KSBi-BIML 2023

Bioinformatics & Machine Learning(BIML)
Workshop for Life Scientists, Data Scientists,
and Bioinformatians

생물정보학&머신러닝 워크샵(온라인)

질량분석을 활용한 단백체 연구

김민식 _ DGIST

SBi 한국생명정보학회
Korean Society for Bioinformatics

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2023 워크샵 온라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2023

Bioinformatics & Machine Learning (BIML)
Workshop for Life Scientists, Data Scientists, and Bioinformatians

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2023에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 9차를 맞이하게 되었습니다. 지난 2년간은 심각한 코로나 대유행으로 인해 아쉽게도 모든 강의가 온라인으로 진행되어 현장 강의에서만 가능한 강의자와 수강생 사이에 다양한 소통의 기회가 없음에 대한 아쉬움이 있었습니다. 다행히도 최근 사회적 거리두기 완화로 현장 강의가 가능해져 올해는 현장 강의를 재개함으로써 온라인과 현장 강의의 장점을 모두 갖춘 프로그램을 구성할 수 있게 되었습니다.

BIML 워크샵은 전통적으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되었습니다. 올해 AI 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 심층학습(Deep learning) 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체 분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 싱글셀오믹스 분석과 메타유전체분석 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다. 이외에 다양한 생명정보학 분야에 대하여 30개 이상의 온라인 강좌가 개설되어 제공되며 온라인 강의의 한계를 극복하기 위해서 실시간 Q&A 세션 또한 마련했습니다. 특히 BIML은 각 분야 국내 최고 전문가들의 강의로 구성되어 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것입니다.

이번 BIML-2023을 준비하기까지 너무나 많은 수고를 해주신 BIML-2023 운영위원회의 남진우, 우현구, 백대현, 정성원, 정인경, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 연사분께 깊은 감사를 드립니다.

2023년 2월

한국생명정보학회장 이 인 석

강의개요

질량분석을 활용한 단백체 연구

(이론) Mass Spectrometry-based Proteomics
(실습) Proteomics Data Analysis

인간의 DNA에 있는 유전자들은 발달 및 노화 과정 동안, 장기 및 조직의 위치에 따라, 그리고 외부 환경의 변화에 대응하여 단백질을 만들어 내고 있는 것은 매우 흥미로운 일이다. 이를 통해 다세포 생물의 하나인 인간 몸속의 수많은 세포가 각기 다른 일을 유기적으로 할 수 있는 것일 것이다. 우리는 많은 경우 유전적이지 않은 상황으로 인해 DNA의 변형을 맞이하고 이를 통해 세포에 병인이 발생하여 때로는 결국 죽음에 이르는 병을 얻게 된다.

최근 질량분석법을 기반으로 하는 단백질 집합(통칭 단백체, Proteome)에 대한 연구 기술이 급격히 발달하고 있으며 가까운 시일 내에 NGS 수준의 방대한 데이터 양을 생산하는 날이 머지 않았다. 그렇다면 우리는 언제 단백체 기반의 빅데이터 연구를 하게 될까? 나아가서 유전체와 단백체 통합에 대한 활발한 연구를 통해 생명체가 시간적, 그리고 공간적으로 어떻게 외부 환경에 반응하고, 어떻게 내부적으로 짜여진 프로그램을 영위해 나가는지 이해할 수 있을 것이다. 그러나 단백체 데이터는 유전체 데이터와는 상이한 방식으로 수집이 되고 이를 이해하는 방법은 매우 다르다. 본 강의에서는 질량분석에 대해 이해하고 단백체 데이터 수집 방식을 공부할 것이며 단백체 데이터 분석을 위해 사용되는 플랫폼에 대해 경험하고 데이터 처리에 대한 예를 다룰 것이다. 이를 통해 빅데이터를 빠르고 손쉽게 처리할 수 있는 핵심 역량을 갖추는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- Mass Spectrometry 개요와 단백체 실험의 개론
- MS 데이터 수집 방법 및 이해
- 데이터 처리 방법과 이해

* 참고강의교재: Min-Sik Kim et al. Nature 2014

* 교육생준비물: 노트북 (메모리 8GB 이상, 디스크 여유공간 30GB 이상)

* 강의 난이도: 초급

Curriculum Vitae

Speaker Name: Min-Sik Kim, Ph.D.

▶ **Personal Info**

Name	Min-Sik Kim
Title	Associate Professor
Affiliation	Department of New Biology, DGIST

▶ **Contact Information**

Address	DGIST, 333 TechnoJungang-daero, Dalseong-gun, Daegu
Email	mkim@dgist.ac.kr
Phone Number	053-785-1630

Research Interest

Mass Spectrometry, Proteomics, Systems Biology, Metabolomics, Multi-Omics

Educational Experience

2002	B.S. in Chemistry, Korea University, Korea
2004	M.S. in Physical Chemistry, Korea University, Korea
2013	Ph.D. in Biological Chemistry, Johns Hopkins University School of Medicine, USA

Professional Experience

2013-2016	Postdoctoral fellow, Institute of Genetic Medicine, Johns Hopkins University School of Medicine
2016-2018	Assistant Professor, Department of Applied Chemistry, Kyung Hee University
2018-present	Assistant, Associate Professor, Department of New Biology, DGIST

Selected Publications (5 maximum)

1. Hyeon, D. Y., Nam, D...., **Kim, M.-S**., ... Hwang, D., Lee, S.-W. (2022) Proteogenomic landscape of human pancreatic ductal adenocarcinoma in an Asian population reveals tumor cell-enriched and immune-rich subtypes. ***Nature Cancer***. Accepted.
2. **Jang, E. W.¶, Park, J. H.¶, ... Kim, M.-S.*** (2022) Cntnap2-dependent molecular networks in autism spectrum disorder revealed through an integrative multi-omics analysis. ***Molecular Psychiatry***. Accepted.
3. Cha, S.-J., **Kim, M.-S.**, Na, C. H., Jacobs-Lorena, M. (2021) Plasmodium sporozoite phospholipid scramblase interacts with mammalian carbamoyl-phosphate synthetase 1 to infect hepatocytes. ***Nature Communications***. 12(1):6773.
4. Park, J.-H., Ryu, S. J., ...,, **Lee, J. H., Park, J. H., ..., Kim, M.-S.***, Hwang, D.*, Lee, Y.-S.*, and Park, S. C.* (2021) Disruption of nucleocytoplasmic trafficking as a cellular senescence driver. ***Experimental & Molecular Medicine***. 53, 1092–1108.
5. Huh, S., Hwang, D.*, **Kim MS*** (2020) Statistical modeling for enhancing discovery power of citrullination from tandem mass spectrometry data. ***Analytical Chemistry***. 92, 19, 12975–12986.

KSBi-BIML

질량 분석을 활용한 단백체 연구

(이론) Mass Spectrometry-based Proteomics

약력

- 김민식, 이학박사
  - 1995-2002 고려대학교 화학과 이학 학사
  - 2002-2004 고려대학교 화학과 질량분석학 석사
  - 2007-2013 존스홉킨스 의과대학 생화학 박사
  - 2013-2016 존스홉킨스 유전체연구소 포스닥
  - 2016-2018 경희대학교 응용화학과 조교수
  - 2018-2020 DGIST 뉴바이올로지학과 조교수
  - 2020-현재 DGIST 뉴바이올로지학과 부교수

2

대구경북과학기술원(DGIST)

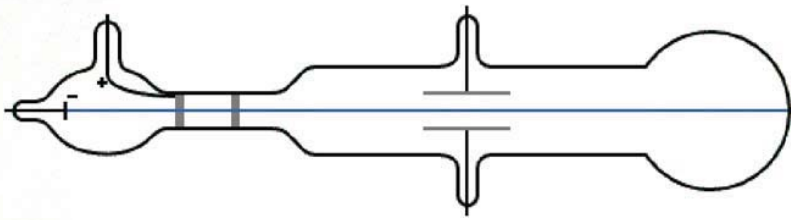
3

Mass Spectrometry

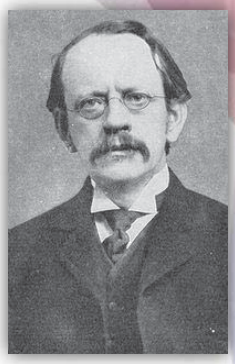
질량 분석기 개발의 기초 아이디어

Nobel Prize in Physics (1906)

"in recognition of the great merits of his theoretical and experimental investigations on the conduction of electricity by gases."



wikipedia



J. J. Thomson (1856~1940)

네온의 질량 스펙트럼

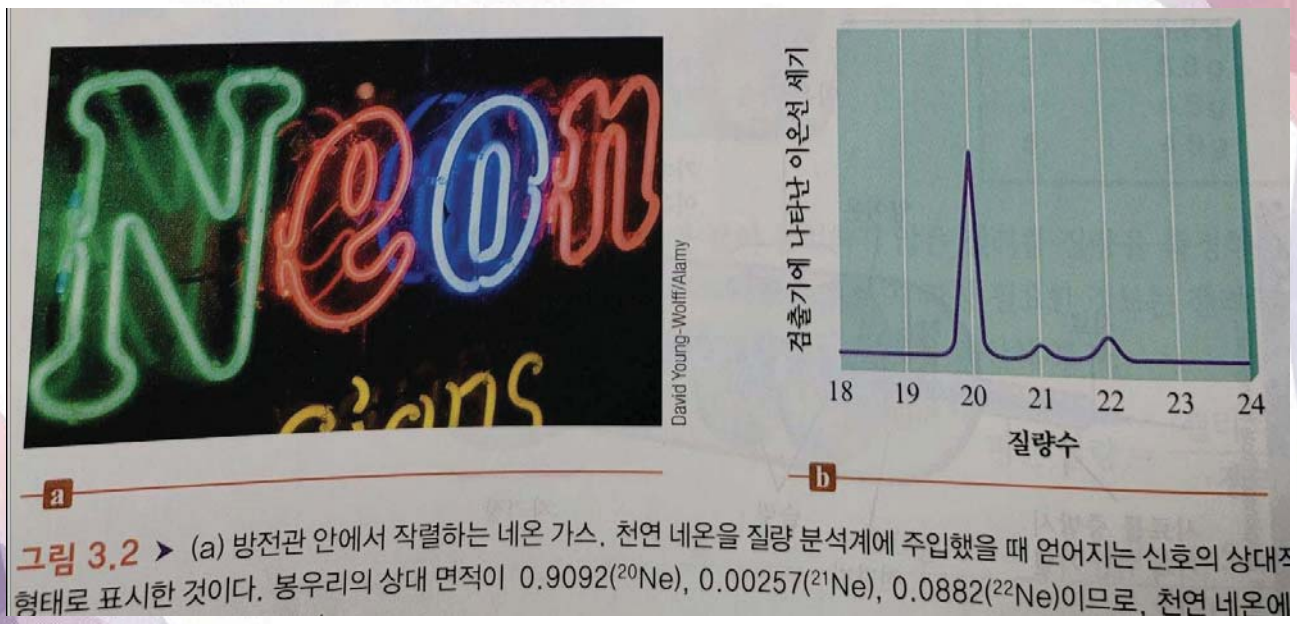
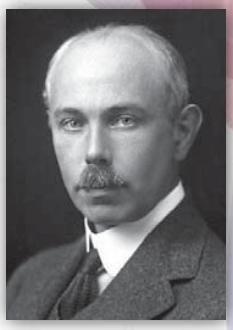


그림 3.2 ➤ (a) 방전관 안에서 작렬하는 네온 가스. 천연 네온을 질량 분석계에 주입했을 때 얻어지는 신호의 상대적 형태로 표시한 것이다. 봉우리의 상대 면적이 0.9092(^{20}Ne), 0.00257(^{21}Ne), 0.0882(^{22}Ne)이므로, 천연 네온에

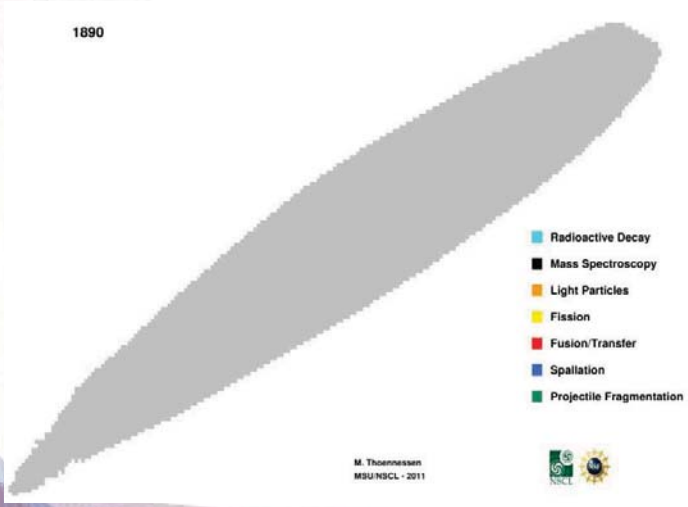
질량분석기의 개발과 안정한 동위원소의 발견

Nobel Prize in Chemistry (1922)

"for his discovery, by means of his mass spectrograph, of isotopes, in a large number of non-radioactive elements, and for his enunciation of the whole-number rule."

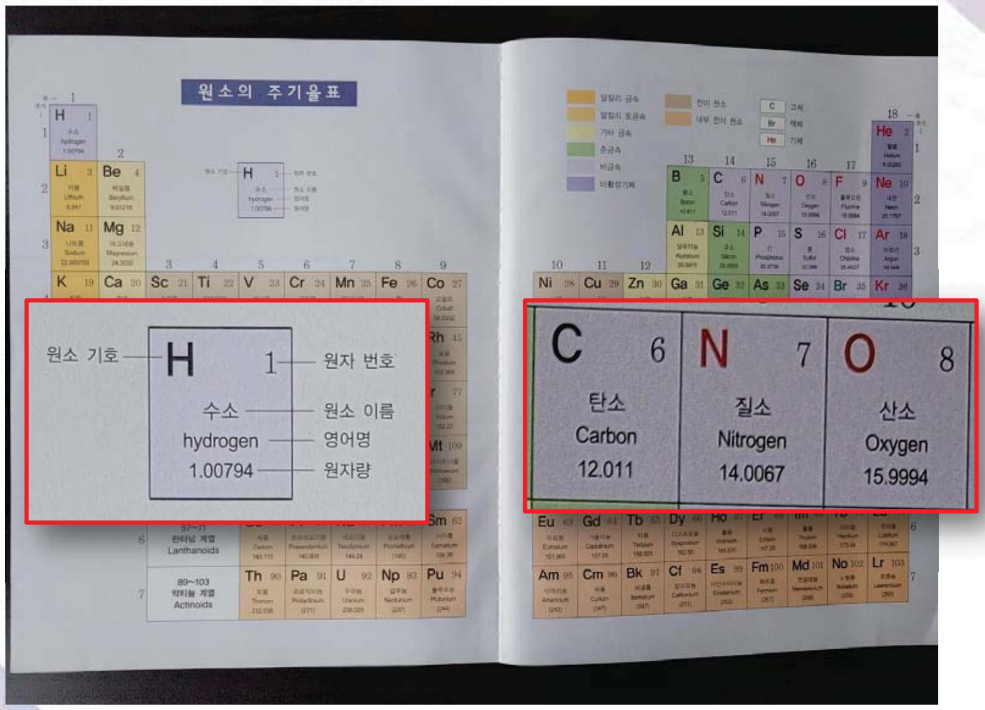


Francis Aston (1877~1945)



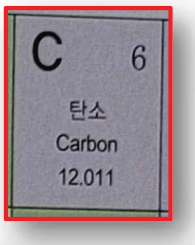
Identified
212 of the 287 naturally occurring isotopes

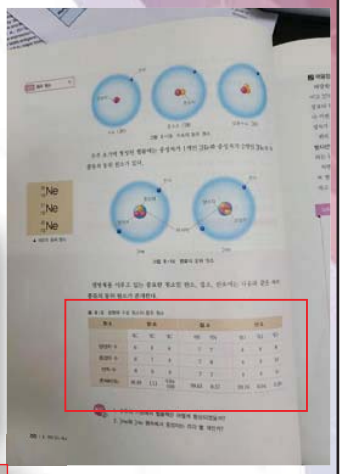
원자의 질량



탄소의 질량

- 탄소, carbon
  - (평균)원자량 12.011





교학사 화학 I

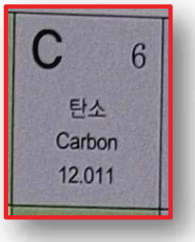
표 Ⅱ-3 생명체 구성 원소의 동위 원소

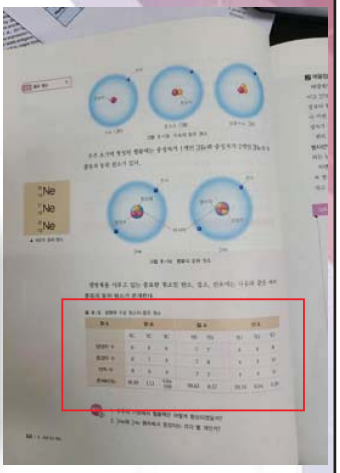
원소	탄소			질소		산소		
	$^{12}_{6}C$	$^{13}_{6}C$	$^{14}_{6}C$	$^{14}_{7}N$	$^{15}_{7}N$	$^{16}_{8}O$	$^{17}_{8}O$	$^{18}_{8}O$
양성자 수	6	6	6	7	7	8	8	8
중성자 수	6	7	8	7	8	8	9	10
전자 수	6	6	6	7	7	8	8	8
존재비(%)	98.89	1.11	0.0x 이하	99.63	0.37	99.76	0.04	0.20

9

탄소의 질량

- 탄소, carbon
  - (평균)원자량 12.011
  - ^{12}C - 98.89%
  - ^{13}C - 1.11%
  - ^{14}C - 거의 없음





