



An agency of the
Provincial Health Services Authority



THE UNIVERSITY
OF BRITISH COLUMBIA

Proteomics in Cancer Research and Precision Medicine

STAT/BIOF/GSAT 540
Statistical Methods for High Dimensional Biology

Dr. Philipp Lange

Canada Research Chair in Translational Proteomics of Pediatric Malignancies

Pathology - University of British Columbia

Michael Cuccione Childhood Cancer Research Program - BC Children's Hospital

Molecular Oncology – BC Cancer

Disclaimer

Images and tables contained within this presentation were obtained from the print or online resources or “inherited” from previous academic years and may be subject to copyright.

Their inclusion herein is for educational/illustrative purposes under the fair use policy only and does not imply ownership and/or authorship by the presenter.

**Includes slides and modified work kindly provided by
Drs Morin, Maurer, Keller, Mohtaram & Huesgen.**

About me

Philipp Lange, PhD

Canada Research Chair in Translational
Proteogenomics of Childhood Malignancies

Research Interests:

Molecular pathology of childhood cancer
Precision oncology
Posttranslational protein modification



Contact:

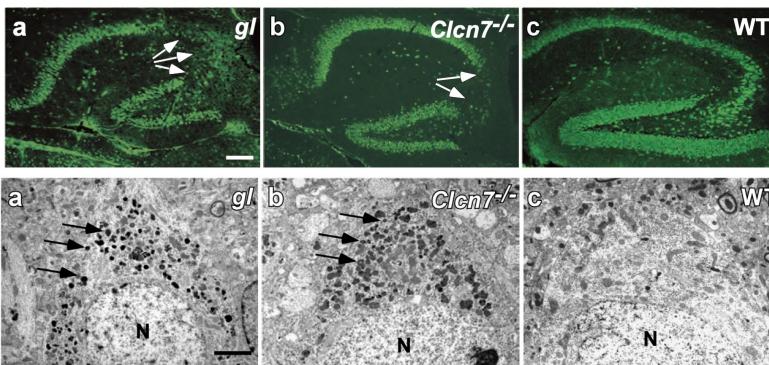
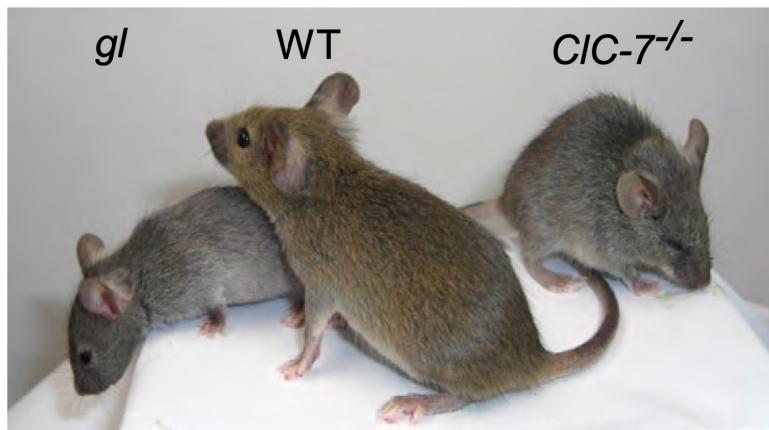
philipp.lange@ubc.ca



What got me here

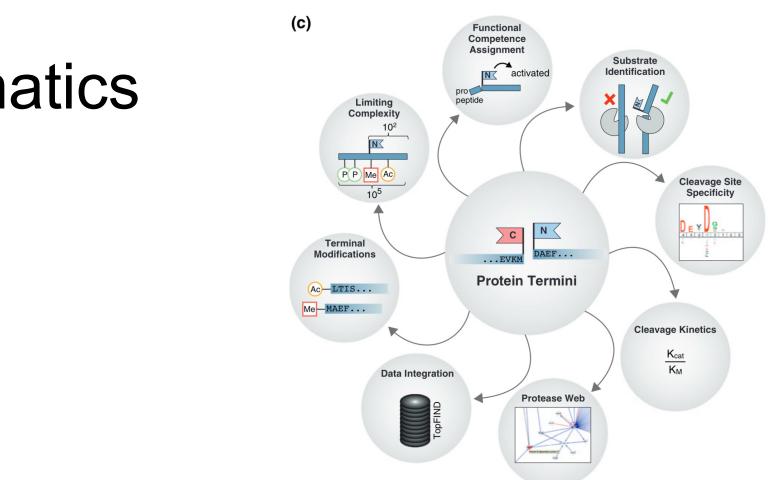
- PhD in Biochemistry

- Autosomal dominant osteopetrosis in kids
- Osteoporosis in adults
- Lots of basic cell biology, biochemistry, biophysics, microscopy and histology



- PostDoc in Proteomics & Bioinformatics

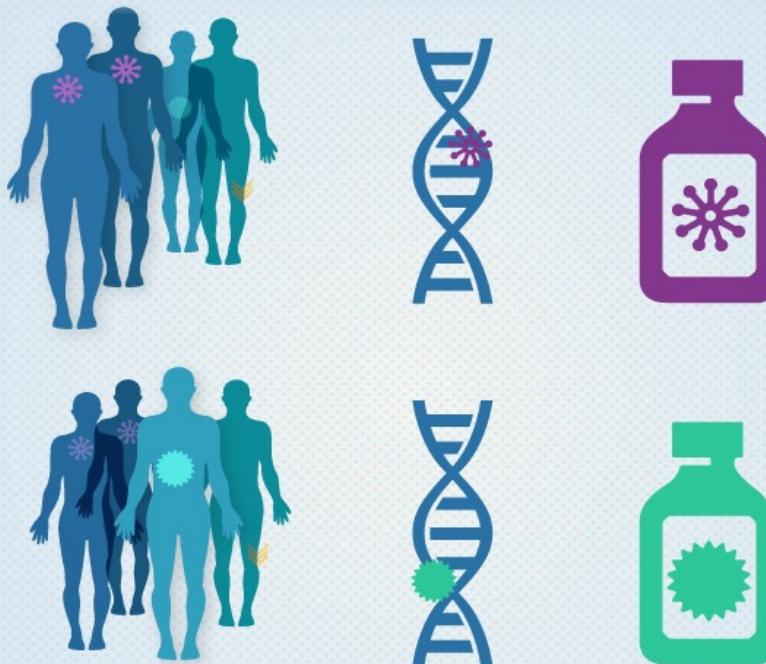
- Breast cancer models
- Post translational modification
- Protein function and stability



Lots of privilege

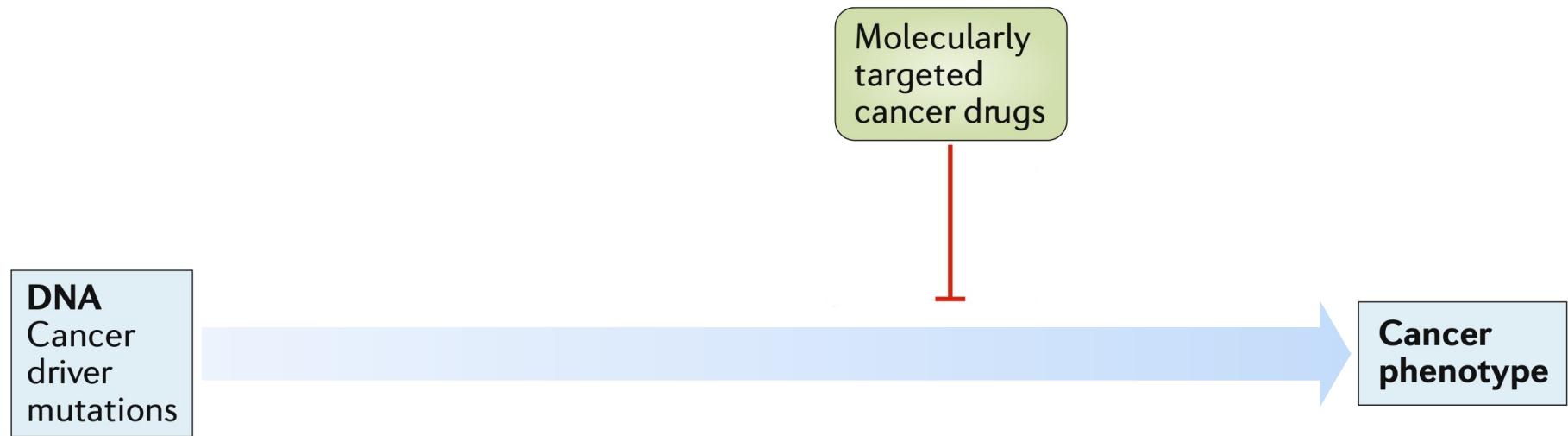
- 3rd generation academic – education was always valued
- White cis-gender male
- Grew up in a system where education was free
- ...

Precision medicine: genome variant – drug paring



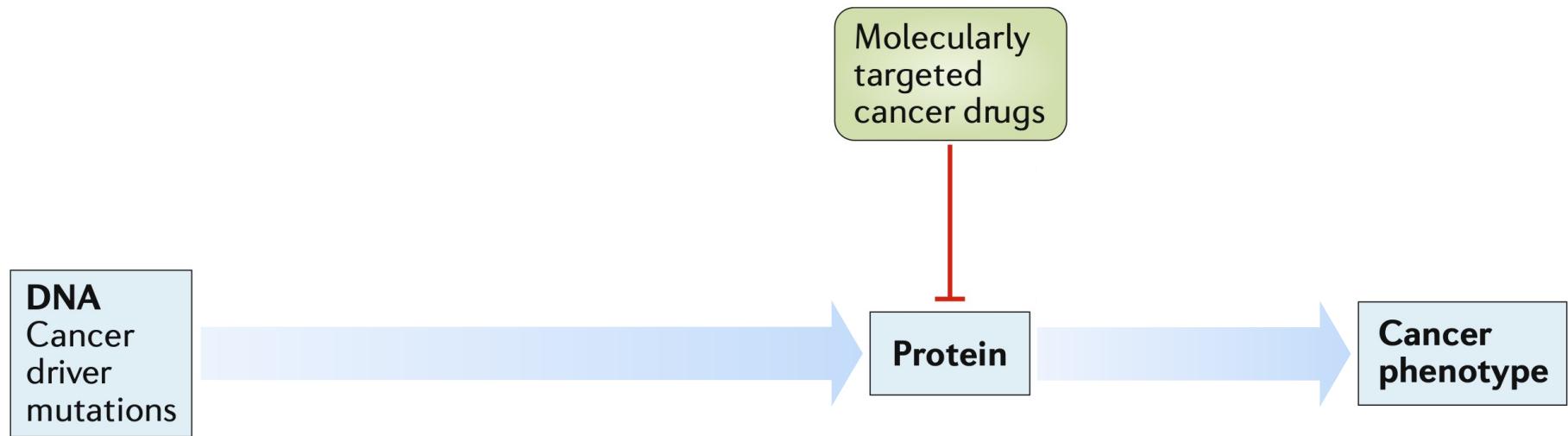
adapted from cancer.gov

Precision medicine: genome variant – drug paring



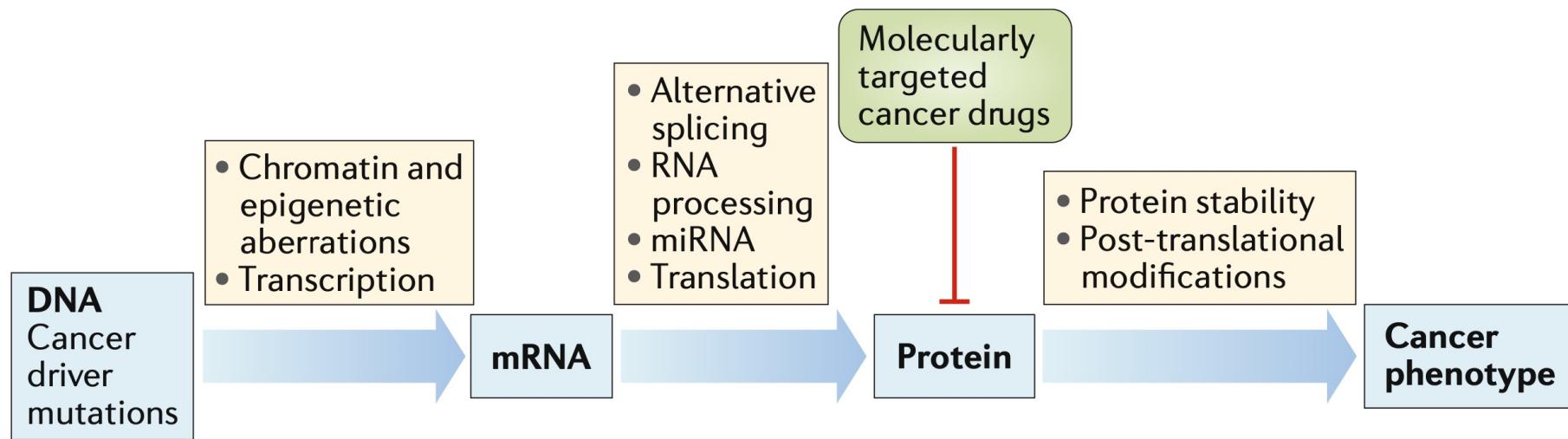
Zang et al (2019), *Nature Clinical Oncology*

Many processes downstream of the genome can affect the tumour phenotype



Zang et al (2019), *Nature Clinical Oncology*

Many processes downstream of the genome can affect the tumour phenotype

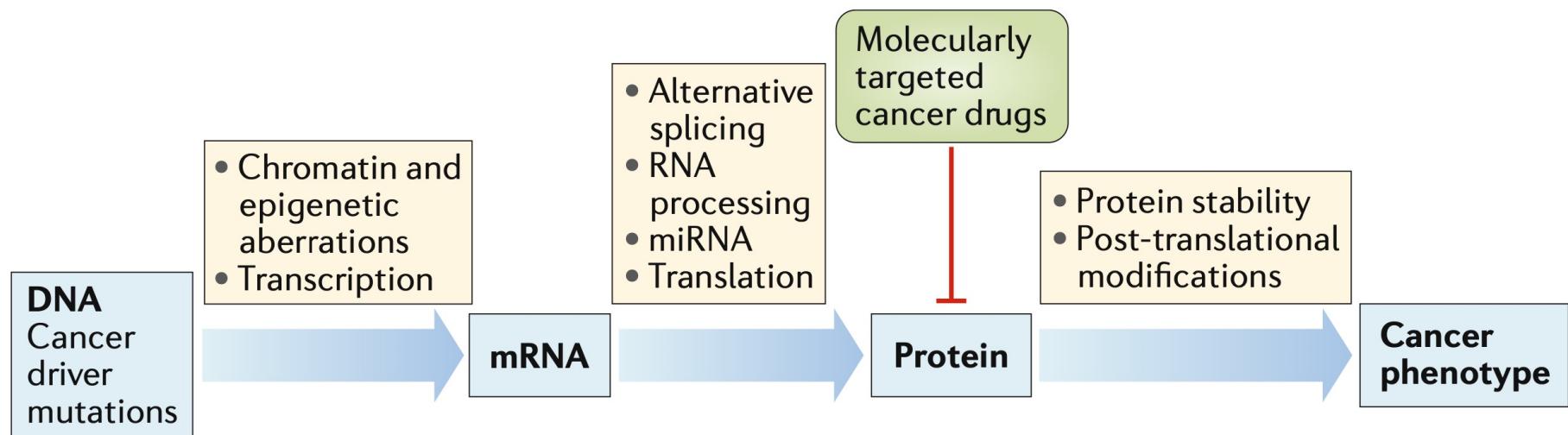


Zang et al (2019), *Nature Clinical Oncology*

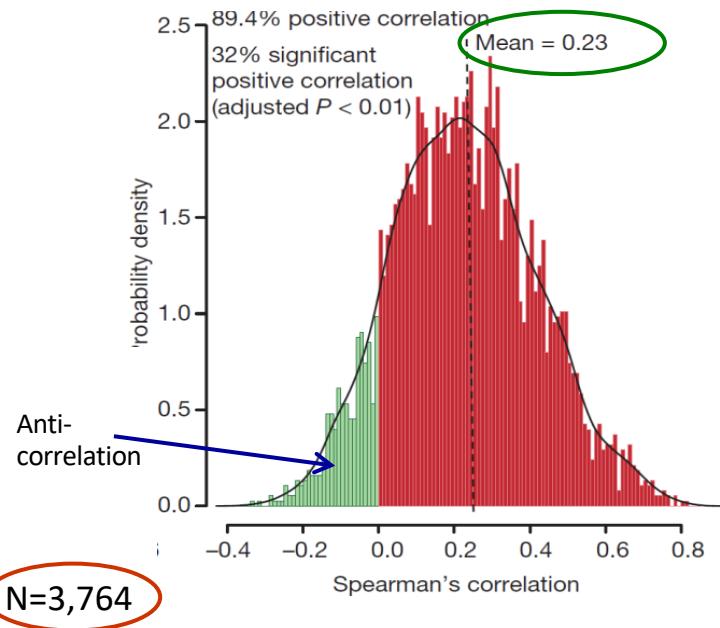
Learning objectives

- Relationship of genome, transcriptome, proteome and metabolome
- Properties of data describing genome, transcriptome, proteome and metabolome
- Understand why and when to study more than one
- Describe ways to study the proteome
- Basic working principle of a mass spectrometer
- Application of proteomics in the context of precision medicine and complex biological systems

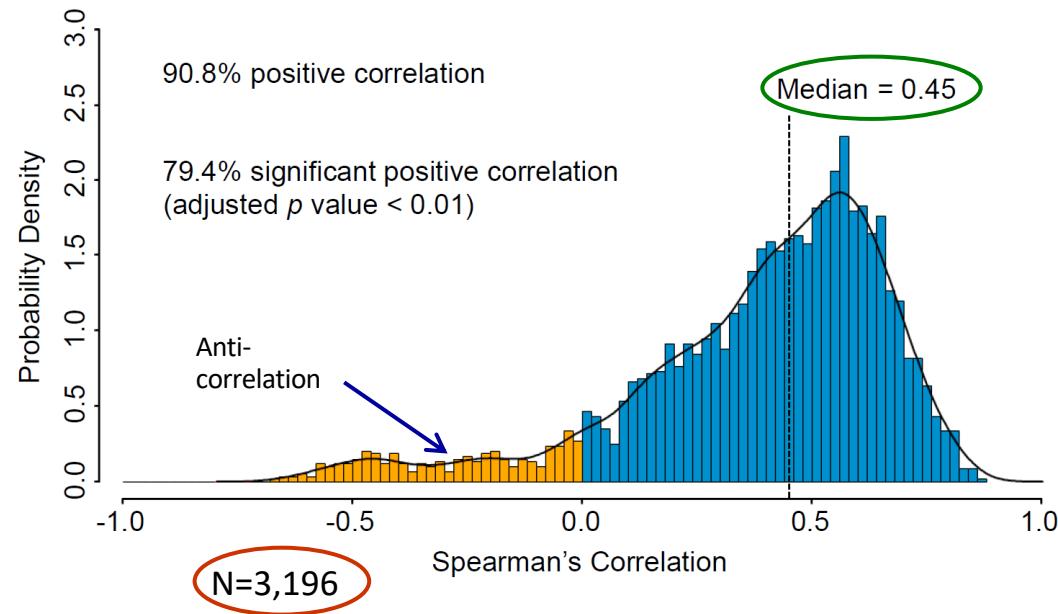
Do DNA or RNA measurements accurately describe protein abundance?



