Autophagy suppresses tumorigenesis through elimination of p62. *Cell* **137**, 1062–1075
47. Barré, B., and Perkins, N. D. (2010) The Skp2 promoter integrates signaling through the NF- $\kappa$ B, p53, and Akt/GSK3 $\beta$  pathways to regulate autophagy and apoptosis. *Mol. Cell* **38**, 524–538
48. de Carné Trécesson, S., Guillemin, Y., Bélanger, A., Bernard, A. C., Preisser, L., Ravon, E., Gamelin, E., Juin, P., Barré, B., and Coqueret, O. (2011) Escape from p21-mediated oncogene-induced senescence leads to cell dedifferentiation and dependency on anti-apoptotic Bcl-xL and Mcl1 proteins. *J. Biol. Chem.* **286**, 12825–12838
49. Bernards, R. (2010) It's diagnostics, stupid. *Cell* **141**, 13–17
50. Chin, L., and Gray, J. W. (2008) Translating insights from the cancer genome into clinical practice. *Nature* **452**, 553–563
51. van't Veer, L. J., and Bernards, R. (2008) Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* **452**, 564–570
52. Sawyers, C. L. (2008) The cancer biomarker problem. *Nature* **452**, 548–552
53. Pichler, P., Köcher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., and Mechtler, K. (2010) Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Anal. Chem.* **82**, 6549–6558
54. Liu, W., Lee, H. W., Liu, Y., Wang, R., and Rodgers, G. P. (2010) Olfactomedin 4 is a novel target gene of retinoic acids and 5-aza-2'-deoxycytidine involved in human myeloid leukemia cell growth, differentiation, and apoptosis. *Blood* **116**, 4938–4947
55. Chen, L., Li, H., Liu, W., Zhu, J., Zhao, X., Wright, E., Cao, L., Ding, I., and Rodgers, G. P. (2011) Olfactomedin 4 suppresses prostate cancer cell growth and metastasis via negative interaction with cathepsin D and SDF-1. *Carcinogenesis* **32**, 986–994
56. Grover, P. K., Hardingham, J. E., and Cummins, A. G. (2010) Stem cell marker olfactomedin 4: Critical appraisal of its characteristics and role in tumorigenesis. *Cancer Metastasis Rev.* **29**, 761–775
57. White, E., and Lowe, S. W. (2009) Eating to exit: Autophagy-enabled senescence revealed. *Genes Dev.* **23**, 784–787
58. Thiele, C. J., Li, Z., and McKee, A. E. (2009) On Trk: The TrkB signal transduction pathway is an increasingly important target in cancer biology. *Clin. Cancer Res.* **15**, 5962–5967
59. Vanhecke, E., Adriaenssens, E., Verbeke, S., Meignan, S., Germain, E., Berteaux, N., Nurcombe, V., Le Bourhis, X., and Hondemarcq, H. (2011) Brain-derived neurotrophic factor and neurotrophin-4/5 are expressed in breast cancer and can be targeted to inhibit tumor cell survival. *Clin. Cancer Res.* **17**, 1741–1752
60. Liu, W., Liu, Y., Zhu, J., Wright, E., Ding, I., and Rodgers, G. P. (2008) Reduced hGC-1 protein expression is associated with malignant progression of colon carcinoma. *Clin. Cancer Res.* **14**, 1041–1049
61. Roninson, I. B. (2003) Tumor cell senescence in cancer treatment. *Cancer Res.* **63**, 2705–2715

### 1.3 Figure complémentaire

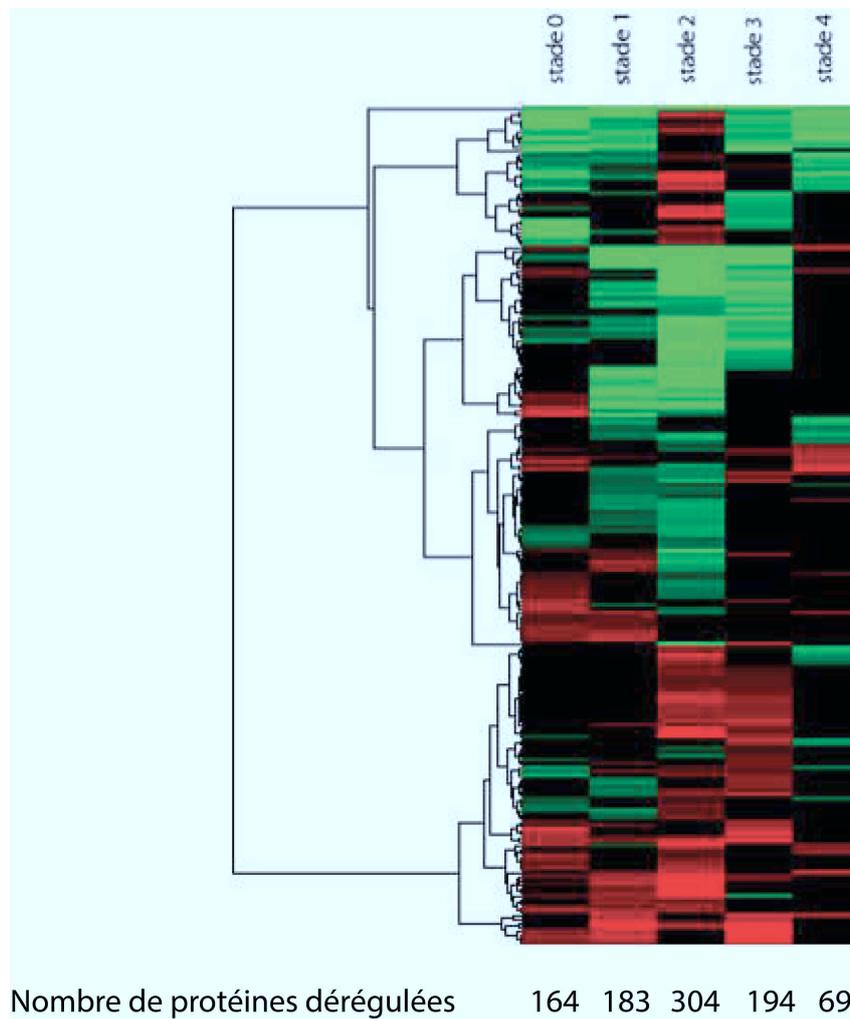


FIGURE 1.1 – **Protéines statistiquement dérégulées dans les différents stades du cancer colorectal** Nous avons 164 protéines dérégulées dans des adénomes, 183 dans les adénocarcinomes de stade 1, 304 dans les adénocarcinomes de stade 2, 194 dans les adénocarcinomes de stade 3 et 69 dans les adénocarcinomes de stade 4.



## Chapitre 2

### Résultats complémentaires concernant l'OLFM4

#### 2.1 L'OLFM4 un marqueur de détection des tumeurs précoces

DANS un premier temps, nous avons souhaité doser l'OLFM4 chez des patients sains et chez des patients présentant des tumeurs précancéreuses (polypes). Nous avons pu obtenir 60 séras de patients sains et 14 séras de patients présentant des polypes. Nous avons dosé l'OLFM4 dans les séras de ces patients à l'aide d'un kit ELISA commercial. Les résultats sont présentés en figure 2.1 p.135.

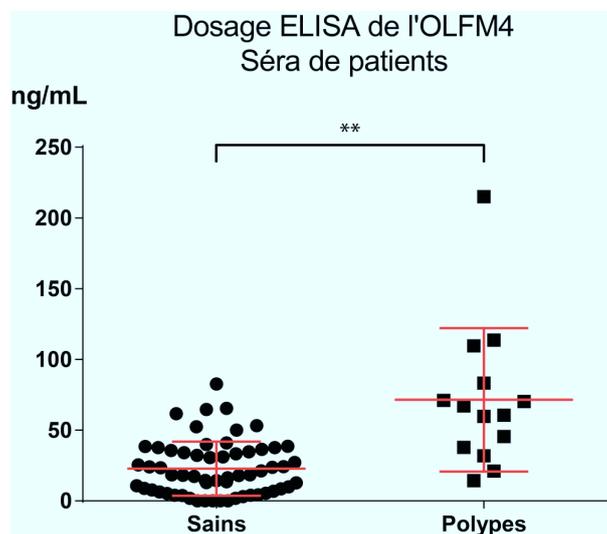


FIGURE 2.1 – L'OLFM4 est une protéine sécrétée, détectable dans les séras de patients et marqueur potentiel de tumeurs précoces. Résultats préliminaires de tests ELISA réalisés sur des séras de patients atteints de polypes. Le taux basal d'OLFM4 est de 20 ng/mL  $\pm$  6 ng/mL.

Bien que préliminaires, ces résultats montrent que l'OLFM4 est surexprimée de façon significative chez les patients porteurs de polypes. Les valeurs obtenues pour les dosages allaient de la limite de détection du kit à une valeur de 215 ng/mL pour le patient

présentant la concentration la plus élevée. Le taux basal se situe autour de 20 ng/mL  $\pm$  6 ng/mL.

## 2.2 L'OLFM4 : une protéine glycosylée

Seule la forme glycosylée semble être capable d'être sécrétée par la cellule d'après les résultats que nous avons précédemment décrit. Nous avons réalisé le développement d'une approche protéomique permettant d'analyser les protéines glycosylées sur des surnageants de COS7 transfectées avec un vecteur d'expression codant pour l'OLFM4. Cela nous a permis d'identifier plusieurs sites de glycosylations présents sur l'OLFM4. Afin de vérifier la pertinence de la technique, et sa robustesse, nous avons également réalisé des mutants de l'OLFM4 sur les sites mis en évidence.

### 2.2.1 Détermination des sites de glycosylation de l'OLFM4

#### 2.2.1.1 Approche bioinformatique

Il existe des outils disponibles gratuitement en ligne permettant de déterminer, à partir d'algorithmes recherchant les séquences consensus, les sites de glycosylation potentiels d'une protéine. Ces logiciels utilisent plusieurs paramètres d'analyse mais le plus important d'entre eux, est la recherche de la séquence consensus de N-glycosylation qui est de la forme N-X-S/T (où X peut être n'importe lequel des acides aminés excepté la proline). Nous avons souhaité comparer au moins deux logiciels ayant pour fonction de proposer des sites potentiels de glycosylation. Le plus connu d'entre eux est le logiciel Net-N-Glyc disponible à l'adresse suivante : <<http://www.cbs.dtu.dk/services/NetNGlyc/>><sup>1</sup>. Il existe également le logiciel EnsemblGly, qui lui est disponible à l'adresse suivante : <<http://turing.cs.iastate.edu/EnsembleGly/>><sup>2</sup>. Ces deux logiciels utilisent des algorithmes différents pour prédire les sites de glycosylation potentiels. Plus le score obtenu pour un site est proche de 1, plus ce site a de chance d'être glycosylé.

L'analyse de l'OLFM4 grâce à ces outils nous a permis de déterminer les sites potentiels de glycosylation présentés dans le tableau 2.1 p.137.

Bien que n'ayant pas les mêmes scores en fonction des logiciels utilisés, les sites de glycosylations prédits par les différents algorithmes de prédiction sont les mêmes. D'autre part, la séquence complète en acides aminés de l'OLFM4 est présentée en figure 2.2 p.137.

---

1. Page consultée le 16 avril 2012

2. Page consultée le 16 avril 2012

Tableau 2.1 – Sites de glycosylation potentiels de l'OLFM4 par des outils bioinformatiques

Net-N-Glyc		EnsembleGly	
Sites	Scores	Sites	Scores
72	0,68	72	0,99
136	0,82	136	0,75
193	0,52	193	0,96
253	0,73	253	0,99
352	0,59	352	0,90
411	0,40	411	0,98

```

>sp|Q6UX06|OLFM4_HUMAN
1 -MRPGLSFLALLFFLGQAAGDLGDVGPPIPSPGFSSFPVDSSTSSSFSSSSRSGSSSSRSL
61 -GSGGSVSQLFSNFTGSVDDRGTCCQSVSLPDTTFPVDRLVERLEFTAHVLSQKFEKELSKV
121-REYVQLISVYEKLLNLTVRIDIMEKDTISYTELDFELIKVEVKEMEKLVIQLKESFGGS
181-SEIVDQLEVEIRNMTLLVEKLETLDKNNVLAIRREIVALKTKLKECEASKDQNTPVVHPP
241-PTPGSCGHGGVVNISKPSVVQLNWRGFSYLYGAWGRDYSPPQHPNKGLYWVAPLNTDGRLL
301-EYYRLYNTLDDLLLYINARELRITYQGSGTAVYNNMYVMYNTGNIARVNLTNTTIAV
361-TQTLPNAAAYNNRFSYANVAWQDIDFAVDENGLWVIYSTEASTGNMVISKLNDDTLQVLNT
421-WYTKQYKPSASNAFMVCGVLYATRTMNRTEEIFYYYDTNTGKEGKLDIVMHKMQEKVQS
481-INYNPFDQKLYVYNDGYLLNYDLSVLQKPQ

```

FIGURE 2.2 – Séquence protéique complète de type FASTA de l'OLFM4 humaine extraite de a base de données protéique Uniprot

### 2.2.1.2 Identification de sites de glycosylations grâce à l'enrichissement basé sur les lectines

Grâce à notre approche, nous avons pu démontrer expérimentalement un certain nombre d'entre eux. Pour cela, nous avons utilisé des surnageants de cellules Cos-7 transfectées avec un plasmide codant pour l'OLFM4 pendant 24 heures et au bout de cette période, les cellules sont déplétées en sérum de veau fœtal. Ainsi, la protéine est produite par les cellules, sécrétée dans le milieu privé de sérum pendant 24 heures. Le milieu est ensuite récupéré, concentré à l'aide de centricons (10 kDa). Nous avons ensuite réalisé les étapes de la technique FASP et les peptides ont été digérés à la trypsine. Nous avons ensuite enrichi ces peptides sur un mélange de lectines composé de la lectine WGA et de la lectine ConA. Le tampon est ensuite échangé par de l'eau lourde afin de permettre l'introduction d'un  $^{18}\text{O}$  lors de la déglycosylation par la PNGase F, dans la chaîne latérale de l'acide aspartique et ainsi mettre en évidence les sites déglycosylés.

Ces expériences ont été réalisées trois fois et les résultats de ces expériences sont mentionnés dans le tableau 2.2 p.138.

Tableau 2.2 – Sites de glycosylation potentiels de l'OLFM4 par une approche expérimentale utilisant les lectines

	Sites de glycosylation	Expérience 1	Expérience 2	Expérience 3
Sites ayant été prédits par les logiciels de bioinformatique	72	Oui	Oui	Oui
	136	-	-	-
	193	-	-	-
	253	Oui	-	Oui
	352	Oui	Oui	Oui
	411	-	-	Oui
Sites non prédits par bioinformatique	356	-	Oui	Oui
	366	-	Oui	-

### 2.2.1.3 Identification de sites de glycosylations grâce à l'enrichissement basée sur les billes d'hydrazine

De la même manière que précédemment, une expérience d'enrichissement en glycopeptides a été réalisée sur les billes hydrazines à partir de surnageants de Cos-7 transfectées avec un vecteur codant pour l'OLFM4. Cette expérience a également été réalisée avec une déglycosylation dans de l'eau lourde. Ainsi, cette expérience semble quant elle confirmer les sites en position 352, 356, et 411, comme le résume le tableau 2.3 p.138.

Tableau 2.3 – Sites de glycosylation potentiels de l'OLFM4 par une approche expérimentale utilisant billes hydrazines

	Sites de glycosylation	Expérience 1	Expérience 2
Sites ayant été prédits par les logiciels de bioinformatique	72	-	-
	136	-	Oui
	193	-	-
	253	-	-
	352	Oui	Oui
	411	Oui	-
Sites non prédits par bioinformatique	356	Oui	-

Lors de notre première expérience, nous avons tout d'abord confirmé le site 356 que nous avons détecté avec l'approche utilisant les lectines, et qui n'était pas prédit par les logiciels bioinformatiques. Le fait que ce site non prédit ait été détecté par les deux

approches semble conforter son existence. Cependant, nous n'avons toujours pas détecté les sites 136 et 193.

Nous avons regardé les sites de digestion *in silico* de la trypsine autour de ces sites (cf tableau 2.4 p.139). Ainsi, nous avons remarqué que les peptides porteurs de ces sites étaient particulièrement petits avec une masse de 828 Da pour le peptide porteur du site 136 et une masse de 947 pour celui porteur du site 196. Or, lors des analyses protéomiques nous réalisons les acquisitions en SM entre 1 000 et 4 000 Da donc ces peptides n'étaient pas détectés par le spectromètre de masse. Pour l'expérience 2 avec l'approche hydrazine, nous avons réglé les paramètres du spectromètre de masse afin de réaliser l'acquisition en SM entre 800 et 4 000. Nous avons alors réussi à identifier le site 136.

Afin d'améliorer la qualité de la détection, nous avons également souhaité utiliser une autre enzyme que la trypsine pour digérer les protéines avant l'analyse. L'enzyme Glu-C est souvent utilisée dans la littérature pour les expériences de glycoprotéomique, et nous avons souhaité tester cette enzyme. Malheureusement, nous n'avons pas réussi à obtenir des conditions de digestion optimales pour permettre l'enrichissement en peptides glycosylés.

Tableau 2.4 – Liste et masse des peptides tryptiques de l'OLFM4 détectables entre 1 000 et 4 000 Da

masse	position	Sequences peptidiques
3673.8655	231-265	DQNTPVVHPPPTPGSCGHGGVVNISKPSVVQLNWR
3086.4095	323-350	ITYGQGSQTAVYNNMYVNMNTGNIAR
3056.4119	21-51	DLGDVGPPIPSPGFSSFPDSSSSSFSSSR
2518.2864	490-510	LYVYNDGYLLNYDLSVLQKPQ
2389.2470	351-372	VNLTTNTIAVTQTLPNAAAYNNR
2217.0418	59-80	SLGSGGSVSQLFSNFTGSVDDR
2206.0783	425-444	QYKPSASNAFMVCGVLYATR
1993.9712	175-192	ESFGGSSEIVDQLEVEIR
1925.8731	81-98	GTCQCSVSLPDTTFPVDR
1809.9745	305-319	LYNTLDDLLLYINAR
1809.9381	410-424	LNDTTLQVLNTWYTK
1743.7748	450-463	TEEIFYYYDTNTGK
1686.8472	147-160	DTISYTELDFELIK
1461.7484	286-298	GLYWVAPLNTDGR
1452.7117	478-489	VQSINYNPFDQK
1370.7202	122-132	EYVQLISVYEK
1276.6109	266-276	GFSYLYGAWGR
1272.6946	102-112	LEFTAHVLSQK
1085.5010	277-285	DYSPQHPNK

À la lumière des résultats de ces deux approches, nous avons tout d'abord observé

Tableau 2.5 – Liste et masse des peptides issus de la digestion pas l'enzyme Glu-C de l'OLFM4

masse	position	Sequences peptidiques
2421.2448	302-320	YYRLYNTLDDLLLYINARE
1685.9982	132-145	KKLLNLTVRIDIME
1541.8758	203-215	TLDKNNVLAIRRE
1513.6845	453-464	IFYYYDTNTGKE
1428.7337	465-476	GKLDIVMHKMQE
1306.6790	104-114	FTAHVLSQKFE
1181.5837	390-399	NGLWVIYSTE
1142.7507	216-225	IVALKTKLKE
1113.5826	123-131	YVQLISVYE
1088.6132	191-199	IRNMTLLVE

que nous identifions deux positions qui n'étaient pas prédites par l'analyse bioinforma-  
tique, et pour la position 356, cela a été observé 2 fois. Cela suggère que la prédiction  
des sites est probablement incomplète lorsque l'on utilise uniquement l'analyse des sé-  
quences consensus. Ces sites sont contenus dans le peptide ayant la séquence suivante :  
VNLTTN<sub>356</sub>TIAVTQTLPN<sub>366</sub>AAYNNR.

### 2.2.2 Validation des sites de glycosylation par mutagenèse dirigée

Grâce à notre approche de glycoprotéomique, nous avons pu analyser plusieurs sites  
de glycosylation potentiels de l'OLFM4 et de la sorte, nous avons souhaité, d'une part  
les confirmer, d'autre part analyser l'effet de la perte de la glycosylation sur sa sécrétion.  
Nous avons choisi de réaliser des mutants sur les asparagines situées en position 72, 352,  
253 et 356 car ces positions semblent être celles pour lesquelles nous sommes le plus sûr  
de la pertinence, étant donné leur redondance. Nous avons également réalisé un double  
mutant sur les positions 72 et 352. Les asparagines de la séquence sauvage de l'OLFM4  
ont été substitués par des alanines. Un séquençage a été réalisé dans le but de vérifier  
que la mutation a bien été intégrée dans la séquence initiale. Afin de confirmer ces sites  
de glycosylation, nous avons transfecté des cellules Cos-7 avec ces mutants et nous avons  
réalisé un *western blot* afin de voir si nous observions des variations de taille de la protéine  
liées à la glycosylation. Comme le montre la figure 2.3C p.141 il y a bien une différence de  
la taille de l'OLFM4 suite à la mutation des différents sites de glycosylation. Cela confirme  
que ces sites sont réellement des sites glycosylés de cette protéine.

Comme le montre la figure, nous avons également réalisé un mutant double sur les  
positions 72 et 352. La sécrétion est très nettement diminuée en comparaison aux mutants

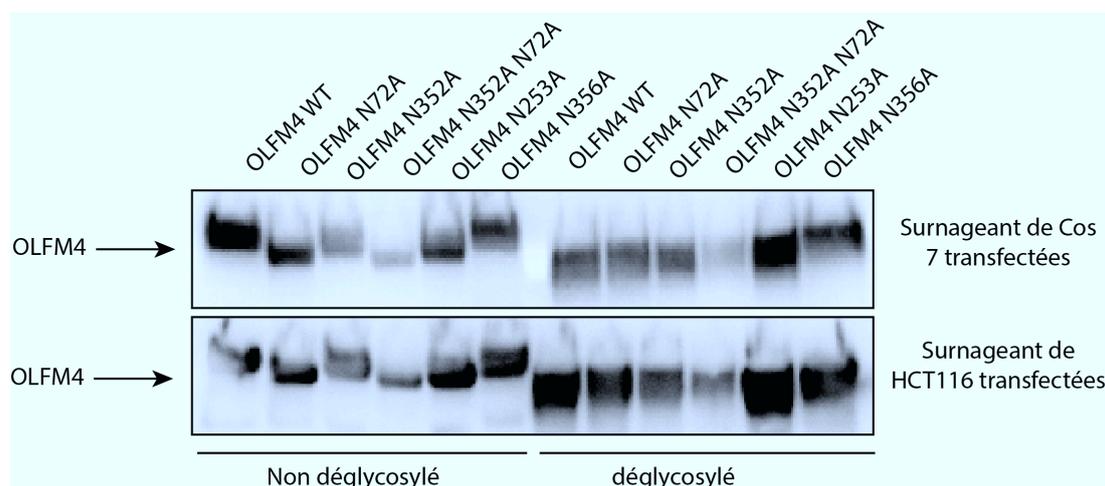


FIGURE 2.3 – **Validation des sites de glycosylation par *western blot*** – Des cellules COS7 et des cellules HCT116 ont été transfectées avec les plasmides codant pour les formes mutantes de l’OLFM4 puis les surnageants ont été analysés par western blot.

simples et au vecteur contenant la forme sauvage de la protéine. Cela semble confirmer le fait que la glycosylation de l’OLFM4 est nécessaire à sa sécrétion.

D’autre part, cela confirme que les sites que nous avons identifiés par spectrométrie de masse sont bien tous des sites de glycosylation car les formes mutantes ont toutes des tailles différentes de la protéine sauvage dû à l’absence du motif glucidique dont la taille est plus ou moins importante.

Enfin, nous avons pu identifier un site de glycosylation non prédit par la bioinformatique, et ne faisant pas parti d’une séquence consensus. Cela suggère qu’il peut exister des N-glycosylations, en dehors des séquences consensus.

## 2.3 L’OLFM4 : une protéine impliquée dans la survie cellulaire

Nous avons ensuite souhaité savoir quels bénéfices pouvaient apporter l’OLFM4 à la tumeur en terme de survie ou de résistance aux traitements. Pour cela, nous avons réalisé des tests de clonogénicité sur les cellules HCT116 ayant la capacité d’induire l’expression de l’OLFM4 sous l’effet de la doxycycline grâce à un système inductible. Tout d’abord, le modèle inductible a été validé en regardant, par *western blot*, la présence d’OLFM4 dans les surnageants de culture après l’induction (cf figure 2.4(A) p.142). Ensuite, nous avons observé une faible augmentation du nombre de colonies sous l’effet de l’induction de l’expression de l’OLFM4 qui est malgré tout significative. De plus, nous avons observé que cette augmentation d’expression conférait aux cellules HCT116 une légère résistance au Sn38, l’un des principaux traitements des cancers colorectaux métastatiques (cf figure

2.4(A) p.142).

Nous avons également souhaité voir l'effet inverse, en inactivant l'expression de l'OLFM4 par ARN interférence dans des cellules présentant une expression basale importante, les LS180. Des tests de clonogénicité ont été réalisés suite à ces expériences d'ARN interférence et ces tests montrent une diminution significative (50%) du nombre de colonies formées dans la condition test, par rapport à la condition contrôle. Des expériences complémentaires de cytométrie en flux ont été réalisées afin d'explorer le mécanisme expliquant cette mort cellulaire (données non présentées). Ces résultats montrent une augmentation du clivage de la caspase 3 suite à la transfection de l'ARN interférence dirigé contre l'OLFM4 suggérant que cette dernière protège les cellules de l'apoptose.

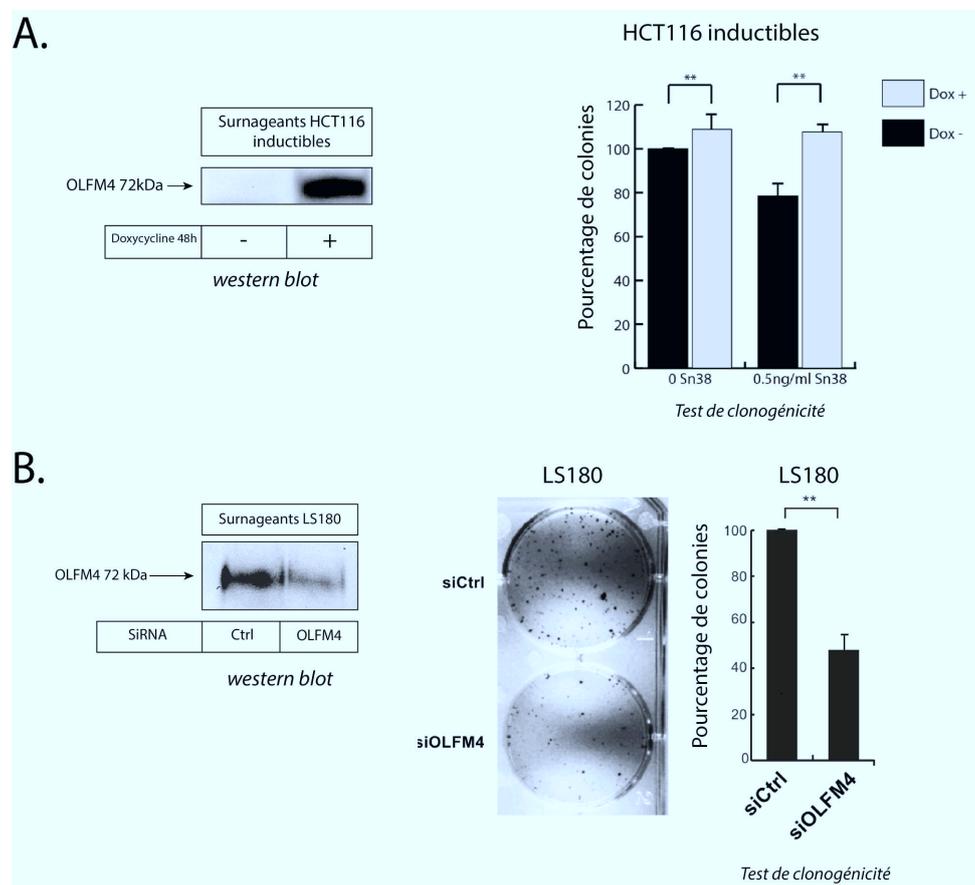


FIGURE 2.4 – **L'OLFM4 améliore la survie des cellules tumorales.** (A) Western blot réalisé sur les surnageants de cellules HCT116 ayant la capacité d'induire l'expression de l'OLFM4 sous l'effet de la doxycycline (n=6). L'induction de l'expression de l'OLFM4 augmente le nombre de colonie dans les puits non traités, et facilite la résistance au Sn38 des puits traités (n=6). (B) Un ARN interférence dirigé contre l'OLFM4 dans des cellules LS180 permet de diminuer la quantité d'OLFM4 présente dans le milieu extracellulaire (n=5). Cette diminution a pour effet d'augmenter la mort de ces cellule par rapport à l'ARN interférence contrôle (n=3).

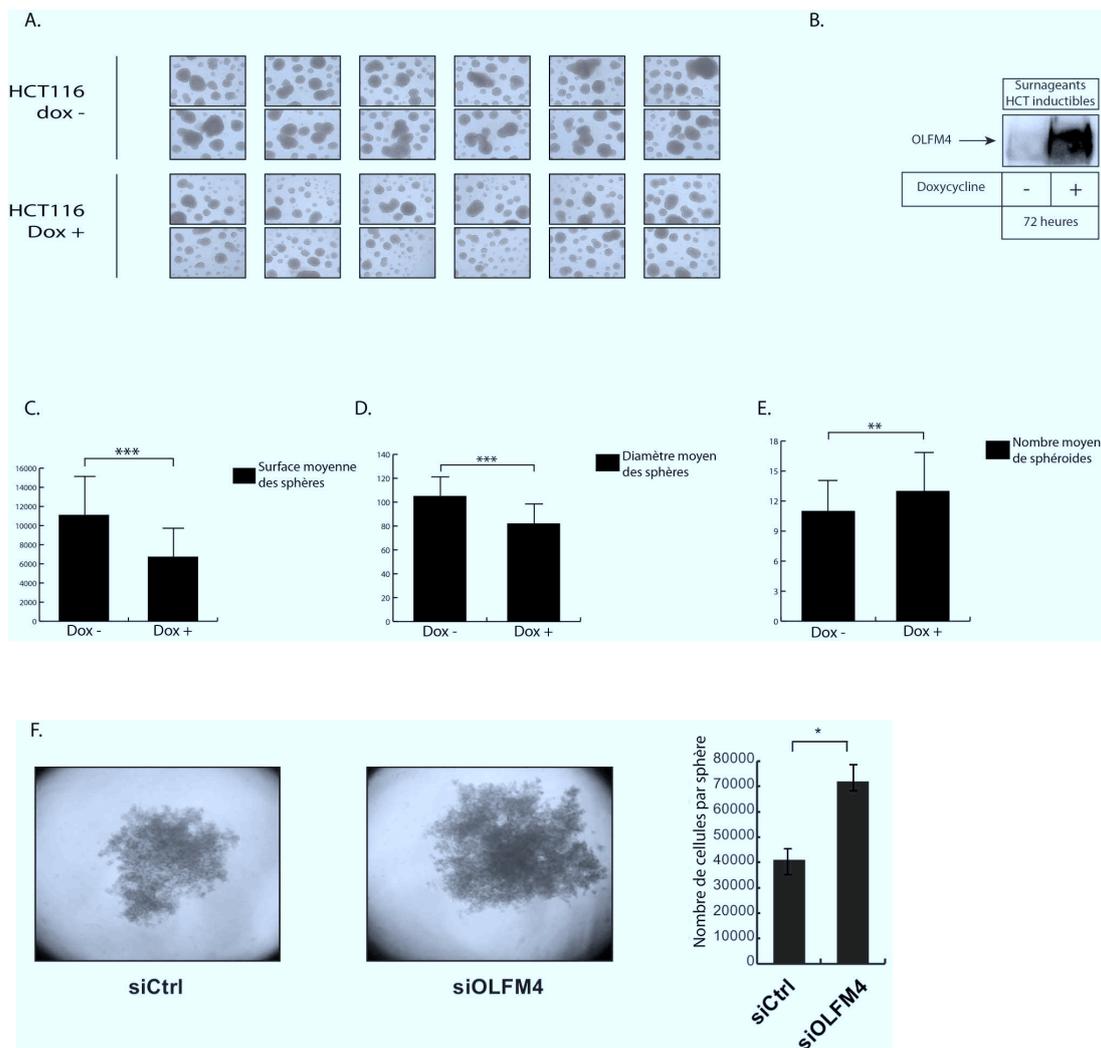
## 2.4 L'OLFM4 : diminue la capacité des cellules à former des structures tridimensionnelles

L'OLFM4 est une protéine connue également pour être une protéine intervenant dans les contacts cellules-cellules, de la même manière que le collagène de type I ou IV, que l'élastine, la laminine ou encore la fibronectine. Nous avons testé la capacité de ces cellules à former des structures en trois dimensions dans des conditions de culture particulières [Reynolds and Weiss, 1992]. De façon surprenante, l'induction de l'olfactomedine 4 dans des cellules HCT116 conduit à une diminution de la taille des structures 3D (tant au niveau de l'aire des structures formées qu'au niveau de leur diamètre) et à une augmentation du nombre de sphères formées (cf figure 2.5 p.144). Au total, ce sont 96 images du type de celles présentées en figure 2.5 p.144, pour chacune des deux conditions, (soit 192 images) réparties selon 3 expérimentations indépendantes de 32 puits. Afin de s'affranchir d'erreurs potentiellement inhérentes au logiciel imageJ, nous n'avons pas utilisé les paramètres automatiques, mais nous avons réalisé les analyses manuellement en délimitant les zones dont nous souhaitions obtenir la surface.

Ces résultats suggèrent que cette protéine intervient probablement dans la régulation de la dissociation cellulaire et dans la formation de structures tridimensionnelles. On peut également supposer que ces propriétés de diminution des contacts cellules-cellules en fait une protéine qui pourrait potentiellement intervenir dans le processus de dissémination métastatique.