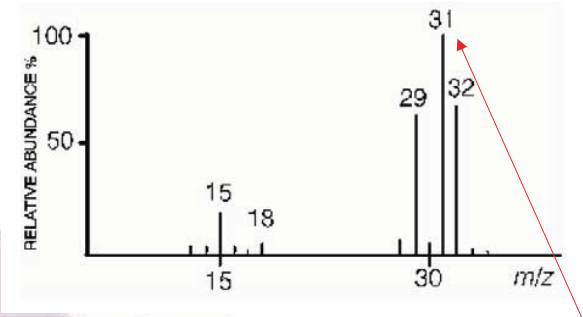
교학사 화학 I

탄소 원자의 (평균)원자량 = 12 X 0.9889 + 13 X 0.0111
= 12.0111

10

질량 스펙트럼

Spectrum



mass-to-charge

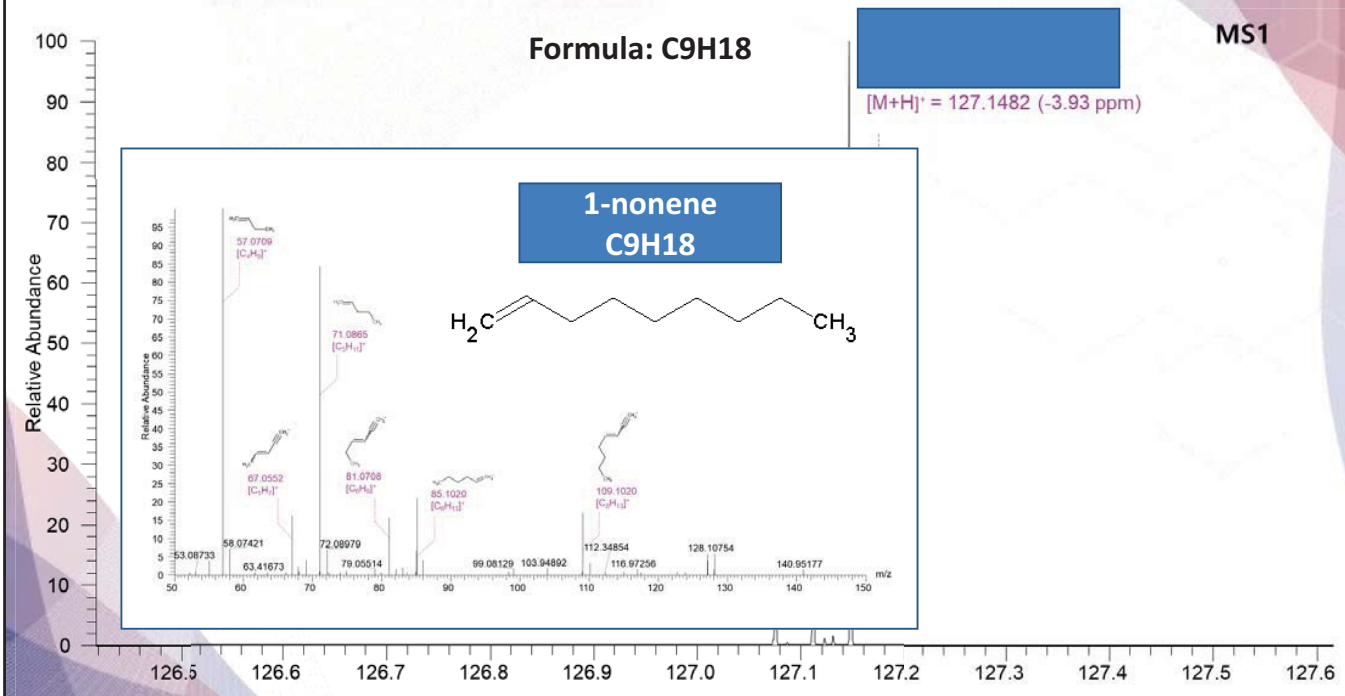
Mass table

m/z	Relative abundance (%)	m/z	Relative abundance (%)
12	0.33	28	6.3
13	0.72	29	64
14	2.4	30	3.8
15	13	31	100
16	0.21	32	66
17	1.0	33	0.73
18	0.9	34	~ 0.1

base peak

11

저분자 물질의 정성분석



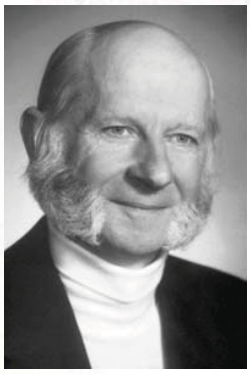
12

Unpublished data

기체 이온의 공간적 트랩핑 기술 개발

Nobel Prize in Physics (1989)

"for the development of the ion trap technique."



Hans G. Dehmelt
(1913~2017)

Penning ion trap

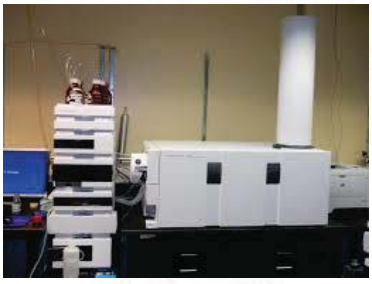
Wolfgang Paul
(1913~1993)

Paul ion trap

www.youtube.com

13

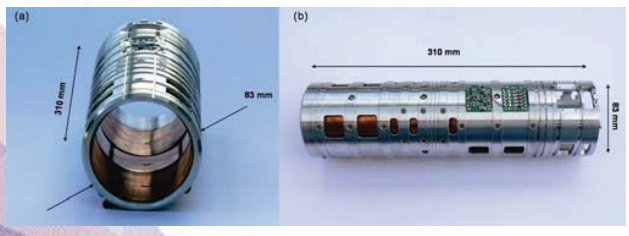
다양한 질량분석기



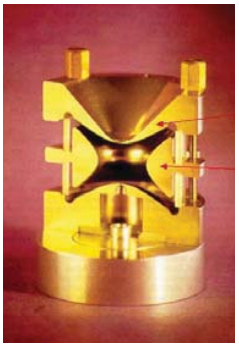
Time-of-Flight

Quadrupole

Orbitrap

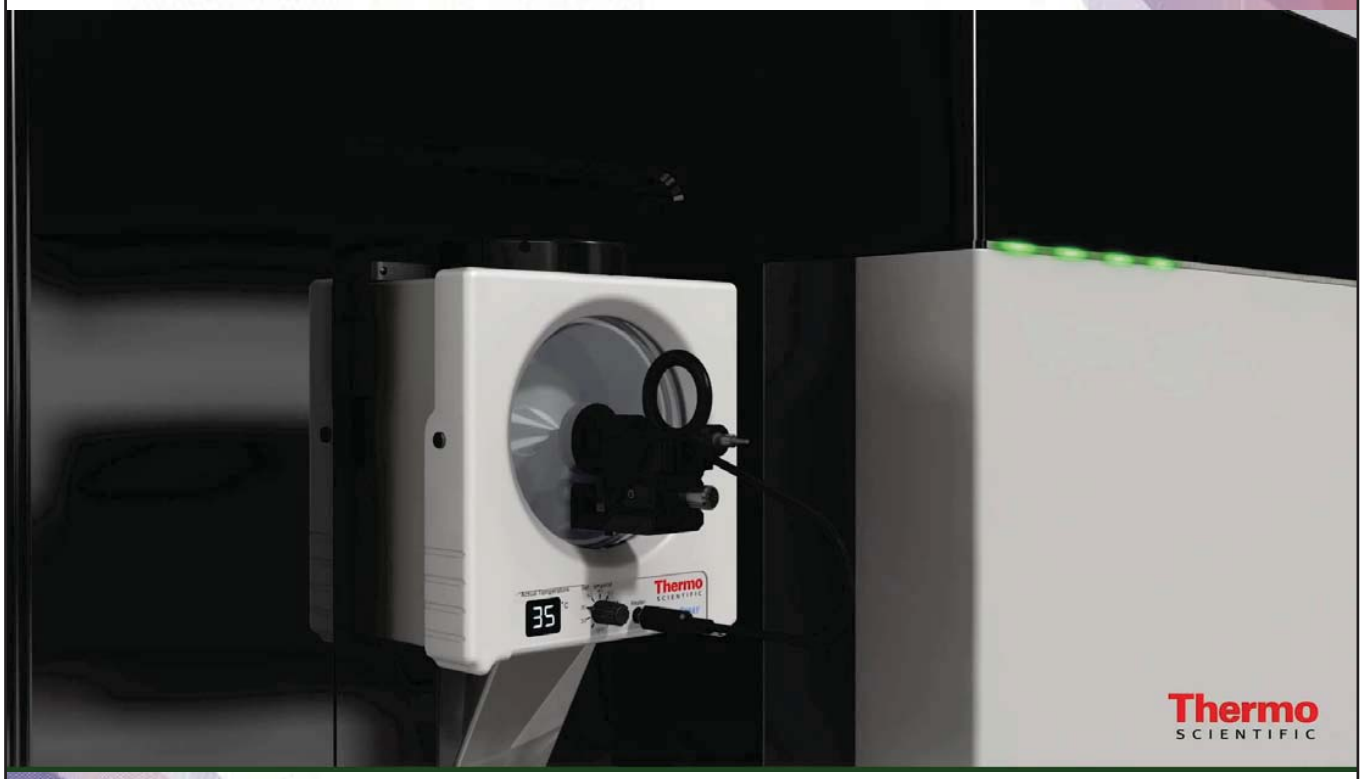


FT-ICR



Ion trap

14



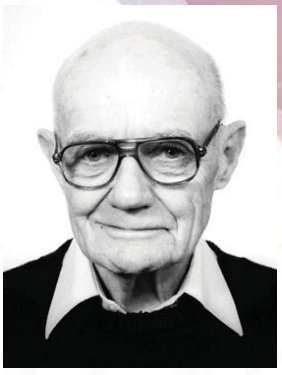
15

저에너지 이온화 기술 개발과 바이오 고분자 분석

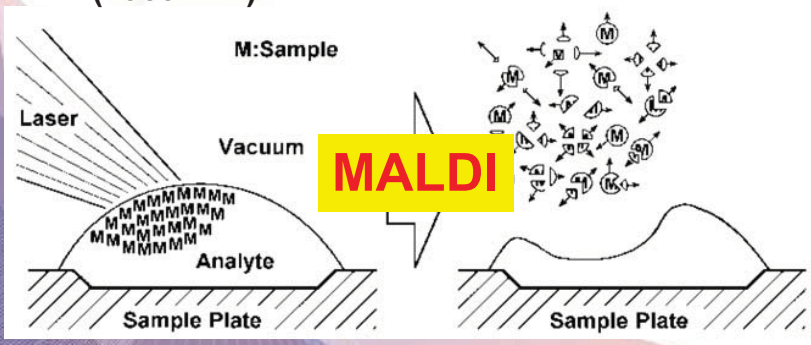
Koichi Tanaka
(1959 -)

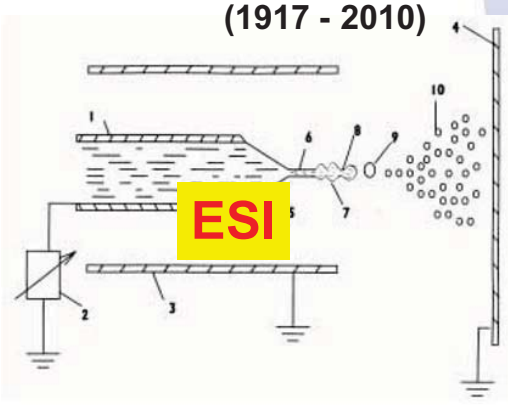
Nobel Prize in Chemistry (2002)

"for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules"

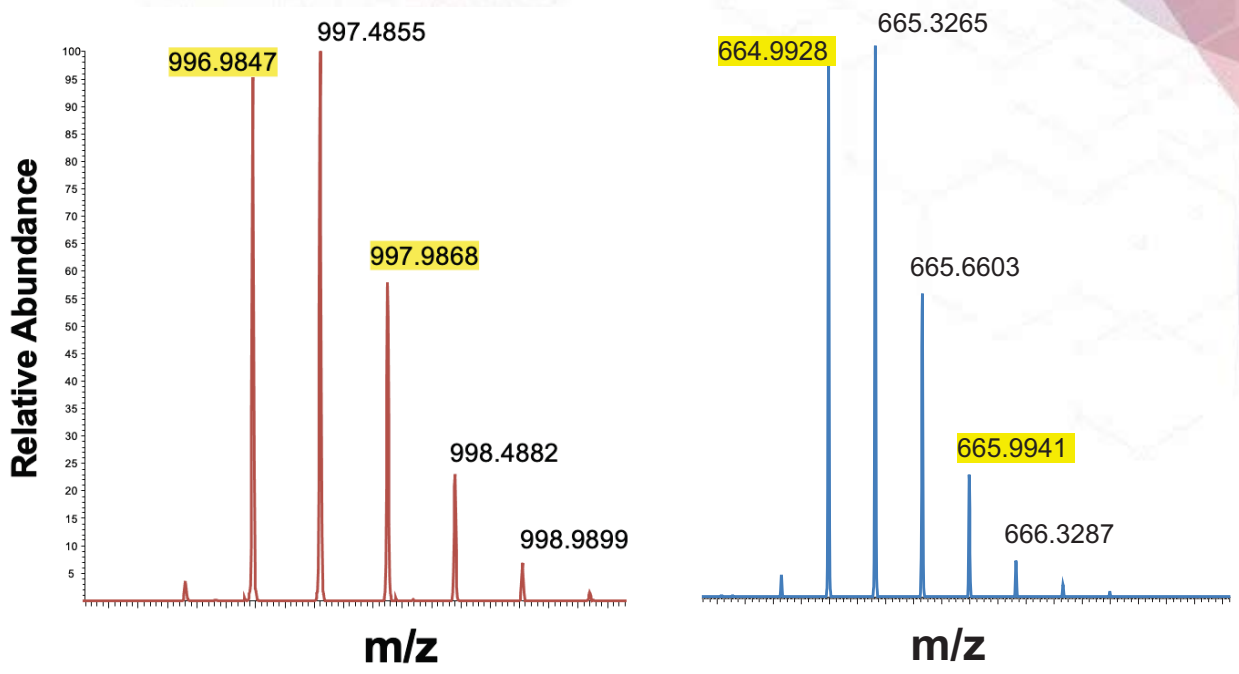


John B. Fenn
(1917 - 2010)



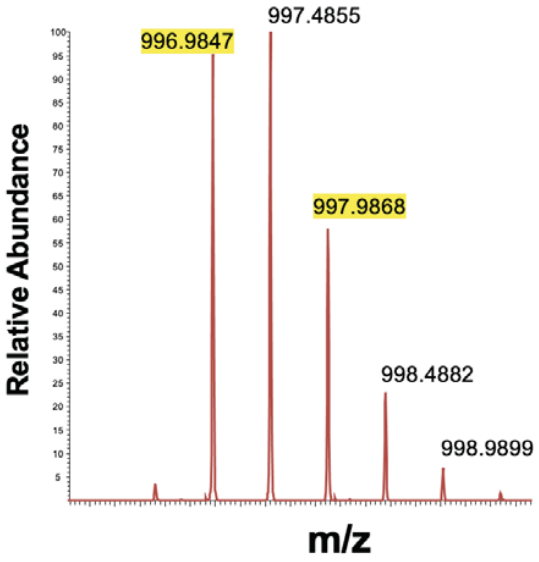


질량스펙트럼의 특성 – isotopic cluster (동위원소분포클러스터)



17

질량스펙트럼의 특성 – isotopic cluster (동위원소분포클러스터)

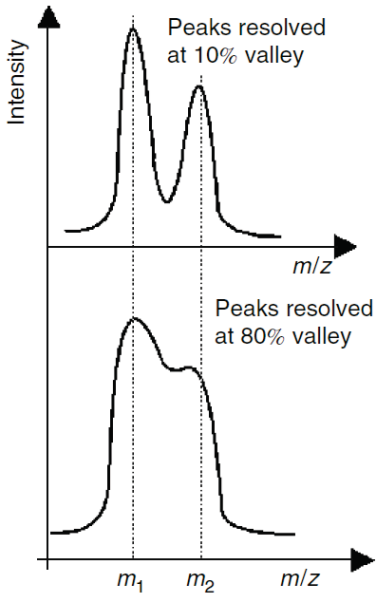


m/z = 996.9847
Charge (z) = +2

→(m) = 996.9847 x (+2) = Peptide + $2H^+$
→ precursor mass
= 996.9847 x (+2) – 2 x 1.0078
= 1991.9538

18

질량스펙트럼의 특성 – resolution(분해능)



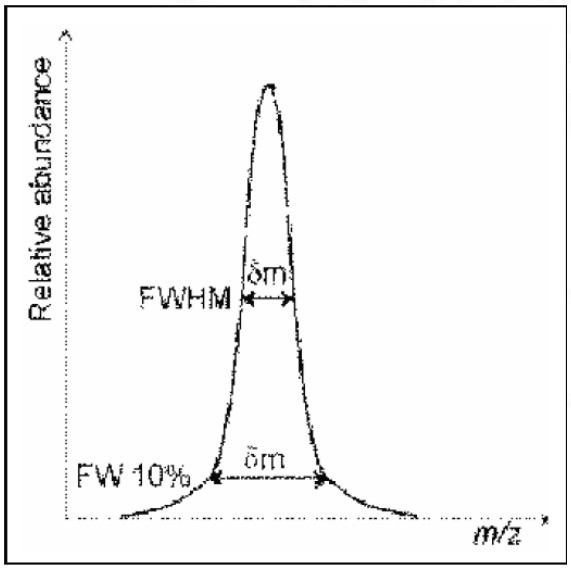
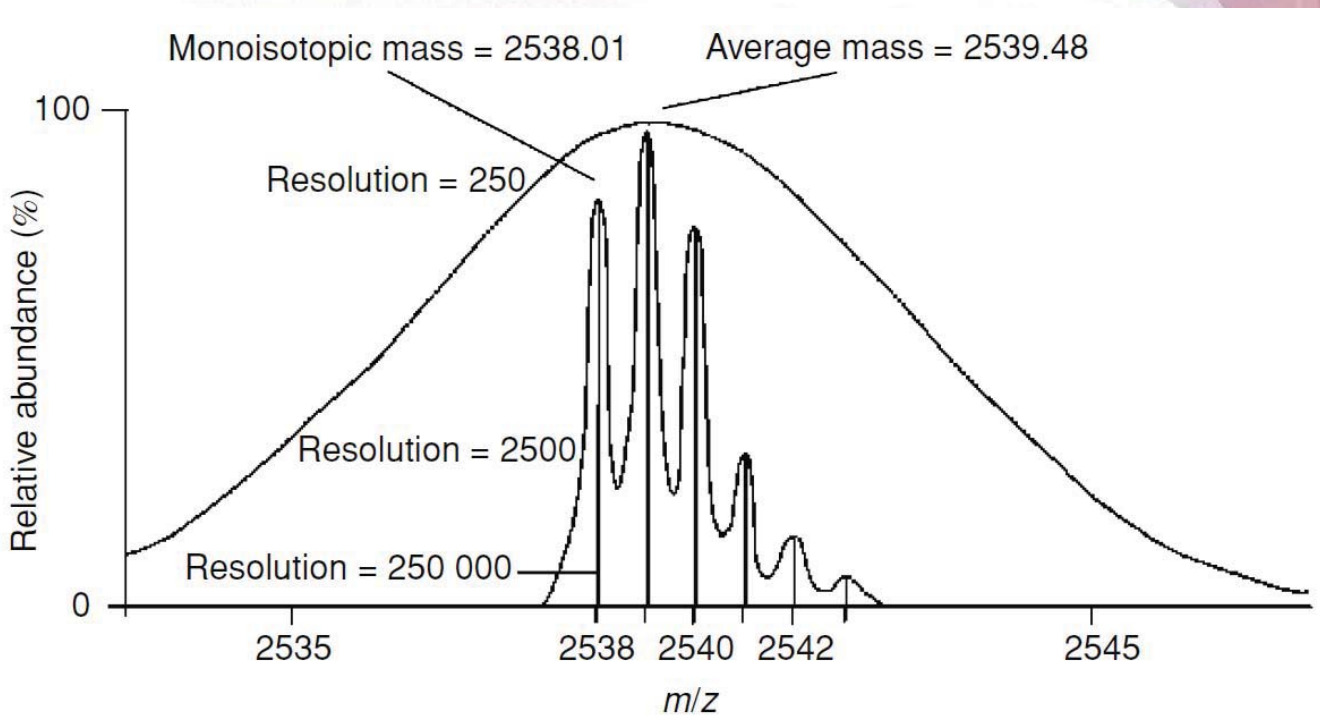


Figure 2.1
Diagram showing the concepts of peak resolution and valley.

