

KSBi-BIML 2021

Bioinformatics & Machine Learning (BIML)
Workshop for Life Scientists

생물정보학 & 머신러닝 워크샵(온라인)

Mass Spectrometry-based Proteomics

김민식

강의개요

Mass Spectrometry-based Proteomics

DNA에 있는 수많은 유전자들은 시간과 위치, 그리고 외부 환경에 따라 단백질을 만들어 내고 있는 것은 매우 흥미로운 일이다. 이를 통해 다세포 생물의 하나인 인간에게 존재하는 수많은 다른 종류의 세포가 각기 다른 일을 유기적으로 할 수 있는 것일 것이다. 우리는 많은 경우 유전적이지 않은 상황으로 인해 DNA의 변형을 맞이하고 이를 통해 세포에 병인이 발생하고 결국 죽음에 이르는 병을 얻게 된다.

최근 질량분석법을 기반으로 하는 단백질 집합체에 대한 연구 기술이 급격히 발달하고 있으며 곧 유전체 수준의 방대한 데이터 양을 생산하는 날이 머지 않았다. 그렇다면 우리는 언제 단백질 기반의 빅데이터 연구를 하게 될까? 아마도 곧 유전체와 단백질의 통합에 대한 활발한 연구를 통해 생명체가 시간적, 그리고 공간적으로 어떻게 외부 환경에 반응하고, 어떻게 내부적으로 짜여진 프로그램을 영위해 나가는지 이해할 수 있을 것이다. 그러나 단백질 데이터는 유전체 데이터와는 상이한 방식으로 수집이 되고 이를 이해하는 방법은 매우 다르다.

본 강의에서는 단백질 데이터 수집 방식을 이해하고 데이터를 효과적이고 빠르게 분석하기 위해 널리 사용되는 플랫폼에 대해 공부하고 데이터 처리에의 응용을 설명하며 이를 통해 빅데이터를 빠르고 손쉽게 처리할 수 있는 핵심 역량을 갖추는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- Mass Spectrometry 개요
- 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: 김민식 교수 (DGIST 뉴바이올로지)

Curriculum Vitae

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



► Personal Info

Name Min-Sik Kim
Title Associate Professor
Affiliation Dep't of New Biology, DGIST

► Contact Information

Address DGIST, 333 TechnoJungang-daero, Dalseong-gun, Daegu, 42988
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
2018-2020 Assistant Professor, Department of New Biology, DGIST
2020- Associate Professor, Department of New Biology, DGIST

Selected Publications (5 maximum)

Huh, S., Hwang, D.*, Kim MS* (2020) Statistical modeling for enhancing discovery power of citrullination from tandem mass spectrometry data. Analytical Chemistry. Accepted.

2. Ntwali P.V M, Heo CE, Han JU, Chae SY, Kim M, Vu, H. M., Kim MS* and Kim, H.* (2020) Mass spectrometry-based proteomics of single cells and organoids: The new generation of cancer research. Trends in Analytical Chemistry. Accepted.

3. Sim SY, Choi YR, Lee JH, Lim JM, Lee SE, Kim KP, Kim JY, Lee SH, and Kim MS* (2019) In-depth proteomic analysis of human bronchoalveolar lavage fluid towards the biomarker discovery for lung cancers. Proteomics-Clinical applications. 13(5):e1900028..

4. Bae J, Kim SJ, Lee SE, Kwon W, Kim H, Han Y, Jang JY, Kim MS* and Lee, S.-W.* (2019) Comprehensive proteome and phosphoproteome profiling shows negligible influence of RNAlater on protein abundance and phosphorylation. Clinical Proteomics. 16:18..

5. Lee, S.-E., Song, J.-G., Bösl, K., Müller, A. C., Vitko, D., Bennett, K. L., Superti-Furga, G, Pandey, A., Kandasamy, R. K.* and Kim MS* (2018) Proteogenomic analysis to identify missing proteins from haploid cell lines. Proteomics. 18(8):e1700386. *Co-correspondence authorship.

Bioinformatics & Machine Learning for Life Scientists

BIML-2021

안녕하십니까?

한국생명정보학회의 동계 워크샵인 BIML-2021을 2월 15부터 2월 19일까지 개최합니다. 생명정보학 분야의 융합이론 보급과 실무역량 강화를 위해 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하였으며 올해로 7차를 맞이하게 되었습니다. 유례가 없는 코로나 대유행으로 인해 올해의 BIML 워크숍은 온라인으로 준비했습니다. 생생한 현장 강의에서만 느낄 수 있는 강의자와 수강생 사이의 상호교감을 가질 수 없다는 단점이 있지만, 온라인 강의의 여러 장점을 살려서 최근 생명정보학에서 주목받고 있는 거의 모든 분야를 망라한 강의를 준비했습니다. 또한 온라인 강의의 한계를 극복하기 위해서 실시간 Q&A 세션 또한 마련했습니다.

BIML 워크샵은 전통적으로 크게 생명정보학과 AI, 두 개의 분야로 구성되어오고 있으며 올해 역시 유사한 방식을 채택했습니다. AI 분야는 Probabilistic Modeling, Dimensionality Reduction, SVM 등과 같은 전통적인 Machine Learning부터 Deep Learning을 이용한 신약개발 및 유전체 연구까지 다양한 내용을 다루고 있습니다. 생명정보학 분야로는, Proteomics, Chemoinformatics, Single Cell Genomics, Cancer Genomics, Network Biology, 3D Epigenomics, RNA Biology, Microbiome 등 거의 모든 분야가 포함되어 있습니다. 연사들은 각 분야 최고의 전문가들이라 자부합니다.

이번 BIML-2021을 준비하기까지 너무나 많은 수고를 해주신 BIML-2021 운영위원회의 김태민 교수님, 류성호 교수님, 남진우 교수님, 백대현 교수님께 커다란 감사를 드립니다. 또한 재정적 도움을 주신, 김선 교수님 (AI-based Drug Discovery), 류성호 교수님, 남진우 교수님께 감사를 표시하고 싶습니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 강의자료를 만드는데 노력하셨을 뿐만 아니라 실시간 온라인 Q&A 세션까지 참여해 수고해 주시는 모든 연사분들께 깊이 감사드립니다.

2021년 2월

한국생명정보학회장 김동섭



KSBi-BIML 2021

Mass Spectrometry-based Proteomics
Min-Sik Kim, Department of New Biology, DGIST

본 강의 자료는 한국생명정보학회가 주관하는 KSBi-BIML 2021 워크숍 온라인 수업을 목적으로 제작된것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다. 수업 목적으로 배포 및 전송 받은 경우에도 이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없습니다.

만약 이러한 사항을 위반할 경우 발생하는 모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고합니다.

- 김민식, 이학박사

- 1995-2002 고려대학교 화학과 이학 학사
- 2002-2004 고려대학교 화학과 질량분석학 석사
- 2007-2013 존스홉킨스 의과대학 생화학 박사
- 2013-2016 존스홉킨스 유전체연구소 박사후연구과정
- 2016-2018 경희대학교 응용화학과 조교수
- 2018-2020 DGIST 뉴바이올로지 조교수
- 2020-현재 DGIST 뉴바이올로지 부교수

3

대구경북과학기술원(DGIST)



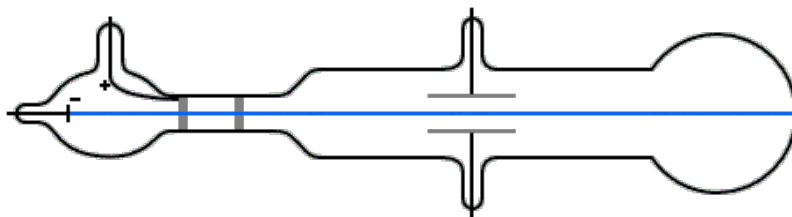
Mass Spectrometry

질량분석기의 원리



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

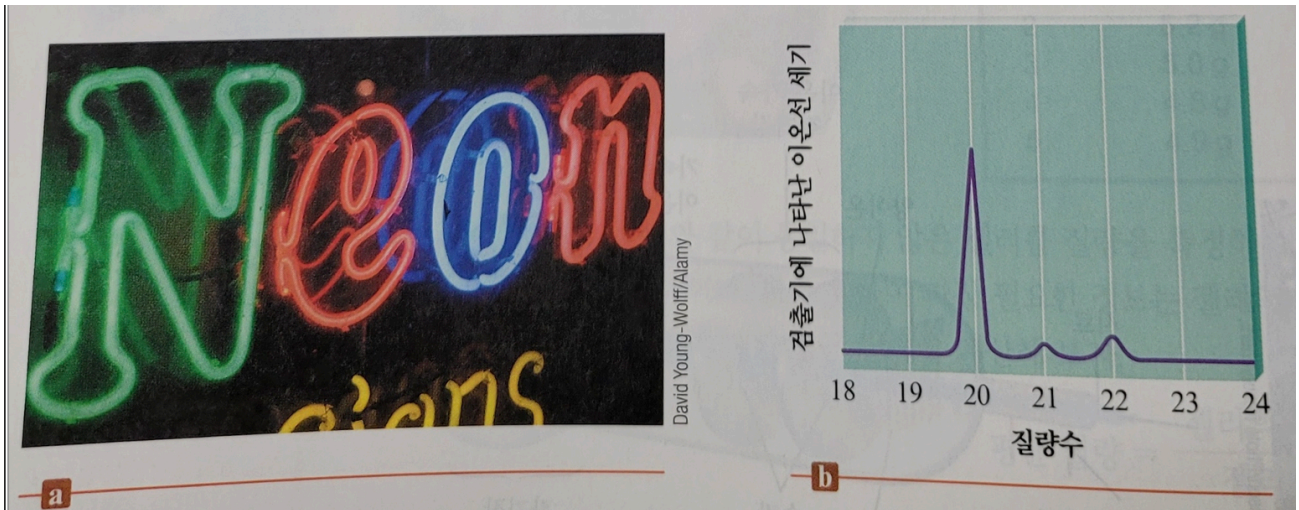


그림 3.2 ▶ (a) 방전관 안에서 작렬하는 네온 가스. 천연 네온을 질량 분석계에 주입했을 때 얻어지는 신호의 상대적 형태로 표시한 것이다. 봉우리의 상대 면적이 0.9092(^{20}Ne), 0.00257(^{21}Ne), 0.0882(^{22}Ne)이므로, 천연 네온에

중달의 일반화학

7

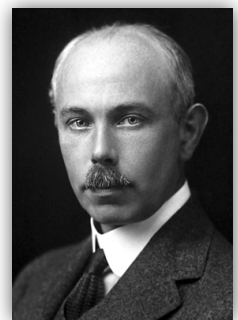
질량분석기의 개발과 안정한 동위원소의 발견

Nobel Prize in Chemistry (1922)

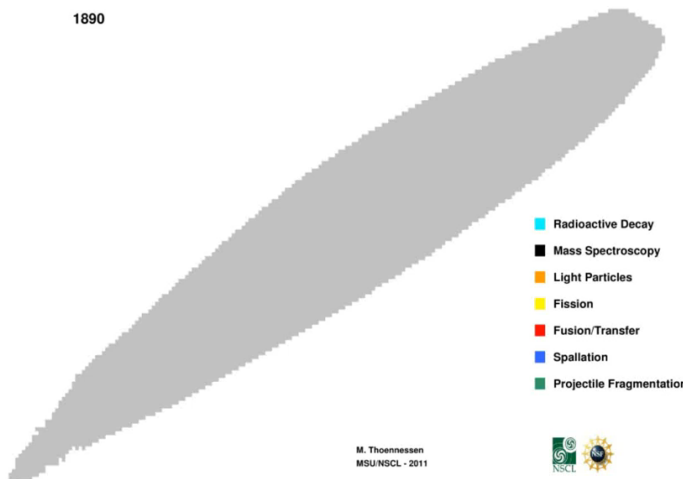


"for his discovery, by means of his mass spectrograph, of isotopes, in a large number of non-radioactive elements, and for his enunciation of the whole-number rule."

1890



Francis Aston
(1877~1945)



- Radioactive Decay
- Mass Spectroscopy
- Light Particles
- Fission
- Fusion/Transfer
- Spallation
- Projectile Fragmentation

M. Thoennessen
MSU/NSCL - 2011



Identified
212 of the 287 naturally
occurring isotopes

원소의 질량

원소의 주기율표

원소 기호: H, 원자 번호: 1, 원소 이름: 수소, 영어명: hydrogen, 원자량: 1.00794

탄소: C, 질소: N, 산소: O

탄소: Carbon, 질소: Nitrogen, 산소: Oxygen

탄소: 12.011, 질소: 14.0067, 산소: 15.9994

교학사 화학 I

9

탄소의 질량

- 탄소, carbon
- (평균)원자량 12.011

C 6

탄소

Carbon

12.011

원소 기호: C, 원자 번호: 6, 원소 이름: 탄소, 영어명: Carbon, 원자량: 12.011

탄소: Carbon, 질소: Nitrogen, 산소: Oxygen

탄소: 12.011, 질소: 14.0067, 산소: 15.9994

교학사 화학 I, 66p

표 II-3 생명체 구성 원소의 동위 원소

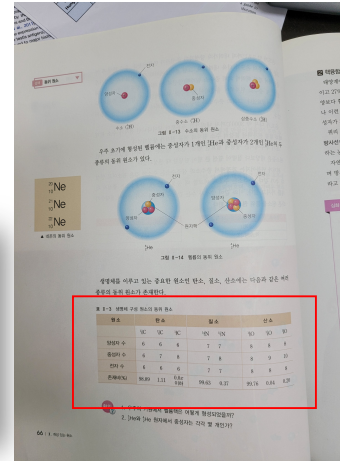
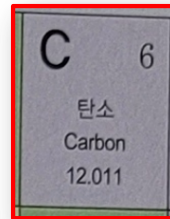
원소	탄소			질소		산소		
	^{12}C	^{13}C	^{14}C	^{14}N	^{15}N	^{16}O	^{17}O	^{18}O
양성자 수	6	6	6	7	7	8	8	8
중성자 수	6	7	8	7	8	8	9	10
전자 수	6	6	6	7	7	8	8	8
존재비(%)	98.89	1.11	0.0x 이하	99.63	0.37	99.76	0.04	0.20

10

탄소의 질량

• 탄소, carbon

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

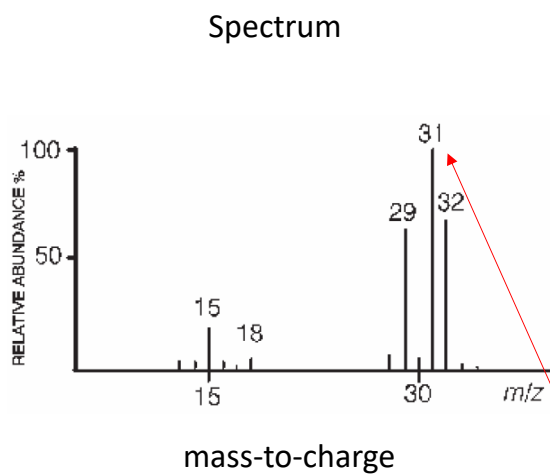


교학사 화학 I

$$\begin{aligned} \text{탄소 원자의 (평균)원자량} &= 12 \times 0.9889 + 13 \times 0.0111 \\ &= \mathbf{12.0111} \end{aligned}$$

11

A Typical Mass Spectrum

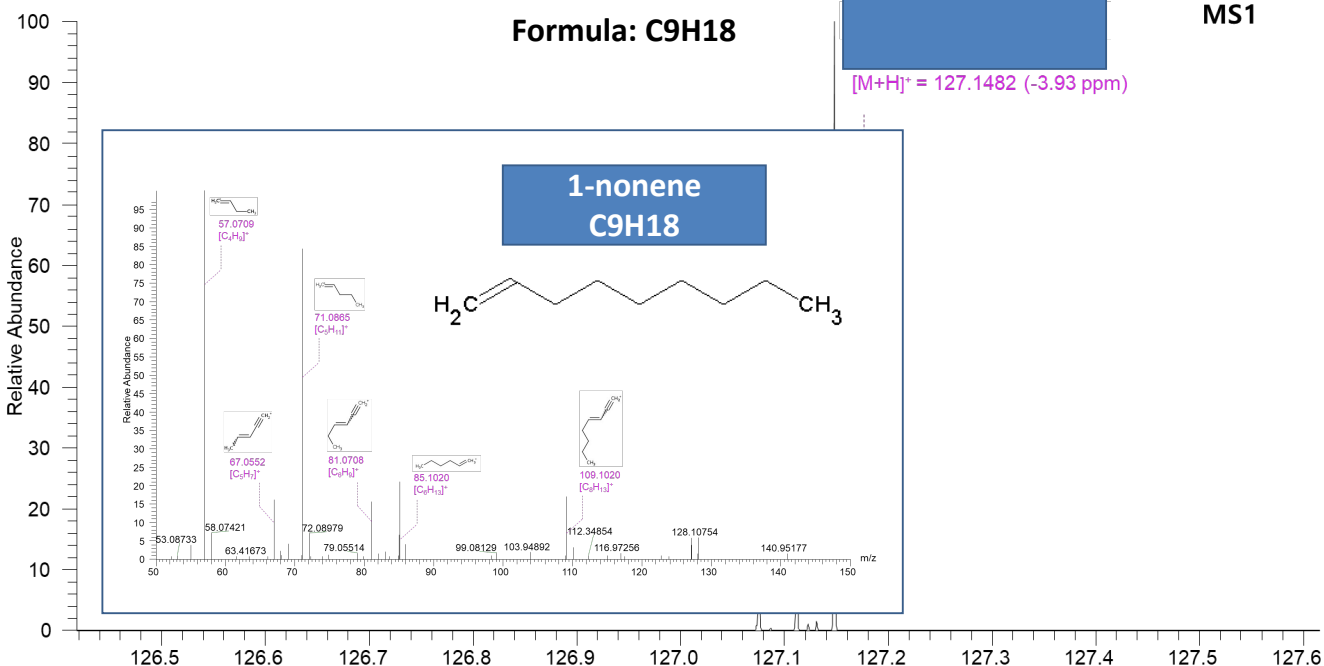
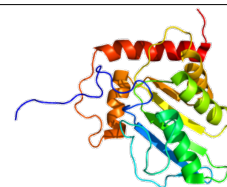