L'expérience inverse a également été réalisée en réalisant des sphères à partir de cellules LS180 sur lesquelles l'expression de l'OLFM4 a, ou non, été inhibée par ARN interférence (cf figure 2.5(F) p.144). Dans cette expérience, la comparaison de la taille des sphères a été réalisée par comptage du nombre de cellules de ces dernières. Cela a été rendu possible par le fait qu'une seule structure était présente dans chaque puits permettant de les comparer les uns aux autres. On remarque ainsi, que lorsque l'expression de l'OLFM4 est diminuée dans les cellules LS180, les sphères formées semblent plus grosses, et semblent contenir plus de cellules.

L'ensemble de ces résultats a été obtenu sans que ne soient discriminé l'effet intracellulaire, et l'effet extracellulaire de l'OLFM4. Avant d'aller plus loin dans ces analyses il pourrait être pertinent d'utiliser les mutants de glycosylation empêchant la sécrétion de l'OLFM4 pour identifier si ces effets sont le fruit de l'action de l'OLFM4 sur un récepteur



**FIGURE 2.5 – L'olfactomedine 4 induit une diminution de la taille des sphéroides**  
 (A) Images de quelques sphéroides obtenus avec des cellules HCT116 dans lesquelles l'expression de l'OLFM4 est induite ou non. (B) *Western Blot* montrant l'induction de l'olfactomedine 4 (n=6). (C) Surface moyenne des sphéroides présents sur chaque image (n=3). (D) Diamètre moyen des sphéroides présents sur chaque image (n=3). (E) Nombre moyen de sphéroides présents sur chaque image (n=3). (F) Images de sphéroides obtenus avec des cellules LS180 transfectées soit avec un ARN interférence contrôlé, soit avec un pool de deux ARN interférence dirigé contre l'OLFM4. Le nombre de cellules constituant les sphères a été compté après 10 jours de culture et dissociation tryptique (n=3).

ou s'il s'agit de l'activation de voies de signalisation en intracellulaire.

## Chapitre 3

### Résultats complémentaires : l'OLFM4 comme potentiel marqueur de la mutation de RAS

#### 3.1 Analyse génétique de tumeurs colorectales

EN préalable de notre étude protéomique, nous avons cherché à caractériser les tumeurs que nous avons à notre disposition d'un point de vue génétique. Nous avons alors cherché, par pyroséquençage, les mutations importantes décrites dans le cancer colorectal sur les gènes KRAS, BRAF, P53, et PI3KCA. Le tableau 3.1 p.145 récapitule les résultats de cette analyse réalisée sur 79 patients.

Tableau 3.1 – Résultats de l'analyse des mutations sur 79 tumeurs du côlon

Gène (proportion de mutation)	Mutation	Nombre de patients
TP53 (20%)	R175H	5
	E282Q	4
	R273C	3
	C176T	2
	R248W	1
	R273H	1
PI3KCA (5%)	E545K	2
	E542K	2
KRAS (33%)	G12D	12
	G13D	7
	G12V	3
	G12C	2
	G12A	1
	Q61K	1
BRAF (8%)	V600E	6
Patients non mutés (44%)		35

Nous avons mis en évidence le fait que 20% des tumeurs colorectales étaient mutées

sur le gène suppresseur de tumeurs TP53, 5% étaient mutées sur le gène de la PI3KCA, 8% étaient mutées sur le gène BRAF, et 33% étaient mutées sur l'oncogène KRAS. Par ailleurs, 44% des tumeurs ne présentaient pas de mutation connues sur ces gènes. Nous avons décidé de focaliser notre attention sur les tumeurs mutées sur l'oncogène KRAS car il s'agit de la mutation la plus fréquente dans le cancer colorectal, et la mutation de cet oncogène implique une résistance aux traitements de chimiothérapie ciblée. Ainsi, nous espérons mettre en évidence un biomarqueur potentiellement sécrété par les tumeurs et qui serait un marqueur de la mutation.

Au final, huit des tumeurs que nous avons analysées par spectrométrie de masse exprimaient des formes sauvage de RAS, et treize exprimaient des formes mutées. L'analyse par pyroséquençage nous a donné six tumeurs présentant une mutation G12D, deux G12V, une G12A et une G12C. Trois tumeurs portaient une mutation de type G13D. Aucune mutation sur le codon 61 n'a été détectée chez ces 21 tumeurs.

Le schéma expérimental que nous avons suivi est présenté en figure 3.1 p.146.

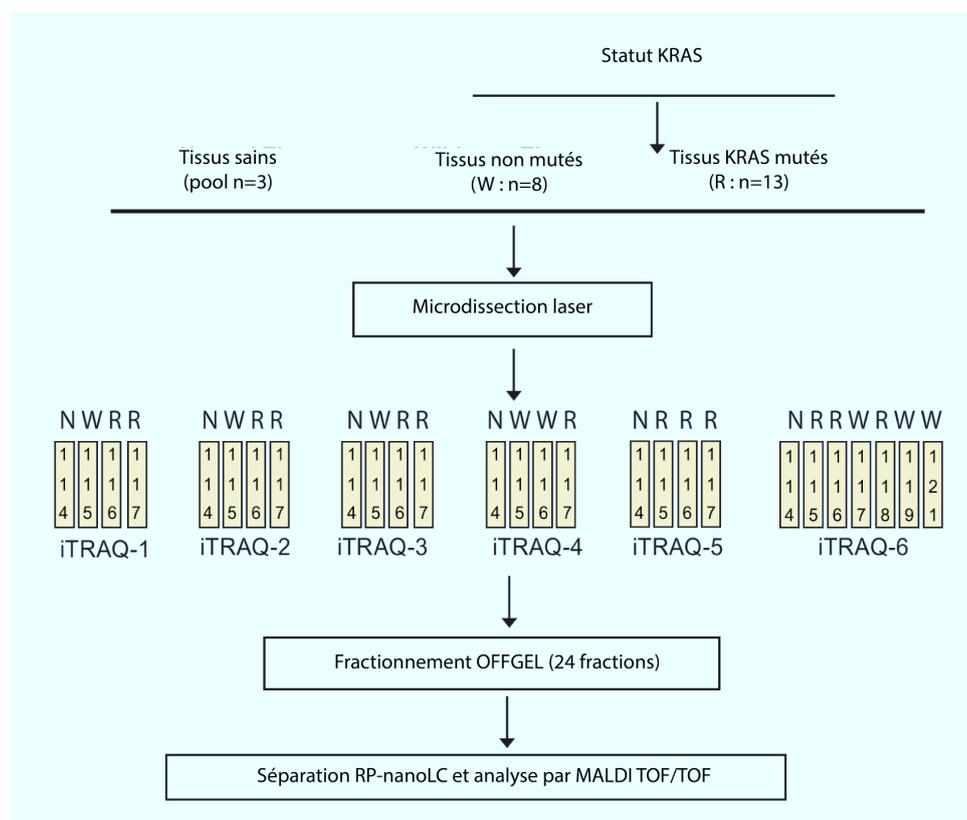


FIGURE 3.1 – Schéma expérimental de l'analyse des tumeurs colorectales en fonction du statut de l'oncogène KRAS

## 3.2 Analyse protéomique de tumeurs mutées ou non sur l'oncogène RAS

Cette analyse nous a permis d'identifier et de quantifier 2 692 protéines avec au moins deux peptides. Grâce à cette approche, nous avons détecté 256 protéines dont l'expression varie significativement entre les tumeurs présentant une forme sauvage de l'oncogène RAS et le tissu sain. De la même façon, 253 protéines varient de façon statistiquement significative entre les tumeurs mutées sur KRAS et le pool de tissu sain. Le détail de protéines surexprimées et sous-exprimées est présenté dans le tableau 3.2 p.147.

Tableau 3.2 – **Résultat de l'analyse protéomique de tumeurs mutées sur KRAS et non mutées comparées à un pool de tissu sain**

	KRAS WT	KRAS mutées
Protéines surexprimées	107	113
Protéines sous-exprimées	149	140

Dans le détail, ces données montrent la surexpression de l'OLFM4 dans les tumeurs mutées sur l'oncogène RAS avec un rapport iTRAQ<sup>TM</sup> de 1,68 (Cf figure 3.3(A) p. 151) associé à un intervalle de confiance statistique de [1,35 ; 2,10].

Au passage, nous pouvons également noter la variation d'autres protéines importantes telles que l'ALDH1B1. Cette protéine est augmentée de près de deux fois par rapport au tissu sain avec un rapport iTRAQ<sup>TM</sup> de 1,9 et un intervalle de [1,59 ; 2,28].

Nous pouvons également observer que la protéine CEACAM5, connue pour être un marqueur de récurrence du cancer colorectal, est également fortement augmentée dans les tumeurs mutées sur KRAS avec un rapport iTRAQ<sup>TM</sup> de 1,69 pour un intervalle de confiance de [1,46 ; 1,95].

On peut également mentionner l'identification de la protéine TGFBI, impliquée dans la modulation des interactions cellule-cellule. Cette dernière aurait un rôle dans la diminution de ces interactions. Elle est sur-exprimée d'un rapport 1,82 pour un intervalle de confiance de [1,59, 2,08].

La protéine sorbitol déshydrogénase est également sur-exprimée dans les tumeurs mutées sur l'oncogène KRAS avec un rapport iTRAQ<sup>TM</sup> de 1,59 associé à un intervalle de confiance de [1,26 ; 2,00]. Cette protéine faisait également partie des quatre candidats biomarqueurs obtenus suite à l'analyse des tumeurs en fonction de leur stade clinique.

Ces données ont pu être confirmées par les marquages immunohistochimiques de

l'OLFM4 réalisés sur des tumeurs colorectales (cf figure 3.3(B) p.151). En effet, les tumeurs présentant une mutation de KRAS, présentent également un important marquage immunohistochimique de l'OLFM4 suggérant une importante expression de la protéine.

D'autre part, des *western blot* ont été réalisés sur des extraits de tumeurs en fonction de leur statut pour l'oncogène RAS. En effet, ce sont au final 29 tumeurs qui ont pu être analysées par *western blot* et des rapports d'intensité des bandes correspondant aux différentes formes ont pu être calculés. Pour chacune des tumeurs nous avons réalisé une quantification de l'intensité des bandes par détermination du nombre de pixel. À chaque fois, la bande supérieure glycosylée (72kDa) était comparée à la bande inférieure non glycosylée (55kDa). Parmi ces tumeurs, il y avait 12 tumeurs mutées sur l'oncogène RAS – 5 de stade 2, 3 de stade 3 et 4 de stade 4 – et 17 tumeurs non mutées – 5 de stade 2, 6 de stade 3 et 6 de stade 4 – réparties relativement uniformément en fonction de leur stade histologique (cf tableau 3.3 p.148). Malgré tout, nous n'avions pas à notre disposition de tumeurs de stade 1 pour procéder à une analyse complète de tous les stades. Cependant, comme le montre la figure 3.3(C) p.151, la mutation implique une forte augmentation de l'expression de l'OLFM4, et plus particulièrement de la forme glycosylée (72kDa) (cf figure 3.3(D) p.151).

Tableau 3.3 – **Caractéristiques des tumeurs analysées par western blot**

Statut KRAS	Stade	Nombre de tumeurs
Muté	2	5
	3	3
	4	4
Non muté	2	5
	3	6
	4	6

### 3.3 Confirmation du lien entre l'OLFM4 et KRAS

Afin de confirmer le lien entre l'expression de l'OLFM4 et celui de l'oncogène RAS, nous avons utilisé des modèles cellulaires. Ainsi des cellules d'adénocarcinome colorectal HT29, transformées pour posséder la capacité d'induire l'expression de la protéine H-RAS<sup>V12</sup> sous la dépendance d'un promoteur inductible à la doxycycline ont été utilisées. Ces cellules ont été stimulées à la doxycycline pendant 48 et 72h afin d'induire l'expression de la protéine Ras. Cette expression est confirmée à 48h par le *western blot* présenté en figure 3.3(E) p.151. Vingt quatre heures avant la récupération des échantillons, le milieu de culture a été

remplacé par du milieu ne contenant pas de sérum de veau fœtal. Ce dernier est en effet incompatible avec une analyse par *western blot* du fait de sa très forte concentration en protéines. Une privation de sérum de 24h suite à une induction préalable de l'expression de la protéine Ras montre, après une concentration équivalente sur des centricons de valeur seuil de 10 kDa une présence de l'OLFM4 dans les surnageants de culture alors que le contrôle non induit n'en possède pas (cf figure 3.3(E) p.151). Le niveau d'expression de l'ARNm de l'OLFM4 a également été analysé sur les lignées d'adénocarcinomes colorectaux présentes au laboratoire. Ces résultats montrent que toutes les lignées mutées sur RAS n'expriment pas nécessairement l'OLFM4, mais en revanche, toutes les lignées exprimant cette dernière sont mutées sur RAS (cf tableau 3.4 p.149).

Tableau 3.4 – Niveau d'expression de l'ARNm de l'OLMF4 dans des lignées de cellules colorectales en fonction du statut de RAS.

Lignée cellulaire	Mutation de RAS	Expression OLFM4	Localisation	Stade clinique
HT29	WT	-	côlon	2
SW48	WT	-	côlon	4
RKO	WT	-	côlon	Non communiqué
SW1417	WT	-	côlon	3
WIDr	WT	-	côlon	Non communiqué
Caco-2	WT	-	côlon	Non communiqué
Colo320	WT	-	côlon	3
HCT116	G13D	-	côlon	Non communiqué
HCT15	G13D	-	côlon	3
SW480	G12V	-	côlon	2
SW620	G12V	-	côlon	4
DLD-1	G13D	-	côlon	3
LS174T	G12D	+	côlon	2
SW403	G12V	++	côlon	3
SW1463	G12C	+++	rectum	3
LS180	G12D	+++	côlon	2

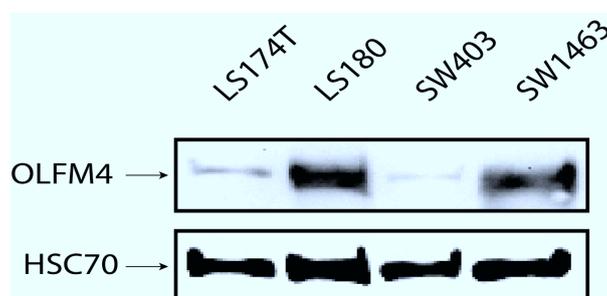


FIGURE 3.2 – Western blot de validation de l'expression de l'OLFM4 dans les lignées LS174t, LS180, SW1463, et SW403

Quatre lignées présentent donc un niveau d'ARNm élevé pour l'OLFM4. Nous avons

souhaité confirmer cette observation au niveau protéique et nous avons réalisé un *western blot* sur les lignées LS174t, LS180, SW403, et SW1463. Les résultats présentés en figure 3.2 p.149 montrent que ces lignées expriment effectivement, non seulement l'ARNm, mais aussi la protéine OLFM4.

### 3.4 Validation de l'OLFM4 comme marqueur de la mutation de KRAS

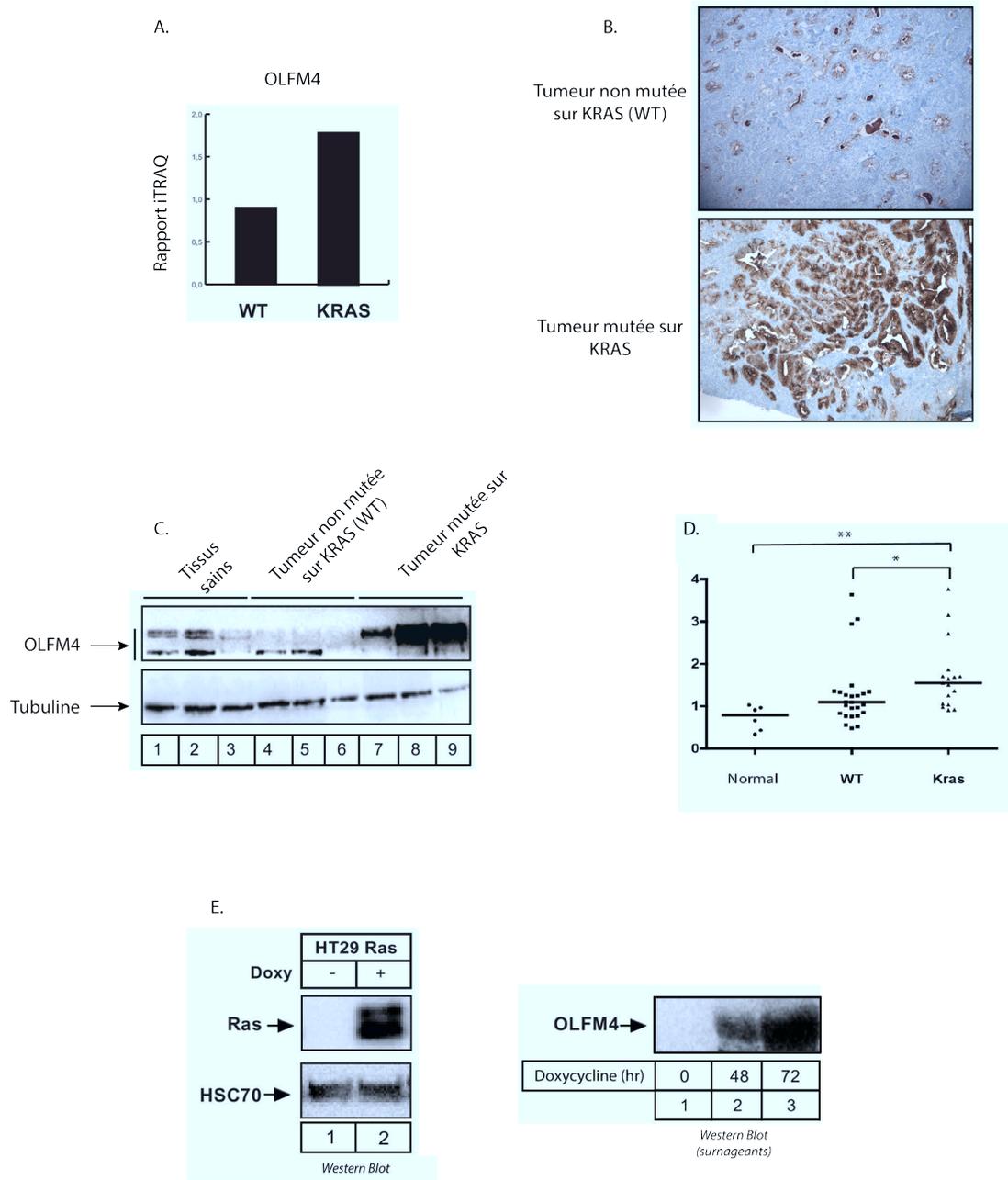
Nous avons pu nous procurer les échantillons de 45 patients que nous avons tester à l'aide d'un kit ELISA commercial. Les données sont présentées en figure 3.4 p.152, et elles montrent qu'il est tout à fait possible de détecter l'OLFM4 dans les séra. Les valeurs actuelles varient des taux inférieurs au seuil de détection de l'ELISA, à 225 ng/mL. Le taux basal semble quant à lui se situer autour d'une valeur de 20 ng/mL  $\pm$  6 ng/mL.

Ces résultats préliminaires demandent à être confirmés sur une cohorte de patient plus importante, mais les premiers résultats à notre disposition semblent montrer la possibilité de doser simplement cette protéine dans le sang des patients, et les dosages révèlent une variation entre les patients non muté sur et KRAS et ceux porteurs d'une tumeur mutée.

L'utilisation de l'OLFM4 comme candidat biomarqueur précoce du développement du cancer colorectal, ou comme candidat biomarqueur de la mutation de KRAS a fait l'objet d'un brevet<sup>1</sup>.

---

1. Use of the olfactomedin-4 protein (OLFM4) in colorectal cancer diagnosis. Référence PCT/IB2011/000787 le 2 mars 2011, Catherine Guette, Olivier Coqueret, Benjamin Barre, Erick Gamelin.



**FIGURE 3.3 – L'OLFM4 est fortement exprimée dans les tumeurs mutées sur l'oncogène RAS** (A) Rapport iTRAQ<sup>TM</sup> de la variation du niveau d'expression de l'OLFM4 en fonction du statut génétique des tumeurs pour l'oncogène RAS. (B) Marquage immunohisto chimique d'une tumeur non mutée sur RAS et d'une tumeur mutée avec un anticorps dirigé contre l'OLFM4 (C) Western blot montrant les deux formes de l'olfactomedine 4 réalisés sur des extraits de tumeurs. (D) Représentation des rapports d'intensité des bandes à 55 et 72 kDa en fonction de la présence ou non de mutation de RAS et en comparaison avec du tissu normal. (E) Des cellules d'adénocarcinome colique HT29 ont été stimulée à la doxycycline afin d'induire l'expression de la protéine Ras. Des cellules HT29 ont été traitées à la doxycycline pendant 48h et 72h afin d'induire l'expression de la protéine Ras. Au bout de 24h pour la première condition, et de 48h pour la deuxième, le milieu de culture a été remplacé avec du milieu ne contenant pas de sérum de veau foetal. Puis un western blot dirigé contre l'OLFM4 a été réalisé sur ces surnageants après concentration sur des centrions ayant une valeur seuil de 10 kDa.

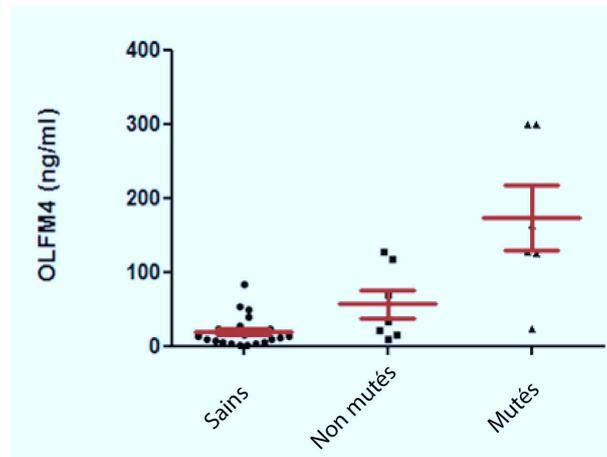


FIGURE 3.4 – L’OLFM4 est une protéine sécrétée, détectable dans les séra de patients et un marqueur potentiel de la mutation de KRAS – Résultats préliminaires de tests ELISA réalisés sur des séra de patients atteints de cancers colorectaux mutés ou non sur KRAS. Le taux basal d’OLFM4 est de 20 ng/mL

# Chapitre 4

## Étude du stroma de tumeurs colorectales

### 4.1 Microdissection laser

DANS le but d'identifier de nouveaux biomarqueurs potentiels, la plupart des analyses protéomiques globales étudient les cellules épithéliales tumorales. Nous avons souhaité réaliser une étude préliminaire portant sur le stroma tumoral. Cette zone conjonctive entourant la tumeur est généralement constituée de cellules moins actives métaboliquement. Il est cependant démontré que ces cellules sont sensibles aux signaux envoyés par la tumeur, et peuvent avoir un effet sur celles-ci.

Nous avons donc analysé le stroma de deux tumeurs du côlon par spectrométrie de masse, après microdissection laser de ce dernier (cf figure 4.1 p.153). La tumeur n°1 est une tumeur de stade 4 et présente une substitution Q61K sur le gène de KRAS. La tumeur n°2 est une tumeur de stade 3, et elle présente une substitution G12D sur de codon 13 de KRAS, et également une substitution R175H sur le gène TP53.



FIGURE 4.1 – **Microdissection laser du stroma tumoral** (A) Image de la coupe tumorale déshydratée avec les impacts laser sur les cellules stromales (B) Image de la coupe après microdissection du stroma. Il ne reste plus que les cellules épithéliales tumorales sur la coupe (C) Cellules du stroma fixées sur la capsule.

## 4.2 Résultats de l'étude par spectrométrie de masse

Les échantillons ont été normalisés à 100  $\mu\text{g}$  de peptides chacun, et analysés séparément. Nous avons ainsi pu identifier 245 protéines au total (cf annexe n°4), avec au moins deux peptides. Un total de 11 102 spectres a été acquis en cumulant les deux analyses, permettant l'identification de 2 925 peptides uniques. La tumeur n°1 a donné lieu à l'identification de 1 326 peptides uniques. À partir de ces données, 187 protéines ont été identifiées avec au moins deux peptides et une confiance supérieure à 95%. La tumeur n°2 a permis l'acquisition de 1 826 peptides uniques. Ces données ont permis l'identification de 156 protéines avec au moins deux peptides et une confiance supérieure à 95%. En comparant les deux listes de protéines identifiées dans chaque tumeur, nous avons observé que 98 protéines étaient communes (cf figure 4.2(A) p.155).

Une analyse ontologique des 245 protéines identifiées a été réalisée à l'aide du logiciel GoMiner (cf figure 4.2(B) p.155). Les résultats indiquent que 70 de ces protéines sont localisées dans le compartiment extracellulaire soit 28%. Nous avons également constaté que 14 de ces protéines étaient des protéines membranaires (soit 6%), et que 189 étaient localisées au niveau des organelles cellulaires soit 77%.

Parmi les 245 protéines identifiées, une analyse bioinformatique réalisée à l'aide des logiciels SecretomP et SignalP montre que 96 protéines sont susceptibles d'être sécrétées soit 39,2% (cf figure 4.2(C) p.155). En effet, 56 de ces protéines sont porteuses d'un peptide signal selon le logiciel SignalP – donc sont sécrétées – et 40 d'entre elles peuvent être sécrétées selon le logiciel SecretomP malgré l'absence de peptide signal. Ici, toutes les protéines pour lesquelles un peptide signal a été mis en évidence par le logiciel SignalP ont été également identifiées par le logiciel SecretomP.

Par ailleurs, à l'aide de deux articles références traitant de l'identification de protéines présentes dans le plasma de patients, nous avons pu établir une base de données de 1 211 protéines sécrétées et détectables [Schenk et al., 2008] [Cao et al., 2013]. Ces études se sont attachées à mettre en évidence les protéines du plasma humain à l'aide de méthode d'identification particulièrement stringentes de sorte qu'ils estiment les protéines identifiées comme sûres. Nous avons donc compilé leurs données, et nous avons comparé le protéome du stroma tumoral obtenu précédemment. De la sorte, ce sont 124 des 245 protéines identifiées qui sont détectables dans le plasma de patients, soit 50,6% (cf figure 4.2(D) p.155). On remarque donc que près d'une protéine sur deux, identifiée dans nos échantillons de stroma tumoral est une protéine identifiable dans le plasma de patients.

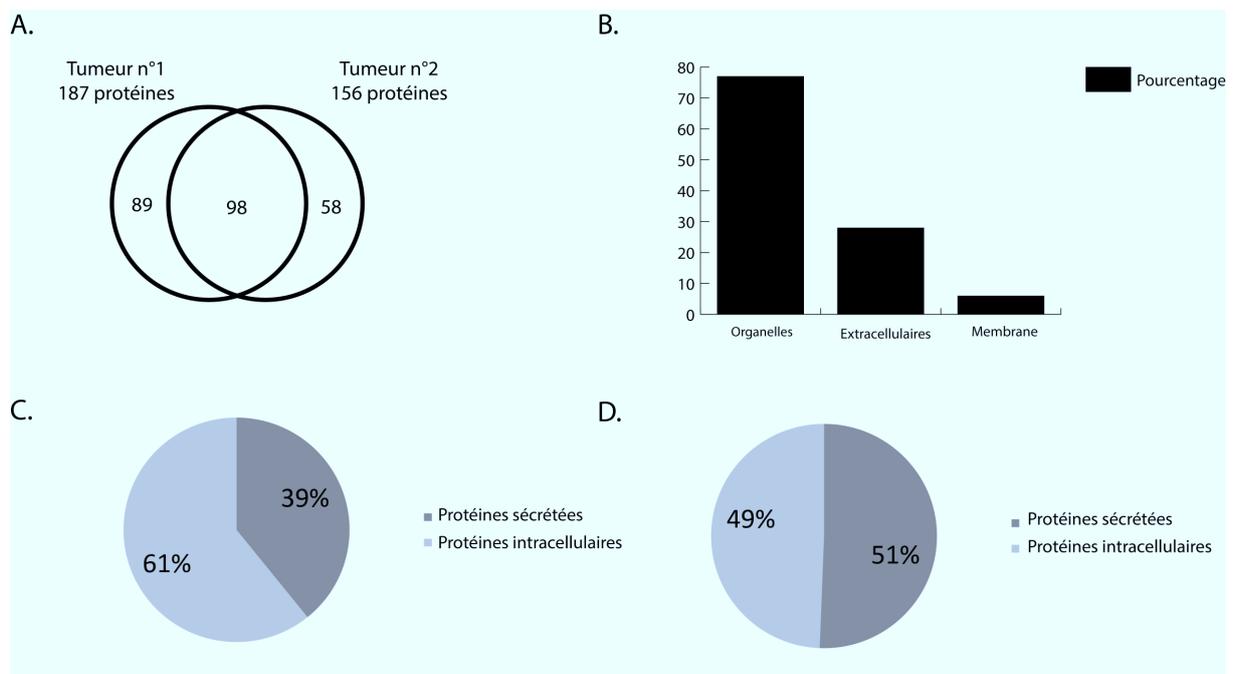


FIGURE 4.2 – **Analyse du protéome issu du stroma de deux tumeurs coliques** (A) Diagramme de Venn présentant les résultats de l'analyse protéomique des deux stromas tumoraux (B) Résultats de l'analyse ontologique des protéines identifiées dans les deux tumeurs (C) Représentation du pourcentage de protéines intracellulaires et extracellulaires selon l'analyse bioinformatique (SignalP et SecretomP) (D) Représentation du pourcentage de protéines intracellulaires et extracellulaires selon la comparaison avec la base de données que nous avons établie à l'aide des deux articles cités plus haut [Schenk et al., 2008] [Cao et al., 2013].

Parmi les protéines identifiées, il est intéressant de noter la présence de l'OLFM4 dont nous avons largement parlé en première partie de ces résultats.

D'autre part, il est intéressant de noter l'identification d'un grand nombre de protéines impliquées dans le remodelage de la matrice extracellulaire, telles que la vimentine, le TGF $\beta$ i, la décorine, ou encore la protéine versican. Ces protéines pourraient avoir un rôle dans le processus à l'origine de la dissémination tumorale et l'apparition de métastases.

La vimentine est une protéine qui est utilisée, en association avec la cadhérine, comme marqueur de la transition épithélio-mésenchymateuse.

La protéine TGF $\beta$ i est, quant à elle, une protéine contenant un domaine RGD lui permettant de se fixer au collagène de type 1, 2 et 4 et participant ainsi à la régulation de la matrice extracellulaire. Cette protéine, lorsqu'elle est présente, diminue l'interaction des cellules entre elles. En effet, une mutation du gène est impliquée dans des formes sévères de dystrophies cornéennes. L'épithélium de cette dernière ne remplit plus son rôle car sa structure est altérée par la mutation.

La décorine est une protéine faisant partie de la matrice extracellulaire puisqu'il s'agit

d'un protéoglycane. Elle semble également être altérée dans certains types de dystrophie de la cornée.

Enfin, le versican est une grosse protéine de la matrice extracellulaire également. Cette protéine est souvent considérée comme limitant l'adhésion entre les cellules.

Nous pouvons également remarquer l'identification de la protéine S100A8 qui est un marqueur potentiel déjà décrit dans le cancer colorectal. Cette protéine semble jouer un rôle important dans la régulation des processus inflammatoire et de la réponse immunitaire.

# Chapitre 5

## Recherche de biomarqueurs à partir de lignées cellulaires

### 5.1 Une approche protéomique globales

**A**FIN de compléter notre étude, nous avons également réalisé une étude protéomique de lignées cellulaires d'adénocarcinomes coliques afin d'établir une cartographie protéomique de ces dernières. De la sorte, nous avons analysé les 10 lignées suivantes obtenues auprès de l'ATCC :

- HCT116, HT29, LS174T, RKO, SW48, SW1417, SW1463, SW480, SW620, SW48.

Cette étude nous a conduit à proposer une cartographie de 1 897 protéines identifiées avec au moins deux peptides et une confiance supérieure à 95%, issues de ces lignées.

Afin de voir s'il était possible d'identifier plus de protéines potentiellement sécrétées nous avons, de la même manière que précédemment, comparé le protéome total des lignées à la base de données de protéines sécrétées que nous avons établit. Grâce à cela, nous mettons en évidence 303 protéines communes, ce qui signifie que parmi les 1 897 protéines identifiées dans les lignées cellulaires, 303 sont susceptibles d'être sécrétées et retrouvées dans le plasma des patients soit, 16%.

Une analyse ontologique des 303 protéines présentes dans la base de données révèle que 181 d'entre elles sont impliquées dans des régulations biologiques (59 sont impliquées dans la transduction de signaux, 68 sont impliqués dans des processus métaboliques, 56 sont impliquées dans la mort cellulaire). Par ailleurs, 64 protéines sont impliquées dans des voies de signalisation cellulaire faisant notamment intervenir la voie Wnt, les récepteurs couplés aux protéines G.

Les fonctions moléculaires des protéines identifiées sont principalement associées à des interactions avec d'autres protéines, ou encore à des activités catalytiques. Parmi ces activités nous avons pu identifier principalement des hydrolases. Nous avons également observé que 50 de ces protéines étaient classées dans la catégorie des protéines ayant un rôle dans la structure. Parmi celles-ci, 17 sont impliquées dans la formation du cytosquelette, et 10 sont des composants de la matrice extracellulaire.

La localisation de ces 303 protéines est principalement intracellulaire selon la classification ontologique avec 257 protéines répondant à cette localisation. Malgré tout, toujours selon la classification ontologique, 117 protéines seraient localisées à la membranes. De plus 112 protéines seraient extracellulaires.

Nous avons observé que la proportion de protéines détectables dans le plasma de patients avec ce type d'approche (16%) était perfectible. Nous avons développé une approche permettant d'élargir le protéome en améliorant la couverture protéomique. L'objectif que nous avions était de mettre en évidence plus de biomarqueurs potentiels par spectrométrie de masse en améliorant la couverture protéomique de nos analyses. Autrement dit, de rendre détectables des protéines qui ne l'étaient pas en mettant en place des techniques spécifiques. En effet, chaque spot de la plaque MALDI contient plusieurs peptides, et parfois, ils en contiennent trop pour pouvoir être tous analysés. C'est la raison pour laquelle nous avons souhaité tester un moyen d'améliorer la couverture protéomique en utilisant une séparation des peptides plus importante en réalisant des séparations séquentielles par OFFGEL. Cette étude a été réalisée sur des lignées LS180 de cancer colorectal (cf chapitre 5.2 p.158).

