Weatherall Institute for Molecular Medicine DPhil Course 2019
Tuesday, November 19th, 2019

UNIVERSITY OF OXFORD

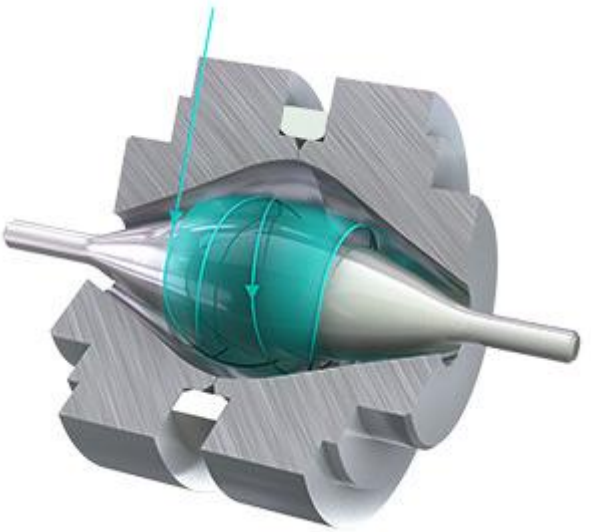
TDI TARGET DISCOVERY INSTITUTE

Basic Principles in Proteomics & Metabolomics

Darragh P. O'Brien, PhD

Target Discovery Institute,
University of Oxford
United Kingdom

darragh.obrien@ndm.ox.ac.uk

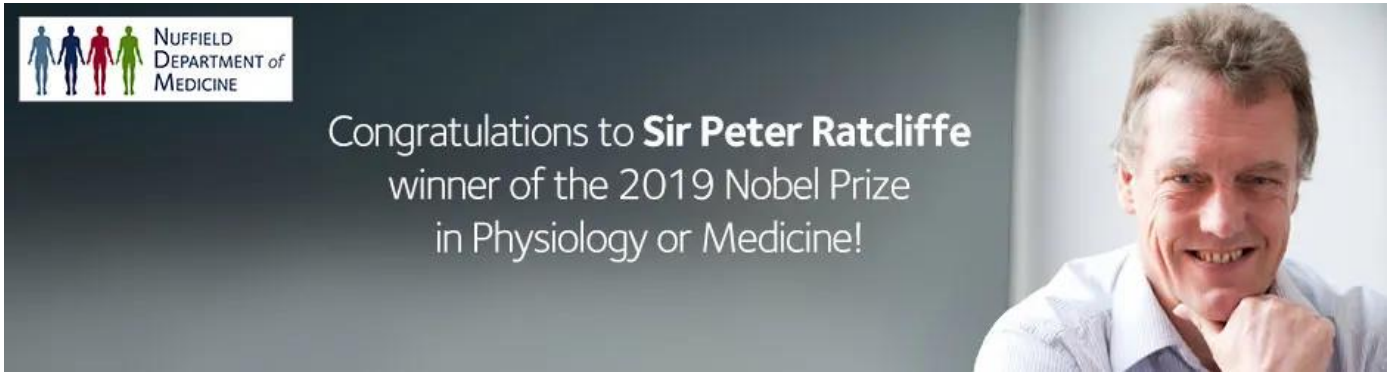


Target Discovery Institute

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

http://www.tdi.ox.ac.uk/mass-spectrometry



Mass Spec @ Old Road Campus

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

TDI-MS Lab

- LC-MS/MS
- GCxGC-MS
- Ion mobility

METABOLOMICS

PROTEOMICS

LIPIDOMICS / METABOLOMICS

Data Analysis

UNIVERSITY OF OXFORD

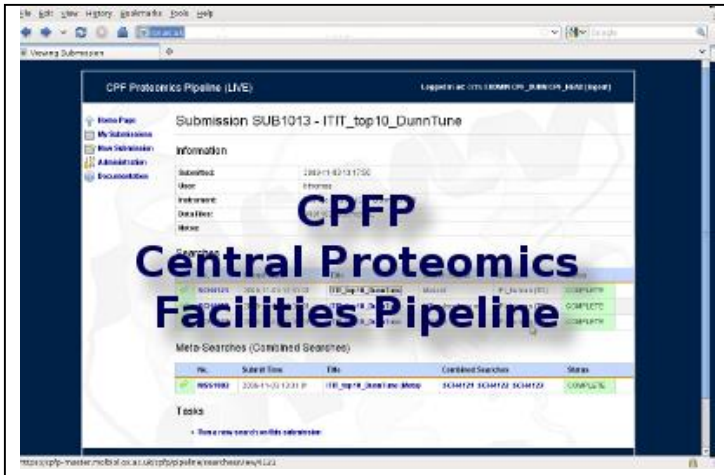
TARGET DISCOVERY INSTITUTE

Mohammed Shabaz
CPF Dunn School/Biochem
Orbitraps

Iolanda Vendrell
ORCRB
LC-MS/MS (ETD)

Rod Chalk
SGC
LC-MS

James McCullagh
Carol Robinson
LC-MS , LC-MS/MS, MALDI



MATRIX SCIENCE

CBRG/Dunn School/Biochemistry

INGENUITY SYSTEMS

Mass Profiler Professional
Proteome Discoverer
Protein Pilot / PEAKS

Proteins/Peptides as Biomarkers: Clinical Proteomics

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

Molecules as marker candidates:

Metabolites →	Metabolomics
Drugs	
Peptides →	Proteomics
Proteins	
Lipids →	Lipidomics
DNA/mRNA →	Transcriptomics

Sources:

Serum/Plasma
Urine
Blood
Synovial fluid
Tears
Other body fluids

Defined cell populations
Cell supernatants
Biopsy material
Semen
Endometrial Lavage
Endometrium / lesion
Tissues

→ **Clinical Proteomics**

PROTEin complement to a genOME

UNIVERSITY OF OXFORD

TARGET DISCOVERY INSTITUTE

One Genome…

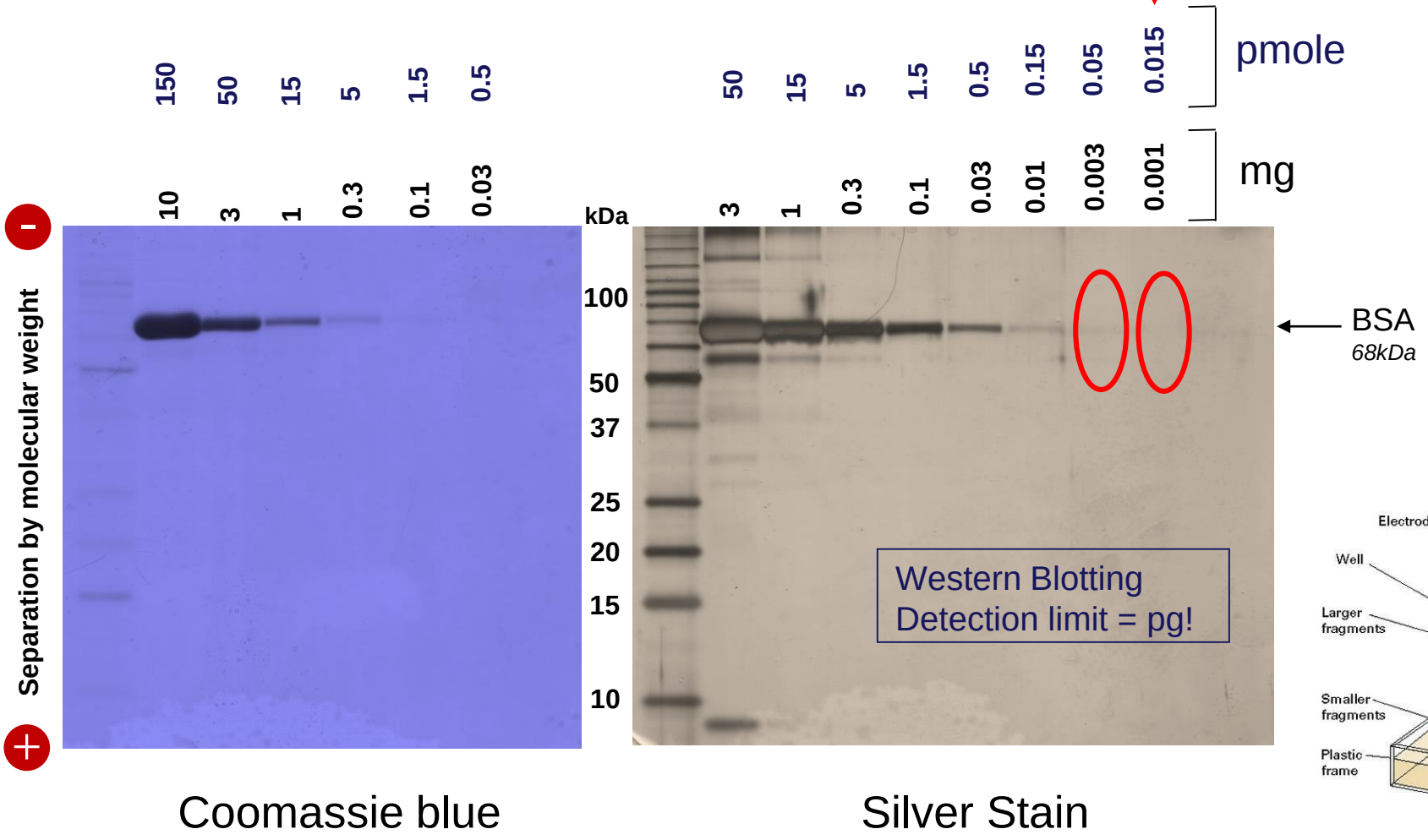
…Two Proteomes

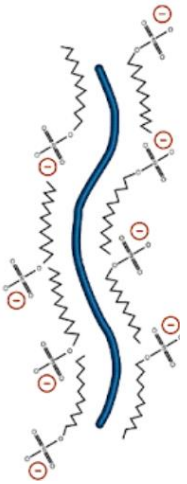
- To characterize functional protein networks and their dynamic alteration during physiological and pathological processes, proteins have to be identified, sequenced, categorized and classified with respect to their function and interaction partners in a protein network
- This is achieved by 'proteomics', the combination of high-resolution protein separation techniques with mass spectrometry and modern sequence database mining tools

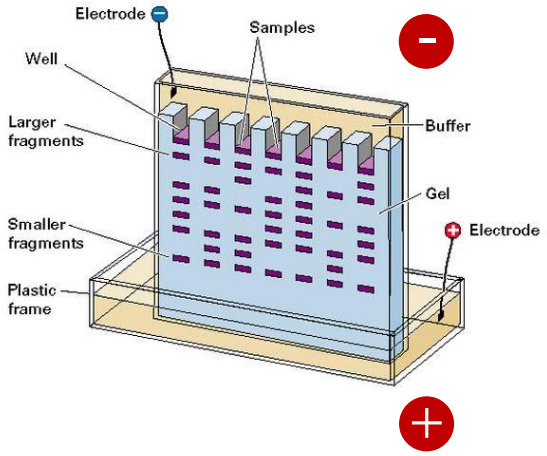
Protein Analysis by SDS-PAGE

Sodium Dodecylsulfate Gel-Electrophoresis: protein separation by molecular Weight along an electric gradient, molecules travel through a poly-acrylamide polymer

Detection limit by mass spectrometry: $3x10^9$ molecules

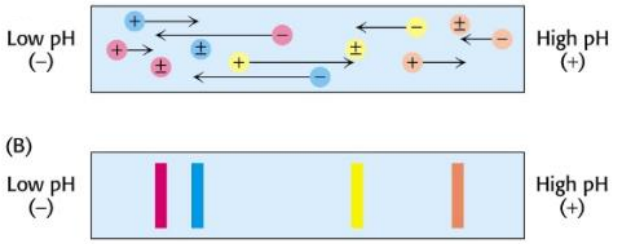




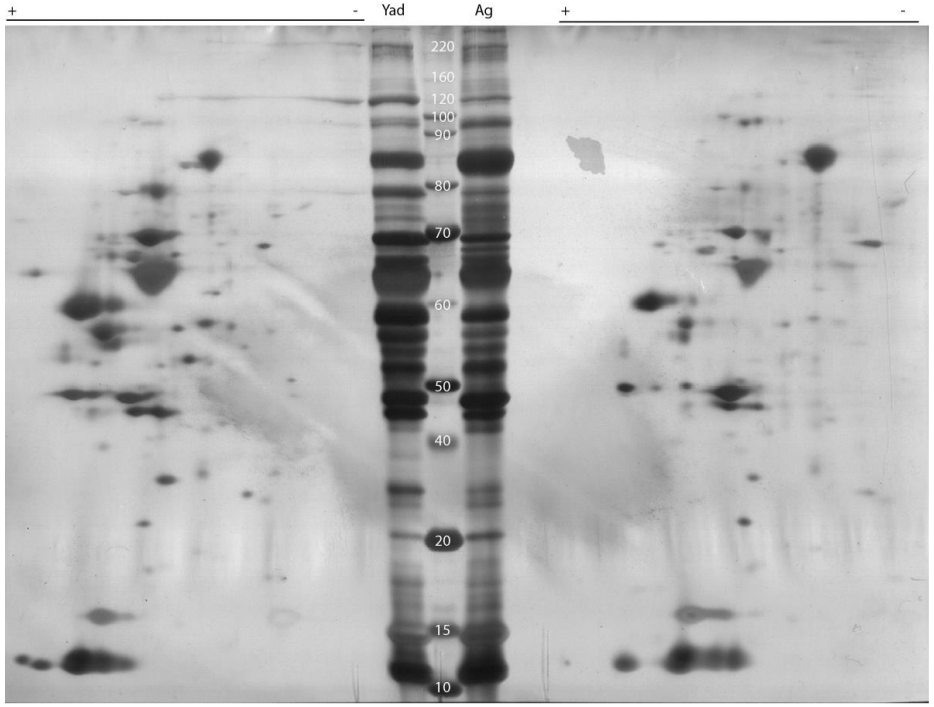


Gene versus Protein Expression

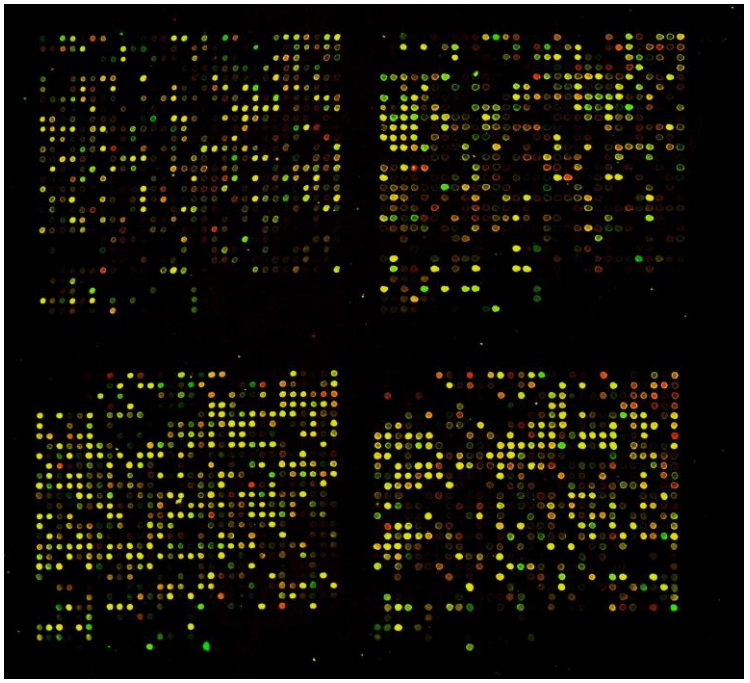
Two-dimensional separation
- 1st dimension – IEF (paper strips, tube gels)
- 2nd dimension – SDS PAGE



