La protéomique pour les biologistes et médecins: techniques et méthodologies

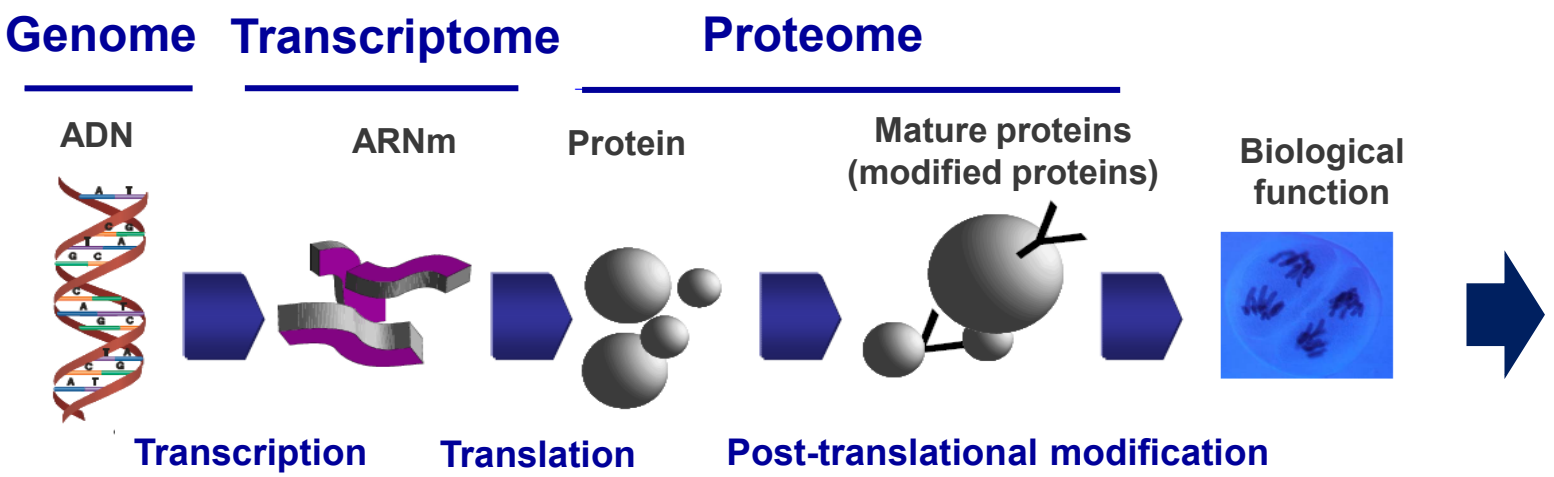
REDEKER Virginie

Inserm, CNRS, CEA, Université Paris-Saclay

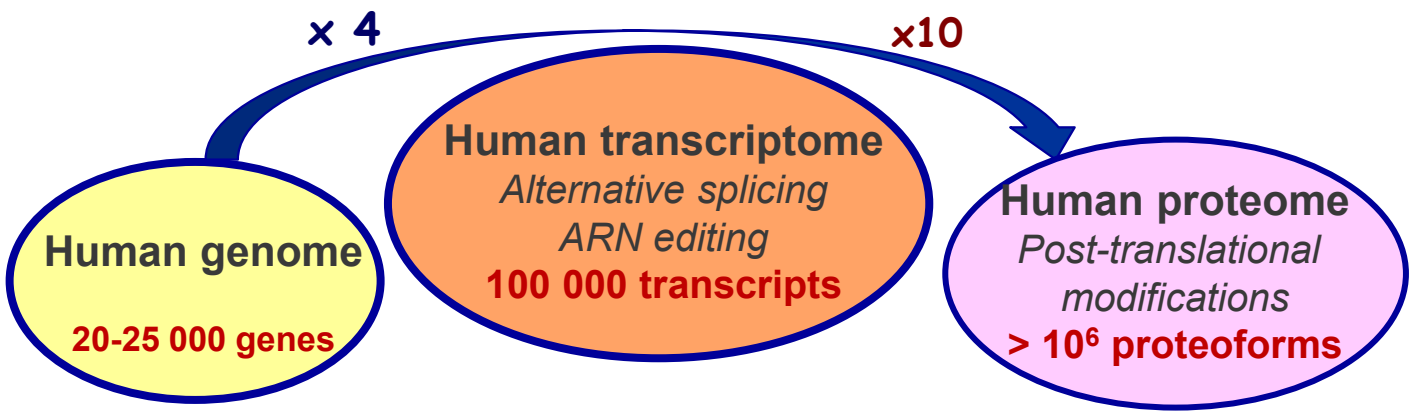
MIRCen, Fontenay-aux-Roses

I2BC, Plateforme Proteomics-Gif, Gif-sur-Yvette

Introduction to proteomics



Analysis of proteomes provide unique insights into cellular function and disease mechanism



Introduction to Proteomics

PROTEOME

the set of PROTEins expressed by a genOME

- In a given cell, cell population or tissue
- At a given time
- In a given environment
- With a given history

Wasinger, Electrophoresis,1995

One genome , many proteomes

Introduction to Proteomics

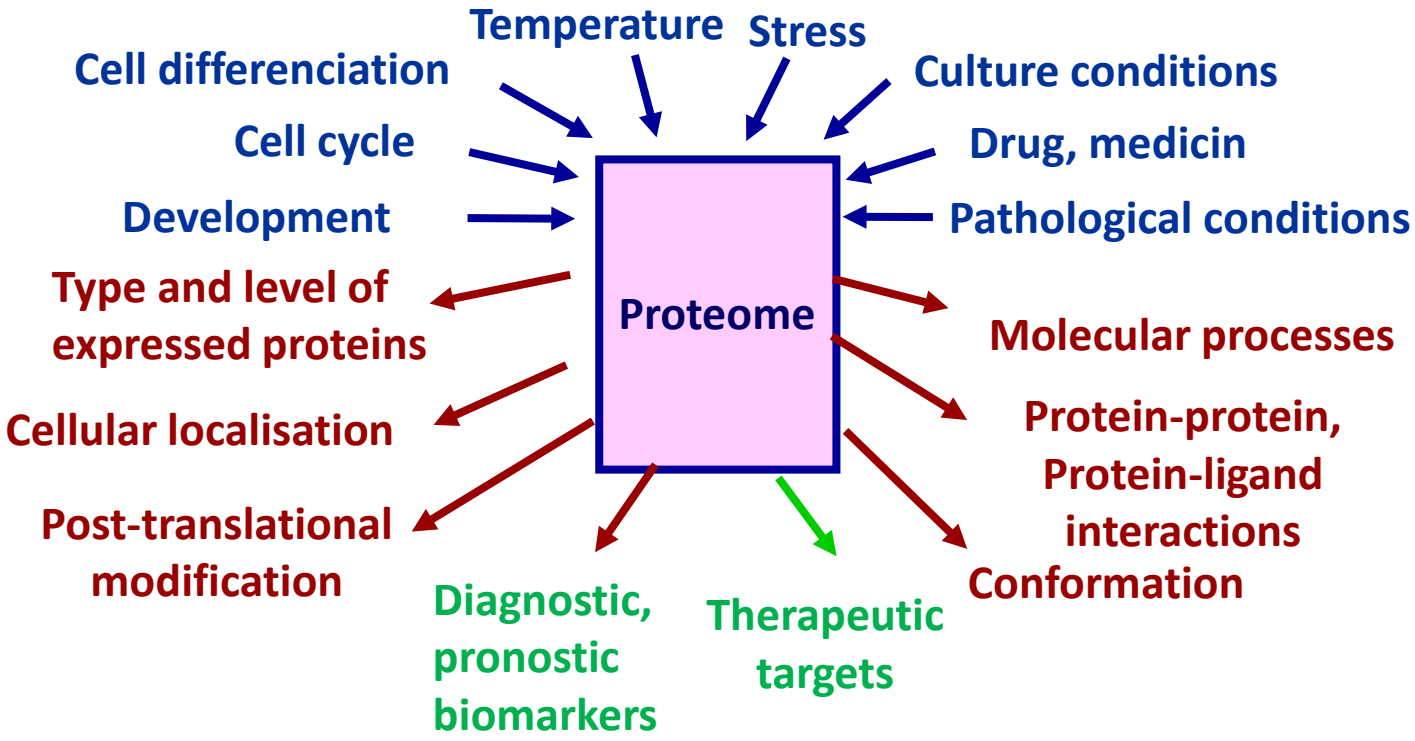
PROTEOME

the set of PROTEins expressed by a genOME

- In a given cell, cell population or tissu
- At a given time
- In a given environment
- With a given history

Wasinger, Electrophoresis,1995

1 genome , many proteomes



Introduction to Proteomics

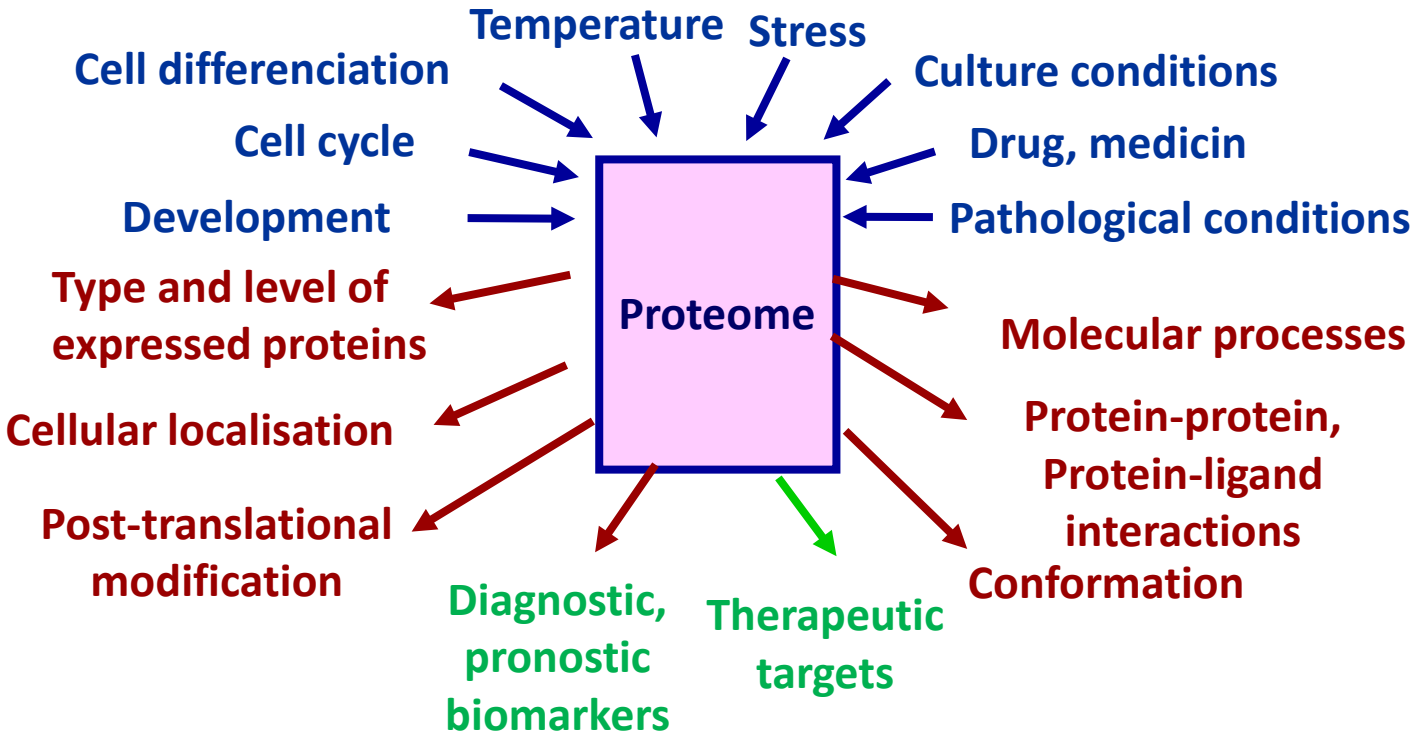
PROTEOME

the set of PROTEins expressed by a genOME

- In a given cell, cell population or tissu
- At a given time
- In a given environment
- With a given history

Wasinger, Electrophoresis,1995

1 genome , many proteomes



➢ **Proteomics analysis :**

- **Aim at identifying, characterizing and quantifying proteins**
- **Contribute to understand normal and pathological processes and identify key protein candiates**

Complexity of the human proteome

- **Highly dynamic in time and space, and condition dependant** (many normal and pathological situations)

- **Highly heterogeneous** (> 10^6 proteoforms)
  - Many co-/post-translational modifications (> 200)
  - Physico-chemical properties heterogeneity
  - More than 200 cell types in human body

- **High range of concentration, abundance** (> 10^9**)**

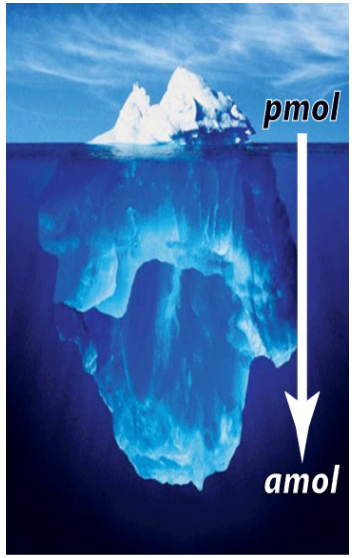
- **Many protein complexes and protein interaction networks**

(One protein -> Multiple functions)

- **Still hidden proteins**

(alternative proteins translated from noncanonical open reading frames)

- **This complexity**
  - **is an analytical challenge**
  - **requires dedicated methodologies**

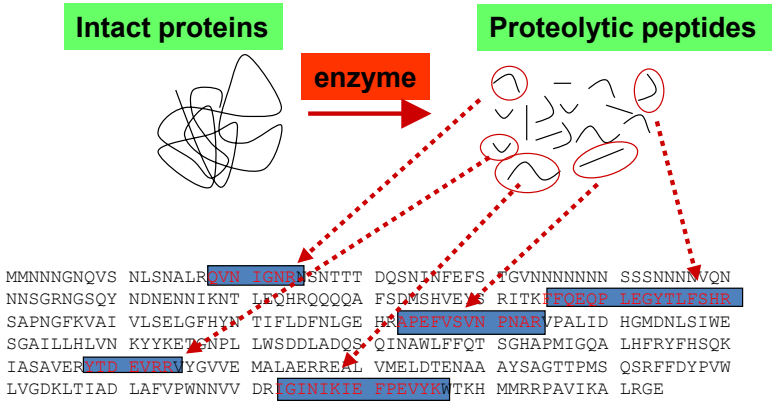


10^{-6} M

10^{-16} M

Large scale MS-based proteomics: Bottom-up approaches

- **Bottom-up approaches** for large scale and in-depth proteomics analysis of low abundant proteins and their modifications in complex biological matrices



- Need to adapt the analytical strategy to the biological question

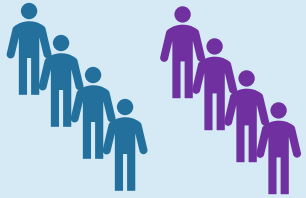
What is <u>THE</u> biological question?

