Défis de la protéomique quantitative label-free

Yohann Couté

20 Novembre 2014

Atelier PROSPECTOM

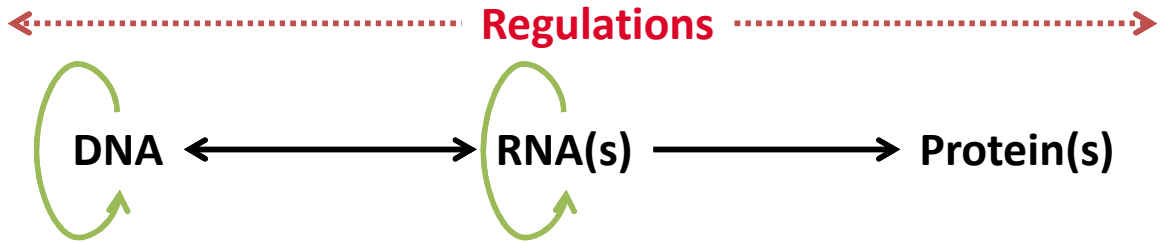
Summary

- **Introduction: MS-based quantitative proteomics**
- **Evaluation of label-free workflows for discovery proteomics**
- **Biological applications:**
  - SLP76 interactome (mast cells signalling)
  - Identification of a novel molecular weapon in a clinical P. *aeruginosa* strain
- **Conclusions and perspectives**

Introduction

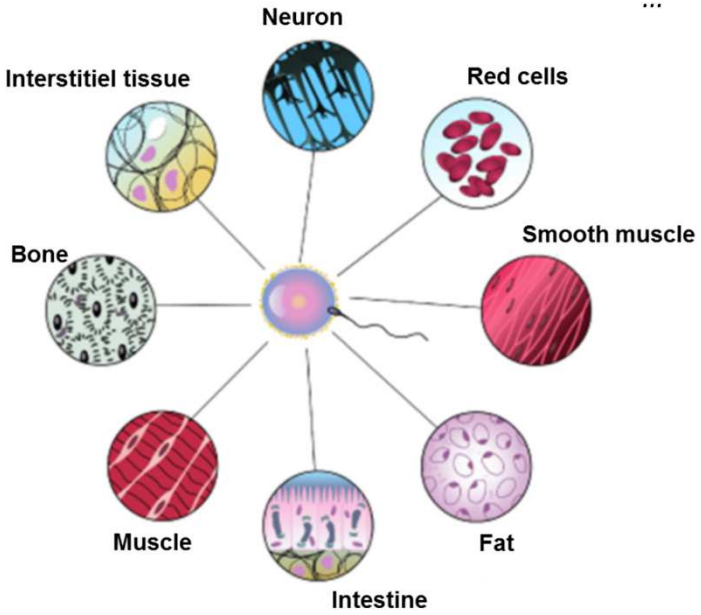
MS-based quantitative proteomics

Proteomics challenges



Alternative splicing
mRNA editing
…

Post-translational modifications
…



➡ **Complexity**

➡ **Dynamic range**

➡ **Dynamics**

Constant genome → different proteomes

Expression proteomics challenges

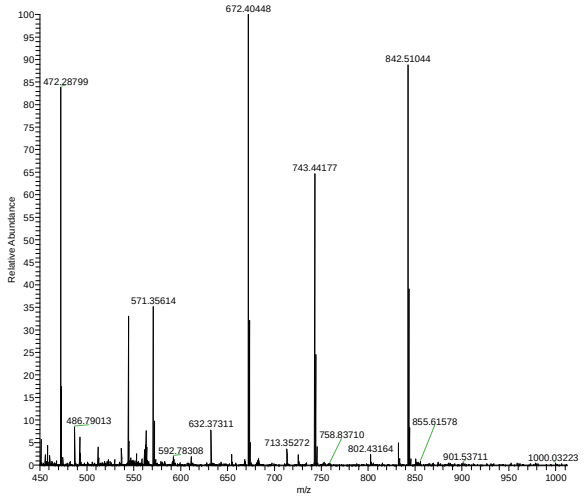
Exhaustively identify and accurately compare protein levels in complex biological systems in various states (if possible, routinely…)

→ MS-based analyses

Mass spectrometry

MS is a technique for separating ions by their mass-to-charge (m/z) ratios

Inlet:
Sample introduction



Computer:
to process the data which produces the mass spectrum

Source Ionisation:
to produce ions from the sample

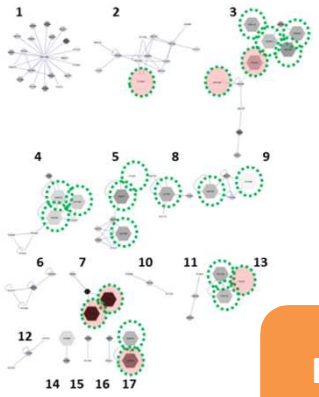
→

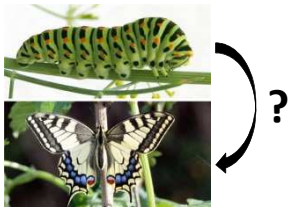
Analyzer(s):
to separate the various ions according to their mass-to-charge ratio

→

Detector:
to count the ions and to measure their abundance

Quantitative MS-based expression proteomics in discovery mode





Biological/ biomedical question

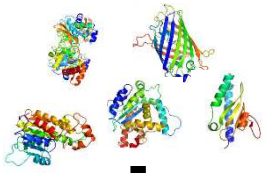
Sample preparation

MS-based measurements

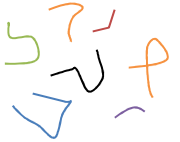
Computation & statistics

Bioanalysis and data mining

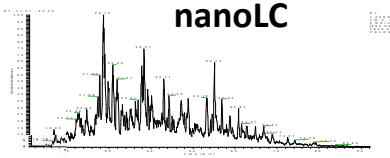
Protein extraction

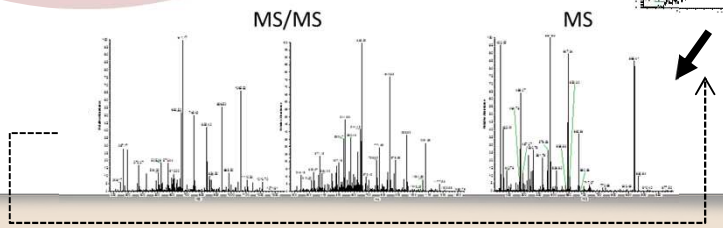


Protein digestion

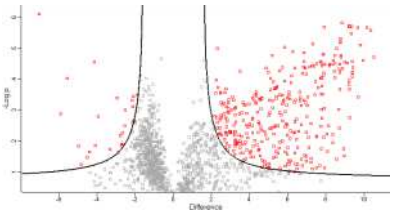


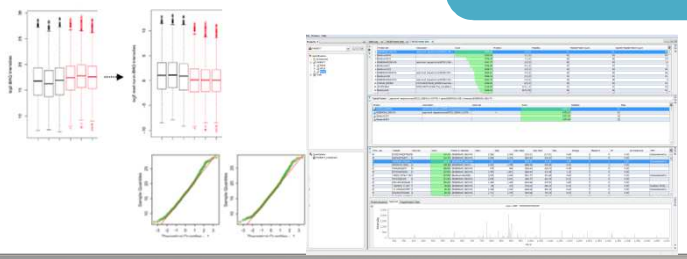
nanoLC





DDA mode





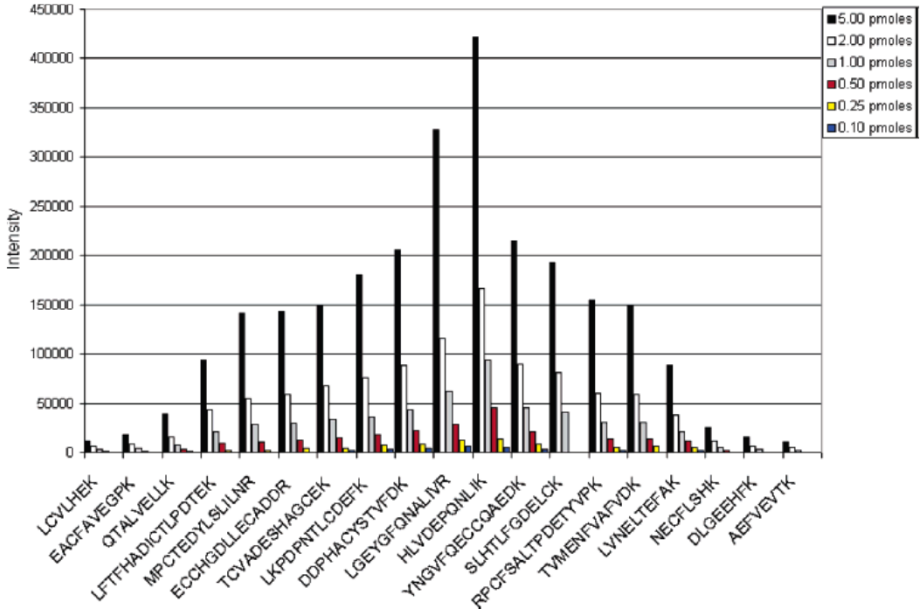
MS-based quantitative proteomics

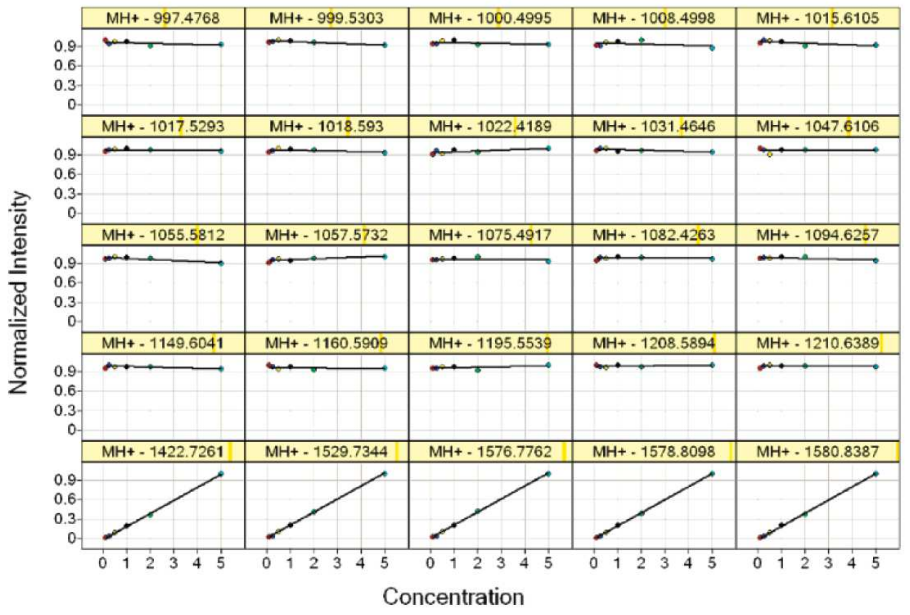
!!! For several reasons, MS is only partially a quantitative tool !!!