Modest mRNA-protein correlation



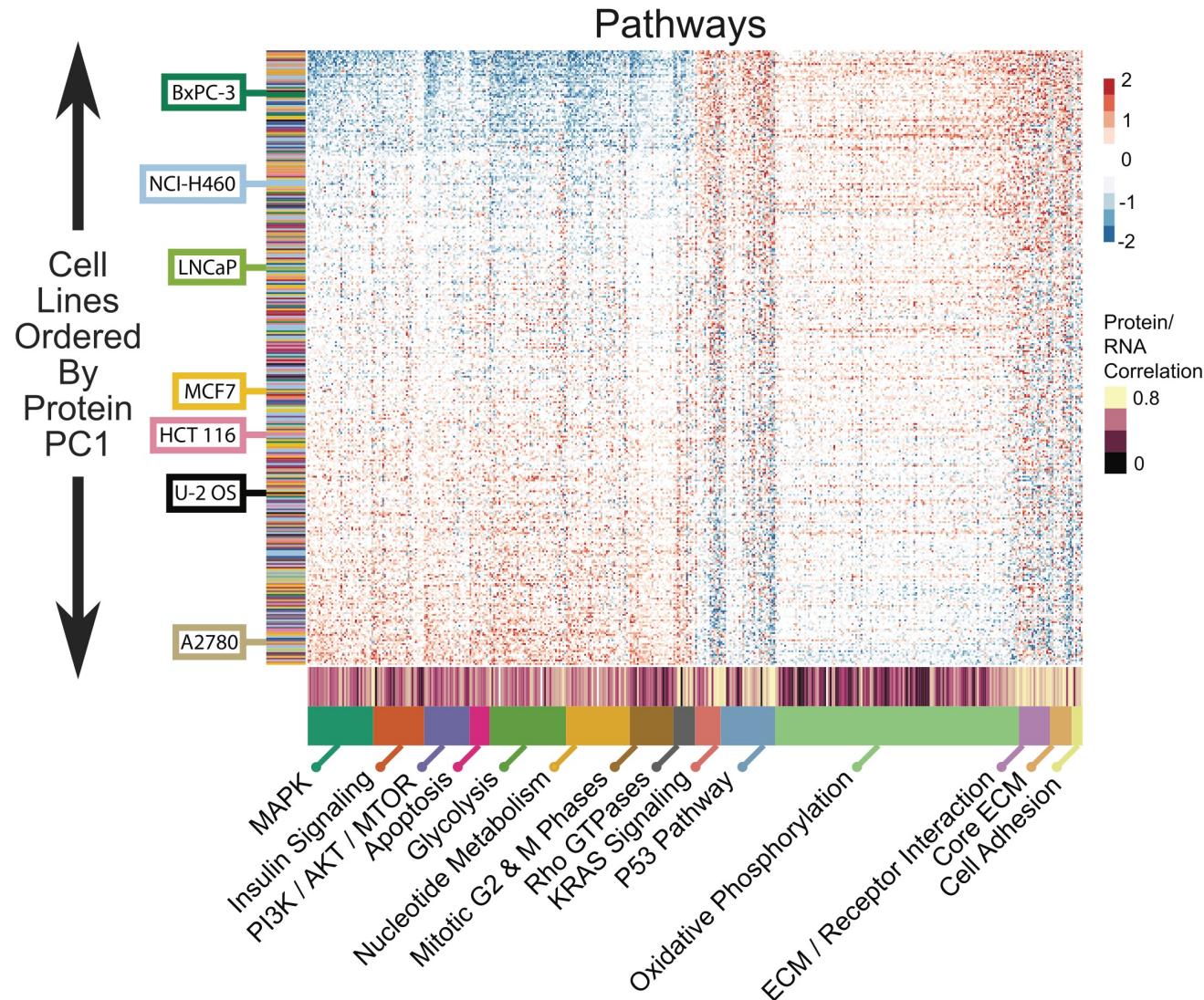
Colorectal Cancer – CPTAC
Zhang, B. et al. (2014) *Nature*



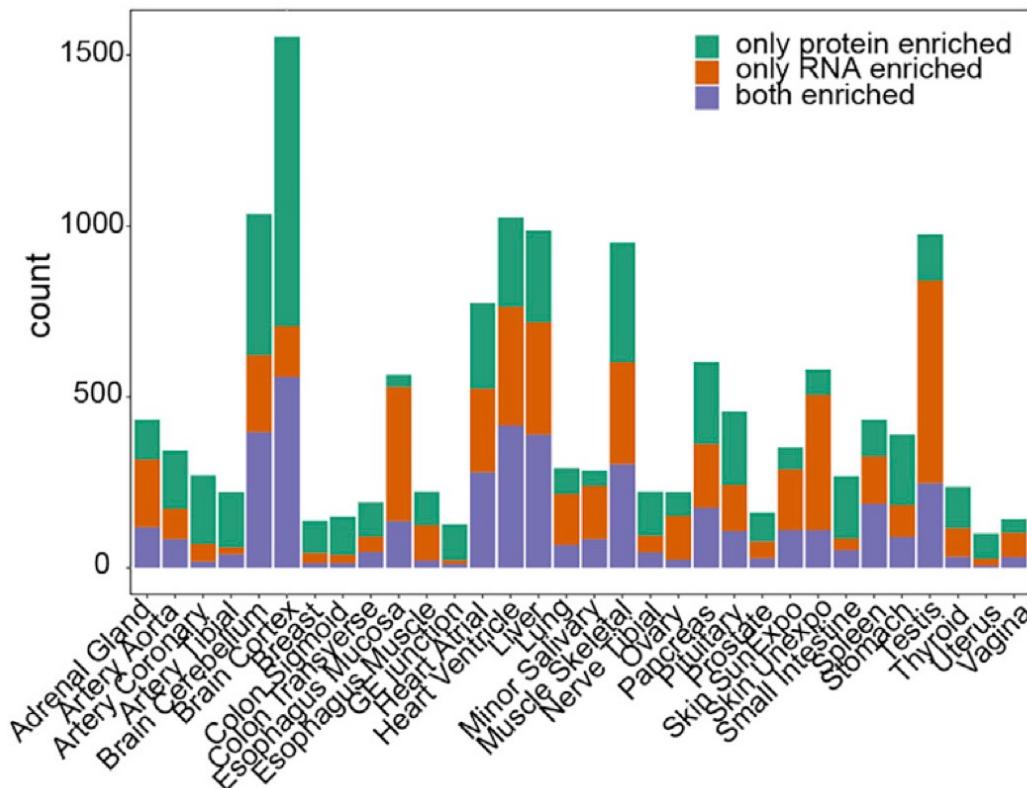
HGS Ovarian Cancer – CPTAC
Zhang, H. et al. (2016) *Cell*

CPTAC = Clinical Proteomics Tumor Analysis Consortium

mRNA-protein correlation is not uniform

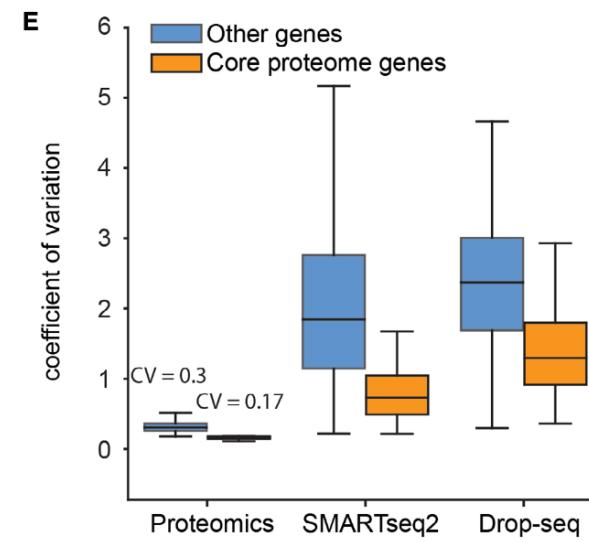
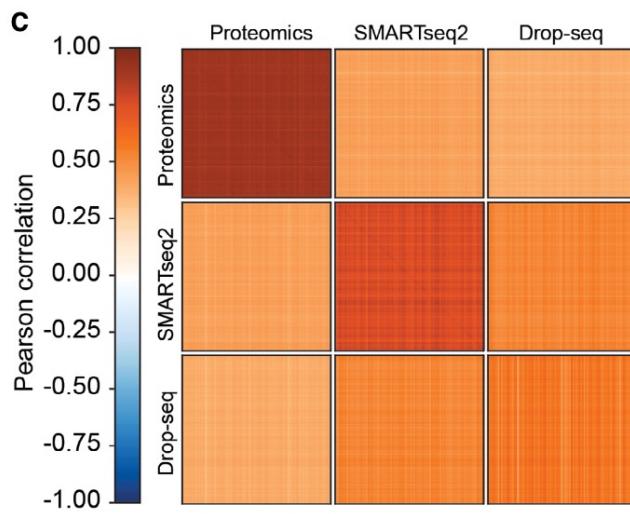


mRNA & protein show different tissue enrichment



Jiang L et al. Cell 183, 269–283, October 1, 2020

Single cells have more similar proteomes than transcriptomes

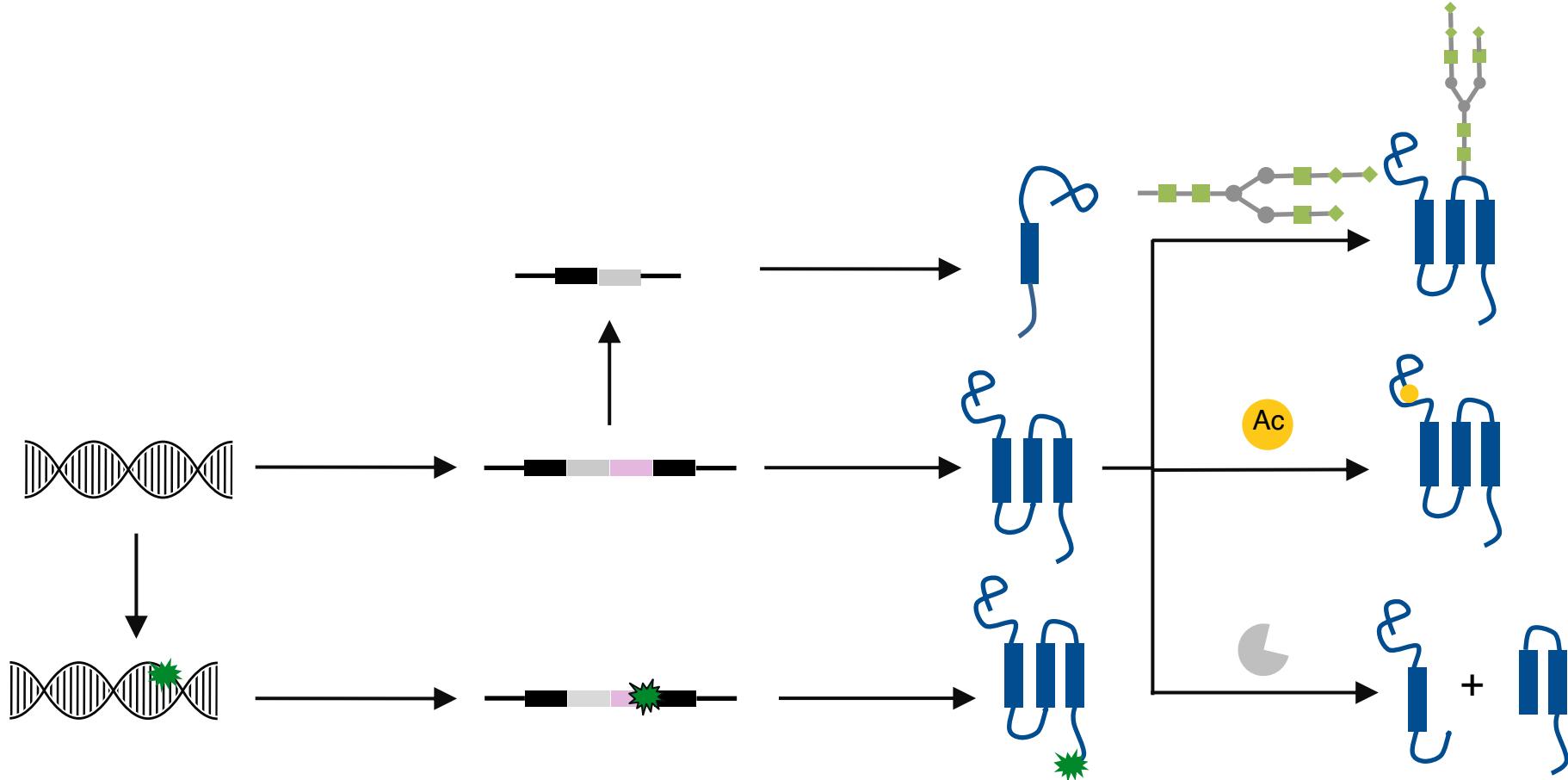


How many different
genes are in a human
cell?

How many different
proteins are in a
human cell?