## **5.2 Amélioration de la couverture protéomique par double séparation OFFGEL**

Les résultats de notre étude ont fait l'objet d'une soumission dans la revue *Proteomics* et sont en cours de révision.

*Two-step OFFGEL approach for effective peptide separation compatible with iTRAQ labelling*

Alexis Rideau, Damien Besson, Alice Boissard, Olivier Coqueret, Catherine Guette.

*Proteomics*

### 5.2.1 Publication

#### **Two-step OFFGEL approach for effective peptide separation compatible with iTRAQ labelling**

##### **Technical Brief**

Alexis Rideau, Damien Besson, Alice Boissard, Olivier Coqueret, Catherine Guette

Institut de Cancérologie de l'Ouest, Paul Papin Cancer Center, INSERM U892, Angers, France.

To whom correspondence should be addressed: Dr Catherine Guette Paul Papin Cancer Center, INSERM U892 2 rue Moll, 49033 Angers, France.

Tel.: 33-2-41-35-29-14; E-mail:[catherine.guette@ico.unicancer.fr](mailto:catherine.guette@ico.unicancer.fr).

Key words : OGE, OFFGEL, iTRAQ, proteomics

##### **Abstract**

The shotgun proteomic analyses are increasingly becoming methods of choice for complex samples. The development of effective methods for fractionating peptides to reduce the complexity of the sample before mass analysis is a key point in this strategy. The OFFGEL technology (OGE) has recently become a tool of choice in proteomic analysis at peptide level. This OGE approach allows the in-solution separation of peptides from various biological sources by isoelectric focusing in highly resolved 24 fractions. It was also demonstrated that OGE technology is a filtering tool for pI-based validation of peptide identification. As peptide OGE is compatible with iTRAQ labelling, OGE is finding valuable applications in quantitative proteomics as well. The aim of this study is to explain a new two-dimensional OGE approach which improves the proteomic coverage of complex mixtures such colorectal cell line lysates, and which is compatible with iTRAQ labelling.

Received: April 26, 2013; Revised: September 04, 2013; Accepted: September 10, 2013

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pmic.201300161.

Accepted Article

**Abbreviations**

OGE : Off-Gel electrophoresis

iTRAQ : Isobaric tags for relative and absolute quantitation

MS : mass spectrometry

IPG IEF : immobilized pH gradient isoelectric focusing

FASP : Filter aided proteome preparation

MMTS : Methyl methanethiosulfonate

TEAB : Triethylammonium bicarbonate

ACN : acetonitrile

### Introduction

The complexity of the human proteome, in terms of its size and dynamic range, is well recognized [1,2]. Because MS methods preferentially identify the most abundant proteins in complex mixtures [3], the ability to identify low-abundance proteins by MS requires a variety of techniques to be applied for fractionating proteins or peptides [4-12].

Of these techniques, OFFGEL electrophoresis (OGE), with the capability to resolve proteins as well as peptides by IPG IEF with subsequent liquid-phase recovery [13], is proving quite powerful in providing greatly improved protein coverage [14,15]. As peptide IPG IEF is compatible with iTRAQ [16], OGE is finding valuable applications in quantitative proteomics as well [17,18].

The objective of this study was to explain a new two-dimensional OGE approach which improved the proteomic coverage of complex mixtures such as colorectal cell line lysates, and was compatible with iTRAQ labelling. The experimental setup employed for fractionation by this two-step OGE approach (2D-OGE) is shown schematically in Figure 1. Approximately  $5 \times 10^6$  cells (colorectal cell line LS180) were lysed in 0.6 ml of 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl, pH 7.6 at room temperature for 30 min and briefly sonicated to reduce viscosity of the lysate. Detergent was removed from the lysates and the proteins were digested with trypsin using the FASP protocol (19) using spin ultrafiltration units of nominal molecular weight cut of 30 000. Briefly, the protein lysate was applied to an YM-30 microcon filter units (Cat No. MRCF0R030, Millipore) spun down and washed three times with 200  $\mu$ L of 8 M urea in 0.1 M Tris/HCl, pH 8.5. Then 6  $\mu$ L of 200 mM MMTS in 8 M urea was added to the filters and the samples were incubated for 20 min. Filters were washed thrice with 200  $\mu$ L of 8 M urea in 0.1 M Tris/HCl, pH 8.5, followed by six washes with 100  $\mu$ L 0.5M TEAB. Finally, trypsin (AB sciex) was added in 100  $\mu$ L 0.5M TEAB to each filter. The protein to enzyme ratio was 100:1. Samples were incubated overnight at 37°C and released peptides were collected by centrifugation. Samples were then dried completely using a Speed-Vac and re-suspended in 100  $\mu$ l of 0.5% trifluoroacetic acid (TFA) in 5% acetonitrile, and desalted via PepClean C-18 spin columns (Pierce Biotechnology, Rockford, IL). Peptide content was determined using a Micro BCA Protein Assay Kit (Pierce -Thermo Scientific, Rockford, IL).

For the iTRAQ labelling, 100  $\mu$ g of each peptide solution was labelled at room temperature for 2h with one iTRAQ reagent vial previously reconstituted with 70 $\mu$ l of ethanol for 4plex

iTRAQ reagent. Labelled peptides were subsequently mixed in a 1:1:1:1 ratio and dried completely using a Speed-Vac.

For pI-based peptide separation, we used the 3100 OFFGEL Fractionator (Agilent Technologies, Böblingen, Germany) with a 12 or 24-well set-up using our protocol [20].

First, samples were desalted onto a Sep-Pak C18 cartridge (Waters). For the 24-well set-up, peptide samples were diluted in the OFFGEL peptide sample solution to a final volume of 3.6 mL. Then, the IPG gel strip of 24 cm-long (GE Healthcare, München, Germany) with a 3-10 linear pH range was rehydrated with the Peptide IPG Strip Rehydration Solution, according to the protocol of the manufacturer, for 15 min. 150  $\mu$ L of sample was loaded in each well. Electrofocusing of the peptides was performed at 20°C and 50  $\mu$ A until the 50 kVh level was reached. After focusing, the 24 peptide fractions were withdrawn and the wells were washed with 200  $\mu$ L of a solution of water/methanol/formic acid (49/50/1). After 15 min, each washing solution was pooled with its corresponding peptide fraction. All fractions were evaporated by centrifugation under vacuum and maintained at -20°C. For the 2D-OFFGEL approach, the peptides were first fractionated in 12 fractions in the pH range 3-10. Then, fractions F1-F2, fraction F3 to F8 and fractions F9 to F12 were pooled and refractionated in 24 fractions in the pH range 3.5-4.5, 4-7 and 6-9, respectively. We obtained 72 fractions which were analysed by nanoLC-MS/MS. Each fraction was resuspended in 20  $\mu$ L of H<sub>2</sub>O with 0.1% (v/v) TFA. The samples were separated on an Ultimate 3,000 nano-LC system (Dionex, Sunnyvale, USA) using a C18 column (PepMap100, 3 $\mu$ m, 100A, 75 $\mu$ m id x 15cm, Dionex) at 300nL/min a flow rate. Buffer A was 2% ACN in water with 0.05% TFA and buffer B was 80% ACN in water with 0.04% TFA. Peptides were desalted for 3 min using only buffer A on the precolumn, followed by a separation for 105 min using the following gradient: 0 to 20% B in 10 min, 20% to 45% B in 85 min and 45% to 100% B in 10 min. Chromatograms were recorded at the wavelength of 214 nm. Peptide fractions were collected using a Probot microfraction collector (Dionex). We used CHCA (LaserBioLabs, Sophia-Antipolis, France) as MALDI matrix. The matrix (concentration of 2mg/mL in 70% ACN in water with 0.1% TFA) was continuously added to the column effluent via a micro "T" mixing piece at 1.2  $\mu$ L/min flow rate. After 12 min run, a start signal was sent to the Probot to initiate fractionation. Fractions were collected for 10 sec and spotted on a MALDI sample plate (1,664 spots per plate, ABSciex, Foster City, CA.).

MS and MS/MS analyses of off-line spotted peptide samples were performed using the 5800 MALDI-TOF/TOF Analyser (ABSciex) and 4000 Series Explorer software, version 4.0. The instrument was operated in positive ion mode and externally calibrated using a mass

Accepted Article

calibration standard kit (ABSciex). The laser power was set between 2800 and 3400 for MS and between 3600 and 4200 for MS/MS acquisition. After screening all LC-MALDI sample positions in MS-positive reflector mode using 2000 laser shots, the fragmentation of automatically-selected precursors was performed at a collision energy of 1kV using air as collision gas (pressure of  $\sim 2 \times 10^{-6}$  Torr) with an accumulation of 3000 shots for each spectrum. MS spectra were acquired between  $m/z$  1000 and 4000. We used the parent ion of Glu1-fibrinopeptide at  $m/z$  1570.677 diluted in the matrix (30 femtomoles per spot) for the internal calibration. Up to 12 of the most intense ion signals per spot position having a  $S/N > 20$  were selected as precursors for MS/MS acquisition. The identification of peptides and proteins was performed by the ProteinPilot™ Software V 4.0 (AB Sciex) using the Paragon algorithm as the search engine [21]. Each MS/MS spectrum was searched for Homo sapiens species against the Uniprot/swissprot database (UniProtKB/Sprot 20110208 release 01, with 525997 sequence entries). The searches were run using the fixed modification of methylmethanethiosulfate labeled cysteine parameter enabled. Other parameters, such as tryptic cleavage specificity, precursor ion mass accuracy and fragment ion mass accuracy, are MALDI 5800 built-in functions of ProteinPilot software. The detected protein threshold (unused protscore (confidence)) in the software was set to 1.3 to achieve 95% confidence, and identified proteins were grouped by the ProGroup algorithm (ABSciex) to minimize redundancy. The bias correction option was executed.

To estimate the false discovery rate (FDR), a decoy database search strategy was used. The FDR is defined as the percentage of decoy proteins identified against the total protein identification. The FDR was calculated by searching the spectral against the Uniprot Homo sapiens decoy database.

For the quantification of the relative protein expression, we employed a customized software package, iQuantitor [20,22,23], to infer the magnitude of change in protein expression. The software infers treatment-dependent changes in expression using Bayesian statistical methods. Basically, this approach was used to generate means, medians, and 95% confidence intervals (upper and lower) for each treatment-dependent change in protein expression by using peptide-level data for each component peptide. For proteins whose iTRAQ ratios were downregulated, the extent of down-regulation was considered further if the higher limit of the credible interval had a value lower than 1. Conversely, for proteins whose iTRAQ ratios were increased, the extent of upregulation was considered further if the lower limit of the confidence interval had a value greater than 1. The width of these confidence intervals depends on the data available for a given protein. The software took into consideration all the

peptides observed and the number of spectra used to quantify the change in expression for a given protein. In these conditions, it is possible to detect small but significant changes in up- or downregulation when many peptides are available. The peptide selection criteria for relative quantification were performed as follows. Only peptides unique for a given protein were considered for relative quantification, excluding those common to other isoforms or proteins of the same family. Proteins were identified on the basis of having at least two peptides with an ion score above 95% confidence.

This study was performed on colorectal cell line lysates. In the label-free approach, we first compared the number of identified peptides in a traditional 24 fractions OGE approach versus the 2D-OGE approach (Figure 2). In the 1D-OGE approach, we identified 3602 peptides in 4 fractions F1 to F4 fractions ; 6074 peptides in 12 fractions F5 to F16 and 2155 peptides in the last 8 fractions F17 to F24 (Supplementary data 1); in the 2D-OGE approach, 18428 peptides were characterized in the acidic pH 3.5-4.5 OFFGEL (Supplementary data 2), 13327 in the pH 4-7 OFFGEL (Supplementary data 3) and finally in the pH 6-9, 5712 peptides were identified (Supplementary data 4). These results showed that 5.1 times more acidic peptides, 2.2 times more peptides in the pH 4-7 range and 2.6 times more basic peptides were identified with the 2D-OGE approach, confirming that this novel approach elicited the decrease of the complexity of peptide fractions and thus the identification of a larger number of peptides. In terms of protein, in the 1D-OGE approach, 1917 proteins were identified (with at least 2 peptides) whereas 2641 proteins were characterized with the 2D-OGE approach confirming an improved proteomic coverage (Figure 3A) (Supplementary data 5). The same study was conducted on an experience of quantitative proteomics with iTRAQ labelling. In the 1D-OGE approach, we identified 2539 peptides in the F1-F4 fractions; 10024 peptides in the F5-F16 fractions and 4680 peptides in the F17-F24 fractions (Supplementary data 6); in the 2D-OGE approach, 18256 peptides were characterized in the acidic pH 3.5-4.5 OFFGEL (Supplementary data 7), 17447 in the pH 4-7 OFFGEL (Supplementary data 8) and finally in the pH 6-9, 6640 peptides were identified (Supplementary data 9) (Figure 2). These results showed that 7.2 times more acidic peptides, 1.7 times more peptides in the pH 4-7 range and 1.4 times more basic peptides were identified with the 2D-OGE approach. In terms of protein, 1784 proteins were identified with 1D approach versus 2407 proteins with 2D-OGE methodology (Figure 3A) (Supplementary data 5). These results are consistent with those obtained in the label-free experiment. Averaging on 4 iTRAQ experiments on cell lines, we identified 2220 +/-99 proteins (with at least 2 peptides) with the 2D-OGE approach versus 1707 +/-190 proteins with the 1D-OGE

fractionation highlighting an increase of the identified proteins of 30%. Finally, we observed that the number of identified peptides were drastically increased for the 2D-OGE approach, 3-fold and 2.5-fold higher for the label-free and iTRAQ labelling experiments, respectively, when the protein number increased only by 30%. This increase of identified proteins is not proportional to the number of identified peptides that highlights the increase in protein coverage.

Several studies have shown that iTRAQ labelling is associated with bias; fold changes are compressed towards one [24]. It has been suggested that this underestimation of fold change is due to co-eluting peptides with similar m/z values that are isolated together, creating mixed iTRAQ intensities in complex samples [25]. In our study, we identified 237 significantly over or under-expressed proteins with 1D-OGE; the number rose to 387 dysregulated proteins with the 2D-OGE approach (Supplementary data 5). Between both experiments, 205 proteins were found in both approaches (Supplementary data 10). In the 1D-OGE approach, 1406 peptides were counted for the protein identification whereas the 2D-OGE approach, 1769 peptides were counted, corresponding to a 26% average increase in protein coverage. For the 182 dysregulated proteins found in 2D-OGE which were not identified as significantly dysregulated in the 1D-OGE, we found 39 novel proteins and 143 proteins identified in both 1D and 2D method (Supplementary data 10). Among these proteins, in the 2D approach, 2% decreased coverage, 45% had no change in protein coverage and 53% increased in protein coverage. These results showed that this approach is fully compatible with iTRAQ labelling and allows a better separation of peptides before MS/MS experiments resulting in a decreased effect of the interference of mixed MS/MS contribution occurring during precursor selection and so, an increase of significant dysregulated identified proteins (versus the 1D-OGE technology).

Finally, an application that we wanted to achieve was the comparison of the cell line LS180 proteome with 1D and 2D-OGE approaches and the proteome of colorectal tumor that we have already published [20]. With fractionation in 24 fractions, 1562 common proteins were exhibited, and this protein number rose to 1777 when 2D-OGE protocol was used (Figure 3B). Gene ontology analysis of common identified proteins using DAVID [26,27] revealed that 1D-OGE successfully identified 168 gene names associated with “extracellular part” ontology, whereas the 2D-OGE technique identified 195 proteins associated with this ontology (Figure 3C). In addition, OLFM4, a secreted protein characterized as a marker of early colon tumor [20] was evaluated in a same colon tumor with ELISA test and with both 1D and 2D-OGE approaches relative to healthy colon tissue. The ratio OLFM4 concentration

in colon tumor tissue versus in healthy tissue was 6.5, 3.5 and 5.1 for ELISA, 1D-OGE and 2D-OGE, respectively (Figure 3D), suggesting that the iTRAQ ratio calculated with the 2D-OGE method underestimated less than the 1D-OGE approach.

In conclusion, we have shown that the 2D-OGE approach allows the identification of more proteins in complex mixtures. The increased number of different peptides being identified is particularly important, not only for the purposes of confidently identifying proteins (increasing the number of protein identifications based on more than one peptide), but also for peptide-based quantification approaches using iTRAQ labelling. The identification and quantification of more peptides per protein provide data for better quantification, and increase sequence coverage, which in turn increases the possibility of distinguishing between protein isoforms. At the very least, better separation between peptides renders iTRAQ quantification more reliable.

#### Acknowledgments

This work was supported by grants from the Ligue Contre le Cancer (comité du Maine et Loire), Cancéropole Grand Ouest and Pays de Loire region.

#### References

- [1] Uhlen, M., Ponten, F., Antibody-based proteomics for human tissue profiling. *Mol Cell Proteomics* 2005, 4, 384-393.
- [2] Anderson, NL., Anderson, NG., The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002, 1,845-867.
- [3] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, 422, 198-207.
- [4] Zolotarjova, N., Martosella, J., Nicol, G., Bailey, J. et al., Differences among techniques for high-abundant protein depletion. *Proteomics* 2005, 5, 3304-3313.
- [5] Weber, G., Islinger, M., Weber, P., Eckerskorn, C. et al., Efficient separation and analysis of peroxisomal membrane proteins using free-flow isoelectric focusing. *Electrophoresis* 2004, 25, 1735-1747.
- [6] Cho, SY., Lee, EY., Lee, JS., et al: Efficient prefractionation of low-abundance proteins in human plasma and construction of a two-dimensional map. *Proteomics* 2005, 5, 3386-3396.

- [7] Stasyk, T., Huber, LA., Zooming in: fractionation strategies in proteomics. *Proteomics* 2004, 4, 3704-16.
- [8] Zuo, X., Speicher, DW., Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional electrophoresis. *Proteomics* 2002, 2, 58-68.
- [9] Herbert, B., Righetti, PG., A turning point in proteome analysis: sample prefractionation via multicompartment electrolyzers with isoelectric membranes. *Electrophoresis* 2000, 21, 3639-48.
- [10] Righetti, PG., Castagna, A., Herbert, B., Candiano, G., How to bring the “unseen” proteome to the limelight via electrophoretic pre-fractionation techniques. *Biosci Rep* 2005, 25, 3-17.
- [11] Bier, M., Recycling isoelectric focusing and isotachopheresis. *Electrophoresis* 1998, 19, 1057-1063.
- [12] Righetti, PG., Castagna, A., Antonioli, P., Boschetti, E., Prefractionation techniques in proteome analysis: the mining tools of the third millennium. *Electrophoresis* 2005, 26, 297-319.
- [13] Horth, P., Miller, CA., Preckel, T., Wenz, C., Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol Cell Proteomics* 2006, 5, 1968-1974.
- [14] Hubner, NC., Ren, S., Mann, M., Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics* 2008, 8, 4862-4872.
- [15] Lam, HT., Jossierand, J., Lion, N., Girault, HH., Modeling the isoelectric focusing of peptides in an OFFGEL multicompartment cell. *J Proteome Res* 2007, 6, 1666-76.
- [16] Lenggqvist, J., Uhlen, K., Lehtio, J., iTRAQ compatibility of peptide immobilized pH gradient isoelectric focusing. *Proteomics* 2007, 7, 1746-1752.
- [17] Ernoult, E., Gamelin, E., Guette, C.: Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation. *Proteome Sci* 2008, 6, 27.
- [18]. Chenu, J., Michelland, S., Sidibe, J., Seve, M., Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. *Proteome Sci* 2008, 6, 9.
- [19] Wiśniewski, JR., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* 2009, 6, 359-362.

[20] Besson, D., Pavageau, A.H., Valo, I., Bourreau, A., et al., A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. *Mol Cell Proteomics*. 2011 12, 1-14.

[21] Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., Hunter, C. L., Nuwaysir, L. M., and Schaeffer, D. A. The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol Cell Proteomics* 2007, 6, 1638-1655.

[22] Schwacke, J. H., Hill, E. G., Krug, E. L., Comte-Walters, S., and Schey, K. L. (2009) iQuantitor: a tool for protein expression inference using iTRAQ. *BMC Bioinformatics* 10, 342.

[23] Grant, J.E., Bradshaw, A.D., Schwacke, J.H, Baicu, C.F., Zile, M.R., Schey, K.L. (2009) Quantification of protein expression changes in the aging left ventricle of *Rattus norvegicus*. *J Proteome Res*. 9, 4252-63.

[24] Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V., and Lilley, K. S. Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics* 2010, 9, 1885-1897.

[25] Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". *J Proteome Res* 2009, 8, 5347-5355.

[26] Huang, D.W., Sherman, B.T., Lempicki, R.A., Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* **2009;4(1):44-57.**

[27] Huang D.W., Sherman B.T., Lempicki R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009, 37, 1-13.

Figure 1 : Workflow of 1D and 2D-OGE approaches

Accepted Article

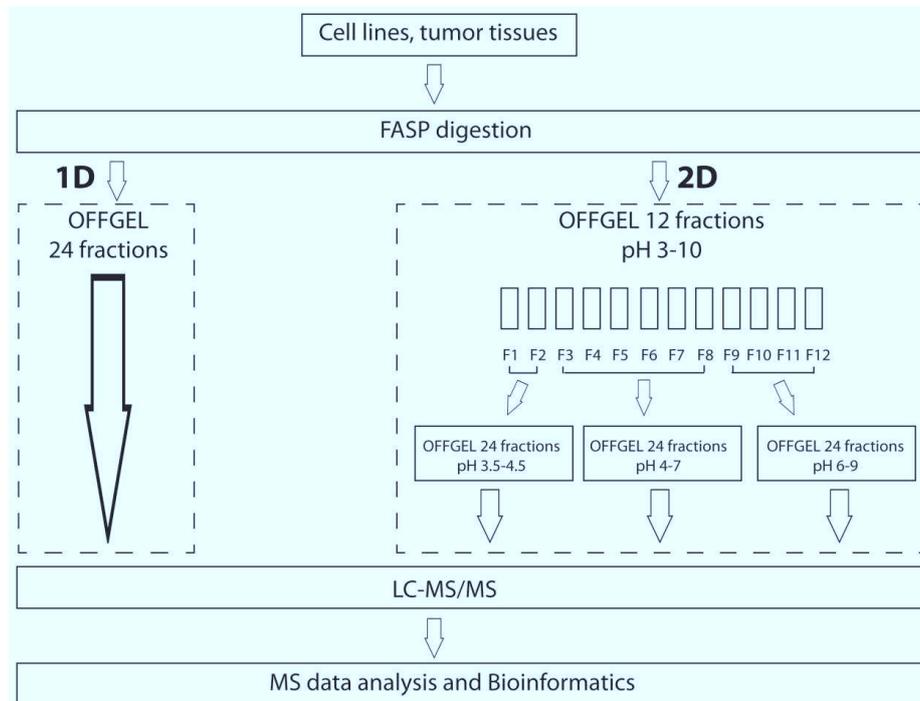


Figure 2 : Analyses of colorectal cell line lysates by 1D and 2D OGE approaches without and with iTRAQ labelling. The total number of identified peptides in each fraction are plotted. A: 1D-OGE; B: 2D-OGE, 3.5-4.5 pH range; C: 2D-OGE, 4-7 pH range; D: 2D-OGE, 6-9 pH range.

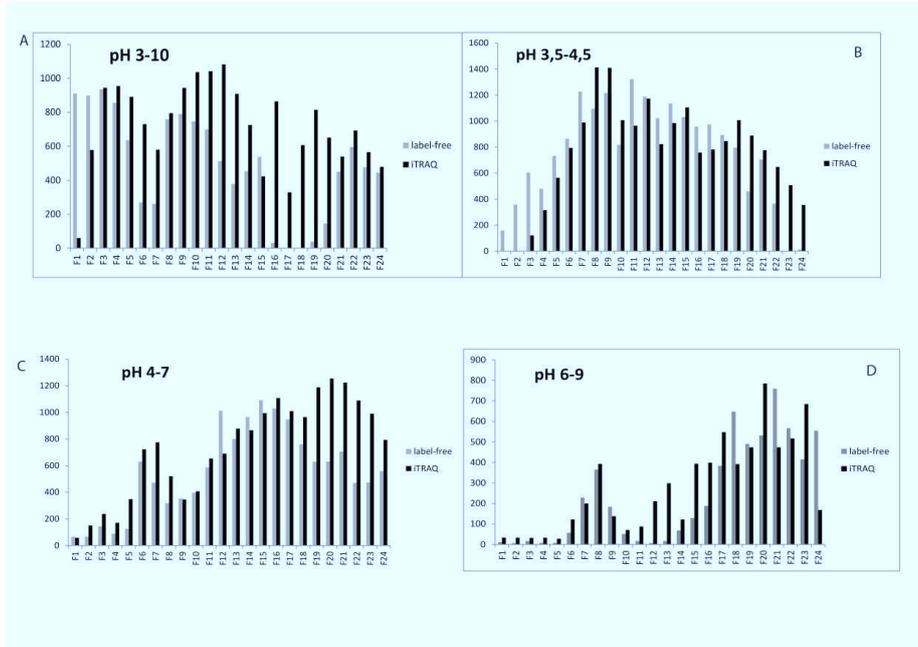
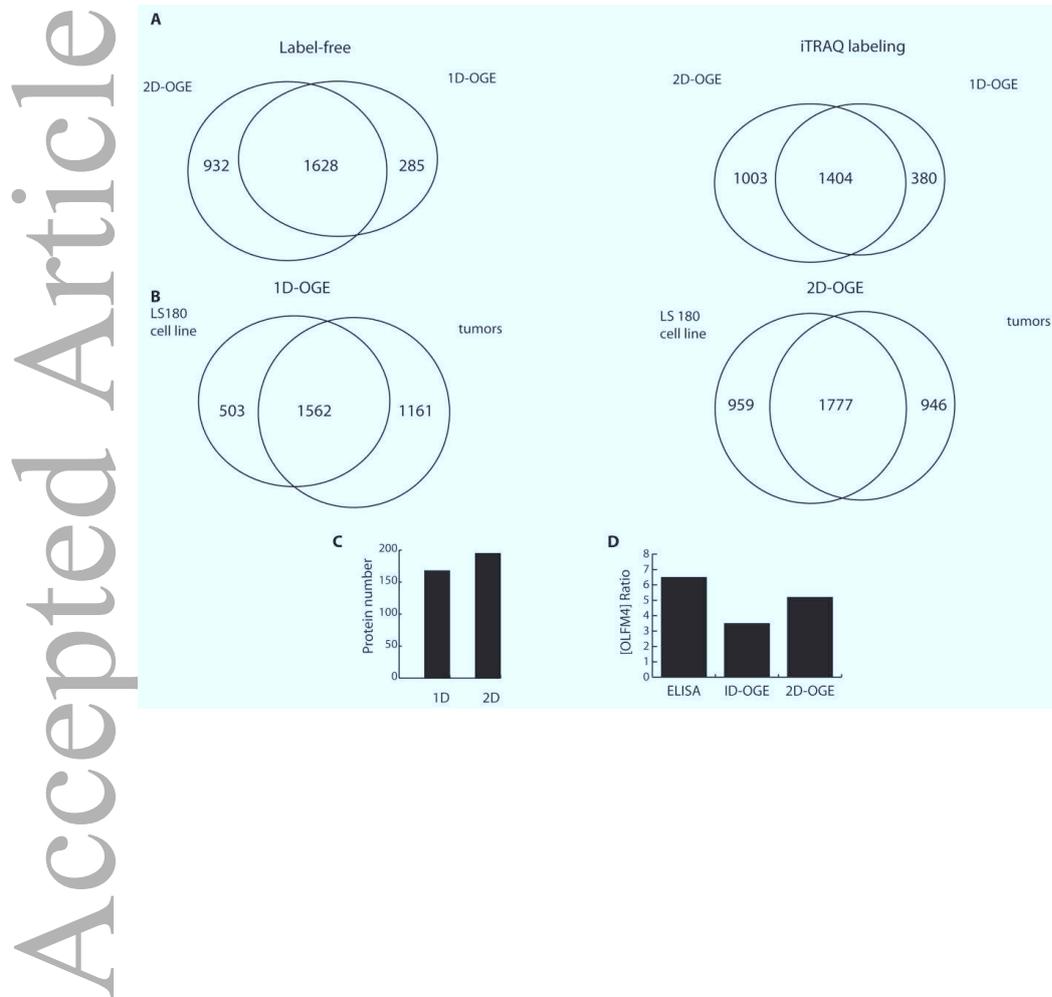


Figure 3: Comparison of proteome coverage at the protein level. A: label-free versus iTRAQ labelling. B: the colorectal tumor proteome is compared with colorectal cell line LS180 by 1D-OGE and 2D-OGE approaches. C: comparison of secreted proteins. D: ratio [OLFM4] tumor tissue/healthy controls by Elisa test, 1D- and 2D-OGE approaches.



### 5.3 Analyse du sécrétome des LS180

Pour améliorer encore la pertinence des protéines identifiées comme potentiels biomarqueurs, nous avons analysé les protéines issues du sécrétome d'une lignée d'adénocarcinome colique (LS180). Cette analyse nous avait permis d'identifier 165 protéines avec au moins deux peptides et une confiance supérieure à 95%. Parmi celles-ci, 94 protéines étaient présentes dans la base de données créée à partir de la littérature, soit près de 57%.

Plusieurs protéines intéressantes peuvent être mentionnées ici. Nous avons, par cette étude, identifié l'antigène CD44. Cette protéine est connue dans le cancer du côlon pour être un marqueur de cellules initiatrices de cancer en association avec le marqueur CD24.

La décorine est également une protéine identifiée dans cette expérience. Comme nous l'avons déjà mentionné cette protéine est une protéine de la matrice extracellulaire impliquée dans certaines pathologie de l'œil. Son rôle dans le remodelage de la matrice extracellulaire en fait également un biomarqueur potentiel intéressant.

La desmoplakine, une protéine de jonction inter-cellulaire mérite également l'attention. Elle est notamment capable d'interagir avec la vimentine, une protéine impliquée dans la transition épithélio-mésenchymateuse et probablement dans les phénomènes de dissémination tumorale. La vimentine est d'ailleurs également présentes parmi les protéines identifiées dans les lignées et détectées dans le plasma de patients.

Les protéines S100A13 et S100A4 sont également importantes dans la mesure où les protéines S100A8 et A9 sont déjà décrites comme biomarqueurs potentiels dans le côlon, ces dernières pourraient venir compléter la liste. Par ailleurs, ces protéines possèdent un rôle cellulaire important pour la tumeur puisqu'elle permet la progression du cycle cellulaire et la différenciation.

Les septines sont également importantes puisqu'elles ont un rôle essentiel dans la cytokinèse, permettant la division cellulaire.

La sorbitol déshydrogénase est également identifiée dans cette expérience, ce qui semble conforter son intérêt déjà aperçu dans la première partie de notre étude.

La protéine TGF $\beta$ 1 est également identifiée dans cette expérience, tout comme dans l'étude du stroma tumoral.

La compilation de l'ensemble de ces résultats nous a conduit à proposer une base de données de protéines issues des cellules ou des lignées colorectales et pouvant être sécrétées. Cette base de données est présentée dans le tableau (cf annexe n°3 p.225).

## Chapitre 6

### Développement d'une approche de glycoprotéomique

Nous avons souhaité utiliser l'approche que nous avons mis au point pour l'étude des sites de glycosylation de l'OLFM4 sur des mélanges complexes, et nous avons pour cela réalisé un enrichissement des protéines glycosylées d'un extrait de tumeur précédemment analysé dans notre étude. Cela nous a permis de comparer les protéines identifiées lors d'une analyse protéomique globale, et celles identifiées lors de l'analyse d'un sous-protéome.

Les protéines de la tumeurs ont été extraites par la technique FASP et digérées à la trypsine. Les peptides générés ont été enrichit sur un mélange deux lectines (Concanavoline A et WGA). Ainsi, 800  $\mu\text{g}$  de peptides issus de la tumeurs ont été incubés avec 800  $\mu\text{g}$  de chacune des deux lectines à 4 ° C pendant une nuit. Les peptides ont ensuite été déglycosylés par l'enzyme PNGase F dans l'eau lourde.

Après enrichissement nous avons récupéré 140  $\mu\text{g}$  de peptides ce qui correspond à 17,5%. Ces peptides ont été ensuite séparés par OFFGEL en 12 fractions. Ces fractions ont ensuite été séparées sur nano-HPLC en phase reverse sur une colonne de C18 avant l'analyse par spectrométrie de masse.

Ainsi, nous avons identifié 607 protéines avec au moins deux peptides et une confiance supérieure à 95%.

Nous avons également comparé ces 607 protéines à la base de données des protéines sécrétées établies à l'aide des données de la littérature. Parmi les 607 protéines, nous avons identifié 53 protéines porteuses d'au moins un peptide possédant un résidu asparagine transformé en résidu d'acide aspartique ayant incorporé l' $^{18}\text{O}$ . Cet enrichissement correspond à un rendement de 8,7%.

Nous avons également constaté que 8 de ces 53 protéines n'avaient jamais été identifiées auparavant dans l'ensemble de nos analyses. Ces huit protéines sont les suivantes :

- Aggrecan core protein (ACAN), une protéine de la matrice extracellulaire ;

- Cell cycle control protein 50B (TMEM30B), une protéine transmembranaire spécifique du cartilage et impliquée dans des pathologies ostéo-articulaires lorsqu'elle est mutée ;
- Glutaminyl-peptide cyclotransferase (QPCT), une protéine impliquée dans la biosynthèse des peptides possédant des cycles pyroglutamiques.
- Integrin alpha-V (ITGAV), une protéine membranaire ayant une fonction de récepteur au fibrinogène, à la vitronectine, à la laminine et à la thrombospondine notamment.
- Polypeptide N-acetylgalactosaminyltransferase 1 (GALNT1), une protéine existant sous une forme soluble et sous une forme soluble sécrétée, et impliquée dans les étapes initiales de glycosylation des protéines ;
- Procollagen galactosyltransferase 1 (GLT25D1), une protéine intracellulaire localisée majoritairement dans la lumière du réticulum endoplasmique, et ayant pour fonction de transférer des groupements galactoses sur les résidus hydroxylysines du collagène ;
- Procollagen-lysine,2-oxoglutarate 5-dioxygénase 2 (PLOD2), une protéine intracellulaire impliquée dans la biosynthèse des hydroxylysines du collagène à partir de la Lysine ;
- Soluble calcium-activated nucleotidase 1 (CANT1), une protéine intracellulaire dépendante du calcium, capable d'hydrolyser des nucléosides diphosphates en monophosphate avec libération d'énergie utilisable par la cellule.