Mass table

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

Qualitative Analysis of Molecules using MS



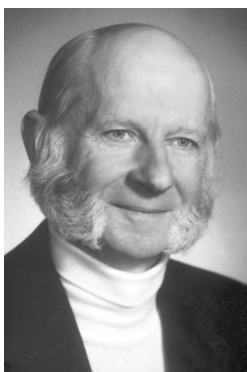
Unpublished data

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

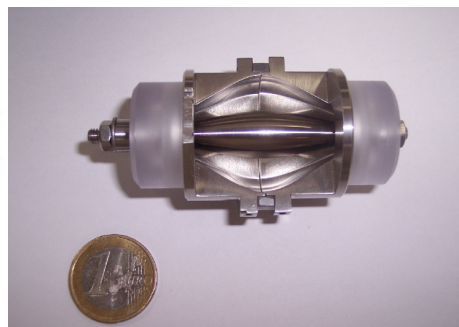
질량분석기



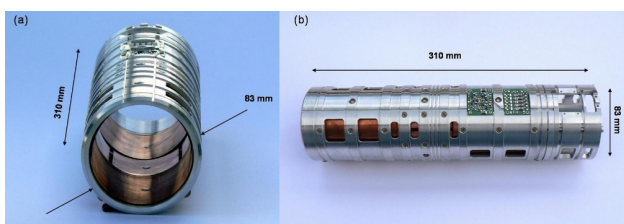
Time-of-Flight



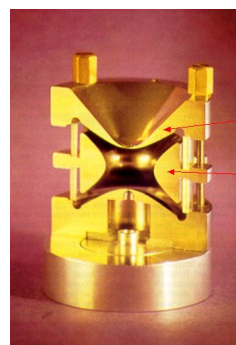
Quadrupole



Orbitrap



FT-ICR



Ion trap

15





Koichi Tanaka
(1959 -)

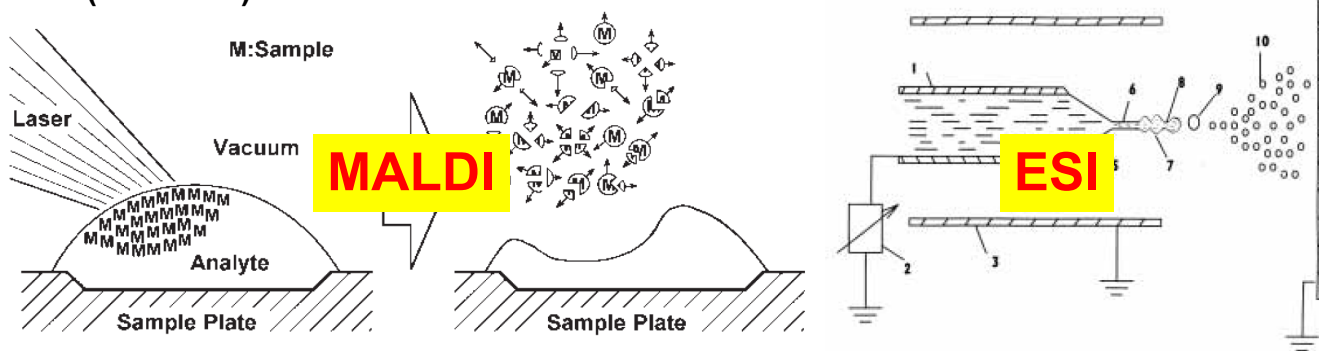


Nobel Prize in Chemistry (2002)

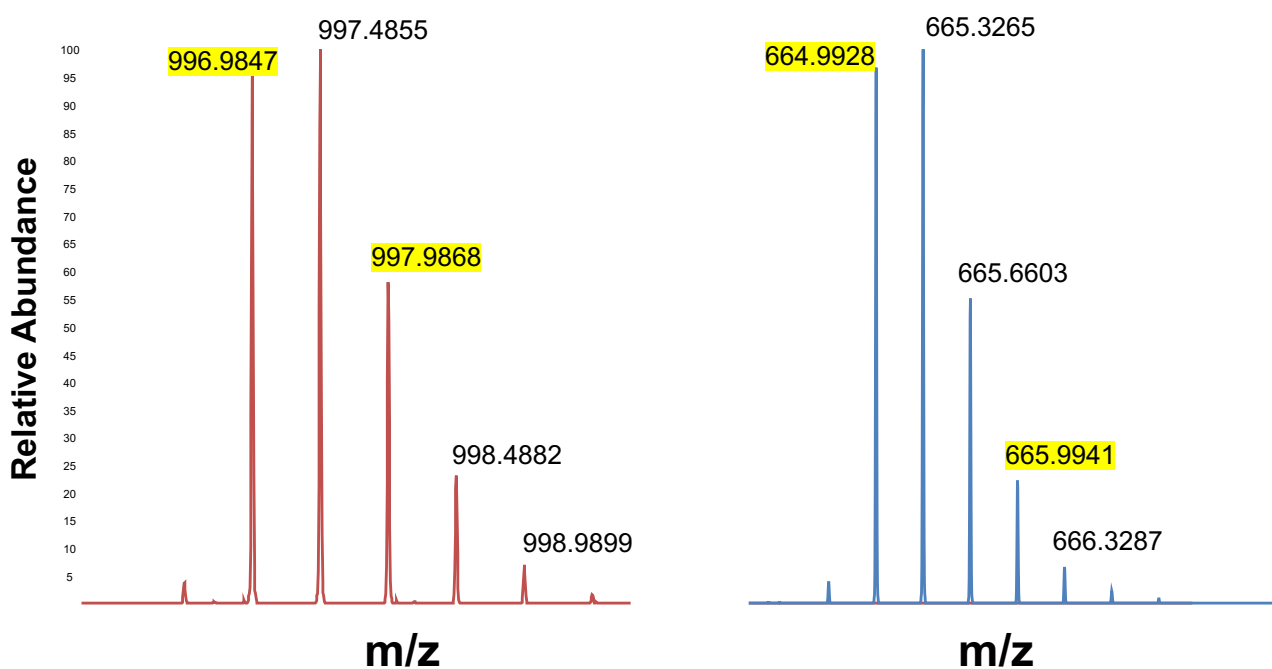
“for their development of soft desorption ionisation
methods for mass spectrometric analyses of
biological macromolecules”



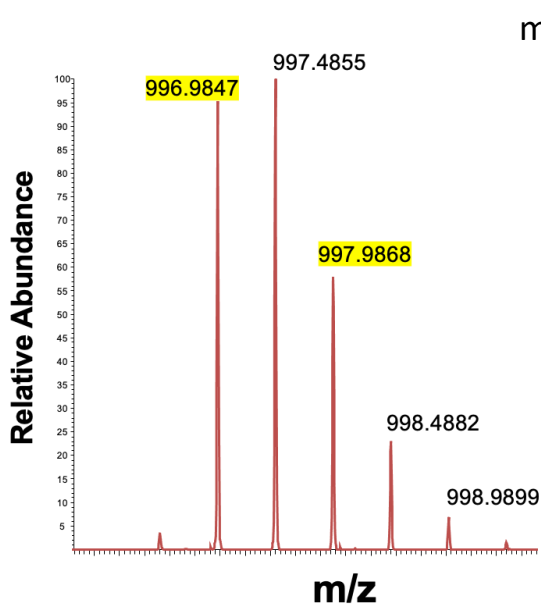
John B. Fenn
(1917 - 2010)



질량스펙트럼의 특성 – isotopic cluster (동위원소분포클러스터)



질량스펙트럼의 특성 – mass to charge ratio (질량전하비)



$$m/z = 996.9847$$

$$\text{charge } (z) = +2$$

$$(m) = 996.9847 \times (+2) = \text{Peptide} + 2\text{H}^+$$

$$\begin{aligned} \text{Peptide mass} &= 996.9847 \times (+2) - 2 \times 1.0078 \\ &= 1991.9538 \end{aligned}$$

19

질량스펙트럼의 특성 – resolution(분해능)

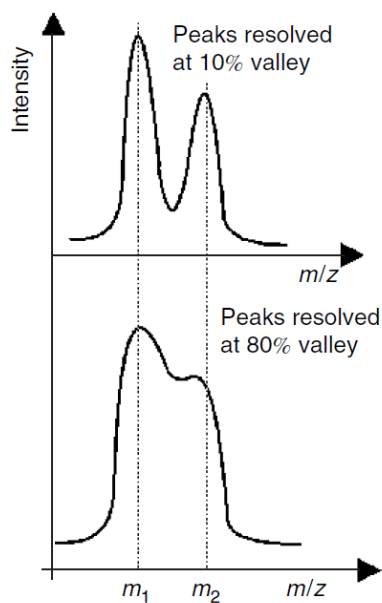
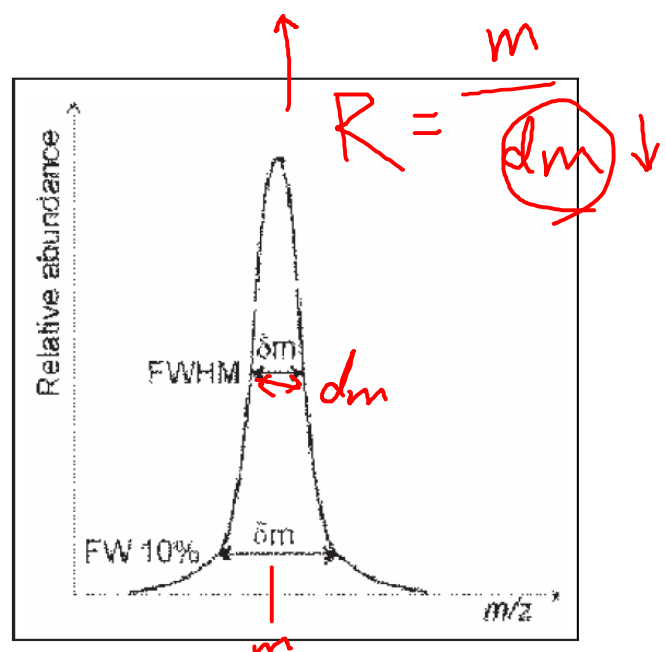
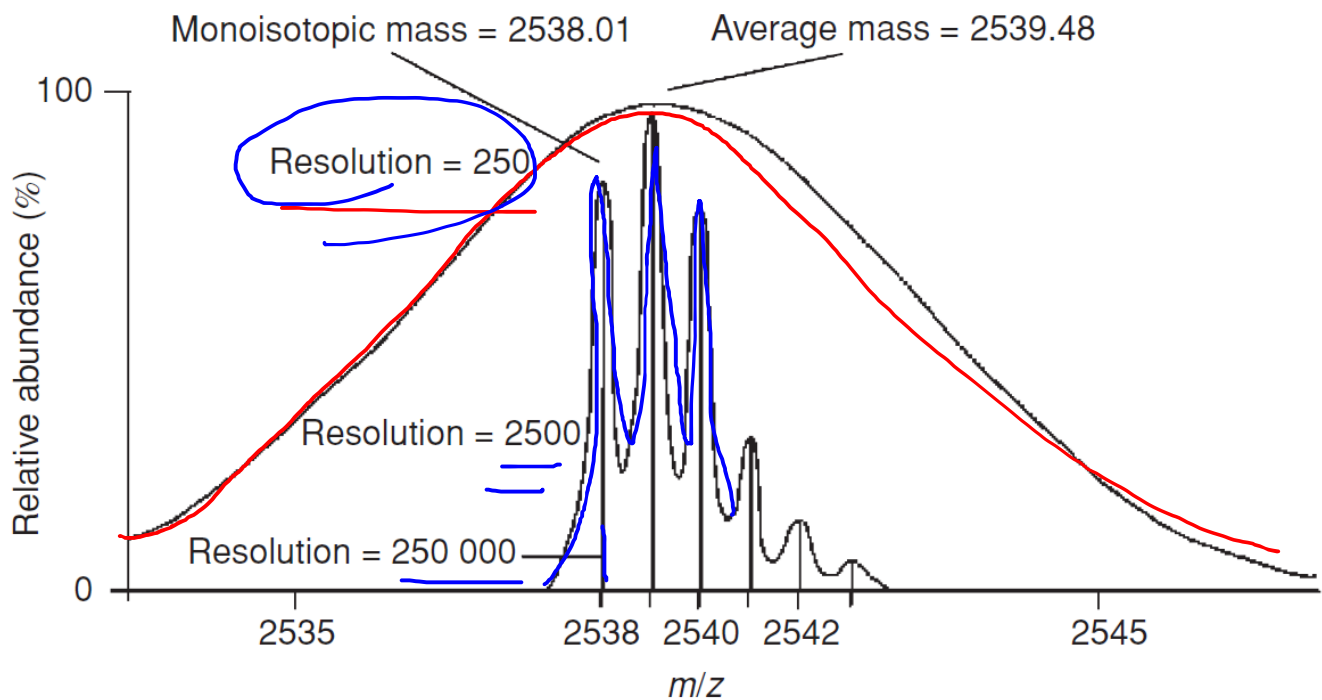


Figure 2.1
Diagram showing the concepts of peak resolution and valley.



20

질량스펙트럼의 특성 – resolution(분해능)



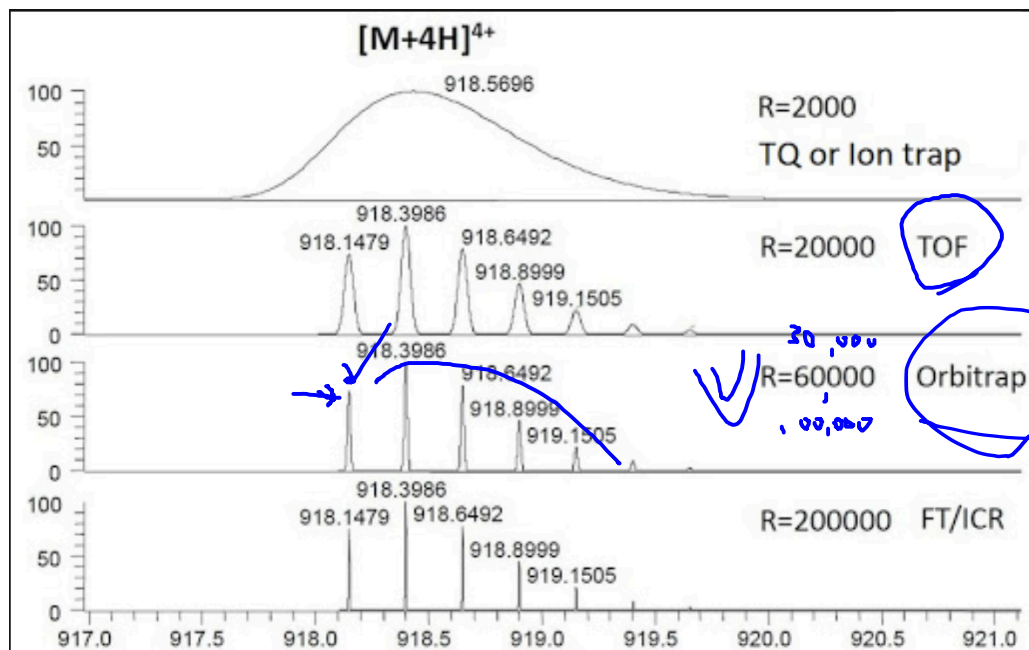
21

분해능의 발전

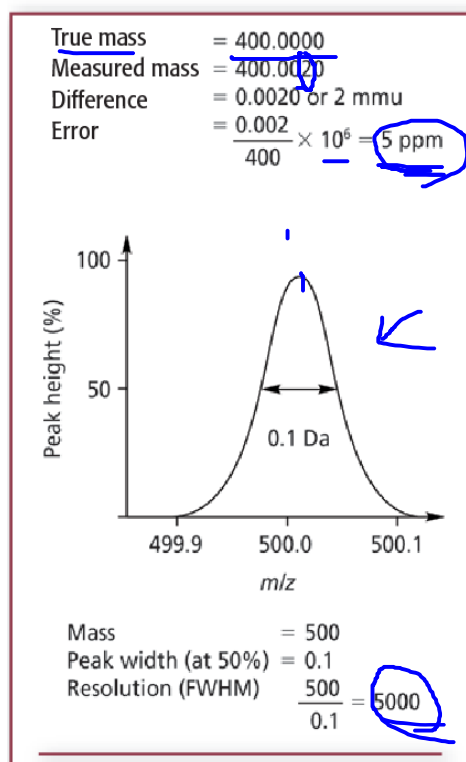
	$m/\delta m$	
1913	<u>13</u>	Thomson ²⁰ Ne ²² Ne
1918	100	Dempster
1919	<u>130</u>	Aston
1937	<u>2000</u>	Aston
1998	<u>8 000 000</u>	Marshall and co-workers

FT-ICR

질량분석기와 분해능

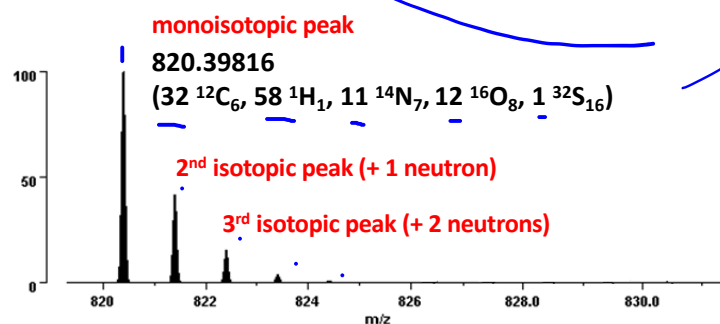
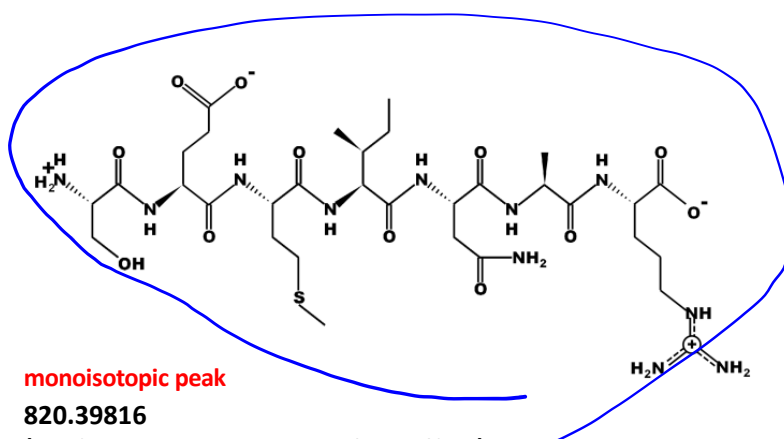
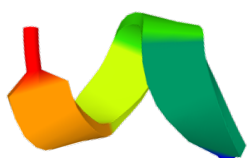