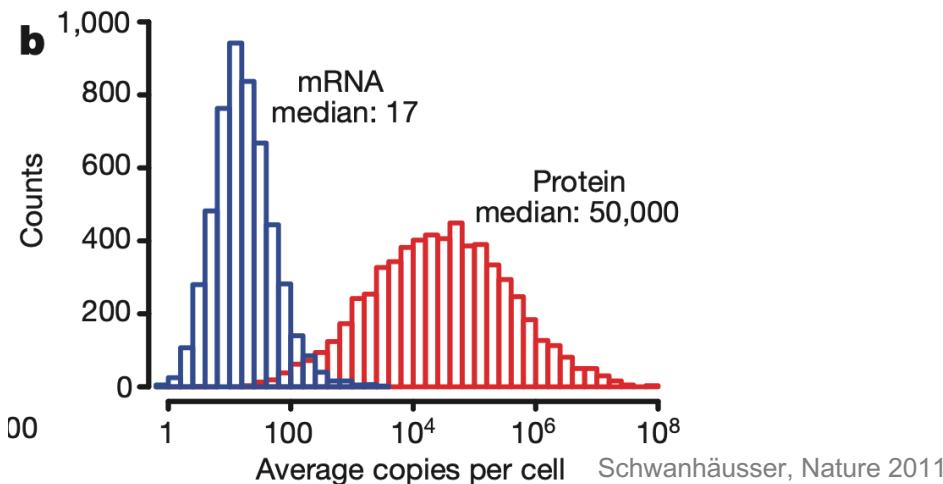
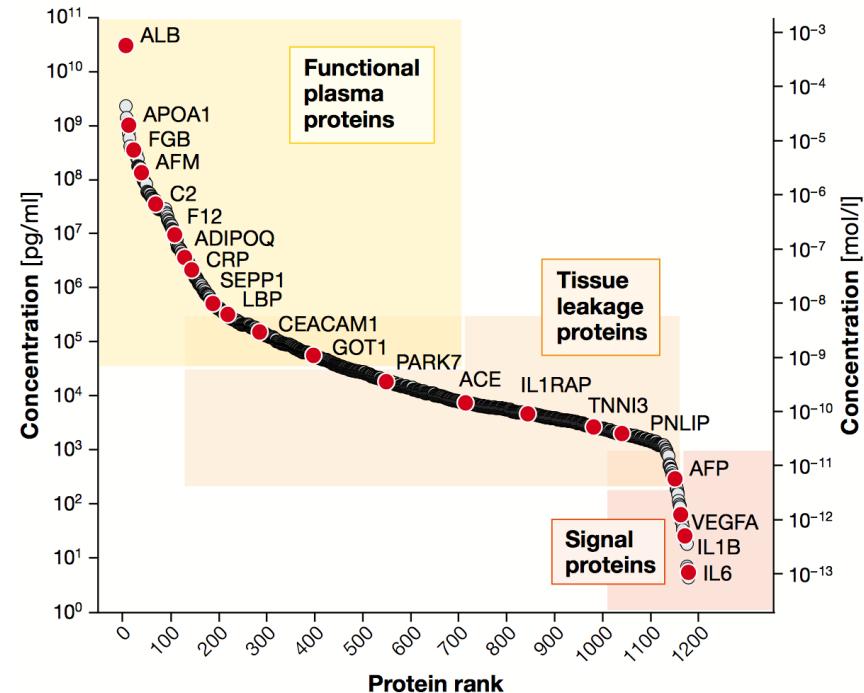
Genome encoded proteins mature to many proteoforms

all of the different molecular forms in which a single gene product exists



The proteome is complex and dynamic

- Concentration spans >10 orders
- Proteoforms
- Localization
- Interaction
- Activation status
- Temporal dynamics



Key features / differences with analytical relevance

	Genome	Proteome		Metabolome
		proteins	proteoforms	
Size				
Abundance range				
Abundance dynamic				
Structural stability				
Chemical composition				

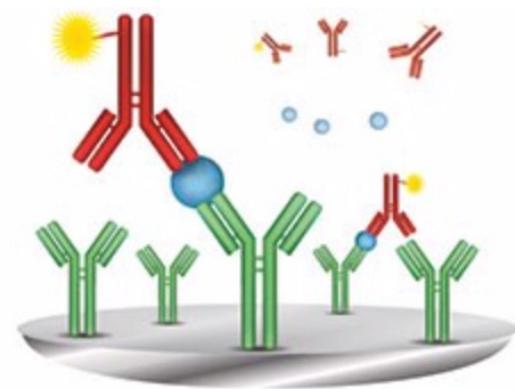
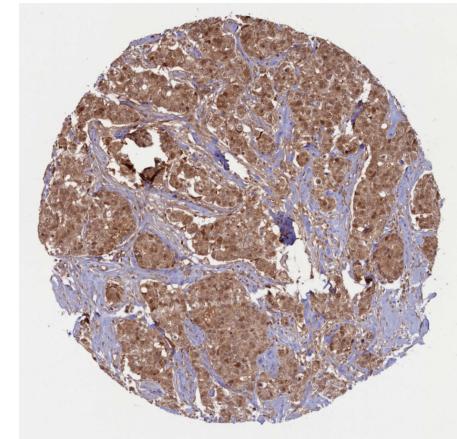
Key features / differences with analytical relevance

	Genome	Transcriptome	Proteome		Metabolome
			proteins	proteoforms	
Size	~finite	~finite	~finite	large unknown total	well understood core unknow total
Abundance range	small	~3 orders	>6 orders		high & variable
Abundance dynamic	static	dynamic	dynamic in space and time	dynamic in space and time	contains highly stable elements and elements changing dynamically in space and time
Structural stability	few / slow changes	few	stable	variable	variable
Chemical composition	defined, low complexity	Defined, low complexity	defined, moderately complex	complex	large chemical diversity

How can we study the
proteome?

Antibody based methods

- Western blot
- Immunohistochemistry
- ELISA
- Reverse Phase Protein Array (RPPA)
- Flow / Mass cytometry



Only as specific & sensitive as the antibody!

Benefits of antibody based detection

- Well established and mature
- Signal amplification possible
- Single cells
- High spatial resolution (even single molecules)

Limitations

- Limited sample and analyte multiplexing capability
- Only as specific & sensitive as the antibody!
- Only detection of known proteins
- Cost & complexity prohibitive for validation of markers for genome / proteome scale studies

Mass Spectrometry

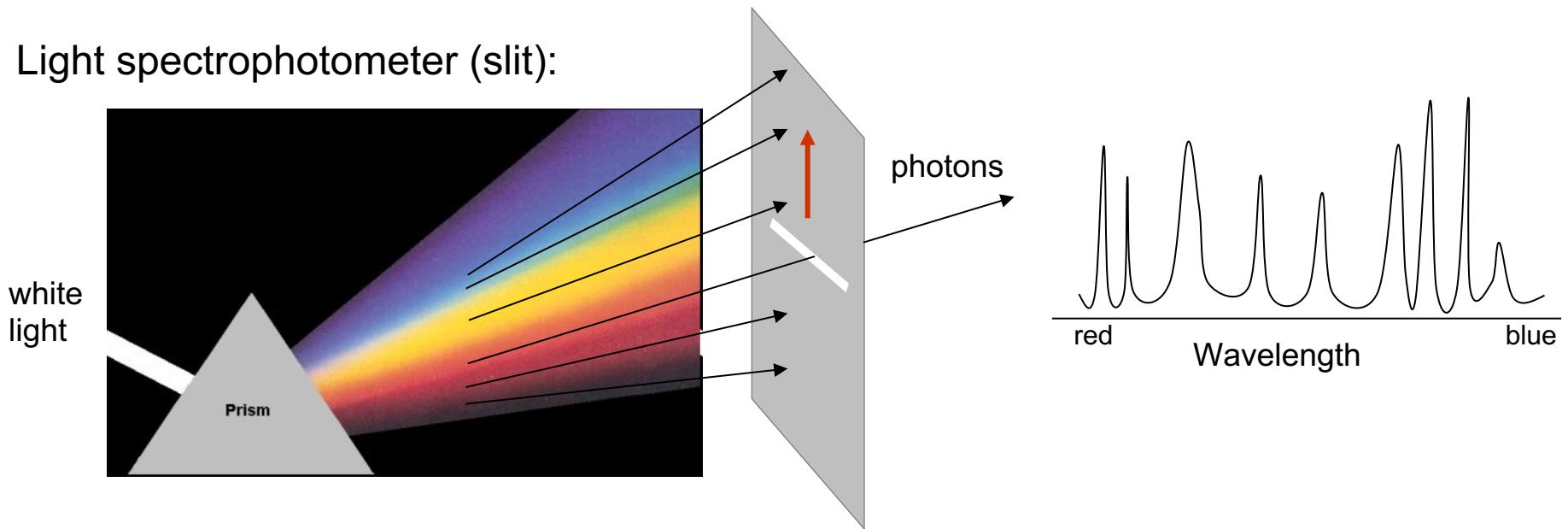
What is mass spectrometry?

Mass spectrometry is the art of **measuring** atoms and **molecules to determine their molecular weight**. Such mass or weight information is sometimes sufficient, frequently necessary, and always useful in determining the identity of a species. To practice this art **one puts charge on the molecules** of interest, i.e., the analyte, **then measures how the** trajectories of the resulting **ions respond in vacuum to** various combinations of **electric and magnetic fields**.

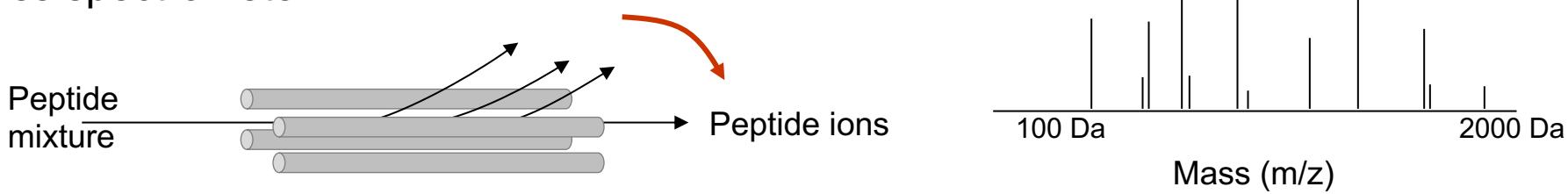
*attributed to John B. Fenn,
2002 Nobel laureate*

What is a mass spectrometer?

Light spectrophotometer (slit):



Mass spectrometer:



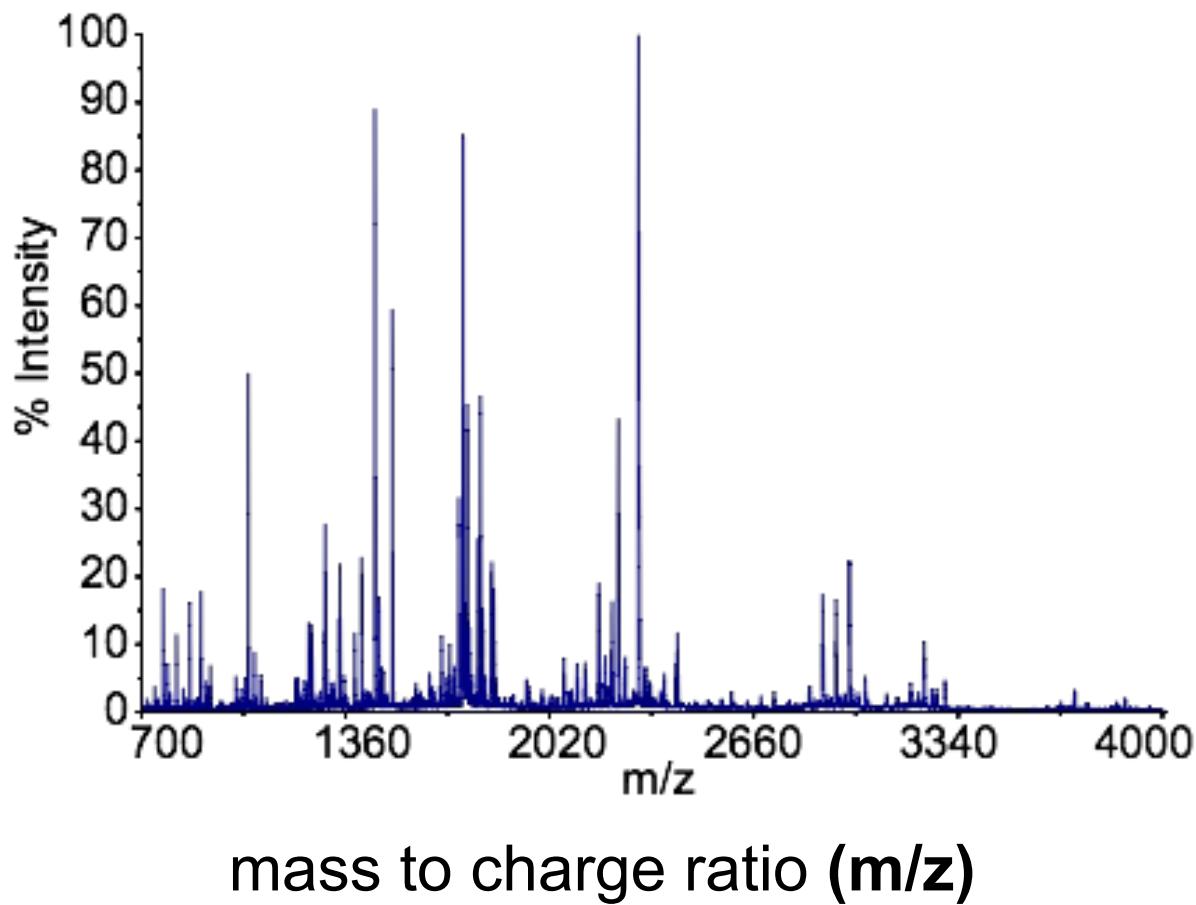
What is measured?

Mass to charge ratio (m/z)

Two particles with the same mass-to-charge ratio move in the same path in a vacuum when subjected to the same electric and magnetic fields.

An ion of 120 atomic mass units ($m = 120$) carrying two charges ($z = 2$) will be observed at $m/z = 60$.

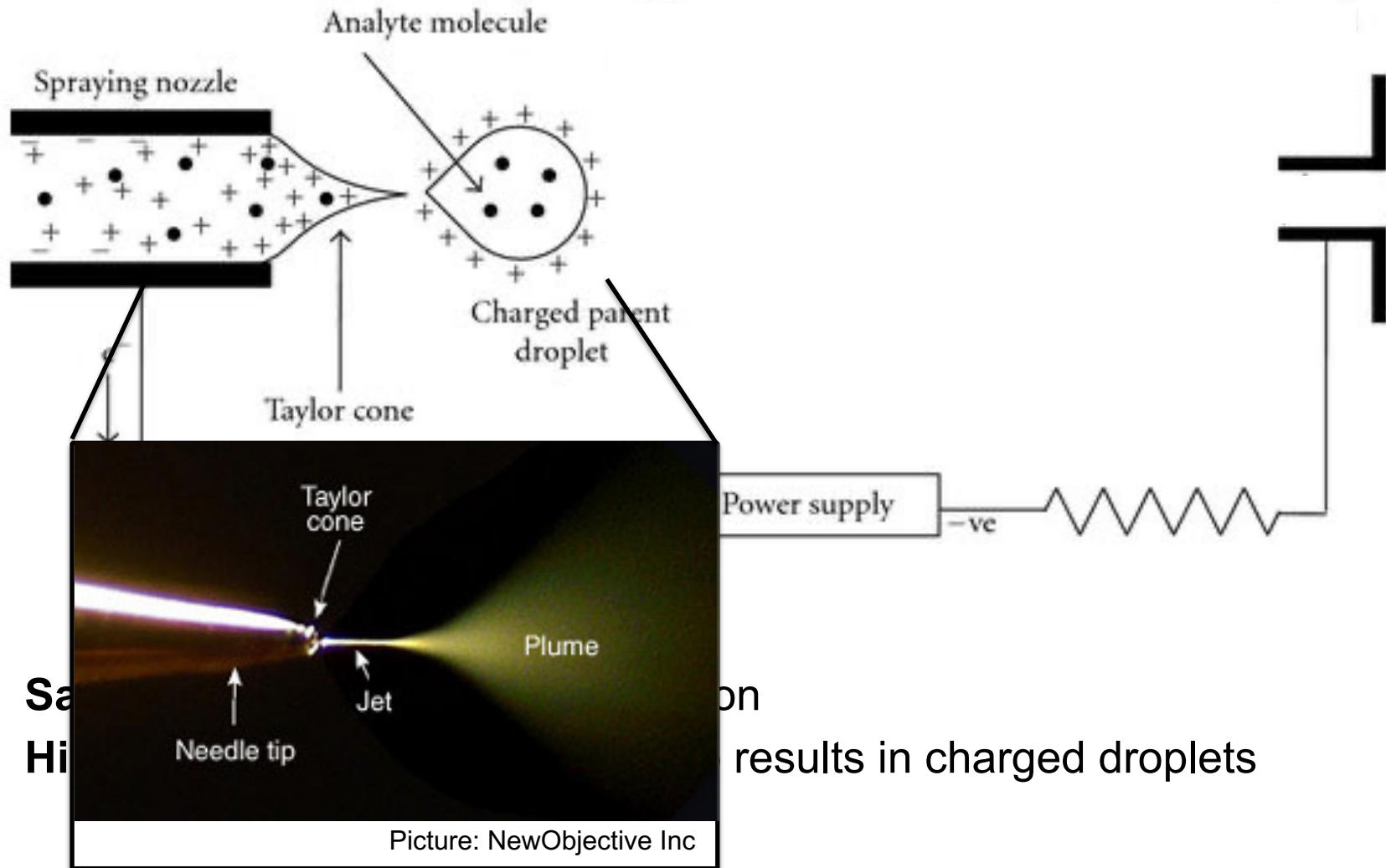
What is measured?



Requires effective ionization of compound!