19

질량스펙트럼의 특성 – resolution(분해능)



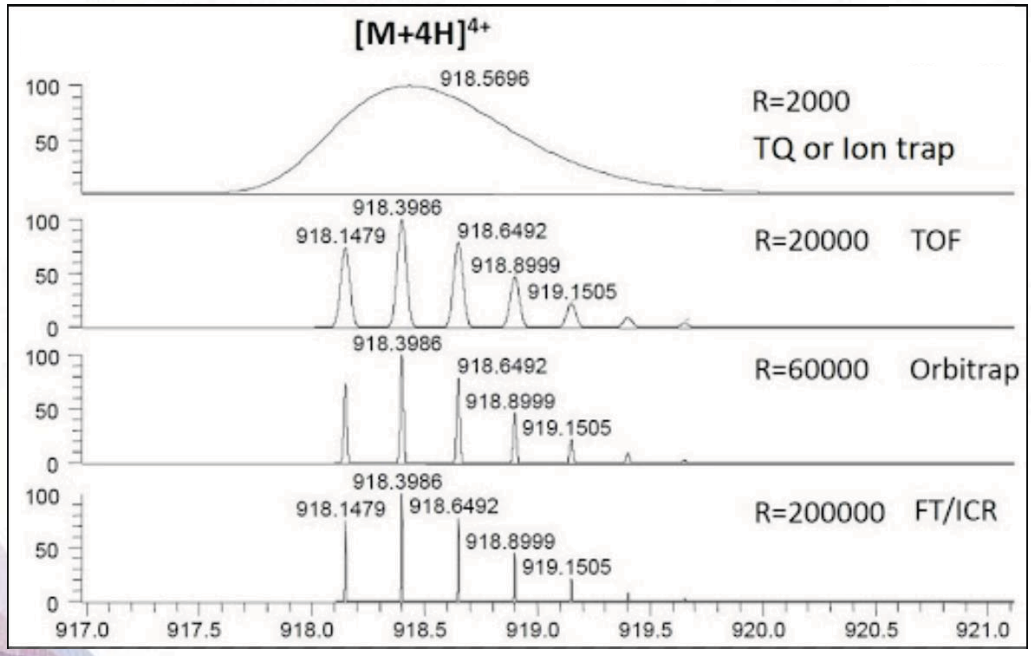
20

분해능의 발전

	$m/\delta m$	
1913	13	Thomson
1918	100	Dempster
1919	130	Aston
1937	2000	Aston
1998	8 000 000	Marshall and co-workers

21

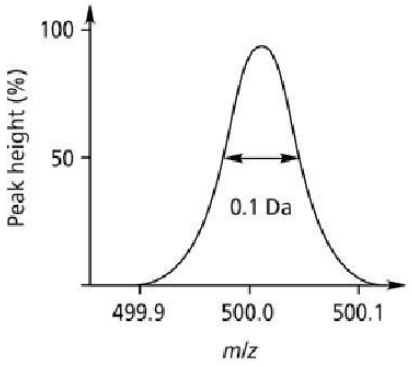
질량분석기별 질량분석 스펙트럼 분해능 차이



22

질량분석스펙트럼의 특성 - mass measurement accuracy(질량측정정확도)

True mass = 400.0000
Measured mass = 400.0020
Difference = 0.0020 or 2 mmu
Error = $\frac{0.002}{400} \times 10^6 = 5$ ppm

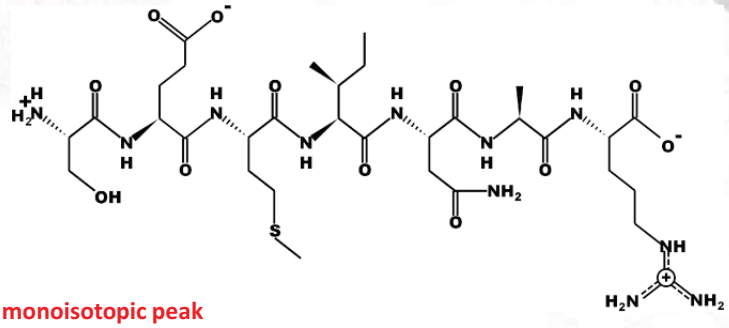


Mass = 500
Peak width (at 50%) = 0.1
Resolution (FWHM) $\frac{500}{0.1} = 5000$

23

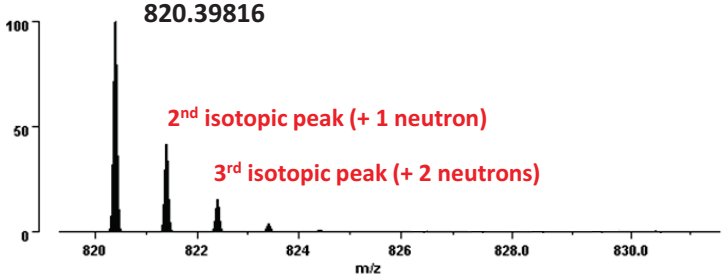
Peptide Sequencing

Example peptide



monoisotopic peak

820.39816

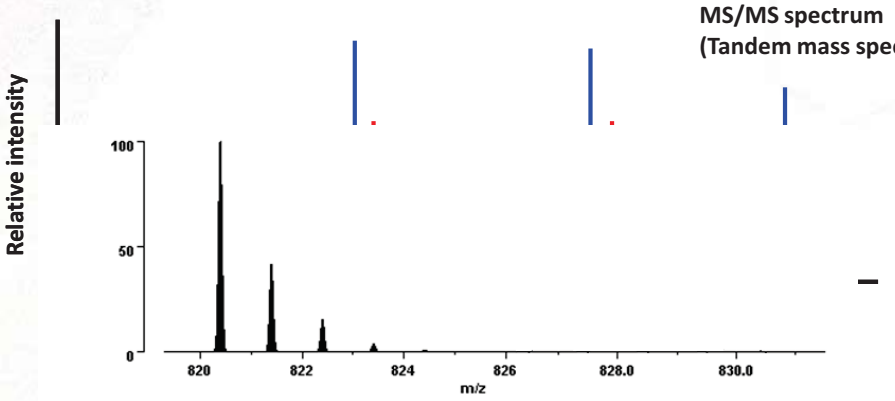


2nd isotopic peak (+ 1 neutron)

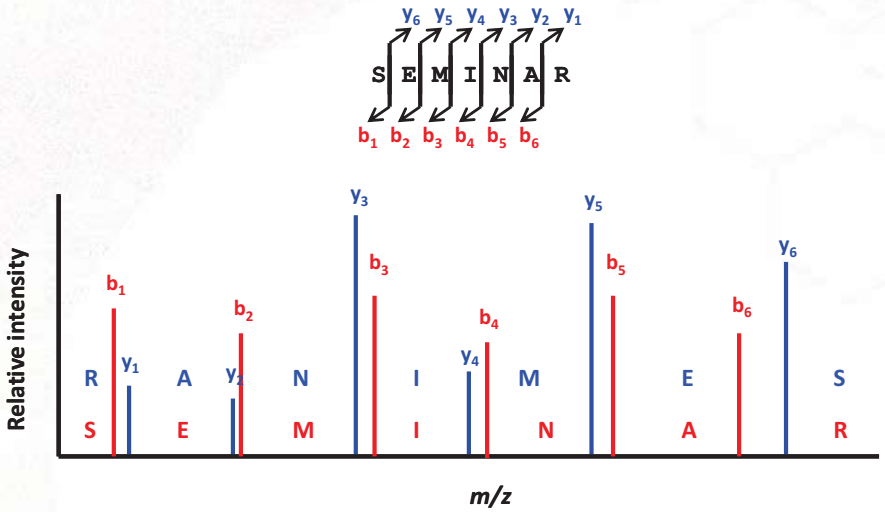
3rd isotopic peak (+ 2 neutrons)

Gas phase dissociation

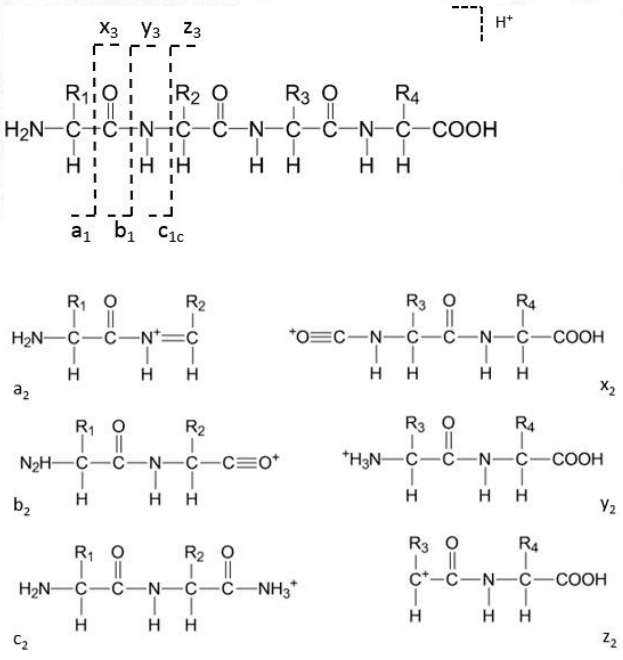
MS/MS spectrum
(Tandem mass spectrum)



Fragment ion assignments

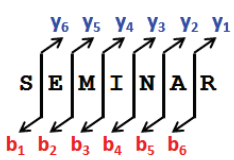


Nomenclature of peptide fragment ions

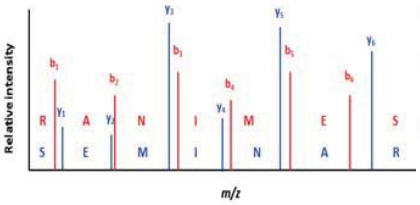


PSM (Peptide-Spectrum Match)

Theoretical precursor mass



Observed precursor mass



Theoretical fragment ion mass list

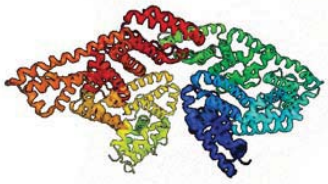
217.0819
348.1224
461.2064
575.2494
646.2865

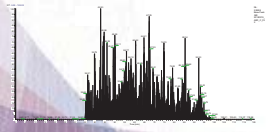
SCORE

Measured fragment ion mass list

217.0823
348.1211
461.2072
575.2490
646.2875

Bottom-Up Proteomics





MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIA
FSQYLQQCPFDEHVK**LVNELTEFAK**TCVADESHAGCEKSLHTLFGDELCK
VASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPK**LKPDPNTLCDEF**
KADEKKFWGKYLYEIARR**HPYFYAPELLYYANK**YNGVFQECCQAEDKGAC
LLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAE
FVEVTK**LVTDLTK**VHK**ECCHGDLLECADDR**ADLAKYICDNQDTISSKLKE
CCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDK**DVCKNYQEAK**DAFL
GSFLYEYSRR**HPEYAVSVLLRLAK**EYEATLEECCAKDDPHACYSTVFDKL
KHLVDEPQNLIKQNCDQFEKLGEYGFQNALIVR**YTRKVPQVSTPTLVEVS**
RSLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTKCC
TESLVNRRPCFSALTPDETYVPKAFDEK**LFTFHADICTLPDTEK**QIKKQT
ALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
STQTALA

ECCHGDLLECADDR

LVNELTEFAK

DVCKNYQEAK

HPYFYAPELLYYANK

HPEYAVSVLLRLAK

LFTFHADICTLPDTEK

Data analysis

Top-Down Proteomics

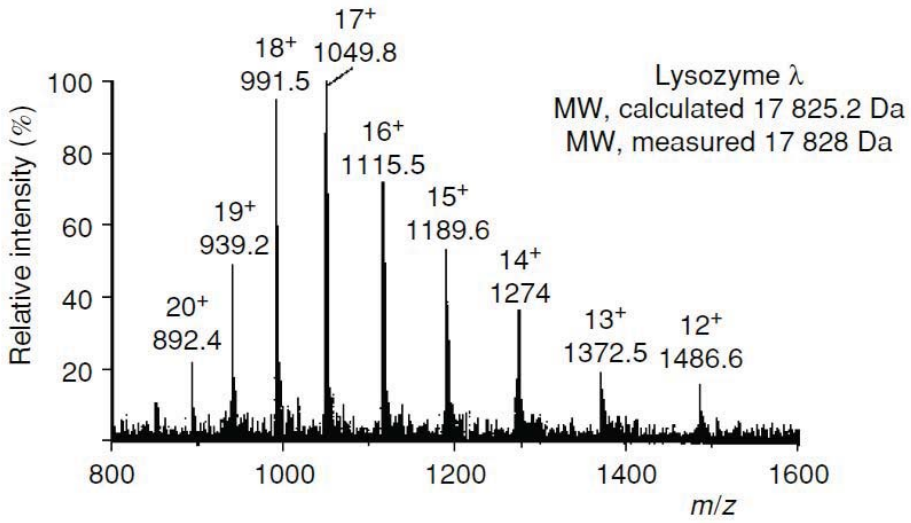
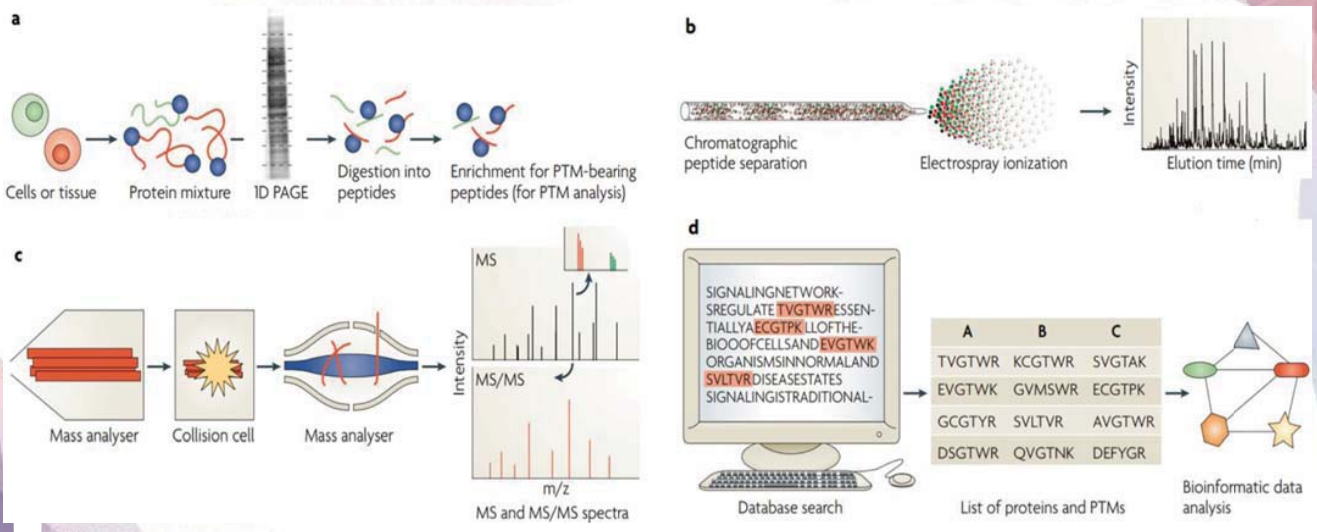


Figure 1.23
ESI spectrum of phage λ lysozyme; *m/z* in Th and the number of charges are indicated on each peak. The molecular mass is measured as being 17 828 ± 2.0 Da.