질량스펙트럼의 특성 – mass measurement accuracy(질량측정정확도)



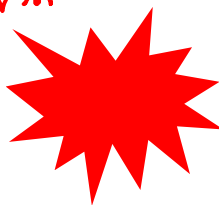
Peptide Sequencing

Precursor spectrum of a peptide 'SEMINAR'



MS 1

precursor

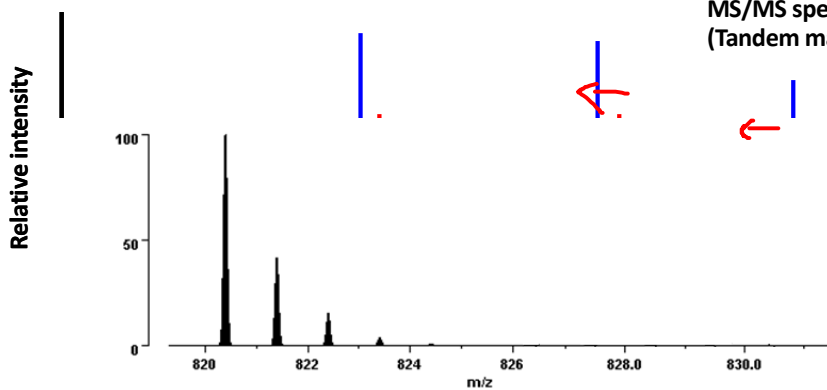


Gas phase dissociation

S/E/M/H/H/R

MS 2

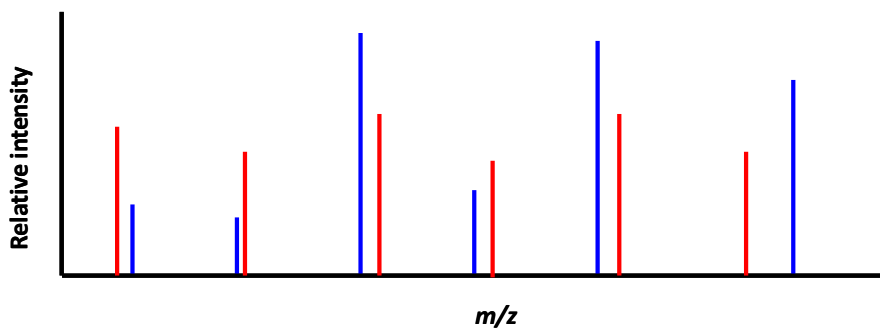
MS/MS spectrum
(Tandem mass spectrum)



27

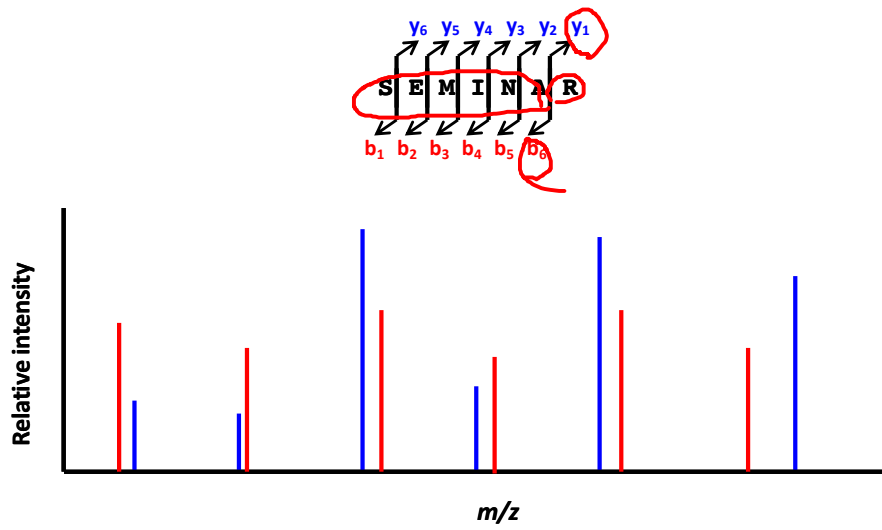
Fragment ion assignments

S E M I N A R



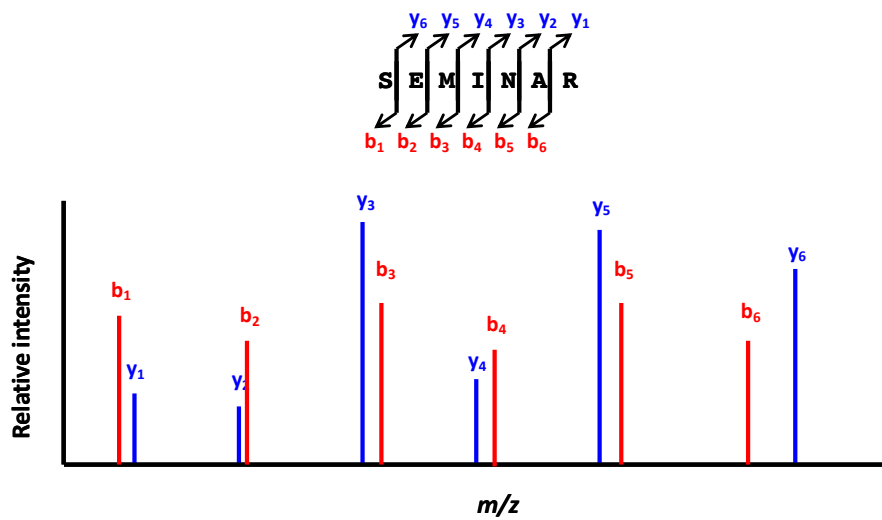
28

Fragment ion assignments



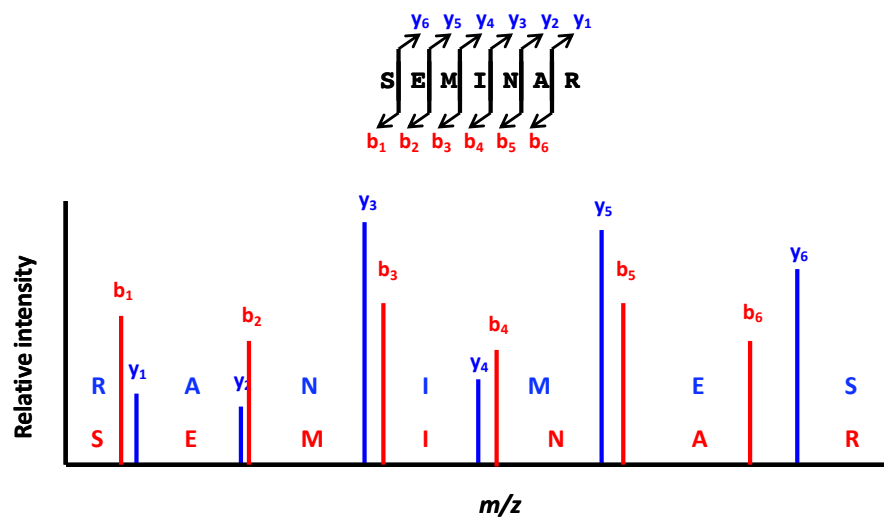
29

Fragment ion assignments



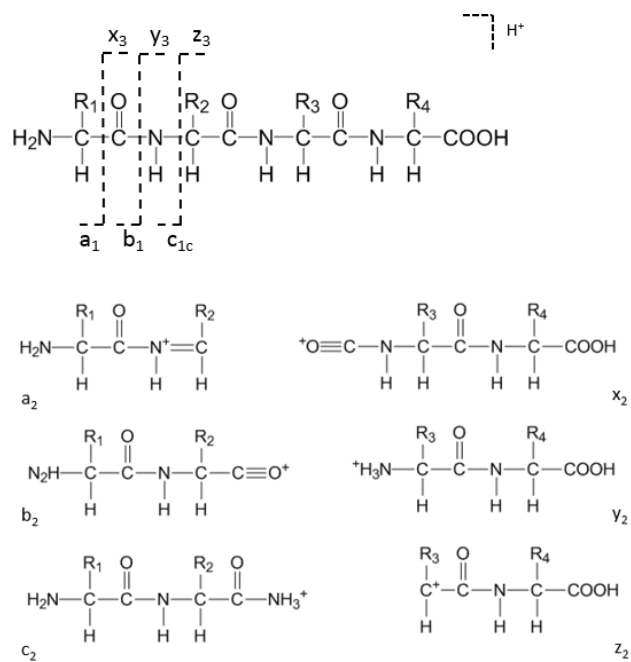
30

Fragment ion assignments



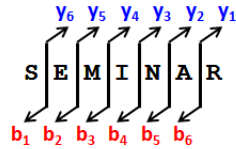
31

Nomenclature of fragment ions

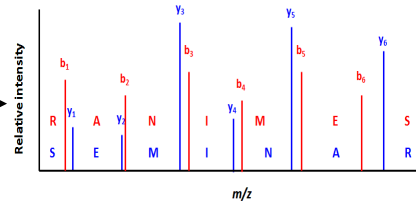


Peptide-spectrum matching

Theoretical precursor mass



Observed precursor mass



Theoretical fragment ion mass list

217.0819
348.1224
461.2064
575.2494
646.2865

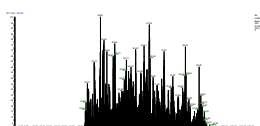
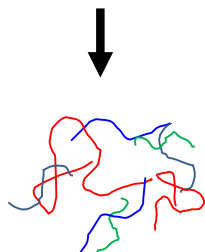
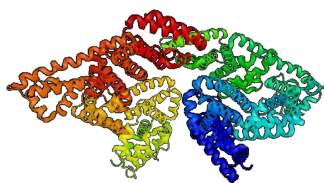
SCORE

Measured fragment ion mass list

217.0823
348.1211
461.2072
575.2490
646.2875

33

Proteomics : Bottom-up approach



MKWVTFISLLLLFSSAYSRGVFRDTHKSEIAHRFKDLGEEHFKGLVLIA
FSQYLQQCPFDEHVK**LVNELTEFAK**TCVADESHAGCEKSLHTLFGDELCK
VASLRETYGDMADCCCKQEPERNECFLSHKDDSPDLPK**LKPDNPNTLCDEF**
KADEKKFWGKYLEIARR**HPYFYAPELLYYANK**YNGVFQECQAEDKGAC
LLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAE
FVEVTK**LVTDLT**KVHK**ECCHGDLLECADDR**ADLAKYICDNQDTISSKLE
CCDKPLEKSHCIAVEKDAIPENLPPLTADFAEDK**DVCKNYQEAK**DAFL
GSFLYEYSRR**HPEYAVSVLLRLAK**EYEATLECCAKDDPHACYSTVFDKL
KHLVDEPQNLIKQNCQFEKLGEYGFQNALIVR**YTRKVPQVSTPTLVEVS**
RSLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTCC
TESLVNRRPCFSALTPDETYVPKAFDEK**LFTFHADICTLPDTEK**QIKKQT
ALVELLKHKPKATEEQLKTVMENFVAFVDKCAADDKEACFAVEGPKLIV
STQTALA

ECCHGDLLECADDR

LVNELTEFAK

DVCKNYQEAK

HPYFYAPELLYYANK

LFTFHADICTLPDTEK

HPEYAVSVLLRLAK

Data analysis

34

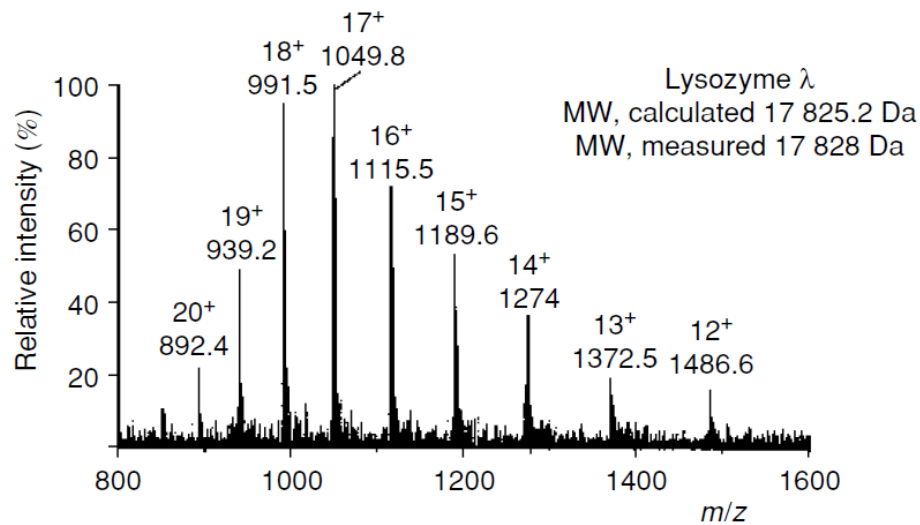
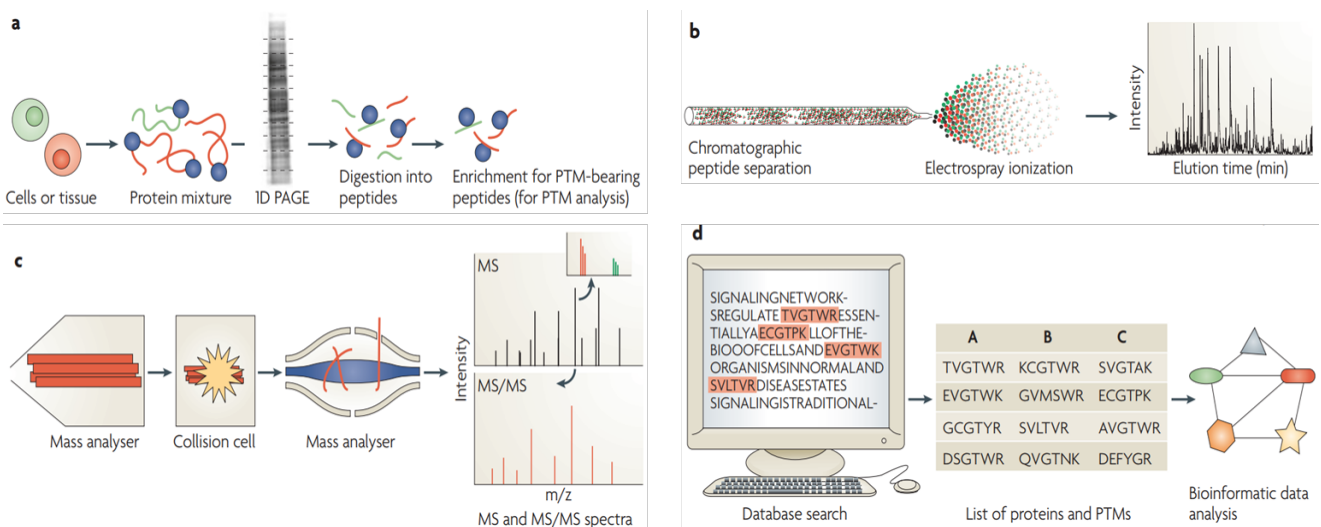


Figure 1.23

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 $17\,828 \pm 2.0$ Da.

