



HAL
open science

Le protéome urinaire : caractérisation et intérêt pour la recherche de biomarqueurs de pathologies

Magali Maizi

► **To cite this version:**

Magali Maizi. Le protéome urinaire : caractérisation et intérêt pour la recherche de biomarqueurs de pathologies. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Grenoble, 2014. Français. NNT : 2014GRENV075 . tel-01556471

HAL Id: tel-01556471

<https://theses.hal.science/tel-01556471>

Submitted on 5 Jul 2017

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Chimie et Science du Vivant / Biotechnologie**

Arrêté ministériel : 7 août 2006

Présentée par

Magali Court

Thèse dirigée par **Christophe Masselon**

préparée au sein du **Laboratoire d'Etude de la Dynamique des protéomes**
dans l'**École Doctorale Chimie et Science du vivant**

Le Protéome Urinaire : Caractérisation et Intérêt pour la Recherche de Biomarqueurs de Pathologies.

Thèse soutenue publiquement le **20 juin 2014**,
devant le jury composé de :

M. Charles PINEAU

Docteur, Université de Rennes, Rapporteur

M. Christophe FLAHAUT

Docteur, Université d'Artois, Rapporteur

M. Yves ALLORY

Professeur, Hôpital Henri Mondor, Examineur

M. Francois BERGER

Professeur, Université de Grenoble, Examineur

M. Alexander SCHERL

Docteur, Hôpitaux Universitaires de Genève, Examineur

M. Christophe MASSELON

Docteur, CEA de Grenoble, Directeur de Thèse

*Université Joseph Fourier / Université Pierre Mendès France /
Université Stendhal / Université de Savoie / Grenoble INP*



A mes enfants Charlène, Elsa et Alexandre

Remerciements

Ce travail a été réalisé au Laboratoire d'Etude de la dynamique des Protéomes (EDyP) sous la direction de Jérôme Garin, puis de Christophe Bruley et sous l'encadrement de Christophe Masselon.

Je tiens à exprimer ma profonde et sincère reconnaissance à Jérôme Garin pour m'avoir initié à la protéomique ainsi que pour ses critiques constructives et son soutien qui m'ont fait évoluer au cours de ces années à ses côtés.

Je tiens également à remercier Christophe Masselon pour m'avoir permis de réaliser cette thèse, pour avoir encadré ce travail, pour ses conseils précieux et toutes nos discussions formelles ou informelles autour d'un café, un immense merci pour tout ce que j'ai appris à tes côtés.

J'adresse mes vifs remerciement au Professeur Francois Berger qui a accepté de présider mon jury de thèse ainsi qu'à Charles Pineau, Christophe Flahaut, Yves Allory et Alexander Scherl pour avoir consacré de leur temps à l'évaluation de ce travail.

Je tiens également à remercier l'ensemble des personnes avec lesquelles j'ai été amenée à collaborer et qui m'ont entraînées dans diverses thématiques : Aurélie Kamoun et Francois Radvanyi pour nos discussions autour de la transcriptomique et du cancer de la vessie. Edwige Lopes et Yves Allory pour nos échanges sur le cancer de la vessie et la clinique en général qui ont permis d'établir un lien fort entre la clinique et la protéomique. Un grand merci à Elodie Duriez, Mariette Matondo, Nahtalie Selevsek et Bruno Domon pour le travail que nous avons réalisé en commun sur la préparation des échantillons et l'évaluation des biomarqueurs. Yann De Ryke et Mourad Mellal pour nos échanges sur la statistique et les applications à la protéomique clinique, deux mondes si lointains au départ.

Je tiens à remercier l'ensemble du laboratoire EdyP et plus particulièrement Véronique Dupierris, Claire Adam, Jérémie Turbet et Christophe Bruley pour le traitement des données et l'élaboration des logiciels qui m'ont aidé tout au long de ce travail, Madalen Le Gorrec pour son aide précieuse à la préparation des échantillons, Agnès Chapel pour ses petits conseils si précieux, Hayet Chebli pour son aide sur le projet DECanBIO et Yves Vandenbrouk pour ses conseils et son aide à la fouille de données.

Un immense merci à Elodie et Véronique pour nos échanges, votre aide si précieuse et votre amitié.

Enfin, je remercie chaleureusement toute ma famille, mes amis, mon mari et mes enfants qui sont ma joie de vivre au quotidien.

Sommaire

Introduction générale 16

I. Chapitre I : Etat de l'art 19

A.	La protéomique.....	19
1.	Des gènes aux protéines	19
2.	Le protéome et la protéomique.....	20
3.	L'analyse Protéomique	21
a)	De la chimie des protéines à la protéomique	21
b)	Spectrométrie de masse, un outil de choix pour la protéomique	22
(1)	Principes de la technique	22
i.	Source d'ionisation.....	23
ii.	Les analyseurs	24
iii.	La spectrométrie de masse en tandem.....	26
(2)	Approches protéomiques par spectrométrie de masse	27
(3)	Couplage chromatographie liquide et spectrométrie de masse en tandem	27
4.	Protéomique quantitative.....	28
a)	Les méthodes de quantification avec marquage isotopique différentiel	30
b)	Les méthodes de quantification avec standard internes (marquage isotopique).....	31
c)	Les méthodes de quantification sans marquage « Label free »	33
(1)	Spectral count	34
(2)	Méthode AMT - Accurate Mass and Time Tags	34
B.	Protéomique clinique.....	35
1.	Les biomarqueurs.....	37
2.	Les fluides biologiques et leur rôle dans la découverte des biomarqueurs	38
a)	L'accessibilité aux cohortes d'échantillons	39
b)	De la collecte au traitement préanalytique des échantillons de fluide biologique	40
C.	Contexte clinique de la thèse (DECanBio).....	43
1.	La Vessie.....	43
2.	Les tumeurs de la vessie	44
3.	Les méthodes de détection des tumeurs vésicales	46
4.	Biomarqueurs urinaires du cancer de la vessie	47
D.	Etat de l'art en caractérisation du protéome urinaire.....	49
E.	Conclusion	50

II. Chapitre II : Matériel et Méthodes 52

A.	Introduction	52
----	--------------------	----

B.	Techniques de Biochimie	52
1.	Protocole préconisé pour la collecte et le prétraitement des échantillons d'urine	52
2.	Dosage des protéines urinaires.....	53
a)	Nécessité d'une normalisation	53
b)	Dosage de protéines	54
3.	Préparation des échantillons en vue de l'analyse protéomique	55
a)	Extraction des protéines urinaires	55
(1)	Précipitation à l'acide trichloroacétique.....	55
(2)	Ultrafiltration	55
b)	Fractionnement - Gel SDS PAGE	56
c)	Digestion des protéines urinaires	56
(1)	Digestion sur gel et extraction des peptides.....	56
(2)	Digestion en solution et dessalage des peptides.....	57
d)	Répétabilité des méthodes d'extraction.....	59
4.	<i>Déplétion et fractionnement des protéines urinaires</i>	60
a)	Fractionnement par isoélectrofocalisation.....	60
b)	Enrichissement Equalizer	61
5.	Analyse en spectrométrie de masse	63
a)	Reprise des échantillons peptidiques	63
b)	Chromatographie liquide	63
c)	Spectrométrie de masse	65
(1)	Analyses MS/MS	66
(2)	Analyses MS	66
d)	Contrôles qualité des analyses.....	66
(1)	Calibration du spectromètre de masse Orbitrap XL	66
(2)	Echantillons de contrôles des performances instrumentales.....	67
(3)	Suivi des analyses.....	68
C.	Traitement des données	69
1.	Méthode AMT	69
a)	Traitement des données MS/MS	70
(1)	Mascot Distiller	70
(2)	Recherche Mascot.....	70
(3)	Bases de données Decoy pour la détermination des Faux Positifs.....	71
(4)	IRMa (interprétation des Résultats Mascot).....	72
(5)	hEIDI (Exploitation et Intégration des Données)	73

b)	Traitement des données de quantification MS	74
(1)	DeconTools.....	74
(2)	Viper.....	74
2.	Qualification des données	75
3.	Analyse statistique « Abundance Index »	75
4.	Analyse des données « data mining ».....	77
D.	Conclusion.....	78
III.	Chapitre III : Travaux préliminaires	79
A.	Etude préanalytique : conditions de collecte de l'urine	79
1.	Utilisation d'additifs et stabilisateurs	80
a)	Interférences potentielles dues aux inhibiteurs de protéases	80
b)	Antibactériens.....	81
2.	Optimisation des conditions de collecte.....	82
a)	Design expérimental	82
b)	Métriques pour l'évaluation	83
c)	Résultats.....	83
B.	Standardisation	84
1.	Design de l'étude	84
2.	Observations initiales.....	87
3.	Etude comparative - AMT	88
4.	Etude comparative - SRM	90
a)	Rendement des méthodes de préparations	90
b)	Gamme dynamique.....	91
5.	Comparaisons AMT-SRM	93
C.	Conclusion.....	93
IV.	Chapitre IV : Caractérisation du protéome urinaire	96
A.	Contexte général – Biomarqueurs de pathologies	96
1.	Recherche de biomarqueurs.....	96
2.	Méthodologie protéomique	97
B.	Exploration du protéome urinaire	98
1.	Fractionnement de l'échantillon.....	99
a)	Au niveau protéique.....	99
b)	Au niveau peptidique (IEF) Hors gel.....	100
2.	Comparaison des techniques de fractionnement.....	106
C.	Base de données des protéines urinaires	114

1.	Description générale	114
2.	Comparaison avec l'état de l'art	117
3.	Intérêt de l'urine pour la découverte de biomarqueurs de pathologies	123
D.	Conclusion	124
V.	Chapitre V : Découverte de candidats biomarqueurs du cancer de la vessie dans l'urine	125
A.	Constitution d'une cohorte de patients.....	125
1.	Sélection des patients.....	126
2.	Classification des patients cancers : facteurs de risque.....	127
3.	Les cohortes multicentriques.....	128
B.	Analyse protéomique quantitative	128
1.	Identification dans la base AMT : réalignement des temps de rétention	128
2.	Quantification par spectral count dans des urines poolées.	131
3.	Analyse semi-quantitative des échantillons individuels par la méthode AMT.....	134
C.	Analyse des données pour la découverte de candidats biomarqueurs.....	135
1.	Qualification des données	135
2.	Filtrage des données	138
3.	Résultats: Quantification d'un millier de protéines dans l'urine	139
4.	Découverte de candidats biomarqueurs: statistiques différentielles.....	142
D.	Candidats biomarqueurs du cancer de la vessie.....	146
1.	Classification fonctionnelle des candidats	147
2.	Les Exosomes	148
3.	Evaluation critique de la liste de candidats	150
E.	Conclusion.....	155
	Conclusion générale	157
	Références Bibliographiques	159

Liste des Figures

Figure I.1: Représentation schématique d'un spectromètre de masse (source wikipedia)

Figure I.2 : Schéma simplifié du fonctionnement d'une source electrospray

Figure I.3: Représentation schématique d'un spectromètre de masse hybride LTQ-Orbitrap.

Figure I.4 : Stratégies de protéomique quantitative tiré de (Bantscheff, Schirle, Sweetman, Rick, & Kuster, 2007)

Figure I.5 : Méthodes de quantification absolue des protéines en utilisant trois types de standards internes PSAQ « Protein Standard Absolute Quantification », QconCAT « absolute Quantification using concatenated signature peptides », AQUA « Absolute QUAntification » (Brun et al., 2007).

Figure I.6 : L'approche AMT en 2 phases : (i) construction d'une base de données, (ii) quantification et identification des peptides.

Figure I.7 : De la découverte à la validation d'un candidat biomarqueur

Figure I.8 : Intervalle de concentration de 70 protéines du plasma, d'après N.L Anderson. (N. L. Anderson & Anderson, 2002)

Figure I.9 : Distribution des protéines représentant (A) 90% et (B) de 91-99% de la quantité totale de protéines du plasma

Figure I.10 : Nombre de protéines identifiées en fonction de leur poids moléculaire avant et après déplétion partielle des protéines majoritaire d'un pool d'urines de 8 patients (1.6 L) par la méthode Equalizer (Tiré de Castagna et al. 2005).

Figure I.11 : Diagramme de Venn illustrant le nombre de protéines identifiées à partir de 50µg de lysat de cellules obtenues après fractionnement des protéines par gel SDS-PAGE et digestion « in gel » ou un fractionnement des peptides par IEF après une digestion des protéines en solution.

Figure I.12 : Schéma descriptif du système urinaire

Figure I.13 : Classification des tumeurs de la vessie

Figure I.14 : La cystoscopie : examen visuel de la paroi vésicale

Figure I.15 : Overlap entre trois études sur la caractérisation du protéome urinaire

Figure II.1 : 1mL d'urine de patients centrifugé, ultrafiltré puis déposé en stacking sur le gel. Les différences de coloration au bleu de Coomassie illustrent les variations de concentration protéique d'un échantillon à l'autre.

Figure II.2 : profil protéique d'un échantillon digéré (piste A) et du même échantillon non digéré (piste B) sur Gel SDS-PAGE.

Figure II.3 : Exemple de corrélation des abondances entre deux répliqués analytiques (à gauche) ou techniques (à droite) R^2 est la valeur du coefficient de corrélation (Pearson).

Figure II.4 : Comparaison des distributions des coefficients de Pearson obtenues pour des répliqués analytiques et des répliqués techniques

Figure II.5 : Principe d'une stratégie visant à diminuer la dynamique de concentration d'un protéome complexe

Figure II.6 : Représentation schématique de la configuration duale du système de chromatographie liquide utilisé

Figure II.7 : Représentation schématique des gradients appliqués sur les colonnes de chromatographie liquide : gradient d'élution des peptides (colonne 1) et gradient de lavage de la colonne (colonne 2)

Figure II.8 : Suivi des temps de rétention de 5 peptides du Cytochrome C sur les deux colonnes de chromatographie liquide

Figure II.9 : Workflow de l'étude de protéomique quantitative en vue de la découverte de biomarqueurs de pathologie uro-génitale

Figure III.1 : Moyenne des rendements obtenus sur les protéines de levure ajoutées dans l'urine et estimés par SRM. Trois peptides AQUA pour l'alcool déshydrogénase, deux pour la carboxypeptidase et un pour l'énolase 1 ont été utilisés pour la quantification.

Figure III.2 : Concentration des 21 peptides issus des 18 protéines humaines (QconCAT) et des 5 peptides issus des 3 protéines de levures (AQUA), dans les trois pools d'urines préparés selon les quatre protocoles. Les peptides quantifiés ont été classés dans l'ordre décroissant de leurs concentrations estimées par SRM dans l'urine.

Figure III.3 : Design de l'étude de l'influence des conditions de collecte sur le résultat d'analyse LC-MS d'échantillons urinaires.

Figure III.4 : Carte en couleur et box plot (centré par échantillon) représentant le coefficient de Pearson de la corrélation des abondances peptidiques avec celles du contrôle pour 6 échantillons distincts traités ou non avec IP et conservés 4 heures à température ambiante ou à froid.

Figure III.5 : Design de l'étude de comparaison des protocoles de préparation des échantillons urinaires. Trois points de contrôles sous la forme de standards internes ont été établis aux étapes clé du traitement des urines afin de pouvoir estimer les performances des quatre protocoles.

Figure III.6 : A : Profil protéique sur Gel SDS-PAGE des trois pools d'urine préparés par ultrafiltration ou précipitation TCA. et B : Carte de couleur représentant le nombre de protéines group identifiées par bande de gel sur une piste « ultrafiltration » et piste « précipitation TCA ».

Figure III.7 : Carte en couleur représentant les coefficients de Pearson moyennés sur les réplicats analytiques de la corrélation des abondances peptidiques pour chaque pool d'urine préparé selon les 4 protocoles (A) et le nombre de peptides communs à chaque paire d'analyse (B).

Figure III.8 : Log de concentration des peptides issus des protéines urinaires estimées à partir des QconCAT en utilisant les plateformes AMT et SRM dans les trois pools d'urines préparés en duplicat selon le protocole standardisé (précipitation TCA – digestion en solution).

Figure IV.1 : Principe général de la stratégie GeLC-MS

Figure IV.2 : Comparaison de la couverture du protéome de levure obtenue après fractionnement sur gel ou par IEF hors-gel. Tiré de Hubner et al. 2008.

Figure IV.3 : Comparaison de la couverture du protéome urinaire après précipitation TCA et digestion en solution avant et après fractionnement par IEF hors gel.

Figure IV.4 : Gel SDS-Page des extraits protéiques obtenus après déplétion partielle des protéines majoritaires par la méthode Equalizer d'un échantillon d'urine de 8 mL à l'aide de quantités variables de billes. Piste 1 : contrôle (échantillon ultrafiltré). Pistes 2, 5, 8 : fractions non retenues (après ultrafiltration). Pistes 3, 6, 9 : lavages. Pistes 4, 7, 10 : Extraits élués des billes.

Figure IV.5 : Périmètre fonctionnel du logiciel hEIDI utilisé pour compiler les résultats de multiples identifications au sein de contextes définis par l'utilisateur. Ces contextes peuvent ensuite être comparés entre eux en termes de protéine groupes et de peptides identifiés.

Figure IV.6 : Comment le fait de regrouper les identification de plusieurs contextes permet de lever l'ambiguïté concernant des protéines appartenant initialement au même protéine groupe par l'ajout d'un peptide non-partagé.

Figure IV.7 : Diagramme de Venn représentant le recouvrement des listes de protéine groupes couverts à l'aide des trois méthodes de fractionnement ou de déplétion utilisées pour la génération de la base de donnée du protéome urinaire.

Figure IV.8 : Nombre cumulé de protéine groupes identifiés par chacune des trois méthodes de fractionnement ou déplétion utilisées pour la génération de la base de donnée urinaire en fonction du nombre de fractions analysées. Exemple de gel SDS-PAGE obtenu en vue de l'analyse GeLC-MS d'un échantillon urinaire.

Figure IV.9 : Gamme dynamique des protéines groupes quantifiés lors de la phase de découverte (voir chapitre V). Les points de couleur correspondent aux protéines groupes identifiés exclusivement par une seule méthode de fractionnement/déplétion.

Figure IV.10 : Représentation bidimensionnelle de la base de données AMT du protéome urinaire compilée dans le cadre de ce travail. Chaque point correspond à un peptide dont les coordonnées analytiques sont la masse monoisotopique et le temps de rétention normalisé. Les peptides correspondant à quelques protéines majoritaires de l'urine ont été mis en évidence par un code couleur.

Figure IV.11 : Résultat de l'analyse des termes GO associés aux protéine groupes de notre base de données du protéome urinaire regroupés par composants cellulaires, fonctions et processus biologiques.

Figure IV.12 : Stratégies de fractionnement et d'analyse suivies dans les publications de Adachi (Adachi et al., 2006) et de Kentsis (Kentsis et al., 2009)

Figure IV.13 : Résultat de l'analyse des termes GO associés aux protéines identifiées dans notre base de données du protéome urinaire et dans les études d'Adachi et Kentsis, regroupées par composants cellulaires, fonctions et processus biologiques. Les répartitions sont très similaires quelles que soient les listes.

Figure IV.14 : Diagramme de Venn traduisant le recouvrement en termes de protéines (panneau de gauche) et de peptides (panneau de droite) en comparant notre liste et celles d'Adachi et de Kentsis

Figure IV.15 : Diagramme de Venn traduisant le recouvrement en termes de protéines identifiées obtenus par notre algorithme (section B.2) entre les listes issues des différentes catégories de patients de la cohorte de découverte de candidats biomarqueurs.

Figure IV.16 : Carte du protéome urinaire obtenu dans le cadre de notre étude mettant en évidence certains candidats biomarqueurs de pathologies rénales ou uro-génitales rapportés dans la littérature.

Figure V.1 : Cohorte globale de plus de 1000 patient ayant consultés pour une suspicion de cancer de la vessie, et sous-cohorte utilisée pour la phase de découverte, comprenant 68 patients avec confirmation de cancer de la vessie et 30 donneurs sains.

Figure V.2 : Design initialement anticipé pour la cohorte de découverte (A) et composition de la cohorte réellement analysée (B).

Figure V.3 : Répartition des patients de la cohorte de découverte suivant leur catégorie (contrôles, incident ou prévalent), leur facteur de risques de progression ou de récurrence et leur origine géographique

Figure V.4 : Algorithme de correction des temps de rétention des peptides dans la base de données constituée lors de la première phase de l'approche AMT

Figure V.5 : Aligement des temps de rétention des peptides détectés dans l'analyse quantitative avec ceux de la base de données AMT avant et après correction des temps de rétention des peptides de la base de données.

Figure V.6 : Cluster de protéines dont l'abondance montre une corrélation à l'augmentation du risque de récurrence/progression.

Figure V.7 : Box plot tracés à partir des valeurs d'abondances mesurées pour chacun des 98 patients de la cohorte, regroupés par catégorie et par risque, dans les trois réplicats analytiques. La zone blanche correspond au deux quartiles centraux, et les bandes des couleurs aux deux quartiles extrêmes.

Figure V.8 : (Panneau supérieur) Box plot des coefficients moyens de Pearson entre les abondances des peptides d'une des trois analyses correspondant à un patient donné et celles des deux autres analyses du même patient, triés par ordre croissant. (Panneau inférieur) Nombre de peptides quantifiés chez un patient donné pour les trois analyses.

Figure V.9 : (Panneau supérieur) Distribution des coefficients de variation calculés sur l'ensemble des peptides d'un patient donnée en considérant l'ensemble des réplicats analytiques. (Panneau inférieur) Distribution des coefficients de variation de ces mêmes peptides en considérant les réplicats deux à deux et en ne conservant, pour chaque peptide chez un patient donné, que les deux points de données présentant le CV le plus faible.

Figure V.10 : Coefficient de Variation selon les permutations deux à deux des trois réplicats analytiques en ayant préalablement éliminé les valeurs manquantes (ONA)

Figure V.11 : Carte en couleur des abondances moyennes des 4648 peptides quantifiés sur l'ensemble des échantillons des 98 patients de la cohorte, classées selon leur abondance moyenne dans le panneau supérieur et occurrence décroissante dans le panneau inférieur.

Figure V.12 : Contribution des protéines majoritaires quantifiées dans les extraits protéiques urinaires

Figure V.13 : Distribution des valeurs d'abondance index calculées à partir des comparaisons des patients incidents et contrôles (panneau supérieur) et des patients prévalents et contrôles (panneau inférieur)

Figure V.14 : Protéines quantifiées classées par abondances médiane décroissante sur l'ensemble des patients de la cohorte de découverte en mettant en évidence les candidats biomarqueurs sur et sous abondants chez les patients cancers

Figure V. 15 : Classification fonctionnelle des termes GO Slim des candidats biomarqueurs plus montrant une augmentation de leur abondance chez les patients cancers par processus biologique, localisation cellulaire et fonctions moléculaire.

Figure V.16 : Graphe d'interaction de nos candidats biomarqueurs surabondants chez les patients cancers (panneau de gauche) et surabondants chez les contrôles (panneau de droite). En rouge, les protéines d'origine exosomales.

Figure V.17 : Protéine APOA1 : profil d'abondance en log de la protéine chez les patients contrôles, cancers incident associés au risque de progression tumoral déterminé dans notre étude quantitative (panneau supérieur) ; Vérification de l'abondance différentielle de cette protéine dans l'étude de Chen par Western blot dans les urines de patients contrôle, de bas grade stade précoce, haut grade stade précoce et stade avancé (panneau inférieur).

Figure V.18 : Concentration de l'APOA1 déterminée par l'analyse iSRM sur des urines de patients contrôles sains et urologiques et de patients atteint d'un cancer incident en fonction du risque de progression tumoral (panneau supérieur) ; visualisation du marquage de cette protéine sur le site HPA sur des coupes de tissu d'urothélium normal et cancer (panneau inférieur).

Figure V.19 : Profil d'abondance de la protéine S100P chez les patients contrôles, cancers incident associés au risque de progression tumoral déterminé dans notre étude quantitative (panneau de gauche) ; visualisation du marquage de cette protéine sur le site HPA sur des coupes de tissu d'urothélium normal et cancer (panneau de droite).

Figure V.20 : Concentration de LAMP1 déterminée par l'analyse iSRM sur des urines de patients contrôles sains et urologiques et cancer incident et prévalent en fonction du risque de progression tumoral.

Liste des Tableaux

Tableau I.1 : Comparaison des performances des différents analyseurs (FWHM : Full Width at Half Maximum)

Tableau I.2 : Groupes de risques des tumeurs superficielles de vessie

Tableau II.1 : Paramètre d'acquisition pour l'analyse MS/MS

Tableau II.2 : Extrait de la liste des analyses LC-MS réalisée sur les 98 patients de la cohorte découverte

Tableau II.3 : Paramètres de recherche Mascot

Tableau II.4 : Filtres utilisés dans le logiciel IRMa

Tableau II.5: Classification des peptides en fonction de l'analyse statistique par « Abundance Index ». La médiane d'AbI est calculée sur la base des valeurs d'AbI observés, σ sur la base des permutations.

Tableau II.6: Classification des protéines selon l'analyse statistique par « Abundance Index »

Tableau III.1 : Protéines identifiées lors de l'analyse de l'inhibiteur de protéases dans de l'eau distillé. (Contaminants grisés)

Tableau III.2. Tableau récapitulatif de la composition des 3 pools utilisés pour tester les protocoles de préparation des urines en vue de l'analyse protéomique.

Tableau III.3. Liste des 3 protéines de levures ajoutées aux échantillons urinaires pour l'estimation des rendements d'extraction.

Tableau III.4. Protéines urinaires représentées dans le standard QconCAT et liste des peptides sélectionnés.

Tableau IV.1 : design de l'optimisation de la quantité de billes Equalizer pour la déplétion des protéines à partir d'un faible volume d'urine.

Tableau IV.2 : Exemple simplifié de comparaison entre deux listes de protéines obtenues par des méthodes de fractionnement différentes.

Tableau IV.3 : résumé de caractéristiques des trois techniques comparées dans le cadre de la caractérisation du protéome urinaire par différentes techniques de fractionnement.

Tableau IV.4 : Nombre d'AMT et couverture de séquence de quatre protéines majoritaires de l'urine.

Tableau IV.5 : Principaux enrichissement en termes GO, basé sur une analyse par le logiciel GORILLA, de notre base de données du protéome urinaire comparée au protéome humain

Tableau V.1 : Protéines représentant le cluster dont l'abondance montre une corrélation à l'augmentation du risque de récurrence/progression

Tableau V.2 Classement des peptides en fonction des résultats de l'analyse statistique par Abondance Index.

Tableau V.3 Nos candidats biomarqueurs classés par ordre décroissant des médianes des valeurs d'abondance index de leurs peptides protéospcifiques dans les comparaisons incident / contrôle et prévalent / contrôle

Tableau V.4 Sélection de nos candidats biomarqueurs qui présentent une abondance différentielle dans les néphropathies à IgA

Introduction générale

Le cancer de la vessie représente le 4^{ième} cancer en Europe chez les hommes. Dans la majorité des cas, les primo tumeurs sont des tumeurs non infiltrantes avec un bon pronostic et qui sont traitées facilement par une résection. Cependant, dans 60% des cas, elles récidivent en tumeur infiltrant le muscle et représentent un caractère beaucoup plus agressif avec un risque de métastases. Il est donc nécessaire de détecter toute récurrence le plus tôt possible. Pour cela, les patients traités pour une tumeur vésicale sont contrôlés régulièrement dans les cinq années qui suivent la résection. A l'heure actuelle, le diagnostic d'un cancer de la vessie est établi à l'issue d'une cystoscopie, et d'une cytologie urinaire. La cystoscopie est un examen très invasif par imagerie médicale au cours duquel un tube doté d'une lentille est introduit dans la vessie par l'urètre. Cette méthode s'est établie comme la référence en termes de sensibilité et de spécificité. Cependant, les petites tumeurs sont difficilement détectables par cette technique que l'on associe souvent à un examen biologique, la cytologie urinaire, afin de rechercher la présence éventuelle de bactéries et de cellules cancéreuses. La cytologie urinaire est, quant à elle, très spécifique, mais peu sensible. Il est donc important, pour le suivi des patients, de trouver de nouvelles méthodes de surveillance sensibles et spécifiques qui réduiraient la contrainte pour le patient et le coût associés à la cystoscopie.

La recherche de biomarqueurs par protéomique a bénéficié de progrès spectaculaire dans le domaine de la spectrométrie de masse. De nouveaux instruments rapides et sensibles sont de mieux en mieux adaptés aux défis posés par l'analyse d'échantillons protéiques complexes. Il est ainsi désormais possible d'accéder à une grande variété de protéines sur une large gamme dynamique de concentration dans un temps d'analyse de l'ordre de l'heure. Cependant, même si la gamme dynamique obtenue avec ce type d'instrument est étendue, cela ne suffit pas pour identifier en une seule analyse l'ensemble des protéines issues d'un échantillon aussi complexe qu'un fluide biologique. Pour cela, on utilise des techniques de fractionnement afin de réduire la complexité de ces échantillons avant leur analyse par spectrométrie de masse. Ces techniques permettent de séparer les protéines ou les peptides en différentes fractions, ou d'extraire partiellement les protéines majoritaires, afin d'accéder aux protéines minoritaires qui sont autant de biomarqueurs potentiels. En parallèle avec ces évolutions expérimentales, des évolutions méthodologiques dans le domaine de la protéomique clinique facilitent le développement de nouveaux biomarqueurs. Ainsi, de nouvelles approches de protéomique quantitatives permettent de comparer les protéomes de plusieurs états physiologiques distincts en mesurant l'apparition, la disparition ou les variations des signaux correspondants aux protéines détectées. En particulier, la stratégie Accurate Mass and Time (AMT) tags couplée à la quantification « label free » suscite un grand intérêt car elle rend possible la comparaison d'échantillons provenant de larges cohortes de patients requises par les études cliniques. Cette approche est une technique de choix pour la découverte de biomarqueurs puisqu'elle permet dans un premier temps de caractériser en profondeur le protéome d'échantillons complexes tels que les fluides biologiques ou les

tissus ; et ensuite de quantifier une grande variété de ces protéines dans des échantillons non fractionnés, approchant ainsi le débit d'analyse des technologies de type profilage protéique comme par exemple le SELDI.

Les cellules relarguent fréquemment des protéines ou des fragments protéiques dans leur micro-environnement. Ces composés sont donc susceptibles de passer du tissu dans les fluides biologiques adjacents. En particulier, il a été montré que les tumeurs sécrètent certains facteurs comme des ARN ou des protéines qui leur sont spécifiques. Ainsi, des protéines d'origine tissulaire retrouvées dans les fluides pourraient refléter l'état physio-pathologique d'un individu. Dans le cas du cancer de la vessie, c'est l'urine qui est directement en contact avec les cellules épithéliales, on parle de fluide proximal. Les protéines relarguées par les tumeurs vésicales sont ainsi susceptibles d'être détectées dans l'urine. De plus, l'urine est un fluide qui peut être collecté de manière non invasive et en grande quantité. L'urine représente donc un fluide idéal pour la recherche de nouveaux biomarqueurs du cancer de la vessie.

Partant de ces constats, nous avons entrepris le travail présenté dans ce manuscrit qui avait pour but de caractériser en profondeur le protéome urinaire et d'y rechercher de nouveaux candidats biomarqueurs du cancer de la vessie et qui s'inscrit dans le cadre d'un projet Européen (DECanBIO) coordonné par notre laboratoire. Dans le premier chapitre, un panorama de la bibliographie reprend les principes généraux de la protéomique et le contexte clinique lié au cancer de la vessie et à la découverte de nouveaux biomarqueurs dans les urines. Dans le deuxième chapitre sont regroupés l'ensemble des méthodes et modes opératoires utilisés pour le traitement biochimique des échantillons d'urines, l'analyse protéomique par chromatographie et spectrométrie de masses, et le traitement des données générées à l'issue des analyses protéomiques qualitative et quantitative. Dans le troisième chapitre, nous verrons comment nous avons défini précisément les conditions optimales de collectes et mis en place un protocole standardisé du traitement des urines en vue de l'analyse protéomique. Dans le quatrième chapitre, nous aborderons la caractérisation en profondeur du protéome urinaire à travers la mise en œuvre de différentes stratégies de fractionnement et de déplétion. Ce chapitre se termine avec la génération de notre base de données de protéines urinaires qui est la fondation de l'étude quantitative présentée dans le cinquième et dernier chapitre. Dans celui-ci, nous aborderons les étapes de l'analyse protéomique comparative réalisée sur une cohorte de 98 patients et qui a permis d'aboutir à une liste finale de 97 candidats biomarqueurs. Finalement, nous discuterons de l'intérêt de cette liste en examinant les liens entre les protéines candidates et le cancer de la vessie. Cet examen nous montrera que l'urine, fluide biologique déchet de l'organisme, présente un potentiel remarquable pour le diagnostic, non seulement du cancer de la vessie, mais aussi d'autres pathologies uro-génitales. Ce travail représente ainsi un espoir certain pour la détection précoce des primo cancer mais aussi des récurrences du cancer de la vessie.

I. Chapitre I : Etat de l'art

Ce chapitre est une étude bibliographique composée de 4 parties présentant les principaux outils et méthodologies de l'analyse protéomique, leurs applications cliniques, le contexte clinique de ce travail et enfin l'état de l'art dans le domaine de la caractérisation du protéome urinaire.

La première partie décrit la genèse de la protéomique et les différentes méthodes utilisées pour identifier les protéines, du séquençage d'Edman jusqu'aux dernières avancées dans le domaine de la spectrométrie de masse. Puis, sont introduites les différentes méthodes de quantification des protéines ou des peptides. Dans la seconde partie, sont présentés les différents aspects de la protéomique quantitative : (1) la définition de biomarqueurs, (2) l'intérêt de la caractérisation et de la quantification de fluides biologiques dans le cadre de la recherche de nouveaux biomarqueurs, (3) la nécessité de fractionner des échantillons très complexes tels que les fluides biologiques et (4) les méthodes disponibles pour réaliser ces fractionnements au niveau peptidique ou protéique. La troisième partie présente le contexte clinique de ce travail de thèse, l'intérêt de découvrir de nouveaux biomarqueurs dans le cadre du suivi de patients ou dans le cadre du diagnostic, ainsi que le potentiel que représente l'urine dans la recherche de biomarqueurs pour différentes pathologies urinaires et les défis analytiques associés. Pour finir, la dernière partie présente différentes études préalables de caractérisation du protéome urinaire avec leurs intérêts et leurs limites.

A. La protéomique

1. Des gènes aux protéines

Les protéines sont des macromolécules impliquées dans une grande variété de processus biologiques. Elles sont synthétisées dans le cytosol des cellules vivantes par des organelles dédiées, les ribosomes. Dans le noyau, les gènes sont d'abord transcrits en acides ribonucléiques messagers (ARNm) qui servent de modèle pour la synthèse des protéines lors du processus de traduction. Le ribosome décode l'ARN messager et assemble les acides aminés codés par les bases nucléotidiques pour créer la protéine correspondante. Les acides aminés ont des propriétés physico-chimiques très différentes (hydrophobes, polaires, acides ou basiques) qui définissent, en fonction de leur enchaînement, les propriétés des protéines.

Le terme protéine est couramment employé quand le nombre de résidus est environ supérieur à 100, alors que le terme peptide pour un nombre de résidus inférieur à 50. Dans la zone intermédiaire, les termes de peptides, de petites protéines ou de polypeptides sont employés. La taille d'une protéine est très variable, et peut aller jusqu'à plusieurs milliers d'acides aminés. Les protéines peuvent être décrites, dans un premier temps, par leur séquence d'acides aminés ordonnée de leur extrémité N-terminale à leur extrémité C-terminale, et qui correspond à leur structure primaire. Les protéines adoptent une structure tridimensionnelle

qui est critique pour leur activité biologique. Ces différentes conformations permettent leur interaction avec diverses molécules telles que d'autres protéines ou des cofacteurs. Du fait de la présence de groupes acides et basiques des acides aminés, les protéines sont des composés amphotères. La charge portée par une protéine dépendra donc du nombre de résidus acides et de résidus basiques ainsi que du pH de la solution dans laquelle elle se trouve. Le point isoélectrique d'une protéine (pI), est défini comme étant le pH pour lequel la charge globale de la protéine est nulle.

Les protéines sont composées de domaines qui ont divers rôles fonctionnels très spécifiques: activité catalytique, reconnaissance d'un substrat, interaction protéine-protéine ou encore protéine-ADN. La production de protéines par la cellule ne consiste pas uniquement en une traduction de la séquence d'acide nucléique ; des modifications post-traductionnelles telles que des acétylations, des phosphorylations et des glycosylations augmentent également la diversité des protéines et donc de leurs activités biologiques. Les protéines constituent donc un ensemble diversifié et complexe de macromolécules impliquées dans une grande variété de mécanisme biologiques

2. Le protéome et la protéomique

Le protéome représente l'ensemble des protéines d'un système biologique (cellule, organites, tissu, organe ou organisme) à un moment donné et sous des conditions données. Ce terme, introduit pour la première fois par Mark Wilkins en 1994 au congrès de Sienne en Italie, est issu de la fusion des termes «protéines » et « génome ».

Depuis 2004, on connaît la quasi-totalité du génome humain, soit environ 22 000 gènes, grâce notamment au projet « Projet Génome Humain » (International Human Genome Sequencing Consortium, 2004). Celui-ci a constitué une ressource essentielle pour aborder l'étude des réseaux d'interactions fonctionnelles qui commandent les processus biologiques normaux et pathologiques. Néanmoins, la caractérisation du génome humain n'a pas permis d'appréhender l'entière complexité des systèmes biologiques. Ceci s'explique en partie par le fait que les véritables unités fonctionnelles de la cellule sont les protéines. L'exemple de la chenille et du papillon illustre la différence fondamentale qui existe entre les niveaux génomique et protéomique. Ces deux organismes, apparemment différents, ont exactement le même génome mais des protéomes distincts. Il est donc important, pour comprendre un organisme, d'étudier son génome mais aussi son protéome.

Alors que le génome est identique dans toutes les cellules d'un organisme quel que soit leur environnement ; le protéome est, quant à lui, une entité dynamique et complexe. Au sein de chaque cellule, le contenu protéique évolue en permanence en fonction des conditions intra ou extra cellulaires. Le protéome contient un nombre beaucoup plus important de protéines que le génome contient de gènes. On estime aujourd'hui qu'il existerait environ un million de protéines chez l'homme (Gaudet et al., 2013). En 2011, le

projet international HPP (Human Proteome Project), à l'initiative de l'organisation HUPO (Human Proteome Organisation), a émergé dans le but de caractériser l'ensemble des protéines humaines à partir de l'analyse des fluides biologiques, des tissus et des organes constituant le corps humain. Il reste à l'heure actuelle, encore 35% des protéines prédites qui n'ont jamais été observées.

3. L'analyse Protéomique

Les principaux objectifs de l'analyse protéomique sont d'identifier et de mesurer l'abondance des protéines présentes dans un échantillon biologique à un instant donné et sous des conditions définies. Sur la base de ces informations, on cherche à dériver des informations fonctionnelles sur ces protéines, telles que leur localisation, l'identification de leurs partenaires ou encore la détermination de leurs modifications post-traductionnelles. Ces informations permettent de mieux comprendre les mécanismes moléculaires impliqués dans les grandes fonctions cellulaires ou dans l'apparition de pathologies comme le cancer.

Les techniques dédiées au séquençage des protéines ont fortement évolué au cours de ces 20 dernières années. Jusqu'au début des années 90, l'électrophorèse bidimensionnelle et le séquençage par la dégradation d'Edman étaient les seules techniques disponibles. Les évolutions technologiques en spectrométrie de masse permettent aujourd'hui d'identifier des milliers de protéines en une seule analyse et donc d'envisager la caractérisation de protéomes extrêmement complexes.

a) *De la chimie des protéines à la protéomique*

La technique de l'électrophorèse en gel de polyacrylamide (PAGE) bidimensionnelle (2DE) (O'Farrell, 1975), (Klose, 1975) et (Scheele, 1975) permet la séparation de mélanges protéiques complexes. Le principe du 2DE repose sur la combinaison d'une séparation des protéines selon leur charge (isoélectrofocalisation) et selon leur taille (migration sur gel SDS PAGE). Ainsi, O-Farrell a démontré la capacité de séparer plus d'un millier de protéines de la bactérie *Escherichia coli*. Ces travaux constituent un premier jalon important vers la caractérisation de protéomes.

Si cette technique permet de détecter un grand nombre de protéines simultanément en estimant leur masse et leur point isoélectrique, elle est difficilement automatisable et donc peu reproductible. De plus, il est difficile de détecter des protéines faiblement abondantes, celles de points isoélectriques extrêmes, de haut poids moléculaire ou très hydrophobes comme celles localisées au niveau des membranes. Cette dernière classe de protéines représente environ un tiers des protéines du protéome humain prédit sur la base du génome et leurs fonctions sont essentielles puisqu'elles assurent la liaison entre les milieux extracellulaire et intracellulaire. Plus de deux tiers de ces protéines sont des cibles de médicaments (Shukla, Vaitiekunas, & Cotter, 2012).

L'utilisation de l'électrophorèse mono-dimensionnelle a permis de contourner une partie de ces difficultés en s'affranchissant de l'étape d'isoélectrofocalisation au cours de laquelle les protéines ont

tendance à précipiter et ainsi de conserver les protéines hydrophobes, de haut poids moléculaire ou de points isoélectriques extrêmes (Edmondson et al., 2002).

Si l'électrophorèse mono ou bidimensionnelle est utile pour séparer les protéines, elle ne permet pas de les identifier. Jusqu'au début des années 90, l'identification d'une protéine inconnue reposait sur le séquençage de quelques acides situés en N-terminal par la méthode d'Edman. Cette méthode, apparue dans les années 50 (EDMAN, 1949), consiste à dériver le N-terminal de la protéine par le phényl isothiocyanate (PITC), à rompre la liaison entre le premier acide aminé et le reste de la chaîne polypeptidique et à déterminer sa nature par chromatographie liquide haute performance (CLHP). C'est une technique longue (une protéine par jour) et peu sensible (20 pmol par protéine) qui permet d'identifier les protéines majeures d'un échantillon séparé sur un gel 2DE. De plus, le séquençage d'Edman ne permet pas de séquencer des chaînes peptidiques de plus d'une cinquantaine d'acides aminés, or, de nombreuses protéines comptent plusieurs centaines d'acides aminés. Pour pallier à ce problème, Sanger a proposé de digérer la protéine d'intérêt avec des protéases et de séquencer par la méthode d'Edman les fragments ainsi obtenus après leur séparation chromatographique. Les fragments chevauchants étaient ensuite utilisés pour reconstruire la séquence de la protéine entière (SANGER and THOMPSON, 1953). Sur la base de ces travaux, de nouvelles voies analytiques pour le séquençage des peptides ont été explorées, et en particulier la spectrométrie de masse.

b) Spectrométrie de masse, un outil de choix pour la protéomique

(1) Principes de la technique

La spectrométrie de masse est une méthode d'analyse physico-chimique basée sur la séparation en phase gazeuse de particules chargées (ions) en fonction de leur rapport masse/charge (m/z). Un spectromètre de masse est schématiquement composé de trois parties : la source qui sert à volatiliser et ioniser les composés d'intérêt, l'analyseur qui permet de collecter et de séparer les ions selon leur rapport m/z et le détecteur qui transforme les ions en un courant électrique qui est enregistré pour être traité afin de déterminer leur rapport m/z et leur intensité.

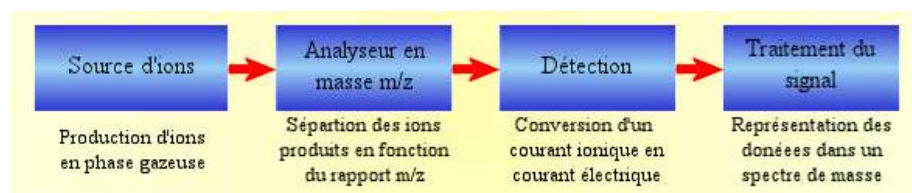


Figure 1.1: Représentation schématique d'un spectromètre de masse (source Wikipédia)

Les caractéristiques principales d'un spectromètre de masse sont sa gamme de mesure, sa vitesse d'acquisition, sa résolution, sa précision et sa sensibilité. La résolution est la capacité à séparer des composés

de masses très proches. Elle est donnée par le rapport $m/\Delta m$, où Δm est par convention la largeur de pic à mi-hauteur. La gamme de rapport m/z correspond aux limites minimales et maximales de m/z au-delà desquelles l'analyseur ne peut plus mesurer le signal. La sensibilité correspond à la quantité d'échantillon nécessaire pour détecter un signal (interprétable) dans le spectre de masse. L'intensité du pic est comparée à l'intensité du bruit de fond et un ion est considéré détectable pour une valeur de signal/bruit (S/N) supérieur à 3. La vitesse d'acquisition (ou vitesse de balayage) correspond au temps que va mettre l'analyseur pour établir un spectre pour une gamme de m/z donnée. La précision de mesure de masse correspond à l'écart relatif entre la masse mesurée d'un composé (M_{exp}) et sa masse théorique ($M_{théo}$) ; elle s'exprime en partie par million (ppm).

i. Source d'ionisation

Jusqu'au début des années 90, il était très difficile, voire impossible, d'analyser directement des protéines ou même des peptides par spectrométrie de masse en raison du caractère thermiquement labile et non volatil de ces composés. Ce n'est qu'à partir de la découverte des méthodes d'ionisation douce que sont l'ionisation par électrospray (Fenn, Mann, Meng, Wong, & Whitehouse, 1989) et l'ionisation laser assistée par matrice (Tanaka et al., 1988), (Karas & Hillenkamp, 1988) que de tels échantillons ont pu être analysés. Ces découvertes ont valu à leurs auteurs, John Fenn et Kiochi Tanaka, le prix Nobel de chimie en 2002.

Une source d'ionisation permet de vaporiser les composés d'intérêt et de les ioniser en vue de leur analyse en masse. La source d'ionisation la plus couramment utilisée pour analyser des mélanges peptidiques complexe en couplage avec la chromatographie liquide est l'électrospray (ESI).

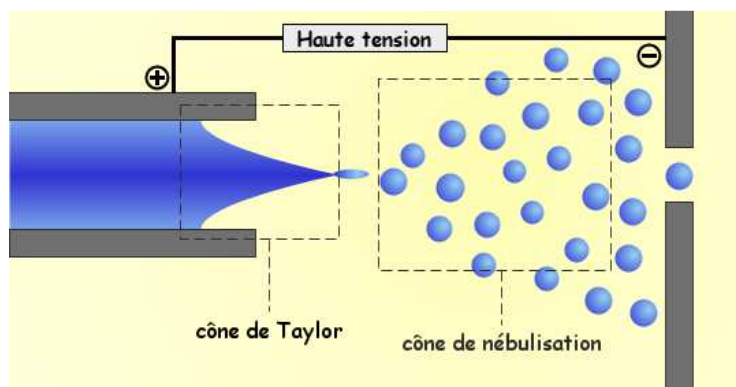


Figure I.2 : Schéma simplifié du fonctionnement d'une source électrospray

L'ESI est une technique d'ionisation douce, capable d'ioniser des biomolécules en phase liquide tels que les peptides ou protéines. C'est un mode d'ionisation à pression atmosphérique, impliquant l'application d'un potentiel électrique de quelques kV sur un échantillon en solution émergeant d'un tube capillaire à un faible débit (de l'ordre du $\mu\text{L}/\text{min}$). Ce champ provoque une accumulation de charges à la surface du liquide situé à l'extérieur du capillaire qui va se rompre pour former des gouttelettes fortement chargées.

L'évaporation du solvant contenu dans ces gouttelettes va provoquer leur rétrécissement jusqu'au moment où les forces coulombiennes répulsives vont approcher le niveau des forces de cohésions des gouttes et provoquer leur dissociation. Ces gouttelettes subissent une cascade de fissions donnant des gouttelettes de plus en plus petites, jusqu'au moment où le champ électrique en surface devient suffisant pour provoquer la désorption des ions. Ce mode d'ionisation permet de coupler au spectromètre de masse un système de chromatographie liquide, pour la séparation des peptides et le dessalage de l'échantillon biologique, ce qui améliore l'efficacité d'ionisation.

ii. Les analyseurs

Les analyseurs permettent de discriminer les ions provenant de la source d'ionisation en fonction de leur rapport masse/charge. Différents types d'analyseurs sont utilisés en protéomique: on peut citer en particulier, les analyseurs quadripolaires et les trappes ioniques qui sont des analyseurs basse résolution et les analyseurs à temps de vol ou les analyseurs à transformée de Fourier (FT-ICR ou Orbitrap) qui sont des analyseurs à haute résolution. Les caractéristiques importantes des analyseurs pour l'analyse protéomique sont : la vitesse d'acquisition afin de maximiser le nombre de peptides détectés ; la résolution qui permet de séparer des peptides de masses proches ; et la précision de mesure de masse qui améliore la spécificité des identifications. Les performances de quelques analyseurs les plus couramment utilisés sont résumées dans le Tableau I.1.

Tableau I.1 : Comparaison des performances des différents analyseurs (FWHM : Full Width at Half Maximum)

Analyseur	Limite en rapport m/z	Résolution (FWHM)	Exactitude en masse (ppm)	Vitesse d'acquisition
Quadripôle	4000	3 000	200	10 - 100 ms
Piège à ions	6000	5 000	200	100 - 300 ms
Temps de vol	Illimitée	5 000-20 000	20-200	1-10 ms
Orbitrap	4000	30 000-100 000	<5	0.3 – 1 s

Comme illustré dans le Tableau I.1, chacun de ces instruments possède des caractéristiques et des performances analytiques qui lui sont propres. Ceci a conduit les constructeurs à développer des instruments hybrides tels que les Quadruple-Temps de vol ou trappe ionique – Orbitrap permettant d'allier les

performances de différents analyseurs. Ces instruments sont particulièrement bien adaptés pour des analyses protéomiques d'échantillons complexes et de large gamme dynamique (e.g. fluide biologique). Pour toutes ces raisons, ce type d'instrument est extrêmement répandu dans la communauté protéomique.

Dans le cadre de cette thèse nous avons utilisé un instrument hybride (Figure I.3) associant une trappe ionique et un Orbitrap. Il combine les avantages de l'Orbitrap (très haute précision de mesure de masse et résolution) et d'une trappe linéaire (vitesse d'acquisition et très grande sensibilité).

L'Orbitrap est un piège à ion électrostatique développé en 1999 (Makarov, 2000), aussi appelé trappe orbitale à transformée de Fourier. Il se compose d'une électrode creuse, à l'intérieur de laquelle est placée coaxialement une électrode en forme de fuseau. Les ions sont injectés tangentiuellement à l'électrode centrale et piégés autour d'elle par la force électrostatique qui compense les forces centrifuges. Le mouvement des ions se décompose alors en un mouvement circulaire autour de l'électrode centrale dans le plan (xy) et un mouvement oscillatoire de va-et-vient selon l'axe z. A chaque passage devant les électrodes, les ions vont générer un courant induit qui est détecté et enregistré. Un calcul par transformée de Fourier permet de calculer la fréquence axiale des ions, inversement proportionnelle à la racine carrée du rapport m/z .

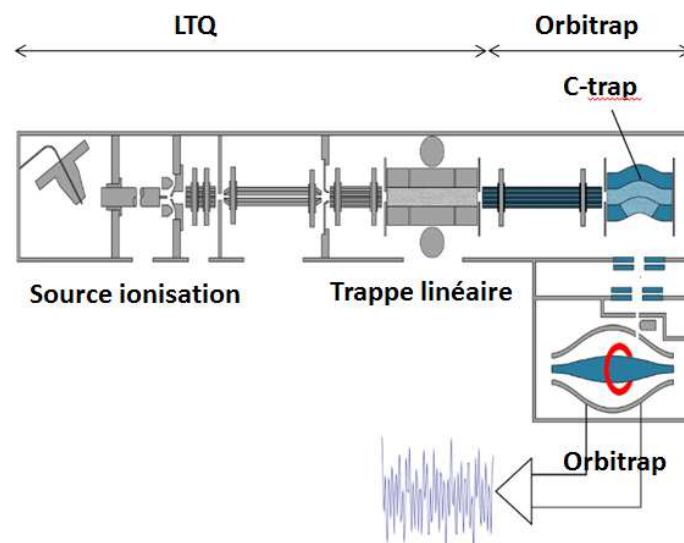


Figure I.3 : Représentation schématique d'un spectromètre de masse hybride LTQ-Orbitrap.

De manière à optimiser l'éjection des ions vers la trappe linéaire, les peptides ionisés dans la source sont transférés dans la trappe linéaire (LTQ) en passant par deux multipôles (quadropôle et octupôle). Une fois les peptides capturés dans la trappe, les ions sont soumis à des champs électriques additionnels qui permettent d'isoler un peptide particulier et de le fragmenter pour mesurer les masses de ses fragments et remonter ainsi à sa séquence en acides aminés. Les ions d'intérêts sont stockés dans la trappe avant d'être

transférés vers la C-Trap où ils vont perdre leur énergie cinétique sous l'action de collisions avec un gaz (azote), avant d'être éjectés vers l'Orbitrap en appliquant une augmentation rapide du voltage sur les électrodes. Les ions sont éjectés tangentiellement dans l'Orbitrap avec une énergie cinétique de quelques keV et se mettent à osciller en orbite autour de l'électrode centrale dans les directions axiales et radiales sous l'action des tensions générées par le champ électrique. Grâce au vide poussé qu'il y a dans l'Orbitrap et à une éjection rapide depuis la C-Trap, les ions sont stables pendant plusieurs secondes dans l'Orbitrap ; ce qui permet d'atteindre des résolutions très élevées et une excellente précision en masse.

Même si la gamme dynamique obtenue avec ce type d'instrument est étendue, cela ne suffit pas pour identifier l'ensemble des protéines issues d'échantillons très complexes comme des tissus ou des fluides biologiques. Pour cela, en amont de l'analyse par spectrométrie de masse, on utilise des techniques de fractionnement afin de réduire la complexité de ces échantillons. Ces techniques permettent de séparer les protéines ou les peptides en différentes fractions et ainsi d'analyser des échantillons moins complexes. La chromatographie liquide est la méthode classique utilisée en amont d'un spectromètre de masse pour séparer les protéines et plus généralement les peptides issus de la digestion enzymatique des protéines.

iii. La spectrométrie de masse en tandem

Il est possible d'associer plusieurs analyseurs en série afin de réaliser une analyse de spectrométrie de masse en tandem (MS/MS), ce qui permet, en particulier, d'obtenir la séquence en acide aminé d'un peptide et d'en déduire l'identité de la protéine associée. La spectrométrie de masse en tandem consiste à isoler un type d'ions qui est ensuite fragmenté et à acquérir un spectre de ses fragments. Selon les analyseurs utilisés, il existe différentes configurations. Dans un premier cas, appelé MS en tandem dans l'espace, les peptides sont sélectionnés dans un premier analyseur, fragmentés dans une chambre de collision et analysés dans un second analyseur avant d'être détectés. Dans une seconde configuration, appelée tandem dans le temps, comme c'est le cas dans la trappe ionique, les peptides sont piégés puis sélectionnés et fragmentés au même endroit pour être ensuite mesurés sur les propres détecteurs de la trappe. Dans le cas d'un instrument hybride (LTQ Orbitrap), les ions fragments peuvent alternativement être mesurés dans l'Orbitrap.

Les modes d'acquisition MS/MS impliquent une activation des ions de laquelle va résulter la fragmentation des molécules. Il existe différents moyens d'activation. Parmi les plus utilisés, on compte : le CID (collision induced dissociation), l'ETD (electron transfert dissociation) ou l'ECD (electron capture dissociation). La fragmentation CID est la méthode la plus répandue pour dissocier les ions et obtenir la séquence des peptides. Son principe repose sur la collision des ions avec des molécules de gaz neutre (argon, azote, hélium). Quand l'énergie de collision est assez haute pour convertir une partie de l'énergie cinétique en énergie vibrationnelle, la fragmentation de l'ion sélectionné est obtenue. On obtient alors le spectre des ions produits résultant de la fragmentation. Les masses des ions qui ont été fragmentés et mesurés sont

comparés avec des listes de masses issues de spectres de fragmentation théoriques de protéines présentes dans les banques de données via des moteurs de recherches.

(2) Approches protéomiques par spectrométrie de masse

Deux approches protéomiques par spectrométrie de masse peuvent être utilisées. La première approche est connue sous le terme anglo-saxon « Top-down » (NL. Kelleher, 2004). Les ions de la protéine intacte sont introduits en phase gazeuse et sont fragmentés et analysés dans le spectromètre de masse, fournissant la masse moléculaire de la protéine ainsi que des portions de sa séquence par les ions fragments. Ces informations peuvent être utilisées pour déduire la structure primaire de la protéine. Ainsi, l'identification se fait par recherche de séquences homologues pour des échantillons inconnus dans les bases de données. La deuxième approche est de type « Bottom-up ». L'échantillon subit au préalable une hydrolyse enzymatique des protéines avant leur analyse par spectrométrie de masse. Deux possibilités s'offrent alors : soit l'échantillon est analysé par MS simple et les protéines sont identifiées par empreinte peptidique massique, soit l'échantillon est analysé par MS/MS. Dans ce cas, les spectres de fragmentation des peptides issus de la digestion enzymatique de l'échantillon permettent l'identification précise des protéines par déduction de sa structure primaire ainsi que de ses modifications post-traductionnelles. Le mélange de peptides obtenus par hydrolyse chimique ou enzymatique d'une protéine est caractéristique de cette protéine. Ainsi, un des principaux avantages de l'approche « Bottom-up » est d'augmenter la précision en masse et la sensibilité pour des ions de faibles poids moléculaires, comparativement aux ions issus de protéines intactes. De plus, la spécificité et la fiabilité de cette approche peuvent être améliorées en augmentant le nombre de peptides mesurés, la précision en masse, et également en utilisant plusieurs enzymes présentant des spécificités différentes.

(3) Couplage chromatographie liquide et spectrométrie de masse en tandem

Lorsque l'analyse par spectrométrie de masse est réalisée par nébulisation directe d'un mélange peptidique, il arrive souvent que des problèmes de compétition à l'ionisation et d'augmentation générale du bruit de fond ne laissent accessible à l'analyse que les peptides les plus intenses du mélange. Par contre, dans une configuration où le spectromètre de masse est couplé à la chromatographie liquide, chaque peptide séparé par la colonne capillaire arrive dans la source du spectromètre à un temps de rétention qui lui est propre. Ainsi, les peptides les moins intenses, séparés des autres, deviennent accessibles à l'analyse. Un grand nombre de peptides peut ainsi être analysé ; ce qui permet d'obtenir un gain d'information de séquence très significatif. En chromatographie liquide, la méthode classique consiste à séparer les peptides sur une colonne de phase inverse. Ce type de colonne permet de dessaler l'échantillon et est compatible avec la spectrométrie de masse. Ce mode de fractionnement repose sur l'interaction hydrophobe entre les peptides

et la phase stationnaire de la colonne constituée de silice greffée avec des groupements alkyles de longueurs variables (C18 est le plus couramment utilisé). Les peptides sont chargés, dessalés et concentrés sur une première colonne de phase inverse (colonne préanalytique) puis ils sont élués sur une seconde colonne de phase inverse plus longue (colonne analytique), classiquement de l'ordre d'une vingtaine de centimètres. Les peptides vont ensuite être élués selon la force de leurs interactions avec la phase stationnaire au fur et à mesure que la concentration du solvant organique augmente.

La chromatographie liquide est classiquement couplée à une source électrospray; il suffit de connecter la colonne de séparation à une aiguille capillaire et d'appliquer le potentiel requis.

4. Protéomique quantitative

La quantification des peptides par spectrométrie de masse permet de déduire la quantité de protéines présentes dans un échantillon biologique. Deux types de quantification sont possibles : relative et absolue.

Dans la Figure I.4 sont résumées les méthodes de quantifications de protéines ou de peptides les plus utilisées à l'heure actuelle.

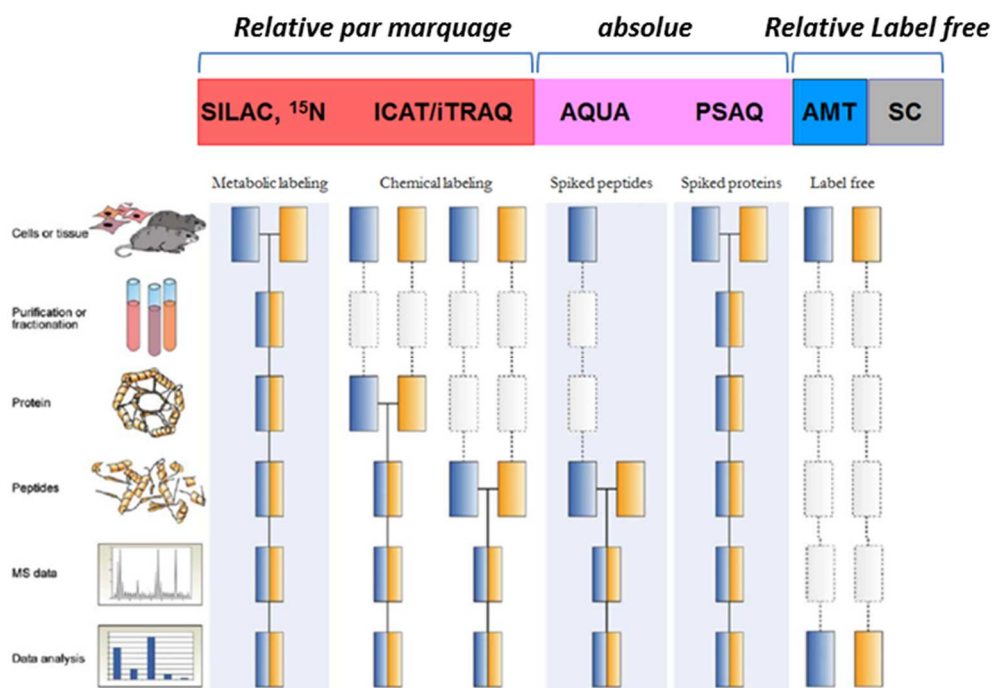


Figure I.4 : Stratégie de protéomique quantitative tiré de (Bantscheff, Schirle, Sweetman, Rick, & Kuster, 2007)

Sur cette figure sont présentées les différentes méthodes de protéomique quantitative en partant des traitements biologiques pour l'extraction des protéines jusqu'aux traitements informatiques des données. Les échantillons comparés sont représentés en bleu et jaune, la ligne pleine horizontale correspond au moment

où les échantillons sont mélangés. Lorsque les échantillons sont traités en parallèle, des erreurs de quantification peuvent être introduites altérant ainsi la précision de la quantification relative.

