Stratégies d'acquisition/interprétation des données en protéomique

Christophe Bruley

Analyse et interprétation de données métabolomiques et protéomiques : points communs et spécificités

Analyses protéomiques par spectrométrie de masse

Préparation

déplétion

fractionnement

enrichissement

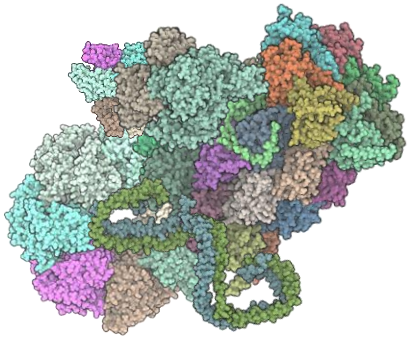
purification

digestion

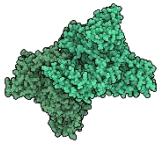
enrichissement

Une information **incomplète**

Complexité ?
Gamme dyn. ?



protéome



protéine

peptide

Une mesure **indirecte**

Qui (id) ?
Proteoformes ?
PTMs ?
Abondance ?
Partenaires ?
Fonction ?

Top-down

Qui (id) ?
Proteoformes?
Abondance ?

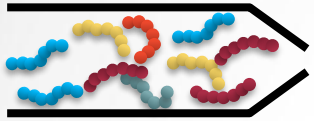
Bottom-up

Qui (id) ?
PTMs ?
Abondance ?

