



HAL
open science

Convergence entre l'analyse biostatistique et les méthodes d'inversion hiérarchique bayésienne pour la recherche et la validation de biomarqueurs par spectrométrie de masse

Pierre Grangeat, Jean-François Giovannelli, Pascal Roy, Vincent Picaud, Caroline Truntzer, Jérôme Lemoine, Pierre Mahé, Jean-Philippe Charrier, Laurent Gerfault, Pascal Szacherski, et al.

► To cite this version:

Pierre Grangeat, Jean-François Giovannelli, Pascal Roy, Vincent Picaud, Caroline Truntzer, et al.. Convergence entre l'analyse biostatistique et les méthodes d'inversion hiérarchique bayésienne pour la recherche et la validation de biomarqueurs par spectrométrie de masse. 24^e Colloque sur le traitement du signal et des images, SFGBM (Société Française du génie biologique et médical), Sep 2013, Brest, France. cea-03996401

HAL Id: cea-03996401

<https://cea.hal.science/cea-03996401v1>

Submitted on 19 Feb 2023

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.

Convergence entre l'analyse biostatistique et les méthodes d'inversion hiérarchique bayésienne pour la recherche et la validation de biomarqueurs par spectrométrie de masse

PIERRE GRANGEAT¹, JEAN-FRANÇOIS GIOVANNELLI², PASCAL ROY³, VINCENT PICAUD⁴, CAROLINE TRUNTZER⁵, JEROME LEMOINE⁶, PIERRE MAHE⁷, JEAN-PHILIPPE CHARRIER⁸, LAURENT GERFAULT¹, PASCAL SZACHERSKI^{1,2}, AUDREY GIREMUS², NOURA DRIDI², AMNA KLICH³, CATHERINE MERCIER³, DELPHINE MAUCORT-BOULCH³, OLIVIER GAL⁴, PATRICK DUCOROY⁵, ARNAUD SALVADOR⁶, TANGUY FORTIN⁸, GENEVIEVE CHOQUET-KASTYLEVSKY⁸, BRUNO LACROIX⁸

¹CEA Leti, MINATEC Campus, DTBS, 17 rue des Martyrs, F-38054 Grenoble cedex 9, France

²Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France

³Hospices Civils de Lyon, Service de Biostatistique, Université Lyon I, CNRS UMR 5558, 162 avenue Lacassagne, F-69424 Lyon, France

⁴CEA, LIST, DCSI, 91191 Gif-sur-Yvette cedex, France

⁵Plateforme protéomique CLIPP, CHU Dijon, 1 rue du professeur Marion, F- 21000 Dijon, France

⁶Institut des Sciences Analytiques, UMR 5280, CNRS / Université de Lyon, Lyon-1, 5 rue de la Doua, F-69100 Villeurbanne, France

⁷bioMérieux, Parc Polytec, 5 rue des Berges, F-38000 Grenoble, France

⁸bioMérieux, 376 Chemin de l'Orme, F-69280 Marcy l'Etoile, France

¹pierre.grangeat@cea.fr, laurent.gerfault@cea.fr, pszacherski@gmail.com,

²giova@ims-bordeaux.fr, audrey.giremus@ims-bordeaux.fr, noura.dridi@ims-bordeaux.fr,

³pascal.roy@chu-lyon.fr, amna.klich@chu-lyon.fr, catherine.mercier@chu-lyon.fr, delphine.maucort-boulch@chu-lyon.fr,

⁴vincent.picaud@cea.fr, olivier.gal@cea.fr,

⁵caroline.truntzer@clipproteomic.fr, patrick.ducoroy@clipproteomic.fr,

⁶jerome.lemoine@univ-lyon1.fr, arnaud.salvador@univ-lyon1.fr,

⁷pierre.mahe@biomerieux.com,

⁸jean-philippe.charrier@eu.biomerieux.com, tanguy.fortin@biomerieux.com, genevieve.choquet@biomerieux.com,

bruno.lacroix@eu.biomerieux.com

Résumé – Dans cette communication, nous présentons une vue d'ensemble des méthodes d'analyse biostatistique et d'inversion hiérarchique Bayésienne que nous étudions sur le projet BHI-PRO pour la recherche et la validation de biomarqueurs par spectrométrie de masse en mode MALDI et MRM. Nous avons aussi explicité les critères de performance utilisés et présenter des premiers résultats expérimentaux. En mode MRM, nous avons montré sur une cohorte de cancer colorectal, en utilisant la LFABP et la PDI comme biomarqueurs, une amélioration significative de la sensibilité pour une spécificité correspondant aux performances demandées pour un test de dépistage.