31

General Proteomics Workflow



Choudhary and Mann, *Nat Rev Mol Cell Biol*, 2010

32

Quantitation

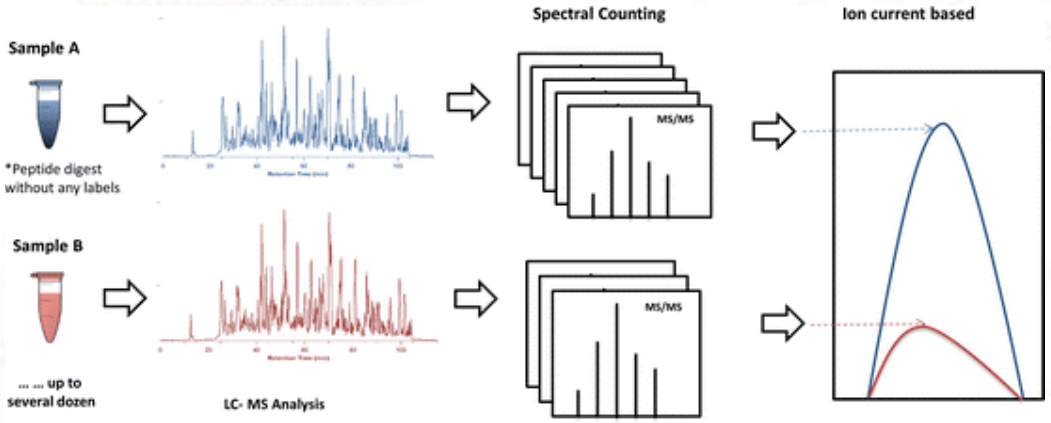
Choice of Quantitation in Proteomics

- Label
  - Metabolic labeling
  - Chemical labeling
- Label-free
  - # spectral counting (ex. # PSM)
  - LC profile

- Model samples (ex, cell)
- Clinical samples (ex, blood)

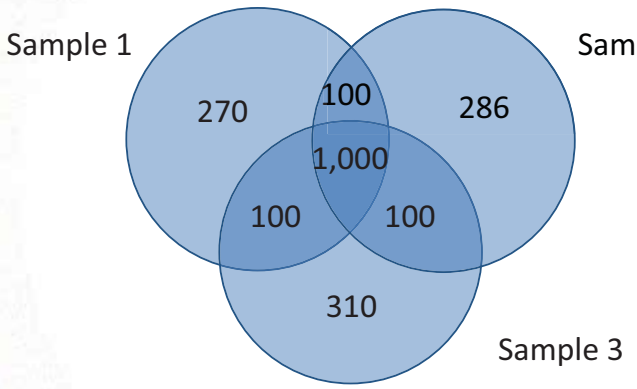
34

Label-free (spectral counting or ion currnet area)



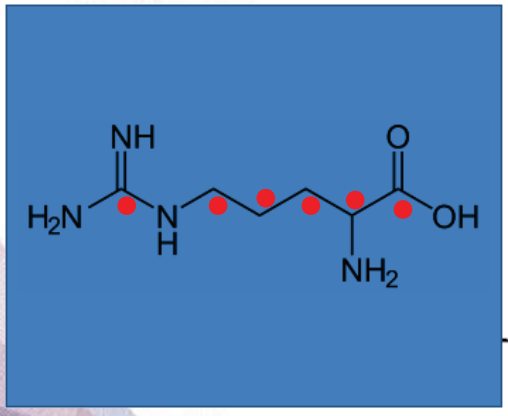
Label-free (spectral counting or ion currnet area)

- Sample 1 : 10 ug peptides -> 1,000,000 MS/MS -> 400,000 PSMs
- Sample 2 : 10 ug peptides -> 1,100,000 MS/MS -> 380,000 PSMs
- Sample 3 : 10 ug peptides -> 950,000 MS/MS -> 390,000 PSMs

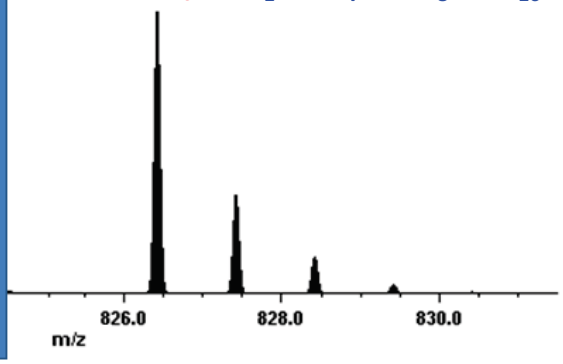


Metabolic labeling

SEMINAR*



$^{13}C_6$ $^{1}H_1$ $^{14}N_7$ $^{16}O_8$ $^{32}S_{16}$



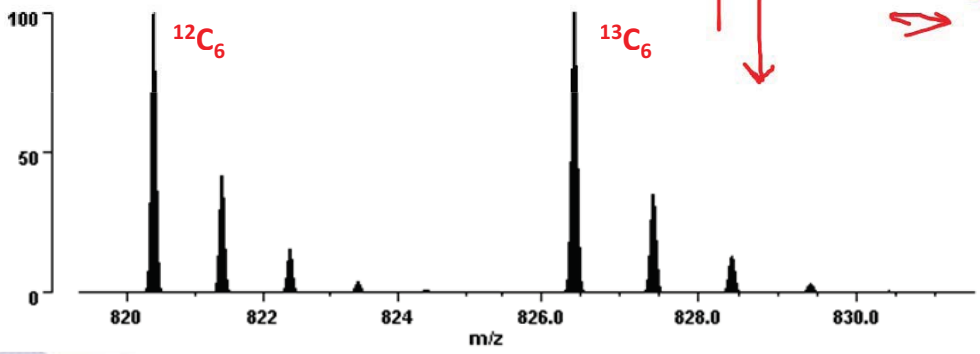
37

SILAC-based Quantitative proteomics

SEMINAR:**SEMINAR*** = 1:1

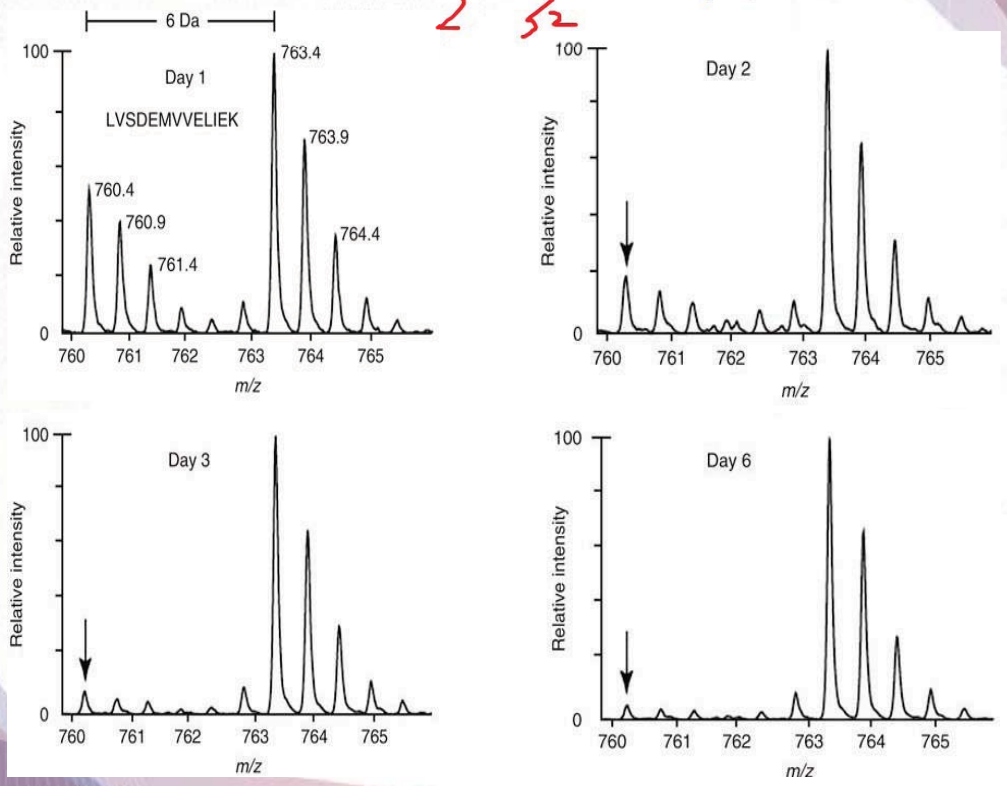
SEMINAR

SEMINAR*



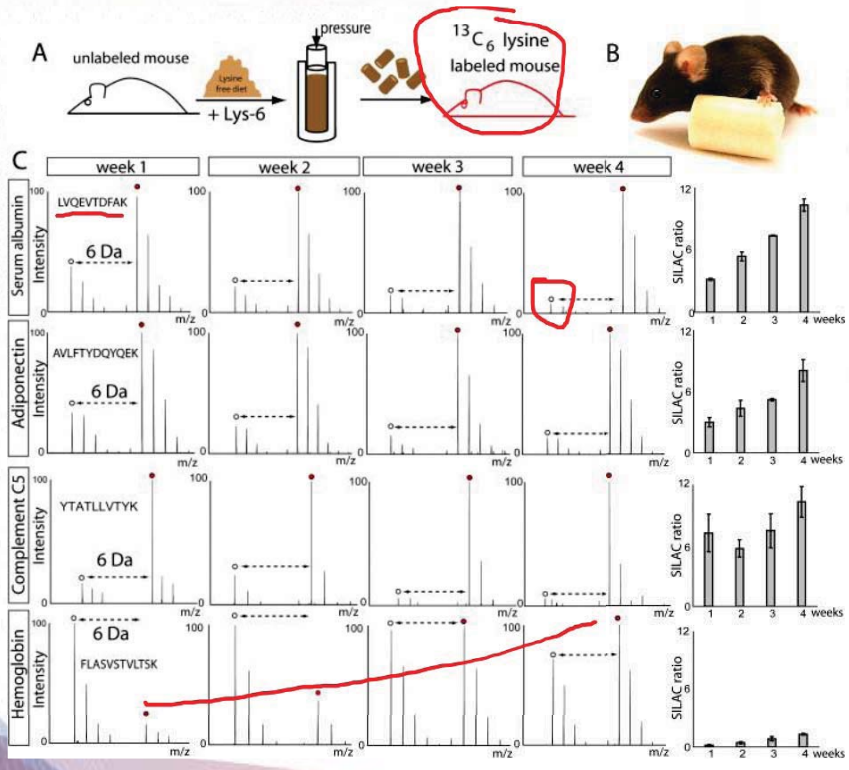
38

Metabolic labeling of the whole cellular proteome

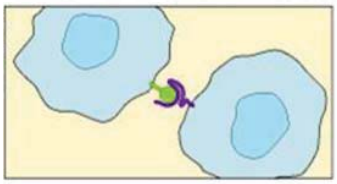


Gowda et al. *Nature Protocol*

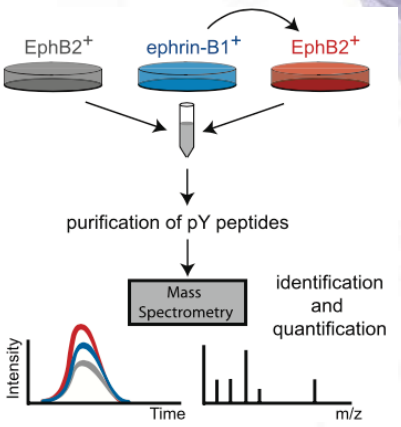
In vivo SILAC labeling

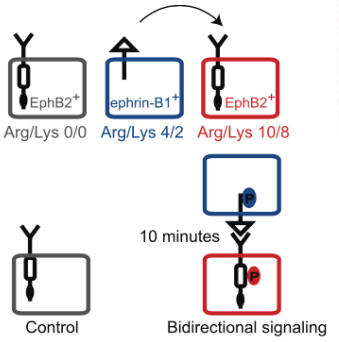


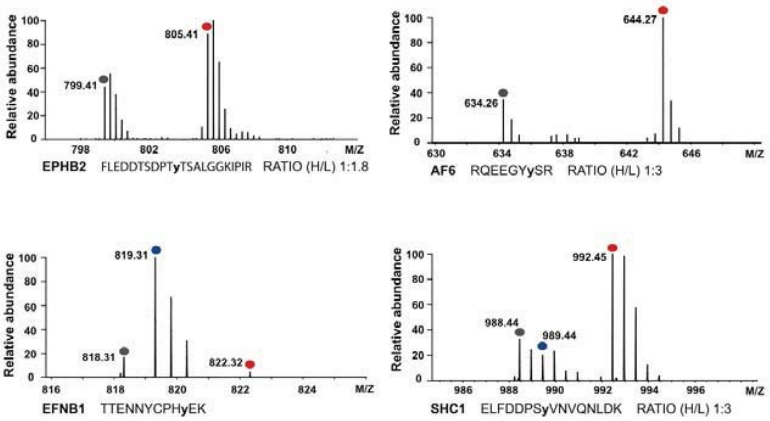
Mann et al. *Cell*

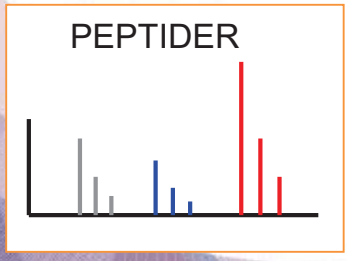


Juxtacrine



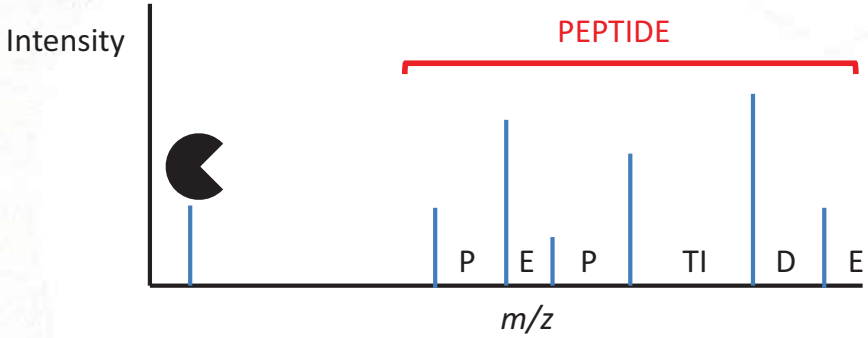




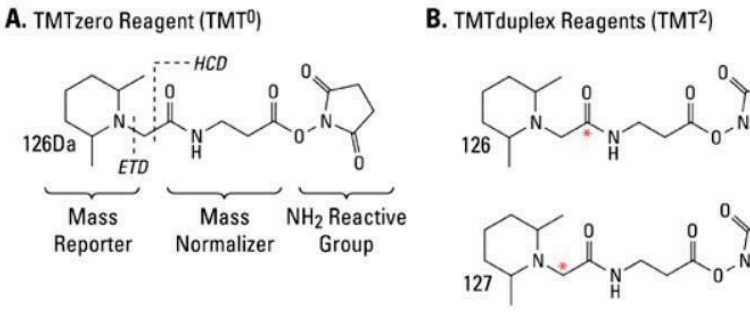


Jyrgensen et al. *Science*

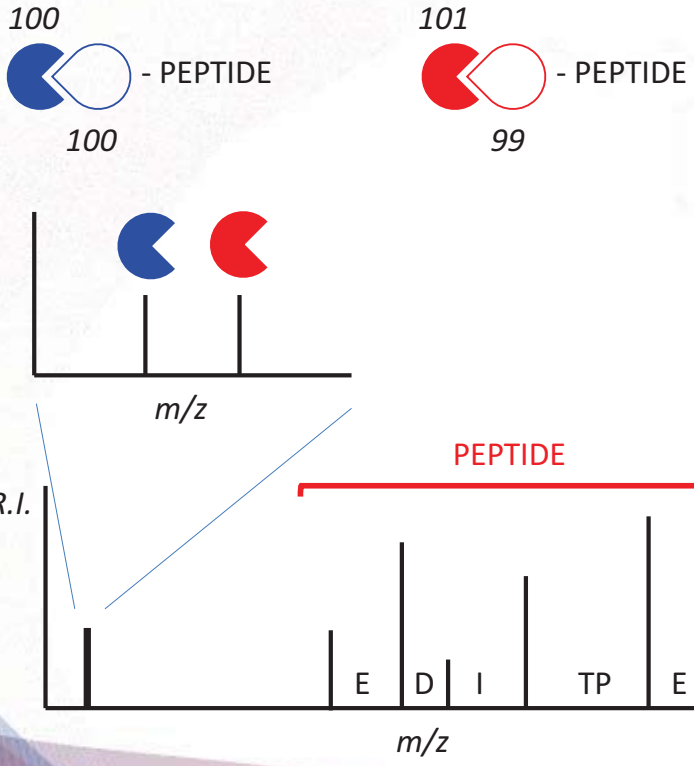
Isotope-coded chemical labeling to clinical samples



TMT-based quantitative proteomics



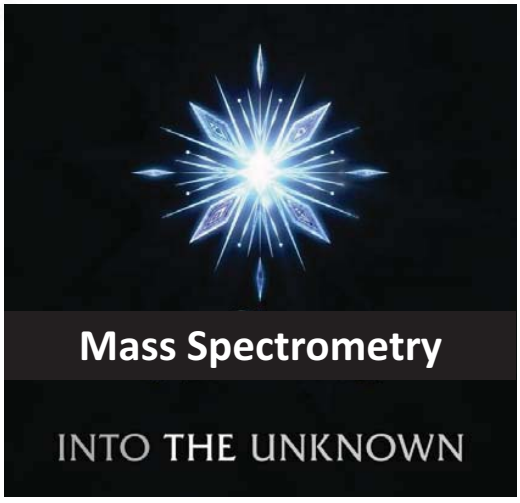
TMT-based quantitative proteomics



Summary

- 질량분석의 기초 원리
- 질량분석 데이터의 핵심 요소
- Peptide sequencing 의 기초 원리
- Peptide 정량 기술

- Principle of Mass Spectrometry and Basics of Proteomics
- Applications to different research fields



Laboratory for QBIO and Precision Medicine (큐바이오 정밀의학 연구실)

단백체, 멀티오믹스 생명정보학
포스닥, 석/박사과정생 모집 중

(김민식, mkim@dgist.ac.kr)

47

KSBi 한국생명정보학회

KSBi-BIML 2022

질량 분석을 활용한 단백체 연구

(실습) Proteomics Data Analysis

약력

- 김민식, 이학박사
  - 1995-2002 고려대학교 화학과 이학 학사
  - 2002-2004 고려대학교 화학과 질량분석학 석사
  - 2007-2013 존스홉킨스 의과대학 생화학 박사
  - 2013-2016 존스홉킨스 유전체연구소 포스닥
  - 2016-2018 경희대학교 응용화학과 조교수
  - 2018-2020 DGIST 뉴바이올로지학과 조교수
  - 2020-현재 DGIST 뉴바이올로지학과 부교수

2

대구경북과학기술원(DGIST)