Dans le cas de la protéomique comparative, il est possible de comparer les protéomes de plusieurs états physiologiques distincts en mesurant l'apparition, la disparition ou les variations des signaux correspondants aux protéines détectées. Il s'agit de protéomique quantitative relative puisqu'il n'est pas possible de connaître la quantité absolue d'une protéine dans un échantillon mais uniquement de savoir si elle est plus abondante dans une condition que dans une autre. On distingue deux grandes familles de méthodes de quantification relative: la quantification par marquage isotopique effectué à différentes étapes de la préparation de l'échantillon (SILAC, ICAT,...) et la quantification sans marquage couramment appelée « Label Free ».

La quantification par marquage isotopique repose sur l'introduction d'éléments isotopiques de masses différentielles mais de composition chimique équivalente aux peptides d'intérêts, garantissant ainsi une co-élution des espèces comparées en chromatographie liquide. L'information quantitative est ensuite obtenue par le rapport d'intensité entre les peptides marqués et les peptides non marqués. L'intérêt de ces méthodes de quantification est que les échantillons à comparer sont analysés en même temps, garantissant ainsi la précision de la quantification en limitant les biais introduits par la perte hétérogène de matériel survenant au cours des différentes étapes de l'analyse protéomique (préparation de l'échantillon, fractionnement et analyse LC-MS). Cependant, cette technique nécessite un marquage isotopique proche de 100% pour éviter les interférences entre peptides non marqués et marqués. De plus cette méthode peut devenir coûteuse si l'on souhaite comparer des cohortes complètes de patients (pouvant atteindre, dans certaines études, jusqu'à des centaines d'échantillons).

La quantification « label free » est basée, quant à elle, sur la comparaison des intensités des signaux MS des peptides ou le comptage des spectres issus de la fragmentation des peptides pour apprécier l'abondance relative des protéines de chaque échantillon à comparer. Cette approche suscite un grand intérêt car il est possible de comparer un nombre illimité d'échantillons correspondant à plusieurs conditions. De plus, elle s'applique à tous types d'échantillons biologiques (cellules en culture, tissus ou fluides biologiques). Cependant, cette méthode est plus délicate à mettre en œuvre car les pertes de matériel hétérogènes entre les échantillons peuvent intervenir tout au long de l'analyse protéomique et générer, si l'on n'y prend garde, des différences artéfactuelles entre les populations à comparer. Il faut donc s'assurer de la reproductibilité des méthodes de préparation en amont de la cohérence des données avant toute interprétation biologique.

a) Les méthodes de quantification avec marquage isotopique différentiel

La méthode SILAC (Stable isotope by amino acids in cell culture), décrite par Oda (Oda, Huang, Cross, Cowburn, & Chait, 1999) et affiné par la suite par Ong (Ong et al., 2002), consiste à cultiver les cellules que l'on souhaite comparer dans deux milieux différents : l'un contenant des acides aminés naturels, l'autre des acides aminés isotopiquement marqués. Un ou plusieurs acides aminés, généralement l'arginine, la lysine ou la méthionine, sont marqués par substitution de certains atomes constitutifs par des isotopes rares stables (^{13}C , ^2H ou ^{15}N). Dans une expérience SILAC, on applique un stimulus ou un stress à une population cellulaire, l'autre servant de référence. Les deux populations sont ensuite mélangées en quantités égales puis traitées et analysées comme un échantillon unique. Ainsi, les deux échantillons sont préparés de manière identique jusqu'à l'analyse par spectrométrie de masse, permettant ainsi une meilleure précision de quantification en évitant les biais introduits lorsque les échantillons sont traités séparément. Sur les spectres de masse, chaque peptide est détecté sous la forme d'une paire constituée de la forme lourde du peptide et de sa forme légère. L'analyse protéomique quantitative différentielle consiste à repérer ces paires peptidiques et à mesurer les rapports des signaux correspondant aux deux versions du peptide détecté. La fragmentation des peptides en mode MS/MS permet de déterminer leur séquence en acide aminés et ainsi d'identifier les protéines d'intérêt. Il est à noter que cette approche est limitée à l'étude de cellules en cultures et ne permet pas l'analyse d'échantillons provenant de patients.

L'approche ICAT (Isotope Coded Affinity Tag-labeled), décrite par Gygi en 1999 (Gygi et al., 1999), consiste à utiliser des réactifs isotopiquement distincts pour marquer chimiquement, et de façon spécifique, les cystéines des protéines des deux échantillons à comparer. Le tag est constitué de 3 parties : (i) un groupement réactif iodeure permettant le marquage spécifique des cystéines par alkylation, (ii) une chaîne carbonée de structure isotopique caractéristique : la forme légère contient des atomes d'hydrogène D_0 et la forme lourde est composée d'atomes d'hydrogène D_8 et (iii) une biotine permettant l'enrichissement ultérieur des peptides marqués par affinité à la streptavidine. Cette technique permet de décomplexifier l'échantillon en ciblant uniquement les peptides contenant des résidus cystéine. Il est donc possible d'améliorer la gamme dynamique et ainsi quantifier des protéines minoritaires. Cependant, comme les cystéines ne sont pas des acides aminés très fréquents (Vinobha, Rajasekaran, & Rajadurai, 2010), on ne quantifie qu'un sous ensemble du protéome.

La stratégie iTRAQ, décrite en 2004 par Ross et collaborateur (Ross et al., 2004), repose sur le marquage chimique des peptides tryptiques. Cette méthode permet d'analyser jusqu'à 8 échantillons en parallèle et d'augmenter ainsi le débit analytique par multiplexage. Les réactifs iTRAQ forment une liaison covalente avec les amines primaires (extrémité N-terminal des peptides et chaînes latérales des résidus lysines). Chaque échantillon est digéré séparément et les peptides sont marqués par un réactif spécifique. Les

réactifs iTRAQ sont constitués de 3 parties (i) un groupement « reporter » (N-méthyl piperazine) de masse 114, 115, 116, 117 (pour le 4-plex) permettant la quantification relative des différents échantillons lors de l'analyse MS/MS, (ii) un groupement « balance » (clivé lors de l'analyse MS/MS) qui permet d'homogénéiser la masse finale du peptide marqué permettant ainsi qu'un même peptide présent dans les différents échantillons apparaisse sous la forme d'un seul massif isotopique améliorant ainsi la détection de peptides minoritaires et (iii) un groupement réactif qui lie de manière covalente les deux autres groupements aux peptides. Cette technique est utilisée pour la découverte de biomarqueurs en protéomique clinique, et pour la compréhension des mécanismes moléculaires de pathologie. Cependant, dès que l'on travaille sur de large cohorte de patients, analysés individuellement, cette technique n'est plus appropriée à cause du nombre limité de rapporteurs disponibles.

b) Les méthodes de quantification avec standard internes (marquage isotopique)

Les méthodes de quantifications décrites dans cette partie suivent toutes le même principe celui de la dilution isotopique. Pour quantifier un peptide, on ajoute en quantités connues un standard interne marqué qui est un composé ayant des propriétés physico-chimiques similaires à celui à doser. L'échantillon est ensuite analysé en spectrométrie de masse (analyse LC-MS). Sur le chromatogramme, les deux pics correspondants au peptide endogène (peptide que l'on souhaite doser) et au peptide standard (peptide marqué) possèdent le même temps de rétention chromatographique. En réalisant une courbe de calibration à partir de la quantité du standard interne que l'on a ajouté et l'intensité des peptides endogènes et exogènes, on en déduit la quantité du peptide endogène.

La quantification avec standards internes marqués s'apparente à de la quantification absolue de protéines ou de peptides. Ces méthodes de quantification (Figure I.5) concernent principalement l'analyse ciblée de certaines protéines définies au préalable, comme c'est le cas lorsque l'on cherche à qualifier, confirmer ou valider des candidats biomarqueurs (cf : B. Protéomique clinique). Différentes méthodes de quantification absolue, utilisant des peptides ou des protéines marquées comme étalons internes, ont été publiées : la stratégie AQUA « absolute quantification » publiée en 2003 par Gerber (Gerber, Rush, Stemman, Kirschner, & Gygi, 2003), l'approche QconCAT « Quantification concatamer » décrite en 2005 par Beynon (Beynon, Doherty, Pratt, & Gaskell, 2005) et la méthode utilisant des protéines marquées PSAQ « Protein Standard Absolute Quantification » décrite en 2008 par Dupuis (Dupuis, Hennekinne, Garin, & Brun, 2008).

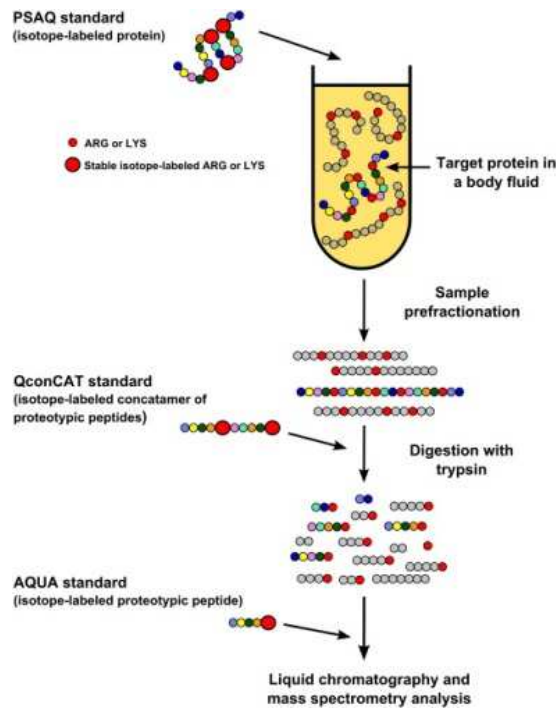


Figure I.5 : méthodes de quantification absolue des protéines en utilisant trois types de standards internes PSAQ « Protein Standard Absolute Quantification », QconCAT « absolute Quantification using concatenated signature peptides », AQUA « Absolute QUantification » (Brun et al., 2007).

La technique des peptides AQUA permet de réaliser la quantification absolue de protéines contenues dans un mélange complexe en rajoutant, juste avant l'analyse par spectrométrie de masse, des peptides marqués dans l'échantillon. L'intégration des signaux issus des peptides « léger » et « lourd » va permettre de déterminer la concentration du peptide cible et donc de sa protéine parente.

Une des difficultés rencontrées avec les standards AQUA est liée à la solubilisation, rendant ainsi la quantification absolue incertaine. En effet, si le peptide étalon est mal solubilisé, il est difficile de connaître la quantité exacte de peptide standard rajoutée à l'échantillon et donc impossible d'en déduire précisément la quantité absolue du peptide endogène. De plus, des erreurs lors de la quantification peuvent être introduites, dues notamment aux rendements de digestion. Une digestion partielle peut entraîner une quantification inexacte puisque la quantité obtenue pour un peptide ne reflète pas la quantité de ce peptide dans l'échantillon initial ; cela se traduit en particulier par des différences de quantité estimées pour les différents peptides d'une même protéine. La concentration obtenue est parfois sous-évaluée par rapport à la concentration initiale de la protéine dans l'échantillon, avant tout traitement (Brun, Masselon, Garin, & Dupuis, 2009).

L'approche QconCAT permet de quantifier plusieurs protéines de manière simultanée à partir d'un seul standard. Un QconCAT est un concatémère de plusieurs peptides issus d'une ou de plusieurs protéines.

Après avoir sélectionné les protéines à doser dans l'échantillon, on sélectionne des peptides « rapporteur » de ces protéines. Un gène synthétique codant pour un concatémère de ces différents peptides est alors façonné. L'expression de ce gène de manière recombinante permettra la production d'une protéine chimérique constituée par l'agrégation bout à bout des peptides de chaque protéine à doser. Comme pour la technique AQUA, cette protéine est synthétisée en y incorporant des formes isotopiques stables de certains acides aminés ; permettant ainsi la quantification à partir de l'analyse des signaux par spectrométrie de masse. Cette protéine chimérique est rajoutée à l'échantillon juste avant l'étape de digestion trypsique afin d'éliminer les biais dus au rendement de digestion. Cependant, cette protéine peut différer, de par ses propriétés biochimiques, des protéines natives, ce qui peut introduire des erreurs de quantification. De plus, la production d'une telle protéine dans des conditions stables n'est pas aisée.

L'approche PSAQ repose sur la biosynthèse d'étalons protéiques marqués isotopiquement. Ces étalons, biochimiquement équivalents aux protéines à doser, sont ajoutés à chaque échantillon avant tout traitement. L'échantillon peut ensuite être fractionné, digéré à la trypsine puis analysé par spectrométrie de masse. Chaque peptide issu de la digestion de la protéine marquée va être traité et analysé de la même manière que dans les deux méthodes précédentes. Cette méthode est plus robuste et apporte une quantification absolue plus précise pour des protéines issues d'échantillons biologiques complexes que les deux autres méthodes de quantification citées précédemment (AQUA et QConCAT). En effet, elle permet de tenir compte des rendements des différentes étapes en amont de la spectrométrie de masse, lors de la détermination de la concentration des protéines cibles.

c) Les méthodes de quantification sans marquage « Label free »

Les méthodologies Label-Free consistent à analyser séparément des échantillons traités en parallèle pour comparer leurs signaux, au terme de leur analyse LC-MS. Contrairement aux méthodes basées sur le marquage isotopique, elles permettent de comparer un nombre quasi-illimité d'échantillons biologiques. Elles peuvent, de plus, s'appliquer à tous types d'échantillons. Ces méthodes contournent certaines difficultés liées à la dilution isotopique telles que l'efficacité du marquage et la précision du mélange des échantillons à comparer. Cependant, elles nécessitent une excellente reproductibilité autant au niveau de la préparation des échantillons (réplicats techniques) qu'au niveau de l'analyse proprement dite (réplicats analytiques). En effet, les variations introduites lors de la préparation peuvent créer des différences fortuites entre les échantillons. De même, les conditions chromatographiques ou la réponse du spectromètre de masse peuvent évoluer au cours des analyses, créant ainsi une dérive des signaux. Il est donc nécessaire de s'assurer de la fiabilité des techniques de préparation des échantillons et de la stabilité de la réponse instrumentale tout au long de l'analyse. On distingue deux approches de quantification sans marquage : la première qui consiste à quantifier les peptides en comptant les spectres à partir d'une analyse LC-MS/MS et la seconde qui repose sur l'analyse des signaux MS à partir d'une analyse LC-MS.

(1) Spectral count

Le “spectral count” consiste à quantifier les protéines en comptant les spectres qui sont générés lors de la fragmentation de leurs peptides lors de l’analyse LC-MS/MS. Il existe une corrélation entre l’abondance d’une protéine et le nombre de spectres obtenus lors de la fragmentation des peptides issus de la digestion trypsique de cette protéine. Cette corrélation a été mise en évidence à mainte reprise dans les données d’identifications lors d’études d’échantillons complexes (Old et al., 2005). Plus une protéine est abondante dans l’échantillon (plus ses peptides sont abondants), plus le spectromètre de masse va générer de spectres de fragmentation de ces peptides. Cependant, la longueur des protéines constitue un biais connu de cette méthode de quantification qui ne donne au mieux qu’une valeur approximative de l’abondance des protéines.

(2) Méthode AMT - Accurate Mass and Time Tags

La stratégie AMT décrite en 2000 par Conrads (Conrads, Anderson, Veenstra, Pasa-Tolić, & Smith, 2000) a été appliqué pour la première fois en 2002 (Lipton et al., 2002) afin de caractériser le protéome de *Deinococcus radiodurans*. Cette méthode de quantification, qui se déroule en deux phases, consiste dans un premier temps, à identifier les protéines d’un mélange complexe préalablement séparées par différentes techniques de fractionnement et, dans un deuxième temps, à quantifier les protéines à partir des intensités des signaux MS des peptides générés par digestion trypsique des protéines des échantillons à comparer (voir Figure I.6).

Durant la première phase, on caractérise de la manière la plus exhaustive possible le protéome de l’échantillon d’intérêt (cellules, fluides biologiques ou tissus). Pour ce faire, on utilise des techniques de préfractionnement de l’échantillon au niveau protéique, tel que le gel SDS-PAGE et/ou peptidique tel que l’IEF-OFFGEL. Les peptides générés par la digestion trypsique des protéines sont séparés sur une colonne de chromatographie liquide et séquencés par une méthode de spectrométrie de masse en tandem. La trappe ionique couplée ou non à un analyseur à transformée de Fourier représente l’instrument le mieux adapté pour cette phase d’identification grâce à sa grande sensibilité, sa gamme dynamique élevée, et son temps de cycle. Le préfractionnement des protéines ou des peptides implique plusieurs analyses pour un même échantillon. L’ensemble des informations issues de ces analyses protéomiques sont stockées dans une base de données nommée « base AMT ». Dans cette « base AMT », les peptides sont définis par (i) leur séquence en acides aminés, (ii) leur masse théorique calculée à partir de leur séquence en acides aminés et (iii) leur temps de rétention chromatographique. Un couple masse théorique - temps de rétention étant nommé AMT tags.

Durant la deuxième phase, les échantillons, sans aucun fractionnement préalable, issus des différentes conditions à comparer, sont quantifiés à partir de l’intensité des signaux MS. En spectrométrie de

masse, on mesure sur le détecteur de courants ioniques. Ces courants ioniques sont proportionnels au nombres de molécules présentes dans la source. Il est possible de mesurer le courant ionique ou le signal MS d'un ion en particulier nommé XIC (Xtracted Ion Chromatogram) ; le XIC étant corrélé à l'abondance du peptide (Ong & Mann, 2005). Cette méthode de quantification consiste donc à extraire le signal MS pour chaque peptide. Cependant, cette seconde phase de quantification des peptides nécessite une grande précision de masse et une excellente résolution. L'information obtenue au terme de l'analyse quantitative contient uniquement la masse (précise) des peptides, leur temps de rétention chromatographique et leurs intensités dans un échantillon donné. La précision de mesure de masse et le temps de rétention chromatographique contrôlé permettent d'obtenir la séquence en acide aminés de ces peptides en alignant les masses mesurées et les masses théoriques contenues dans la base de données ainsi que les temps de rétention chromatographiques expérimentaux et ceux définis dans la base de donnée.

L'approche AMT ne nécessite pas de fragmenter les peptides durant cette seconde phase, ce qui permet d'obtenir un débit d'analyse plus élevé que dans les techniques basées sur le préfractionnement (p. ex. MudPIT), notamment pour la comparaison d'échantillon provenant de cohortes de patients.

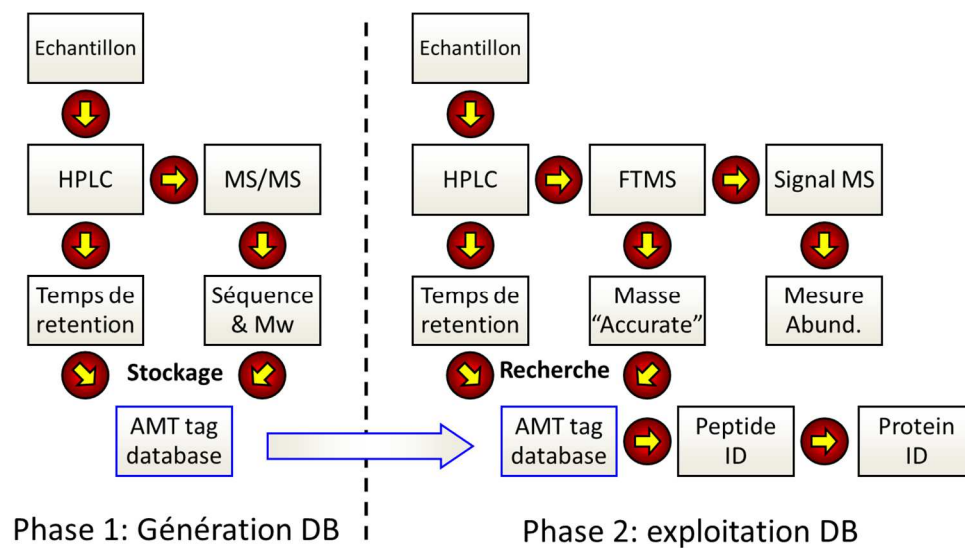


Figure 1.6: L'approche AMT en 2 phases : (i) construction d'une base de données, (ii) quantification et identification des peptides.

B. Protéomique clinique

La protéomique clinique concerne l'étude du protéome pour la recherche de marqueurs diagnostiques, pronostiques et de suivi thérapeutique des pathologies humaines. L'étude du protéome permet également d'éclairer les cliniciens sur de potentielles cibles thérapeutiques. Ce type d'étude nécessite

souvent l'analyse d'un grand nombre d'échantillons et le suivi d'abondance de plusieurs centaines, voire de milliers, de protéines par échantillon. On assiste, ces dernières années, à des évolutions technologiques dans le domaine de la protéomique clinique focalisées sur la découverte de nouveaux biomarqueurs. C'est le cas en particulier d'approches couplant préfractionnement chromatographique et spectrométrie de masse telle que l'approche AMT. L'intérêt de ces technologies réside dans la découverte de biomarqueurs à partir d'échantillons biologiques complexes.

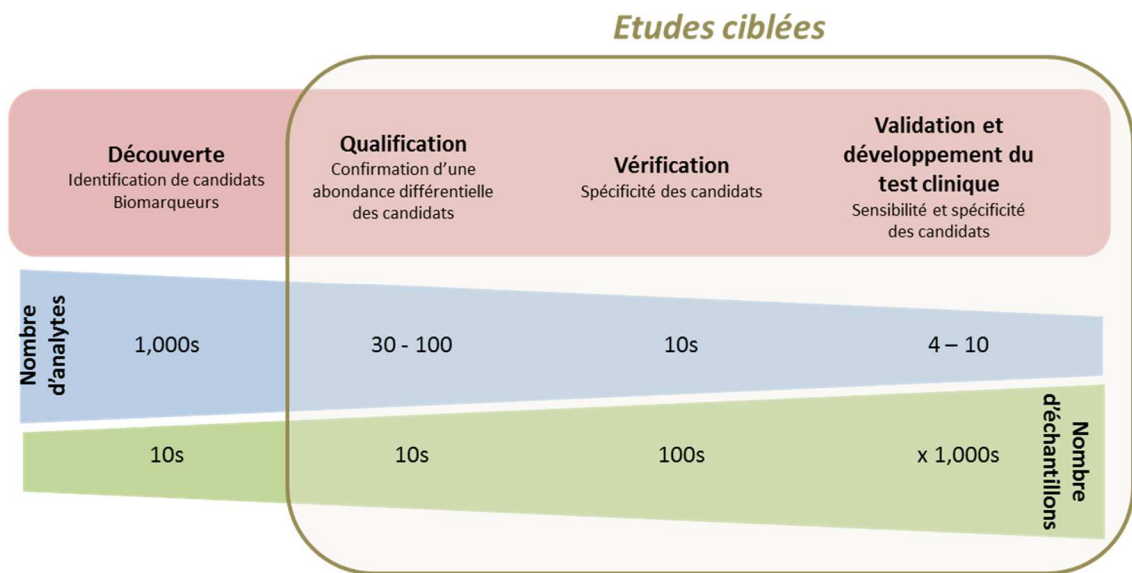


Figure I.7 : De la découverte à la validation d'un candidat biomarqueur

Un biomarqueur, suit un certain nombre d'étapes pour être considéré comme validé. Rifai et al. ont décrit un processus complet et détaillé de la validation de biomarqueurs à travers quatre étapes successives (Rifai, Gillette, & Carr, 2006) (Figure 1.7) : (i) l'étape de découverte des candidats, (ii) l'étape de qualification, (iii) l'étape de vérification et enfin (iv) l'étape de validation des candidats biomarqueurs. L'étape de découverte vise à identifier des biomarqueurs potentiels, par analyse différentielle semi quantitative (quantification comparative) entre différents états (sain et malade). L'étape de qualification consiste en la vérification d'une abondance différentielle des protéines entre les différentes populations (malades et contrôles). L'étape de vérification consiste en la vérification de la spécificité du biomarqueur à la pathologie étudiée. Cette notion est expliquée plus précisément dans la partie suivante. Et enfin l'étape de Validation consiste à établir la spécificité et la sensibilité des candidats biomarqueurs. Comme illustré dans la figure I.7, la taille des cohortes et donc le nombre d'échantillons requis croit au fur et à mesure de l'avancement des candidats biomarqueurs au travers des différentes étapes de découverte, de qualification, de vérification et de validation. En effet, un pouvoir statistique de plus en plus discriminant est nécessaire pour établir des preuves de spécificités et de sensibilités, ce qui induit un nombre de patients à doser de plus

en plus important. Parallèlement, le nombre de candidats biomarqueurs diminue à mesure que l'on avance dans le processus puisque l'on élimine les biomarqueurs non spécifiques ou peu sensibles.

Dans le processus complet de la découverte à la validation de biomarqueurs protéiques, trois obstacles majeurs sont à surmonter: (i) la complexité et la gamme dynamique étendue des protéines dans le type d'échantillon étudié (tissus, cellules, plasma, urine, etc.), (ii) la concentration présumée très faible de nombreux biomarqueurs spécifiques de maladies et (iii) la variabilité des individus et des pathologies. Aujourd'hui, aucune technique analytique utilisée seule n'a la capacité de surmonter ces trois obstacles simultanément. Ainsi, les approches expérimentales combinées à des outils analytiques adaptés sont propres à chaque phase du processus de développement des biomarqueurs.

1. Les biomarqueurs

Un biomarqueur est une protéine dont l'abondance ou les modifications (par clivage ou modification post-traductionnelle) sont corrélées à un état pathologique ou à la réponse à un traitement pharmacologique (Boja et al., 2011). En clinique, pour le diagnostic, le pronostic ou le suivi des patients, il est nécessaire de disposer de tests fiables, c'est-à-dire qui ont une bonne sensibilité et une bonne spécificité. Un test qui a une bonne sensibilité reconnaît la présence de la maladie chez les individus qui en sont atteints. Si la sensibilité du test est mauvaise, certains individus malades seront considérés comme sains. Un test qui a une bonne spécificité reconnaît comme "non atteints" les patients qui ne sont pas malades. Si la spécificité du test est mauvaise, certains individus sains seront considérés comme malades et seront alors soumis à des examens complémentaires entraînant des coûts psychologiques et financiers inutiles.

La recherche de biomarqueurs diagnostiques, pronostiques ou évolutifs est devenue un enjeu crucial pour la prise en charge des patients atteints de pathologies complexes et en particulier de cancers. La protéomique clinique est une approche permettant l'identification de ces biomarqueurs. Pour ce faire, il existe plusieurs techniques disponibles en protéomique pour la découverte de biomarqueurs telles que l'électrophorèse bidimensionnelle couplée à la spectrométrie de masse (2DE-MS) ou la technologie SELDI-TOF MS (Surface Enhanced Laser Desorption/Ionization - Time Of Flight Mass Spectrometry). Cependant, même si ces méthodes permettent d'obtenir un profil protéique avec un haut débit d'analyses, elles présentent des inconvénients tels que l'incapacité d'accéder à certaines catégories de protéines ou liées à l'identification des protéines détectées différenciellement abondantes lors de la comparaison d'échantillon. En conséquence, l'approche AMT apparaît comme une technique de choix pour la découverte de biomarqueurs puisqu'elle permet (i) de caractériser le protéome en profondeur et (ii) de quantifier une grande variété de protéines à partir d'échantillons complexes tels que les fluides biologiques ou les tissus, sans aucun fractionnement préalable, approchant ainsi le débit d'analyse des technologies de type profilage protéique (SELDI-ToF). Dans une première phase, on établit un catalogue protéique représentant une source d'information indispensable par la suite, pour l'analyse quantitative réalisée dans une seconde phase.

Cette analyse quantitative permet d'obtenir des candidats biomarqueurs à partir de la comparaison de différentes populations suivie d'une analyse statistique permettant de mettre en évidence des protéines différenciellement abondantes entre les populations « malades » et « saines ».

2. Les fluides biologiques et leur rôle dans la découverte des biomarqueurs

Les fluides biologiques ou fluides corporels désignent d'une manière générale des liquides produits par un organisme vivant. On distingue deux types de fluides : les liquides au sein d'un circuit fermé tels que le sang ou le liquide cébrospinal et les fluides dans un circuit ouvert tels que l'urine ou la sueur. Les deux fluides les plus couramment utilisés en protéomique clinique sont le sang et l'urine car il est facile de les collecter de manière peu invasive et en quantité suffisante pour réaliser différentes études à partir d'un seul prélèvement.

Les fluides présentent un défi analytique considérable pour la protéomique en raison de leur gamme dynamique des concentrations protéiques, leur complexité (Adkins et al., 2002) ou la présence de composés interférents (tels que les sels ou les débris cellulaires,...). L'extrême complexité des fluides biologiques, tels que le plasma ou l'urine, représente un enjeu majeur pour la protéomique (N. L. Anderson & Anderson, 2002). Dans le cas du plasma, le rapport de concentration entre les protéines les plus abondantes (albumine) et les moins abondantes (cytokines) peut ainsi atteindre une valeur estimée à 10^{12} , comme l'illustre la Figure I.8. L'albumine, à elle seule représente 55 % des protéines du plasma et les 9 protéines majoritaires constituent 90 % des protéines plasmatiques, alors qu'une vingtaine de protéines représentent, à elles seules, les 99% du contenu protéique total du plasma (Figure I.9).

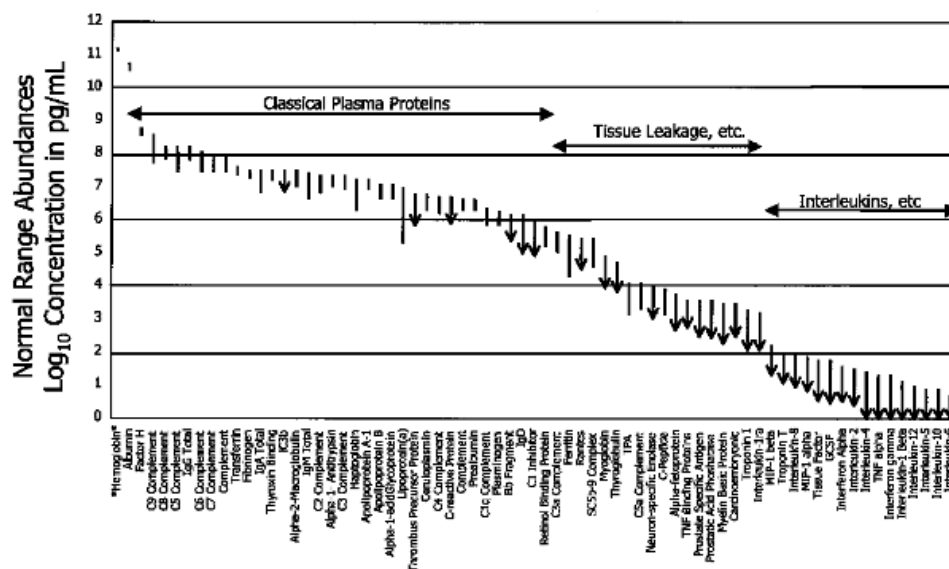


Figure I.8 : Intervalle de concentration de 70 protéines du plasma, d'après N.L Anderson. (N. L. Anderson & Anderson, 2002)

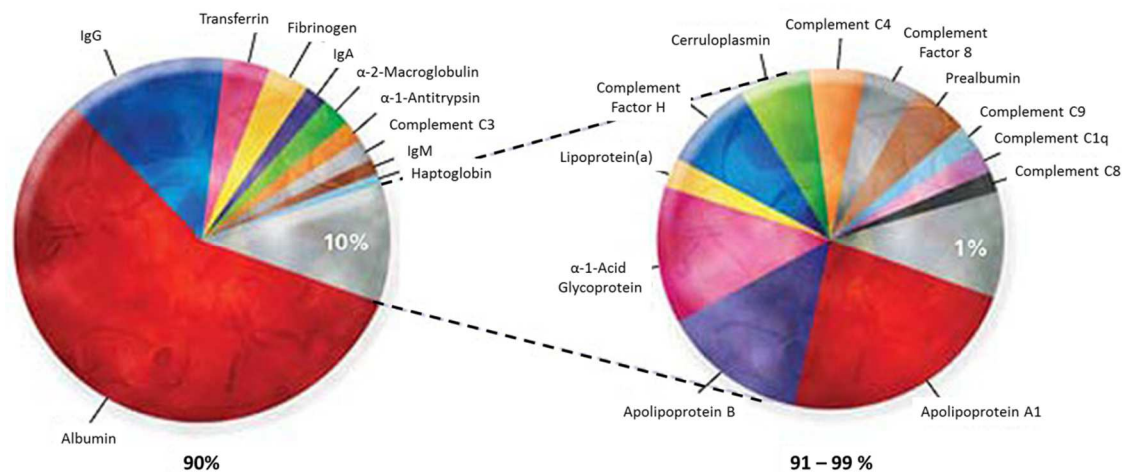


Figure 1.9 : Distribution des protéines représentant (A) 90% et (B) de 91-99% de la quantité totale de protéines du plasma.

a) L'accessibilité aux cohortes d'échantillons

Une des difficultés inhérentes à la protéomique clinique repose sur la constitution de cohortes en accord avec les exigences de l'analyse protéomique quantitative, en particulier lorsque l'on utilise des stratégies « Label Free ». Pour la phase de découverte des candidats biomarqueurs, il est nécessaire, d'analyser plusieurs dizaines d'échantillons (Rifai et al., 2006), collectés dans des conditions strictement contrôlées et de minimiser la variabilité entre les patients d'une même population en appariant l'âge, le sexe, les caractéristiques des tumeurs pour les populations pathologiques. Dans la plupart des études cliniques, il est difficile d'obtenir des cohortes de patients qui réunissent l'ensemble de ces caractéristiques. De plus, les performances des outils statistiques utilisés pour discriminer les protéines différemment abondantes dépendent du nombre d'analytes mesurés (les peptides) et du nombre d'échantillons considérés. Lorsque le nombre de peptide détectés est grand (typiquement plusieurs milliers), il est souhaitable d'avoir une cohorte importante (centaines de patients). Des méthodes statistiques ont été développées pour tenir compte de la variabilité des échantillons au sein d'une même population et du déséquilibre des matrices peptides-patients (Fu et al., 2008).

Le fait que beaucoup d'études de protéomique clinique pour la découverte de biomarqueurs s'appuient sur l'analyse d'une ou deux dizaines d'échantillons (Chen et al., 2010) met en évidence la difficulté de constituer des cohortes, mais aussi le défi de satisfaire les exigences de la protéomique tout en considérant les contraintes de la clinique. Le travail présentée dans cette thèse, s'appuie sur l'analyse d'une centaine d'échantillons en accord avec les recommandations de Rifai (Rifai et al., 2006). Ceci a constitué un travail considérable des équipes de protéomique et de clinique, pour établir un protocole qui satisfasse les

exigences des statisticiens et des protéomiciens en considérant la difficulté de préparer des échantillons en même temps que se déroule la collecte. Il est à noter, qu'en général, c'est la même personne qui assure la collecte et les premiers traitements de l'échantillon.

b) De la collecte au traitement préanalytique des échantillons de fluide biologique

Les fluides biologiques sont très sensibles à divers paramètres qui ne sont pas toujours en lien direct avec la pathologie. En particulier, les facteurs préanalytiques et notamment les protocoles de collecte peuvent avoir un rôle déterminant pour les analyses de protéomique quantitative en induisant des différences d'abondance de protéines indépendantes de la pathologie. Il en va de même pour les conditions de conservation des échantillons avant l'analyse. Par ailleurs, la variabilité physiologique inter et intra individuelle doit être prise en compte (Lambert, Berrahmoune, Herbeth, Siest, & Visvikis-Siest, 2007) lors de la conception du design d'une étude clinique. Pour cela, on évalue la variabilité inter individu en analysant une large cohorte de patients de manière à minimiser cet effet et à inclure cette variabilité dans l'analyse statistique. Le projet PPP Plasma Proteome Project, initié en 2002 par l'organisation HUPO, a permis, dans le cas du plasma, de définir des stratégies de collecte, des prérequis pour la préparation et l'analyse protéomique (Rai et al., 2005) tels que l'utilisation d'inhibiteurs de protéase, la température de stockage ou la durée écoulée avant le stockage des échantillons. En ce qui concerne l'urine, la plupart des recommandations sur les traitements préanalytiques se basent sur des études par gel2D (Thongboonkerd, 2007), et comme nous allons le voir par la suite, ne sont pas nécessairement applicables aux études basées sur la MS.

Pour obtenir une large couverture de protéomes de fluides biologiques, il apparaît indispensable de fractionner l'échantillon avant l'analyse par spectrométrie de masse de manière à réduire la gamme dynamique. Ce problème de gamme dynamique est une caractéristique de tous les fluides biologiques et, même si elle est moindre dans l'urine, elle n'en reste pas moins un handicap pour les analyses protéomiques. Pour répondre à cette difficulté, il existe différentes approches basées sur le fractionnement protéique comme la séparation sur Gel 1D, l'isoélectrofocalisation ou sur l'immunodéplétion des protéines majoritaires ou des méthodes plus originales comme la technologie d'égalisation : « Equalizer ».

L'immunodéplétion :

Pour la découverte de biomarqueurs dans les fluides complexes comme le plasma, le sérum ou le liquide céphalo-rachidien, il est parfois conseillé de dépler les protéines de plus hautes abondances pour permettre de détecter les protéines minoritaires qui sont autant de potentiels biomarqueurs. La technique d'immunodéplétion qui consiste en la déplétion par immunoaffinité des protéines majoritaires d'un échantillon est réalisée à l'aide d'un mélange d'anticorps spécifiques immobilisés sur une matrice de

chromatographie. Pour cela plusieurs kits commerciaux sont disponibles pouvant dépler jusqu'à 20 protéines majoritaires du plasma selon le kit utilisé. Bien que l'intérêt de cette approche ait été démontré dans différentes études de protéomique (Pieper et al., 2003), cette méthode de déplétion peut également éliminer d'autres protéines que celles visées à l'origine, certaines protéines abondantes pouvant jouer un rôle de transporteur. et Pour le fractionnement de l'urine, il n'existe actuellement pas de kit commercial permettant d'éliminer une des protéines majoritaires qu'est l'uromoduline.

La technologie « Equalizer » :

Une autre stratégie, décrite pour la première fois par Thulasiraman en 2005 (Thulasiraman et al., 2005), est basée sur le principe d'affinité : les protéines de l'échantillon se lient à une banque combinatoire d'hexapeptides (20^6 soit 64 millions ligands) greffés sur des billes de polyméthacrylate, pouvant théoriquement sélectionner toutes les protéines. Chaque ligand est présent en quantité identique, permettant de fixer un nombre identique de chacune des protéines du mélange. Sous un ratio billes/échantillon contrôlé, les protéines majoritaires saturent rapidement leur site et sont donc éliminées par lavage, tandis que les protéines moins représentées s'adsorbent sur les billes et se retrouvent plus concentrées dans l'éluat final. Cette technologie permet d'atténuer les différences de concentrations protéiques au sein des fluides biologiques et ainsi d'identifier un plus grand nombre de protéines (Righetti, Castagna, Antonioli, & Boschetti, 2005),(Righetti, Boschetti, Lomas, & Citterio, 2006).

Castagna et collaborateurs ont utilisé cette méthode pour « égaliser » les concentrations protéiques d'un pool d'échantillons urinaires (Castagna et al., 2005). A partir de 200mL d'urine de 8 volontaires (soit au total 1.6 L d'urine), ils ont identifié 383 protéines non redondantes. En comparant leurs résultats avant et après égalisation (figure I.10), il apparaît que le traitement de l'urine par cette technique améliore de manière significative la couverture du protéome urinaire.

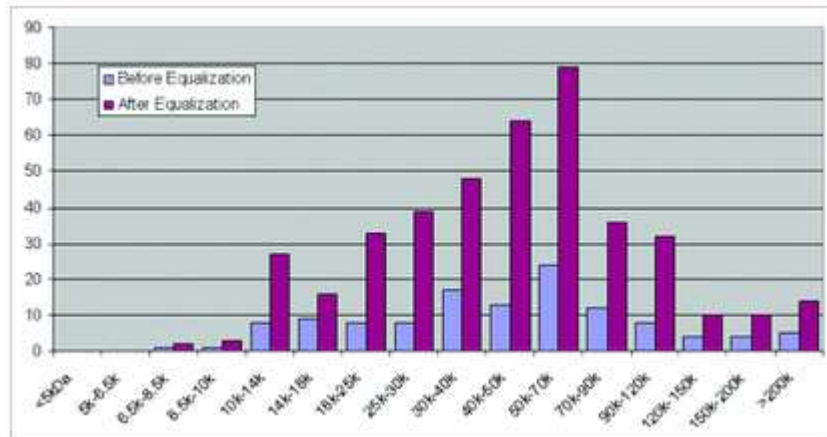


Figure I.10 : Nombre de protéines identifiées en fonction de leur poids moléculaire avant et après déplétion partielle des protéines majoritaire d'un pool d'urines de 8 patients (1.6 L) par la méthode Equalizer (Tiré de Castagna et al. 2005).

La technologie IEF-OFFGEL :

Dans la plupart des cas, le fractionnement d'un mélange peptidique complexe repose sur la combinaison de deux types de séparations chromatographiques d'affinité : les peptides sont tout d'abord séparés selon leur charge sur une colonne SCX (Strong Cation eXchange) sous l'application d'un gradient de sels puis ils sont séparés selon leur hydrophobicité sur une deuxième colonne de phase inverse (RP) sous l'application d'un gradient de phase organique. Ainsi, les technologies DALPC (Direct Analysis of Large Protein Complex) (Link et al., 1999) et MudPIT (multidimensional protein identification technology) (Wolters, Washburn, & Yates, 2001) ont émergé vers la fin des années 90, combinant ces deux types de phases afin de caractériser des protéomes complexes comme ceux du plasma humain ou encore du liquide céphalo-rachidien humain. Cependant, des gradients de sels combinés à des gradients de phases organiques sont bien souvent difficiles à concilier sur une même chaîne de chromatographie liquide et posent des problèmes de robustesse de la méthode analytique développée, tels que des dérives de temps de rétention, bouchages de colonnes.

La technique d'IEF-OFFGEL introduite en 1982 par Bjellqvist (Bjellqvist et al., 1982), représente une alternative intéressante aux techniques de fractionnement MudPIT, permettant de contourner les problèmes de robustesses dues à l'utilisation combinée de gradients de sels et de phase organique. Elle représente la première dimension de l'électrophorèse bidimensionnelle permettant de séparer les peptides ou protéines selon leurs points isoélectriques dans un gradient de pH sous l'effet d'un champ électrique. Hubner et collaborateurs (Hubner, Ren, & Mann, 2008) ont utilisé cette technologie de fractionnement en amont de l'analyse LC-MS sur un lysat de cellules HeLa et sur un digestat de levure. Après une digestion en solution de l'échantillon, les peptides ont été séparés selon leur point isoélectrique. Une comparaison des résultats

obtenus par cette technique à une méthode plus conventionnelle qui consiste à séparer les protéines sur un gel SDS PAGE 1D et à les digérer avec la trypsine a été réalisée. La Figure (I.11), tirée de cette publication, montre une nette augmentation du nombre de protéines identifiées en utilisant la technologie IEF-OFFGEL, comparée à la technique plus classique de gel1D.

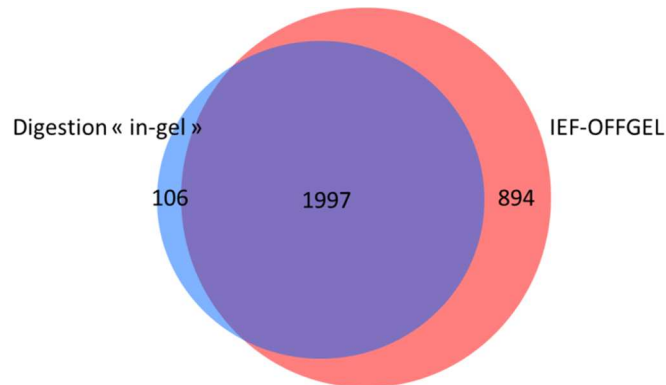


Figure I.11 : Diagramme de Venn illustrant le nombre de protéines identifiées à partir de 50µg de lysat de cellules obtenues après fractionnement des protéines par gel SDS-PAGE et digestion « in gel » ou un fractionnement des peptides par IEF après une digestion des protéines en solution.

C. Contexte clinique de la thèse (DECanBio)

Ce travail de thèse s'inscrit dans le cadre d'un projet européen DECanBio « Discover and Evaluate Cancer Biomarkers ». Ce projet, qui s'est déroulé sur 4 ans, avait pour principal objectif de mettre en place une stratégie permettant la découverte et l'évaluation de biomarqueurs du cancer de la vessie dans les urines, en s'appuyant sur l'utilisation des méthodes récentes d'analyse des protéomes par spectrométrie de masse.

Le cancer de la vessie est le 4ème cancer en terme d'incidence, chez les hommes, en Europe et aux Etats Unis. Chez les femmes, il se place au 9ème rang aux Etats Unis et 14ème en Europe (Ferlay, Parkin, & Steliarova-Foucher, 2010), (Jemal, Siegel, Xu, & Ward, 2010). En règle générale, ce cancer survient en moyenne vers l'âge de 65 ans. Le ratio homme/femme qui développe un cancer de la vessie est approximativement de 3 pour 1. Les facteurs de risques sont le tabagisme et l'exposition à certains composés chimiques comme les amines aromatiques dans l'industrie chimique (Guha et al., 2010).

1. La Vessie

La vessie est un organe creux situé dans le petit bassin et qui se remplit progressivement avec l'urine produite par les reins pour la stocker avant son élimination au cours de la miction par l'urètre. L'urine émise par les reins est évacuée dans la vessie par l'intermédiaire des uretères.

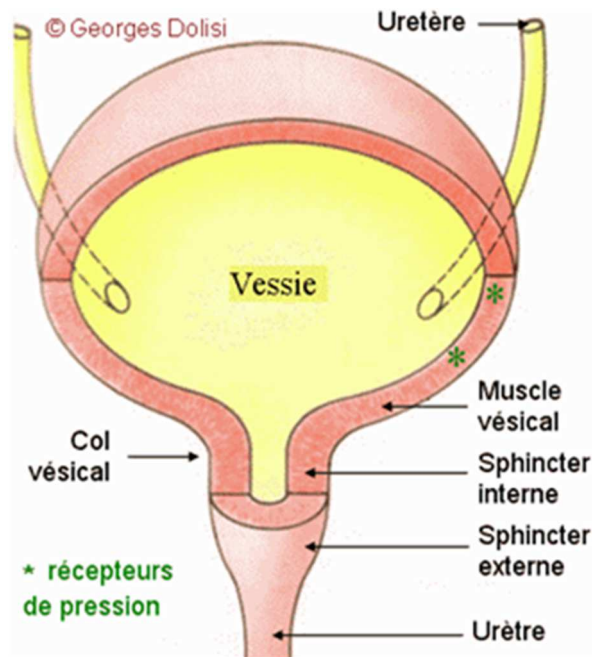


Figure I.12: Schéma descriptif du système urinaire

La figure I.12 décrit le système urinaire qui comprend : les uretères, des canaux du système urinaire de 25 à 30 cm de long qui conduisent l'urine des reins vers la vessie. L'urètre, quant à lui, permet l'évacuation de l'urine vers l'extérieur de l'organisme. L'intérieur de la vessie est recouvert d'une couche d'épithélium : l'urothélium qui représente un tissu composé de cellules juxtaposées disposées en une ou plusieurs couches tapissant la face interne des organes. La plupart des tumeurs de la vessie affectent les cellules épithéliales.

2. Les tumeurs de la vessie

Les tumeurs de la vessie sont classées en deux grandes catégories selon leur stade d'infiltration croissante dans l'épaisseur de la paroi vésicale (Figure I.13) : les tumeurs non infiltrantes ou classées superficielles qui ne touchent que la partie superficielle de la paroi vésicale et les tumeurs infiltrantes qui atteignent le muscle de la paroi vésicale.

Les tumeurs superficielles appelées aussi polypes ont comme risque principal la récurrence. Les tumeurs Ta pénètrent superficiellement l'urothélium sans aller au-delà de la membrane basale. Dans 50% des cas, les primo tumeurs sont des tumeurs Ta, de bas grade avec un bon pronostic. Ce type de tumeur affiche un taux élevé de récurrence 60% à 5 ans (Knowles, 2008). Cependant, seulement 10% à 15% des patients développent ultérieurement une tumeur infiltrante. Le carcinome in situ (Tis) est une lésion superficielle de haut grade qui ne pénètre pas au-delà de la membrane basale. Ce type de tumeur est souvent rencontré avec d'autres tumeurs urothéliales. Enfin, les tumeurs T1 représentent 20% des tumeurs de la vessie, elles pénètrent au-delà de la membrane basale et rentrent dans le chorion sans envahir le muscle. Les tumeurs T1 sont encore considérées superficielles mais elles indiquent des tumeurs à plus haut risque de

progression dont l'évolution est difficile à prévoir. Ces tumeurs présentent un fort risque de progression vers des tumeurs infiltrantes.

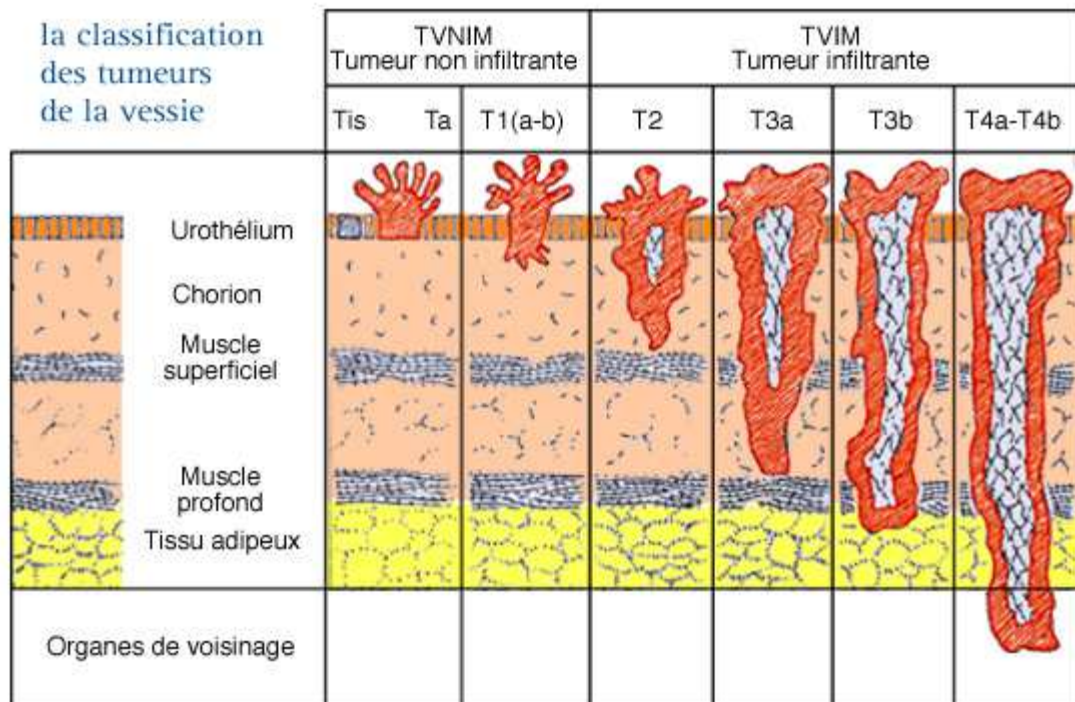


Figure I.13 : Classification des tumeurs de la vessie

Les tumeurs infiltrantes (T2 à T4), quant à elles, présentent un fort risque d'évolution vers des métastases qui touchent essentiellement les ganglions, les os, les poumons et le foie. Elles représentent 30% des tumeurs et envahissent le muscle superficiel et profond. Dans toutes les classes de tumeurs, on retrouve un fort taux de récurrence après traitement entre 50% et 80% des cas à 5 ans. Cette récurrence peut apparaître au niveau de la vessie ou dans d'autres organes de l'appareil urinaire comme les reins ou l'urètre.

Les tumeurs vésicales sont, par ailleurs, classées par grade cytologique allant de G1 à G3. Le grade cytologique est basé sur la description d'anomalies architecturales et cyto-nucléaires de l'urothélium et ne tient pas compte du caractère invasif ou non de la tumeur. Toutefois, il est lié à l'agressivité de la tumeur et représente un facteur pronostique très important, prédictif de l'invasion de la paroi vésicale. Les critères cytologiques de malignité sont la taille du noyau, la présence de chromatine irrégulièrement répartie, des contours irréguliers, ou encore une augmentation de la taille du nucléole avec des anomalies de forme. Ainsi, le grade G1 est associé à un faible risque de malignité et ne concerne que les tumeurs non infiltrantes. Le grade G2 est associé à un carcinome de bas grade et est retrouvé dans les mêmes proportions dans les tumeurs infiltrantes et non infiltrantes. Le grade G3 est associé à un carcinome de haut grade qui est majoritairement rencontré dans les tumeurs infiltrantes.

Nous venons de voir que les tumeurs vésicales superficielles présentent un risque important de récurrence et surtout de progression. Une surveillance régulière et efficace de l'appareil urinaire est donc

cruciale pour les déceler. Le traitement des tumeurs superficielles de vessie (TSV) doit prendre en compte un certain nombre de facteurs de risque de récurrence et de progression. Ces facteurs pronostiques de récurrence sont par ordre décroissant : le nombre de tumeurs, l'existence d'une récurrence à trois mois après la résection initiale, la fréquence des récurrences, la taille de la tumeur, le grade. Le risque de progression est lié au grade et au stade tumoral. Selon ces facteurs, les TSV se présentent cliniquement en trois groupes de risque :

Tableau I.2 : Groupes de risques des tumeurs superficielles de vessie

Faible risque : <i>Ta unique, bas grade ou LMP (tumeur à faible potentiel de malignité) (grade 1), diamètre inférieur à 3 cm, non récidivé</i>
Risque intermédiaire : <i>Ta bas grade ou LMP multifocal e/ou récidivante. T1 bas grade ou LMP (grade 1-2)</i>
Haut risque : <i>Ta haut grade (grade 3), T1 récidivante, T1 grade 3 (haut grade), CIS</i>

3. Les méthodes de détection des tumeurs vésicales

Le plus souvent, les cancers de la vessie sont détectés parce que le patient vient consulter pour des saignements qui apparaissent lors de la miction. La plupart des nouveaux cancers détectés concernent des tumeurs non infiltrantes (75 – 85% des cas). Ces tumeurs sont facilement traitées par résection mais le principal problème associé est le fort taux de récurrence ou l'apparition de nouvelles tumeurs. Pour cette raison, le patient est étroitement surveillé après le traitement et en particulier durant la première année qui suit la résection de la tumeur. Selon les recommandations de L'Association européenne d'Urologie (EAU), chez les patients ayant développé des tumeurs à faible risque de récurrence, une cystoscopie doit être effectuée à 3 mois et dans le cas où elle est négative, à 9 mois, puis annuellement pendant 5 ans si le patient ne développe plus de tumeur (Jacobs, Lee, & Montie, 2010). La référence actuelle pour la détection des cancers de la vessie et pour le suivi des patients traités par résection est la cystoscopie (Figure I.14). La cystoscopie est un examen par imagerie médicale au cours duquel un tube doté d'une lentille est introduit dans la vessie par l'urètre. Elle permet donc un examen visuel de la paroi vésicale. Cette méthode est considérée comme le meilleur outil de détection à l'heure actuelle car elle est sensible et spécifique. Cependant, les résultats de la cystoscopie se révèlent souvent négatifs et l'on pourrait donc bien souvent éviter cet examen agressif pour le patient. Cependant, lorsque la tumeur est petite, il est difficile de la visualiser par cette méthode. Dans ce cas les urologues utilisent une technique complémentaire pour poser un diagnostic : la cytologie. La cytologie urinaire est un examen de biologie destiné à déceler la présence éventuelle de bactéries et de cellules cancéreuses. Elle est non invasive et spécifique dans plus de 90% des cas de cancer de la vessie. Cependant, elle est peu sensible (33%) dans les cas de cancers de bas grades et nécessite l'expertise de pathologistes pour l'interprétation des résultats (Nakamura et al., 2009). Il est donc souhaitable d'utiliser d'autres

méthodes en complément de ces techniques et l'espoir de nouvelles techniques s'appuyant sur des marqueurs moléculaires est important.

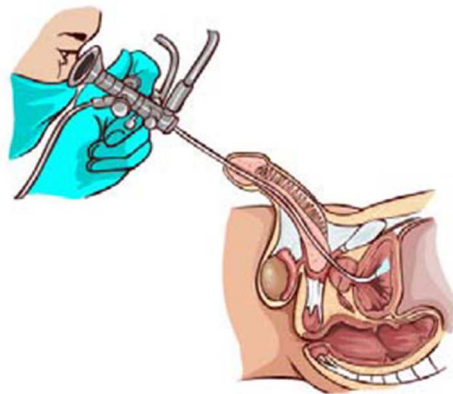


Figure I.14 : La cystoscopie : examen visuel de la paroi vésicale

La prévalence du cancer de la vessie est trop faible pour qu'il soit envisagé un dépistage systématique de l'ensemble de la population. Il est donc fortement souhaitable de trouver une méthode non invasive, peu coûteuse, sensible et spécifique pour améliorer le pronostic des patients et diminuer les coûts associés aux examens de surveillance. Le coût et la contrainte pour le patient de la cystoscopie, soulignent l'intérêt d'un test discriminatif fiable et peu agressif.

4. Biomarqueurs urinaires du cancer de la vessie

L'urine est un fluide biologique composé de déchets de l'organisme. L'élaboration de l'urine se fait grâce aux néphrons, les unités fonctionnelles du rein. Le rôle de l'urine consiste en l'élimination des déchets solubles de l'organisme: urée, créatinine, médicaments, toxines. L'urine est composée essentiellement d'eau, environ 95 %. Les 5 % restant correspondent à des sels, des métabolites et à une faible quantité de protéines provenant de protéines sériques ou du rein. La composition protéique de l'urine n'excède pas, dans des conditions normales, 150mg par jour. A l'état normal, 60% des protéines urinaires excrétées proviennent d'une filtration des protéines plasmatiques par le glomérule. Il s'agit principalement de protéines de bas poids moléculaire, inférieure à 20 kDa comme le lysozyme ou la Beta 2 microglobuline. Les 40% restants proviennent de la sécrétion tubulaire d'uromoduline (ou protéine de Tamm-Horsfall) ou de l'urothélium des voies urinaires. La technique la plus utilisée pour mesurer la protéinurie est la bandelette urinaire. Cette technique permet de détecter la présence d'albumine lorsque celle-ci dépasse 50 à 100 mg/mL. Cependant, elle ne permet pas de détecter les protéines de bas poids moléculaires présentes dans les urines. D'autres méthodes sont utilisées comme l'électrophorèse des protéines urinaires sur SDS-PAGE permettant une étude qualitative de la protéinurie. L'analyse du protéome urinaire soulève plusieurs difficultés liées à la faible

concentration des protéines urinaires comparées aux protéines plasmatiques, à l'activité de protéases, à la présence de sels et à la variation de la concentration des protéines en fonction de l'état d'hydratation.

Les tissus relarguent constamment des protéines ou des fragments protéiques susceptibles d'être détectés dans les fluides biologiques. Ainsi, une tumeur peut relarguer des protéines spécifiques ou modifier l'abondance des protéines normalement sécrétées dans les fluides biologiques. Dans le cas du cancer de la vessie, l'urine représente donc, l'analyte idéal puisque elle est directement en contact avec les cellules épithéliales de la vessie.

Certains tests urinaires à visée diagnostique sont aujourd'hui disponibles. Les plus utilisés sont le BTA Trak ou le NMP 22.

Le BTA Track est un test immuno-enzymatique qui permet la détection du BTA (Bladder Tumor Antigen), une protéine caractéristique de cellules vésicales cancéreuses. Elle a été isolée spécifiquement dans des urines de patients atteints de cancer de la vessie et non dans les urines de sujets sains. Cette protéine est un analogue structural du facteur H du complément hémolytique humain (CFH). Tout comme le facteur H, le BTA est capable d'inhiber l'activité de la voie alterne du complément. Cette activité pourrait permettre aux cellules cancéreuses d'échapper à la lyse. La sensibilité de ce test est d'environ 70%, donc bien meilleure que celle de la cytologie classique qui se situe autour de 33% (Thomas et al., 1999). Quant à sa spécificité (~ 70%), elle est moindre que celle de la cytologie (~ 90%). Le test urinaire NMP22 (Urinary Nuclear Matrix Protein) est basé sur la détection d'une protéine de la matrice nucléaire impliquée dans la formation des fuseaux au cours de la mitose. La production excessive de certaines protéines de la matrice nucléaire qui définit la structure tridimensionnelle du noyau cellulaire et le contrôle de sa morphologie pourrait indiquer une croissance cellulaire excessive et suggérer la présence de néoplasie. Chez les patients atteints de cancer de la vessie, elle est libérée dans l'urine à partir des cellules tumorales. Le NMP-22 est un dosage quantitatif qui repose sur la méthode ELISA (van Rhijn, van der Poel, & van der Kwast, 2005). Les objectifs de ces deux marqueurs tumoraux sont, d'une part de dépister les lésions urothéliales asymptomatiques dans une population à risque et d'autre part, de simplifier les modalités de surveillance des tumeurs vésicales connues et traitées en réduisant les indications de l'endoscopie vésicale. Ces marqueurs pourraient être associés à la classique cytologie urinaire afin d'en améliorer la sensibilité. Cependant, aucun de ces tests n'associe en routine une sensibilité et une spécificité proches de 100% (Abd El Gawad et al., 2005).

