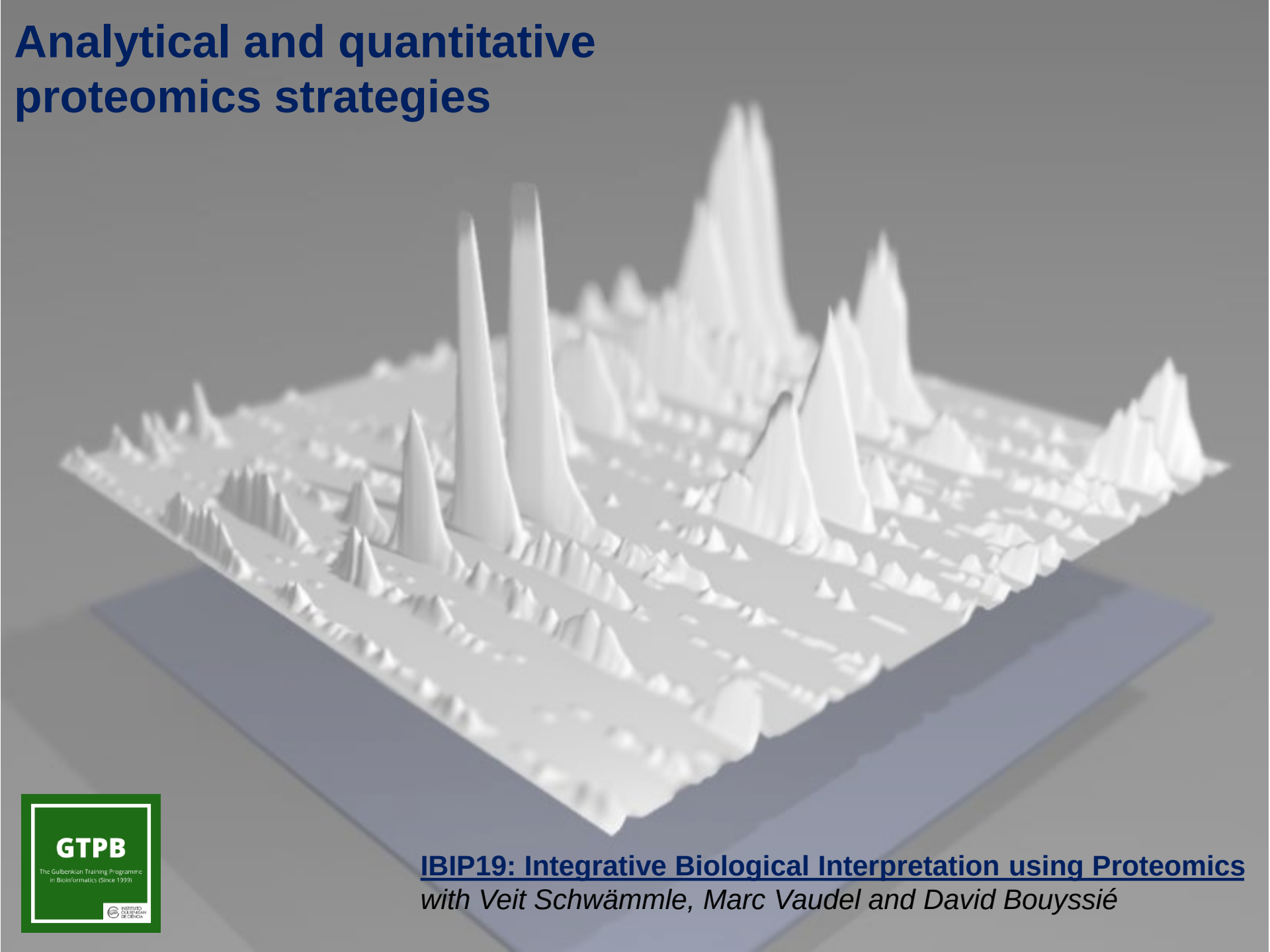
Analytical and quantitative proteomics strategies



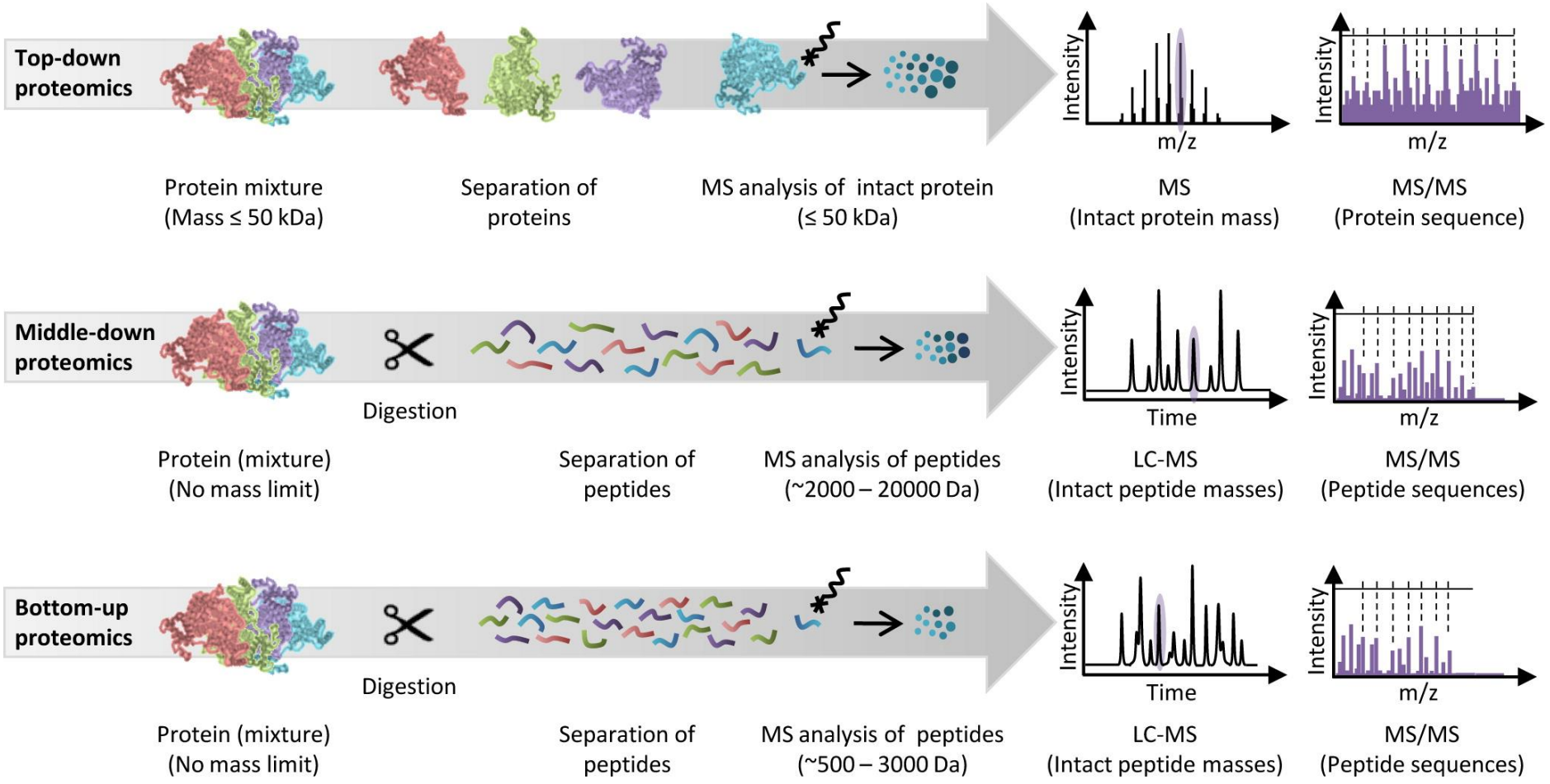
GTPB
The Gulbenkian Training Programme in Bioinformatics (Since 1999)
INSTITUTO GULBENKIAN DE CIÊNCIA