Spectrométrie

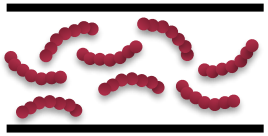
Modes d'acquisition MS

chromatographie

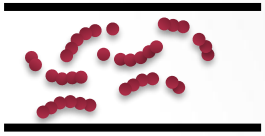


analyseur 1

fragmentation



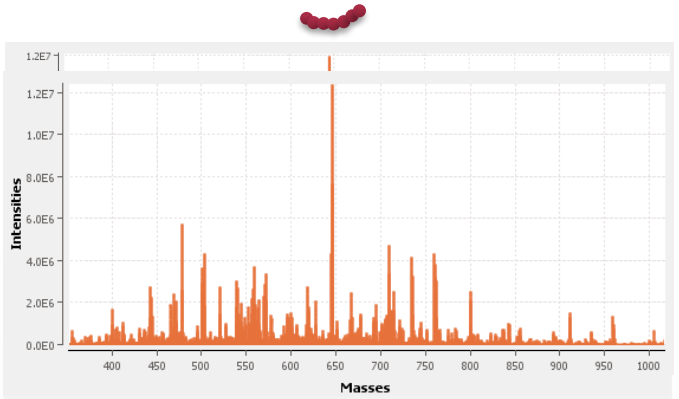
analyseur 2