Nous avons également observé que 64 protéines parmi les 607 identifiées étaient présentes dans la base de données, et donc détectables dans le sang de patients. Cela correspond à 10,5% des protéines. Dans un premier temps, nous avons trouvé ce pourcentage faible, puisqu'il était même inférieur à celui que l'on trouvait dans les lignées cellulaires (16%). Nous avons alors décidé de regarder quelle était la proportion de protéines détectables dans les plasmas de patients identifiées parmi les protéines glycosylées. C'est alors que nous vîmes que 27 des 53 protéines glycosylées étaient présentes dans notre base de données des protéines sécrétées et détectables dans le sang, ce qui correspond à 51%. Autrement dit, en analysant uniquement les protéines glycosylées, nous obtenons un pourcentage de protéines sécrétées environ équivalent à celui de l'analyse de sécrétome de lignées cellulaires.

Ces données ont été comparées à l'analyse globale du protéome de la tumeur du patient. Cette analyse nous avait permis de mettre en évidence 1 336 protéines avec au moins deux peptides et une confiance supérieure à 95%. L'analyse ontologique de ces protéines révèle que 164 étaient présentes dans le milieu extracellulaire ce qui correspond à 12,3%. Enfin, la comparaison avec la base de données de protéines sécrétées a montré que 294 des protéines

étaient communes soit 22%.

Tableau 6.1 – Comparaison des résultats de protéomique issus de l'analyse globale du protéome d'un patient et de son glycoprotéome correspondant

	Protéome total (1336 protéines)	Glycoprotéome (607 protéines)
Gene Ontology ( <i>extracellular part</i> )	12,3%	15,5%
Base de données	22%	10,5%



# Quatrième partie

## Discussion



# Chapitre 1

## Discussion des résultats obtenus

### 1.1 L'analyse protéomique par stade tumoraux

L'INSTABILITÉ génétique qui règne au sein des cancers colorectaux conduit à une forte hétérogénéité des tumeurs qui explique probablement pourquoi la réponse aux traitements est variable d'un patient à l'autre. C'est la raison pour laquelle le classement moléculaire des tumeurs est devenu un enjeu essentiel dans la prise en charge du traitement des patients, non seulement d'un point de vue diagnostique et pronostic, mais aussi pour identifier les voies oncogéniques qui doivent être précisément ciblées chez un patient particulier si l'on veut avoir le plus de chance possible de traiter son cancer [van't Veer and Bernards, 2008]. Des résultats récents ont montré l'intérêt d'utiliser ce type de signature génomique pour identifier les voies dérégulées et caractériser des éventuels marqueurs pronostics. Ces derniers peuvent également être utilisés pour tenter de prédire la réponse au traitement d'un patient [Nevins and Potti, 2007]. La protéomique quantitative est également une approche intéressante pour établir ce type de classification, prédire l'échappement tumoral, ou encore pour caractériser de nouveaux biomarqueurs. C'est cette dernière application qui a été le moteur de nos investigations.

Dans notre étude protéomique, nous avons pu établir un protéome total 3 138 protéines avec un taux de faux positif de 2,1% et avec une identification à l'aide d'au moins deux peptides sur un total de 28 tumeurs coliques. Il s'agit de l'une des plus grande base de données protéomique établie sur le cancer du colon, ce qui montre la pertinence de la combinaison de la microdissection laser, de la séparation OFFGEL couplée au marquage iTRAQ et à l'analyse sur MALDI TOF/TOF. Nous avons ainsi pu, pour chaque patient identifier et quantifier plus de 1100 protéines.

Parmi ces protéines, 555 varient de façon statistiquement significative entre le tissu

sain et les tumeurs ou les adénomes ce qui représente 16% du protéome total. Nous proposons donc la première signature protéomique correspondant aux différents stades histologiques des tumeurs colorectales. Il est intéressant de noter que les tumeurs de stade II sont celles qui présentent le nombre de protéines statistiquement dérégulées le plus important. En effet, c'est entre ce stade et le stade III que les tumeurs acquièrent la capacité de coloniser les tissus environnants – les tumeurs de stade III sont définies comme présentant des invasions ganglionnaires –, ce qui pourrait expliquer le fait que beaucoup de processus cellulaires soient dérégulés dans ces tumeurs. Cela suggère également que les protéines dérégulées dans ces tumeurs soient, pour une partie d'entre elles, impliquées dans l'agressivité tumorale ; permettant notamment à la tumeur d'acquérir des capacités de dégradation de la membrane basale de l'épithélium intestinal, et des capacités de migration. En effet, si on analyse les protéines du stade II qui varient, on observe que sur les 307 protéines, 21 protéines – soit 6,8% – sont annotées par *Gene Ontology* comme étant impliquées dans la migration cellulaire, et 59 – soit 19,2% – font parti de l'organisation du cytosquelette. Pour le stade I, elles ne sont plus que 10 sur 196 – soit 5,1% – pour ce qui est de la migration cellulaire, et 33 font parti de l'organisation du cytosquelette – soit 16,8% –. Pour le stade IV, ces proportions représentent 3,8% pour les protéines impliquées dans la migration cellulaire, et 13,5% pour les protéines du cytosquelette ; et pour le stade IV elle représente 13% pour la migration cellulaire et 33% pour l'organisation du cytosquelette.

Nous avons souhaité analyser la base de données que nous avons établit afin de voir si nous aussi nous pouvions faire ressortir un biomarqueur potentiel pertinent et si possible spécifique de l'avancée tumorale. Pour cela, nous avons restreint notre analyse aux protéines statistiquement dérégulées qui sont au nombre de 555 au total. Nous avons ensuite regardé quelles étaient les protéines sur-exprimées dans les stades 1 et 2 et dans les adénomes, et qui n'étaient plus sur-exprimées dans les stades 3 et 4. Cette restriction nous permet de nous intéresser à des biomarqueurs étant potentiellement spécifiques des stades initiaux du développement tumoral, avant la mise en place des phénomènes de dissémination. Quatre protéines répondaient à ces critères : l'aldéhyde déshydrogenase (ALDH1A1), la heat shock protein 1 (HSPE1), la sorbitol déshydrogenase (SORD) et l'olfactomedine 4 (OLFM4). Les deux premières protéines (ALDH1A1 et HSPE1) sont des protéines intervenant dans le métabolisme cellulaire, mais ne sont pas retrouvées dans la base de données des protéines présentes dans le sang de patients. La sorbitol déshydrogenase est quant à elle présente dans notre base de données de protéines détectables dans le sang de patients, mais son expression ubiquitaire nous a laissé supposer qu'il s'agirait probablement d'une protéine

dont la spécificité serait difficile à démontrer. En revanche, l'OLFM4 paraît être un bon candidat biomarqueur spécifique du cancer colorectal.

La première protéine de cette famille a été isolée et caractérisée en 1993 comme étant un des constituants essentiels de la matrice muqueuse extracellulaire du neuroépithélium respiratoire chez la grenouille, d'où le nom de la famille [Bal and Anholt, 1993]. Cette dernière comprend 13 membres décrits chez les mammifères et seulement cinq identifiés chez l'Homme [Zeng et al., 2005] [Mukhopadhyay et al., 2004] [Kulkarni et al., 2000]. Il a cependant fallu attendre 2002 pour que le gène de l'olfactomedine 4 soit cloné à partir de cellules myéloïdes humaines [Zhang et al., 2002a]. Il est localisé sur le chromosome 13 au niveau de la région 13q14.3. Le gène est composé de 5 exons qui couvrent un total de 23kb et pour lesquels aucun épissage alternatif n'est connu. La transcription de ce gène conduit à la production d'un ARNm de 1530 nucléotides dont la traduction génère une protéine de 510 acides aminés comprenant un peptide signal en N-terminal et un domaine olfactomedine en C-terminal.

Les travaux de l'équipe de Rodgers ont été d'une importance capitale dans l'étude de la protéine ; ils ont notamment montré la présence de 5 résidus cystéines en position 83, 85, 226, 246 et 437 intervenant dans les structures tertiaires et quaternaires de la protéine [Liu et al., 2006]. Bien que n'affectant pas la formation de complexes multimériques, la substitution des cystéines en position 83, 85, 246 et 437 empêche la sécrétion de la protéine. En revanche, des mutants présentant une forme tronquée du domaine olfactomedine sont toujours sécrétés suggérant que le domaine n'intervient pas dans la régulation de la sécrétion, contrairement à la conformation tridimensionnelle de la protéine.

D'autre part, il s'agit d'une protéine dont le niveau d'expression est important au niveau de la prostate, de l'intestin grêle et du colon, et dans une moindre mesure au niveau de l'œsophage, de la moelle osseuse et de l'estomac [Zhang et al., 2002a] [Zhang et al., 2004] [Liu et al., 2007]. Elle semble donc avoir une distribution limitée dans l'organisme, ce qui en fait un potentiel biomarqueur spécifique d'un ou quelques types tissulaires.

L'équipe de Clevers a montré en 2009 que l'olfactomedine 4 était un marqueur robuste des cellules souches intestinales, et a confirmé ces observations en 2010 en montrant un lien étroit entre l'olfactomedine 4 et la protéine *Leucin-rich repeat-containing G-protein coupled receptor 5* Lgr5 [van der Flier et al., 2009] [Barker and Clevers, 2010].

Il est également montré que la protéine Lgr5 est une cible de la voie Wnt [Barker and Clevers, 2010]. Cette dernière est donc un marqueur de cellules souches intestinales, potentiellement régulée par la voie Wnt, et il a été montré que les cellules souches étaient très

probablement à l'origine du développement des cancers intestinaux chez la souris [Barker et al., 2009].

D'autre part, il a été montré par la technique de double-hybride que l'olfactomedine 4 pouvait interagir avec la protéine *genes associated with retinoid-interferon-induced-mortality-19* GRIM-19 [Zhang et al., 2004]. Cette protéine est connue pour intervenir dans la régulation de l'apoptose médiée par l'action combinée de l'interferon-beta (IFN- $\beta$ ) et de l'acide rétinoïque (AR) [Angell et al., 2000], qui sont capables d'inhiber la croissance tumorale. En effet, une sur-expression de cette protéine dans des cellules de carcinomes mammaires est capable d'induire l'apoptose des cellules en réponse aux deux traitements. Inversement, la sous-expression de GRIM-19 induit une résistance à ces traitements. Selon les auteurs, GRIM-19 pourrait donc être un suppresseur de tumeurs. Cela suggère que l'olfactomedine 4 est probablement impliquée dans les mécanismes d'apoptose par son interaction avec GRIM-19. Plus récemment, le lien entre l'olfactomedine 4 et GRIM-19 a été étudié plus en détails [Huang et al., 2010] et il en ressort que la sur-expression de GRIM-19 dans des cellules de carcinomes gastriques (SGC-7901) induit une diminution de l'olfactomedine 4, et une diminution de la fixation de NF $\kappa$ B sur le promoteur de celle-ci. En outre, cette protéine est capable d'interagir avec l'oncogène STAT3 supprimant ainsi ses fonctions transcriptionnelles en supprimant la possibilité de translocation dans le noyau [Lufei et al., 2003]. Une autre équipe a montré que GRIM-19 inhibait l'expression des gènes cibles de STAT3 sans modifier son activation ou sa capacité de se lier à l'ADN. Grâce à des expériences de mutagénèse dirigée, ils ont aussi montré que la phosphorylation de la sérine 727 était essentielle à la fixation de GRIM-19 et de STAT3 [Zhang et al., 2003b].

Toutes ces observations nous ont conduit à penser que cette protéine pouvait avoir un rôle important dans les processus initiaux du développement tumoral et nous avons souhaité découvrir plus en détails cette protéine.

D'un point de vue moléculaire, nous avons démontré que son expression était régulée par la protéine NF $\kappa$ B suggérant un rôle primordial dans les étapes initiales du développement tumoral.

Nous avons démontré qu'il s'agissait d'une protéine sécrétée et N-glycosylée. Nos résultats démontrent expérimentalement les sites de glycosylation de l'OLFM4 aux positions 72, 136, 253, 352, et 411. Ces sites sont des sites qui étaient inclus dans la séquence consensus connue pour les N-glycosylation à savoir : N-X-S/T (où X peut être n'importe lequel des acides aminés excepté la proline). Nos résultats nous ont également amené à montrer un

site glycosylé en position 356 qui n'était pas inclut dans cette séquence consensus. Il existe donc probablement des N-glycosylations qui ne répondent pas à la séquence consensus connue. Ce site se situe dans une zone entourée de trois thréonines.

Enfin, nous avons montré, grâce à des tests ELISA qu'il s'agissait d'une protéine détectable dans le sérum de patients. Nous avons, de façon préliminaire, analysé les séra de 60 patients sains, et 14 de patients présentant des polypes. Cette étude montre que l'OLFM4 est surexprimée chez les patients présentant des polypes. Cela montre qu'il s'agit d'un bon candidat biomarqueur de tumeurs précoces du cancer colorectal.

## 1.2 Analyse des tumeurs et statut de l'oncogène KRAS

Depuis la découverte de la mutation de KRAS comme facteur de résistance aux thérapies ciblant l'EGFR, l'analyse de cette mutation est devenu une pratique de routine dans tous les centres prenant en charge des patients atteint de cancer colorectaux. La mutation est systématiquement recherchée afin de savoir s'il est possible ou non de donner ce traitement au patient.

Dans cet optique, nous proposons une signature protéomique permettant de discriminer les tumeurs mutées sur KRAS de celles qui ne le sont pas. Cette signature, issue d'une base de données de 2 692 protéines, contient des biomarqueurs de la mutation de KRAS. Parmi les protéines dérégulées de façon statistique, 112 sont sur-exprimées et 140 sont sous-exprimées.

D'autres protéines viennent s'ajouter à celle-ci, notamment la protéine ALDH1B1 dont l'expression est augmentée d'un facteur 1,9 entre les tumeurs non mutées et les tumeurs mutées. Cette protéine est connue pour être un marqueur de cellules souches colorectales. Associée à l'OLFM4, la signature protéomique de la mutation de KRAS semble donc révéler que les tumeurs mutées sur KRAS activent des voies de signalisation leur permettant d'acquérir un phénotype équivalent à des cellules souches, qui sont les cellules initiatrices du cancer dans le côlon.

## 1.3 Analyse du stroma tumoral

Le stroma tumoral constitue l'environnement le plus proche de la tumeur dans lequel se trouve théoriquement la plupart des molécules sécrétées par la tumeur. Cela en fait une source de biomarqueurs potentiels pertinentes. Nous avons donc souhaité analyser ce

stroma dans l'espoir de créer une base de données des protéines qu'il contient. À l'aide de la microdissection laser nous avons pu l'étudier spécifiquement, et ainsi mettre en évidence 245 protéines.

En comparant avec la base de données des protéines sécrétées dans les plasmas de patients, nous avons obtenu plus de 50% de protéines sécrétées. Cela montre que l'analyse du stroma semble particulièrement pertinente pour mettre en évidence des protéines sécrétées par la tumeur et détectables dans le plasma de patients. Il ne s'agit ici que d'une analyse d'identification, et il ne nous est pas possible de conclure sur la dérégulation de ces protéines par rapport au tissu conjonctif sain puisqu'il n'a pas été fait de quantification. On peut cependant observer qu'identifier, sur une expérience, plus de 50% de protéines potentiellement utilisables comme biomarqueur, est une bonne proportion. Il pourrait être intéressant, pour compléter nos travaux de réaliser une étude plus complète et quantitative du stroma tumoral afin de déterminer si ces protéines voient leur expression modifiée et si ce pourrait être des candidats biomarqueurs.

D'autre part, comme nous l'avons mentionné, certaines de ces protéines sont intéressantes par leur rôle dans la modulation de la matrice extracellulaire (notamment la Desmine, le TGF $\beta$ 1, la Vimentine, ou le Versican). En effet, ces protéines se situent généralement dans l'environnement de la tumeur et participe à son développement. Elles peuvent aussi, par exemple, intervenir dans la modulation de la matrice extracellulaire pour permettre à la tumeur de coloniser les tissus éloignés. De plus, encore une fois, dans cette étude, nous avons identifié l'OLFM4 parmi les protéines du stroma montrant l'importance de cette protéine. Qui plus est, le fait de la détecter dans plusieurs approches différentes conforte sa pertinence. En effet, cela signifie probablement que c'est une protéine facile à détecter, donc facile à doser, ce qui en fait un bon candidat biomarqueur.

## 1.4 Analyse du glycoprotéome tumoral

Nous avons souhaité analyser un sous-protéome tumoral afin de mettre en évidence de nouveaux biomarqueurs potentiels. Nous avons choisi d'étudier le glycoprotéome car la plupart des protéines sécrétées sont glycosylées, ce qui en fait une source de biomarqueurs intéressantes. De la sorte, nous avons identifié 607 protéines parmi lesquelles 64 étaient présentes dans la base de données des protéines présentes dans les plasmas de patients.

Contrairement à ce que nous nous attendions, ce rapport nous est apparu assez faible. C'est alors que nous regardé combien de protéines étaient glycosylées, et nous en avons

trouvé 53. Ce rendement de 8,7% est relativement faible ce qui peut expliquer pourquoi peu de protéines identifiées sont susceptibles d'être détectées dans le plasma de patients. Ces observations montrent qu'il sera nécessaire pour la suite d'améliorer la spécificité de la technique en augmentant par exemple le nombre de lavages des lectines pour bien éliminer les peptides non spécifiques. Il peut également être envisagé de tester une enzyme de déglycosylation différente, plus robuste.

Ces résultats sont donc préliminaires et mériteraient d'être consolidés. Pour autant, parmi les 53 protéines glycosylées, 8 sont de nouvelles protéines (cf paragraphe 6 p.173), que nous n'avions jamais identifiées auparavant dans le côlon que ce soit chez les tumeurs ou dans les lignées. Cela montre la pertinence de l'analyse d'un sous-protéome pour identifier de nouvelles protéines, et donc de potentiels nouveaux marqueurs.

Par ailleurs, parmi ces nouvelles protéines, l'aggreccan et l'integrin alpha-V sont des protéines impliqués dans la modulation de la matrice extracellulaire et de l'interaction entre les cellules et entre les cellules et la matrice. Ces protéines sont donc probablement impliquées dans les phénomènes de dissémination tumorale.

## 1.5 Analyse de lignées cellulaires

Toujours dans le but d'augmenter notre couverture protéomique et notre liste de biomarqueurs potentiels, nous avons réalisé une étude protéomique de 10 lignées d'adénocarcinome colique. Cette étude nous a permis de réaliser une cartographie du protéome des lignées colorectales, nous permettant ainsi de constituer une base de données de 1 897 protéines.

Comme nous nous y attendions, l'analyse ontologique de ces protéines révèle majoritairement des protéines intracellulaires. Pourtant la comparaison avec la base de données de protéines détectées dans les plasmas de patients révèlent 303 protéines communes. Ces 303 protéines sont donc probablement sécrétées par les lignées cellulaires.

L'analyse ontologique des protéines sécrétées révèle également 257 protéines cellulaires. Cela montre que les analyses ontologiques sont très probablement imprécises et qu'elles sont utiles pour réaliser un premier tri des données. Cependant, l'ensemble des informations apportées doivent être prise avec beaucoup de précautions.

Parmi les 303 protéines nous pouvons relever plusieurs protéines d'intérêt, notamment la protéines CD44, la vimentine ou encore la desmoplakine qui sont des protéines impliquées dans la transition épithélio-mésenchymateuse et dans le phénotype de cellules initiatrices

de cancers.

La décorine et le TGF $\beta$ 1 sont des protéines impliquées dans la modulation de la matrice extracellulaire et pourrait constituer de bons biomarqueurs potentiels. D'autre part, par leur rôle de diminution des contacts entre la matrice et les cellules, ces protéines pourraient également être des marqueurs de dissémination tumorale. Leur présence pourrait être associée à un phénotype métastatique de la tumeur.

Enfin, les septines, protéines impliquées dans la cytokinèse, sont impliquées dans la division cellulaire. Par leur action, elles permettent la libération des deux cellules formées lors de la mitose. La septine 9 est par exemple très étudiée notamment par l'industrie pour proposer d'ici quelques temps un test diagnostique basé sur cette protéine [Misselwitz et al., 2011]. Peut être qu'en couplant cette protéine à celles que nous avons mises en évidence, il serait possible de proposer un test diagnostique précis du cancer colorectal.

## Synthèse

**A** LA lumière de nos résultats, plusieurs choses se dégagent. Nous avons réalisé ce qui nous semble être une des études les plus complètes du protéome associé au cancer colorectal. Nous avons souhaité, par notre analyse, mettre en évidence des candidats biomarqueurs potentiels utilisables en pratique clinique.

Les caractéristiques d'un biomarqueur nous ont conduit à nous intéresser particulièrement aux protéines qu'il était possible de retrouver dans le plasma de patients afin d'être sûr de pouvoir les doser chez ces derniers. C'est la raison pour laquelle nous avons réalisé, à l'aide des données de la littérature, une base de données des protéines identifiées les plasmas de patients. Ces études ont été réalisées avec des paramètres d'analyse particulièrement stringents, de sorte que les protéines ainsi identifiées dans la base de données soient sûres.

Nous avons donc analysé le protéome total de 28 tumeurs du côlon par protéomique quantitative, afin de mettre en évidence des biomarqueurs de l'état d'avancement tumoral. Cela nous a conduit à proposer une signature protéomique de protéines dérégulées dans chacun des quatre stades tumoraux du cancer colorectal. Quatre protéines répondaient aux critères qui nous intéressaient pour mettre en évidence des biomarqueurs précoces du développement tumoral, et nous avons focalisé notre attention sur l'une d'entre elle, l'OLFM4. Le protéome global de tumeur que nous avons pu établir par cette analyse (3 138 protéines) comprend 874 protéines communes avec la base de données de protéines détectables dans les plasmas de patient ce qui correspond à 27%. Autrement dit, nous mettons ici en évidence le fait que lors d'une analyse globale sans *a priori*, plus d'un quart des protéines sont des protéines susceptibles d'être sécrétées et détectées dans les plasmas de patients.

Nous avons également analysé les protéomes de tumeurs en fonction du statut génétique de l'oncogène KRAS. Nous avons alors constaté que l'OLFM4 était également une protéine dont l'expression était modulée en présence de la mutation de KRAS dans la tumeur.

Cette protéine pourrait donc présenter un double rôle, à la fois diagnostique dans le cadre du dépistage des tumeurs précoces, mais également un rôle pronostic chez les patients pour qui le cancer a été diagnostiqué.

En lien avec ces travaux, nous avons également souhaité compléter notre analyse en nous intéressant au stroma des tumeurs. Nous pensions que cette zone, peu étudiée pour le moment dans la littérature, constituait un réservoir des protéines sécrétées par la tumeur. Effectivement, une analyse sans *a priori* du stroma tumoral permet d'identifier plus de 50% de protéines qui sont présentes dans la base de données des protéines plasmatiques que nous avons établie. L'étude de cette zone tumorale semble donc particulièrement intéressante.

Dans l'espoir d'affiner la qualité des biomarqueurs identifiés, nous avons souhaité développer une approche d'analyse du glycoprotéome. Nous avons réussi à mettre en place un protocole qui permet de mettre en évidence des protéines glycosylées dans un mélange complexe. Ce protocole est encore imparfait, mais les résultats préliminaires sont prometteurs puisqu'ils ont permis d'identifier 8 protéines que nous n'avions pas identifiées auparavant, parmi toutes nos analyses.

Toujours avec l'objectif d'étayer notre liste de biomarqueurs potentiels, nous avons également utilisé des modèles cellulaires pour tenter de mettre en évidence des protéines que nous n'avions jusqu'alors pas identifiées. L'analyse de ces lignées nous a permis de mettre en évidence 16% de protéines communes avec la base de données des protéines plasmatiques. Ce chiffre, plus faible que dans les tumeurs peut s'expliquer par le fait qu'il n'y a pas de matrice extracellulaire lors de la culture de cellules dans des boîtes en plastique. Lors de l'extraction protéique, nous ne récupérons que les protéines intracellulaires. Même si certaines d'entre elles sont destinées à être sécrétées plus tard, il ne s'agit pas de la majorité des protéines. D'autre part, ces modèles sont relativement éloignés des tumeurs humaines telles qu'elles se présentent chez les patients.

Par ailleurs, nous avons également réalisé l'analyse du surnageant de culture de l'une de ces lignées (LS180) et nous avons pu identifier 57% de protéines incluses dans la base de données des protéines plasmatiques. À l'inverse du résultat précédent, nous ne récupérons ici que les protéines sécrétées par les cellules, ce qui explique la forte proportion de protéines pouvant être retrouvées dans le sang des patients.

Tout au long de ces expériences, nous avons mis en évidence des protéines qui méritaient qu'on s'y intéresse. C'est notamment le cas de l'OLFM4 qui constitue selon nous un excellent candidat biomarqueur pour toutes les raisons déjà évoquées. Bien que nous

nous soyons attardés sur celui-ci, d'autres peuvent aussi être considérés. C'est notamment le cas du TGF $\beta$ i, de la vimentine, de la décorine, du versican qui sont mis en lumière dans plusieurs de nos expériences et qui constitue également des protéines d'intérêt. Nous avons fait le choix de n'en étudier qu'un seul afin de ne pas nous disperser, mais des études complémentaires pourront par la suite être réalisées sur ces autres marqueurs. Il est probable que la solution la plus efficace réside dans l'étude cumulée de plusieurs biomarqueurs pour obtenir une réponse plus robuste, un diagnostic plus sûr, un pronostic plus précis.



## Cinquième partie

### Conclusion et perspectives



PAR la présente étude, nous proposons une analyse globale du protéome associé à des tumeurs du côlon. Dans la première partie de notre étude, nous avons souhaité analyser le plus en détail possible, le protéome de tumeurs colorectales. Pour cela, nous avons décidé de séparer notre analyse en étudiant d'une part le protéome des tumeurs, et celui du stroma qui leur est associé. D'autre part, nous avons analysé le protéome de lignées pour compléter nos données et nous avons également tester l'étude du sous-protéome constitué par les protéines glycosylées.

L'ensemble de ces données nous a conduit à proposer plusieurs bases de données. D'une part, nous avons réalisé une étude globale des tumeurs pour laquelle nous avons identifié plus de 3 000 protéines. Nous avons également cartographié le protéome de lignées colorectales. Nous avons au passage noté que ces protéomes étaient significativement différents (données non présentées). Le développement d'une technique permettant d'améliorer la couverture protéomique des analyses globales par OFFGEL successifs a également été proposé. Parmi nos bases de données, nous avons focalisé notre attention sur les protéines sécrétées pouvant être retrouvées dans la circulation sanguine des patients. Un très grand nombre de ces protéines sont impliquées dans la régulation de matrice extracellulaire.

Nos analyses nous ont permis de mettre en évidence des candidats biomarqueurs déjà décrits dans la littérature tels que le CEACAM5 ou encore les protéines S100A8 et S100A9 indiquant la pertinence de notre approche. Nous avons alors proposé l'OLFM4 comme candidat biomarqueur potentiel de la détection précoce du cancer colorectal. Par ailleurs, nous avons noté que son niveau d'expression était plus fortement dérégulé dans les tumeurs mutées sur l'oncogène KRAS. Nos études préliminaires chez les patients par ELISA semble montrer une pertinence de ce biomarqueur dans la détection précoce du cancer colorectal. Son taux plasmatique est également supérieur chez la patient présentant des tumeurs mutées sur KRAS. Nous avons également cherché à comprendre certaines de ces fonctions, et nous avons constaté qu'elle semblait augmenter la résistance des tumeurs notamment à l'irinotécan. Son expression semble régulée par le facteur de transcription NF- $\kappa$ B. Elle semble également intervenir dans la régulation des contacts cellulaires en les diminuant, favorisant ainsi probablement la dissémination tumorale. Par ailleurs, l'analyse de ses sites de glycosylation nous a conduit à montrer que ces dernières étaient nécessaires à sa sécrétion, et nous avons également mis en évidence de fait qu'il pouvait y avoir des glycosylations sur des sites ne correspondant pas à la séquence consensus connue.

D'autres protéines issues de nos analyses pourraient également constituer de bons candidats biomarqueurs, et mériteraient également d'être étudiés dans ce contexte. C'est

notamment le cas de la protéine TGF $\beta$ 1 dont l'expression est très souvent modifiée dans nos analyses. La Vimentine, la Décorine ou encore le Versican sont des protéines qui apparaissent dérégulées dans toutes nos analyses et elles pourraient également constituer de bons candidats.

La suite de ce travail pourrait consister à réaliser des dosages de ces marqueurs dans le sang de personnes malades et non malades, afin de tenter de déterminer leur relevance. Ces marqueurs pourraient, s'ils s'avéraient pertinents, être associés pour proposer une aide diagnostique, et pronostique aux cliniciens.

## Références bibliographiques

- Mechanisms of tumor escape : role of tumor microenvironment in inducing apoptosis of cytolytic effector cells.
- Aebersold, R., Mann, M., and Others (2003a). Mass spectrometry-based proteomics. *Nature*, 422(6928) :198–207.
- Aebersold, R., Mann, M., and Others (2003b). Mass spectrometry-based proteomics. *Nature*, 422(6928) :198–207.
- Alazzouzi, H., Alhopuro, P., Salovaara, R., Sammalkorpi, H., Järvinen, H., Mecklin, J. P., Hemminki, A., Schwartz, S., Aaltonen, L. A., and Arango, D. (2005). SMAD4 as a prognostic marker in colorectal cancer. *Clinical cancer research*, 11(7) :2606–2611.
- Alcindor, T. and Beauger, N. (2011). Oxaliplatin : a review in the era of molecularly targeted therapy. *Current Oncology*, 18(1) :18.
- Alhopuro, P., Alazzouzi, H., Sammalkorpi, H., Dávalos, V., Salovaara, R., Hemminki, A., Järvinen, H., Mecklin, J. P., Schwartz, S., and Aaltonen, L. A. (2005). SMAD4 levels and response to 5-fluorouracil in colorectal cancer. *Clinical cancer research*, 11(17) :6311–6316.
- Alvarez-Manilla, G., Atwood, Guo, Y., Warren, N. L., Orlando, R., and Pierce, M. (2006). Tools for Glycoproteomic Analysis Size Exclusion Chromatography Facilitates Identification of Tryptic Glycopeptides with N-linked Glycosylation Sites. *Journal of Proteome Research*, 5(3) :701–708.
- Ambartsumian, N., Sandelin, K., Rank, F., and Gromova, I. (2004). Proteomic Characterization of the Interstitial Fluid Perfusing the Breast Tumor Microenvironment. *Molecular & Cellular Proteomics*, 3 :327.
- Anderson, L. and Hunter, C. L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Molecular & cellular proteomics : MCP*, 5(4) :573–88.
- Anderson, N. L. and Anderson, N. G. (2002). The human plasma proteome history, character, and diagnostic prospects. *Molecular & Cellular Proteomics*, 1(11) :845–867.
- Ang, C. S., Rothacker, J., Patsiouras, H., Burgess, A. W., and Nice, E. C. (2010). Murine fecal proteomics : A model system for the detection of potential biomarkers for colorectal cancer. *Journal of Chromatography A*, 1217(19) :3330–3340.
- Angell, J. E., Lindner, D. J., Shapiro, P. S., Hofmann, E. R., and Kalvakolanu, D. V. (2000). Identification of GRIM-19, a novel cell death-regulatory gene induced by the interferon-beta and retinoic acid combination, using a genetic approach. *J Biol Chem*, 275(43) :33416–33426.
- Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database1. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1473(1) :4–8.
- Bal, R. S. and Anholt, R. R. (1993). Formation of the extracellular mucous matrix of olfactory neuroepithelium : identification of partially glycosylated and nonglycosylated precursors of olfactomedin. *Biochemistry*, 32(4) :1047–1053.
- Balazy, M. (2004). Eicosanomics : targeted lipidomics of eicosanoids in biological systems. *Prostaglandins & other lipid mediators*, 73(3) :173–180.
- Baldwin, R. (1971). Tumour-associated antigens and tumour-host interactions. *Proceedings of the Royal Society of Medicine*, 64(7).
- Baldwin, R. W. (1966). Tumour-specific immunity against spontaneous rat tumours. *International Journal of Cancer*, 1(3) :257–264.
- Barker, N. and Clevers, H. (2010). Leucine-rich repeat-containing G-protein-coupled receptors as markers of adult stem cells. *Gastroenterology*, 138(5) :1681–1696.

- Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*, 457(7229) :608–611.
- Bartel, D. P. (2004). MicroRNAs : genomics, biogenesis, mechanism, and function. *Cell*, 116(2) :281–297.
- Beavis, R. C., Chait, B. T., and Fales, H. M. (1989). Cinnamic acid derivatives as matrices for ultraviolet laser desorption mass spectrometry of proteins. *Rapid Communications in Mass Spectrometry*, 3(12) :432–435.
- Beavis, R. C., Chaudhary, T., and Chait, B. T. (1992).  $\alpha$ -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Organic Mass Spectrometry*, 27(2) :156–158.
- Becker, J. W., Reeke Jr, G. N., Wang, J. L., Cunningham, B. A., and Edelman, G. M. (1975). The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *Journal of Biological Chemistry*, 250(4) :1513–1524.
- Belot, A., Velten, M., Grosclaude, P., Bossard, N., Launoy, G., Remontet, L., Benhamou, E., Chérié-Challine, L., and Others (2008). *Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2005*. Institut de veille sanitaire.
- Ben-Porath, I. and Weinberg, R. A. (2005). The signals and pathways activating cellular senescence. *The International Journal of Biochemistry & Cell Biology*, 37(5) :961–976.
- Bendall, S. C., Hughes, C., Stewart, M. H., Doble, B., Bhatia, M., and Lajoie, G. A. (2008). Prevention of amino acid conversion in SILAC experiments with embryonic stem cells. *Molecular & Cellular Proteomics*, 7(9) :1587–1597.
- Besson, D., Pavageau, A. H., Valo, I., Bourreau, A., Bélanger, A., Eymerit-Morin, C., Moulière, A., Chassevent, A., Boisdrion-Celle, M., Morel, A., and Others (2011). A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. *Molecular & Cellular Proteomics*, 10(12).
- Blagoev, B. and Mann, M. (2006). Quantitative proteomics to study mitogen-activated protein kinases. *Methods (San Diego, Calif.)*, 40(3) :243–50.
- Boccaccio, C., Ande, M., Tamagnone, L., Bardelli, A., Michieli, P., Battistini, C., and Comoglio, P. M. (1998). Induction of epithelial tubules by growth factor HGF depends on the STAT pathway. *Nature*, 391(6664) :285–288.
- Boisvert, F. M., Lam, Y. W., Lamont, D., and Lamond, A. I. (2010). A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Molecular & Cellular Proteomics*, 9(3) :457–470.
- Bose, R., Verheij, M., Haimovitz-Friedman, a., Scotto, K., Fuks, Z., and Kolesnick, R. (1995). Ceramide synthase mediates daunorubicin-induced apoptosis : an alternative mechanism for generating death signals. *Cell*, 82(3) :405–14.
- Bromberg, J. (2002). Stat proteins and oncogenesis. *J Clin Invest*, 109(9) :1139–1142.
- Bromberg, J. F., Horvath, C. M., Besser, D., Lathem, W. W., and Darnell, J. E. (1998). Stat3 activation is required for cellular transformation by v-src. *Mol Cell Biol*, 18(5) :2553–2558.
- Bromberg, J. F., Wrzeszczynska, M. H., Devgan, G., Zhao, Y., Pestell, R. G., Albanese, C., and Darnell, J. E. (1999). Stat3 as an oncogene. *Cell*, 98(3) :295–303.
- Brown, J. P. (1997). Bypass of Senescence After Disruption of p21CIP1/WAF1 Gene in Normal Diploid Human Fibroblasts. *Science*, 277(5327) :831–834.
- Brun, V., Masselon, C., Garin, J., and Dupuis, A. (2009). Isotope dilution strategies for absolute quantitative proteomics. *Journal of Proteomics*, 72(5) :740–749.
- Brusniak, M.-Y. K., Kwok, S.-T., Christiansen, M., Campbell, D., Reiter, L., Picotti, P., Kusebauch, U., Ramos, H., Deutsch, E. W., Chen, J., Moritz, R. L., and Aebersold, R. (2011). ATAQS : A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. *BMC bioinformatics*, 12(1) :78.
- Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., and Winiewski, J. R. (2004). Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*, 4(2) :454–465.
- Calin, G. a. and Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature reviews. Cancer*, 6(11) :857–66.
- Campisi, J. and d'Adda di Fagagna, F. (2007). Cellular senescence : when bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, 8(9) :729–740.

- Cao, Z., Yende, S., Kellum, J. a., and Robinson, R. a. a. S. (2013). Additions to the Human Plasma Proteome via a Tandem MARS Depletion iTRAQ-Based Workflow. *International journal of proteomics*, 2013 :654356.
- Cargile, B. J., Talley, D. L., and Stephenson, J. L. (2004). Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis*, 25(6) :936–945.
- Chaerkady, R. and Pandey, A. (2007). Quantitative proteomics for identification of cancer biomarkers. *Proteomics-Clinical Applications*, 1(9) :1080–1089.
- Chang, B. D., Swift, M. E., Shen, M., Fang, J., Broude, E. V., and Roninson, I. B. (2002). Molecular determinants of terminal growth arrest induced in tumor cells by a chemotherapeutic agent. *Proceedings of the National Academy of Sciences*, 99(1) :389–394.
- Chang, B. D., Watanabe, K., Broude, E. V., Fang, J., Poole, J. C., Kalinichenko, T. V., and Roninson, I. B. (2000). Effects of p21Waf1/Cip1/Sdi1 on cellular gene expression : implications for carcinogenesis, senescence, and age-related diseases. *Proceedings of the National Academy of Sciences*, 97(8) :4291–4296.
- Chen, J., Hunter, D. J., Stampfer, M. J., Kyte, C., Chan, W., Wetmur, J. G., Mosig, R., Selhub, J., and Ma, J. (2003). Polymorphism in the thymidylate synthase promoter enhancer region modifies the risk and survival of colorectal cancer. *Cancer Epidemiology Biomarkers & Prevention*, 12(10) :958–962.
- Chen, J.-S., Chen, K.-T., Fan, C.-W., Han, C.-L., Chen, Y.-J., Yu, J.-S., Chang, Y.-S., Chien, C.-W., Wu, C.-P., and Hung, R.-P. (2010). Comparison of membrane fraction proteomic profiles of normal and cancerous human colorectal tissues with gel-assisted digestion and iTRAQ labeling mass spectrometry. *FEBS Journal*, 277(14) :3028–3038.
- Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J., and Guo, X. (2008). Characterization of microRNAs in serum : a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research*, 18(10) :997–1006.
- Chenau, J., Michelland, S., and Seve, M. (2008). Le sécrétome : définitions et intérêt biomédical. *La Revue de Médecine Interne*, 29(7) :606–608.
- Chevallet, M., Luche, S., and Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. *Nat. Protocols*, 1(4) :1852–1858.
- Choe, L., D’Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., and Lee, K. H. (2007). 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer’s disease. *Proteomics*, 7(20) :3651–60.
- Choi, J., Shendrik, I., Peacocke, M., Peehl, D., Buttyan, R., Ikeguchi, E. F., Katz, a. E., and Benson, M. C. (2000). Expression of senescence-associated beta-galactosidase in enlarged prostates from men with benign prostatic hyperplasia. *Urology*, 56(1) :160–6.
- Cleves, A. E. (1997). Protein transport : the nonclassical ins and outs. *Current Biology*, 7(5) :R318–R320.
- Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H., and Young, R. A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev*, 22(6) :746–755.
- Collado, M., Gil, J., Efeyan, A., Guerra, C., Schuhmacher, A. J., Barradas, M., Benguría, A., Zaballos, A., Flores, J. M., Barbacid, M., Beach, D., and Serrano, M. (2005). Tumour biology : senescence in premalignant tumours. *Nature*, 436(7051) :642.
- Cooper, D. N. and Barondes, S. H. (1990). Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. *The Journal of cell biology*, 110(5) :1681–1691.
- Coppe, J.-P., Boisen, M., Sun, C. H., Wong, B. J. F., Kang, M. K., Park, N.-H., Desprez, P.-Y., Campisi, J., and Krtolica, A. (2008). A role for fibroblasts in mediating the effects of tobacco-induced epithelial cell growth and invasion. *Molecular cancer research : MCR*, 6(7) :1085–98.
- Coppé, J.-P., Desprez, P.-Y., Krtolica, A., and Campisi, J. (2010). The Senescence-Associated Secretory Phenotype : The Dark Side of Tumor Suppression. *Annual Review of Pathology : Mechanisms of Disease*, 5(1) :99–118.
- Coppé, J.-P., Kauser, K., Campisi, J., and Beauséjour, C. M. (2006). Secretion of vascular endothelial growth factor by primary human fibroblasts at senescence. *The Journal of biological chemistry*, 281(40) :29568–74.
- Coppé, J.-P. P., Patil, C. K., Rodier, F., Sun, Y., Muñoz, D. P., Goldstein, J., Nelson, P. S., Desprez, P.-Y. Y., and Campisi, J. (2008). Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. *PLoS biology*, 6(12) :e301.

- Cullen, P. J. and Lockyer, P. J. (2002). Integration of calcium and Ras signalling. *Nature reviews. Molecular cell biology*, 3(5) :339–48.
- David Cunningham, M. D., Heinz-Josef Lenz, M. D., Lynch, H. T., Bruce Minsky, M. D., Bernard Nordlinger, M. D., and Starling, N. (2012). Lancet Seminar : Colorectal cancer. *The Lancet*, 375 :1030–1047.
- Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., and Bottomley, W. (2002). Mutations of the BRAF gene in human cancer. *Nature*, 417(6892) :949–954.
- de Carne Trecesson, S., Guillemin, Y., Belanger, A., Bernard, A.-C., Preisser, L., Ravon, E., Gamelin, E., Juin, P., Barre, B., and Coqueret, O. (2011). Escape from p21-mediated Oncogene-induced Senescence Leads to Cell Dedifferentiation and Dependence on Anti-apoptotic Bcl-xL and MCL1 Proteins. *Journal of Biological Chemistry*, 286(15) :12825–12838.
- de Hoffmann, E. (1996). Tandem mass spectrometry : a primer. *Journal of mass spectrometry*.
- de Wit, M., Jimenez, C. R., Carvalho, B., Belien, J. A. M., Delis-van Diemen, P. M., Mongera, S., Piersma, S. R., Vikas, M., Navani, S., and Pontén, F. (2012). Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. *Gut*, 61(6) :855–864.
- Dempsey, E. and Rudd, P. M. (2012). Acute phase glycoproteins : bystanders or participants in carcinogenesis? *Annals of the New York Academy of Sciences*, 1253(1) :122–132.
- Dennis, J. W., Nabi, I. R., and Demetriou, M. (2009). Metabolism, Cell Surface Organization, and Disease. *Cell*, 139(7) :1229–1241.
- Desmetz, C., Lacombe, J., and Mange, A. (2011). Autoanticorps et diagnostic précoce des cancers. *MS. Médecine Sciences*, 27 :633–638.
- Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre', M., Nuciforo, P. G., Bensimon, A., Maestro, R., Pelicci, P. G., and d'Adda di Fagagna, F. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature*, 444(7119) :638–642.
- Di Nicolantonio, F., Martini, M., Molinari, F., Sartore-Bianchi, A., Arena, S., Saletti, P., De Dosso, S., Mazzucchelli, L., Frattini, M., Siena, S., and Bardelli, A. (2008). Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 26(35) :5705–12.
- Di Palma, S., Stange, D., van de Wetering, M., Clevers, H., Heck, A. J. R., and Mohammed, S. (2011). Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. *Journal of proteome research*, 10(8) :3814–3819.
- Doroudgar, S. and Glembotski, C. C. (2011). The cardiokine story unfolds : ischemic stress-induced protein secretion in the heart. *Trends in molecular medicine*, 17(4) :207–214.
- Dowling, P. and Clynes, M. (2011). Conditioned media from cell lines : A complementary model to clinical specimens for the discovery of disease-specific biomarkers. *PROTEOMICS*, 11(4) :794–804.
- Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nature Reviews Cancer*, 3(1) :11–22.
- Dube, D. H. and Bertozzi, C. R. (2005). Glycans in cancer and inflammation—potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery*, 4(6) :477–488.
- Dupuis, A., Hennekinne, J.-A., Garin, J., and Brun, V. (2008). Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*, 8(22) :4633–6.
- Edelman, G. M. and Gall, W. E. (1971). Cascade chromatography and automated multidimensional fractionation. *Proceedings of the National Academy of Sciences*, 68(7) :1444–1449.
- El-aneed, A., Cohen, A., Banoub, J., Projects, S., and Canada, O. (2009). Applied Spectroscopy Reviews. (February 2013) :210–230.
- Elliott, M. H., Smith, D. S., Parker, C. E., and Borchers, C. (2009). Current trends in quantitative proteomics. *Journal of Mass Spectrometry*, pages n/a–n/a.
- Elsaleh, H. and Iacopetta, B. (2001). Microsatellite instability is a predictive marker for survival benefit from adjuvant chemotherapy in a population-based series of stage III colorectal carcinoma. *Clinical colorectal cancer*, 1(2) :104–109.
- Elsaleh, H., Powell, B., McCaul, K., Grieu, F., Grant, R., Joseph, D., and Iacopetta, B. (2001). P53 alteration and microsatellite instability have predictive value for survival benefit from chemotherapy in stage III colorectal carcinoma. *Clinical cancer research*, 7(5) :1343–1349.

- Ernoul, E., Bourreau, A., Gamelin, E., and Guette, C. (2009). A proteomic approach for plasma biomarker discovery with iTRAQ labelling and OFFGEL fractionation. *BioMed Research International*, 2010.
- Ernoul, E., Gamelin, E., and Guette, C. (2008). Improved proteome coverage by using iTRAQ labelling and OFFGEL fractionation. *Proteome Science*, 6(27).
- Essader, A. S., Cargile, B. J., Bundy, J. L., and Stephenson, J. L. (2005). A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics*, 5(1) :24–34.
- Evans, C., Noirel, J., Ow, S. Y., Salim, M., Pereira-Medrano, A. G., Couto, N., Pandhal, J., Smith, D., Pham, T. K., Karunakaran, E., Zou, X., Biggs, C. A., and Wright, P. C. (2012). An insight into iTRAQ : where do we stand now? *Analytical and Bioanalytical Chemistry*.
- Fang, D. D., Kim, Y. J., Lee, C. N., Aggarwal, S., McKinnon, K., Mesmer, D., Norton, J., Birse, C. E., He, T., and Ruben, S. M. (2010). Expansion of CD133+ colon cancer cultures retaining stem cell properties to enable cancer stem cell target discovery. *British journal of cancer*, 102(8) :1265–1275.
- Farrell, a. M., Uchida, Y., Nagiec, M. M., Harris, I. R., Dickson, R. C., Elias, P. M., and Holleran, W. M. (1998). UVB irradiation up-regulates serine palmitoyltransferase in cultured human keratinocytes. *Journal of lipid research*, 39(10) :2031–8.
- Fijneman, R. J. A., de Wit, M., Pourghiasian, M., Piersma, S. R., Pham, T. V., Warmoes, M. O., Lavaei, M., Piso, C., Smit, F., and Delis-van Diemen, P. M. (2012). Proximal fluid proteome profiling of mouse colon tumors reveals biomarkers for early diagnosis of human colorectal cancer. *Clinical Cancer Research*, 18(9) :2613–2624.
- Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs : are the answers in sight? *Nature reviews. Genetics*, 9(2) :102–14.
- Fischer, F. and Poetsch, A. (2006). Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome science*, 4(1) :2.
- Flavin, P., Redmond, A., McBryan, J., Cocchiglia, S., Tibbitts, P., Fahy-Browne, P., Kay, E., Treumann, A., Perrem, K., and McIlroy, M. (2011). RuvB12 cooperates with Ets2 to transcriptionally regulate hTERT in colon cancer. *FEBS letters*, 585(15) :2537–2544.
- Fletcher, R. H. (1986). Carcinoembryonic antigen. *Ann Intern Med*, 104(1) :66–73.
- Freue, G. V. C., Sasaki, M., Meredith, A., Günther, O. P., Bergman, A., Takhar, M., Mui, A., Balshaw, R. F., Ng, R. T., and Opushneva, N. (2010). Proteomic signatures in plasma during early acute renal allograft rejection. *Molecular & Cellular Proteomics*, 9(9) :1954–1967.
- Gan, C. S., Chong, P. K., Pham, T. K., and Wright, P. C. (2007). Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *Journal of proteome research*, 6(2) :821–827.
- Garfinkel, S., Brown, S., Wessendorf, J. H., and Maciag, T. (1994). Post-transcriptional regulation of interleukin 1 alpha in various strains of young and senescent human umbilical vein endothelial cells. *Proceedings of the National Academy of Sciences*, 91(4) :1559–1563.
- Gauci, V. J., Padula, M. P., and Coorssen, J. R. (2013). Coomassie blue staining for high sensitivity gel-based proteomics. *Journal of proteomics*, pages 1–11.
- Gerber, S. a., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12) :6940–5.
- Gevaert, K., Impens, F., Ghesquière, B., Van Damme, P., Lambrechts, A., and Vandekerckhove, J. (2008). Stable isotopic labeling in proteomics. *Proteomics*, 8(23-24) :4873–4885.
- Ghosh, D., Yu, H., Tan, X. F., Lim, T. K., Zubaidah, R. M., Tan, H. T., Chung, M. C., and Lin, Q. (2011). Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. *Journal of proteome research*, 10(10) :4373–4387.
- Goldkorn, T., Balaban, N., Shannon, M., Chea, V., Matsukuma, K., Gilchrist, D., Wang, H., and Chan, C. (1998). H2O2 acts on cellular membranes to generate ceramide signaling and initiate apoptosis in tracheobronchial epithelial cells. *Journal of cell science*, 111 ( Pt 2 :3209–20.

- Gorgoulis, V. G. and Halazonetis, T. D. (2010). Oncogene-induced senescence : the bright and dark side of the response. *Curr Opin Cell Biol*, 22(6) :816–827.
- Grady, W. M. and Markowitz, S. (2000). Genomic instability and colorectal cancer. *Current opinion in gastroenterology*, 16(1) :62.
- Green, J., Timmcke, A., and Mitchell, W. (1993). Mucinous carcinoma—just another colon cancer? *Diseases of the colon & . . .*
- Gregorieff, A. and Clevers, H. (2005). Wnt signaling in the intestinal epithelium : from endoderm to cancer. *Genes Dev*, 19(8) :877–890.
- Gut, I. G. (2004). DNA analysis by MALDI-TOF mass spectrometry. *Human mutation*, 23(5) :437–441.
- Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences*, 97(17) :9390–9395.
- Hägglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., and Roepstorff, P. (2004). A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *Journal of proteome research*, 3(3) :556–566.
- Halvey, P. J., Zhang, B., Coffey, R. J., Liebler, D. C., and Slebos, R. J. C. (2011). Proteomic consequences of a single gene mutation in a colorectal cancer model. *Journal of Proteome Research*, 11(2) :1184–1195.
- Hamelin, C., Cornut, E., Poirier, F., Pons, S., Beaulieu, C., Charrier, J.-P., Haïdous, H., Cotte, E., Lambert, C., Piard, F., Ataman-Önal, Y., and Choquet-Kastylevsky, G. (2011). Identification and verification of heat shock protein 60 as a potential serum marker for colorectal cancer. *The FEBS journal*, 278(24) :4845–59.
- Han, C., Abel, P. W., and Minneman, K. P. (1987). Heterogeneity of alpha 1-adrenergic receptors revealed by chlorethyl-clonidine. *Molecular pharmacology*, 32(4) :505–510.
- Han, C.-L., Chen, J.-S., Chan, E.-C., Wu, C.-P., Yu, K.-H., Chen, K.-T., Tsou, C.-C., Tsai, C.-F., Chien, C.-W., Kuo, Y.-B., Lin, P.-Y., Yu, J.-S., Hsueh, C., Chen, M.-C., Chan, C.-C., Chang, Y.-S., and Chen, Y.-J. (2011). An informatics-assisted label-free approach for personalized tissue membrane proteomics : case study on colorectal cancer. *Molecular & cellular proteomics : MCP*, 10(4) :M110.003087.
- Han, X., Aslanian, A., and Yates, J. R. (2008). Mass spectrometry for proteomics. *Current Opinion in Chemical Biology*, 12(5) :483–490.
- Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer : the next generation. *Cell*, 144(5) :646–674.
- Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A., and Burlingame, A. L. (2003). Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable <sup>13</sup>C-Isotope-coded Affinity Tag and Multidimensional Chromatography. *Molecular & Cellular Proteomics*, pages 299–314.
- Hao, J., Li, T.-G., Qi, X., Zhao, D.-F., and Zhao, G.-Q. (2006). WNT/beta-catenin pathway up-regulates Stat3 and converges on LIF to prevent differentiation of mouse embryonic stem cells. *Dev Biol*, 290(1) :81–91.
- Harris, L. R., Churchward, M. A., Butt, R. H., and Coorsen, J. R. (2007). Assessing detection methods for gel-based proteomic analyses. *Journal of proteome research*, 6(4) :1418–1425.
- Harvey, D. J. (2003). Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *International Journal of Mass Spectrometry*, 226(1) :1–35.
- Haura, E. B., Turkson, J., and Jove, R. (2005). Mechanisms of disease : Insights into the emerging role of signal transducers and activators of transcription in cancer. *Nat Clin Pract Oncol*, 2(6) :315–324.
- Hayden, M. S. and Ghosh, S. (2004). Signaling to NF-kappaB. *Genes Dev*, 18(18) :2195–2224.
- Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, 37(3) :614–636.
- He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998). Identification of c-MYC as a target of the APC pathway. *Science*, 281(5382) :1509–1512.
- Herbst, R. S. (2004). Review of epidermal growth factor receptor biology. *International Journal of Radiation Oncology\* Biology\* Physics*, 59(2) :S21–S26.
- Hillenkamp, F. and Karas, M. (2007). The MALDI process and method. *MALDI MS*, page 1785.

- Ho, S. B., Hyslop, A., Albrecht, R., Jacobson, A., Spencer, M., Rothenberger, D. A., Niehans, G. A., D’Cunha, J., and Kratzke, R. A. (2004). Quantification of colorectal cancer micrometastases in lymph nodes by nested and real-time reverse transcriptase-PCR analysis for carcinoembryonic antigen. *Clin Cancer Res*, 10(17) :5777–5784.
- Hoare, M. and Narita, M. (2013). Transmitting senescence to the cell neighbourhood. *Nature cell biology*, 15(8) :887–9.
- Hoffert, J. D., Pisitkun, T., Saeed, F., Song, J. H., Chou, C.-L., and Knepper, M. A. (2012). Dynamics of the G protein-coupled vasopressin V2 receptor signaling network revealed by quantitative phosphoproteomics. *Molecular & Cellular Proteomics*, 11(2).
- Hu, J., Qian, J., Borisov, O., Pan, S., Li, Y., Liu, T., Deng, L., Wannemacher, K., Kurnellas, M., and Patterson, C. (2006). Optimized proteomic analysis of a mouse model of cerebellar dysfunction using amine-specific isobaric tags. *Proteomics*, 6(15) :4321–4334.
- Huang, C., Nakatsuji, T., Liu, Y., and Shi, Y. (2008). In vivo tumor secretion probing via ultrafiltration and tissue chamber : implication for anti-cancer drugs targeting secretome. *Patents on anti-cancer drug discovery*, 3(1) :48–54.
- Huang, Y., Yang, M., Yang, H., and Zeng, Z. (2010). Upregulation of the GRIM-19 gene suppresses invasion and metastasis of human gastric cancer SGC-7901 cell line. *Exp Cell Res*, 316(13) :2061–2070.
- Hundt, S., Haug, U., and Brenner, H. (2007). Blood markers for early detection of colorectal cancer : a systematic review. *Cancer Epidemiology Biomarkers & Prevention*, 16(10) :1935–1953.
- Hussain, S. M., Outwater, E. K., and Siegelman, E. S. (1999). Mucinous versus nonmucinous rectal carcinomas : differentiation with MR imaging. *Radiology*, 213(1) :79–85.
- Hwang, S. I., Lundgren, D. H., Mayya, V., Rezaul, K., Cowan, A. E., Eng, J. K., and Han, D. K. (2006). Systematic Characterization of Nuclear Proteome during Apoptosis A Quantitative Proteomic Study by Differential Extraction and Stable Isotope Labeling. *Molecular & Cellular Proteomics*, 5(6) :1131–1145.
- Iliopoulos, D., Hirsch, H. A., and Struhl, K. (2009). An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell*, 139(4) :693–706.
- Ionov, Y., Peinado, M., and Malkhosyan, S. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*.
- Ishida, Y. (1997). Inhibition of Carnitine Palmitoyltransferase I Augments Sphingolipid Synthesis and Palmitate-induced Apoptosis. *Journal of Biological Chemistry*, 272(6) :3324–3329.
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics*, 4(9) :1265–1272.
- Jafari, M., Primo, V., Smejkal, G. B., Moskovets, E. V., Kuo, W. P., and Ivanov, A. R. (2012). Comparison of in-gel protein separation techniques commonly used for fractionation in mass spectrometry-based proteomic profiling. *Electrophoresis*, 33(16) :2516–2526.
- James, P. (1997). Protein identification in the post-genome era : the rapid rise of proteomics. *Quarterly reviews of biophysics*, 30(4) :279–331.
- Janković, M. (2011). Glycans as Biomarkers : Status and Perspectives. *Journal of Medical Biochemistry*, 30(3) :213–223.
- Jass, J. R., Biden, K. G., Cummings, M. C., Simms, L. A., Walsh, M., Schoch, E., Meltzer, S. J., Wright, C., Searle, J., and Young, J. (1999). Characterisation of a subtype of colorectal cancer combining features of the suppressor and mild mutator pathways. *Journal of clinical pathology*, 52(6) :455–460.
- Jass, J. R., Do, K. A., Simms, L. A., Iino, H., Wynter, C., Pillay, S. P., Searle, J., Radford-Smith, G., Young, J., and Leggett, B. (1998). Morphology of sporadic colorectal cancer with DNA replication errors. *Gut*, 42(5) :673–679.
- Jones, H. B. (1848). On a new substance occurring in the urine with mollities ossium. *Phil Trans R Soc Lond*, 138 :55–62.
- Jover, R., Zapater, P., Castells, A., Llor, X., Andreu, M., Cubiella, J., Balaguer, F., Sempere, L., Xicola, R. M., and Bujanda, L. (2009). The efficacy of adjuvant chemotherapy with 5-fluorouracil in colorectal cancer depends on the mismatch repair status. *European Journal of Cancer*, 45(3) :365–373.
- Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K.-i., Takahashi, N., and Isobe, T. (2003). Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat Biotech*, 21(6) :667–672.

- Kaneto, H., Fujii, J., Myint, T., Miyazawa, N., Islam, K. N., Kawasaki, Y., Suzuki, K., Nakamura, M., Tatsumi, H., Yamasaki, Y., and Taniguchi, N. (1996). Reducing sugars trigger oxidative modification and apoptosis in pancreatic. *863* :855–863.
- Karas, M., Glückmann, M., and Schäfer, J. (2000). Ionization in matrix-assisted laser desorption/ionization : singly charged molecular ions are the lucky survivors. *Journal of mass spectrometry*, 35(1) :1–12.
- Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical chemistry*, 60(20) :2299–2301.
- Karin, M. (2006a). NF-kappaB and cancer : mechanisms and targets. *Mol Carcinog*, 45(6) :355–361.
- Karin, M. (2006b). Nuclear factor-kappaB in cancer development and progression. *Nature*, 441(7092) :431–436.
- Karin, M., Cao, Y., Greten, F. R., and Li, Z.-W. (2002). NF-kappaB in cancer : from innocent bystander to major culprit. *Nat Rev Cancer*, 2(4) :301–310.
- Karp, N. A., Huber, W., Sadowski, P. G., Charles, P. D., Hester, S. V., and Lilley, K. S. (2010). Addressing accuracy and precision issues in iTRAQ quantitation. *Molecular & Cellular Proteomics*, 9(9) :1885–1897.
- Kawada, M., Seno, H., Uenoyama, Y., Sawabu, T., Kanda, N., Fukui, H., Shimahara, Y., and Chiba, T. (2006). Signal transducers and activators of transcription 3 activation is involved in nuclear accumulation of beta-catenin in colorectal cancer. *Cancer Res*, 66(6) :2913–2917.
- Keidel, E.-M., Dosch, D., Brunner, A., Kellermann, J., and Lottspeich, F. (2011). Evaluation of protein loading techniques and improved separation in OFFGEL isoelectric focusing. *Electrophoresis*, 32(13) :1659–1666.
- Kenrick, K. G. and Margolis, J. (1970). Isoelectric focusing and gradient gel electrophoresis : A two-dimensional technique. *Analytical Biochemistry*, 33(1) :204–207.
- Kerr, J., Wyllie, A., and Currie, A. (1972). Apoptosis : a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British journal of cancer*.
- Keshamouni, V. G., Michailidis, G., Grasso, C. S., Anthwal, S., Strahler, J. R., Walker, A., Arenberg, D. a., Reddy, R. C., Akulapalli, S., Thannickal, V. J., Standiford, T. J., Andrews, P. C., and Omenn, G. S. (2006). Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *Journal of proteome research*, 5(5) :1143–54.
- Kim, H. J., Kang, U. B., Lee, H., Jung, J. H., Lee, S. T., Yu, M. H., Kim, H., and Lee, C. (2011). Profiling of differentially expressed proteins in stage IV Colorectal cancers with good and poor outcomes. *Journal of proteomics*.
- Klein-Scory, S., Kübler, S., Diehl, H., Eilert-Micus, C., Reinacher-Schick, A., Stühler, K., Warscheid, B., Meyer, H. E., Schmiegel, W., and Schwarte-Waldhoff, I. (2010). Immunoscreening of the extracellular proteome of colorectal cancer cells. *BMC cancer*, 10(1) :70.
- Kolarich, D., Jensen, P. H., Altmann, F., and Packer, N. H. (2012). Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protocols*, 7(7) :1285–1298.
- Konishi, M., Kikuchi-Yanoshita, R., Tanaka, K., Muraoka, M., Onda, a., Okumura, Y., Kishi, N., Iwama, T., Mori, T., Koike, M., Ushio, K., Chiba, M., Nomizu, S., Konishi, F., Utsunomiya, J., and Miyaki, M. (1996). Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. *Gastroenterology*, 111(2) :307–17.
- Krishnan, S., Gaspari, M., Corte, A. D., Bianchi, P., Crescente, M., Cerletti, C., Torella, D., Indolfi, C., de Gaetano, G., and Donati, M. B. (2011). OFFgel-based multidimensional LC-MS/MS approach to the cataloguing of the human platelet proteome for an interactomic profile. *Electrophoresis*, 32(6-7) :686–695.
- Krtolica, a., Parrinello, S., Lockett, S., Desprez, P. Y., and Campisi, J. (2001). Senescent fibroblasts promote epithelial cell growth and tumorigenesis : a link between cancer and aging. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21) :12072–7.
- Kuilman, T., Michaloglou, C., Vredeveld, L. C., Douma, S., van Doorn, R., Desmet, C. J., Aarden, L. A., Mooi, W. J., and Peeper, D. S. (2008). Oncogene-Induced Senescence Relayed by an Interleukin-Dependent Inflammatory Network. *Cell*, 133(6) :1019–1031.
- Kulkarni, N. H., Karavanich, C. A., Atchley, W. R., and Anholt, R. R. (2000). Characterization and differential expression of a human gene family of olfactomedin-related proteins. *Genet Res*, 76(1) :41–50.

- Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics : a tutorial. *Molecular systems biology*, 4(222) :222.
- Larsen, M. R., Hø jrup, P., and Roepstorff, P. (2005). Characterization of gel-separated glycoproteins using two-step proteolytic digestion combined with sequential microcolumns and mass spectrometry. *Molecular & Cellular Proteomics*, 4(2) :107–119.
- Lee, C. H., Jeon, Y.-T., Kim, S.-H., and Song, Y.-S. (2007). NF-kappaB as a potential molecular target for cancer therapy. *Biofactors*, 29(1) :19–35.
- Li, B., An, H. J., Kirmiz, C., Lebrilla, C. B., Lam, K. S., and Miyamoto, S. (2008). Glycoproteomic analyses of ovarian cancer cell lines and sera from ovarian cancer patients show distinct glycosylation changes in individual proteins. *Journal of proteome research*, 7(9) :3776–3788.
- Li, F. Y. and Lai, M.-D. M.-d. (2009). Colorectal cancer, one entity or three. *Journal of Zhejiang University*, 10(3) :219–229.
- Li, W. Q., Kawakami, K., Ruzskiewicz, A., Bennett, G., Moore, J., and Iacopetta, B. (2006). BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status. *Molecular cancer*, 5(1) :2.
- Lievre, A., Bachet, J.-B., Boige, V. A., Cayre, A., Le Corre, D., Buc, E., Ychou, M., BouchÃ©, O., Landi, B., Louvet, C., AndrÃ©, T., Bibeau, F. A., Diebold, M.-D. A., Rougier, P., Ducreux, M., Tomasic, G., Emile, J.-F. A., Penault-Llorca, F. A. A., and Laurent-Puig, P. (2008). KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol*, 26(3) :374–379.
- Lievre, A., Bachet, J.-B., Le Corre, D., Boige, V., Landi, B., Emile, J.-F., Côté, J.-F., Tomasic, G., Penna, C., and Ducreux, M. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer research*, 66(8) :3992–3995.
- Lin, J.-K., Lin, C.-C., Yang, S.-H., Wang, H.-S., Jiang, J.-K., Lan, Y.-T., Lin, T.-C., Li, A. F.-Y., Chen, W.-S., and Chang, S.-C. (2011). Early postoperative CEA level is a better prognostic indicator than is preoperative CEA level in predicting prognosis of patients with curable colorectal cancer. *International journal of colorectal disease*, 26(9) :1135–41.
- Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nature biotechnology*, 17(7) :676–682.
- Liu, H., Sadygov, R., and Yates III, J. (2004). A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal Chem*, 76 :4193–4201.
- Liu, W., Chen, L., Zhu, J., and Rodgers, G. P. (2006). The glycoprotein hGC-1 binds to cadherin and lectins. *Experimental cell research*, 312(10) :1785–97.
- Liu, W., Zhu, J., Cao, L., and Rodgers, G. P. (2007). Expression of hGC-1 is correlated with differentiation of gastric carcinoma. *Histopathology*, 51(2) :157–165.
- Longley, D. B., Harkin, D. P., and Johnston, P. G. (2003). 5-Fluorouracil : Mechanisms of Action and Clinical Strategies. *Nature reviews. Cancer*, 3(5) :330–8.
- Lothe, R. A., Peltomäki, P., Meling, G. I., Aaltonen, L. A., Nyström-Lahti, M., Pylkkänen, L., Heimdal, K., Andersen, T. I., Mø ller, P., and Rognum, T. O. (1993). Genomic instability in colorectal cancer : relationship to clinicopathological variables and family history. *Cancer Research*, 53(24) :5849–5852.
- Lu, S.-Y. (2006). Ripe areca nut extract induces G1 phase arrests and senescence-associated phenotypes in normal human oral keratinocyte. *Carcinogenesis*, 27(6) :1273–1284.
- Lufei, C., Ma, J., Huang, G., Zhang, T., Novotny-Diermayr, V., Ong, C. T., and Cao, X. (2003). GRIM-19, a death-regulatory gene product, suppresses Stat3 activity via functional interaction. *EMBO J*, 22(6) :1325–1335.
- Luo, X., Burwinkel, B., Tao, S., and Brenner, H. (2011). MicroRNA signatures : novel biomarker for colorectal cancer ? *Cancer Epidemiology Biomarkers & Prevention*, 20(7) :1272–1286.
- Luque-García, J. L., Martínez-Torrecuadrada, J. L., Epifano, C., Cañamero, M., Babel, I., and Casal, J. I. (2010). Differential protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. *Proteomics*, 10(5) :940–52.
- Lurje, G., Zhang, W., Yang, D., Groshen, S., Hendifar, A. E., Husain, H., Nagashima, F., Chang, H. M., Fazzone, W., and Ladner, R. D. (2008). Thymidylate synthase haplotype is associated with tumor recurrence in stage II and stage III colon cancer. *Pharmacogenetics and genomics*, 18(2) :161–168.