IBIP19: Integrative Biological Interpretation using Proteomics

with Veit Schwämmle, Marc Vaudel and David Bouyssié

Proteomics strategies

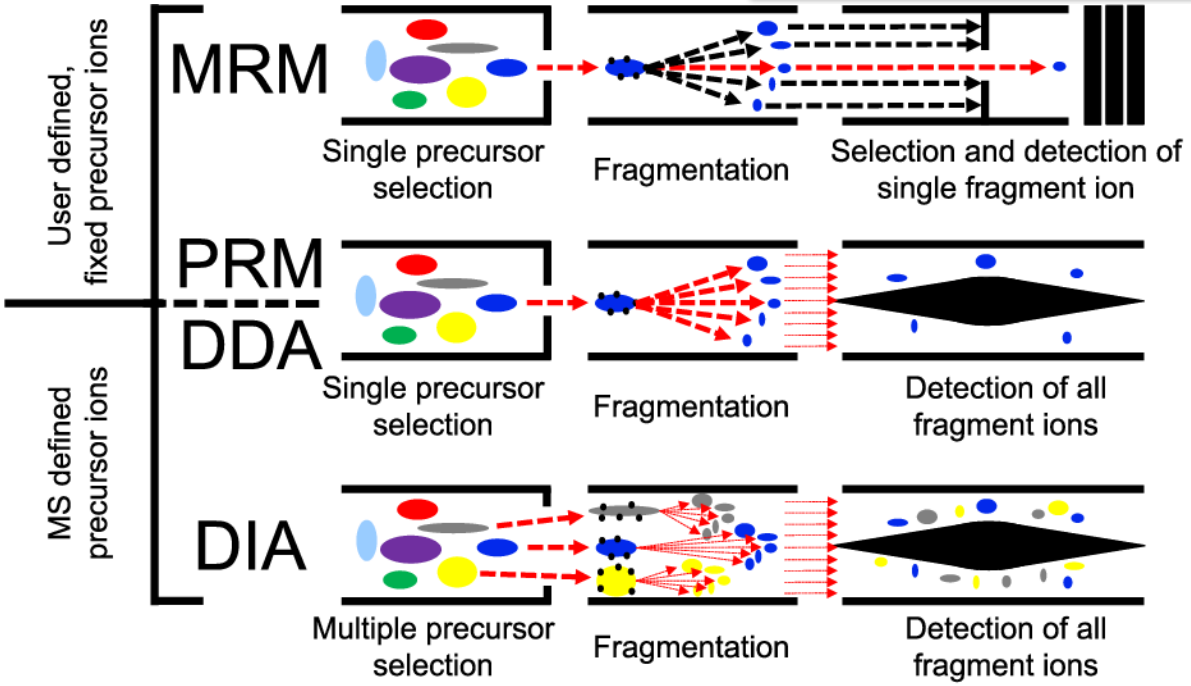
Bottom-Up, Middle-down and Top-Down



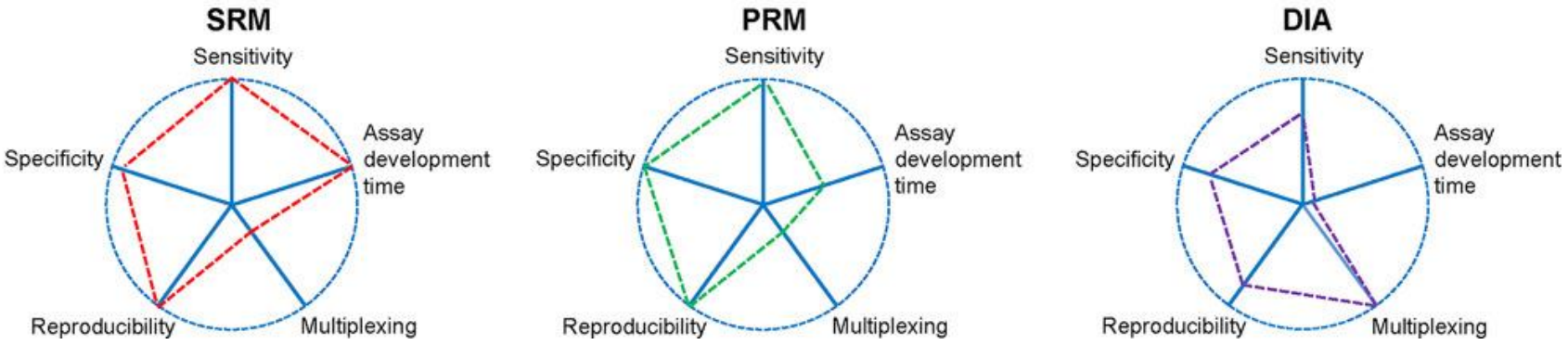
Linda Switzar, J. Proteome Res., 2013

Proteomics strategies

Targeted *versus* untargeted (shotgun)



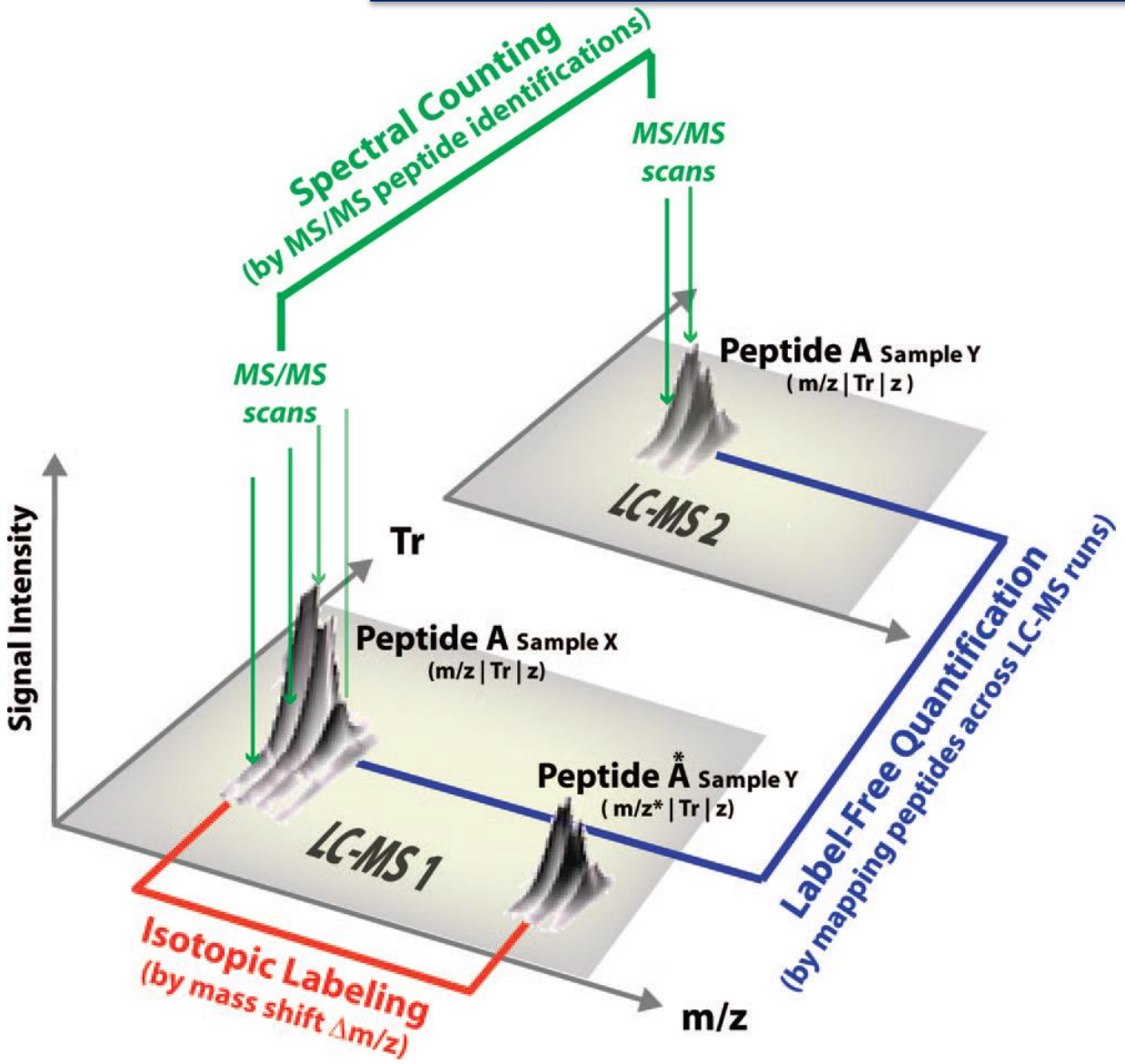
Hu A, Noble WS, Wolf-Yadlin A. F1000Res. 2016 Mar.
Technical advances in proteomics: new developments in data-independent acquisition.