What are the characteristics of the sample?

Basic MS-based proteomics workflow

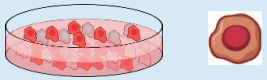
1- Sample preparation

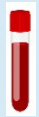
1- Experimental design



2- Type of sample

Tissue 

Cells 

Body fluids 

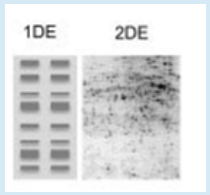
Proteins 

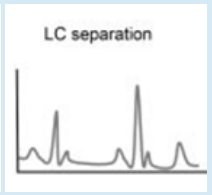
3- Sample recovery

Collection,
Preservation,
Protein extraction,
solubilisation

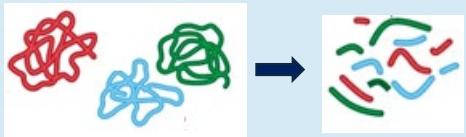
4- Proteins separation

Enrichment, Fractionation





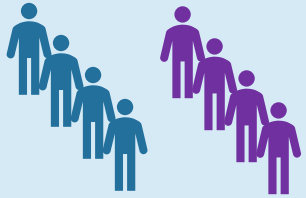
5- Proteins digestion



Basic MS-based proteomics workflow

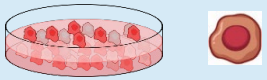
1- Sample preparation

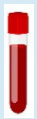
1- Experimental design



2- Type of sample

Tissue 

Cells 

Body fluids 

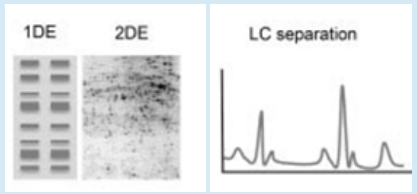
Proteins 

3- Sample recovery

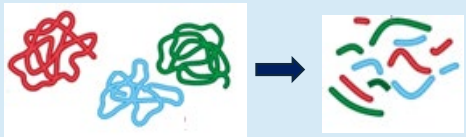
Collection,
Preservation,
Protein extraction,
solubilisation

4- Proteins separation

Enrichment, Fractionation

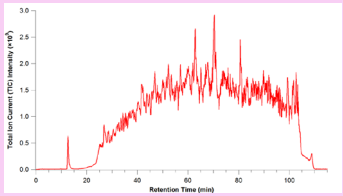


5- Proteins digestion



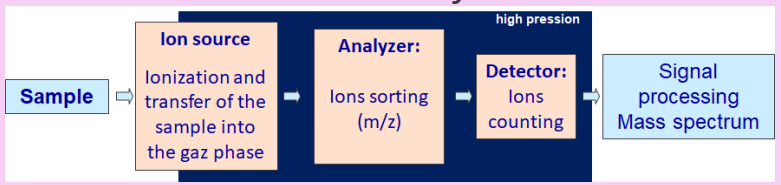
2- Mass spectrometry analysis

1- Peptide separation (nanoLC)



2- MS technology and acquisition method

The science of ions

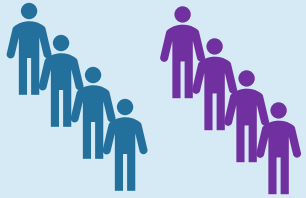


- **Molecular mass**
- **Sequence information**
- **Quantification**

Basic MS-based proteomics workflow

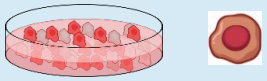
1- Sample preparation

1- Experimental design



2- Type of sample

Tissue 

Cells 

Body fluids 

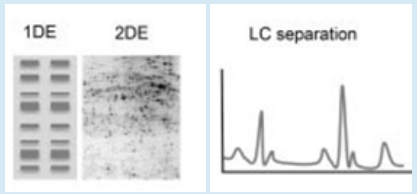
Proteins 

3- Sample recovery

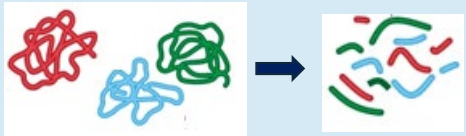
Collection,
Preservation,
Protein extraction,
solubilisation

4- Proteins separation

Enrichment, Fractionation

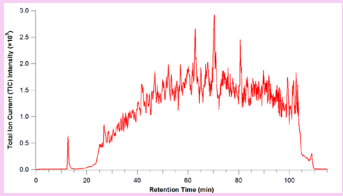


5- Proteins digestion



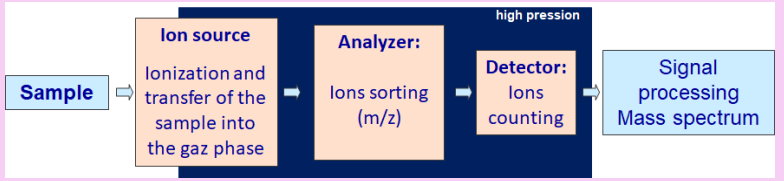
2- Mass spectrometry analysis

1- Peptide separation (nanoLC)



2- MS technology and acquisition method

The science of ions



- **Molecular mass**
- **Sequence information**
- **Quantification**

3- Bioinformatic data Analysis

Database searches

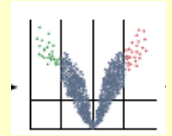
Protein identification

Protein quantification

Statistical analysis

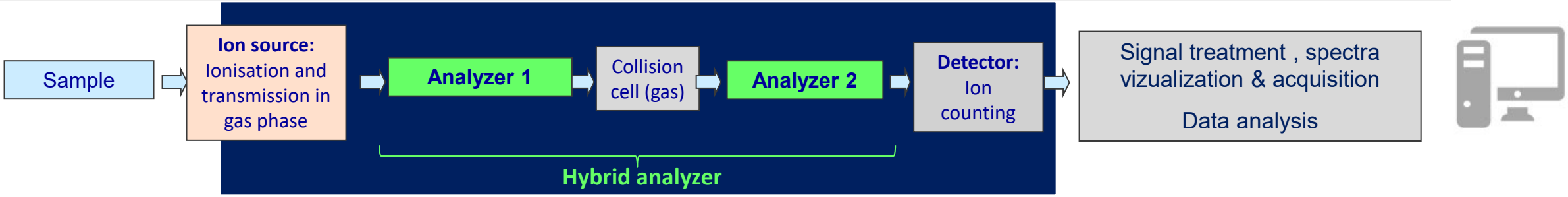
Data interpretation & integration:

Enrichment analysis
Protein interaction network

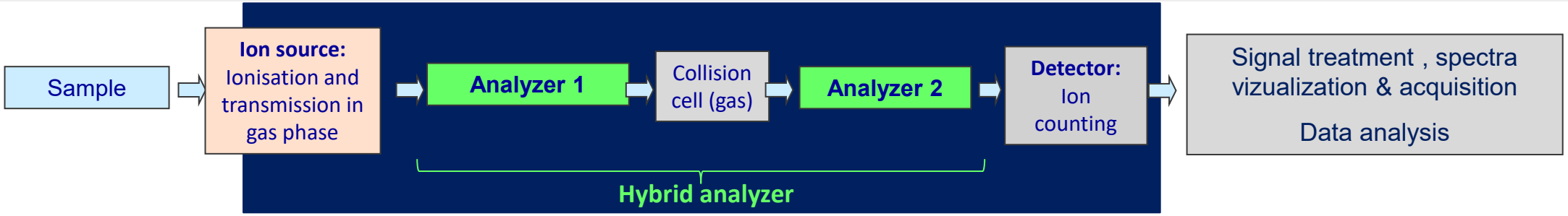


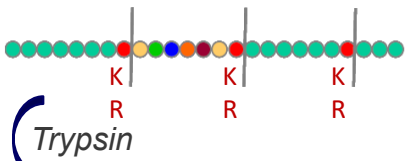
4- Data validation

Principle of nanoLC-MS/MS Mass spectrometry

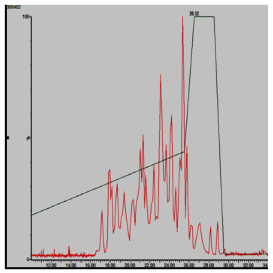


Principle of nanoLC-MS/MS Mass spectrometry





NanoLC separation of peptides



ESI IonizationI

1 MS

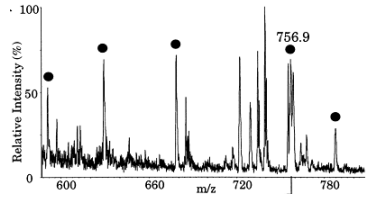
MS mode

Transmission of ions within a m/z range

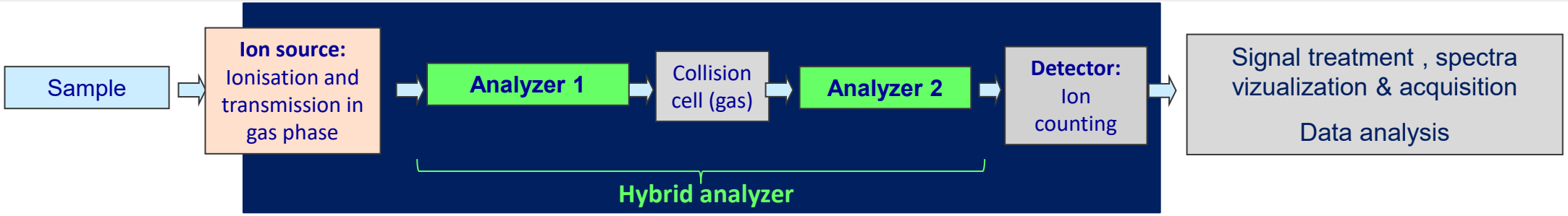
Ion transmission

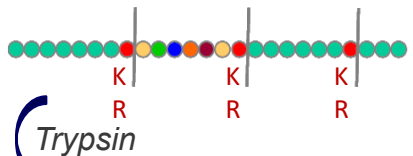
MS1

m/z mass measurement of intacts ions (MS)

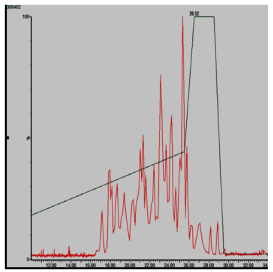


Principle of nanoLC-MS/MS Mass spectrometry





NanoLC separation of peptides



ESI

1 MS

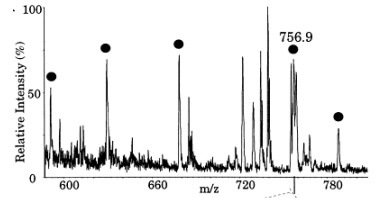
2,3,..100s MS/MS

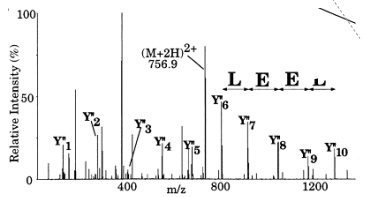
MS mode

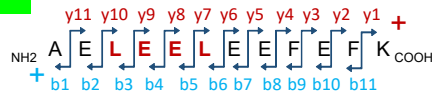
Transmission of ions within a m/z range → Ion transmission → **MS1** m/z mass measurement of intacts ions (MS)

MS/MS mode

Selection of one m/z ion → Fragmentation (Collision Induced Dissociation, CID) → **MS2** m/z mass measurementof fragments (MS/MS)







Identification

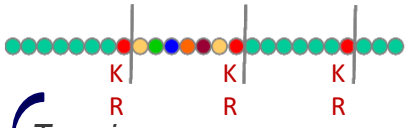
- 10- 20 MS/MS in 1-2sec *(LTQ-Orbitrap)*
- 100 MS/MS i 1 sec *(timsTOF Pro)*

➢ Fragmentation of thousands of peptides

Principle of nanoLC-MS/MS Mass spectrometry

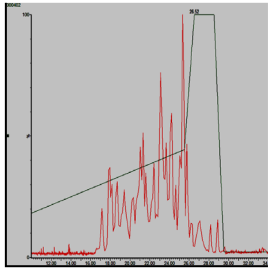
Sample → **Ion source:** Ionisation and transmission in gas phase → **Analyzer 1** → Collision cell (gas) → **Analyzer 2** → **Detector:** Ion counting → Signal treatment , spectra vizualization & acquisition → Data analysis

Hybrid analyzer



Trypsin

NanoLC separation of peptides



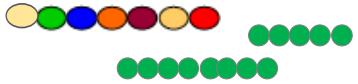
ESI

1 MS

2,3,..100s MS/MS

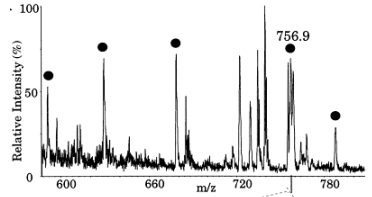
MS mode

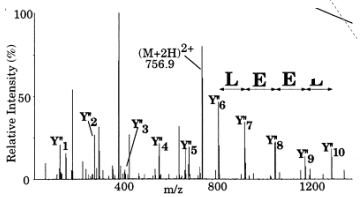
Transmission of ions within a m/z range → **Ion transmission** → **MS1** **m/z mass measurement of intact ions (MS)**

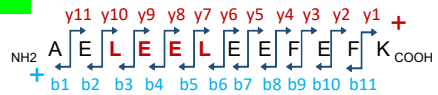


MS/MS mode

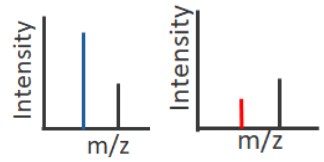
Selection of one m/z ion → **Fragmentation (Collision Induced Dissociation, CID)** → **MS2** **m/z mass measurementof fragments (MS/MS)**