35

Widely employed, conventional proteomics workflow



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

36

Human Proteome Map and Genome Annotation

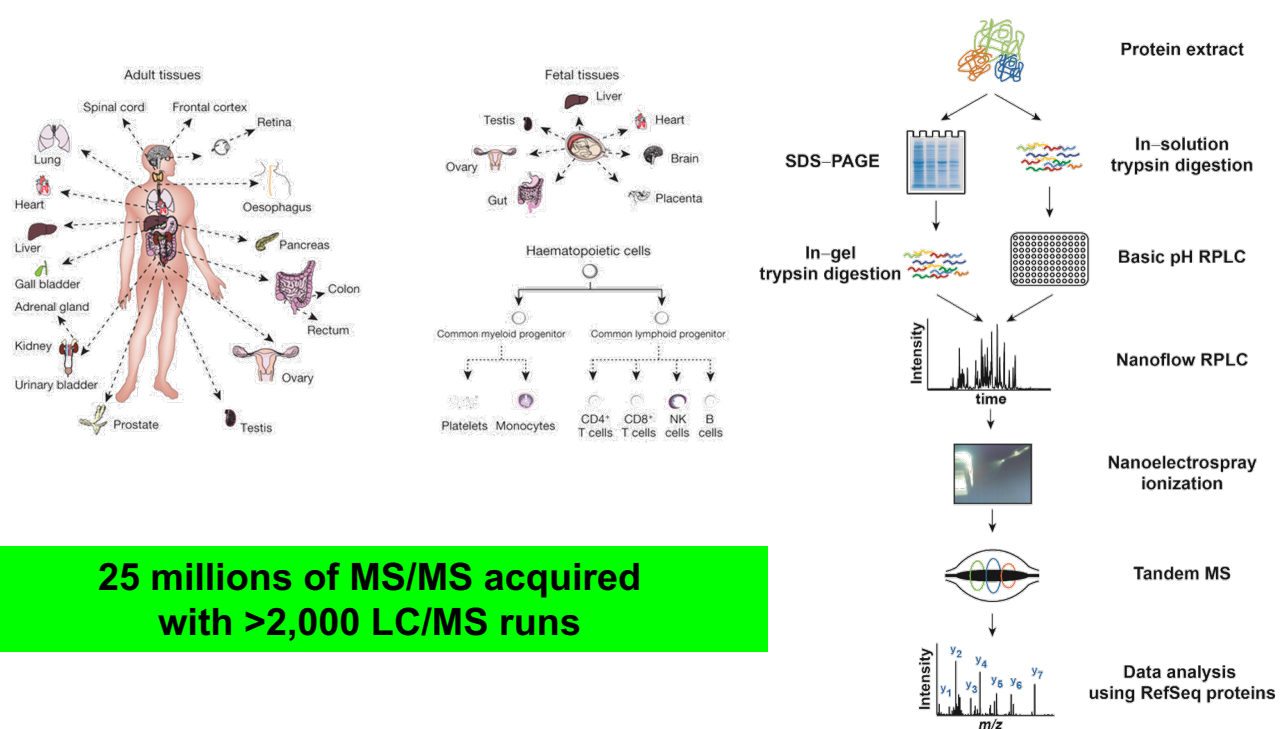
The Human Genome Map



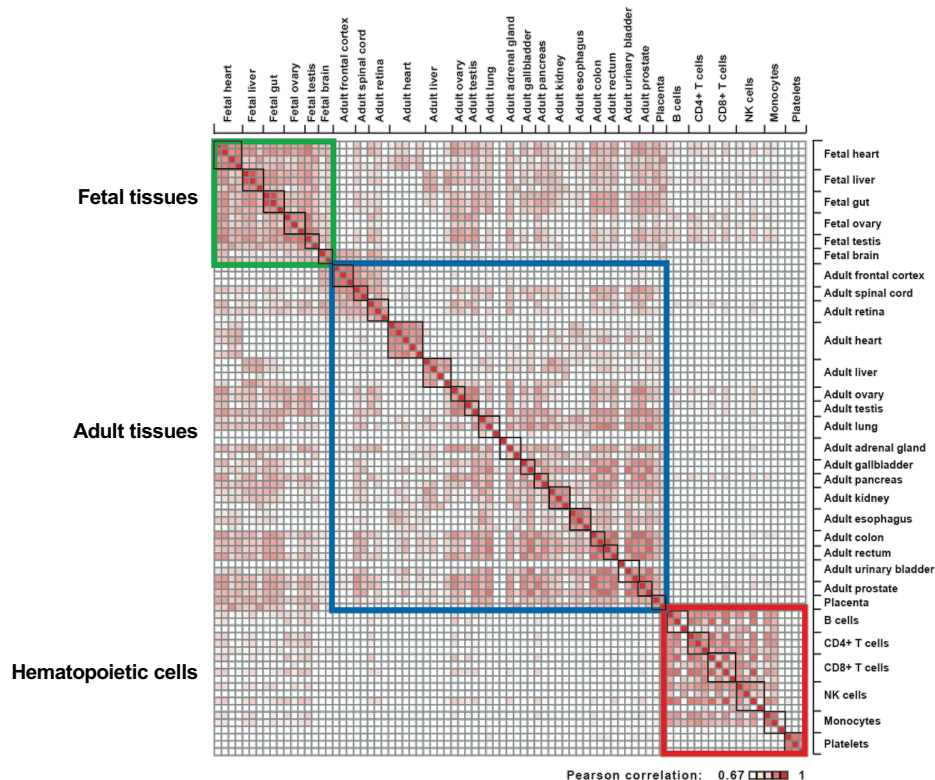
A Draft Map of the Human Proteome

39

Samples and Workflow



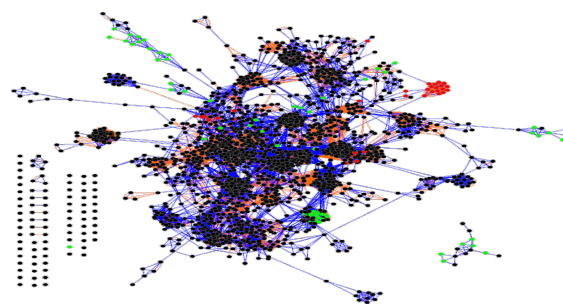
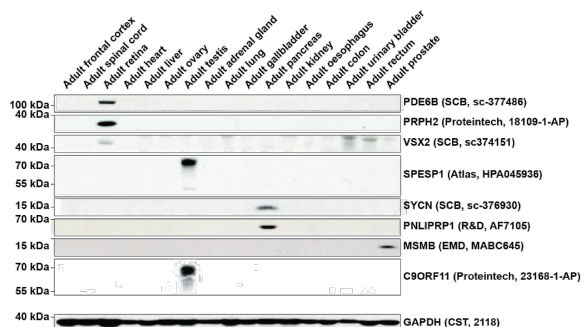
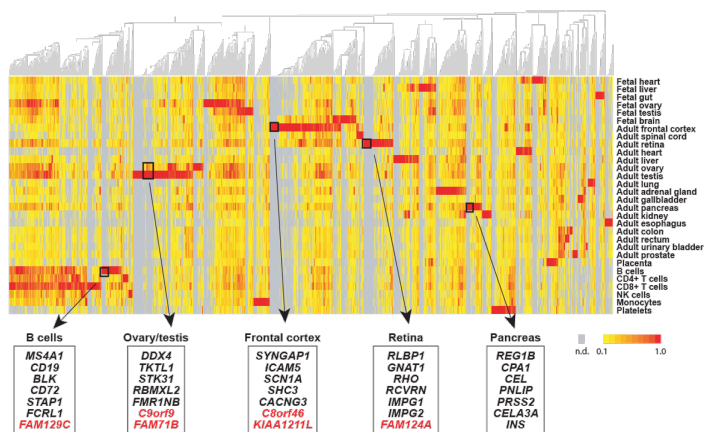
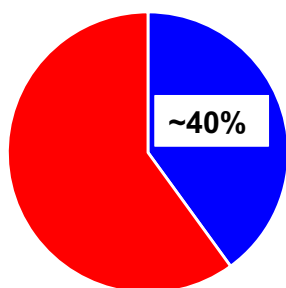
Proteome Correlation



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

41

Tissue-specific proteome expression



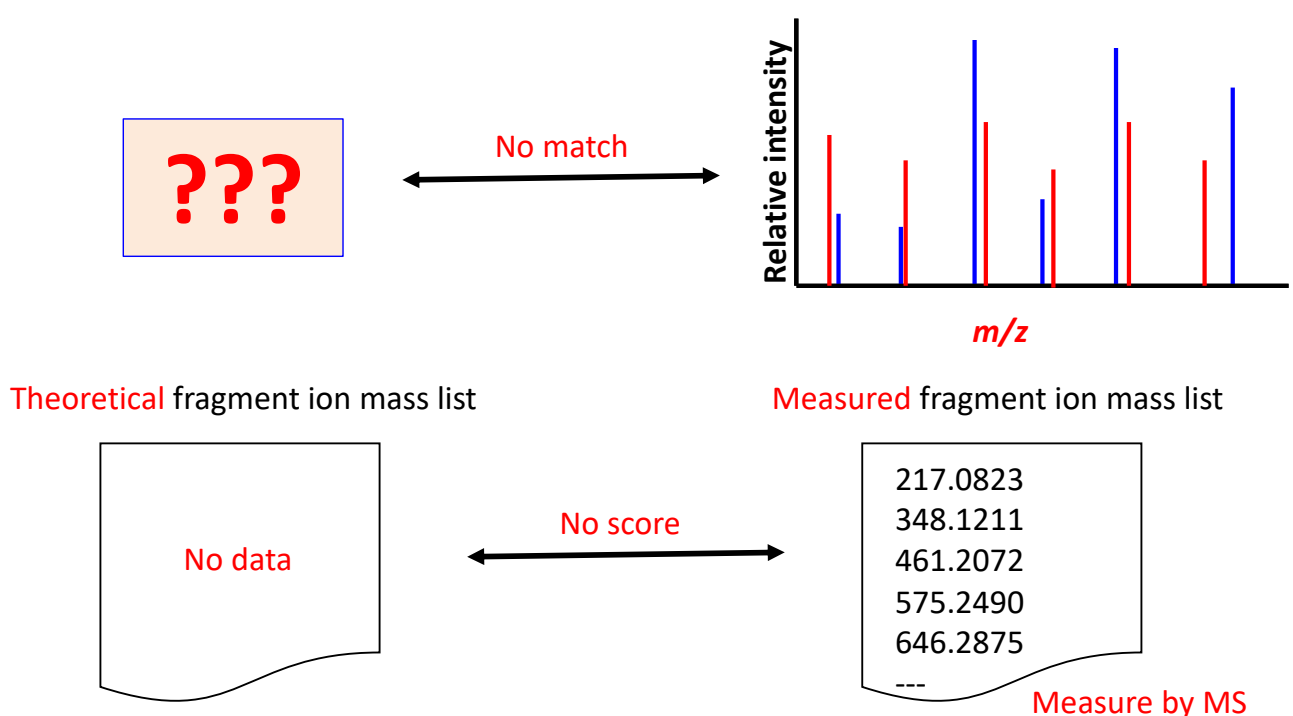
42

Modified from Kim et al. Nature 2014

- 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?**

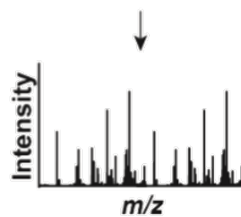
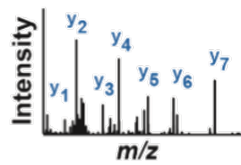
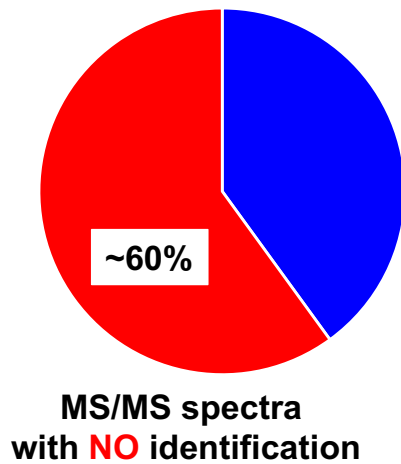
43

A good tandem mass spectrum remains unmatched!



44

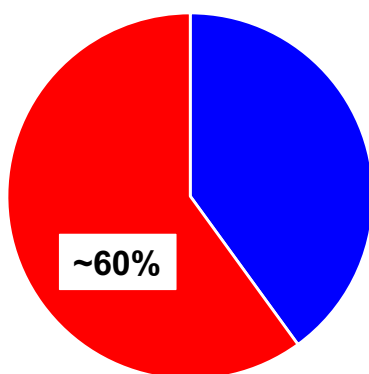
~60% of unidentified MS/MS collected for proteogenomic analysis



Collection of 'unidentified' MS/MS spectra

45

Possible explanations



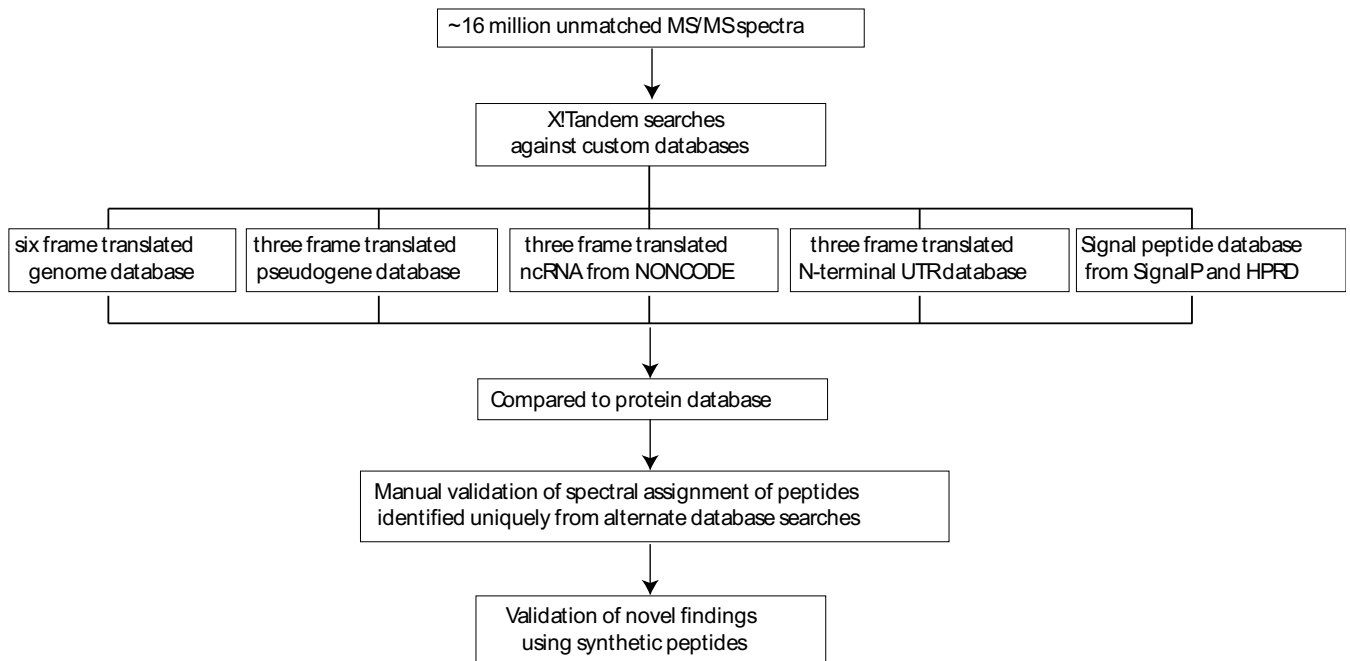
- They may be just poor quality MS/MS spectra
- Good quality MS/MS spectra may be originated from
 1. Novel PTMs
 2. Novel Isoforms
 3. **Novel Genes**

Gene

Non-coding RNA

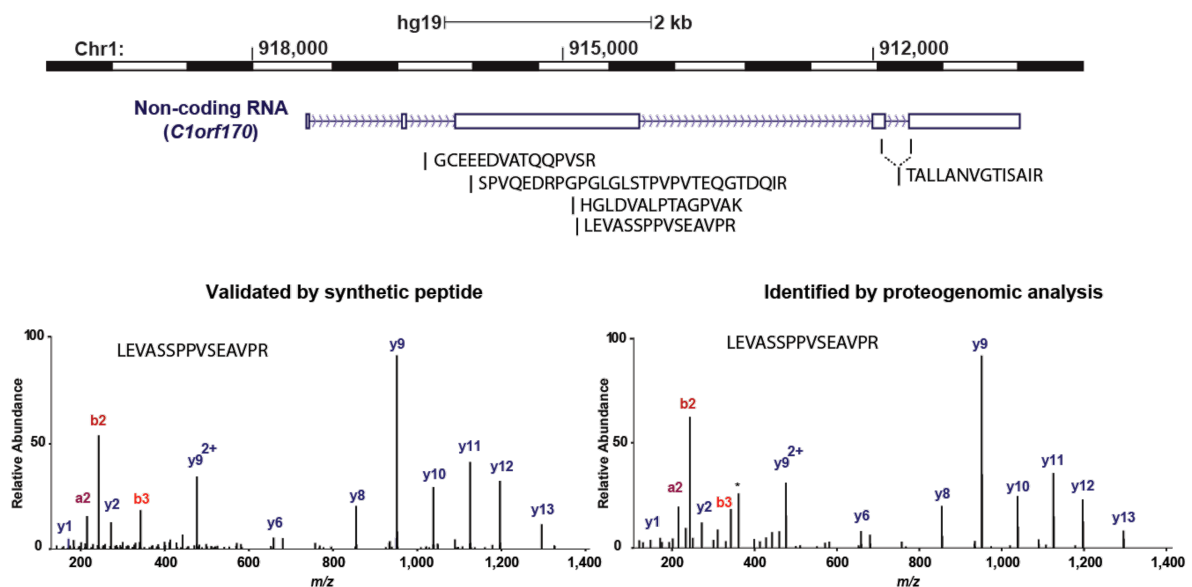
46

Proteogenomics analysis pipeline



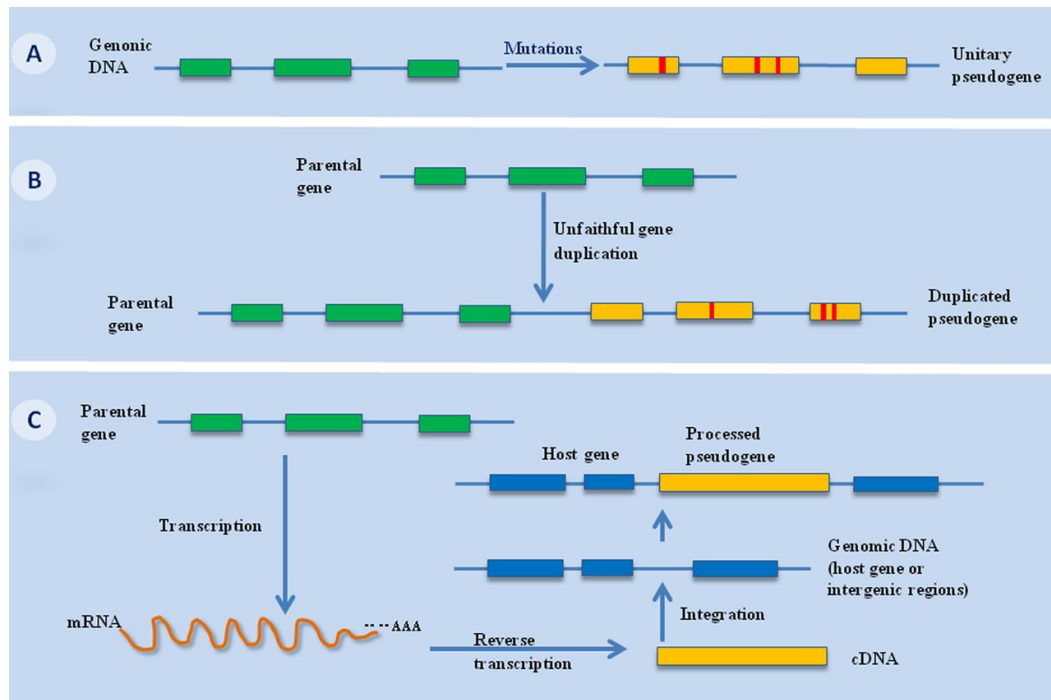
47

Non-coding RNA?