Impossible to infere quantification values by comparing intensities of peptides with different sequences.

but

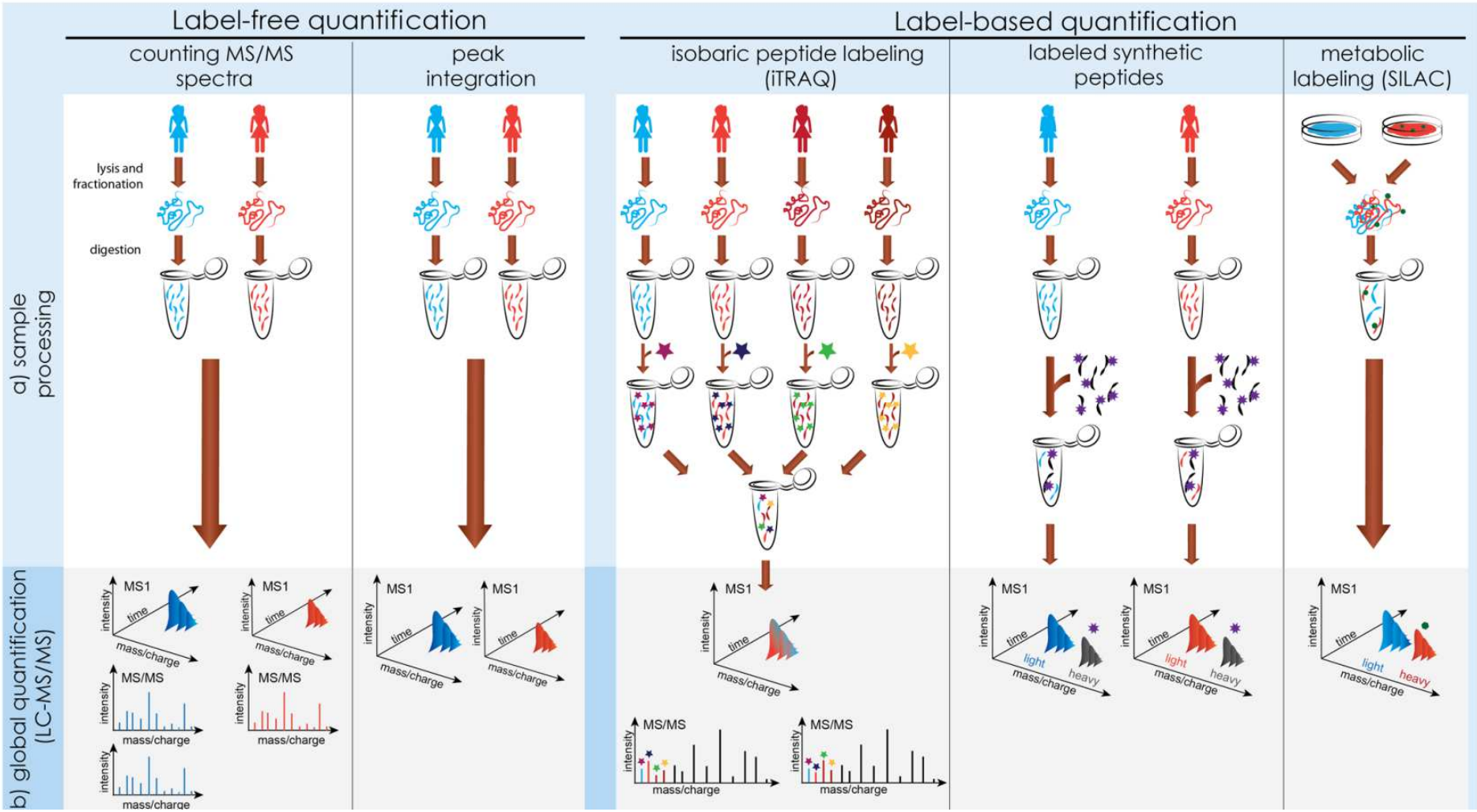
The same peptide can be quantified relatively in separate analyses





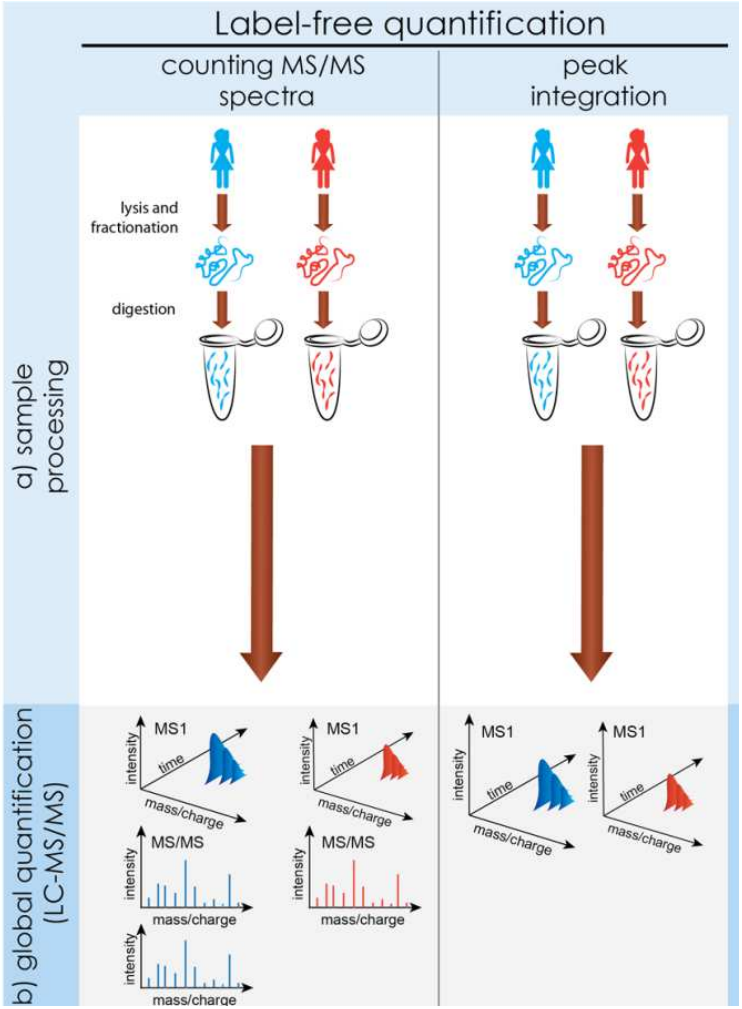
Silva JC et al., 2005, Anal. Chem

Strategies in discovery quantitative proteomics



Käll L & Vitek O, 2011, PLoS Comput Biol

Potential limitations for label-free approaches

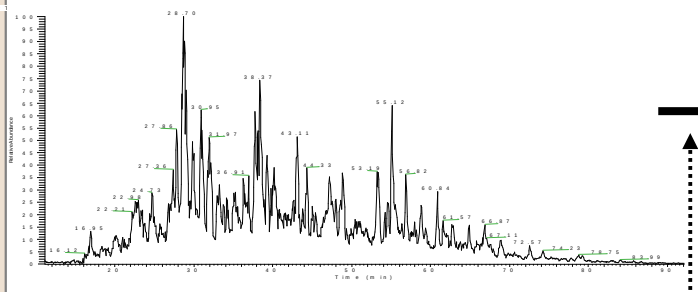


- Limited number of biological and analytical replicates
- Irreproducibility of biological samples
- Multiple fractionation before nanoLC-MS/MS
- Irreproducibility of analytical platform
- Big amounts of data
- …

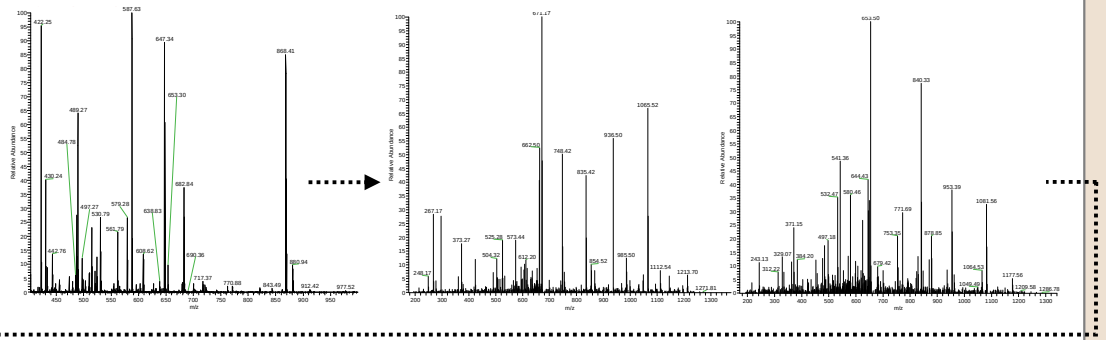
Label-free strategies

Liquid Chromatography

Reverse-phase separation of peptides



Mass spectrometry (DDA)