Abstract – In this communication, we are presenting an overview on the methods in biostatistical analysis and Bayesian hierarchical inversion we are studying on the BHI-PRO project for research and validation of biomarkers based on mass spectrometry in MALDI and MRM mode. We are also expliciting the performance criteria used and giving some first experimental results. In MRM mode, we are showing on a colorectal cancer cohort using the LFABP and PDI biomarkers a relevant improvement of the sensibility for a specificity satisfying the minimum requirement for a screening test.

1 Introduction

Des efforts de recherche importants sont consacrés au niveau mondial pour développer des chaînes d'analyse reposant sur la spectrométrie de masse pour la découverte, la validation et la quantification de biomarqueurs protéiques dans des matrices complexes comme l'urine ou le sang. Cependant, maîtriser la variabilité technologique sur ces chaînes d'analyse est un point critique. Ceci nécessite de développer des techniques de traitement de l'information adaptées pour prendre en compte la complexité du mélange analysé, pour améliorer la fiabilité des mesures et pour faciliter l'usage de ces technologies.

Une chaîne d'analyse protéomique est un enchaînement de traitements moléculaires qui peut être décrit par une structure de graphe hiérarchique, chaque nœud représentant un niveau d'analyse dans la chaîne [1-3]. Chaque branche correspond à

une décomposition moléculaire définissant un modèle de mélange hiérarchique. Dans ce projet BHI-PRO (ANR 2010 BLAN 0313), nous proposons d'introduire des modèles hiérarchiques dédiés pour décrire les chaînes d'analyse MALDI et MRM. Les nouveaux algorithmes d'inversion hiérarchique bayésiens reposent sur deux innovations : les associations protéomique - problèmes inverses bayésiens - échantillonnage stochastique d'une part et la convergence biostatistique - problèmes inverses d'autre part.

Les enjeux scientifiques et économiques sont d'améliorer l'efficacité des phases de découverte et de validation de biomarqueurs protéiques. Nous proposons de contrôler la variabilité technologique sur ces chaînes d'analyse grâce à des algorithmes d'inversion Bayésiens adaptatifs innovant pour retrouver la concentration des protéines et le statut clinique des individus.

2 Convergence entre l'analyse biostatistique et les problèmes inverses

2.1 L'analyse biostatistique pour la recherche et la validation de biomarqueurs

Pour définir le problème, nous supposons que nous disposons de U prélèvements sanguins indicés par u , et d'une liste de P protéines candidates, indicées par p . Pour chaque protéine d'indice p dans l'échantillon d'indice u , nous supposons qu'il existe un instrument qui délivre une mesure $x_{p,u}$ représentative de la concentration de cette protéine. Le profil moléculaire d'un individu u est le vecteur regroupant ces concentrations. Nous définissons deux étiquettes b_u et d_p représentant respectivement l'état de l'individu u , sain ou pathologique, et de la protéine p , suivant sa contribution au pouvoir de discrimination pour séparer les classes de profils moléculaires. L'objectif de l'analyse biostatistique pour la recherche et la validation de biomarqueurs est de proposer des méthodes pour estimer ces étiquettes, connaissant ces profils moléculaires, après un apprentissage statistique des paramètres des classes, et d'en caractériser les performances. Ceci revient à étudier une double classification sur les protéines et sur les individus.