2D-gel profile of
young and aged muscle

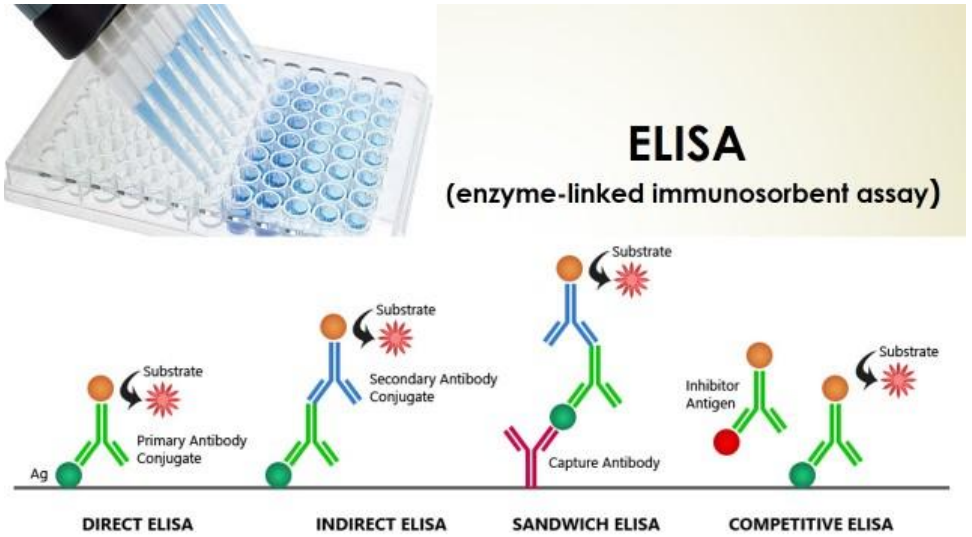


GeneChip array of
young and aged muscle

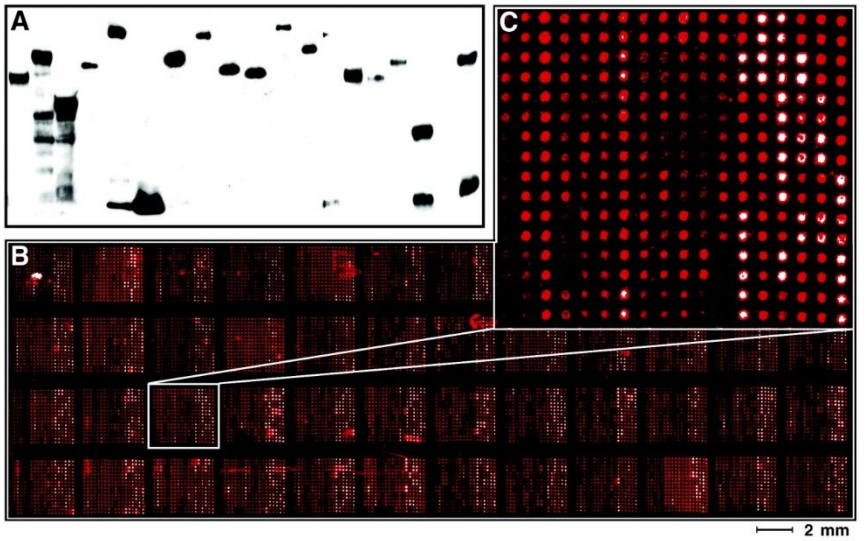


M. Altun, E.Edström, E.Spooner, A. Flores-Moralez E. Bergman, P. Tollet-Egnell, G.Norstedt, B. Kessler and B.Ulfhake. Muscle & Nerve, 2007, 36(2): 223-33.

Protein and Antibody Microarrays



http://www.biosciencenotes.com/elisa-enzyme-linked-immunosorbent-assay/



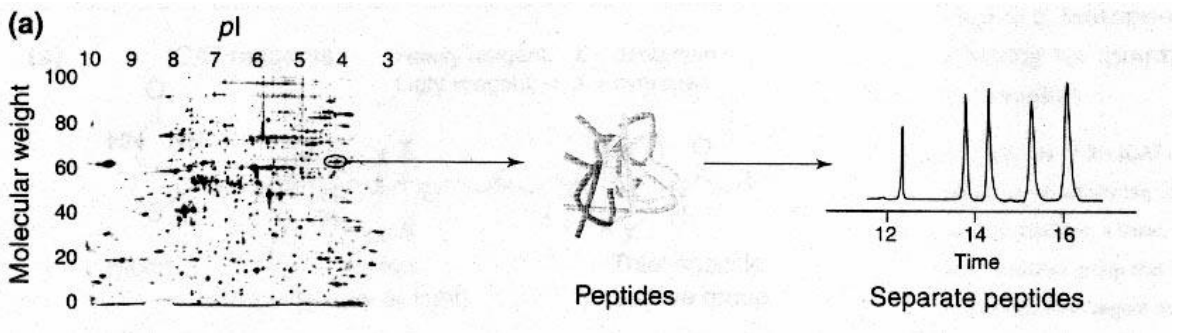
- LUMINEX (different protein families: cytokines, kinome, growth factors etc.)
- Protein arrays
- Antibody arrays
- Nucleic Acid Programmable Protein Array (NAPPA)
- Peptide arrays

Proteomics Applications

- Standard proteome analysis
  - ✓ MW, Protein ID
- Analysis of Multi-Protein Complexes
  - ✓ Protein Tagging, Tandem Affinity Purification (TAP)
- Analysis of post-translational modifications (PTMs)
  - ✓ > 400 known modifications
  - ✓ Phosphorylation, Acetylation, Methylation, Ubiquitination
  - ✓ Glycosylation, Hydroxylation, Lipid modification
- Quantitative Analysis of proteins in health and disease
  - ✓ Clinical Proteomics
- Systems Biology
  - ✓ Combining large proteomic & genomic datasets
  - ✓ Mathematical modelling

Standard Proteome Analysis

Protein identification by MS



- 2D gel
- Silver stain
- Spot isolation
- Enzymatic digestion (Trypsin)
- Chromatography

MALDI-TOF

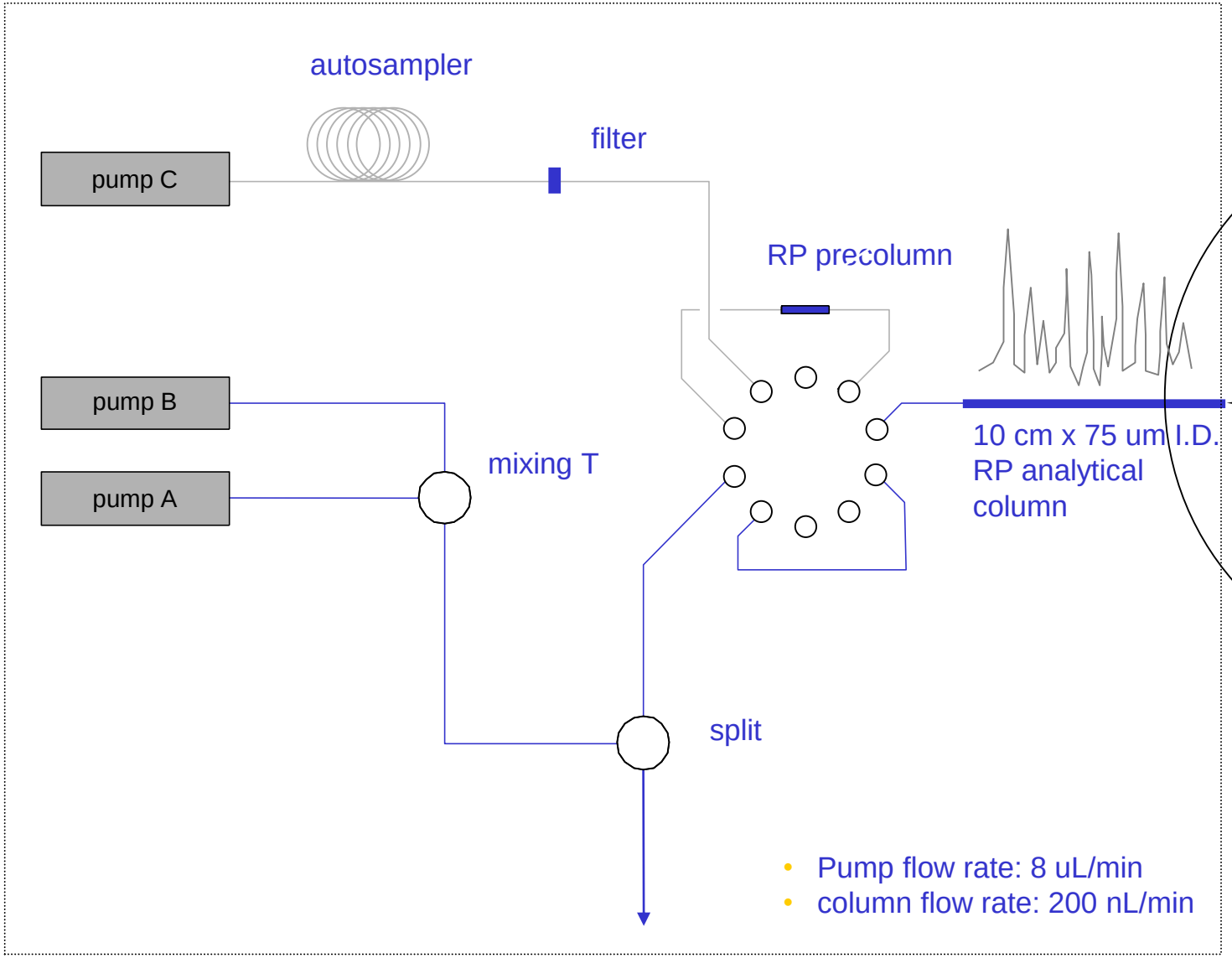
LC-MS/MS **?**

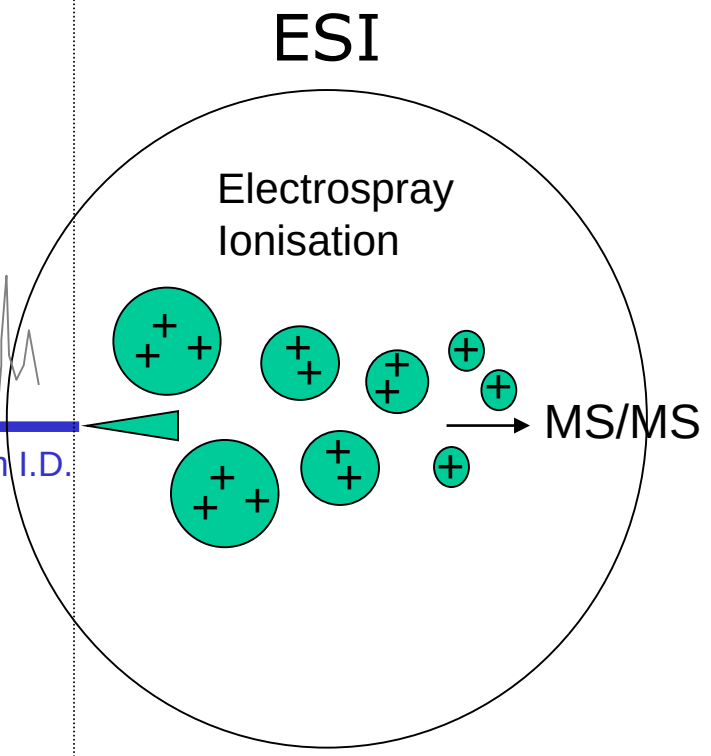
nLC-ESI-MS/MS: Nanoscale LC Set up

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

LC Nano-Liquid Chromatography





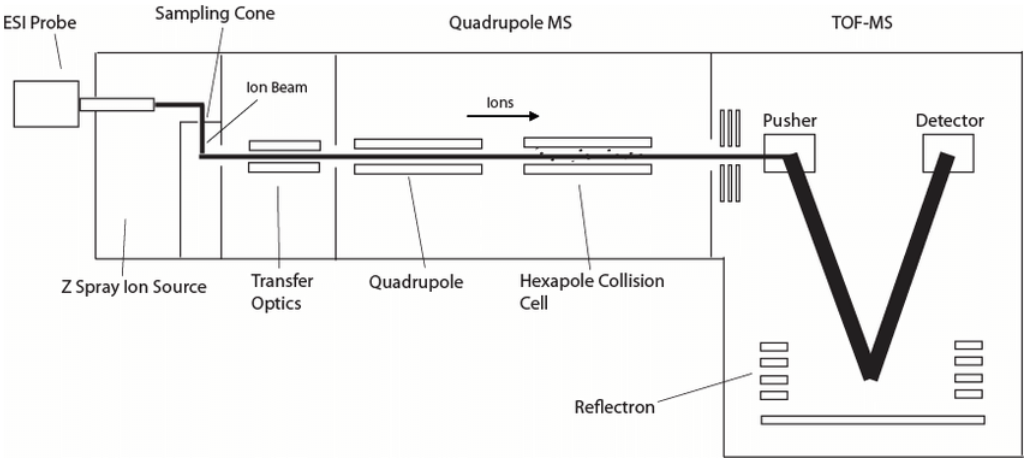
Tandem Mass Spectrometers

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

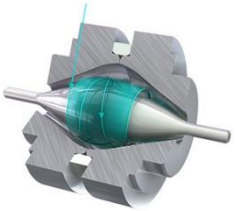
Quadrupole Time-of-flight (QTOF)

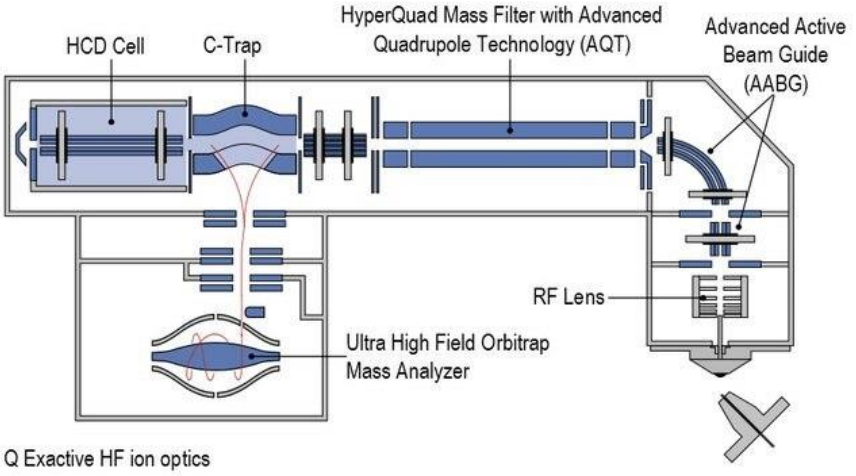
Resolution 20,000-40,000
Mass Accuracy 1-5ppm
Sensitivity 1fmol