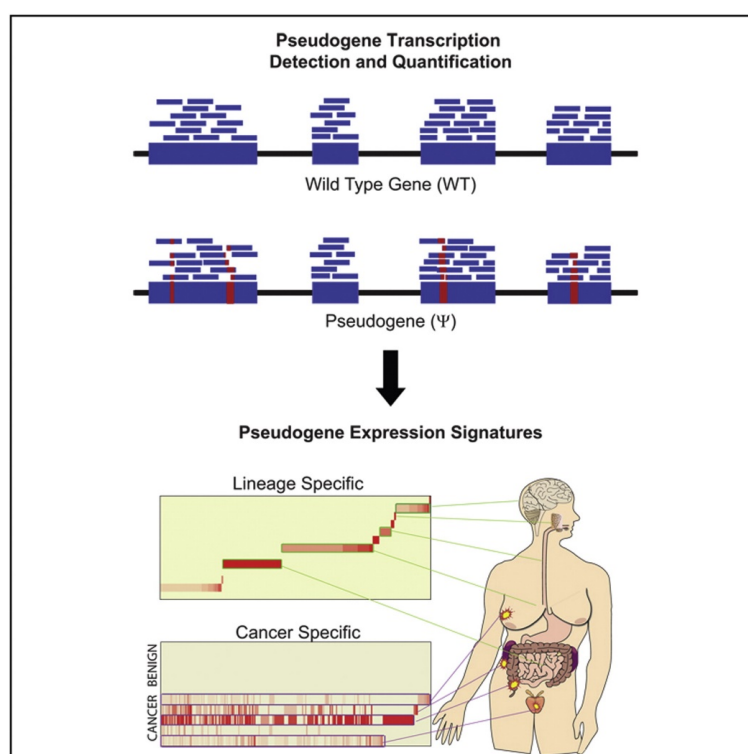


48

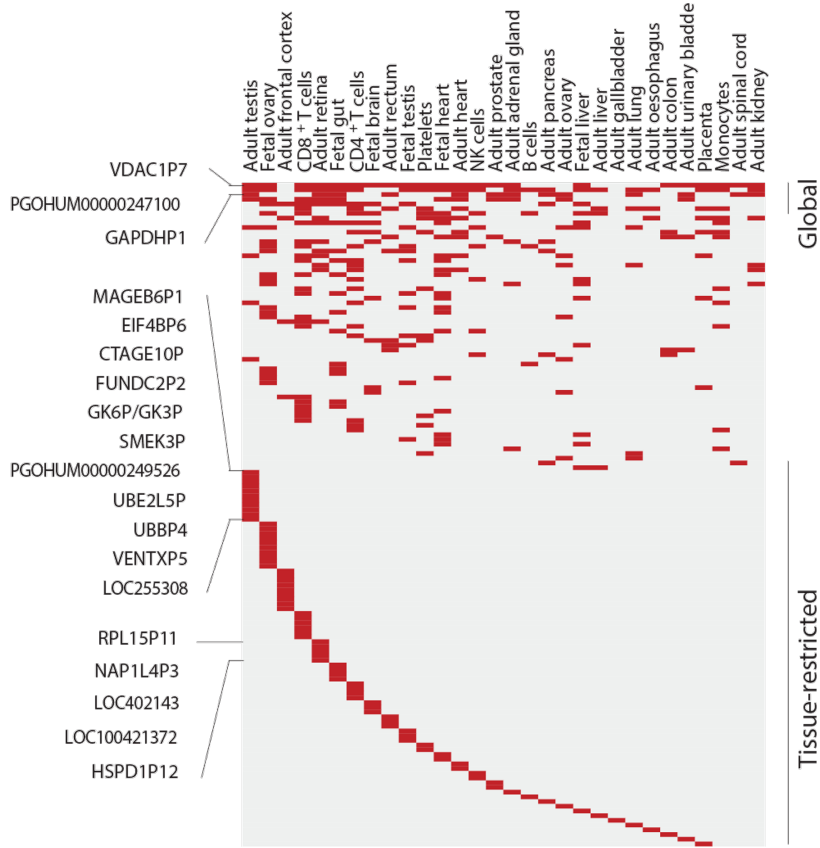
Pseudogene은 화석 유전자?



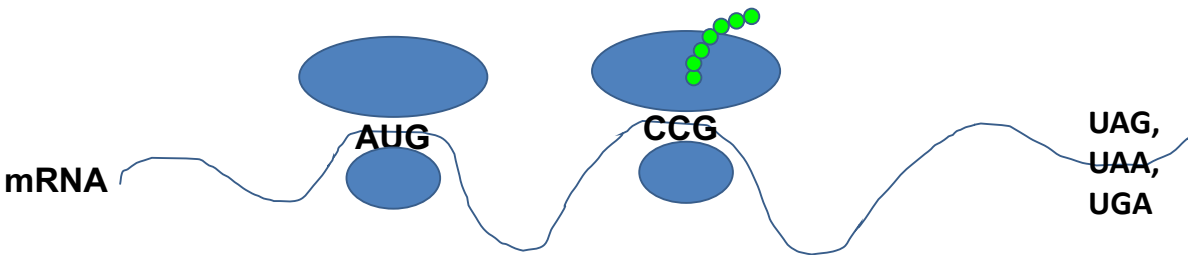
Transcription of pseudogenes



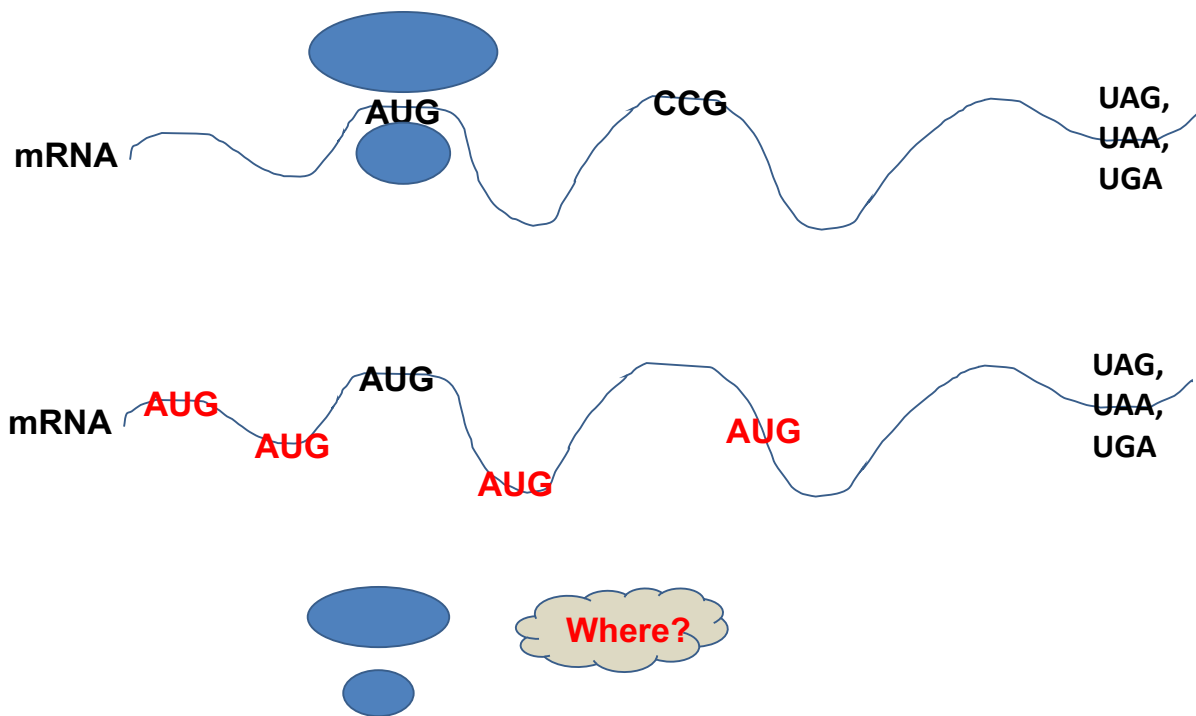
Translation of pseudogenes



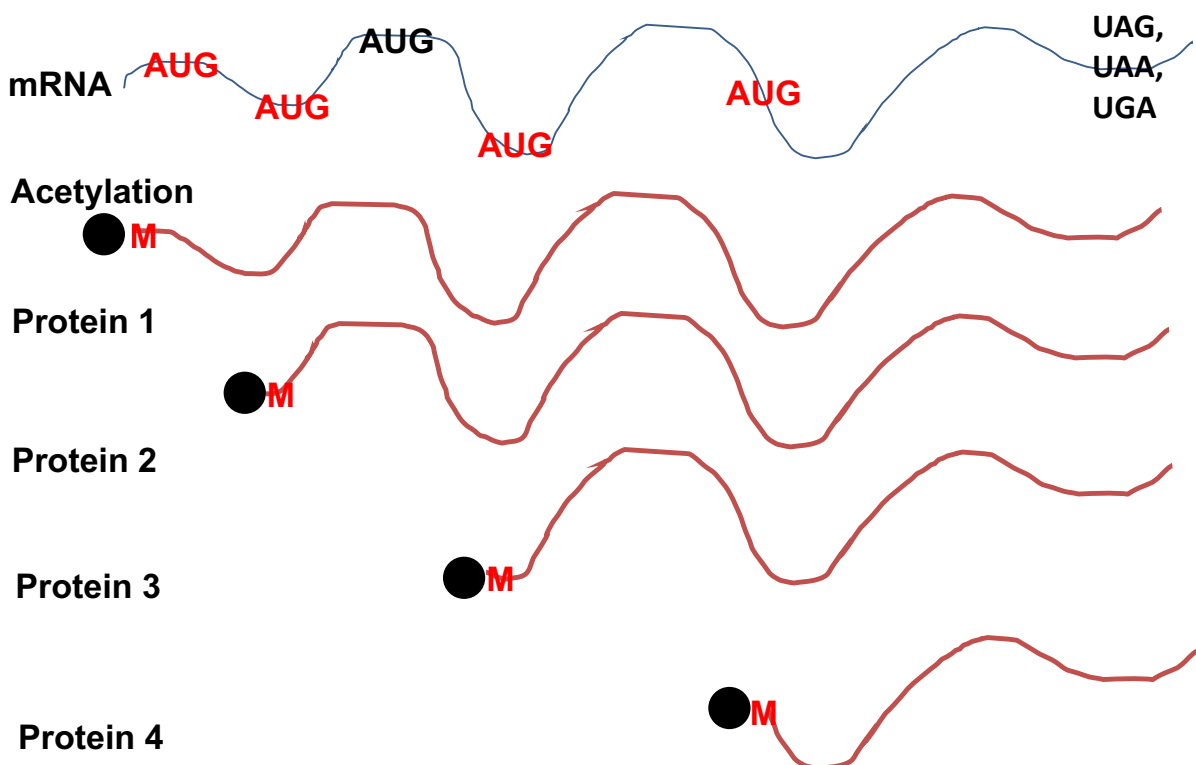
Translation



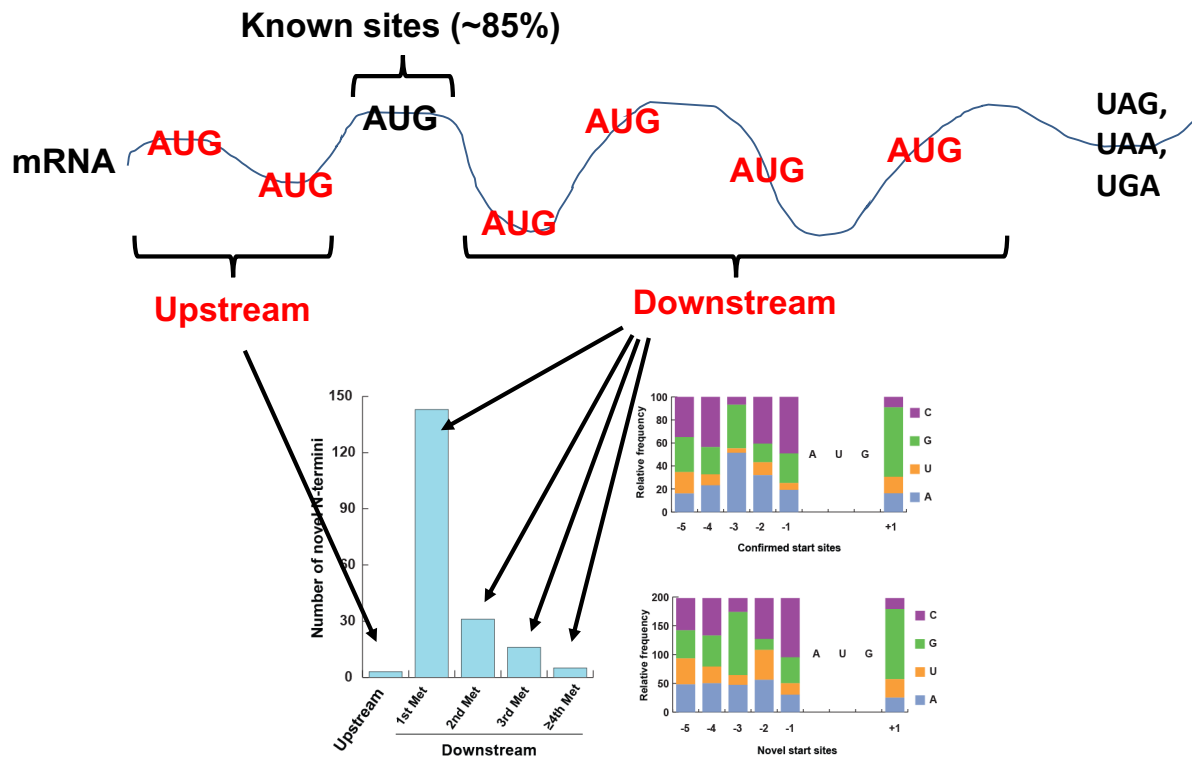
Translation start site?



Method of a newer protein sequence database

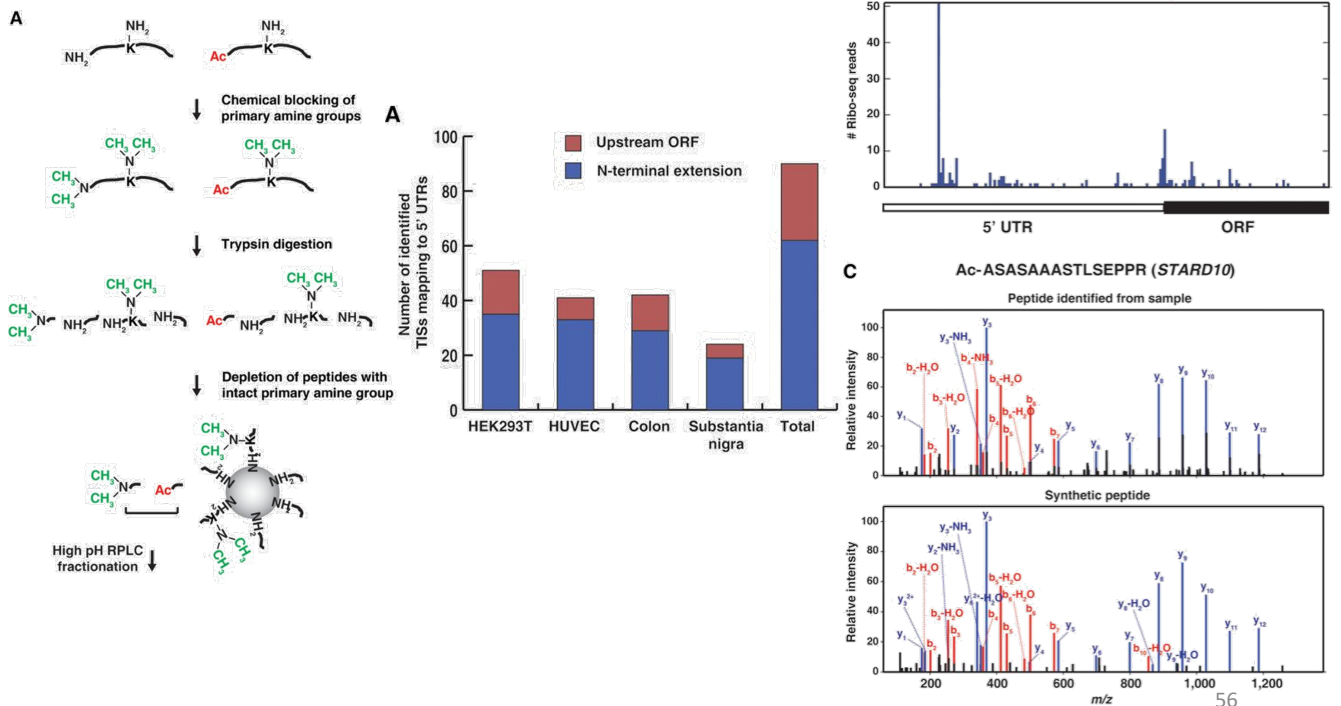


Heterogeneous translation start sites of a messenger



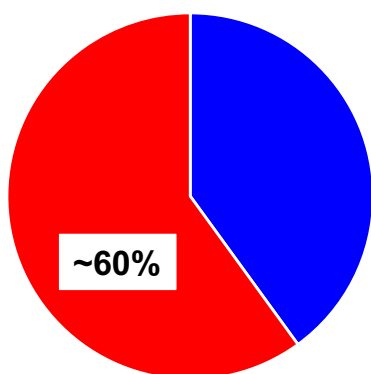
55

Further confirmation of the heterogeneous starts



Post-translational Modifications

Possible explanations



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

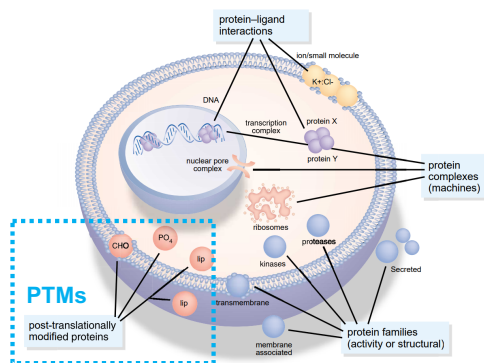
1. **Novel PTMs**
2. **Novel Isoforms**
3. **Novel Genes**

Gene

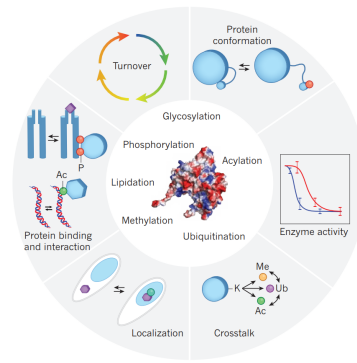
Protein-coding mRNA



Post-translational modifications, unexplored bioinformatics field



Patterson and Aebersold, *Nat Gen*, 2003

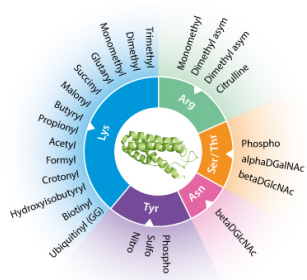


Aebersold and Mann, *Nature*, 2016

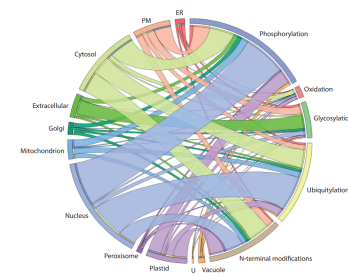
Systems-wide study of PTMs is indispensable for a more comprehensive understanding of the human proteome.