Il est donc essentiel de poursuivre les efforts vers la découverte de nouveaux biomarqueurs qui amèneraient un bénéfice pour la surveillance des tumeurs vésicales superficielles, sans remplacer pour autant complètement la cystoscopie qui resterait de toute façon indiquée en cas de test positif. Ces tests devront surtout permettre d'éviter les cystoscopies négatives qui sont bien plus nombreuses que celles mettant en évidence une tumeur et qui engendrent ainsi des désagréments indus pour les patients et des coûts superflus.

La protéomique représente un atout majeur dans la recherche de biomarqueurs. En effet, des variations dans les taux d'excrétion des protéines en lien avec la pathologie pourraient avoir une valeur prédictive dans le diagnostic précoce de pathologies uro-génitales. Cependant, même si l'urine représente un analyte idéal, de par sa proximité avec les cellules tumorales, elle représente un véritable défi pour les études de protéomique quantitative de par sa composition.

D. Etat de l'art en caractérisation du protéome urinaire

On note, dans la littérature, plusieurs études relatives à la caractérisation du protéome urinaire. Ces études s'appuient sur différentes méthodes de préparation, de séparation ou d'enrichissement des protéines ou des peptides. L'une des premières études sur le sujet date de 1979 (N. G. Anderson, Anderson, & Tollaksen, 1979). Les auteurs ont obtenu un profil de protéines urinaires humaines en visualisant environ 250 spots sur un gel bidimensionnel. Cependant, en 1979, comme nous l'avons vu précédemment, il était difficile d'obtenir l'identification des protéines à partir d'un gel bidimensionnel. En 2004, Pieper et collaborateur (Pieper et al., 2004) ont pu observer 1400 spots, toujours par la technique d'électrophorèse bidimensionnelle suivie cette fois-ci d'une identification des spots par MALDI-TOF. Cependant, seulement 30% de ces spots ont été identifiables, permettant la détermination de 150 protéines. Sur la base de ces résultats, l'intérêt de ce fluide pour la découverte de biomarqueurs était considéré jusque-là comme limité.

L'apparition de la chromatographie liquide couplée à des spectromètres de masse à hautes performances a permis d'augmenter de manière significative le nombre de protéines urinaires identifiées. Ainsi, en 2006, Adachi a publié une étude dans laquelle (Adachi, Kumar, Zhang, Olsen, & Mann, 2006), à partir d'urines poolées (9 patients « sains ») ou de patient individuels, différentes méthodes de séparations et de digestion ont été utilisées afin d'identifier un maximum de protéines urinaires. Après avoir éliminé les débris cellulaires grâce à une centrifugation basse vitesse, les protéines urinaires étant extraites par ultrafiltration permettant dans un même temps de concentrer et de dessaler l'urine. Les protéines urinaires ainsi extraites ont été séparées sur un gel SDS PAGE 1D puis digérées par la trypsine. Les peptides ont ensuite été extraits du gel avant d'être analysés par spectrométrie de masse. Parallèlement, le kit Vivapure anti HSA a été utilisé pour dépléter l'albumine. Finalement, Adachi a identifié plus de 1500 protéines urinaires. Cette étude a été un révélateur du potentiel de la protéomique de l'urine pour la recherche de biomarqueurs.

Dans le même temps que ce travail de thèse, Kentsis a publié une étude sur la caractérisation du protéome urinaire en fractionnant les protéines urinaires par ultracentrifugation pour purifier les exosomes. Les exosomes sont des vésicules de 40 à 100 nm qui sont sécrétés par une cellule dans son environnement. Plusieurs études suggèrent que certaines des protéines contenues dans ces vésicules pourraient être des biomarqueurs de pathologies (Pisitkun, Shen, & Knepper, 2004),(Wang, Hill, Luther, Hachey, & Schey,

2012). Au final, à partir d'urine d'enfants et de jeunes adultes (12 au total), Kentsis a identifié 2362 protéines en utilisant une méthode basée sur l'électrophorèse 1D et le fractionnement SCX.

Marimuthu et collaborateurs (Marimuthu et al., 2011) se sont intéressés à une méthode d'enrichissement des glycoprotéines pour caractériser le protéome urinaire. Dans leur étude, après avoir poolées les urines de 24 patients « sains », ils ont effectué une centrifugation à basse vitesse afin d'éliminer les débris cellulaires. Après avoir concentré et dessalé les protéines par ultrafiltration, Les protéines ou glycoprotéines ont été fractionnées selon leurs poids moléculaires par gel SDS-PAGE. Les glycoprotéines ont été capturées sur des billes d'agaroses sur lesquelles sont greffées 3 lectines (la concanavoline A, l'agglutinine de germe de blé et la jacaline). Cette étude a conduit à l'identification de 1823 protéines urinaires.

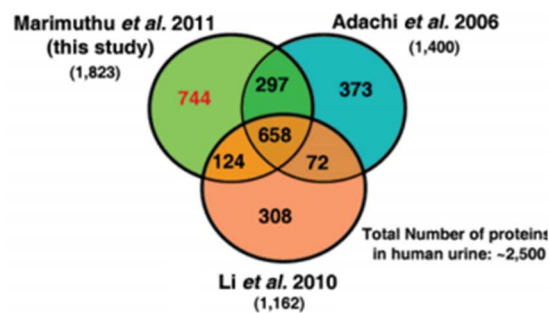


Figure I.15 : Overlap entre trois études sur la caractérisation du protéome urinaire

Etant donné le faible recouvrement qui représente environ un quart de la totalité des protéines identifiées sur trois études récentes de caractérisation du protéome urinaire, on peut se demander si la couverture du protéome urinaire est complète. Une base de référence de l'ensemble des protéines urinaires permettrait l'identification de nouveaux biomarqueurs plus spécifiques et plus sensibles pour différentes pathologies uro-génitales, tel que le cancer de la vessie.

E. Conclusion

Depuis plusieurs années, il y a eu de nombreuses avancées dans le domaine de la spectrométrie de masse avec l'arrivée sur le marché d'instruments de plus en plus rapides et précis permettant l'identification et la quantification d'un grand nombre de peptides et de protéines. De la même façon, de nouvelles méthodes de séparation des protéines ou des peptides sont apparues permettant, en amont de la spectrométrie de masse, de mieux séparer les peptides ou les protéines. Et enfin, dans le domaine de la quantification sans marquage, ont émergé de nouvelles méthodes d'extraction du signal en sortie des spectromètres de masse, des méthodes pour traiter ces données et pour extraire des informations pertinentes pour la découverte de nouveaux

biomarqueurs - notamment dans le domaine de la statistique. Il nous a donc semblé opportun de réunir un certain nombre de ces avancées dans le but de caractériser le protéome urinaire de manière très exhaustive.

II. Chapitre II : Matériel et Méthodes

A. Introduction

Une étude de protéomique clinique telle que celle présentée dans cette thèse nécessite de faire appel à de nombreuses techniques et protocoles pour la préparation des échantillons, la qualification des protéines de ces échantillons, leur analyse par spectrométrie de masse, et le traitement des données. Ce chapitre décrit en détail les méthodes et modes opératoires utilisés au cours de ce travail de thèse, du traitement biochimique des échantillons d'urine et de leurs analyses protéomiques par spectrométrie de masse, aux traitements statistiques des données générées par l'analyse en spectrométrie de masse.

B. Techniques de Biochimie

1. Protocole préconisé pour la collecte et le prétraitement des échantillons d'urine

La protéolyse non contrôlée est un des artéfacts majeurs qui peut survenir lors de la collecte et le stockage des échantillons biologiques. Pour pallier à ce problème, l'utilisation d'inhibiteur de protéases lors de la collecte des fluides biologiques est parfois recommandée. Cependant, ces recommandations sont controversées car les inhibiteurs de protéases peuvent interférer lors des analyses protéomiques (Marshall et al., 2003). Par conséquent, une partie de notre étude a été dédiée aux conditions de collecte de l'urine (voir Chapitre III), et plus précisément, à l'influence de la température de conservation des échantillons d'urine avant leur congélation à -80°C , et à l'utilisation d'inhibiteurs de protéases. L'utilité d'ajouter des inhibiteurs de protéases a été démontrée ainsi que la nécessité de limiter le temps de conservation des échantillons urinaires avant centrifugation.

Les échantillons d'urines ont été collectés dans des tubes de polycarbonate (n°352070, Dutscher, Brumath, France) dans lesquels a été ajouté préalablement une pastille d'inhibiteur de protéases « Complete EDTA Free » (Roche réf. : 14967900). Un premier bilan chimique des urines a été réalisé, pour chaque patient, à l'aide de bandelettes réactives Combur test 10 strips (Roche Diagnostics, Basel, Switzerland). Ces bandelettes permettent de déterminer simultanément dix paramètres : la densité et le pH, le taux approximatif ou la présence de leucocytes, nitrites, protéines, glucose, cétones, urobilinogène, bilirubine et de sang (hémoglobine en érythrocytes). Ce bilan chimique a permis aux cliniciens de sélectionner les patients de la cohorte de découverte. Une fois le bilan effectué, l'échantillon d'urine attend sur la pailasse à température ambiante au maximum 4 heures avant d'être traité. Cette attente est inhérente au fait que la collecte a lieu dans le service d'urologie d'un hôpital. Le technicien en charge de la collecte doit s'occuper des patients qui viennent en consultation, et du traitement des échantillons pour le programme de recherche. La priorité numéro 1 est la prise en charge du patient et le travail de recherche doit interférer le moins possible avec cette responsabilité. L'urine est ensuite centrifugée à basse vitesse (2,000 g), pendant 10

minutes et à température ambiante, de manière à éliminer les débris cellulaires. Le surnageant est finalement aliquoté dans 2 Falcons de 15 mL, 7 cryotubes de 2 mL et un cryotube de 500 μ L, qui sont congelés à -80°C . Le transfert d'échantillons à analyser au laboratoire de protéomique a été effectué à cette même température à -80°C .

2. Dosage des protéines urinaires

a) Nécessité d'une normalisation

Pour comparer des échantillons entre eux, une étape de normalisation est nécessaire afin de réduire autant que possible la variabilité des échantillons. Cette normalisation peut s'effectuer à posteriori sur les données issues des analyses par spectrométrie de masse ou en amont lors de la préparation des échantillons (en normalisant par exemple les volumes d'urine extraits, ou les quantités de protéines traitées).

La concentration des protéines dans les urines peut être très variable d'un individu à l'autre ou chez un même individu selon l'heure de la journée, son alimentation, ou encore son niveau d'activité. Cette variabilité, souvent évoquée, a été comparée à la variabilité technique dans une publication de Nagaraj (Nagaraj & Mann, 2011). Dans ce rapport, ils ont comparé parmi 7 donneurs sains la variabilité inter et intra individus à la variabilité technique de l'analyse protéomique en mesurant les coefficients de variation des abondances mesurées (CV). La valeur calculée des CV était de 66%, 48% pour la variabilité inter-individu, et 18% pour les réplicats techniques. Nous avons, nous aussi, vérifié cette variabilité en normalisant les échantillons d'urine par le volume. Pour cela, nous avons préparé 1 mL d'urine de 7 donneurs sains sur membrane Vivaspine. Après avoir ultrafiltré et dessalé les urines, nous avons déposé les protéines sur un gel SDS-PAGE pour purifier et concentrer l'échantillon protéique dans un mode que l'on appelle « Stacking ». Pour cela on laisse migrer, juste assez de temps pour que l'ensemble des protéines entrent dans le gel. Sur la Figure II.1, on observe des différences majeures de coloration sur les différents échantillons. Ces résultats ont été confirmés par l'analyse protéomique. Ainsi, les échantillons entourés en rouge, présentant la plus forte coloration des gels n'ont permis de quantifier qu'un tiers des peptides détectés dans les autres échantillons.



Figure II.1 : 1mL d'urine de patients centrifugé, ultrafiltré puis déposé en stacking sur le gel. Les différences de coloration au bleu de Coomassie illustrent les variations de concentration protéique d'un échantillon à l'autre.

La normalisation par le volume d'échantillon traité a donc été écartée et nous avons repensé la normalisation des urines en considérant cette fois-ci les quantités de protéines présentes dans l'urine.

b) Dosage de protéines

Le dosage des protéines urinaires est un examen couramment pratiqué dans les laboratoires d'analyses biologiques et médicales. Actuellement, les méthodes les plus utilisées sont les dosages colorimétriques. Le principe de ces dosages repose sur le déplacement du maximum d'absorption suite à la réaction entre certains groupements chimique des protéines et un colorant. La différence d'absorption entre les deux spectres d'absorptions est proportionnelle à la concentration en protéines. Dans le cadre de cette thèse nous avons utilisé le dosage de Bradford (Bradford, 1976).

Le dosage de Bradford est sans aucun doute la méthode la plus répandue en protéomique pour le dosage de protéines. Cette quantification des protéines totales est basée sur la modification de la longueur d'onde maximale d'absorption de 460nm à 595 nm du bleu de Coomassie une fois fixé sur les protéines. La quantification des protéines est réalisée par référence à une gamme étalon de sérum albumine bovine (BSA). L'absorbance lue à 595 nm après 10 minutes de réaction est directement proportionnelle à la concentration protéique. Cette méthode de dosage a été référencé comme étant une des méthodes les plus sensible pour le dosage de complexe protéiques (Chutipongtanate, Watcharatanyatip, Homvises, Jaturongkakul, & Thongboonkerd, 2012). Nous avons utilisé cette technique de dosage après l'étape d'extraction des protéines urinaires. En effet, cette méthode de dosage est sensible aux solutions tamponnées et n'est donc pas approprié sur des urines brutes.

Dans un premier temps, nous avons réalisé la gamme étalon avec une quantité croissante de BSA (0.15, 0.31, 0.635, 1.250 et 2 mg/mL). Puis nous avons prélevé 4 µL de cette gamme auxquels on a ajouté 200 µL de réactif Bradford dilué 4 fois dans de l'eau. On a fait de même pour les échantillons urinaires à doser en répétant 3 fois l'expérience de manière à établir une moyenne de l'absorbance mesurée. Si l'échantillon urinaire est trop concentré, il est possible de le diluer avant le dosage en rapportant le facteur de dilution au moment du calcul de la concentration protéique. Parallèlement, on a réalisé un blanc en prélevant 4 µL d'eau auxquels nous avons ajouté 200 µL de réactif Bradford.

Après une incubation de 10 minutes à température ambiante, la quantité de protéines a été estimée par la mesure de l'absorption à 595nm à l'aide d'un spectrophotomètre. La quantité de protéines est déterminée par rapport à la gamme d'étalon obtenue selon la formule :

$$\frac{[Moyenne (Abs \text{ échantillon}) - Abs \text{ blanc}] - b}{a} \times \text{facteur dilution}$$

Où a et b représente respectivement le coefficient directeur et l'ordonnée à l'origine de la droite $y = ax + b$ obtenue à partir de la gamme étalon de BSA.

3. Préparation des échantillons en vue de l'analyse protéomique

a) Extraction des protéines urinaires

Plusieurs protocoles ont été décrits pour extraire les protéines d'un échantillon complexe. La méthode de précipitation en milieu acide représente la méthode la plus utilisée et la plus pratique pour la préparation des échantillons (Rajalingam, Loftis, Xu, & Kumar, 2009). En effet, les protéines sont solubles à l'intérieur d'une gamme restreinte de pH (entre 4 et 6). Le fait d'acidifier l'échantillon rend les protéines insolubles et induit leur dénaturation et leur précipitation.

(1) Précipitation à l'acide trichloroacétique

Les protéines sont précipitées grâce à l'ajout de 2.5mL d'une solution de TCA à 30% dans 10mL d'urine afin d'obtenir une concentration en acide finale de 6%. L'échantillon est ensuite incubé pendant 2 heures à 4°C. Les protéines ayant précipité sont récupérées par une étape de centrifugation à 14000 g pendant 15 minutes à 4°C. Afin d'éliminer les traces de TCA résiduels, on réalise 2 lavages successifs avec 2.5mL d'acétone froid (conservé à -20°C). Pendant ces étapes de lavage, le culot est remis en suspension dans 1.5mL d'acétone en opérant plusieurs allers-retours dans un cône de micropipette. Le mL d'acétone restant est utilisé pour rincer le cône qui a servi pour resuspendre le culot. Cette étape est critique pour l'élimination de l'acide qui peut affecter l'efficacité de digestion enzymatique des protéines. On centrifuge ensuite l'échantillon à 14000 g à 4°C pendant 15 minutes. Au final, on resuspend le culot avec 300µL d'une solution d'urée 8M, 0.1mM de bicarbonate d'ammonium. De manière à faciliter la solubilisation des protéines, on incube l'échantillon pendant 15 minutes à 37°C.

(2) Ultrafiltration

L'ultrafiltration (UF) est une filtration à travers une membrane semi-perméable grâce à une différence de pression. Les particules de haut poids moléculaire, en solution ou en suspension, sont retenues tandis que l'eau et les molécules de faible poids moléculaire passent à travers la membrane.

Dans notre étude, les protéines urinaires ont été concentrées en utilisant des concentrateurs centrifuges (vivaspin 2, 5kDa Mw cutoff, PES membrane, Sartorius Stedim, Aubagne, France). Après avoir déposé 2mL d'urine dans le concentrateur, le tube est centrifugé à 5000g jusqu'à réduire le volume final à 100µL. On ajoute ensuite à deux reprises 2mL d'eau pour laver l'extrait des sels contenus dans les urines. On centrifuge à nouveau de manière à atteindre un volume final de 100µL. Pour décrocher les protéines de la membrane, on dépose 30µL de tampon Laemmli 5X, on agite au vortex afin de resuspendre les protéines. Après avoir retourné le concentrateur et centrifugé l'extrait dans la partie haute, on ajuste le volume final à 75µL avec de l'eau milliQ.

b) Fractionnement - Gel SDS PAGE

L'électrophorèse en gel de polyacrylamide contenant du laurylsulfate de sodium ou SDS-PAGE est une technique qui permet de séparer les protéines selon leur poids moléculaire. Le SDS (Sodium dodecyl sulfate) empêche le repliement des protéines et leur confère une charge négative permettant la migration des protéines dans la matrice sous l'effet d'un potentiel électrique. La séparation s'effectue ainsi uniquement selon le poids moléculaire apparent, les protéines ayant un poids moléculaire plus faible migrent plus rapidement dans le gel. Les protéines sont préalablement dénaturées par chauffage de l'échantillon protéique à 95°C pendant 5 minutes. L'extrait est ensuite déposé sur un gel précoulé Novex NuPAGE Bis-Tris gels (4–12% acrylamide gradient) (Invitrogen, Cergy Pontoise, France). Parallèlement, on dépose sur une autre piste du même gel un marqueur protéique de poids moléculaires (Sigma). Pour une migration complète des protéines sur le gel on applique un voltage de 200V pendant environ 30 minutes jusqu'à ce que le front de migration approche de l'extrémité du gel. Le gel est ensuite démoulé et incubé dans une solution de fixation contenant 30% d'éthanol et 7.5% d'acide acétique pendant 30 minutes. Cette étape est renouvelée une fois. On colore ensuite le gel avec une solution de bleu de Coomassie R-250 (Bio-Rad, Marnes-La-Coquette, France).

c) Digestion des protéines urinaires

(1) Digestion sur gel et extraction des peptides

Après la coloration des gels au bleu de Coomassie et les étapes de lavage des gels avec une solution d'éthanol 30%, acide acétique 7.5%, les bandes de gel sont découpées à l'aide d'un scalpel puis traités comme décrit ci-dessous.

Décoloration des bandes de gel : Les bandes de gels sont lavées pendant 30 minutes et sous agitation à température ambiante successivement avec (1) 200µL de bicarbonate d'ammonium 25mM, 50% acétonitrile, (2) 200µL d'eau et pour finir (3) 200µL d'acétonitrile 100%. Les bandes de gels sont ensuite passées aux ultrasons et séchées au SpeedVac.

Réduction et alkylation : Après l'étape de décoloration, les ponts disulfures des protéines sont réduits avec du Dithiothréitol (DTT). Pour cela, on incube chaque bandes de gel avec 200µL d'une solution de DTT 10mM pendant 45 minutes à 56°C. Les cystéines sont ensuite alkylées par iodoacétamide qui conduit à une carboxyamidométhylation. Chaque bande de gel est incubée avec 200µL d'une solution d'iodoacétamide 55mM pendant 35 minutes à température ambiante dans le noir. Pour finir la réaction d'alkylation est bloquées avec 200µL d'une solution de DTT 10mM pendant 10 minutes à température ambiante. Pour finir, on rince les bandes de gels avec 200µl d'une solution de bicarbonate d'ammonium 25mM.

Oxydation des bandes de gel : Une alternative à la réduction et l'alkylation consiste en l'oxydation des protéines contenues dans le gel. Ce protocole est plus simple et beaucoup moins long, cependant les méthionines peuvent être oxydées sous différentes formes (méthionine sulfoxyde ou méthionine sulfone). Les bandes de gel préalablement lavées et séchées sont réhydratées avec 200µl d'une solution H₂O₂ à 7% et incubées pendant 30 minutes à température ambiante. Après élimination du surnageant, les bandes de gel sont ensuite lavées avec 200µl d'eau milliQ puis 200µl d'une solution à 100% d'acétonitrile. Pour finir les bandes de gels sont séchées au SpeedVac.

Digestion des protéines : La digestion des protéines est réalisée sur gel avec une solution de trypsine. La solution de trypsine (0.4µg/µL) est obtenue par solubilisation des 20µg de lyophilisat dans un volume de 50µL de solution d'acide acétique 50mM, selon les consignes du mode opératoire Promega Part 9PIV511. Ensuite la solution est diluée à 0.006µg/µL avec du bicarbonate d'ammonium 25mM et on ajoute ensuite 25µL de cette solution sur chaque bande de gel. Le gel se réhydrate pendant 15 minutes à 4°C, avant de rajouter 30µL de bicarbonate d'ammonium 25mM pour réhydrater complètement le gel et avoir un excédent de tampon. L'échantillon est ensuite incubé à 37°C sous agitation pendant toute une nuit.

Extraction des peptides : Dans un premier temps on transfère le surnageant de digestion dans un tube Multi low-adsorption (Dutscher, Brumath, France). Puis 200µL d'acétonitrile 50% sont ajoutés dans le tube contenant les bandes de gel. L'échantillon est incubé pendant 15 minutes à température ambiante et sous agitation, et on récupère le surnageant que l'on ajoute au tube contenant le premier surnageant de digestion. Le même protocole est utilisé pour une solution de 200µL d'acide formique 5% puis 200µL d'une solution à 100% d'acétonitrile. Suite à ce dernier traitement, l'échantillon est soniqué avant que le surnageant soit transféré dans un tube. L'échantillon final est ensuite séché au SpeedVac et conservé à -80°C.

(2) Digestion en solution et dessalage des peptides

Cette étape est réalisée sur les échantillons protéiques repris dans l'urée 8M et le bicarbonate d'ammonium 100mM suite à la précipitation TCA. (cf. paragraphe 3.a.i.)

Réduction et alkylation des protéines: Les ponts disulfures des protéines sont réduits à l'aide de Dithiothréitol (DTT). On ajoute 5.45 µL de DTT 700mM à l'échantillon de manière à obtenir une concentration finale de 12.5mM ; on incube cette solution à 37°C sous agitation (800 rpm) pendant 30 minutes. Les cystéines sont ensuite alkylées à l'iodoacétamide, ce qui conduit à une carboxyamidométhylation. Pour ce faire, 18.5 µL d'Iodoacétamide 700mM sont ajoutés à l'échantillon de manière à obtenir une concentration finale de 40mM et on incube à 25°C, sous agitation (800rpm) dans le noir pendant 30 minutes. Le volume est ajusté à 1 mL avec une solution de bicarbonate d'ammonium 100mM.

Digestion des protéines en solution: Pour digérer les protéines en solution, la trypsine (Promega, Charbonnières-les-bains, France) est ajoutée à l'échantillon dans un ratio 1/100 (poids / poids) et l'échantillon est incubé 12h à 37°C sous agitation (1000 rpm). Après la première digestion, un 2e ajout de trypsine dans un même ratio 1/100 (poids/poids) est effectué et l'échantillon est incubé 30 min à 37°C et sous agitation (1000 rpm). Lorsque les protéines sont digérées en solution, il est possible de vérifier que la digestion est complète sur un gel SDS-PAGE. Pour cela, nous avons déposé sur un gel (Figure II.2), la même quantité de protéines avant (échantillon B) et après digestion en solution (échantillon A), et visualisé sur la piste de l'échantillon digéré s'il restait des protéines intactes ou des fragments de haut poids moléculaire. Cette vérification a été effectuée sur un échantillon après extraction des protéines urinaires par précipitation TCA : après avoir dosé les protéines de l'échantillon par la méthode de Bradford, nous avons déposé l'équivalent de 30µg de protéines digérées sur la piste A et le même échantillon précipité et non digéré B.

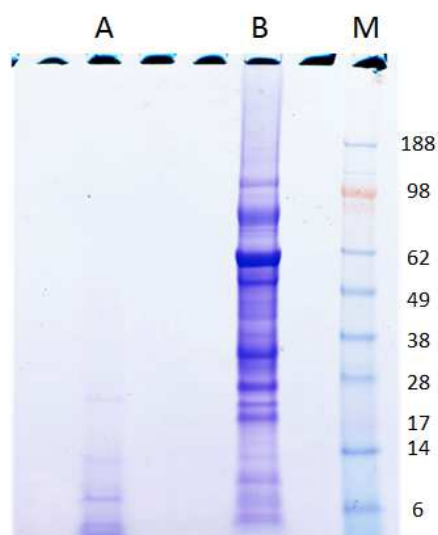


Figure II.2 : Profil protéique d'un échantillon digéré (piste A) et du même échantillon non digéré (piste B) sur Gel SDS-PAGE.

Sur la piste A correspondant à l'échantillon digéré, aucune bande correspondant à des protéines non digérées n'est observée. Seuls quelques fragments de faibles poids moléculaires (< 14 kDa) sont observés. Ce résultat démontre une digestion quasi-complète de l'échantillon urinaire. Cette vérification n'est malheureusement pas triviale dans le cas de la digestion des protéines sur gel, et n'a pu être effectuée.

Dessalage des peptides : La solution de peptides obtenue après digestion des protéines en solution est dessalée sur une colonne de phase inverse greffées en C18 (Sep-Pak tC18 reverse phase cartridges, 100 mg (référence WAT036820, Waters, Milford, MA). Pour cela, la colonne est réhydratée avec 1mL de méthanol, puis lavée avec 1mL d'une solution d'acétonitrile 80%. La colonne est ensuite équilibrée avec 4mL d'une solution d'acide formique 0.1%. L'échantillon est acidifié avec 90µL d'une solution d'acide

formique 10%. On vérifie que le pH de l'échantillon peptidique est approximativement entre 2 et 3 avec du papier pH, avant son dépôt sur la colonne. Le premier effluent est récupéré pour le redéposer sur la colonne. Puis la colonne est lavée avec 6mL d'une solution d'acide formique 0.1%. Enfin, les peptides sont élués avec une solution composée d'acide formique 0.1% et d'acétonitrile 50%. L'échantillon est ensuite séché au SpeedVac et conservé à -80°C. Pour une conservation optimale des peptides, les échantillons peptidiques sont stockés à -80°C dans des tubes Multi low-adsorption avant de les solubiliser en vue de l'analyse par LC-MS ou LC-MS/MS (Kraut et al., 2009).

d) Répétabilité des méthodes d'extraction

Afin de vérifier la bonne répétabilité des méthodes d'extraction des protéines, nous avons comparés les réplicats analytiques et les réplicats techniques obtenus à partir des analyses LC-MS. Pour ce faire, nous avons préparé un pool d'urine à partir de deux donneurs sains que nous avons divisé en 8 aliquots traités en parallèle par deux protocoles. Dans le premier, les protéines extraites par TCA ont été réduites, alkylées, puis digérées à la trypsine en solution, les peptides dessalés ont été analysés par LC-MS. Dans le second, les protéines extraites par ultrafiltration, ont été digérées sur gel puis les peptides ont été analysés en LC-MS. Les réplicats analytiques correspondent au même échantillon injecté plusieurs fois sur la chromatographie liquide et analysé sur le spectromètre de masse.

A partir de l'abondance des peptides quantifiés, nous avons calculés les corrélations entre paires d'échantillons pour les réplicats analytiques et techniques (Figure II.3). Le coefficient de corrélation représente la relation linéaire entre deux variables. On observe une dispersion plus importante entre les abondances mesurées dans la paire de réplicats techniques comparée à celle des réplicats analytiques. Cette dispersion est due à la préparation de l'échantillon.

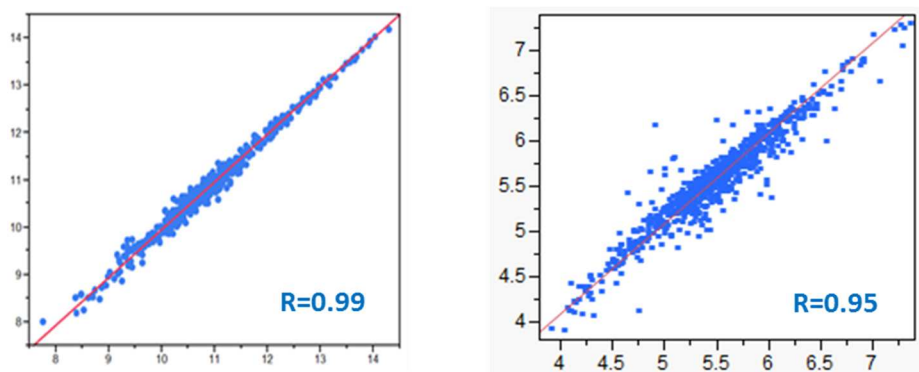


Figure II.3 : Exemple de corrélation des abondances entre deux réplicats analytiques (à gauche) ou techniques (à droite) R^2 est la valeur du coefficient de corrélation (Pearson).

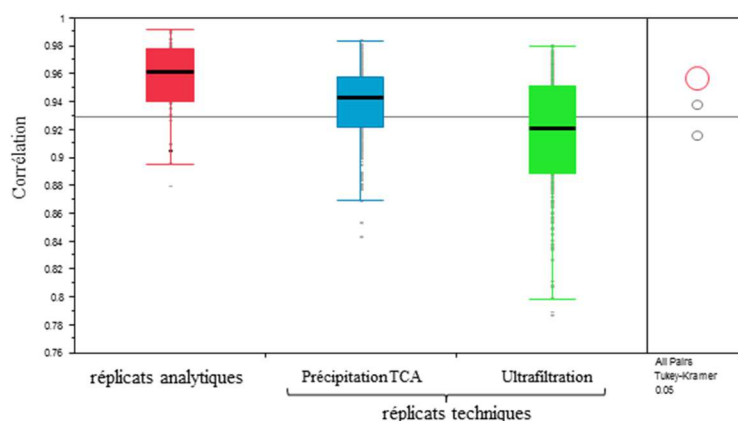


Figure II.4 : Comparaison des distributions des coefficients de Pearson obtenues pour des réplicats analytiques et des réplicats techniques

La Figure II.4, montre les dispersions des coefficients de Pearson obtenues entre des réplicats analytiques (même échantillon analysé deux fois), et des réplicats techniques pour les deux protocoles étudiés. On observe sur cette Figure, que les médianes des coefficients de corrélation sont supérieures à 0,9, ce qui représente une valeur élevée et donc une bonne répétabilité des techniques de préparations des urines que ce soit par précipitation TCA ou par ultrafiltration. La variabilité introduite par la technique utilisée pour la préparation et l'analyse des échantillons doit être minimale, afin de permettre la détection de faibles variations d'abondances protéiques dues spécifiquement à la pathologie. Ainsi, les deux « pipeline » reposant sur l'extraction des protéines par précipitation ou ultrafiltration, semblent suffisamment répétable pour être utilisée dans le cadre d'une étude de découverte des biomarqueurs. Cependant, les résultats obtenus par précipitation TCA sont meilleurs que ceux obtenus par la méthode d'ultrafiltration. Le test de Wilcoxon, nous permet de conclure un effet significatif du traitement ($p < 0,0001$), le seuil de significativité étant fixé à 0,05.

4. Déplétion et fractionnement des protéines urinaires

a) Fractionnement par isoélectrofocalisation

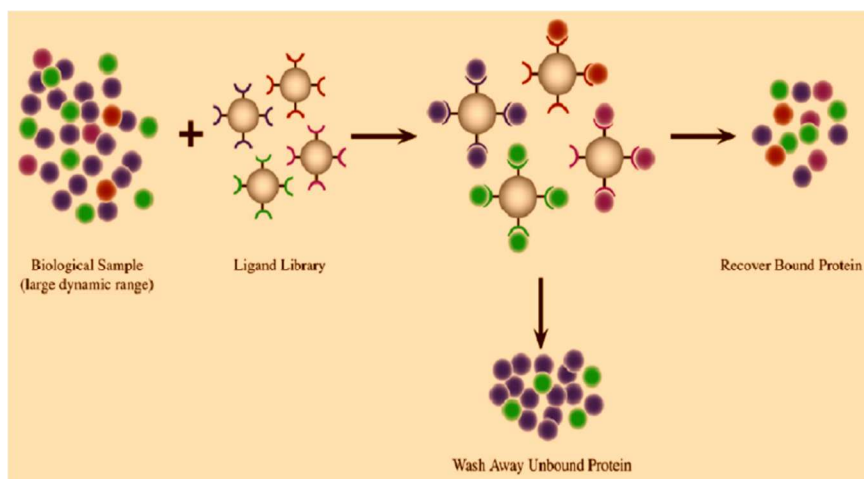
Les peptides ont été séparés en fonction de leur points isoélectriques en utilisant le « 3100 OFFGEL Fractionator » commercialisé par Agilent Technologies avec des bandes « Immobiline DryStrip » de longueur 13 cm, couvrant une gamme de pH de 3 à 10. Nous avons utilisé le protocole fourni par le constructeur (http://www.chem.agilent.com/Library/usermanuals/Public/G3100-90113_OFFGEL_Fractionator_QuickStart.pdf). Une solution stock « Peptide OFFGEL 1.25X » est préalablement préparée pour la reprise des peptides et la réhydratation des strips. Cette solution stock contient 6mL d'une solution de Glycérol, et 600 μ L de tampon « OFFGEL Buffer » (pH 3-10). Ce mélange est ensuite ajustée jusqu'à 50mL avec de l'eau milliQ. Cette solution est ensuite aliquoté puis stockée à -20°C dans une limite de 4 mois. Les strips sont réhydratés dans une solution : « Peptide IPG Strip Rehydration solution » contenant 336 μ L de la

solution « Peptide OFFGEL 1.25X » et 84µL d'eau milliQ. Le système est ensuite installé selon la procédure décrite dans le guide du constructeur. Avant l'isoélectrofocalisation, le mélange peptidique est tout d'abord dessalé sur une cartouche de phase inverse C18 (Sep-Pak C18, Waters Corporation, Milford, MA, USA) (voir ci-dessus). Les échantillons sont ensuite repris avec une solution contenant 1.44 mL de solution stock « Peptide OFFGEL 1.25X » et 360µL d'eau milliQ. Après avoir réhydraté les « strips » avec 40µL de la solution « Peptide IPG Strip Rehydration solution », 150 µL d'échantillon peptidique sont déposés dans chacun des 12 puits. Pour la focalisation des peptides, nous avons utilisé la méthode par défaut pour les peptides : OG12PE00. Cette étape se déroule à 20°C, et jusqu'à ce que 20 kVh soient délivrés. Une fois le fractionnement terminé, les 12 fractions peptidiques sont collectées et dessalées sur cartouche de phase inverse C18 (Harvard Apparatus Micro SpinColumns).

Dessalage des peptides sur colonne Harvard Apparatus : Afin d'éliminer les ampholytes présents dans les échantillons fractionnés par IEF, on effectue un dessalage des peptides sur des colonnes C18. Dans un premier temps, on acidifie l'échantillon peptidique avec quelques µL d'acide trifluoroacétique 1%, jusqu'à un pH compris entre 2 et 3 (vérification effectuée avec du papier pH). On prémouille ensuite la colonne Harvard avec 150 µL d'une solution d'acide formique 0.1%, acétonitrile 50%. On centrifuge à 750 g pendant 90 secondes. On lave la colonne avec 150 µL d'une solution d'acide formique 0.1%, acétonitrile 5%, et on centrifuge (750 g, 90 secondes). L'échantillon peptidique acidifié, est chargé puis centrifugé (750 g, 90 s). On lave la colonne avec 150µL d'une solution d'acide formique 0.1%, acétonitrile 5%, suivi d'une centrifugation (750 g, 90 s). Pour finir, on élue les peptides de la colonne avec 50µL d'une solution d'acide formique 0.1%, acétonitrile 50% avec une étape de centrifugation à 1000 g pendant 90 seconde. On renouvelle l'étape d'élution avec le même volume de cette solution. Au final, on sèche les échantillons qui sont ensuite conservés à -80°C.

b) Enrichissement Equalizer

Cette technologie permet de réduire la gamme dynamique en égalisant les concentrations protéiques d'un fluide biologique. Pour cela, le fluide biologique est incubé en présence d'une banque combinatoire d'hexapeptides (6 acides aminés, soit 206 ou 64 millions de peptides ligands différents possibles) greffés sur des microbilles. Chaque bille porte des millions de copies d'un unique ligand. Théoriquement, chaque protéine de l'échantillon complexe possède un ligand qui lui est propre. Un des paramètres essentiels est que le ratio entre la capacité de charge des billes et la quantité totale de protéines de l'échantillon soit contrôlé pour permettre une déplétion efficace. Une fois les sites des microbilles saturés, les protéines en excès sont éliminées avec le surnageant, permettant ainsi de réduire les différences de concentration entre les protéines les plus abondantes et les moins abondantes de l'échantillon (voir Figure II.5).



Thulasiraman et al. *Electrophoresis* 2005, 26: 3561–3571

Figure II.5 :Principe de la stratégie Equalizer visant à diminuer la dynamique de concentration d'un protéome complexe

Ce type d'enrichissement est commercialisé par Bio-Rad sous le nom de ProteoMiner™. La déplétion partielle des protéines urinaires a été réalisée à l'aide de ce kit d'enrichissement. Pour l'enrichissement des échantillons d'urine, nous nous sommes basés sur le protocole original publié par Castagna (Castagna et al., 2005). Cependant, n'ayant pas accès à de tels volumes d'urines, nous avons dû l'adapter à nos conditions (voir Chapitre IV).

Traitement de l'urine avec les billes Equalizer : Dans un premier temps, les bouchons sont retirés des deux extrémités de la colonne dans laquelle sont stockées les billes du ProteoMiner™. La colonne contenant les billes est ensuite centrifugée à 1000g pendant 2 minutes de manière à éliminer le liquide de stockage. Un volume de 1mL d'eau milliQ est déposé sur la colonne que l'on place sur un agitateur rotatif à tambour, pendant 5 minutes. La colonne est ensuite centrifugée pendant 2 minutes à 1000 g, et l'éluat jeté. L'étape de centrifugation est renouvelée une fois. Les billes sont lavées avec 1mL de tampon de lavage, fourni dans le kit, sous agitation (agitateur rotatif à tambour) pendant 5 minutes puis la colonne est centrifugée pendant 2 minutes à 1000 g et l'éluat jeté. L'étape de centrifugation est renouvelée une fois. De nouveau, on ajoute 1mL de tampon de lavage sur les billes mais cette fois-ci, après avoir agité les billes pendant 5 minutes sur le tambour, on transfère les 16µL de billes sur les volumes d'urine. Les tubes (falcons) contenant l'urine et les billes sont alors agités pendant 3 heures à température ambiante sur le tambour. Les tubes sont ensuite centrifugés à 1000g pendant 2 minutes puis on laisse sédimenter les billes dans le fond du Falcon jusqu'à ce que le liquide paraisse limpide. Cette étape critique pourrait être améliorée en remplaçant les billes de polyméthacrylate par des billes magnétiques, ce qui faciliterait ainsi la récupération des billes. La quasi-totalité du surnageant est éliminé en laissant approximativement 1 mL d'urine sur les billes pour retransférer les billes dans les colonnes « Bio-Rad » prévues avec le kit. On centrifuge la colonne à 1000g pendant 2 minutes, en jetant l'éluat. On réalise deux nouvelles étapes de

lavage comme précédemment avec le tampon de lavage du kit (agitation et centrifugation). A chaque étape de lavage on récupère l'éluat pour le transférer sur des concentrateurs Vivaspin, afin d'évaluer la quantité de protéines non retenues sur un gel SDS-PAGE. On réalise ensuite une étape de lavage des billes avec 1mL d'eau milliQ (agitation et centrifugation). Puis, on ajoute 40µL de bleu de Laemmli 2.5X afin de décrocher les protéines de leurs ligands. Enfin, on centrifuge les colonnes, sur une centrifugeuse de paillasse, de manière à faire tomber le bleu de manière afin d'éviter qu'il ne s'adsorbe sur le filtre de la colonne. On chauffe le bleu de Laemmli à 95°C pendant 4 minutes, puis on dépose 10 µL sur un gel SDS-PAGE (4-12%). Le tampon utilisé pour la migration est le tampon MOPS (MOPS 1M, TrisBase 1M, SDS 70mM, EDTA 20mM, solution 20X commercialisée par Novex, USA).

5. Analyse en spectrométrie de masse

a) Reprise des échantillons peptidiques

Les échantillons peptidiques obtenus par digestion des extraits de protéines urinaires sont repris avec une solution 1% TFA et 5% acétonitrile et soniqués. L'échantillon est ensuite dilué 10 fois dans le tampon de chargement du système de chromatographie liquide (tampon C, constitué de 0.2% d'acide formique, 2% d'acétonitrile et 97.8% d'eau milliQ). Il est important, particulièrement, lorsque l'on utilise une méthode basée sur les temps de rétentions et la masse pour identifier les peptides, de reprendre les échantillons dans exactement les mêmes conditions pour éviter des décalages de temps de rétention.

b) Chromatographie liquide

Les échantillons peptidiques sont séparés sur une chaîne de chromatographie liquide à nano débit «Ultimate 3000» (Dionex), pilotée par le logiciel Chromeleon (Dionex). La chaîne de chromatographie liquide est montée dans une configuration « duale » représentée sur la Figure II.6. Cette configuration permet de monter en parallèle deux colonnes de séparation et ainsi d'analyser plus rapidement de grands nombre d'échantillons.

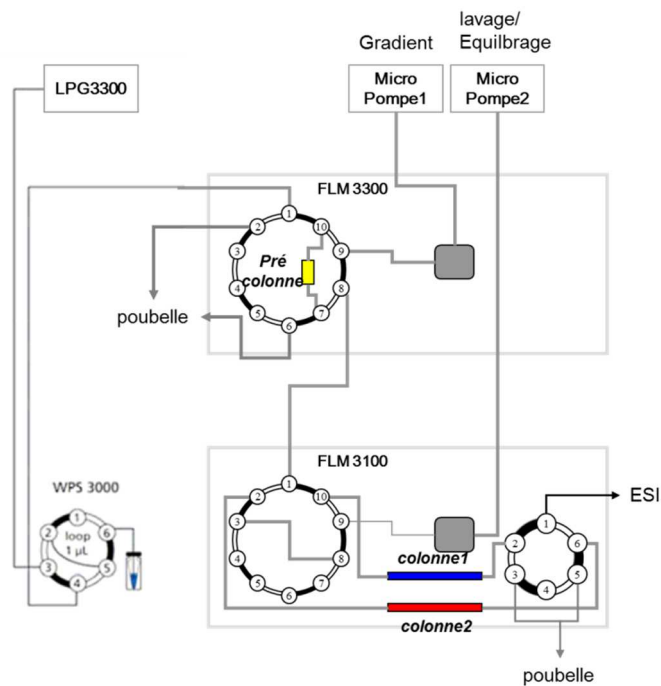


Figure II.6 : Représentation schématique de la configuration duale du système de chromatographie liquide utilisé.

Sur la chaîne de chromatographie liquide, on dispose de plusieurs modules, un premier qui comprend la pompe de chargement (LPG3300) de l'échantillon sur la précolonne avec un débit de 20 μ L/min. Un second module comprenant les deux pompes d'élution : annotées « Micro Pompe 1 » et « Micro Pompe 2 » sur la Figure II.6. Dans cette configuration, la « Micro Pompe 1 » est utilisée pour l'élution des peptides sur les deux colonnes de phase inverse C18, et la « Micro Pompe 2 » est utilisée pour le lavage des colonnes C18. Lors d'une analyse, pendant que l'on élue les peptides d'une des deux colonnes avec un débit de 300nL/min, on procède en parallèle au lavage de l'autre colonne avec le même débit. Après le chargement des peptides sur la précolonne, on applique un gradient de solvant décrit dans la Figure II.7 qui permet de séparer et d'éluer les peptides de la colonne de phase inverse. Les peptides ainsi élués sont ionisés dans la source nano-électrospray et détectés dans le spectromètre de masse. Les différentes phases mobiles utilisées pour le chargement et la séparation des peptides sur les colonnes de phase inverse sont les tampons A et B et C (C étant le tampon utilisé pour le chargement des peptides sur la précolonne) :

- La phase A étant constituée de 0.1% d'acide formique, 2% d'ACN et 97.8% d'eau milliQ
- La phase B étant constituée de 0.08% d'acide formique, 79.92% d'ACN et 20% d'eau milliQ.
- Le tampon C étant constitué de 0.2% d'acide formique, 2% d'acétonitrile et 97.8% d'eau milliQ.

Un volume de 10 μ L d'échantillon peptidique est injecté sur une boucle de 20 μ L en s'assurant que l'échantillon soit placé au milieu de la boucle d'injection encadré par du solvant de transport, puis l'échantillon est transféré sur une précolonne (300 μ m ID, 5 mm long, 5 μ m, 100 Å pores) à un débit de

20 μ L/min, de manière à le dessaler et le concentrer pendant 1 minute. Les peptides sont ensuite transférés sur une colonne de phase inverse C18 (75 μ m ID, 15 cm long, 3 μ m, 100 Å pores) afin d'être séparés. Un gradient linéaire (Figure II.7) comprenant une phase d'élution, un palier isocratique et enfin une rééquilibration de la colonne est appliqué. Pendant cette même période, la seconde colonne est lavée et rééquilibrée. Ainsi, entre chaque échantillon, la colonne est lavée et la configuration duale permet de ne pas perdre le temps d'une analyse pour effectuer ce lavage. La précolonne, quant à elle est lavée à la fin de la séparation des peptides pendant 3 minutes à 90% de solvant.

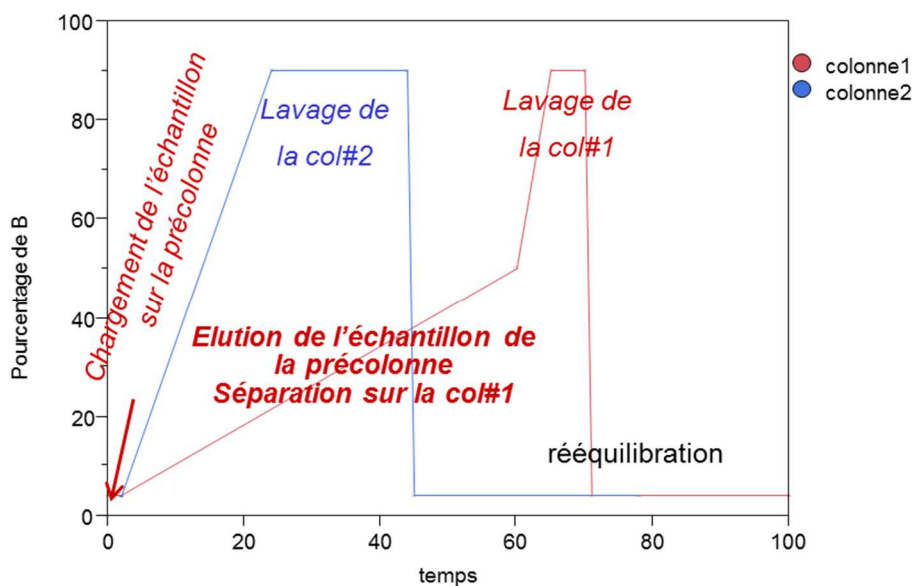


Figure II.7 : Représentation schématique des gradients appliqués sur les colonnes de chromatographie liquide : gradient d'élution des peptides (colonne 1) et gradient de lavage de la colonne (colonne 2)

c) Spectrométrie de masse

La stratégie d'analyse protéomique qui a été retenue pour ce travail de thèse est une méthode de quantification « Label Free » par une approche AMT. Brièvement, cette méthode consiste dans un premier temps à identifier les peptides présents dans un mélange complexe en utilisant une combinaison de techniques de fractionnement, nécessitant des analyses en mode MS/MS. Dans une deuxième phase, les peptides identifiés lors de la première phase sont quantifiés dans des analyses en mode MS. L'extraction des signaux MS est réalisée grâce à des logiciels adaptés (Decon2LS et VIPER). Ce type d'analyses requiert des performances instrumentales spécifiques qu'une vitesse d'acquisition, une résolution et une précision en masse élevées. Ainsi, l'ensemble des analyses de cette étude ont été réalisées sur un LTQ-Orbitrap XL (Thermo Fisher Scientific, Bremen, Germany).

(1) Analyses MS/MS

Dans la première phase de l'approche AMT, qui consiste à caractériser le protéome urinaire, nous avons réalisé des analyses par spectrométrie de masse en mode tandem (MS/MS). Dans le chapitre I, nous avons décrit le principe de l'instrument hybride LTQ-Orbitrap. La fragmentation des peptides est réalisée dans la trappe ionique. Pour cela, les trois ions les plus intenses du spectre d'ion précurseurs sont sélectionnés et fragmentés. Afin de ne pas fragmenter plusieurs fois le même peptide, une exclusion dynamique a été spécifiée, selon les paramètres listés dans le Tableau II.1.

Tableau II.1 : Paramètre d'acquisition pour l'analyse MS/MS

<i>mass tolerance</i>	$\pm 2 \text{ ppm}$
<i>repeat MS/MS acquisition</i>	<i>2 times</i>
<i>Repeat duration</i>	<i>30 s</i>
<i>exclusion duration</i>	<i>180 s</i>
<i>exclusion list size</i>	<i>500 entries</i>
<i>AGC (automated gain control) target</i>	<i>1e4</i>
<i>Maximum injection times</i>	<i>500ms</i>

(2) Analyses MS

Les analyses quantitatives ont été acquises en mode MS, dans l'Orbitrap. La gamme de masse sélectionnée s'étend de 400 à 1850 m/z avec une résolution de 60 000. L'AGC target (automated gain control) et le temps maximum d'injection des ions dans la trappe linéaire ont été fixés à $1e6$ et 500 ms. Ces paramètres avaient été optimisés auparavant pour permettre une meilleure couverture du protéome (Masselon, Kieffer-Jaquinod, Brugière, Dupierris, & Garin, 2008). Toutes les analyses ont été réalisées en triplicat par blocs et les échantillons ont été analysés dans un ordre aléatoire à l'intérieur de chaque réplicat.

d) Contrôles qualité des analyses

(1) Calibration du spectromètre de masse Orbitrap XL

Une calibration hebdomadaire du spectromètre de masse LTQ-Orbitrap XL a été effectuée afin d'analyser les échantillons dans des conditions optimales : calibration de la masse et de tous les paramètres qui peuvent avoir une influence sur la qualité du signal. La solution de calibration utilisée est un mélange de caféine, MRFA (Met-Arg-Phe-Ala) et Ultramark 1621 selon le protocole fourni par le constructeur Thermo Fisher Scientific. Cette solution est injectée en infusion et la calibration est réalisée en mode semi-automatique.

(2) Echantillons de contrôles des performances instrumentales

Cytochrome C : Evaluation des temps de rétention et de la qualité du signal : Afin de contrôler la reproductibilité des temps de rétention et de l'intensité des peptides tout au long d'une étude qui peut durer plusieurs semaines voire plusieurs mois, on analyse un échantillon contrôle qualité, au début de l'étude puis tous les 4 échantillons, et à la fin de l'étude, lorsque l'ensemble des échantillons ont été analysés. L'échantillon injecté pour ce contrôle est un digestat de cytochrome C (Thermo Scientific Dionex Cytochrome C Digest lyophilisat) repris dans 200µL d'une solution à 5% d'acétonitrile et 0.05% de TFA. Cet échantillon est aliquoté (5µL par aliquot soit 40pmol) et conservé à -20°C. Au moment de l'analyse on ajoute 195µL du tampon de chargement décrit dans la partie chromatographie liquide aux 5µL de Cyt C, on dilue ensuite 10 fois le mélange toujours dans le même tampon C. On injecte ensuite 10µL de cet échantillon dilué sur la précolonne de chromatographie liquide. Les temps de rétentions et l'intensité des peptides sont ensuite suivis à l'aide du logiciel XCalibur et retranscrit sur un fichier Excel (Figure II.8) pour vérifier qu'ils rentrent dans les critères d'acceptations que l'on a définis au préalable (+/- 3 min pour les mesures des temps de rétention et un facteur < 2.25 pour l'intensité des peptides). Ainsi, l'analyse de cet échantillon de « contrôle qualité » permet d'identifier rapidement d'éventuel décalage des temps de rétention ou la chute de sensibilité du spectromètre de masse.

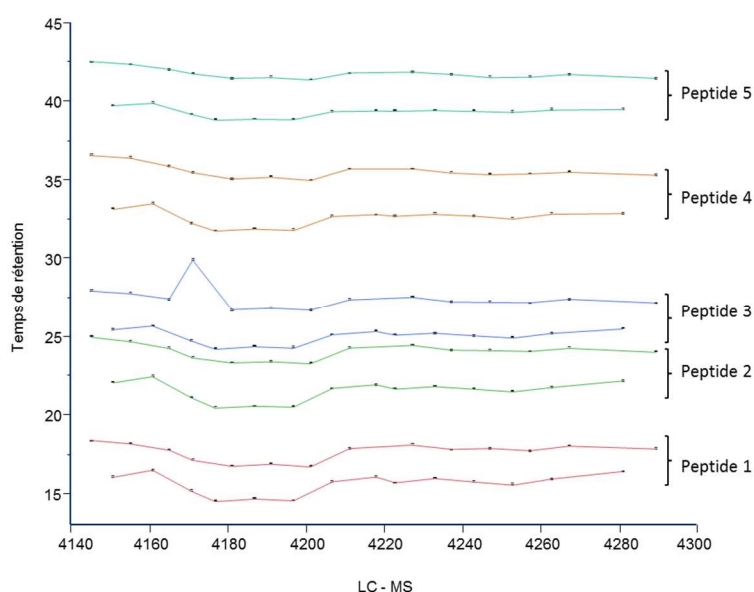


Figure II.8 : Suivi des temps de rétention de 5 peptides du Cytochrome C sur les deux colonnes de chromatographie liquide

Chloroplaste : Evaluation de la gamme dynamique : Le Cytochrome C permet un contrôle rapide des temps de rétention et de l'intensité du signal au cours d'une série d'analyse. Cependant avant d'analyser une série d'échantillons complexes, il est nécessaire d'évaluer la gamme dynamique que l'instrument permet

de couvrir en observant le nombre de peptides identifiés ou quantifiés. Pour ce faire, 1 µg d'un échantillon contrôle de chloroplaste d'*Arabidopsis Thaliana* est injecté sur la colonne de chromatographie et les peptides mesurés en mode MS. Le fichier de données brutes est ensuite traité avec la suite de logiciels Decon2LS et VIPER et les espèces détectées sont comparées à une base de donnée AMT du chloroplaste (Ferro et al., 2010) afin d'évaluer le nombre de peptides détectés et quantifiés. Pour pouvoir lancer la série d'analyse, il faut que le nombre de peptides quantifiés soit supérieur à 2300 sur cet échantillon. Ce critère a été défini sur la base d'analyses réalisées à la suite d'intervention du constructeur au cours desquelles la gamme dynamique du spectromètre de masse avait été optimisée.

(3) Suivi des analyses

L'ensemble des informations relatifs aux analyses des patients lors de la phase de quantification a été compilé dans un tableau, pour consultations ultérieures si observations de comportements aberrants de peptides ou d'analyses lors du traitement des données. On peut voir dans le tableau II.2, un extrait de ce tableau qui contient le nom du fichier pour les analyses des échantillons de patients ou d'échantillons « contrôles qualités », le nom des patients associés aux échantillons analysés, la méthode instrumentale utilisée, le volume de reprise des échantillons pour l'analyse d'1 µg de protéine totale, le nom de l'expérimentateur qui a préparé les échantillons, de celui en charge de l'analyse LC-MS, le jour et l'heure de la préparation de l'échantillon et enfin des événements survenus lors de l'analyse.

Tableau II.2 : Extrait de la liste des analyses LC-MS réalisée sur les 98 patients de la cohorte découverte

fichier	echantillon	Méthode instrument	volume de reprise	Operator	user	date heure prep	commentaires
CAVEN4141	INDIV_188011307	85min_1ITM SMS	224.2016361	court	court	16/07/10 9h30 AM	
CAVEN4142	INDIV_188111182	85min_1ITM SMS	405.4994567	court	court	16/07/10 9h30 AM	
CAVEN4143	INDIV_188112263	85min_1ITM SMS	764.5441225	court	court	16/07/10 9h30 AM	
CAVEN4144	INDIV_188111148	85min_1ITM SMS	200.581161	court	court	16/07/10 9h30 AM	
CAVEN4145	controlelc	85min_1ITM SMS	cytoc	court	court	15/07/10 9h30 AM	
CAVEN4146	INDIV_188011246	85min_1ITM SMS	5134.866486	court	court	16/07/10 9h30 AM	
CAVEN4147	INDIV_188011109	85min_1ITM SMS	2566.009006	court	court	16/07/10 9h30 AM	
CAVEN4148	INDIV_188011320	85min_1ITM SMS	762.4763167	court	court	16/07/10 9h30 AM	
CAVEN4149	INDIV_188011373	85min_1ITM SMS	374.5831997	court	court	16/07/10 9h30 AM	
CAVEN4150	controlelc	85min_1ITM SMS	cytoc	court	court	15/07/10 9h30 AM	
CAVEN4151	INDIV_188112177	85min_1ITM SMS	508.5416667	court	court	16/07/10 9h30 AM	
CAVEN4152	INDIV_188011284	85min_1ITM SMS	976.3709756	court	court	16/07/10 9h30 AM	
CAVEN4153	INDIV_188011335	85min_1ITM SMS	419.6400111	court	court	16/07/10 9h30 AM	
CAVEN4154	INDIV_188011121	85min_1ITM SMS	6749.139074	court	court	16/07/10 9h30 AM	
CAVEN4155	controlelc	85min_1ITM SMS	cytoc	court	court	15/07/10 9h30 AM	
CAVEN4156	INDIV_188011221	85min_1ITM SMS	390.5371561	court	court	17/07/10 10h20 AM	
CAVEN4157	INDIV_188012320	85min_1ITM SMS	2453.38311	court	court	17/07/10 10h20 AM	
CAVEN4158	INDIV_188111177	85min_1ITM SMS	1850.263871	court	court	17/07/10 10h20 AM	
CAVEN4159	INDIV_188011045	85min_1ITM SMS	363.3118721	court	court	17/07/10 10h20 AM	
CAVEN4160	controlelc	85min_1ITM SMS	cytoc	court	court	17/07/10 10h20 AM	
CAVEN4161	INDIV_188011313	85min_1ITM SMS	2516.290478	court	court	17/07/10 10h20 AM	
CAVEN4162	INDIV_188011344	85min_1ITM SMS	339.5908102	court	court	17/07/10 10h20 AM	
CAVEN4163	INDIV_188111185	85min_1ITM SMS	5161.976712	court	court	17/07/10 10h20 AM	
CAVEN4164	INDIV_188112201	85min_1ITM SMS	998.5131082	court	court	17/07/10 10h20 AM	
CAVEN4165	controlelc	85min_1ITM SMS	cytoc	court	court	17/07/10 10h20 AM	
CAVEN4166	INDIV_188112148	85min_1ITM SMS	3598.860088	court	court	17/07/10 10h20 AM	
CAVEN4167	INDIV_188112184	85min_1ITM SMS	6102.141102	court	court	17/07/10 10h20 AM	
CAVEN4168	INDIV_188011304	85min_1ITM SMS	3622.737892	court	court	17/07/10 10h20 AM	
CAVEN4169	INDIV_188111256	85min_1ITM SMS	6416.797694	court	court	17/07/10 10h20 AM	
CAVEN4170	controlelc	85min_1ITM SMS	cytoc	court	court	17/07/10 10h20 AM	remplissage A et B
CAVEN4171	controlelc	85min_1ITM SMS	cytoc	court	court	17/07/10 10h20 AM	
CAVEN4172	INDIV_188112257	85min_1ITM SMS	3048.75	court	court	18/07/10 11h50 AM	

On peut observer, par exemple, sur la Figure II.8, représentant le suivi des temps de rétention des peptides du cytochrome C que le temps de rétention du peptide 3 a été affecté par le changement des solvants A et B lors de l'analyse numéro 4170, et que ce phénomène observé est revenu à la normale dans l'analyse suivante.

C. Traitement des données

1. Méthode AMT

la génération de la base de données AMT pour « Accurate Mass and time Tags » nécessite des outils spécifiques pour identifier les peptides à partir des spectres MS/MS et pour compiler les informations dans la base de données. Dans un deuxième temps, pour la phase de quantification des peptides d'autres programmes vont permettre d'extraire le signal MS, de confronter les informations extraites des analyses LC-MS (masses expérimentales et temps de rétentions des peptides mesurés) aux informations compilées dans la base de données AMT (masses calculées et temps de rétentions corrigés). Toutes ces étapes sont décrites en détail dans les parties qui suivent.

a) *Traitement des données MS/MS*

Dans cette partie, sont décrites toutes les étapes qui conduisent à l'identification des peptides et des protéines correspondantes. On retrouve également les différents outils qui permettent de filtrer les résultats d'identification et d'optimiser le calcul des temps de rétentions des peptides et enfin d'archiver l'ensemble de ces informations dans une base de données que l'on appelle : « base AMT ».