Pour répondre à cette problématique dans le cadre des données de grande dimension (i.e. nombre de protéines candidates supérieur au nombre d'individus inclus dans l'étude), une approche classique est de construire des modèles multivariés permettant d'estimer l'ampleur d'effet de chacune des protéines candidates et de sélectionner simultanément celles dont l'ampleur d'effet est significativement différente de 0. Les modèles obtenus sont dits parcimonieux ; le LASSO en étant un exemple.

L'apprentissage statistique permet d'estimer les amplitudes d'effet des protéines candidates au travers des coefficients qui leur sont affectés dans le modèle multivarié. Cette étape est effectuée sur un jeu de données dit d'apprentissage. Un jeu de données indépendant dit de validation est utilisé pour caractériser ensuite les performances du modèle estimé, et ainsi permettre de le valider.

Différents critères peuvent être appliqués pour caractériser ces performances. Notamment, pour la classification sur les individus, il est d'usage de se baser sur le taux d'erreur obtenu sur le jeu de données de validation, notamment la sensibilité Se , la probabilité que le statut estimé soit malade si la maladie est présente, et la spécificité Sp , la probabilité que le statut estimé soit sain si la maladie est absente. Pour l'étude de biomarqueurs, un critère important est le FDR (False Discovery Rate) correspondant à la probabilité de déclarer discriminant un biomarqueur qui ne le serait pas.

2.2 L'approche problème inverse pour quantifier et classifier

Le défi dans notre cas est double. D'une part, la chaîne d'analyse par spectrométrie de masse ne délivre pas directement la concentration de la protéine, mais un signal sous forme de spectre représentant une mesure indirecte, et d'autre part, elle fournit un ensemble multidimensionnel de signaux. Par ailleurs, les différentes interactions qui interviennent sur la protéine induisent une variabilité technologique sur les signaux délivrés. L'approche standard consiste à extraire un paramètre représentatif du pic caractéristique de chaque signal, comme l'aire sous le pic, et à calculer un paramètre global au niveau de la protéine, en prenant la valeur moyenne ou la valeur médiane des paramètres au niveau des spectres élémentaires. Sur le projet

BHI-PRO, nous étudions une approche alternative où nous reformulons cette problématique de recherche et validation de biomarqueurs comme un problème d'inversion hiérarchique bayésien associé à deux modélisations de la chaîne d'analyse par spectrométrie de masse en modes MALDI et MRM.

2.2.1 Approche MALDI en mode découverte et méthode d'inversion hiérarchique bayésienne associée

La chaîne d'acquisition MALDI comporte principalement deux étapes. Une première étape appelée traitement pré-analytique qui consiste à simplifier les échantillons pour ne retenir que certains peptides. Cette purification est basée sur les propriétés physico-chimiques des peptides sans *a priori* de fonction. Dans ce projet, l'étape de purification est faite en deux temps : 1 - purification par la technologie ProteoMiner qui permet de diminuer artificiellement la concentration des protéines fortement concentrées et d'augmenter celle des protéines faiblement concentrées. 2 - purification par billes magnétiques de type C8 qui permettent de ne retenir que les peptides à caractère hydrophobe et d'éliminer les contaminants chimiques pouvant perturber l'acquisition en spectrométrie de masse.

La deuxième étape, dite analytique, correspond à l'acquisition de spectres par un spectromètre de masse de type MALDI-TOF. Les échantillons sont déposés sur une plaque en acier puis recouvert d'une matrice ionique. Celle-ci permet d'ioniser les peptides présents dans les échantillons par l'intermédiaire d'un laser qui vient exciter chacun des dépôts « matrice/échantillon » une fois la plaque insérée dans le spectromètre. L'application d'un champ électrique permet ensuite d'accélérer les peptides ionisés qui vont alors traverser le tube de vol du spectromètre jusqu'au détecteur. Le détecteur placé au bout du tube permet de mesurer la quantité de peptides arrivant sur celui-ci à chaque instant et de rendre l'information sous forme de spectre. Le temps de vol est ensuite transformé en une information masse-sur-charge (m/z) par une relation quadratique simple.