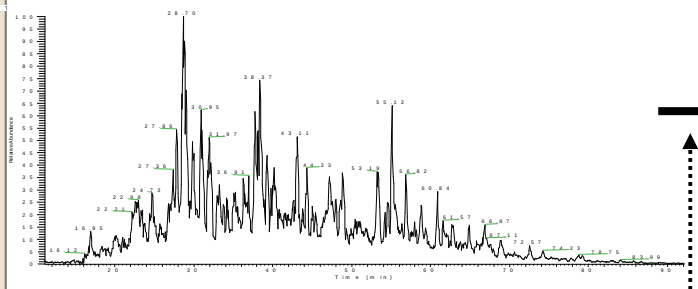
Quantification signal extraction

Statistical treatment

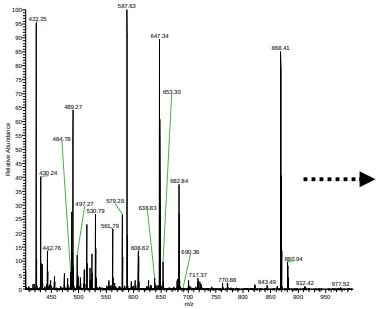
Label-free strategies: quantification signal extraction

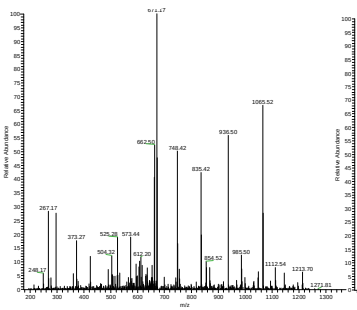
Liquid Chromatography

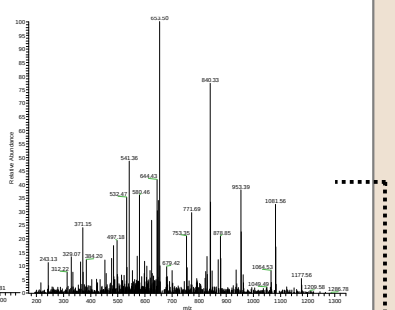
Reverse-phase separation of peptides

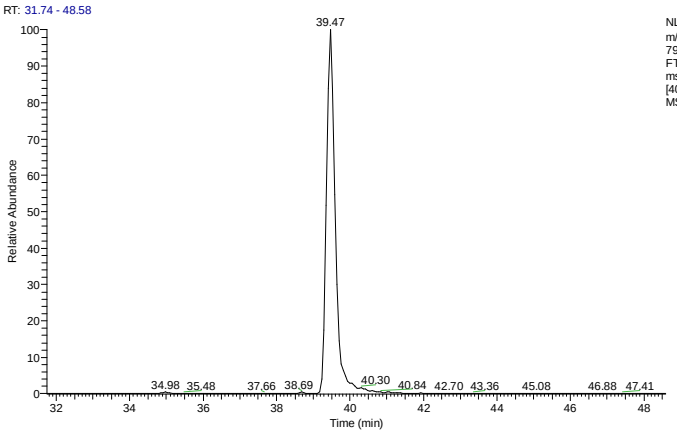


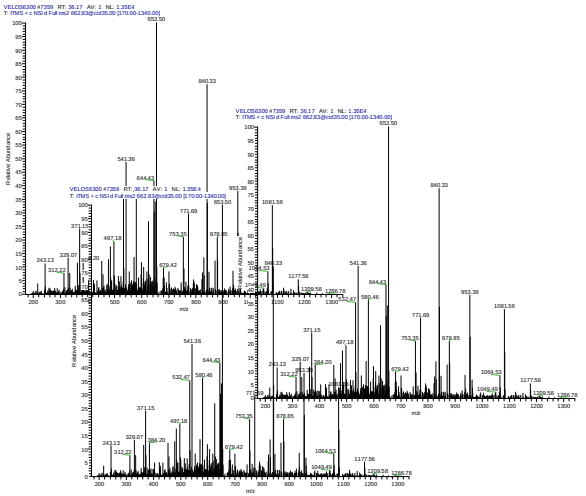
Mass spectrometry (DDA)











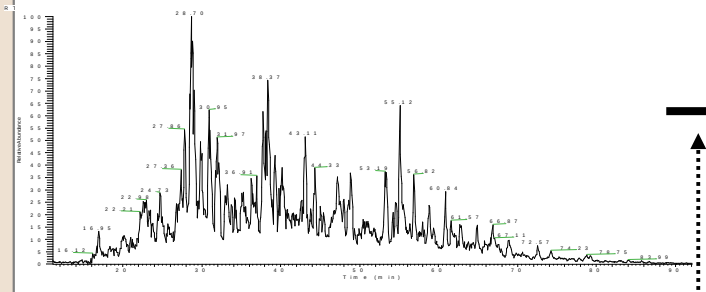
Peak integration

Spectral counting

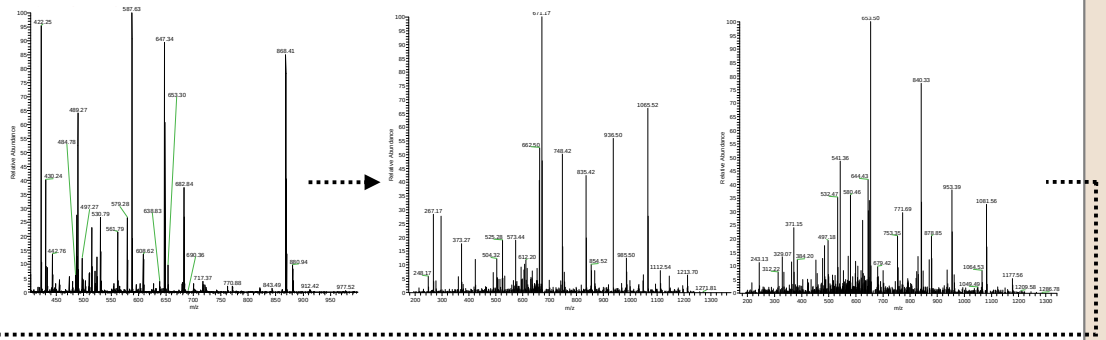
Label-free strategies: quantification signal extraction

Liquid Chromatography

Reverse-phase separation of peptides



Mass spectrometry (DDA)



MS/MS data

↓ **Signal processing**

↓ **Database search**

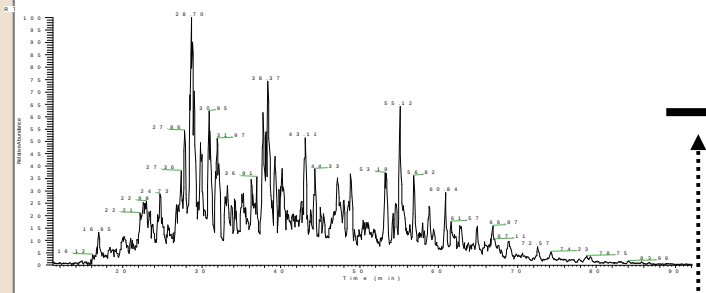
↓ **Peptide identification (+ filtering)**

↓ **Protein grouping**

↓ **Quantitative protein report**

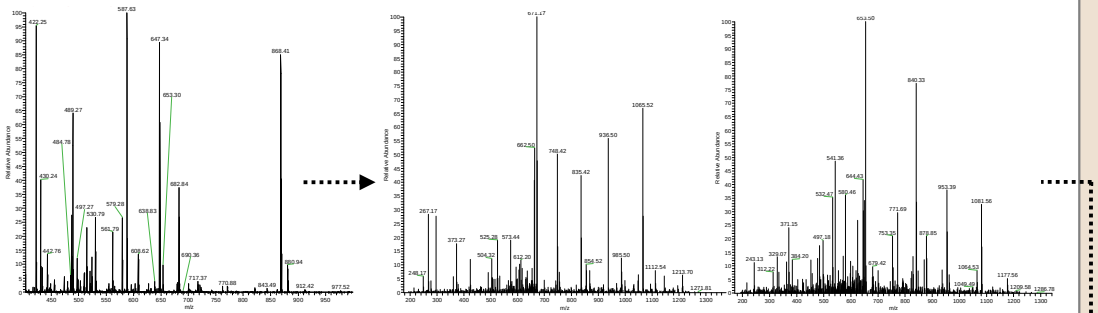
Spectral counting

Label-free strategies: quantification signal extraction

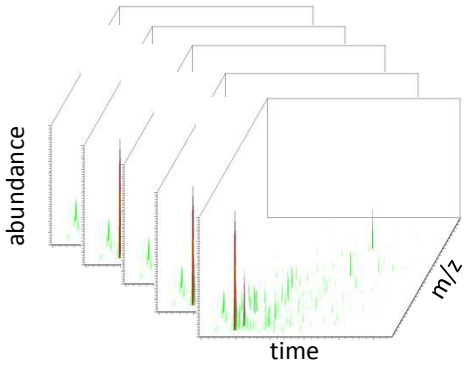


Liquid Chromatography

Reverse-phase separation of peptides



Mass spectrometry (DDA)



MS data

High resolution

Signal processing

Feature detection

Map alignment

Quantitative feature report

Quantitative peptide/protein report

MS/MS data

Signal processing

Database search

Peptide identification (+ filtering)

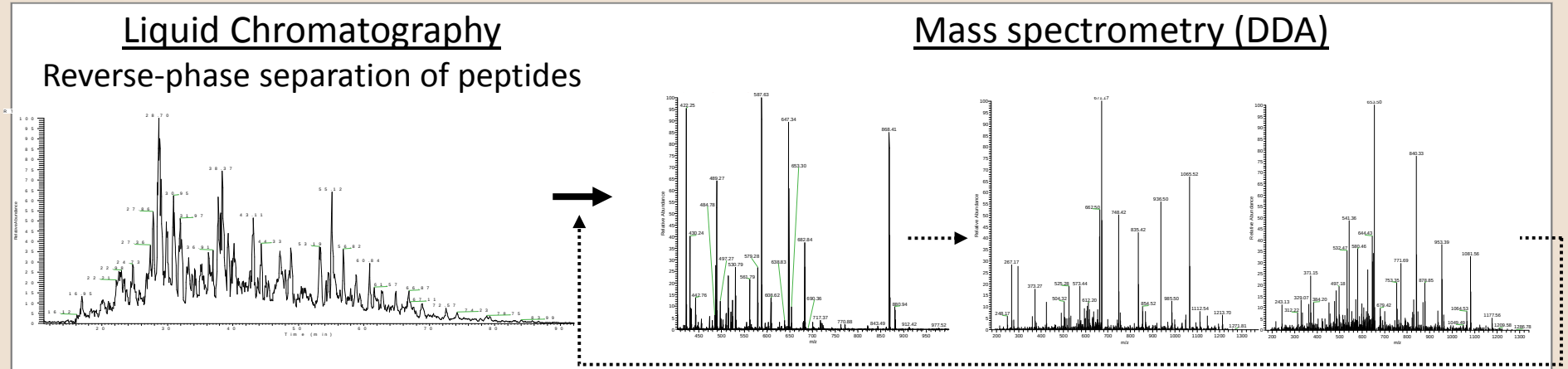
Protein grouping

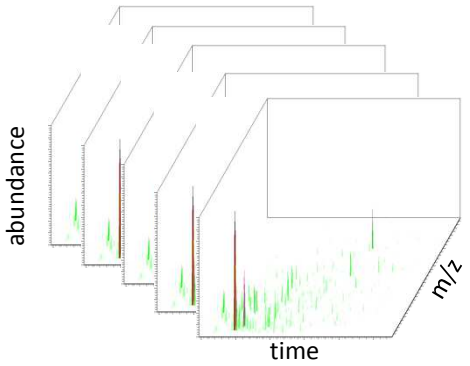
Quantitative protein report

Peak integration

Spectral counting

Label-free strategies: quantification signal extraction





MS data

High resolution

- Signal processing
- Feature detection
- Map alignment

Quantitative feature report

MS/MS data

- Signal processing
- Database search
- Peptide identification (+ filtering)
- Protein grouping
- Quantitative protein report