Orbitrap

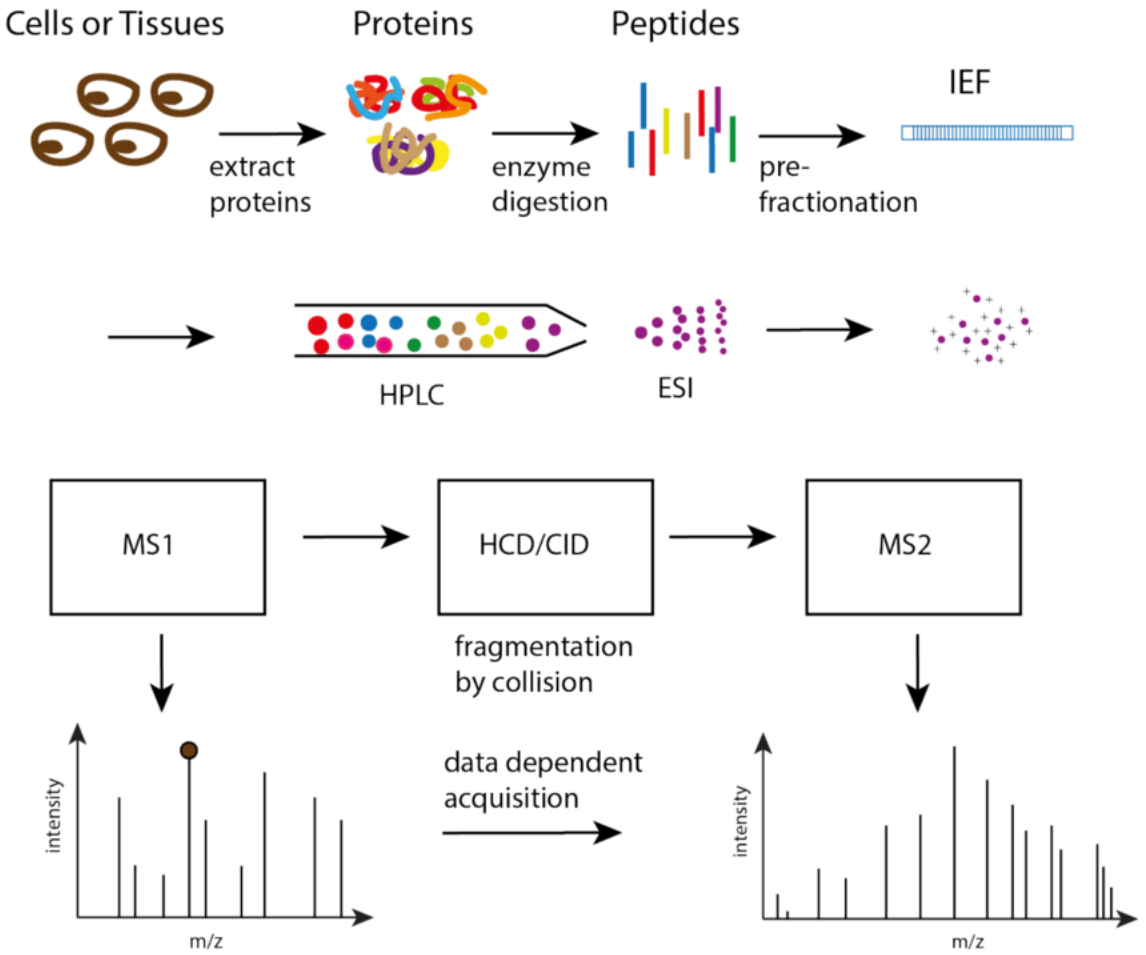
Resolution 50,000-500,000
Mass Accuracy <1ppm
Sensitivity <1fmol





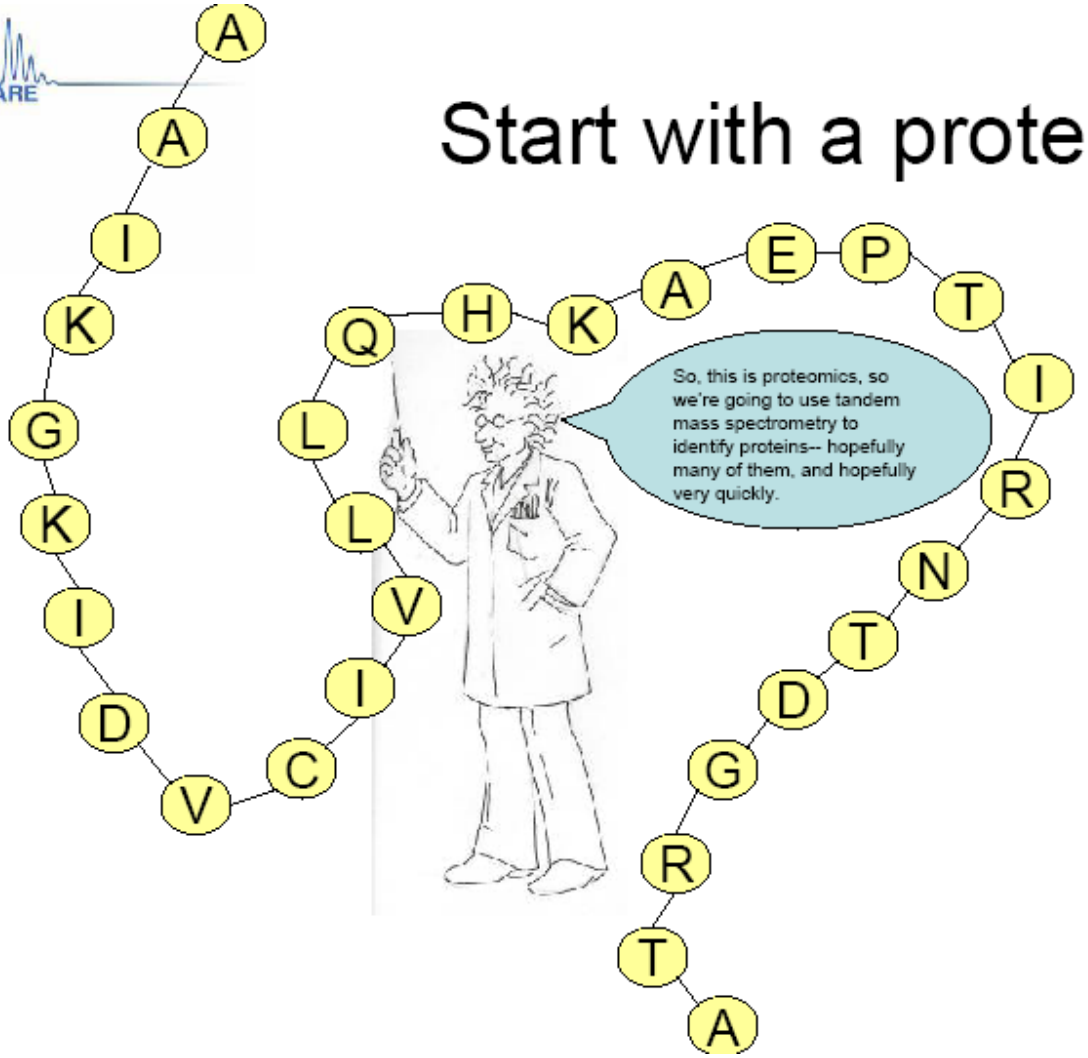
How a mass spectrometer works:

Protein Identification by MS/MS

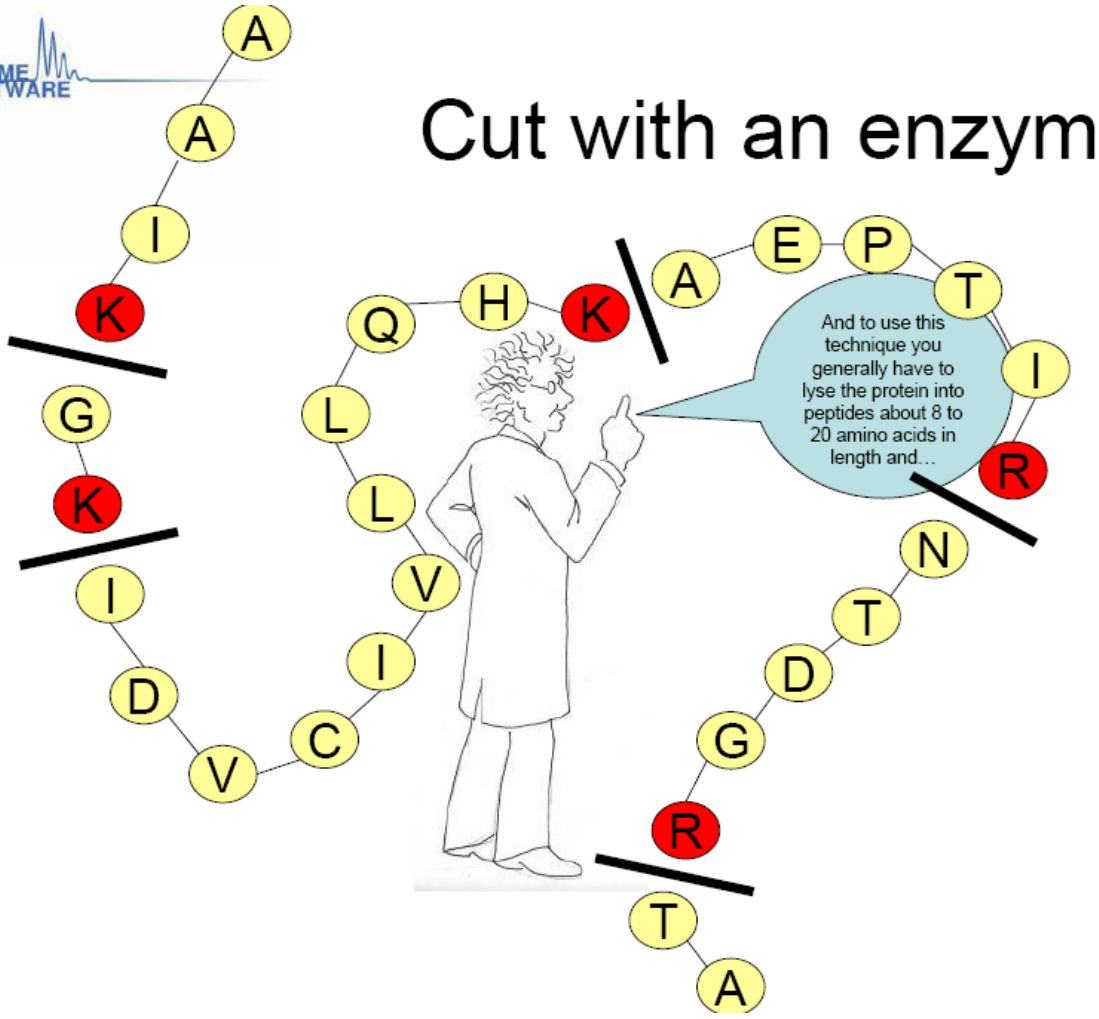


- Data analysis
- Theoretical sequences from a defined mass
- Generation of a theoretical spectrum
- Matching with the acquired spectrum
- *De novo* sequencing