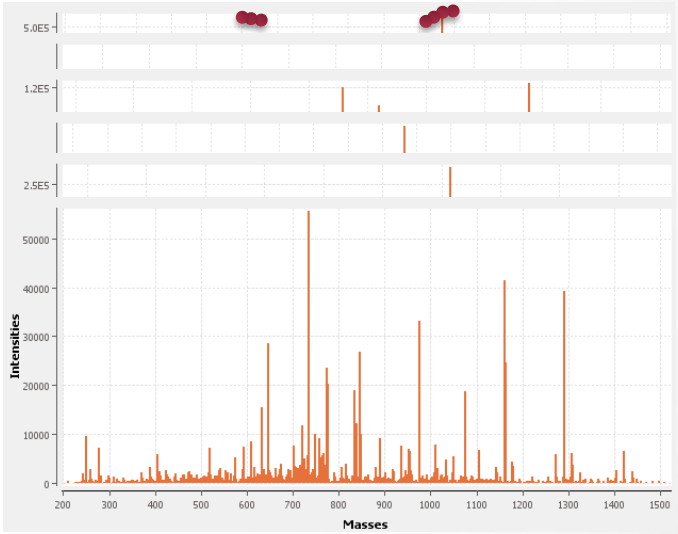


Spectre MS

Spectre MS/MS

~ 4-5 ordres de grandeur

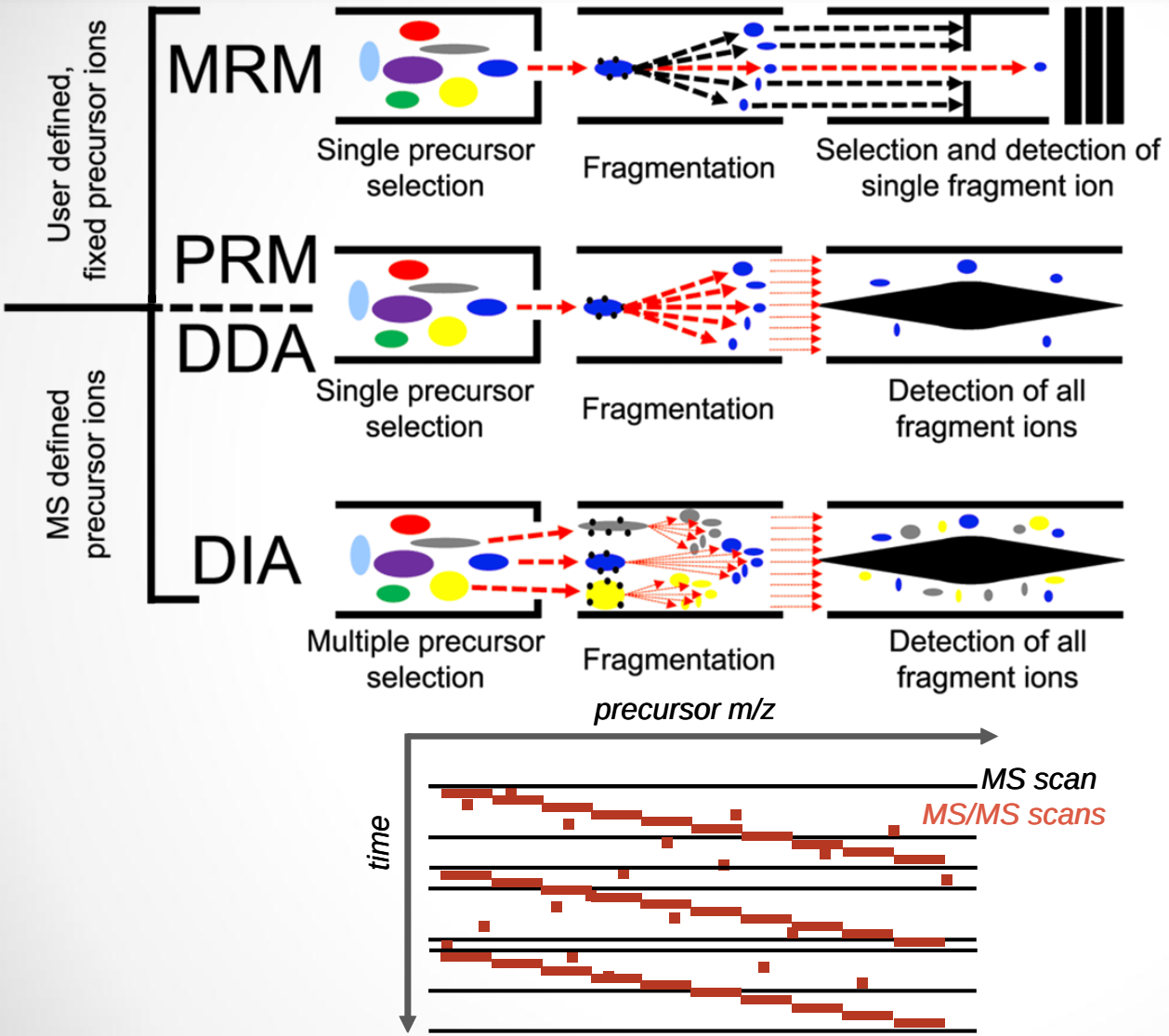




1h de gradient chromatographique :

- ~ 7 000 cycles
- ~ 7 000 spectres MS
- ~ 25 000 spectres MS/MS
- ~ 350 Mo données