(1) Mascot Distiller

A l'issue d'une analyse LC-MS/MS, un fichier de données brutes (raw file) est généré. Celui-ci contient un chromatogramme (intensité totale du courant ionique) et des spectres de masses représentant les rapports m/z , des ions détectés et leur abondance relative. Les ions précurseurs sont détectés dans l'Orbitrap, et les ions issus de leur fragmentation dans la trappe ionique. Pour extraire l'information contenue dans ces différents spectres nous avons utilisé le logiciel Mascot Distiller (Matrix Science, London, United Kingdom) qui permet de (1) de déterminer l'état de charge d'un ion, (2) de déconvoluer les spectres MS permettant d'obtenir des ions monochargés monoisotopiques, et (3) de réunir dans un fichier appelé .mgf pour Mascot Generic Format le rapport masse sur charge d'un ion précurseur, sa charge, son temps de rétention chromatographique ainsi que la liste des masses de ses fragments et leur intensité.

(2) Recherche Mascot

Un fois le fichier « mgf » obtenu, il est ensuite transféré au logiciel Mascot pour l'identification de peptides à partir d'une base de données de séquences protéiques. Chaque spectre obtenu est matché à une banque de spectres théoriques, pour établir une correspondance soumise à un score. Le modèle de calcul de score de MASCOT permet de discriminer de manière itérative la meilleure séquence théorique pour chaque spectre de fragmentation en considérant l'intensité des ions fragments expérimentaux attribués, le nombre de fragments attribués, le nombre d'ions consécutifs d'une série de fragmentation, et l'erreur de masse sur la masse des fragments attribués. Pour notre étude, l'ensemble des recherches a été réalisée contre la banque SwissProt TrEMBL (releases 2010_07) contenant 23 254 972 entrées en utilisant la taxonomie "Homo sapiens". Elle est constituée de 2 sous-banques : SwissProt et TrEMBL. SwissProt contient des séquences annotées manuellement avec des informations extraites de la littérature et corrigées à l'aide d'outils d'analyse informatiques. En outre, les isoformes d'une même protéine sont rassemblées sous une séquence référence et sont indexées lorsque cela est possible. La sous-banque TrEMBL (« Translation from EMBL ») contient les séquences protéiques traduites de toutes les séquences codantes des banques nucléotidiques EMBL/GenBank/DDBJ à l'exception de celles déjà présentes dans SwissProt. Les séquences redondantes sont par ailleurs combinées. Des scripts de pré-annotation automatique des entrées permettent d'intégrer des informations primitives qui aideront à l'annotation manuelle avant l'indexation des séquences dans SwissProt. La banque TrEMBL a été créée en complément de SwissProt pour permettre de suivre l'accroissement rapide de la taille des banques nucléotidiques (traduites automatiquement) sans

compromettre la qualité et l'homogénéité de la banque SwissProt à très haute valeur ajoutée, mais plus longue et délicate à enrichir.

Par ailleurs, les recherches Mascot ont été réalisées en incluant des modifications post-traductionnelles fixes (Carbamidométhyl ou Trioxydation des Cystéines) en adéquation avec le protocole de préparation d'échantillon, i.e, DTT et Iodoacétamide ou H₂O₂, et des modifications variables (Acétylation des protéines en N-terminal, Dioxydation et oxydation des méthionines, et oxydation des prolines). Enfin, les critères de tolérance pour les masses des ions précurseurs ou des ions fragments ont été déterminés en fonction des analyseurs utilisés. Ainsi, une tolérance de ± 10 ppm pour les ions précurseurs analysés par Orbitrap, et une tolérance de ± 0.8 Da pour les ions fragments analysés par la trappe ionique ont été sélectionnées. L'ensemble des paramètres de recherches Mascot sont listés dans la Tableau II.3.

Tableau II.3 : Paramètres de recherche Mascot

Parameter	Value
Enzyme	Trypsin/P
Variable modifications	Acetyl(Protein N-term), Dioxydation (M) Oxydation (M) Oxydation (P)
Fixed modifications	Carbamidomethyl (C) pour les échantillons réduits et alkylés (DTT iodoacétamide) Ou Trioxydation (C) pour les échantillons oxydés avec H ₂ O ₂
Mass values	Monoisotopic
Protein Mass	Unrestricted
Peptide Mass tolerance	± 10 ppm
Fragment Mass Tolerance	± 0.8 da
Max Missed Cleavages	2
Instrument type	ESI-TRAP

(3) Bases de données Decoy pour la détermination des Faux Positifs

Le taux de faux positifs FDR est la propriété d'un ensemble de spectres MS/MS définie comme la proportion d'identifications incorrectes parmi toutes les identifications obtenues. Il représente le nombre de faux positifs rapporté à la somme des identifications. Afin d'estimer le FDR, différentes stratégies ont été développées. La stratégie « Target-Decoy » (Elias & Gygi, 2007) repose sur la recherche des résultats d'identifications par un moteur de recherche directement dans une banque protéique classique «Target »

concaténée avec une banque « Decoy » qui contient toutes les séquences inversées de la même banque. L'approche suppose que les identifications de peptides dans la banque « Decoy » et les identifications de faux positifs dans la banque « Target » originale suivent la même distribution. Le taux de faux positifs (FDR) peut être directement estimé en calculant le rapport du nombre de peptides identifiés dans la banque «Decoy » par le nombre total de peptides identifiés selon la formule :

$$FP = \frac{2 \times \text{nombre de peptides Decoy}}{\text{nombre de peptides totaux}}$$

(4) IRMa (interprétation des Résultats Mascot)

Les résultats de recherche obtenus par le logiciel Mascot sont automatiquement validés grâce au logiciel IRMa (Interprétation des Résultats Mascot), développé au laboratoire (Dupierris, Masselon, Court, Kieffer-Jaquinod, & Bruley, 2009). Les résultats générés par Mascot ont ainsi été filtrés selon les paramètres résumés dans le Tableau II.4. A chaque peptide identifié MASCOT attribue un « score d'identification » représentatif de la probabilité que les fragments aient pu être attribués par hasard ; et un « score d'homologie » qui dépend de la taille de l'espace de recherche. Un « score d'identification » supérieur au «score d'homologie» signifie que la probabilité pour que l'identification soit le résultat d'un phénomène aléatoire est inférieure à 5 %. Une fois les peptides filtrés, les protéines couvertes par le même ensemble de peptides sont regroupées dans un *protéine groupe*, et celles qui n'ont plus de peptides au sortir du filtre sont éliminées de la liste.

Tableau II.4 : Filtres utilisés dans le logiciel IRMa

<i>Report settings</i>	<i>Auto p-value < 0.05</i> <i>Peptides cutoff : 0</i> <i>Sub sets threshold : 1^a</i>
<i>Parser properties</i>	<i>Select « Read Only master protein sequences »</i>
<i>Filter results</i>	<i>Homology threshold pvalue < 0.05</i> <i>Rank threshold : 1</i>
	<i>Delete Protein Group (with no relevant peptides)</i> <i>Merge Protein Group (with same set of peptides)</i>

^a ce paramètre permet de considérer l'ensemble des protéines

Après avoir validé l'ensemble des peptides et des protéines d'une analyse LC-MS/MS, les résultats sont compilés dans une base de données relationnelle MSIdb (Mass Spectrometry Identification database).

(5) hEIDI (Exploitation et Intégration des Données)

Cette application développée au sein du laboratoire permet d'explorer les données d'identification stockées dans une base de données. Elle permet également de compiler différentes analyses dans un « contexte » pour extraire une liste de protéines non-redondantes grâce à un algorithme analogue de celui utilisé par Mascot. Cette application permet également de comparer des listes de protéines, d'aligner les temps de rétentions de peptides dans un « contexte » donné ou d'appliquer des filtres supplémentaires (différents des filtres implémentés dans Mascot) au niveau des protéines ou des peptides. Un « contexte » représente un regroupement d'analyses selon différents critères : des analyses LC-MS/MS issues d'une même méthode de fractionnement ou correspondant à différents états d'un compartiment subcellulaire donné, etc...

En pratique, nous avons, dans un premier temps, créé un contexte global regroupant l'ensemble des études relatives à la caractérisation du protéome urinaire. Les paramètres de filtre appliqués étaient :

- une tolérance en masse de 0.001 ppm, pour regrouper toutes les instances du même peptide identifié à multiples reprises dans les analyses LC-MS/MS.
- une longueur minimale de séquence peptidique supérieure ou égale à 6 résidus
- élimination des protéines « reverse » (cf stratégie Target-Decoy)
- élimination des protéines identifiées par un seul peptide de score inférieur à 60.

Pour compiler les résultats de plusieurs analyses dans une base AMT, il est nécessaire d'avoir (1) une mesure de masse fiable et (2) des temps de rétentions reproductibles entre les différentes analyses. Si la première condition est facilement remplie grâce aux spectromètres de masse utilisés à l'heure d'aujourd'hui, la deuxième condition qui concerne la reproductibilité des temps de rétention est plus difficile à obtenir. En effet, de nombreux facteurs plus ou moins contrôlés peuvent être à l'origine de variations de mesures. Il est donc nécessaire d'aligner les acquisitions LC-MS/MS afin d'apparier les signaux qui se correspondent dans les diverses analyses. Pour cela, nous avons implémenté dans hEIDI une application qui permet de réaligner les temps de rétentions chromatographiques de plusieurs analyses MS dans un contexte. Brièvement, cette application s'appuie sur un logiciel développé à PNNL : « Normalized Elution Time (NET) Prediction Utility ». Dans un premier temps, on considère uniquement des peptides ne possédant aucune modification post-traductionnelle. Les temps de rétention de ces peptides sont calculés grâce au logiciel « NET Prediction Utility ». Pour chaque peptide du contexte, on procède suivant les schémas ci-après: (1) on calcule la droite de régression entre les temps de rétentions calculés par rapport aux temps de rétention expérimentaux (si un peptide a été fragmenté plusieurs fois on ne conserve que le temps de rétention de l'occurrence de meilleur score Mascot), (2) on récupère les coefficients de la droite de régression (a - la pente et b - l'ordonnée à l'origine), et (3) on recalcule le temps de rétention de tous les peptides de l'analyse dans le référentiel de

l'algorithme de prédiction à partir des coefficients a et b. Au final, on a des temps de rétention plus homogènes entre les analyses et l'on peut ainsi les utiliser pour l'analyse quantitative.

Pour finir, les résultats ainsi filtrés et réalignés sont compilés dans une nouvelle base de données au format Access nommée « base AMT ».

b) Traitement des données de quantification MS

Cette partie décrit les différents logiciels utilisés pour la seconde phase de la méthode AMT, la quantification des peptides dans les échantillons urinaires. Ces échantillons n'ont pas été fractionnés afin d'augmenter le débit d'analyse. L'objectif est d'analyser un grand nombre d'échantillons pour gagner en puissance statistique, pour la découverte de biomarqueurs. En évitant de fractionner les échantillons, on évite que des peptides se retrouvent dans plusieurs analyses d'un même échantillon, et ce qui risque de fausser l'information quantitative correspondante.

Dans un premier temps, il s'agit d'extraire les signaux MS des données brutes (raw files), puis de retrouver l'identifiant des espèces détectées dans la base de données AMT constituée lors de la première phase de l'étude. Pour ce faire, nous avons utilisé une suite de logiciels développés au laboratoire du Dr Richard D. Smith au Pacific Northwest National Laboratory: les logiciels DeconTools et VIPER successivement.

(1) DeconTools

Ce logiciel, permet à partir de données brutes d'une analyse LC-MS, de déconvoluer les spectres et d'extraire ainsi le rapport masse sur charge, la masse monoisotopique ainsi que l'intensité de toutes les espèces détectées dans chaque scan. Nous avons, pour cela utilisé l'algorithme « Horn mass Transform ». Le programme renvoie deux fichiers dans un format .csv : un fichier iso et un fichier scan. Dans le fichier iso, on retrouve pour chaque scan les informations relatives à la masse, la charge et l'abondance des espèces détectées. Dans le fichier scan, les informations relatives au temps de rétention correspondant à chaque spectre de masse de l'acquisition.

(2) Viper

Une fois les spectres déconvolués, le logiciel VIPER permet, (1) de regrouper les espèces détectées en peptides, (2) aligner les peptides détectés avec les peptides présents dans la base de données AMT correspondante (3) identifier les peptides détectés grâce aux informations présentes dans la base AMT après avoir recalibré les analyses en masse et en temps de rétention. A l'issue de ce traitement, un dossier regroupant différents fichiers relatifs à l'analyse LC-MS est récupéré. Parmi ces fichiers, celui qui nous intéresse préférentiellement est le fichier « ConglomerateUMCsWithNET.txt », dans lequel sont réunies toutes les informations utiles à l'analyse statistique et descriptive qui va suivre. On retrouve notamment les informations relatives à l'identification des peptides : l'identifiant AMT (MassTagID : permettant de

retrouver la séquence en acide aminé du peptide, sa masse et son temps de rétention dans la base de donnée AMT), son intensité (UMCAbundance), et des informations sur la qualité de l'identification tel que le SLIC-Score qui va nous permettre de filtrer les peptides. Pour la suite de l'analyse seront conservés uniquement les peptides avec un SLIC-Score ≥ 0.9 .

2. Qualification des données

Une fois les informations relatives aux peptides (intensité et identité de séquence) extraites à partir de l'analyse des échantillons urinaires, la qualité des données est examinée par une analyse descriptive. Les analyses descriptives ont été réalisées grâce au logiciel JMP commercialisé par SAS.

L'une des premières étapes de cette analyse consiste en la constitution d'un fichier contenant l'ensemble des intensités des peptides dans différents échantillons urinaires à comparer. Ce tableau permet ensuite de visualiser un certain nombre de caractéristiques tel que le nombre de peptides et de protéines quantifiés dans toutes les analyses réalisées. Il permet ainsi de voir rapidement si une analyse n'a pas fonctionné, par exemple quand le nombre de peptides quantifiés dans cette analyse est nettement inférieur à ceux des autres analyses et éventuellement de l'éliminer de l'analyse statistique. Afin d'évaluer la consistance des réplicats analytiques, les coefficients de variation des abondances entre tous les réplicats analytiques ont été calculés. Ainsi, pour chaque peptide mesuré chez chaque patient, les coefficients de variation ont été déterminés en considérant à chaque fois 2 réplicats sur 3: CVR1R2, CVR2R3 et CVR1R3. Le CV minimum a été retenu, si celui-ci était inférieur ou égal à 30%. Nous avons ensuite retenu comme valeur d'abondance du peptide la valeur moyenne des deux réplicats en question. (voir Chapitre V)

3. Analyse statistique « Abundance Index »

Afin de détecter des candidats biomarqueurs du cancer de la vessie, nous avons mis en place une méthode statistique basée sur la méthode décrite par Fu et al. (Fu et al., 2008). Nous avons adapté cette méthode à nos données de quantification peptidiques, puisqu'elle utilisait initialement des données de spectral count sur les protéines. En pratique, pour chaque peptide, l'« Abundance Index » AbI est calculé selon la formule :

$$AbI = \frac{\bar{A}}{\bar{A} + \bar{B}} \times \frac{N_A^D}{N_A^T} - \frac{\bar{B}}{\bar{A} + \bar{B}} \times \frac{N_B^D}{N_B^T}$$

\bar{A} étant la valeur moyenne d'abondance des peptides dans la population contrôle et \bar{B} la valeur moyenne d'abondance des peptides dans la population cancer.

N_A^D et N_B^D représentent respectivement le nombre de fois qu'est détecté le peptide dans la condition contrôle et cancer B.

N_A^T et N_B^T représentent le nombre total de patients dans les conditions contrôle et cancer respectivement.

Dans un deuxième temps, on trace la distribution des AbI, et on extrait la médiane de ces AbI. Les valeurs d'AbI sont ensuite soumises à une permutation des patients à travers les catégories à comparer. Cette permutation nous permet de déterminer les distributions d'AbI correspondant à l'hypothèse H_0 : tous les patients sont équivalents. Cela permet ensuite de calculer les bornes au-delà desquelles les peptides peuvent être considérés comme différentiellement abondants entre les deux populations à comparer. De ces permutations on extrait l'écart type sous l'hypothèse d'équivalence des sous populations. On calcule la valeur des bornes selon la formule : *médiane* $\pm 1.96\sigma$ (intervalle de confiance). On détermine pour chacun des peptides s'il est ou non discriminant à partir des bornes, et dans quelle catégorie se situe le peptide, les possibilités étant récapitulées dans le Tableau II.5.

Tableau II.5: Classification des peptides en fonction de l'analyse statistique par « Abundance Index ». La médiane d'AbI est calculée sur la base des valeurs d'AbI observés, σ sur la base des permutations.

Abundance Index	Classement des peptides	catégorie des peptides
$AbI < \text{médiane} - 1.96\sigma$	<i>Significativement plus abondant chez les cancers</i>	<i>Cancer</i>
$AbI > \text{médiane} + 1.96\sigma$	<i>Significativement plus abondant chez les contrôles</i>	<i>Contrôles</i>
$\text{médiane} - 1.96\sigma < AbI < \text{médiane} + 1.96\sigma$	<i>Non discriminant</i>	<i>non discriminants</i>
N/A	<i>Présents uniquement chez les cancers</i>	<i>cancers only</i>
N/A	<i>Présents uniquement chez les contrôles</i>	<i>contrôles only</i>

Pour les peptides présents uniquement dans une condition (cancer ou contrôle), on ne conserve que les peptides qui sont présents chez au moins 15% de la population, cette valeur ayant été définie par la pratique clinique. On établit ensuite une correspondance entre le tableau comprenant l'identité des peptides et des protéines et le tableau contenant les informations de quantification des peptides ainsi que leurs

catégories résumées dans le tableau II.5. On regroupe les protéines en sommant les peptides en conservant l'information sur les catégories. Ainsi, on obtient des protéines cancers, des protéines contrôles et des protéines non discriminantes (tableau II.6).

Tableau II.6: Classification des protéines selon l'analyse statistique par « Abundance Index »

$A = N(\text{Peptides cancers}) + N(\text{Peptides cancers only}^*)$
$B = N(\text{Peptides contrôles}) + N(\text{Peptides contrôles only}^*)$
$C = N(\text{Peptides non discriminants})$

**Vérifiant le critère de présence chez au moins 15% de la population concernée*

Finally, on considèrera en fonction des statuts A, B ou C (Tableau II.6) qu'une protéine est un candidat biomarqueur si elle vérifie la condition : $A / (A+B+C) \geq 0.5$ et $B=0$. En d'autres termes, la majorité des peptides spécifiques (A+B+C) d'un candidat biomarqueur sont plus abondants chez les patients cancer ou cancer only, et aucun de ses peptides n'est plus abondant chez les patients contrôles ou contrôle only.

L'ensemble des étapes de cette analyse statistique, a été inclus dans un script JMP pour automatiser la procédure.

4. Analyse des données « data mining »

Les « Gene Ontologies » (GO) sont classées selon trois hiérarchies indépendantes pour décrire les processus biologiques, les fonctions moléculaires et les compartiments cellulaires des protéines. Depuis quelques années et grâce aux avancées dans le domaine de la spectrométrie de masse et des techniques de fractionnement, de plus en plus de protéines sont détectées par conséquent, l'analyse de ces listes de protéines est devenue de plus en plus complexe. Il y a quelques années, il était facile de regarder en détail chaque protéine identifiée grâce à des bases de données comme Uniprot ou des logiciels comme quickGO (www.ebi.ac.uk/QuickGO/) utilisés pour l'analyse des catégories fonctionnelles et les localisations subcellulaires des gènes associées aux protéines identifiées. Mais, désormais, il n'est plus possible d'observer individuellement l'ensemble des protéines identifiées par une analyse protéomique. De nouveaux logiciels ont donc été développés pour analyser ces listes de protéines visant à faire ressortir les catégories fonctionnelles, les localisations subcellulaires mais aussi les voies métaboliques dans lesquels étaient impliquées les protéines identifiées. Il existe un large choix d'outils tel que les logiciels d'accès libre comme DAVID - Annotation, Visualization and Integrated Discovery (<http://david.abcc.ncifcrf.gov/>), les logiciels

commerciaux comme Protein Center (Thermo) (représentant un outil global d'analyse informatique des données) et Ingenuity Pathway Analysis, qui sont plus facile à utiliser, mais dont le cout est important. Par ailleurs, il est noté par Louie et al. (Louie, Higdon, & Kolker, 2010) que l'interprétation des résultats obtenus par une analyse d'enrichissement des termes GO sur des listes de protéines est souvent délicate et que les postulats posés au départ pour l'analyse statistique des données sont partiellement biaisés ou inappropriés aux jeux de données issus d'analyse protéomique par spectrométrie de masse. Il faut donc veiller à ne pas sur interpréter ces résultats qui ne donnent au mieux que des pistes qualitatives d'interprétations biologiques de nos listes. Dans notre étude, les analyses des données « Gene Ontology » associés aux protéines identifiées ont été réalisées grâce au logiciel « Protein Center » commercialisé par la société Thermo Fisher.

D. Conclusion

Les protocoles présentés dans ce chapitre, couvrent l'intégralité de notre étude de protéomique quantitative pour la caractérisation du protéome urinaire en vue de la découverte de nouveaux biomarqueurs de pathologie uro-génitales. La Figure II.9 reprend l'ensemble de ces protocoles de la préparation des urines couvrant la collecte des échantillons urinaires, l'extraction des protéines urinaires (voir Chapitre III), la caractérisation du protéome urinaire (voir Chapitre IV), et enfin l'établissement d'une liste de candidats biomarqueurs du cancer de la vessie à travers une étude comparative d'échantillons individuels d'urine de patients atteints du cancer de la vessie et de contrôles sains (voir Chapitre V).

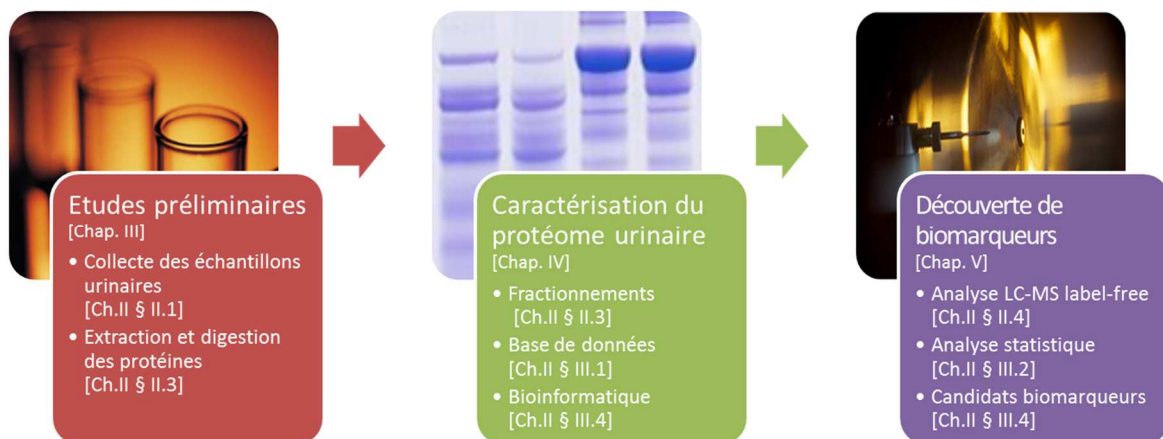


Figure II.9 : Workflow de l'étude de protéomique quantitative en vue de la découverte de biomarqueurs de pathologie uro-génitale

III. Chapitre III : Travaux préliminaires

Après le sérum, l'urine est le fluide biologique le plus étudié pour la découverte de biomarqueurs de pathologies. Les raisons principales sont qu'il est disponible en grande quantité (jusqu'à ~100mL par collecte), la collecte en est très facile, et qu'il est en contact direct avec les protéines potentiellement relarguées par la tumeur dans le cas de pathologies du système urinaire telles que le cancer de la vessie. Cependant, certaines difficultés rencontrées lors de l'analyse du protéome urinaire font que peu de biomarqueurs issus d'études protéomiques de l'urine sont arrivés jusqu'à validation finale. Ces difficultés proviennent entre autre de la faible concentration protéique de l'urine (~ 100 µg/ml), comparée à celle du plasma qui est 500 fois supérieure et de sa forte teneur en sels (~ 10g/l) qui peut gêner l'analyse protéomique, mais aussi de sa grande variabilité inter et intra individus. Visith Thongboonkerd, dans sa revue « Practical points in Urinary proteomics » (Thongboonkerd, 2007), a tenté de répondre à certaines questions pratiques relatives à l'analyse protéomique d'échantillons urinaires, en considérant plus particulièrement les techniques basées sur les gels 2D. Il s'est notamment intéressé aux conditions de collecte des échantillons, à l'utilisation de conservateurs ou d'inhibiteurs de protéases, ou encore à l'élimination des débris cellulaire. S'il est parvenu à définir un certain nombre de recommandation relatives à certaines de ces questions, d'autres sont restées sans réponses ou incomplètes. Ainsi, lorsque nous avons débuté ce travail, nous avons sélectionné certains points qui nous paraissaient important : la pertinence d'ajouter des conservateurs ou inhibiteurs de protéases au moment de la collecte, la normalisation des échantillons, ou encore les méthodes d'extraction et de digestion des protéines de l'urine. Nous avons évalué ces aspects afin de mettre en place une méthode de collecte et de préparation standardisée en vue de la découverte de candidats biomarqueurs dans les urines et leur évaluation par des méthodes de protéomique quantitatives basées sur la spectrométrie de masse.

A. Etude préanalytique : conditions de collecte de l'urine

Un certain nombre de recommandations concernant l'utilisation d'inhibiteurs de protéase, la température de stockage ou la durée écoulée entre la collecte et le stockage des échantillons qui ont définies dans le cadre du projet Plasma Proteome Project lancé par l'organisme international HUPO (HUMAN Proteome Organisation), ont été étendues à d'autres fluides biologiques, et notamment à l'urine. Cependant, certains de ces traitements ne sont pas forcément applicables à ce fluide. Ainsi, l'ajout d'inhibiteur de protéase (IP) dans les urines a été largement discuté notamment dans la revue « Practical Points in Urinary Proteomics » (Thongboonkerd, 2007). En effet, si les IP permettent de limiter la protéolyse des échantillons, ils sont également susceptibles d'interférer avec l'analyse protéomique. De plus étant donné que l'urine est stockée dans la vessie pendant plusieurs heures avant l'urination, les protéases sont censées avoir eu amplement le temps d'agir, et l'ajout d'IP après la collecte, peut être considéré superflu, car trop tardif.

Havanapan (Havanapan & Thongboonkerd, 2009) a traité cette question en centrant son étude sur la nécessité d'ajouter ou non des IP lorsqu'il y a plusieurs cycles de congélation/décongélation des échantillons urinaires. Dans ce cas précis, il est apparu lors de l'analyse différentielle de gel bidimensionnels, que l'ajout d'IP n'était pas indispensable.

Dans la même période, nous avons, nous aussi, étudié cette question par une analyse protéomique utilisant l'approche AMT. L'étude portait sur l'intérêt ou non d'ajouter des IP au moment de la collecte des urines, mais aussi sur la température de conservation des échantillons d'urine avant leur centrifugation à basse vitesse. Par analogie avec le traitement du plasma, de nombreuses études d'analyse protéomique de l'urine conseillent de stocker l'urine à froid ou dans la glace, pour éviter la dégradation des protéines. Toutefois, il est connu que les protéines peuvent précipiter à froid, en particulier en présence de sels. Ce processus est même utilisé pour isoler les protéines d'un échantillon biologique par une approche nommée « salting-out ». Le paramètre de la température nous est donc également apparu important.

1. Utilisation d'additifs et stabilisateurs

Les protéases sont des enzymes qui permettent de réguler, par clivage spécifique ou non-spécifique des protéines ciblées, notamment dans le contexte de fonctions biologiques essentielles telles que le cycle cellulaire ou l'apoptose. Un grand nombre de ces protéases sont présentes dans les fluides biologiques et peuvent y entraîner la dégradation de certaines protéines. Cette dégradation non contrôlée est susceptible de gêner l'analyse protéomique en créant des peptides non-tryptiques qui ne seront pas reconnus par les moteurs de recherche. C'est pour ces raisons qu'il est parfois conseillé d'introduire des IP très tôt dans la préparation des échantillons. Parce que les protéases, en fonction de leurs caractéristiques physico-chimiques, ont des cibles spécifiques, il est recommandé d'utiliser une combinaison ou un cocktail de différents IP.

a) *Interférences potentielles dues aux inhibiteurs de protéases*

Dans le cadre de cette thèse, nous avons testé un cocktail d'IP commercialisé par Roche « Complete Protease Inhibitor Cocktail Tablet, EDTA-free ». Il est difficile d'obtenir la composition exacte de ce produit commercial, si ce n'est qu'il s'agit d'un cocktail d'inhibiteurs de protéases à sérine et à cystéine, mais pas contre les métalloprotéase (absence d'EDTA). Pour déterminer la capacité de cet IP à interférer avec nos mesures, nous avons analysé une tablette de cet inhibiteur dans 40mL d'eau ultrapure. Nous avons suivi pour cela un protocole basé sur l'ultrafiltration utilisé pour la préparation des échantillons d'urine. Après ultrafiltration, l'échantillon a été séparé sur gel SDS-PAGE (migration de type « stacking ») et digéré sur gel. Le mélange résultant a été ensuite analysé par LC-MS/MS. Nous avons effectué une recherche MASCOT dans la base de données UNIPROT en incluant l'ensemble des taxonomies disponibles. Les protéines identifiées ont été classées par ordre décroissant de valeur d'emPAI (exponentially modified Protein Abundance Index). Cet indice permet d'estimer la quantité des protéines au sein d'un échantillon. En

utilisant le nombre de peptides identifiés par rapport au nombre de peptides qu'il est possible de détecter dans cette protéines (Ishihama et al., 2005).

$$PAI = \frac{\text{nobserved}}{\text{nobservable}} \quad \text{emPAI} = 10^{\text{PAI}} - 1$$

Au total, 87 peptides correspondant à 10 protéines ont été identifiés dans cet échantillon (voir Tableau III.1). La grande majorité de ces protéines (9 sur 10) étaient des contaminants (kératines, trypsine) provenant de la manipulation des échantillons. Ce sont en effet des protéines fréquemment identifiées lorsque l'échantillon est très peu concentré. En revanche, nous avons identifié, mais avec seulement 2 peptides, une protéine bovine, l'isoaprotinine G2, dont la fonction (serine-type endopeptidase inhibitor activity) indique sa présence possible dans l'IP. Etant donné le nombre très limité de peptides identifiés dans l'échantillon contenant seulement l'IP, nous avons considéré que cet ajout ne gênerait en rien la détection de protéines minoritaires lors de l'analyse de nos échantillons urinaires.

Tableau III.1 : Protéines identifiées lors de l'analyse de l'inhibiteur de protéases dans de l'eau distillé. (Contaminants grisés)

description	accession	pl	score	mass	coverage	#peptides	emPAI
Isoaprotinin G2 - Bos indicus x Bos taurus (hybrid cattle)	Q7M312_9CETA	9.24	117.28	6527	27.59	2	1.66
Keratin, type II cytoskeletal 1 - Homo sapiens (Human)	K2C1_HUMAN	8.16	1189.19	65978	28.88	16	1.32
Keratin 1 - Pan troglodytes verus	A5A6M6_PANTR	7.61	1156.71	65450	24.96	15	1.22
Hair acidic keratin 1 - Capra hircus (Goat)	Q6R651_CAPHI	4.79	485.90	46847	20.10	8	0.80
Keratin, type I cytoskeletal 10 - Homo sapiens (Human)	K1C10_HUMAN	5.13	827.18	59475	21.42	10	0.79
Trypsin precursor - Sus scrofa (Pig)	TRYP_PIG	7.00	247.95	24394	21.65	3	0.75
Keratin, type II cuticular Hb3 - Bos taurus (Bovine)	A4FUZ0_BOVIN	5.43	498.99	53955	12.98	8	0.67
Similar to keratin, hair, basic, 1 - Bos taurus (Bovine)	Q148H4_BOVIN	5.62	502.80	54578	12.40	8	0.66
Hair type II keratin intermediate filament protein - Ovis aries (Sheep)	Q28582_SHEEP	6.08	490.86	55210	12.80	8	0.65
Keratin 2A - Pan troglodytes verus	A5A6P6_PANTR	7.61	635.70	65944	15.50	9	0.61

b) Antibactériens

Des antibactériens, tels que l'Azide de Sodium ou l'acide borique, sont parfois ajoutés dans les urines pour éviter une éventuelle contamination bactérienne. Pour la protéomique, on conseille notamment l'ajout de tels conservateurs lorsque les urines sont conservées plus de 8 heures à température ambiante ou plus de 16 heures à 4°C (Thongboonkerd, 2007). Dans notre cas, ayant déterminé au préalable avec les cliniciens que nos échantillons ne seraient pas stockés plus de 4 heures avant leur congélation à -80°C, nous avons décidé de ne pas ajouter d'antibactériens. Pour vérifier cependant l'absence de toute contamination bactérienne des urines collectées suivant ce principe, nous avons tout de même réalisé une recherche dans les banques de données avec des données issues d'une de nos séries d'extraits urinaires, en spécifiant comme taxonomies « Ecoli » et « Eubacteria ». Avec les paramètres de filtres que nous utilisons en routine, aucune protéine n'a été identifiée dans ces deux taxonomies. Nous avons donc considéré que la contamination bactérienne était négligeable dans nos conditions de collecte.

2. Optimisation des conditions de collecte

Lors d'une précédente étude portant sur un faible nombre de cas, nous avons observé que l'ajout d'IP dans l'urine au moment de la collecte pouvait induire une légère augmentation du nombre de protéines identifiées dans l'échantillon. Nous avons donc décidé de lancer une étude plus complète sur l'intérêt d'utiliser des IP, en intégrant cette fois un plus grand nombre d'échantillons. De plus, nous avons inclus un autre facteur important qui est la température de conservation des urines avant leur centrifugation à basse vitesse. Lors de l'élaboration de l'étude, nous avons aussi considéré les contraintes du milieu hospitalier. En effet, la personne en charge de la collecte dans le service d'urologie n'a pas toujours la possibilité d'effectuer la centrifugation de l'échantillon dans les minutes qui suivent la collecte. Nous avons alloué un temps maximum de 4 heures d'attente pour la préparation des urines. En ce qui concerne les IP, nous sommes convenus avec les cliniciens qu'une tablette d'IP serait introduite au préalable dans le tube utilisé pour la collecte.

a) Design expérimental

Nous avons réalisé une analyse comparative à partir d'échantillons urinaires de 12 donneurs sains (6 hommes et 6 femmes) poolées par paires. La figure III.3 illustre le design de cette étude. Différents aliquots ont été traités ou non avec des IP, et stockés sur la glace fondantes ou à température ambiante pendant 4 heures avant centrifugation à basse vitesse. Pour chaque échantillon urinaire, nous avons réalisé un contrôle, qui a été centrifugé immédiatement après la collecte. Les protéines urinaires ont été extraites par précipitation au TCA immédiatement après la centrifugation à basse vitesse, puis digérées en solution. Les échantillons peptidiques ont ensuite été dessalés sur colonnes C18, avant d'être analysés en mode LC-MS sur l'Orbitrap XL. Nous avons réalisé une analyse quantitative « Label-free » en utilisant une approche AMT pour évaluer l'influence du traitement préanalytique sur le résultat de l'analyse protéomique semi-quantitative.

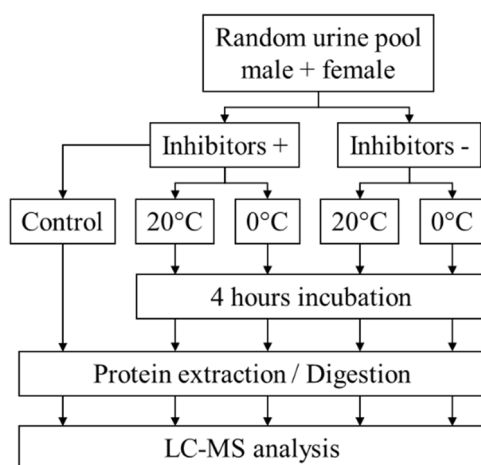


Figure III.3 : Design de l'étude de l'influence des conditions de collecte sur le résultat d'analyse LC-MS d'échantillons urinaires.

b) Métriques pour l'évaluation

Si les traitements à évaluer ont un effet sur l'intensité de certains peptides, la comparaison de ces intensités avec celles des mêmes peptides dans l'échantillon contrôle devrait révéler des non concordances. Cela doit donc se traduire par des points qui s'éloignent de la droite de régression entre les abondances du traitement et celles du contrôle. Ceci peut être évalué sur la base du coefficient de corrélation de Pearson. Nous avons donc évalué la corrélation de Pearson entre les abondances des peptides dans les échantillons traités avec ou sans IP, à température ambiante ou à 0°C, et les contrôles correspondants. Cette corrélation devrait être proche de 1 lorsque l'échantillon traité est semblable au contrôle, et s'écarter de cette valeur lorsqu'ils présentent des dissimilarités. Cette mesure est également insensible aux légères différences de mesures provenant par exemple de la quantité injectée car toutes les abondances des peptides varient alors d'un même facteur.

c) Résultats

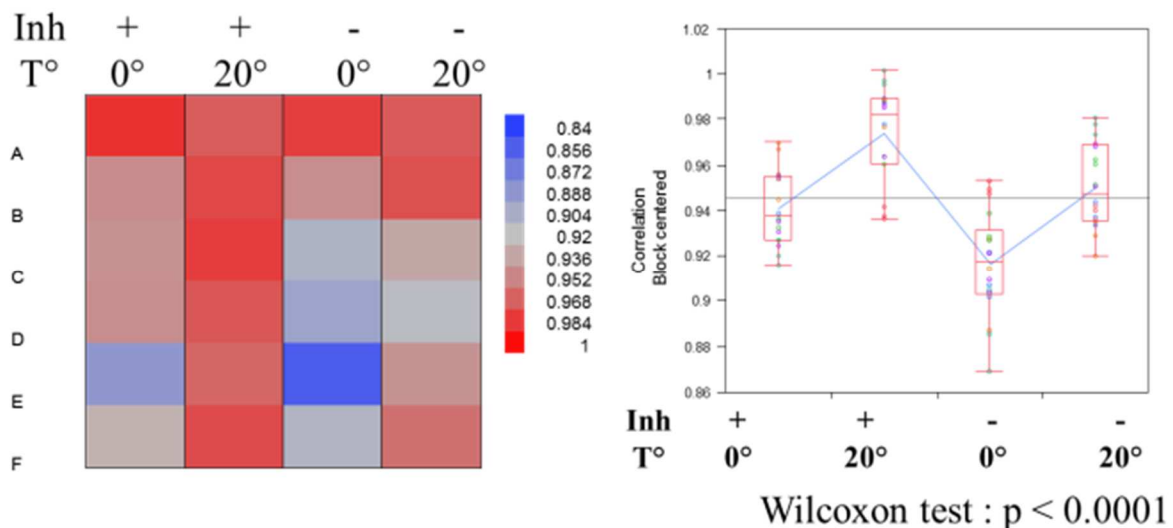


Figure III.4 : Carte en couleur et box plot (centré par échantillon) représentant le coefficient de Pearson de la corrélation des abondances peptidiques avec celles du contrôle pour 6 échantillons distincts traités ou non avec IP et conservés 4 heures à température ambiante ou à froid.

La figure III.4 (panneau de gauche) représente la carte en couleur des coefficients de Pearson de la comparaison au contrôle correspondant pour chacun des traitements appliqués aux six échantillons (A à F). Sur le panneau de droite de cette même Figure III.4, on a représenté les « box plot » correspondants, centrés par échantillon afin d'éliminer les effets dus aux échantillons. Les corrélations observées varient de 0,86 à 0,99, ce qui indique dans l'ensemble une faible dispersion des signaux d'abondances quel que soit le traitement. Néanmoins, le test de Wilcoxon (p -value < 0.0001) montre qu'il y a un effet significatif du traitement sur la corrélation avec le contrôle. De plus, le traitement consistant en l'addition d'IP et la conservation à température ambiante montre quasi-systématiquement le meilleur coefficient de Pearson. La

condition la plus défavorable correspond, quant à elle, au traitement sans inhibiteur de protéases dans les urines et une conservation à 4°C.

Il est à noter que cette étude a été répliquée en utilisant une autre méthode de préparation des urines, l'ultrafiltration suivie d'une digestion in-gel. Les mêmes tendances ont été observées, ce qui nous a permis d'établir que les conditions optimales étaient indépendantes de la méthode de préparation.

Contrairement aux résultats de Havanapan (Havanapan & Thongboonkerd, 2009), nous avons constaté que le traitement préanalytique pouvait introduire une variabilité en termes d'abondances peptidiques. Cette variabilité ne nous semble pas suffisante pour générer des protéines significativement différentielles, mais elle pourrait masquer de réelles variations dans les analyses semi-quantitatives. Il est donc important de minimiser cette variance en employant le protocole le plus robuste. Ces résultats ont permis d'émettre de nouvelles recommandations sur la collecte de l'urine en vue d'une étude protéomique quantitative qui ont été publiées sur le site de l'INCa : « http://www.e-cancer.fr/component/docman/doc_download/4826-prelevement-et-conservation-de-lurine ».

B. Standardisation

Il existe dans la littérature, une grande variété de méthodes pour préparer les urines dans le but d'en caractériser le protéome. Ainsi, il est assez rare de trouver deux laboratoires utilisant la même technique de préparation d'échantillons urinaires. Dans les études cliniques, et particulièrement dans un projet allant de la découverte jusqu'à l'évaluation de biomarqueurs dans un fluide biologique, il est important de minimiser les sources de variabilités extrinsèques à l'échantillon, et notamment celles due à la préparation. Il est ainsi souhaitable d'avoir un protocole unique et standardisé pour la préparation des échantillons à toutes les étapes du développement des biomarqueurs. En particulier, des facteurs tels que la reproductibilité, la précision et l'exactitude des mesures obtenues doivent pouvoir être évaluées objectivement (Rifai et al., 2006). Dans ce contexte, la standardisation de la préparation de l'échantillon nous est parue essentielle. Nous avons donc travaillé avec le partenaire du projet en charge de la vérification des biomarqueurs, l'ETH Zurich, pour définir ensemble un protocole standardisé de préparation des échantillons urinaires en vue de leur analyse protéomique.

1. Design de l'étude

Pour la préparation de l'urine en vue d'une analyse protéomique, on distingue deux étapes : la première consiste à extraire les protéines urinaires, et la seconde à digérer enzymatiquement les protéines extraites. Initialement, deux protocoles distincts avaient été optimisées dans chacun des deux laboratoires : l'un combinant une précipitation acide et une digestion en solution, l'autre comprenant une ultrafiltration suivie d'une digestion sur gel. Nous avons combiné les deux méthodes d'extraction des protéines et les deux méthodes de digestion pour établir quatre protocoles distincts que nous avons comparés entre eux. Des

critères d'évaluation ont ensuite été établis et une étude protéomique a été menée dans les deux laboratoires en utilisant deux stratégies d'analyses quantitatives distinctes : une approche AMT, qui sera utilisée par la suite pour la découverte de biomarqueurs et une analyse SRM qui servira à l'évaluation de candidats biomarqueurs dans le projet DECanBio.

Les quatre protocoles ainsi mis au point ont été testés sur trois pools d'urines détaillés dans le Tableau III.2. Ces pools ont été préparés à partir d'urines de patients sains collectées en clinique selon le protocole optimisé précédemment (voir Partie A). Les quantités de protéines de chaque pool ont été ajustées de façon à ce que chaque patient soit représenté équitablement et que les concentrations protéiques totales soient semblables.

Tableau III.2 : Tableau récapitulatif de la composition des 3 pools utilisés pour tester les protocoles de préparation des urines en vue de l'analyse protéomique.

Pool	Patients	Concentration protéines µg/mL	Volume d'urine mL	Concentration protéique µg/mL	Volume total mL
Pool1	188111001	29,29	22,73	41,29	64,47
Pool1	188111014	46,27	14,38		
Pool1	188111010	46,65	14,27		
Pool1	188111056	50,84	13,09		
Pool2	188111021	30,81	21,60	44,25	60,17
Pool2	188111038	40,82	16,30		
Pool2	188111072	46,13	14,43		
Pool2	188111054	84,94	7,84		
Pool3	188111012	26,62	25,00	39,38	67,61
Pool3	188111029	31,95	20,83		
Pool3	188111057	52,23	12,74		
Pool3	188111008	73,66	9,04		

Afin d'évaluer les mérites de chaque protocole, nous avons réalisé une étude utilisant deux stratégies de quantification basées sur la spectrométrie de masse, AMT et SRM en introduisant trois points de contrôle à différentes étapes du traitement de l'échantillon (voir Figure III.5).

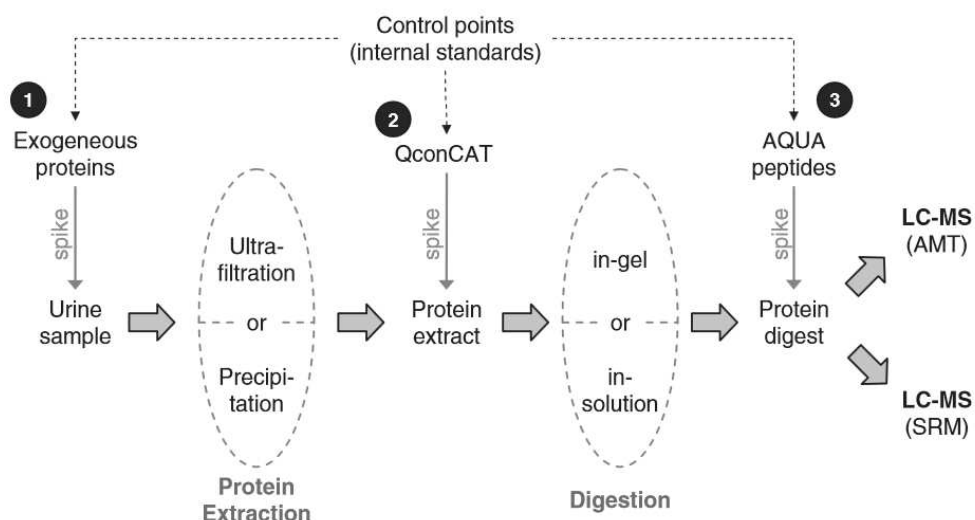


Figure III.5 : Design de l'étude de comparaison des protocoles de préparation des échantillons urinaires. Trois points de contrôles sous la forme de standards internes ont été établis aux étapes clé du traitement des urines afin de pouvoir estimer les performances des quatre protocoles.

Les trois points de contrôles étaient les suivants : (i) Un mélange contenant 3 protéines de levure (tableau III.3) de différents poids moléculaires en quantités équimolaires a été ajouté dans les pools d'urine de départ afin d'estimer les rendement d'extraction des différents protocoles ; (ii) un concatémère (QconCAT) contenant 42 peptides marqués avec des isotopes correspondant à 18 protéines urinaires compilées dans le tableau III.4 et couvrant une large gamme de concentrations dans l'urine humaine a été ajouté dans chaque extrait protéique pour évaluer les protocoles analytiques ; (iii) des peptides synthétiques marqués avec des isotopes (AQUA peptides), correspondant aux protéines de levures (premier point de contrôle), ont été ajoutés avant l'analyse en LC-MS pour estimer précisément le rendement global de chaque protocole.

Tableau III.3. Liste des 3 protéines de levures ajoutées aux échantillons urinaires pour l'estimation des rendements d'extraction.

Accession	Description	Nombre d'AA	Poids moléculaire
P00330	Alcool déshydrogénase 1	348	36 849 Da
P00729	Carboxypeptidase Y	532	59 802 Da
P00924	Enolase 1	437	46 816 Da

Tableau III.4 : Protéines urinaires représentées dans le standard QconCAT et liste des peptides sélectionnés.

Accession	Protéines	Peptides
CUBN_HUMAN	Cubilin	SDNSPTHVGFK AGFDASFHK
LG3BP_HUMAN	Galectin-3-binding protein	LADGGATNQGR SDLAVPSELALLK ALGFEDATQALGR GLDLTEDTYKPR AAIPSALDTDSSK
UROM_HUMAN	Uromodulin	DWVSVVTPAR VGGTGMFTVR QDFDITDISLLEHR DETHATYSNTLYLADEIIIR
CO6A1_HUMAN	Collagen alpha-1(VI) chain	DFTAADWQSR
EGF_HUMAN	Pro-epidermal growth factor	LFWIQYNR GDNSHILLSALK
CD44_HUMAN	CD44 antigen	FAGVFHVEK YGFIEGHVVIPR AFDSTLPTMAQMEK DPDHSEGSTTLLEGYTSHPHTK
KNG1_HUMAN	Kininogen-1	TVGSDTFYSFK LNAENDATFYFK
ANAG_HUMAN	Alpha-N-acetylglucosaminidase	LLLTSAPSLATSPAFR VFPQVDVTK
TRFE_HUMAN	Serotransferrin	DGAGDVAFVK QQQHFLFGSDVTDCSGNFCLFR
CAD11_HUMAN	Cadherin-11	DADDPTYGNSAK EETAWLIDITVFAAEIHNR
ATRN_HUMAN	Attractin	LTLTPWVGLR IDSTGDVTNELR
PGS1_HUMAN	Biglycan	IQAIELEDLLR LLQVVYLHNSNDITK
CALR_HUMAN	Calreticulin	VHVIFNYK SGTIFDNFLITNDEAYAEFGDETWGVT
A1BG_HUMAN	Alpha-1B-glycoprotein	HQFLLTGDTQGR PLADVTLTCQAR
IL33_HUMAN	Interleukin-33	TDPGVFIGVK
BIRC5_HUMAN	Apoptosis inhibitor survivin	QFEELTLGEFLK
K1C19_HUMAN	Cytokeratin-19	AALEDTLAETEAR SLLEGQEDHYNDLSASK LEQEIATYR
CADH1_HUMAN	Epithelial cadherin	NTGVISVVTGLDR PPPHQASVSGIQAEELLTFPDSSPGLR

Les mélanges peptidiques des échantillons préparés selon les quatre protocoles, ont été analysés sur les deux plateformes LC-MS (AMT et SRM) situées dans deux laboratoires différents (Grenoble pour l'AMT et Zürich pour la SRM).

2. Observations initiales

L'un des premiers constats sur les différents protocoles testés est que la méthode de préparation de l'échantillon a fortement influencé le résultat de l'analyse protéomique, à la fois qualitativement (différentes protéines ont été détectées) et quantitativement. Juste après l'extraction, chaque échantillon a été repris dans

du tampon Laemmli et l'extrait protéique a été déposé sur un gel 1D pour une première évaluation visuelle. A partir de ces gels, il apparaît sur la Figure III.6.A que les deux méthodes d'extraction ont abouti à un enrichissement différentiel du protéome urinaire.

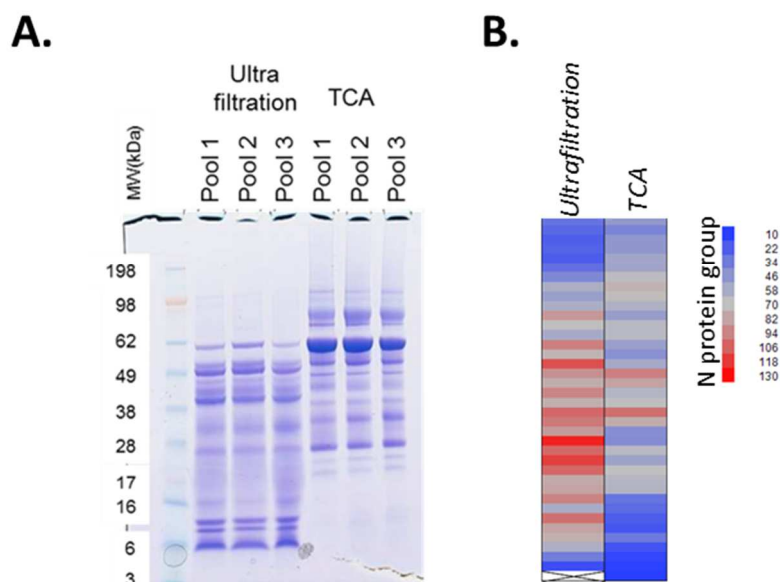


Figure III.6. A : Profil protéique sur Gel SDS-PAGE des trois pools d'urine préparés par ultrafiltration ou précipitation TCA. et B : Carte de couleur représentant le nombre de protéines group identifiées par bande de gel sur une piste « ultrafiltration » et piste « précipitation TCA ».

On observe sur la figure III.6.B le nombre de protein group qui ont été identifiées dans chacune des bandes correspondant à une des pistes de l'ultrafiltration et une piste de précipitation TCA. De la même manière que sur la figure A, on observe peu de protéines identifiées dans les hauts poids moléculaire pour l'ultrafiltration et peu de protéines identifiées dans les bas poids moléculaire pour la précipitation TCA. De plus, on compte globalement plus de protéines identifiées sur la piste correspondant à l'ultrafiltration. Mais ces deux figures ne représentent qu'une première étape de la comparaison entre les deux méthodes de préparation et méritent d'être complétées par une analyse quantitative plus approfondie.

3. Etude comparative - AMT

Afin de comparer les quatre protocoles, nous avons dans un premier temps constitué une base de données AMT à partir de 136 analyses LC-MS/MS issues des échantillons préparés selon ces quatre protocoles. Cette base contenait 3102 peptides couvrant 529 protéines non redondantes, avec une couverture moyenne de 5 peptides par protéine. La redondance des protéines dans la liste finale a été minimisée en utilisant un algorithme de regroupement des protéines correspondant aux mêmes ensembles de peptides.

Chacun des trois pools d'urines, traité selon les quatre protocoles de préparation a été analysé en mode LC-MS. Toutes les analyses ont été réalisées en triplicat par blocs. Plusieurs mesures ont été extraites afin d'évaluer les effets de chaque traitement sur la réponse analytique : (1) le coefficient de corrélation de Pearson basé sur l'abondance des peptides mesurés dans les différents protocoles testés (2) le nombre de peptides communs entre deux analyses LC-MS et (3) le nombre total de peptides identifiés et quantifiés dans chacune des analyses LC-MS. Comme indiqué précédemment, le coefficient de Pearson nous permet d'apprécier la similitude entre les différentes analyses. De plus, nous avons représenté sur une carte en couleur le nombre de peptides commun à chaque paire d'analyse et qui participaient au calcul du coefficient de Pearson. De la même manière, le nombre total de peptides mesurés dans chaque analyse permet d'évaluer le potentiel de chaque protocole d'accéder au protéome le plus exhaustif possible.

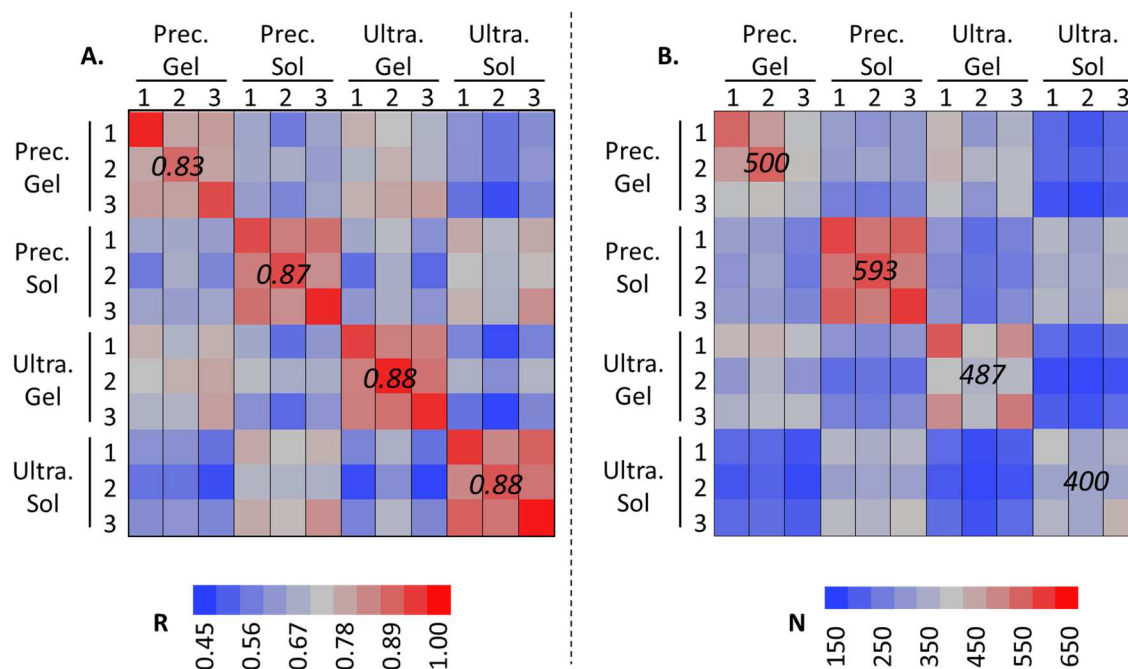


Figure III.7 : Carte en couleur représentant les coefficients de Pearson moyennés sur les réplicats analytiques de la corrélation des abondances peptidiques pour chaque pool d'urine préparé selon les 4 protocoles (A) et le nombre de peptides communs à chaque paire d'analyse (B).

La Figure III.7 représente la carte en couleurs des coefficients de Pearson et le nombre de peptides communs (moyennés sur les trois réplicat analytiques) pour les trois pools d'urines préparés selon les quatre méthodes de préparation (précipitation TCA - digestion sur gel, précipitation TCA - digestion en solution, ultrafiltration - digestion sur gel et ultrafiltration - digestion en solution) analysés en mode LC-MS. Les corrélations observées varient à l'intérieur de chaque protocole de 0.83 à 0.88 indiquant une faible dispersion des abondances peptidiques quel que soit le protocole utilisé. Par contre lorsque l'on observe les corrélations entre différentes méthodes de préparation, les corrélations varient de 0.45 à 0.99 avec une moyenne de 0.58. Ceci met en évidence que la dispersion des abondances peptidiques s'aggrave de façon significative lorsque

les échantillons sont préparés selon différents protocoles. Néanmoins le fait de digérer les protéines selon la même méthode de digestion, que ce soit en solution ou sur gel, semble améliorer cette corrélation en obtenant une moyenne de 0.74. Ainsi, la digestion des protéines représente une étape beaucoup plus critique que la méthode d'extraction (précipitation TCA ou ultrafiltration).

Nous avons, dans un deuxième temps, observé le nombre de peptides communs à deux analyses en considérant l'ensemble des pools préparés selon les quatre protocoles (Figure III.7.B). Sur cette figure, on observe de grandes variations du nombre de peptides communs à deux analyses préparés selon le même protocole. Ainsi, lorsque l'on compare des échantillons dont les protéines ont été extraites par une précipitation en milieu acide et digérées en solution, on obtient en moyenne 593 peptides communs entre deux analyses LC-MS. Lorsque l'on fait la même comparaison cette fois-ci sur des échantillons dont les protéines ont été extraites par ultrafiltration et digérées en solution on obtient en moyenne 400 peptides en communs. Ainsi, le choix du protocole pour extraire ou digérer les protéines semble affecter de manière significative le nombre de peptides communs entre deux analyses LC-MS de deux réplicats biologiques.

Dans le but de minimiser les variabilités dues à la préparation des échantillons urinaires, il est préférable, aux vu de ces résultats, d'extraire les protéines par une précipitation en milieu acide (TCA) et de digérer les protéines en solution. Cette méthode de préparation apporte en effet la meilleure reproductibilité en termes de nombre de peptides communs entre deux analyses (593) et de corrélation des abondances peptidiques (coefficient de Pearson de 0.87 : élevé). Il est à noter qu'il ne s'agit pas de comparaison entre réplicats, mais bien d'analyse de trois pools d'urines différentes.

4. Etude comparative - SRM

a) Rendement des méthodes de préparations

Le rendement d'extraction des protéines est mesuré en comparant la quantité initiale de protéines urinaires à celle que l'on estime après le traitement de l'échantillon. Pour évaluer l'effet de la préparation des échantillons sur ce rendement, nous avons ajouté dans les urines des protéines de levures en quantités connues et l'on a évalué, par une méthode quantitative ciblée (SRM), la quantité de protéines récupérée après le traitement global des urines comprenant l'étape d'extraction et de digestion des protéines. Le mélange de protéines était composé de trois protéines de levure de masse moléculaire différentes (36, 46 et 59 kDa), représentatives des masses des protéines présentes dans les urines (<60 kDa en raison du filtre du rein). Les quantités de protéines ont été déterminées par SRM en utilisant comme standards de quantification des peptides AQUA qui avaient été rajoutés dans les échantillons avant l'analyse LC-MS, à des concentrations de 100 fmol/ μ l. Le rendement d'extraction des protéines de levure a été calculé pour tous les peptides issus de ces protéines. Les résultats ont été moyennés sur les trois pools d'urine (Figure III.1).

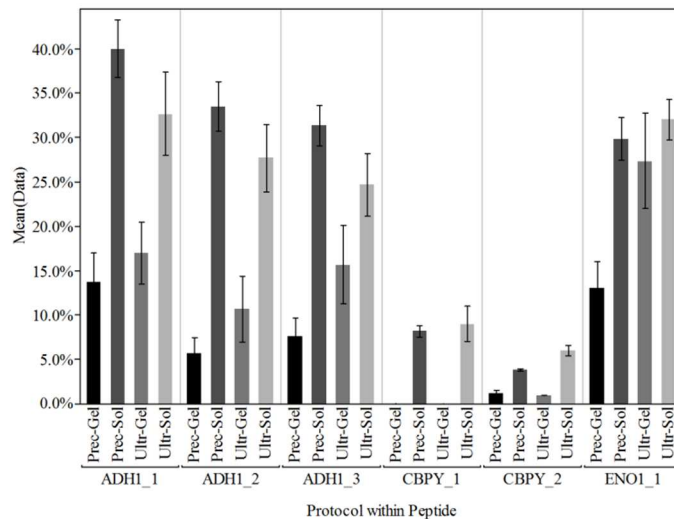


Figure III.1 : Moyenne des rendements obtenus sur les protéines de levure ajoutées dans l'urine et estimés par SRM. Trois peptides AQUA pour l'alcool déshydrogénase, deux pour la carboxypeptidase et un pour l'énolase 1 ont été utilisés pour la quantification.