Chaque pic dans le spectre correspond à un peptide et son aire reflète la quantité du peptide dans l'échantillon d'origine. Les spectres reflètent ainsi le profil protéique des échantillons.

A partir de ces spectres, la découverte de biomarqueurs se décompose également en deux étapes. Dans un premier temps, les spectres sont traités pour extraire les caractéristiques des pics. Tout d'abord, les spectres obtenus doivent être prétraités pour en retirer la partie du signal qui correspond à du bruit. L'étape de débruitage supprime le bruit blanc aléatoire ; l'étape de soustraction de la ligne de base retire le bruit chimique ; l'étape de normalisation ramène tous les spectres étudiés sur une même échelle ; enfin l'étape de détection des pics sélectionne dans le signal les pics qui correspondent réellement à des peptides, i.e. à des biomarqueurs candidats. Une difficulté supplémentaire est l'enchaînement séquentiel de ces traitements [4]. En particulier si un artefact apparaît lors d'une étape, aucun retour en arrière n'est possible. Dans le projet BHI-PRO, un pas vers un traitement global et bayésien de spectres MALDI est réalisé en introduisant une méthode qui effectue simultanément et de façon cohérente, la recherche de la ligne de base et la déconvolution du spectre. La méthode s'accommode de spectres bruités et l'étape de lissage disparaît également. La méthode proposée cherche une ligne de base non-paramétrique (définie pour chaque canal), mais dont les variations sont contraintes à une certaine douceur.

Simultanément, la méthode réalise une déconvolution du spectre (à forme de pic connue et fixée). Lors de la déconvolution, les pics sont sélectionnés parcimonieusement grâce à une contrainte de type L1.

Dans second temps, les marqueurs pouvant discriminer les individus sains et malades sont identifiés. A cet effet, un modèle bayésien hiérarchique reliant variables biologiques et caractéristiques des pics est proposé [6,7]. Ensuite, une fonction de sélection optimale au sens d'un risque bayésien est construite. La décision est fondée sur une comparaison des probabilités a posteriori de chaque combinaison de marqueurs. Une contribution importante est liée à la possibilité d'inclure d'éventuelles corrélations entre les marqueurs. Un autre aspect, remarquable, est que les calculs sont entièrement explicites, ce qui permet une mise en œuvre efficace et une interprétation directe.

2.2.2 Approche MRM en mode validation et méthode d'inversion hiérarchique bayésienne associée

Une chaîne d'acquisition MRM comporte principalement deux étapes. La première appelée pré-analytique consiste à découper les protéines qui sont de grandes molécules en plusieurs molécules plus petites, appelées peptides, grâce à un traitement appelé digestion, à les concentrer et en sélectionner une partie par un traitement de fractionnement. Cette étape est décrite par une matrice reliant la concentration de protéines à la concentration de peptides. Chaque coefficient est constitué d'un gain de digestion théorique multiplié par un rendement expérimental. La deuxième étape appelée analytique rassemble une colonne de chromatographie liquide qui va séparer les peptides en fonction de leur temps de rétention dans la colonne, suivi d'une ionisation et d'une injection dans le spectromètre de masse. Le spectromètre est un dispositif à trois quadripôles : le premier quadripôle sert à sélectionner les ions en fonction de leur masse, le deuxième sert à les fragmenter en molécules plus petites par collision avec des molécules d'azote, et le troisième sert à sélectionner de nouveau les fragments en fonction de leur masse. Chaque couple (peptide, fragment) est appelé une transition. Le signal mesuré (y) rassemble plusieurs traces correspondant à l'intensité d'arrivée des fragments en sortie du troisième quadripôle en fonction du temps. Un peptide va donner lieu à plusieurs traces associées à chaque transition. Pour décrire la réponse théorique du système, nous utilisons un modèle gaussien paramétré par son temps de rétention, son écart-type et son gain. Le signal est entaché d'un bruit que l'on modélise par un bruit blanc gaussien d'inverse variance indéterminée. Pour aider à identifier les paramètres de gain, des mesures de calibration sont réalisées : des peptides isotopiquement alourdis en quantités connues sont ajoutés à la solution en entrée de l'étape analytique et les traces associées sont mesurées (y^*). Par ailleurs, quotidiennement, des mesures de référence sur les peptides natifs (y^{CQ}) et sur les peptides alourdis (y^{*CQ}) sont réalisées sur un mélange étalon servant aussi au contrôle qualité.