Quantitative peptide/protein report

Peak integration

Spectral counting

Label-free strategies: quantification signal extraction concerns

Peak integration

Complexity of data (overlapping signals, various isotopic distribution, multiple charge states, non-linear distortion in RT dimension, …)

Sequence of complex tasks (error control)

Inference of protein quantification

Both

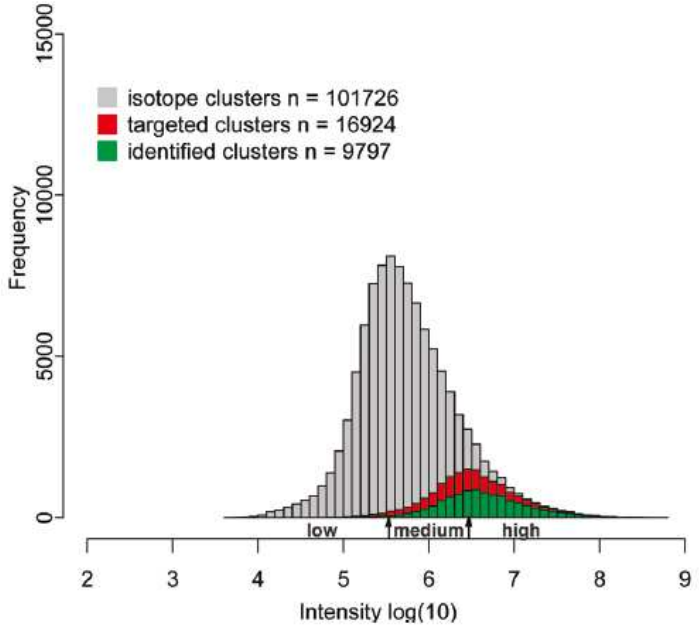
Peptide ID errors

Shared peptides

Low abundance peptides/proteins

Spectral counting

Dynamic exclusion

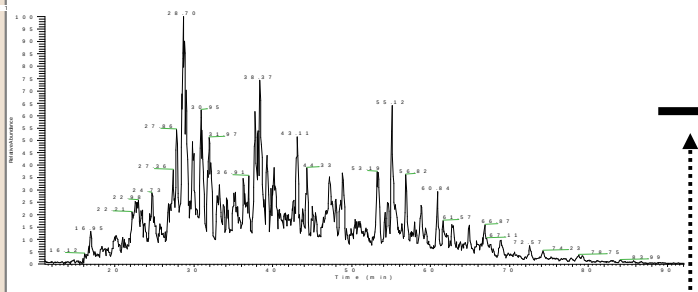


Michalski et al., 2011, J Proteom Res

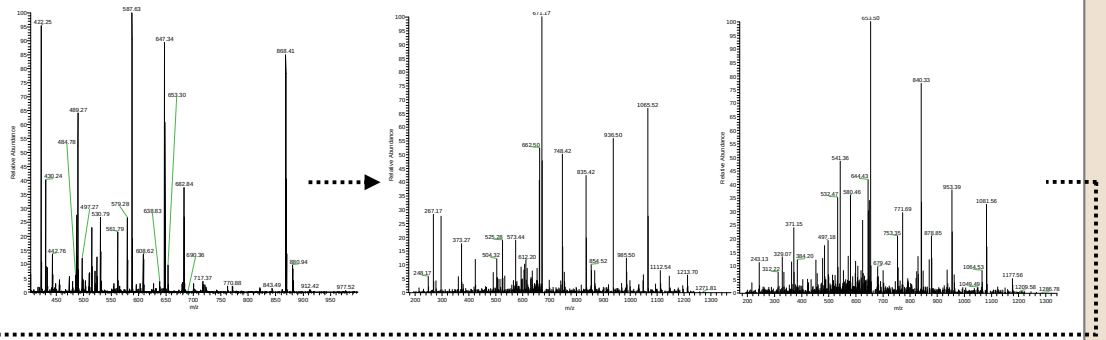
Label-free strategies

Liquid Chromatography

Reverse-phase separation of peptides



Mass spectrometry (DDA)



Quantification signal extraction

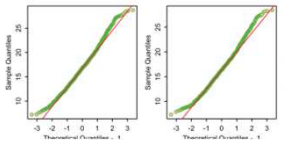
Statistical treatment

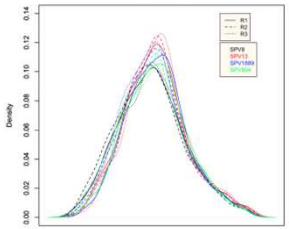
Label-free strategies: road to statistical assessment

Peak integration **Spectral counting**

	Control			Test		
	R1	R2	R3	R1	R2	R3
Prot/Pep A	15	12	18	11		14
Prot/Pep B			1	123	145	138
Prot/Pep n	…	…	…	…	…	…

↓ **Log**



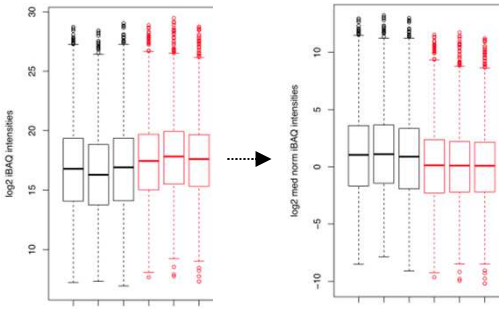


Data filtering

Data qualification

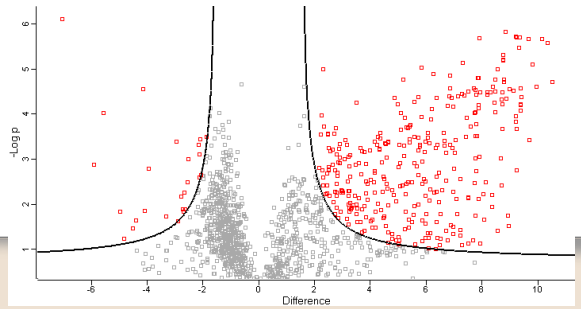
Data normalisation

Missing value imputation



↓

Statistical test

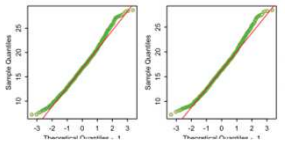


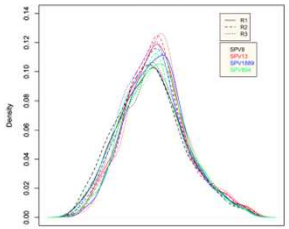
Label-free strategies: road to statistical assessment

Peak integration **Spectral counting**

	Control			Test		
	R1	R2	R3	R1	R2	R3
Prot/Pep A	15	12	18	11		14
Prot/Pep B			1	123	145	138
Prot/Pep n	…	…	…	…	…	…

↓ **Log**



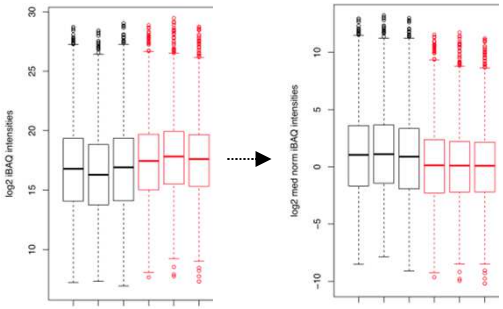


Data filtering

Data qualification

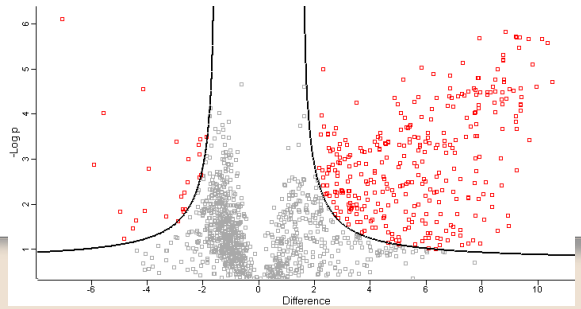
Data normalisation

Missing value imputation



↓

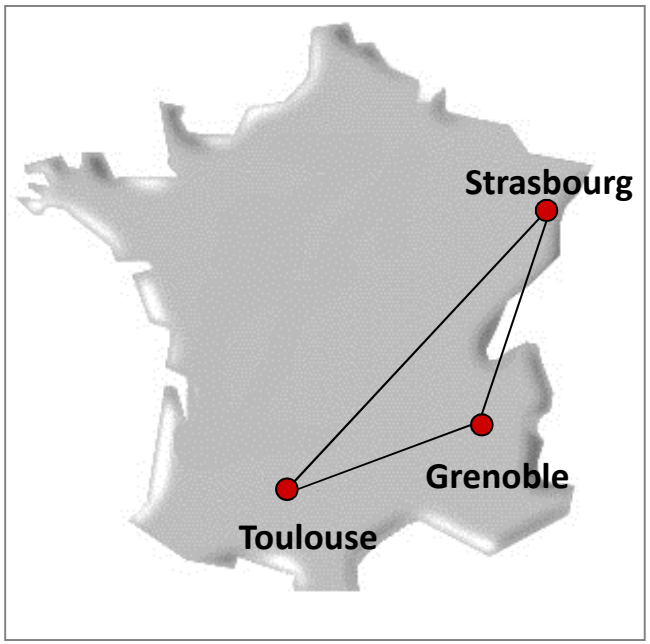
Statistical test



Evaluation of label-free workflows

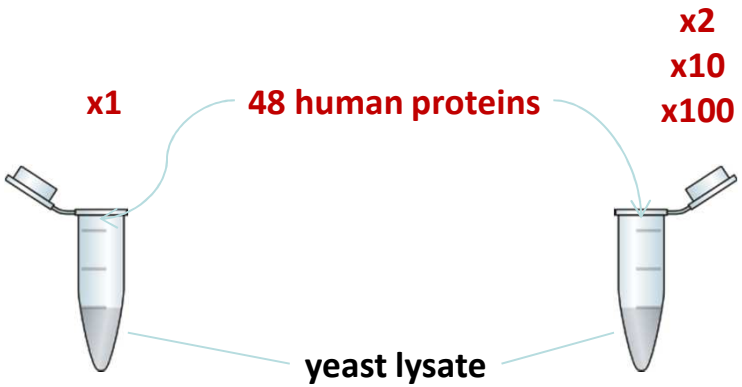
Development of a consistent set of integrated methods to tackle ambitious biological questions