Start with a protein

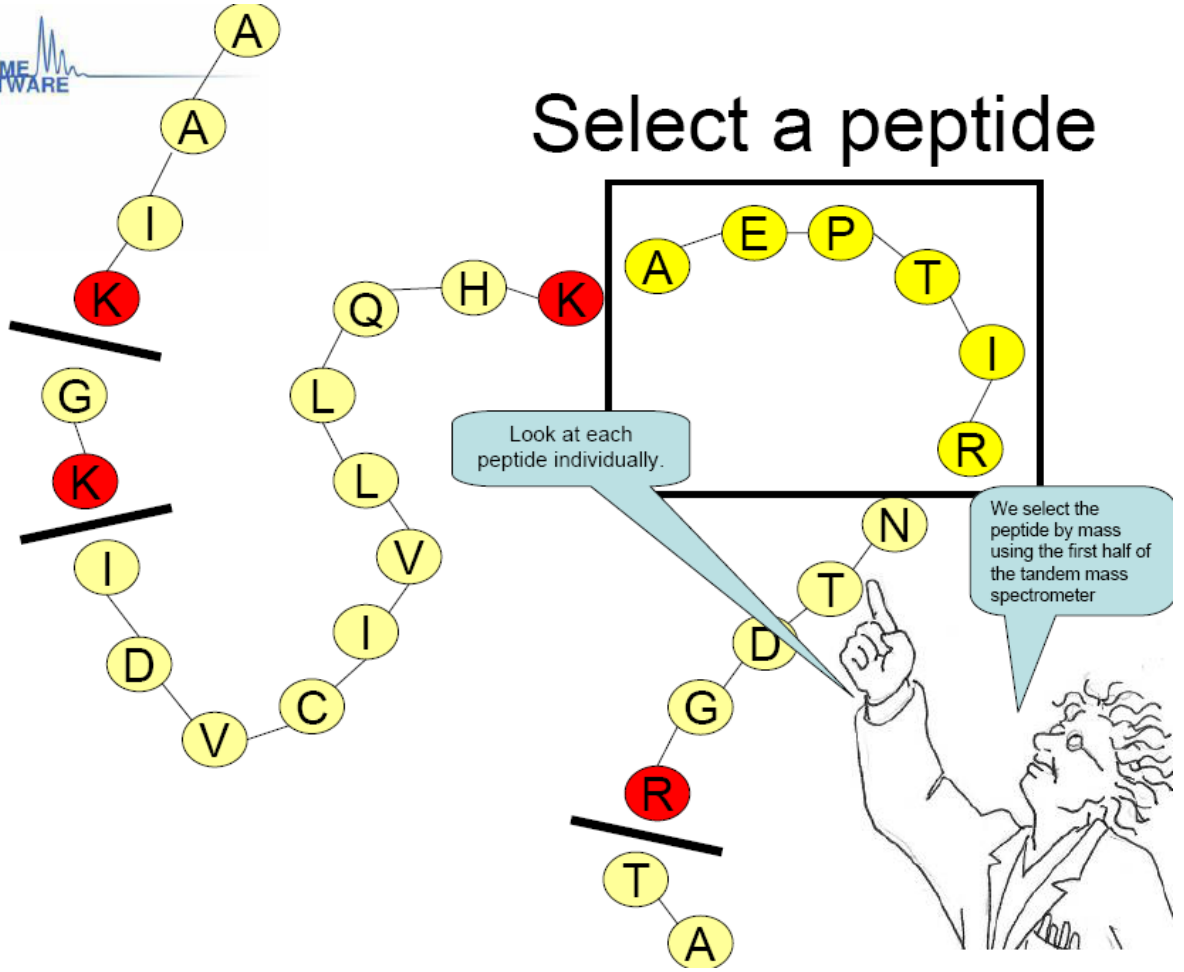


Cut with an enzyme

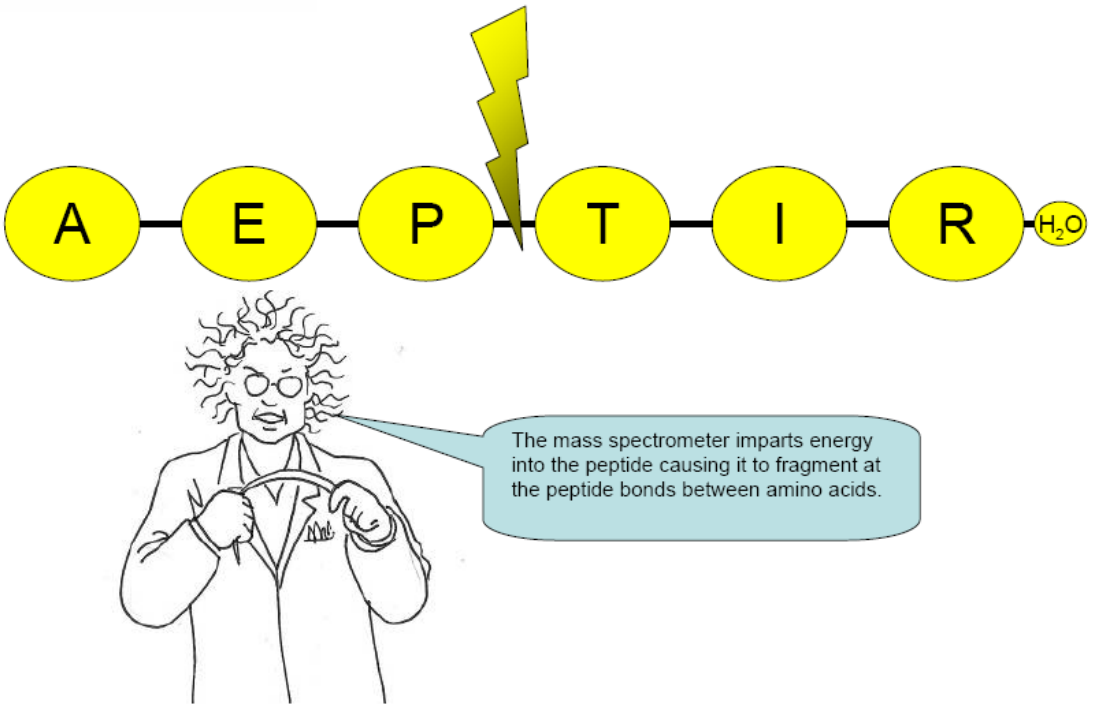


PROTEOME SOFTWARE

Select a peptide

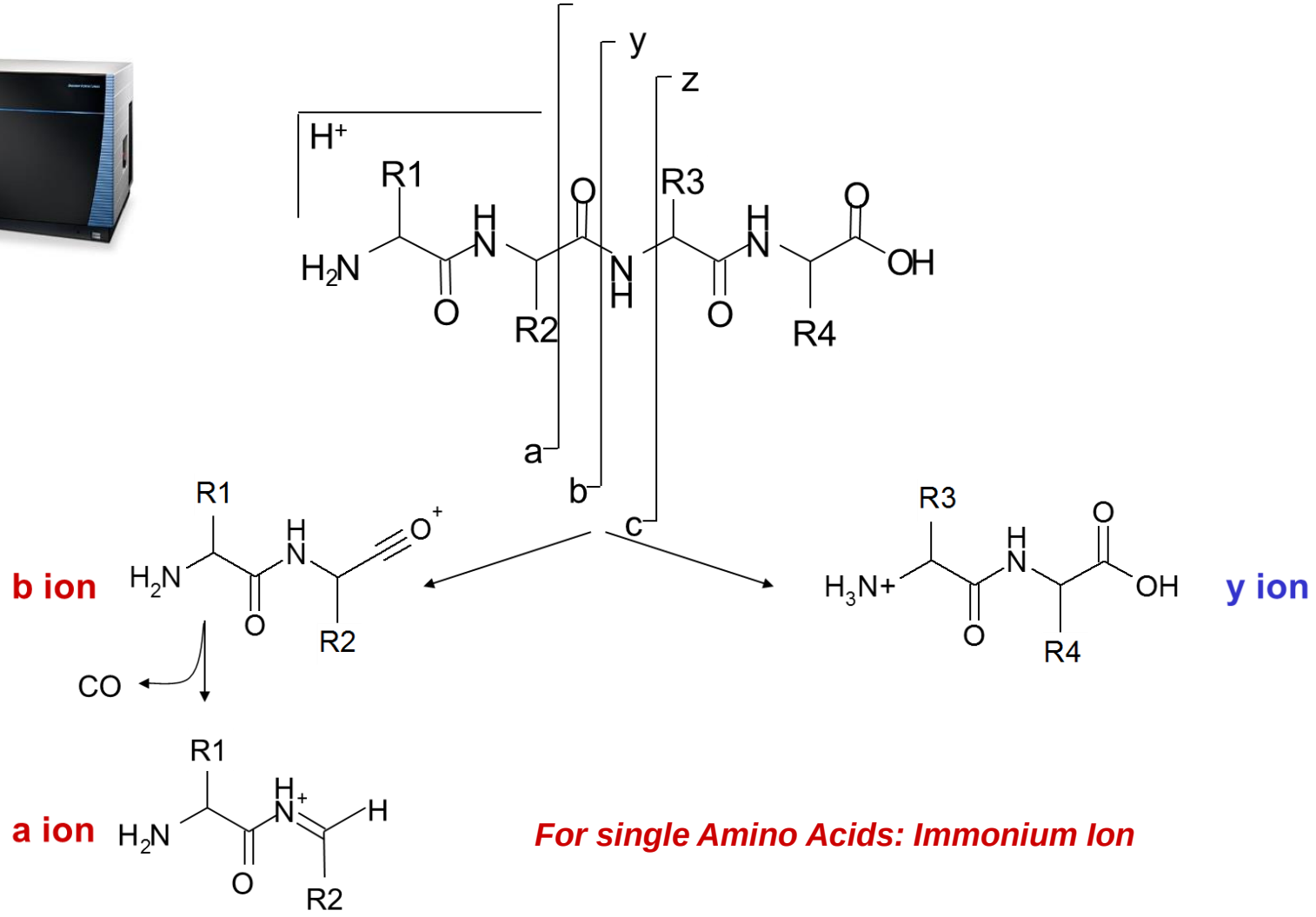


Impart energy in collision cell



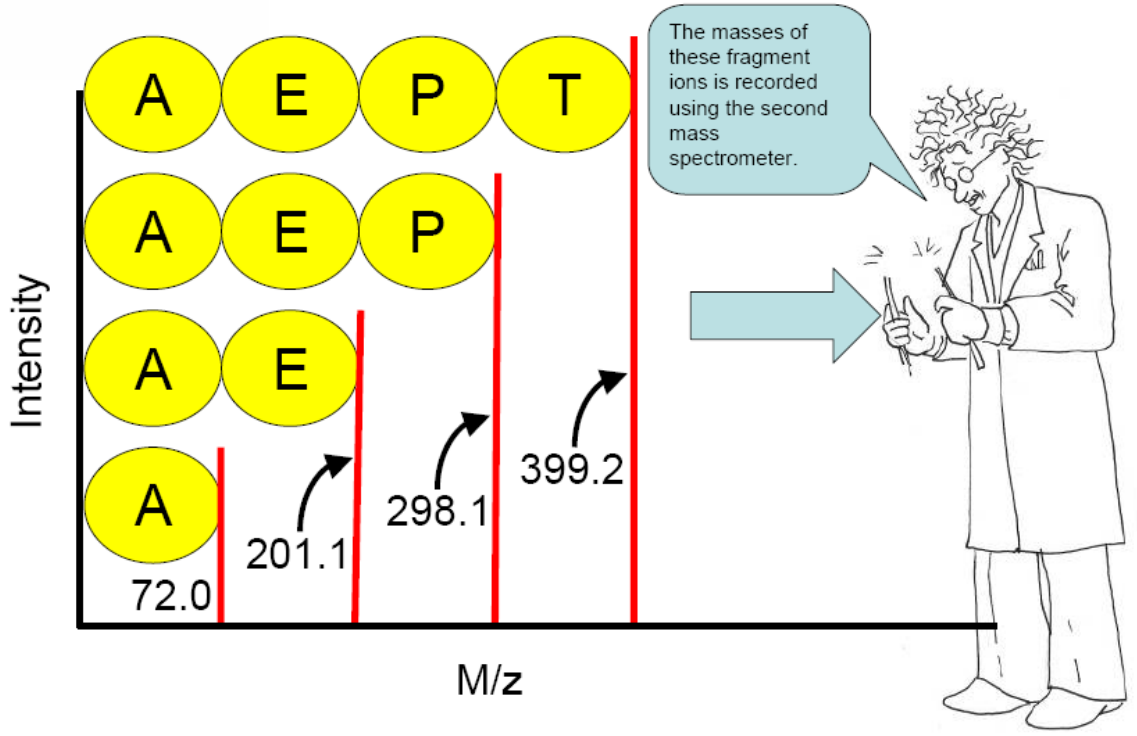
Protein Identification by MS/MS

Peptide Fragmentation Pattern



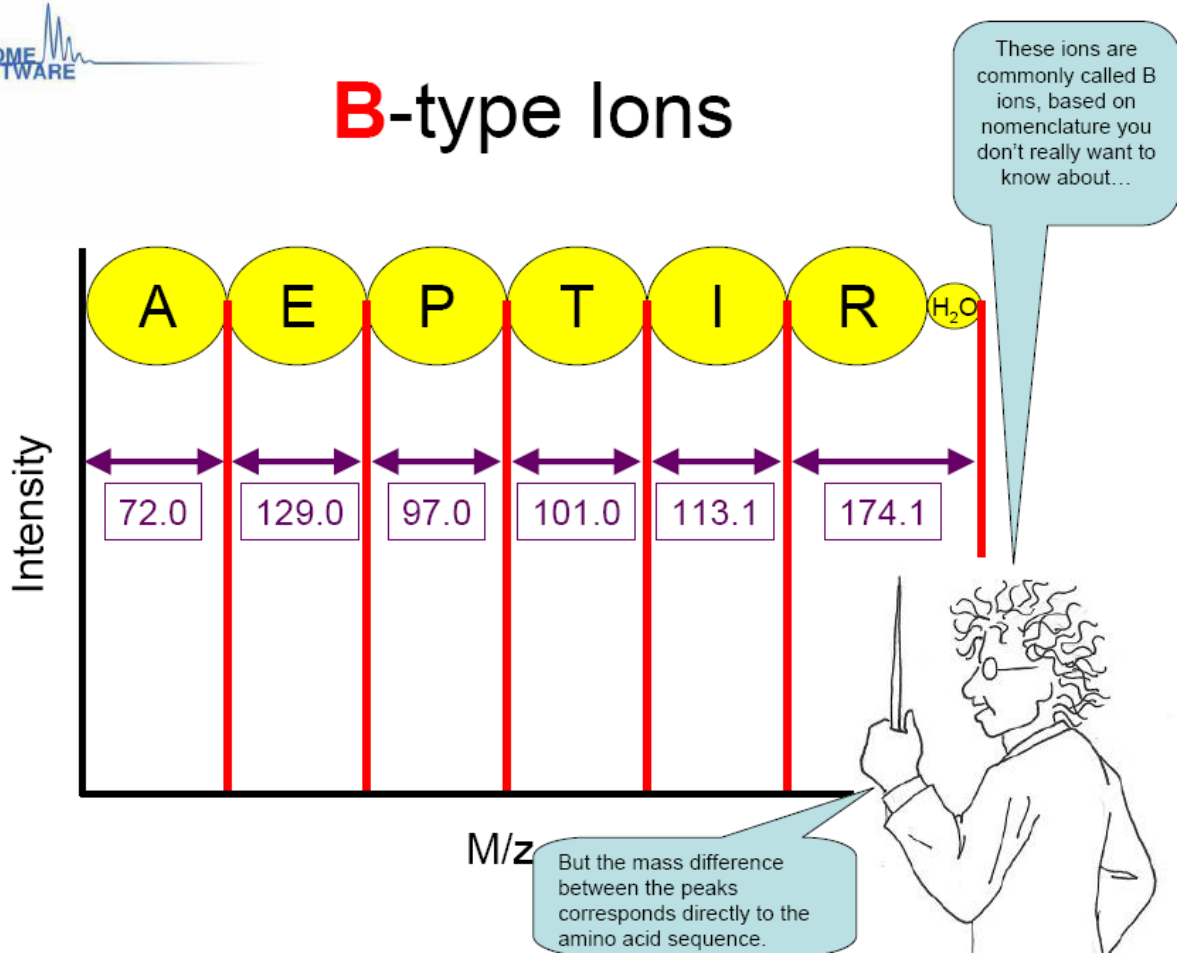
For single Amino Acids: Immonium Ion

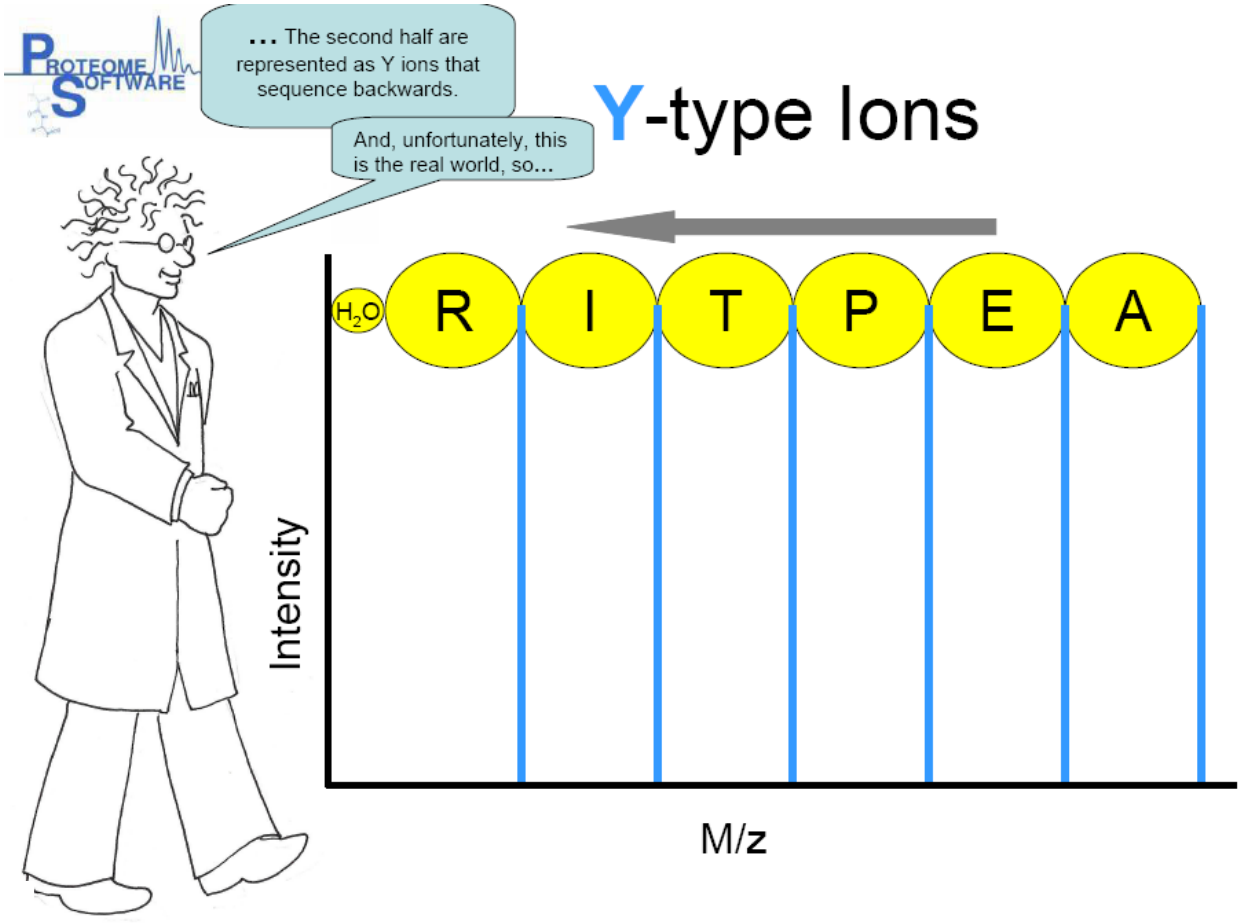
Measure mass of daughter ions



PROTEOME SOFTWARE

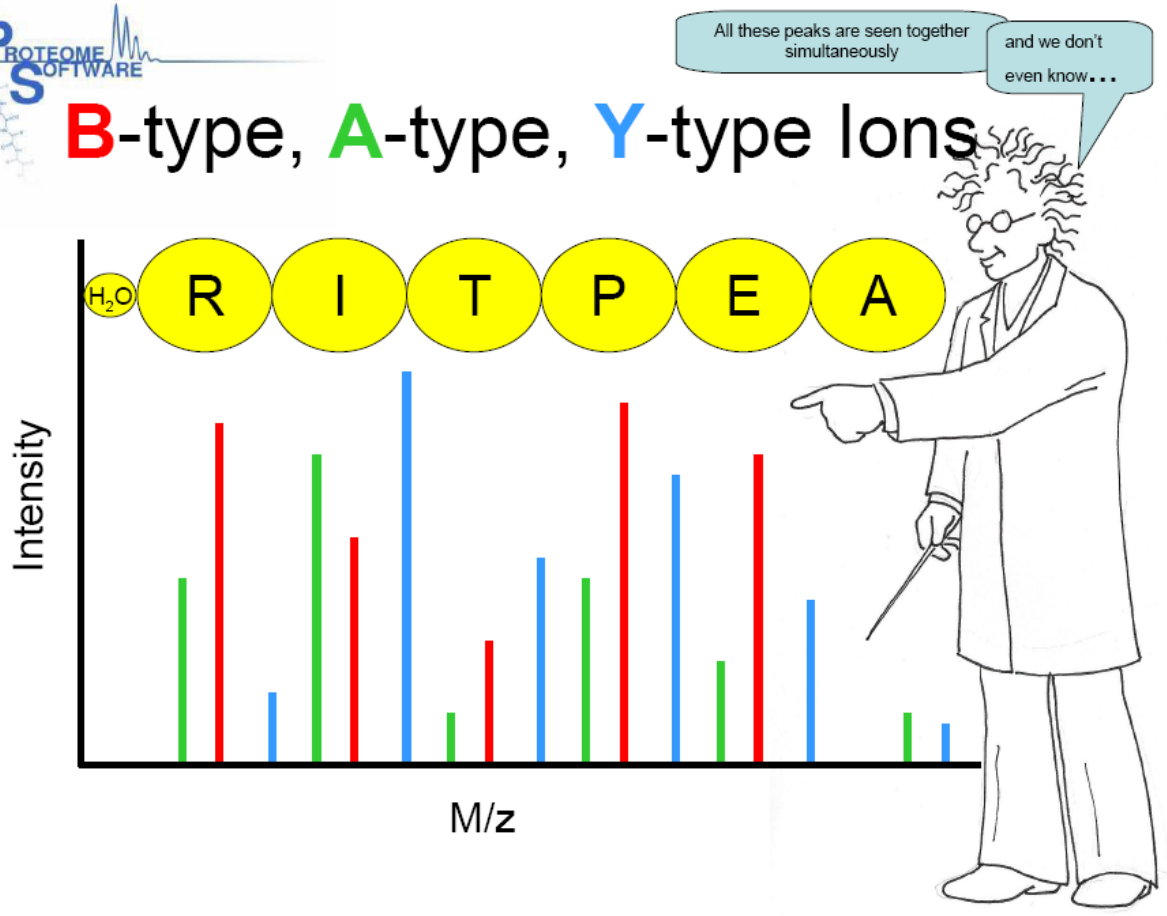
B-type Ions



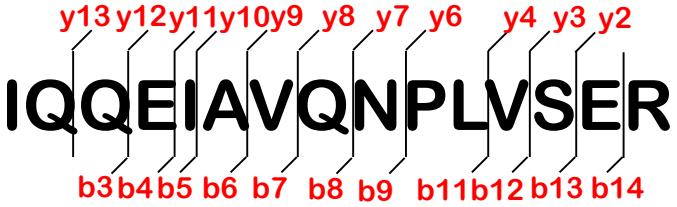

PROTEOME SOFTWARE
... The second half are represented as Y ions that sequence backwards.
And, unfortunately, this is the real world, so...
Y-type Ions
H2O
R
I
T
P
E
A
Intensity
M/z

PROTEOME SOFTWARE

B-type, A-type, Y-type Ions



…Is The Basis for Sequence Information



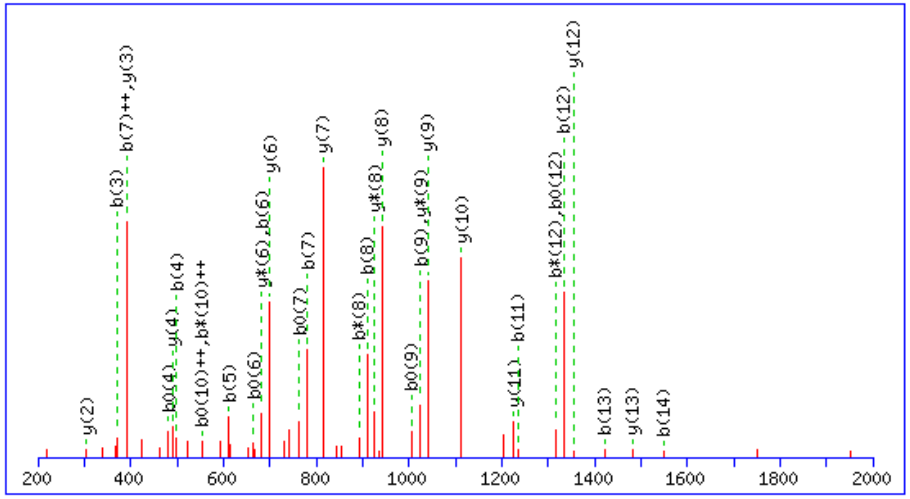
MATRIX SCIENCE Mascot Search Results

Peptide View

MS/MS Fragmentation of **IQQEIAVQNPLVSER**
Found in **OTUB1_HUMAN**, (Q96FW1) Ubiquitin thioleserase protein OTUB1 (EC 3.4.-.-) (Otubain 1) (OTU domain-containing ubiq

Match to Query 88: 1722.985448 from(862.500000,2+) intensity(22678.0000)
Cmpd 32, +MSn(863.13), 10.4 min
From data file D:\Data\October2005\MB 2_1-B,2_01_1888.d\Analysis.mgf

Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 2000 Da



Monoisotopic mass of neutral peptide Mr(calc): 1722.93
Ions Score: 112 **Expect:** 1e-07
Matches (Bold Red): 35/158 fragment ions using 38 most intense peaks

#	b	b^{++}	b*	b^{*++}	b^{0}	b^{0++}	Seq.	y	y^{++}	y*	y^{*++}	y^{0}	y^{0++}	#
1	114.09	57.55					I							15
2	242.15	121.58	225.12	113.07			Q	1610.85	805.93	1593.82	797.42	1592.84	796.92	14
3	**370.21**	185.61	353.18	177.09			Q	**1482.79**	741.90	1465.76	733.39	1464.78	732.89	13
4	**499.25**	250.13	482.22	241.62	**481.24**	241.12	E	**1354.73**	677.87	1337.71	669.36	1336.72	668.86	12
5	**612.34**	306.67	595.31	298.16	594.32	297.67	I	**1225.69**	613.35	1208.66	604.84	1207.68	604.34	11