Ces résultats montrent que la digestion en solution permet d'obtenir de meilleurs rendements de récupération des protéines : les valeurs sont comprises entre 25 et 40% pour les peptides de l'alcool déshydrogénase et l'énolase 1. En revanche, les rendements obtenus pour les protocoles avec une digestion sur gel se situent entre 5 et 25% pour ces mêmes protéines. Concernant la carboxypeptidase, un rendement beaucoup plus faible a été systématiquement obtenu. Les barres d'erreur montrent que ces rendements sont reproductibles quel que soit le pool analysé.

b) Gamme dynamique

Afin d'évaluer les performances des quatre protocoles en terme de gamme dynamique nous avons utilisé une protéine chimérique « QconCAT » ajoutée dans les urines juste avant l'étape de digestion, ainsi que les 3 protéines de levure ajoutées dans les urines. Les protéines ciblées avec le QconCAT correspondent à 18 protéines (21 peptides marqués isotopiquement) couvrant une large gamme de concentrations des protéines urinaires dans les échantillons humains. Les concentrations relatives de ces 18 protéines humaines et 3 protéines de levure dans les trois pools d'urines ont été déterminées par SRM dans les quatre protocoles et sont représentés dans la figure III.2.

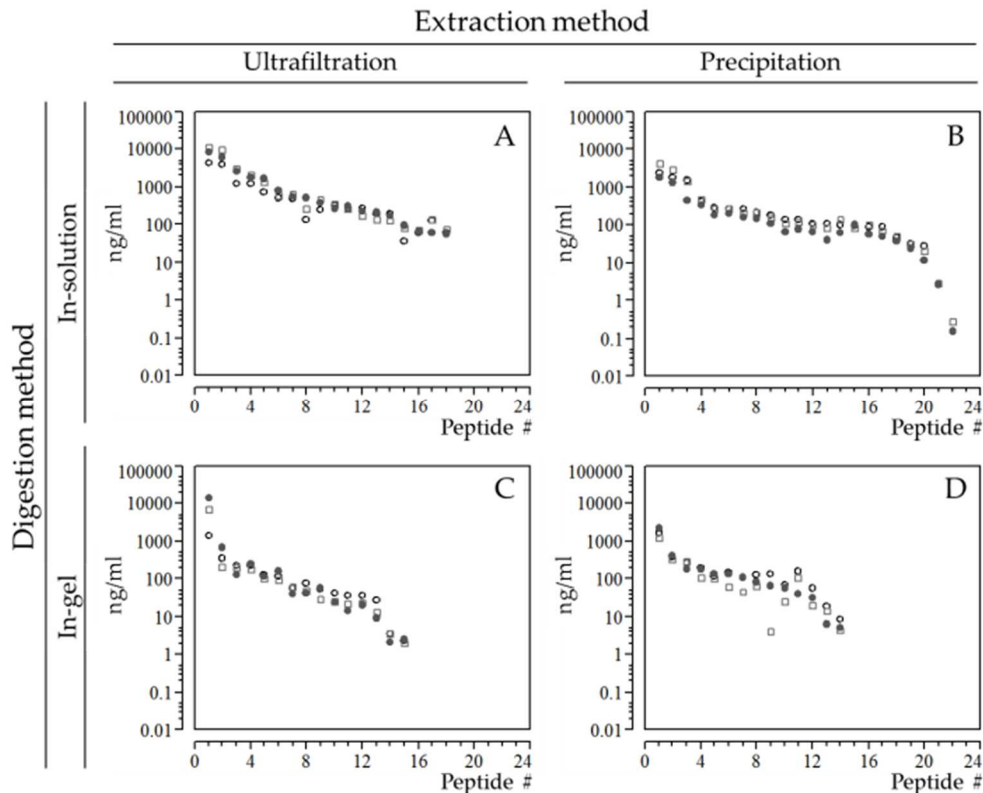


Figure III.2 : Concentration des 21 peptides issus des 18 protéines humaines (QconCAT) et des 5 peptides issus des 3 protéines de levures (AQUA), dans les trois pools d'urines préparés selon les quatre protocoles. Les peptides quantifiés ont été classés dans l'ordre décroissant de leurs concentrations estimées par SRM dans l'urine.

La meilleure gamme dynamique de concentration des protéines a été obtenue avec la méthode de précipitation TCA suivie d'une digestion en solution. Cette gamme dynamique comprend environ 4 ordres de grandeur entre la protéine la moins concentrée et la plus concentrée (Figure III.2). Seize peptides correspondant à 13 protéines urinaires et 6 peptides des protéines de levure ont été quantifiés dans les trois pools d'urines par cette méthode de préparation. La concentration de la protéine la plus abondante (l'uromoduline) a été estimée dans les trois pools d'urines entre 1,5 et 3,3 mg/mL, et la concentration de la protéine la moins abondante (la survivine) a été estimée entre 0,25 et 0,15 ng/mL. En revanche, seulement 9 protéines urinaires et 6 peptides provenant des protéines de levure ont pu être quantifiés dans les autres protocoles de préparation, faisant ainsi diminuer la gamme dynamique à trois ordres de grandeur.

En conclusion, sur les quatre protocoles qui ont été testés, les résultats et les métriques qui ont été utilisés, ont permis systématiquement de révéler la précipitation TCA suivie d'une digestion en solution comme étant la meilleure méthode de préparation des urines.

5. Comparaisons AMT-SRM

Afin d'établir une validation croisée entre les deux plateformes d'analyse (AMT et SRM), nous avons réalisé une dernière expérience dans laquelle les trois pools d'urines ont été précipités en milieu acide (TCA) puis digérés en solution. Trois opérateurs (A, B et C) se sont partagés les pools qui ont été préparés en duplicats. Des standards QconCAT ont été ajoutés dans les échantillons juste avant l'étape de digestion pour estimer les concentrations des 18 protéines urinaires du tableau III.4. Les analyses ont été réalisées dans les deux laboratoires en utilisant les deux méthodes de quantification : l'approche AMT et l'approche SRM. L'extraction des signaux MS (AMT) et des transitions (SRM) ont permis de quantifier les 18 protéines urinaires en utilisant le rapport entre l'abondance des peptides endogènes et les peptides issus des QconCAT marqués isotopiquement. La figure III.8 montre le graphe des estimations de log de concentrations obtenues par les deux approches. Ce graphe se caractérise par un coefficient de corrélation élevé ($R^2 = 0.9945$). Ce résultat démontre la cohérence des résultats de quantification obtenus par les plateformes AMT et SRM. Ceci valide la qualité des mesures d'abondances obtenues par la méthodologie AMT que nous avons utilisée pour l'étape de découverte de candidats (voir Chapitre V).

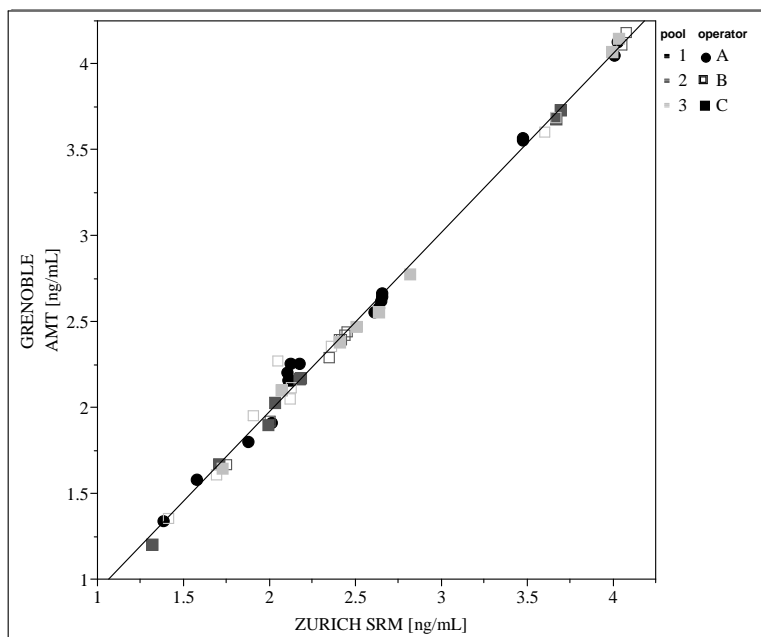


Figure III.8 : Log de concentration des peptides issus des protéines urinaires estimées à partir des QconCAT en utilisant les plateformes AMT et SRM dans les trois pools d'urines préparés en duplicat selon le protocole standardisé (précipitation TCA – digestion en solution).

C. Conclusion

Au cours des travaux préliminaires décrits dans ce chapitre, nous avons mis en place, en collaboration avec les partenaires cliniques du projet, des conditions de collectes standardisées en accord avec les exigences de l'analyse protéomique et les contraintes de la clinique. Ainsi après la collecte, les

échantillons d'urines dans lesquels ont été rajoutés des inhibiteurs de protéases, sont conservés un maximum de quatre heures à température ambiante. Par ailleurs, nous avons établis, en collaboration avec nos partenaires en charge de l'évaluation des biomarqueurs, un protocole standardisé d'extraction des protéines urinaires basés sur la précipitation TCA suivie d'une digestion en solution. La combinaison de ces deux méthodes (précipitation TCA et digestion en solution) a fourni des données quantitatives de meilleure qualité que trois autres protocoles testés. Ainsi, cette méthode a été employée pour l'étude quantitative en vue de déterminer les candidats biomarqueurs du cancer de la vessie dans les urines (Chapitre V).

IV. Chapitre IV : Caractérisation du protéome urinaire

L'une des premières étapes essentielle dans le déroulement d'une étude de protéomique clinique repose sur une caractérisation approfondie du fluide biologique en lien avec la pathologie ou fluide proximal (tel que l'urine pour le cancer de la vessie). Ces dernières années, plusieurs études (Adachi et al., 2006), (Kentsis et al., 2009), ont été menées pour tenter de caractériser le protéome urinaire. Par ailleurs, de nouvelles technologies basées sur le fractionnement ou la déplétion des protéines ont été développées pour atteindre « le protéome profond » de fluides biologiques comme le sérum ou le liquide céphalo rachidien (Mouton-Barbosa et al., 2010). Ainsi dans notre étude, afin de caractériser l'urine en profondeur en vue d'établir notre base de donnée AMT qui nous servira par la suite à rechercher des nouveaux biomarqueurs du cancer de la vessie, nous avons mis en œuvre diverse méthodes de déplétion ou de fractionnement. Nous avons été en mesure d'évaluer à posteriori les mérites respectifs de ces différentes méthodes en comparant les listes de protéines ainsi identifiées. Nous avons également réalisé une analyse bioinformatique de ces listes pour vérifier l'existence de biais analytiques associés aux différents fractionnements, en déterminant s'il y avait un enrichissement de catégories fonctionnelles ou de localisations subcellulaires particulières. Par ailleurs, nous avons comparé notre base de données avec les études montrant la meilleure couverture du protéome urinaire publiées dans la littérature scientifique. Nous avons finalement évalué le potentiel de notre base de données pour la découverte de biomarqueurs en confirmant la présence de certains biomarqueurs possibles du cancer de la vessie ou d'autres pathologies urinaire telle que l'insuffisance rénale aigue, précédemment rapportés dans la littérature.

A. Contexte général – Biomarqueurs de pathologies

1. Recherche de biomarqueurs

Les protéines biomarqueurs sont par définition celles dont l'abondance diffère en fonction d'un état pathologique. Depuis maintenant plusieurs années, la découverte de biomarqueurs par des approches protéomiques suscite beaucoup d'intérêt et d'espoir chez les cliniciens car elle ouvre des perspectives de développement de nouveaux tests cliniques plus sensibles et plus spécifiques. Grâce aux avancées des technologies d'analyse protéomiques, il est désormais possible de détecter des milliers de protéines et de les quantifier dans des tissus ou des fluides biologiques. Même si le tissu tumoral reste incontestablement la ressource la plus appropriée pour repérer des signaux spécifiques de la tumeur, il y a un réel intérêt à rechercher les biomarqueurs dans les fluides biologiques car ils impliquent des collectes beaucoup moins invasives pour le patient qu'une biopsie du tissu tumoral. En tant que milieu physiologique potentiellement représentatif de la composition des tissus environnants (fluide proximal) et des processus pathologiques, le plasma ou l'urine représentent ainsi des matrices de choix pour la mesure de biomarqueurs,. La découverte de biomarqueurs dans le plasma représente un réel défi pour la protéomique clinique. En effet, des biomarqueurs protéiques sont dosés dans le sang depuis de nombreuses années, mais très peu de marqueurs

issus de la protéomique, ont été approuvés par les autorités de santé comme la FDA (Food and Drug Administration). Ce n'est que très récemment, que la FDA a approuvé le premier test diagnostique développé par une approche protéomique : l'OVA1. Ce test repose sur le dosage d'un panel de 5 biomarqueurs pour le diagnostic du cancer de l'ovaire (Fung, 2010). Pour être validé par ces autorités, un biomarqueur doit être fiable et reproductible. Il doit en plus présenter une sensibilité (proportion de cas diagnostiqués par le marqueur, c'est-à-dire de vrais positifs) et une spécificité (proportions de sujets indemnes reconnus par le marqueur, c'est-à-dire de vrais négatifs) très élevées. Cependant, dans un fluide proximal, les protéines caractéristiques d'une tumeur sont moins concentrées que dans le tissu tumoral, et diluées dans une grande variété de composés sans liens avec la pathologie étudiée. Il était donc nécessaire de caractériser en profondeur le protéome urinaire pour accéder aux protéines minoritaires spécifiques de la pathologie étudiée.

2. Méthodologie protéomique

D'une manière générale, les études protéomiques impliquent une analyse par MS des peptides générés par digestion trypsique des protéines. Dans ce contexte, de telles analyses impliquent à la fois: (i) la mesure de masses peptidiques et (ii) des expériences de MS en tandem (MS / MS) afin d'obtenir des informations de séquence permettant l'identification des protéines dont sont issus ces peptides. En outre, lorsque les mélanges de peptides à analyser sont très complexes, le couplage avec la chromatographie liquide (LC) permet d'améliorer la caractérisation du protéome en limitant le nombre de peptides soumis au spectromètre de masse au même moment.

Il a été montré que, lorsque la masse d'un peptide est mesurée avec une précision telle qu'elle est unique parmi toutes les masses peptidiques prédites à partir d'une séquence génomique déterminée, elle peut être utilisée comme marqueur spécifique d'identification d'une protéine, marqueur que l'on nomme AMT pour « Accurate Mass Tag ». La précision de mesure requise (partie par million) est actuellement accessible en utilisant la spectrométrie de masse à transformée de Fourier (FTMS). Cette propriété essentielle est mise à profit dans une méthode d'identification de peptides appelée approche AMT qui se déroule en deux phases consécutives. La première est une étape classique de LC-MS/MS sur des échantillons fractionnés pour générer une base de données. Cette base de données nommée « base AMT » contient toutes les masses des peptides calculées sur la base de leurs séquences en acide aminés et leurs temps de rétention chromatographiques observés. Dans la seconde phase, ces « AMT tags » sont utilisés, lors d'analyses nanoLC-FT-MS, comme des signatures des peptides précédemment caractérisés sans recourir systématiquement à la MS/MS pour l'identification. Par conséquent, une fois la base de données AMT générée, il devient possible d'identifier et de quantifier des milliers de peptides dans une analyse. Cette approche AMT permet en particulier d'utiliser les méthodes de quantification sans marquage ou « Label Free » qui consistent en la comparaison des intensités des signaux MS pour apprécier l'abondance relative des protéines issues de chaque échantillon afin de mettre en évidence celles dont l'abondance varie en fonction de perturbations dues à une pathologie. Il est ainsi possible de comparer plusieurs protéomes, sans

limitations de nombre comme dans le cas des méthodes de quantification avec marquage isotopique. Pour ces différentes raisons, l'approche AMT représente une méthode de choix, dans le cadre d'une étude sur la découverte de biomarqueurs.

B. Exploration du protéome urinaire

La caractérisation du protéome urinaire représente ainsi la première phase de l'approche AMT. La base de données qui en résulte va ensuite permettre d'accéder à l'identifiant des peptides dont l'intensité sera mesurée pendant l'analyse quantitative des échantillons provenant de patients sains ou pathologique en vue de découvrir de nouvelles protéines biomarqueurs du cancer de la vessie. Dans cette optique de caractérisation, de récentes études ont démontré que le protéome urinaire contenait une bien plus grande variété de protéines que ce qui était supposé auparavant (Adachi et al., 2006), (Kentsis et al., 2009). Ces études ont permis en particulier de montrer qu'il était possible de détecter des protéines minoritaires dans l'urine. En extrapolant ces travaux, nous avons pressenti qu'il pouvait être possible d'accéder à des marqueurs potentiels du cancer de la vessie grâce à une analyse approfondie, qualitative et quantitative, du protéome urinaire de patients sains ou malades.

Comme nous l'avons vu précédemment, l'urine est un fluide biologique extrêmement complexe contenant une grande variété de protéines, et bien que sa gamme dynamique soit moins élevée que celle du plasma, elle reste malgré tout importante. En dépit des avancées récentes en spectrométrie de masse, comme la vitesse d'acquisition des spectres MS/MS et la sensibilité, il reste à l'heure actuelle nécessaire d'utiliser des méthodes de décomplexification de l'échantillon pour accéder aux protéines minoritaires qui sont autant de biomarqueurs potentiels. A cet effet, plusieurs stratégies sont disponibles. Les protéines et les peptides peuvent être fractionnés suivant différents paramètres physico-chimiques : leur poids moléculaire, leur hydrophobicité, leur charge ou leur point isoélectrique. Il est possible d'effectuer ces fractionnements avant ou après l'étape de digestion. Dans le cas des protéines non digérées, la méthode la plus répandue est basée sur la séparation sur gel SDS-PAGE. Au niveau des peptides obtenus par digestion, la méthode la plus classique est celle de la séparation multidimensionnelle par chromatographie liquide par échange de cation couplée à la phase inverse connue sous le nom de MudPIT (H Liu, Lin, & Yates, 2002). Cependant les gradients de sels combinés aux gradients de phases organiques sont bien souvent difficiles à concilier sur une même chaîne de chromatographie liquide et ceci peut poser des problèmes de robustesse (Slebos et al., 2009). Une méthode alternative à MudPIT, basée sur la séparation des peptides suivant leur point isoélectriques (IEF OFFGEL), a également montré d'excellents résultats pour l'analyse d'échantillons complexes (Hubner et al., 2008).

Pour améliorer la détection des protéines minoritaires, il est également possible d'agir sur la gamme dynamique de l'échantillon en procédant à la déplétion totale ou partielle de protéines majoritaires. Ainsi, dans le cas du sérum ou du plasma, les meilleures couvertures de protéomes ont été obtenues suite à la

déplétion des protéines les plus abondantes telle que l'albumine, et les immunoglobulines à l'aide d'un mélange d'anticorps spécifiques immobilisés sur une matrice chromatographique (Stempfer, Kubicek, Lang, Christa, & Gerner, 2008). Il est à noter que dans le cas de l'urine, les kits commerciaux élaborés pour le sérum ne sont pas totalement adaptés car d'autres protéines que celles du sérum, telles que l'uromoduline, y sont présentes à haute concentration. Par ailleurs, la méthode Equalizer, que nous verrons plus en détail par la suite, constitue une stratégie originale qui permet d'éliminer partiellement les protéines les plus abondantes par saturation de leurs sites de fixation sur une librairie combinatoire de ligands hexapeptides greffés sur des billes de polyméthacrylate.

Dans le cadre de ce travail, afin d'accéder à la meilleure couverture du protéome urinaire, nous avons sélectionné trois méthodes complémentaires de fractionnement ou de déplétion des protéines. Nous avons réalisé, pour référence, un fractionnement protéique par SDS-PAGE. Nous avons eu recours ensuite à deux technologies innovantes basés sur des principes physico-chimiques très différents : un fractionnement des peptides par focalisation isoélectrique et une déplétion partielle des protéines majoritaires par la méthode Equalizer.

1. Fractionnement de l'échantillon

a) *Au niveau protéique*

Le gel SDS-PAGE permet de séparer les protéines en fonction de leur poids moléculaire mais aussi d'éliminer les impuretés présentes dans les échantillons urinaires. C'est une technique facile et rapide à mettre en œuvre, et qui permet d'analyser une grande variété de protéines de caractéristiques physico-chimiques diverses (protéines basiques ou hydrophobes). Comme son pouvoir de séparation est limité, ainsi qu'en témoigne la présence d'un grand nombre de protéines dans certaines bandes de gel 1D lors de l'analyse d'échantillon complexes (fluide biologique), elle est typiquement couplée à une seconde étape de séparation des peptides par chromatographie liquide avant l'analyse MS. Ce couplage est souvent désigné sous le terme GeLC-MS. La figure IV.1 décrit succinctement le protocole expérimental d'une analyse GeLC-MS. L'échantillon biologique est dans un premier temps fractionné sur gel 1D. Ensuite, le gel est découpé en une quarantaine de bandes. Chacune des bandes est lavée et les protéines contenues dans le gel sont digérées à la trypsine. Les peptides ainsi générés sont extrudés du gel par traitements successifs en solvants organique et aqueux en vue de leur analyse par nanoLC-MS/MS.

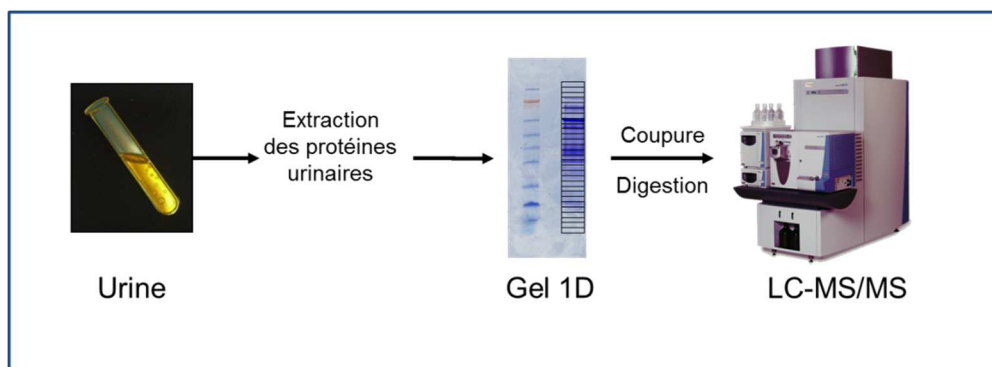


Figure IV.1 : Principe général de la stratégie GeLC-MS

La technologie GeLC-MS est une méthode de référence en protéomique. En 2006, la publication d'Adachi (Adachi et al., 2006), a révolutionné la protéomique urinaire. Avant cette étude, l'estimation du nombre de protéines urinaires basés sur une compilation des résultats de différentes équipes était d'environ 800 (Castagna et al., 2005). En utilisant la technique GeLC-MS et une instrumentation avancée de spectrométrie de masse (LTQ-Orbitrap), conduisant à la détection de plus de 1500 protéines urinaires, Adachi a démontré la complexité véritable de ce fluide.

Dans le cadre de cette thèse, nous avons donc utilisé cette technique pour caractériser les protéines extraites d'échantillons d'urines individuels ou poolés issus de 60 patients atteints de cancer de la vessie ou de donneurs sains. Nous avons analysé un total de 693 fractions SDS-PAGE conduisant à l'identification de 12868 peptides de séquences uniques correspondant à 1595 protéines non redondantes. Ce résultat, très similaire à celui obtenu par Adachi, représente un premier jalon vers la caractérisation du protéome urinaire dans le contexte du cancer de la vessie. En particulier, on pouvait se demander si c'était là une limite de la technologie de fractionnement utilisée. Le fractionnement par gel SDS-PAGE présente, en effet, certaines limitations telles qu'une perte de débit analytiques liée au grand nombre de fractions qu'il engendre (40 bandes /échantillons) et des problèmes de récupération de l'échantillon liés à la digestion ou à l'extraction des peptides du gel (Brun et al., 2007).

Dans le chapitre précédent nous avons établis un protocole standardisé basé sur l'extraction des protéines par précipitation TCA suivie d'une digestion en solution. Ce protocole a été sélectionné sur la base d'observations quantitatives en vue de l'étude sur la recherche de nouveaux candidats biomarqueurs du cancer de la vessie dans les urines. Ce protocole donnant directement accès à des échantillons de peptides en solution ; nous nous sommes intéressés aux techniques de fractionnement peptidiques, et en particulier à l'Isoélectrofocalisation.

b) Au niveau peptidique (IEF) Hors gel

L'approche IEF hors gel représente une alternative au fractionnement par Gel SDS-PAGE qui permet de s'affranchir des limitations soulevées précédemment. Ainsi, en travaillant en solution le

rendement de digestion est meilleur et peut être vérifié (voir chapitre II). De plus, le nombre de fractions (12 ou 24) suivant le système utilisé est nettement inférieur à celui du Gel SDS-PAGE, impliquant un débit analytique plus élevé. Cette technique commercialisée, par Agilent sous le nom « OFFGEL Fractionator », a fait l'objet d'une comparaison approfondie avec le fractionnement par gel SDS-PAGE dans une publication d'Hubner en 2008 (Hubner et al., 2008) (voir Figure IV.2).

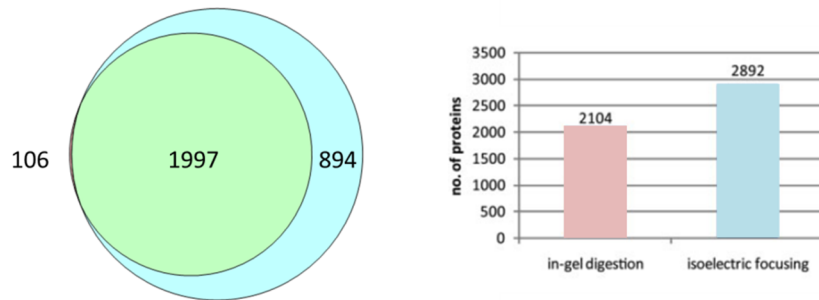


Figure IV.2 : Comparaison de la couverture du protéome de levure obtenue après fractionnement sur gel ou par IEF hors-gel. Tiré de Hubner et al. 2008.

A partir de 50µg d'extraits de levures, Hubner a identifié 2892 protéines en fractionnant l'échantillon par IEF et 2104 protéines en fractionnant ce même échantillon par Gel-SDS-PAGE. De plus, il a observé que la quasi-totalité des protéines identifiées par les deux méthodes l'étaient par fractionnement IEF, alors que seulement quelques protéines étaient identifiées uniquement par GeLC-MS. Ces observations ont été répliquées à partir de différentes quantités de protéines déposés séparées par gel SDS-PAGE et IEF. Il est ainsi apparu que la technique d'isoélectrofocalisation donnait systématiquement de meilleurs résultats en termes de couverture protéique ou peptidique que la GeLC-MS. Les performances de la technique IEF sont donc supérieures à celles du gel sur plusieurs critères : une meilleure couverture du protéome, une digestion plus contrôlée avec un meilleur rendement que la digestion in gel et un débit d'analyse supérieur à celui de la GeLC-MS. Nous avons donc testé cette technique sur un échantillon d'urine.

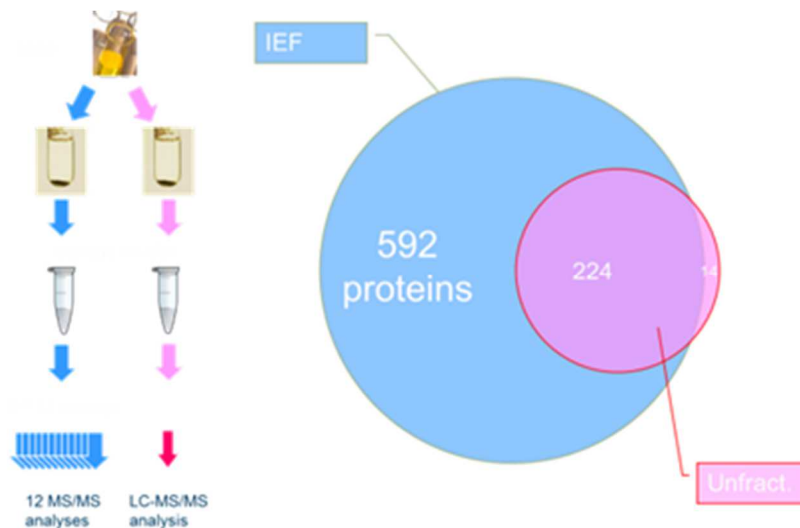


Figure IV.3 : Comparaison de la couverture du protéome urinaire après précipitation TCA et digestion en solution avant et après fractionnement par IEF hors gel.

Après avoir extrait les protéines urinaires par précipitation TCA, les protéines ont été dénaturées à l'urée et digérées en solution. Après un premier dessalage des peptides obtenus sur colonne C18, l'échantillon a été repris dans une solution compatible avec le fractionnement par isoélectrofocalisation hors gel et analysé par LC-MS/MS. Nous avons ensuite compilé l'ensemble des résultats obtenus dans une base de donnée, groupé les protéines identifiées dans les 12 fractions et confronté cette liste de protéine à la liste obtenue à partir du même échantillon non fractionné et analysé par LC-MS/MS (voir Figure IV.3). Nous avons ainsi identifiés un total de 830 protéines avec les critères de validation décrits au chapitre II, dont la quasi-totalité (816) a été identifiée à partir des 12 fractions IEF. Ces résultats nous ont permis de conclure qu'il était possible d'utiliser cette technique sur des échantillons urinaires préparé selon notre protocole standardisé défini dans le chapitre III : précipitation TCA suivi d'une digestion en solution. De plus, le nombre de protéines identifié à partir d'un échantillon contrôle fractionné par Gel SD-PAGE était d'environ 700 protéines. Le résultat obtenu sur un échantillon fractionné par IEF était donc similaire, voire légèrement supérieur à celui obtenu par GeLC-MS.

Par ailleurs, dans le cadre de ce projet, nous disposons d'une cohorte de patients décrite en détail dans le chapitre suivant. Cette cohorte était composée de patients sains et de patients atteints de tumeurs de la vessie. Les échantillons urinaires de ces patients ont été regroupés en 10 pools constitués par les cliniciens selon des critères de risque de progression ou de récurrence de la tumeur. Chaque pool contenait chacun les urines de 10 patients. Les protéines urinaires ont été extraites par précipitation TCA et digérées en solution selon notre protocole standardisé, les peptides ainsi obtenus ont été fractionnés en 12 fractions par IEF. A l'issue de l'analyse des 120 fractions obtenues à partir des 10 pools de patients sains ou malades, nous avons identifiées un total de 8903 peptides correspondant à 1357 protéines non redondantes. Ce résultat,

comparable en profondeur d'analyse à celui obtenu avec le fractionnement par Gel SDS-PAGE et à celui d'Adachi correspond à une bonne couverture du protéome urinaire.

Réduction de la gamme dynamique : Equalizer

Parallèlement, à l'évaluation des techniques de fractionnement, nous avons testé une méthode permettant de réduire la gamme dynamique d'échantillons complexes appelée Equalizer. Cette technologie, parue en 2008, a été développée par Biorad et commercialisée sous le nom de ProteoMiner™. Cette approche permet de réduire la gamme dynamique en égalisant les concentrations protéiques d'un échantillon biologique. Pour cela, l'échantillon est incubé en présence d'une banque combinatoire d'hexapeptides (soit 20^6 ou 64 millions de ligands différents) greffés sur des microbilles de polyméthacrylate. Chaque bille porte des millions de copies d'un unique ligand. Théoriquement, chaque protéine de l'échantillon complexe possède un ligand pour lequel son affinité est maximale. Une fois les sites de leurs ligand préférentiel saturés, les protéines en excès sont éliminées avec le surnageant, permettant ainsi de réduire les différences de concentration entre les protéines les plus et les moins abondantes de l'échantillon.

L'Equalizer a d'abord démontré son potentiel pour la caractérisation du sérum (Thulasiraman et al., 2005) qui présente une très importante gamme dynamique. Il a ensuite été testé sur l'urine par Castagna en 2005 (Castagna et al., 2005). Dans cette étude en partant d'un même échantillon urinaire, 134 protéines ont été identifiées après traitement d'urine par un protocole conventionnel (GeLC-MS), et 383 protéines après l'étape d'égalisation. Ils ont ainsi quasiment triplé le taux d'identification de protéines urinaires grâce à cette déplétion. Sur la base de ce résultat, nous avons estimé que cette technique avait le potentiel d'augmenter de façon substantielle notre couverture du protéome urinaire et d'accéder ainsi à des candidats biomarqueurs très faiblement abondants. Néanmoins, nous notons que le volume d'urine utilisé par Castagna et al. pour la déplétion était considérable (1,6 litre) et difficilement compatible avec les contraintes d'une étude clinique telle que la nôtre. En considérant un volume moyen d'urine par patient de 80mL, il faudrait collecter les urines d'une vingtaine de patients de même pathologie (stade, grade, risque de progression ou de récurrence) pour pouvoir obtenir une telle quantité d'urine. Or, en fonction de la prévalence de la maladie étudiée, cela peut prendre plusieurs mois voire des années pour réunir un aussi large panel de patients, et c'est parfois tout simplement infaisable. Ainsi, dans notre cas, nous disposons d'un volume d'urine de seulement 40 mL par patients à partager entre trois centres de recherches impliqués dans la découverte et l'évaluation des biomarqueurs. Nous avons donc été obligés d'adapter le protocole en conséquence.

D'après les concepteurs de la méthode Equalizer, le paramètre le plus critique pour l'égalisation est le taux de saturation des billes, c'est à dire le ratio entre la capacité de charge des billes et la quantité totale de protéines dans l'échantillon. Pour permettre une bonne interaction entre les protéines et leurs ligands, il est préconisé de travailler avec un taux de saturation de 50x pour des performances optimales, soit 50mg de

protéines pour un volume de 100 μ L de billes. Si cette quantité de protéines est aisément atteinte avec 1mL de plasma, elle correspond à environ 500mL d'urine. S'il est possible de jouer sur le volume d'urine, il est également possible de réduire le volume de bille. Cependant, le risque encouru est que cela induise un sous-échantillonnage de la librairie de ligands, c'est-à-dire que la sous-population de billes utilisée n'est plus représentative de la complexité de la librairie. Ceci pourrait engendrer une perte de reproductibilité entre les analyses. De plus, réduire le volume de billes a pour résultat de diminuer la capacité de charge et donc la sensibilité de l'analyse. Pour évaluer ces différents facteurs, nous avons fait varier le volume de billes en fixant le volume d'urine à 8mL (Tableau IV.1). Partant d'un taux de saturation proche de celui qui était recommandé (20x), nous sommes descendus à des valeurs très inférieures (2.5x). Il ne nous a pas été possible d'atteindre le taux préconisé de 50x, car nous étions limités par le volume trop faible de billes à prélever (inférieur à 2 μ L).

Tableau IV.1 : design de l'optimisation de la quantité de billes Equalizer pour la déplétion des protéines à partir d'un faible volume d'urine.

Volume initial d'urine	Volume de billes Equalizer	Taux de saturation des billes
8mL	2 μ L	20 x
8mL	4 μ L	10 x
8 mL	16 μ L	2.5 x

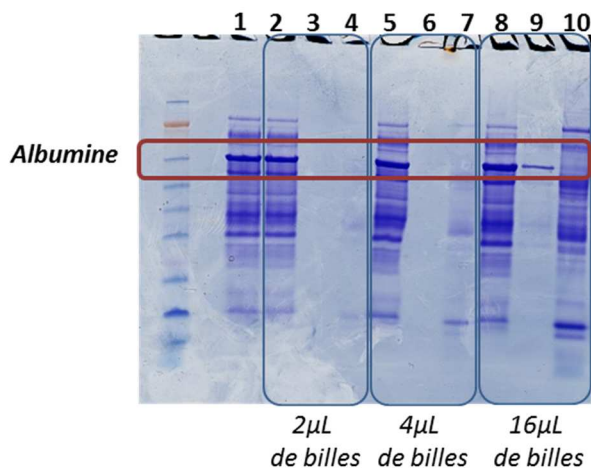


Figure IV.4 : Gel SDS-Page des extraits protéiques obtenus après déplétion partielle des protéines majoritaires par la méthode Equalizer d'un échantillon d'urine de 8 mL à l'aide de quantités variables de billes. Piste 1 : contrôle (échantillon ultrafiltré). Pistes 2, 5, 8 : fractions non retenues (après ultrafiltration). Pistes 3, 6, 9 : lavages. Pistes 4, 7, 10 : Extraits élués des billes.

On observe sur la figure IV.4 les profils protéiques obtenus à partir d'un même échantillon traité dans les différentes conditions de déplétion décrites dans le tableau IV.1 comparés au profil du même échantillon non égalisé correspondant à notre condition contrôle (piste 1). Les piste 2 à 10 correspondent à 8 mL d'urine égalisée avec : 2 μ L de billes Equalizer (pistes 2 à 4), 4 μ L de billes (pistes 5 à 7) et 16 μ L de billes (pistes 8 à 10). Après avoir incubé les différents volumes de billes dans 8 mL d'urine et les avoir lavées de manière à éliminer les protéines en excès (pistes 2-3, 5-6 et 8-9), nous avons élué les protéines fixées sur ces billes à l'aide de tampon de Laemmli (Pistes 4, 7 et 10).

On observe que les profils protéiques correspondant aux fractions non retenues (piste 2, 5 et 8) sont très similaires à l'échantillon initial, ce qui permet de confirmer que l'on n'a pas extrait une grande proportion des protéines en excès dans l'urine. Par contre, à l'exception d'une bande correspondant à l'albumine sur l'échantillon à 16 μ L de billes, très peu de protéines sont élués lors de l'étape de lavage (pistes de gel 3, 6 et 9). De plus, les profils protéiques d'éluion correspondants aux volumes de billes 2 et 4 μ L (piste 4 et 7) sont très faiblement colorés. Ceci confirme que la capacité de charge des billes quand une trop faible quantité est utilisée, entraîne une perte de sensibilité de l'analyse. Cependant, les résultats obtenus avec un volume de billes de 16 μ L sont très encourageants. En effet, on observe un enrichissement du profil protéique et une diminution très significative de la bande correspondant à l'albumine (protéine majoritaire de l'urine). En conclusion, il semble donc possible de dépléter efficacement les protéines majoritaires avec un taux de saturation 20 fois moindre que celui recommandé en partant de seulement 8mL d'urine.

Cependant, étant donné que nous avons utilisé une fraction de la librairie de ligands (16 μ L d'un volume globale de 100 μ L), nous nous sommes inquiétés du risque de sous-échantillonnage de la librairie et du risque d'irreproductibilité ainsi encouru. Nous avons donc évalué la répétabilité de la mesure en utilisant le même taux de saturation (2.5X) mais en incorporant, cette fois-ci, la totalité de la librairie de ligands à l'échantillon d'urine. Pour ce faire, nous avons incubé la totalité des billes dans 50 mL d'urine pour obtenir le même taux de saturation. Les résultats obtenus sur le profil protéique du gel SDS-PAGE étaient très similaires à ceux obtenus à partir de 8 mL d'urine incubé avec 16 μ L de billes. Cependant, pour vérifier que nous obtenions des résultats homogènes à partir des deux volumes urines (50 et 8 mL), nous avons confronté les listes de protéines obtenues à partir des analyses LC-MS/MS correspondant à ces deux conditions. Les deux échantillons ainsi traités ont permis d'identifier 1103 protéines au total : 993 ont été obtenues à partir de 50 mL d'urine égalisée avec 100 μ L de billes et 954 à partir de 8 mL d'urine égalisée avec 16 μ L de billes. Au total, on obtient 77% de recouvrement entre les deux méthodes, ce qui est tout à fait comparable au taux de recouvrement entre deux réplifications d'un même échantillon protéique. Ceci démontre que le faible volume de bille utilisé pour les 8 mL d'urine ne génère pas de biais expérimental dû au sous échantillonnage de la librairie de ligands.

En conclusion, il a été démontré qu'il était possible de travailler avec un taux de saturation bien moindre que celui préconisé par Biorad (2.5X), et que cela n'affecte pas la reproductibilité. Cela permet

d'égaliser les protéines urinaires à partir de plus faibles volumes d'urine (8 mL) compatibles avec des études de protéomique clinique. Nous avons ainsi réalisé 120 analyses LC-MS/MS avec cette technique, à partir d'échantillons urinaires issus de 30 donneurs sains. L'ensemble de ces analyses a permis d'identifier un total de 12035 peptides correspondant à 1467 protéines non redondantes.

Cette approche présente, cependant, quelques limitations. En effet, il est extrêmement difficile de récupérer la totalité des billes par sédimentation après l'étape d'incubation. Une amélioration possible de la méthode serait d'utiliser des billes magnétiques afin d'améliorer la reproductibilité des analyses mais également la couverture protéique. De plus, il n'a pas encore été démontré la faisabilité de digérer en solution les protéines fixées sur les billes. Développer un tel protocole permettrait de coupler les avantages de l'égalisation aux avantages du fractionnement par IEF hors gelet d'obtenir une meilleure couverture protéique.

On note également que les résultats obtenus en termes de couverture du protéome urinaire sont nettement supérieurs à ceux de Castagna. En effet on triple quasiment le nombre de protéines qu'ils ont identifiées. Cette différence peut s'expliquer en partie par l'évolution de l'instrumentation (LTQ-Orbitrap) mais aussi par le traitement préanalytique des échantillons. Lors de notre analyse, nous avons tenté d'ultrafiltrer de plus grand volumes d'urines et nous avons observé que les protéines urinaires précipitaient lors de la préconcentration de l'urine dans les tubes vivaspin. Nous n'avons donc pas conservé ces échantillons pour le reste de l'étude. Ainsi, lorsque Castagna ultrafiltre 1.6 litre d'urines jusqu'à 45 ml, il est possible qu'il ait perdu une grande proportion des protéines à cette étape de préconcentration.

2. Comparaison des techniques de fractionnement

Afin de déterminer si une méthode était meilleure que les autres en termes de couverture protéique, de gamme dynamique et de cout associé à l'obtention de cette couverture protéique, nous avons comparé les techniques de fractionnement et de déplétion que nous avons utilisé. Enfin, nous avons analysé les termes de la Gene Ontology associés aux protéines identifiées, afin d'évaluer s'il n'y avait pas de biais en termes de classes de protéines identifiées, introduit par ces méthodes de fractionnement ou de déplétion.

En partant des seules entités qui peuvent être comparées entre elles, c'est à dire les peptides identifiés, nous avons élaboré une méthode permettant de comparer des listes de protéines à partir de différentes analyses LC-MS/MS. Lorsque l'on veut comparer des listes de protéines qui proviennent de plusieurs analyses LC-MS/MS, il faut un algorithme qui n'était pas disponible à l'époque où cette étude a été réalisée. Nous avons donc développé un tel algorithme qui a été, par la suite, intégré dans un logiciel dédié développé au laboratoire : hEIDI (Exploration and Integration of Data and Identification). Ce logiciel open-source est maintenant disponible à la communauté (<http://biodev.extra.cea.fr/docs/heidi>). Quelques définitions sont nécessaires afin de mieux comprendre le fonctionnement du logiciel. Ainsi, une *identification* est un résultat généré par le logiciel Mascot et compilé dans une base de données

relationnelle appelée Mass Spectral Identification database (MSIdb). Typiquement, dans notre pipeline actuel, les résultats de recherche de Mascot sont validés à l'aide du logiciel IRMa, et exportés vers la MSIdb où ils sont représentés comme des *identifications* (Figure IV.5). Le plan d'expérience peut être représenté comme une structure hiérarchique où chaque « nœud », appelé *contexte*, correspond à un stade expérimental particulier, incluant les *identifications* correspondantes. Ceci permet de regrouper au niveau d'un *contexte* l'ensemble des fractions d'un même échantillon. Par exemple, dans notre étude, un *contexte* appelé GeLC-MS pourra regrouper toutes les *identifications* obtenues à la suite des analyses d'extraits urinaire après fractionnement sur gel.

Succinctement, hEIDI, permet de créer, d'explorer et de comparer différents *contextes*, de filtrer les peptides et les protéines selon le nombre de peptides qui permettent de les identifier. Ce logiciel permet également de regrouper les protéines identifiées par un même ensemble ou sous ensemble de peptides (appelé aussi protéine groupe). Enfin, il est possible de comparer des listes de protéines et d'exporter les résultats dans une base de données AMT, au format Access compatible avec les logiciels utilisés pour la quantification.

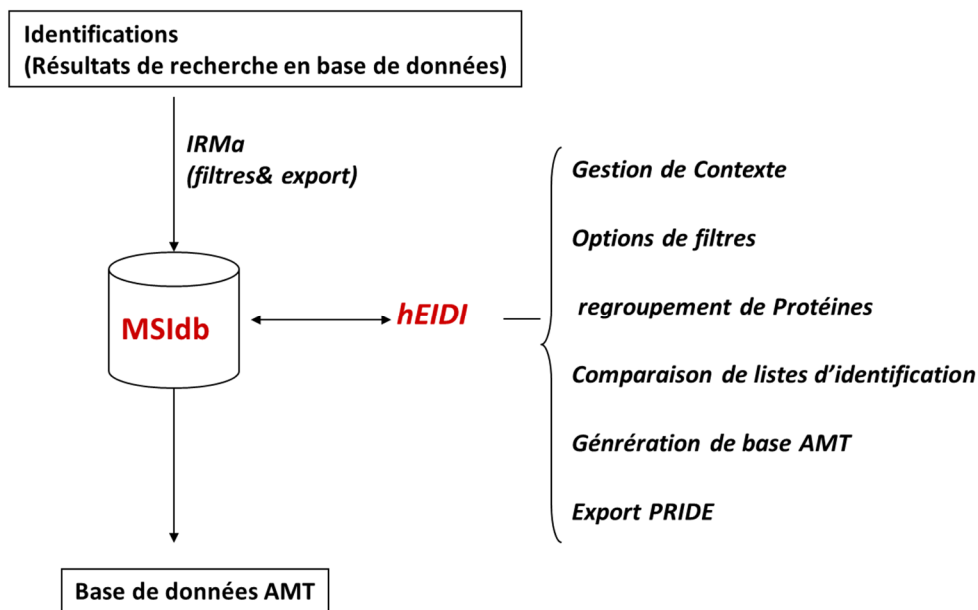


Figure IV.5 :

Périmètre fonctionnel du logiciel hEIDI utilisé pour compiler les résultats de multiples identifications au sein de contextes définis par l'utilisateur. Ces contextes peuvent ensuite être comparés entre eux en termes de protéine groupes et de peptides identifiés.

La méthode de comparaison que nous avons utilisée, consiste en premier lieu à combiner ensemble plusieurs *identifications* dans un *contexte*. Ainsi nous avons regroupé les analyses LC-MS/MS réalisées dans les différentes méthodes de fractionnement ou de déplétion, dans un seul *contexte*. Dans un deuxième temps, l'ensemble des protéines identifiées ont été regroupées en *protéine groupes*. On exporte ensuite la liste des

peptides non-redondants, les protéines et *protéines groupes* correspondants dans une base de données AMT. On établit ensuite la correspondance entre chaque AMT tags et les peptides de la MSIdb sur la base de leur séquence et de leur masse calculée pour retrouver quelle méthode avait permis leur identification. Enfin, on attribue alors à chaque *protéine groupe* la ou les méthodes de fractionnement ou de déplétion qui ont permis de l'identifier.

Tableau IV.2 : Exemple simplifié de comparaison entre deux listes de protéines obtenues par des méthodes de fractionnement différentes.

Protein	Equalizer	IEF	Résultat final
Protein P1	● ● ● ● ● ●	● ● ● ● ● ●	●● ●●● ●● ●
Protein P2	● ● ● ● ● ● ● ●	● ● ● ● ● ● ●	●●● ●●● ●●● ●●
Protein P3	● ● ● ●		●● ●●
Protein P4		● ● ●	●● ●

- Peptide identifié par Equalizer
- Peptide identifié par IEF

Le tableau IV.2 présente un exemple simplifié de comparaison de deux contextes. Ainsi, si on a identifié les protéines P1, P2 et P3 avec la méthode Equalizer et que la méthode IEF a permis d'identifier les protéines P1, P2 et P4, on regroupe toutes ces protéines dans un contexte global (colonne Résultat final). L'interrogation du tableau complet permet de déduire que les protéines P1 et P2 ont été identifiées par les deux méthodes Equalizer et IEF, que la protéine P3 a été identifiée uniquement avec la méthode Equalizer et P4 uniquement avec la méthode IEF. Il est de plus important de comprendre que, lors du regroupement des protéines en *protéine groupes* dans un *contexte* plus large, des protéines identifiées dans les sous-contexte par un même sous-ensemble de peptides peuvent être discriminées en plusieurs *protéines groupes* sur les base de nouveaux peptides identifiés. La figure IV.6 illustre ce phénomène à travers l'exemple d'un premier *contexte*, dans lequel on identifie les protéines P1 et P2. P1* constitue alors le seul *protéine groupe* puisque dans le set de peptides qui permet de l'identifier, il y en a au moins un qui lui est spécifique (p1). Par contre P2 n'est pas un *protéine groupe* mais une protéine « sub-set » du *protéine groupe* P1* puisqu'on n'a pas identifié de peptide spécifique de cette protéine (p2 et p3 permettant d'identifier également P1). Dans le second contexte, P1 et P2 sont tous les deux des *protéine groupes*, puisqu'ils ont au moins un peptide

spécifique (p1 pour P1 et p4 pour P2). Lorsque l'on regroupe les deux *sous-contextes* dans un *contexte* plus global, P1 et P2 deviennent ainsi toutes deux des *protéine groupes*.

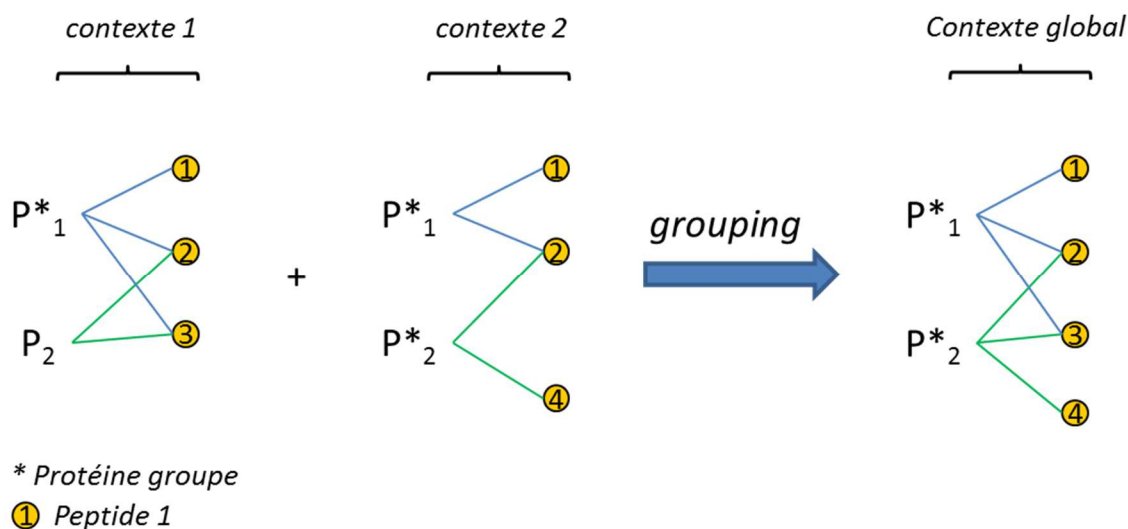


Figure IV.6. Comment le fait de regrouper les identifications de plusieurs contextes permet de lever l'ambiguïté concernant des protéines appartenant initialement au même protéine groupe par l'ajout d'un peptide non-partagé.

A partir des analyses des échantillons traités selon les différentes techniques de fractionnement et de déplétion (GeLC-MS, IEF OFF GEL et Equalizer), nous avons constitué une base de données contenant l'ensemble des protéines urinaires identifiées. Cette base comprend 18 551 peptides correspondant à 13 729 protéines regroupées en 2014 protéines non redondantes ou *protéine groupes*.

Tableau IV.3 : résumé de caractéristiques des trois techniques comparées dans le cadre de la caractérisation du protéome urinaire par différentes techniques de fractionnement.

Paramètres	Gel SDS-PAGE 1D	Equalizer	Précipitation TCA - IEF
Donneurs	Sains et malades	sains	sains et malades
Nombres de donneurs approx.	60	30	100
Quantité d'urine	2mL	8-50mL	10mL
Nombres d'analyses LC-MS/MS	693	121	120
Nombres de protéines identifiées	1593	1467	1614

Dans le tableau IV.3, on a regroupé les performances et caractéristiques des trois méthodes utilisées pour accéder aux 2014 protéines urinaires. Il est à noter que les fractionnements GeLC-MS ou IEF ont été effectués sur des échantillons de patients sains et malades, alors que la déplétion Equalizer n'a été effectuée que sur des donneurs sains. Les patients dont les urines ont été fractionnées par la technique IEF sont les patients de la cohorte qui a ensuite servi pour la quantification et la découverte de nouveaux biomarqueurs de cancer de la vessie (voir Chapitre V). En ce qui concerne la technologie Equalizer, nous avons collecté les

urines de personnels du laboratoire pour tester les différentes conditions citées plus haut. Le fractionnement par gel SDS-PAGE a été réalisé à partir de 2 mL d'urine, tandis que l'Equalizer ou l'IEF ont nécessité un plus grand volume 8 et 50 mL pour l'Equalizer et 10mL pour l'IEF. Ces différents volumes d'urine, mis à part les 50 mL correspondant au contrôle de reproductibilité de l'Equalizer, sont tout à fait compatibles avec un projet clinique dans lequel il est commun de partager les échantillons entre différents partenaires sur différents axes de recherche. Enfin le nombre de protéines identifiées est sensiblement identiques quel que soit la méthode utilisée pour caractériser le protéome urinaire. Cependant, les nombres d'analyses qu'il a fallu pour atteindre ces résultats sont quant à eux, très disparates allant de 120 jusqu'à environ 700 analyses. Les résultats en termes de nombre de protéines identifiées sont légèrement différents de ceux décrits dans les paragraphes dédiés à la description de chaque méthode. Cette différence est due au fait que l'on a créé un *contexte* global regroupant toutes les protéines urinaires identifiées quelle que soit la méthode de fractionnement ou de déplétion et que le regroupement des protéines en *protéine groupes* est réalisé à ce niveau. Ainsi des peptides identifiés dans les autres traitements vont permettre de différencier des *protéine groupes* qui étaient initialement impossible à distinguer dans le contexte d'une seule technique, comme illustré sur la Figure IV.6.

Lorsque l'on analyse le recouvrement obtenu en termes de *protéine groupes* identifiés grâce aux méthodes de fractionnement et de déplétion, on observe que le nombre de *protéine groupes* communs représente 52% de la totalité des protéines identifiées (Figure IV.7). Si on considère maintenant les *protéine groupes* communs à au moins deux méthodes, on aboutit à environ 80% de recouvrement. Ceci indique que les méthodes utilisées pour caractériser le protéome urinaire conduisent plus ou moins à la même couverture du protéome urinaire, bien qu'elles soient conceptuellement très différentes : fractionnement des protéines, fractionnement des peptides et réduction de la gamme dynamique. On note de plus un fort recouvrement (70%) en comparant les deux méthodes de fractionnement et la technique permettant de réduire la gamme dynamique semble légèrement différentes (140 protéines communes avec la méthode IEF ou GeLC-MS, alors que l'on obtient 269 protéines communes entre les protéines identifiées par GeLC-MS et celles identifiées par IEF). Pour autant, après avoir déplété partiellement les protéines par Equalizer nous avons utilisé la GeLC-MS pour fractionner et digérer les protéines ainsi extraites.

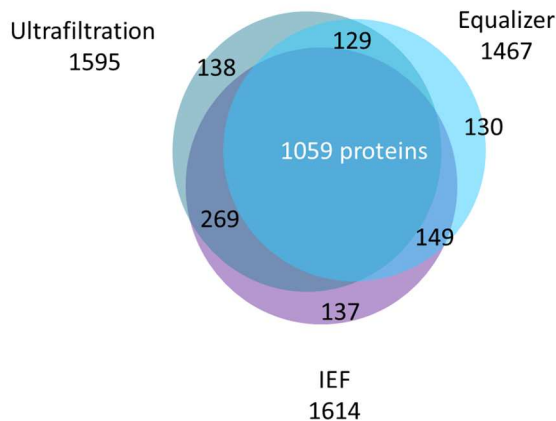


Figure IV.7 : Diagramme de Venn représentant le recouvrement des listes de protéine groupes couverts à l'aide des trois méthodes de fractionnement ou de déplétion utilisées pour la génération de la base de donnée du protéome urinaire.

Par ailleurs, le nombre d'analyse nécessaires pour atteindre les différentes couvertures protéiques doit être pris en considération. Dans une étude de protéomique, le coût moyen d'une analyse d'une heure réalisée sur une plateforme de service représente approximativement 100 euros. Ainsi il est important d'évaluer si la technique choisie pour explorer le protéome permet d'accéder rapidement à une couverture du protéome suffisante pour accéder aux protéines minoritaires qui sont autant de biomarqueurs potentiels.

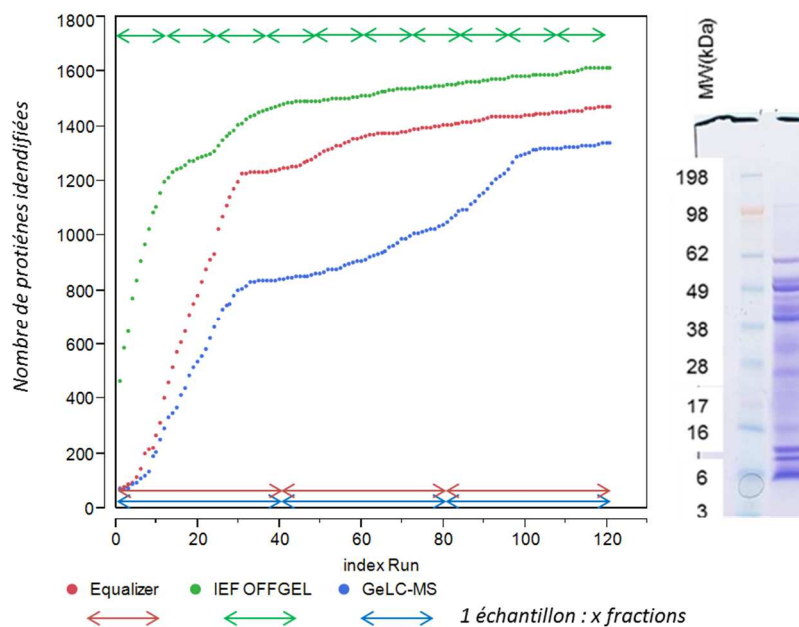


Figure IV.8 : Nombre cumulé de protéine groupes identifiés par chacune des trois méthodes de fractionnement ou déplétion utilisées pour la génération de la base de donnée urinaire en fonction du nombre de fractions analysées. Exemple de gel SDS-PAGE obtenu en vue de l'analyse GeLC-MS d'un échantillon urinaire.

La figure IV.8 représente le nombre cumulé de protéine groupes identifiés en fonction du nombre de fractions analysées par LC-MS/MS pour chacune des méthodes utilisées. On observe une croissance plus rapide de la couverture du protéome urinaire quand on utilise le fractionnement peptidique par IEF, tandis que celle associée au fractionnement par GeLC-MS est la plus modérée. Les analyses LC-MS/MS après déplétion partielle des protéines les plus abondantes par Equalizer donnent, quant à elles, des résultats intermédiaires. Dès la première analyse, on identifie environ 400 protéines avec l'IEF alors que l'on en identifie moins d'une centaine avec le fractionnement GeLC-MS avec ou sans déplétion. Ceci s'explique en partie par le fait que les premières bandes de gel analysées correspondent aux protéines de haut poids moléculaire, et lorsque l'on observe la coloration du gel (Figure IV.8), on note une faible coloration dans cette zone, indiquant qu'il y a sans doute peu de protéines dans ces bandes. Si l'on considère de surcroît que l'on ne contrôle pas le rendement de digestion ainsi que l'efficacité de l'extraction des peptides du gel, on récupère au final un peu moins de protéines que par la technique IEF. Lorsque l'on analyse un échantillon complexe en fractionnant les peptides par IEF, on a dans un premier temps vérifié que l'étape de digestion s'était bien déroulée et qu'il ne reste pas de fragments trop importants sur un Gel SDS-PAGE. De plus, étant donné que l'on sépare ensuite les peptides et non les protéines, on couvre potentiellement une large gamme de protéines dès les premières fractions. Par ailleurs, dans une bande de Gel SDS-PAGE, les protéines sont séparées selon leur poids moléculaires, et si une bande contient des protéines majoritaires, celles-ci créent une gamme dynamique défavorable à l'identification des protéines les moins abondantes. C'est pourquoi le fait d'avoir égalisé auparavant les concentrations protéiques avec la méthode Equalizer permet d'améliorer le nombre d'identifications par rapport au GeLC/MS.

Dans une publication datant de 2004, Liu (Hongbin Liu, Sadygov, & Yates, 2004) a développé un modèle capable d'estimer le nombre d'analyses qu'il faut réaliser pour atteindre la saturation du nombre de protéines identifiées sur un extrait de levure digéré à la trypsine. Ils concluent qu'au bout de 10 analyses du même échantillon, on atteint une saturation de 95% du total de protéines identifiées. Même s'il est difficile de faire un strict parallèle avec nos données, il apparaît que l'analyse d'un seul échantillon que ce soit avec un fractionnement par IEF ou une déplétion Equalizer permet d'atteindre entre 75% et 85% des protéines identifiées. En ce qui concerne la GeLC-MS, par contre, on atteint seulement 50% de la couverture totale à l'issue de l'analyse des quarante bandes du premier échantillon analysé. En conclusion, d'après la croissance du nombre de protéines en fonction des analyses LC-MS/MS, c'est le fractionnement peptidique par IEF qui permet d'accéder le plus rapidement à la meilleure couverture du protéome urinaire (1614 protéines en 120 analyses).

Si l'on ne devait choisir qu'une seule méthode de fractionnement, il serait important d'évaluer si les protéines que l'on perd ne seraient pas issues d'une catégorie fonctionnelle ou une localisation particulière. Ainsi, en faisant le choix d'utiliser uniquement le fractionnement IEF, nous avons vérifié que les protéines perdues ne permettaient pas d'enrichir une catégorie fonctionnelle, ou une localisation cellulaire.