59

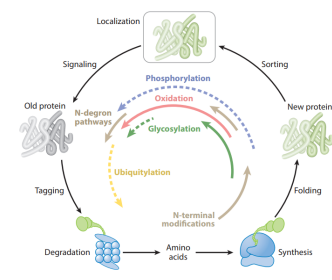
PTMs play a central role in regulating biological processes



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**.

60

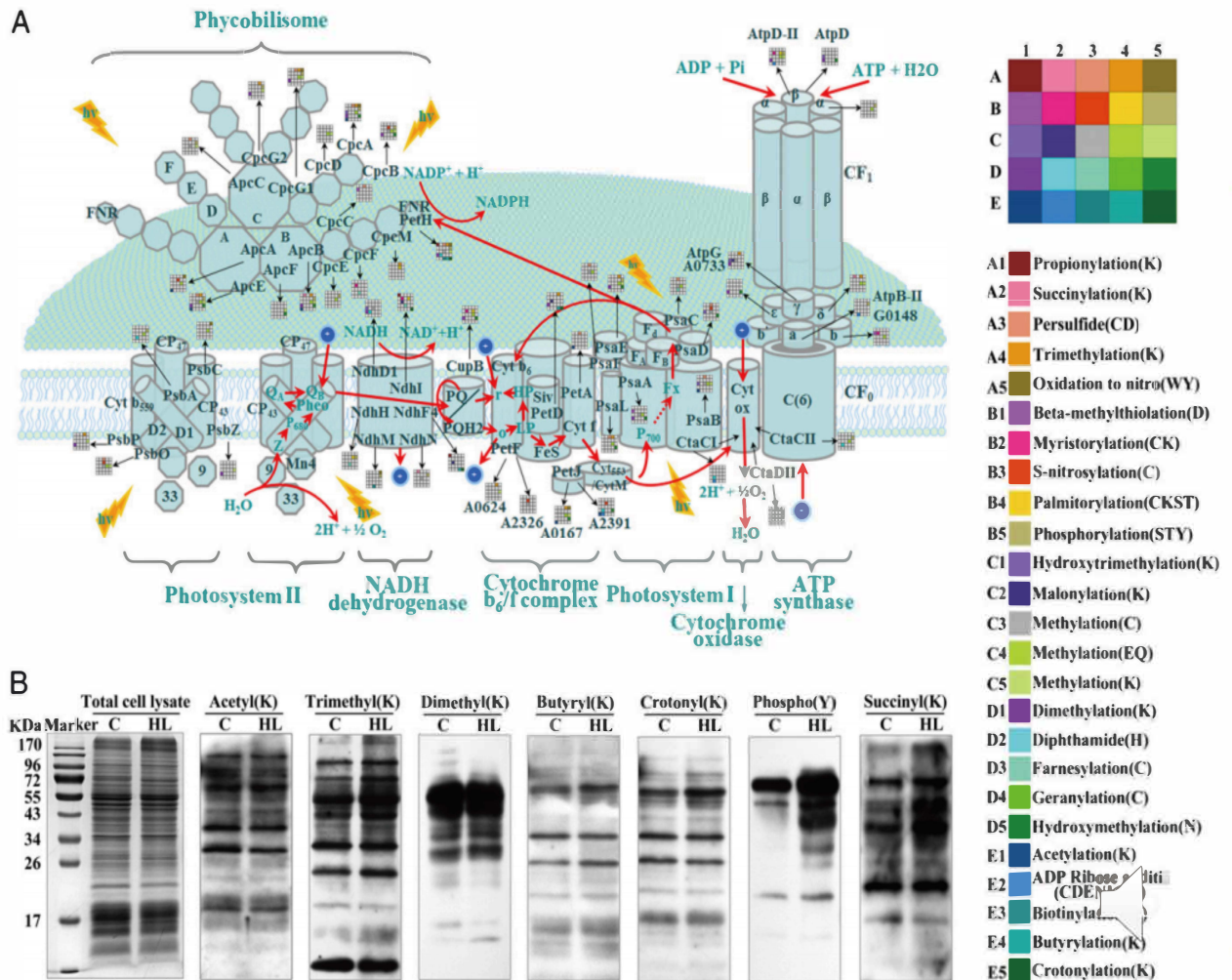
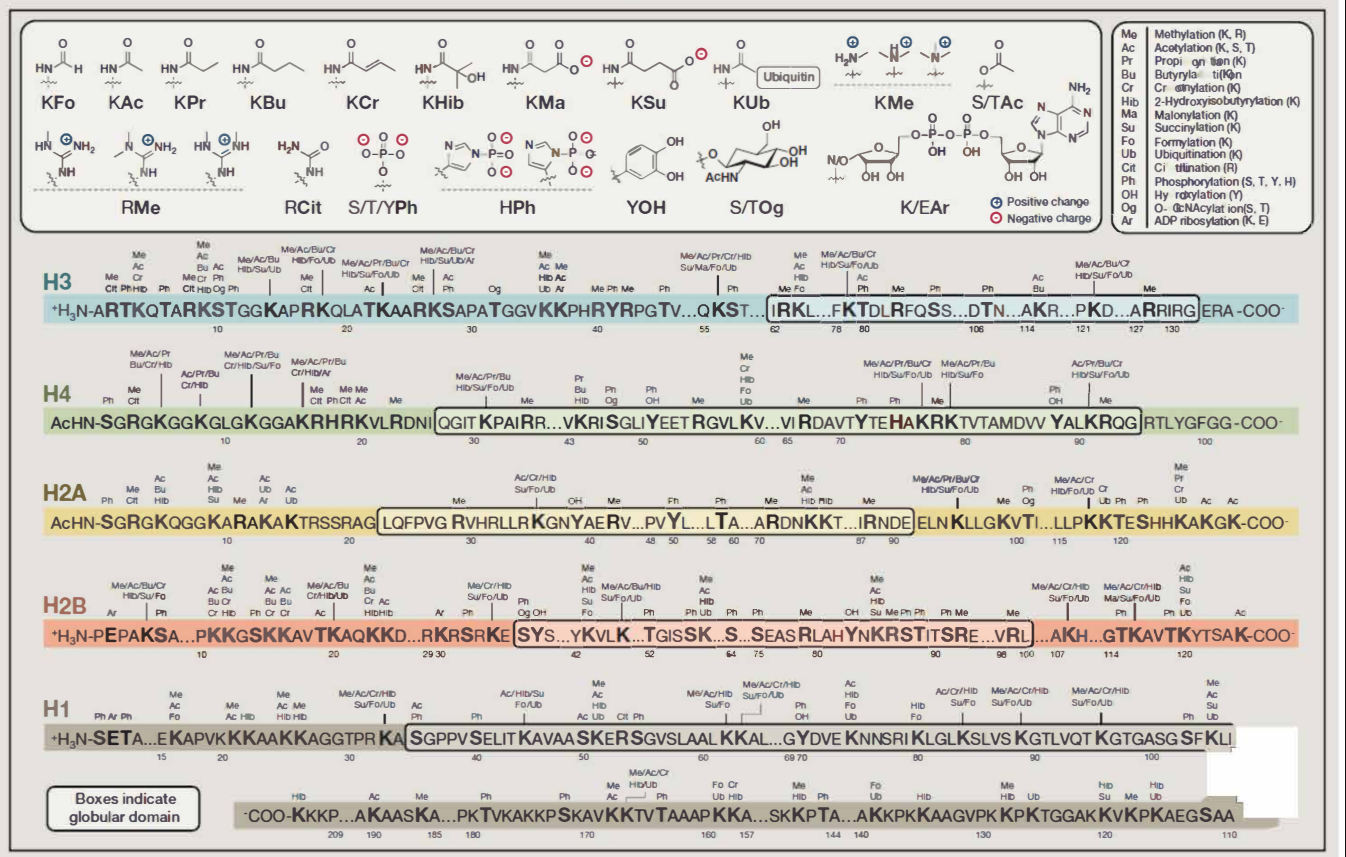
SnapShot: Histone Modifications

He Huang,¹ Benjamin R. Sabari,² Benjamin A. Garcia,³ C. David Allis,² and Yingming Zhao¹

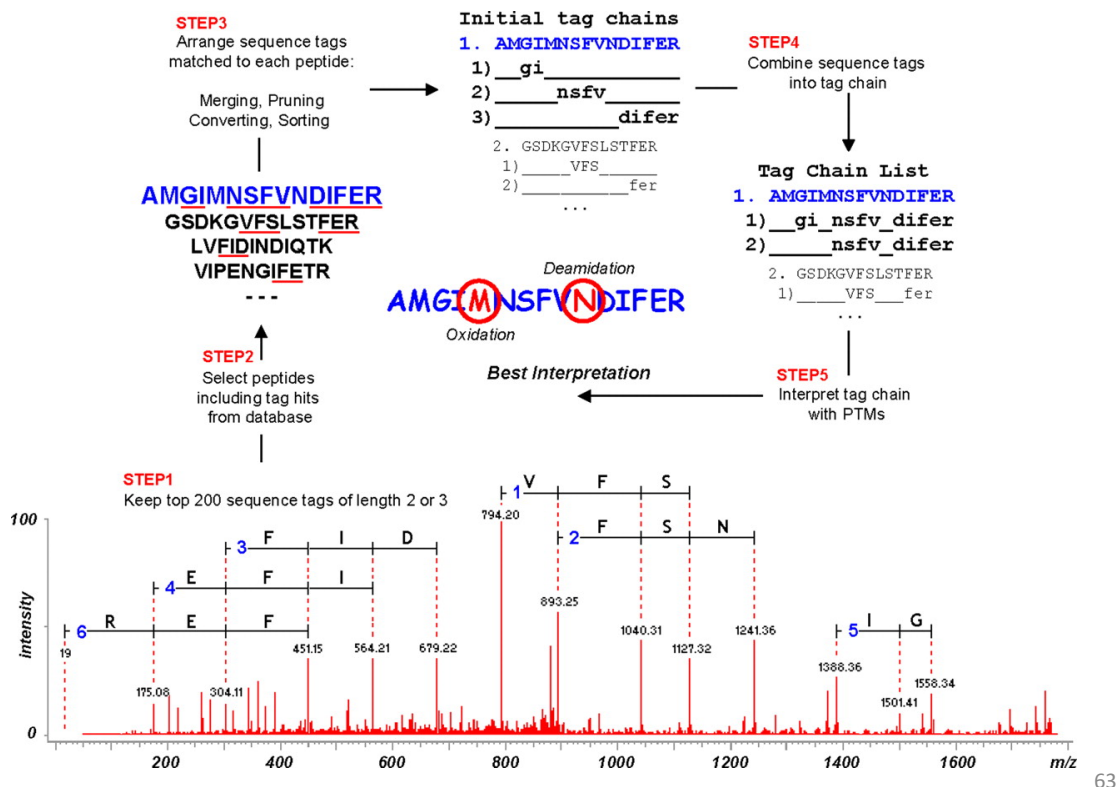
¹Ben May Department of Cancer Research, The University of Chicago, Chicago, IL 60637, USA

²Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10021, USA

³Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA



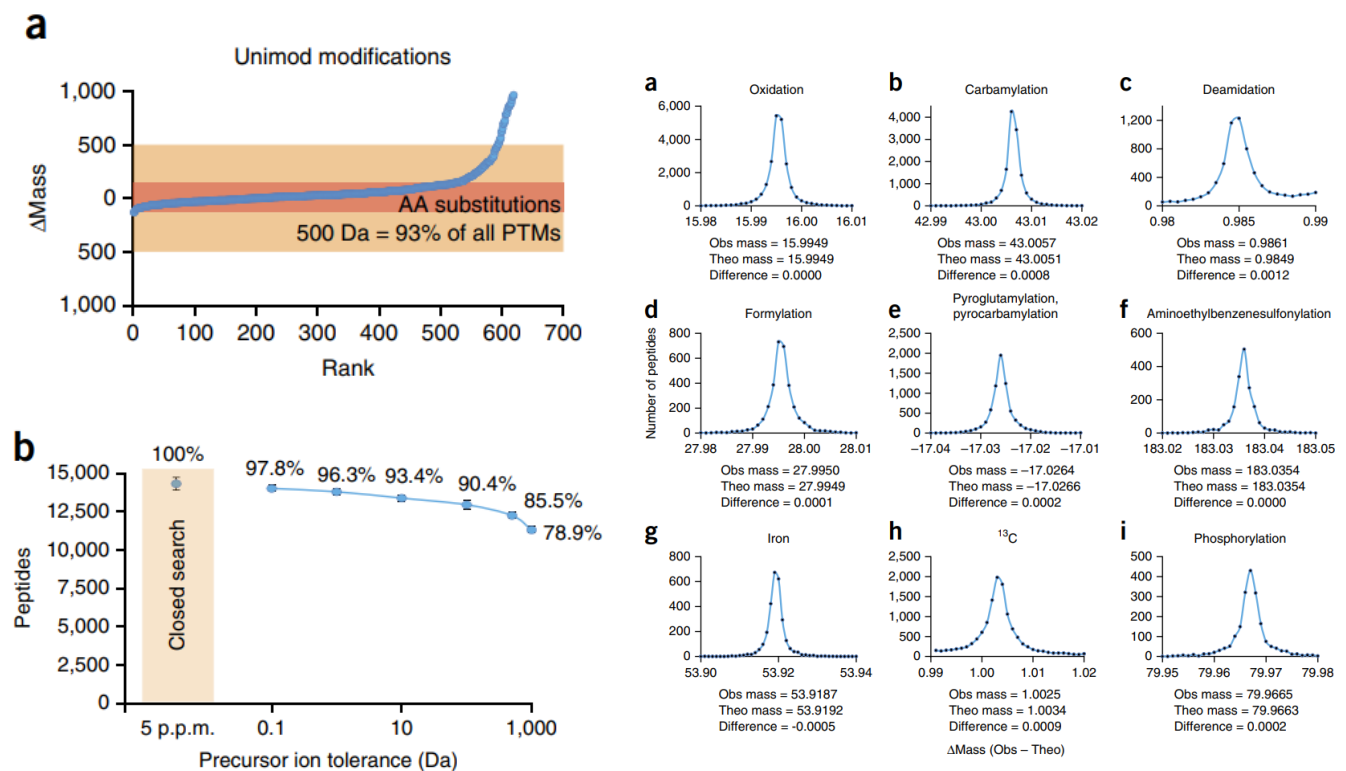
PTM analysis – MODi algorithm



Na et al. *Molecular and Cellular Proteomics*

63

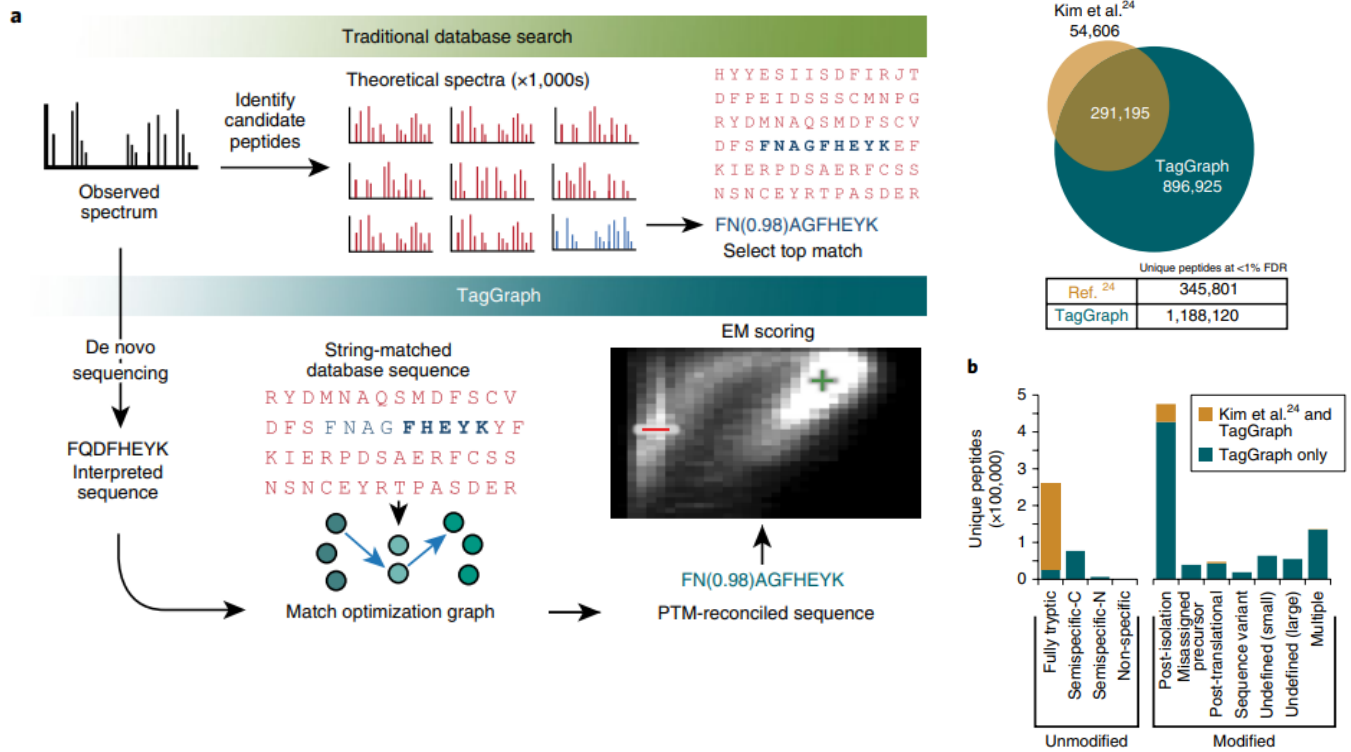
Open Search Strategy



Gygi et al. *Nature Biotechnology*

64

TagGraph algorithm



Ellias et al. *Nature Biotechnology*

Deep learning-based PTM ID

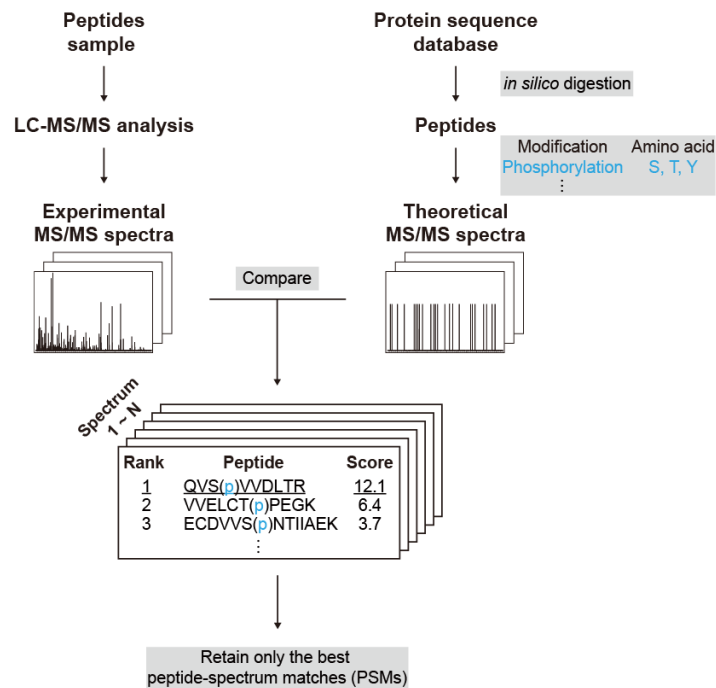


Yang et al. *JPR*

Conventional method of PTM identification

- Database search: identifying the peptide sequences of experimental MS/MS spectra.