Proteomics French Infrastructure



Joint Research Activities
Access
Networking/Dissemination

Realisation of a standard



Proteomic analyses

in-solution digestion

↓

NanoLC-MS/MS (LTQ-Orbitrap Velos)
3 replicates

↓

Label-free workflows

Tools:

MSn Extract
MaxQuant
MFPaQ
Mascot Distiller
Mascot
hEIDI
LC-Progenesis
Viper
Skyline
IRMa
DeconTools
JMP
Perseus
Andromeda
Scaffold

Tool selection

MSn Extract
MaxQuant
MFPaQ
Mascot Distiller
Mascot
hEIDI
LC-Progenesis
Viper
Skyline
IRMa
Perseus
DeconTools
JMP
Andromeda
Scaffold

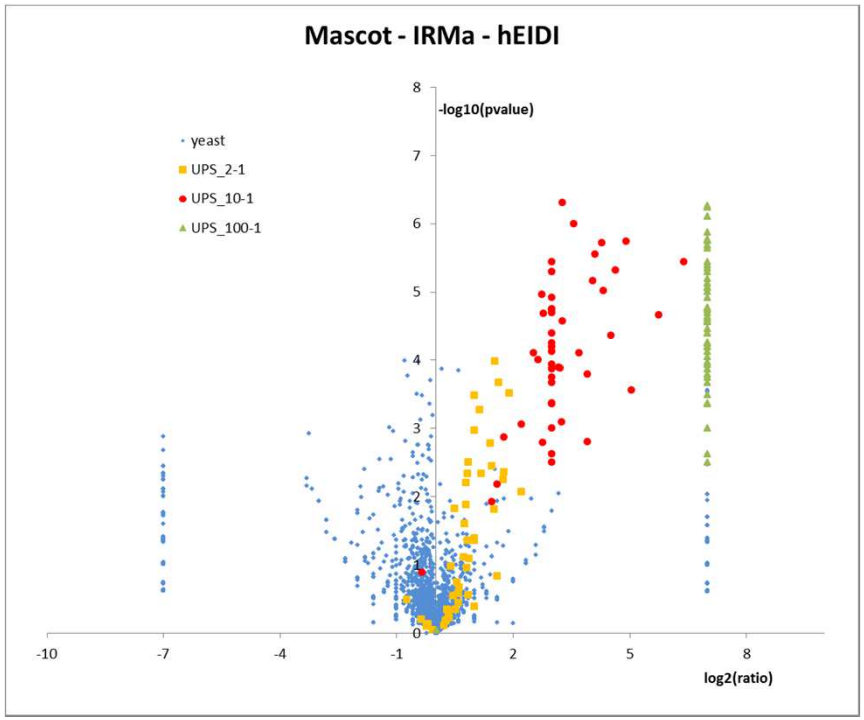
- Type of quantification
- Open-source / freeware / commercial
- Usability
- …

↓

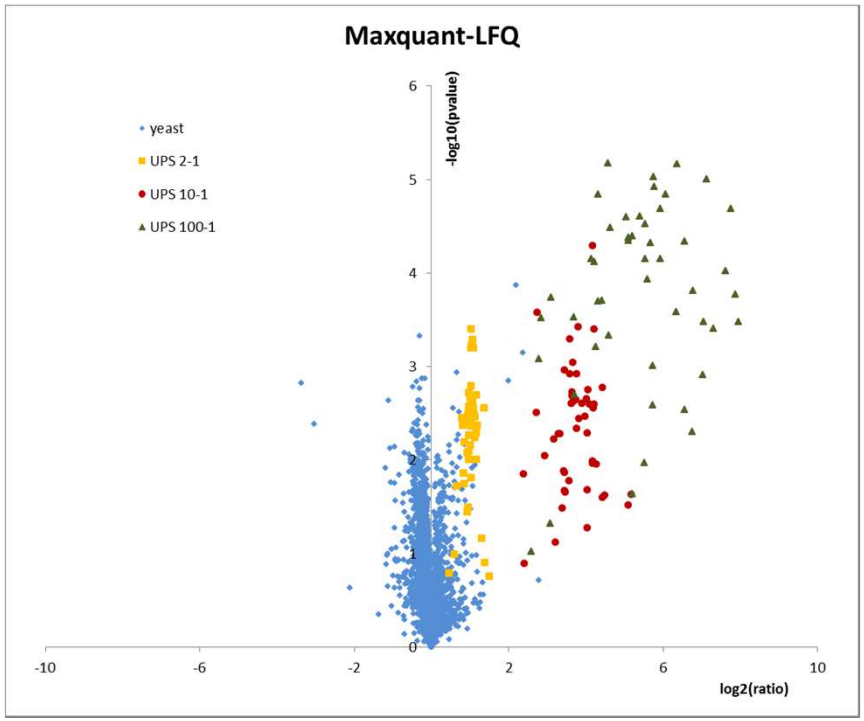
Same statistical assessments used per quantification type

Best results obtained with ProFi standard

Spectral counting



Peak integration

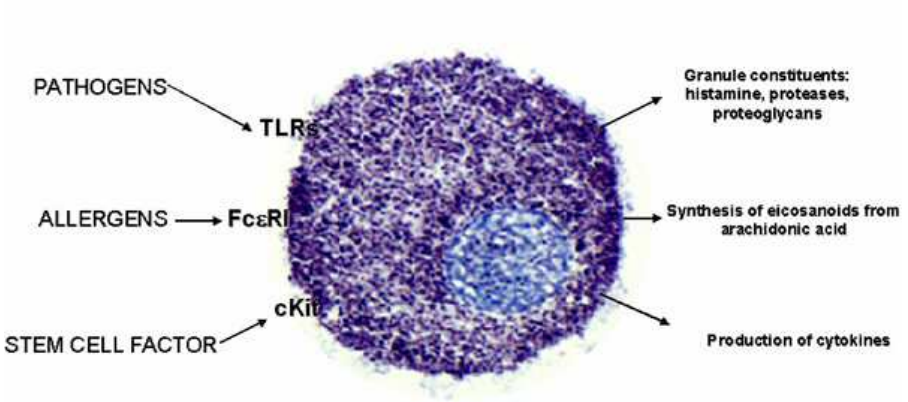


→ **Spectral-count analysis is globally less sensitive but shows to be efficient for successfully detecting proteins with high fold in a « clean » way (low FDP)**

→ **Very good sensitivity can be achieved by label-free MS signal analysis at the cost of a relatively high FDP → pre-statistical assessment procedures? Statistical test?**

Application of label-free MS-based proteomics to « real-life » samples

Immunoreceptor signalling dissection in mast cells



Institut Pasteur
Immunology department

IBENS
Institut de Biologie de l'Ecole Normale Supérieure

Inflammation and signalling in mastocytes

- **Mast cells play critical roles in the initiation of IgE-dependent allergic inflammation**
- **How signalling in mast cells propagates and provides the condition for inflammation to start and develop?**