Shi T *et al.* Proteomics. 2016 Aug. Advances in targeted proteomics and applications to biomedical research.

Proteomics strategies

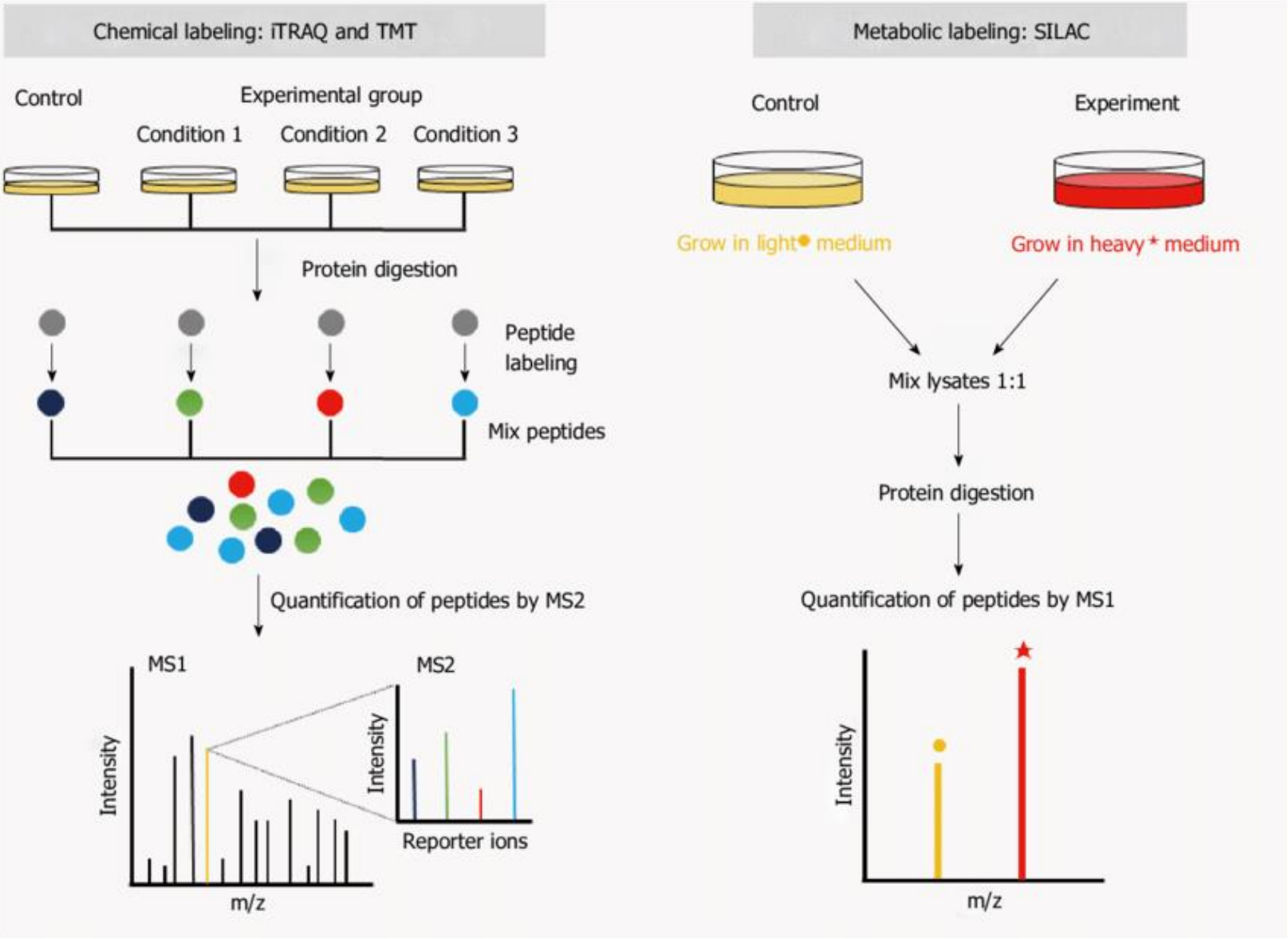
Isotopic labeling *versus* label-free



Mueller LN et al. *J. Proteome Res*. 2008, 7, 51-61.

Proteomics strategies

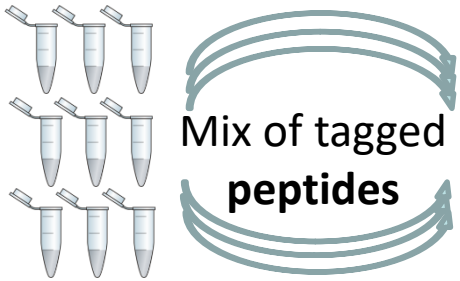
Isotopic labeling *versus* isobaric tagging



Kang C *et al.* World J Gastroenterol. 2016 Oct

Quantitative proteomics

Isobaric tagging: typical workflow

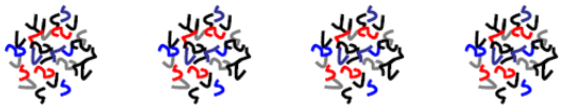


Mix of tagged **peptides**

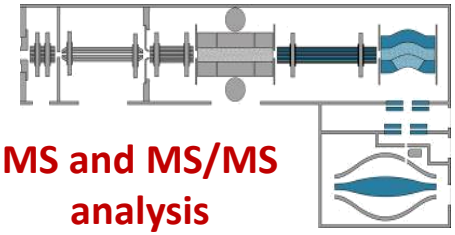
Peptide fractionation

(optional)