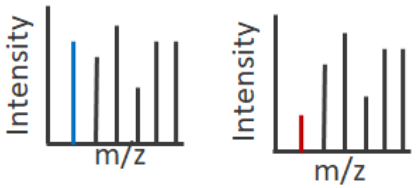


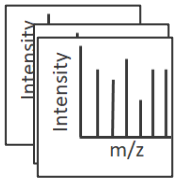


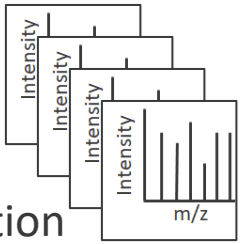
MS1



MS2







Identification

Quantification

- 10- 20 MS/MS in 1-2sec *(LTQ-Orbitrap)*
- 100 MS/MS i 1 sec *(timsTOF Pro)*

➢ Fragmentation of thousands of peptides

Database searches

Entry: List of precursor masses and corresponding mass fragments

Parameters:

- Mass tolerances in MS1 and MS2
- Enzyme digestion: specificity and number of miscleavages
- Protein modifications: list of fixed or variable modifications
- Type of instrument used
- Decoy database

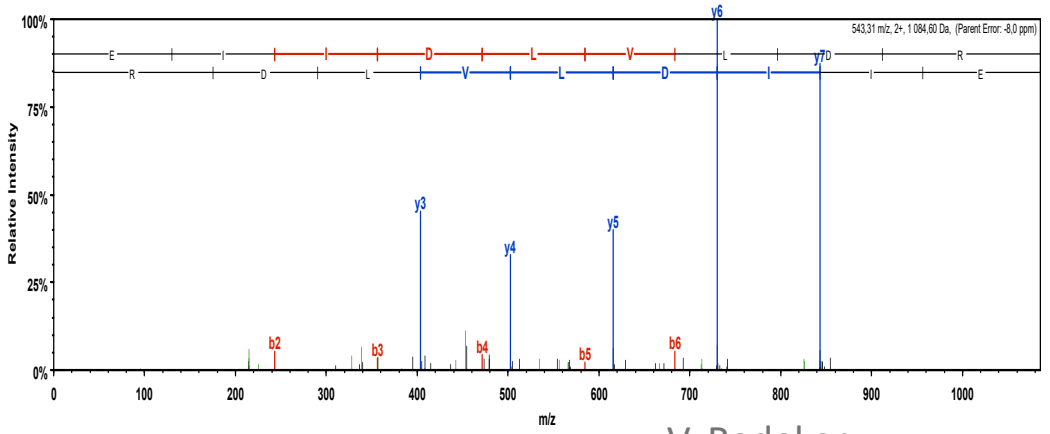
Public Databases	Database search engines
NCBI RefSeq	SEQUEST
UniProt	Mascot
Ensembl	OMSSA
UniRef	X!tandem
OpenProt	Andromeda

Results

P05213 (100 %), 50 151,7 Da
Tubulin alpha-1B chain OS=Mus musculus OX=10090 GN=Tuba1b PE=1 SV=2
2 exclusive unique peptides, 2 exclusive unique spectra, 5 total spectra, 290/451 amino acids (64 % coverage)

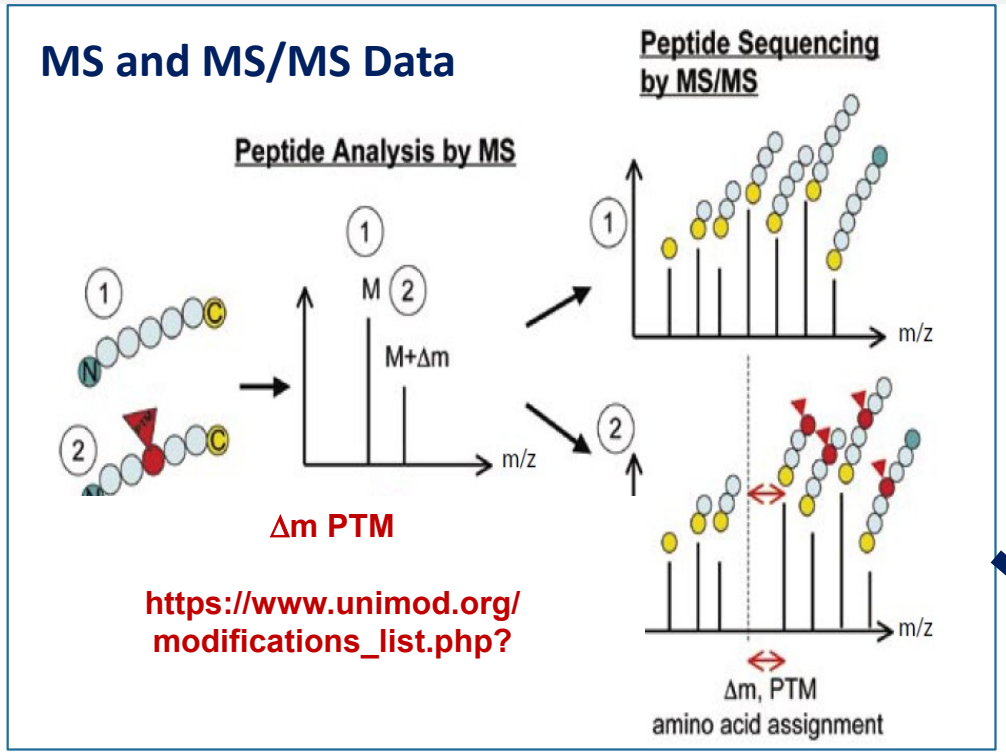
MRECISIHVG QAGVQIGNAC WELYCLEHGI QPDGQMPSDK TIGGGDDSFN TFFSETGAGK HVPRAVFVDL EPTVIDEVRT
GTYRQLFHPE QLITGKEDAA NNYARGHYTI GKEIIDLVLD RIRKLADQCT GLQGFLVFHS FGGGTGSGFT SLLMERLSVD
YGKKSKLEFS IYPAPQVSTA VVEPYNSILT THTTLEHSDC AFMVDNEAIY DICRRNLDIE RPTYTNLNRL ISQIVSSITA
SLRFDGALNV DLTEFQTNLV PYPRIHFPLA TYAPVISAEK AYHEQLSVAE ITNACFEPAN QMVKCDPRHG KYMACCLLYR
GDVVPKDVNA AIATIKTKRS IQFVDWCPTG FKVGINYQPP TVVPGGDLAK VQRAVCMLSN TTAIAEAWAR LDHKFDLMYA
KRAFVHWYVG EGMEEGEFSE AREDMAALEK DYEEVGVDSV EGEGEEEEGE EY

FDR



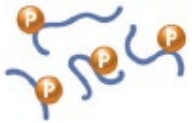
Post-translational modification identification: low abundant, specific physico-chemical properties

Shotgun analysis of proteolytic peptides



Modification specific enrichment strategies

Phosphorylation specific strategies



- Immuno-affinity (pS, pT, pY)
- TiO2, IMAC Chromatography
- Chemical Modification & capture

MS and MS/MS Data

Database search

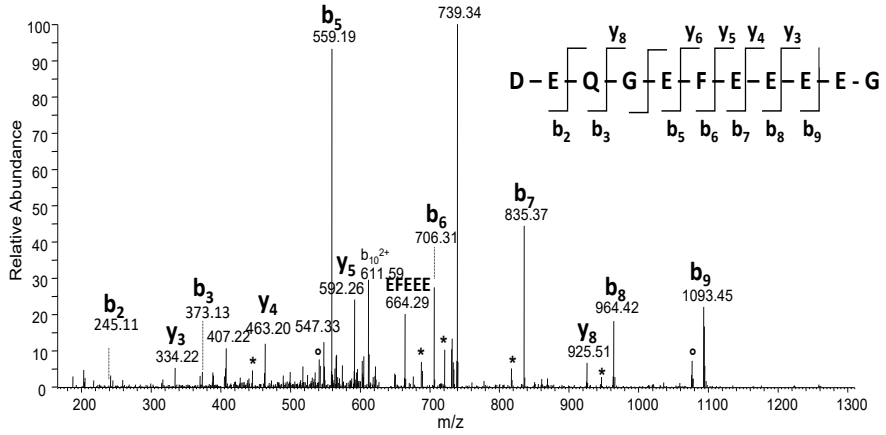
PTM assignment by bottom-up methods

Proteoforms toechiometry by top-down MS

Larsen MR, BioTechniques, 2006

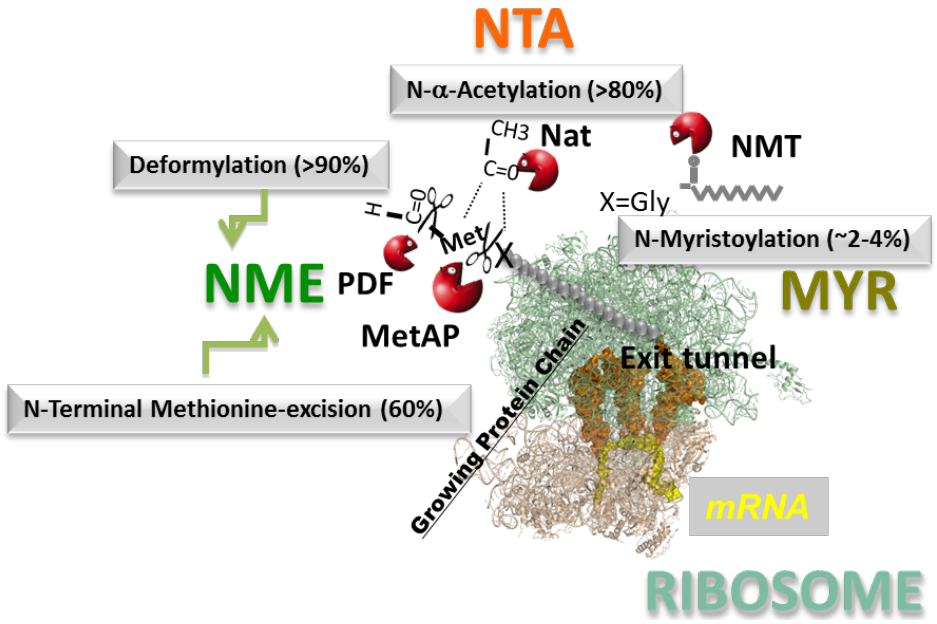
Example of post-translational modification identification

New C-terminal truncation of tubulin



Aillaud et al, Mol Biol Cell, 2016

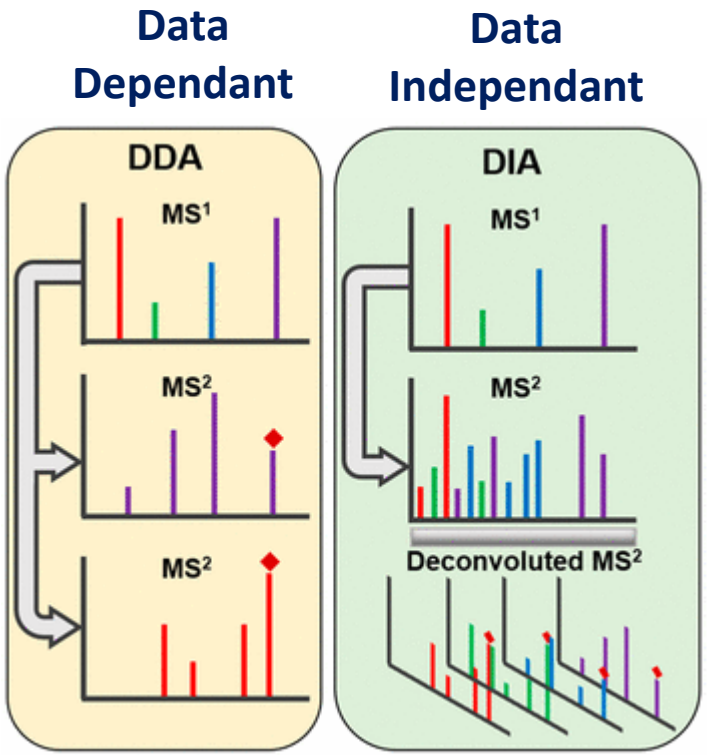
Large scale N-terminomics: from identification to quantification



Guesdes et al, Mol Cell Biol, 2024; Armbruster, Plant Physiol 2020.; Linster, New Physiologist, 2020; Dian, Nat Comm, 2020; Bienvenut, Mol Syst Biol. 2020; Castrec, Nat. Chem. Biol, 2018; Bienvenut, Meth Mol Biol. 2017 ;

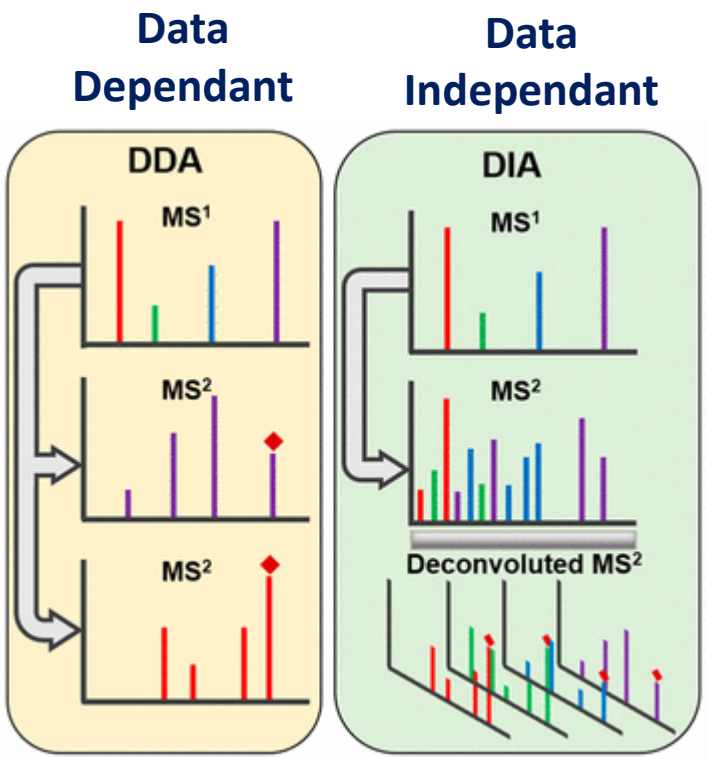
MS-based identification of proteins for global discovery proteomics