→ **Dissection of the pathway using AP-MS**

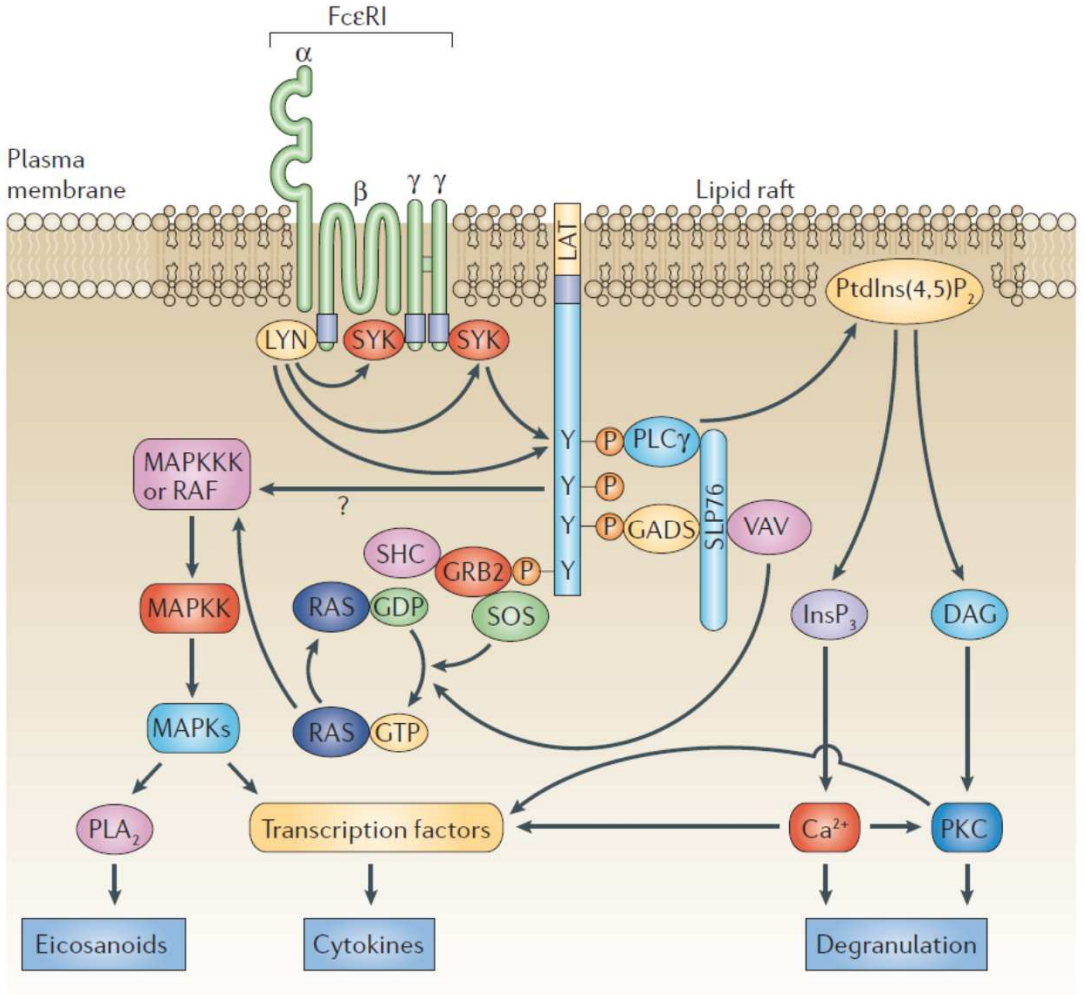
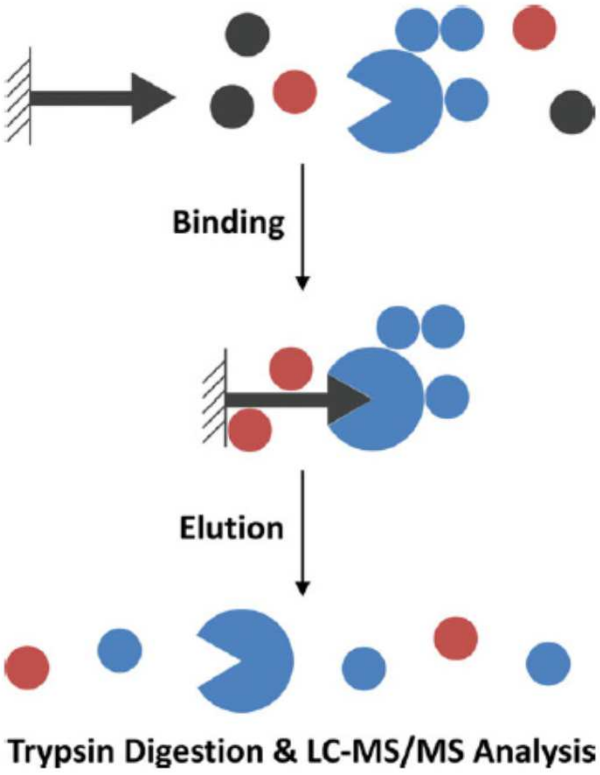


Figure 2 | **The 'principle' signalling cascade in activated mast cells.**

Gilfillan & Tkaczyk, Nat Rev Immunol , 2006

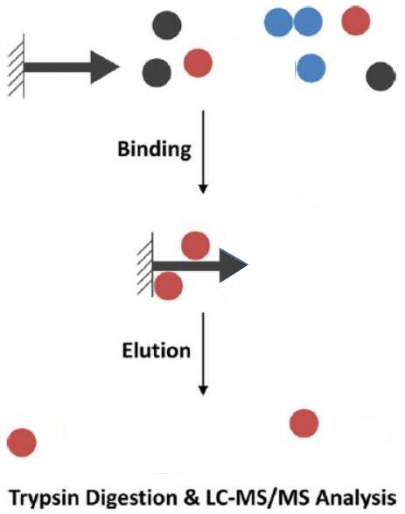
AP-MS



Marcilla M & Albar JP, 2013, IUBMB Life

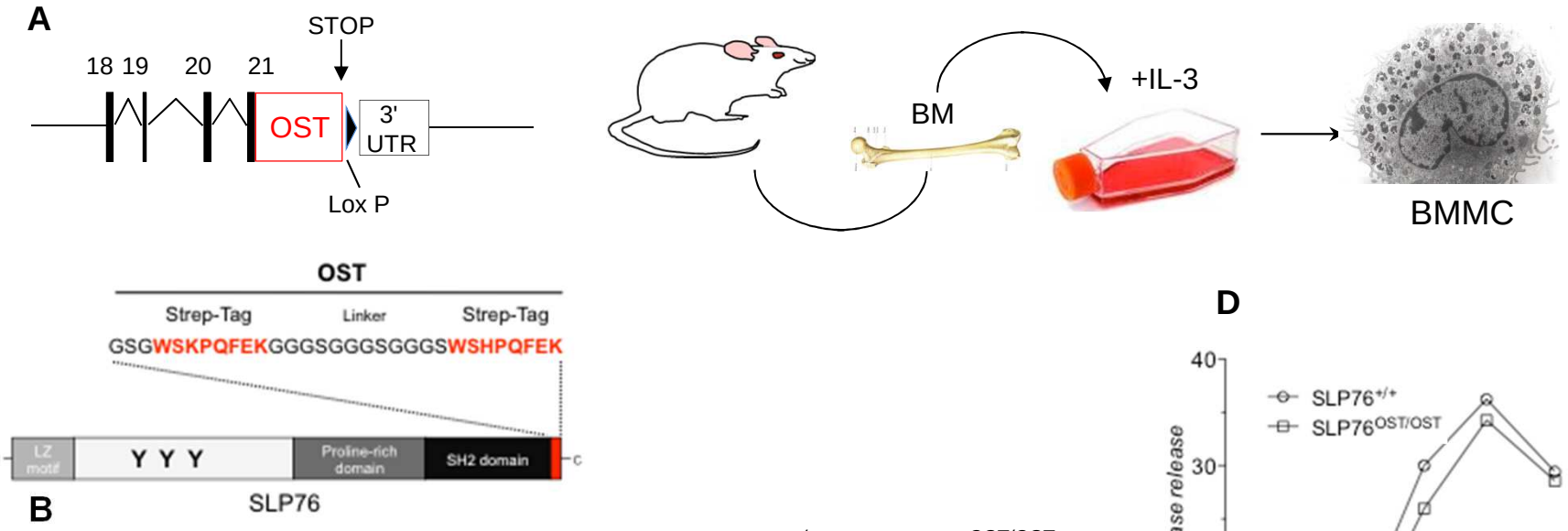
How to discriminate between real partners and unspecific background?

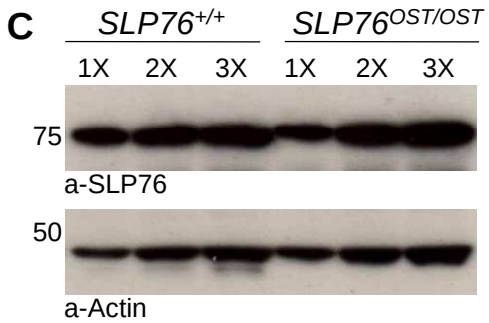
→ **Quantitative proteomics** comparing control and test samples