6	**683.37**	342.19	666.35	333.68	**665.36**	333.18	A	**1112.61**	556.81	1095.58	548.29	1094.60	547.80	10
7	**782.44**	**391.72**	765.41	383.21	**764.43**	382.72	V	**1041.57**	521.29	**1024.54**	512.77	1023.56	512.28	9
8	**910.50**	455.75	**893.47**	447.24	892.49	446.75	Q	**942.50**	471.75	**925.47**	463.24	924.49	462.75	8
9	**1024.54**	512.77	1007.52	504.26	**1006.53**	503.77	N	**814.44**	407.72	797.42	399.21	796.43	398.72	7
10	1121.59	561.30	1104.57	**552.79**	1103.58	**552.30**	P	**700.40**	350.70	**683.37**	342.19	682.39	341.70	6
11	**1234.68**	617.84	1217.65	609.33	1216.67	608.84	L	603.35	302.18	586.32	293.66	585.34	293.17	5
12	**1333.75**	667.38	**1316.72**	658.86	**1315.74**	658.37	V	**490.26**	245.63	473.24	237.12	472.25	236.63	4
13	**1420.78**	710.89	1403.75	702.38	1402.77	701.89	S	**391.19**	196.10	374.17	187.59	373.18	187.10	3
14	**1549.82**	775.41	1532.80	766.90	1531.81	766.41	E	**304.16**	152.58	287.13	144.07	286.15	143.58	2
15							R	175.12	88.06	158.09	79.55			1

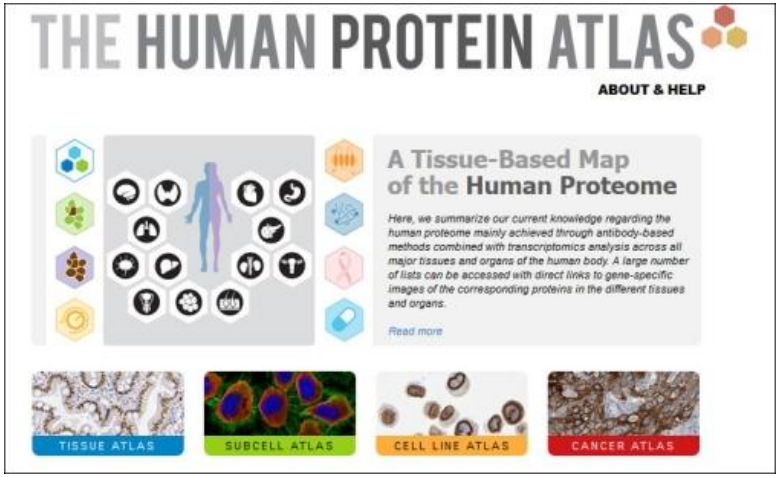
Batycka and Kessler, 2005

Progress in Proteome Analysis
"Proteome/Sub-Proteome Screens"

Human Protein Atlas (tissue based map of human proteome)
>17,005 unique proteins, 217 tissues
(www.proteinatlas.org)

Human Proteome Map (HPM)
>17,294 proteins (84% of human genes, 17 adult, 7 fetal tissues)
(http://humanproteomemap.org)

"Proteomes" of 11 cell lines analysed
(~10,500 proteins each; >90% of the total proteome ??)
(Geiger T *et al.,* MCP 2012)

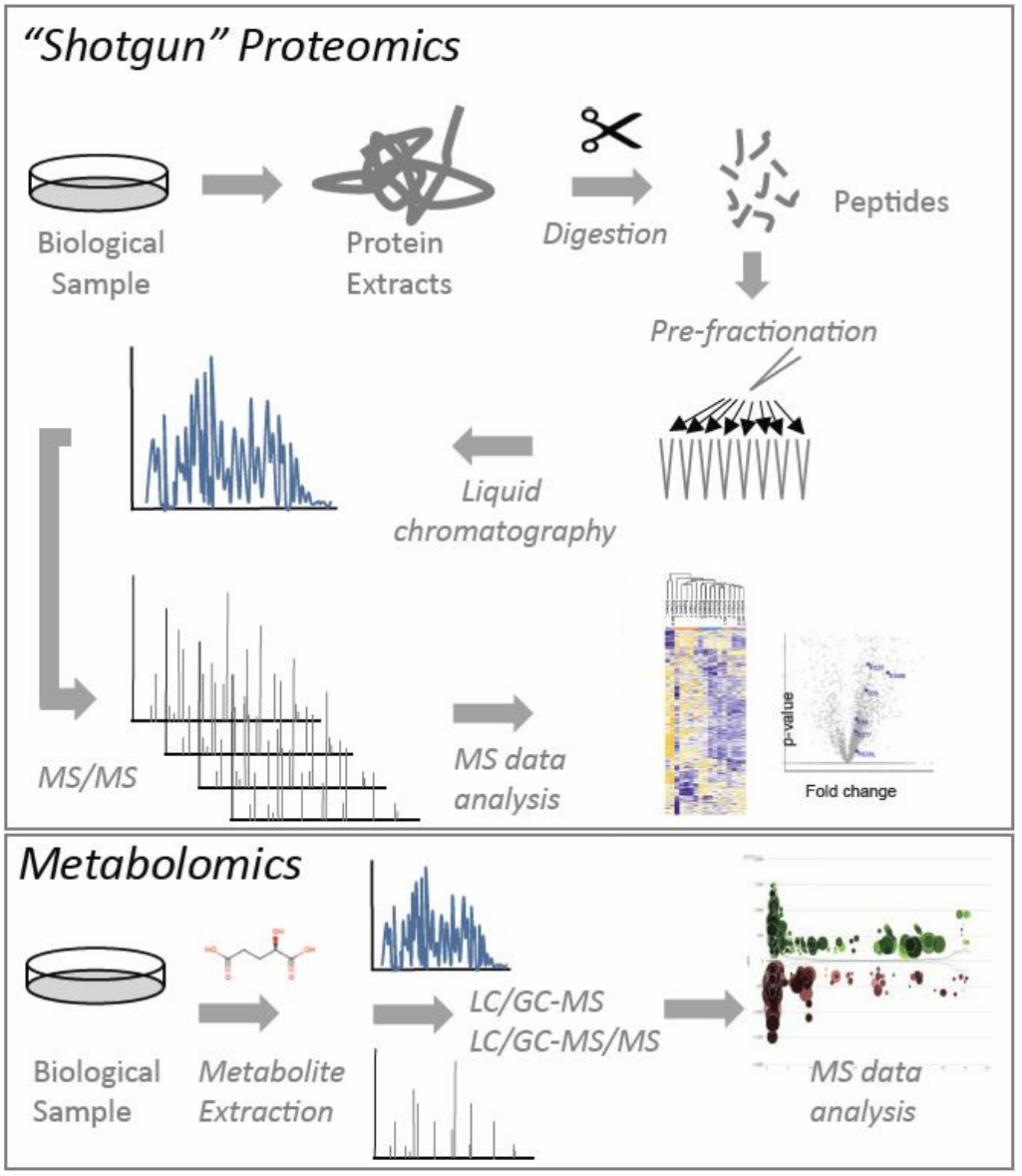


>50,000 phosphorylation sites known
>16,000 acetylation sites known
>6,000 glycosylation sites known
(www.phosida.com)

>17,000 ubiquitylation sites known
(https://gygi.med.harvard.edu/ggbase)

>8,000 proteins known in human plasma
Proteomic inventory & quantity ~2,000 (Farrah *et al*., MCP 2011)

Modern Biological MS



- >10,000 proteins
- >50,000 phosphosites
- >30,000 ubiquitylation sites
- >16,000 acetylation sites
- >400 different protein PTMs
- >1,000 metabolites