MS Acquisition methods



MS-based identification of proteins for global discovery proteomics

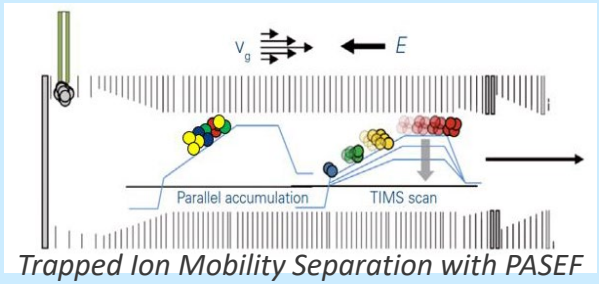
MS Acquisition methods



Last generation MS instruments

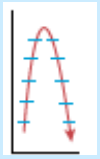
timsTOF

- TIMS Ions separation

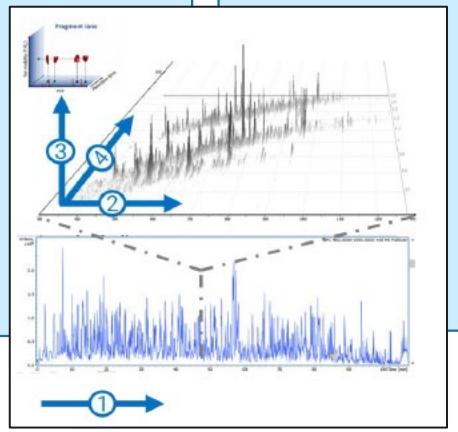


Trapped Ion Mobility Separation with PASEF

- Analyzer 1 : Quadrupole
- Analyzer 2: Time-of- flight



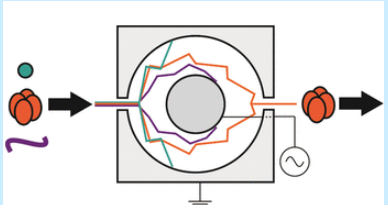
1. RT
2. m/z
3. CCS
4. MS/MS



4D proteomics

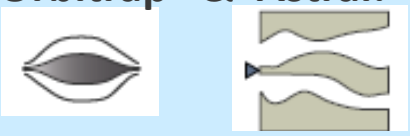
Astral

- FAIMS Ions fractionation



High-field asymmetric waveform ion mobility spectrometry

- Analyzer 1 : Quadrupole
- Analyzer 2 : Orbitrap & Astral:



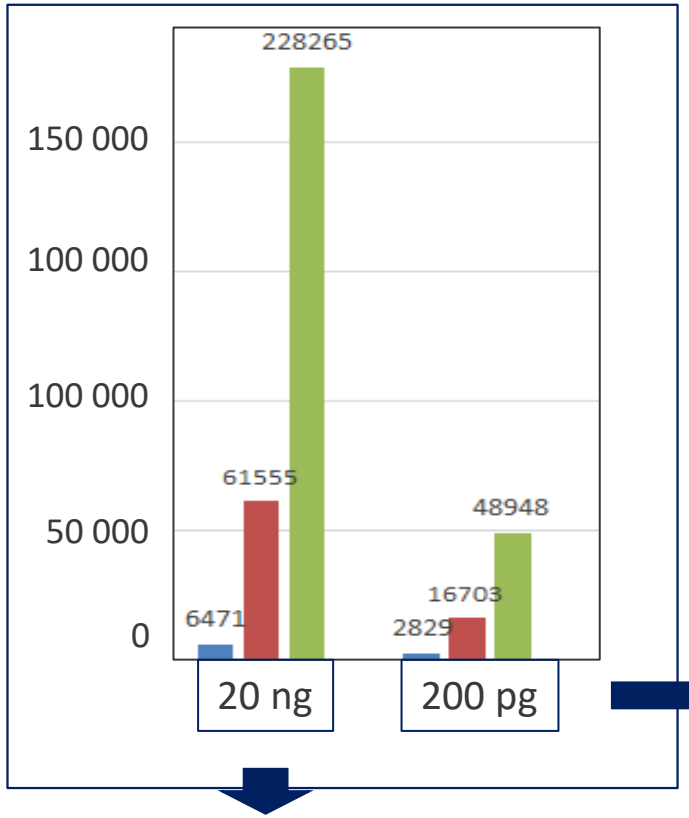
1. RT
2. m/z
4. MS/MS

3D proteomics

Example of DIA-PASEF protein identifications with the timsTOF Ultra2 MS

HeLa extract

(20 ng , 200 pg; 22 min nLC gradient)



1.9

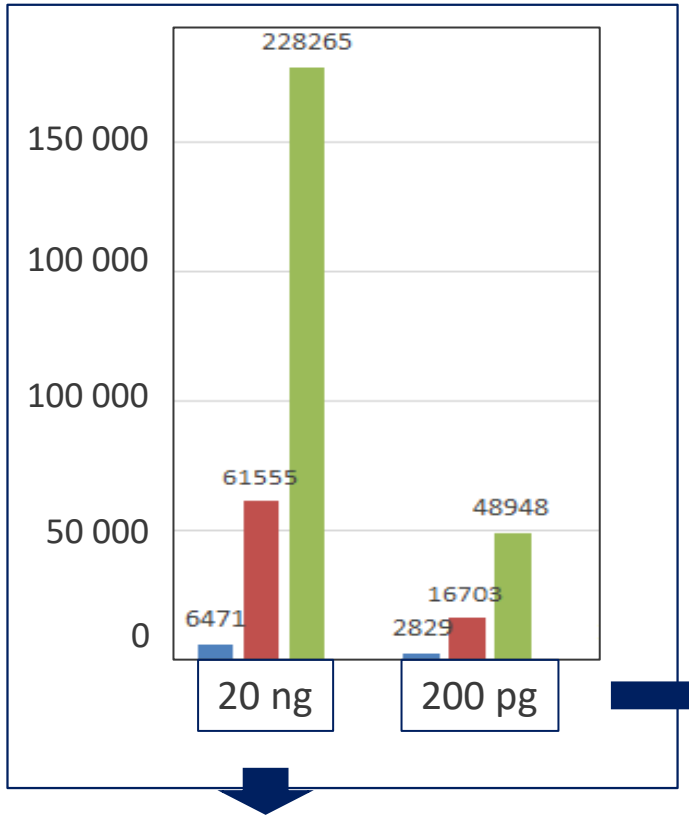
Single cell sensitivity

In DDA mode: ID of 998 phosphorylation peptide
Without any enrichment

Example of DIA-PASEF protein identifications with the timsTOF Ultra2 MS

HeLa extract

(20 ng , 200 pg; 22 min nLC gradient)



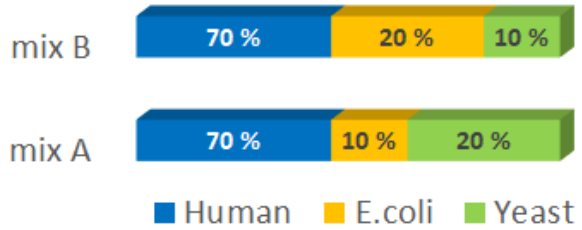
1.9

Single cell sensitivity

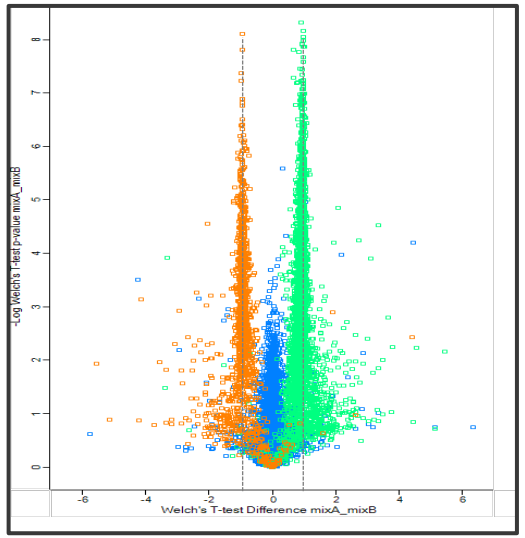
In DDA mode: ID of 998 phosphorylation peptide
Without any enrichment

Mix of 3 proteomes

(20ng, 45 min nLC gradient)



In total:
Protein groups:
11 855 to 11 936
Precursor ions:
139 988 to 143 343



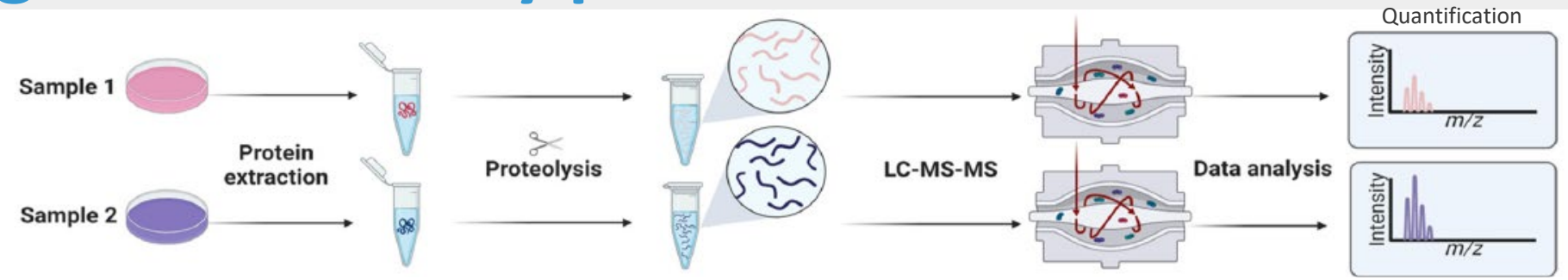
Protein groups
Human: 8 412
E. Coli: 1 150
Yeast: 3 450

Precursors
Human: 136 506
E. Coli: 33 568
Yeast: 7 562

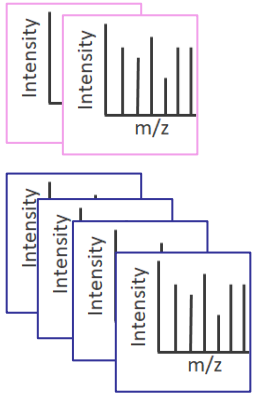
Volcanoplot (Label-free quantification)

MS-based relative quantification of proteins for global discovery proteomics

Label-free quantification



MS1 intensity



MS2 counts

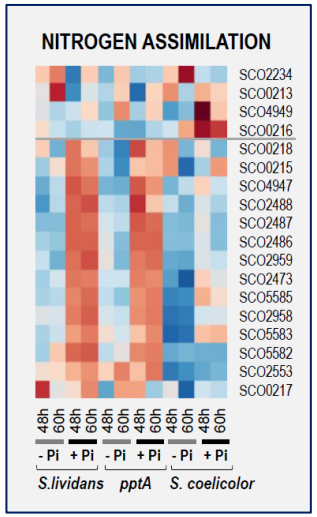
Jiang et al, ACS Meas. Sci. Au 2024

Example of proteome comparisons using label-free quantification

Streptomyces strains

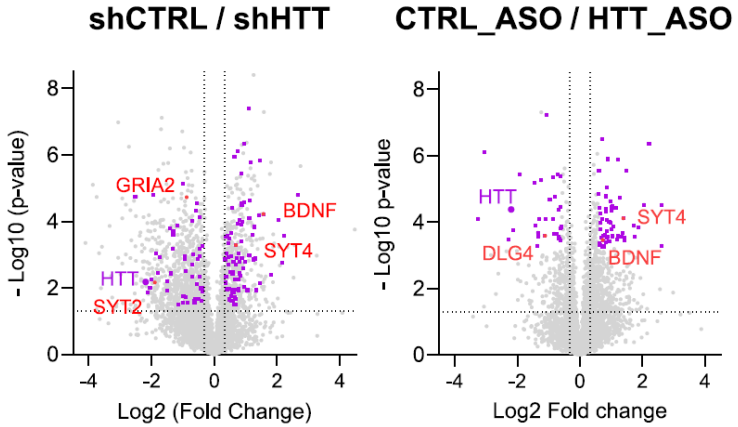
Comparisonf of 3 strains and different culture conditions using multivariable statistical analysis

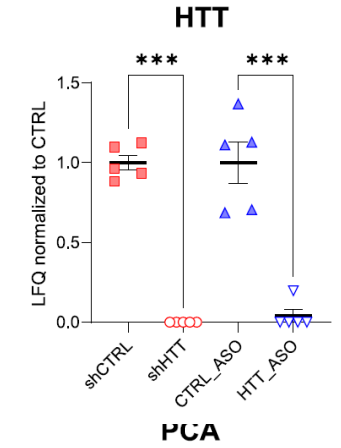
Lejeune et al, Res Microbiol 2024

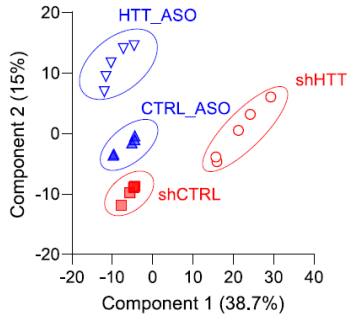


iPSC cells

Effect of lowering the expression of one protein







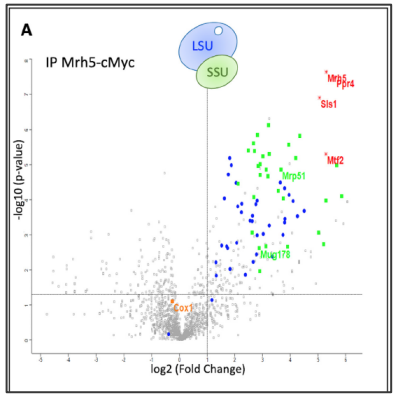
Louça et al, Neurobiol Dis. 2024

➢ **Exploration of protein network changes and defects**

Example of interactome identification and comparison

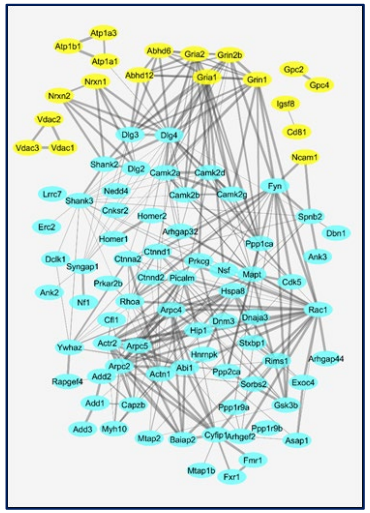
- **Proteins interacting with mitochondrial proteins** (cMyc tag)