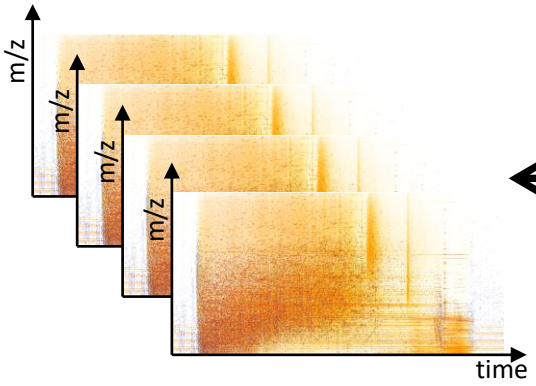
Several peptides fractions



nanoLC



MS and MS/MS analysis



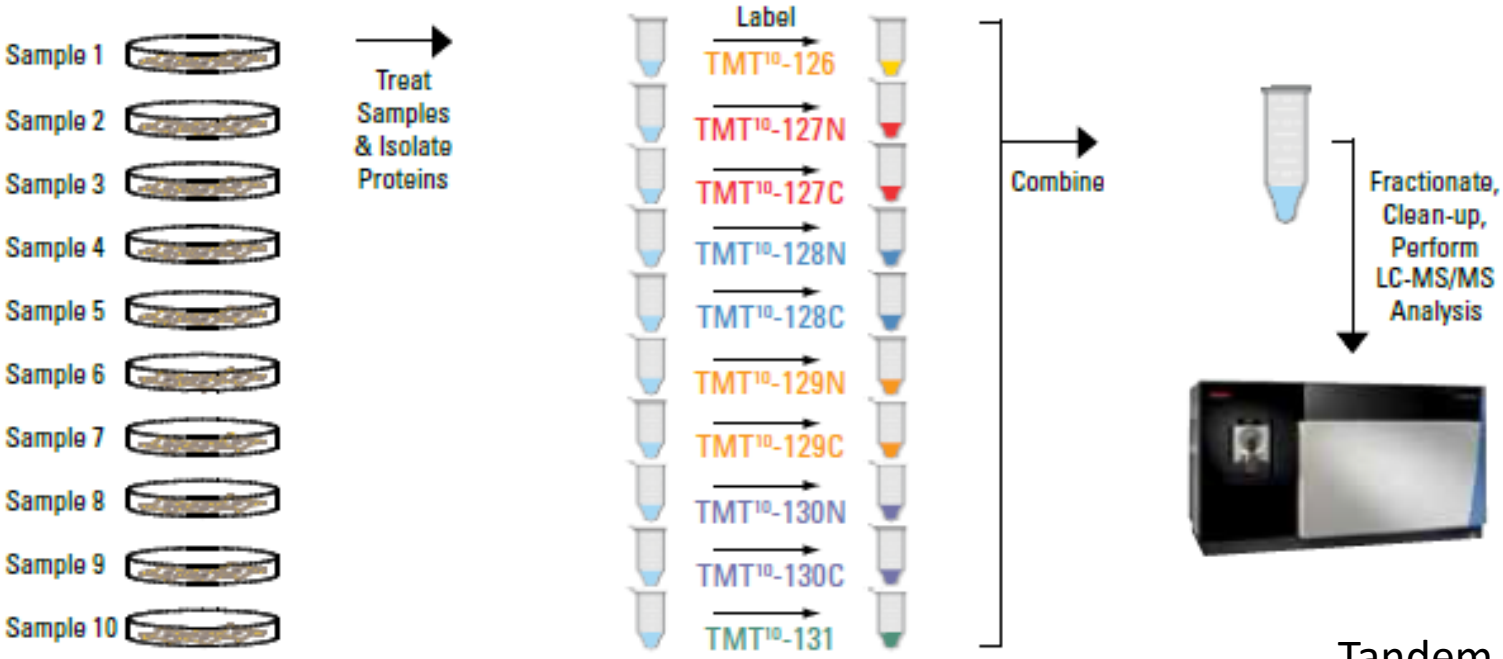
RAW files of analyzed fractions

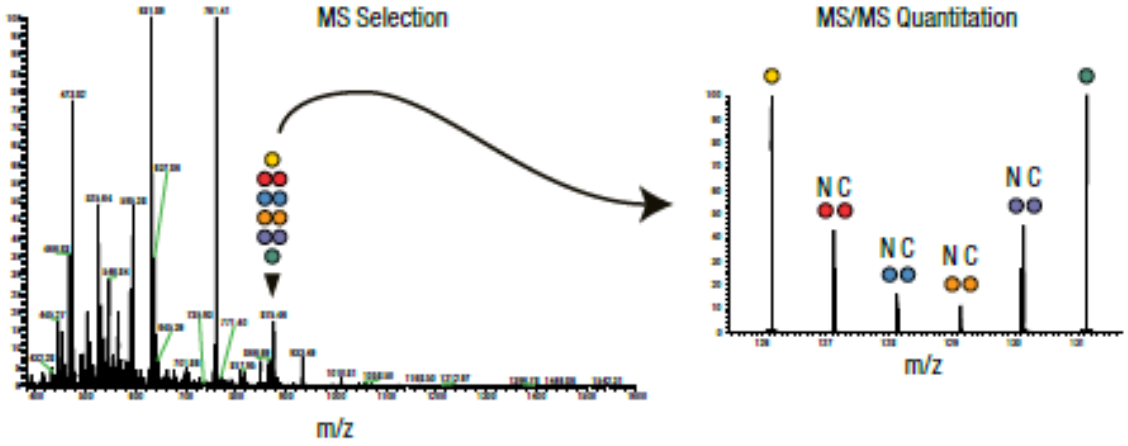
Software Analysis

Lists of quantified peptides and proteins

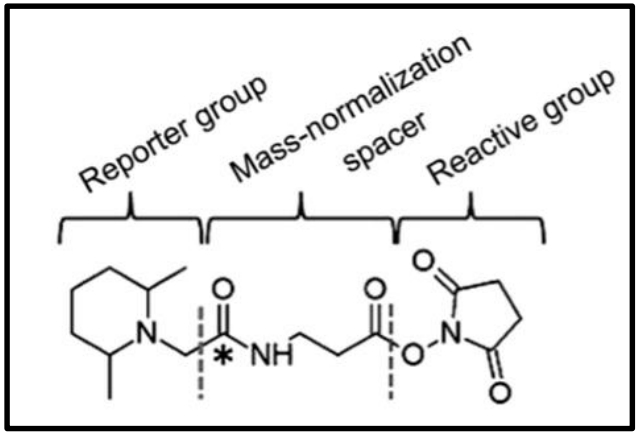
Quantitative proteomics

Isobaric tagging: Tandem Mass Tags





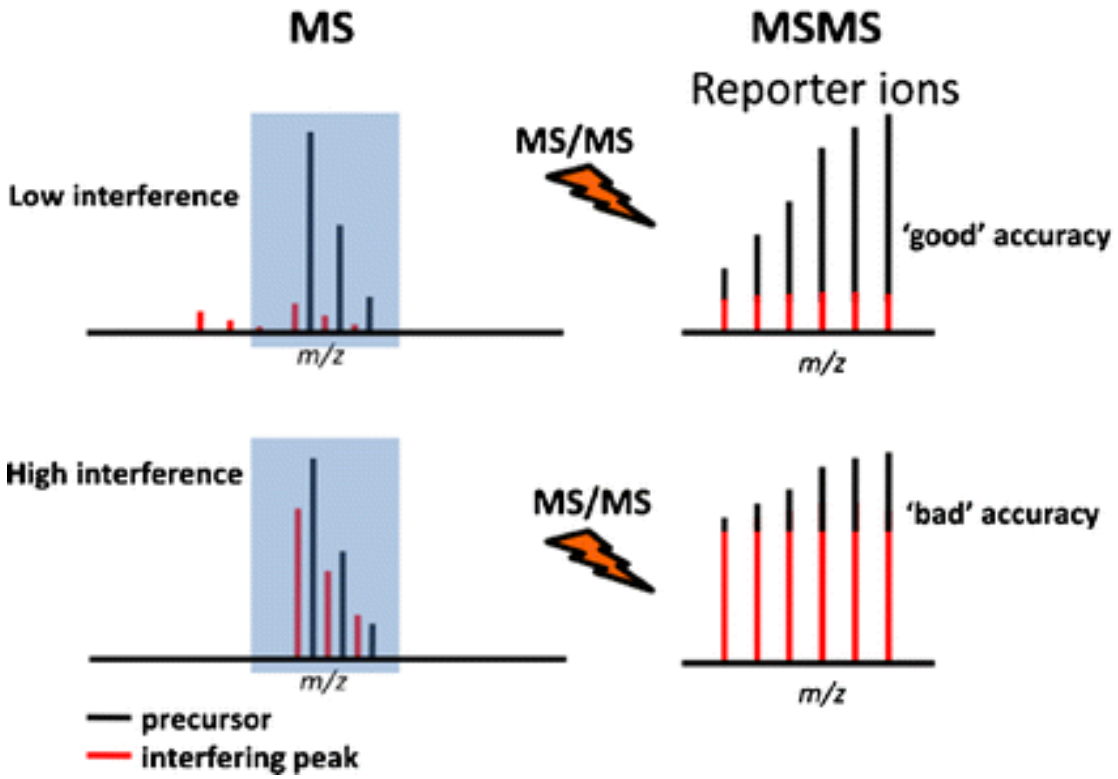
Tandem Mass Tag



Quantitative proteomics

Isobaric tagging: main issues

Impact of sample complexity and signal intensity on quantification accuracy.