3

Proteomics data

MaxQuant

https://www.maxquant.org/

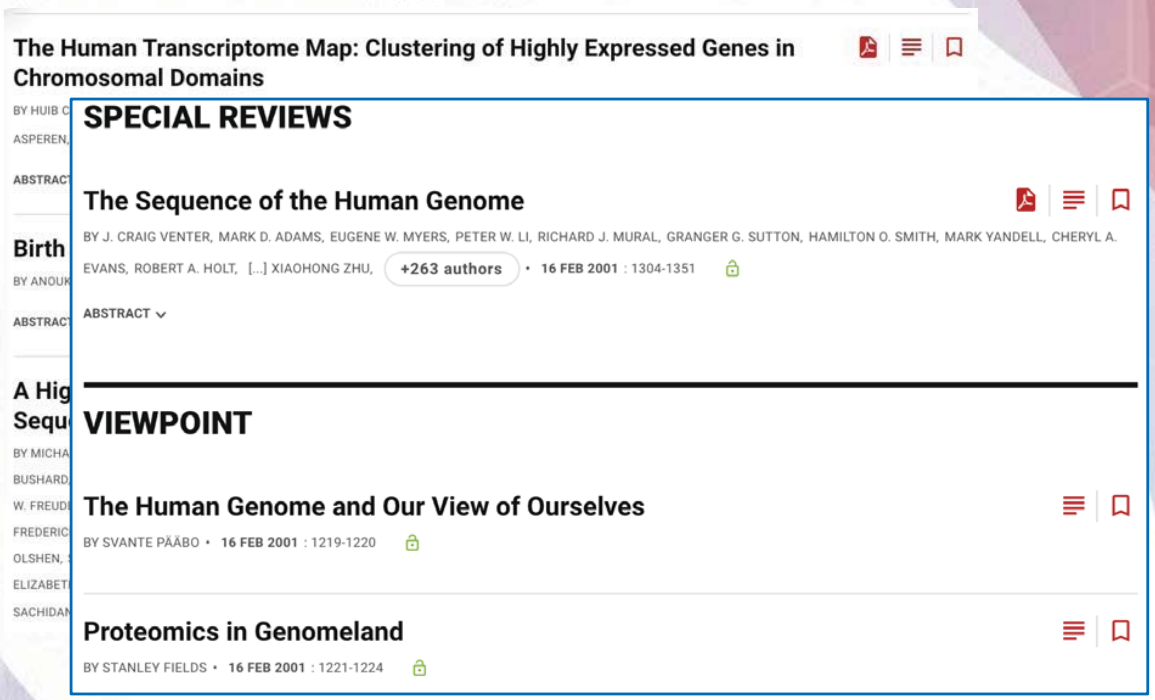
5

Human Proteome Map and Genome Annotation

The Human Genome

2

- The Human Transcriptome
- The draft human genome map

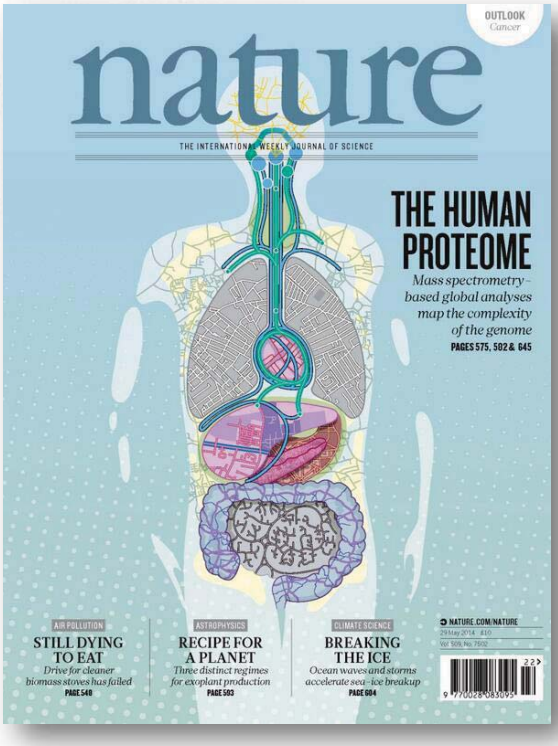


3

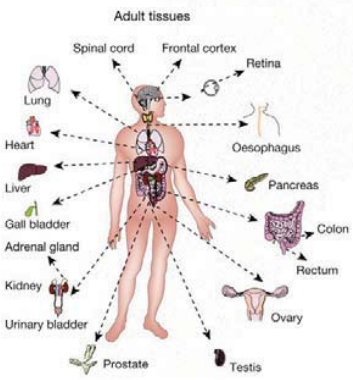
CREDIT: JOE SUTLIFF

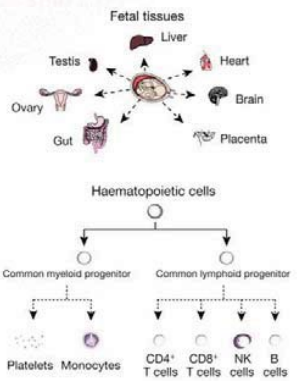
The shift in thinking from genomics to proteomics comes with an appreciation of the difficulty of the task: Proteins are much more complicated than nucleic acids. Unlike the decoratively challenged DNA, proteins get phosphorylated, glycosylated, acetylated, ubiquitinated, farnesylated, sulphated, linked to glycophosphatidylinositol anchors, and embellished in numerous other ways. A single gene can encode multiple different proteins—these can be produced by alternative splicing of the mRNA transcript, by varying translation start or stop sites, or by frameshifting during which a different set of triplet codons in the mRNA is translated. All of these possibilities result in a proteome estimated to be an order of magnitude more complex than the genome. (So it may be fortunate for proteomicists that humans might have as few as six times the number of genes that yeast have!) What is more, proteins respond to altered conditions by changing their location within the cell, getting cleaved into pieces, and adjusting their stability as well as changing what they bind to (other proteins, nucleic acids, lipids, small molecules, or other ligands). Protein levels often do not reflect mRNA levels (1), and even the presence of an open reading frame does not guarantee the existence of a protein. Lastly, a single protein may be involved in more than one

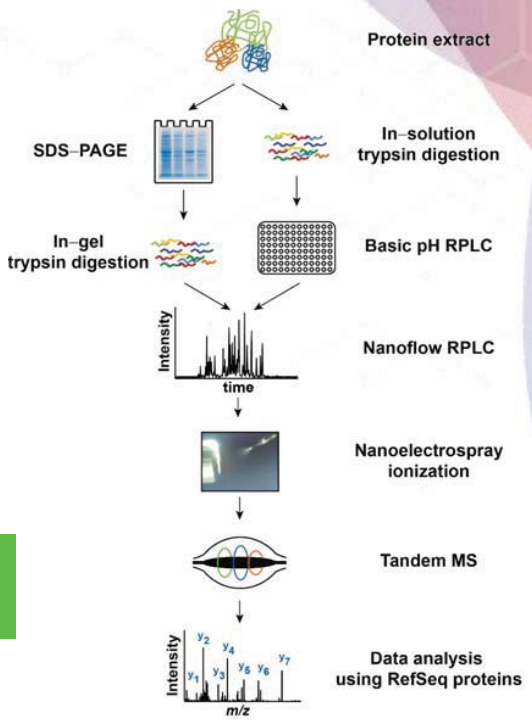
A draft map of the human proteome



Sampling and workflow

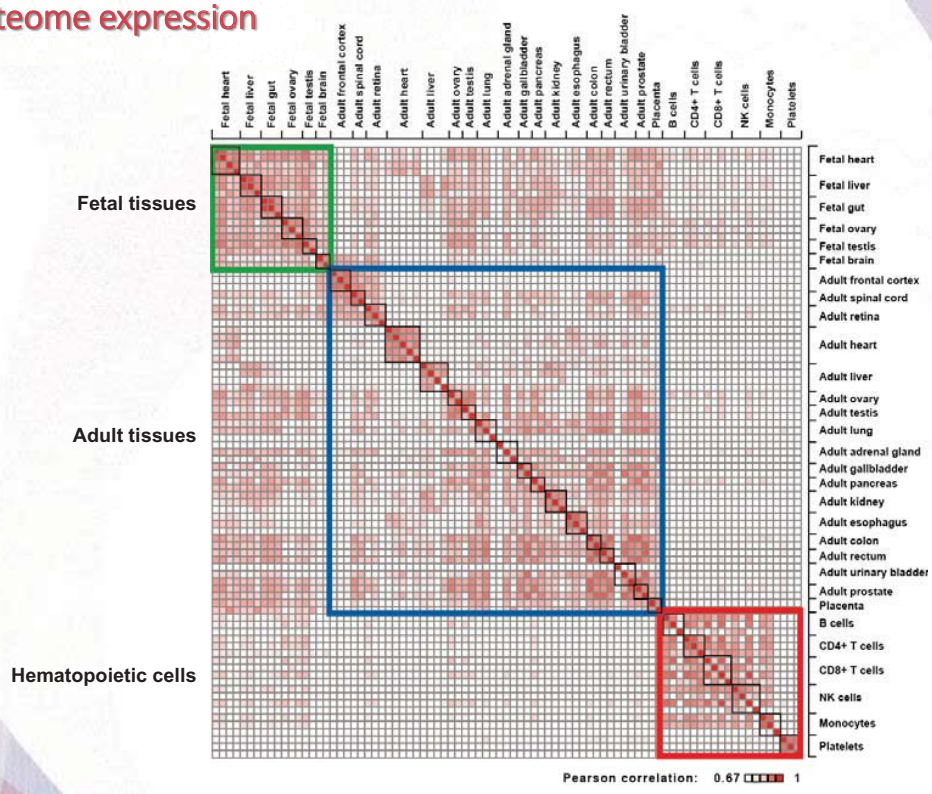




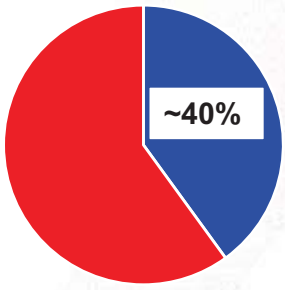


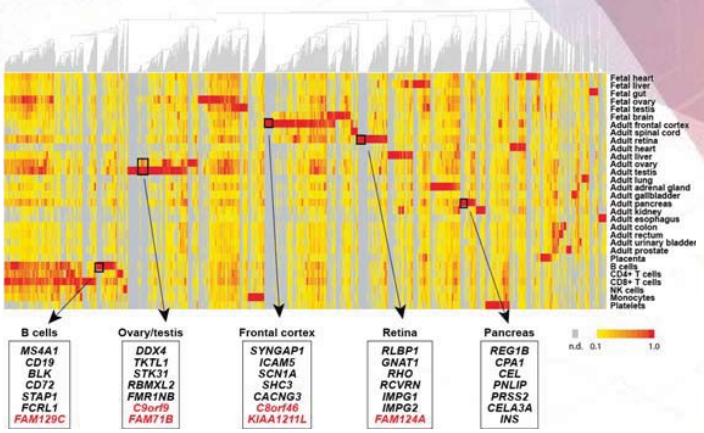
25 millions of MS/MS acquired with >2,000 LC/MS runs

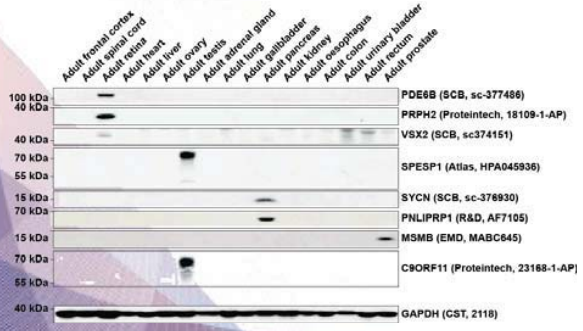
Proteome expression

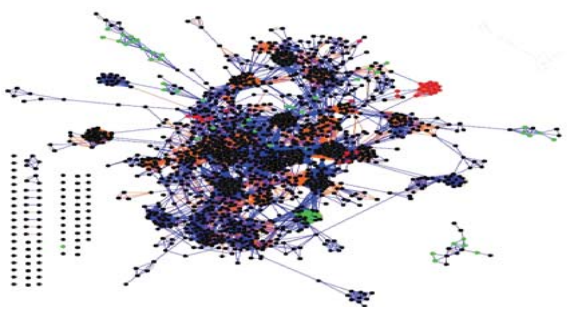


Tissue-specific expression









Modified from Kim et al. Nature 2014

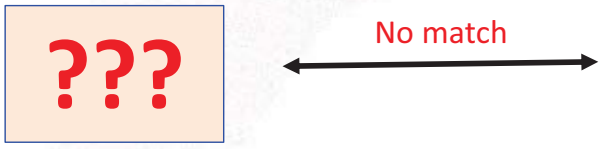
Current limitation of proteome analysis

- **The peptide sequencing is fully based on a protein sequence database**
- **Currently most of protein sequencing analyses are based on the bottom-up approach instead of analyzing intact proteins**

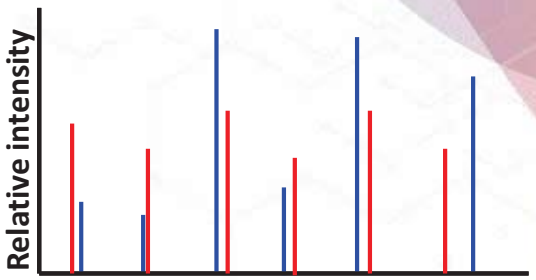
✓ **What if database is inaccurate?**

9

A good tandem mass spectrum remains unmatched!



No match



Theoretical fragment ion mass list

No data

No score

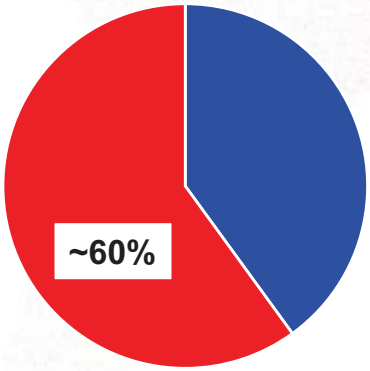
Measured fragment ion mass list

217.0823
348.1211
461.2072
575.2490
646.2875

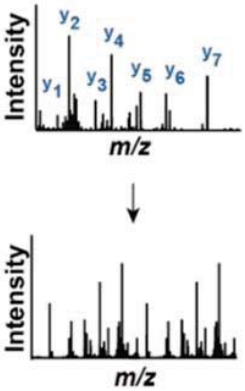
Measure by MS

10

~60% of unidentified MS/MS collected for proteogenomic analysis

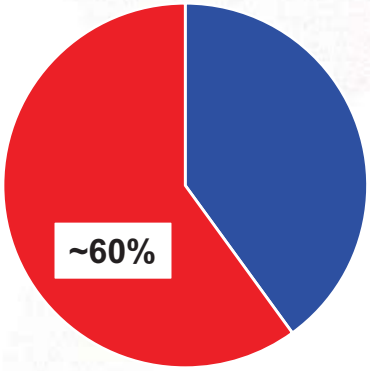


MS/MS spectra with NO identification

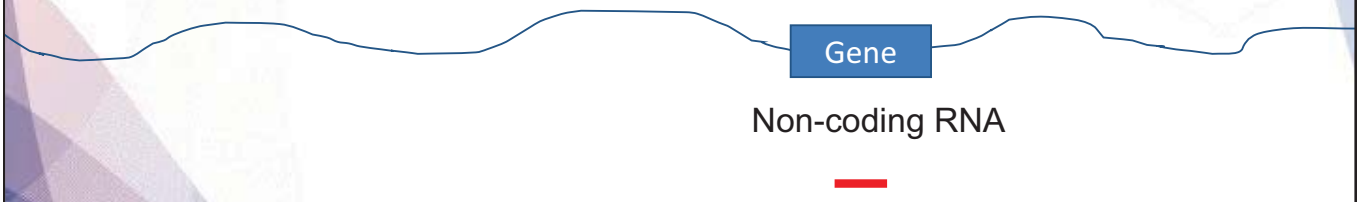


Collection of 'unidentified' MS/MS spectra

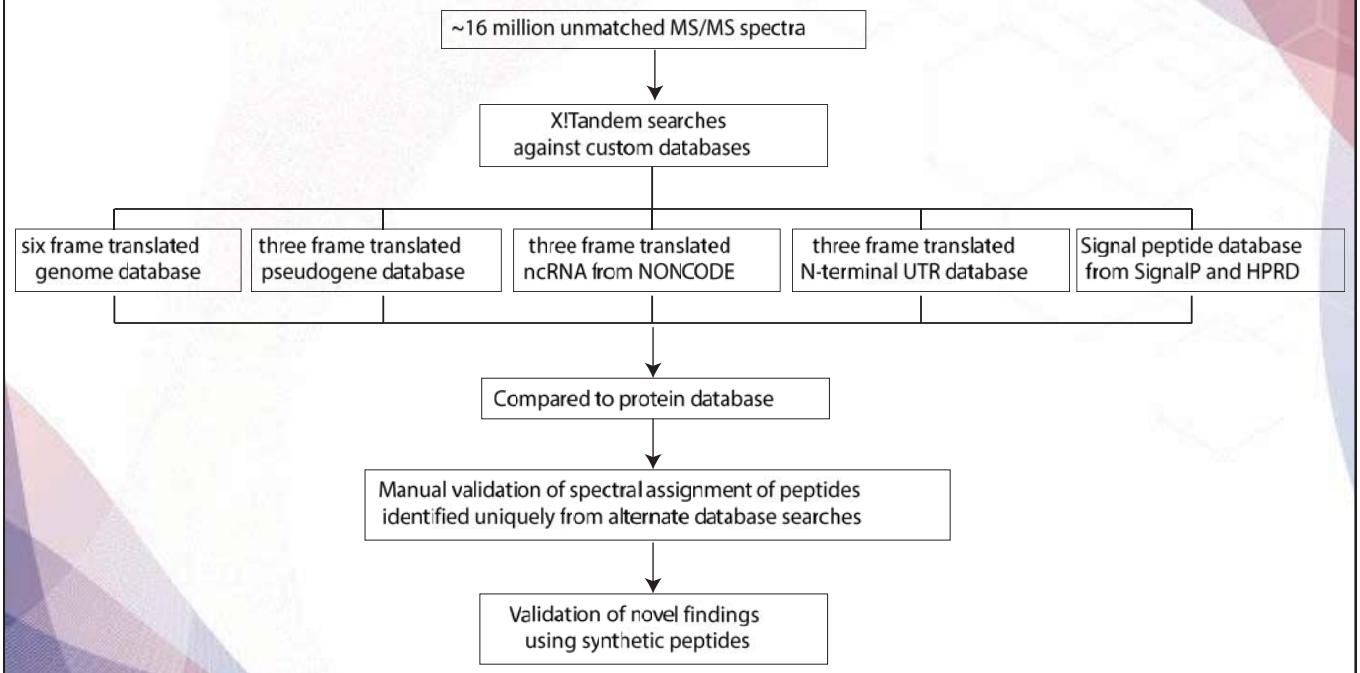
Possible scenarios



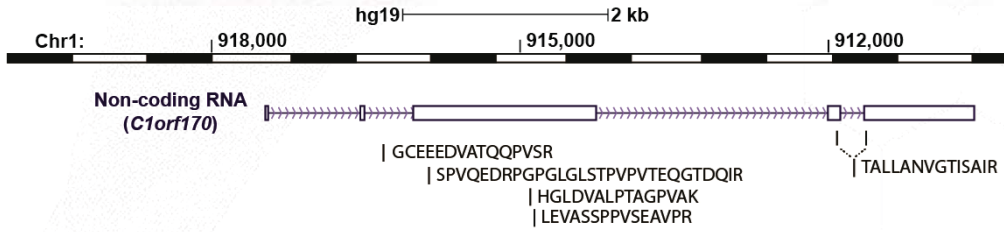
- **They may be just poor quality MS/MS spectra**
- **Good quality MS/MS spectra may be originated from**
  1. **Novel PTMs**
  2. **Novel Isoforms**
  3. **Novel Genes**