Herbert et al, NAR, 2021

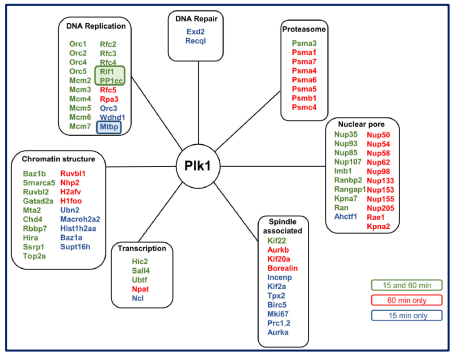


- **Receptors of protein aggregates at the surface of the cell membrane** (+/- crosslinking)

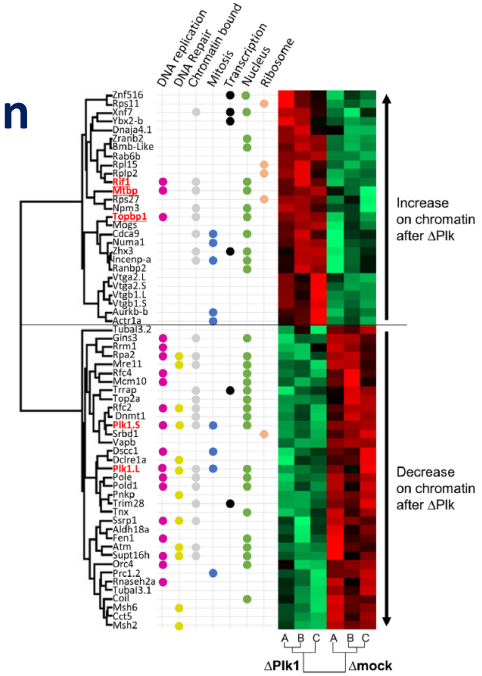
Shrivastava, Redeker et al EMBO J. 2015; 2019



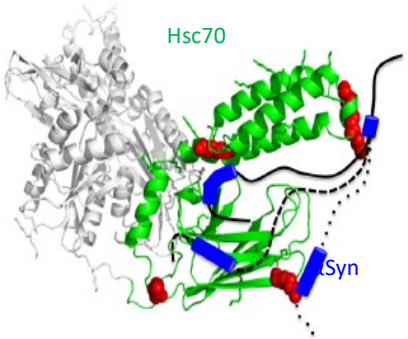
- **Chromatin bound protein dynamics during replication WT vs ΔPlk1**

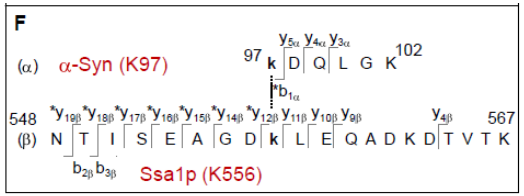


Ciardo et al , NAR, 2021



- **Protein-protein surface interactions** (covalent cross-linking)

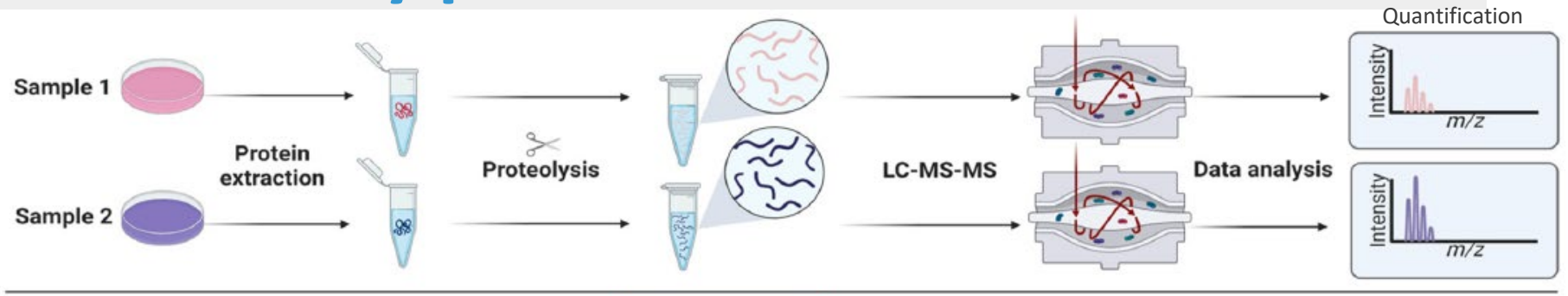




Nury, Anal Chem. 2015; Bendifallah, BBRC, 2020;

MS-based relative quantification of proteins for global discovery proteomics

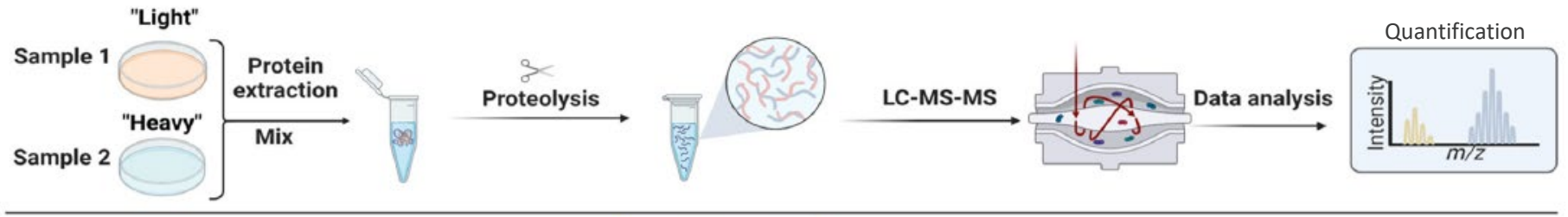
Label-free quantification



MS1

Metabolic labeling

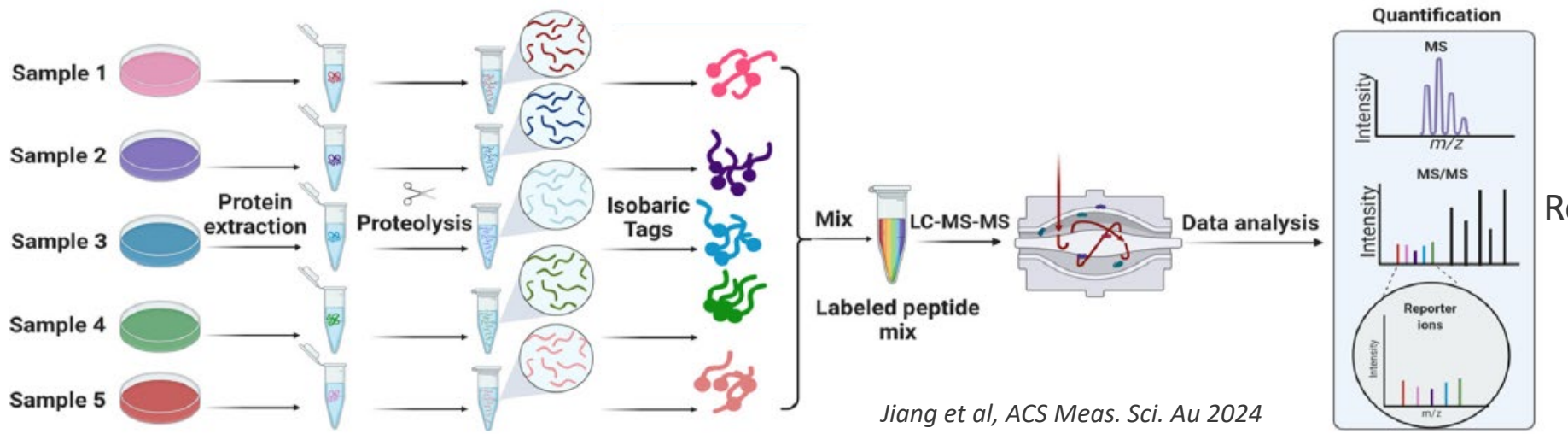
- **SILAC** *(Stable Isotope Labeling of Amino acids in cell Culture)*



MS1

Isobaric tags

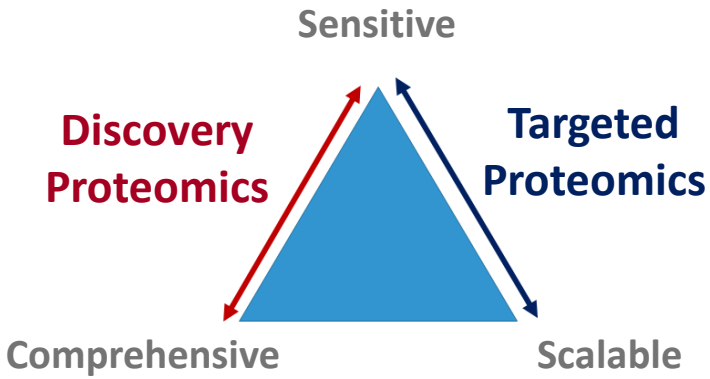
- **iTRAQ** *(Isobaric Tagging Reagents for Accurate Quantification)*
- **TMT** *(Tandem Mass Tags)*

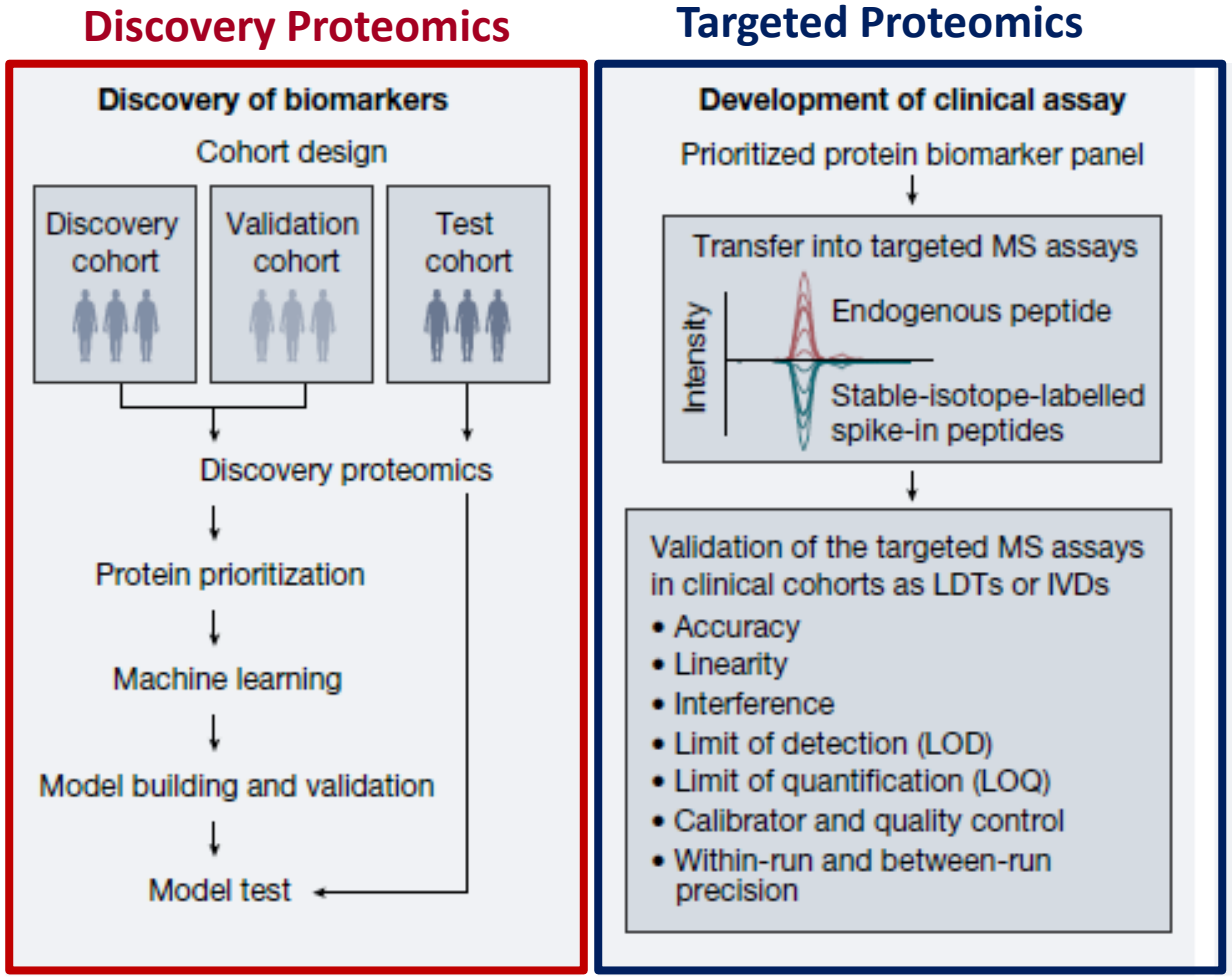


MS2 Reporter ion

Jiang et al, ACS Meas. Sci. Au 2024

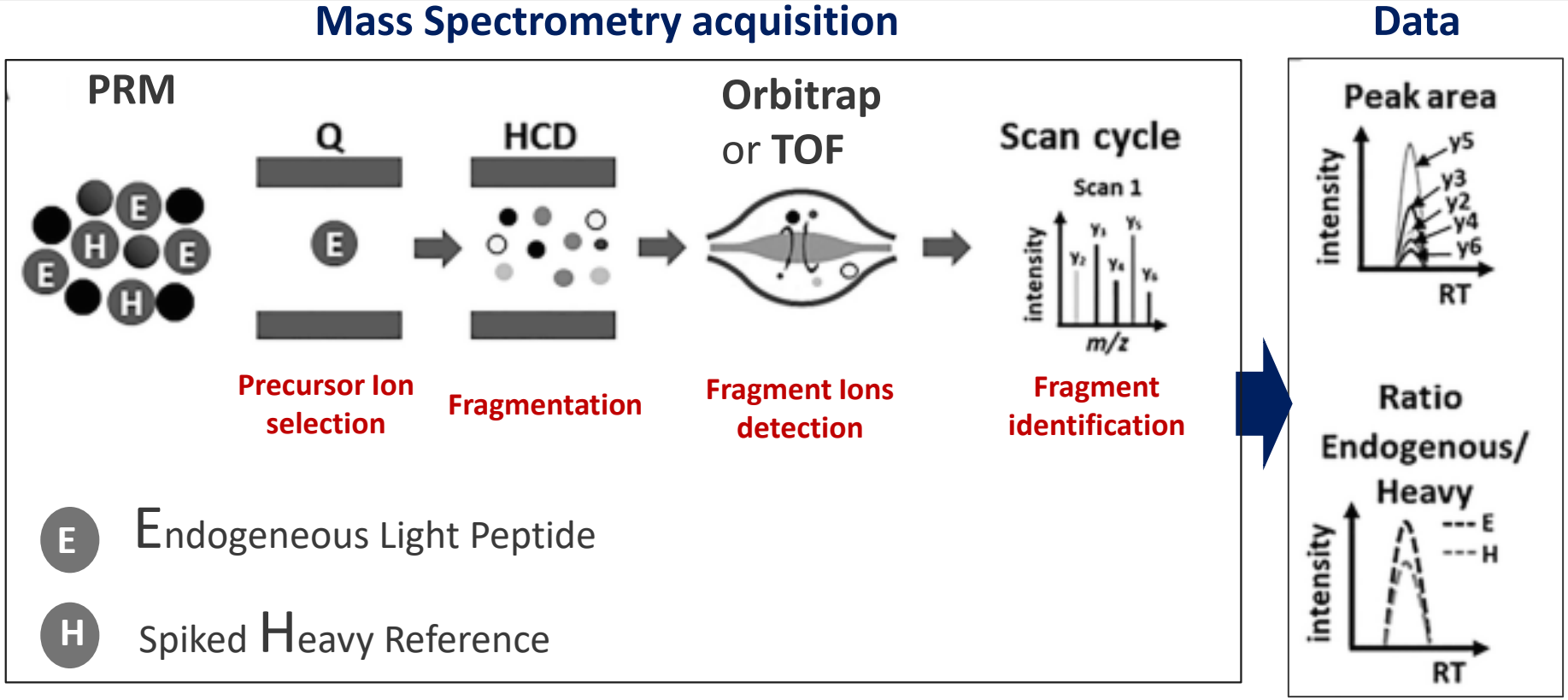
From discovery proteomics to clinical proteomics