Cutillas P & Kessler BM. 2017, OTCB. In press

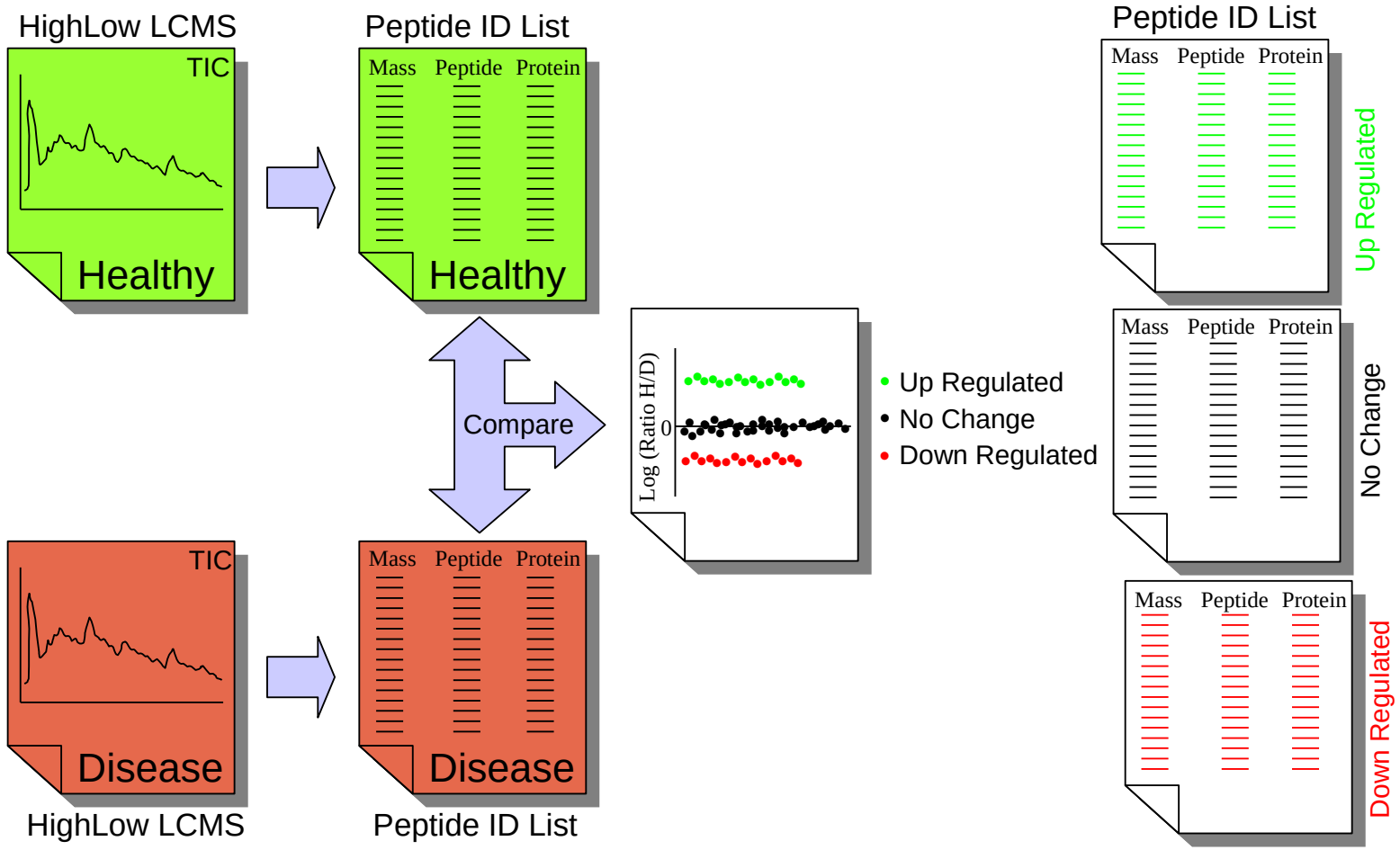
Proteome Depth

Method	Protein amount	Protein groups	Time
‘Ultra depth’ Multi Enzyme, Liquid Chromatography High-pH reverse phase pre-fractionation, less demanding fraction concatenation.	2 – 5 mg $>10^8$ cells	10,000 – 13,000	++++++…
‘High depth’ Liquid Chromatography high-pH reverse phase pre-fractionation, demanding fraction concatenation.	500 µg – 1 mg 20-50x10^7 cells	8,000 – 10,000	++++
‘Medium depth’ Spin column or StageTip based high-pH pre-fractionation. No concatenation	10 µg – 400 µg (spin column) 500,000 to few million cells	4,000 – 8,000	++
‘Single shot’ No pre-fractionation.	< 10 µg Few thousand cells	3,000 – 4,000	+

Protein group numbers: ~20% less for SILAC experiments

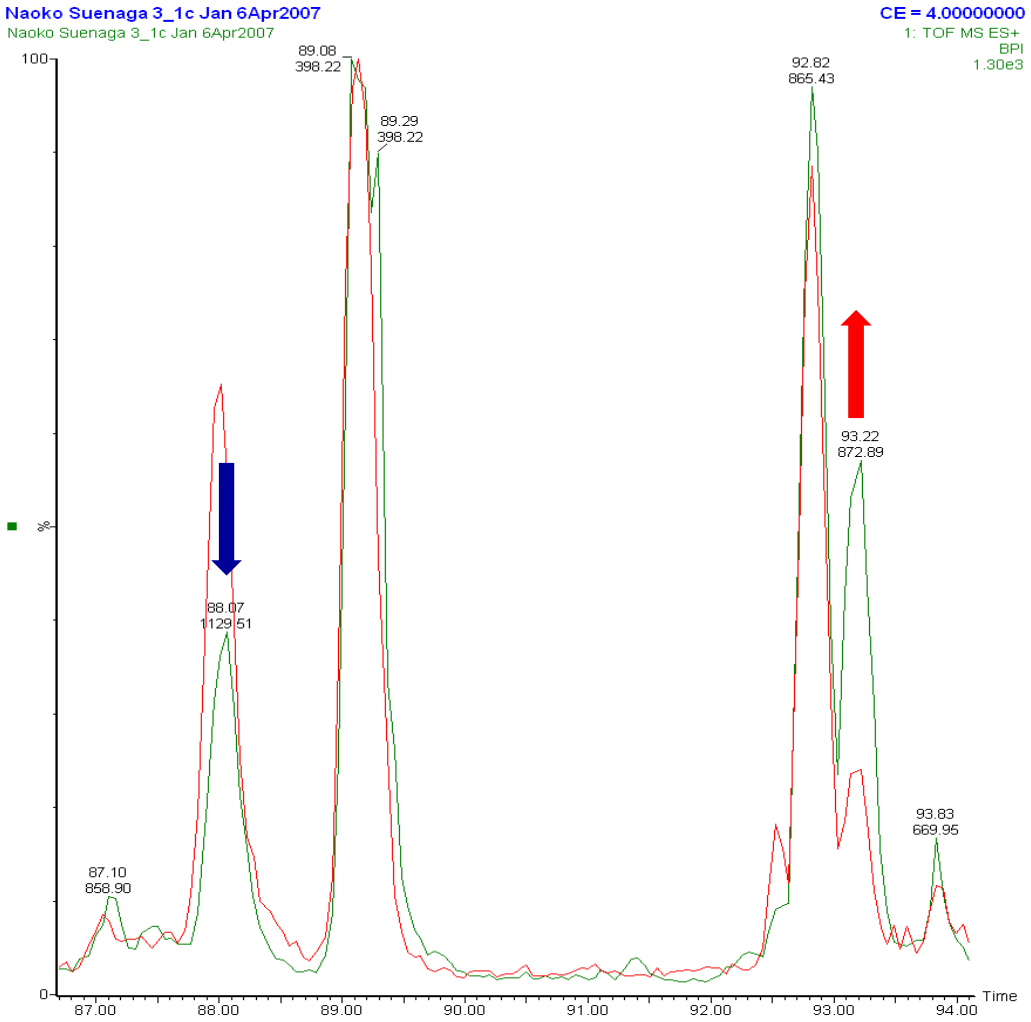
Quantitation of Differentially-Regulated Proteins

LC/MS/MS^E



LC/MS/MS^E

Label-Free Quantitation by MS^E

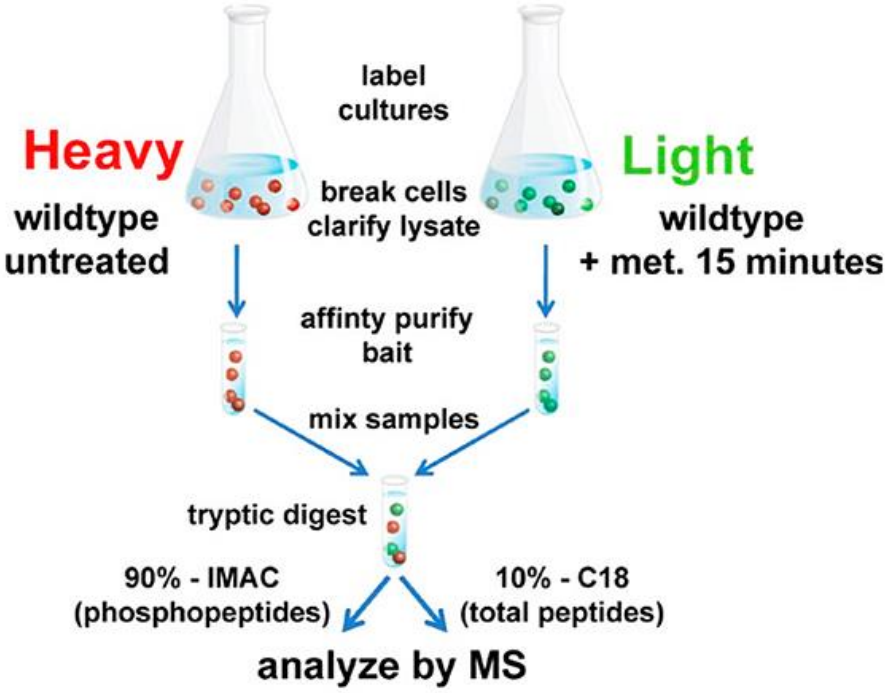


- **MS for quantitative Information (peak Intensities)**
- **MS^E for identification (assignment of precursor and fragment ions)**
- **Experiment performed in triplicate**
- **Each peak defined by: exact Mass + retention Time (EMRT)**

Xu D, Suenaga N, Edelmann MJ, Fridman R, Muschel RJ, Kessler BM. Mol Cell Proteomics 2008 7(11): 2215-2228

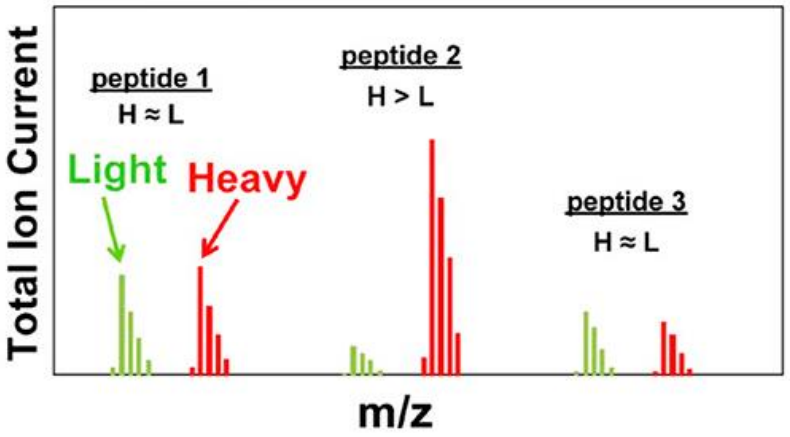
SILAC

Stable isotope labeling of amino acids in cell culture



Amino Acids:

Lys $^{13}C_6$ $^{2}H_2$
Arg $^{13}C_6$ $^{15}N_4$



Other MS Quant Methods:

LFQ
iTRAQ / TMT
SWATH
PRM/MRM
AQUA

UNIVERSITY OF OXFORD

TARGET DISCOVERY INSTITUTE

Biomarker - Definition

- ✓ A characteristic that is objectively measured and evaluated as an indicator of normal biologic or pathogenic processes or pharmacological responses to a therapeutic intervention

Definition of Biomarker

Biomarker: A biochemical feature or facet that can be used to measure the progress of disease or the effects of treatment.

MedicineNet Authored
Webster's New World Medical Dictionary
Learn more »

...tionary

The American Heritage® dic·tion·ar·ies

Library > Words > Dictionary

bi·o·mark·er (bī'ō-mär'kər)

n. Medicine.

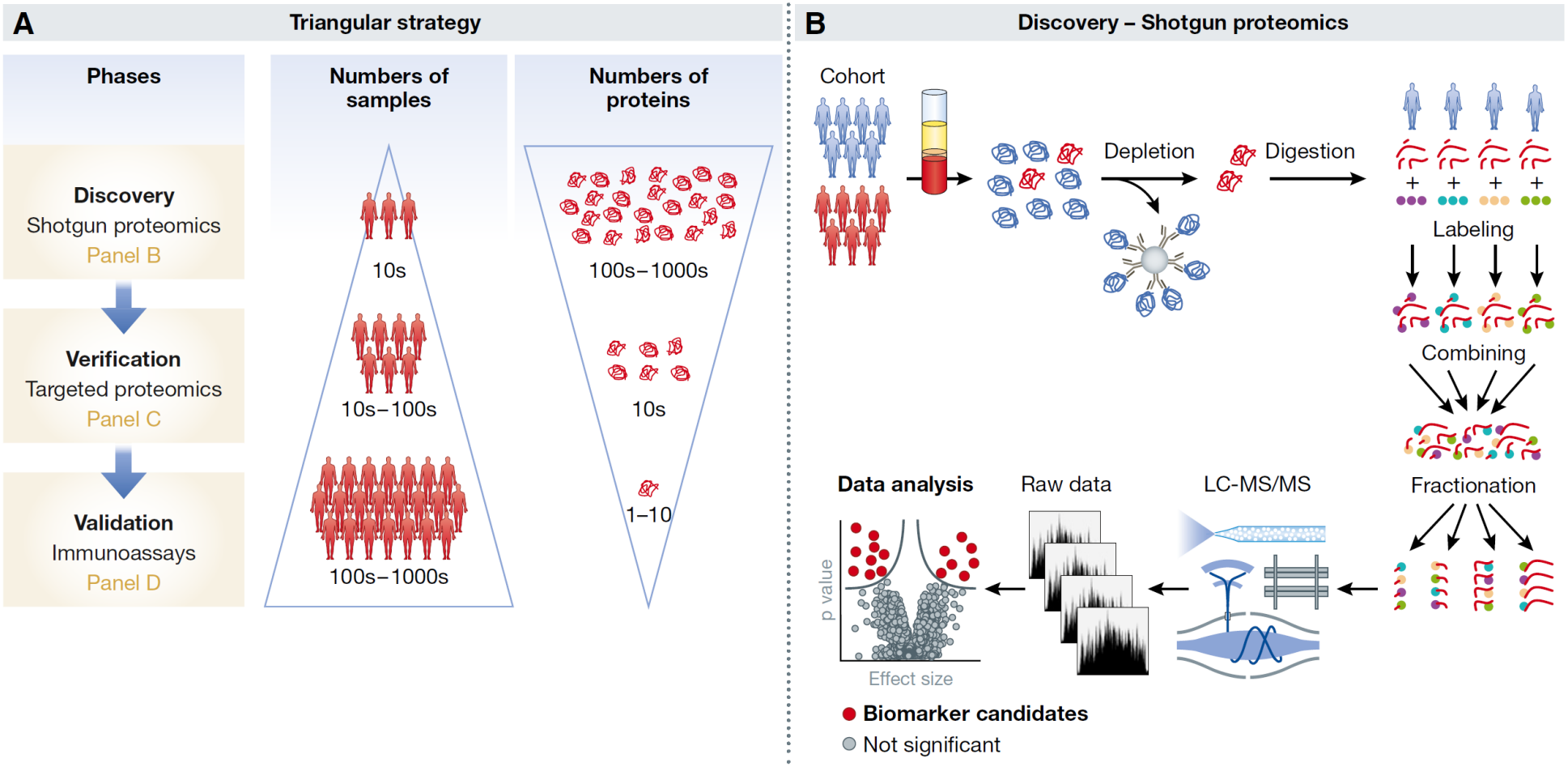
1. See marker (sense 9).
2. A specific physical trait used to measure or indicate the effects or progress of a disease or condition: *Biomarkers of aging include thinning of the hair and diminished elasticity of the skin.*

Examples of Biomarkers

Type of Biomarker	Examples
Biochemical markers	high-sensitivity, C-reactive peptide, plasma B-type natriuetic peptide, plasma homocysteine, urinary albumin excretion, plasma renin activity
Physiologic markers	Blood pressure, heart rate, pulmonary artery pressure, pulmonary capillary wedge pressures, ventricular premature beats
Anatomic measures	Coronary vessel diameter (by coronary angiography), carotid intima media thickness (ultrasound), atherosclerotic plaque burden (intravascular ultrasound)
Histologic markers	Tissue biopsy specimens
Physical markers	Skin color, weight, height