- M, U., ME, M., and JS., M. (1997). Difference gel electrophoresis : A single gel method for detecting changes in protein extracts. *Electrophoresis*, 18(11) :2071–2077.
- Mahoney, D. W., Therneau, T. M., Heppelmann, C. J., Higgins, L., Benson, L. M., Zenka, R. M., Jagtap, P., Nelsestuen, G. L., Bergen, H. R., and Oberg, A. L. (2011). Relative quantification : characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *Journal of proteome research*, 10(9) :4325–4333.
- Makridakis, M. and Vlahou, A. (2010). Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics*, 73(12) :2291–2305.
- Mamyryn, B. A. and Shmikk, D. V. (1979). The linear mass reflectron 6. 1505(October 1978) :1978–1980.
- Mann, M., Hendrickson, R. C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annual review of biochemistry*, 70(1) :437–473.
- Mantovani, A., Locati, M., Vecchi, A., Sozzani, S., and Allavena, P. (2001). Decoy receptors : a strategy to regulate inflammatory cytokines and chemokines. *Trends in immunology*, 22(6) :328–336.
- Marouga, R., David, S., and Hawkins, E. (2005). The development of the DIGE system : 2D fluorescence difference gel analysis technology. *Analytical and bioanalytical chemistry*, 382(3) :669–678.
- Martinou, I., Desagher, S., Eskes, R., Antonsson, B., André, E., Fakan, S., and Martinou, J. C. (1999). The release of cytochrome c from mitochondria during apoptosis of NGF-deprived sympathetic neurons is a reversible event. *The Journal of cell biology*, 144(5) :883–9.
- Marusyk, A., Almendro, V., and Polyak, K. (2012). Intra-tumour heterogeneity : a looking glass for cancer? *Nature Reviews Cancer*.
- McDonald, C. A., Yang, J. Y., Marathe, V., Yen, T. Y., and Macher, B. A. (2009). Combining results from lectin affinity chromatography and glyco-capture approaches substantially improves the coverage of the glycoproteome. *Molecular & Cellular Proteomics*, 8(2) :287–301.
- McNeil, P. L., Muthukrishnan, L., Warder, E., and D’Amore, P. A. (1989). Growth factors are released by mechanically wounded endothelial cells. *The Journal of cell biology*, 109(2) :811–822.
- Meany, D. L., Zhang, Z., Sokoll, L. J., Zhang, H., and Chan, D. W. (2009). Glycoproteomics for Prostate Cancer Detection : Changes in Serum PSA Glycosylation Patterns. *Journal of Proteome Research*, 8(2) :613–619.
- Medzihradszky, K. F., Campbell, J. M., Baldwin, M. a., Falick, a. M., Juhasz, P., Vestal, M. L., and Burlingame, a. L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical chemistry*, 72(3) :552–8.
- Michel, P. E., Crettaz, D., Morier, P., Heller, M., Gallot, D., Tissot, J.-D., Reymond, F., and Rossier, J. S. (2006). Proteome analysis of human plasma and amniotic fluid by Off-Gel™ isoelectric focusing followed by nano-LC-MS/MS. *Electrophoresis*, 27(5-6) :1169–1181.
- Miquel, C., Jacob, S., Grandjouan, S., Aime, A., Viguier, J., Sabourin, J. C., Sarasin, A., Duval, A., and Praz, F. (2007). Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. *Oncogene*, 26(40) :5919–5926.
- Mirgorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Körner, R., Sönksen, C. P., and Roepstorff, P. (2000). Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using 18O-labeled internal standards. *Rapid Communications in Mass Spectrometry*, 14(14) :1226–1232.
- Mishra, A. and Verma, M. (2010). Cancer Biomarkers : Are We Ready for the Prime Time? *Cancers*, 2(1) :190–208.
- Misselwitz, B., Becker, A., Bauerfeind, P., and Vavricka, S. R. (2011). Dépistage du cancer du côlon. 11(6) :103–107.
- Mitulovic, G. and Mechtler, K. (2006). HPLC techniques for proteomics analysis—a short overview of latest developments. *Briefings in functional genomics & proteomics*, 5(4) :249–260.
- Mukhopadhyay, A., Talukdar, S., Bhattacharjee, A., and Ray, K. (2004). Bioinformatic approaches for identification and characterization of olfactomedin related genes with a potential role in pathogenesis of ocular disorders. *Mol Vis*, 10 :304–314.
- Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A., and Hynes, R. O. (2012). The matrisome : in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & Cellular Proteomics*, 11(4).

- Nadiminty, N., Lou, W., Lee, S. O., Lin, X., Trump, D. L., and Gao, A. C. (2006). Stat3 activation of NF- $\kappa$ B p100 processing involves CBP/p300-mediated acetylation. *Proc Natl Acad Sci U S A*, 103(19) :7264–7269.
- Nanavati, D., Gucek, M., Milne, J. L., Subramaniam, S., and Markey, S. P. (2008). Stoichiometry and absolute quantification of proteins with mass spectrometry using fluorescent and isotope-labeled concatenated peptide standards. *Molecular & cellular proteomics*, 7(2) :442–447.
- Nandan, M. O. and Yang, V. W. (2011). An Update on the Biology of RAS/RAF Mutations in Colorectal Cancer. *Current colorectal cancer reports*, 7(2) :113–120.
- Nevins, J. R. and Potti, A. (2007). Mining gene expression profiles : expression signatures as cancer phenotypes. *Nat Rev Genet*, 8(8) :601–609.
- Niccoli, P., Wion-Barbot, N., Caron, P., Henry, J. F., de Micco, C., Saint Andre, J. P., Bigorgne, J. C., Modigliani, E., and Conte-Devolx, B. (1997). Interest of routine measurement of serum calcitonin : study in a large series of thyroidectomized patients. *Journal of Clinical Endocrinology & Metabolism*, 82(2) :338–341.
- Nickel, W. (2005). Unconventional Secretory Routes : Direct Protein Export Across the Plasma Membrane of Mammalian Cells. *Traffic*, 6(8) :607–614.
- Obeid, L. M., Linardic, C. M., Karolak, L. a., and Hannun, Y. a. (1993). Programmed cell death induced by ceramide. *Science (New York, N.Y.)*, 259(5102) :1769–71.
- Ochsenbein, A. F. (2002). Principles of tumor immunosurveillance and implications for immunotherapy. *Cancer gene therapy*, 9(12) :1043–55.
- O’Connell, J. B., Maggard, M. a., and Ko, C. Y. (2004). Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J Natl Cancer Inst*, 96(19) :1420–1425.
- O’Dwyer, D., Ralton, L. D., O’Shea, A., and Murray, G. I. (2011). The proteomics of colorectal cancer : identification of a protein signature associated with prognosis. *PloS one*, 6(11) :e27718.
- O’Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of biological chemistry*, 250(10) :4007–4021.
- Ogino, S. and Goel, A. (2008). Molecular Classification and Correlates in Colorectal Cancer. *The Journal of Molecular Diagnostics*, 10(1) :13–27.
- Ogino, S., Kawasaki, T., and Kirkner, G. J. (2006). CpG island methylator phenotype-low (CIMP-low) in colorectal cancer : possible associations with male sex and KRAS mutations. *The Journal of . . .*, 38(4) :511.
- Ogino, S., Nosho, K., Kirkner, G. J., Kawasaki, T., Meyerhardt, J. A., Loda, M., Giovannucci, E. L., and Fuchs, C. S. (2009). CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut*, 58(1) :90–96.
- Ohuchida, K., Mizumoto, K., Murakami, M., Qian, L.-W., Sato, N., Nagai, E., Matsumoto, K., Nakamura, T., and Tanaka, M. (2004). Radiation to stromal fibroblasts increases invasiveness of pancreatic cancer cells through tumor-stromal interactions. *Cancer research*, 64(9) :3215–22.
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5) :376–386.
- Ow, S. Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P. C. (2009). iTRAQ underestimation in simple and complex mixtures :“the good, the bad and the ugly”. *Journal of proteome research*, 8(11) :5347–5355.
- Ow, S. Y., Salim, M., Noirel, J., Evans, C., and Wright, P. (2011). Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *Proteomics*, 11(11) :2341–2346.
- Pan, S., Chen, R., Aebbersold, R., and Brentnall, T. A. (2011). Mass spectrometry based glycoproteomics—from a proteomics perspective. *Molecular & cellular proteomics : MCP*, 10(1) :R110.003251.
- Pan, S., Wang, Y., Quinn, J. F., Peskind, E. R., Waichunas, D., Wimberger, J. T., Jin, J., Li, J. G., Zhu, D., Pan, C., and Zhang, J. (2006). Identification of Glycoproteins in Human Cerebrospinal Fluid with a Complementary Proteomic Approach. *Journal of Proteome Research*, 5(10) :2769–2779.
- Parham, D. (1923). Colloid carcinoma. *Annals of Surgery*.
- Perlis, R. H. (2011). Translating biomarkers to clinical practice. *Mol Psychiatry*, Online pub :1–12.

- Phelps, R. A., Chidester, S., Dehghanizadeh, S., Phelps, J., Sandoval, I. T., Rai, K., Broadbent, T., Sarkar, S., Burt, R. W., and Jones, D. A. (2009). A two-step model for colon adenoma initiation and progression caused by APC loss. *Cell*, 137(4) :623–634.
- Pichler, P., KÄ¶cher, T., Holzmann, J., Mazanek, M., Taus, T., Ammerer, G., and Mechtler, K. (2010). Peptide labeling with isobaric tags yields higher identification rates using iTRAQ 4-plex compared to TMT 6-plex and iTRAQ 8-plex on LTQ Orbitrap. *Anal Chem*, 82(15) :6549–6558.
- Pinedo, H. and Peters, G. (1988). Fluorouracil : biochemistry and pharmacology. *Journal of clinical oncology*.
- Pocsfalvi, G., Votta, G., De Vincenzo, A., Fiume, I., Raj, D. A. A., Marra, G., Stoppelli, M. P., and Iaccarino, I. (2011). Analysis of secretome changes uncovers an autocrine/paracrine component in the modulation of cell proliferation and motility by c-Myc. *Journal of proteome research*, 10(12) :5326–5337.
- Popat, S., Hubner, R., and Houlston, R. S. (2005). Systematic review of microsatellite instability and colorectal cancer prognosis. *Journal of Clinical Oncology*, 23(3) :609–618.
- Raimondo, F., Morosi, L., Chinello, C., Magni, F., and Pitto, M. (2011). Advances in membranous vesicle and exosome proteomics improving biological understanding and biomarker discovery. *Proteomics*, 11(4) :709–720.
- Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome research*, 12(8) :1231–1245.
- Ren, G., Okerberg, C., and Mathews, S. (2012). Ultrasensitive Protein Detection and Imaging : Comparison of Lumitein, ProteoSilver, SYPRO Ruby, and Coomassie Brilliant Blue Gel Stains. In Kurien, B. T. and Scofield, R. H., editors, *Protein Electrophoresis*, volume 869 of *Methods in Molecular Biology*, pages 621–632. Humana Press.
- Reya, T. and Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nature*, 434(7035) :843–850.
- Reynolds, B. A. and Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science*, 255(5052) :1707–1710.
- Roepstorff, P. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*, 11(11) :601–606.
- Rogers, S., Girolami, M., Kolch, W., Waters, K. M., Liu, T., Thrall, B., and Wiley, H. S. (2008). Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. *Bioinformatics*, 24(24) :2894–2900.
- Ros, A., Faupel, M., Mees, H., van Oostrum, J., Ferrigno, R., Reymond, F., Michel, P., Rossier, J. S., and Girault, H. H. (2002). Protein purification by Off-Gel electrophoresis.
- Rosenzweig, D., Smith, D., Myler, P. J., Olafson, R. W., and Zilberstein, D. (2008). Post-translational modification of cellular proteins during *Leishmania donovani* differentiation. *Proteomics*, 8(9) :1843–1850.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004). Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and cellular proteomics*, 3(12) :1154–1169.
- Roth, J. (2002). Protein N -Glycosylation along the Secretory Pathway : Relationship to Organelle Topography and Function, Protein Quality Control, and Cell Interactions. *Chemical Reviews*, 102(2) :285–304.
- Rubartelli, A., Cozzolino, F., Talio, M., and Sitia, R. (1990). A novel secretory pathway for interleukin-1 beta, a protein lacking a signal sequence. *The EMBO Journal*, 9(5) :1503.
- Ruhaak, L. R., Miyamoto, S., and Lebrilla, C. B. (2013). Developments in the identification of glycan biomarkers for the detection of cancer. *Molecular & cellular proteomics : MCP*, pages 846–855.
- Russell, D., Oldham, N. J., and Davis, B. G. (2009). Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydrate research*, 344(12) :1508–1514.
- Salcedo, R., Young, H. a., Ponce, M. L., Ward, J. M., Kleinman, H. K., Murphy, W. J., and Oppenheim, J. J. (2001). Eotaxin (CCL11) induces in vivo angiogenic responses by human CCR3+ endothelial cells. *Journal of immunology (Baltimore, Md. : 1950)*, 166(12) :7571–8.
- Sarkar, D. (2004). Human Polynucleotide Phosphorylase (hPNPaseold-35) : A Potential Link between Aging and Inflammation. *Cancer Research*, 64(20) :7473–7478.

- Schachter, H. and Freeze, H. H. (2009). Glycosylation diseases : Quo vadis? *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1792(9) :925–930.
- Schenk, S., Schoenhals, G. J., de Souza, G., and Mann, M. (2008). A high confidence, manually validated human blood plasma protein reference set. *BMC medical genomics*, 1 :41.
- Schulze, W. and Usadel, B. (2010). Quantitation in mass-spectrometry-based proteomics. *Annual review of plant biology*, 80 :1–44.
- Schwacke, J. H., Hill, E. G., Krug, E. L., Comte-Walters, S., and Schey, K. L. (2009a). iQuantitator : a tool for protein expression inference using iTRAQ. *BMC Bioinformatics*, 10(Lc) :342.
- Schwacke, J. H., Hill, E. G., Krug, E. L., Comte-Walters, S., and Schey, K. L. (2009b). iQuantitator : a tool for protein expression inference using iTRAQ. *BMC Bioinformatics*, 10 :342.
- Sears, P. and Wong, C. H. (1998). Enzyme action in glycoprotein synthesis. *Cellular and molecular life sciences*, 54(3) :223–252.
- Seliger, B. (2005). Strategies of tumor immune evasion. *BioDrugs*.
- Shen, Y. and Smith, R. D. (2002). Proteomics based on high-efficiency capillary separations. *Electrophoresis*, 23(18) :3106–24.
- Shinya, H. and Wolff, W. I. (1979). Morphology, anatomic distribution and cancer potential of colonic polyps. *Annals of surgery*, 190(6) :679.
- Soreide, K., Janssen, E. A. M., SÄ¶iland, H., KÄ¶rner, H., and Baak, J. P. A. (2006). Microsatellite instability in colorectal cancer. *Br J Surg*, 93(4) :395–406.
- Sparbier, K., Wenzel, T., and Kostrzewa, M. (2006). Exploring the binding profiles of {ConA}, boronic acid and {WGA} by {MALDI-TOF/TOF} {MS} and magnetic particles. *Journal of Chromatography B*, 840(1) :29–36.
- Spiro, R. G. (2002). MINI REVIEW Protein glycosylation : nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. 12(4).
- Stastna, M. and Van Eyk, J. E. (2012). Secreted proteins as a fundamental source for biomarker discovery. *PROTEOMICS*, 12(4-5) :722–735.
- Stein, G. H., Drullinger, L. F., Robetorye, R. S., Pereira-Smith, O. M., and Smith, J. R. (1991). Senescent cells fail to express cdc2, cycA, and cycB in response to mitogen stimulation. *Proceedings of the National Academy of Sciences*, 88(24) :11012–11016.
- Stephens, W. E. (1946). Proceedings of the American Physical Society. *Phys. Rev.*, 69(11-12) :674.
- Strupat, K., Karas, M., and Hillenkamp, F. (1991). 2,5-Dihydroxybenzoic acid : a new matrix for laser desorption—ionization mass spectrometry. *International Journal of Mass Spectrometry and Ion Processes*, 111(0) :89–102.
- Sun, P., Yoshizuka, N., New, L., Moser, B. a., Li, Y., Liao, R., Xie, C., Chen, J., Deng, Q., Yamout, M., Dong, M.-Q., Frangou, C. G., Yates, J. R., Wright, P. E., and Han, J. (2007). PRAK is essential for ras-induced senescence and tumor suppression. *Cell*, 128(2) :295–308.
- Surinova, S., Schiess, R., Hüttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011). On the development of plasma protein biomarkers. *Journal of proteome research*, 10(1) :5–16.
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. (1988). Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(8) :151–153.
- Tejpar, S., Bertagnolli, M., Bosman, F., Lenz, H.-J., Garraway, L., Waldman, F., Warren, R., Bild, A., Collins-Brennan, D., Hahn, H., Harkin, D. P., Kennedy, R., Ilyas, M., Morreau, H., Proutski, V., Swanton, C., Tomlinson, I., Delorenzi, M., Fiocca, R., Van Cutsem, E., and Roth, A. (2010). Prognostic and Predictive Biomarkers in Resected Colon Cancer : Current Status and Future Perspectives for Integrating Genomics into Biomarker Discovery. *The Oncologist*, 15(4) :390–404.
- ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E., and Nusse, R. (2008). Wnt signaling mediates self-organization and axis formation in embryoid bodies. *Cell Stem Cell*, 3(5) :508–518.
- Tetsu, O. and McCormick, F. (1999). Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature*, 398(6726) :422–426.

- Thingholm, T. E., Palmisano, G., Kjeldsen, F., and Larsen, M. R. (2010). Undesirable charge-enhancement of isobaric tagged phosphopeptides leads to reduced identification efficiency. *Journal of proteome research*, 9(8) :4045–4052.
- Tjalsma, H., Bolhuis, A., Jongbloed, J. D. H., Bron, S., and van Dijk, J. M. (2000). Signal Peptide-Dependent Protein Transport in *Bacillus subtilis* : a Genome-Based Survey of the Secretome. *Microbiology and Molecular Biology Reviews*, 64(3) :515–547.
- Turchinovich, A., Weiz, L., Langheinz, A., and Burwinkel, B. (2011). Characterization of extracellular circulating microRNA. *Nucleic acids research*, 39(16) :7223–7233.
- Tweedle, E. M., Khattak, I., Ang, C. W., Nedjadi, T., Jenkins, R., Park, B. K., Kalirai, H., Dodson, A., Azadeh, B., and Terlizzo, M. (2010). Low molecular weight heat shock protein HSP27 is a prognostic indicator in rectal cancer but not colon cancer. *Gut*, 59(11) :1501–1510.
- Ueland, F. R., Desimone, C. P., Seamon, L. G., Miller, R. A., Goodrich, S., Podzielinski, I., Sokoll, L., Smith, A., van Nagell Jr, J. R., and Zhang, Z. (2011). Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstetrics & Gynecology*, 117(6) :1289.
- Ueno, H., Mochizuki, H., Hashiguchi, Y., Shimazaki, H., Aida, S., Hase, K., Matsukuma, S., Kanai, T., Kurihara, H., Ozawa, K., Yoshimura, K., and Bekku, S. (2004). Risk factors for an adverse outcome in early invasive colorectal carcinoma. *Gastroenterology*, 127(2) :385–394.
- Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., and Hamelin, R. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute*, 96(4) :261–268.
- Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., and Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, 9(6) :654–659.
- Van Cutsem, E., Nordlinger, B., and Cervantes, A. (2010). Advanced colorectal cancer : ESMO Clinical Practice Guidelines for treatment. *Annals of Oncology*, 21(suppl 5) :v93–v97.
- Van Cutsem, E. and Oliveira, J. (2009). Advanced colorectal cancer : ESMO clinical recommendations for diagnosis, treatment and follow-up. *Annals of Oncology*, 20(suppl 4) :iv61–iv63.
- van der Flier, L. G., Haegebarth, A., Stange, D. E., van de Wetering, M., and Clevers, H. (2009). OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology*, 137(1) :15–17.
- van't Veer, L. J. and Bernards, R. A. (2008). Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, 452(7187) :564–570.
- Varki, A. (1993). Biological roles of oligosaccharides : all of the theories are correct. *Glycobiology*, 3(2) :97–130.
- Vaughn, C. P., Crockett, D. K., Lim, M. S., and Elenitoba-Johnson, K. S. J. (2006). Analytical characteristics of cleavable isotope-coded affinity tag-LC-tandem mass spectrometry for quantitative proteomic studies. *The Journal of molecular diagnostics : JMD*, 8(4) :513.
- Vilar, E. and Gruber, S. B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. *Nat Rev Clin Oncol*, 7(3) :153–162.
- Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319(9) :525–532.
- Vogelstein, B., Lane, D., and Levine, A. J. (2000). Surfing the p53 network. *Nature*, 408(6810) :307.
- Voyksner, R. D. and Lee, H. (1999). Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry. *Rapid Communications in Mass Spectrometry*, 13(14) :1427–1437.
- Wanebo, H. J., Rao, B., Pinsky, C. M., Hoffman, R. G., Stearns, M., Schwartz, M. K., and Oettgen, H. F. (1978). Preoperative Carcinoembryonic Antigen Level as a Prognostic Indicator in Colorectal Cancer. *New England Journal of Medicine*, 299(9) :448–451.
- Wang, H., Alvarez, S., and Hicks, L. M. (2011). Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two *Chlamydomonas reinhardtii* strains of interest for biofuels engineering. *Journal of Proteome Research*, 11(1) :487–501.

- Washburn, M. P., Wolters, D., and Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature biotechnology*, 19(3) :242–247.
- Wegenka, U. M., Buschmann, J., Letticken, C., Heinrich, P. C., and Horn, F. (1993). Acute-phase response factor, a nuclear factor binding to acute-phase response elements, is rapidly activated by interleukin-6 at the posttranslational level. *Mol Cell Biol*, 13(1) :276–288.
- Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., Kang, G. H., Widschwendter, M., Weener, D., and Buchanan, D. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature genetics*, 38(7) :787–793.
- Wells, L. (2001). Glycosylation of Nucleocytoplasmic Proteins : Signal Transduction and O-GlcNAc. *Science*, 291(5512) :2376–2378.
- Whitehouse, C. M., Dreyer, R. N., Yamashita, M., and Fenn, J. B. (1985). Electrospray interface for liquid chromatographs and mass spectrometers. *Analytical Chemistry*, 57(3) :675–679.
- Wilkins, M. R., Pasquali, C., Appel, R. D., and Ou, K. (1996). From proteins to proteomes : Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nature . . .*
- Wisniewski, J. R. (2010a). Tools for phospho- and glycoproteomics of plasma membranes. *Amino Acids*, 41(2) :223–233.
- Wisniewski, J. R. (2010b). Tools for phospho- and glycoproteomics of plasma membranes. *Amino Acids*, 41(2) :223–233.
- Wisniewski, J. R., Ostasiewicz, P., and Mann, M. (2011a). High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *Journal of proteome research*, 10(7) :3040–3049.
- Wisniewski, J. R., Zielinska, D. F., and Mann, M. (2011b). Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Analytical biochemistry*, 410(2) :307–9.
- Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, 6(5) :359–362.
- Wu, C. C., Chen, H. C., Chen, S. J., Liu, H. P., Hsieh, Y. Y., Yu, C. J., Tang, R., Hsieh, L. L., Yu, J. S., and Chang, Y. S. (2008). Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics*, 8(2) :316–332.
- Wüthrich, K. and Shulman, R. G. (1969). Proton magnetic resonance studies of porphyrin iron (III) cyanides. *Proceedings of the . . .*, (Iii) :636–643.
- Xie, L. Q., Zhao, C., Cai, S. J., Xu, Y., Huang, L. Y., Bian, J. S., Shen, C. P., Lu, H. J., and Yang, P. Y. (2010). Novel Proteomic Strategy Reveal Combined  $\alpha$ 1 Antitrypsin and Cathepsin D as Biomarkers for Colorectal Cancer Early Screening. *Journal of Proteome Research*, 9(9) :4701–4709.
- Xu, B. J., Yan, W., Jovanovic, B., Shaw, A. K., An, Q. A., Eng, J., Chytil, A., Link, A. J., and Moses, H. L. (2011). Microdialysis combined with proteomics for protein identification in breast tumor microenvironment in vivo. *Cancer Microenvironment*, 4(1) :61–71.
- Xue, H., Lu, B., Zhang, J., Wu, M., Huang, Q., Wu, Q., Sheng, H., Wu, D., Hu, J., and Lai, M. (2010). Identification of Serum Biomarkers for Colorectal Cancer Metastasis Using a Differential Secretome Approach. *Journal of Proteome Research*, 9(1) :545–555.
- Yang, H. Y., Kwon, J., Park, H. R., Kwon, S. O., Park, Y. K., Kim, H. S., Chung, Y. J., Chang, Y. J., Choi, H. I., and Chung, K. J. (2012). Comparative proteomic analysis for the insoluble fractions of colorectal cancer patients. *Journal of proteomics*.
- Yang, Z. (2006). Multilectin Affinity Chromatography for Characterization of Multiple Glycoprotein Biomarker Candidates in Serum from Breast Cancer Patients. *Clinical Chemistry*, 52(10) :1897–1905.
- Yeatman, T. J. (2004). A renaissance for SRC. *Nat Rev Cancer*, 4(6) :470–480.
- Yu, H., Kortylewski, M., and Pardoll, D. (2007). Crosstalk between cancer and immune cells : role of STAT3 in the tumour microenvironment. *Nat Rev Immunol*, 7(1) :41–51.
- Yu, Y., Anjum, R., Kubota, K., Rush, J., Villen, J., and Gygi, S. P. (2009). A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. *Proc Natl Acad Sci U S A*, 106(28) :11606–11611.

- Zeng, L.-C., Han, Z.-G., and Ma, W.-J. (2005). Elucidation of subfamily segregation and intramolecular coevolution of the olfactomedin-like proteins by comprehensive phylogenetic analysis and gene expression pattern assessment. *FEBS Lett*, 579(25) :5443–5453.
- Zhang, H., Li, X.-J., Martin, D. B., Aebersold, R., and Others (2003a). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature biotechnology*, 21(6) :660–666.
- Zhang, J., Liu, W. L., Tang, D. C., Chen, L., Wang, M., Pack, S. D., Zhuang, Z., and Rodgers, G. P. (2002a). Identification and characterization of a novel member of olfactomedin-related protein family, hGC-1, expressed during myeloid lineage development. *Gene*, 283(1-2) :83–93.
- Zhang, J., Lou, X., Yang, S., He, S., Yang, L., Liu, M., Zhu, H., Shan, Q., Su, S., and Zhan, Q. (2011). BAG2 is a target of the c-Myc gene and is involved in cellular senescence via the p21<sup>></sup> CIP1<sup><</sup> pathway. *Cancer Letters*.
- Zhang, J., Yang, J., Roy, S. K., Tininini, S., Hu, J., Bromberg, J. F., Poli, V., Stark, G. R., and Kalvakolanu, D. V. (2003b). The cell death regulator GRIM-19 is an inhibitor of signal transducer and activator of transcription 3. *Proc Natl Acad Sci U S A*, 100(16) :9342–9347.
- Zhang, R., Sioma, C. S., Thompson, R. a., Xiong, L., and Regnier, F. E. (2002b). Controlling deuterium isotope effects in comparative proteomics. *Analytical chemistry*, 74(15) :3662–9.
- Zhang, T., Otevrel, T., Gao, Z., Ehrlich, S. M., Fields, J. Z., and Boman, B. M. (2001). Evidence that APC regulates survivin expression : a possible mechanism contributing to the stem cell origin of colon cancer. *Cancer Res*, 61(24) :8664–8667.
- Zhang, X., Huang, Q., Yang, Z., Li, Y., and Li, C.-Y. (2004). GW112, a novel antiapoptotic protein that promotes tumor growth. *Cancer Res*, 64(7) :2474–2481.
- Zhu, W., Smith, J. W., and Huang, C.-M. (2010). Mass spectrometry-based label-free quantitative proteomics. *Journal of biomedicine & biotechnology*, 2010 :840518.
- Zielinska, D. F., Gnad, F., Schropp, K., Wisniewski, J. R., and Mann, M. (2012). Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Molecular Cell*, 46(4) :542–548.
- Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*, 141(5) :897–907.
- Zybailov, B., Coleman, M. K., Florens, L., and Washburn, M. P. (2005). Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Analytical chemistry*, 77(19) :6218–6224.

# Annexes



## Annexe n ° 1

Tableau 1: Tableau des protéines dont l'expression varie significativement (selon l'analyse par iQuantitator) en fonction des stades tumoraux – Les valeurs égales à 1 correspondent à une variation de la protéine qui est non significative par rapport au tissu sain.