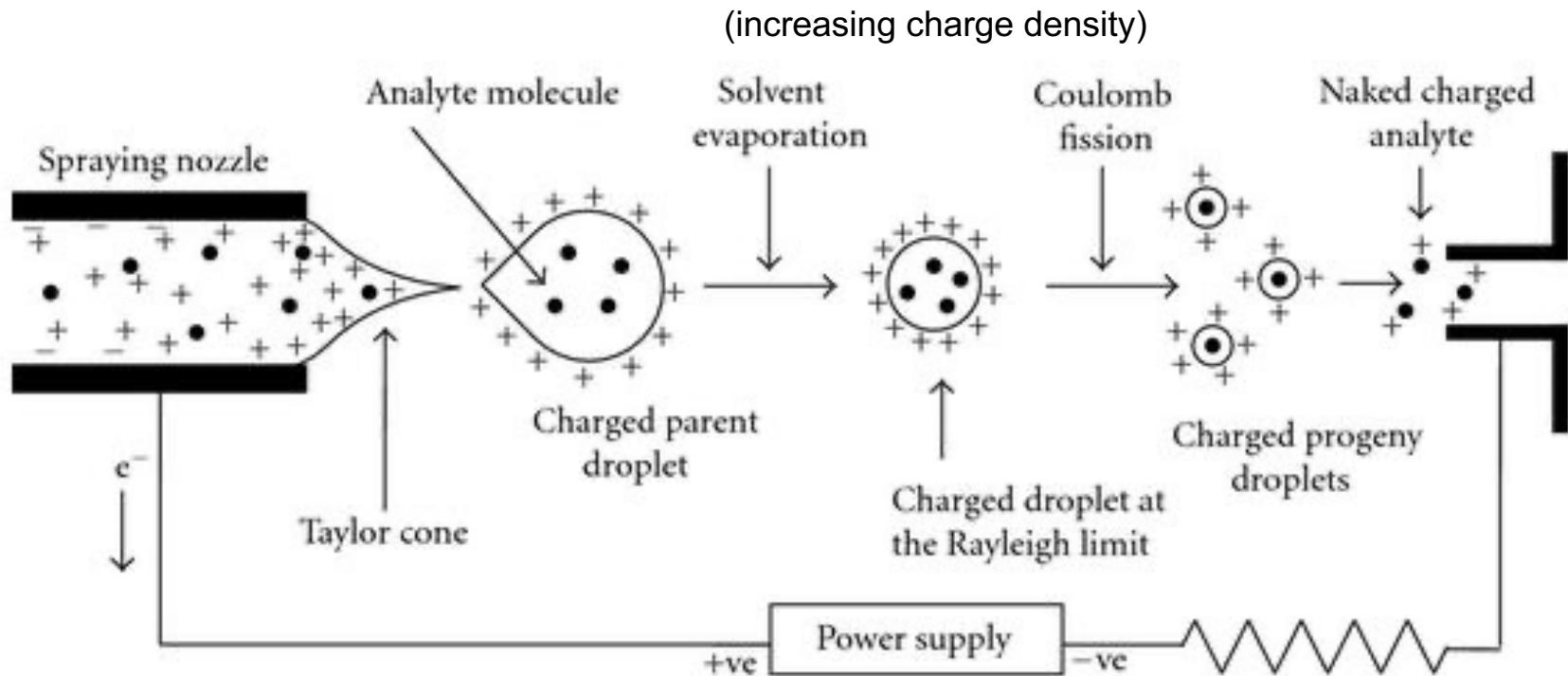
How do we get a charge on the analyte?

ElectroSpray Ionization



How do we get a charge on the analyte?

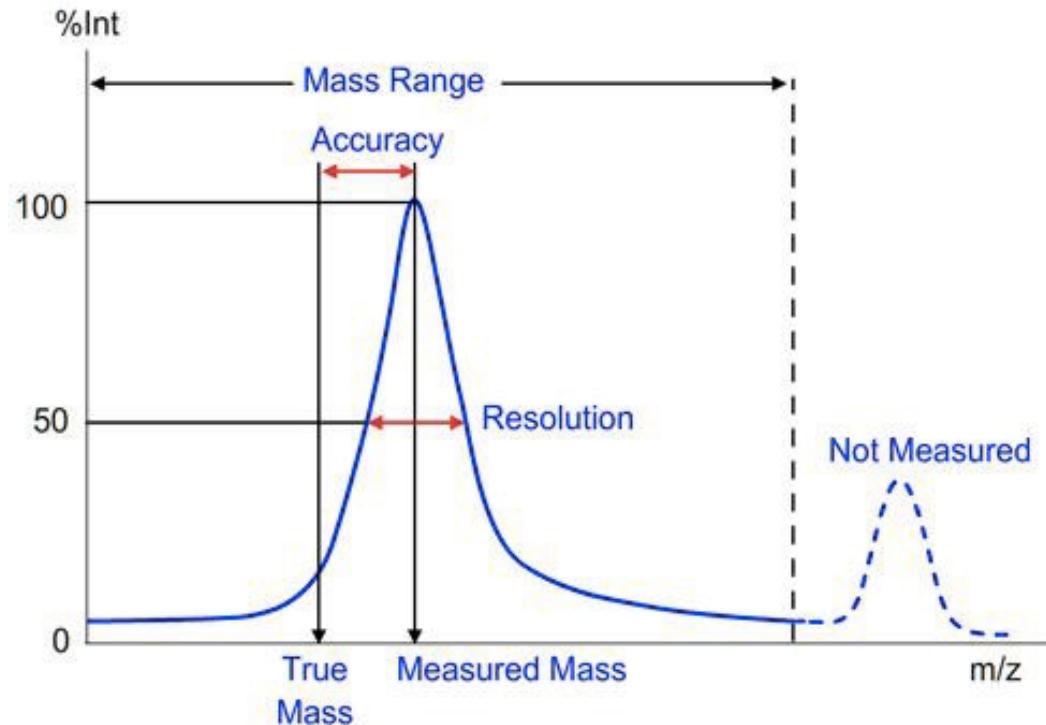
ElectroSpray Ionization



- **Sample** introduced as **liquid solution**
- **High voltage** applied **at needle tip** results in charged droplets
- Droplets sequentially dried until molecule ion remains

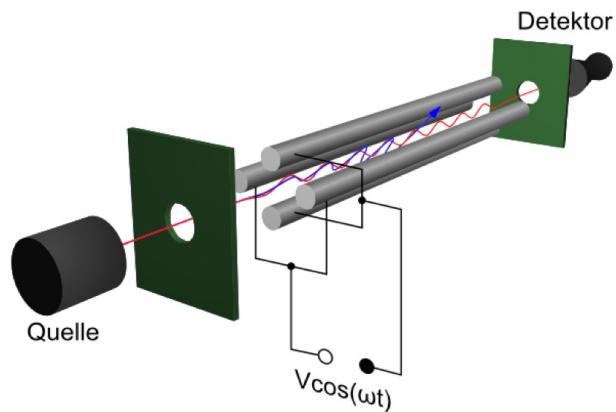
Basic properties of mass analyzers

- **Speed:** Expressed as spectrum acquisition rate (in Hz) or cycle time
- **Mass range:** The m/z range in which ions are detected
- **Mass accuracy:** Difference between the measured mass and true (calculated) mass of a compound, usually given in ppm (parts per million) or Da
- **Resolution:** A measure of the ability to resolve two distinct peaks

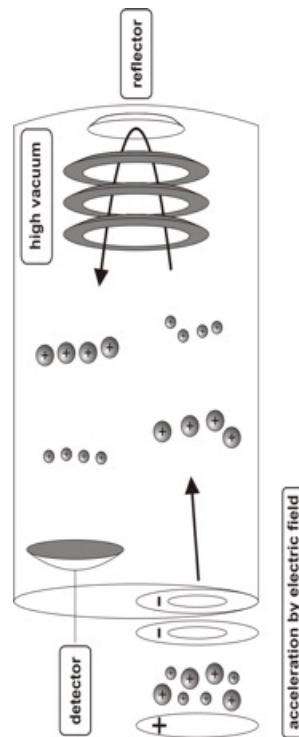


Mass analyzers

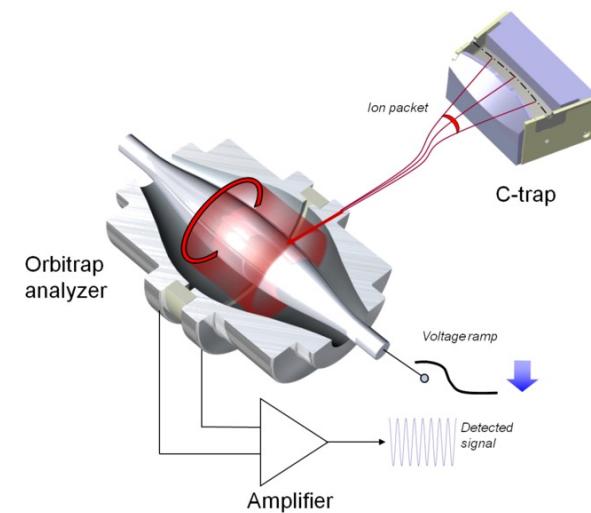
quadrupol



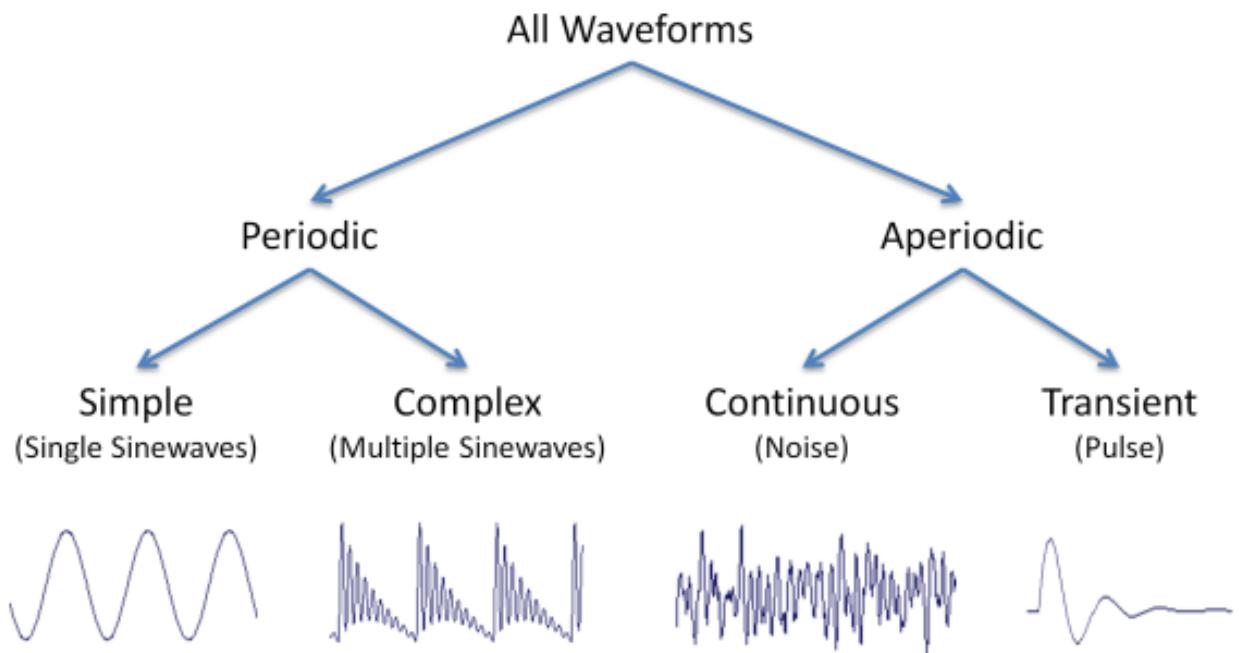
time of flight



orbitrap



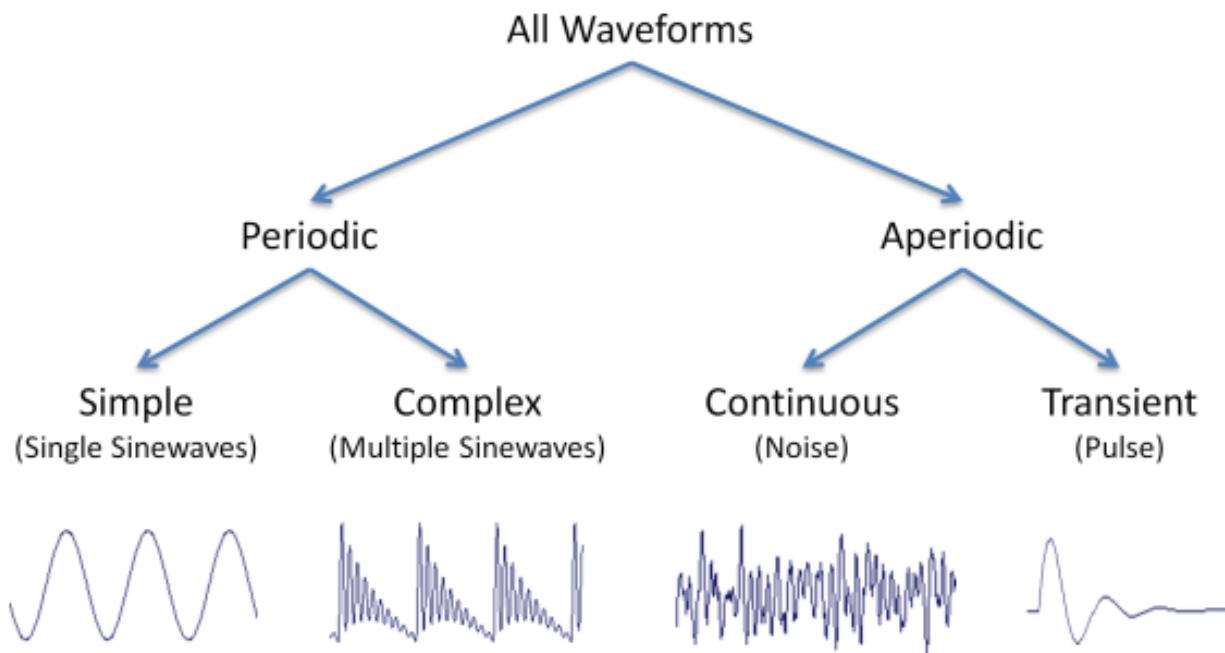
Complex waveforms



Fourier Transformation



Jean Baptiste Joseph Fourier
(1768 – 1830)



Any periodic waveform with frequency f can be constructed as a superposition of sine waves with frequencies $f, 2f, 3f, \dots$

Fourier transformation

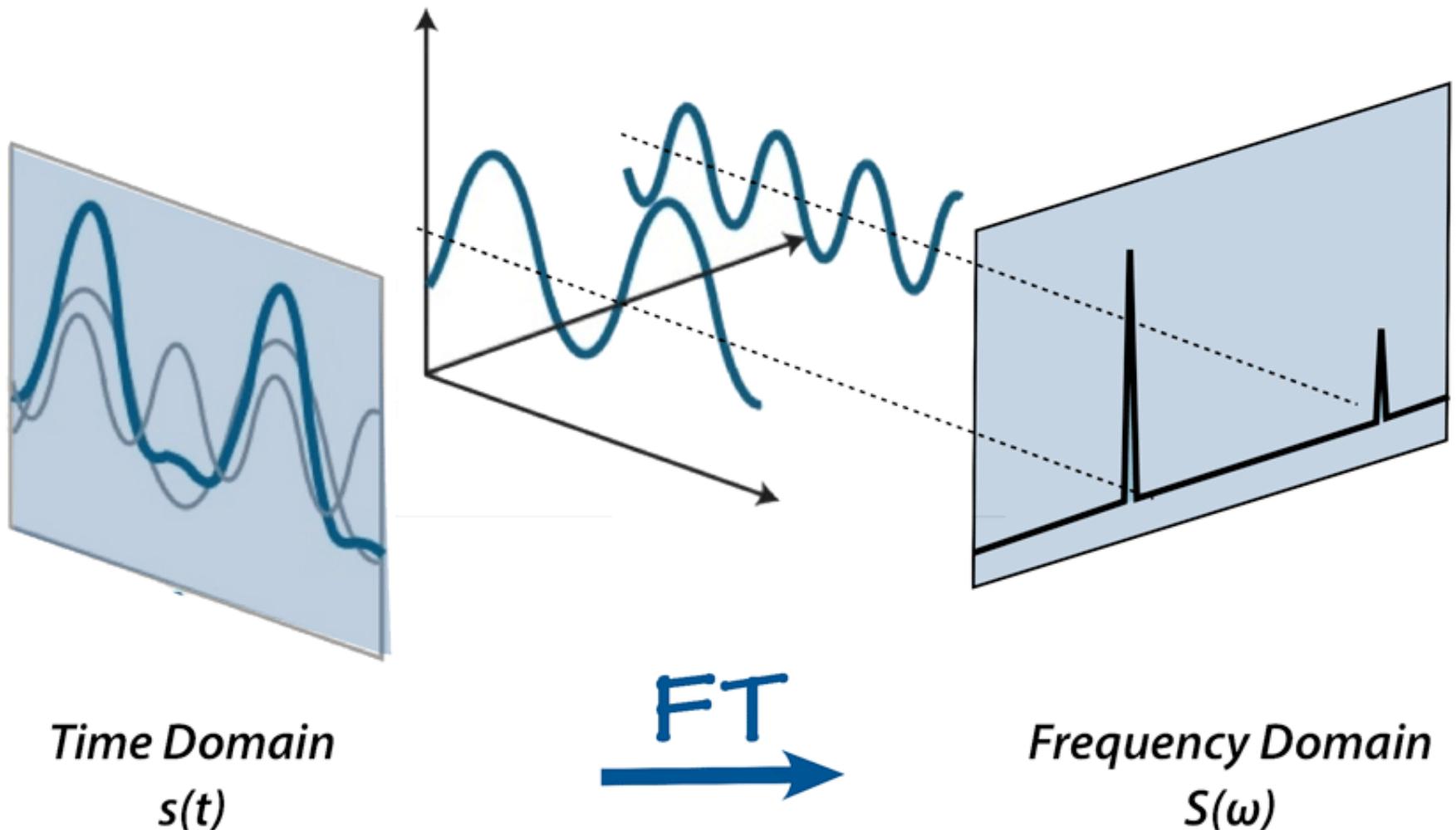
The **Fourier transformation** converts waveform data in the time domain into the frequency domain.