Modes d'acquisition MS



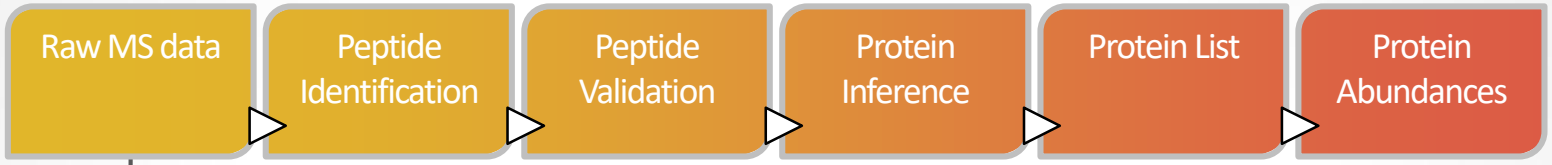
Sélection systématique
Exhaustivité faible

Sélection aléatoire
Exhaustivité moyenne

Sélection systématique
Exhaustivité élevée

Hu A, Noble WS and Wolf-Yadlin. F1000Research 2016, 5:419 (doi: 10.12688/f1000research.7042.1)

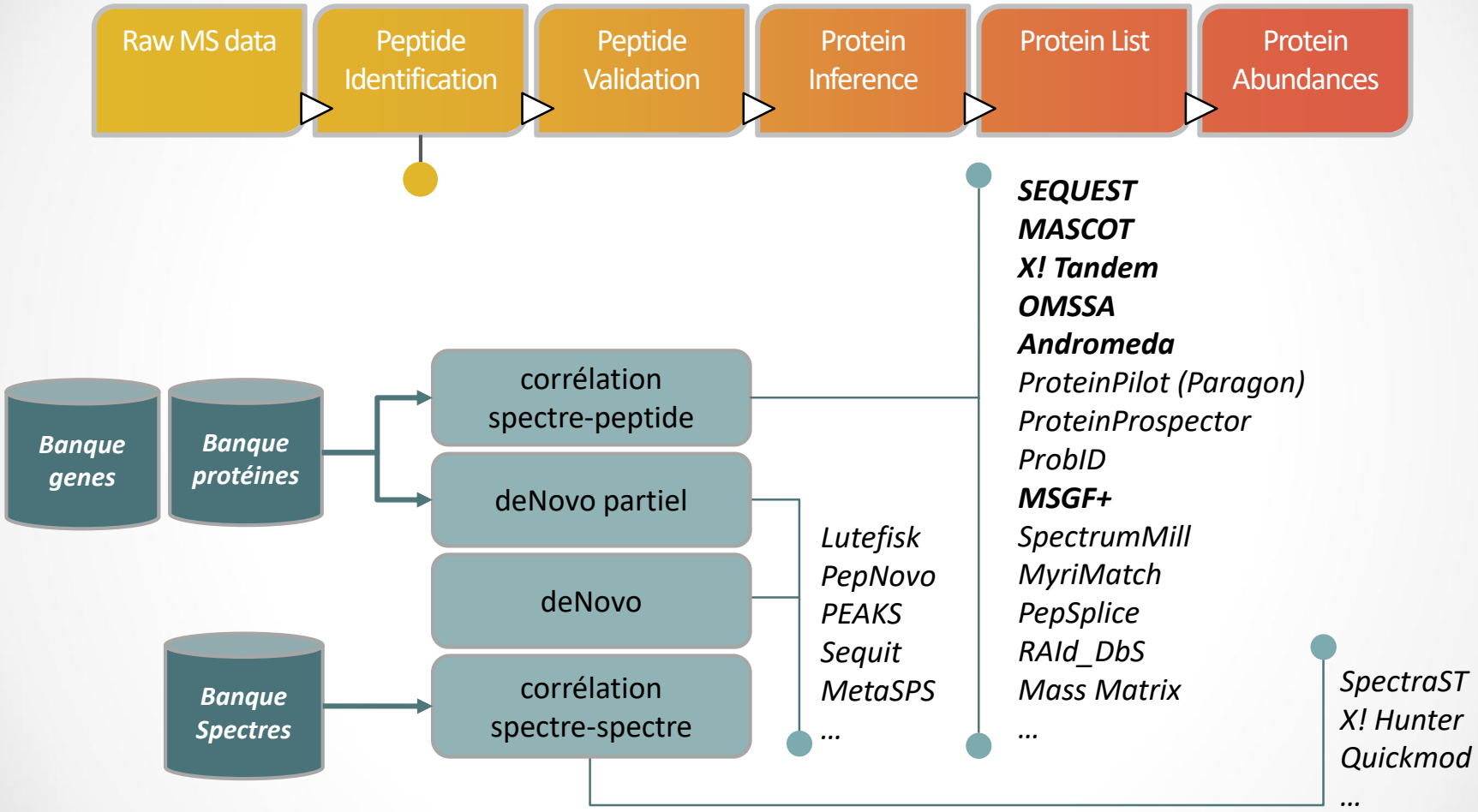
Interprétation des données DDA



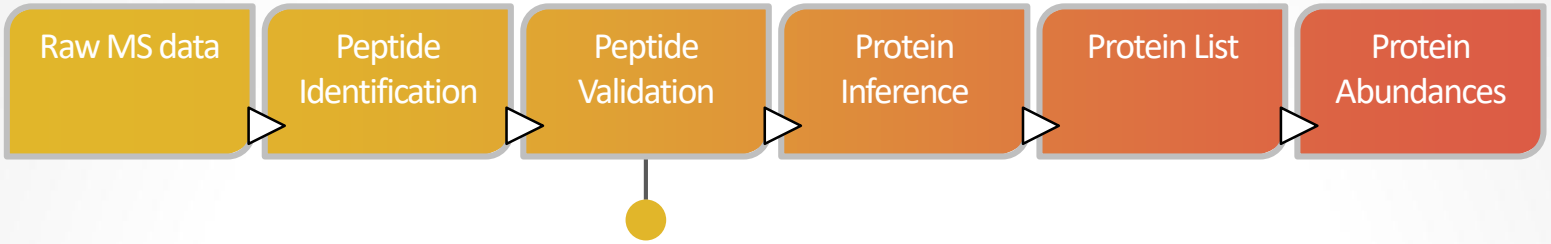
Extract_msn
Msconvert (pwiz)
Mascot distiller
Proteome discoverer
DeconMsn
…

- Traitement des spectres MS2
  - Extraction vs process des spectres
    - de…. (denoising, deisotoping)
    - redétermination de la masse et état charge du précurseur
    - somme des spectres contigus
- Impact sur le résultat
  - Meta données disponibles (output format)
  - Nombre et **score** des identifications