Analytical and Bioanalytical Chemistry
September 2012, Volume 404, Issue 4, pp 939–965 | Cite as

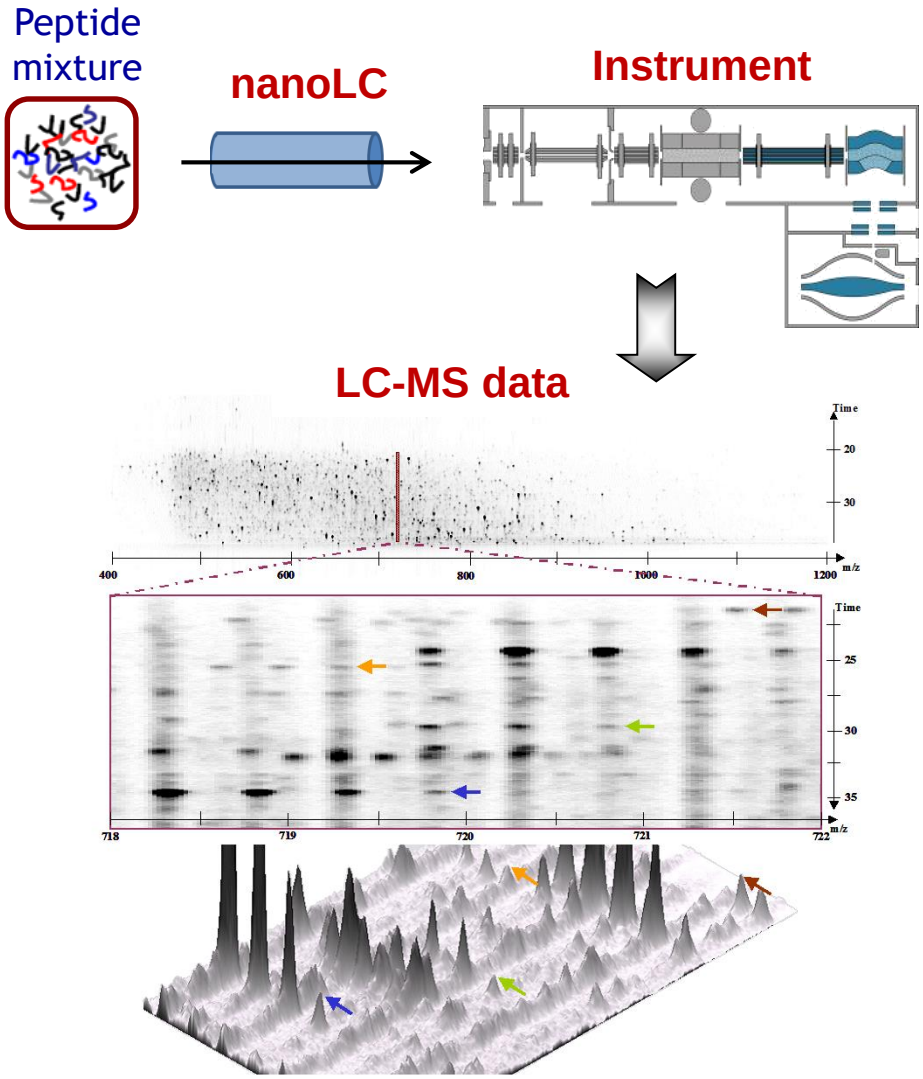
Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present

Authors Authors and affiliations

Marcus Bantscheff, Simone Lemeer, Mikhail M. Savitski, Bernhard Kuster

Quantitative proteomics

Label-free quantification: main issues



Source : MSight (Swiss Institute of Bioinformatics, www.expasy.org)

- **Global scale analysis**
  - Thousands of peptides are monitored
  - Simultaneaous information about peptide identity (MS/MS) and quantity (MS)
- **LC-MS data**
  - Three dimensionnal data (elution time, m/z and intensity)
  - Peak intensities are directly correlated to peptide abundance
  - Very complex: each peptide exhibits one or more 3D isotopic patterns (charge states)

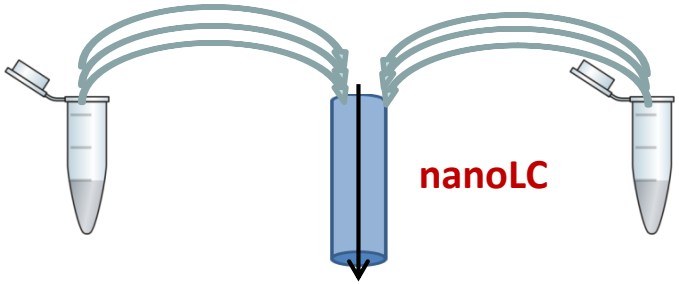
Requires sophisticated algorithms to provide high-quality data

