Karolinska Institutet

Department of Oncology-Pathology

Clinical Proteomics – Quantitative Analysis and Biological Interpretation

AKADEMISK AVHANDLING
som för avläggande av medicine doktorsexamen vid Karolinska Institutet offentligen försvaras i Hörsal Rockefeller, Nobels väg 11, Karolinska Institutet Solna.

Fredagen den 2:a november 2012, kl. 9.15

av
AnnSofi Sandberg

Huvudhandledare:
Dr. Jenny Forshed
Karolinska Institutet
Institutionen för Onkologi-Patologi
Cancer Proteomik, Masspektrometri
SciLife lab Stockholm

Bihandledare:
Docent Janne Lehtiö
Karolinska Institutet
Institutionen för Onkologi-Patologi
Cancer Proteomik, Masspektrometri
SciLife lab Stockholm

Professor Magnus Sköld
Karolinska Institutet
Institutionen för Medicin
Enheten för Lungmedicin

Fakultetsopponent:
Professor Anthony Whetton
University of Manchester
Stem Cell and Leukemia Proteomics Laboratory, Manchester Academic Health Sciences Centre

Betygsnämnd:
Professor Jacob Odeberg
Kungliga Tekniska Högskolan
Avdelningen för Proteomik

Professor Johan Gottfries
Göteborgs Universitet
Institutionen för Medicinsk Kemi

MD. Dr. Magnus Nisell
Karolinska Institutet
Institutionen för Medicin
Lung-Allergikliniken, Karolinska Universitetssjukhuset

Stockholm 2012

ABSTRACT

The main expectations of applying proteomics technologies to clinical questions are the discovery of disease related biomarkers. Despite technological advancement to increase proteome coverage and depth to meet these expectations the number of generated biomarkers for clinical use is small. One of the reasons is that found potential biomarkers often are false discoveries. Small sample sizes, in combination with patient sample heterogeneity increase the risk of false discoveries. To be able to extract relevant biological information from such data, high demands are put on the experimental design and the use of sensitive and quantitatively accurate technologies.

The overall aim of this thesis was to apply quantitative proteomics methods for biomarker discovery in clinical samples. A method for reducing bias by controlling for individual variation in smoking habits is described in paper I. The aim of the method was objective assessment of recent smoking in clinical studies on inflammatory responses. In paper II, the proteome of alveolar macrophages obtained from smoking subjects with and without the inflammatory lung disease chronic obstructive pulmonary disease (COPD) were quantified by two-dimensional gel-electrophoresis (2-DE). A gender focused analysis showed protein level differences within the female group, with down-regulation of lysosomal pathway and up-regulation of oxidative pathway in COPD patients. Paper III, a mass spectrometry based proteomics analysis of tumour samples, contributes to the molecular understanding of vulvar squamous cell carcinoma (VSCC) and we identified a high risk patient subgroup of HPV-negative tumours based on the expression of four proteins, further suggesting that this subgroup is characterized by an altered ubiquitin-proteasome signalling pathway. Paper III describes a data analysis workflow for the extraction of biological information from quantitative mass spectrometry based proteomics data. High patient-to-patient tumour proteome variability was addressed by using pathway profiling on individual tumour data, followed by comparison of pathway association ranks in a multivariate analysis. We show that pathway data on individual tumour level can detect subpopulations of patients and identify pathways of specific importance in pre-defined clinical groups by the use of multivariate statistics. In paper IV, the potentials and limits of quantitative mass spectrometry on clinical samples was evaluated by defining the quantitative accuracy of isobaric labels and label-free quantification. Quantification by isobaric labels in combination with p*I* pre-fractionation showed a lower limit of quantification (LOQ) than a label-free analysis without p*I* pre-fractionation, and 6-plex TMT were more sensitive than 8-plex iTRAQ. Precursor mixing measured by isolation interference (MS1 interference) is more linked to the quantitative accuracy of isobaric labels than reporter ion interference (MS2 interference). Based on that we could define recommendations for how much isolation interference that can be accepted; in our data <30% isolation interference had little effect the quantitative accuracy.

In conclusion, getting biological knowledge from proteomics studies requires a careful study design, control of possible confounding factors and the use of clinical data to identify disease subtypes. Further, to be able to draw conclusions from the data, the analysis requires accurate quantitative data and robust statistical tools to detect significant protein alterations. Methods around these issues are developed and discussed in this thesis.

ISBN 978-91-7457-914-7