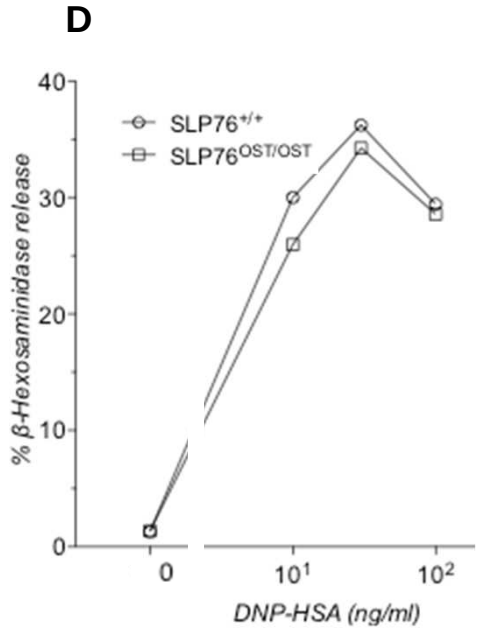


Model of homozygous tagged mice

Slp76$^{ost/ost}$ homozygous mice

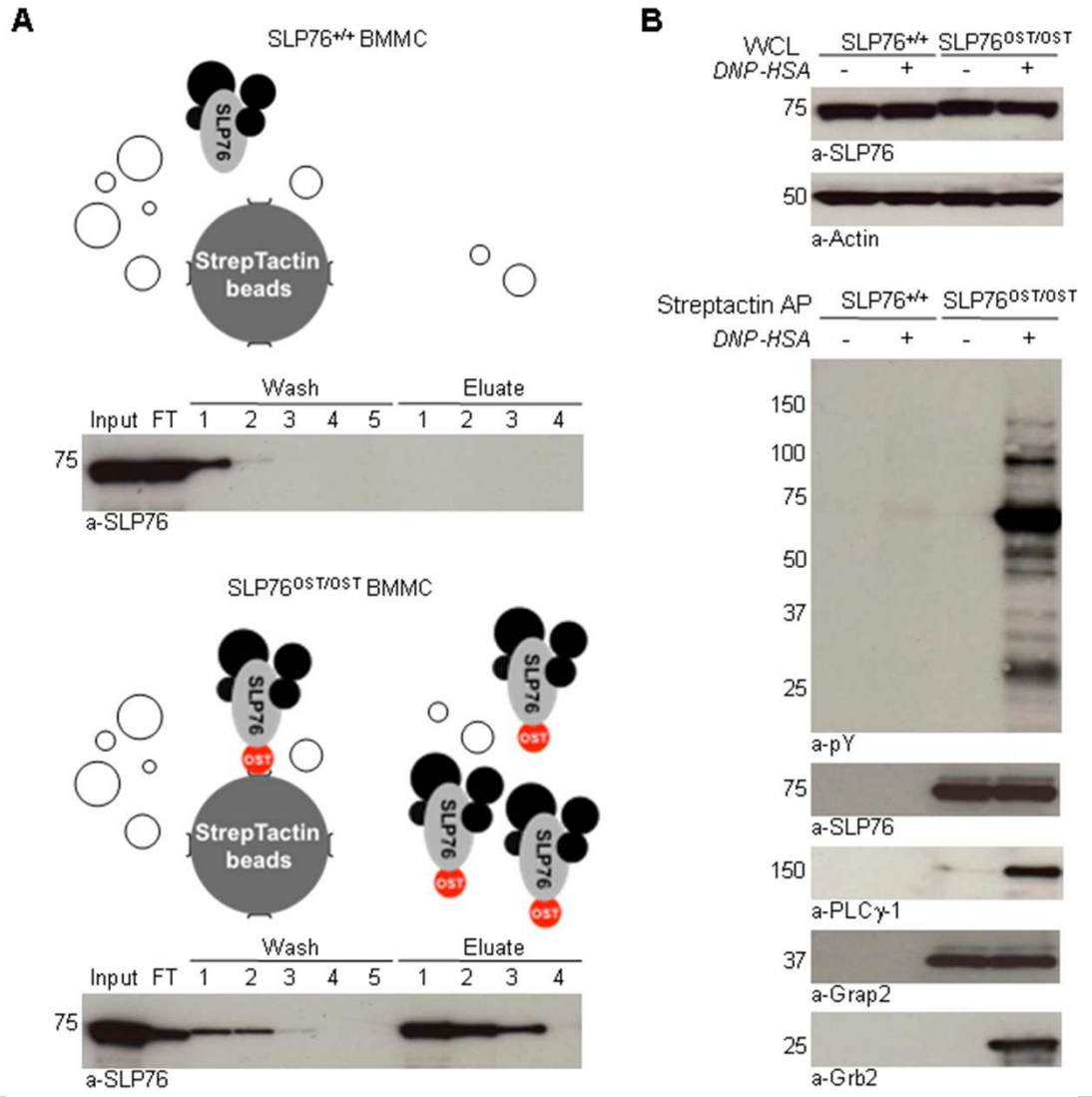






Bounab, Hesse et al., 2013, Mol Cell Proteomics

Affinity purification efficiency



SLP76 interactome in resting and activated mast cells

Resting mast cells WT and OST

Activated mast cells WT and OST

↓

AP

↓

Stacking

↓

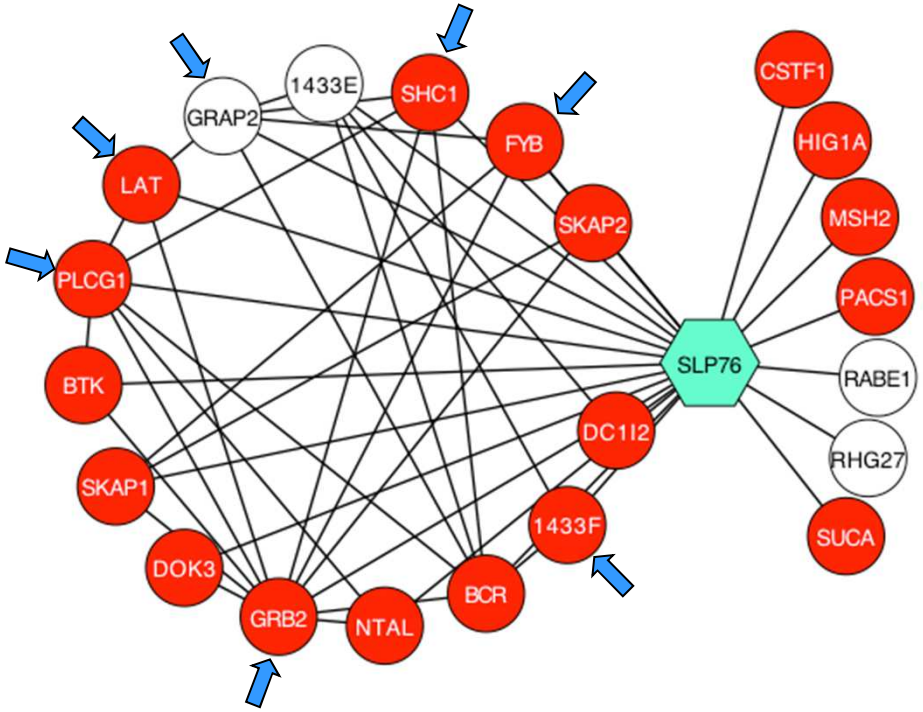
In-gel trypsin digestion

↓

NanoLC-MS/MS
(120 min gradients,
LTQ-Orbitrap Velos pro)

↓

Identification and
quantification of proteins
(*MaxQuant + post-processing*)

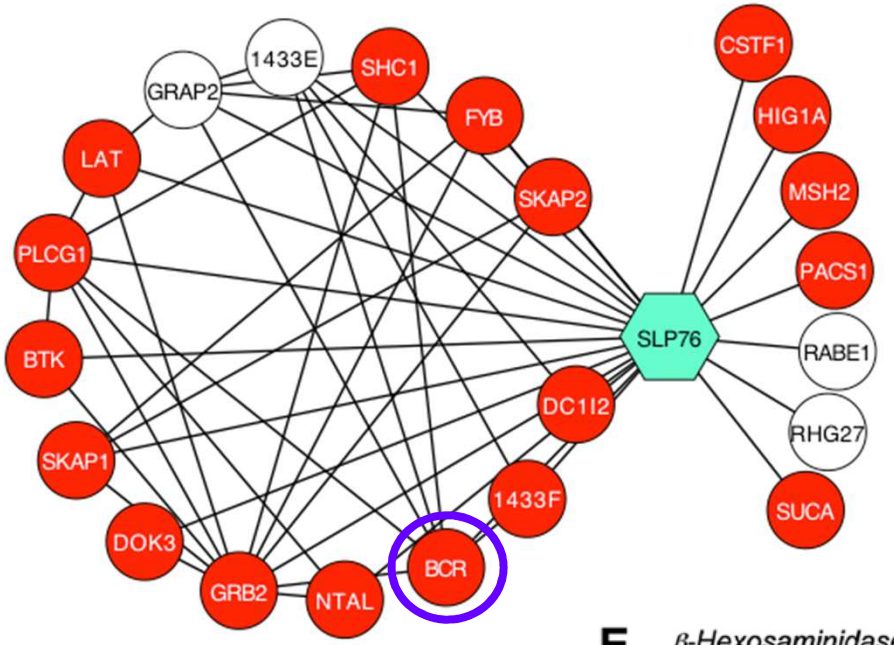


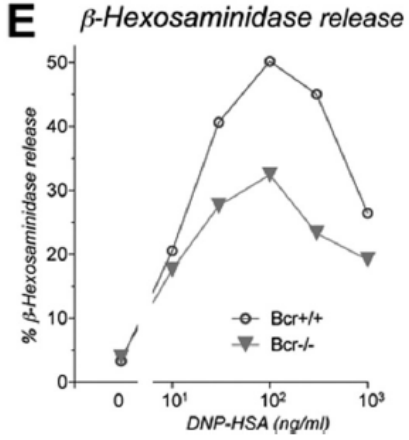
4 binding partners

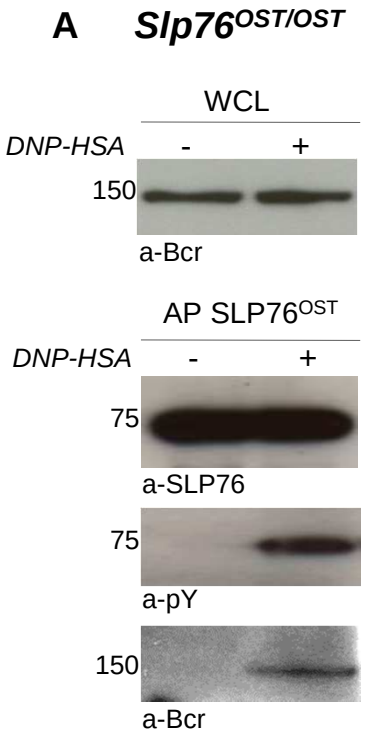
18 binding partners

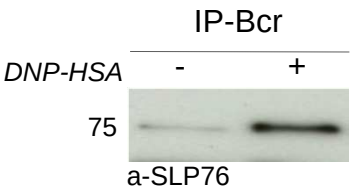
Slp76 interactome uncovered both partners already described in T-cells and novel partners seen only in mast cells.

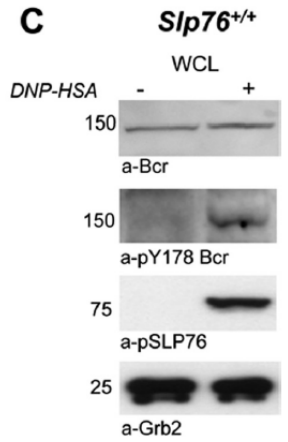
Validation of the SLP76-Bcr interaction in activated mast cells