The Fourier transform accomplishes this by breaking down the original time-based waveform into a series of sinusoidal terms, each with a unique

magnitude, frequency, and phase.

This process, in effect, converts a waveform in the time domain that is difficult to describe mathematically into a more manageable series of sinusoidal functions that when added together, **exactly reproduce the original waveform.**

Fourier Transformation



Time Domain
 $s(t)$

FT

Frequency Domain
 $S(\omega)$

Why is mass spectrometry powerful?

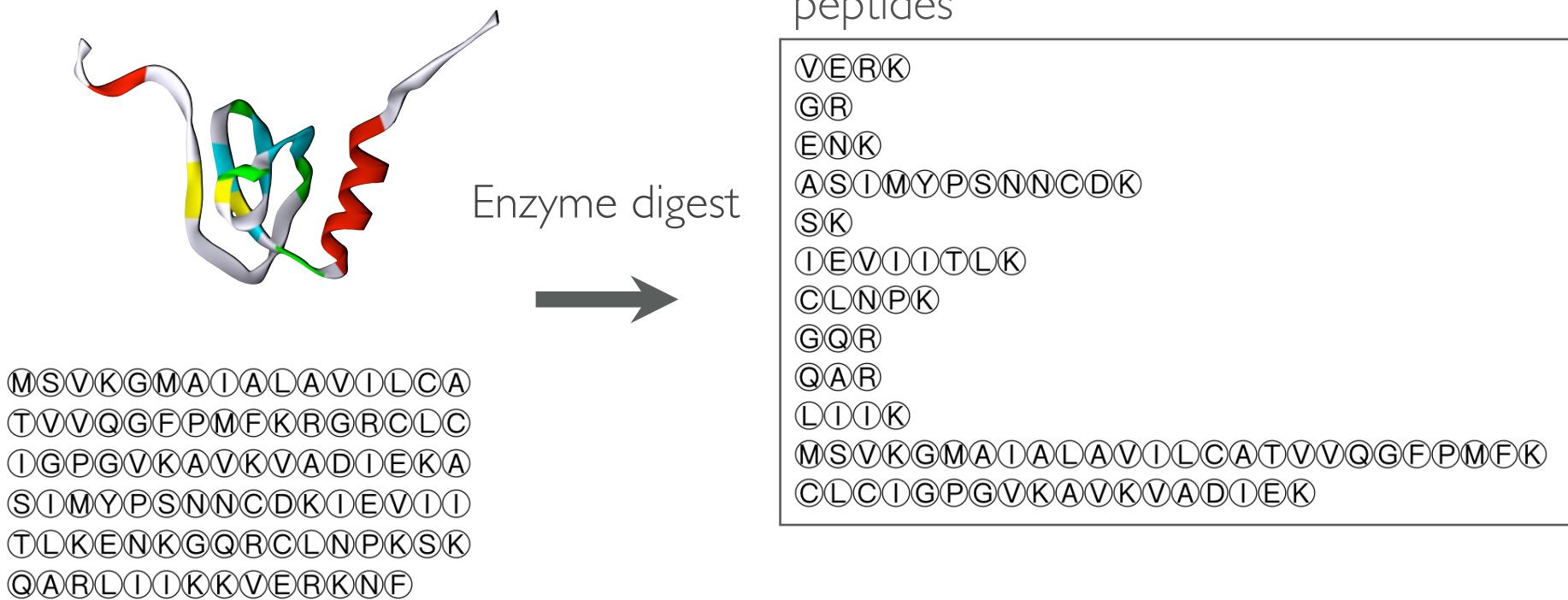
- Detector for all kinds of samples:
 - Elements
 - Synthetic molecules and polymers
 - DNA/RNA
 - Metabolites
 - Lipids
 - Peptides
 - Proteins
- Often fast and convenient
- Very sensitive, attomoles of analyte may be sufficient
- **Enables identification of unknown compounds**
- Relative and absolute quantification possible

Mass spectrometry for proteomics

in a nutshell

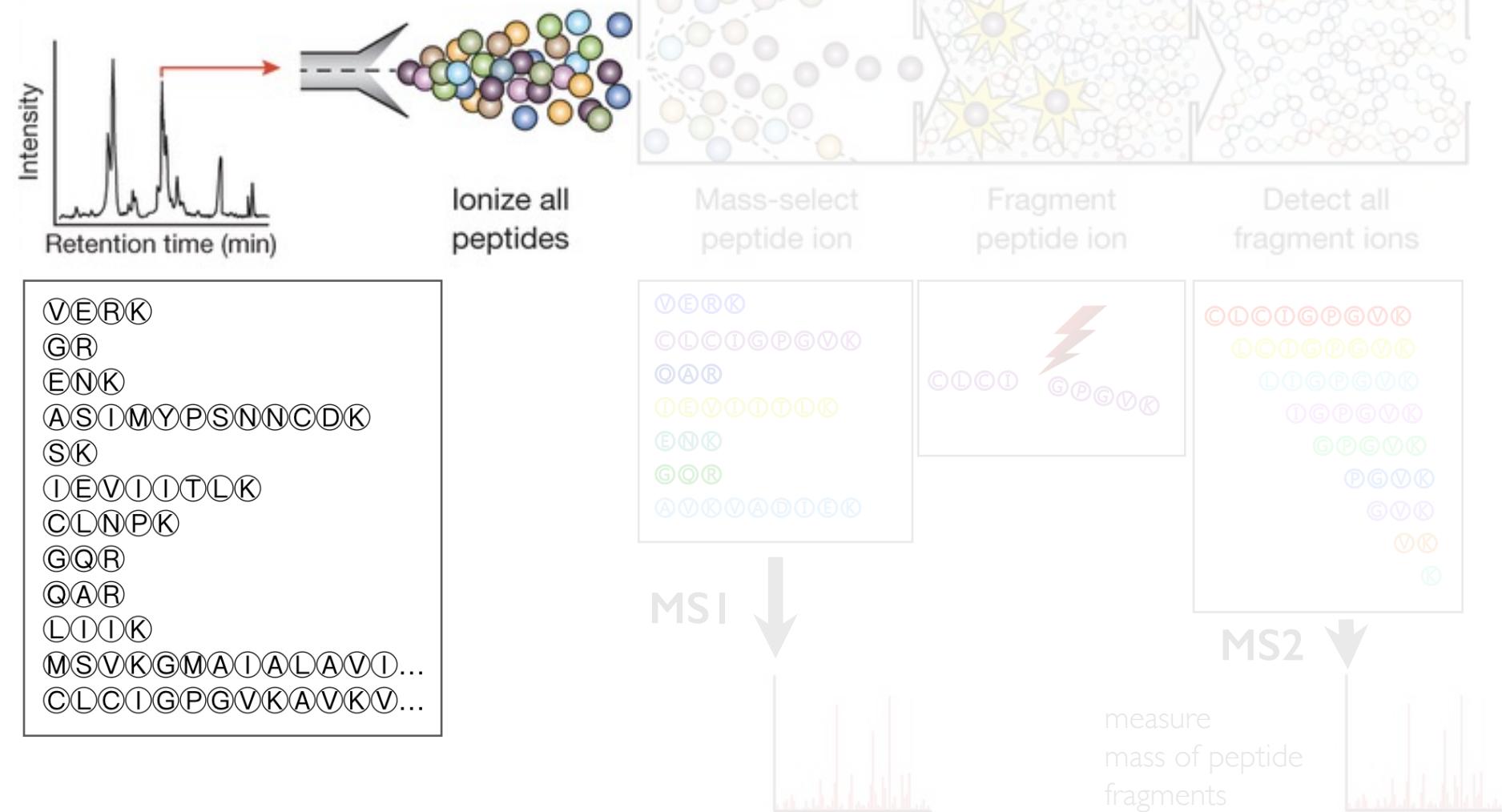
Large molecules are more difficult to measure

> digest to peptides



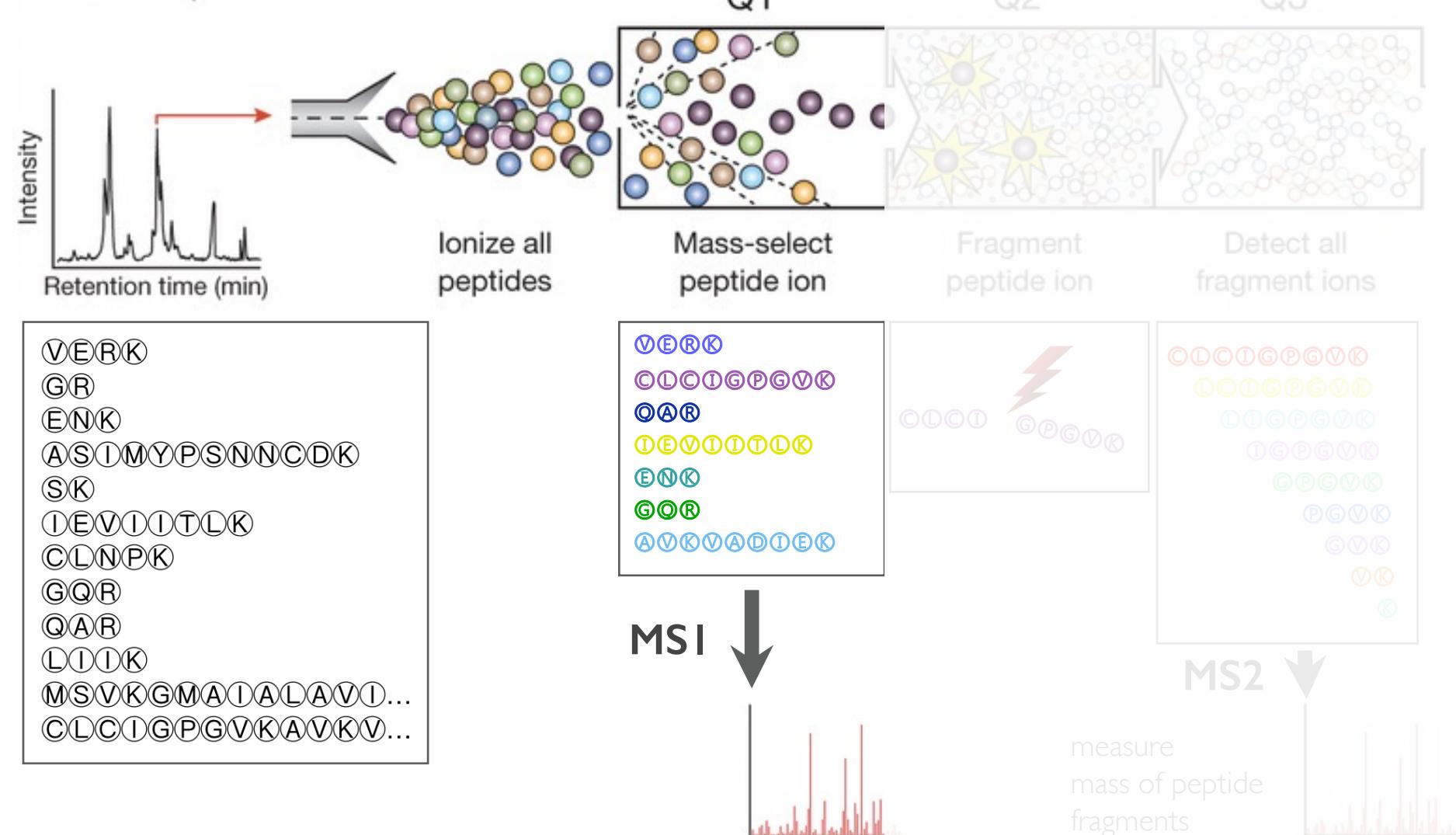
Ionization

Data-dependent MS/MS



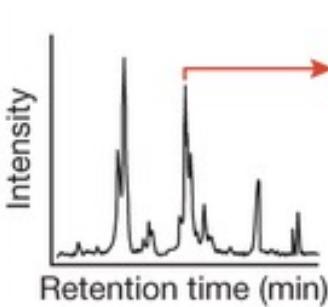
Measure peptide mass / select one

Data-dependent MS/MS



Fragment selected peptide

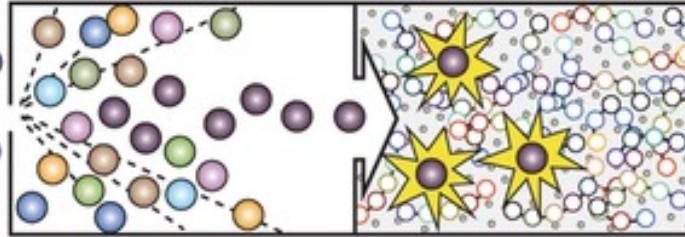
Data-dependent MS/MS



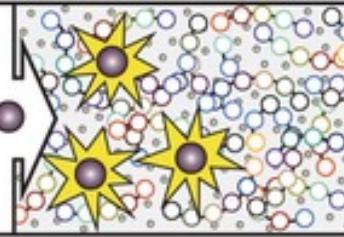
Ionize all peptides

V E R K
G R
E N K
A S I M Y P S N N C D K
S K
I E V I I T L K
C L N P K
G Q R
Q A R
L I I K
M S V K G M A I A L A V I ...
C L C I G P G V K A V K V ...

Q1



Q2



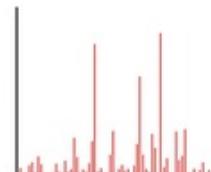
Q3



V E R K
C L C I G P G V K
Q A R
I E V I I T L K
E N K
G Q R
A V K V A D I E K

C L C I G P G V K
I E V I I T L K
E N K
G Q R
A V K V A D I E K

MS1



C L C I G P G V K
I E V I I T L K
E N K
G Q R
A V K V A D I E K

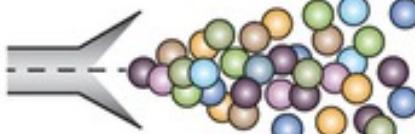
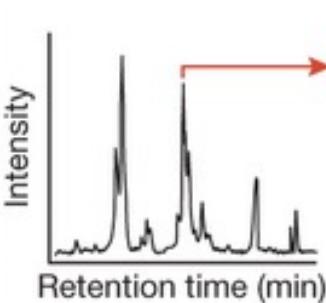
MS2



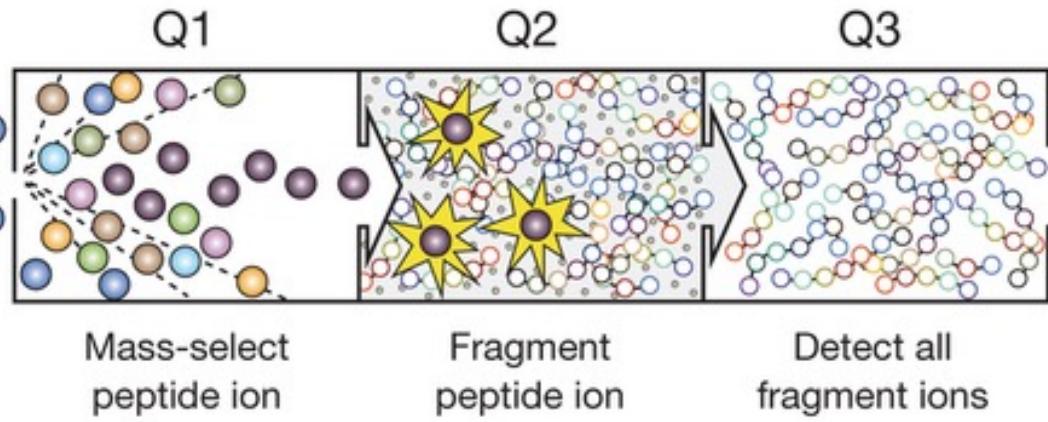
measure mass of peptide fragments

Detect fragment ions

Data-dependent MS/MS



Ionize all peptides



Mass-select peptide ion

Fragment peptide ion

Detect all fragment ions

V E R K
G R
E N K
A S I M Y P S N N C D K
S K
I E V I I T L K
C L N P K
G Q R
Q A R
L I I K
M S V K G M A I A L A V I ...
C L C I G P G V K A V K V ...