En conclusion, nos résultats montrent qu'il n'existe pas de biais analytiques considérables dus aux méthodes de fractionnement ou de déplétion. En effet, on obtient une couverture équivalente du protéome urinaire quelle que soit la méthode utilisée. Pour autant, nous avons sélectionné pour cette phase de caractérisation du protéome, trois stratégies très diverses : une classique et deux innovantes et très prometteuses d'après la littérature. En conséquence, le fait d'opter pour une seule méthode de fractionnement ou de déplétion, ne remettrait pas en cause de manière drastique la profondeur de caractérisation du protéome urinaire. Il est à noter tout de même que l'isoélectrofocalisation présente un léger avantage avec une croissance plus rapide du nombre d'identifications en fonction du nombre d'analyses comparé aux deux méthodes utilisant le fractionnement des protéines par gel SDS-PAGE, grâce en particulier à son débit analytique accru (12 fractions à analyser comparé à une quarantaine de bandes par échantillon pour la GeLC-MS). De plus avec à cette méthode, il est possible de traiter jusqu'à 16 échantillons en parallèle. Toutes ces raisons, font que nous pourrions recommander cette méthode pour l'analyse de grand nombre d'échantillons complexes.

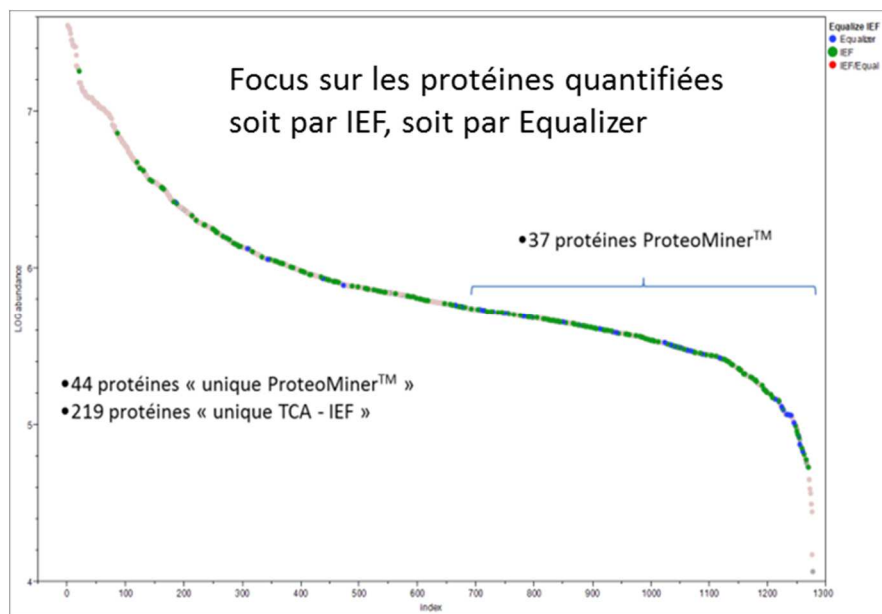


Figure IV.9 : Gamme dynamique des protéines groupes quantifiés lors de la phase de découverte (voir chapitre V). Les points de couleur correspondent aux protéines groupes identifiés exclusivement par une seule méthode de fractionnement/déplétion.

Après la phase de caractérisation du protéome urinaire, nous avons réalisé une étude de quantification des protéines urinaires pour établir une liste de candidats biomarqueurs du cancer de la vessie. Comme nous allons le voir par la suite (voir Chapitre V) nous avons ainsi pu quantifier 1378 protéines sur les 2014 protéines identifiées lors de la phase préliminaire. Nous avons également pu déterminer quelle méthode de préparation et de fractionnement avait permis l'identification de ces protéines quantifiées ainsi que leur abondance relative chez les patients analysés. La Figure IV. 9 montre la gamme d'abondance des protéines ainsi quantifiées. On note sur ce graphe que parmi l'ensemble des protéines quantifiées, 44

protéines avaient été identifiées uniquement par la technique Equalizer et, parmi ces 44 protéines, une large majorité (37) se situe parmi les protéines de plus faible abondance. Ces protéines ne se regroupent pas dans une catégorie fonctionnelle ni dans une localisation cellulaire particulière. Ce résultat confirme l'hypothèse selon laquelle la technique Equalizer permet de réduire la gamme dynamique d'échantillon complexe et donc d'accéder à l'identification de protéines minoritaires. Il est important de noter également, que bien que ces protéines aient nécessité un technique de déplétion pour être identifiées par MS/MS, elles ont pu être quantifiées par la suite dans des échantillons non-fractionnés sur la base de leur signal MS. D'après ces résultats, il nous semblerait intéressant après avoir réduit la gamme dynamique protéique à l'aide des billes Equalizer, de pouvoir digérer les protéines fixées sur les billes en solution et de fractionner les peptides résultants par IEF. Ceci permettrait en effet de coupler l'augmentation de gamme dynamique due à la déplétion avec les avantages en termes de débit et de couverture du protéome obtenus à l'aide du fractionnement par isoélectrofocalisation.

C. Base de données des protéines urinaires

1. Description générale

Une base de données urinaire a été compilée à partir de 1233 analyses LC-MS/MS d'extraits protéiques issus d'urines de patients sains et malades. Au final, nous avons identifiées 18551 peptides correspondant à 2014 protéines urinaires non redondantes. Cette base de données est représentée dans la Figure IV.10, où chaque point représente un peptide dont les coordonnées sont sa masse monoisotopique calculées sur la base de sa séquence en acide aminées et ses modifications post-traductionnelles, et le temps de rétention chromatographique observé pour ce peptide lors de la compilation de la base de donnée normalisé sur la durée de la séparation. Sur cette figure, nous avons mis en évidence quelques protéines majoritaires de l'urine telles que l'albumine, l'hémopexine, la sérotransferrine ou l'uromoduline. Ainsi, nous pouvons visualiser pour l'albumine, 247 AMT correspondant à 119 séquences en acides aminés uniques et une couverture protéique de 95.6%. Les données concernant la sérotransferrine, l'hémopexine et l'uromoduline sont regroupées dans le tableau IV.4 La médiane du nombre de peptides par protéine est de 6, et pour la plupart des protéines majoritaires de l'urine, telles que l'albumine, la sérotransferrine ou l'hémopexine, on obtient une excellente couverture protéique, supérieure à 75%. En ce qui concerne l'uromoduline, il n'a pas été possible d'identifier l'ensemble des peptides de cette protéine et en particulier le peptide signal et la partie N-terminale de la protéine ne contenant aucun site de digestion, ce qui explique la couverture relativement plus modeste de cette protéine (58.4%), alors qu'elle représente l'une des principales protéines de l'urine.

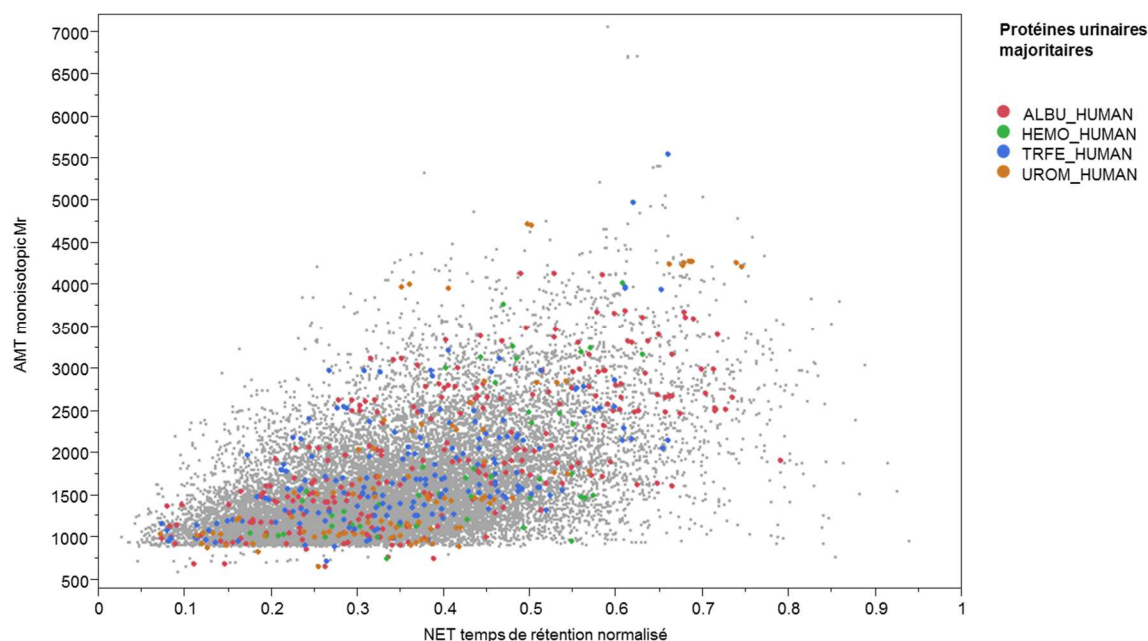


Figure IV.10 : Représentation bidimensionnelle de la base de données AMT du protéome urinaire compilée dans le cadre de ce travail. Chaque point correspond à un peptide dont les coordonnées analytiques sont la masse monoisotopique et le temps de rétention normalisé. Les peptides correspondant à quelques protéines majoritaires de l'urine ont été mis en évidence par un code couleur.

Tableau IV.4 : Nombre d'AMT et couverture de séquence de quatre protéines majoritaires de l'urine.

Protéine	Nombre AMT	Nombre de séquences uniques (acides aminés)	Couverture protéique
Albumine	247	119	95.6%
sérot transferrine	177	95	91.5%
hémopexine	49	34	74.7%
uromoduline	93	38	58.4%

Au niveau de la MSIdb utilisée pour compiler notre base AMT, le taux de faux positif, estimé par la formule de Gygi (Elias & Gygi, 2007) était de 1.53%. Cependant il est à noter que ce taux de faux positif est estimé avant les filtres sur la longueur des peptides (longueur ≥ 6 AA) et l'élimination des protéines identifiées par un seul peptide de score inférieur à 60. Ainsi, ce taux de faux positif est une surestimation de celui que l'on a réellement dans la base de données AMT.

L'analyse des termes de la Gene Ontology pour observer les classifications fonctionnelles des 2014 protéines identifiées a été faite à l'aide du logiciel ProteinCenter (ThermoFisher).

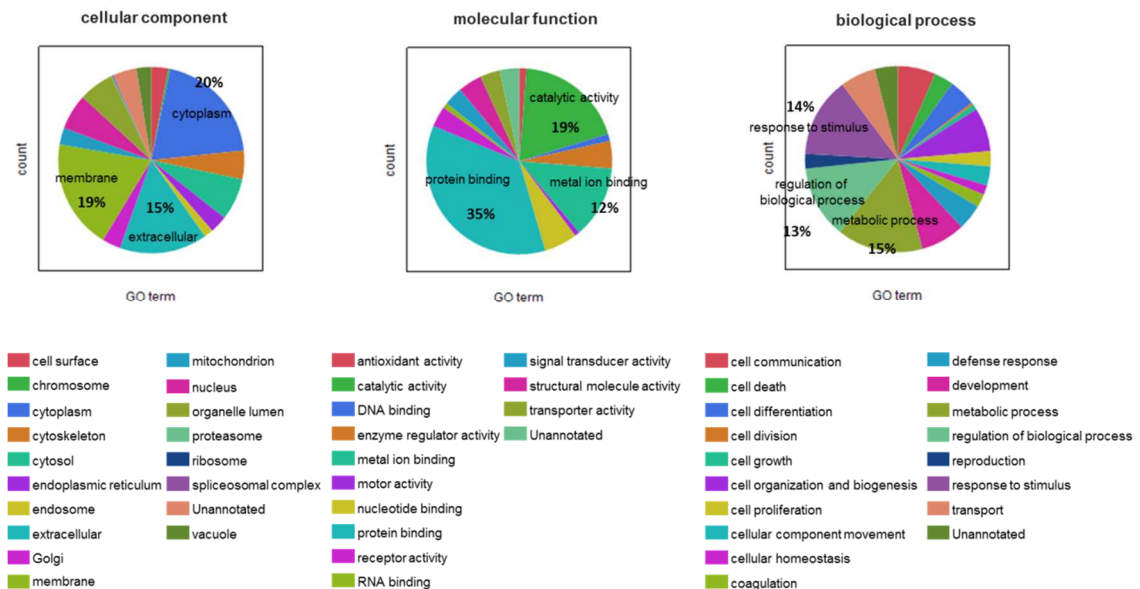


Figure IV.11 : Résultat de l'analyse des termes GO associés aux protéine groupes de notre base de données du protéome urinaire regroupés par composants cellulaires, fonctions et processus biologiques.

A partir de ces résultats, nous avons observés les différents compartiments cellulaires, fonctions moléculaires et processus biologiques mis en évidence sur la base de nos protéines urinaires (Figure IV.11). La classification des comportements cellulaires révèle une majorité de protéines cytoplasmique (20%), de protéines membranaires (19%) et de protéines sécrétées (15%). Ces observations, proches des celles précédemment publiées dans des études antérieures par Marimuthu (Marimuthu et al., 2011) ou Adachi (Adachi et al., 2006), s'expliquent par le fait que l'urine est un fluide biologique dans lequel on retrouve les déchets de l'organisme et également des protéines sécrétées par l'épithélium qui tapisse l'intérieur de la vessie. Les fonctions moléculaires, quant à elles, révèlent des protéines impliquées dans des activités catalytiques (19%) ou qui interagissent avec des ions métalliques (12%). La plupart de ces annotations, déjà observées par ailleurs, sont trop génériques pour en tirer des conclusions informatives. Les similarités trouvées avec les études d'Adachi et de Marimuthu, nous permettent de conclure qu'il n'y a pas de dérive sur ces classifications dues à nos méthodes de préparations ou de fractionnement.

D'autre part, une analyse de notre liste de protéines urinaires, à l'aide du logiciel GORILLA (<http://cbl-gorilla.cs.technion.ac.il>) a permis de calculer son enrichissement en termes GO particulier par rapport à l'ensemble du protéome humain. Il ressort de cette analyse que notre base de données du protéome urinaire est significativement enrichie en protéines impliquées dans le métabolisme lipidique et en lipoprotéines (et plus particulièrement celles du catabolisme), ainsi qu'en protéines impliquées dans le transport et la dégradation (voir Tableau IV.5). Cet enrichissement très significatif du protéome urinaire n'a

jamais été rapporté par ailleurs et il serait intéressant de le mettre en lien avec les processus physiologiques du rein et de la vessie. Néanmoins, il est important de noter que pour l'analyse d'enrichissement, seulement 41% des protéines ont été utilisés par le logiciel. Ceci est dû au fait que GORILLA utilise des banques de données dans lesquelles un terme GO n'est pas systématiquement associé à chacune des entrées. Ceci peut donc induire des biais dans l'interprétation des résultats. De plus les processus surexprimés représentent les processus les mieux documentés à l'heure actuelle, ce qui peut engendrer un biais supplémentaire dans l'analyse.

Tableau IV.5 : Principaux enrichissement en termes GO, basé sur une analyse par le logiciel GORILLA, de notre base de données du protéome urinaire comparée au protéome humain.

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0019216	regulation of lipid metabolic process	4.08E-08	0.000199	8.41 (810,20,53,11)
GO:0006869	lipid transport	0.00000013	0.000316	7.71 (810,21,55,11)
GO:0034375	high-density lipoprotein particle remodeling	0.000000137	0.000223	12.89 (810,8,55,7)
GO:0045833	negative regulation of lipid metabolic process	0.000000168	0.000204	15.28 (810,6,53,6)
GO:0033700	phospholipid efflux	0.000000168	0.000163	15.28 (810,6,53,6)
GO:0044765	single-organism transport	0.000000329	0.000267	2.83 (810,138,58,28)
GO:0006810	transport	0.000000395	0.000275	2.58 (810,168,58,31)
GO:0051234	establishment of localization	0.00000055	0.000335	2.55 (810,170,58,31)
GO:0034377	plasma lipoprotein particle assembly	0.000000752	0.000407	11.45 (810,9,55,7)
GO:0065005	protein-lipid complex assembly	0.000000752	0.000366	11.45 (810,9,55,7)
GO:0033344	cholesterol efflux	0.000000752	0.000333	11.45 (810,9,55,7)
GO:0015711	organic anion transport	0.000000824	0.000335	8.60 (810,16,53,9)
GO:0015748	organophosphate ester transport	0.00000133	0.000497	13.10 (810,7,53,6)
GO:0015914	phospholipid transport	0.00000133	0.000462	13.10 (810,7,53,6)
GO:0043691	reverse cholesterol transport	0.00000212	0.00069	12.62 (810,7,55,6)
GO:0034367	macromolecular complex remodeling	0.00000272	0.000829	10.31 (810,10,55,7)
GO:0034368	protein-lipid complex remodeling	0.00000272	0.00078	10.31 (810,10,55,7)
GO:0034369	plasma lipoprotein particle remodeling	0.00000272	0.000737	10.31 (810,10,55,7)
GO:0006820	anion transport	0.00000683	0.00175	7.24 (810,19,53,9)

2. Comparaison avec l'état de l'art

Après les efforts conséquents qui ont été consentis, en particulier au sein du consortium international HUPO (Human Proteome Organization), pour caractériser le plasma, l'urine a suscité un grand engouement dans les laboratoires de protéomiques de par sa disponibilité et son potentiel pour la découverte de biomarqueurs. Plusieurs études ont ainsi été entreprises et publiées ces dernières années. Nous avons comparé notre étude aux deux études les plus citées et pertinentes en termes de couverture du protéome: celle d'Adachi (Adachi et coll. 2006) qui représente la première étude ayant permis d'identifier plus de 1500 protéines urinaires, et celle de Kentsis (Kentsis et coll. 2009) qui est celle ayant rapporté la plus grande couverture du protéome urinaire à ce jour (2362 protéines non redondantes). Les études précédentes, basées essentiellement sur le fractionnement des protéines par Gel-2D, n'avaient jamais atteints de tels niveaux de profondeur d'analyse. Pour réaliser la comparaison des listes de protéines identifiées dans ces deux études et la nôtre, nous avons utilisé le logiciel ProteinCenter commercialisé par la société Thermo Fisher Scientific. Ce logiciel permet de comparer rapidement et facilement différentes listes de protéines provenant de différentes banques de données comme NCBI (National Center for Biotechnology Information), UniProtKB/Swiss-Prot ou encore International Protein Index (IPI). Cependant, comme beaucoup de logiciels commerciaux, l'algorithme qui permet de comparer les listes de protéines n'est pas décrit avec suffisamment

de détails et nous verrons par la suite que le résultat peut dépendre du niveau auquel on se place pour cette comparaison (protéique ou peptidique).

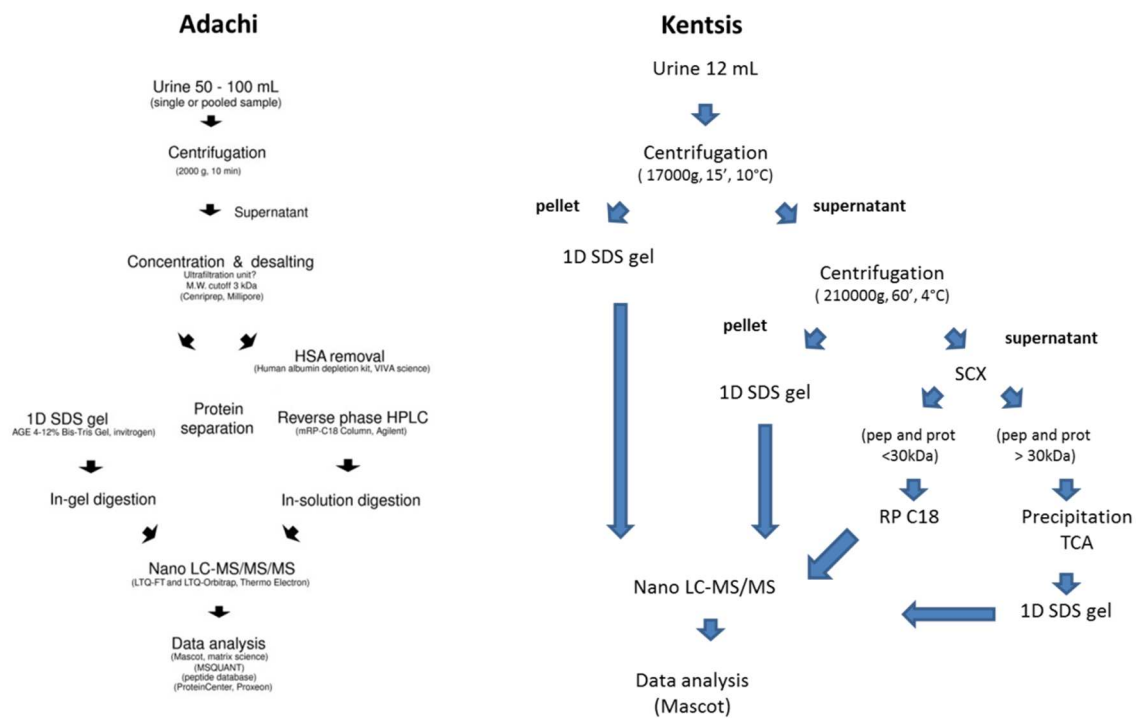


Figure IV.12 : Stratégies de fractionnement et d'analyse suivies dans les publications de Adachi (Adachi et al., 2006) et de Kentsis (Kentsis et al., 2009)

La stratégie de préparation et de fractionnement de l'urine utilisée par Adachi se rapproche d'une des technologies dont nous nous sommes servis (voir Figure IV.12) : les protéines ont été extraites par ultrafiltration en utilisant des dispositifs similaires aux vivaspins utilisés dans notre étude. Les analyses LC-MS/MS ont été réalisées sur un système hybride LTQ-Orbitrap et c'est le logiciel Mascot qui a été utilisé pour l'identification dans la banque de données IPI (International Protein Index). Cependant, dans leur protocole, un kit de déplétion de l'albumine a été utilisé, après l'étape d'extraction des protéines. Cette étape de déplétion a été contestée dans le cadre de l'analyse du sérum, car d'autres protéines potentiellement intéressantes peuvent être éliminées en raison de leur interaction avec les protéines que l'on souhaite exclure (Thulasiraman et al., 2005). Enfin, les protéines ont été fractionnées par Gel 1D et par chromatographie liquide en phase inverse suivie d'une digestion des protéines sur gel ou en solution. Pour l'analyse LC-MS/MS, l'auteur s'est appuyé sur la MS3 pour valider des protéines identifiées avec un seul peptide alors que dans notre cas ces protéines identifiées par un peptide unique n'ont été validées que si le score Mascot du peptide était supérieur à une valeur seuil de 60, (cette valeur ayant été précédemment définie lors d'une étude pilote au laboratoire (Ferro et al., 2010).

L'étude de Kentsis était, quant à elle, basée sur une extraction des protéines par précipitation TCA suivi d'un fractionnement des protéines sur Gel 1D ou sur une colonne échangeuse de cations. De plus, Kentsis s'est employé à séparer les exosomes urinaires par une centrifugation à très haute vitesse (210 000g) en vue de les analyser plus spécifiquement. Les exosomes sont de petites vésicules membranaires pouvant être sécrétés par la plupart des cellules, et ils pourraient jouer des rôles de « messagers intercellulaires », transférant divers signaux entre les cellules (Théry, Ostrowski, & Segura, 2009). Les protéines ont ensuite été séparées sur un gel SDS-PAGE ou sur colonne C18. Après digestion des protéines in gel ou en solution, les peptides résultants ont été analysés par LC-MS, les données ont été soumises au moteur de recherche Mascot pour effectuer les recherches dans la banque de données IPI.

Une première comparaison basée sur les classifications des termes de la Gene Ontology a été réalisée avec le logiciel Protein Center en fonction des compartiments cellulaires, des processus biologiques ou des fonctions moléculaires des protéines issus des listes d'Adachi, de Kentsis et de notre base de données.

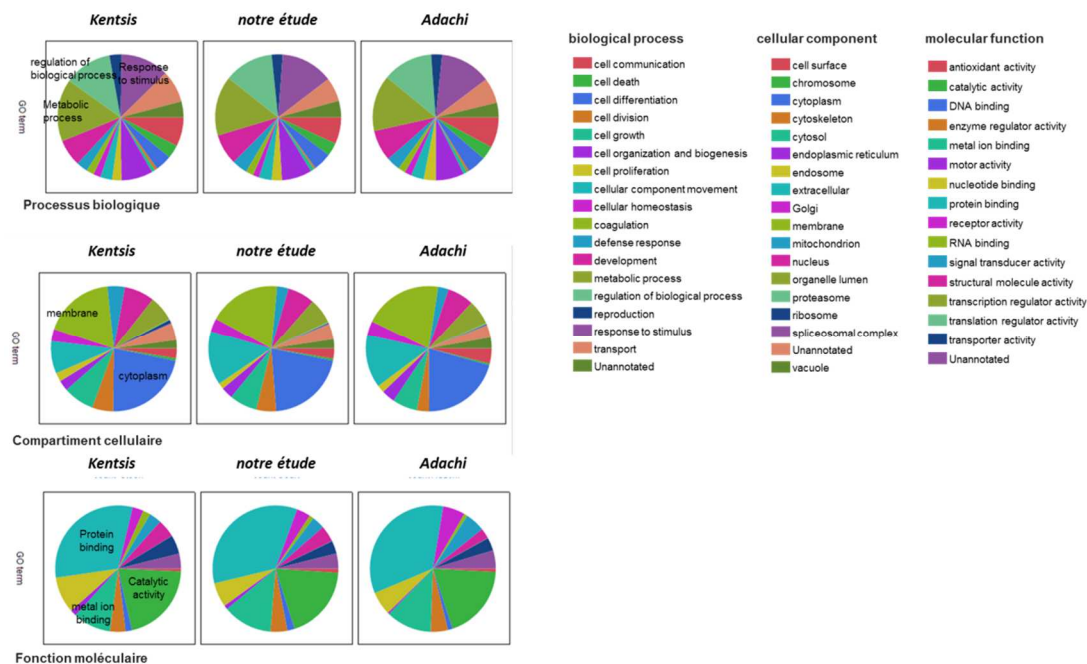


Figure IV.13 : Résultat de l'analyse des termes GO associés aux protéines identifiées dans notre base de données du protéome urinaire et dans les études d'Adachi et Kentsis, regroupées par composants cellulaires, fonctions et processus biologiques. Les répartitions sont très similaires quelles que soient les listes.

Sur cette figure, on peut observer pour les trois études des répartitions très similaires des termes de Gene Ontology que ce soit sur les compartiments cellulaires, les processus biologiques ou les fonctions moléculaires. Il n'existe donc apparemment pas de biais significatif en termes de couverture du protéome

urinaire entre les trois études, ce qui est surprenant, notamment parce que Kentsis et al ont enrichi les protéines d'origine exosomales.

Lorsque l'on observe la cohérence des trois listes en termes de protéines, (Figure IV.14) on est surpris de constater un faible recouvrement entre les études prises deux à deux (~33%) et un très faible recouvrement entre les trois études (11%).

Sur la base de la comparaison des listes de protéines par le logiciel Protein Center, ces trois études semblent présenter des différences très significatives qui n'avaient pas été mises en évidence lors de l'analyse des catégories fonctionnelles ou des compartiments cellulaires (Figure IV.14). Pour autant, en comparant les trois méthodes de préparation que nous avons employées pour caractériser le protéome urinaire, le recouvrement que nous avons obtenu était nettement supérieur : ~80% de recouvrement par paires et 52% entre les trois préparations. Ces différences sont donc difficilement imputables aux méthodes de préparations des échantillons urinaires. Le faible recouvrement doit donc s'expliquer par d'autres raisons : Tout d'abord, il est difficile de donner du sens à une comparaison des listes de protéines quand les identifications initiales proviennent de banques de données différentes (NCBI, Uniprot, IPI). Dans notre cas les identifications de protéines ont été effectuées en utilisant la banque SwissProt TrEMBL, réputée pour sa forte valeur ajoutée grâce à la qualité du travail de tri des entrées de la banque, alors que les protéines identifiées par Adachi ou Kentsis ont été identifiées en utilisant la banque IPI. Cette banque qui était l'une des plus utilisées dans les approches de protéomique est depuis fermée et a été remplacée par UniProtKB (Griss et al., 2011). La correspondance entre les identifiants dans ces différentes bases n'est pas toujours établie de façon non ambiguë. Cela est encore aggravé par le fait que plusieurs protéines qui partagent le même ensemble de peptides identifiés sont regroupées sous un *protéine groupe* dont le choix de l'identifiant est arbitraire. C'est pourquoi les taux d'intersections entre les différents sets de données sont à considérer d'une manière purement qualitative. Pour éviter ce genre de biais et avoir une estimation plus fiable des similarités et différences entre les trois études du protéome urinaire, nous avons également comparé les listes de peptides identifiés.

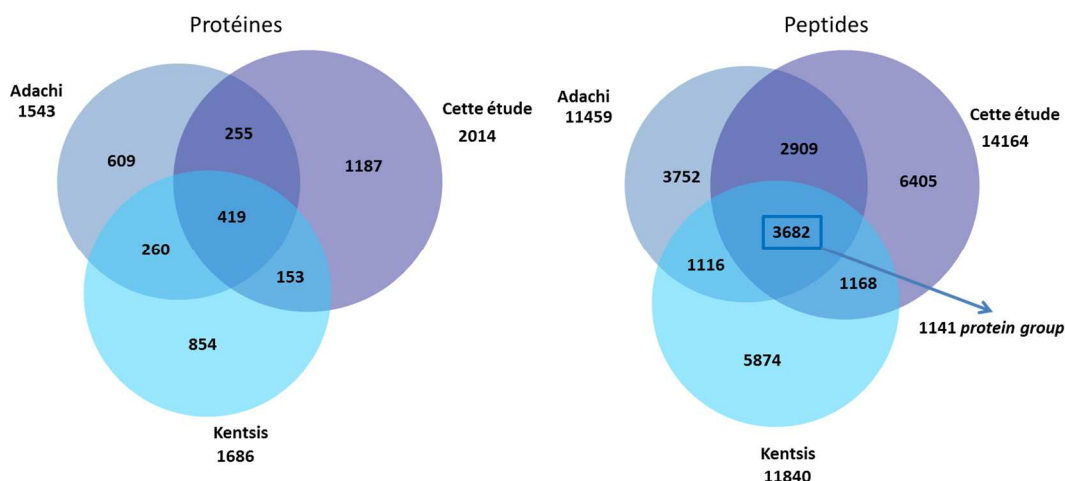


Figure IV.14 : diagramme de Venn traduisant le recouvrement en termes de protéines (panneau de gauche) et de peptides (panneau de droite) en comparant notre liste et celles d'Adachi et de Kentsis

Le recouvrement est assez similaire à celui obtenu sur la base de la comparaison des protéines. En effet, on obtient 16% de recouvrement entre les trois études en comparant les séquences des peptides identifiés, au lieu de 11% en considérant uniquement les protéines. Néanmoins, le recouvrement entre l'étude d'Adachi et la nôtre est légèrement supérieur au niveau peptidique (34% au lieu de 23% sur les protéines). Si maintenant on observe dans notre base de données les peptides communs entre les trois études (3682), ils correspondent à 1141 *protéine groups*, dont 662 sont identifiées par au moins 1 peptide protéo-spécifique. Par conséquent, le recouvrement obtenu par le logiciel Protein Center, ne reflète peut être pas le véritable recouvrement qu'il y a entre les trois études qui est sans doute plus élevé que 11%. Cependant il reste difficile d'estimer ce recouvrement qui est sans doute compris entre 17% (obtenu sur la base de la comparaison peptidiques) et 50% (sur la base de la couverture de nos *protéines groups* avec les peptides de l'intersection des trois études).

Par ailleurs, le recouvrement limité que l'on observe peut s'expliquer par les populations qui ont été analysées pour la caractérisation des différents protéomes urinaires. Dans notre étude, nous avons une forte proportion de patients atteints d'un cancer de la vessie (68) dont les âges variant de 43 à 90 ans avec une moyenne de 70 ans. L'âge moyen de nos patients contrôles était de 50 ans. Les autres études avaient été réalisées à partir d'urine de jeunes patients (12) souffrant de douleurs abdominales aiguës dans le cas de Kentsis et de patients sains (9) de moyennes d'âge 41.5 ans dans le cas d'Adachi. Ces différences de populations (nombre de patients, âge, patients sains et malades) peuvent donc aussi expliquer en partie les différences observées sur les listes de protéines identifiées.

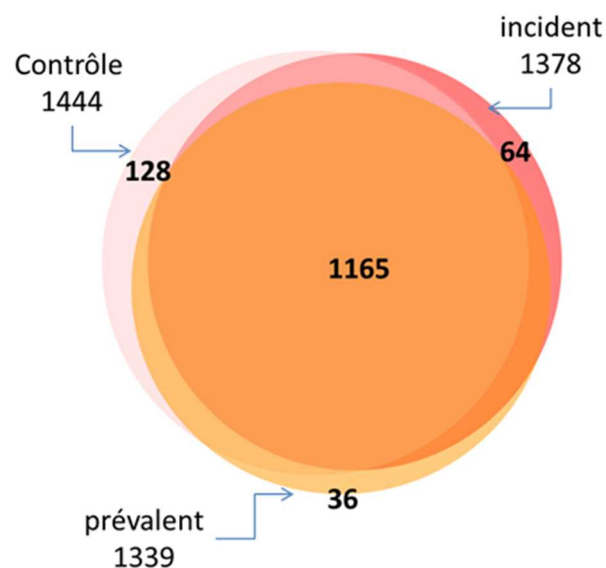


Figure IV.15 : diagramme de Venn traduisant le recouvrement en termes de protéines identifiées obtenus par notre algorithme (section B.3) entre les listes issues des différentes catégories de patients de la cohorte de découverte de candidats biomarqueurs.

Dans notre base de données, nous avons des patients sains et des patients atteints de cancer de la vessie classés suivant deux catégories : incident (dans le cas où la tumeur est détectée pour la première fois) et prévalent (patients avec une récurrence du cancer de la vessie). Nous avons donc comparé les listes de protéines issues de ces trois populations pour évaluer s'il y avait des différences significatives entre les protéomes identifiées. L'ensemble des échantillons avaient été préparés selon le même protocole, standardisé consistant en une précipitation TCA suivi d'une digestion des protéines en solution. La comparaison des listes de protéines à partir de l'algorithme décrit précédemment (voir section B.3), permet de visualiser un taux de recouvrement de 72% (voir Figure IV.15). En conclusion, en utilisant notre algorithme de comparaison, il semble y avoir peu de différences entre les listes de protéines identifiées à partir d'urines de patients sains ou malade, sachant que les classes d'âges dans les 3 sous-cohortes sont équivalentes (70 ans en moyenne).

Une dernière raison pourrait expliquer les différences entre notre liste de protéine et celles d'Adachi et de Kentsis : les traitements préanalytiques. Nous avons pu observer dans le chapitre précédent que les conditions de conservation de l'urine avant la première centrifugation à basse vitesse et la congélation à -80°C pour le stockage des échantillons pouvaient avoir une influence sur le nombre de peptides identifiés. Dans l'étude d'Adachi et de Kentsis, les échantillons d'urines ont été conservés à froid avant la première centrifugation à basse vitesse. Dans le chapitre III, nous avons pu évaluer que le fait de conserver l'urine dans ces conditions avait pour conséquence la perte de signaux peptidiques probablement dû à la précipitation de certaines protéines qui sont ensuite perdues lors de la centrifugation à basse vitesse.

En conclusion, différents facteurs pourraient expliquer que les protéomes urinaires soient différents les uns des autres : (1) les banques de séquences protéiques qui ont servi à l'identification des protéines (IPI vs SwissProt TrEMBL) et les algorithmes de regroupement des protéines en protéine groupe, (2) les populations de patients (pathologies, classe d'âge) et (3) les conditions de collectes (première urine vs aléatoire, condition de conservation). Néanmoins pour mettre ces résultats en perspective, nous rappelons que la première compilation de quatre études indépendantes du plasma humain avait révélé une concordance de seulement 4% entre les listes de protéines identifiées (N. L. Anderson et al., 2004).

3. Intérêt de l'urine pour la découverte de biomarqueurs de pathologies

La caractérisation du protéome urinaire présentée tout au long de ce chapitre constitue la première phase de l'approche AMT qui va ensuite nous permettre de quantifier les protéines extraites d'échantillons d'urine de patients sains et atteints d'un cancer de la vessie afin d'identifier de nouveaux biomarqueurs de cette pathologie. Pour évaluer le potentiel du protéome urinaire pour la recherche de biomarqueurs, nous avons cherché à vérifier si des biomarqueurs de pathologie urinaires déjà connus étaient présents dans notre base de données contenant 2014 protéines urinaires. Pour ce faire, nous avons extraits de la littérature, différents marqueurs du cancer de la vessie, et nous avons recherché les peptides correspondant à ces protéines dans notre base de données (voir Figure IV.16). Parmi ces biomarqueurs on trouve le catechol-o-methyltransférase, la calreticuline et la Gamma-synucléine, toutes trois identifiées comme candidats biomarqueurs du cancer de la vessie dans une étude de protéomique quantitative par Gel 2DE à partir d'urine de patients contrôle et malades (Iwaki et al., 2004). On trouve également d'autres marqueurs tels que l'Adipocyte type Fatty Acid-binding Protein (Celis et al., 1996), l'Ezrin (Palou et al., 2009) et enfin la Filamine A (Yue, Huhn, & Shen, 2013). De plus, nous avons trouvé dans notre base de données urinaire des marqueurs d'autres pathologies urinaires comme certaines protéines associées à l'insuffisance rénale aiguë : la lipocaline neutrophile associée à la gélatinase (NGAL_HUMAN) (Noto, Cibecchini, Fanos, & Mussap, 2013) ou une protéine liant les acides gras (FaBP) (Belcher et al., 2013). La présence de ces protéines dans notre base de données conforte l'idée que l'urine est un fluide attractif pour la détection de biomarqueurs de pathologies, et non seulement dans le contexte du cancer de la vessie. Notre base de données a donc le potentiel d'être mise à contribution dans le cadre d'autres études du protéome urinaire dédiées à la caractérisation d'autres pathologies rénales ou uro-génitales.

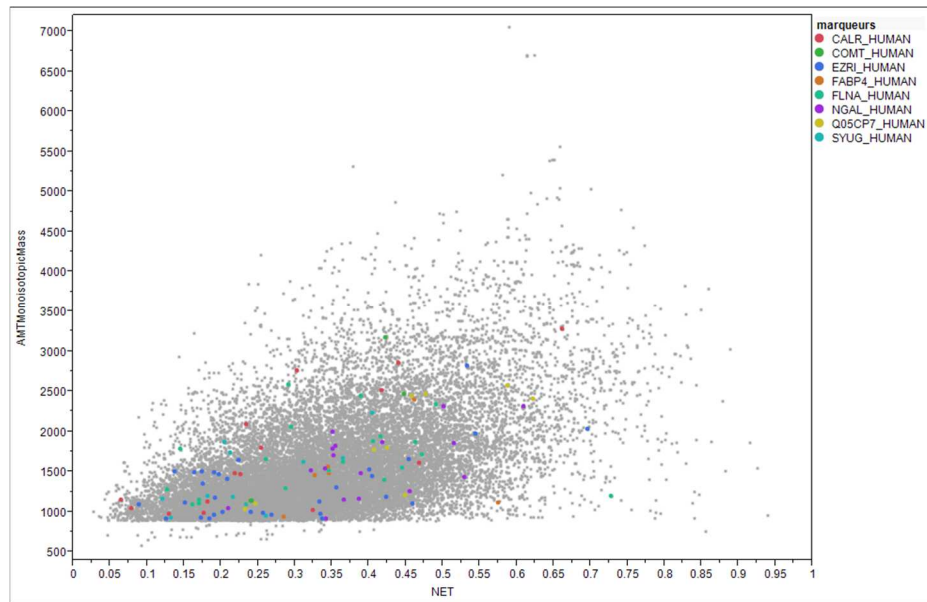


Figure IV.16 : carte du protéome urinaire obtenu dans le cadre de notre étude mettant en évidence certains candidats biomarqueurs de pathologies rénales ou uro-génitales rapportés dans la littérature.

D. Conclusion

Nous avons caractérisé l'urine à partir de différentes méthodes de préparation et de fractionnement, pour aboutir à une base de données contenant 2014 protéines non redondantes. Nous avons ensuite pu comparer les méthodes de fractionnement ou de déplétion entre elles pour évaluer la meilleure stratégie d'exploration du protéome urinaire. Les résultats que nous avons obtenus (croissance du nombre de protéines identifiées en fonction du nombre d'analyse, facilité de préparation des échantillons, automatisation possible de la méthode de fractionnement), ont démontré que la méthode de fractionnement la plus favorable était l'isoélectrofocalisation hors gel. Cependant, nous avons également pu constater, sur la base des résultats des mesures quantitatives que nous aborderons au chapitre suivant, que la méthode Equalizer permettant d'accéder à des protéines minoritaires de l'urine. Ainsi, lorsque l'on en a la possibilité, il est tout de même préférable de combiner plusieurs stratégies de fractionnement ou de déplétion pour bénéficier des avantages de chacune de ces méthodes. L'exploration préliminaire des protéines de la base de données nous laissent entrevoir le potentiel de cette ressource pour la suite de notre travail.

V. Chapitre V : Découverte de candidats biomarqueurs du cancer de la vessie dans l'urine

Dans les chapitres précédents, nous avons défini précisément les conditions optimales de collecte, mis en place un protocole standardisé du traitement des urines, et enfin caractérisé en profondeur le protéome urinaire par diverses stratégies de fractionnement ou de déplétion. Ces travaux avaient pour but ultime de permettre l'analyse semi-quantitative d'échantillons urinaires de dizaines de patients afin de découvrir de nouveaux candidats biomarqueurs du cancer de la vessie dans les urines. A présent, nous abordons l'analyse protéomique comparative réalisée pour aboutir à une liste finale de candidats biomarqueurs : de la constitution de la cohorte de patients, à l'analyse LC-MS proprement dites, et jusqu'à l'exploration statistique des données obtenues. Finalement, nous mettrons cette liste en regard de celles issues d'efforts analogues qui ont été publiées très récemment et nous évaluerons de manière critique certaines des protéines de la liste.

A. Constitution d'une cohorte de patients

Pour obtenir une cohorte de patients qui satisfasse les besoins d'études protéomiques quantitatives couvrant les phases de découverte et d'évaluation initiale de candidats biomarqueurs, plus d'un millier d'échantillons urinaires de patients convoqués à un examen médical pour une suspicion de cancer de la vessie ont été collectés sur une période totale de quatre années. Dans le Chapitre III, nous avons décrit les conditions de collecte et de traitement des urines que nous avons mises en place pour minimiser les variations artefactuelles d'abondances des protéines. Pour améliorer encore le contrôle des échantillons issus de la cohorte, nous avons établis, en collaboration avec les cliniciens, des recommandations qui ont été adressées aux patients quelques jours avant la consultation au cours de laquelle la collecte devait avoir lieu. Ces recommandations consistaient en éviter l'excès de fluide et éviter les boissons alcoolisées ou le sport intensif la veille de la cystoscopie. Le jour de la cystoscopie, il était demandé au patient d'uriner au lever, puis d'attendre d'être arrivé au service d'urologie pour collecter la seconde miction avant l'examen médical. Ces recommandations avaient pour but d'éviter l'obtention d'échantillons urinaires trop concentrés (première miction du matin riches en albumine) ou trop dilués (excès de fluide).

1. Sélection des patients

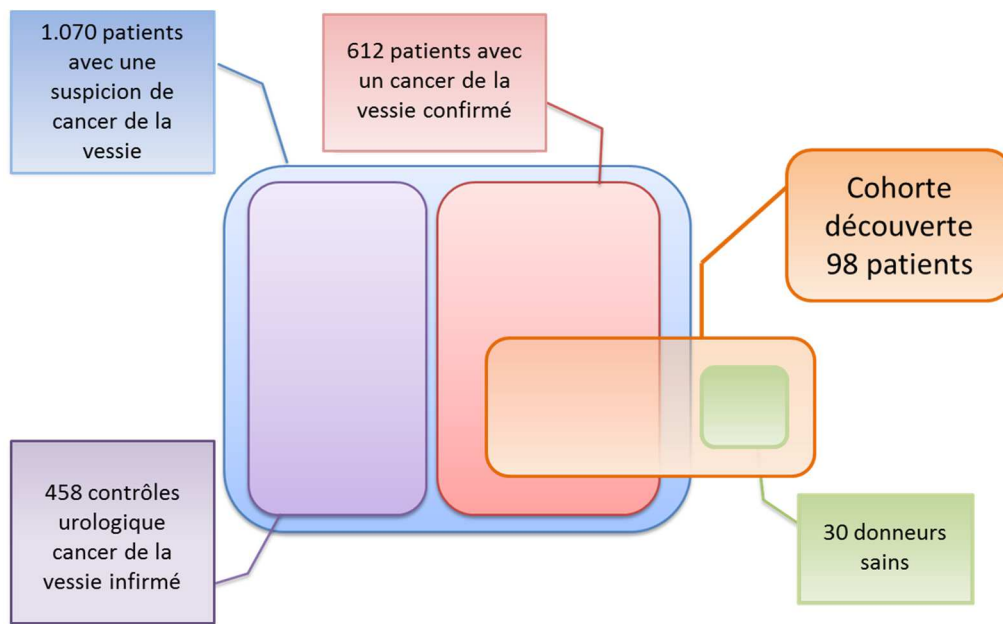


Figure V.1 : Cohorte globale de plus de 1000 patient ayant consultés pour une suspicion de cancer de la vessie, et sous-cohorte utilisée pour la phase de découverte, comprenant 68 patients avec confirmation de cancer de la vessie et 30 donneurs sains.

Lors de la phase de découverte de biomarqueurs, nous avons comparé les profils des protéines urinaires de patients atteints d'un cancer de la vessie à ceux d'une sous-population de patients contrôles (voir figure V.1). Dans notre étude, la catégorie contrôle était constituée de patients hospitalisés dans un autre service que celui d'urologie, par exemple pour une fracture. La cohorte de malades comprenait des patients du service d'urologie venus consulter pour une suspicion d'un cancer de la vessie. Lors de la consultation, une cystoscopie était pratiquée, et si la présence d'une tumeur vésicale était avérée après examen clinique, les patients étaient inclus dans la cohorte de découverte. Parmi ces patients ayant une pathologie avérée, nous avons distingué deux classes: les patients qui présentaient une tumeur de vessie pour la première fois ou patients incidents ; et les patients qui avaient déjà eu un cancer de la vessie traité par résection, et qui déclaraient pour la seconde fois un cancer de la vessie ou patients prévalents. Cette dernière catégorie a été incluse pour permettre de mettre en évidence des biomarqueurs de récurrence.

Au total, plus d'une centaine de patients ont été sélectionnés pour la phase de découverte dans la sous-population de 612 patients pour lesquels la présence d'une tumeur avait été confirmée, et à laquelle on a associé les urines de 30 donneurs sains. Cette sélection a été réalisée en écartant les patients pour lesquels l'examen clinique avait détecté une pathologie cardiaque ou rénale. Les urines ont ensuite toutes été préparées en vue de l'analyse protéomique « label-free ». Parmi ces échantillons, 25 ont été rejetés dès la préparation car leur concentration protéique était trop élevée (supérieure à 3mg/mL). Cette concentration

pouvait en effet atteindre jusqu'à 10 mg/mL chez certains patients. Au final, les extraits protéiques urinaires de 98 patients ont été utilisés pour les analyses quantitatives.

2. Classification des patients cancers : facteurs de risque

Les cancers de la vessie peuvent être classés en fonction de leur facteur de risque de progression tumorale ou de récurrence selon la classification de Babjuk (Babjuk et al., 2008). Le facteur de risque est déterminé à partir du nombre de tumeurs, de leur taille, de leur grade, du stade d'infiltration et du taux de récurrence. Cette classification permet d'associer aux cancers un score de récurrence ou de progression. En fonction de ce score, on calcule une probabilité (<http://www.eortc.be/tools/bladdercalculator/>) qui permet de classer les cancers en trois catégories : (1) les cancers de « faible risque », (2) les cancers de « risque intermédiaire » et enfin (3) les cancers de « risque élevé ». Lors de la phase de design de l'étude (voir Figure V.2.A), nous avons anticipé, sur la base de l'historique des consultations dans les hôpitaux participants, que la cohorte de découverte pourrait être constituée de 30 patients contrôles, 30 patients incidents (10 par catégorie de risque) et 30 patients prévalents (10 par catégorie de risque). De plus, les patients contrôles devaient être appariés selon leur âge et leur sexe aux patients incidents et prévalents. On peut voir sur la figure V.2.B, que notre design initial n'a pas pu être strictement respecté. Ceci témoigne de la difficulté de constituer une telle cohorte. Ainsi, nous avons obtenu au final une plus grande proportion de patients incidents à haut risque de progression (54% des patients incidents au lieu d'un tiers prévu initialement) et nous avons eu plus de mal à recruter des patients incidents à risque intermédiaire (7 au lieu de 10). Ceci s'explique en partie par le fait que les signes cliniques de cancer de la vessie ne permettent pas une prise en charge rapide des patients. Lorsque le patient se présente au service d'urologie, le cancer a souvent déjà progressé et le facteur de risque augmenté.

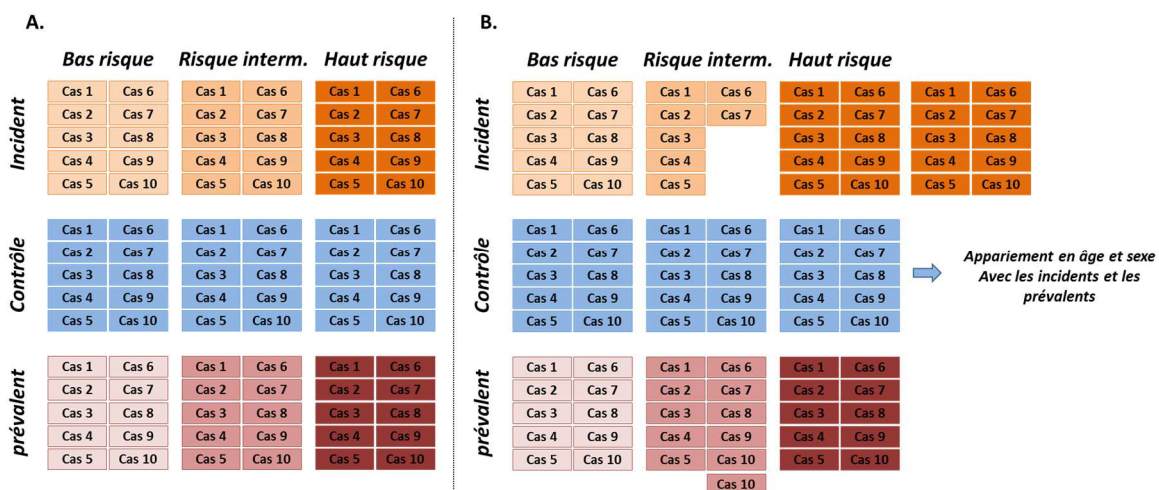


Figure V.2 : Design initialement anticipé pour la cohorte de découverte (A) et composition de la cohorte réellement analysée (B).

3. Les cohortes multicentriques

Quand cela est possible, il est souhaitable de constituer une cohorte multicentrique pour que l'échantillon soit représentatif de la pathologie étudiée et pas spécifique d'un site. La collecte des échantillons urinaires s'effectue alors simultanément dans plusieurs hôpitaux. Cependant, le fait de collecter une cohorte dans plusieurs pays peut engendrer des difficultés supplémentaires. Ainsi, dans notre cas, en raison de lois d'éthiques propres à chaque pays, il nous a fallu gérer des échantillons qui n'étaient pas anonymisés de la même façon suivant le pays où les urines ont été collectées.

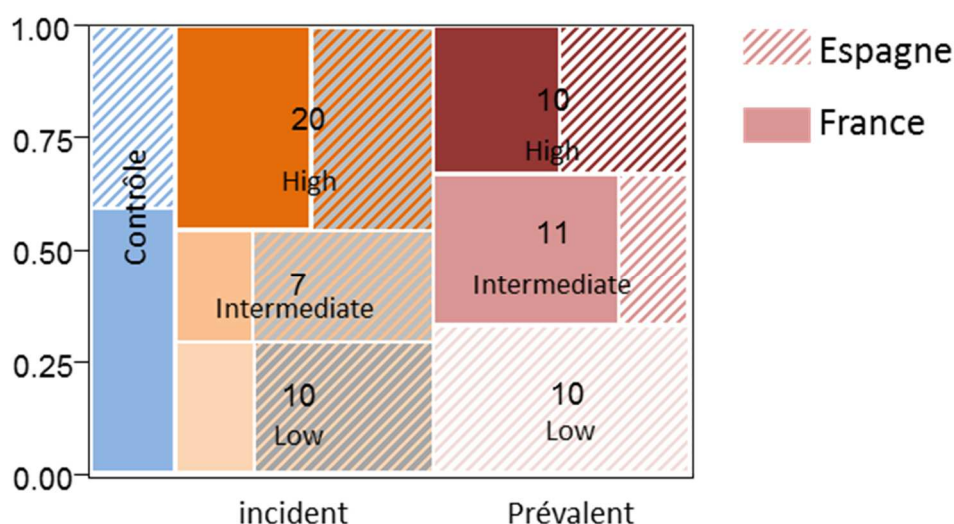


Figure V.3 : Répartition des patients de la cohorte de découverte suivant leur catégorie (contrôles, incident ou prévalent), leur facteur de risques de progression ou de récurrence et leur origine géographique.

Dans cette étude, les urines de patients ont été collectées en France, à l'hôpital Henri Mondor de Créteil (45 patients) et en Espagne à l'hôpital Del Mar de Barcelone (53 patients). La figure V.3 montre la répartition des patients inclus dans notre analyse suivant leurs catégories en fonction du facteur de risque de progression tumorale ou de récurrence et leur origine géographique. Les parties unies représentent la proportion des échantillons provenant de l'hôpital Mondor et les parties hachurées celle des échantillons en provenance de l'hôpital Del Mar. Les répartitions sont relativement équilibrées entre les deux pays, mis à part pour les patients prévalents à bas risques pour lesquels nous n'avons obtenu que des urines de patients espagnols.

B. Analyse protéomique quantitative

1. Identification dans la base AMT : réalignement des temps de rétention

Lors de la génération de notre base de données du protéome urinaire, plus d'un millier d'analyses LC-MS/MS avaient été réalisées sur les échantillons urinaires provenant de différentes catégories de patients et de divers traitements de l'urine comme nous avons pu voir dans le chapitre précédent. Cette première

phase, qui nous a permis de caractériser en profondeur le protéome urinaire, s'est déroulée sur plusieurs années. Or, sur une période aussi longue, les conditions chromatographiques peuvent dériver, avec une répercussion sur les temps de rétention des peptides. En effet, les fluctuations de température, de pression ou de débit de la phase mobile, l'évaporation différentielle des solvants, ou encore le vieillissement de la colonne, peuvent être à l'origine de variations de temps de rétention difficilement prédictibles.

Lors de la première phase de l'approche AMT, qui consiste en la constitution de la base de données, les temps de rétention de chacun des peptides de la base sont calculés en moyennant l'ensemble des temps de rétention de toutes les observations de ce peptide dans les différentes acquisitions où il a été identifié. Ainsi en utilisant des analyses LC-MS/MS qui s'étendent sur une plage de temps étendue, on prend le risque d'avoir des coefficients de variation des temps de rétention importants et des temps moyens éloignés des temps de rétention observés pour ces mêmes peptides lors de l'analyse quantitative.

Dans la seconde phase de l'approche AMT, afin d'identifier les peptides détectés dans les analyses LC-MS, on utilise deux logiciels décrit dans le chapitre II. L'un de ces logiciels, VIPER, permet de réaligner les temps de rétention chromatographiques des peptides détectés lors de l'analyse quantitative avec ceux des peptides de la base de données AMT. Entre le moment où l'on a réalisé les analyses LC-MS/MS et l'analyse quantitative, on devrait idéalement se trouver dans les mêmes conditions chromatographiques. Cette condition n'est malheureusement pas toujours remplie pour les raisons que l'on vient de décrire, avec pour principale conséquence, un alignement médiocre des temps de rétention des peptides détectés avec ceux de la base de données AMT. Il est donc nécessaire de mettre en place des protocoles de traitement de données qui permettent de corriger les temps de rétention observés dans toutes les analyses LC-MS/MS, utilisées pour compiler la base AMT.

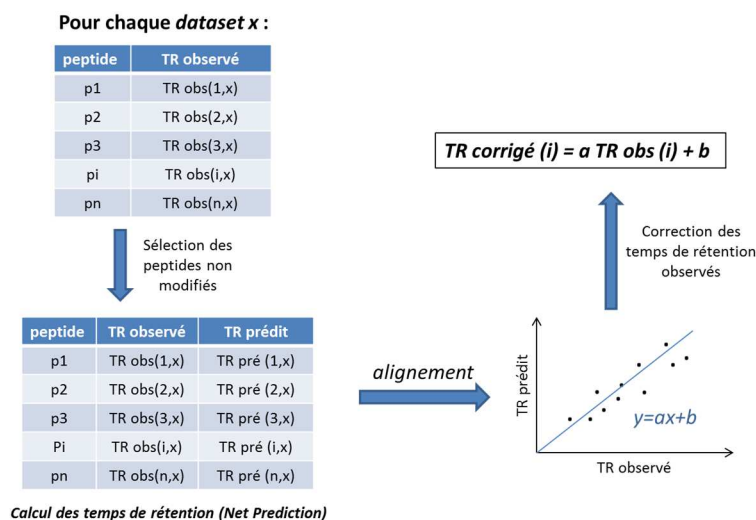


Figure V.4 : Algorithme de correction des temps de rétention des peptides dans la base de données constituée lors de la première phase de l'approche AMT

Les informaticiens du laboratoire ont implémenté le logiciel « NET Prédiction» dans hEIDI, notre logiciel de gestion des bases de données d'identification. Cette extension permet de prédire les temps de rétention des peptides sur la base d'un modèle basé sur un réseau de neurones (Petritis et al., 2006). Ceci peut notamment permettre de standardiser les temps de rétention de tous les peptides d'une analyse LC-MS/MS donnée en se servant des prédictions comme référentiel fixe. Cet algorithme, décrit dans la Figure V.4, corrige les temps de rétention des peptides de chaque analyse LC-MS/MS en se basant sur des temps de rétention prédits pour les peptides ne portant pas de modifications post-traductionnelle. Pour chaque analyse LC-MS/MS, les temps de rétention ainsi prédits sont alignés avec les temps de rétention observés. On en déduit ensuite les paramètres de régression linéaire (ordonnée à l'origine et coefficient directeur) entre les valeurs prédites et les valeurs observées qui sont ensuite utilisés pour recalculer les temps de rétention de tous les peptides de l'analyse LC-MS/MS. On fait de même pour chaque analyse LC-MS/MS et, au final, un temps de rétention est associé à chaque peptide correspondant à la médiane des temps de rétention standardisés observés pour ce peptide dans toutes les analyses LC-MS/MS où il a été identifié. Les temps de rétention standardisés des peptides de la base de données sont ainsi indépendants des variations des conditions chromatographiques au cours des acquisitions LC-MS/MS.

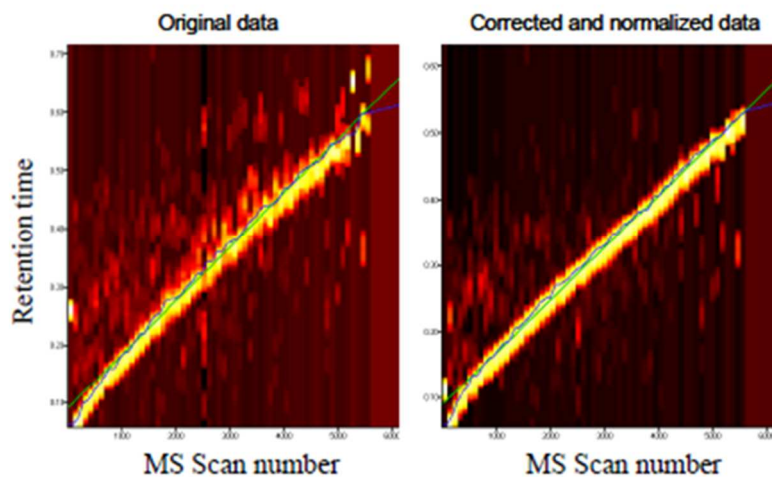


Figure V.5 : Alignement des temps de rétention des peptides détectés dans l'analyse quantitative avec ceux de la base de données AMT avant et après correction des temps de rétention des peptides de la base de données.

Pour évaluer l'impact sur l'analyse quantitative de la correction des temps de rétention de la base AMT, nous avons comparé le nombre de peptides identifiés et quantifiés par VIPER sur une même acquisition LC-MS d'un extrait urinaire avant et après correction. La figure V.5 illustre l'alignement des peptides détectés dans une analyse LC-MS avec les peptides de la base de données avant et après correction des temps de rétention. Sur le panneau de gauche, on visualise la densité des correspondances entre les peptides détectés lors de l'analyse LC-MS et les AMT présents dans une base de données dont les temps de

rétenion n'avaient pas été corrigés. Sur la figure de droite, on peut voir la densité de correspondance entre ces mêmes peptides avec les AMT d'une base de données après traitement par l'algorithme de correction. La courbe est significativement plus dense dans la figure de droite car les temps de rétention présents dans la base de données sont plus cohérents avec les valeurs observées. Grâce à cette procédure, nous avons observé, en moyenne, une augmentation de 20 % du nombre de peptides quantifiés dans les échantillons d'urines.

2. Quantification par spectral count dans des urines poolées.

Vers la fin de l'étape de caractérisation du protéome urinaire (voir Chap. IV), lorsque nous avons testé la méthode de fractionnement des peptides par isoélectrofocalisation, les échantillons des patients de la cohorte de découverte sont arrivés au laboratoire. Afin d'inclure ces patients dans la base de données, nous avons décidé de constituer 10 pools en rassemblant les urines de patients de même catégorie. Les pools ont été constitués suivant la répartition représentée figure V.2.B. Nous avons ainsi constitué 3 pools d'urines de patients contrôles, 2 pools de patients incident à haut risque, et un pool pour chacune des autres catégories de patients. Les urines ainsi poolées ont été traitées selon le protocole standardisé, et les peptides fractionnés par isoélectrofocalisation. Chacune des 12 fractions d'un pool donné a ensuite été analysée sur le spectromètre de masse en mode LC-MS/MS.