Partant de ces 4 ensembles de traces y , y^* , y^{CQ} , y^{*CQ} , nous devons résoudre le problème inverse, c'est-à-dire estimer conjointement la concentration des protéines natives dans le sérum et tous les paramètres de nuisance intervenant dans le modèle. Cette inversion est réalisée par une approche bayésienne nécessitant de choisir des distributions a priori pour chaque paramètre. Nous utilisons principalement des lois normales, sauf pour les temps de rétention et la largeur des pics où nous utilisons des lois uniformes, et pour l'inverse

variance du bruit où nous utilisons une loi gamma. Les concentrations de protéines $x_{p,u}$ et les facteurs de nuisance sont estimés par un estimateur de la moyenne a posteriori. Cet estimateur est calculé par un algorithme de Gibbs.

Dans le cadre de la convergence avec les biostatistiques, nous avons étendu cette approche problèmes inverses en introduisant un niveau hiérarchique supplémentaire correspondant à l'état de l'individu, sain ou pathologique. Nous avons proposé deux nouveaux traitements [3]:

1. l'apprentissage supervisé à partir d'une cohorte multi-classe, vu comme l'estimation des paramètres du modèle des classes,
2. la classification à partir du modèle des classes, vue comme l'estimation de l'étiquette b_u de la classe d'un nouvel individu.

Ces traitements peuvent être réalisés conjointement avec la quantification, ou séparément dans une deuxième étape. Une version simplifiée accélérée reposant sur un modèle de classe gaussien a été évaluée. Celle-ci est réalisée sous la forme d'une analyse quadratique discriminante (QDA) [5].

3 Résultats expérimentaux pour le mode MALDI

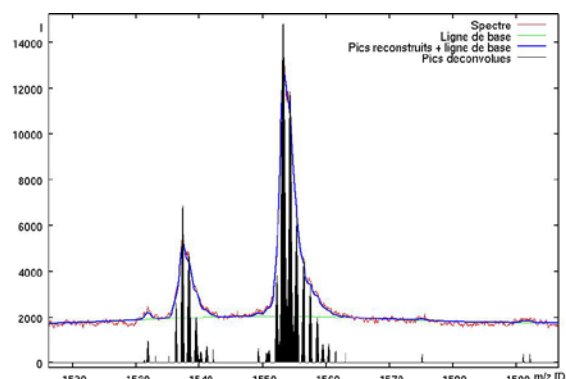


Figure 1: spectre MALDI mesuré et spectre estimé après soustraction de la ligne de base et déconvolution

Nous présentons sur la figure 1 un exemple de spectre mesuré sur un spectromètre de masse MALDI TOF Xtreme de Bruker et le résultat des traitements après soustraction de la ligne de base et déconvolution. Nous montrons aussi la sortie du modèle de signal rassemblant les pics reconstruits et la ligne de base estimée.

Pour l'étape d'identification des biomarqueurs, nous l'avons testée et validée sur des données simulées correspondant à des cohortes de différentes tailles (de 100 à 1000 individus). Dans le cas univarié, nous avons montré que notre approche présente des connections avec les tests classiques de Student et de Fisher [6]. Dans le cas multivarié, nous avons évalué et comparé la méthode au test classique de Student [7]. Nous montrons en théorie et au travers des résultats de simulation que la fonction de sélection proposée est supérieure à ces tests classiques au sens de l'erreur de classification moyenne. Il semble que cette fonction ait tendance à produire davantage de faux négatifs que de faux positifs. Un travail ultérieur pourra être consacré à la construction d'un nouveau risque permettant de pénaliser plus ou moins fortement chaque type d'erreur. Ce développement est assez naturel dans le cadre bayésien hiérarchique adopté.