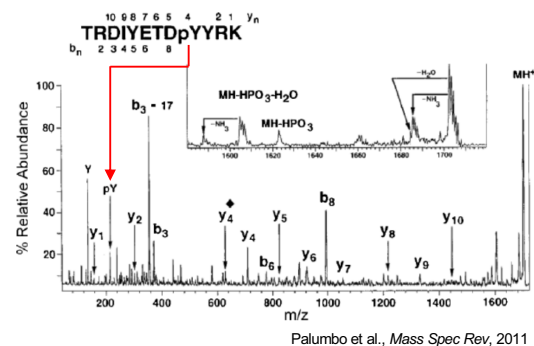
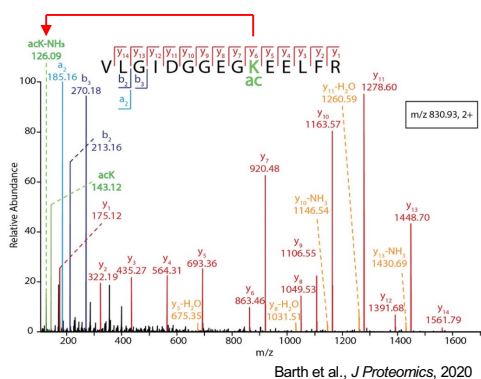
Database search



67

PTM signature in MS/MS spectra

- Diagnostic ions can be signatures for PTMs.

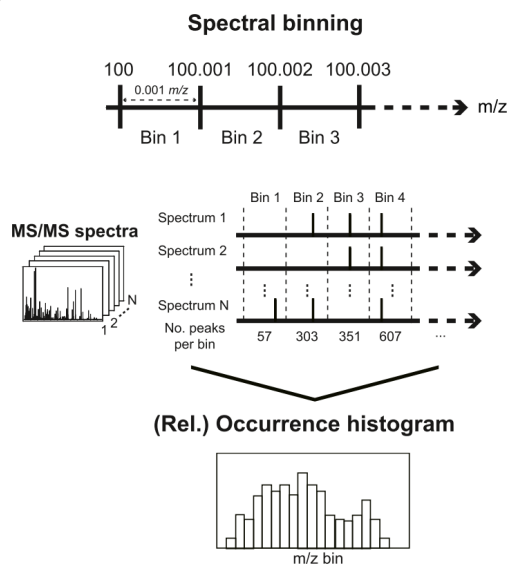


Idea: PTM diagnostic ions can be utilized to facilitate the identification of PTMs in MS/MS data.

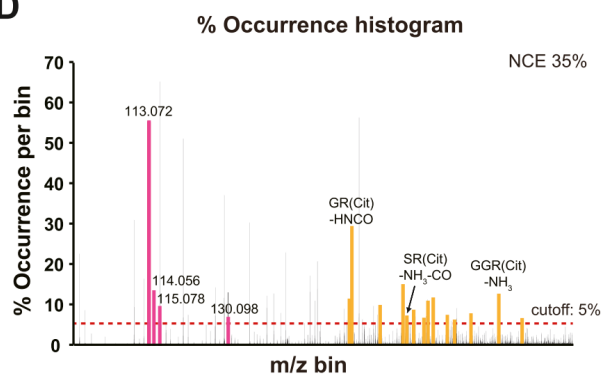
68

Screening of citrullination signatures

C



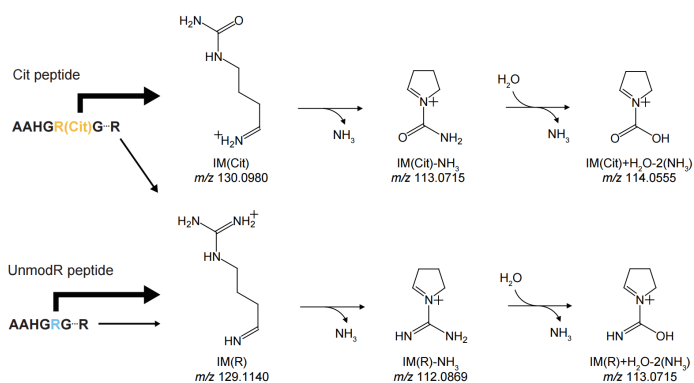
D



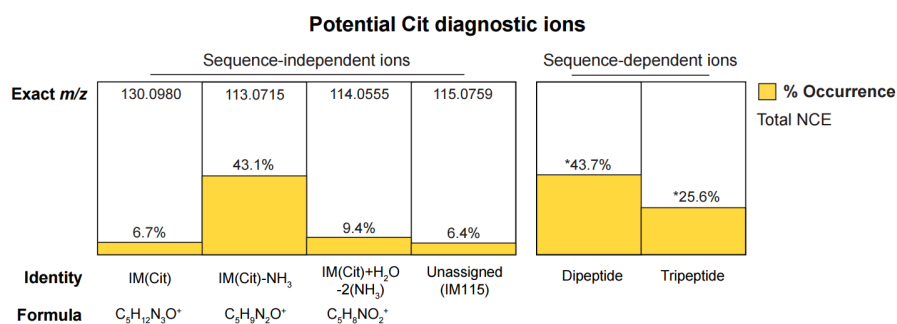
69

Discovery of citrullination diagnostic ions

C

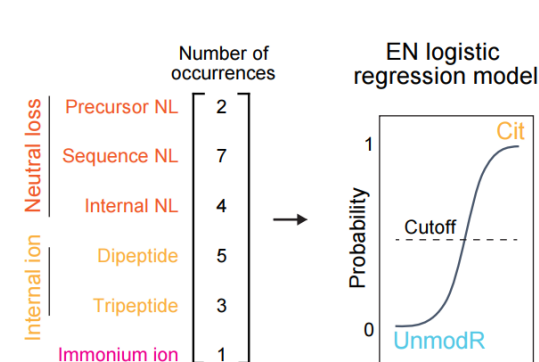


E



70

Development of statistical model

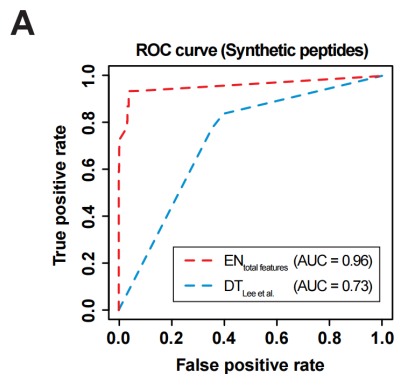


Model

Classification of Cit PSMs

Logistic regression model with elastic net (EN) regularization

'Valid' if probability > 0.5
'Invalid' otherwise

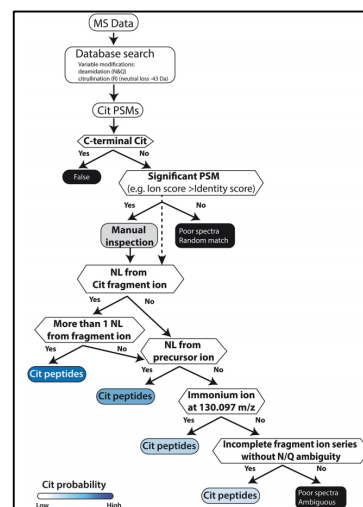


EN model

Sensitivity: 0.94

Specificity: 0.96

Accuracy: 0.95

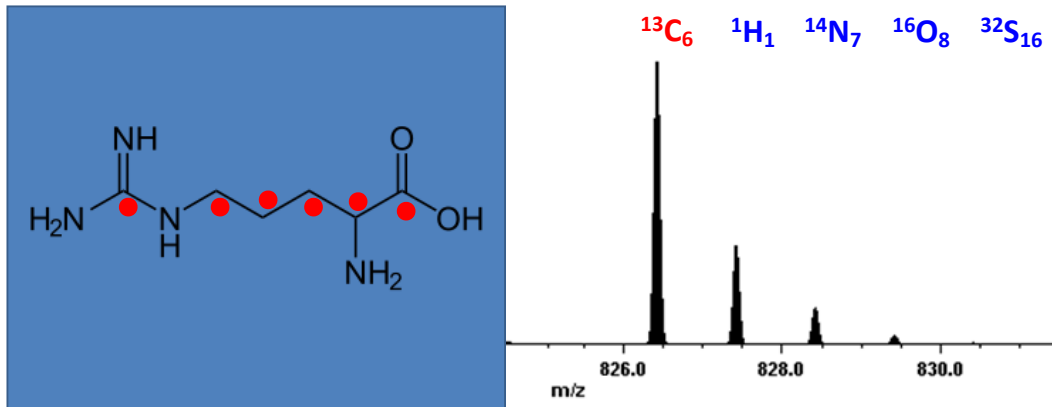


71

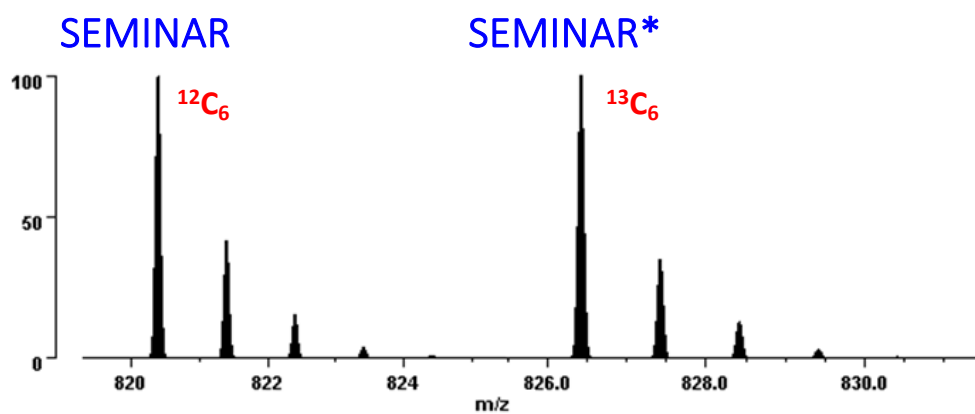
Huh et al. *Analytical Chemistry*

Quantitation

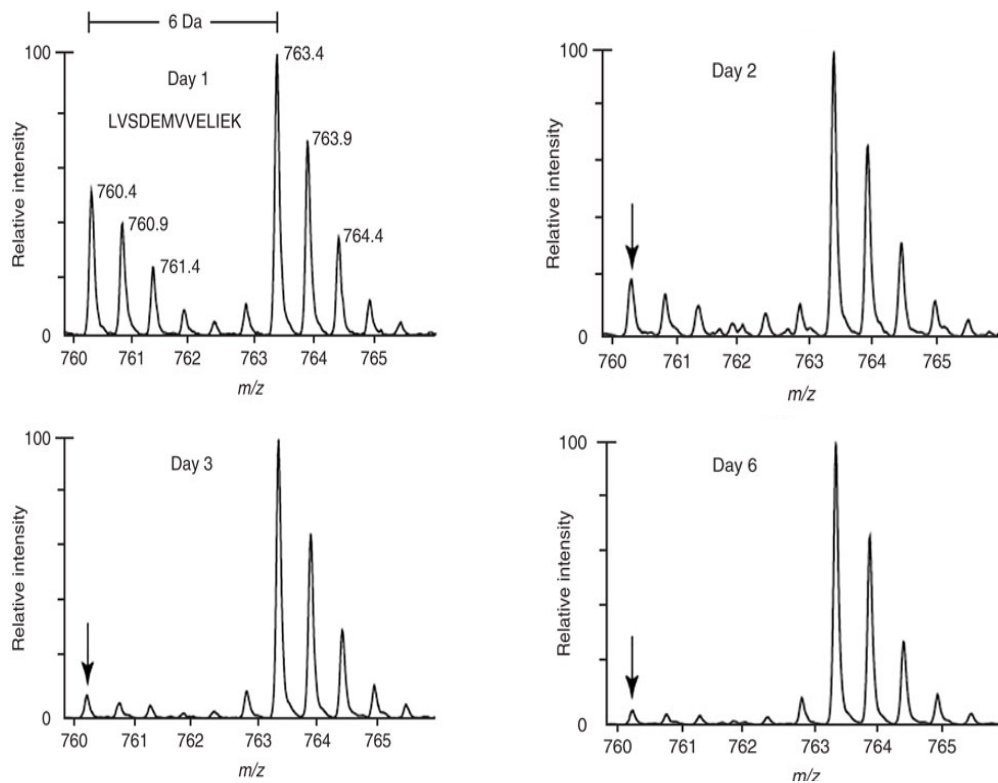
SEMINAR*



SEMINAR:SEMINAR* = 1:1



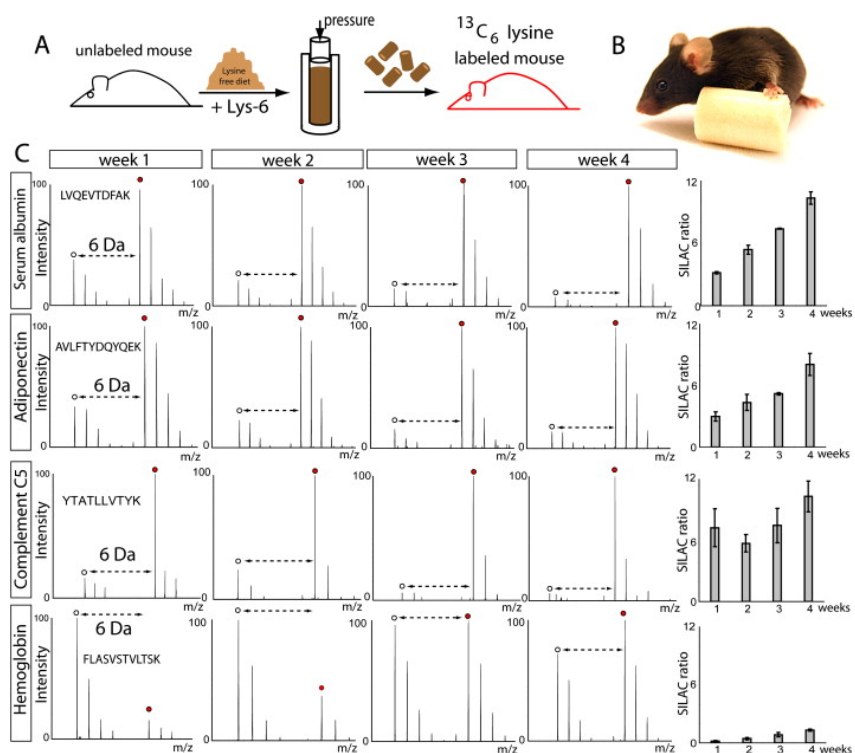
Metabolic labeling of the whole cellular proteome