Guo et al, Nature 2025

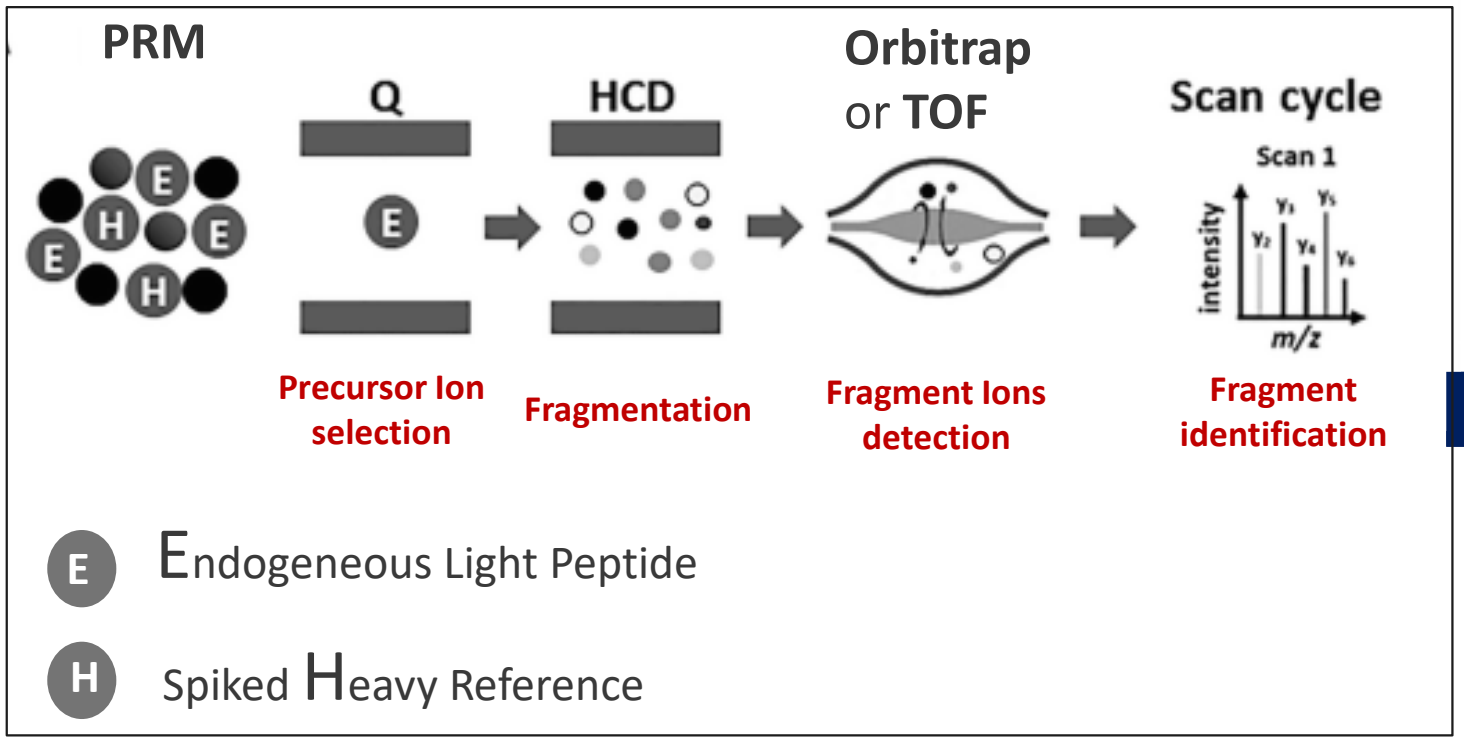
MS-based absolute quantification of proteins for targeted proteomics



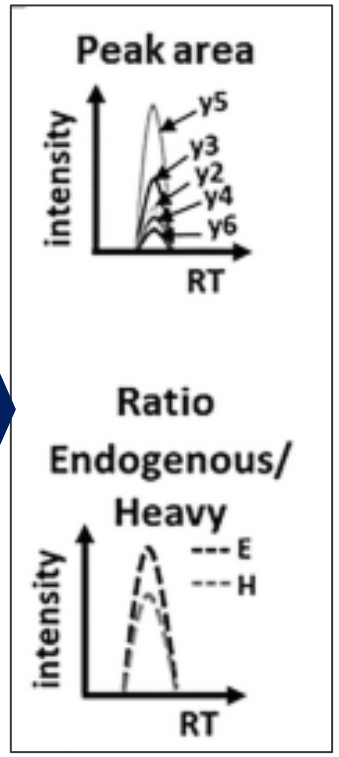
Barkovits et al, Meth Mol Biol, 2021

MS-based absolute quantification of proteins for targeted proteomics

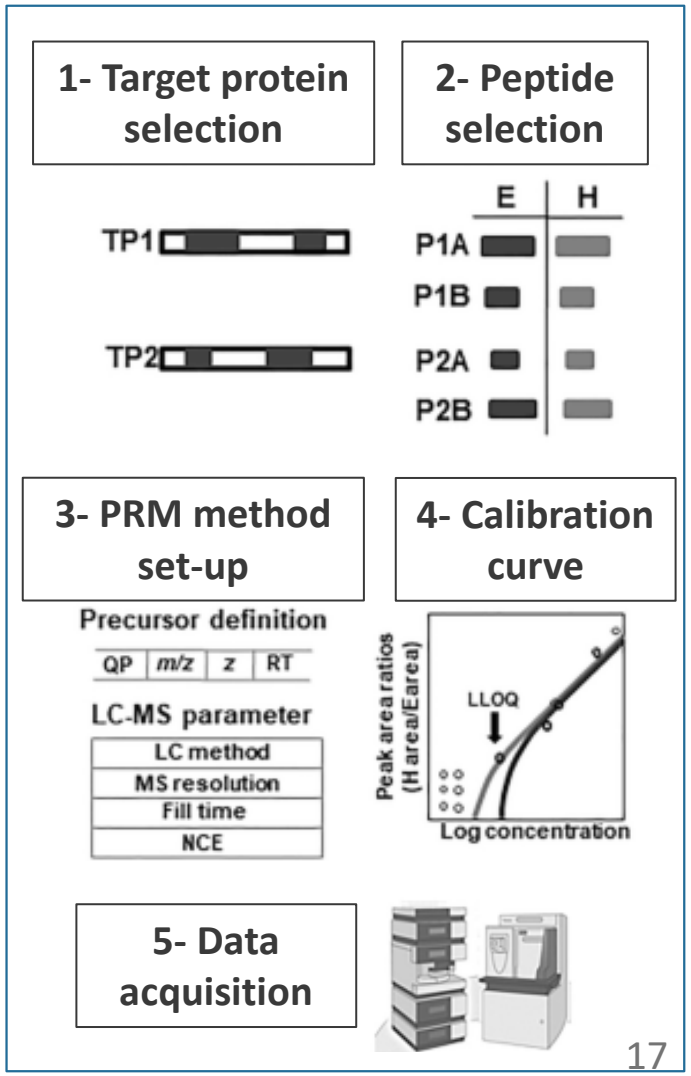
Mass Spectrometry acquisition



Data



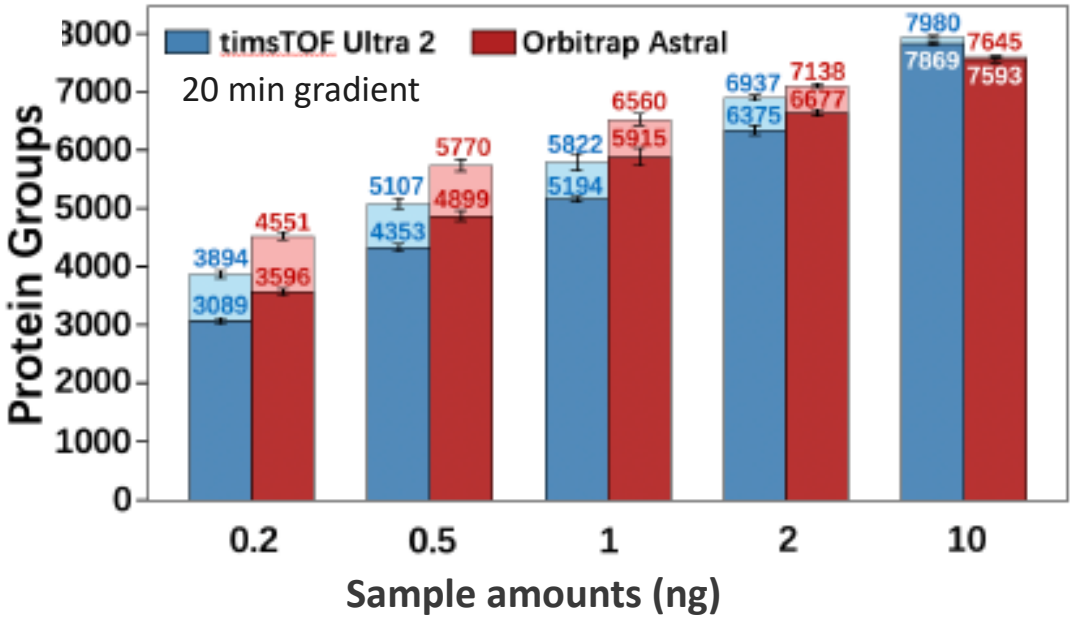
Method



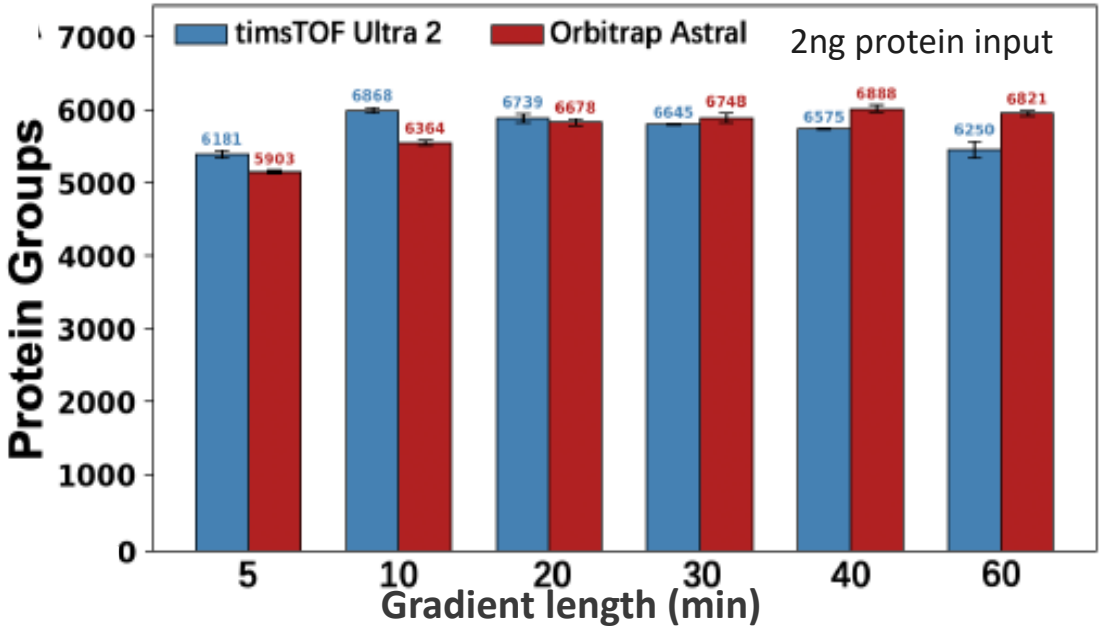
Barkovits et al, Meth Mol Biol, 2021

Advances in DIA protein identifications with the latest generation instruments

High sensitivity

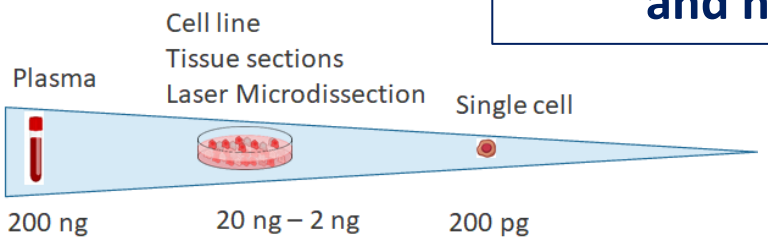


High speed



Towards single cell proteomics and spatial proteomics

Towards limited run times per sample (5-10min) and high throughput sample analysis