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
ABHD10	1.464	1.691	1	1	1
ABHD14B	1	1	0.656	1	1
ABP1	1	1	0.64	1	1
ACAD9	1	1	1.51	1	1
ACADS	1	1	0.561	1	1
ACADVL	1	1.49	1	1	1
ACAT2	1	0.673	1	1	1
ACLY	1	1	1	1.398	1
ACO1	1.328	1	0.739	1	1
ACO2	1.459	1.364	1	1	1
ACOT9	1	1	1.8	1	1
ACSL5	1	1	0.681	0.597	1
ACTA1	1	1	1	1.735	1
ACTA2	0.425	1	1.28	1	1
ACTB	0.83	1	1	1.341	1
ACTBL2	0.835	1	0.738	1	1
ACTC1	1	1	1.96	1	0.686
ACTG1	1	1	0.87	1	1
ACTG2	0.737	1	2.02	0.626	1
ACTN1	1	1	1.68	1.324	1
ACTR2	1	1	1	1.385	1
ADH1B	0.622	1	0.447	0.545	1
ADH1C	1	0.28	0.368	0.318	1
AGR2	0.793	0.706	0.358	0.592	1
AHCY	1	1.294	1.42	1.506	1
AHNAK	0.774	1	0.734	1	1
AIF1	1	0.431	0.559	1	1
AIFM1	1	1	1.65	1	1
AKR1B1	1	1	0.614	1	1
AKR1B10	1	0.657	0.331	0.516	1
AKR7A3	1	1	0.507	1	1
ALDH18A1	1	1	1.38	1	1.4
ALDH1A1	1.344	1.523	2.07	0.639	1
ALDOA	1	0.7	0.69	1.499	1.4
ANXA1	1	1	1	1.401	1
ANXA2	1	1	0.79	1.183	1
ANXA2P2	1	1	0.784	1	1
ANXA3	1	1.719	1	1.53	1
ANXA4	1.298	1	1	1	1
ANXA5	1	1	1.48	1	0.772
ANXA6	0.724	1	1.39	1	1
AOC3	0.473	0.571	1	0.579	0.461
APCS	0.62	1	1	1	1
APEX1	1.448	1	1	1	1
APOA1	1	1	1	1.495	0.7
APOB	1	1	1.27	1	1
APOE	1	1	2.13	1	1
ASAHI	1	1	1	1	1
ASAH1	0.756	0.754	0.623	1	1
ASL	1	1	0.681	1	1
ASPN	1	1	1.93	1	1
ATAD3A	1	1.39	1.36	1	1
ATIC	1	1.28	1	1	1
ATP1A1	1	1	0.734	1	1
ATP5A1	1.485	1	1	1	1
ATP5B	1.18	1	0.832	1	1
ATP5H	1.438	1	1	1	1
BAT1	1	1	0.683	1	1
BCAR1	1	1	1	1.593	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
BCAT2	1	1	1.53	1	1
BCLAF1	1	1	1	1.552	1
BDH1	1.575	1	1	1	1
BGN	0.505	1	1.79	1	0.61
BPHL	1	1.677	1	1	1
BXDC1	1	1	1.7	1	1
C11	1	0.653	1	1	1
C19orf21	1	1	0.616	1	1
C1QBP	1.501	1	1	1	1
C21	1.385	1	1	1	1
C3	0.748	1	1	1	1
C4B	1	1	1.38	1	1
C4BPA	1	1.554	1	1	1
CA1	1	1	0.611	1	1
CA12	1	1	0.489	0.618	1
CA2	0.726	0.328	0.301	0.46	1
CAD	1	1	1	1.445	1
CADPS	2.028	1	1	1	1
CALD1	1	1	1.64	1	0.555
CALM1	1	1	0.703	1	1
CALR	1	0.79	0.572	1.541	1
CALU	1	1	1.54	1.445	1
CANX	1	1	1	1.482	1
CAP1	0.708	1	0.631	1	1
CAPG	1	1.553	1	1	1
CAT	1.308	1	0.686	1	1
CCT2	1	1	1	1.244	1
CCT3	1	1	1	1.348	1
CDC2	1	1	1	2.088	1
CDC42	1	1	1	1.521	1
CDH17	1	1.248	0.595	1	1
CEACAM3	1	1	2.57	1	1
CEACAM5	1	2.954	1.75	2.901	1
CES2	0.747	0.614	0.474	0.546	1
CFL1	1	1	0.761	1	1
CHD4	1	1	1	2.031	1
CHDH	1.662	1	1	1	1
CHDH	1	1.41	1	1	1
CHGA	0.34	1	1	1	1
CKB	1.622	0.691	0.673	1	1.99
CKMT1A	1.458	1.56	0.712	1	1.42
CLC	1	0.46	0.363	1	1
CLCA1	1.352	0.807	0.29	0.51	1
CLTC	1	1	0.869	1.207	1
CMPK1	1.387	1.264	1	1	1
CNDP2	1.708	1	0.746	1	1
CNN1	0.549	0.663	2.32	1	0.604
CNN2	1	1	1	1.645	1
COL11A2	1	1	1	0.342	1
COL12A1	1.59	1.419	1.9	2.08	1
COL14A1	0.355	0.669	1	0.564	0.499
COL18A1	1	1	1.8	1	1
COL1A1	0.289	0.604	1.27	0.575	0.485
COL1A2	0.308	0.621	1.31	0.581	0.532
COL2A1	0.413	1	1	0.559	1
COL3A1	0.237	0.641	1	0.388	1
COL4A2	1	0.743	1.31	1	1
COL5A1	0.35	1	1	1	1
COL5A2	0.429	0.681	1.32	1	1
COL6A1	1	0.769	1	1	0.684
COL6A2	0.692	0.706	1	1	0.592
COL6A3	0.786	0.739	1.18	1	0.45
COL8A1	1	1	1.87	1	1
COMT	1.635	1	1	1	1.65
COX4I1	1.784	1	1	1	1
COX5A	1.774	0.683	0.619	1	1
COX6A1	1.349	1	1	1	1
COX6B1	2.593	0.42	0.74	1	1
CP	0.724	1	1	1	1
CPNE1	1.395	1.619	2.07	1.393	1
CPS1	1	2.167	1	1	1
CPT1A	1	1	0.587	1	1
CS	1	1.319	1	1	1
CSE1L	1	1	1.59	1	1
CSTF2T	0.616	1	1	1	1
CTNNA1	1	1.444	1	1	1
CTSB	1.45	1	1	1	1
CTSC	1	0.502	0.384	1	1
CTSD	1	1	0.658	1	1
CTSG	1	1	1	1.708	0.628
CTTN	1	1	1	1.755	1
CYB5B	1	1	1.69	1	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
CYC1	1.864	1	1	1	1
DBNL	1	1	1	1.426	1
DCN	0.613	0.561	1	0.554	0.574
DDT	1	1	0.647	1	1
DEFA3	1	1	1.48	1	1
DENND4C	1.707	1	1	1	1
DES	0.336	0.477	1.79	0.538	0.391
DHX9	1	1	1.28	1.264	1
DMBT1	1	4.238	1	1	1
DPEP1	1	1	1.59	1.824	1
DPP7	1	0.631	1	1	1
DPY30	1	1	1.66	1	1
DSG2	1	1.465	1	1	1
DSP	1	1.34	1	1	1
DYNC1H1	1.275	1	1	1	1
ECHS1	1.366	1	0.728	1	1
EEF1A1	1.296	1	1	1	1
EEF1A3	1	1	0.705	1	1
EEF1D	1	1	1	1.519	1
EEF1G	1	0.755	1	1.46	1
EEF2	0.683	1	0.761	1	1
EFEMP1	1	1.453	2.81	1	1
EHD2	1	0.769	1	1	1
EHHADH	1	1	1	0.529	1
EIF4A1	1	1	0.804	1	1
EIF5A	1	0.723	0.737	1	1
ELN	0.365	1	1.68	0.657	1
EMILIN1	1	1	1.66	1	1
ENO1	1	1	1	1.338	1
EPHX2	1	1	0.464	0.612	1
EPPK1	1	1	0.725	1	1
EPRS	1.406	1	1	1	1
EPX	1.358	0.498	0.449	1	1
ERAP1	1	0.757	1	1	1
ERAP2	1	1	1	1.528	1
ERP29	1	1	0.737	1	1
ERP44	1.595	1	1	1	1
ETFA	1	1.472	0.656	1	1
ETFB	1.597	1	1	1	1
ETFDH	1	1	0.598	1	1
ETHE1	1.349	0.563	0.613	1	1
EZR	1	1	0.707	1	1
F13A1	0.422	1	1	0.625	1
F2	1.383	1	1	1	1
FABP1	1	0.633	0.258	0.465	1
FABP5	1	1	0.65	1	1
FAM129A	0.681	1	1	1	1
FASN	1	1	1	1	1.41
FBL	1.602	1	1	1	1
FBLN1	1	1.358	2.61	1	1
FBLN5	1	1	1.71	1	1
FBN1	1.54	0.796	2.09	1	1.43
FCGBP	1.413	0.319	0.31	0.451	2.08
FDFT1	1	1	1	0.593	1
FERMT2	1	1	1.57	1	1
FGA	0.67	1	1.23	1.378	0.725
FGB	0.618	1	1.37	1.47	1
FGG	1	1	1	1.428	1
FHL1	1	0.511	1	1	1
FHL2	1	1	1	1.55	1
FKBP1A	1	1	1	1.629	1
FLNA	0.696	1	1.8	1	0.569
FLNB	1	0.822	0.635	1	1
FLNC	0.648	1	1.57	1	0.59
FN1	0.806	1	2.25	1.499	1
FTL	1	1	1.97	1.359	1
FUCA1	0.574	0.28	0.199	0.254	0.535
GALNT7	1	1	0.627	1	1
GAPDH	1.205	1	0.74	1.298	1.49
GBAS	1	1.585	1	1	1
GCS1	1	1	1	1	1.64
GFPT1	1	1	0.614	1	1
GMDS	1.373	1	0.732	1	1
GNA11	1	1	1	0.594	1
GOT1	1	1	0.612	1	1
GOT2	1.312	1	1	1	1
GPX2	1	1.459	1	1	1
GPX3	1	1	1.72	1	1
GSTP1	1	1.291	0.65	1.403	1
GTF2I	1	1	1.56	1	1
H2AFX	1	0.726	1	0.531	1
H2AFY	1.273	1	1	1	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
H3F3A	1	1	0.514	1	1
H3F3C	1	0.576	1	1	1
HADH	1	1	0.677	1	1
HADHA	1	1	0.827	1	1
HBA1	1	0.663	0.821	1	1
HDLBP	1	1	1	1	1.67
HIBADH	1.4	1	1.45	1	1
HIST1H1C	1	1	1	1.391	1
HIST1H1D	1	1	0.596	1	1
HIST1H2AA	1	1	1	0.551	1
HIST1H2BA	1	1	1	2.673	1
HIST1H2BL	1	1	0.746	1	1
HIST1H3A	1.498	1	1.7	1	1
HLA	1	0.527	1	1	1
HLA	1	0.543	1	1	1
HLA-A	1	1	0.576	1	1
HLA-C	1	1	1	0.589	1
HLA-C	1	1	1	2.49	1
HLA-DRA	1	1	1	1.605	1
HLA-DRB1	1	1	1	0.321	1
HLA-DRB1	1	1	1	2.633	1
HLA-DRB1	1	1	1	6.784	1
HMGA2	1	1	1	3.339	1
HMGB1	1	1.395	1.68	1	2.08
HMGCS2	1	1	0.66	0.488	1
HNRNPA1	1	1	1.2	1	1
HNRNPA2B1	0.804	1	1.19	1	1
HNRNPA3	1	1	1.32	1	1
HNRNPU	1	1	1.4	1	1
HNRNPUL1	1	1	1.86	1	1
HNRNPUL2	1	1	1.44	1	1
HP	1	0.507	1	1	1
HPX	0.76	1	1	1	1
HSD11B2	1	0.574	0.594	0.533	1
HSD17B10	1	1.259	1	1	1
HSD17B10	1	1	1.77	1	1
HSD17B11	1	1	0.685	1	1
HSP90AB3P	1	1	0.483	1	1
HSP90B1	1	0.757	0.648	1	1
HSPA2	1	1.656	1	1	1
HSPA5	0.777	0.845	0.598	1	1
HSPB1	1	1	1.63	1	1
HSPD1	0.721	1.286	1.32	1.232	1
HSPE1	1.511	1.341	1.5	1	1
HSPG2	0.659	0.65	1.29	1	0.559
HSPH1	1	1	1.35	1	1
HYOU1	1	0.685	1	1	1
IARS2	1.471	1	1	1	1
IDH2	1.249	1.307	1	1	1
IGHA1	1	1.623	0.571	1	1
IGHA2	1	1	0.436	0.47	1
IGHG1	1	1	1	1.516	1
IGHG3	1	1.699	1	1	1
IGHG4	1	1	1	1.544	1
IGHM	1	1	0.728	1	1
IGKC	0.704	1	1	1	1
ILF3	1.503	1	1	1	1
ILK	1	1	1.63	1	1
INTS5	1	1	1	0.582	1
IPO5	1	1	1.58	1	1
IQGAP1	1	1	0.855	1	1
IRS2	1	1	2.08	1	1
ITLN1	1	0.552	0.431	1	1
KIAA0564	1.335	1	1	1	1
KLC4	1	1	1	0.479	1
KRT18	1.309	1.202	1	1	1.62
KRT19	1	0.679	0.654	1	1.27
KRT20	1	0.567	0.373	0.387	1
KRT72	0.619	1	1	1	1
KRT8	1.268	1	0.634	1	1.71
LAMA4	1	1	1	1	0.672
LAP3	1	1	1	1.695	1
LASP1	1	1	1	1.461	1
LCN2	1	1.683	1	2.078	1.9
LCP1	1	1	0.721	1	0.678
LDHA	1	1	1	1.831	1
LDHB	1	1	1	1.547	1
LGALS1	1	0.712	1.43	1.391	1
LGALS3	0.804	0.755	0.49	0.588	1
LGALS3BP	1	1	1	1.477	1
LGALS4	0.749	0.701	0.423	0.399	1
LGALS9	1	1	0.722	1	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
LMAN2	1	0.632	1	1	1
LOH11CR2A	1	1	0.685	1	1
LONP1	1	1.39	1	1	1
LPP	1	1	1	1	1
LPP	0.611	1	1.4	1	1
LRRC59	1	1	0.717	1	1
LTF	1.403	1.727	1	3.388	1
LUM	0.559	0.558	1	0.638	0.391
LYPLA1	1	1.57	1.49	1	1
LYZ	1	1	1	1.513	1
MAOA	1	0.747	0.64	0.497	1
MCAM	0.669	1	1	1	1
MCM6	1	1	1	1.67	1
MDH1	1	1	1	1	0.713
MDH2	1.581	1	1.27	1	1
MIF	1.418	1	1	1	1
MPO	1.514	1.669	1	2.199	1
MRPL19	1	1	1.41	1	1
MRPL44	1.442	1	1	1	1
MSN	1	1	1	1	0.732
MT-CO2	1.3	1	0.677	1	1
MTTP	1	1	1	0.647	1
MUC2	1.246	0.381	0.293	1	1
MVP	1	0.659	0.616	1	1
MX1	1	1	1.79	1	1
MYBBP1A	1	1	1.38	1.339	1
MYH10	1	1	1.52	1	1
MYH11	0.339	0.373	1.53	0.669	0.354
MYH14	1	1	0.814	1	1.44
MYL9	1	1	1.4	1	0.592
MYO1A	1	1	0.569	1	1
MYO1D	1	1	0.724	1	1
MYO7A	1.702	1	1	1	1
NACA	1	1	0.674	1	1
NAMPT	1	1	1	1.519	1
NANS	1	1	0.557	1	1
NARS	1.457	1	1	1	1
NDRG1	1	1	0.577	1	1
NDUFA13	1	0.649	1	1	1
NDUFB8	1	0.693	1	1	1
NIPSNAP1	1	1.358	1	1	1
NIPSNAP3A	1	1.649	1	1	1
NMES1	1	0.54	0.366	0.542	1
NOL5	1	1	1.5	1	1
NOL5A	1	1	1.46	1	1
NONO	1	1	1.39	1.255	1
NP	1	1	0.756	1	1
NPM1	1.556	1.292	1.47	1.705	1
NQO1	1	1.52	1	1	1
NSUN2	1	1	1	1.649	1
NUP188	1	0.631	1	1	1
NUP93	1	1	1.54	1	1
OAS3	1	1	1	1.37	1
OGN	0.469	0.412	1	0.415	0.566
OLFM4	1.227	1.731	2.19	1	1
OPA1	1	1	1.36	1	1
P4HB	1	0.743	0.662	1	1
PABPC4	1	1	1	1.374	1
PACAP	1	0.739	0.564	1	0.693
PADI2	1	0.648	0.437	0.564	1
PAFAH1B3	1	1	1.8	1	1
PAICS	1	1	1	1	1.51
PAPSS2	1	0.524	0.459	0.669	1
PARP1	1	1	1.47	1	1
PARP14	1	1	1	1.541	1
PDCD4	1	1	0.601	1	1
PDCD6IP	1	1	0.73	1	1
PDIA3	1	0.749	0.82	1	0.822
PDIA4	1	0.774	0.686	1.352	1
PDIA6	1	1	0.793	1	1
PDLIM1	1	0.709	1	1	1
PDLIM4	1	1	1.7	1	1
PDXDC1	1	1	0.648	1	1
PEA15	1	1	1.5	1	1
PEBP1	1	1	0.782	1	1
PEX1	0.708	1	1	1	1
PFN1	1	1	0.68	1	1
PGK1	1.268	1	1.25	1	1.46
PGM5	1	1	1.75	0.637	1
PGRMC1	1	0.723	1.43	1.625	1
PHB2	1	1.252	1	1	1
PHGDH	1	1	1.52	1	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
PIGR	1.232	1.231	0.431	1	1
PKM2	1	1	1	1.31	1.65
PLD3	1	0.574	0.654	0.604	1
PLEC1	1	1	1	1	1.19
PLG	0.579	1	1.52	1	1
PLOD3	1	1	1.47	1	1
PLS1	1	0.723	0.469	0.566	1
POSTN	0.621	0.521	1.64	0.56	0.467
PPA1	1	1.553	1	1	1
PPA2	1	1.726	1	1	1
PPIA	1.235	1	0.813	1	1.23
PPIB	1	1	0.74	1	1
PPP2CA	1	1	0.738	1	1
PRDX1	1.653	1	1	1.28	1
PRDX3	1	1.534	1	1	1
PRDX5	1.224	1	1.33	0.631	1
PRDX6	1	0.764	1	1	0.704
PRELP	1	1	2.15	0.642	1
PRG2	1	1	0.471	1	1
PRKDC	1	1	1.4	1.21	1
PRKG1	1	1	1.54	1	1
PRSS12	0.493	1	1	1	1
PRTN3	1	2.27	1.42	2.681	1
PSME1	1	0.779	0.771	1	1
PSME2	1	1	0.748	1	1
PTGES3	1	0.709	1	1.441	1
PYGB	1	1.428	0.652	1	1
PYGM	1	1.637	1	1	1
RAB31	1	1	1.49	1	1
RAN	1	1	1	1.655	1
RBM47	1	1	0.524	1	1
RCC2	1	1	1.49	1.408	1
RCN1	1	1	1	1.507	1
RCN3	1	1	1	1.489	1
REG4	1	1	0.473	0.663	1
RHOA	1	1	1	2.097	1
RNASE2	1	0.363	0.393	0.472	1
RNASE3	1	0.526	0.491	1	1
RNF213	1	1	1	1.528	1
RNH1	1	1	1.34	1	1
RNPEP	1.338	1	1	1	1
RPL13A	0.666	1	1	1	1
RPL38	1	0.584	1	1	1
RPL5	1	1	1	1.402	1
RPL6	1	1	1	1.311	1
RPL8	1.886	1	1	1	1
RPN2	1	1	0.795	1	1
RPS14	1	1	1	1.417	1
RPS15	1	1	0.438	1	1
RPS27A	1	1	1	1.382	1
RPS3A	1	0.696	1	1	1
RPS4X	1.59	0.78	1	1	1
RPS4Y2	1	1	1.5	1	1
RPS7	1	1	0.638	1	1
RRBP1	1	1	0.687	1	1
RSL1D1	1	1	1.46	1	1
S100A10	1	1	0.566	1	1
S100A14	1	0.662	1	1	1
S100A6	1.312	1	1	1.559	1
S100A8	1.49	1.568	1	3.693	1
S100A9	1.508	1.855	1	3.625	1
SAA1	1	1	2.98	1	1
SAFB	1	1	1	1.651	1
SAMM50	1	1.568	1	1	1
SARS	1	0.755	1	1	1
SCP2	1	1	0.646	1	1
SDHA	1.222	1.345	1	1	1
SEC23B	1	1	0.753	1	1
SELENBP1	1.393	1	0.361	0.534	1.43
SERBP1	1	1	1	1.728	1
SERPINA1	0.556	0.832	1	1	0.646
SERPINB6	1	1	0.633	1	1
SERPINH1	1	1	1.79	1.549	1
SET	1	1	1	1.405	1
SF3B2	1	1	1	1.319	1
SF3B3	1	1	1.39	1	1
SFPQ	1	1	1	1.251	1
SFRS3	1	1	1	1.393	1
SHMT2	1	1	1	1.442	1
SKIV2L2	1	1	1	1.407	1
SLC12A2	1.443	1.623	1	1	1
SLC1A5	0.678	1	1	1	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
SLC25A24	1	1.47	1	1	1
SLC25A3	1	1.249	1	1	1
SLC25A5	1	1	1.43	1	1
SLC2A1	1	1	1.77	1.477	1
SOD1	1	1	1	1.534	1
SOD2	1.304	1.319	1	1.568	1
SON	0.67	1	1	1	1
SORD	1.537	1.658	1.52	1	1
SPTB	1	1	1.46	1	1
SQRDL	1	1	0.518	0.66	1
SRI	1	1.382	1	1	1
SSB	1	1	1.45	1.504	1
ST6GALNAC1	1	1	0.318	0.596	1
STAT1	1	1	1	1.669	1
SUCLA2	1	1.46	1	1	1
SUCLG2	1	1	0.7	1	1
SULT1B1	1	1	0.409	1	1
TAGLN	0.51	0.772	2.32	1	0.482
TAGLN2	0.823	1	1	1	1
TAP1	1	1	1	1.498	1
TARDBP	1	1	1.43	1	1
TF	1	1	1	1	1.26
TFF3	1.55	0.442	0.306	1	1
TGFB1I1	1	1	1.59	1	0.699
TGFBI	1	1.59	2.19	1.295	1
TGM2	1	1	1.37	1	1
THBS1	1	1	1	1.388	1
THY1	1	1	1.94	1	1
TINAGL1	1	1	1	1	0.68
TKT	1	1	0.765	1	1
TLN1	1	0.808	1.37	1	1
TLN2	1	1.736	1	1	1
TMPO	1	1	0.71	1.359	1
TMPRSS1D	0.333	1	1	1	1
TMPRSS13	2.024	1	1	1.705	1
TNC	1	1	0.24	1.701	1
TOMM40	1	1	1.43	1	1
TPI1	1	1	1	1	1.38
TPM1	1	0.672	1	1	0.565
TPM2	1	0.622	1	1	0.567
TPM4	1	1	1.8	1	1
TPP1	0.659	1	1	1	1
TPSAB1	1	1	0.388	0.548	1
TRAP1	1	1	1	1.399	1
TRIM28	1	1	1.42	1	1
TST	1	1	0.765	1	1
TTN	1	0.286	1	1	1
TUBA1A	1.272	1	1.49	1	1
TUBA1B	1	1	1	1.528	1
TUBA1C	1	1	1	1.402	1
TUBA4A	1	1	1	1	1.47
TUBB	1	1.251	1	1.363	1
TUBB3	1	1	1.91	1	1
TUBB6	1	1	1.61	1	1
TUFM	1.363	1	1	1	1
TXNDC17	1.53	1	1	1	1
TXNDC4	1	0.711	1	1	1
TXNDC5	1	0.707	0.62	1	0.601
TYMP	1	0.667	1	1.453	1
UBA1	1	1	1.33	1.192	1
UBE2L3	1	1	0.469	1	1
UGDH	1	0.572	0.481	1	1
UGP2	1	0.76	0.619	1	1
UGT2B15	1	1	0.321	1	1
UGT2B17	1	0.26	0.322	0.631	1
UQCRB	1.613	1	1	1	1
UQCRC1	1.541	1	1	1	1
UQCRC2	1.646	1	1	1	1
VARS	1	1	1.27	1.271	1
VASP	1	1	0.679	1	1
VCAN	1	1	2.29	1.418	1
VCL	0.788	1	1.22	1	0.743
VDAC1	1.45	1.663	1	1	1
VIL1	1	0.683	0.611	1	1
VILL	1	1	0.59	1	1
VIM	0.669	0.861	1	1.22	0.567
VTN	0.741	0.643	1.28	1	1
WARS	1.61	1	1	2.089	1
XRCC5	1	1	1.33	1.291	1
XRCC6	1	1	1	1.328	1
XTP3TPA	1	1	1	1.444	1
YBX1	1	1	1	1.716	1

Suite page suivante ...

Nom du gène	Adenome	stade I	stade II	stade III	stade IV
... suite de la page précédente					
YWHAH	1	1	1	1.344	1
YWHAZ	1	1	1	1.252	1
ZG16	0.74	0.361	0.284	0.46	1
ZYX	1	1	1.54	1.496	1
					Fin

## Annexe n ° 2

Tableau 2: Tableau des protéines dont l'expression varie significativement (selon l'analyse par iQuantitator) chez les tumeur présentant une mutation de l'oncogène KRAS

Nom du gène	Ratio iTRAQ	Intervalle de confiance	Peptides
EFEMP1	2,08	(1.61, 2.72)	3(16)
ALDH1B1	1,9	(1.59, 2.28)	12
PHGDH	1,82	(1.46, 2.28)	4(48)
TGFBI	1,82	(1.59, 2.08)	17
TNC	1,78	(1.53, 2.07)	14
CPNE1	1,75	(1.35, 2.27)	5
IRS2	1,74	(1.19, 2.53)	3
CEACAM5	1,69	(1.46, 1.95)	10
OLFM4	1,68	(1.35, 2.10)	3
COL12A1	1,67	(1.40, 1.98)	19
CEACAM3	1,63	(1.21, 2.19)	2
BXDC1	1,63	(1.24, 2.15)	3
GTF2I	1,59	(1.17, 2.15)	2
SORD	1,59	(1.26, 2.00)	4
FN1	1,57	(1.42, 1.73)	46
TPM4	1,53	(1.1, 2.14)	2
TPM2	1,53	(1.1, 2.14)	2
SAA1	1,52	(1.21, 1.89)	4
SERPINH1	1,52	(1.28, 1.81)	9
SLC25A4	1,51	(1.05, 2.17)	2
CALD1	1,51	(1.14, 1.99)	4
HS1	1,51	(1.32, 1.74)	7
FBLN1	1,5	(1.20, 1.87)	5
ACTC1	1,5	(1.21, 1.85)	8
ACAD9	1,48	(1.01, 2.19)	2
PAFAH1B3	1,48	(1.16, 1.89)	2
GBAS	1,47	(1.12, 1.95)	4
BGN	1,47	(1.18, 1.84)	4
ACOT9	1,47	(1.12, 1.94)	4
HMGB1	1,46	(1.10, 1.95)	7
IPO5	1,46	(1.19, 1.8)	8
CSE1L	1,44	(1.23, 1.70)	9
ANXA3	1,43	(1.09, 1.88)	3
HNRNPA1L	1,43	(1.12, 1.84)	5
ACTG2	1,43	(1.20, 1.69)	12
HSD17B10	1,43	(1.24, 1.66)	7
SERPINA3	1,42	(0.981, 2.06)	1(3)
CHDH	1,42	(1.03, 1.95)	3
PPA2	1,42	(1.03, 1.98)	2
LYPLA1	1,42	(1.04, 1.96)	3
THY1	1,42	(1.08, 1.86)	2
AHCY	1,41	(1.15, 1.73)	
TAGLN	1,41	(1.23, 1.61)	16
TPT1	1,4	(1, 1.98)	2
APOE	1,4	(1.03, 1.90)	2
PRTN3	1,4	(1.09, 1.79)	2
CNN1	1,4	(1.09, 1.8)	4
RSL1D1	1,4	(1.09, 1.81)	3
CTPS	1,39	(0.98, 1.97)	2
SLC2A1	1,39	(1.06, 1.81)	2
PPA1	1,39	(1.11, 1.73)	9
ACTA2	1,38	(1.08, 1.76)	3
VCAN	1,38	(1.08, 1.78)	3
NOL5	1,38	(1.16, 1.66)	7
PARP1	1,38	(1.17, 1.62)	9
NPM1	1,38	(1.20, 1.59)	11
FGB	1,38	(1.21, 1.57)	20
COL8A1	1,37	(0.981, 1.91)	2
ASPN	1,37	(1.02, 1.84)	2
BPHL	1,37	(1.06, 1.76)	3
DMBT1	1,36	(1.08, 1.71)	5
SLC25A24	1,36	(1.12, 1.64)	8

Suite page suivante ...

Nom du gène	Ratio iTRAQ	Intervalle de confiance	Peptides
... suite de la page précédente			
HIBADH	1,36	(1.12, 1.67)	3
MTCH2	1,35	(1, 1.84)	2
GPX3	1,35	(1.03, 1.77)	2
TRIM28	1,35	(1.11, 1.62)	7
DPY30	1,34	(0.997, 1.77)	2
SLC25A10	1,34	(1.02, 1.75)	3
LP1	1,34	(1.04, 1.72)	2
NOL5A	1,34	(1.07, 1.67)	4
PRELP	1,34	(1.08, 1.65)	6
VDAC1	1,34	(1.11, 1.60)	8
TUBB6	1,33	(1.01, 1.75)	4
GPX2	1,33	(1.07, 1.65)	6
PRDX3	1,32	(0.999, 1.74)	3
TPR	1,32	(1.01, 1.72)	3
ATAD3A	1,32	(1.06, 1.65)	5
LPP	1,32	(1.06, 1.65)	5
RCC2	1,32	(1.08, 1.62)	4
HUWE1	1,32	(1.08, 1.64)	4
UBA1	1,32	(1.19, 1.47)	27
PRKDC	1,32	(1.20, 1.45)	37
NIPSNAP3A	1,31	(0.999, 1.72)	3
TARDBP	1,31	(1.02, 1.69)	4
CALU	1,31	(1.05, 1.65)	7
MSH6	1,3	(0.989, 1.71)	4
NUP133	1,3	(1.02, 1.66)	3
PGM5	1,3	(1.04, 1.62)	4
ELN	1,3	(1.08, 1.57)	6
ACTN1	1,3	(1.16, 1.45)	23
AIFM1	1,29	(0.953, 1.74)	3
TUBB3	1,29	Gene name(1.00, 1.65)	4
SLC12A2	1,29	(1.05, 1.59)	10
MRPL19	1,29	(1.08, 1.54)	4
FGA	1,29	(1.10, 1.51)	14
FLNA	1,28	(1.17, 1.39)	65
SLC4A1	1,27	(1.04, 1.55)	6
IARS	1,27	(1.05, 1.54)	9
MYBBP1A	1,27	(1.05, 1.56)	7
FTL	1,26	(1.01, 1.57)	3
CAPG	1,25	(1.01, 1.55)	4
PTBP1	1,25	(1.11, 1.41)	14
HSPD1	1,25	(1.12, 1.39)	20
ALDH18A1	1,23	(1.04, 1.45)	8
PHB2	1,23	(1.05, 1.44)	8
PDHB	1,22	(1.01, 1.48)	
HNRNPU	1,22	(1.03, 1.44)	15
XPO1	1,21	(1, 1.47)	5
FBN1	1,21	(1.09, 1.34)	32
DES	1,2	(1.05, 1.37)	14
TUBB	1,2	(1.08, 1.33)	20
C3	1,18	(1.01, 1.38)	24
PLOD3	1,02	(1.05, 1.61)	5
YWHAE	0,9	(0.809, 1.00)	9
COL1A2	0,839	(0.754, 0.931)	59
ACTG1	0,838	(0.747, 0.943)	24
LDHB	0,832	(0.702, 0.986)	10
HSPA1A	0,83	(0.717, 0.965)	18
PYGB	0,826	(0.724, 0.94)	21
PFN1	0,823	(0.707, 0.957)	9
EIF5A	0,82	(0.688, 0.981)	6
ALDH1A1	0,817	(0.655, 1.02)	5
PDCD6IP	0,816	(0.664, 1.00)	7
EPPK1	0,815	(0.733, 0.906)	24
PIIB	0,807	(0.688, 0.943)	8
COL1A1	0,807	(0.723, 0.898)	49
TPP1	0,806	(0.699, 0.931)	11
KRT8	0,806	(0.726, 0.899)	50
COPB1	0,803	(0.661, 0.976)	8
CALM1	0,801	(0.647, 0.99)	5
OAT	0,799	(0.65, 0.98)	4
PSME2	0,796	(0.667, 0.952)	9
MIF	0,792	(0.639, 0.981)	6
MYH14	0,79	(0.697, 0.896)	20
NDUFV1	0,789	(0.636, 0.978)	4
NDRG1	0,788	(0.63, 0.988)	4
DCN	0,788	(0.665, 0.934)	6
C19orf21	0,785	(0.63, 0.978)	6
IGKC	0,784	(0.636, 0.961)	6
NDUFA9	0,78	(0.633, 0.965)	4
CNDP2	0,777	(0.625, 0.967)	8
CAT	0,777	(0.663, 0.912)	11
ASL	0,776	(0.613, 0.977)	3
EZR	0,774	(0.682, 0.878)	14

Suite page suivante ...

Nom du gène	Ratio iTRAQ	Intervalle de confiance	Peptides
... suite de la page précédente			
PRDX6	0,774	(0.682, 0.881)	11
CA1	0,771	(0.607, 0.986)	5
NDUFB8	0,77	(0.604, 0.978)	2
PDXDC1	0,769	(0.589, 1.00)	2
GAPDH	0,766	(0.69, 0.85)	31
ABP1	0,763	(0.582, 0.998)	3
PPP2R1A	0,763	(0.628, 0.926)	5
PDIA4	0,762	(0.643, 0.902)	11
TST	0,757	(0.652, 0.877)	13
AHNAK	0,755	(0.668, 0.85)	23
FABP5	0,754	(0.597, 0.952)	3
DDT	0,753	(0.588, 0.96)	3
GOT1	0,753	(0.609, 0.924)	5
AKR1B1	0,75	(0.555, 1.02)	2
UGP2	0,75	(0.606, 0.926)	6
UBE2L3	0,749	(0.568, 0.986)	3
HSPA4	0,747	(0.563, 0.988)	4
F13A1	0,747	(0.592, 0.945)	4
TXNDC4	0,745	(0.604, 0.92)	5
SQRDL	0,744	(0.656, 0.844)	16
MYO1D	0,741	(0.617, 0.892)	12
ERP29	0,734	(0.589, 0.917)	6
ACSL5	0,73	(0.554, 0.968)	2
TYMP	0,729	(0.595, 0.893)	7
HSP90AB3	0,728	(0.604, 0.874)	9
TXNDC5	0,727	(0.616, 0.861)	11
COX5A	0,726	(0.549, 0.962)	4
RRBP1	0,724	(0.585, 0.894)	6
LCP1	0,723	(0.635, 0.823)	17
LOH11CR2A	0,722	(0.562, 0.927)	3
P4HB	0,722	(0.638, 0.82)	21
FLNB	0,721	(0.647, 0.802)	33
AKR7A3	0,716	(0.54, 0.957)	3
CTSD	0,708	(0.62, 0.808)	15
GFPT1	0,702	(0.587, 0.835)	11
VILL	0,699	(0.523, 0.933)	3
EPHX2	0,699	(0.546, 0.9)	3
MAOA	0,698	(0.573, 0.851)	8
HSP90B1	0,698	(0.635, 0.768)	28
HSPA5	0,696	(0.638, 0.76)	31
NANS	0,693	(0.535, 0.894)	6
VIL1	0,693	(0.578, 0.83)	7
RBM47	0,692	(0.515, 0.922)	2
KRT19	0,691	(0.609, 0.785)	28
GPA33	0,69	(0.477, 0.998)	2
TUBA1B	0,681	(0.557, 0.827)	4
TUBA1C	0,681	(0.557, 0.827)	4
CALR	0,672	(0.57, 0.791)	12
ALDOA	0,668	(0.577, 0.774)	15
DPP7	0,666	(0.51, 0.867)	3
SULT1B1	0,662	(0.495, 0.88)	2
CA12	0,656	(0.479, 0.902)	2
ASAH1	0,656	(0.523, 0.823)	6
MT-CO2	0,647	(0.508, 0.817)	5
TXN	0,643	(0.492, 0.85)	4
HSD11B2	0,641	(0.428, 0.954)	3
OGN	0,641	(0.492, 0.833)	3
HP	0,64	(0.526, 0.78)	7
COL3A1	0,623	(0.545, 0.712)	30
CMA1	0,622	(0.453, 0.85)	3
CTSB	0,619	(0.458, 0.838)	2
NUP188	0,616	(0.471, 0.806)	3
PLS1	0,614	(0.490, 0.773)	5
IGHA1	0,613	(0.497, 0.76)	5
TTN	0,599	(0.46, 0.783)	6
MVP	0,59	(0.519, 0.673)	23
TPSAB1	0,588	(0.506, 0.782)	6
COX6B1	0,587	(0.456, 0.758)	4
MYO1A	0,581	(0.441, 0.761)	3
AIF1	0,576	(0.428, 0.778)	2(118)
ETHE1	0,568	(0.458, 0.699)	5
PRG2	0,565	(0.448, 0.714)	3
PAPSS2	0,561	(0.463, 0.681)	9
PIGR	0,561	(0.49, 0.642)	14
LGALS3	0,552	(0.482, 0.632)	10
PLD3	0,544	(0.428, 0.69)	2
CES2	0,544	(0.474, 0.626)	14
HLA-A	0,54	(0.348, 0.853)	3
SELENBP1	0,54	(0.483, 0.602)	21
HLA-DRA	0,536	(0.357, 0.811)	3
PADI2	0,533	(0.44, 0.647)	6
UGDH	0,522	(0.44, 0.623)	9

Suite page suivante ...