Quantitative proteomics

Label-free quantification: typical workflow

Multiple injections of the same sample or of biological replicates

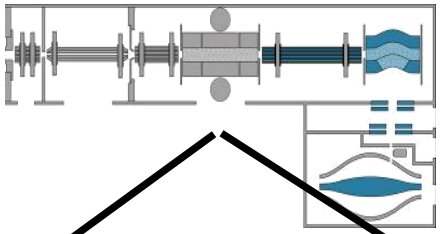
Peptide mixture from **condition A**

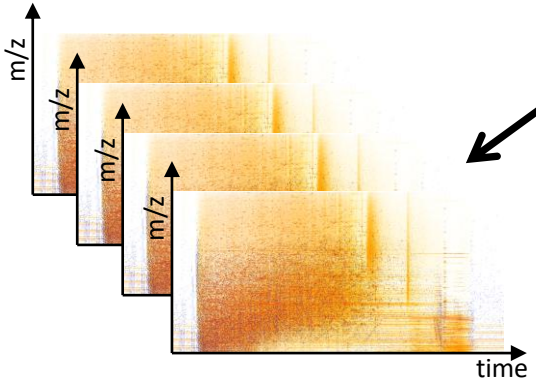


nanoLC

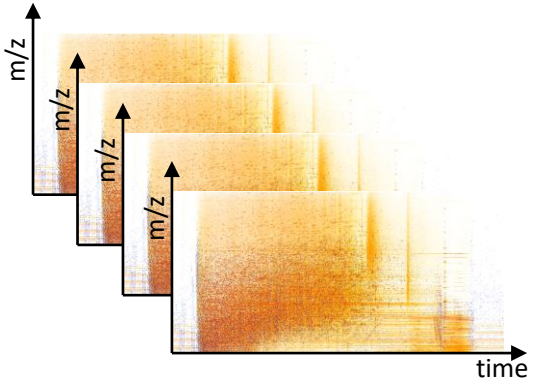
Peptide mixture from **condition B**

Instrument





Software Analysis



Condition A RAW files

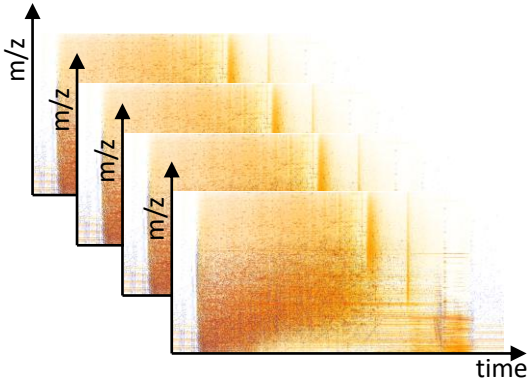
Condition B RAW files

Lists of quantified peptides and proteins

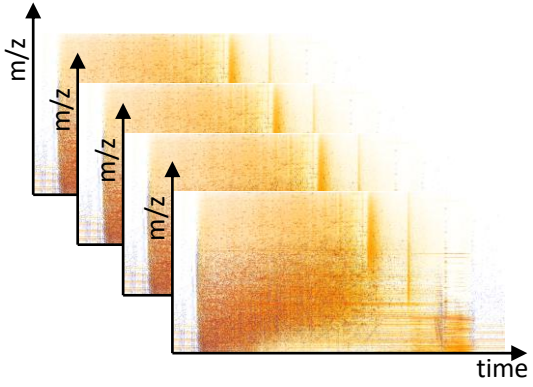
Quantitative proteomics

Label-free quantification: typical data processing workflow

Condition A RAW files



Condition B RAW files



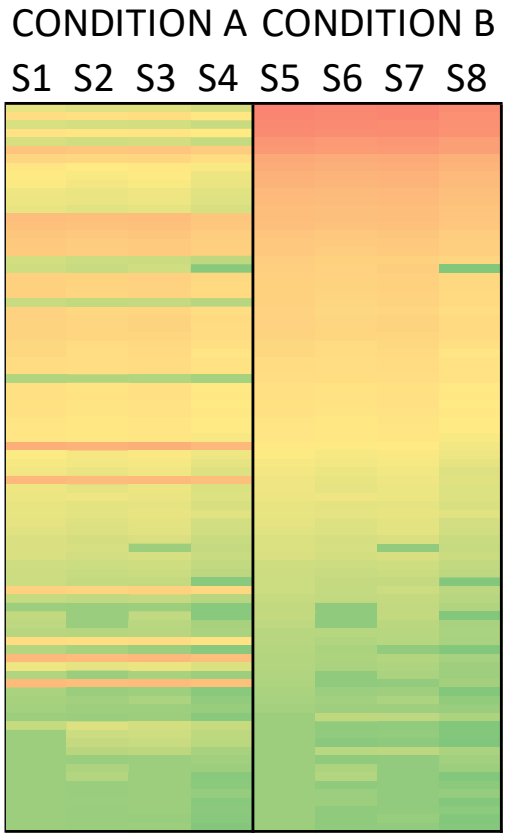
Data-processing steps:

1. Features detection
2. Features annotation
3. LC-MS maps alignment (chromatographic reproducibility)
4. Intensity normalization

(Various sources: samples, protocols, instrument)

5. Peptides -> proteins (intensity summarization)
6. Missing value inference (MS/MS reproducibility)

Final protein intensity matrix

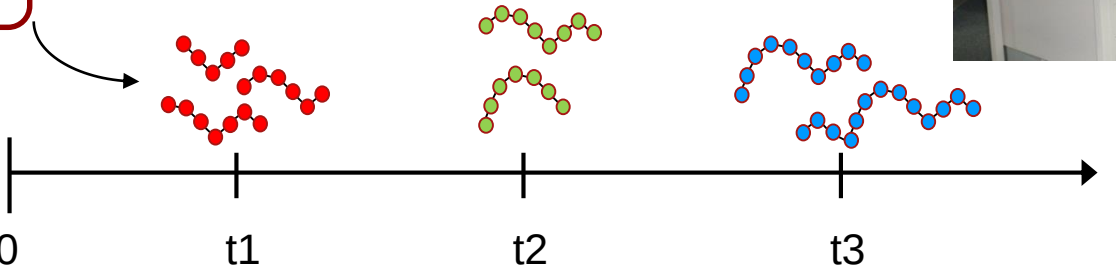


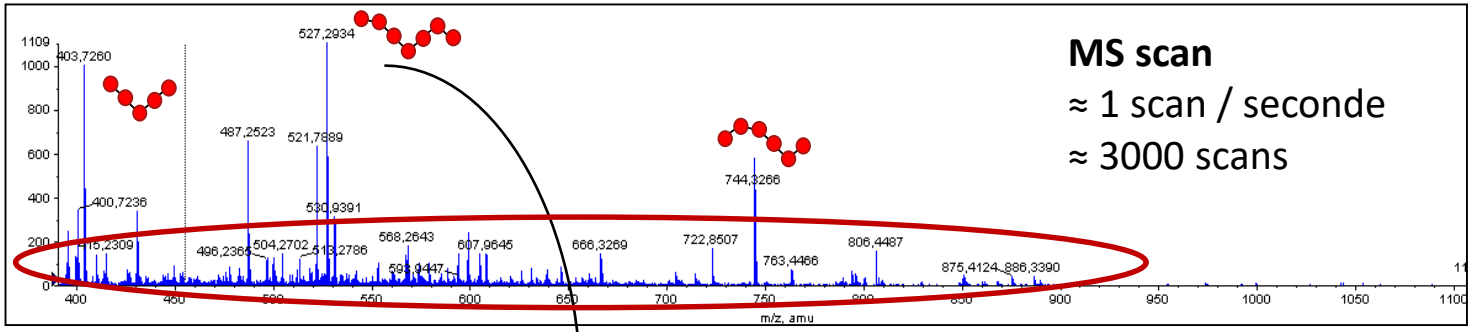
Quantitative proteomics

Label-free quantification: the stochastic selection of precursor ions

Peptide mixture

Peptides chromatographic separation

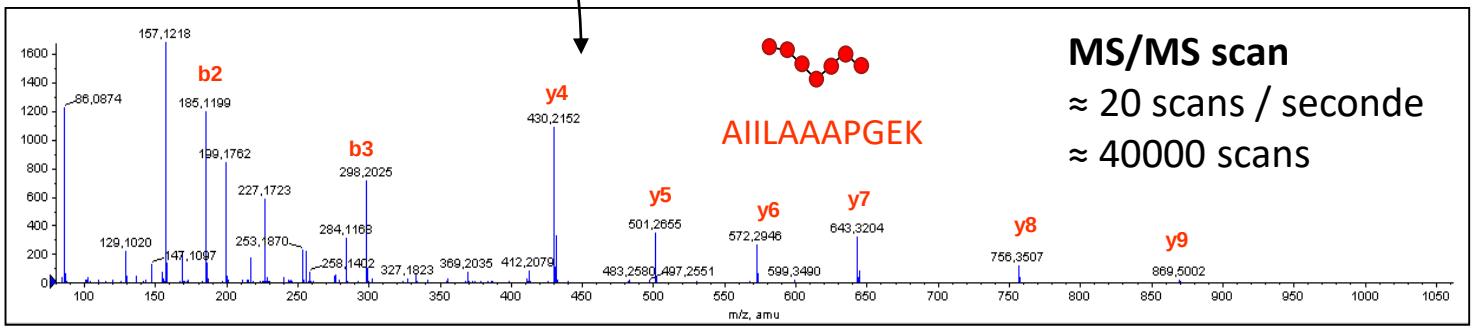




MS scan

≈ 1 scan / seconde

≈ 3000 scans



MS/MS scan

≈ 20 scans / seconde

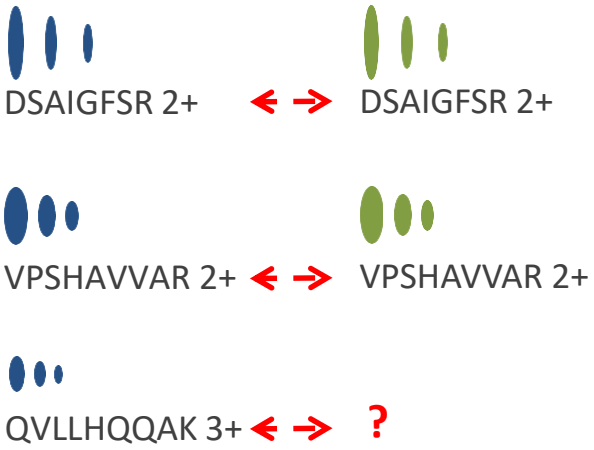
≈ 40000 scans

Only a subset of the observed MS features are selected for MS/MS fragmentation

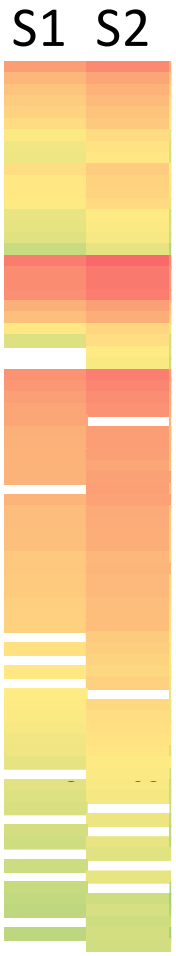
Quantitative proteomics

Label-free quantification: main issues

Identified /quantified features in two samples to be compared



Once annotated it is really easy to associate features corresponding to the same peptide ion, allowing thus to compare intensities.