V E R K
C L C I G P G V K
Q A R
I E V I I T L K
E N K
G Q R
A V K V A D I E K

MS1

measure mass of peptides



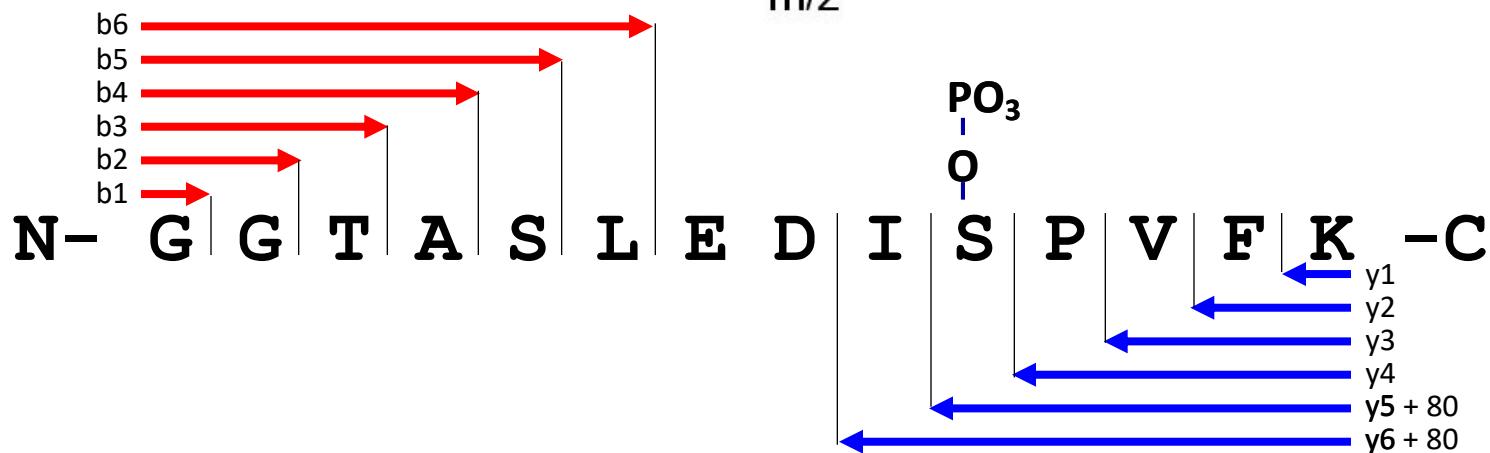
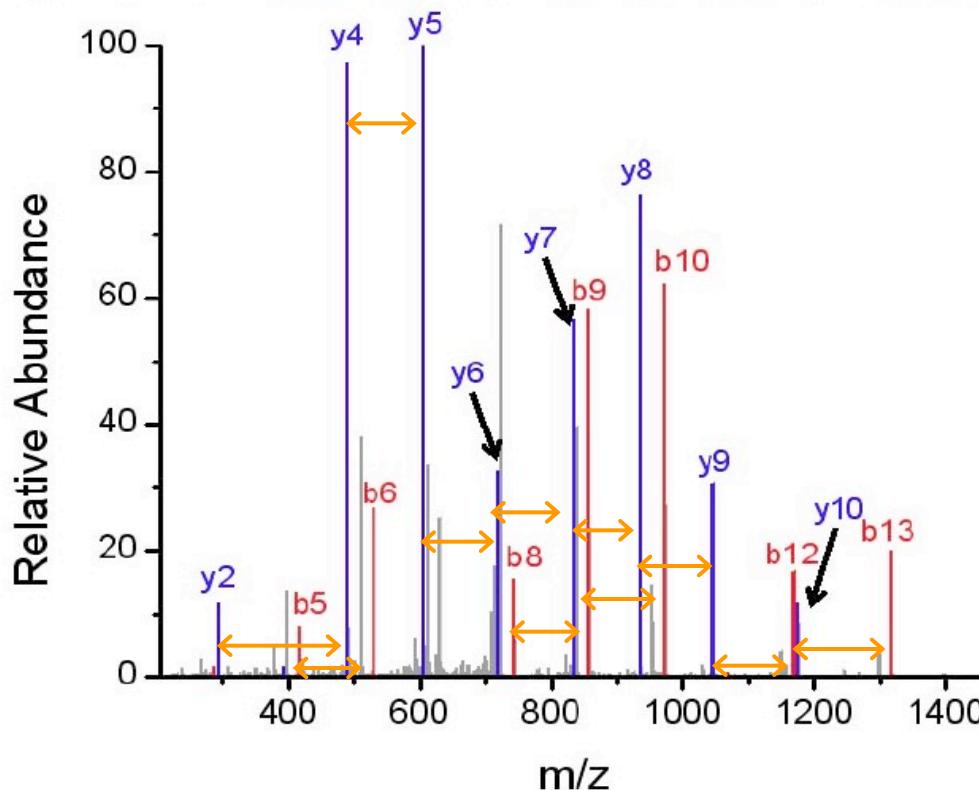
C L C I G P G V K
I C I G P G V K
L I G P G V K
I G P G V K
G P G V K
P G V K
G V K
V K
K

MS2

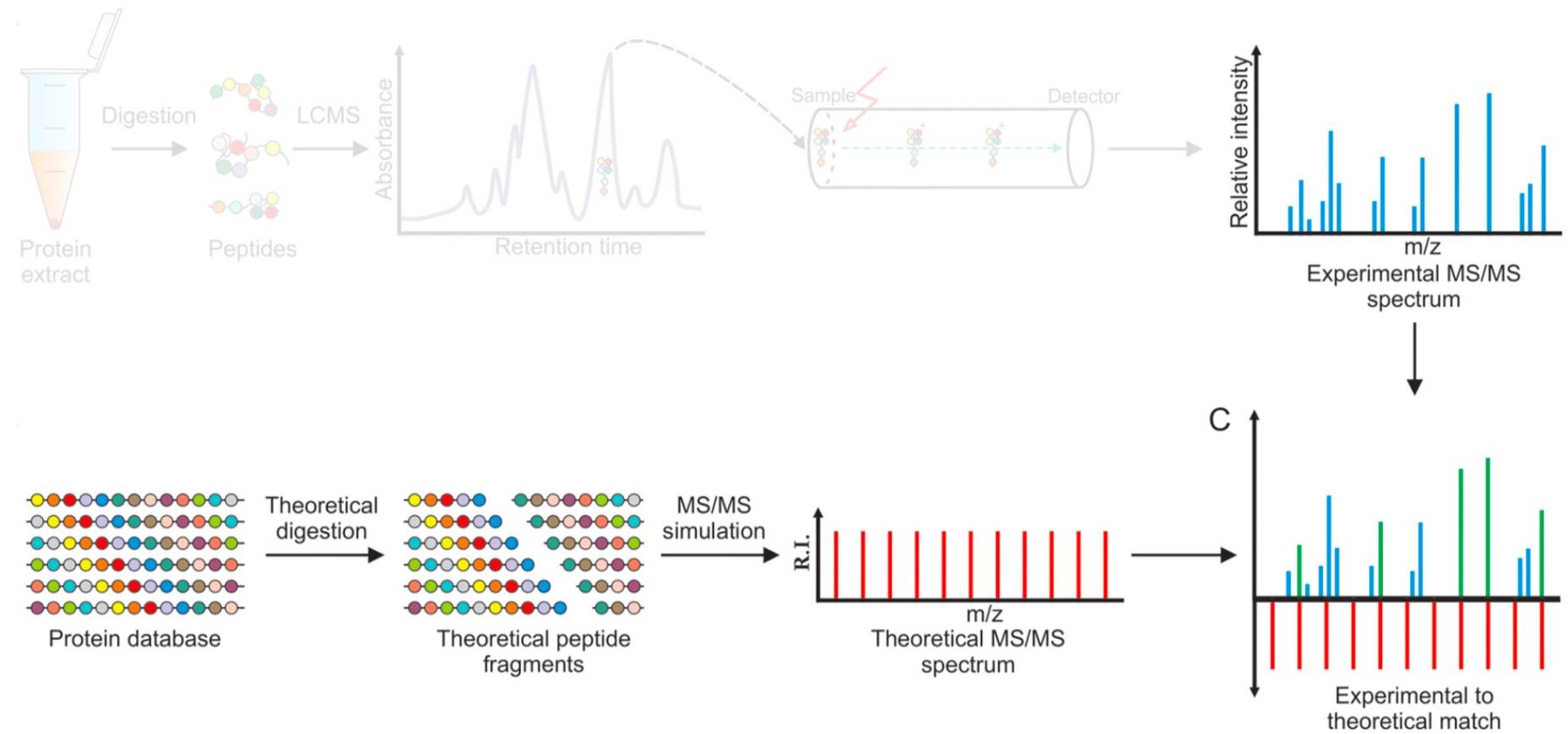
measure mass of peptide fragments



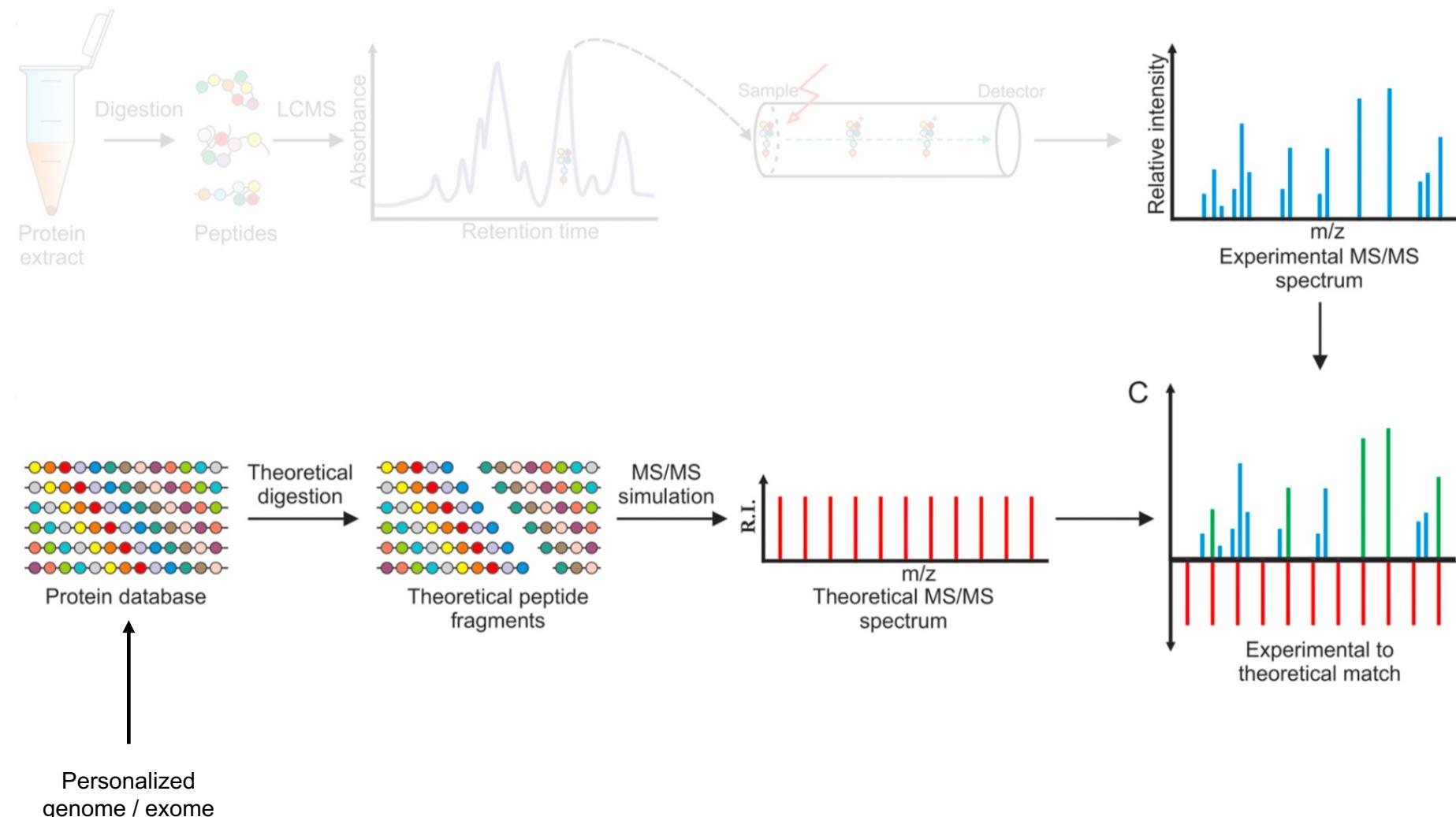
MS/MS Spectrum Analysis



Peptide identification by database search



Proteogenomics personalized proteome databases



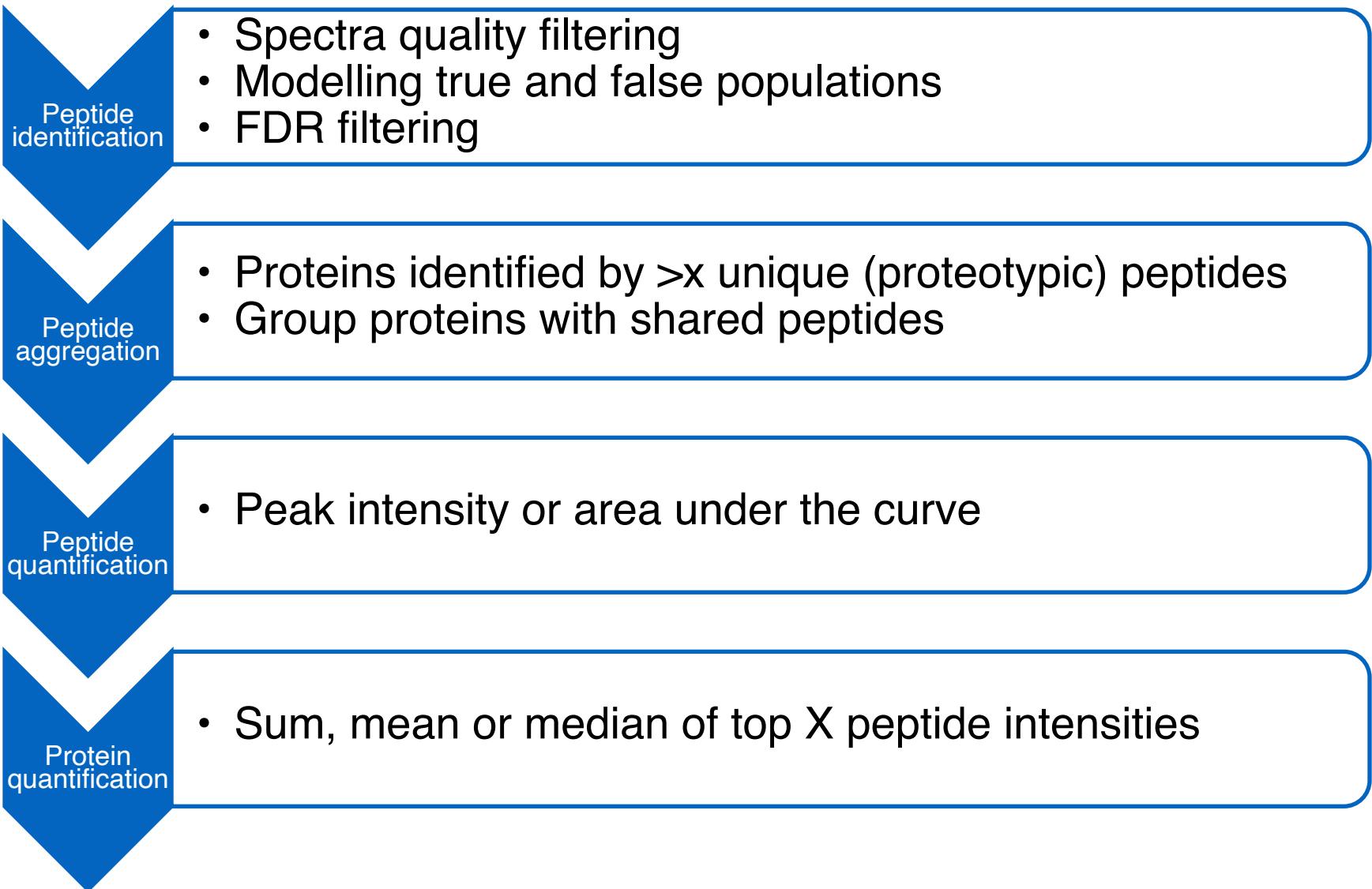
Benefits of mass spec based detection

- Comprehensive (proteome wide) and unbiased
- Detection of unknown proteins and modifications
- High specificity
- Quantitation across multiple orders of magnitude