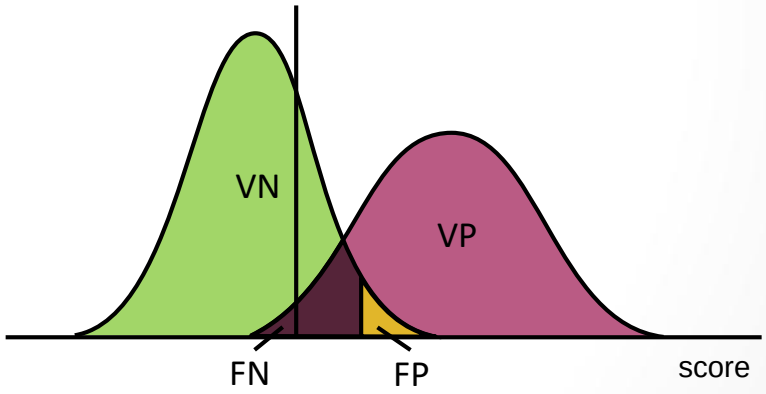
Interprétation des données DDA



Interprétation des données DDA

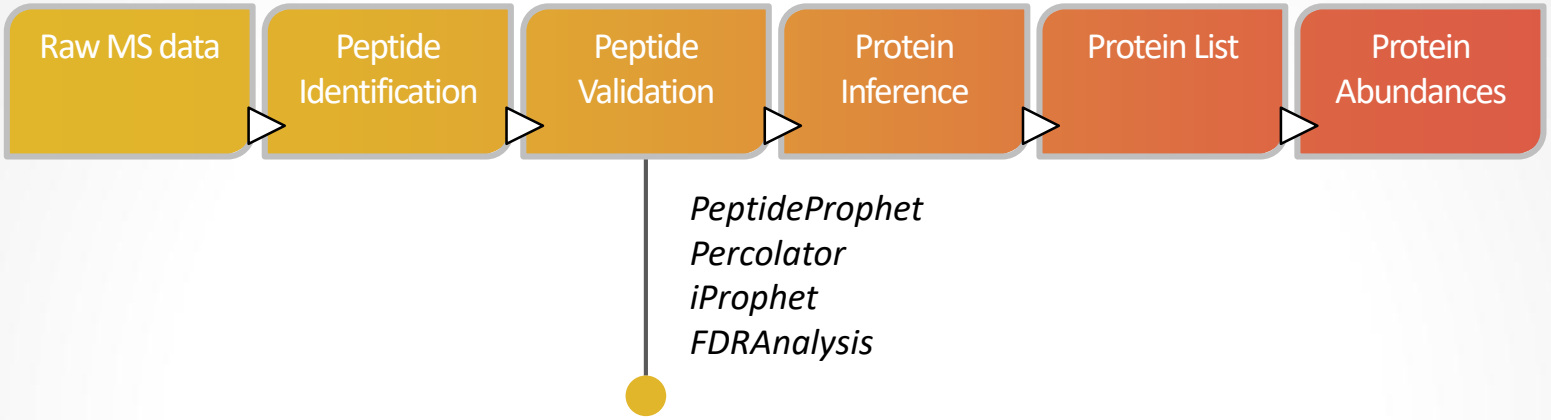


- Score des PSM
  - Le score traduit la qualité du match spectre/peptide
  - La fonction de score tente de discriminer au mieux entre match corrects et incorrects.



- Filtrage et FDR(s)
  - Choisir un seuil de score qui sépare les deux distributions
    - Estimation du taux de PSM incorrect par stratégie target/decoy
  - Utiliser d'autres informations pour calculer une fonction de score plus discriminante

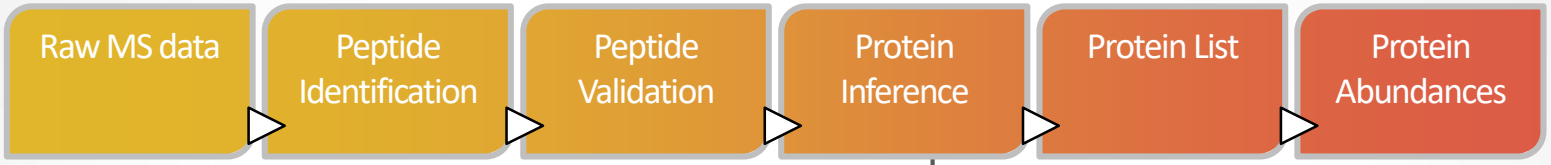
Interprétation des données DDA



- Difficultés d'estimation du FDR
  - pour les stratégies « multi-stage » ou encore la combinaison de dataset
  - Prendre en compte les les match incorrects (peptides homologues)
- FDR & Spectres haute résolution
  - Limites de l'approche target/decoy
  - Choix des paramètres de recherche
- PTMs
  - Présence et probabilité de localisation des modifications

Survey : Nesvizhskii, (2010). Journal of Proteomics

Interprétation des données DDA

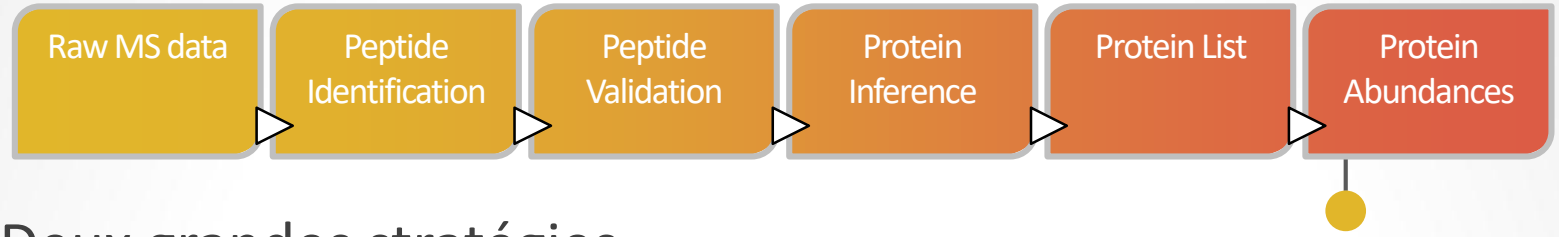


- Inférence de protéines
  - Principe de parcimonie
  - Modèle statistique
  - Validation combinées PSM & Protéines
- FDR protéique
  - Estimation FDR par approche Target/Decoy
    - Agrégation non aléatoire des PSM