Comparison of features intensities

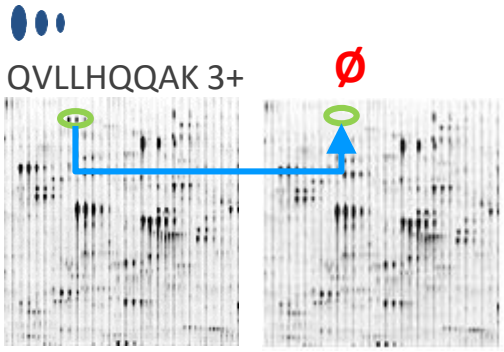
Issue: we may observe features in a given sample but not in the other.

Multiple reasons:

- peptide really absent from the sample
- non detected MS signal
- no MS/MS triggered by this instrument

To be certain that we have a real absence of material in the considered sample, it is needed to recover these missing observations.

This procedure is called "cross-assignment" or "match between runs" (MaxQuant).

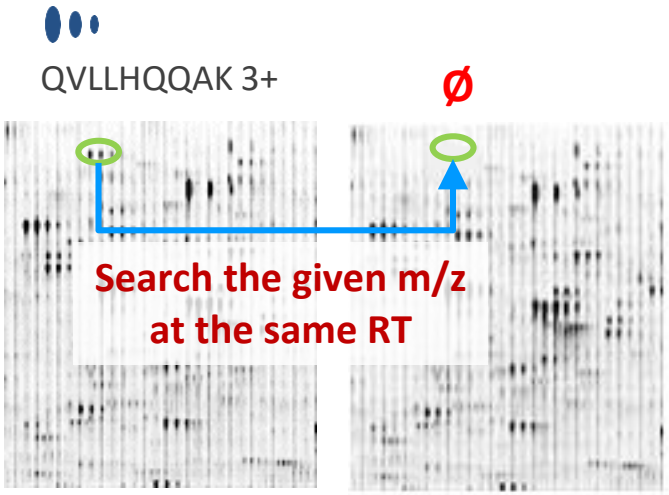


Quantitative proteomics

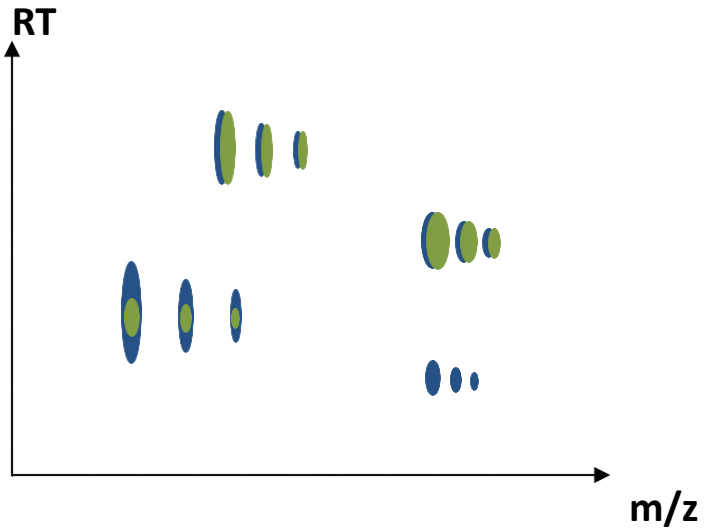
LFQ: “cross-assignment” and issues of chromatographic reproducibility

The **“cross-assignment”** procedure recovers features by matching their **m/z and RT coordinates**.

Feature “cross-assignment”



But RT dimension is not reliable



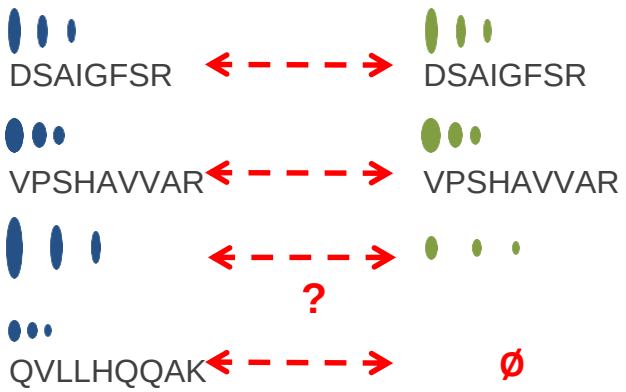
Major issue of label-free quantitative analysis:
chromatographic elution of peptides is not always reproducible

Quantitative proteomics

Label-free quantification: retention time alignment of LC-MS runs

Two main methods were used so far to compute LC-MS maps alignments:

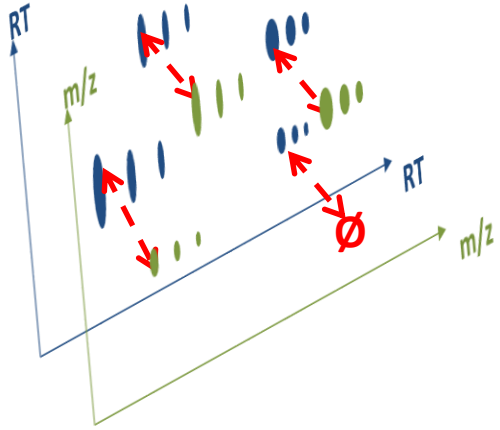
Peptide sequence mapping



Uses MS/MS identification knowledge.
It's the most used method nowadays.

Very robust: can deal with very complex LC-MS maps and important chromatographic time deviations.

Feature coordinates mapping

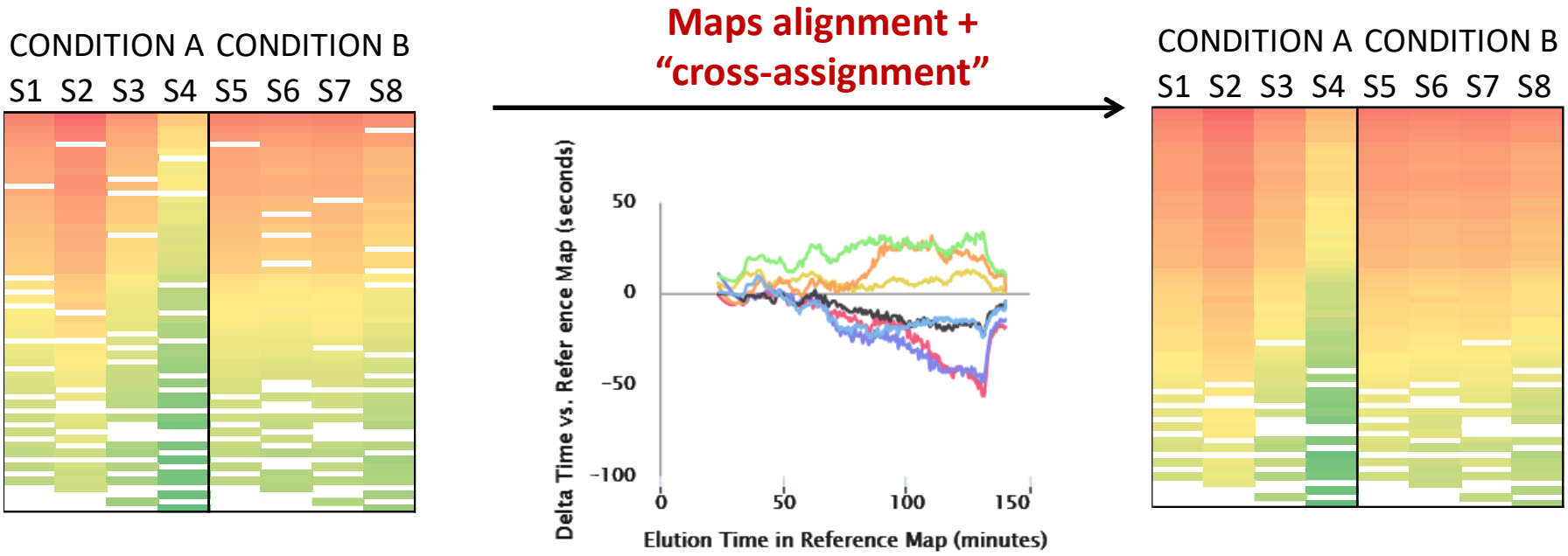


Uses m/z+RT coordinates of the detected features.
Implemented in pioneer tools (OpenMS, VIPER).