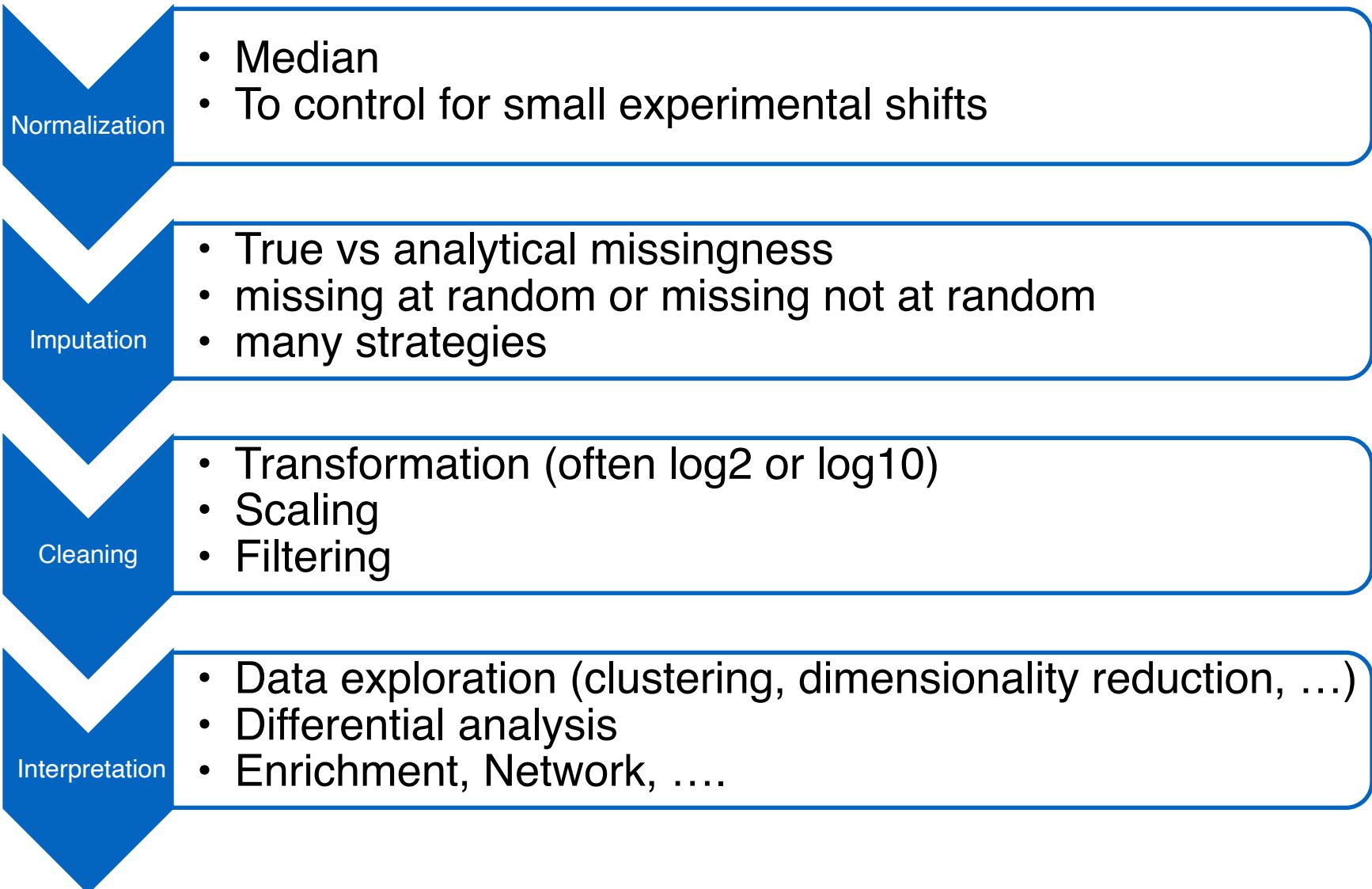
Limitations

- Limited sample multiplexing capability
- No amplification possible
- Limited spatial resolution
- Destructive (no live measurement)

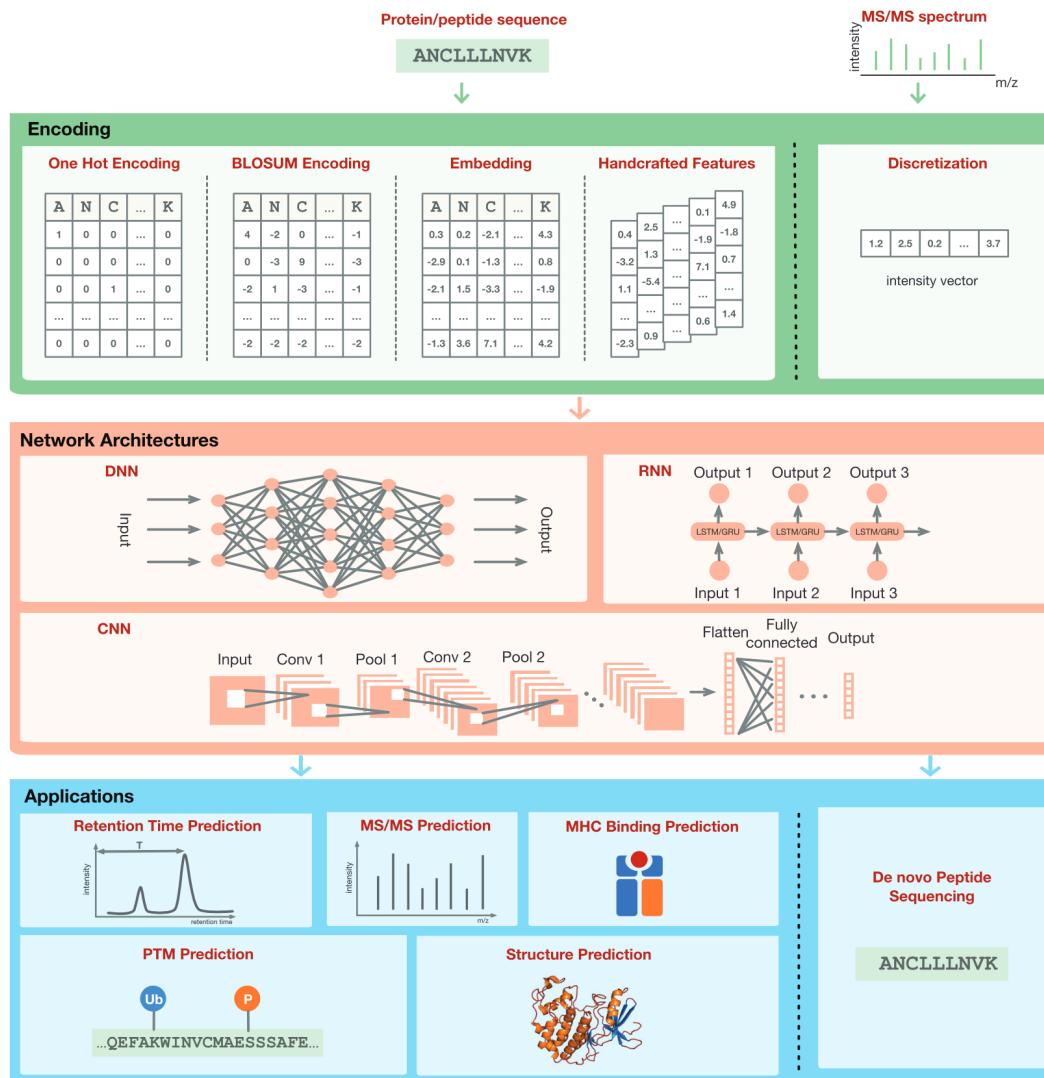
General data analysis workflow



General data analysis workflow



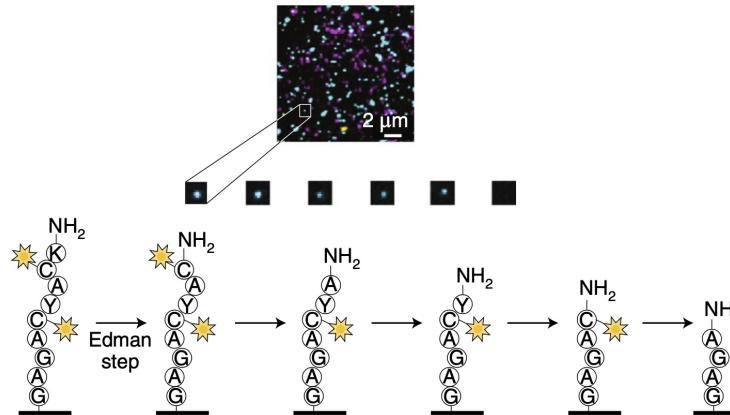
Machine learning in proteomics



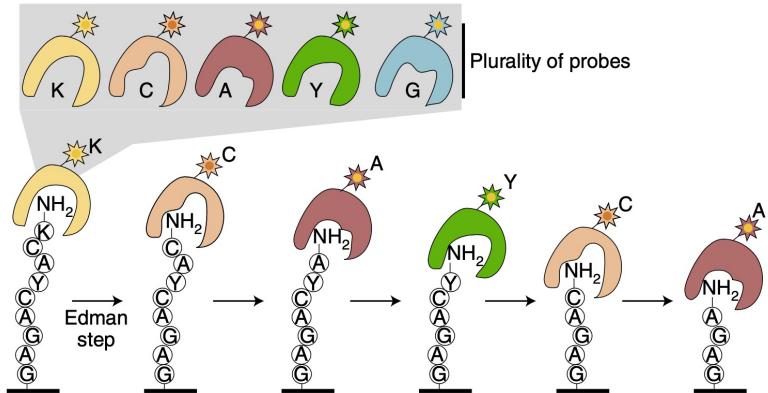
NGS revolutionized genomics

-
can it do the same for
proteomics?

a Fluorosequencing by chemical modification

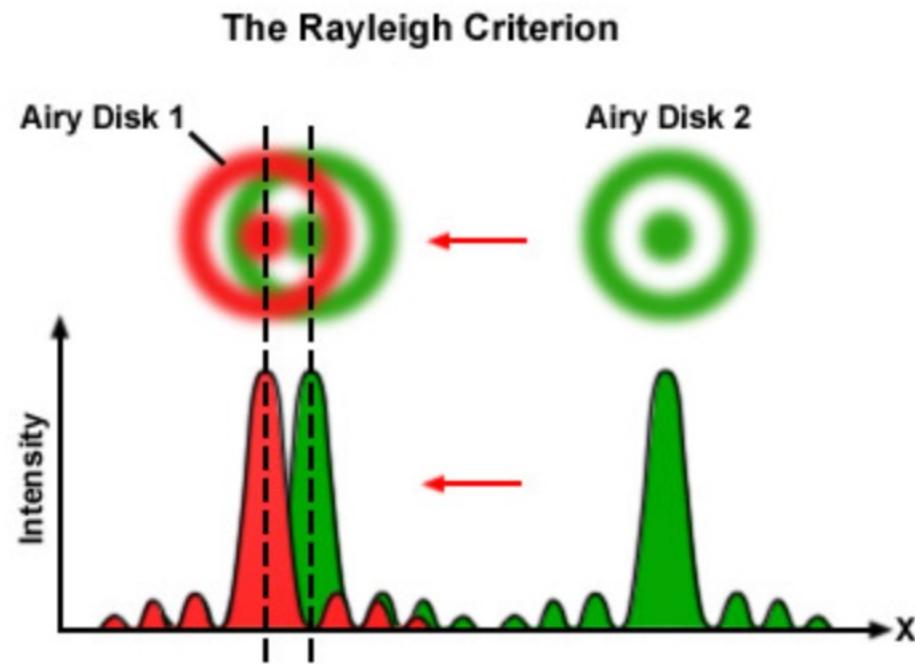


b Sequencing by N-terminal probes



Resolution of light microscopy limited

- Diffraction limits resolution (Ernst Abbe in 1873)
- A good approximation of the resolution attainable is the full width at half maximum (FWHM) of the point spread function
- A precise widefield microscope with high numerical aperture and visible light usually reaches a **resolution of ~ 250 nm**



Deep proteome analysis is hard but possible

400,000 mRNA molecules per cell = 40 million reads

4,000,000,000 protein molecules per cell = 400 billion reads

Flow cell based

- Fluorescent wavelength = 200-750nm
 - leads to practical density of ~1µm
- Flow cell size:
 - 1 million reads: 1mm²
 - 100 million reads: 1cm²
 - 1 billion reads: 10cm²
 - 1 trillion reads: 1m²
- Current fluorescence based flow cells limit reads based on particle density and wide field microscopy limits

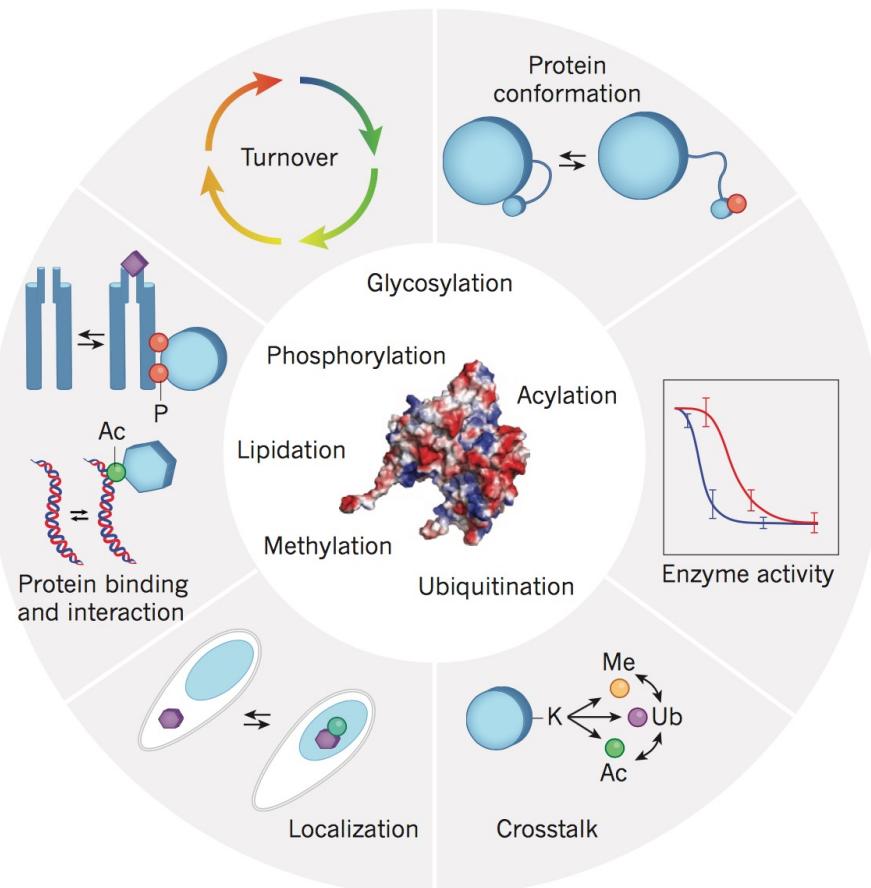
Mass spectrometry based

- 500,000 ions in <1sec
- 30,000,000 ions/min
- 1.8 billion reads/hour
- <\$100 per hour

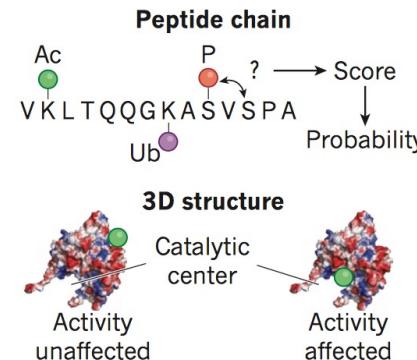
What can we learn
from comprehensive
protein mass
spectrometry