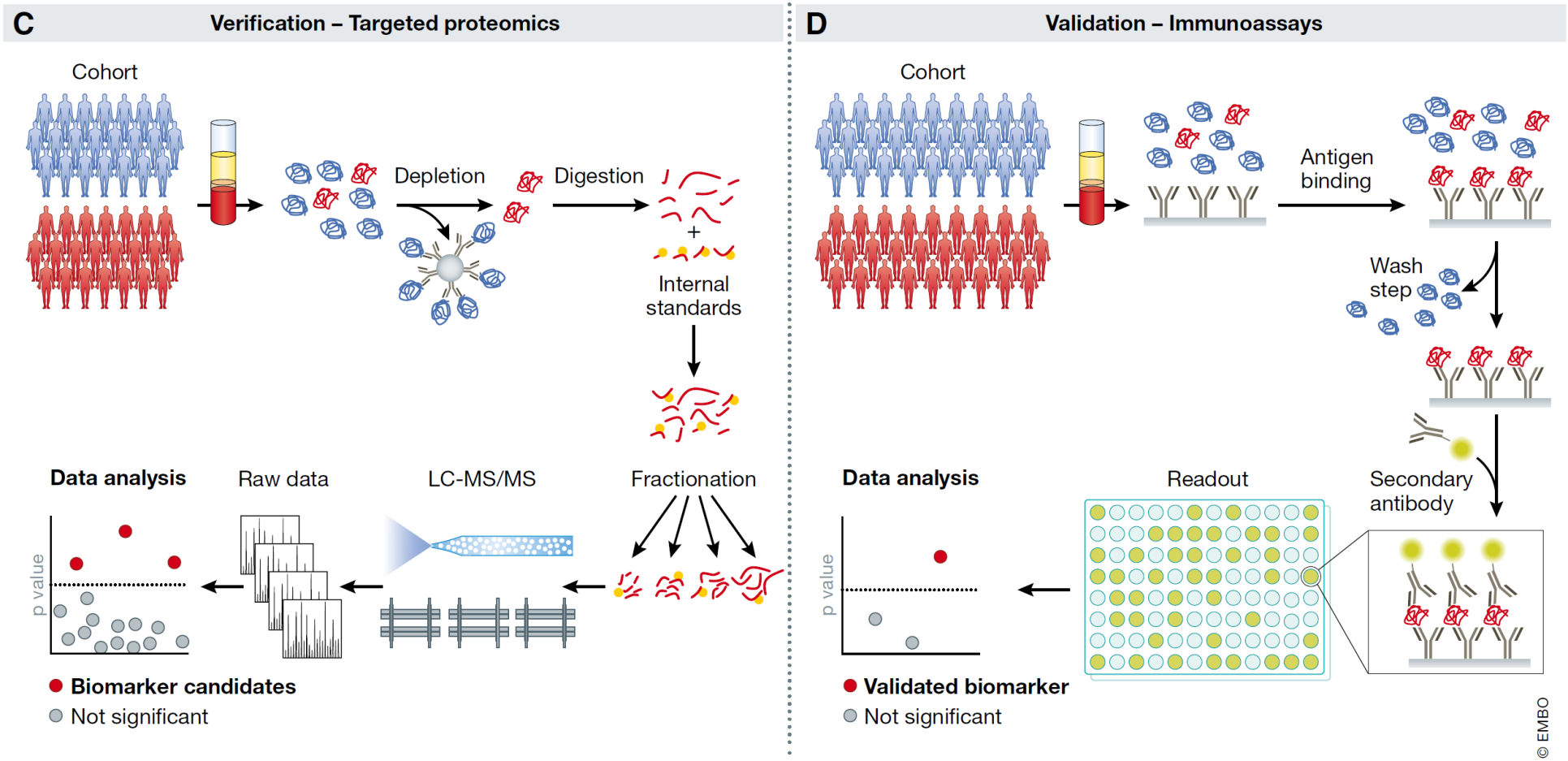
Mehul Desai, Norman Stockbridge, Robert Temple. The AAPS Journal 2006; 8 (1) Article 17

Classical Clinical Proteomics
(Triangular Approach)



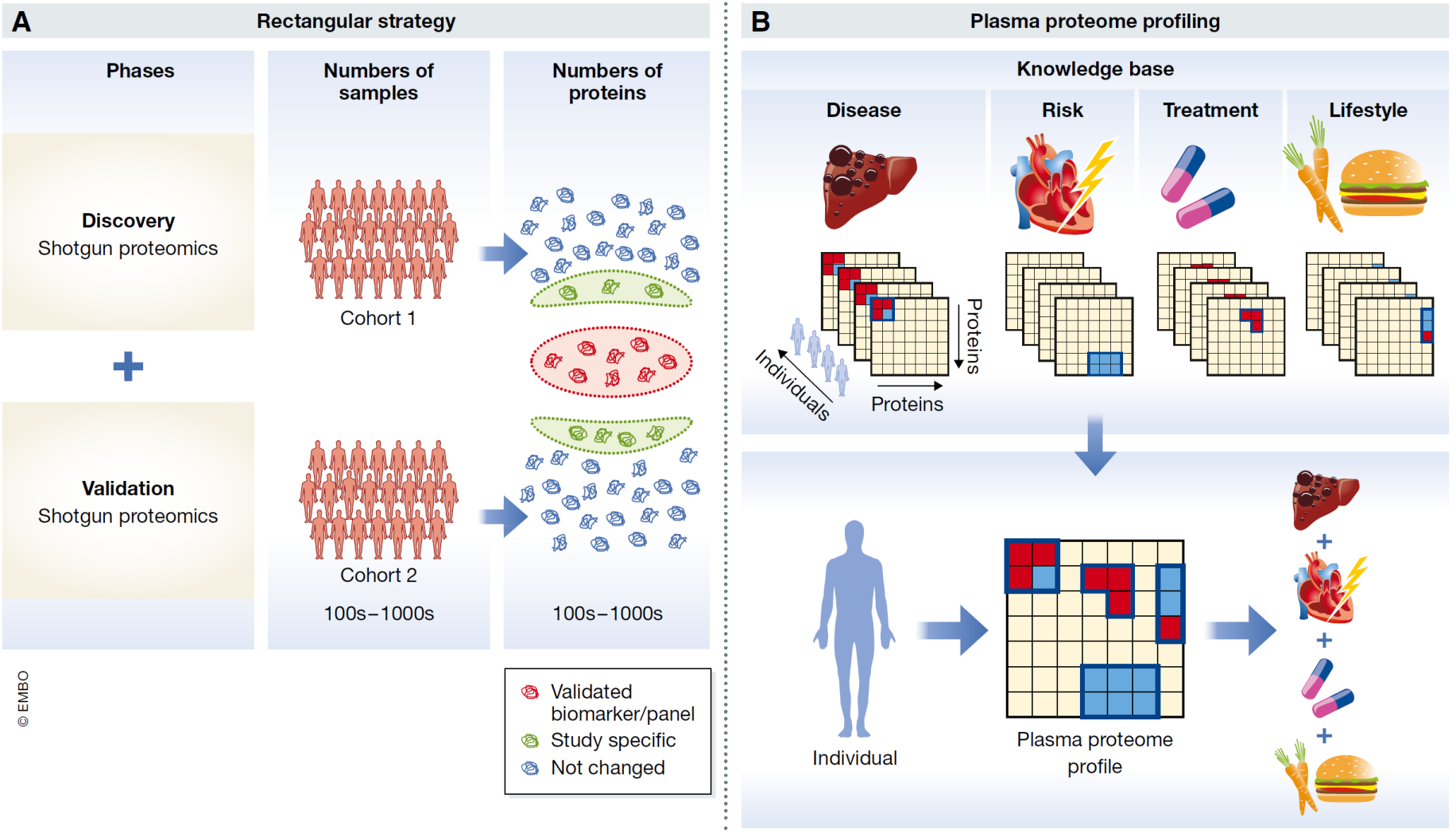
Geyer et al., Mol Sys Biol 2017

Classical Clinical Proteomics
(Triangular Approach)



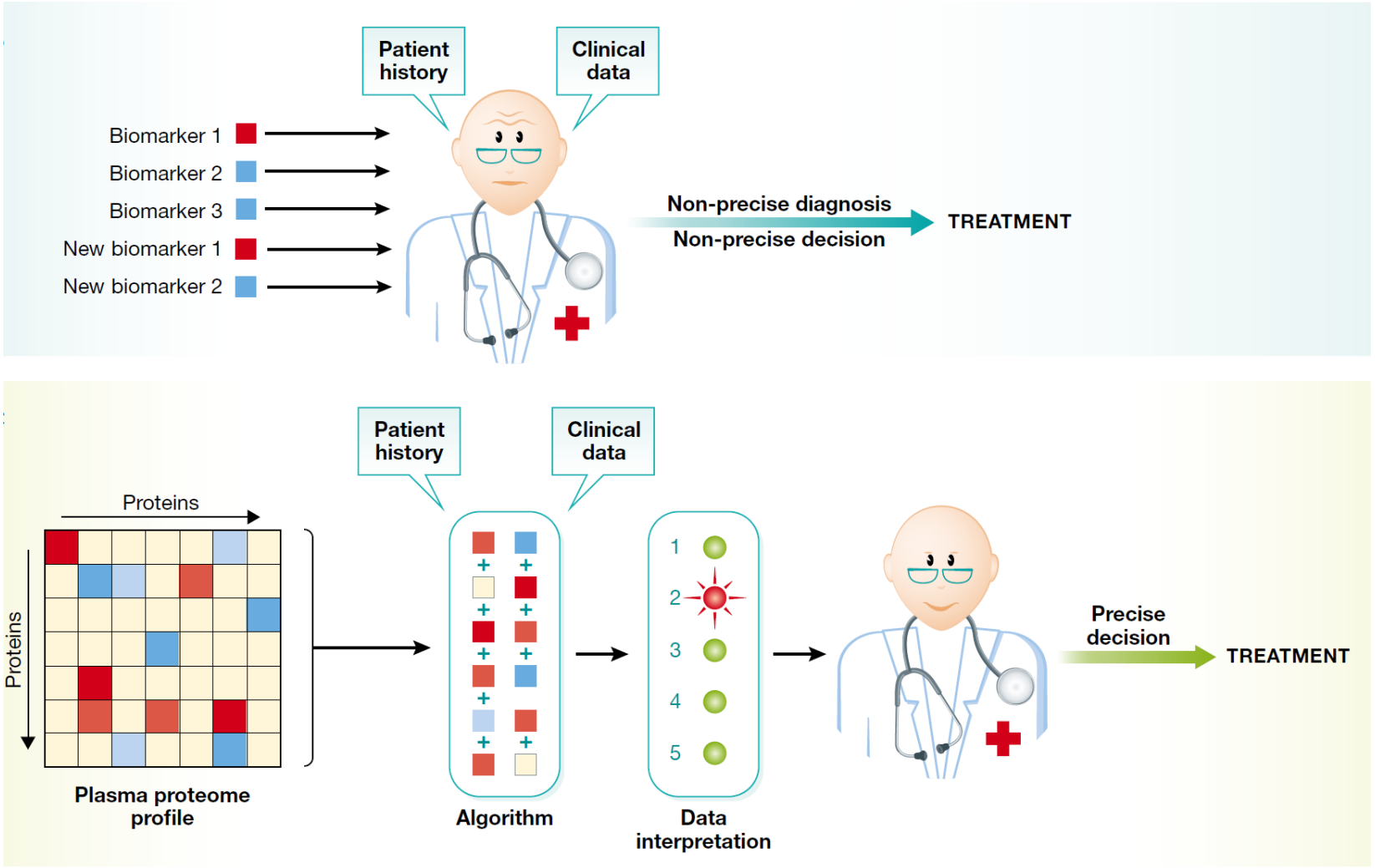
Geyer et al., Mol Sys Biol 2017

Advanced Clinical Proteomics
(Rectangular Approach)



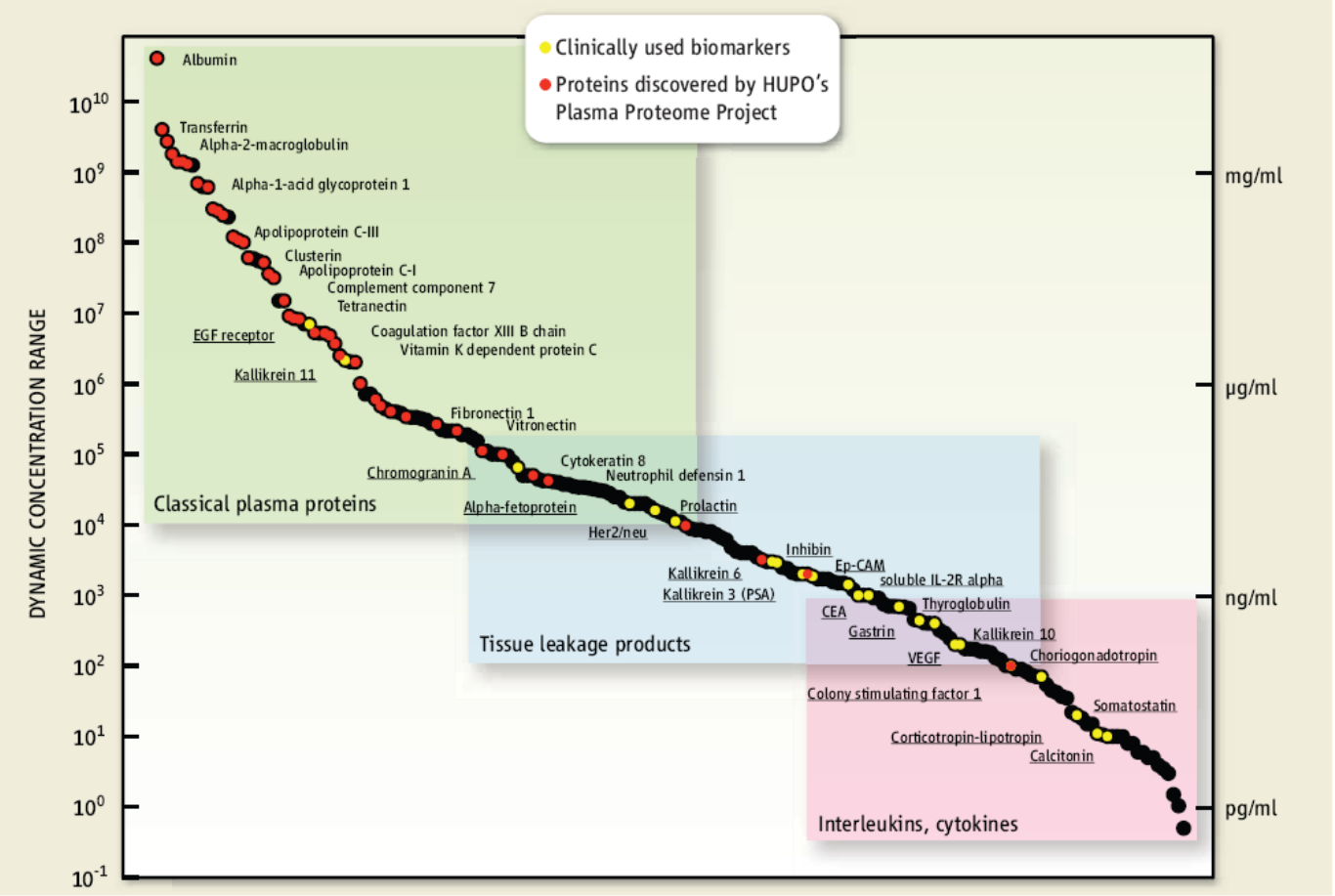
Geyer et al., Mol Sys Biol 2017

More Accurate Prediction of Diagnosis / Treatment



Geyer et al., Mol Sys Biol 2017

Current Limitations:
Abundance of Candidate Biomarkers



Needles in a haystack. The abundance of different proteins in blood varies by more than 10 orders of magnitude. Most commercially used biomarkers (yellow dots) are present in only minute quantities in blood, below the level at which most proteins are detected (red dots).