Less robust (many features have very close m/z and RT coordinates) but can be used when MS/MS information is not available (lipidomics studies for instance).

Quantitative proteomics

Label-free quantification: recovering data with the “cross-assignment”



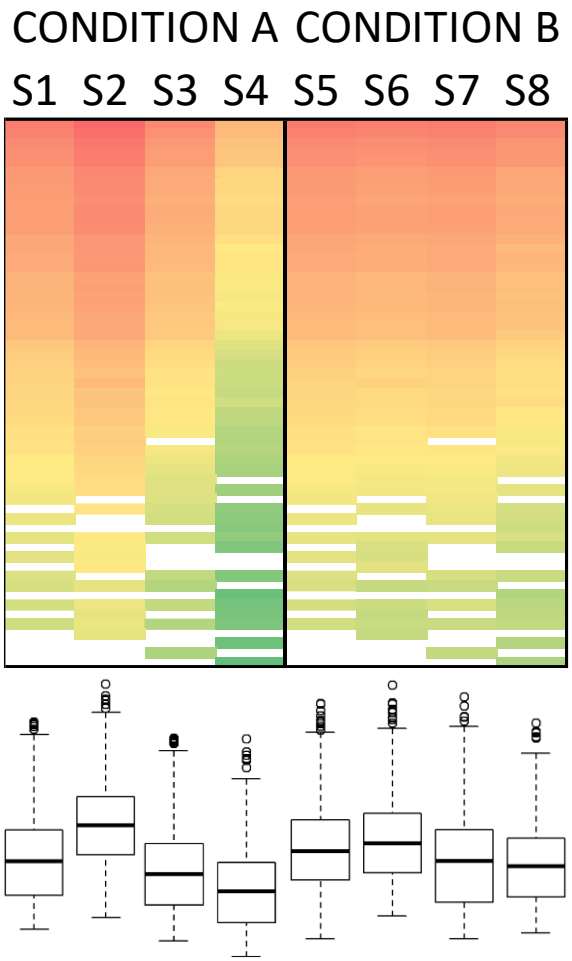
This processing is essential to recover the intensity of peptide ions which are detected by the instrument but not identified (no MS/MS or poor identification score).

The **m/z and RT tolerances** entered in the software configuration must be carefully chosen:

- the m/z tolerance must be determined in function of your instrument resolution
- the time tolerance is more complicated to configure (check software advices)

Quantitative proteomics

Label-free quantification: normalization

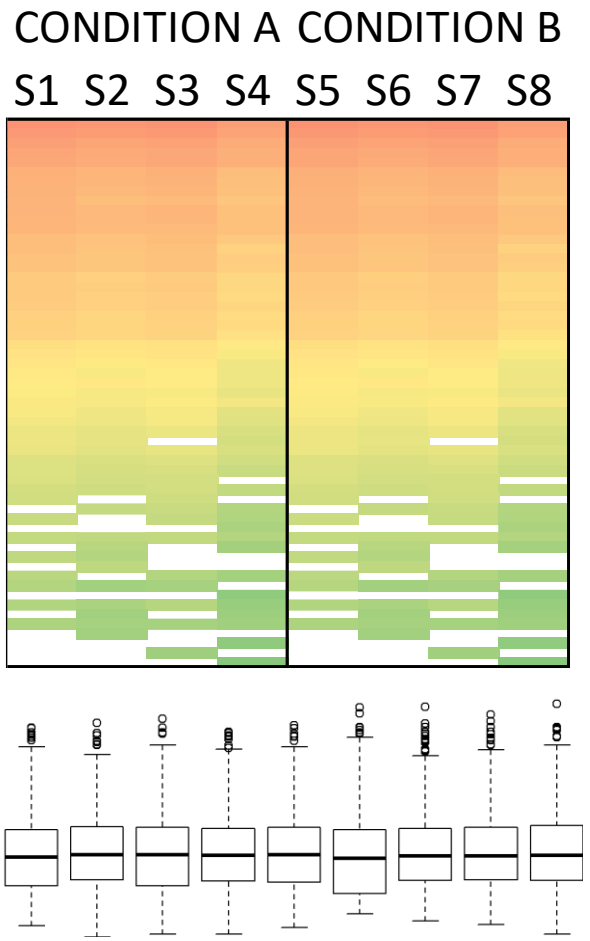


Intensity normalization

Classical methods:

- median centering
- mean centering
- sum equalization

But other methods exist: quantile normalization, ratios based normalization (MQ LFQ), etc…

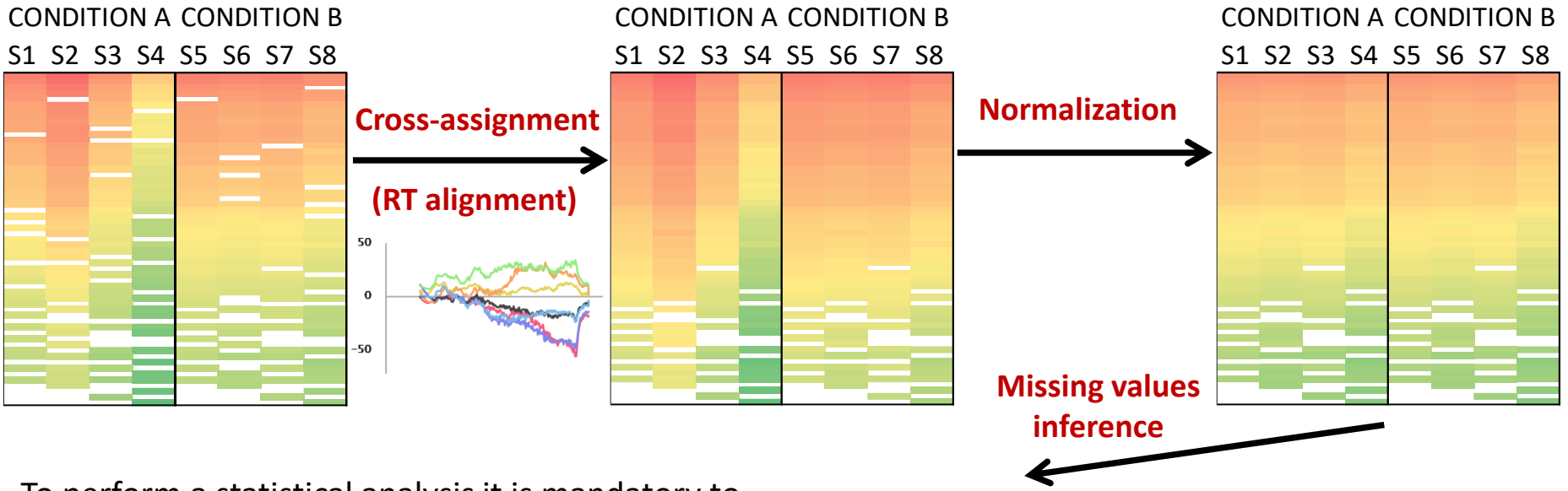


<u>Important assumption</u>: the majority of the proteome typically does not change between compared samples, so the average behavior can be used as a relative standard

Suggested reference: O'Rourke MB *et al*. What is Normalization? Proteomes. 2019 Aug 22;7(3)

Quantitative proteomics

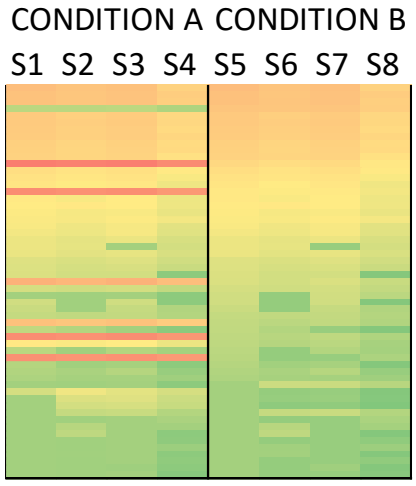
Label-free quantification: summary



To perform a statistical analysis it is mandatory to provide fully defined values.

We need to replace empty values by something else :

- custom noise => low percentile (e.g. 1%)
- random noise values (gaussian model)
- based on statistical models



Statistical analysis