No single answer – experimental design determines meaning of data

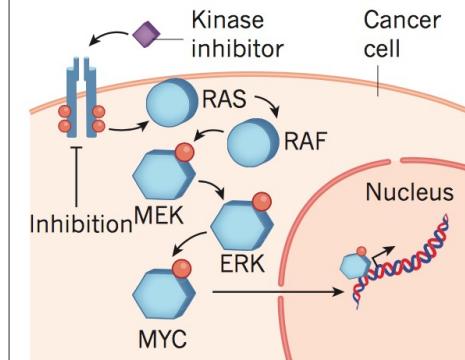
a



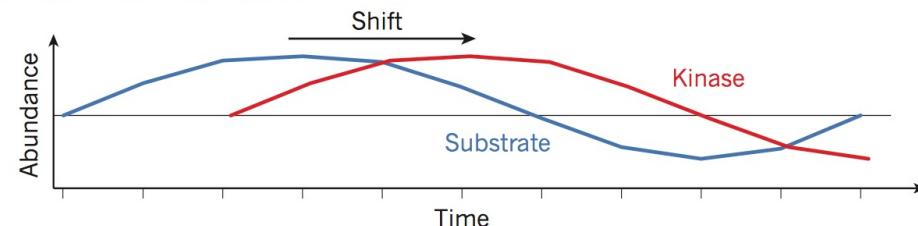
b Location of modification



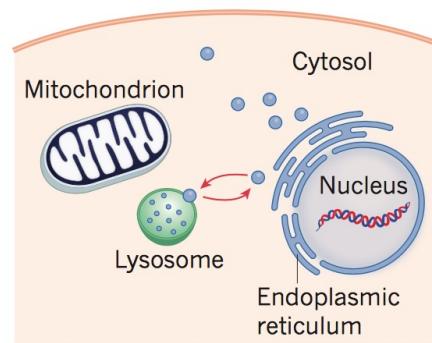
c Cancer signalling



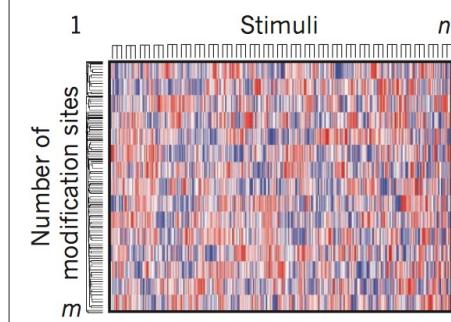
d Time-course experiments



e Subcellular localization



f Large-scale perturbation matrix



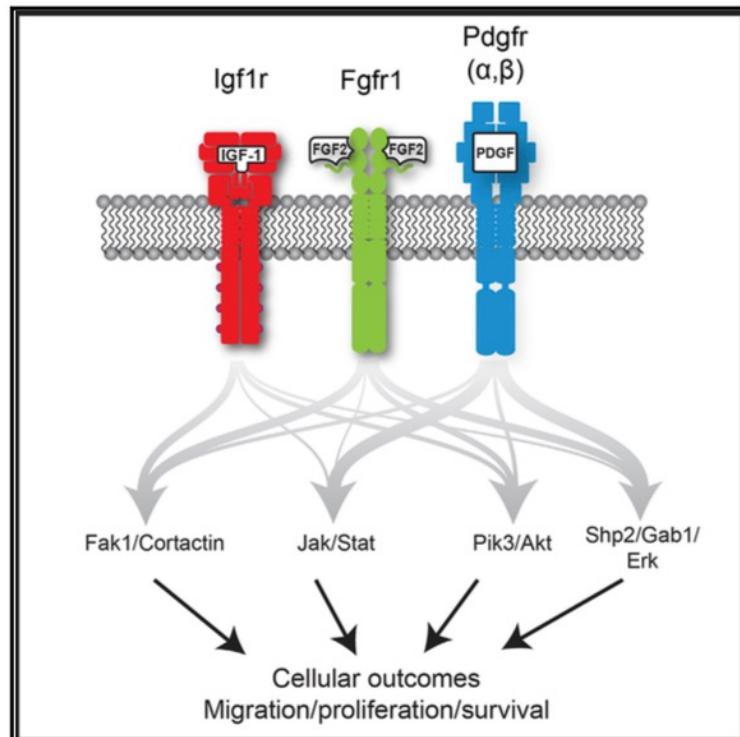
Deconvolution of phospho-signaling pathways

Cell Reports

Resource

Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling

Graphical Abstract



Authors

Tanveer S. Bath, Moreno Papetti,
Anamarija Pfeiffer, Maxim A.X. Tollenaere,
Chiara Francavilla, Jesper V. Olsen

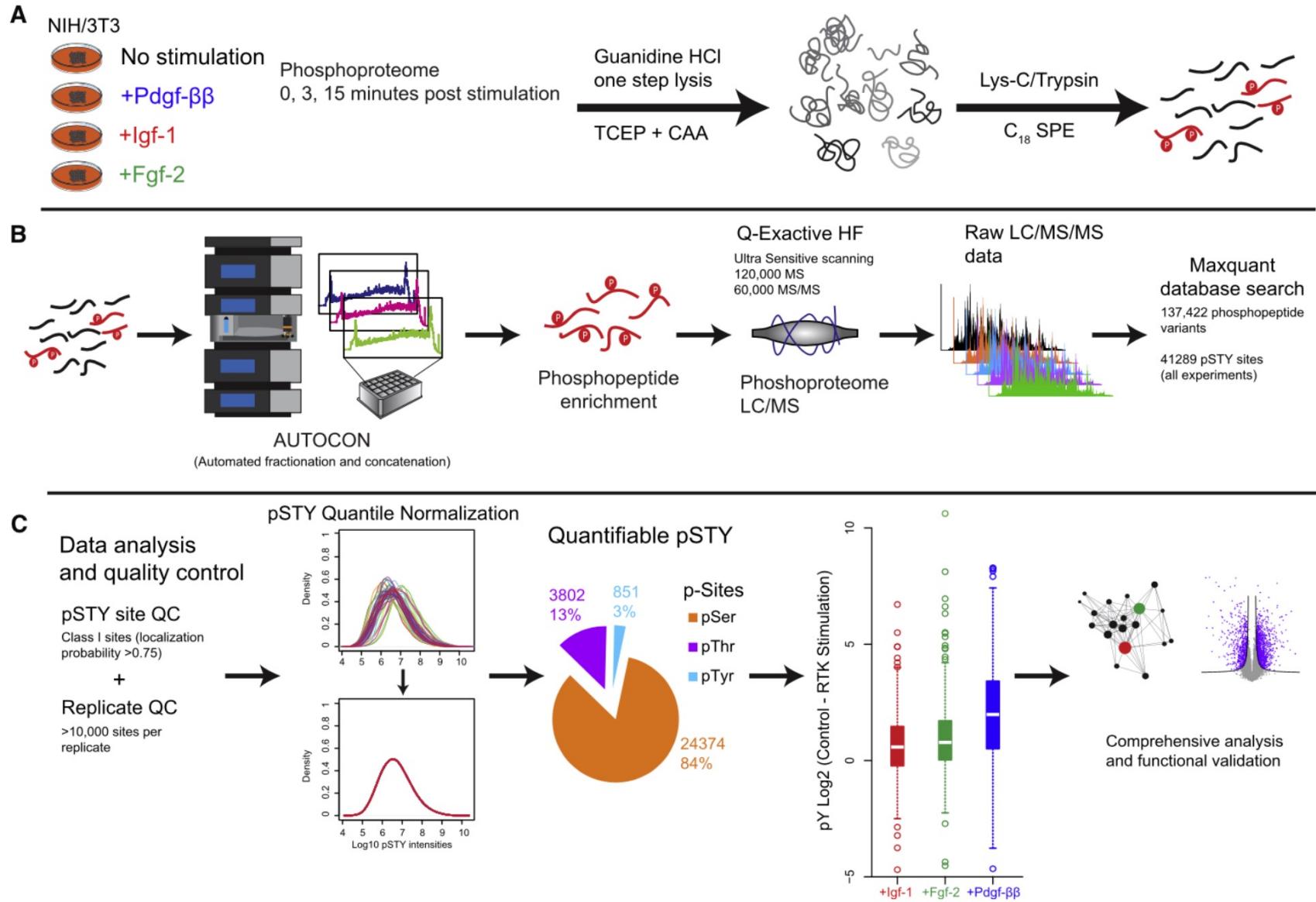
Correspondence

[\(C.F.\),](mailto:chiara.francavilla@manchester.ac.uk)
[\(J.V.O.\)](mailto:jesper.olsen@cpr.ku.dk)

In Brief

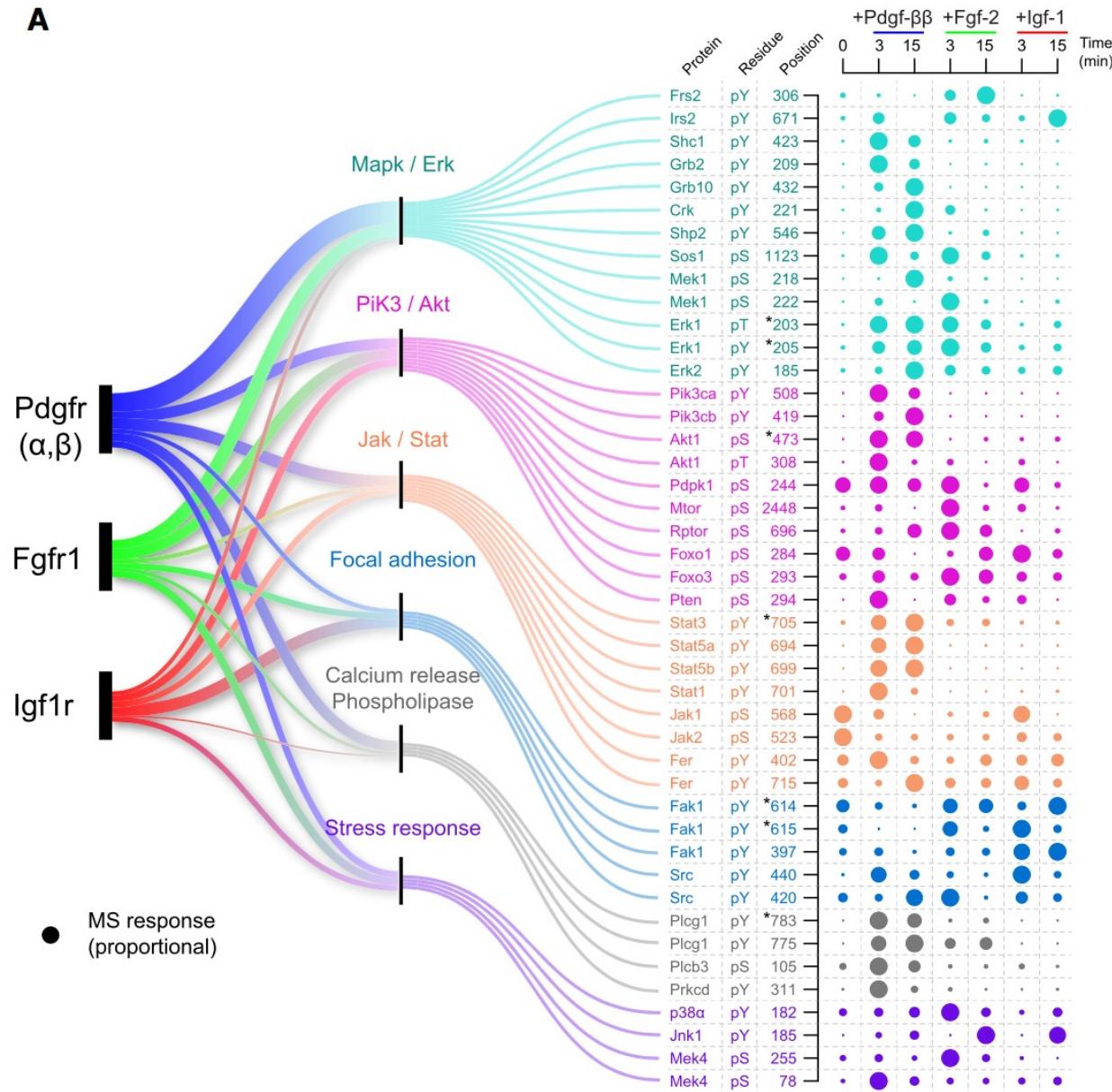
Bath et al. use mass spectrometry-based phosphoproteomics to analyze receptor tyrosine kinase signaling activated by different ligands, identifying hundreds of differentially regulated phosphotyrosine sites. Tyrosine phosphatase Shp-2 regulates global tyrosine phosphorylation in a Pdgf-receptor-dependent manner, affecting cellular outcomes.

Deconvolution of phospho-signaling pathways



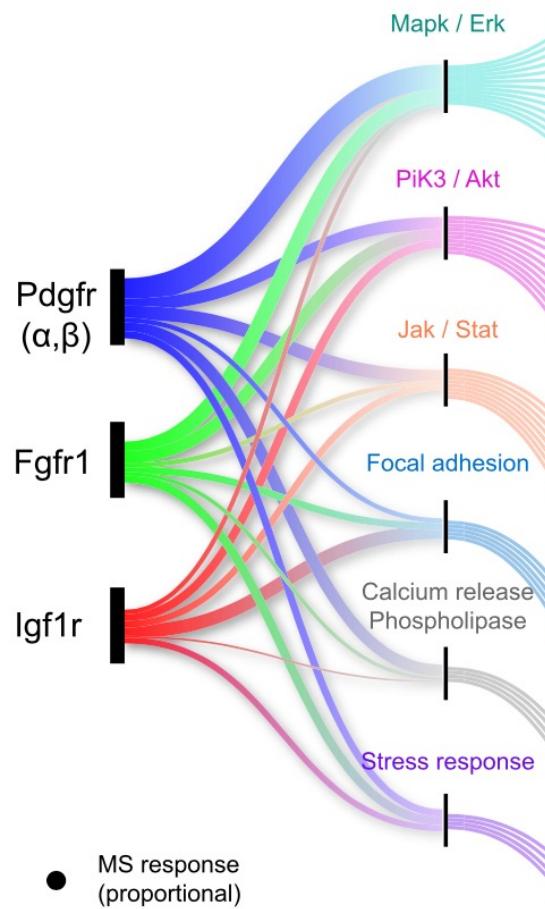
Deconvolution of phospho-signaling pathways

A



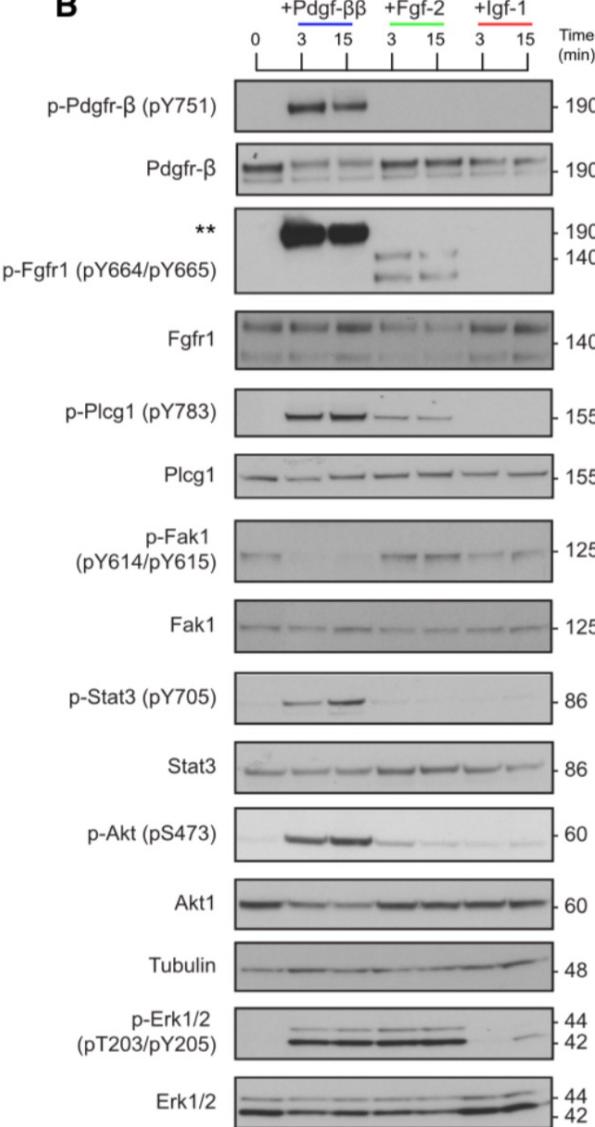
Deconvolution of phospho-signaling pathways

A



Protein	Residue	Position	+Pdgf- $\beta\beta$	+Fgf-2	+Igf-1	Time (min)
Frs2	pY	306	.	.	.	0
Irs2	pY	671	●	.	.	3
Shc1	pY	423	●	.	.	15
Grb2	pY	209	●	.	.	3
Grb10	pY	432	●	.	.	15
Crk	pY	221	●	.	.	3
Shp2	pY	546	●	.	.	15
Sos1	pS	1123	●	.	.	3
Mek1	pS	218	●	.	.	15
Mek1	pS	222	●	.	.	3
Erk1	pT	*203	●	.	.	15
Erk1	pY	*205	●	.	.	3
Erk2	pY	185	●	.	.	15
Pik3ca	pY	508	●	.	.	3
Pik3cb	pY	419	●	.	.	15
Akt1	pS	*473	●	.	.	3
Akt1	pT	308	●	.	.	15
Pdkp1	pS	244	●	.	.	3
Mtor	pS	2448	●	.	.	15
Rptor	pS	696	●	.	.	3
Foxo1	pS	284	●	.	.	15
Foxo3	pS	293	●	.	.	3
Pten	pS	294	●	.	.	15
Stat3	pY	*705	●	.	.	3
Stat5a	pY	694	●	.	.	15
Stat5b	pY	699	●	.	.	3
Stat1	pY	701	●	.	.	15
Jak1	pS	568	●	.	.	3
Jak2	pS	523	●	.	.	15
Fer	pY	402	●	.	.	3
Fer	pY	715	●	.	.	15
Fak1	pY	*614	●	.	.	3
Fak1	pY	*615	●	.	.	15
Fak1	pY	397	●	.	.	3
Src	pY	440	●	.	.	15
Src	pY	420	●	.	.	3
Plcg1	pY	*783	●	.	.	3
Plcg1	pY	775	●	.	.	15
Plcb3	pS	105	●	.	.	3
Prkcd	pY	311	●	.	.	15
p38 α	pY	182	●	.	.	3
Jnk1	pY	185	●	.	.	15
Mek4	pS	255	●	.	.	3
Mek4	pS	78	●	.	.	15

B



Summary

- The proteome is complex and dynamic
- Genome, transcriptome, proteome and metabolome data differ in dimensionality, scale, distribution, ...
- Limited correlation between gene, transcript and protein
- Antibody and mass spectrometry based methods enable proteome quantification with different benefits and limitations
- Machine learning is now employed at all levels of mass spectrometry analysis
- Next gen sequencing like proteome analysis is limited by resolution
- Good experimental design is key