4 Résultats expérimentaux pour le mode MRM

A titre d'exemple, nous donnons les résultats sur une cohorte d'échantillons du cancer colorectal, pour le couple de biomarqueurs LFABP et PDI, en sélectionnant 72 contrôles

sains appartenant à un seul groupe et 98 cas de cancer colorectal allant du stade 1 au stade 4 [8,9]. Les acquisitions MRM ont été réalisées sur un spectromètre de masse triple quadripôle 5500 QT d'AB Sciex en mode MRM. Pour la LFABP, nous avons considéré 8 transitions associées à 3 peptides, et pour la PDI, 3 transitions associées à 1 peptide. Pour cette évaluation, nous avons comparé l'approche inversion hiérarchique bayésienne (BHI) pour la quantification et une approche non-linéaire (NLP) reposant sur une sélection supervisée des pics par l'opérateur, suivi d'une normalisation non-linéaire des valeurs des transitions appliquée à l'aire sous le pic et l'attribution à la protéine de la valeur médiane des valeurs calculées pour chaque transition.

4.1 Analyse expérimentale des performances en quantification

L'analyse des performances en quantification sur données expérimentales pose des problèmes méthodologiques car la concentration absolue des protéines mesurées n'est pas connue. Nous décrivons 2 types d'approches réalisées soit sur des rampes de dilution, soit sur une cohorte de patients et comparaison avec une méthode de dosage de référence par reconnaissance par anticorps appelée ELISA (Enzyme Linked Immunoabsorbent Assay).

Sur les rampes de dilution, l'analyse se fait en calculant les coefficients de régression linéaire entre les dilutions estimées et les dilutions vraies, le coefficient de détermination et les coefficients de variation. Nous avons obtenu des coefficients de variation inférieurs à 5% pour plusieurs protéines dont la protéine LFABP [1,3].

Pour l'évaluation sur la cohorte de patients, nous avons utilisé la protéine LFABP et comparé les résultats obtenus avec le dosage ELISA. Avec l'approche hiérarchique bayésienne BHI, nous avons obtenu un coefficient de corrélation de Pearson de 0.95 [9]. En comparaison, l'approche classique supervisée NLP donne un coefficient de corrélation de 0.93.

4.2 Analyse expérimentale des performances en classification

L'évaluation des performances est réalisée par validation croisée. Dans notre cas, la cohorte est divisée en 10 groupes, chacun d'entre eux ayant 17 échantillons. Dans les 2 cas, nous avons utilisé comme méthode de classification une analyse discriminante quadratique reposant sur des modèles gaussiens pour chacune des classes [9]. Cette classification a été réalisée pour plusieurs valeurs de la probabilité a priori d'appartenir à la classe pathologique, de façon à faire varier le compromis sensibilité/spécificité, et à pouvoir tracer la courbe ROC associée. Les courbes ROC empiriques ont été lissées pour supprimer l'effet de quantification lié à la taille limitée des cohortes.

Nous présentons sur la figure 2 les courbes ROC pour les 2 approches. Si on considère sur les courbes lissées les valeurs pour une spécificité de 0.95, les résultats obtenus sont les suivants :

- inversion hiérarchique bayésienne BHI : $Se=0.59Sp=0.95$
- approche non-linéaire NLP : $Se=0.47Sp=0.95$

Sur cet exemple, la courbe ROC de l'approche BHI est systématiquement au-dessus de celle pour l'approche NLP, ce qui signifie qu'elle est plus sensible pour une spécificité donnée. Cette différence est encore plus sensible pour les fortes spécificités. Cette différence est significative pour un test de dépistage pour lequel une spécificité minimum de 0.95

est demandée. Ceci montre aussi que notre algorithme a su gérer les variabilités technologiques.