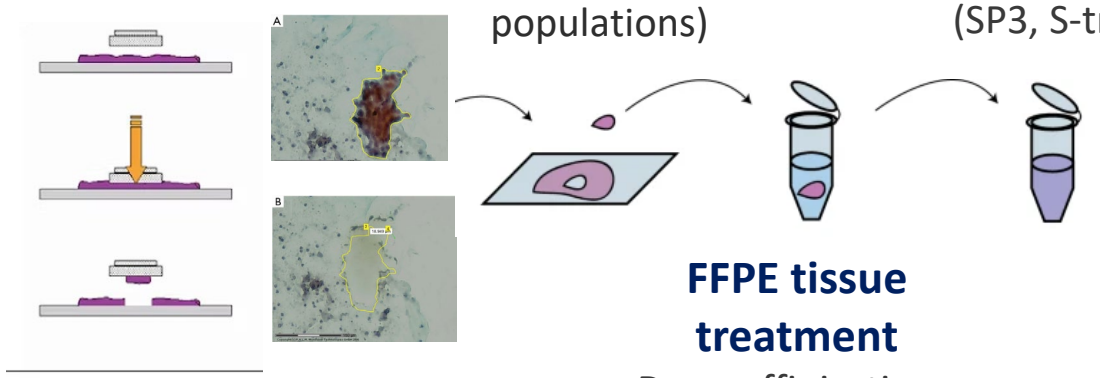
Huang et al BioRxiv, 2025

Laser microdissection (LMD) toward spatial MS proteomics within tissues

Laser-capture microdissection (LCM)

Isolation of cells (distinct sub-populations)

Sample clean-up
reduction/alkylation
digestion
(SP3, S-trap)



Buczak, Nature methods, 2020

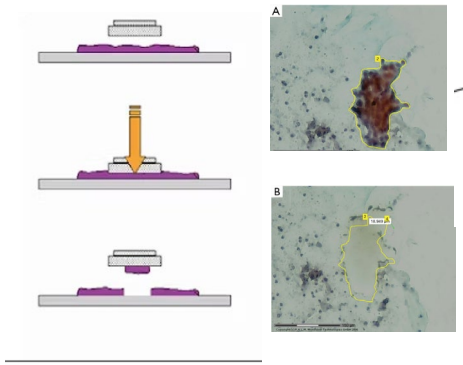
FFPE tissue treatment

- Deparaffinization
- Decrosslinking (90°C, SDS>2%)
- Homogeneization

➢ **Formalin-fixed and paraffin-embedded FFPE tissues: a gold mine for retrospective discovery and clinical proteomics studies**

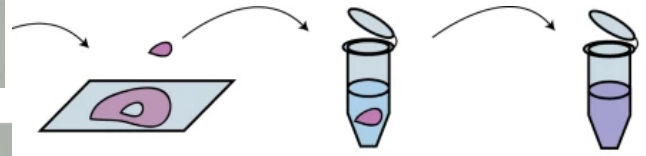
Laser microdissection (LMD) toward spatial MS proteomics within tissues

Laser-capture microdissection (LCM)



Buczak, Nature methods, 2020

Isolation of cells (distinct sub-populations)

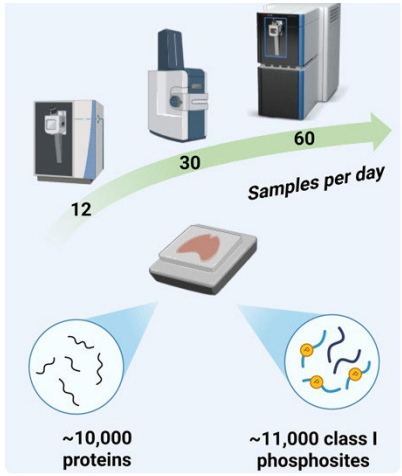


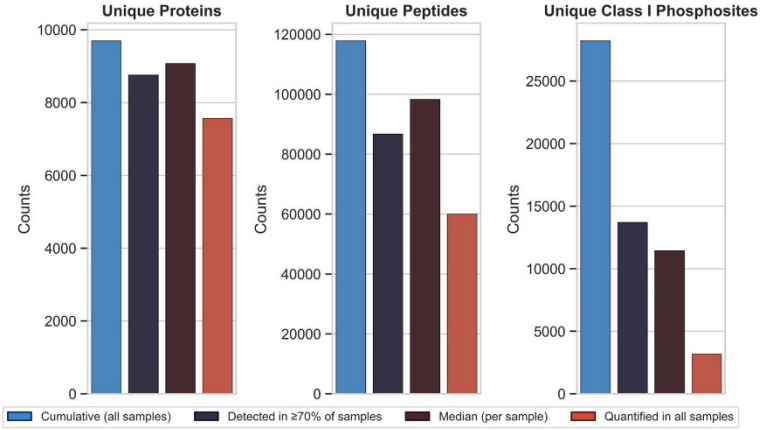
Sample clean-up
reduction/alkylation
digestion
(SP3, S-trap)

FFPE tissue treatment

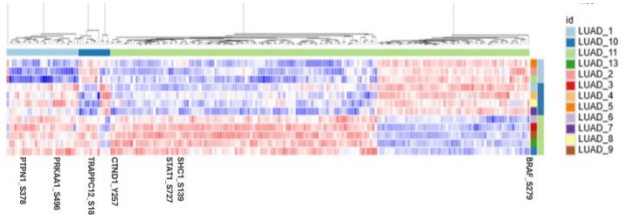
- Deparaffinization
- Decrosslinking (90°C, SDS>2%)
- Homogeneization

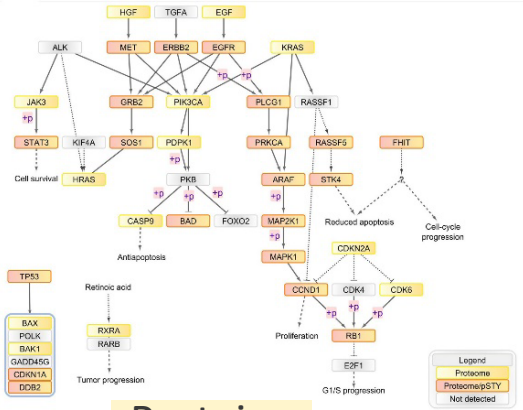
Last generation LC-MS proteomics and phosphoproteomics





Application to clinical lung cancer FFPE samples (Haines, MCP, 2025)



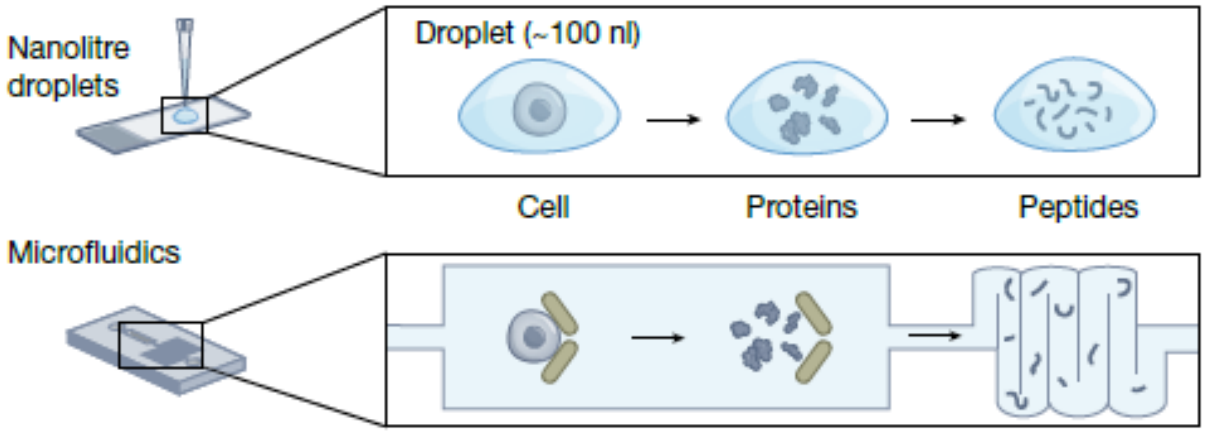


Proteins
Phosphoproteome

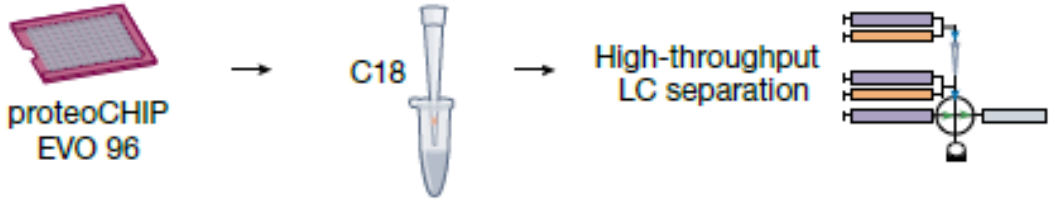
➢ **Formalin-fixed and paraffin-embedded FFPE tissues: a gold mine for retrospective discovery and clinical proteomics studies**