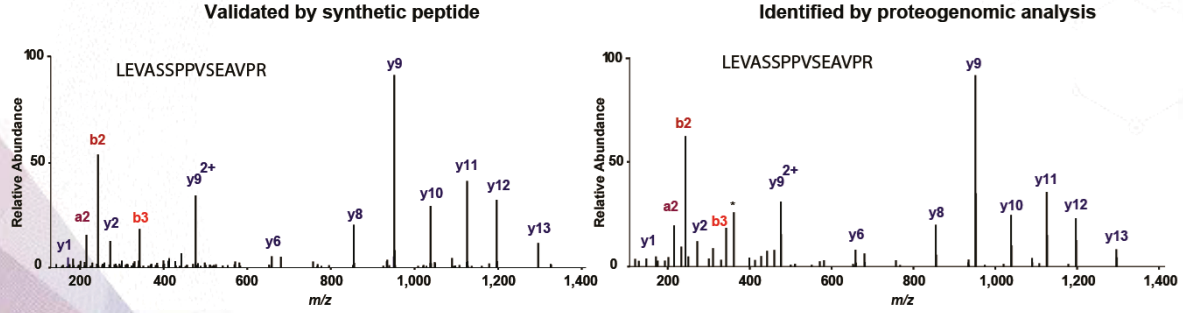


Proteogenomic analysis pipeline

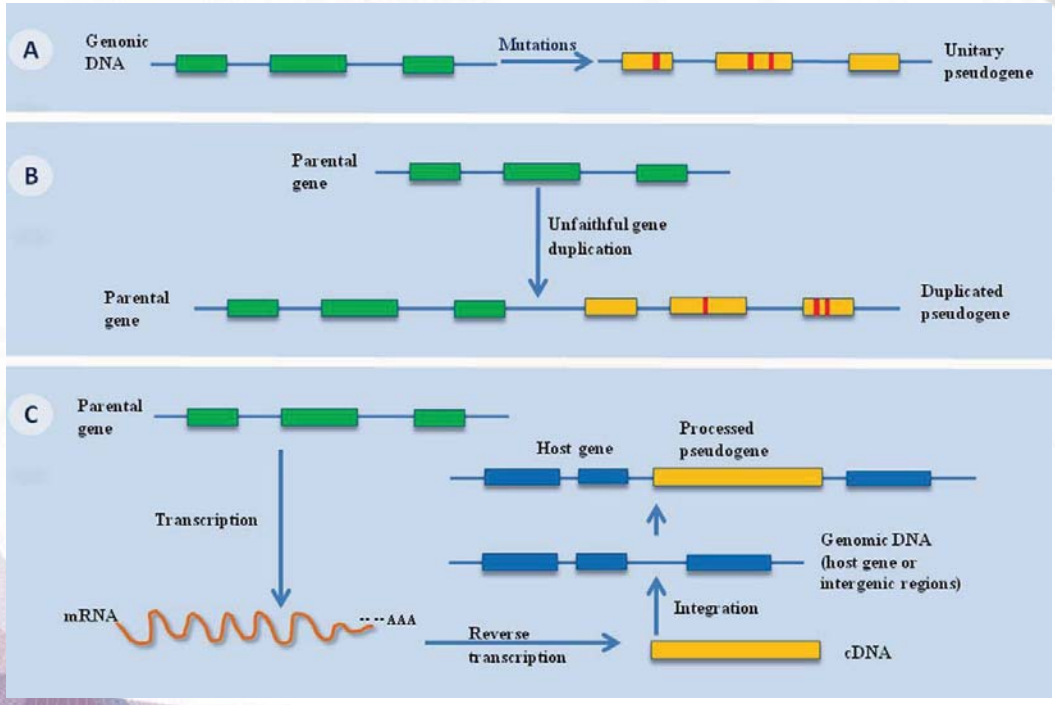


Evidence of translation of non-coding RNA



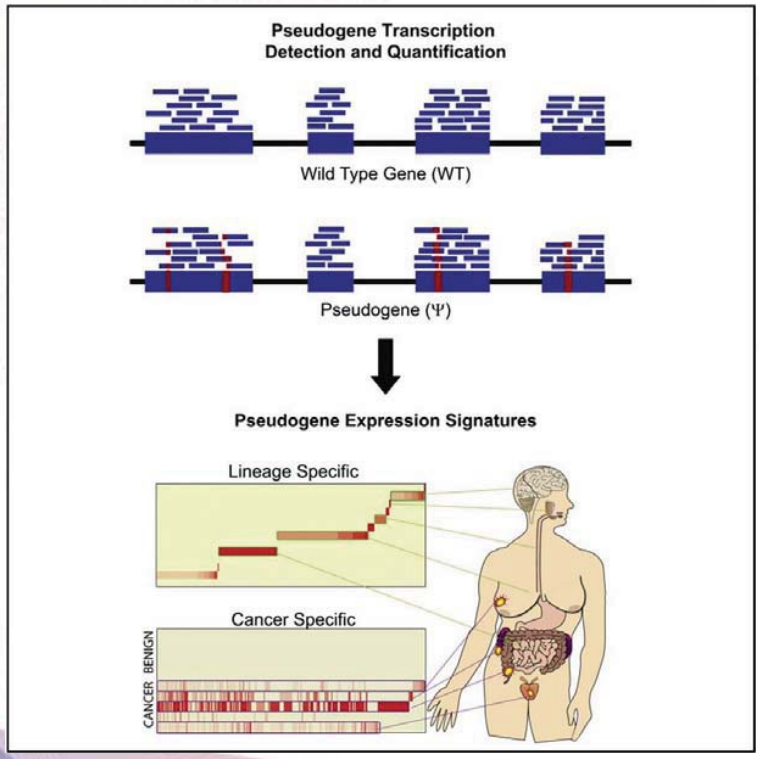


Pseudogene



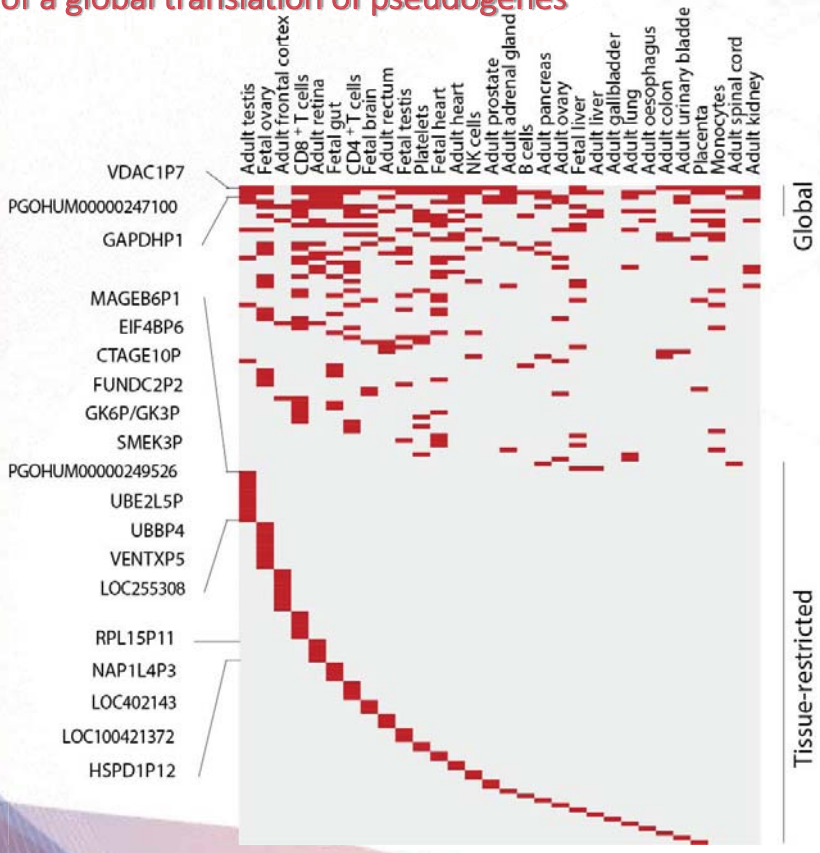
15

Evidence of a global transcription of pseudogenes

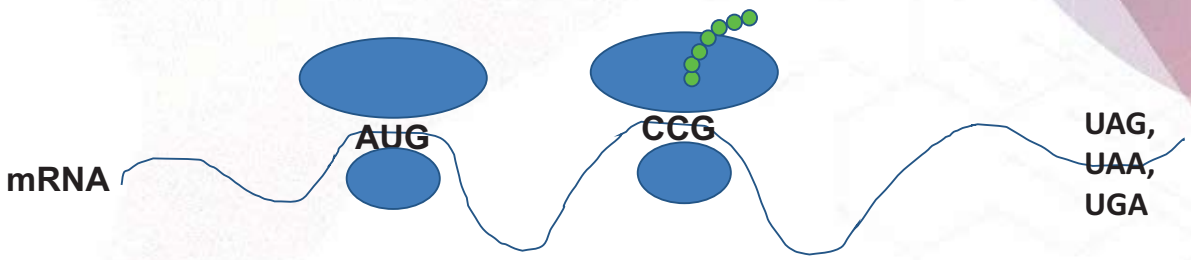


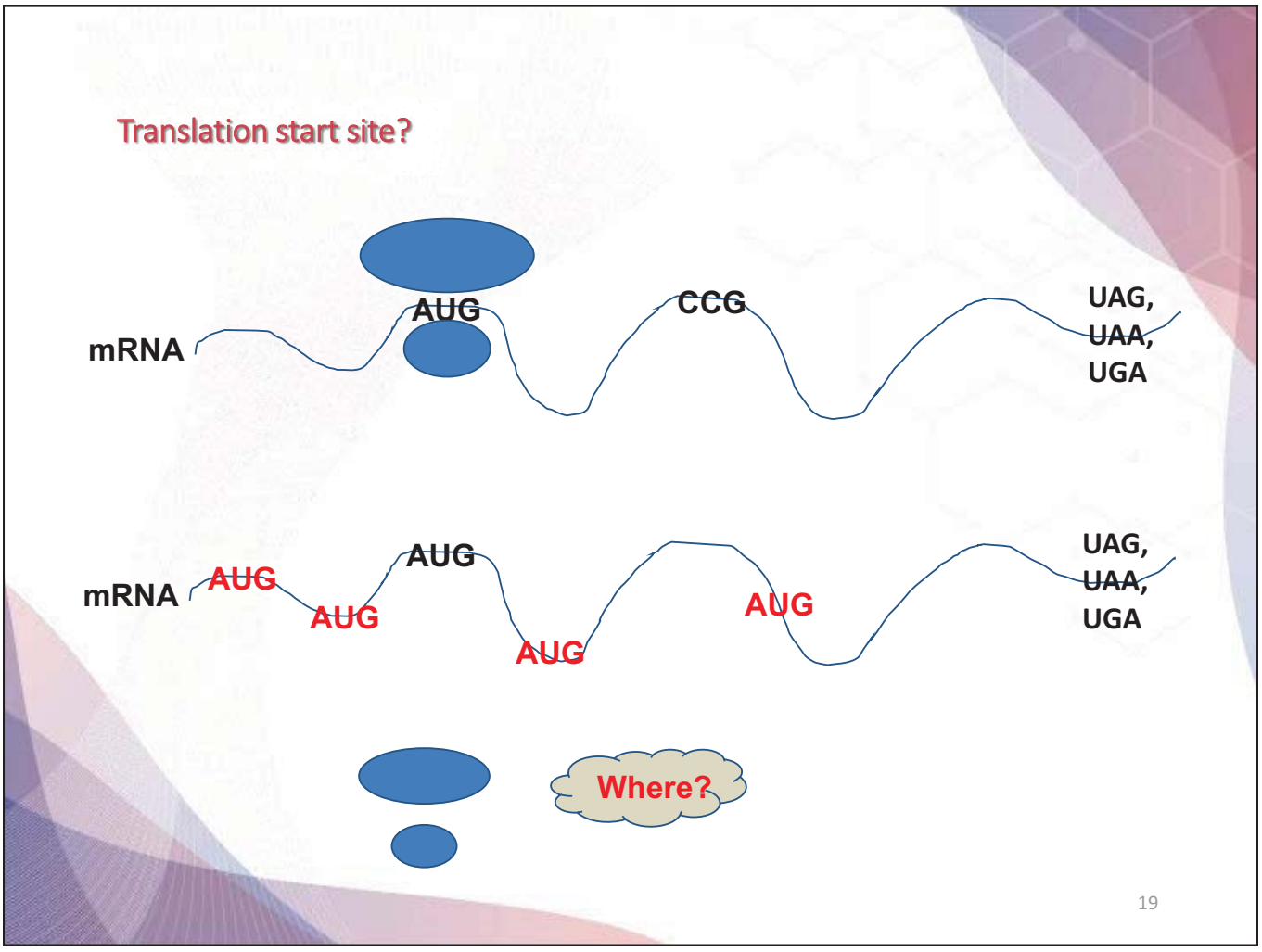
16

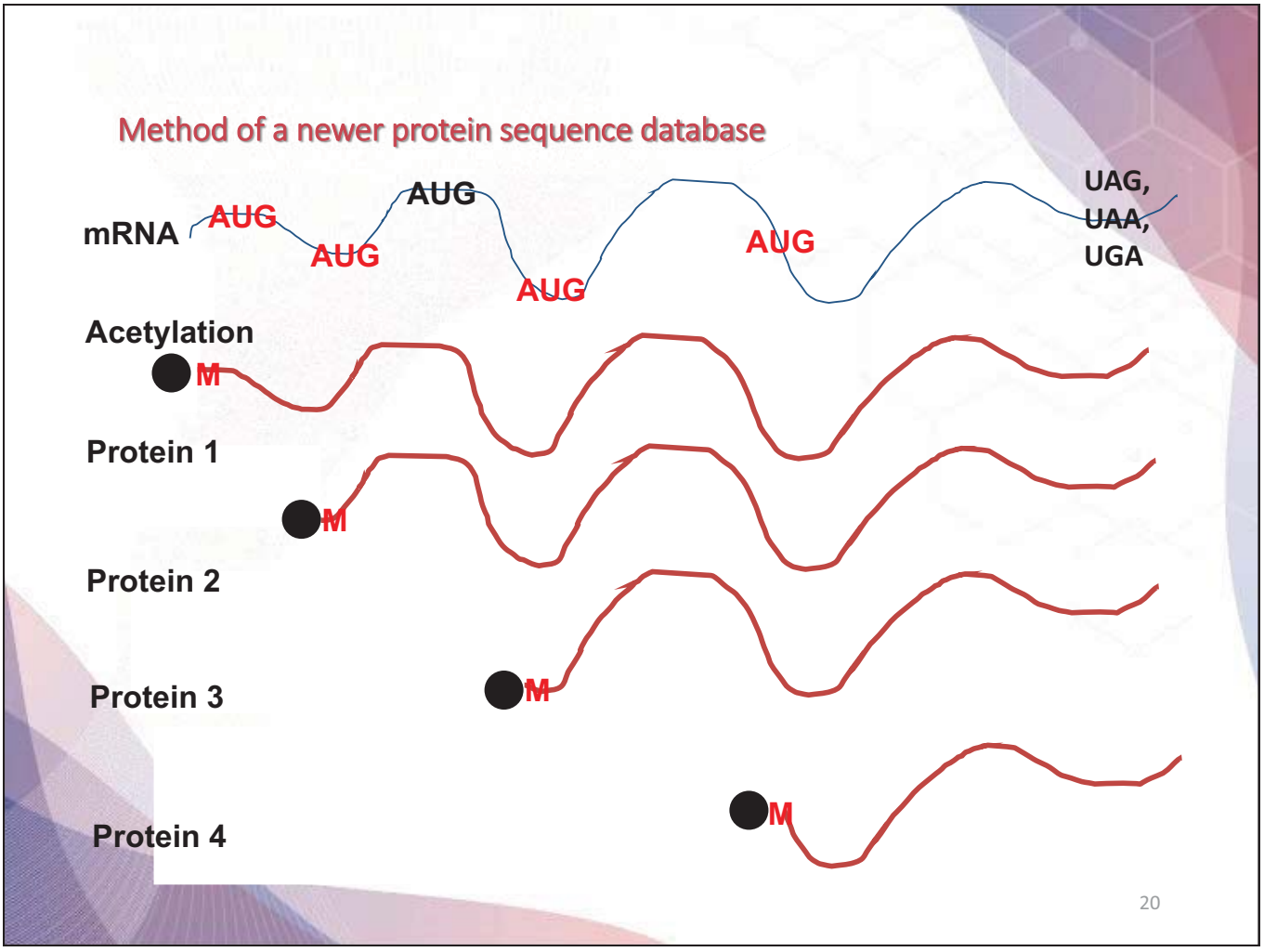
Evidence of a global translation of pseudogenes