Conclusions

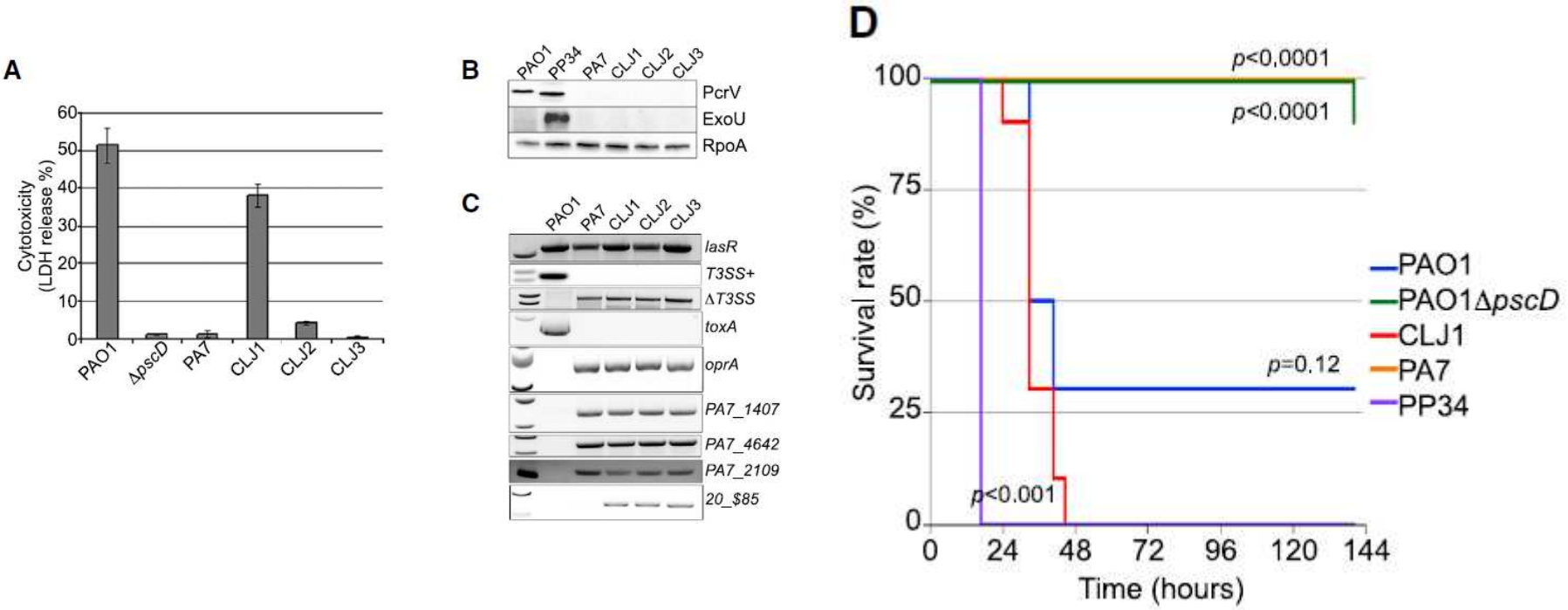
- **First proteomic analysis of FcεRI signalling in primary mouse mast cells.**
- **Description of the SLP76 interactomes in resting and activated cultured mast cells.**
- **Label-free quantitative proteomics successfully identified a novel important molecule in FcεRI signalling.**

Novel molecular weapon in P. *aeruginosa*

Equipe PBRC
iRTSV/BCI
Grenoble

Isolation of a novel clinical strain

Isolated from patient with fatal hemorrhagic pneumonia.

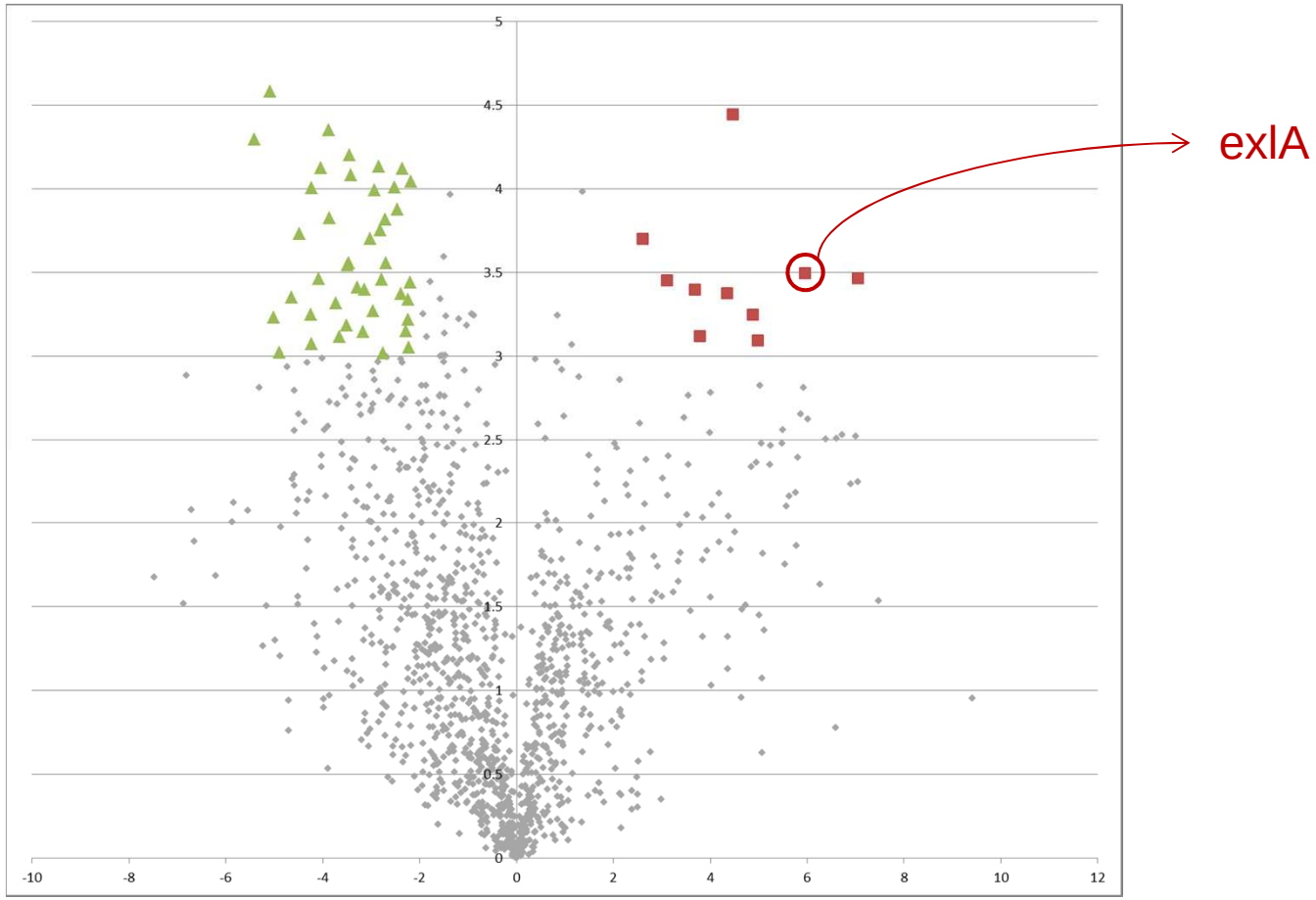


Highly cytotoxic strain lacking classical toxins from T3SS

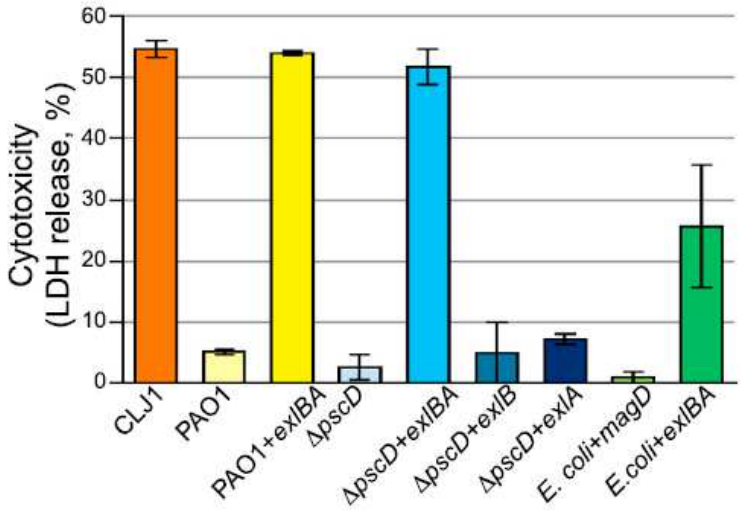
Elsen et al., 2014, Cell Host Microbe

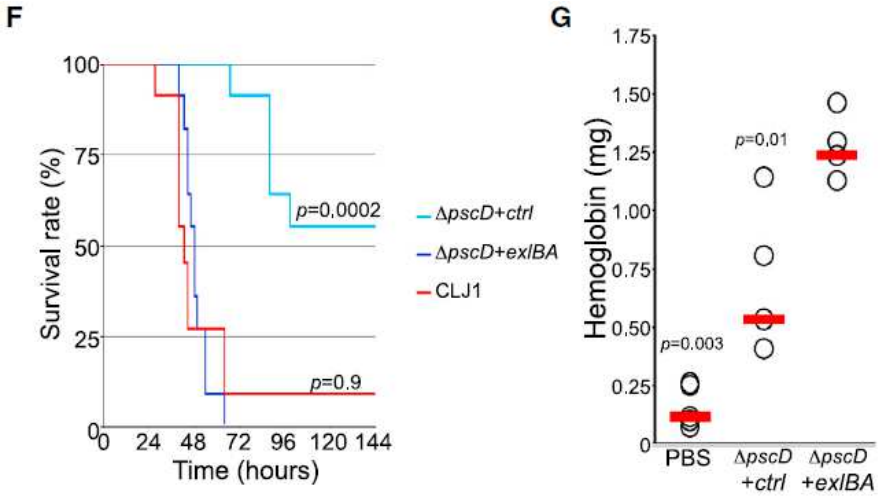
Label-free proteomic analyses

Comparison of PA7 and CLJ1 secretomes



Validation of exlA role





exlA is required and sufficient to provoke cytolysis and fatal hemorrhage in mice infected by P. *aeruginosa* lacking T3SS.

Conclusions

- **Proteomic analysis to find factors of virulence in a novel clinical strain**
- **Identification of several differentially expressed proteins between moderately toxic and highly toxic strains**
- **Uncovering of a novel virulence mechanism**

Conclusions and perspectives

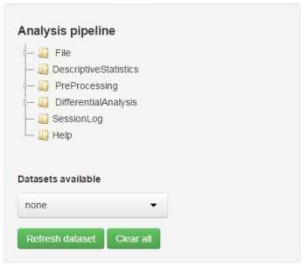
General conclusions

- **Label-free analysis: 2 main approaches in quantitative data extraction**
- **Peak integration suggested to be (become) the most valuable for accurate comparison of proteomes**
- **Complexity of the workflows that can lead to identification and quantification errors**
- **Valuable information obtained from analyses of biological samples**

Perspectives

- Combination of algorithms from different softwares
- Statistical assessment procedures (normalisation, missing values imputation, differential expression testing with error control, …)
- Work at peptide or protein level?
- Increase in data size
- Expanding proteome quantification coverage: moving from DDA to DIA?

ProSta - a GUI for DAPAR package



Acknowledgments

Claire Ramus
Anne-Marie Hesse

Bottom-up PF

Informatics group

KDPD group

Myriam Ferro
Christophe Bruley