Miniaturization toward single cell proteomics

Single Cell Proteomics isolated by FACS or CellONE



High-throughput label-free single cell proteomics

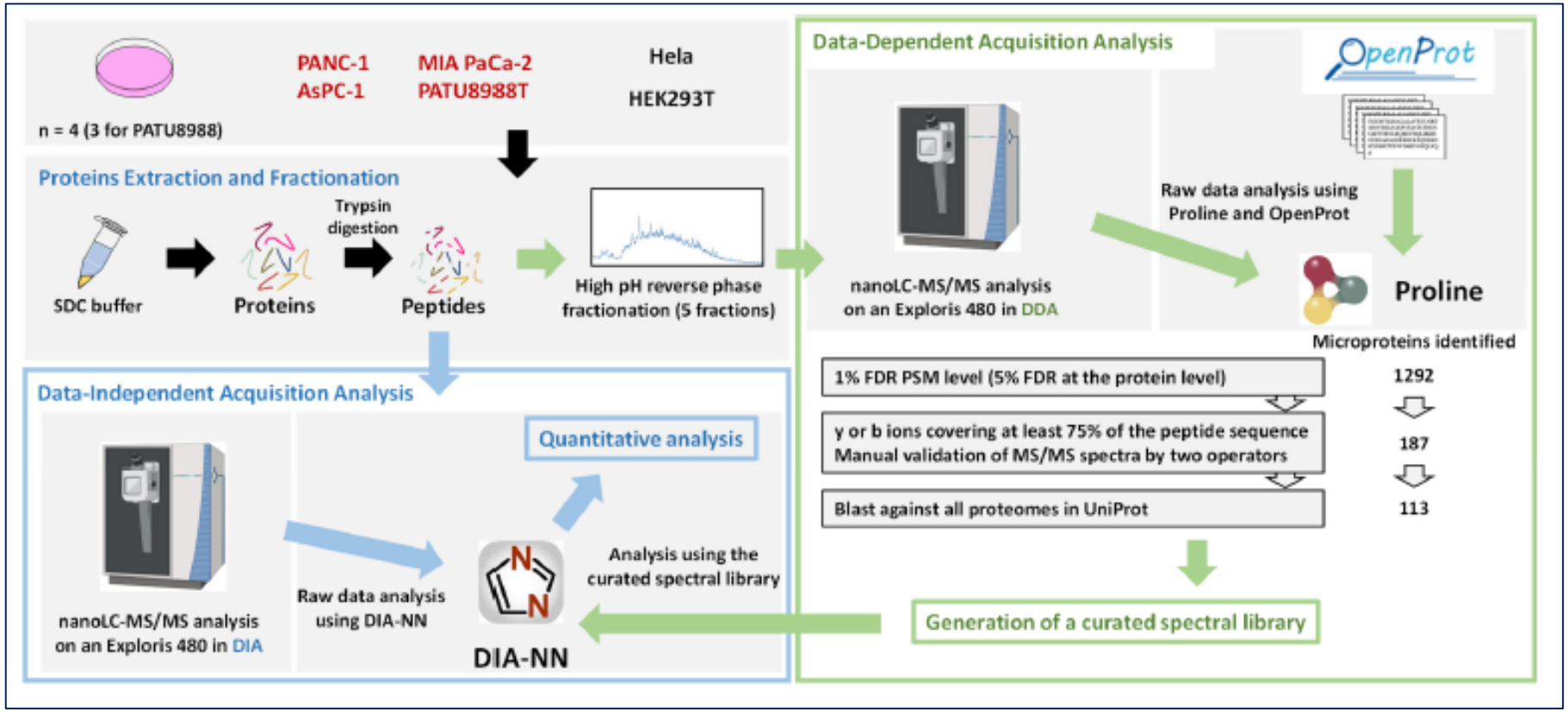


Guo et al, Nature 2025

➢ **Exploration of cell heterogeneity**

Identification of alternative proteins

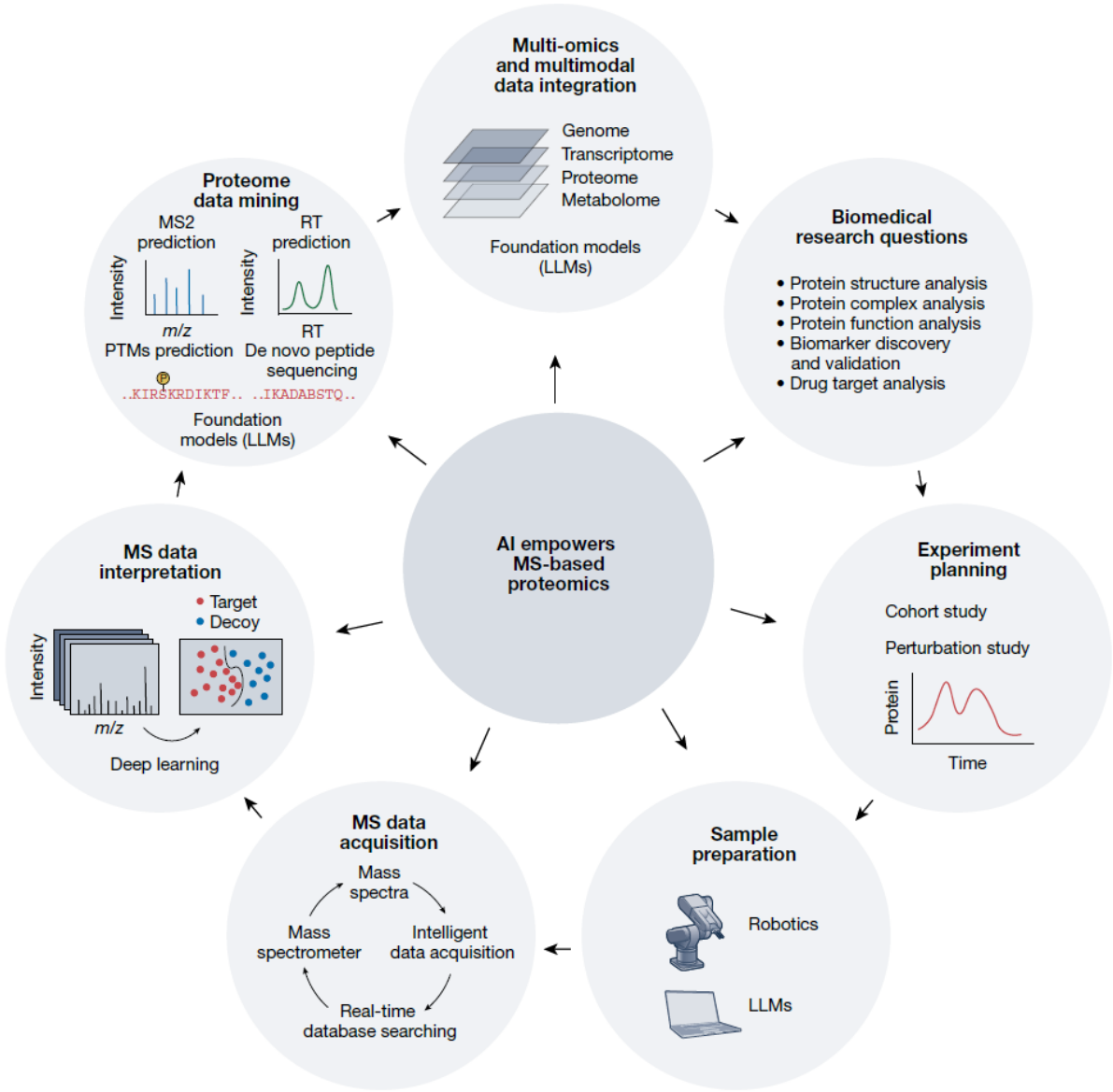
MS-based workflow to identify alternative proteins in pancreatic cancer cell lines



Guillon, Cells, 2024

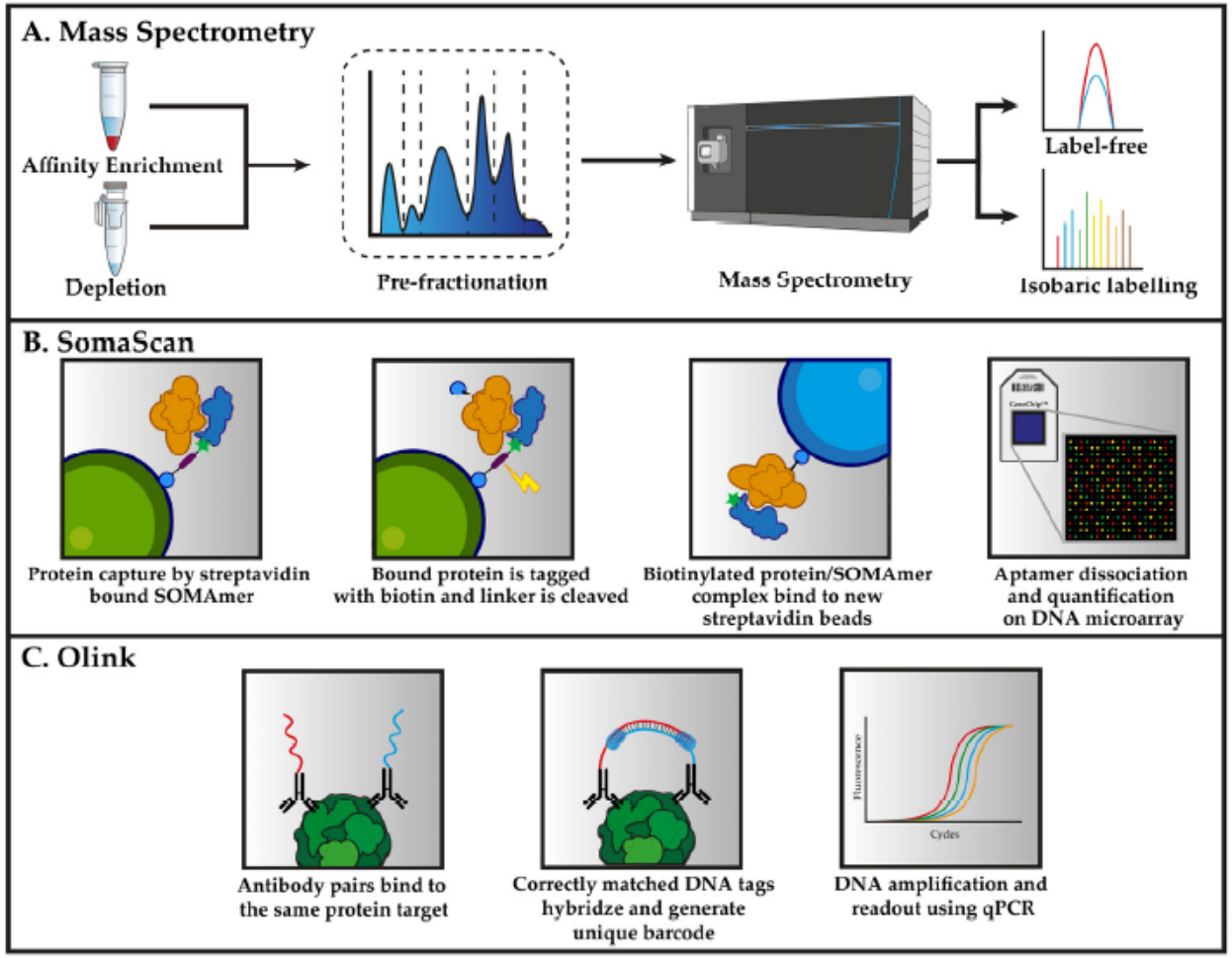
- **Expression of RNAs coding for some of the alternative proteins identified here is altered in Pancreatic tumors versus normal tissues and correlates with patient survival**

AI empowers MS-based proteomics



Guo et al, Nature 2025

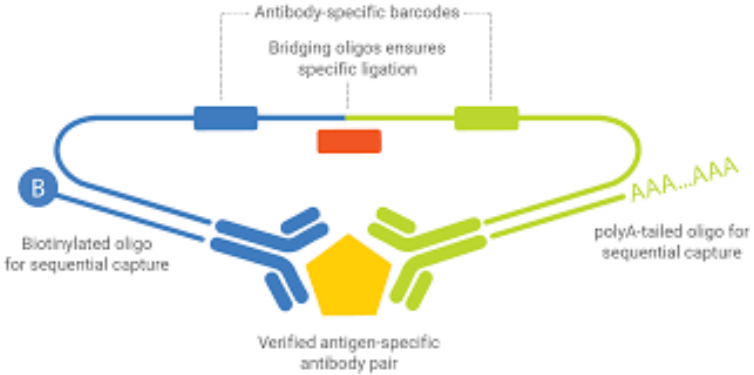
Next generation proteomics technologies



Bjødstrup Palstrøm, Biomedicines, 2022

D- NULISA

(NUcleic acid-Linked Immuno-Sandwich Assay)



1- Immuno-complex formation
2- Capture of immun-complexes to dT beads
3- Release & recapture of immuno-complexes on Streptavidin beads
4- DNA strand ligation generates reporter DNA
5- Quantification of reporter DNA by qPCR / NGS

- Protein identification: 10 – 12 000 proteins : MS
- Dynamic range: 10-12 log (6-7 log)MS
- PTMs identification: MS
- Specificity: MS
- Required expertise: MS
- Proteome coverage in plasma: MS

Conclusion and perspectives of MS proteomics

Large scale MS-proteomics offers a powerful tool for

- **in-depth analysis** characterization and quantification of proteins in highly complex samples
- deciphering the molecular composition and functional heterogeneity of **complex biological systems such as the tumor microenvironment**.
- **Biologists:** insights into cellular proteome complexity, protein function, and signalling in tumor progression.
- **Physicians:** biomarker discovery, understanding and prediction of treatment response and resistance mechanisms, and potential for personalized medicine

Last generation technologies and methodologies advances have enabled significant progress

Sensitivity Speed High-throughput Robustness
Sample Preparation Spatial analysis
Integration of artificial intelligence
Innovative data acquisition

Future developments

- Spatial proteomics and single cell proteomics
- Alternative proteins and immunopeptidomics
- Protein interactions and proximities in cellulo and in tissue
- Multi-omics integration
- AI-based bioinformatic processing of massive MS and data integration
- Combination with affinity-based proteomics approaches

Acknowledgments

Proteomic-Gif Platform
I2BC, CNRS, CEA, Université Paris-Saclay
(Gif-sur-Yvette)

David Cornu,
Laila Sago
Hélène Jean-Jacques

contact-proteomique@i2bc.paris-saclay.fr

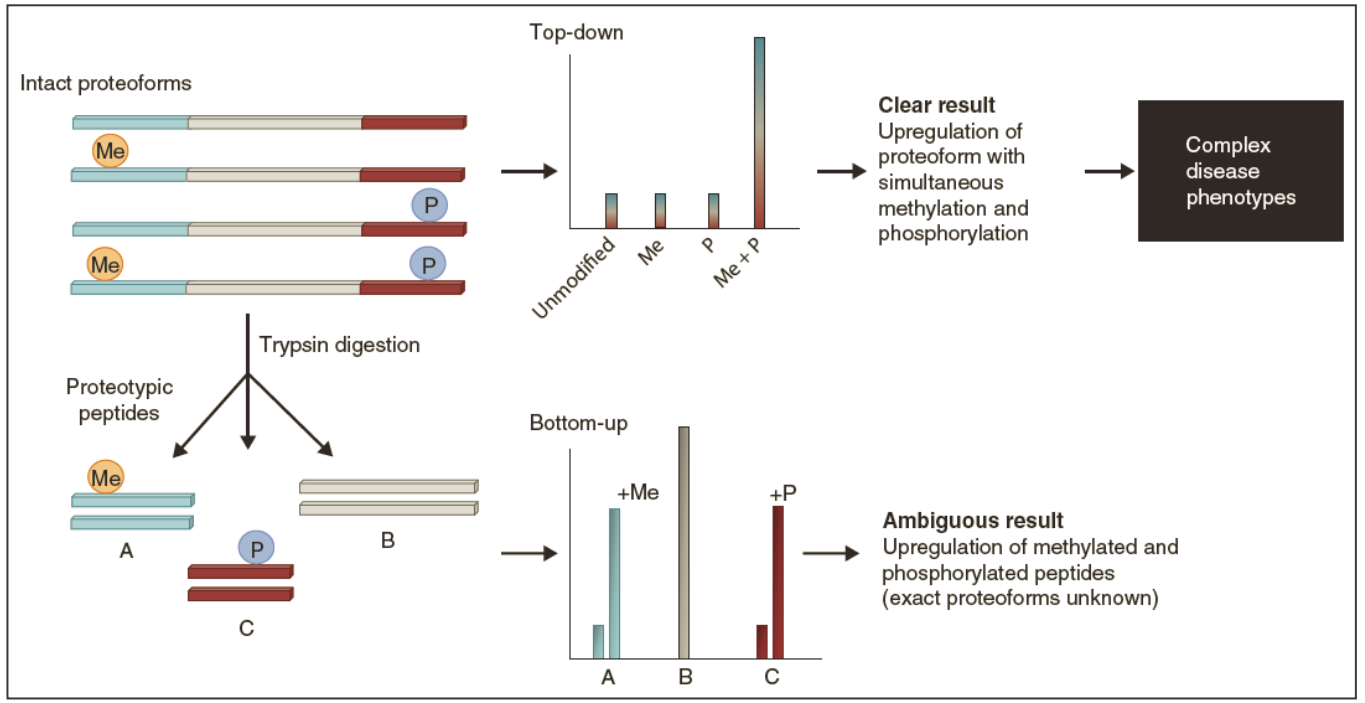
MIRCen (CEA, Fontenay-aux-Roses)

Ronald Melki
Nolwen Rey
Anselme Perrier
Mathilde Louça

Funding

- Mise en perspective de l'apport et des limites des techniques".
- *Quelle sont les techniques ou stratégies analytiques disponibles? Méthodes biochimiques, chimiques ou spectroscopiques? Comment identifier des biomarqueurs inconnus? Rechercher des biomarqueurs cibles? Quelle technique permet de répondre à quelle question ? Spectrométrie de masse, techniques séparatives, spatialomique, séquençage, détection des biomarqueurs, effets cocktails, analyse multiplex*
- *Que peut-on faire à partir de quels échantillons ? (congelés / FFPE) Quelles sont les différences et les limitations de l'analyse d'échantillons cliniques vs les modèles cellulaires? Comment analyser une biopsie solide ou un prélèvement liquide? Comment traiter le souci de la complexité biologique?*
- *Quelles contraintes ? Temps, débit d'analyse, coût, échantillonnage, délai, traitements, conservation…*
- *Quelles conséquences de l'accélération du développement de nouvelles technologies ? Est-ce que à terme la spectrométrie de masse va être remplacée par ces nouvelles approches qui promettent des sensibilité au niveau de la molécule unique?*
- *Que dire des avantages et inconvénient des approches single cell? Sont elles toujours adaptées?*
- *Analyse / traitement des données : autonomie / spécialistes ? Collaborations ? Que décider? Complexité et rapidité des développement bioinformatiques, complémentarité? Quand doit-on faire du top-down ou du bottom-up? Comment traiter les indormations provenant d'organismes peu ou non séquencés comme dans le cas du microbiote?*
- *Intérêt de l'IA : où est-ce que ça peut entrer, quelles sont ses limites ? Existe-t-il des solutions déjà disponibles?*
- *Où les Omics se rejoignent-elles ? (protéomique, peptidomique, métabolomique, points communs, différentes, corrélations génomique / protéomique …) Quel est l'intérêt d'une approche multi-omique et est-elle toujours pertinente? Comment intégrer des données multiomiques?*

ApprocheTop-Down (fragmentation de protéines entières)



Savaryn JP, Genome Med, 2013, 5:53

Other spatial proteomics approaches