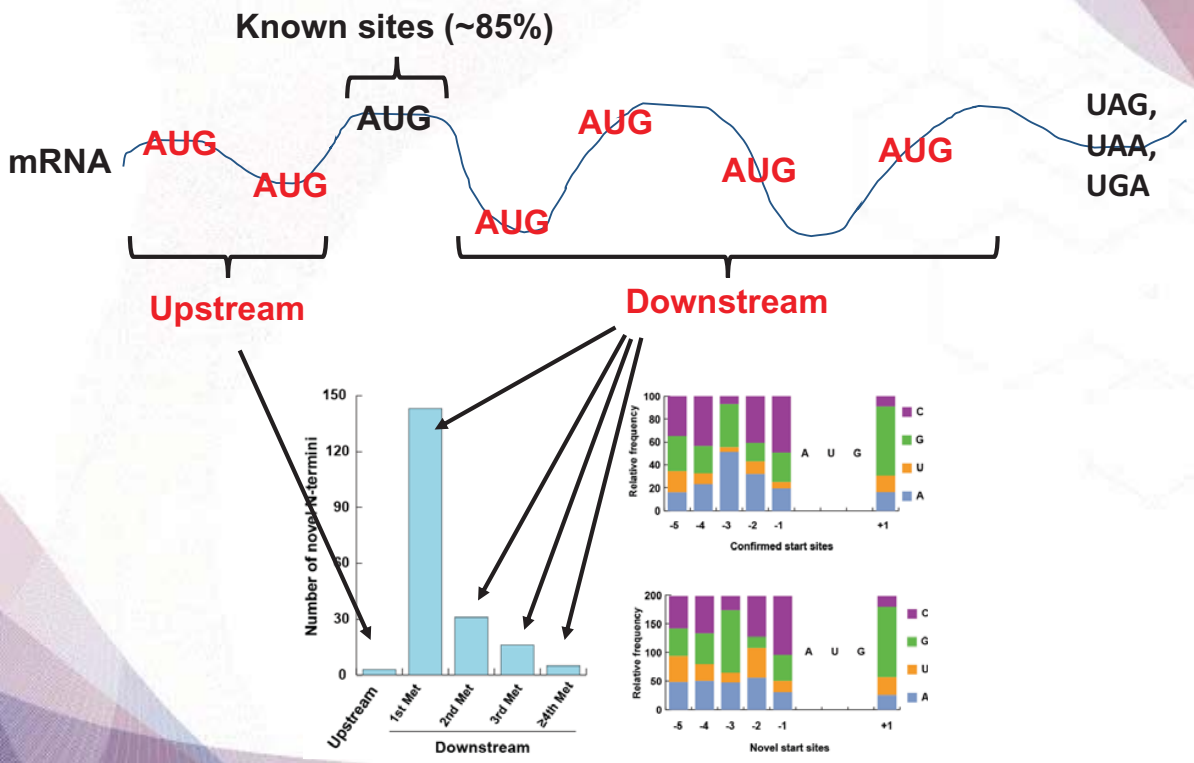
Translation




Translation start site?
mRNA
AUG
CCG
UAG,
UAA,
UGA
mRNA
AUG
AUG
AUG
AUG
AUG
UAG,
UAA,
UGA
Where?
19

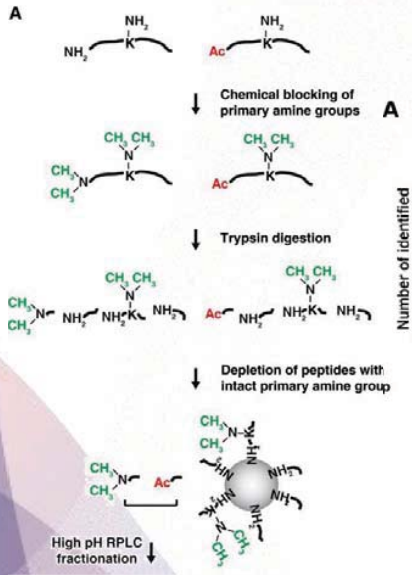

Method of a newer protein sequence database
mRNA
AUG
AUG
AUG
AUG
AUG
UAG,
UAA,
UGA
Acetylation
M
Protein 1
M
Protein 2
M
Protein 3
M
Protein 4
20

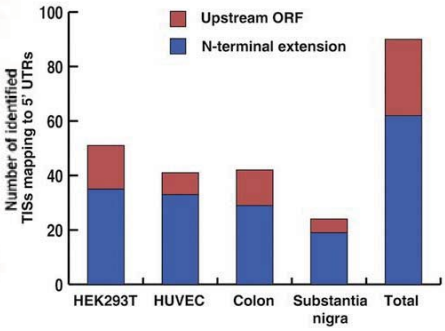
Heterogeneous translation start sites of a messenger

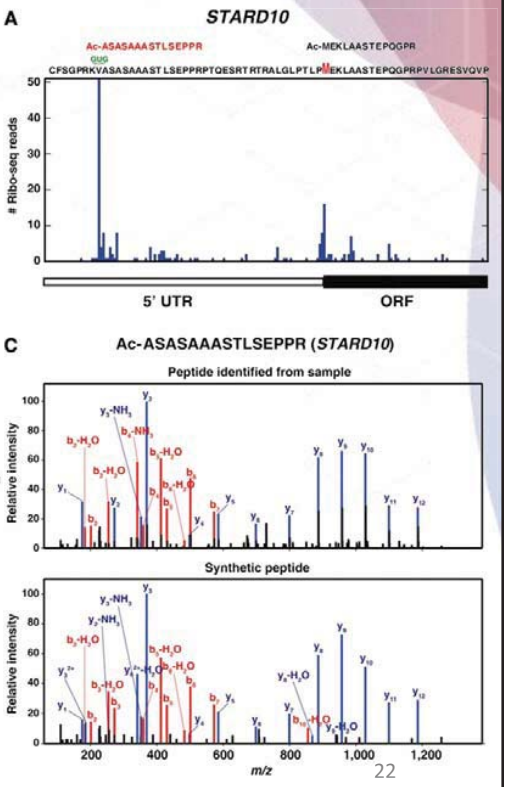


21

Evidence of non-canonical start site for translation



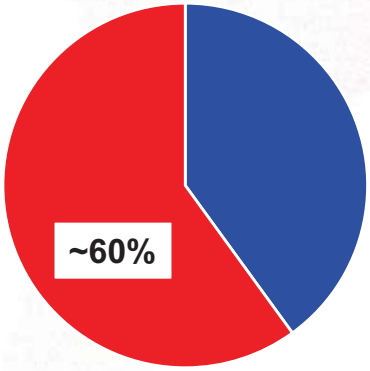




22

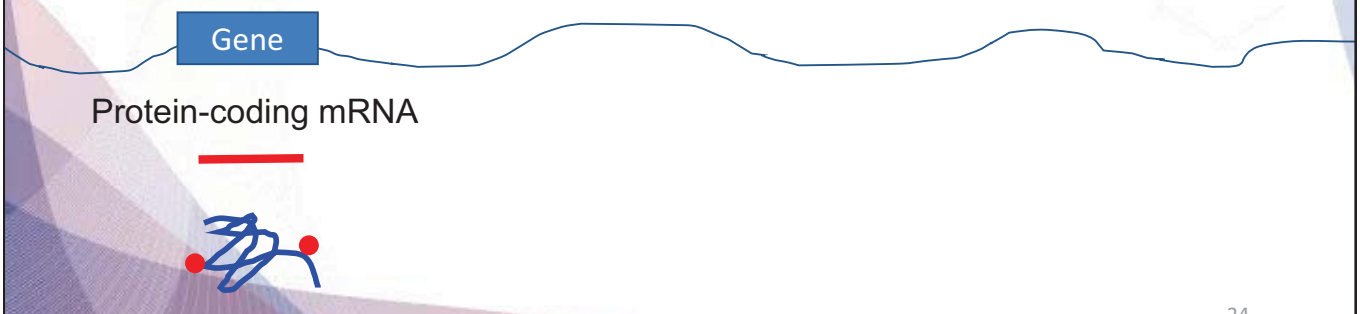
Post-translational Modifications

Possible scenarios



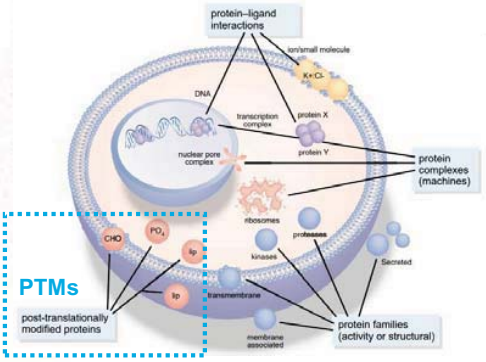
- They may be just poor quality MS/MS spectra
- Good quality MS/MS spectra may be originated from

1. Novel PTMs
2. Novel Isoforms
3. Novel Genes

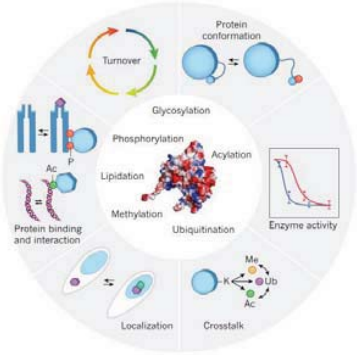


24

Importance of PTM in cellular processes



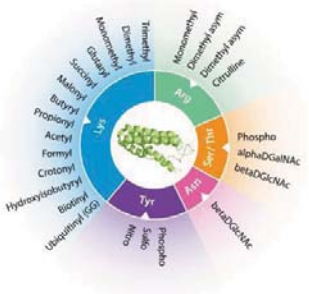
Patterson and Aebersold, *Nat Gen*, 2003



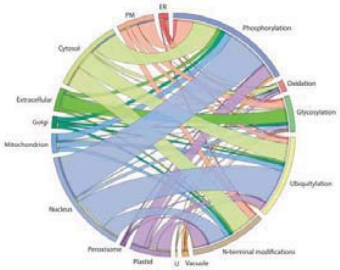
Aebersold and Mann, *Nature*, 2016

Systems-wide study of PTMs is indispensable for a more comprehensive understanding of the human proteome.

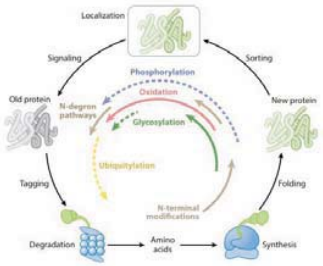
Complicated PTMs



Variety
>300 types of *in vivo* PTMs



Substrate specificity
Different sets of substrate proteins/sites for different PTM types



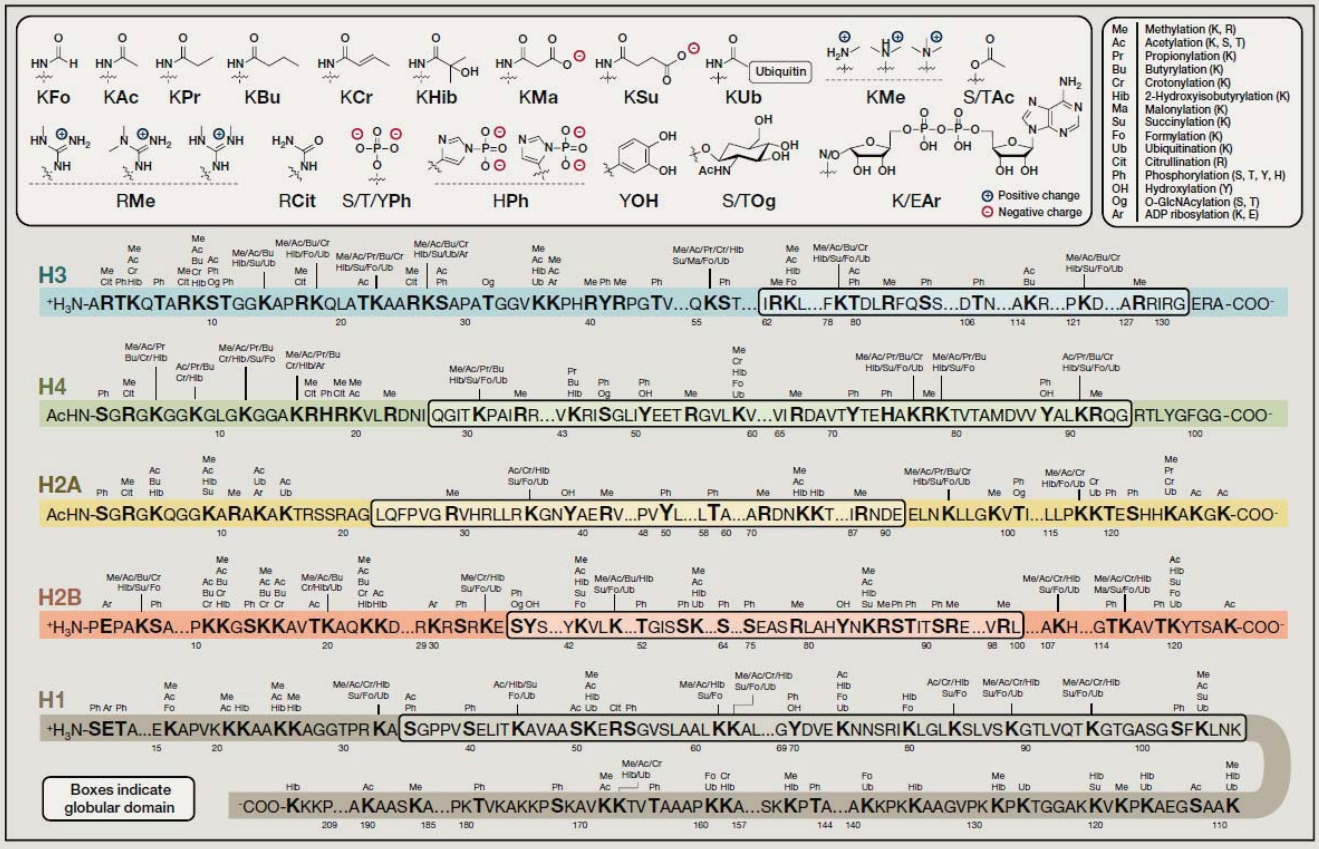
Dynamics
Changes in responsive to intra- and extracellular signals

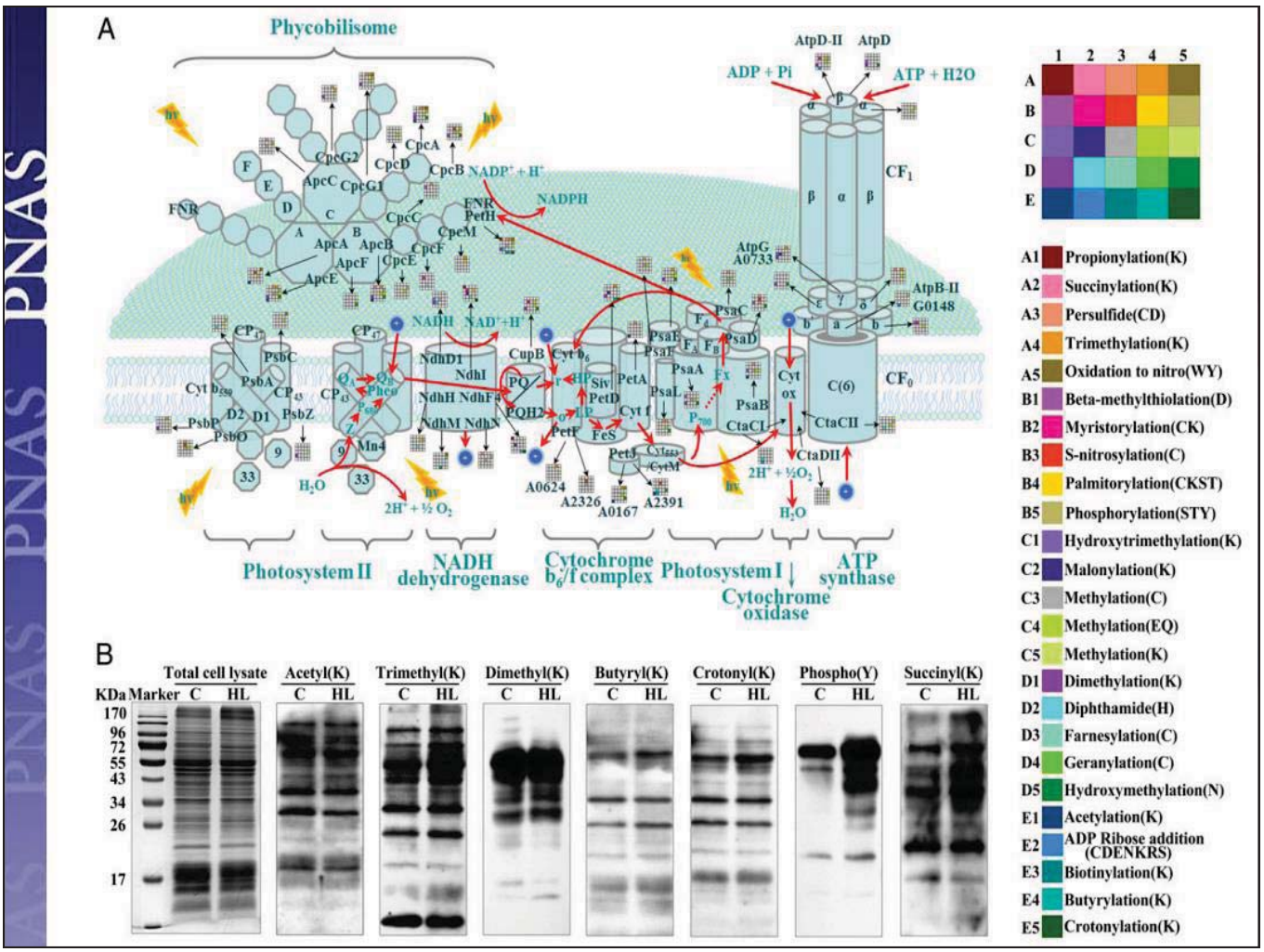
Above all else, a taxonomy of the types of PTMs and their substrate proteins/sites is necessary.
➔ requires *large-scale* & *confident identification* of PTMs.