Après avoir identifié les protéines de chacune des fractions, nous avons regroupé les protéines correspondant aux différentes fractions de chaque pool de patients grâce à l'algorithme de regroupement décrit dans le chapitre IV et implémenté dans le logiciel hEIDI. Ce logiciel permet, entre autre chose, de calculer les « spectral count » des protéines dans un contexte. Le « spectral count » d'une protéine correspond au nombre de spectres MS/MS ayant résulté en l'identification d'un peptide de cette protéine. Il a été montré que cette valeur est liée à la quantité de protéine en solution (Old et al., 2005). Les valeurs de spectraux counts ont été ajustées en fonction de la protéo-spécificités des peptides dans un protéine groupe suivant la formule :

$$SC_{i,weighted} = SC_{i,proteo} + \sum_p (\omega_{p,i} \times SC_p) \quad \omega_{p,i} = \frac{SC_{i,proteo}}{SC_{j,proteo}}$$

Pour chaque protéine groupe *i*, on calcule un « spectral count » (SC) ajusté ($SC_{i,weighted}$), basé sur le « spectral count » des peptides protéo-spécifiques ($SC_{i,proteo}$) auquel on ajoute la contribution pondérée des peptides partagés (*p*) entre différents *protéine groupes* du contexte. Cette pondération ($\omega_{p,i}$) est effectuée en divisant le SC des peptides protéo-spécifiques du protéine groupe *i* par la contribution de l'ensemble des peptides protéo-spécifiques des protéines qui partagent le peptide *p*.

Les valeurs de « spectraux count » ajustés ont été calculées pour l'ensemble des protéines groupes de chaque *contexte*, correspondant à un pool des patients. Les valeurs ont ensuite été utilisées pour mettre en évidence les différences éventuelles entre les catégories de patients. Cette première analyse quantitative que

nous avons effectuée comme un test préliminaire, est à considérer avec réserve pour différentes raisons : premièrement, la corrélation entre le « spectral count » et l'abondance des protéines dans un échantillon donné est assez grossière et dépend de la matrice analysée. Deuxièmement, la quantification effectuée sur des pools de patients ne permet pas d'estimer la dispersion des abondances protéiques mais seulement la moyenne de l'abondance des protéines urinaires au sein d'un même pool. Il est ainsi impossible de juger de la significativité statistique des différences observées entre pools car nous ne disposons que d'un nombre très limité de pools par sous-population (en général un seul).

La liste de *protéines groupes* avec leurs valeurs de « spectral count » ajustés a été analysée à l'aide du logiciel ProteinCenter pour tenter de regrouper des protéines présentant des comportements similaires à travers les différentes catégories de patients (pools). Cette analyse a permis de regrouper les comportements des protéines en 10 clusters. Parmi eux, un cluster particulier a attiré notre attention car les abondances protéiques montraient une corrélation avec risque de progression tumoral, que ce soit pour les patients prévalents ou incidents (voir Figure V.6) Les protéines de ce cluster sont regroupées dans le tableau V.1.

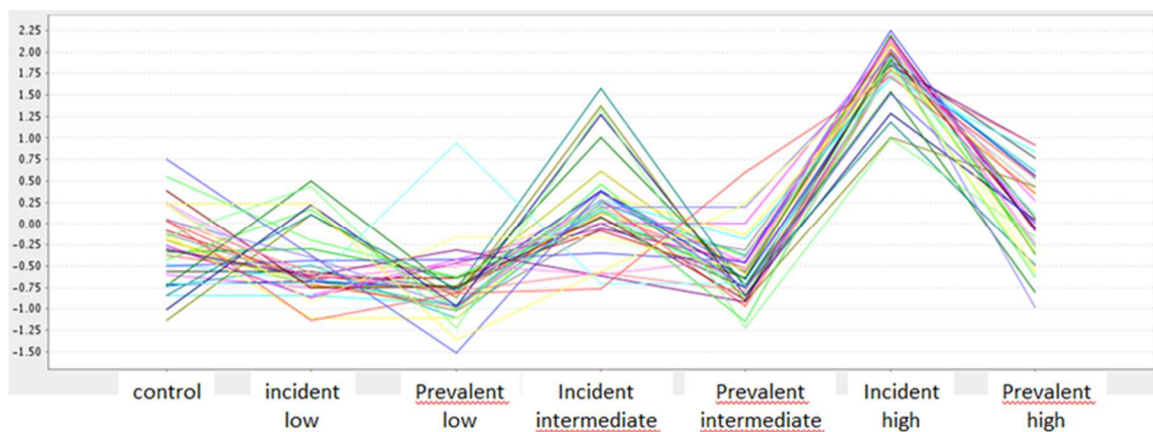


Figure V.6 : Cluster de protéines dont l'abondance montre une corrélation à l'augmentation du risque de récurrence/progression.

Tableau V.1 : Protéines représentant le cluster dont l'abondance montre une corrélation à l'augmentation du risque de récurrence/progression.

Accession	Description
RHOA_HUMAN	Transforming protein RhoA
CO5_HUMAN	Complement C5
CO8A_HUMAN	Complement component C8 alpha chain
FBLN1_HUMAN	Fibulin-1
APOA2_HUMAN	Apolipoprotein A-II*
KLKB1_HUMAN	Plasma kallikrein
PRDX2_HUMAN	Peroxiredoxin-2*
B3VL17_HUMAN	Beta globin
B4DPQ0_HUMAN	cDNA FLJ54471, highly similar to Complement C1r subcomponent (EC 3.4.21.41)
CD5L_HUMAN	CD5 antigen-like
HBD_HUMAN	Hemoglobin subunit delta*
Q86TT1_HUMAN	Full-length cDNA clone CSODD006YL02 of Neuroblastoma of Homo sapiens (human)
HPT_HUMAN	Haptoglobin
CO3_HUMAN	Complement C3*
Q5VVQ8_HUMAN	C4b-binding protein alpha chain
FIBG_HUMAN	Fibrinogen gamma chain*
HEP2_HUMAN	Heparin cofactor 2*
A2MG_HUMAN	Alpha-2-macroglobulin*
D1MGQ2_HUMAN	Hemoglobin subunit alpha*
Q9NSD0_HUMAN	Protein S
CFAH_HUMAN	Complement factor H*
ITI1H2_HUMAN	Inter-alpha-trypsin inhibitor heavy chain H2*
APOC3_HUMAN	Apolipoprotein C-III*
K2C8_HUMAN	Keratin, type II cytoskeletal 8
B3VL31_HUMAN	Beta globin
FIBB_HUMAN	Fibrinogen beta chain*
D3DR07_HUMAN	Catalase
TRFE_HUMAN	Serotransferrin*
APOA1_HUMAN	Apolipoprotein A-I*
A8K2T4_HUMAN	cDNA FLJ78207, highly similar to Human complement protein component C7 mRNA
B1N7B6_HUMAN	Cryocryoglobulin CC1 heavy chain variable region
H2B1D_HUMAN	Histone H2B type 1-D
HPTR_HUMAN	Haptoglobin-related protein*
B7Z549_HUMAN	cDNA FLJ56821, highly similar to Inter-alpha-trypsin inhibitor heavy chain H1
K2C7_HUMAN	Keratin, type II cytoskeletal 7
CRAC1_HUMAN	Cartilage acidic protein 1
A8K8Z4_HUMAN	cDNA FLJ78071, highly similar to Human MHC class III complement component C6 mRNA
C9JMH6_HUMAN	Alpha-2-antiplasmin

Sur les 38 protéines représentées sur la Figure V.6, on note que la plupart sont des protéines sériques (hémoglobine, serotransferrine, kallikrein) ou impliquées dans la cascade du complément (Complément C3, C5, factor H), traduisant sans doute les effets de l'inflammation liée à la pathologie. De plus on note que 15 d'entre elles sont des protéines situées dans des vésicules extracellulaires ou exosomes, connus pour être impliqués dans le trafic cellulaire et ainsi jouer un rôle dans la progression tumorale (Jeppesen et al., 2014). Malheureusement, l'algorithme utilisé dans le logiciel Protein Center pour normaliser les données ou pour

effectuer le regroupement des protéines en cluster n'est pas décrit, et il nous est donc impossible de répéter l'analyse indépendamment pour en vérifier les étapes, et même d'extraire le comportement d'une protéine particulière sur ce graphe. Cette première analyse, malgré ses limites, laisse entrevoir qu'il existe des différences entre les catégories de patients analysés. Nous pressentons que l'étude suivante basée sur l'extraction des signaux MS à partir des analyses d'échantillons individuels, nous permettra d'évaluer plus précisément l'ampleur et la significativité de ces différences, et d'extraire ainsi des candidats biomarqueurs plus robustes qui pourront alors être validés par d'autres méthodes quantitatives (SRM, Elisa).

3. Analyse semi-quantitative des échantillons individuels par la méthode AMT

Lors d'une analyse de protéomique quantitative, pour obtenir des informations permettant de détecter de faibles variations d'abondances de protéines ou de peptides entre deux conditions (malade ou saine), il faut s'assurer que les conditions d'analyses en spectrométrie de masses sont optimales avant de démarrer l'analyse. Puis, tout au long de la série d'échantillons, il faut s'assurer que ces conditions soient conservées. Il faut en particulier vérifier la bonne reproductibilité des temps de rétention ainsi que la gamme dynamique qui est liée au nombre de peptides détectés dans un échantillon standard. On doit également vérifier que les intensités de certains peptides connus soient conformes à des valeurs seuils préalablement fixées. A ce titre, nous avons mis en place une procédure de contrôle qualité décrite plus en détail au Chapitre II. En particulier, pour évaluer les dérives possibles des temps de rétention chromatographiques traduisant une usure de la colonne ou un incident analytique (bouchage de colonne, défaut de pompe), mais également la qualité du signal MS, un standard de cytochrome C digéré a été injecté (1µg) avant l'analyse des échantillons et ensuite à intervalles réguliers (tous les 4 échantillons). On a ainsi suivi l'intensité et le temps de rétention de plusieurs peptides spécifiques de cette protéine. Ce standard nous a permis de vérifier tout au long de l'analyse que les conditions chromatographiques et du spectromètre de masse variaient dans une gamme acceptable : +/- 3 min pour les mesures des temps de rétention et un facteur < 2.25 pour l'intensité des peptides.

Nous avons estimé au préalable que l'analyse en triplicat de toute la cohorte et des contrôles qualité associés (3 x 130 LC-MS/MS de 100 minutes), nécessiterait plus d'un mois de mesures en continu. Sur cette échelle de temps, il est inévitable d'avoir des dérives instrumentales en dépit des contrôles qualité effectués. Dans ce contexte, il nous a semblé inapproprié d'analyser successivement les réplicats d'un même échantillon. De plus, il est souhaitable d'analyser les individus dans un ordre aléatoire. Afin de capturer la variance globale de la mesure sur toute la durée de la campagne, nous avons choisi d'analyser les réplicats par blocs contenant chacun tous les échantillons dans un ordre aléatoire différent à chaque fois. Ainsi, l'échantillon du patient numéro 188 011 045 a été analysé en 48^{ième} position dans la première série de mesure, en 91^{ième} position dans la deuxième série et en 16^{ième} position dans la dernière.

C. Analyse des données pour la découverte de candidats biomarqueurs

Une fois les données acquises nous avons procédé en plusieurs étapes en vue d'aboutir à une liste de candidats. Dans une première phase de qualification des données, nous avons tout d'abord évalué la cohérence des valeurs d'abondance obtenues à travers les répliquats analytiques. Nous avons ensuite effectué un filtrage et une réduction des données avant de passer à l'analyse statistique proprement dite. Pour ce faire nous nous sommes servis des répliquats. Certaines des procédures qui vont être présentés dans cette partie ont été automatisées dans un script écrit en JMP Scripting Language.

1. Qualification des données

Le traitement des données issues de l'analyse des 98 échantillons urinaires en triplicat a permis la mesure de 432 914 valeurs d'abondances peptidiques mesurées qui s'étendent sur presque 6 ordres de grandeurs ($\log(\text{Ab})$ de 3,6 à 9,2). Au total 6162 peptides distincts ont été quantifiés dans ces analyses, correspondant à 1400 protéines groupées. Ces chiffres traduisent bien la complexité des données ainsi que la profondeur d'analyse obtenues. La figure V.7 représente, sur une échelle logarithmique, l'ensemble des valeurs d'abondances mesurées pour chacun des 98 patients de la cohorte dans les trois répliquats analytiques. Sur ce graphe, les données sont résumées sous forme de box-plot, la zone blanche correspondant au deux quartiles centraux, et les bandes des couleurs aux deux quartiles extrêmes. Les valeurs extrêmes ont également été représentées, et les patients ont été regroupés par catégorie et par risque. Ceci permet en particulier de visualiser qu'il n'existe pas dans ces données de biais notable entre les catégories de patients. On note que les trois répliquats d'un même patient, représentés côtes à côtes, sont relativement homogènes. Il est important de souligner encore une fois que les répliquats d'un même échantillon ont été réalisés à plusieurs semaines d'intervalle, et qu'au sein de chaque série les patients avaient été analysés par ordre aléatoire. On note que la normalisation des quantités injectées sur la base des concentrations protéiques déterminées par le dosage Bradford a globalement permis de compenser la grande variabilité inter-individuelle. En effet, si l'on compare ces variations à celles que nous avons obtenues dans une étude précédente en normalisant sur la base du volume d'urine, on note que les résultats sont nettement plus homogènes, et cela quelle que soit la catégorie de patient. Cependant, certains cas, comme le contrôle N° 188011226 indiqué par la flèche sur la figure V.7, présentent des abondances significativement décalées par rapport à la moyenne de la cohorte représentée par la ligne horizontale (6.078). Il est à noter que cela représente quelques cas isolés, et que ces écarts sont obtenus de façon cohérente dans les 3 répliquats analytiques de ces échantillons.

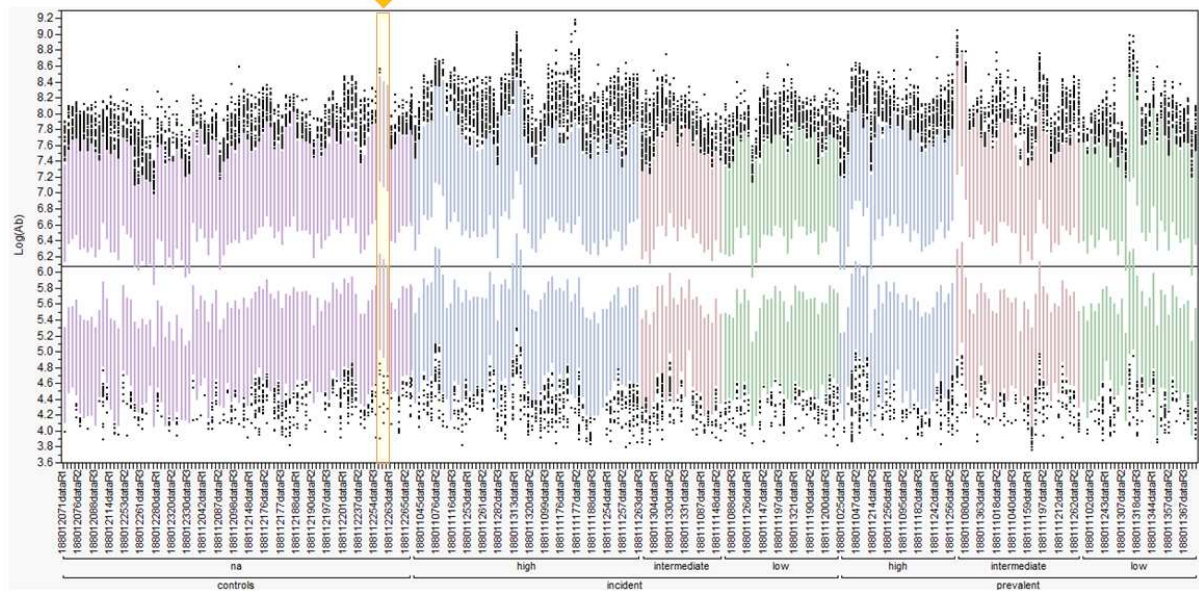


Figure V.7 : Box plot tracés à partir des valeurs d'abondances mesurées pour chacun des 98 patients de la cohorte, regroupés par catégorie et par risque, dans les trois répliquats analytiques. La zone blanche correspond au deux quartiles centraux, et les bandes des couleurs aux deux quartiles extrêmes.

Nous avons cherché à évaluer de façon plus approfondie la cohérence des résultats obtenus dans les répliquats analytiques d'un même patient. Pour ce faire, nous avons focalisé notre attention sur deux métriques : le coefficient de corrélation Pearson entre les valeurs d'abondances mesurées dans deux répliquats et le nombre de peptides quantifiés dans chaque analyse. La figure V.8 représente, pour chaque patient, l'ensemble des valeurs des coefficients de Pearson obtenus entre les paires de répliquats ainsi que le nombre de peptides quantifiés dans chaque répliquat. Sur cette figure, les individus ont été triés par moyenne croissante des coefficients de Pearson entre les abondances mesurées dans les trois répliquats. Le code couleur indique le risque de progression ou de récurrence, n/a réfère aux contrôles (pas de facteurs de risques). On peut observer que les coefficients de Pearson sur les répliquats analytiques sont en moyenne de 0.94, ce qui correspond à une forte corrélation. Cependant, on remarque aussi que chez certains patients, les valeurs d'abondances peptidiques sont moins cohérentes, avec des coefficients de Pearson inférieurs à 0.8, ceci traduisant la présence potentielle de valeurs aberrantes dans les données. Il est à noter que cela est souvent dû à un des trois répliquats, et que les deux autres analyses donnent en général des valeurs plus adéquates. Il apparaît donc qu'il n'est pas toujours souhaitable de réaliser une moyenne des valeurs d'abondances peptidiques sur les 3 répliquats pour tous les patients. Un filtre préalable semble nécessaire afin de pouvoir regrouper les valeurs d'abondances mesurées dans ces répliquats.

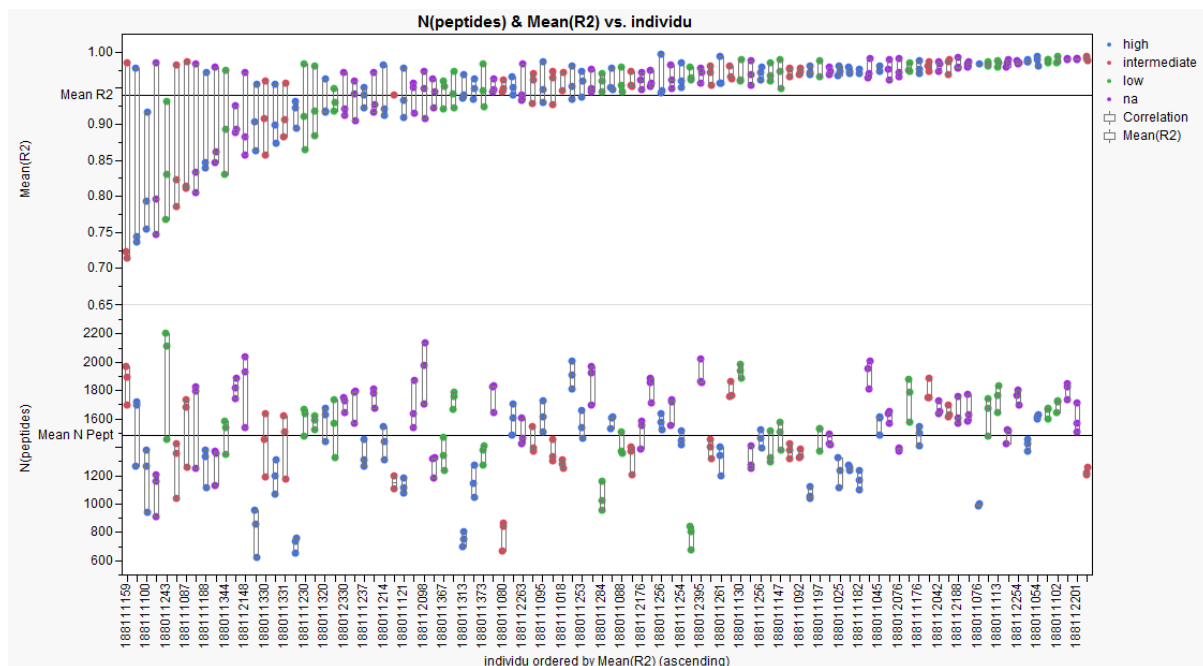


Figure V.8 : (Panneau supérieur) Box plot des coefficients moyens de Pearson entre les abondances des peptides d'une des trois analyses correspondant à un patient donné et celles des deux autres analyses du même patient, triés par ordre croissant. (Panneau inférieur) Nombre de peptides quantifiés chez un patient donné pour les trois analyses.

Lorsque l'on observe le nombre de peptides obtenus (voir Figure V.8 panneau inférieur), on note que l'on quantifie entre 622 à 2204 peptides par échantillon urinaire, avec 1500 peptides quantifiés en moyenne. Ces chiffres sont bien inférieurs à la totalité des peptides quantifiés dans tous les échantillons (plus de 6000). Ce phénomène ne s'explique pas sur la base de la concentration protéique de ces urines et semble plus lié à la variabilité inter-individuelle des échantillons. Dans la plupart des cas on obtient un nombre similaire de peptides dans les trois répliquats analytiques d'un patient donné. Cependant, on observe parfois une dispersion du nombre de peptides quantifiés chez un même patient. Cela va en général de paire avec un coefficient de Pearson réduit, confortant l'idée que l'analyse d'un des répliquats s'est moins bien déroulée. Même si l'on examine plus attentivement les données brutes correspondant à ces analyses, ainsi qu'aux contrôles qualifiés les ayant précédés ou suivis, on ne trouve pas de raison particulière à cela.

Les résultats obtenus dans les répliquats sont donc à manier avec prudence. En effet, l'analyse des coefficients de Pearson et des nombres de peptides quantifiés nous a montré que des valeurs aberrantes peuvent exister dans ces données et qu'il faudra mettre en œuvre des techniques de regroupement plus complexe que simplement faire une moyenne. Nous avons donc procédé à un examen plus approfondi des coefficients de variation des valeurs de quantifications pour tenter de limiter l'influence de ces valeurs aberrantes.

2. Filtrage des données

Le coefficient de variation (CV) également nommé écart type relatif, est une mesure de dispersion relative. Il est défini comme le rapport entre l'écart-type et la moyenne. Nous avons évalué les CV des 4950 peptides détectés dans au moins deux répliquats analytiques d'un même patient (voir figure V.9 panneau supérieur). On observe que les coefficients de variation s'étendent jusqu'à 165% avec une moyenne à 32% et une médiane de 27%. Ces importantes variations sont probablement dues à des biais inhérents à l'analyse protéomique (biais introduit lors de l'injection, erreur de quantification du peptide...) et traduisent la présence de valeurs aberrantes. Il est communément admis que des CV supérieurs à 30% ne peuvent pas être considérés pour une analyse statistique. Dans ce cas, nous devrions rejeter presque la moitié des peptides quantifiés.

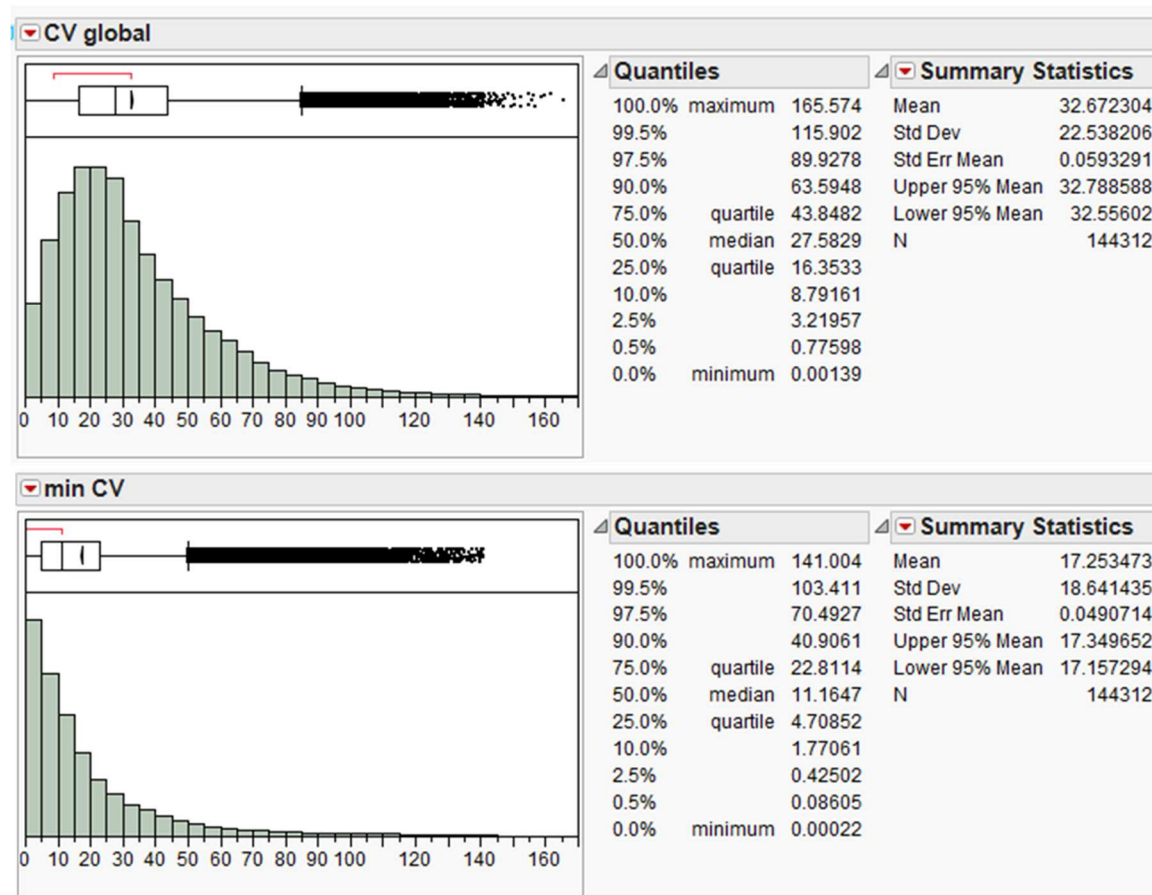


Figure V.9 : (Panneau supérieur) Distribution des coefficients de variation calculés sur l'ensemble des peptides d'un patient donnée en considérant l'ensemble des répliquats analytiques. (Panneau inférieur) Distribution des coefficients de variation de ces mêmes peptides en considérant les répliquats deux à deux et en ne conservant, pour chaque peptide chez un patient donné, que les deux points de données présentant le CV le plus faible.

Pour diminuer ces influences sur le résultat de quantification, nous avons optés pour une méthode de filtrage des données sur la base de la cohérence entre répliquats. Pour ce faire, nous avons évalué les coefficients de variation de chacun des peptides dans les répliquats du même patient pris deux à deux. La figure V.10 illustre le fait que la considération des trois valeurs obtenues tend à élargir la distribution des CV. Si par contre, on ne considère que les couples de valeurs procurant le CV minimal, la distribution est nettement plus resserrée. Lorsque les trois répliquats sont consistants, cette procédure a pour effet de biaiser légèrement l'estimateur de la moyenne. Par contre, lorsque l'un des points s'écarte fortement des deux autres, elle permet de s'affranchir de cette valeur probablement aberrante. La figure V.9 (panneau inférieur) montre le résultat de cette procédure sur la distribution des CV. On note que la médiane est à présent diminuée de 27% à 11%, et qu'une grande majorité des peptides (4648 sur 4950) qualifie à l'étude statistique sur la base du critère CV min inférieur à 30% (Fan, Tam, Vande Woude, & Ren, 2004).

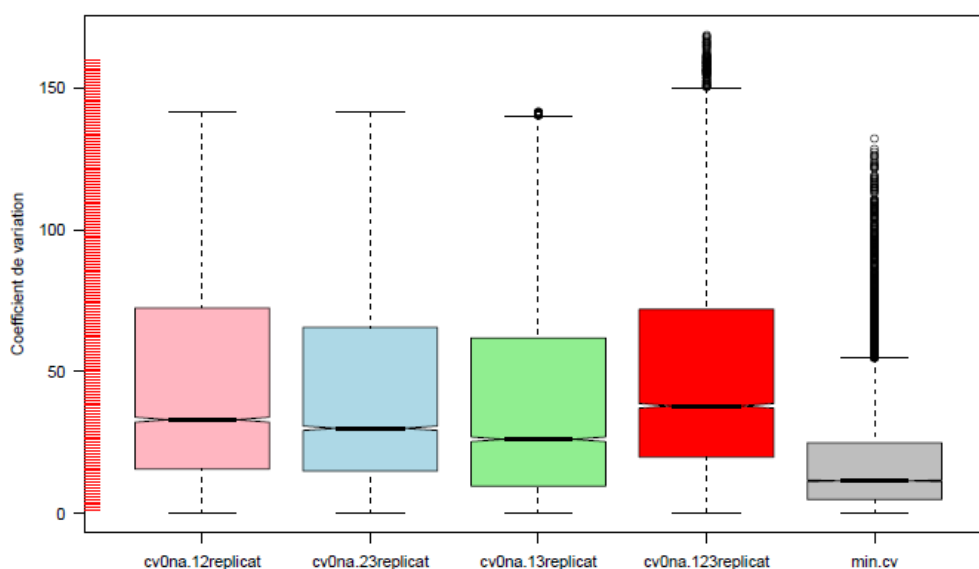


Figure V.10 : Coefficient de Variation selon les permutations deux à deux des trois répliquats analytiques en ayant préalablement éliminé les valeurs manquantes (ONA)

3. Résultats: Quantification d'un millier de protéines dans l'urine

Après avoir appliqué le filtre CV min décrit précédemment, nous avons conservé 4648 peptides correspondant à 1180 protéines groupes.

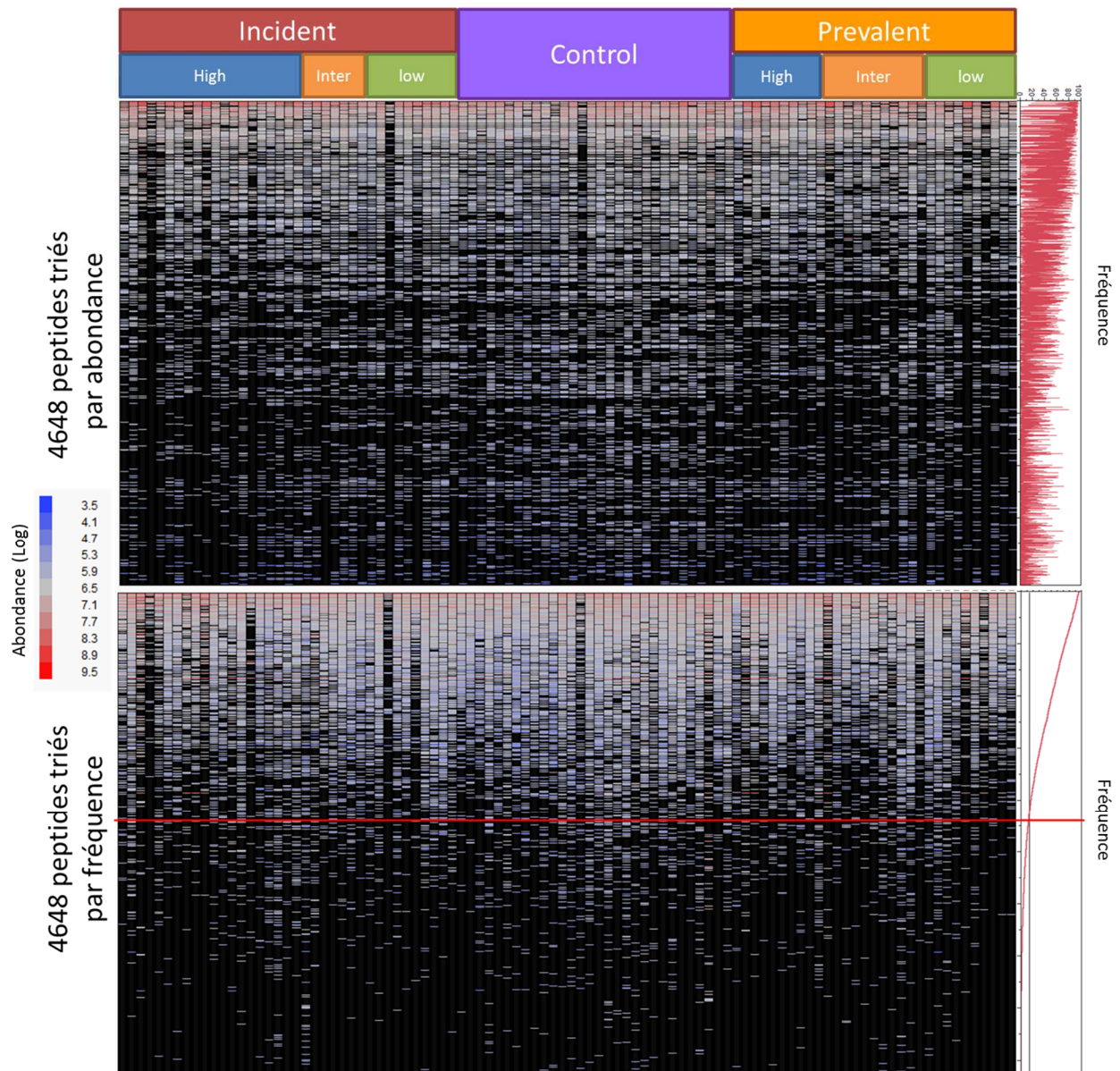


Figure V.11 : Carte en couleur des abondances moyennes des 4648 peptides quantifiés sur l'ensemble des échantillons des 98 patients de la cohorte, classées selon leur abondance moyenne dans le panneau supérieur et occurrence décroissante dans le panneau inférieur.

En vue de l'analyse statistique différentielle, nous avons mis en regard toutes les valeurs réduites de l'abondance de chaque peptide à travers tous les individus de la cohorte. La Figure V.11 représente, sur une échelle logarithmique, les abondances moyennes des peptides quantifiés dans au moins deux réplicats analytiques mesurées pour l'ensemble des patients classés selon leurs catégories (incident, contrôle et prévalent) et les risques de progression ou de récurrence. Dans le panneau supérieur, les peptides ont été classés selon leur abondance moyenne à travers la population, alors que dans le panneau inférieur, ils ont été classés par occurrence décroissante. On note tout d'abord sur cette figure la variabilité inter individuelle en termes de variété des peptides détectés. Bien que les peptides les plus abondants aient en moyenne une plus

grande probabilité de détection, certains sont rarement détectés. De même, certains peptides de très faible abondance sont détectés chez un grand nombre de patients, comme le montre les courbes de fréquences situées à droite des graphes sur la figure V.11. On peut visualiser sur ces graphes un nombre important de données manquantes représentées par les cases noires. Cet aspect spécifique des données de protéomique, nécessite la mise en œuvre de méthodes statistiques capable d'intégrer la détectabilité des espèces dans leur procédure de sélection. Néanmoins, un filtre préalable des peptides les plus rares est souhaitable. Sur suggestion des cliniciens, nous avons donc décidé d'éliminer tous les peptides détectés chez moins de 15% des patients de la cohorte. Sur la figure V.11 (panneau inférieur), la ligne horizontale rouge illustre cette limite de 15% de détection. Ce filtre a permis d'éliminer plus des deux tiers des valeurs manquantes de notre tableau de données. De plus, on visualise sur la figure V.11, la gamme dynamique des abondances peptidiques qui s'étend sur 4.5 ordres de grandeur. Cette profondeur d'analyse témoigne du fait que les protéines quantifiées ne sont pas seulement les protéines majoritaires de l'urine. L'effort réalisé pour peupler la base de données de protéines urinaires, a donc permis de couvrir un large spectre de protéines en terme d'abondance.

Sur la base de ces observations on peut tenter d'établir une classification des protéines majoritaires de l'urine afin de pouvoir vérifier certaines conceptions répandues concernant le protéome urinaire. En particulier, nous nous attendions à une large contribution de l'uromoduline et de l'albumine, mais à une gamme dynamique plus favorable que celle du plasma. Pour cela, nous avons estimé l'abondance médiane des protéines au travers de la cohorte en sommant les intensités médianes de leurs peptides. La figure V.12 représente la contribution des protéines majoritaires quantifiées dans les extraits protéiques urinaires. On retrouve parmi les protéines majeures, des protéines du plasma telles que l'albumine, l'alpha 1 microglobuline, mais aussi la présence de l'uromoduline. Cette dernière est cependant bien moins abondante que rapporté dans la littérature (Rampoldi, Scolari, Amoroso, Ghiggeri, & Devuyst, 2011). Cette observation peut s'expliquer en partie par sa tendance à l'agrégation. En effet, il a été montré que seuls 10% de cette protéine sont retrouvés sous forme libres dans l'urine, le reste étant présent sous forme de fibrilles qui sont très probablement éliminées lors de la centrifugation à basse vitesse. De plus, des glycosylations et l'absence de résidus lysine et arginine sur une partie de sa séquence limitent la couverture obtenue. Par ailleurs, on note une différence marquée entre l'urine et le plasma dans lequel les 10 protéines majoritaires représentent 90% de la teneur totale. Dans les extraits d'urine analysés, les protéines minoritaires représentent presque 50% de la concentration totale. Ces observations valident l'hypothèse que la gamme dynamique de l'urine est bien moindre que celle du plasma, et confirment l'intérêt de ce fluide pour la recherche de biomarqueurs.

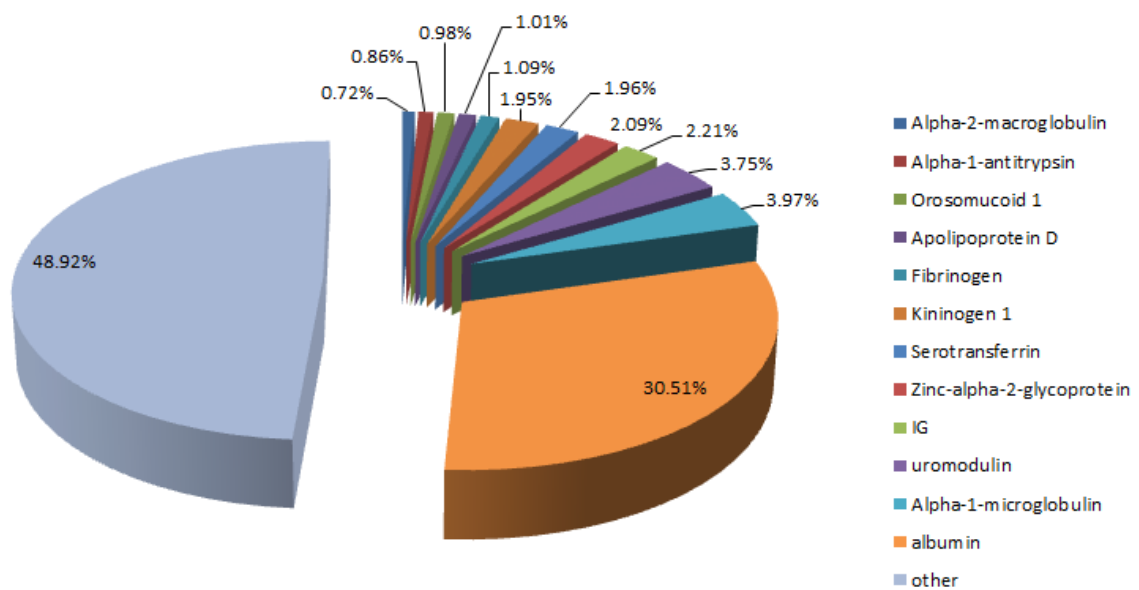


Figure V.12 : Contribution des protéines majoritaires quantifiées dans les extraits protéiques urinaires

4. Découverte de candidats biomarqueurs: statistiques différentielles

Après avoir filtré et vérifié la qualité des données quantitatives issues de l'analyse par spectrométrie de masse, nous les avons traitées selon une méthode statistique basée sur une variante de l'approche Spectral Index proposée par Fu et al (Fu et al., 2008). Cette méthode, appelée abondance index, est basée sur l'estimation d'un index combinant l'information concernant la détection des espèces analysées et leurs variations d'abondances entre les deux populations à comparer. Par conséquent, elle permet de tenir compte des données manquantes. Des permutations entre les populations sont ensuite effectuées afin de déterminer la distribution qu'aurait cet index si l'on considérait les deux populations comme équivalentes. Ceci permet enfin de définir les protéines d'abondances différentielles entre deux conditions (malades et saines). Cette méthode avait été précédemment éprouvée dans une étude de découverte de biomarqueurs de cholangiocarcinomes à partir de cellules isolées par microdissection laser (Dos Santos et al., 2010).

Dans notre étude nous avons remplacé les valeurs de « spectral count » protéiques de l'étude de Fu, par les valeurs d'abondances des peptides quantifiées grâce à l'approche AMT. Nous avons ensuite déterminé pour l'ensemble des peptides quantifiés une valeur d'abondance index définie par la formule :

$$AbI = \frac{\bar{A}}{\bar{A} + \bar{B}} \times \frac{N_A^D}{N_A^T} - \frac{\bar{B}}{\bar{A} + \bar{B}} \times \frac{N_B^D}{N_B^T}$$

Où \bar{A} est la valeur moyenne d'abondance des peptides dans la population contrôle et \bar{B} est la valeur moyenne d'abondance des peptides dans la population cancer (incident ou prévalent).

N_A^D et N_B^D représentent le nombre de fois qu'est détecté le peptide dans la condition contrôle A et cancer respectivement.

N_A^T et N_B^T représentent respectivement le nombre total de patients dans la condition contrôle et cancer.

Pour gagner en puissance statistique et ne pas déséquilibrer le design de l'étude, nous avons décidé de ne pas séparer les patients selon le risque de progression ou de récurrence et de comparer les échantillons urinaires des patients sains et cancers (incidents ou prévalents). En effet, dans la catégorie contrôle, nous avons 30 individus et dans chacune des populations incident et prévalent, nous avons un nombre équivalent de patients (37 et 31). La distribution des ABI observés pour les 2172 peptides quantifiés et filtrés est représentée sur la figure V.13 pour la comparaison patients incidents versus patient contrôles (panneau supérieur) et pour les patients prévalents versus patient contrôles (panneau inférieur) :

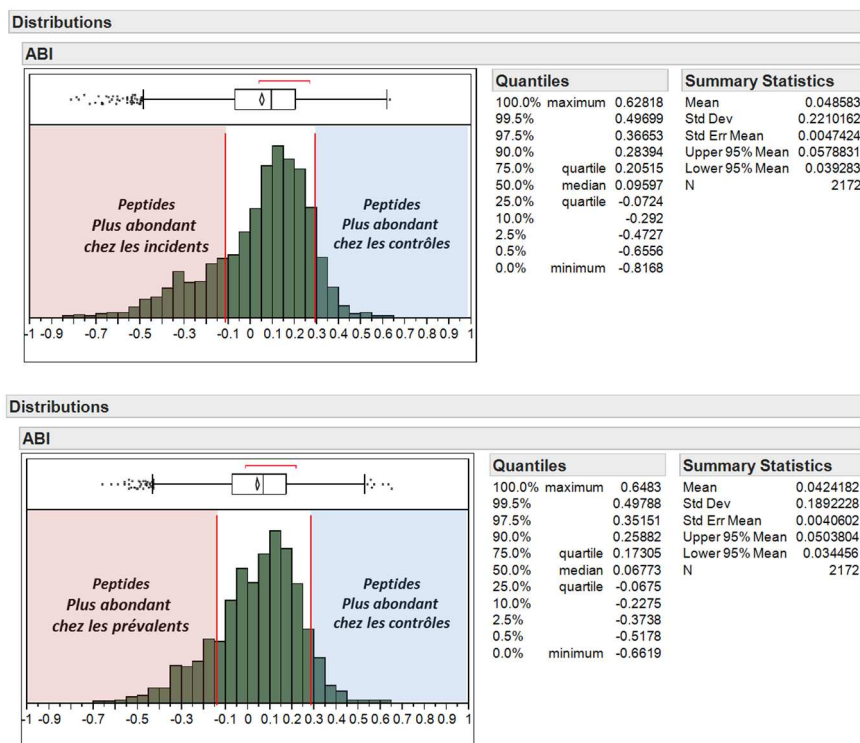


Figure V.13 : Distribution des valeurs d'abondance index calculées à partir des comparaisons des patients incidents et contrôles (panneau supérieur) et des patients prévalents et contrôles (panneau inférieur)

Nous avons déterminé les seuils à partir desquels les peptides peuvent être considérés comme significativement dérégulés dans l'une ou l'autre des populations. Pour ce faire, des permutations aléatoires

des patients entre les classes (contrôle vs cancer) ont été effectuées pour définir l'intervalle de confiance en dehors du gel on peut différencier les variations dues à la pathologie des variations dues à la variabilité des mesures d'abondances. Cette permutation a été réalisée grâce à une macro développée sous Excel. Les bornes, que l'on visualise en rouge sur les Figures V.13 ont ensuite été déterminées en multipliant l'écart type de la distribution obtenue après permutation par 1.96 pour obtenir un intervalle de confiance de 95% autour de la valeur médiane de l'abondance Index observée. Les peptides situés à l'intérieur de cet intervalle ont donc 95% de probabilité d'être non discriminants. De part et d'autre, on trouve les peptides significativement surabondants chez les cancers ou les contrôles. En comparant les conditions incident et contrôle, l'intervalle de confiance à 95% s'étend de -0.11 à 0.3. Pour la comparaison prévalent et contrôle, les bornes de l'intervalle de confiance étaient -0.14 et 0.27. Par ailleurs, nous avons décidé d'inclure aux résultats de l'analyse statistique les peptides détectés exclusivement dans l'une ou l'autre condition. Ces peptides sont nommés « cancer only» ou « contrôle only». Ces peptides sont comptés en tant que tels s'ils sont détectés chez au moins 15% des patients de la population dont il est question, cela concerne au total une dizaine de peptides. Au final, à partir des 2188 peptides qui ont été traités dans cette étude statistique par Abondance index, nous avons obtenu une large majorité de peptides ne variant pas de façon significative : 70% et 75% pour les comparaisons incident / contrôle et prévalent / contrôle respectivement (Tableau V.1). De plus, on trouve une plus large proportion de peptides surabondants dans les deux conditions cancers (22% et 16%) que dans la condition contrôle (8%).

Tableau V.2 Classement des peptides en fonction des résultats de l'analyse statistique par Abondance Index.

	incident vs contrôle	prévalent vs contrôle
contrôle only	1 (0.04%)	1 (0.04%)
plus abondant contrôle	166 (8%)	175 (8%)
non variant	1530 (70%)	1643 (75%)
plus abondant cancer	476 (22%)	357 (16%)
cancer only	15 (0.7%)	12 (0.5%)

A partir des informations sur la classification des peptides, nous sommes passés au niveau protéique en comptant pour chaque protéine le nombre de peptides dans les différentes catégories telles que définies dans le tableau V.1. Pour considérer qu'une protéine est différentiellement abondante, nous exigeons que la majorité de ses peptides protéospécifiques montrent une dérégulation dans le même sens et qu'aucun de ses peptides ne montre de variation contradictoire. Nous avons ainsi extraits une liste de 64 candidats biomarqueurs surabondants chez les patients atteints de cancer et de 33 candidats biomarqueurs surabondants chez les contrôles. La liste de ces candidats triés par médiane d'abondance inde est compilée dans le tableau V.2.

Accession	ID	name	Gene name	N/AMT/ spe
ORNF17_HUMAN	ORNF17	FLJ00385 protein (Fragment)	FLJ00385	4
APQAF1_HUMAN	P02647	Apoptoprotein A1	APQAF1	21
S10BP1_HUMAN	P25816	Protein S100-P	S10BP1	2
APQAF2_HUMAN	P02652	Apoptoprotein A-II	APQAF2	4
AKM3_HUMAN	P01023	Alpha-2-macroglobin	AKM3	24
ICAM2_HUMAN	P13368	Intercellular adhesion molecule 2	ICAM2	1
KCRA2_HUMAN	P19013	Keratin, type II cytoskeletal 4	KCRA2	1
HBA_HUMAN	P09365	Hemoglobin subunit alpha	HBA2	9
FERB_HUMAN	P02676	Ferritinogen beta chain	FERB	13
FERG_HUMAN	P02679	Ferritinogen gamma chain	FERG	13
AAM2_HUMAN	P19201	Peptidyl-glycine alpha-synthesizing monooxygenase	PAM	16
AAM3_HUMAN	P19201	Peptidyl-glycine alpha-synthesizing monooxygenase	PAM	16
CO3_HUMAN	P01028	Complement C3	CO3	47
HPT1_HUMAN	P00738	Haptoglobin	HPT	11
IGH33_HUMAN	P33043	Ig gamma 3 chain C region	IGH33	4
BLVFB_HUMAN	P03043	Blauv protein	BLVFB	1
OSNV90_HUMAN	OSNV90	V2-17 protein (Fragment)	KELV325	2
AKCUB_HUMAN	AKCUB	cDNA FLJ2848, highly similar to human lactoferrin		2
A1AG1_HUMAN	P02763	Alpha-1-acid glycoprotein 1	ORA1	3
SC9L_HUMAN	Q02673	Scavenger receptor class C type 1	SC9L	1
Q9LUB8_HUMAN	Q9LUB8	Myosin-reactive immunoglobulin heavy chain variable region (Fragment)		1
VTB3_HUMAN	P02774	Vitamin D-binding protein	GC	12
S10A6_HUMAN	P06703	S100 calcium binding protein A6 (Calcyon), isoform CRA_a	S100A6	1
BDPNT1_HUMAN	BDPNT1	Protein basic protein	PPP	1
OSGANS_HUMAN	OSGANS	cDNA FLJ5893c, highly similar to Keratin, type II cytoskeletal 4		1
HEMO_HUMAN	P02799	Hemopexin	IGH	1
APQAF3_HUMAN	P41222	Prostaglandin-H2 isomerase	HFX	17
APRW93_HUMAN	APRW93	Selenium binding protein 1 (Fragment)	PLG35	5
APRW92_HUMAN	APRW92	Selenium binding protein 1 (Fragment)	SELENBP1	12
A1BG_HUMAN	P04217	Alpha-1B-glycoprotein	A1BG	1
Q17251_HUMAN	Q17251	Putative uncharacterized protein DKF2688N0239	DKF2688N0	1
TTR_HUMAN	P02765	Transthyretin	TTR	1
PPAP_HUMAN	D30K05	Acid phosphatase, prostatic, isoform CRA_c	ACPP	14
OSTRH5_HUMAN	OSTRH5	Osteocalcin (Fragment)	OSN	1
ZNGC_HUMAN	P25311	Zinc-alpha-2-glycoprotein	AZGP1	23
TRE_HUMAN	P02787	Serotonectin	TR	1
OSV08_HUMAN	OSV08	Complement component 4 binding protein, alpha	CBPA	2
ALDOA_HUMAN	P04075	Fructose-bisphosphate aldolase A	ALDOA	5
AFM1_HUMAN	P43652	Ablum	AFM	3
COL4A1_HUMAN	OSNV1	Collagen alpha-1(I) chain	COL4A1	5
OSV02_HUMAN	OSV02	Gelsolin (Kunyadocin, Finnish type)	GSN	1
OSV03_HUMAN	OSV03	Basement membrane-specific heparan sulfate proteoglycan core protein variant (Fragment)		1
AVP05_HUMAN	P01597	Arginine vasopressin V1 region DEE	AVP03	2
APQCC3_HUMAN	P02666	Apoptoprotein C-III	APQCC3	2
TH2_HUMAN	P19423	Inter-alpha-trypsin inhibitor heavy chain H2	TH2	2
KPYM_HUMAN	P14618	Pyruvate kinase isozyme M1/M2	PKM2	2
SFPG_HUMAN	Q15772	Striated muscle preferentially expressed protein kinase	SFPG	1
SBP1_HUMAN	Q13226	Selenium-binding protein 1	SELENBP1	1
FBA_HUMAN	D3DP14	Fibrinogen alpha chain, isoform CRA_b	FBA	21
B72D01_HUMAN	B72D01	cDNA FLJ57374		1
B25R92_HUMAN	B25R92	cDNA, FLJ36571, highly similar to Homo sapiens serine (or cysteine) protease	SERPNA6	2
A1AT_HUMAN	P01009	Alpha-1-antitrypsin	SERPNA1	25
A3KST3_HUMAN	A3KST3	Superoxide dismutase [Cu-Zn]	SOD1	2
OSG_HUMAN	P08185	Corticosteroid-binding globulin	SERPNA6	2
CE9A1_HUMAN	P09460	Ceruloplasmin	CP	10
N1FG2_HUMAN	P19662	Alpha-1-acid glycoprotein 2	ONM2	2
N4R7_HUMAN	Q14746	Nk(-)H(-) exchange regulatory cofactor NHE-RF1	SLC9A3H1	2
BCE124_HUMAN	BCE124	cDNA FLJ5673, highly similar to Complement factor B (FC3.4.21.47)		8
A3RTV6_HUMAN	A3RTV6	Inter-alpha (Globulin) inhibitor H2	TH2	2
GS1P1_HUMAN	P09211	Galectin-1	GS1P1	3
GLULH1_HUMAN	GLULH1	Galectin-1-like protein		2
BCE1B2_HUMAN	BCE1B2	cDNA FLJ53931, highly similar to Serotonectin	SOD1	2
SODC3_HUMAN	D3DCE4	Superoxide dismutase [Cu-Zn]	SOD1	2
PRDX1_HUMAN	Q09830	Peroxiredoxin-1	PRDX1	2
GAS1_HUMAN	P54626	Growth arrest-specific protein 1	GAS1	1
HEPC_HUMAN	P81172	Heparin	HAMP	1
B2K6_HUMAN	P81789	Beta-2-microglobulin	B2M	4
CTTN_HUMAN	Q15829	Cytoslin-M	CTN1B	3
RE14_HUMAN	P02735	Retinol binding protein 4	REB4	5
AUD8_HUMAN	P05262	Fructose-bisphosphate aldolase B	AUD8	9
ICD3_HUMAN	Q91095	IQ domain-containing protein G	ICD3	5
AKTIA1_HUMAN	P14650	Alcohol dehydrogenase [NAD(P)+]	AKR1A1	1
UBIQ_HUMAN	B2RQV1	Eubiquitin	UBI1	1
AMPY_HUMAN	P15144	Aminopeptidase N	RPS27A	7
PIP_HUMAN	P12273	Protein-tyrosine phosphatase N	PTNP	3
GF15_HUMAN	Q09886	Growth/differentiation factor 15	GF15	3
MTIG_HUMAN	P13640	Metallothionein-G	MTIG	1
CAB3_HUMAN	P07369	Cathepsin B	CTSB	3
LAVER1_HUMAN	P11279	Lysosome-associated membrane glycoprotein 1	LAMP1	1
IES_HUMAN	P06373	Eukaryotic translation initiation factor 6	EIF6	1
UDHB_HUMAN	P07196	Lactate dehydrogenase B chain	UDHB	4
PPDL_HUMAN	P11117	Lysosomal acid phosphatase	ACP2	4
REG3A_HUMAN	P06461	Urostatinase (sialin)	REG3A	5
B2R8G3_HUMAN	B2R8G3	glutaryl aminopeptidase (aminopeptidase A)		7
RETN_HUMAN	Q41089	Retinol	RETN	2
NAPSA_HUMAN	P61916	Epididymal secretory protein E1	NAPSA	4
COMP_HUMAN	P49747	Cartilage oligomeric matrix protein	COMP	1
CD14_HUMAN	P08671	Monocyte differentiation antigen CD14	CD14	16
FABP5_HUMAN	Q01489	Fatty acid-binding protein, epidermal	FABP5	1
GNS_HUMAN	P15986	N-acetylglucosamine-6-sulfatase	GNS	7
PGLYRP1_HUMAN	Q75294	Peptidoglycan recognition protein 1	PGLYRP1	6
EGF_HUMAN	Q12905	EGF-containing domain like extracellular matrix protein 1	EHEM1	3
OPN2_HUMAN	P22262	Chemokine receptor 3	OPN2	2
ABP3_HUMAN	Q0914	Alphabeta hydrolase domain-containing protein 14B	ABP3	2
NCC31_HUMAN	Q02818	Nucleoside triphosphatase	NCC31	2
PL16_HUMAN	Q80086	Peptidase inhibitor 16	PI16	2

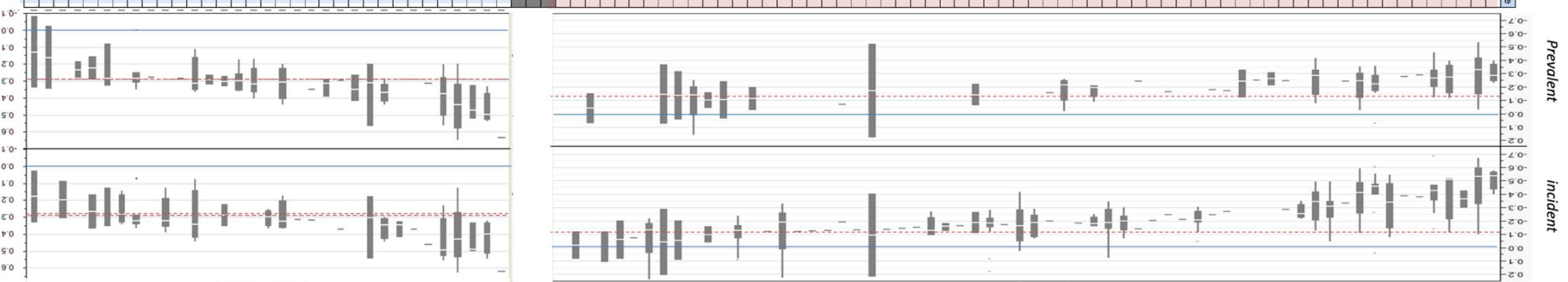


Tableau V.3 : Nos candidats biomarqueurs classés par ordre décroissant des médianes des valeurs d'abondance index de leurs peptides protéosécificques dans les comparaisons incident / contrôle et prévalent / contrôle

Dans ce tableaux, les candidats biomarqueurs ont été classés par ordre décroissant de médiane des valeurs d'abondance index de leurs peptides protéosécifics dans les comparaisons incident / contrôle et prévalent / contrôle. Les box plots correspondant à ces valeurs d'Abondances Index sont représentés sur le graphe à droite du tableau. On note une large dispersion des valeurs médianes d'abondance Index, allant de -0.54 à -0.02 pour les protéines surabondantes chez les patients cancers et de 0.13 à 0.63 pour les protéines qui présentent une diminution de leur abondance chez les patients cancers. Pour rappel, les bornes déterminées lors de l'analyse statistique sont représentées par des lignes pointillées rouges sur les graphes correspondants. De plus, on observe que certains candidats biomarqueurs comme la Vitamin D binding par exemple sont détectés uniquement dans l'une au l'autre des catégories de patients cancers (incident ou prévalent). Ceci suggère que les primo tumeurs associées aux patients incident et les tumeurs récidivantes (patients prévalents) sont peut-être différentes et expliquerait le fait que les tumeurs récidivantes ont des caractères différents (infiltrant le muscle, métastatique) des primo tumeurs.

D. Candidats biomarqueurs du cancer de la vessie

La liste des candidats biomarqueurs maintenant déterminée, nous allons l'analyser d'un point de vue fonctionnel (voies métaboliques, localisation tissulaire ou subcellulaire) en utilisant différents outils bioinformatiques disponibles à la communauté.

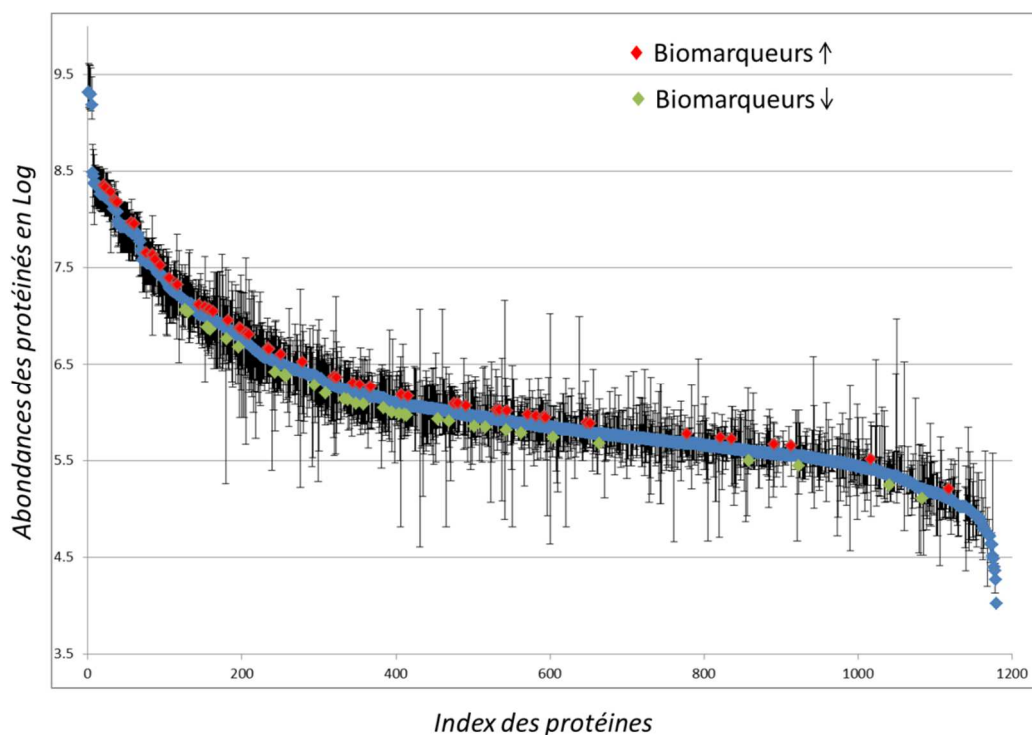


Figure V.14 : Protéines quantifiées classées par abondances médiane décroissante sur l'ensemble des patients de la cohorte de découverte en mettant en évidence les candidats biomarqueurs sur et sous abondants chez les patients cancers

Dans un premier temps, nous avons mis en évidence ces candidats biomarqueurs sur une courbe représentant l'ensemble des protéines quantifiées à partir de l'analyse LC-MS de l'ensemble des échantillons urinaires (Figure V.14). Pour calculer l'abondance des protéines, nous avons sommé les valeurs d'abondances de leurs peptides pour chaque patient et calculé la médiane sur l'ensemble des 98 patients. Les protéines ainsi quantifiées ont été classées par abondance décroissante. Les barres d'erreurs correspondent aux 1^{er} et 3^{ème} quartiles de la distribution d'abondance de cette protéine sur la cohorte totale. On remarque que les variations sont fortement protéines dépendantes et que l'abondance de certaines protéines varie de plus de 2 ordres de grandeurs, même en ne considérant que les 50% centraux de la population. Ces variations semblent dépendantes de l'abondance de la protéine et les protéines majoritaires montrent en général une meilleure consistance. Au final, on couvre une gamme dynamique protéique d'environ 5.5 ordre de grandeurs. On observe que les candidats biomarqueurs sur ou sous abondants chez les patients cancers sont répartis sur la quasi-totalité de la courbe. Ceci confirme que les candidats biomarqueurs découverts dans cette étude, ne sont pas uniquement des protéines majoritaires de l'urine.

