MRM protein quantification and serum sample classification

Pascal Szacherski, Laurent Gerfault, Jean-François Giovannelli, Audrey Giremus, Pierre Mahé, Tanguy Fortin, Geneviève Choquet, Amna Klich, Catherine Mercier, Pascal Roy, et al.

MRM protein quantification and serum sample classification

Pascal SZACHERSKI[12], Laurent GERFAULT[1], Jean-François GIOVANNELLI[2], Audrey GIREMUS[2], Pierre MAHE[3], Tanguy FORTIN[5], Geneviève CHOQUET-KASTYLEVSKY[5], Amna KLICH[4], Catherine MERCIER[4], Pascal ROY[4], Arnaud SALVADOR[6], Jérôme LEMOINE[6], Jean-Philippe CHARRIER[5], Bruno LACROIX[5], Pierre GRANGEAT[1]

[1] CEA Leti, MINATEC Campus, DTBS, 17 rue des Martyrs, F-38054 Grenoble cedex 9, France.
[2] Université de Bordeaux 1 – CNRS - IPB, IMS, 351 Cours de la Libération, F-33405, Talence cedex, France.
[3] bioMérieux, Parc Polytec, 5 rue des Berges, F-38000 Grenoble, France
[4] Hospices Civils de Lyon, Service de Biostatistique; Université Lyon I, CNRS UMR 5558, LBBE, BSS, PRABI, 162 avenue Lacassagne, F-69424 Lyon.
[5] bioMérieux, 376 Chemin de l'Orme, F-69280 Marcy l'Etoile, France
[6] Institut des Sciences Analytiques, UMR 5280, CNRS / Université de Lyon, Lyon-1, 5 rue de la Doua, F-69100 Villeurbanne

Introduction : 120 words
Quantification and classification are key points for differential analysis of proteomic studies and diagnostic tests. A MRM analytical chain is a cascade of molecular events depicted by a graph structure, each node being associated to a molecular state such as protein, peptide or ion and each branch to a molecular processing. Each protein is associated to a set of transition measurements. One key question is how to infer the protein level and the class label. We propose to compare a hierarchical model based Bayesian Hierarchical Inversion combining all transitions and a non-linear processing based on logarithmic transformation of standardized peak value combined with a median filter. Classification performances are evaluated on a colorectal cancer cohort for LFABP and PDI biomarkers.

Methods : 120 words
For Bayesian Hierarchical Inversion [1], a full graphical hierarchical model of MRM acquisition chain is proposed combining respectively sample status, protein, peptide and ion levels, including biological and technological parameters. The Bayesian estimation delivers automatically the protein concentration, including AQUA labelled peptides and control quality samples concentrations for gain estimation. For non-linear processing, an operator-supervised selection of peak position is achieved and a transition value is computed taking the logarithm of one plus the ratio between the native and AQUA transition peak area. Then, the median over all protein transition values is assigned as protein value. In both cases, classification is achieved using a Quadratic Discriminant Analysis [2] based on a Gaussian model for healthy and pathologic classes.

Preliminary data : 300 words
From a list of biomarker candidates selected from another biomarker colorectal cancer research cohort, we have selected 2 of them, the LFABP and the BPI proteins to evaluate the performances of the 2 processing strategies. The evaluation is performed on a second cohort used for validation [3] which includes 91 control cases and 115 colorectal cancer cases starting from grade 1 up to grade 4. MRM acquisitions have been achieved using an AB Sciex

QT5500 Triple Quadrupole mass spectrometer in MRM mode [3]. For LFABP, 8 transitions and 3 peptides have been considered, and for PDI, 3 transitions and 1 peptide.
Evaluation of classification performances has been achieved by cross-validation, using one 10-fold process. The cohort is divided into 10 groups, each having nearly 20 samples. For each configuration, 9 groups are used for training the classifier, and one for testing the classifier performances. Each group is used once as test group during the 10-fold process.
Evaluated classification performances are accuracy Ac, sensitivity Se, and specificity Sp:

- Bayesian Hierarchical Inversion: Ac=0.76 Se=0.60 Sp=0.96
- Non-linear processing: Ac=0.76 Se=0.62 Sp=0.94

Those performances are nearly the same. In particular, in both cases, we get a very high specificity. This demonstrates that using the Bayesian Hierarchical Inversion, we are able to quantify the protein content in an automatic way with the same classification performances than the operator-supervised non-linear processing currently used. The algorithm has been able to manage the technological variability. This open the way towards robust automatic processing of larger cohorts, in order to enhance the statistical power of biomarker studies and to allow the development of automatic test for diagnosis.

[1] Szacherski et al. (2012), IEEE GENSIPS'12 Workshop.
[2] Hastie et al. (2009), The Elements of Statistical Learning, Springer.
[3] Fortin et al. (2011), 10th Annual HUPO World Congress.

Novel aspect : 20 words
Bayesian Hierarchical Inversion, graphical model, median filter, automatic protein quantification, classification, quadratic discriminant analysis, colorectal cancer, serum biomarker, MRM.

Topic for oral session:
Informatics: Protein quantification