SnapShot: Histone Modifications

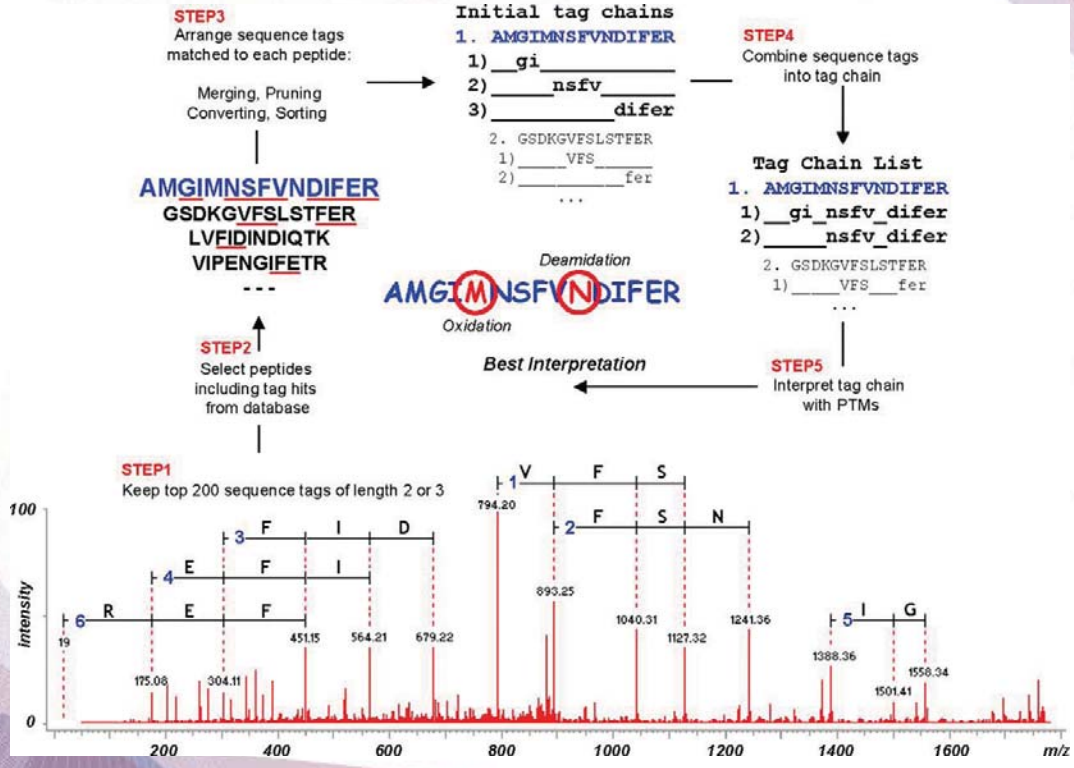
Cell

He Huang,[1] Benjamin R. Sabari,[2] Benjamin A. Garcia,[3] C. David Allis,[2] and Yingming Zhao[1]
[1]Ben May Department of Cancer Research, The University of Chicago, Chicago, IL 60637, USA
[2]Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10021, USA
[3]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA



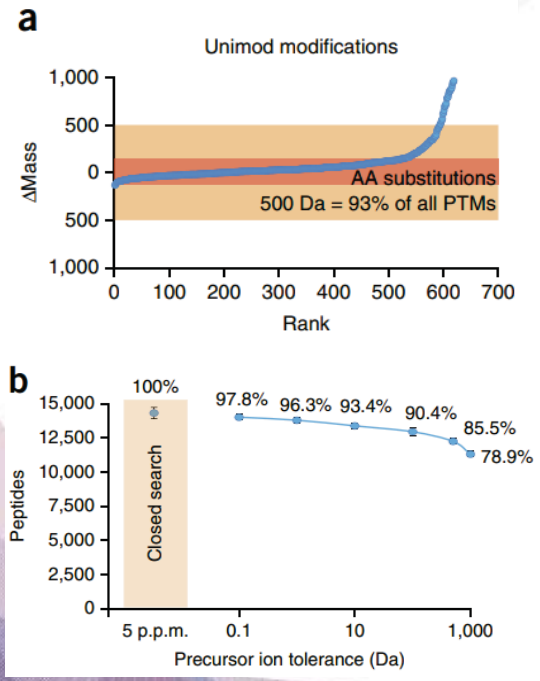


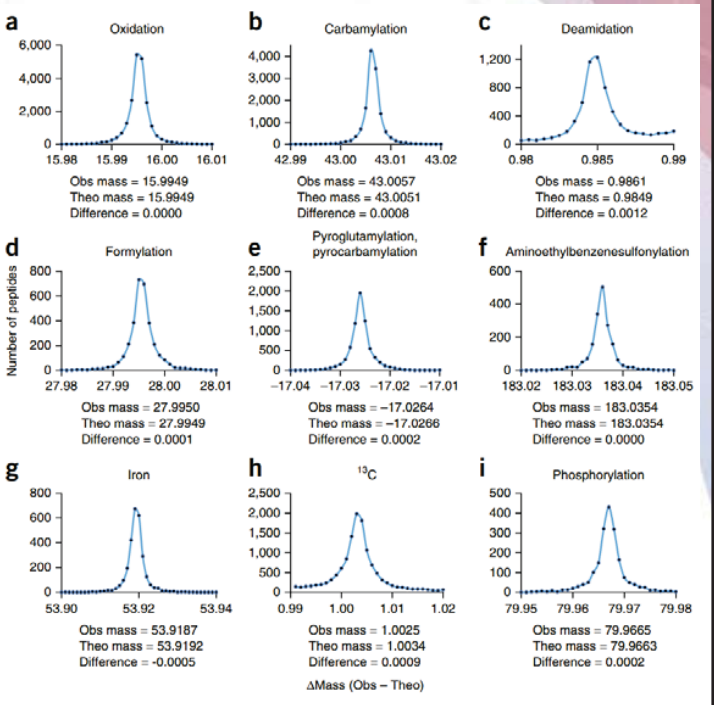
PTM analysis – MODi algorithm



Na et al. *Molecular and Cellular Proteomics*

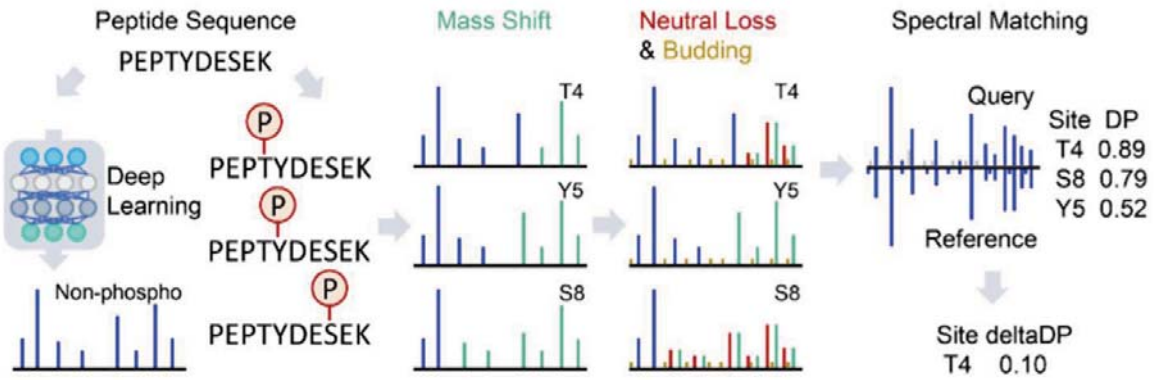
Open Search Strategy



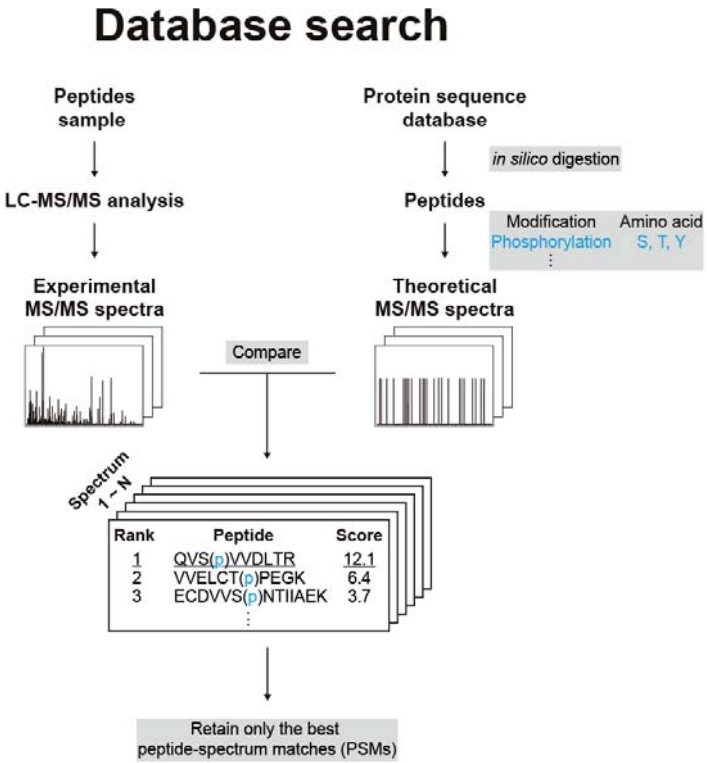


Gygi et al. *Nature Biotechnology*

Deep learning for PTMs

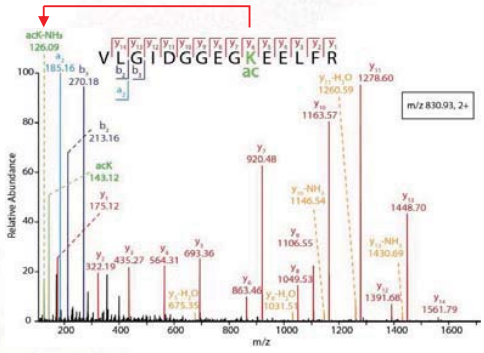


Conventional method of PTM identification

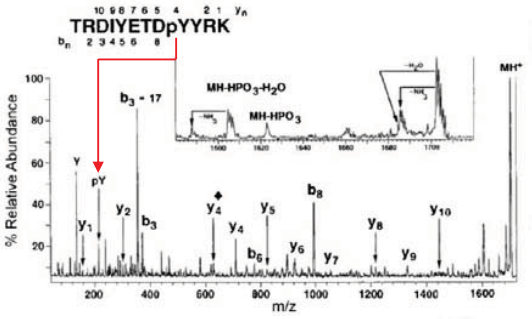


PTM signature ions

- Diagnostic ions can be signatures for PTMs.



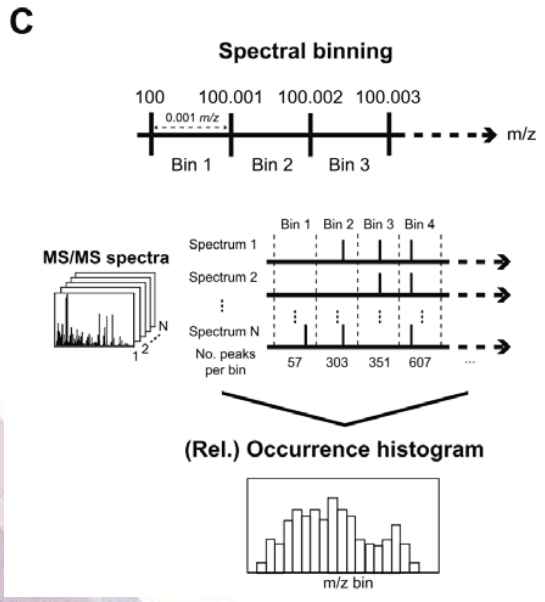
Barth et al., *J Proteomics*, 2020

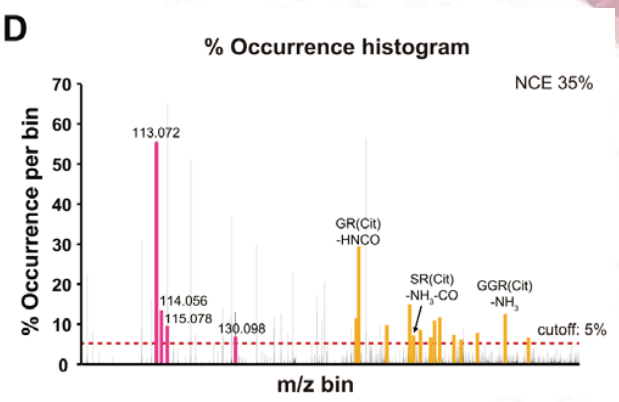


Palumbo et al., *Mass Spec Rev*, 2011

Idea: PTM diagnostic ions can be utilized to facilitate the identification of PTMs in MS/MS data.

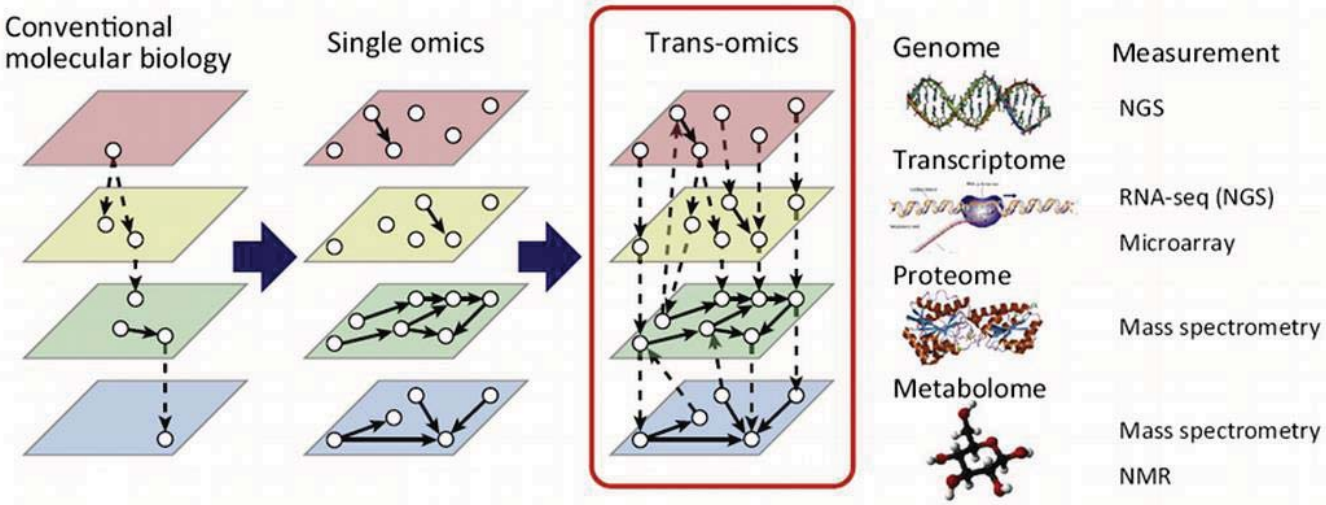
Spectral binning to identify signature ions



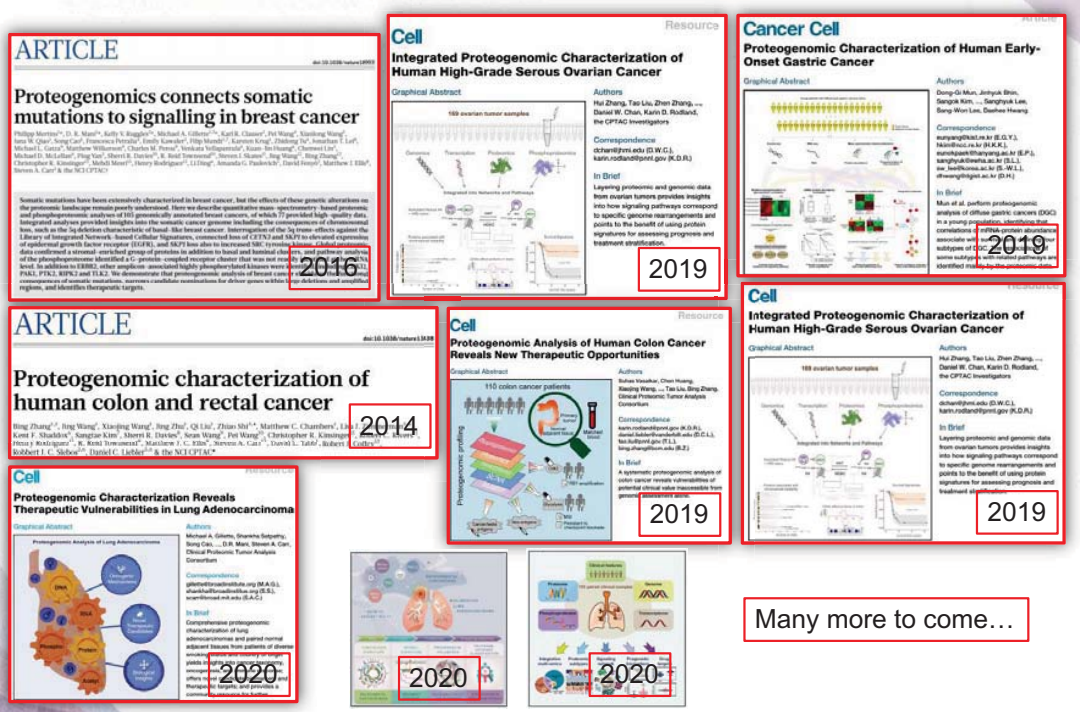


Huh et al. *Analytical Chemistry 2020*

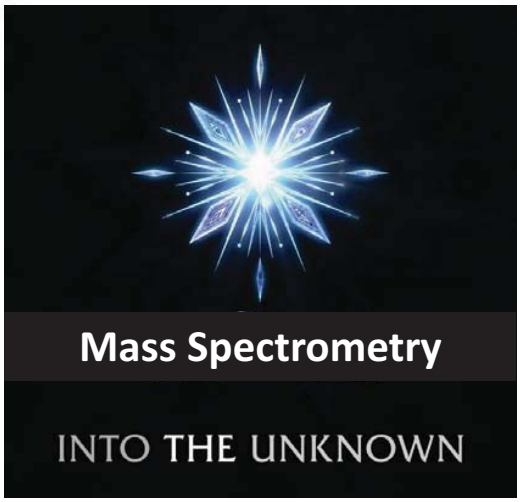
Multi-Omics Integration



Proteome + Transcriptome + Genome



- Principle of Mass Spectrometry and Basics of Proteomics
- Applications to different research fields



Laboratory for QBIO and Precision Medicine (큐바이오 정밀의학 연구실)

단백체, 멀티오믹스 생명정보학
포스닥, 석/박사과정생 모집 중

(김민식, mkim@dgist.ac.kr)