Nom du gène	Ratio iTRAQ	Intervalle de confiance	Peptides
... suite de la page précédente			
PACAP	0,511	(0.418, 0.623)	4
AGR2	0,497	(0.429, 0.577)	9
AKR1B10	0,494	(0.363, 0.668)	3
MTTP	0,486	(0.345, 0.679)	2
ST6GALNAC1	0,484	(0.386, 0.605)	3
H3F3A	0,483	(0.343, 0.677)	4
IGHA2	0,479	(0.373, 0.62)	4
ALDOB	0,457	(0.333, 0.627)	2
NMES1	0,455	(0.358, 0.581)	4
CLCA1	0,454	(0.389, 0.528)	16
LGALS4	0,441	(0.4, 0.485)	36
RNASE3	0,432	(0.363, 0.514)	6
ITLN1	0,431	(0.315, 0.588)	3
CLC	0,428	(0.328, 0.558)	4
EPX	0,421	(0.376, 0.474)	19
FABP1	0,407	(0.328, 0.507)	7
CTSC	0,407	(0.348, 0.476)	9
TFF3	0,373	(0.27, 0.511)	3
RNASE2	0,357	(0.283, 0.449)	3
ADH1C	0,354	(0.267, 0.474)	5
UGT2B17	0,351	(0.248, 0.496)	2
ZG16	0,326	(0.265, 0.405)	4
MUC2	0,319	(0.283, 0.36)	23
CA2	0,312	(0.255, 0.386)	4
FCGBP	0,286	(0.253, 0.325)	30
TMPRSS11D	0,236	(0.175, 0.319)	2
FUCA1	0,222	(0.182, 0.271)	6
			Fin

## Annexe n ° 3

Tableau 3: Tableau des protéines sécrétées par les tumeurs et les lignées du cancer colorectal détectées dans la bases de données issue de la littérature (226 protéines)

Nom du gène	Protéines
A2M	Alpha-2-macroglobulin OS
ACSL5	Long-chain-fatty-acid-CoA ligase 5 OS
ACTA1	Actin, alpha skeletal muscle OS
ACTBL2	Beta-actin-like protein 2 OS
ACTC1	Actin, alpha cardiac muscle 1 OS
ACTG2	Actin, gamma-enteric smooth muscle OS
ADSL	Adenylosuccinate lyase OS
AGR2	Anterior gradient protein 2 homolog OS
AGT	Angiotensinogen OS
AHCY	Adenosylhomocysteinase OS
AHSG	Alpha-2-HS-glycoprotein OS
AIMP1	Aminoacyl tRNA synthase complex-interacting multifunctional protein 1 OS
AK1	Adenylate kinase isoenzyme 1 OS
AKR1C3	Aldo-keto reductase family 1 member C3 OS
ALAD	Delta-aminolevulinic acid dehydratase OS
ALDH18A1	Delta-1-pyrroline-5-carboxylate synthase OS
ANXA1	Annexin A1 OS
ANXA2	Annexin A2 OS
APCS	Serum amyloid P-component OS
APMAP	Adipocyte plasma membrane-associated protein OS
APOA1	Apolipoprotein A-I OS
APOC1	Apolipoprotein C-I OS
APOC3	Apolipoprotein C-III OS
APOE	Apolipoprotein E OS
APOH	Beta-2-glycoprotein 1 OS
ARF1	ADP-ribosylation factor 1 OS
ARPC3	Actin-related protein 2/3 complex subunit 3 OS
ARPC4	Actin-related protein 2/3 complex subunit 4 OS
ARPC5	Actin-related protein 2/3 complex subunit 5 OS
ARSA	Arylsulfatase A OS
ASPN	Asporin OS
ATP2A3	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3 OS
AZU1	Azurocidin OS
B2M	Beta-2-microglobulin OS
BGN	Biglycan OS
BLVRB	Flavin reductase OS
BOLA2	BolA-like protein 2 OS
BTD	Biotinidase OS
C1QB	Complement C1q subcomponent subunit B OS
C1R	Complement C1r subcomponent OS
C2	Complement C2 OS
C3	Complement C3 OS
C4BPA	C4b-binding protein alpha chain OS
C5	Complement C5 OS
C8A	Complement component C8 alpha chain OS
C8B	Complement component C8 beta chain OS
C9	Complement component C9 OS
CA1	Carbonic anhydrase 1 OS
CALM1	Calmodulin OS
CALU	Calumenin OS
CAPNS1	Calpain small subunit 1 OS
CAPZB	F-actin-capping protein subunit beta OS
CASP1	Caspase-1 OS
CBR1	Carbonyl reductase [NADPH] 1 OS
CD14	Monocyte differentiation antigen CD14 OS
CD59	CD59 glycoprotein OS
CDC42	Cell division control protein 42 homolog OS
CES1	Liver carboxylesterase 1 OS
CFL1	Cofilin-1 OS
CLTA	Clathrin light chain A OS
CLU	Clusterin OS
CMA1	Chymase OS

Suite page suivante ...

Nom du gène	Protéines
... suite de la page précédente	
CMPK1	UMP-CMP kinase OS
CNN2	Calponin-2 OS
COQ5	Ubiquinone biosynthesis methyltransferase COQ5, mitochondrial OS
CORO1A	Coronin-1A OS
COTL1	Coactosin-like protein OS
CP	Ceruloplasmin OS
CRIP2	Cysteine-rich protein 2 OS
CTSB	Cathepsin B OS
CTSD	Cathepsin D OS
CTSG	Cathepsin G OS
CTSS	Cathepsin S OS
CTSZ	Cathepsin Z OS
DCN	Decorin OS
DEFA1	Neutrophil defensin 1 OS
DEFA3	Neutrophil defensin 3 OS
DNAJC10	DnaJ homolog subfamily C member 10 OS
DPP7	Dipeptidyl peptidase 2 OS
DSTN	Dextrin OS
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1 OS
ENDOD1	Endonuclease domain-containing 1 protein OS
ENO2	Gamma-enolase OS
ERAP1	Endoplasmic reticulum aminopeptidase 1 OS
ETFDH	Electron transfer flavoprotein-ubiquinone oxidoreductase, mitochondrial OS
F2	Prothrombin OS
FABP1	Fatty acid-binding protein, liver OS
FAH	Fumarylacetoacetase OS
FBLN1	Fibulin-1 OS
FBLN2	Fibulin-2 OS
FBLN5	Fibulin-5 OS
FDPS	Farnesyl pyrophosphate synthase OS
FGB	Fibrinogen beta chain OS
FMOD	Fibromodulin OS
FUCA1	Tissue alpha-L-fucosidase OS
FXR1	Fragile X mental retardation syndrome-related protein 1 OS
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2 OS
GGH	Gamma-glutamyl hydrolase OS
GLA	Alpha-galactosidase A OS
GLB1	Beta-galactosidase OS
GPX3	Glutathione peroxidase 3 OS
GRN	Granulins OS
GSN	Gelsolin OS
GSTP1	Glutathione S-transferase P OS
HIST1H2AH	Histone H2A type 1-H OS
HIST1H2AJ	Histone H2A type 1-J OS
HMGCS2	Hydroxymethylglutaryl-CoA synthase, mitochondrial OS
HMGN2	Non-histone chromosomal protein HMG-17 OS
HPCAL1	Hippocalcin-like protein 1 OS
HPX	Hemopexin OS
HSPA5	78 kDa glucose-regulated protein OS
HSPB1	Heat shock protein beta-1 OS
HTRA1	Serine protease HTRA1 OS
IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic OS
IGFBP4	Insulin-like growth factor-binding protein 4 OS
IGHG1	Ig gamma-1 chain C region OS
IGHG2	Ig gamma-2 chain C region OS
IGHG3	Ig gamma-3 chain C region OS
IGHG4	Ig gamma-4 chain C region OS
IPO9	Importin-9 OS
ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1 OS
ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2 OS
ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3 OS
KCTD12	BTB/POZ domain-containing protein KCTD12 OS
KPNB1	Importin subunit beta-1 OS
KRT19	Keratin, type I cytoskeletal 19 OS
LAD1	Ladinin-1 OS
LARP4B	La-related protein 4B OS
LASP1	LIM and SH3 domain protein 1 OS
LBP	Lipopolysaccharide-binding protein OS
LCN2	Neutrophil gelatinase-associated lipocalin OS
LDHA	L-lactate dehydrogenase A chain OS
LDHB	L-lactate dehydrogenase B chain OS
LGALS3	Galectin-3 OS
LOXL1	Lysyl oxidase homolog 1 OS
LRG1	Leucine-rich alpha-2-glycoprotein OS
LUM	Lumican OS
LYZ	Lysozyme C OS
MFAP4	Microfibril-associated glycoprotein 4 OS
MFGE8	Lactadherin OS
MIF	Macrophage migration inhibitory factor OS
MPO	Myeloperoxidase OS
MRPL11	39S ribosomal protein L11, mitochondrial OS
MSN	Moiesin OS

Suite page suivante ...

Nom du gène	Protéines
... suite de la page précédente	
MYL9	Myosin regulatory light polypeptide 9 OS
NDUFA8	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8 OS
NID1	Nidogen-1 OS
NID2	Nidogen-2 OS
NIF3L1	NIF3-like protein 1 OS
NOP2	Putative ribosomal RNA methyltransferase NOP2 OS
NUTF2	Nuclear transport factor 2 OS
OLFM4	Olfactomedin-4 OS
OLFML3	Olfactomedin-like protein 3 OS
ORM1	Alpha-1-acid glycoprotein 1 OS
ORM2	Alpha-1-acid glycoprotein 2 OS
P4HB	Protein disulfide-isomerase OS
PDCD6	Programmed cell death protein 6 OS
PDIA3	Protein disulfide-isomerase A3 OS
PDIA6	Protein disulfide-isomerase A6 OS
PDXK	Pyridoxal kinase OS
PGLYRP2	N-acetylmuramoyl-L-alanine amidase OS
PLBD1	Putative phospholipase B-like 1 OS
PLP2	Proteolipid protein 2 OS
PON2	Serum paraoxonase/arylesterase 2 OS
PPT1	Palmitoyl-protein thioesterase 1 OS
PRDX1	Peroxiredoxin-1 OS
PRDX2	Peroxiredoxin-2 OS
PRDX5	Peroxiredoxin-5, mitochondrial OS
PRELP	Prolargin OS
PROS1	Vitamin K-dependent protein S OS
PSAP	Proactivator polypeptide OS
PSMD9	26S proteasome non-ATPase regulatory subunit 9 OS
PTRF	Polymerase I and transcript release factor OS
PZP	Pregnancy zone protein OS
QSOX1	Sulfhydryl oxidase 1 OS
QSOX2	Sulfhydryl oxidase 2 OS
RAC2	Ras-related C3 botulinum toxin substrate 2 OS
RAP1B	Ras-related protein Rap-1b OS
REG1A	Lithostathine-1-alpha OS
RHOC	Rho-related GTP-binding protein RhoC OS
RNASE3	Eosinophil cationic protein OS
RNASE4	Ribonuclease 4 OS
RNH1	Ribonuclease inhibitor OS
RNPEP	Aminopeptidase B OS
RPL8	60S ribosomal protein L8 OS
RPS4X	40S ribosomal protein S4, X isoform OS
RSU1	Ras suppressor protein 1 OS
S100A14	Protein S100-A14 OS
S100A4	Protein S100-A4 OS
SAA1	Serum amyloid A protein OS
SAR1A	GTP-binding protein SAR1a OS
SERPINA3	Alpha-1-antichymotrypsin OS
SERPINB9	Serpin B9 OS
SERPINC1	Antithrombin-III OS
SERPIND1	Heparin cofactor 2 OS
SERPINF1	Pigment epithelium-derived factor OS
SERPINF2	Alpha-2-antiplasmin OS
SERPING1	Plasma protease C1 inhibitor OS
SH3BGRL3	SH3 domain-binding glutamic acid-rich-like protein 3 OS
SLC3A2	4F2 cell-surface antigen heavy chain OS
SOD1	Superoxide dismutase [Cu-Zn] OS
SOD3	Extracellular superoxide dismutase [Cu-Zn] OS
SORD	Sorbitol dehydrogenase OS
SPINT2	Kunitz-type protease inhibitor 2 OS
TAGLN2	Transgelin-2 OS
TFRC	Transferrin receptor protein 1 OS
TGFB1I1	Transforming growth factor beta-1-induced transcript 1 protein OS
THBS2	Thrombospondin-2 OS
TINAGL1	Tubulointerstitial nephritis antigen-like OS
TPM1	Tropomyosin alpha-1 chain OS
TPSAB1	Tryptase beta-1 OS
TPT1	Translationally-controlled tumor protein OS
TRY6	Putative trypsin-6 OS
TTR	Transthyretin OS
TUBA4A	Tubulin alpha-4A chain OS
TUBB1	Tubulin beta-1 chain OS
UBE2L3	Ubiquitin-conjugating enzyme E2 L3 OS
UBE2N	Ubiquitin-conjugating enzyme E2 N OS
UBE2V1	Ubiquitin-conjugating enzyme E2 variant 1 OS
UQCRB	Cytochrome b-c1 complex subunit 7 OS
VIM	Vimentin OS
VTN	Vitronectin OS
VWA1	von Willebrand factor A domain-containing protein 1 OS
WDR1	WD repeat-containing protein 1 OS
WIBG	Partner of Y14 and mago OS
ZYX	Zyxin OS



## Annexe n° 4

Tableau 4: Tableau des 245 protéines identifiées par l'analyse du stroma tumoral – Les protéines en gras correspondent aux protéines identifiées dans le stroma tumoral et présentes dans notre base de données des protéines sécrétées.

Nom du gène	Protéines
ALB	Serum albumin OS
COL1A1	Collagen alpha-1(I) chain OS
MYH9	Myosin-9 OS
FLNA	Filamin-A OS
COL1A2	Collagen alpha-2(I) chain OS
COL6A3	Collagen alpha-3(VI) chain OS
FN1	Fibronectin OS
KRT8	Keratin, type II cytoskeletal 8 OS
VIM	Vimentin OS
ACTB	Actin, cytoplasmic 1 OS
ACTG1	Actin, cytoplasmic 2 OS
HP	Haptoglobin OS
TF	Serotransferrin OS
TPM4	Tropomyosin alpha-4 chain OS
HBA1	Hemoglobin subunit alpha OS
COL12A1	Collagen alpha-1(XII) chain OS
PGK1	Phosphoglycerate kinase 1 OS
ANXA2	Annexin A2 OS
POSTN	Periostin OS
HIST1H4A	Histone H4 OS
C3	Complement C3 OS
EZR	Ezrin OS
RDX	Radixin OS
HSP90B1	Endoplasmin OS
AHNAK	Neuroblast differentiation-associated protein AHNAK OS
PRSS1	Trypsin-1 OS
ACTN4	Alpha-actinin-4 OS
TUBB	Tubulin beta chain OS
HIST1H2BM	Histone H2B type 1-M OS
HIST1H2BN	Histone H2B type 1-N OS
HIST1H2BH	Histone H2B type 1-H OS
HIST2H2BF	Histone H2B type 2-F OS
HIST1H2BC	Histone H2B type 1-C/E/F/G/I OS
HIST1H2BD	Histone H2B type 1-D OS
HIST1H2BK	Histone H2B type 1-K OS
HIST1H2BL	Histone H2B type 1-L OS
H2BFS	Histone H2B type F-S OS
HBB	Hemoglobin subunit beta OS
PPIA	Peptidyl-prolyl cis-trans isomerase A OS
HSPA5	78 kDa glucose-regulated protein OS
PDIA3	Protein disulfide-isomerase A3 OS
KRT1	Keratin, type II cytoskeletal 1 OS
FGB	Fibrinogen beta chain OS
TGFBI	Transforming growth factor-beta-induced protein ig-h3 OS
FBN1	Fibrillin-1 OS
BGN	Biglycan OS
OLFM4	Olfactomedin-4 OS
CFL1	Cofilin-1 OS
TUBA1A	Tubulin alpha-1A chain OS
TUBA1B	Tubulin alpha-1B chain OS
TUBA1C	Tubulin alpha-1C chain OS
TAGLN	Transgelin OS
IGHG2	Ig gamma-2 chain C region OS
IGKC	Ig kappa chain C region OS
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase OS
IGHA2	Ig alpha-2 chain C region OS
IGHA1	Ig alpha-1 chain C region OS
EEF1AL3	Putative elongation factor 1-alpha-like 3 OS
EEF1A1	Elongation factor 1-alpha 1 OS
EEF1A2	Elongation factor 1-alpha 2 OS
PRDX1	Peroxisredoxin-1 OS

Suite page suivante ...

Nom du gène	Protéines
... suite de la page précédente	
<b>ORM1</b>	<b>Alpha-1-acid glycoprotein 1 OS</b>
<b>DCN</b>	<b>Decorin OS</b>
<b>FGA</b>	<b>Fibrinogen alpha chain OS</b>
<b>THBS1</b>	<b>Thrombospondin-1 OS</b>
<b>ENO1</b>	Alpha-enolase OS
<b>RPS11</b>	40S ribosomal protein S11 OS
<b>HIST1H1C</b>	<b>Histone H1.2 OS</b>
<b>HIST1H1D</b>	Histone H1.3 OS
<b>HIST1H1E</b>	<b>Histone H1.4 OS</b>
<b>CALD1</b>	<b>Caldesmon OS</b>
<b>ALDOA</b>	<b>Fructose-bisphosphate aldolase A OS</b>
<b>PPIB</b>	<b>Peptidyl-prolyl cis-trans isomerase B OS</b>
<b>PKM2</b>	Pyruvate kinase isozymes M1/M2 OS
<b>CLCA1</b>	Calcium-activated chloride channel regulator 1 OS
<b>HSPD1</b>	<b>60 kDa heat shock protein, mitochondrial OS</b>
<b>YWHAZ</b>	<b>14-3-3 protein zeta/delta OS</b>
<b>KRT18</b>	Keratin, type I cytoskeletal 18 OS
<b>VTN</b>	<b>Vitronectin OS</b>
<b>COL6A1</b>	<b>Collagen alpha-1(VI) chain OS</b>
<b>HSP90AB1</b>	Heat shock protein HSP 90-beta OS
<b>HSP90AB3P</b>	Putative heat shock protein HSP 90-beta-3 OS
<b>CALR</b>	<b>Calreticulin OS</b>
<b>LUM</b>	<b>Lumican OS</b>
<b>HNRNPA1</b>	Heterogeneous nuclear ribonucleoprotein A1 OS
<b>HNRNPA1L2</b>	Heterogeneous nuclear ribonucleoprotein A1-like 2 OS
<b>HNRNPA1L3</b>	Putative heterogeneous nuclear ribonucleoprotein A1-like 3 OS
<b>RAN</b>	GTP-binding nuclear protein Ran OS
<b>S100A9</b>	<b>Protein S100-A9 OS</b>
<b>KRT10</b>	<b>Keratin, type I cytoskeletal 10 OS</b>
<b>RPL3</b>	60S ribosomal protein L3 OS
<b>TXNDC5</b>	Thioredoxin domain-containing protein 5 OS
<b>COL6A2</b>	<b>Collagen alpha-2(VI) chain OS</b>
<b>PRDX2</b>	<b>Peroxiredoxin-2 OS</b>
<b>MYL6</b>	<b>Myosin light polypeptide 6 OS</b>
<b>TPM1</b>	<b>Tropomyosin alpha-1 chain OS</b>
<b>SERPINA1</b>	<b>Alpha-1-antitrypsin OS</b>
<b>C4B</b>	<b>Complement C4-B OS</b>
<b>C4A</b>	<b>Complement C4-A OS</b>
<b>KRT9</b>	<b>Keratin, type I cytoskeletal 9 OS</b>
<b>EEF2</b>	<b>Elongation factor 2 OS</b>
<b>MYH11</b>	Myosin-11 OS
<b>THBS2</b>	<b>Thrombospondin-2 OS</b>
<b>HMGB1</b>	<b>High mobility group protein B1 OS</b>
<b>RPL15</b>	60S ribosomal protein L15 OS
<b>UBC</b>	Polyubiquitin-C OS
<b>UBB</b>	Polyubiquitin-B OS
<b>UBA52</b>	Ubiquitin-60S ribosomal protein L40 OS
<b>RPS27A</b>	<b>Ubiquitin-40S ribosomal protein S27a OS</b>
<b>P4HB</b>	<b>Protein disulfide-isomerase OS</b>
<b>S100A8</b>	<b>Protein S100-A8 OS</b>
<b>COX6C</b>	Cytochrome c oxidase subunit 6C OS
<b>PFN1</b>	<b>Profilin-1 OS</b>
<b>RPL21</b>	60S ribosomal protein L21 OS
<b>RPL5</b>	60S ribosomal protein L5 OS
<b>WDR1</b>	<b>WD repeat-containing protein 1 OS</b>
<b>H2AFJ</b>	Histone H2A.J OS
<b>HIST1H2AJ</b>	<b>Histone H2A type 1-J OS</b>
<b>HIST1H2AH</b>	<b>Histone H2A type 1-H OS</b>
<b>HIST2H2AA3</b>	Histone H2A type 2-A OS
<b>HIST2H2AC</b>	Histone H2A type 2-C OS
<b>HIST1H2AD</b>	Histone H2A type 1-D OS
<b>HIST1H2AG</b>	Histone H2A type 1 OS
<b>HSPG2</b>	<b>Basement membrane-specific heparan sulfate proteoglycan core protein OS</b>
<b>HIST3H3</b>	Histone H3.1t OS
<b>HIST2H3A</b>	Histone H3.2 OS
<b>HIST1H3A</b>	Histone H3.1 OS
<b>H3F3A</b>	Histone H3.3 OS
<b>REG3A</b>	Regenerating islet-derived protein 3-alpha OS
<b>GSN</b>	<b>Gelsolin OS</b>
<b>LGALS1</b>	<b>Galectin-1 OS</b>
<b>ACTN1</b>	<b>Alpha-actinin-1 OS</b>
<b>APOA1</b>	<b>Apolipoprotein A-I OS</b>
<b>ANXA5</b>	Annexin A5 OS
<b>TPPI</b>	<b>Triosephosphate isomerase OS</b>
<b>VCAN</b>	<b>Versican core protein OS</b>
<b>RPL7A</b>	60S ribosomal protein L7a OS
<b>RPS16</b>	40S ribosomal protein S16 OS
<b>FLNB</b>	<b>Filamin-B OS</b>
<b>RPL8</b>	<b>60S ribosomal protein L8 OS</b>
<b>RNASE3</b>	<b>Eosinophil cationic protein OS</b>
<b>A2M</b>	<b>Alpha-2-macroglobulin OS</b>
<b>PLEC</b>	Plectin OS

Suite page suivante ...

Nom du gène	Protéines
... suite de la page précédente	
<b>YWHAE</b>	<b>14-3-3 protein epsilon OS</b>
HADHA	Trifunctional enzyme subunit alpha, mitochondrial OS
RPL31	60S ribosomal protein L31 OS
RPS23	40S ribosomal protein S23 OS
RPL26L1	60S ribosomal protein L26-like 1 OS
RPL26	60S ribosomal protein L26 OS
<b>CAP1</b>	<b>Adenylyl cyclase-associated protein 1 OS</b>
PDIA4	Protein disulfide-isomerase A4 OS
<b>OGN</b>	<b>Mimecan OS</b>
MYH2	Myosin-2 OS
<b>MYH8</b>	<b>Myosin-8 OS</b>
MYH1	Myosin-1 OS
<b>TNC</b>	<b>Tenascin OS</b>
NCL	Nucleolin OS
HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2 OS
CNN1	Calponin-1 OS
<b>FGG</b>	<b>Fibrinogen gamma chain OS</b>
<b>ARPC2</b>	<b>Actin-related protein 2/3 complex subunit 2 OS</b>
HNRNPU	Heterogeneous nuclear ribonucleoprotein U OS
<b>ARHGDI</b>	<b>Rho GDP-dissociation inhibitor 1 OS</b>
SERPINH1	Serpin H1 OS
EIF4A2	Eukaryotic initiation factor 4A-II OS
EIF4A3	Eukaryotic initiation factor 4A-III OS
EIF4A1	Eukaryotic initiation factor 4A-I OS
RPL18	60S ribosomal protein L18 OS
HSPE1	10 kDa heat shock protein, mitochondrial OS
<b>DEFA1</b>	<b>Neutrophil defensin 1 OS</b>
<b>DEFA3</b>	<b>Neutrophil defensin 3 OS</b>
MDH2	Malate dehydrogenase, mitochondrial OS
PLA2G2A	Phospholipase A2, membrane associated OS
ATP5B	ATP synthase subunit beta, mitochondrial OS
LPP	Lipoma-preferred partner OS
DPYSL2	Dihydropyrimidinase-related protein 2 OS
EEF1B2	Elongation factor 1-beta OS
NME2	Nucleoside diphosphate kinase B OS
<b>NME1</b>	<b>Nucleoside diphosphate kinase A OS</b>
CYCS	Cytochrome c OS
LGALS4	Galectin-4 OS
RPL34	60S ribosomal protein L34 OS
<b>LGALS3</b>	<b>Galectin-3 OS</b>
MYL12A	Myosin regulatory light chain 12A OS
MYL12B	Myosin regulatory light chain 12B OS
<b>TXN</b>	<b>Thioredoxin OS</b>
<b>VCL</b>	<b>Vinculin OS</b>
RPL35	60S ribosomal protein L35 OS
<b>KRT2</b>	<b>Keratin, type II cytoskeletal 2 epidermal OS</b>
<b>KRT6B</b>	<b>Keratin, type II cytoskeletal 6B OS</b>
<b>KRT6C</b>	<b>Keratin, type II cytoskeletal 6C OS</b>
<b>KRT6A</b>	<b>Keratin, type II cytoskeletal 6A OS</b>
KRT4	Keratin, type II cytoskeletal 4 OS
RPS8	40S ribosomal protein S8 OS
<b>VIL1</b>	<b>Villin-1 OS</b>
HIST2H2BE	Histone H2B type 2-E OS
HIST1H2BB	Histone H2B type 1-B OS
HIST1H2BO	Histone H2B type 1-O OS
HIST1H2BJ	Histone H2B type 1-J OS
HIST2H2BD	Putative histone H2B type 2-D OS
HIST2H2BC	Putative histone H2B type 2-C OS
HIST3H2BB	Histone H2B type 3-B OS
ANXA6	Annexin A6 OS
<b>DSTN</b>	<b>Destrin OS</b>
IGHM	Ig mu chain C region OS
MUCB	Ig mu heavy chain disease protein OS
<b>CSTB</b>	<b>Cystatin-B OS</b>
<b>NID2</b>	<b>Nidogen-2 OS</b>
<b>CLTC</b>	<b>Clathrin heavy chain 1 OS</b>
<b>HSPA8</b>	<b>Heat shock cognate 71 kDa protein OS</b>
CLC	Eosinophil lysophospholipase OS
RPL13A	60S ribosomal protein L13a OS
<b>GANAB</b>	<b>Neutral alpha-glucosidase AB OS</b>
HSP90AA1	Heat shock protein HSP 90-alpha OS
<b>LMNB1</b>	<b>Lamin-B1 OS</b>
LMNA	Lamin-A/C OS
ECHS1	Enoyl-CoA hydratase, mitochondrial OS
<b>ACTC1</b>	<b>Actin, alpha cardiac muscle 1 OS</b>
<b>ACTG2</b>	<b>Actin, gamma-enteric smooth muscle OS</b>
ACTA2	Actin, aortic smooth muscle OS
<b>ACTA1</b>	<b>Actin, alpha skeletal muscle OS</b>
TLN1	Talin-1 OS
<b>KRT19</b>	<b>Keratin, type I cytoskeletal 19 OS</b>
TUBB2A	Tubulin beta-2A chain OS
HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1 OS

Suite page suivante ...

Nom du gène	Protéines
... suite de la page précédente	
HIST1H2AC	Histone H2A type 1-C OS
HIST3H2A	Histone H2A type 3 OS
HIST1H2AB	Histone H2A type 1-B/E OS
HV102	Ig heavy chain V-I region HG3 OS
<b>MSN</b>	<b>Moesin OS</b>
POTEE	POTE ankyrin domain family member E OS
POTEI	POTE ankyrin domain family member I OS
POTEKP	Putative beta-actin-like protein 3 OS
POTEJ	POTE ankyrin domain family member J OS
<b>TPM2</b>	<b>Tropomyosin beta chain OS</b>
<b>LTF</b>	<b>Lactotransferrin OS</b>
<b>ORM2</b>	<b>Alpha-1-acid glycoprotein 2 OS</b>
<b>COL4A1</b>	<b>Collagen alpha-1(IV) chain OS</b>
MX2	Interferon-induced GTP-binding protein Mx2 OS
SLC25A4	ADP/ATP translocase 1 OS
SLC25A6	ADP/ATP translocase 3 OS
SLC25A5	ADP/ATP translocase 2 OS
PLS3	Plastin-3 OS
<b>PARK7</b>	<b>Protein DJ-1 OS</b>
<b>IGHG1</b>	<b>Ig gamma-1 chain C region OS</b>
	Fin

## Annexe n° 5

Tableau 5: Tableau des 53 protéines glycosylées identifiées dans une tumeur de côlon par enrichissement sur des lectines

Accessions	Nom du gène	Protéines
Q13510	ASAH1_HUMAN	Acid ceramidase OS
P63267	ACTH_HUMAN	Actin, gamma-enteric smooth muscle OS
Q9HDC9	APMAP_HUMAN	Adipocyte plasma membrane-associated protein OS
P16112	PGCA_HUMAN	Aggrecan core protein OS
P19801	ABP1_HUMAN	Amiloride-sensitive amine oxidase [copper-containing] OS
P01008	ANT3_HUMAN	Antithrombin-III OS
P06865	HEXA_HUMAN	Beta-hexosaminidase subunit alpha OS
P21810	PGS1_HUMAN	Biglycan OS
Q12864	CAD17_HUMAN	Cadherin-17 OS
P13688	CEAM1_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 1 OS
P06731	CEAM5_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 5 OS
P40199	CEAM6_HUMAN	Carcinoembryonic antigen-related cell adhesion molecule 6 OS
P07339	CATD_HUMAN	Cathepsin D OS
Q3MIR4	CC50B_HUMAN	Cell cycle control protein 50B OS
P10909	CLUS_HUMAN	Clusterin OS
P12109	CO6A1_HUMAN	Collagen alpha-1(VI) chain OS
P12110	CO6A2_HUMAN	Collagen alpha-2(VI) chain OS
P12111	CO6A3_HUMAN	Collagen alpha-3(VI) chain OS
P01024	CO3_HUMAN	Complement C3 OS
Q9UGM3	DMBT1_HUMAN	Deleted in malignant brain tumors 1 protein OS
Q9UHL4	DPP2_HUMAN	Dipeptidyl peptidase 2 OS
P46977	STT3A_HUMAN	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3A OS
Q8TCJ2	STT3B_HUMAN	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3B OS
P14625	ENPL_HUMAN	Endoplasmic reticulum protein OS
Q96HE7	ERO1A_HUMAN	ERO1-like protein alpha OS
P35555	FBN1_HUMAN	Fibrillin-1 OS
Q08380	LG3BP_HUMAN	Galectin-3-binding protein OS
Q16769	QPCT_HUMAN	Glutamyl-peptide cyclotransferase OS
P00738	HPT_HUMAN	Haptoglobin OS
P30480	1B42_HUMAN	HLA class I histocompatibility antigen, B-2 alpha chain OS
Q07000	1C15_HUMAN	HLA class I histocompatibility antigen, Cw-15 alpha chain OS
Q9Y4L1	HYOU1_HUMAN	Hypoxia up-regulated protein 1 OS
P01859	IGHG2_HUMAN	Ig gamma-2 chain C region OS
P06756	ITAV_HUMAN	Integrin alpha-V OS
P11047	LAMC1_HUMAN	Laminin subunit gamma-1 OS
P11117	PPAL_HUMAN	Lysosomal acid phosphatase OS
O00754	MA2B1_HUMAN	Lysosomal alpha-mannosidase OS
P05164	PERM_HUMAN	Myeloperoxidase OS
Q6UX06	OLFM4_HUMAN	Olfactomedin-4 OS
Q15063	POSTN_HUMAN	Periostin OS
P01833	PIGR_HUMAN	Polymeric immunoglobulin receptor OS
Q10472	GALT1_HUMAN	Polypeptide N-acetylgalactosaminyltransferase 1 OS
Q8NBJ5	GT251_HUMAN	Procollagen galactosyltransferase 1 OS
O00469	PLOD2_HUMAN	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 OS
O60568	PLOD3_HUMAN	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 OS
Q07954	LRP1_HUMAN	Prolow-density lipoprotein receptor-related protein 1 OS
Q6P4A8	PLBL1_HUMAN	Putative phospholipase B-like 1 OS
P05026	AT1B1_HUMAN	Sodium/potassium-transporting ATPase subunit beta-1 OS
Q8WVQ1	CANT1_HUMAN	Soluble calcium-activated nucleotidase 1 OS
Q9Y5Y6	ST14_HUMAN	Suppressor of tumorigenicity 14 protein OS
P24821	TENA_HUMAN	Tenascin OS
P02786	TFR1_HUMAN	Transferrin receptor protein 1 OS
O14773	TPP1_HUMAN	Tripeptidyl-peptidase 1 OS

Fin

## Résumé :

Les avancées récentes dans l'étude des phénomènes aboutissant au cancer montrent que chaque tumeur possède des caractéristiques physiopathologiques qui lui sont propres. L'apport de biomarqueurs, permettant de réaliser des diagnostics moléculaires précis est donc un enjeu primordial.

Une analyse protéomique quantitative globale basée sur l'utilisation du marquage iTRAQ<sup>TM</sup> a été réalisée sur des tumeurs colorectales congelées. Ces tumeurs ont été analysées en fonction de leur stade TNM d'une part, et de leur statut concernant l'oncogène KRAS d'autre part. Afin de compléter notre analyse nous avons également analysé le protéome de lignées cellulaires colorectales. Nous avons également examiné le stroma de ces tumeurs, et nous avons développé une technique d'analyse d'un sous-protéome : le glycoprotéome.

L'ensemble de ces résultats nous a conduit à proposer une base de données des protéines identifiables dans ce modèle constituée de 3168 protéines. 513 des protéines que nous avons identifiées sont susceptibles d'être sécrétées par les tumeurs et constituent une base de données de candidats biomarqueurs pour le cancer colorectal.

Nous avons également validé l'un d'entre eux : l'OLFM4. Il s'agit d'un candidat biomarqueur potentiel pour la détection précoce des cancers colorectaux, et nous avons également mis en évidence que son niveau d'expression était modifié lorsque la tumeur était mutée sur l'oncogène KRAS. Nous nous sommes également intéressés à ses fonctions cellulaires, et nous avons montré qu'elle participait à la résistance des tumeurs aux traitements, et il est probable qu'elle joue un rôle dans la dissémination tumorale.

## Mots-clés :

Protéomique, iTRAQ, OFFGEL, Biomarqueurs, Cancer colorectal, Olfactoméline 4.

## Abstract :

Recent advances in the study of phenomena leading to cancer show that each tumor has their own pathophysiological characteristics. The contribution of biomarkers to achieve specific molecular diagnostics is a key issue.

A comprehensive quantitative proteomic analysis based on the use of iTRAQ labeling was performed on frozen colorectal tumors. These tumors were analyzed according to their TNM stage on one hand, and their status of the KRAS oncogene in another hand. To complete our analysis we also analyzed the proteome of colorectal cell lines. We also examined the stroma of these tumor, and we have developed a technique for analyzing a subproteome : the glycoproteome.

All these results led us to propose a database of proteins identified in this model with 3168 proteins. 513 proteins we identified are likely to be secreted by tumors and constitute a database of candidate biomarkers for colorectal cancer.

We also validated one of them : OLFM4. This is a potential candidate biomarker for the early detection of colorectal cancer, and we also demonstrated that the expression level was changed when the tumor has expressed KRAS oncogene. We are also interested in its cellular functions, and we showed that it was involved in tumor resistance to treatment, and it is likely to play a role in tumor dissemination.

## Mots-clés :

Proteomics, iTRAQ, OFFGEL, Biomarkers, Colorectal cancer, Olfactoméline 4.