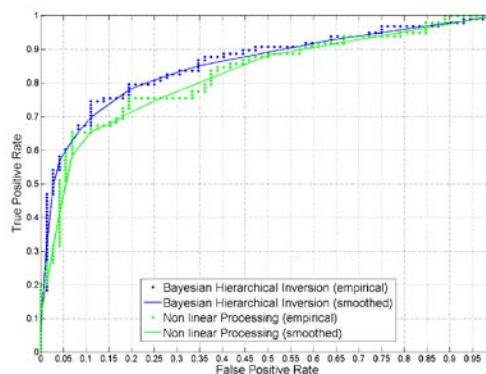


Figure 2: Courbes ROC empiriques et lissées pour les méthodes BHI et NLP

5 Résumé et perspectives

Cette communication présente une vue d'ensemble des méthodes d'analyse biostatistique et d'inversion hiérarchique bayésienne étudiées sur le projet BHI-PRO pour la recherche et la validation de biomarqueurs par spectrométrie de masse. Cette communication met en évidence les nombreux liens avec des domaines plus classiques du traitement du signal et des images. Nous avons aussi explicité les critères de performance utilisés et présenter des premiers résultats expérimentaux. En mode MRM, nous avons montré notamment l'amélioration de la qualité statistique des résultats. Comme l'analyse des données peut être réalisée de manière non-supervisée, il devient possible d'envisager d'appliquer ces traitements sur des cohortes de grande taille afin d'améliorer la puissance statistique des études de biomarqueurs. Ceci va aussi nous amener à étudier la question de l'optimisme lié au taux de faux positifs dans l'étape de sélection des biomarqueurs.

6 Références

- [1] GERFAULT L., ET AL. (2012), "A HIERARCHICAL SRM ACQUISITION CHAIN MODEL FOR IMPROVED PROTEIN QUANTIFICATION IN SERUM SAMPLES", RESEARCH IN COMPUTATIONAL MOLECULAR BIOLOGY (RECOMB) SATELLITE CONFERENCE ON COMPUTATIONAL PROTEOMICS 2012 (RECOMB CP), SAN DIEGO, USA.
- [2] SZACHERSKI P., ET AL. (2012), "ROBUST MS SERUM SAMPLE CLASSIFICATION IN PROTEOMICS BY THE USE OF INVERSE PROBLEMS", 2012 IEEE INTERNATIONAL WORKSHOP ON GENOMIC SIGNAL PROCESSING AND STATISTICS, WASHINGTON, DC, USA.
- [3] SZACHERSKI P. (2012), RECONSTRUCTION DE PROFILS PROTEIQUES POUR LA RECHERCHE DE BIOMARQUEURS, THESE DE DOCTORAT, UNIV. BORDEAUX 1.
- [4] HU, JIANHUA, ET AL. (2005), THE IMPORTANCE OF EXPERIMENTAL DESIGN IN PROTEOMIC MASS SPECTROMETRY EXPERIMENTS: SOME CAUTIONARY TALES. BRIEFINGS IN FUNCTIONAL GENOMICS & PROTEOMICS, VOL. 3, NO 4, P. 322-331.
- [5] HASTIE T., ET AL. (2009), THE ELEMENTS OF STATISTICAL LEARNING, SPRINGER
- [6] N. DRIDI, ET AL. (2013), "SELECTION BAYESIENNE DE BIOMARQUEURS : APPLICATION A UN PROBLEME DE PROTEOMIQUE", GRETSI 2013, BREST, FRANCE.
- [7] F. ADJED, ET AL. (2013), "VARIABLE SELECTION FOR A MIXED POPULATION APPLIED IN PROTEOMICS", ICASSP, VANCOUVER, CANADA.
- [8] FORTIN ET AL. (2011), "BIOMARKER VALIDATION FOR COLORECTAL CANCER USING SRM AND MRM³ AND CORRELATION WITH ELISA TEST", 10TH ANNUAL HUPO WORLD CONGRESS, GENEVA, SWITZERLAND.
- [9] SZACHERSKI P. ET AL (2013), "MRM PROTEIN QUANTIFICATION AND SERUM SAMPLE CLASSIFICATION", 61ST CONFERENCE ON MASS SPECTROMETRY AND ALLIED TOPICS, MINNEAPOLIS, MINNESOTA, USA