Modified from Science 2008, Vol 321, 1758-1761

Single Cell Proteomics – CyTOF

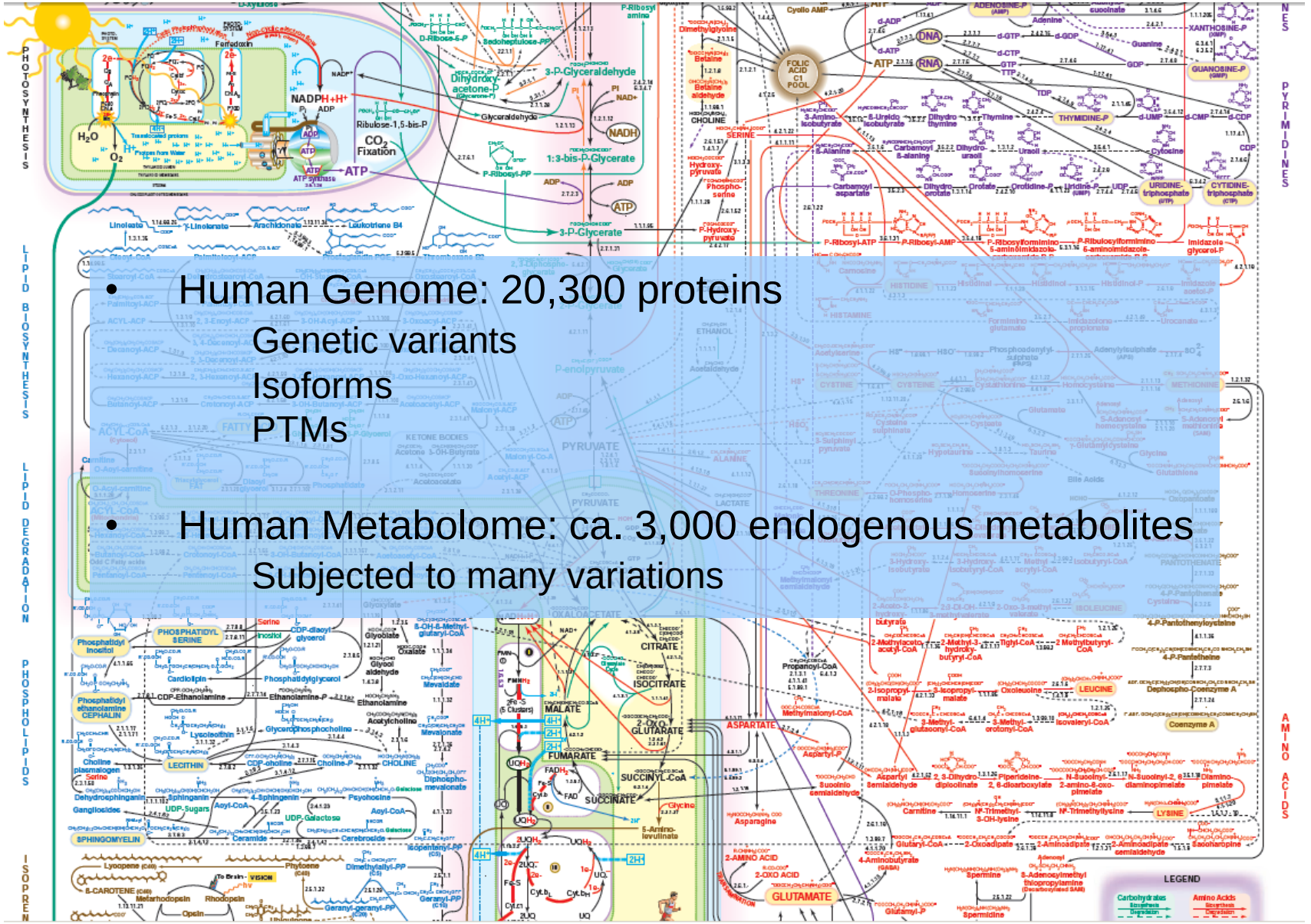
- Not discovery, but targeted proteomics!
- Available at the WIMM

Sean C et al., Nature Biotechnology 2012

Complexity of Metabolic Pathways and Cellular Processes

UNIVERSITY OF OXFORD

TDI TARGET DISCOVERY INSTITUTE

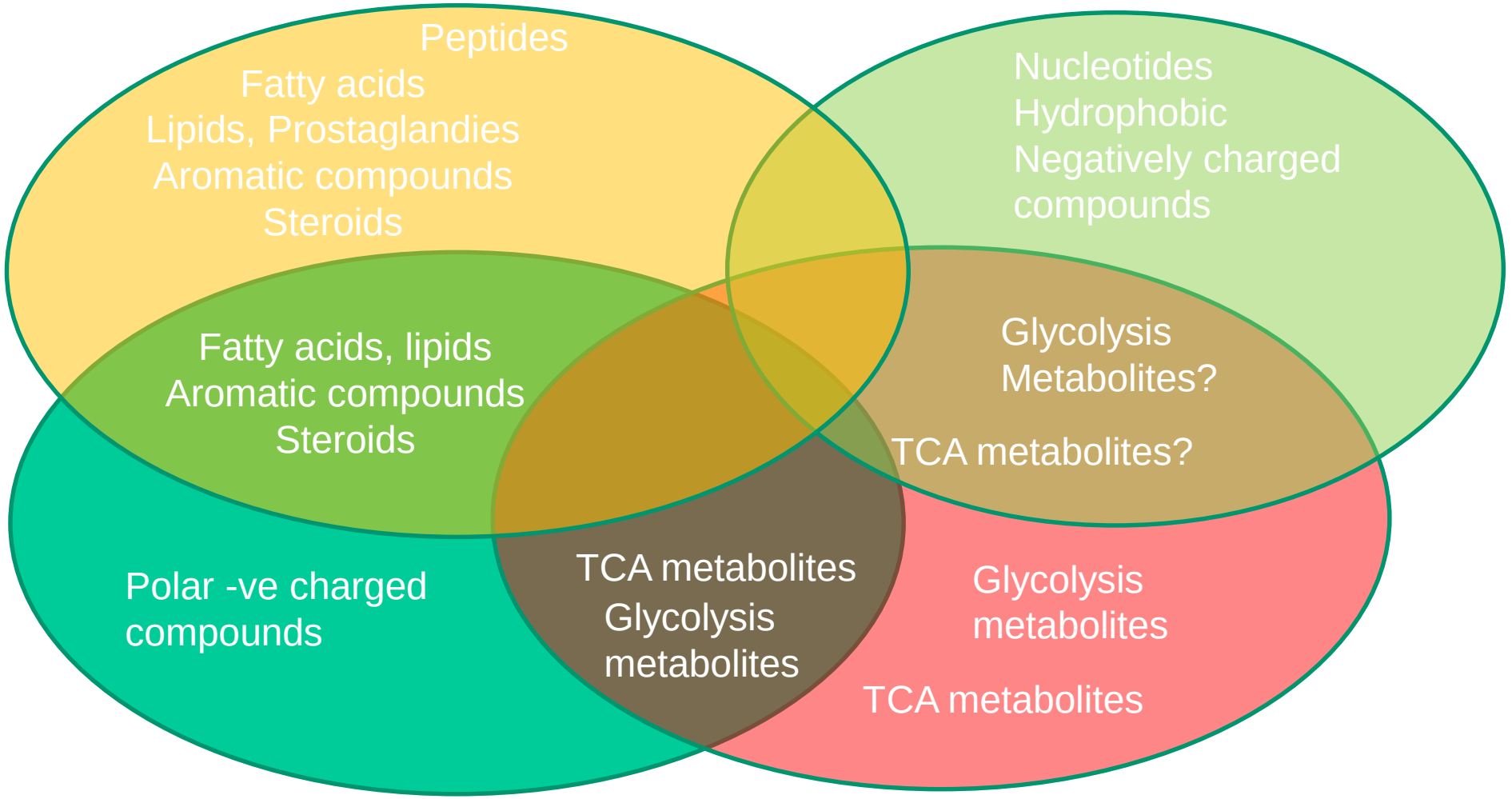


- Human Genome: 20,300 proteins
  - Genetic variants
  - Isoforms
  - PTMs
- Human Metabolome: ca. 3,000 endogenous metabolites
  - Subjected to many variations

LC-MS Strategies for Metabolomics

C18 Reversed Phase — positive mode
C18 Reversed Phase — negative mode
C18 Ion pairing (pH 4.95) — negative mode
HILIC Luna NH2 — negative mode

- ✓ Discovery metabolomics *(AMRT, MS/MS)*
- ✓ Targeted metabolomics *(MRM/SRM)*

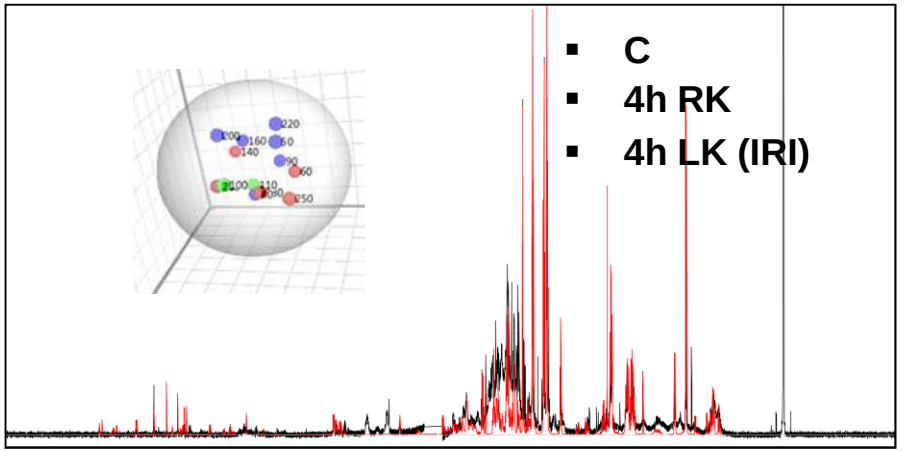


2-Dimensional Gas Chromatography

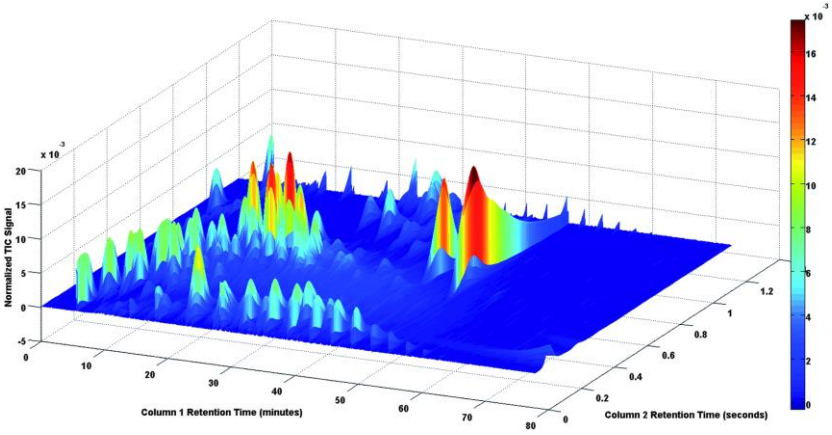
GCxGC-qMS

Detection of >600 Molecular Features ~200 Targeted Metabolites

NMR



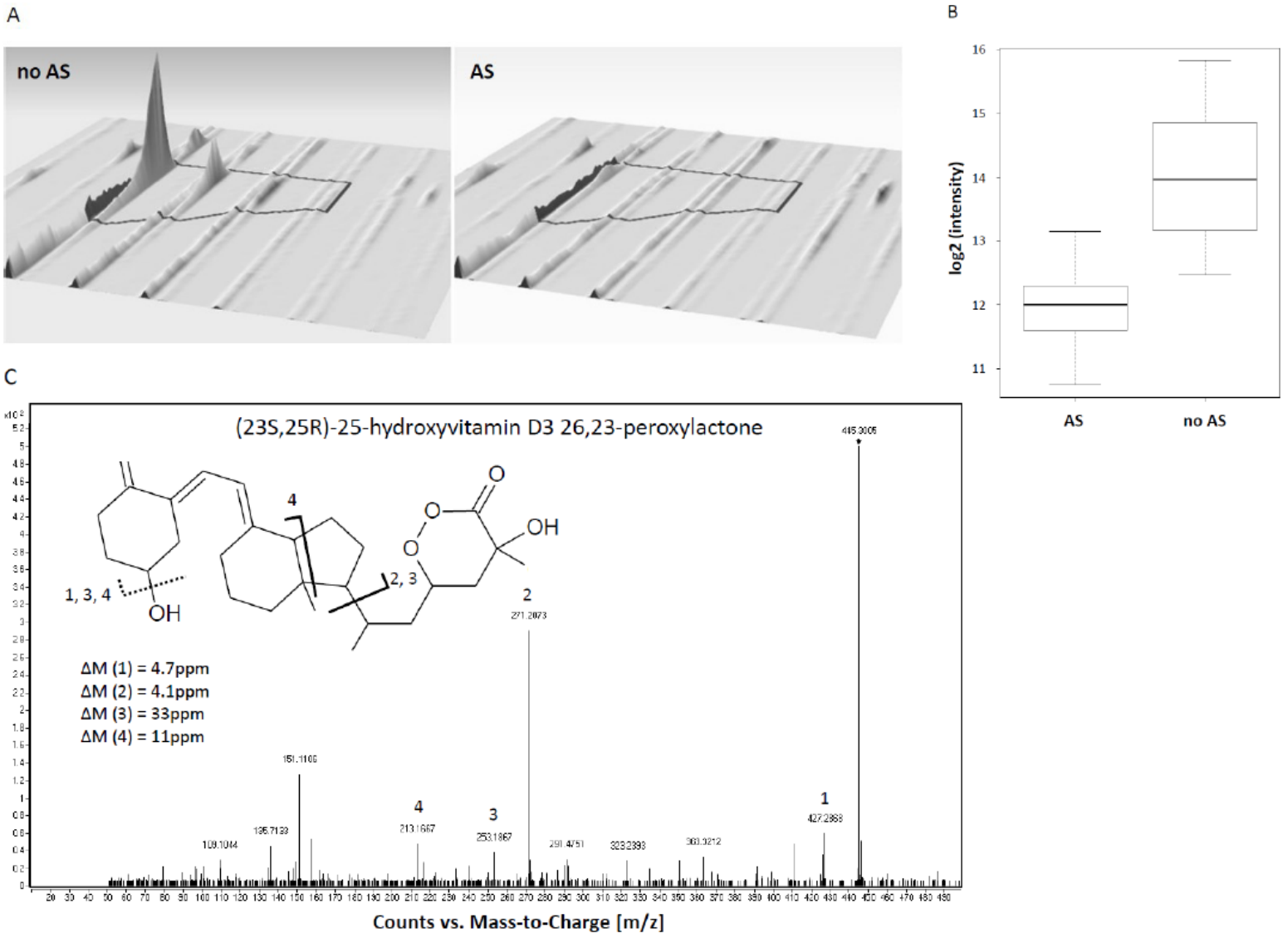
GCxGC-MS



Zhanru Yu et al., TALANTA 2017

Metabolite Profiling & Identification

VERY DIFFICULT !!



Fischer et al., Mol. Cell. Proteomics, 2012

Further Reading

Aebersold AR, Goodlett DR
Mass spectrometry in proteomics ; Chem Rev. 2001 Feb;101(2):269-95. Review.

* Steen H, Mann M.
The ABC's (and XYZ's) of peptide sequencing; Nat Rev Mol Cell Biol. 2004 Sep;5(9):699-711. Review.

Low TY, Heck AJ Reconciling proteomics with next generation sequencing**.**
Curr Opin Chem Biol. 2015 Nov 16;30:14-20.

Stable isotope labeling by amino acids in cell culture for quantitative proteomics.
Ong SE, Mann M.
Methods Mol Biol. 2007;359:37-52.

* Cravatt BF, Simon GM, Yates JR 3rd
The biological impact of mass-spectrometry-based proteomics.
Nature 2007; Dec 13; 450 (7172):991-1000. Review

* Generating and navigating proteome maps using mass spectrometry.
Ahrens CH, Brunner E, Qeli E, Basler K, Aebersold R.
Nat Rev Mol Cell Biol. 2010 Nov;11(11):789-801. Review

Please feel free to get in contact to discuss questions/experiments:

Email: darragh.obrien@ndm.ox.ac.uk
http://www.tdi.ox.ac.uk/mass-spectrometry