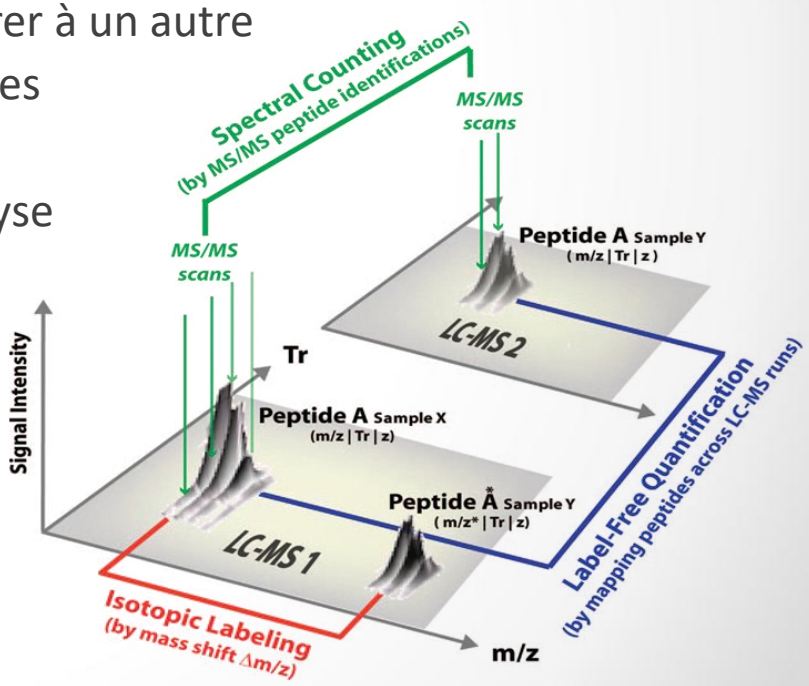
DTASelect
IDPicker
MassSieve
LDFA
ProteinProphet
EBP
PROVALT
MSBayesPro
Fido
Barista
Scaffold
PSC
MSNet
CEA
MSpresso
PeptideClassifier

Review : Huang, T.et al. (2012) Brief Bioinform 13

Stratégies de quantification



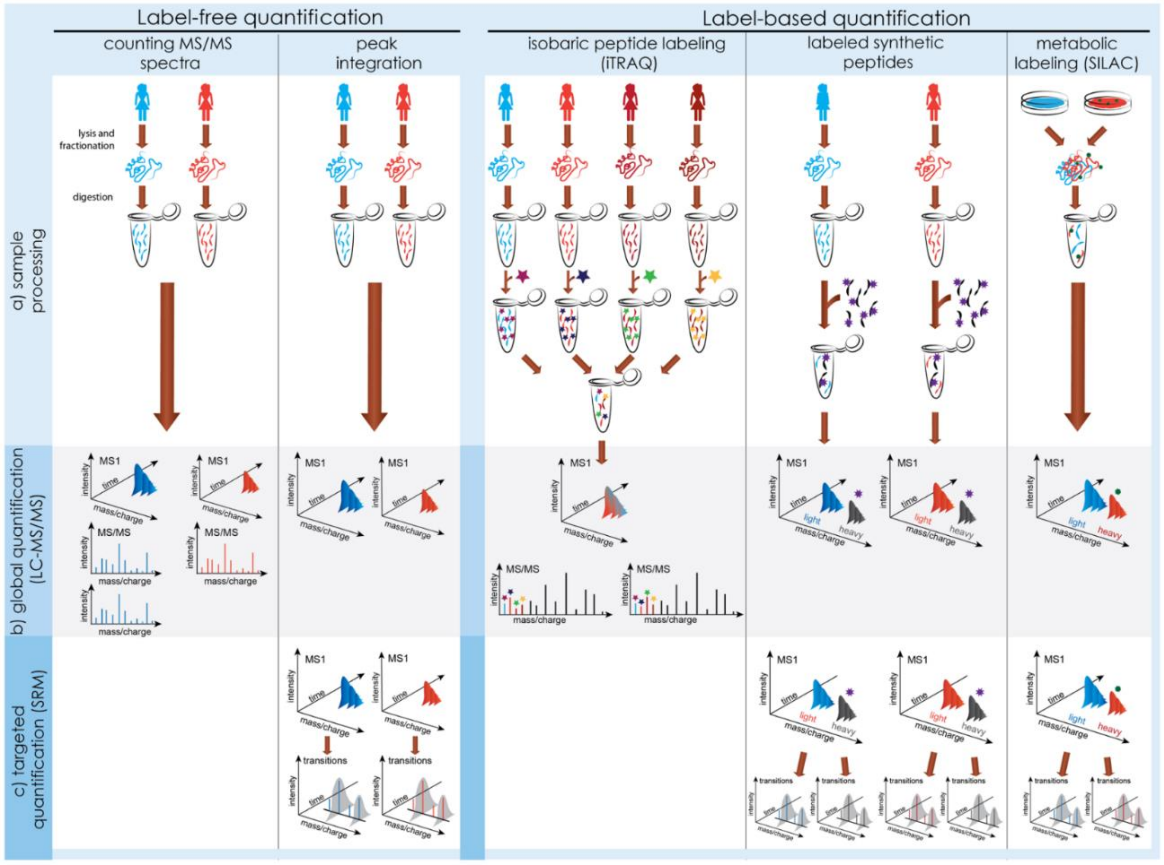
- Deux grandes stratégies
  - Par marquage (label-based)
    - Marquer un échantillon pour le comparer à un autre
    - Utiliser des peptides/protéines marquées
      - dilution isotopique
    - Combiné à différentes stratégies d'analyse
  - Sans marquage (label-free)
    - Via les spectres MS ou MS/MS