1. Classification fonctionnelle des candidats

Nous nous sommes ensuite intéressés aux fonctions moléculaires, aux compartiments cellulaires et aux processus biologiques dans lesquels pourraient être impliqués nos candidats biomarqueurs. Pour ce faire

nous avons analysé les termes d'ontologie des gènes (GO) associés à ces protéines à l'aide du logiciel « Protein Center ». Cette ontologie est extrêmement vaste et ce logiciel permet un regroupement de ces termes en un nombre limité de catégories appelées GO Slim.

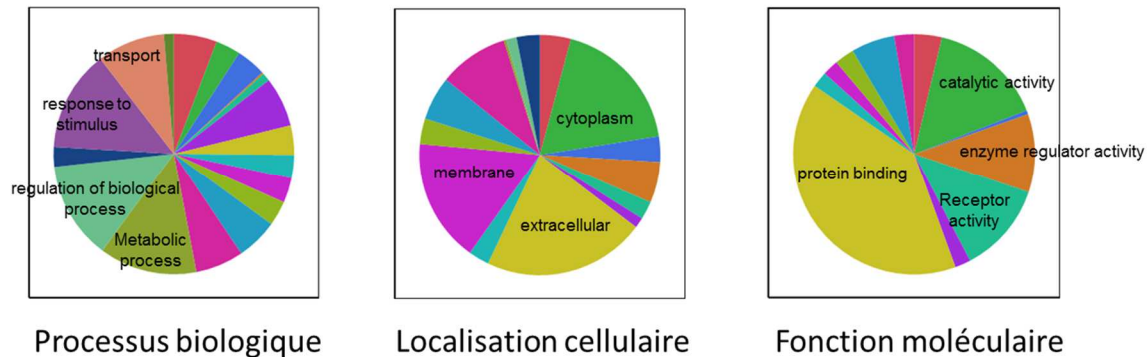


Figure V. 15 : Classification fonctionnelle des termes GO Slim des candidats biomarqueurs plus montrant une augmentation de leur abondance chez les patients cancers par processus biologique, localisation cellulaire et fonctions moléculaire.

La figure V.15 montre la répartition des termes GO associée à nos candidats biomarqueurs par processus, localisation et fonction. Si l'on compare cette distribution à celle obtenue précédemment sur l'ensemble des protéines urinaires identifiées dans notre base AMT, nous pouvons constater un enrichissement des protéines impliquées dans les processus de transport et les fonctions associées à la régulation de l'activité enzymatique ou à l'activité des récepteurs. Ceci corrobore l'hypothèse de départ selon laquelle le tissu tumoral relarguerait des protéines susceptible d'être détectées dans l'urine. Les biomarqueurs seraient ainsi des protéines de la surface des cellules, des protéines sécrétées, des récepteurs ou des transporteurs impliqués dans le trafic cellulaires, la dégradation ou l'externalisation. Un bon nombre de ces protéines sont en effet annotées comme exosomales.

2. Les Exosomes

Les exosomes sont des vésicules membranaires d'origine endocytaire, de 40 à 100 nm, sécrétés par les cellules dans leur environnement. Leurs fonctions ne sont pas totalement élucidées, mais on parle de leur implication dans la communication intercellulaire, et plusieurs sources (Théry, Zitvogel, & Amigorena, 2002), (Mathivanan, Ji, & Simpson, 2010) invoquent un rôle de ces vésicules dans des processus pathologiques, notamment dans le cas de pathologies uro-génitales (Zhou et al., 2008). Ces exosomes contiennent des signatures génomiques et protéiques propres aux cellules qui leur ont donné naissance (Henderson & Azorsa, 2012). Depuis quelques années, de nombreuses études ont montré la présence d'exosomes dans les fluides biologiques et notamment l'urine (Pisitkun et al., 2004). Il a été suggéré que les cellules tumorales sécrétaient des exosomes en plus grande quantité que les cellules normales. De plus, les exosomes sécrétés par ces cellules peuvent participer à la progression du cancer, en favorisant l'irrigation du

foyer tumoral, ou en supprimant la réponse immunitaire. Les protéines d'origines exosomales sont donc particulièrement intéressantes comme candidats biomarqueurs car on les retrouve à l'extérieur des cellules et que leur signature est spécifique de la tumeur dont ils sont originaires. Par exemple, FGF-2 (fibroblast growth factor) qui intervient dans la prolifération cellulaire pendant l'angiogenèse a été détecté dans les exosomes de cellules tumorales (Ceccarelli et al., 2007).

Lors de l'analyse bioinformatique de notre liste de biomarqueurs, il est apparu qu'une forte proportion des protéines candidates était annotée comme exosomales. La Figure V.16 présente les graphes d'interaction obtenus à l'aide du logiciel String (Know and Predicted Protein-Protein Interaction) (string-db.org) des protéines sur- et sous-abondantes dans les urines de patients atteints de cancer. L'analyse par String des termes GO relatifs à la localisation cellulaire de ces protéines révèle également un enrichissement en protéines exosomales, représentées en rouge sur la figure. Cet enrichissement, très significatif, puisqu'il concerne la moitié des protéines de notre liste (48/97) est très surprenant. En effet, nous n'avions fait aucun effort particulier pour purifier ces sous-compartiments. C'est donc très probablement en raison de leur spécificité au processus pathologique que ces protéines se retrouvent à ce point enrichies.

Typiquement, la préparation des exosomes requiert l'ultracentrifugation (>100 000 rcf) de longue durée (1 heure) à partir de grands volumes d'urines, ce qui peut être compliqué pour une étude clinique. Dans notre cas, nous avons obtenu un pool de protéines différentielles fortement enrichi en protéines d'origines exosomales sans avoir eu recours à la purification de ces compartiments. Ceci illustre la facilité d'accès à ces candidats et préfigure bien de leur potentielle utilité clinique.

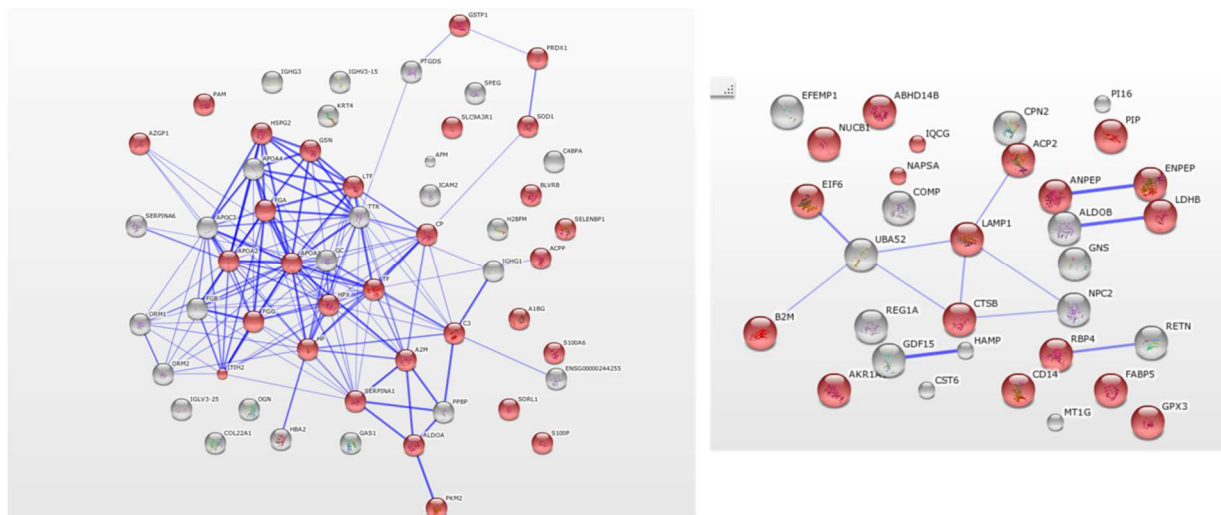


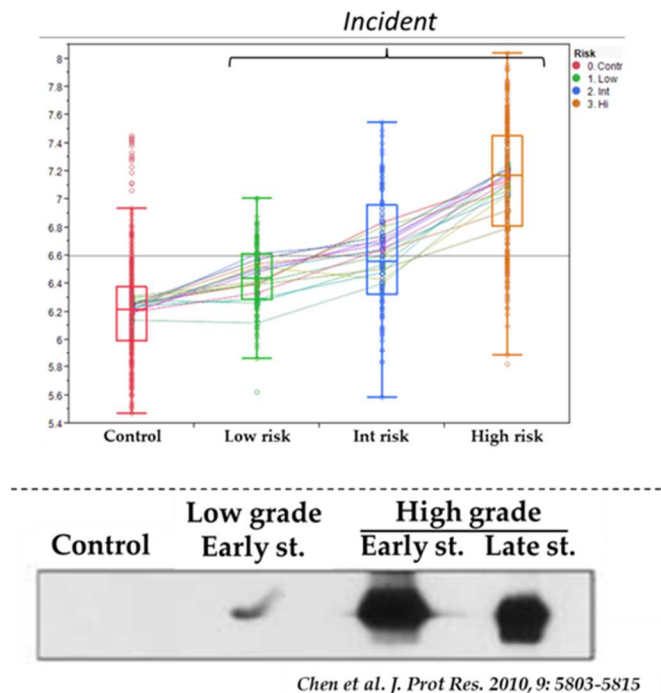
Figure V.16 : Graphe d'interaction de nos candidats biomarqueurs surabondants chez les patients cancers (panneau de gauche) et surabondants chez les contrôles (panneau de droite). En rouge, les protéines d'origine exosomales.

3. Evaluation critique de la liste de candidats

Pour l'instant, nous avons considéré la liste des candidats dans sa globalité à travers des analyses bioinformatiques qui ont révélées des traits collectifs d'intérêts. La quête des candidats les plus pertinents passe nécessairement par une évaluation critique des protéines de la liste. Pendant le déroulement de ce projet, des études sur la découverte de biomarqueurs du cancer de la vessie dans les urines ont été publiées dans la littérature. Celles de Linden (Lindén et al., 2012) et de Chen (Chen et al., 2010), nous apparaissent les plus pertinentes pour affiner nos résultats.

Linden et collaborateurs ont identifié 29 candidats biomarqueurs du cancer de la vessie, à partir d'une analyse de protéomique quantitative « Label-free » d'urine de 11 patients cancers et 11 patients contrôles sains. Une analyse par western Blot a confirmé une abondance différentielle pour 4 de ces candidats biomarqueurs, dont deux figurent dans notre liste : le Fibrinogen β chain (FGB) et l' α -1-antitrypsin (SERPINA1). De plus l'exploration des coupes tissulaires disponibles dans Human Protein Atlas (HPA, <http://www.proteinatlas.org>) a montré une surexpression de ces 2 candidates sur des tissus cancers de l'urothélium par rapport à des tissus normaux.

Dans l'étude de Chen, une analyse protéomique quantitative des urines de 14 patients contrôles sains et 23 patients cancers a été réalisée en utilisant la technologie iTRAQ. Les urines des patients ont été poolées en trois groupes représentatifs de différentes classes de cancer : bas grade-stade précoce, haut grade-stade précoce et haut grade-stade avancé. Cette étude a permis d'identifier 638 protéines dont 55 ont été déterminées comme différentiellement abondantes entre les patients cancers et les patients contrôles. Une évaluation par Western Blot a permis de vérifier l'abondance différentielle de quelques-uns de ces candidats biomarqueurs. Une validation par la technique ELISA a aussi confirmé l'accroissement d'abondance de l'APOA1 chez les patients cancer comparés à des patients contrôles atteints d'infections urinaires ou d'hématurie. Les auteurs soulignent le potentiel de cette protéine pour le diagnostic du cancer de la vessie.



Chen et al. J. Prot Res. 2010, 9: 5803-5815

Figure V.17 : Protéine APOA1 : profil d'abondance en log de la protéine chez les patients contrôles, cancers incident associés au risque de progression tumoral déterminé dans notre étude quantitative (panneau supérieur) ; Vérification de l'abondance différentielle de cette protéine dans l'étude de Chen par Western blot dans les urines de patients contrôle, de bas grade stade précoce, haut grade stade précoce et stade avancé (panneau inférieur).

Dans notre liste, nous avons 23 protéines en commun avec celles de Chen et 7 avec l'étude de Linden. Parmi ces protéines, l'APOA1 a été détectée dans les trois études comme l'une des plus dérégulées. Cette protéine associée aux exosomes est une composante majeure des lipoprotéines de haute densité (HDL) responsable du transport du cholestérol vers le foie où il pourra être éliminé. La Figure V.17 montre le profil d'abondance que nous avons obtenu pour cette protéine en fonction des risques de progression et de récurrence chez les patients incidents (panneau supérieur). Ce graphique est mis en parallèle avec les résultats de la vérification par Western blot des profils quantitatifs obtenus par Chen dans son étude (panneau inférieur). On observe dans les deux cas une corrélation entre l'abondance de cette protéine et la sévérité de la pathologie. Ce résultat est également corroboré par le profil quantitatif obtenu par Linden.

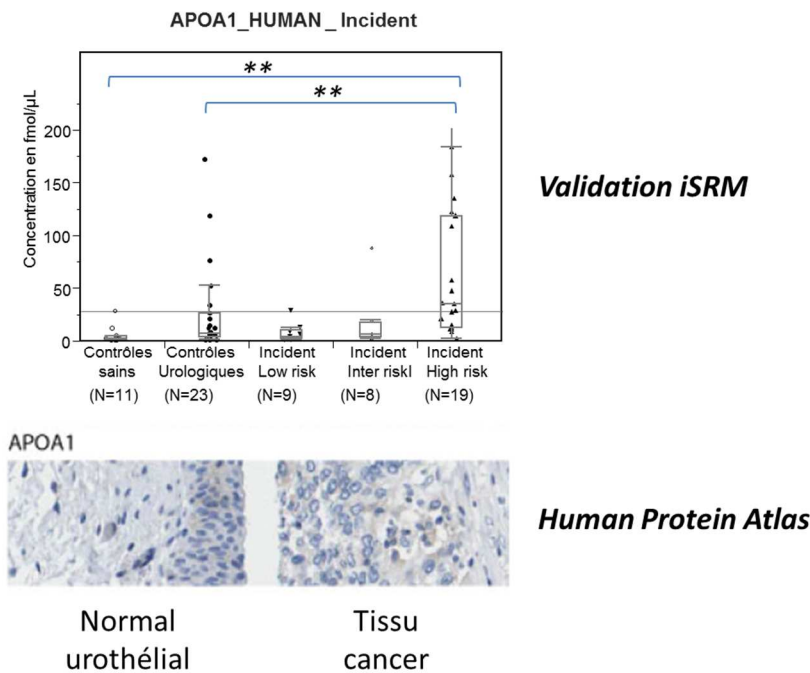


Figure V.18 : Concentration de l'APOA1 déterminée par l'analyse iSRM sur des urines de patients contrôles sains et urologiques et de patients atteint d'un cancer incident en fonction du risque de progression tumorale (panneau supérieur) ; visualisation du marquage de cette protéine sur le site HPA sur des coupes de tissu d'urothélium normal et cancer (panneau inférieur).

A la suite de notre étude, un certain nombre de candidats ont été évalué par i-SRM chez notre partenaire à CRP-Santé Luxembourg. Il est important de souligner que pour cette évaluation la cohorte analysée incluait des contrôles urologiques. Cette nouvelle classe était composée de patients venus consulter pour une suspicion d'une tumeur vésicale et pour lesquels la tumeur n'avait pas été confirmée suite à la cystoscopie. Ces patients contrôles représentent les véritables cas auxquels sont confrontés les cliniciens lorsqu'ils doivent établir un diagnostic. Le profil de concentration urinaire de l'APOA1 obtenu par iSRM sur la cohorte d'évaluation est représenté Figure V.18 (panneau supérieur). Ces résultats confirment l'abondance différentielle entre les patients cancer et les contrôles sains. Cependant, lorsque l'on compare ces concentrations avec celles trouvées chez les contrôles urologiques, on perd le pouvoir discriminant de cette protéine pour les risques intermédiaire et faible. De plus, si l'on examine les coupes tissulaires disponibles sur le site Human Protein Atlas (Figure V.18 – panneau inférieur), on peut émettre des doutes sur le fait que ce marqueur ait pour origine le tissu tumoral. L'augmentation de l'abondance de l'APOA1 dans l'urine semble donc due à une réponse de l'organisme au processus pathologiques, et non pas lié uniquement au développement tumoral. Ceci souligne le caractère potentiellement non spécifique de ce candidat biomarqueur que l'on retrouve dérégulé dans différentes pathologies confondantes.

Les biomarqueurs que l'on recherche sont idéalement issus spécifiquement de la tumeur. Pour filtrer notre liste de candidats, nous nous sommes donc intéressés, à des études de découverte de biomarqueurs

urinaires d'autres pathologies. Une étude en particulier a retenu notre attention, en raison du fort recouvrement avec notre liste. Il s'agit du travail de Kalantari et al (Kalantari et al., 2013) sur la néphropathie IgA, une maladie auto-immune atteignant le rein. En comparant par une méthode de protéomique « label-free » les urines de patients atteints d'une néphropathie IgA à différents stades, ces auteurs ont extrait une liste de protéines dont les abondances étaient corrélées à la sévérité de cette pathologie. Le Tableau V.3 compile la liste des 20 candidats communs avec notre étude. On note en particulier la présence de l'APOA1, ce qui confirme encore une fois son caractère peu spécifique. On remarque également que cette liste contient des protéines associées à la cascade du complément et de la coagulation (C3, A2M, C4BP), ce qui suggère une réponse à l'inflammation. Ces candidats semblent donc être d'un moindre intérêt pour le diagnostic du cancer de la vessie.

Tableau V.4 Sélection de nos candidats biomarqueurs qui présentent une abondance différentielle dans les néphropathies à IgA

ID	name	Gene name
P02647	Apolipoprotein A-I	APOA1
P01023	Alpha-2-macroglobulin	A2M
P01024	Complement C3	C3
P00738	Haptoglobin	HP
P02763	Alpha-1-acid glycoprotein 1	ORM1
P02774	Vitamin D-binding protein	GC
P02790	Hemopexin	HPX
P06727	Apolipoprotein A-IV	APOA4
P04217	Alpha-1B-glycoprotein	A1BG
P02766	Transthyretin	TTR
P25311	Zinc-alpha-2-glycoprotein	AZGP1
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	ITIH2
A2RTY6	Inter-alpha (Globulin) inhibitor H2	ITIH2
P02753	Retinol-binding protein 4	RBP4
Q15828	Cystatin-M	CST6
P61769	Beta-2-microglobulin	B2M
P06703	S100 calcium binding protein A6 (Calcyclin), isoform CRA_a	S100A6
P43652	Afamin	AFM
P00450	Ceruloplasmin	CP
P07858	Cathepsin B	CTSB

Les protéines candidates de notre étude, méritent désormais d'être validée par d'autres techniques de protéomique quantitative ou par ELISA. Pour démontrer la pertinence de cette liste réduite, nous nous focalisons maintenant sur deux protéines annotées comme exosomales qui pourraient être directement liées à la tumeur et permettraient d'établir un diagnostic des cancers incidents et/ou prévalents.

La protéine S100P figure parmi les plus dérégulées chez les cancers incidents. Cette protéine appartient à la famille des protéines S100 localisées dans le cytoplasme ou le noyau et impliquée dans de nombreux processus cellulaires tels que les cycles de progression ou de différenciation. En 2007, il a été démontré que le gène S100P, permettrait de distinguer les tumeurs urothéliales des tumeurs de la prostate et

du rein (Higgins et al., 2007). Selon ces sources, la surexpression du gène S100P permettrait de différencier les tumeurs urothéliales, d'autres tumeurs uro-génitales.

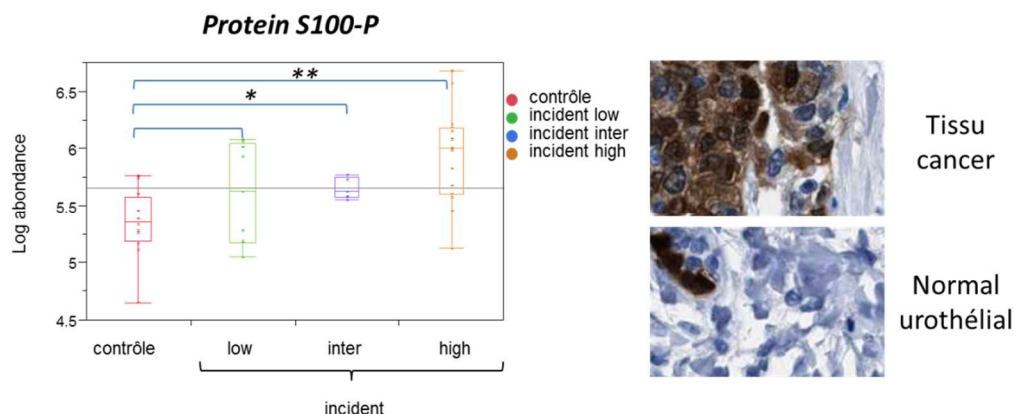


Figure V.19 : Profil d'abondance de la protéine S100P chez les patients contrôles, cancers incident associés au risque de progression tumoral déterminé dans notre étude quantitative (panneau de gauche) ; visualisation du marquage de cette protéine sur le site HPA sur des coupes de tissu d'urothélium normal et cancer (panneau de droite).

Dans notre étude, nous ajoutons une évidence protéique à ces observations (Figure V.19 panneau de gauche), en démontrant que l'abondance différentielle de cette protéine dans l'urine est corrélée au facteur de risque de progression. En effet, on note une différence significative (test de Wilcoxon) entre les patients contrôles et incident (risques intermédiaires et élevés, marqué par un astérisque). Dans son étude, Chen a également détectée cette protéine comme dérégulée, mais n'a pas évalué sa pertinence par western Blot ou par ELISA. Lorsque l'on examine les immunohistologies de tissus normaux et cancéreux, dans Human Protein Atlas, on note une différence notable de marquage (Figure V.19 panneau de droite). Cette protéine représente ainsi un candidat biomarqueur ayant un fort potentiel pour la détection d'une tumeur vésicale chez les patients n'ayant aucun antécédent. De plus par l'exploration dans HPA, on confirme l'origine tumorale de cette protéine. Enfin les résultats obtenus par Higgins en 2007 apportent une preuve supplémentaire sur la spécificité du gène correspondant à cette protéine. Nous n'avons malheureusement pas de validation par SRM ou ELISA de cette protéine.

Par ailleurs, parmi l'ensemble des protéines dont l'abondance est affectée négativement par le cancer, on note la présence d'une protéine connue comme marqueur de la membrane lysosomale : LAMP1. Cette protéine a été impliquée dans la capacité métastatique et l'agressivité des tumeurs sur des modèles murins (Fehrenbacher et al., 2008). Dans notre étude, cette protéine a été détectée surabondante chez les patients contrôles par rapport aux patients cancers (incident et prévalent). Contrairement à la précédente, cette protéine a été évaluée par SRM chez notre partenaire au CRP Santé Luxembourg, la figure V.20 représente le profil de concentration de LAMP1 dans la cohorte d'évaluation incluant les contrôles

confondants. L'évaluation par SRM confirme la dérégulation observée dans notre étude et valide sa généralisation aux contrôles urologiques mais uniquement chez les patients prévalents. De plus, on note que la décroissance de la concentration de LAMP1 est corrélée au facteur de risque. Ces résultats suggèrent que LAMP1 serait capable de discriminer les patients ayant déjà été opéré d'un cancer de la vessie, de patient prévalents pour qui la tumeur n'est pas confirmée par la cystoscopie. En d'autre terme LAMP1 pourrait être un biomarqueur diagnostique du cancer de la vessie chez les patients prévalents. Une analyse ELISA sur une plus grande cohorte indépendante permettrait de confirmer ces résultats. Ce type de biomarqueur est très important car il permettrait, s'il est validé, de faciliter le suivi des patients ayant subi une résection en limitant le nombre de cystoscopie de contrôle.

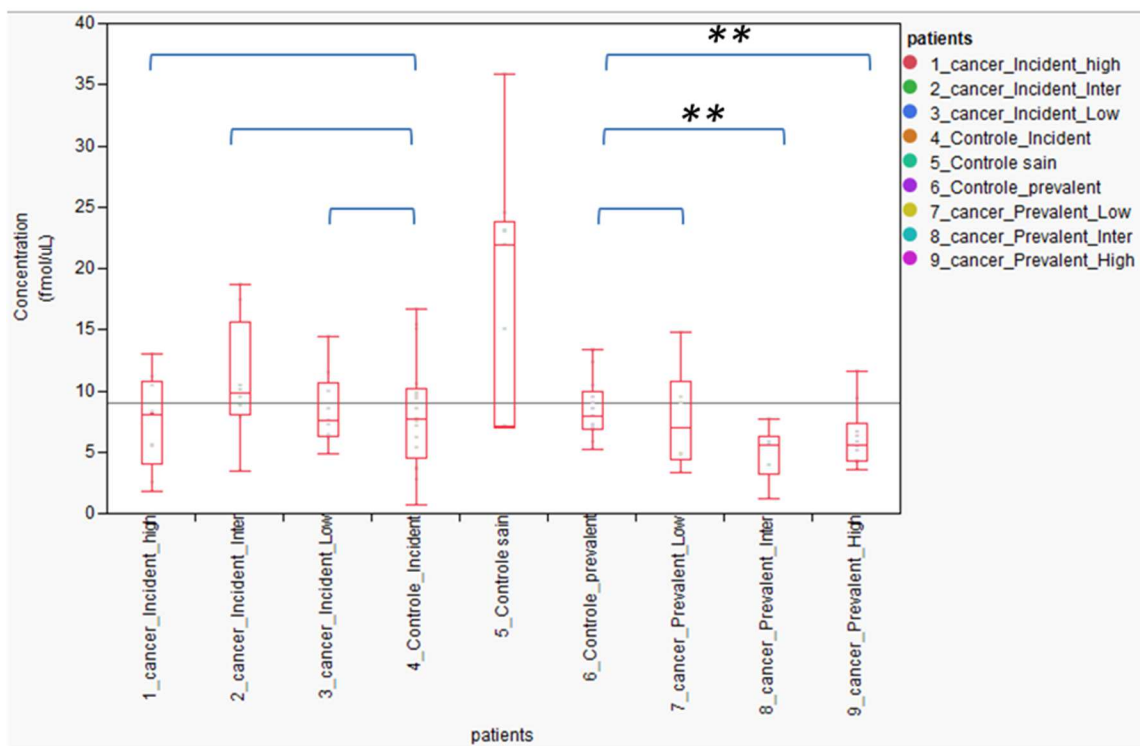


Figure V.20 Concentration de LAMP1 déterminée par l'analyse iSRM sur des urines de patients contrôles sains et urologiques et cancer incident et prévalent en fonction du risque de progression tumoral.

E. Conclusion

Lors de cette étude quantitative, nous avons analysé en triplicat les extraits urinaires de 30 contrôles et de 60 patients ségrégués par type de cancer (incident ou prévalent) et par facteur de risque de progression tumorale ou de récurrence. L'analyse statistique des données a permis de dégager un ensemble de 64 candidats surabondants chez les patients atteints de cancer et de 33 candidats dont l'abondance était diminuée chez les mêmes patients. Il est important de souligner qu'il s'agit d'une liste de candidats et non pas de biomarqueurs. Ceux-ci devront encore subir plusieurs évaluations avant leur potentielle validation finale. L'exploration de notre liste a mis en

évidence plusieurs caractéristiques intéressantes. Tout d'abord, ces protéines couvrent une large gamme d'abondances, ce qui est encourageant et démontre le fait que ces protéines ne sont pas uniquement des protéines majoritaires de l'urine. De plus, on note une forte proportion de protéines d'origine exosomales qui pourraient être sécrétées par les tumeurs. Il est important de noter que nous n'avons pas spécifiquement ciblés cette catégorie de protéines lors de cette étude, leur enrichissement a très probablement pour origine leur spécificité au processus pathologique ; En conclusion, ces résultats très encourageants mettent en lumière le potentiel que combine le protéome urinaire pour la détection précoce du cancer de la vessie.

Conclusion générale

Les avancées méthodologiques en protéomique permettent désormais d'accéder à une très grande variété de protéines dans des échantillons aussi complexes que les fluides biologiques, et ce dans des temps d'analyse toujours plus réduits. Ceci permet de considérer des applications dans le domaine de la santé humaine, où la variabilité inter-individuelle rend nécessaire la comparaison de grands nombres d'échantillons. Dans ce contexte, la recherche de biomarqueurs protéiques présente un intérêt clinique majeur pour le diagnostic, le pronostic ou le suivi thérapeutique de patients atteints de pathologies aussi diverses que les cancers, les maladies auto-immunes ou neurodégénératives.

Le diagnostic précoce du cancer de la vessie sans passer systématiquement par la cystoscopie est un enjeu de santé important, en particulier pour les patients suivis à la suite de la résection d'une première tumeur urothéliale. En raison de sa facilité d'accès et de la diversité de protéines qui s'y retrouvent, l'urine est un fluide particulièrement intéressant pour la recherche de biomarqueurs de pathologies uro-génitales. Ainsi, une simple analyse d'urine dans laquelle on pourrait suivre un ou plusieurs biomarqueurs permettrait d'améliorer de façon conséquente la prise en charge des patients pour lesquels on suspecte un cancer de la vessie. A l'heure actuelle, il n'existe pas de tels biomarqueurs qui soient capables d'allier une sensibilité et une spécificité permettant de s'affranchir de la cystoscopie. Nous nous sommes donc intéressés à l'urine dans le but de déterminer des candidats biomarqueurs de cancer de la vessie.

Au début de ces travaux, peu d'exemples de caractérisation exhaustive du protéome urinaire avaient été publiés, et encore moins qui concernaient la découverte de biomarqueurs de pathologies par des méthodes de protéomique à haut débit utilisant l'urine comme analyte. Bien que l'urine soit un ultrafiltrat du sang, ce fluide comporte des spécificités, comme une bien plus faible concentration protéique ou la présence de protéines majoritaires que l'on ne trouve pas dans le sang. Ainsi, les méthodologies développées pour le plasma ou le sérum n'étaient pas très bien adaptées. Nous avons donc développé un protocole expérimental couvrant l'ensemble du processus de découverte de nouveaux biomarqueurs dans l'urine : de la collecte et la préparation des échantillons, à l'évaluation de méthodes de fractionnement en vue d'aboutir à la caractérisation la plus exhaustive possible du protéome urinaire, et jusqu'à l'application de méthodologies de protéomique quantitative et de méthodes statistiques dédiées.

Ainsi, au cours de ces travaux, nous avons mis en place, en concertation avec nos partenaires cliniciens, un protocole de collecte spécifique répondant aux exigences de l'analyse protéomique et aux contraintes cliniques. Par ailleurs, nous avons optimisé, avec nos collègues en charge de l'évaluation des candidats biomarqueurs, un protocole d'extraction des protéines urinaires basé sur la précipitation TCA suivie d'une digestion en solution qui a démontré une bonne adéquation avec l'analyse protéomique par deux méthodes de quantification. De plus, nous avons caractérisé l'urine à partir de différentes méthodes de

préparation et de fractionnement, pour aboutir à une base de données contenant 2014 protéines urinaires non redondantes. Nous avons ensuite pu comparer ces méthodes de fractionnement ou de déplétion entre elles pour évaluer la meilleure stratégie d'exploration du protéome urinaire. Les résultats obtenus ont ainsi démontré que la méthode de fractionnement la plus favorable était l'isoélectrofocalisation hors gel. Cependant, nous avons également pu constater que la méthode Equalizer avait permis d'accéder, lors de l'étape de quantification, à des protéines minoritaires de l'urine, inaccessibles par d'autres méthodes. Enfin, l'analyse quantitative des extraits urinaires de 98 patients dont les 2/3 étaient atteints de cancer de la vessie, et le traitement statistique des données issues de cette analyse, ont permis de dégager un ensemble de 64 protéines candidates surabondantes chez les patients atteints de cancers et de 33 protéines candidates dont l'abondance était diminuée chez ces mêmes patients comparés aux contrôles. Ces biomarqueurs devront encore subir plusieurs évaluations avant leur éventuelle validation. Néanmoins, l'examen de cette liste a fait ressortir un enrichissement surprenant en protéines associées aux exosomes, dont certaines paraissent être sécrétées spécifiquement par la tumeur. Ces résultats ont démontré de façon claire le potentiel du protéome urinaire pour la détection du cancer de la vessie. De plus, il est important de noter que cette méthode pourrait également être appliquée à d'autres pathologies uro-génitales.

A l'heure actuelle, la principale difficulté pour la recherche de nouveaux biomarqueurs de pathologies est l'intégration de l'ensemble des données générées et leur mise en perspective avec d'autres résultats, qu'ils soient protéomiques ou autres. Ainsi, quand nous avons voulu comparer notre couverture du protéome urinaire à deux autres études les plus notables, nous nous sommes heurtés à des difficultés dans la mise en parallèle des listes d'identification, et l'issue de cette comparaison laisse plusieurs questions sans réponse. Ces résultats, ainsi que d'autres exemples de la littérature montrent les limites des approches de protéomique dites « bottom-up » basées sur l'inférence des protéines sur la base des peptides détectés après digestion protéolytique de l'échantillon. Ceci explique l'intérêt croissant des nouvelles approches de type « top down » dans lesquelles les protéines sont analysées dans leur intégralité sans digestion préalable. Dans le cadre de l'évaluation de nos biomarqueurs urinaire, ces approches pourraient jouer un rôle pour démontrer l'existence et déterminer la forme exacte (protéolyse, modifications) de certaines protéines d'intérêt.

Références Bibliographiques

- Abd El Gawad, I. A., Moussa, H. S., Nasr, M. I., El Gemae, E. H., Masooud, A. M., Ibrahim, I. K., & El Hifnawy, N. M. (2005). Comparative study of NMP-22, telomerase, and BTA in the detection of bladder cancer. *Journal of the Egyptian National Cancer Institute*, 17(3), 193–202. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16799657>
- Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., & Mann, M. (2006). The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biology*, 7(9), 1465–6914. doi:10.1186/gb-2006-7-9-R80
- Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H., Smith, R. D., ... Pounds, J. G. (2002). Toward a Human Blood Serum Proteome Analysis By Multidimensional Separation Coupled With Mass Spectrometry. *Molecular & Cellular Proteomics*, 1(12), 947–955. doi:10.1074/mcp.M200066-MCP200
- Anderson, N. G., Anderson, N. L., & Tollaksen, S. L. (1979). Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. *Clinical Chemistry*, 25(7), 1199–1210. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/455639>
- Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Molecular & Cellular Proteomics: MCP*, 1(11), 845–867. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12488461>
- Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., ... Lobley, A. (2004). The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular & Cellular Proteomics : MCP*, 3(4), 311–26. doi:10.1074/mcp.M300127-MCP200
- Babjuk, M., Oosterlinck, W., Sylvester, R., Kaasinen, E., Böhle, A., & Palou-Redorta, J. (2008). EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder. *European Urology*, 54(2), 303–14. doi:10.1016/j.eururo.2008.04.051
- Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4), 1017–31. doi:10.1007/s00216-007-1486-6
- Belcher, J. M., Sanyal, A. J., Peixoto, A. J., Perazella, M. a, Lim, J., Thiessen-Philbrook, H., ... Parikh, C. R. (2013). Kidney biomarkers and differential diagnosis of patients with cirrhosis and acute kidney injury. *Hepatology (Baltimore, Md.)*, 1–37. doi:10.1002/hep.26980
- Beynon, R. J., Doherty, M. K., Pratt, J. M., & Gaskell, S. J. (2005). Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods*, 2(8), 587–589. doi:10.1038/nmeth774
- Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Westermeier, R., & Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*, 6(4), 317–339. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7142660>

- Boja, E., Hiltke, T., Rivers, R., Kinsinger, C., Rahbar, A., Mesri, M., & Rodriguez, H. (2011). Evolution of clinical proteomics and its role in medicine. *Journal of Proteome Research*, 10(1), 66–84. doi:10.1021/pr100532g
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72, 248–254. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/942051>
- Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., ... Garin, J. (2007). Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & Cellular Proteomics : MCP*, 6(12), 2139–49. doi:10.1074/mcp.M700163-MCP200
- Brun, V., Masselon, C., Garin, J., & Dupuis, A. (2009). Isotope dilution strategies for absolute quantitative proteomics. *Journal of Proteomics*, 72(5), 740–749. doi:10.1016/j.jprot.2009.03.007
- Castagna, A., Cecconi, D., Sennels, L., Rappsilber, J., Guerrier, L., Fortis, F., ... Righetti, P. G. (2005). Exploring the Hidden Human Urinary Proteome via Ligand Library Beads research articles, 1917–1930.
- Ceccarelli, S., Visco, V., Raffa, S., Wakisaka, N., Pagano, J. S., & Torrisi, M. R. (2007). Epstein-Barr virus latent membrane protein 1 promotes concentration in multivesicular bodies of fibroblast growth factor 2 and its release through exosomes. *International Journal of Cancer. Journal International Du Cancer*, 121(7), 1494–506. doi:10.1002/ijc.22844
- Celis, J. E., Ostergaard, M., Basse, B., Celis, a, Lauridsen, J. B., Ratz, G. P., ... Rasmussen, H. H. (1996). Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Research*, 56(20), 4782–90. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8840999>
- Chen, Y.-T., Chen, C.-L., Chen, H.-W., Chung, T., Wu, C.-C., Chen, C.-D., ... Yu, J.-S. (2010). Discovery of Novel Bladder Cancer Biomarkers by Comparative Urine Proteomics Using iTRAQ Technology. *Journal of Proteome Research*, 9(11), 5803–5815. doi:10.1021/pr100576x
- Chutipongtanate, S., Watcharatanyatip, K., Homvises, T., Jaturongkakul, K., & Thongboonkerd, V. (2012). Systematic comparisons of various spectrophotometric and colorimetric methods to measure concentrations of protein, peptide and amino acid: detectable limits, linear dynamic ranges, interferences, practicality and unit costs. *Talanta*, 98, 123–129. doi:10.1016/j.talanta.2012.06.058
- Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolić, L., & Smith, R. D. (2000). Utility of accurate mass tags for proteome-wide protein identification. *Analytical Chemistry*, 72(14), 3349–3354. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10939410>
- Dos Santos, A., Court, M., Thiers, V., Sar, S., Guettier, C., Samuel, D., ... Masselon, C. D. (2010). Identification of cellular targets in human intrahepatic cholangiocarcinoma using laser microdissection and accurate mass and time tag proteomics. *Molecular & Cellular Proteomics : MCP*, 9(9), 1991–2004. doi:10.1074/mcp.M110.000026
- Dupierriis, V., Masselon, C., Court, M., Kieffer-Jaquinod, S., & Bruley, C. (2009). A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. *Bioinformatics (Oxford, England)*, 25(15), 1980–1981. doi:10.1093/bioinformatics/btp301

- Dupuis, A., Hennekinne, J.-A., Garin, J., & Brun, V. (2008). Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*, 8(22), 4633–4636. doi:10.1002/pmic.200800326
- EDMAN, P. (1949). A method for the determination of amino acid sequence in peptides. *Archives of Biochemistry*, 22(3), 475. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18134557>
- Edmondson, R. D., Vondriska, T. M., Biederman, K. J., Zhang, J., Jones, R. C., Zheng, Y., ... Ping, P. (2002). Protein kinase C epsilon signaling complexes include metabolism- and transcription/translation-related proteins: complimentary separation techniques with LC/MS/MS. *Molecular & Cellular Proteomics: MCP*, 1(6), 421–433. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12169683>
- Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, 4(3), 207–214. doi:10.1038/nmeth1019
- Fan, J., Tam, P., Vande Woude, G., & Ren, Y. (2004). Normalization and analysis of cDNA microarrays using within-array replications applied to neuroblastoma cell response to a cytokine. *Proceedings of the National Academy of Sciences of the United States of America*, 101(5), 1135–40. doi:10.1073/pnas.0307557100
- Fehrenbacher, N., Bastholm, L., Kirkegaard-Sørensen, T., Rafn, B., Bøttzauw, T., Nielsen, C., ... Jäättelä, M. (2008). Sensitization to the lysosomal cell death pathway by oncogene-induced down-regulation of lysosome-associated membrane proteins 1 and 2. *Cancer Research*, 68(16), 6623–33. doi:10.1158/0008-5472.CAN-08-0463
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.)*, 246(4926), 64–71. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2675315>
- Ferlay, J., Parkin, D. M., & Steliarova-Foucher, E. (2010). Estimates of cancer incidence and mortality in Europe in 2008. *European Journal of Cancer (Oxford, England: 1990)*, 46(4), 765–781. doi:10.1016/j.ejca.2009.12.014
- Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., ... Rolland, N. (2010). AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Molecular & Cellular Proteomics: MCP*, 9(6), 1063–84. doi:10.1074/mcp.M900325-MCP200
- Fu, X., Gharib, S. a, Green, P. S., Aitken, M. L., Frazer, D. a, Park, D. R., ... Heinecke, J. W. (2008). Spectral index for assessment of differential protein expression in shotgun proteomics. *Journal of Proteome Research*, 7(3), 845–54. doi:10.1021/pr070271+
- Fung, E. T. (2010). A recipe for proteomics diagnostic test development: the OVA1 test, from biomarker discovery to FDA clearance. *Clinical Chemistry*, 56(2), 327–9. doi:10.1373/clinchem.2009.140855
- Gaudet, P., Argoud-Puy, G., Cusin, I., Duek, P., Evalet, O., Gateau, A., ... Lane, L. (2013). neXtProt: Organizing Protein Knowledge in the Context of Human Proteome Projects. *Journal of Proteome Research*, 12(1), 293–298. doi:10.1021/pr300830v
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., & 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), 6940–6945. doi:10.1073/pnas.0832254100

- Griss, J., Martín, M., O'Donovan, C., Apweiler, R., Hermjakob, H., & Vizcaíno, J. A. (2011). Consequences of the discontinuation of the International Protein Index (IPI) database and its substitution by the UniProtKB “complete proteome” sets. *Proteomics*, *11*(22), 4434–8. doi:10.1002/pmic.201100363
- Guha, N., Steenland, N. K., Merletti, F., Altieri, A., Cogliano, V., & Straif, K. (2010). Bladder cancer risk in painters: a meta-analysis. *Occupational and Environmental Medicine*, *67*(8), 568–73. doi:10.1136/oem.2009.051565
- Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, *17*(10), 994–999. doi:10.1038/13690
- Havanapan, P., & Thongboonkerd, V. (2009). Are Protease Inhibitors Required for Gel-Based Proteomics of Kidney and Urine ? research articles, 3109–3117.
- Henderson, M. C., & Azorsa, D. O. (2012). The genomic and proteomic content of cancer cell-derived exosomes. *Frontiers in Oncology*, *2*(April), 38. doi:10.3389/fonc.2012.00038
- Higgins, J., Kaygusuz, G., Wang, L., Montgomery, K., Mason, V., Zhu, S., ... Brooks, J. (2007). Placental S100 (S100P) and GATA3: markers for transitional epithelium and urothelial carcinoma discovered by complementary DNA microarray. *The American Journal Surgical Pathology*, *5*(31), 673–680.
- Hubner, N. C., Ren, S., & Mann, M. (2008). Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *PROTEOMICS*, *8*(23-24), 4862–4872. doi:10.1002/pmic.200800351
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. doi:10.1038/nature03001
- Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., & 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: MCP*, *4*(9), 1265–1272. doi:10.1074/mcp.M500061-MCP200
- Iwaki, H., Kageyama, S., Isono, T., Wakabayashi, Y., Okada, Y., Yoshimura, K., ... Yoshiki, T. (2004). Diagnostic potential in bladder cancer of a panel of tumor markers (calreticulin, γ -synuclein, and proteomic analysis, 95(12), 955–961.
- Jacobs, B. L., Lee, C. T., & Montie, J. E. (2010). Bladder cancer in 2010: how far have we come? *CA: A Cancer Journal for Clinicians*, *60*(4), 244–272. doi:10.3322/caac.20077
- Jemal, A., Siegel, R., Xu, J., & Ward, E. (2010). Cancer statistics, 2010. *CA: A Cancer Journal for Clinicians*, *60*(5), 277–300. doi:10.3322/caac.20073
- Jeppesen, D. K., Nawrocki, A., Jensen, S. G., Thorsen, K., Whitehead, B., Howard, K. a., ... Ostensfeld, M. S. (2014). Quantitative proteomics of fractionated membrane and lumen exosome proteins from isogenic metastatic and nonmetastatic bladder cancer cells reveal differential expression of EMT factors. *Proteomics*, *14*(6), 699–712. doi:10.1002/pmic.201300452
- Kalantari, S., Rutishauser, D., Samavat, S., Nafar, M., Mahmudieh, L., Rezaei-Tavirani, M., & Zubarev, R. a. (2013). Urinary prognostic biomarkers and classification of IgA nephropathy by high resolution mass spectrometry coupled with liquid chromatography. *PloS One*, *8*(12), e80830. doi:10.1371/journal.pone.0080830

- Karas, M., & Hillenkamp, F. (1988). Laser Desorption Ionization of Proteins with Molecular Masses I ! *Analytical Chemistry*, 2301(29), 2299–2301.
- Kentsis, A., Monigatti, F., Dorff, K., Campagne, F., Bachur, R., & Steen, H. (2009). Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics. Clinical Applications*, 3(9), 1052–1061. doi:10.1002/prca.200900008
- Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, 26(3), 231–243. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1093965>
- Knowles, M. A. (2008). Molecular pathogenesis of bladder cancer. *International Journal of Clinical Oncology*, 13(4), 287–297. doi:10.1007/s10147-008-0812-0
- Kraut, A., Marcellin, M., Adrait, A., Kuhn, L., Louwagie, M., Kieffer-Jaquinod, S., ... Gallagher-Gambarelli, M. (2009). Peptide storage: are you getting the best return on your investment? Defining optimal storage conditions for proteomics samples. *Journal of Proteome Research*, 8(7), 3778–85. doi:10.1021/pr900095u
- Lambert, D., Berrahmoune, H., Herbeth, B., Siest, G., & Visvikis-Siest, S. (2007). [Preanalytical variations of proteic biomarkers]. *Médecine Sciences: M/S, 23 Spec No*, 9–12. doi:10.1051/medsci/2007231s9
- Lindén, M., Lind, S. B., Mayrhofer, C., Segersten, U., Wester, K., Lyutvinskiy, Y., ... Pettersson, U. (2012). Proteomic analysis of urinary biomarker candidates for nonmuscle invasive bladder cancer. *Proteomics*, 12(1), 135–44. doi:10.1002/pmic.201000810
- Link, a J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., ... Yates, J. R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology*, 17(7), 676–82. doi:10.1038/10890
- Lipton, M. S., Pasa-Tolic', L., Anderson, G. A., Anderson, D. J., Auberry, D. L., Battista, J. R., ... Smith, R. D. (2002). Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17), 11049–11054. doi:10.1073/pnas.172170199
- Liu, H., Lin, D., & Yates, J. R. (2002). Multidimensional separations for protein/peptide analysis in the post-genomic era. *Biotechniques*, 32(4), 898–902.
- Liu, H., Sadygov, R. G., & Yates, J. R. (2004). A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics proteolytic digestion and liquid chromatography in com-, 76(14), 4193–4201.
- Louie, B., Higdon, R., & Kolker, E. (2010). The necessity of adjusting tests of protein category enrichment in discovery proteomics. *Bioinformatics (Oxford, England)*, 26(24), 3007–11. doi:10.1093/bioinformatics/btq541
- Makarov, a. (2000). Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical Chemistry*, 72(6), 1156–62. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10740853>
- Marimuthu, A., O'Meally, R. N., Chaerkady, R., Subbannayya, Y., Nanjappa, V., Kumar, P., ... Pandey, A. (2011). A comprehensive map of the human urinary proteome. *Journal of Proteome Research*, 10(6), 2734–2743. doi:10.1021/pr2003038

- Marshall, J., Kupchak, P., Zhu, W., Yantha, J., Vrees, T., Furesz, S., ... Jackowski, G. (2003). Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. *Journal of Proteome Research*, 2(4), 361–372. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12938926>
- Masselon, C. D., Kieffer-Jaquinod, S., Brugière, S., Dupierris, V., & Garin, J. (2008). Influence of mass resolution on species matching in accurate mass and retention time (AMT) tag proteomics experiments. *Rapid Communications in Mass Spectrometry: RCM*, 22(7), 986–992. doi:10.1002/rcm.3447
- Mathivanan, S., Ji, H., & Simpson, R. J. (2010). Exosomes: extracellular organelles important in intercellular communication. *Journal of Proteomics*, 73(10), 1907–20. doi:10.1016/j.jprot.2010.06.006
- Mouton-Barbosa, E., Roux-Dalvai, F., Bouyssié, D., Berger, F., Schmidt, E., Righetti, P. G., ... Gonzalez de Peredo, A. (2010). In-depth exploration of cerebrospinal fluid by combining peptide ligand library treatment and label-free protein quantification. *Molecular & Cellular Proteomics : MCP*, 9(5), 1006–21. doi:10.1074/mcp.M900513-MCP200
- Nagaraj, N., & Mann, M. (2011). Quantitative Analysis of the Intra- and Inter-Individual Variability of the Normal Urinary Proteome research articles. *Journal of Proteome Research*, 10, 637–645.
- Nakamura, K., Kasraeian, A., Iczkowski, K. A., Chang, M., Pendleton, J., Anai, S., & Rosser, C. J. (2009). Utility of serial urinary cytology in the initial evaluation of the patient with microscopic hematuria. *BMC Urology*, 9, 12. doi:10.1186/1471-2490-9-12
- NL. Kelleher. (2004). Top-down proteomics. *Analytical Chemistry*, 11(76), 197A–203–A.
- Noto, A., Cibecchini, F., Fanos, V., & Mussap, M. (2013). NGAL and Metabolomics : The Single Biomarker to Reveal the Metabolome Alterations in Kidney Injury, 2013.
- O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry*, 250(10), 4007–4021. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/236308>
- Oda, Y., Huang, K., Cross, F. R., Cowburn, D., & Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences*, 96(12), 6591–6596. doi:10.1073/pnas.96.12.6591
- Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., ... Ahn, N. G. (2005). Comparison of Label-free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Molecular & Cellular Proteomics*, 4(10), 1487–1502. doi:10.1074/mcp.M500084-MCP200
- Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & 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. doi:10.1074/mcp.M200025-MCP200
- Ong, S.-E., & Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology*, 1(5), 252–262. doi:10.1038/nchembio736
- Palou, J., Algaba, F., Vera, I., Rodriguez, O., Villavicencio, H., & Sanchez-Carbayo, M. (2009). Protein expression patterns of ezrin are predictors of progression in T1G3 bladder tumours treated with nonmaintenance bacillus Calmette-Guérin. *European Urology*, 56(5), 829–36. doi:10.1016/j.eururo.2008.09.062

- Petritis, K., Kangas, L. J., Yan, B., Strittmatter, E. F., Monroe, M., Qian, W., ... Smith, R. D. (2006). Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. *Analytical Chemistry*, 78(14), 5026–5039.
- Pieper, R., Gatlin, C. L., McGrath, A. M., Makusky, A. J., Mondal, M., Seonarain, M., ... Steiner, S. (2004). Characterization of the human urinary proteome: a method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics*, 4(4), 1159–1174. doi:10.1002/pmic.200300661
- Pieper, R., Su, Q., Gatlin, C. L., Huang, S.-T., Anderson, N. L., & Steiner, S. (2003). Multi-component immunoaffinity subtraction chromatography: An innovative step towards a comprehensive survey of the human plasma proteome. *PROTEOMICS*, 3(4), 422–432. doi:10.1002/pmic.200390057
- Pisitkun, T., Shen, R.-F., & Knepper, M. A. (2004). Identification and proteomic profiling of exosomes in human urine. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36), 13368–13373. doi:10.1073/pnas.0403453101
- Rai, A. J., Gelfand, C. A., Haywood, B. C., Warunek, D. J., Yi, J., Schuchard, M. D., ... Chan, D. W. (2005). HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*, 5(13), 3262–3277. doi:10.1002/pmic.200401245
- Rajalingam, D., Loftis, C., Xu, J. J., & Kumar, T. K. S. (2009). Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate. *Protein Science : A Publication of the Protein Society*, 18(5), 980–93. doi:10.1002/pro.108
- Rampoldi, L., Scolari, F., Amoroso, A., Ghiggeri, G., & Devuyst, O. (2011). The rediscovery of uromodulin (Tamm-Horsfall protein): from tubulointerstitial nephropathy to chronic kidney disease. *Kidney International*, 80(4), 338–47. doi:10.1038/ki.2011.134
- Rifai, N., Gillette, M. A., & Carr, S. A. (2006). Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology*, 24(8), 971–983. doi:10.1038/nbt1235
- Righetti, P. G., Boschetti, E., Lomas, L., & Citterio, A. (2006). Protein Equalizer™ Technology : The quest for a “democratic proteome.” *PROTEOMICS*, 6(14), 3980–3992. doi:10.1002/pmic.200500904
- Righetti, P. G., Castagna, A., Antonioli, P., & Boschetti, E. (2005). Prefractionation techniques in proteome analysis: The mining tools of the third millennium. *ELECTROPHORESIS*, 26(2), 297–319. doi:10.1002/elps.200406189
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., ... Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics: MCP*, 3(12), 1154–1169. doi:10.1074/mcp.M400129-MCP200
- Scheele, G. A. (1975). Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins. *The Journal of Biological Chemistry*, 250(14), 5375–5385. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1141235>
- Shukla, H. D., Vaitiekunas, P., & Cotter, R. J. (2012). Advances in membrane proteomics and cancer biomarker discovery: current status and future perspective. *Proteomics*, 12(19-20), 3085–3104. doi:10.1002/pmic.201100519

- Slebos, R. J. C., Brock, J. W. C., Winters, N. F., Stuart, S. R., Martinez, M. A., Li, M., ... Liebler, D. C. (2009). NIH Public Access, 7(12), 5286–5294. doi:10.1021/pr8004666.Evaluation
- Stempfer, R., Kubicek, M., Lang, I. M., Christa, N., & Gerner, C. (2008). Quantitative assessment of human serum high-abundance protein depletion. *Electrophoresis*, 29(21), 4316–23. doi:10.1002/elps.200800211
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., & 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. doi:10.1002/rcm.1290020802
- Théry, C., Ostrowski, M., & Segura, E. (2009). Membrane vesicles as conveyors of immune responses. *Nature Reviews. Immunology*, 9(8), 581–93. doi:10.1038/nri2567
- Théry, C., Zitvogel, L., & Amigorena, S. (2002). Exosomes: composition, biogenesis and function. *Nature Reviews. Immunology*, 2(8), 569–79. doi:10.1038/nri855
- Thomas, L., Leyh, H., Marberger, M., Bombardieri, E., Bassi, P., Pagano, F., ... Ishak, L. (1999). Multicenter trial of the quantitative BTA TRAK assay in the detection of bladder cancer. *Clinical Chemistry*, 45(4), 472–477. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10102906>
- Thongboonkerd, V. (2007). Practical Points in Urinary Proteomics. *Journal of Proteome Research*, 6(10), 3881–3890. doi:10.1021/pr070328s
- Thulasiraman, V., Lin, S., Gheorghiu, L., Lathrop, J., Lomas, L., Hammond, D., & Boschetti, E. (2005). Reduction of the concentration difference of proteins in biological liquids using a library of combinatorial ligands. *ELECTROPHORESIS*, 26(18), 3561–3571. doi:10.1002/elps.200500147
- Van Rhijn, B. W. G., van der Poel, H. G., & van der Kwast, T. H. (2005). Urine Markers for Bladder Cancer Surveillance: A Systematic Review. *European Urology*, 47(6), 736–748. doi:10.1016/j.eururo.2005.03.014
- Vinobha, C. S., Rajasekaran, E., & Rajadurai, M. (2010). Compositional analysis of protein sequences of different species. *International Journal of Bioinformatics Research*, 2(2), 7–11.
- Wang, Z., Hill, S., Luther, J. M., Hachey, D. L., & Schey, K. L. (2012). Proteomic analysis of urine exosomes by multidimensional protein identification technology (MudPIT). *Proteomics*, 12(2), 329–38. doi:10.1002/pmic.201100477
- Wolters, D., Washburn, M., & Yates, J. R. (2001). An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry*, 73(7), 5683–90.
- Yue, J., Huhn, S., & Shen, Z. (2013). Complex roles of filamin-A mediated cytoskeleton network in cancer progression. *Cell & Bioscience*, 3(1), 7. doi:10.1186/2045-3701-3-7
- Zhou, H., Pisitkun, T., Aponte, A., Yuen, P. S. T., Hoffert, J. D., Hu, X., ... Star, A. (2008). Exosomal Fetuin-A identified by proteomics: a novel urinary biomarker for detecting acute kidney injury. *Kidney International*, 70(10), 1847–1857.

Le cancer de la vessie représente le 4^{ième} cancer de l'homme en Europe. Dans la majorité des cas, les primo-tumeurs sont traitées facilement par résection mais dans 60% des cas, il y a récurrences sous formes plus agressives. Il est donc nécessaire de détecter au plus tôt ces récurrences. A ce jour, l'examen de référence pour le diagnostic d'un cancer de la vessie est la cystoscopie. Cet examen permet d'examiner l'intérieur de la vessie à l'aide d'un système optique introduit via l'urètre. Cette méthode est spécifique et sensible mais inconfortable et invasive pour le patient. Il est donc important de trouver de nouvelles méthodes de diagnostic de tumeurs de la vessie et, de surveillance aussi sensibles et spécifiques pour réduire l'inconfort chez le patient et par la même occasion, le coût associés à la cystoscopie. La recherche de biomarqueurs protéiques dans l'urine représente une alternative majeure pour le diagnostic, le pronostic et le suivi thérapeutique de patients atteints de pathologies uro-génitales. Dans le cas présent du cancer de la vessie, la vessie joue le rôle de réservoir de cellules relarguées par la tumeur, ce qui fait de l'urine, fluide dit "proximal", le fluide idéal pour la recherche de biomarqueurs protéiques. Ces dernières années, la recherche de biomarqueurs protéiques a bénéficié de progrès spectaculaires dans le domaine de la spectrométrie de masse et de la biochimie, permettant d'accéder à une large variété de protéines sur une large gamme dynamique de concentration. De plus, l'apparition de nouvelles approches de protéomique quantitatives, telle que la stratégie "Accurate Mass and Time (AMT) tags" couplée à la quantification « label free », permet la comparaison de protéomes d'états physiologiques distincts par mesures d'abondances des protéines. Dans le cadre de ce travail de thèse, nous avons ainsi développé un protocole expérimental couvrant l'ensemble du processus de découverte de nouveaux biomarqueurs associés au cancer de la vessie dans l'urine, *i.e.*, de la collecte et préparation des échantillons d'urine à l'évaluation de méthodes de fractionnement pour la caractérisation la plus exhaustive possible du protéome urinaire, en utilisant des méthodes de protéomiques quantitatives et des méthodes de statistiques dédiées. Ce travail nous a permis de constituer une base de données contenant 2014 protéines urinaires. Des variations d'abondances ont été mesurées pour plus de la moitié d'entre elles, au travers d'une cohorte de 98 patients constituée de patients atteints de cancers de la vessie et de patients contrôles. Une liste finale composée de 97 candidats biomarqueurs du cancer de la vessie a été établit. Cette liste contient un grand nombre de protéines exosomales potentiellement sécrétées de façon spécifique par la tumeur.

Bladder cancer is the 4th type of cancer causing man death in Europe. In most cases, primary tumors can be easily removed by resection but in 60% of the cases, the tumors regrowth in more aggressive forms. Therefore, it is essential to detect early recurrence of bladder cancer. To date, the gold standard for the diagnosis of bladder cancer is cystoscopy. It allows the examination of the inside of the bladder using an optical system which is inserted in the urethra. This method is sensitive and specific but extremely uncomfortable and invasive for the patient. Therefore, it is crucial to find new diagnosis and monitoring methods for bladder tumour more comfortable for the patient, in a cost-efficient way. The search for clinically useful protein biomarkers in the urine is a major alternative for the diagnosis, the prognostic and the therapeutic treatment of patients with urea-genital pathologies. In the specific case of the bladder cancer, the bladder contains the cells left by the tumour, and subsequently urine becomes the ideal fluid for biomarker investigation as a “proximal fluid”. In the last few years, the search for protein biomarkers has benefited of significant progress in the field of mass spectrometry and bio-chemistry, allowing the detection of a wide variety of proteins in a large dynamic range of concentrations. In addition, new approaches of quantitative proteomic, such as the “Accurate Mass and Time (AMT) tags” approach, coupled with the “label free” quantification, allows the comparison of proteomes from distinct physiological state by measuring protein abundances. During this thesis, we developed an experimental protocol covering the whole process of discovering new biomarkers associated to bladder cancer in urines, *i.e.*, from the collection and preparation of urine samples, to the evaluation of the best fractioning method to define the urinary proteome using quantitative approaches and dedicated statistical methodologies. This work enables the population of a database containing 2014 urinary proteins. Abundance variations were measured for more than the half through a cohort of 98 healthy and bladder cancer patients. A final list of 97 biomarker candidates has been established. This list holds a significant number of exosomal proteins that are potentially secreted by the tumour.