Mueller et al. J Proteome Res. 2008. doi: 10.1021/pr700758r

Stratégies de quantification

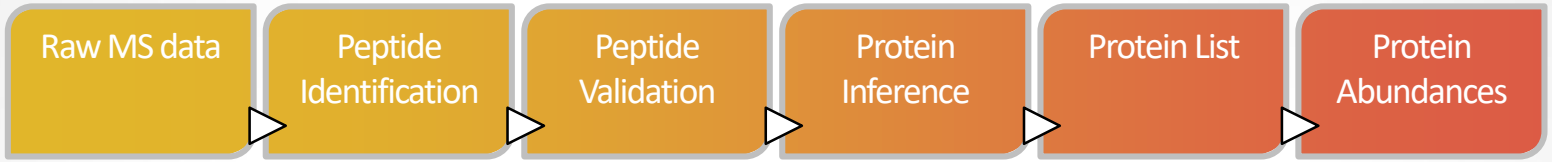
Raw MS data → Peptide Identification → Peptide Validation → Protein Inference → Protein List → Protein Abundances



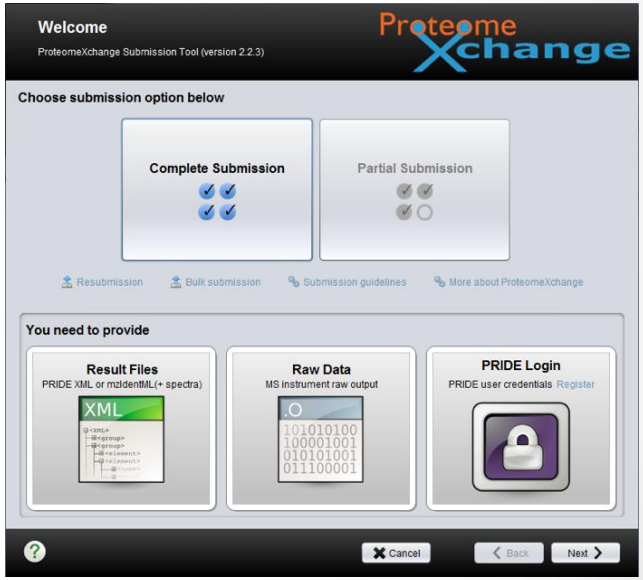
Proline
massChroQ
MaxQuant
OpenMS
Demix-Q
Skyline
…

Käll, Lukas & Vitek, Olga. (2011). PLoS computational biology. doi. 10.1371/journal.pcbi.1002277

Publication



- Dépôt des dataset obligatoire pour publier
  - ProteomeExchange / PRIDE
- Données déposées
  - Utilisation de formats standard
    - mzIdentML ou PrideXML
    - MzML pour les données de spectres



http://www.proteomexchange.org/