Gowda et al. *Nature Protocol*

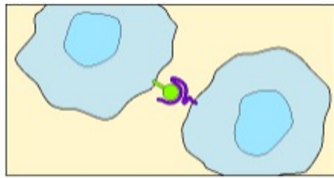
75

In vivo SILAC labeling

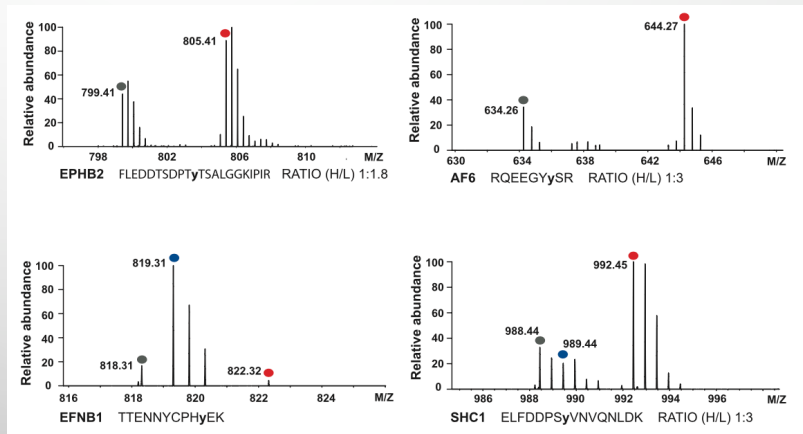
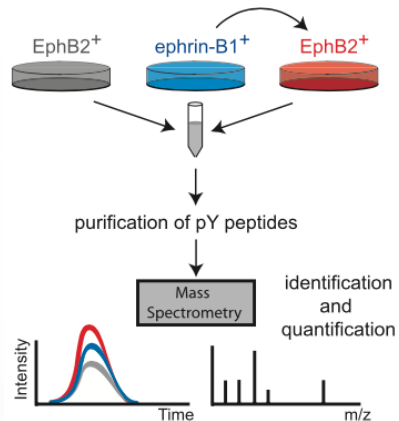
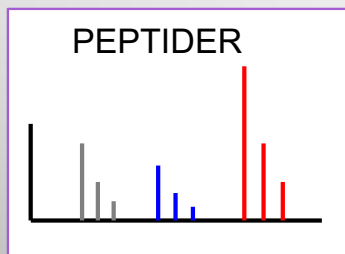
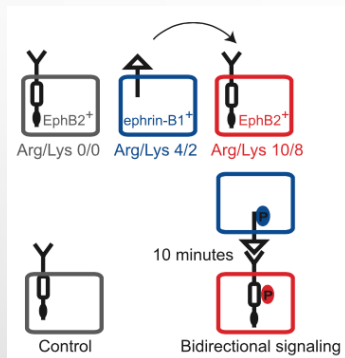


Mann et al. *Cell*

76

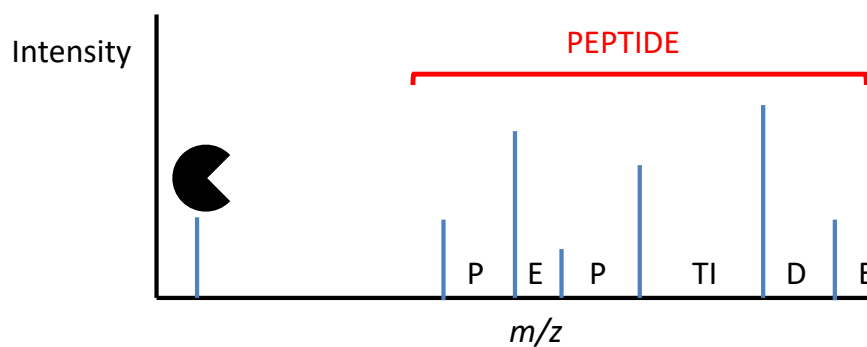
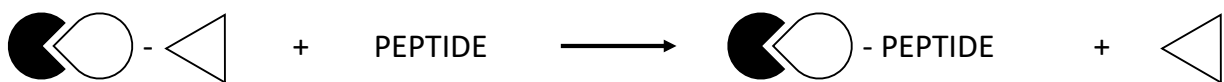


Juxtacrine

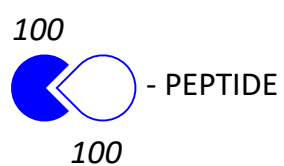
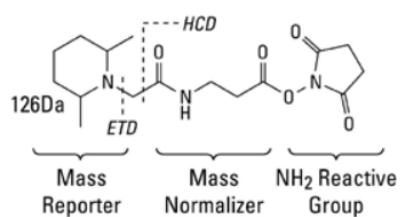


Jørgensen et al. *Science*

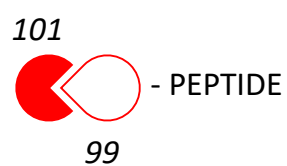
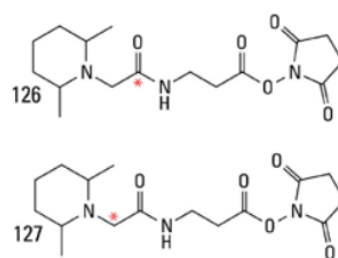
Isotope labeling to clinical samples



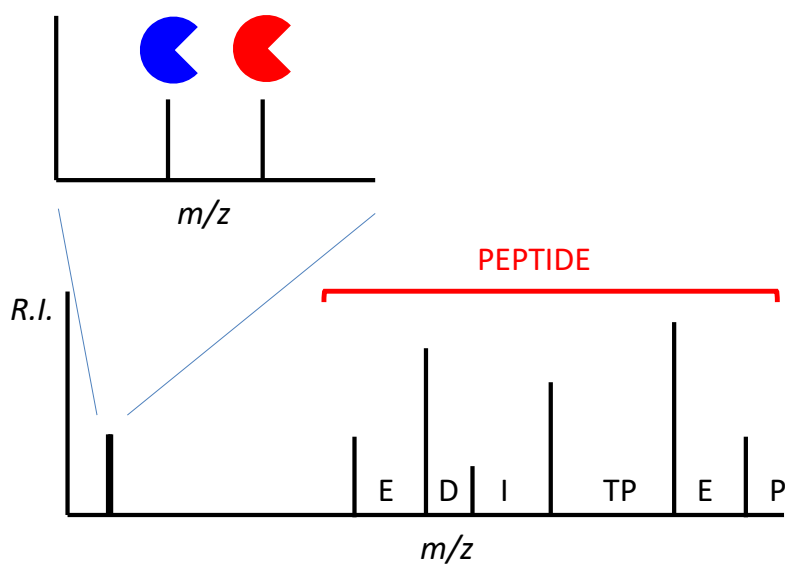
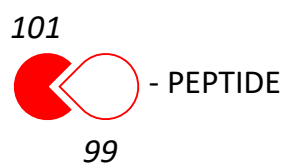
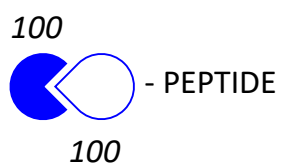
A. TMTzero Reagent (TMT⁰)



B. TMTduplex Reagents (TMT²)



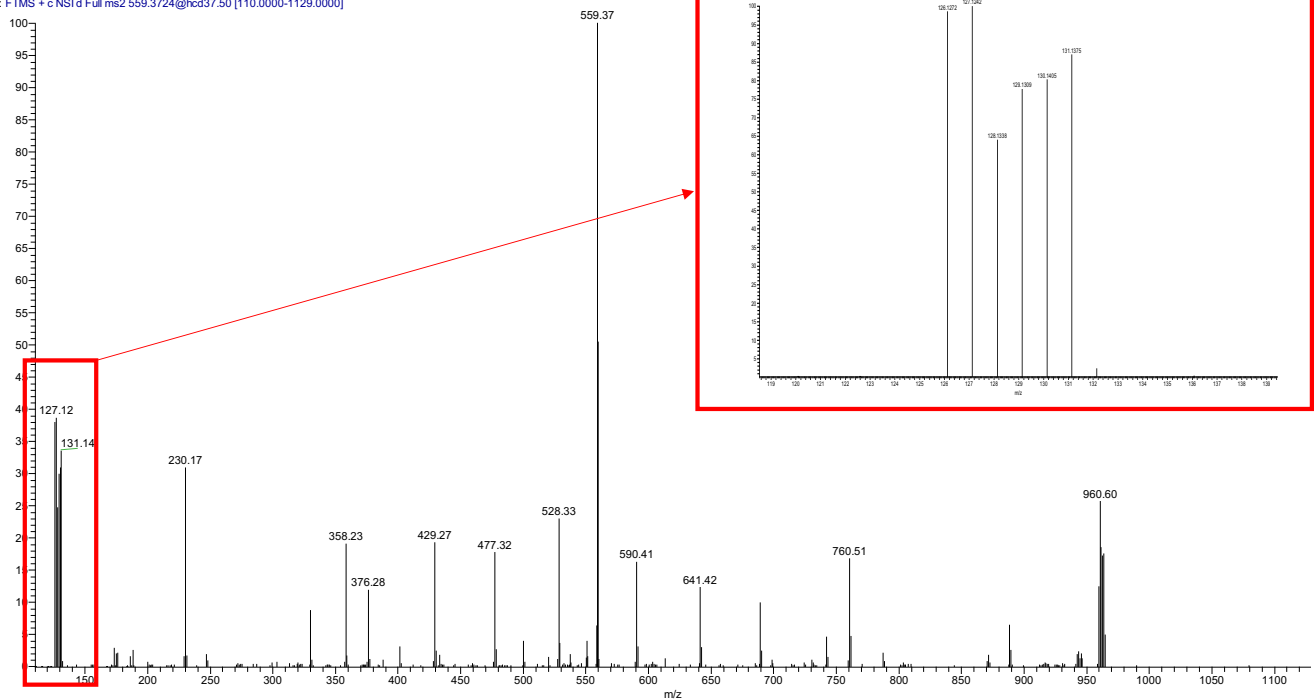
79



80

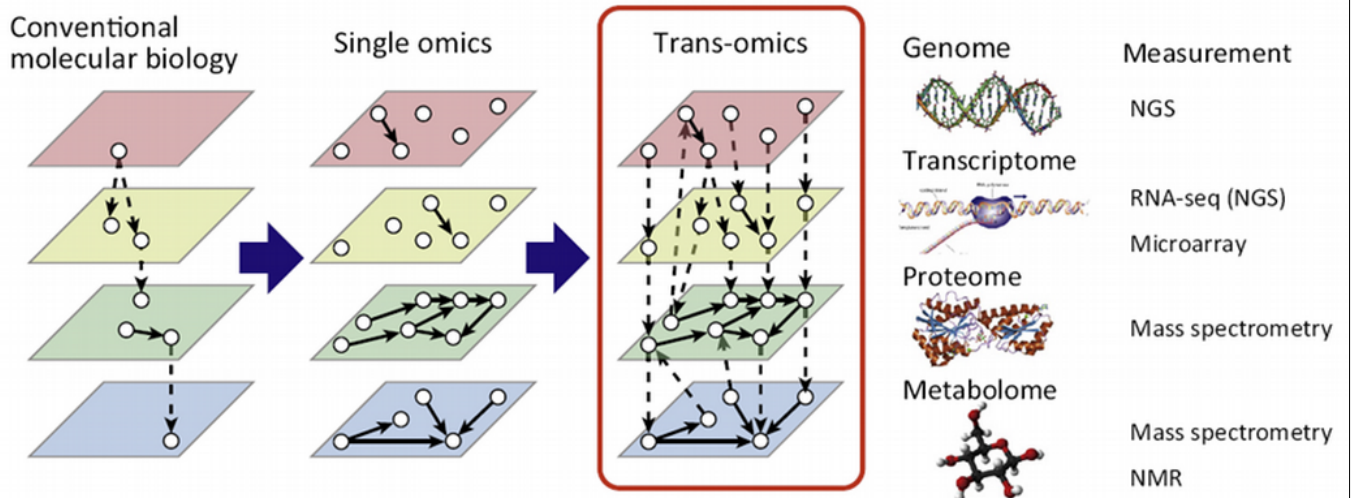
An example MS/MS of TMT-labeled peptide

20191004_DGIST_cere_mix_Frac_03 #35293 RT: 67.61 AV: 1 NL: 1.56E8
T: FTMS + c NSI d Full ms2 559.3724@hcd37.50 [110.0000-1129.0000]



81

Multi-Omics study



82

Proteogenomics

ARTICLE

Proteogenomics connects somatic mutations to signalling in breast cancer

Philip M. Martin^{1,2}, D. R. Mani³, Kelly V. Ruggles⁴, Michael A. Gillette^{5,6}, Karl R. Clauser⁷, Pei Wang⁸, Xianlong Wang⁹, Jina W. Qiao¹⁰, Song Cao¹¹, Francesca Petralia¹², Emily Kawaler¹³, Filip Mundi¹⁴, Kristen Krag¹⁵, Zhihong Tu¹⁶, Jonathan T. Lee¹⁷, Michael L. Gatta¹⁸, Matthew Wilkerson¹⁹, Charles M. Perou²⁰, Venkata Yellapantula²¹, Kuan-lin Huang²², Chenwei Lin²³, Michael D. McLellan²⁴, Ping Yan²⁵, Sherri R. Davies²⁶, R. Reid Townsend²⁷, Steven A. Skates²⁸, Jing Wang²⁹, Bing Zhang³⁰, Christopher R. Kinsinger³¹, Mohd Meer³², Henry Rodriguez³³, Li Ding³⁴, Amanda G. Paulovich³⁵, David Ferry³⁶, Matthew J. Ellis³⁷, Steven A. Carr³⁸ & the NCI CPTAC³⁹

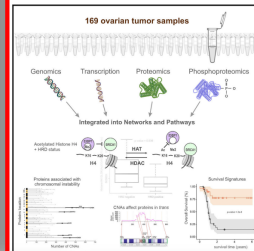
Somatic mutations have been extensively characterized in breast cancer, but the effects of these genetic alterations on the proteomic landscape remain poorly understood. Here we describe quantitative mass spectrometry-based proteomic and phosphoproteomic analyses of 105 genomically annotated breast cancers, of which 77 provided high-quality data. Integrated analyses provided insights into the somatic cancer genome including the consequences of chromosomal loss, such as the 5q deletion characteristic of basal-like breast cancer. Interrogation of the 5q minus-effects against the Library of Integrated Network-based Cellular Signatures, connected loss of *CEP350* and *SKP1* to elevated expression of epidermal growth factor receptor (EGFR), and *SKP1* loss also to increased SRC tyrosinase, and pathway analysis of the phosphoproteome identified a G-protein-coupled receptor cluster that was not readily identifiable from RNA level. In addition to ERBB2, other amplicon-associated highly phosphorylated kinases were identified, including PAK1, PTK2, RIPK2 and TLK2. We demonstrate that proteogenomic analysis of breast cancer can identify consequences of somatic mutations, narrow candidate nominations for driver genes within regions, and identifies therapeutic targets.

2016

Cell

Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer

Graphical Abstract



Authors
Hui Zhang, Tao Liu, Zhen Zhang, ..., Daniel W. Chan, Karin D. Rodland, the CPTAC Investigators

Correspondence
dchan@bcm.edu (D.W.C.), karin.rodland@pnnl.gov (K.D.R.)

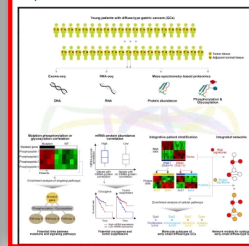
In Brief
Layering proteomic and genomic data from ovarian tumors provides insights into how signaling pathways correspond to specific genome rearrangements and points to the benefit of using protein signatures for assessing prognosis and treatment stratification.

2019

Cancer Cell

Proteogenomic Characterization of Human Early-Onset Gastric Cancer

Graphical Abstract



Authors
Dong-Gi Mun, Jinhyuk Bhn, Sangok Kim, ..., Sanghyuk Lee, Sang-Won Lee, Daehae Hwang

Correspondence
eunyang@kist.re.kr (E.G.Y.), hkim@kist.re.kr (H.K.K.), eunok@kist.re.kr (E.P.), sanghyuk@kist.re.kr (S.L.), se.lee@kist.ac.kr (S.-W.L.), dhwang@kist.ac.kr (D.H.)

In Brief
Mun et al. perform proteogenomic analysis of diffuse gastric cancers (DGC) in a young population. They identify correlations of mRNA-protein abundance associated with specific genomic subtypes of DGC. Some subtypes with related pathways are identified using the proteomic data.

2019

ARTICLE

Proteogenomic characterization of human colon and rectal cancer

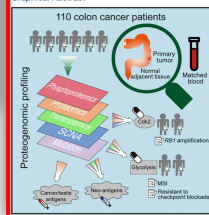
Bing Zhang^{1,2}, Jing Wang³, Xiaojing Wang⁴, Jing Zhu¹, Qi Liu¹, Zhao Shi^{4,5}, Matthew C. Chambers⁶, Lisa J. Kent⁷, Shaddox⁸, Sangtae Kim⁹, Sherri R. Davies¹⁰, Sean Wang¹¹, Pei Wang¹², Henry Rodriguez¹³, R. Reid Townsend¹⁴, Matthew J. C. Ellis¹⁵, Steven A. Carr¹⁶, David L. Tabb¹⁷, Robert J. Collier¹⁸, Robert J. C. Slebos¹⁹, Daniel C. Liebler²⁰ & the NCI CPTAC²¹

2014

Cell

Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities

Graphical Abstract



Authors
Suhay Vasaikar, Chen Huang, Xiaojing Wang, ..., Tao Liu, Bing Zhang, Clinical Proteomic Tumor Analysis Consortium

Correspondence
karin.rodland@pnnl.gov (K.D.R.), daniel.liebler@vanderbilt.edu (D.C.L.), bob.fallgren@pnnl.gov (T.L.), bing.zhang@bcm.edu (B.Z.)

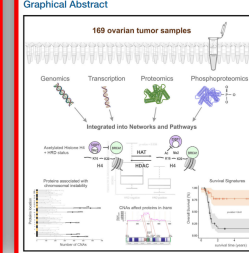
In Brief
A systematic proteogenomic analysis of colon cancer reveals vulnerabilities of potential clinical value inaccessible from genomic assessment alone.

2019

Cell

Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer

Graphical Abstract



Authors
Hui Zhang, Tao Liu, Zhen Zhang, ..., Daniel W. Chan, Karin D. Rodland, the CPTAC Investigators

Correspondence
dchan@bcm.edu (D.W.C.), karin.rodland@pnnl.gov (K.D.R.)

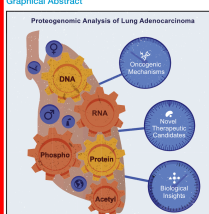
In Brief
Layering proteomic and genomic data from ovarian tumors provides insights into how signaling pathways correspond to specific genome rearrangements and points to the benefit of using protein signatures for assessing prognosis and treatment stratification.

2019

Cell

Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma

Graphical Abstract

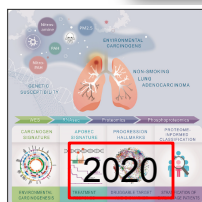


Authors
Michael A. Gillette, Shaikha Satpathy, Song Cao, ..., D.R. Mani, Steven A. Carr, Clinical Proteomic Tumor Analysis Consortium

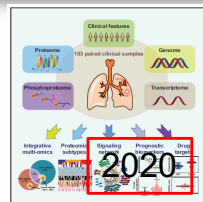
Correspondence
gillette@broadinstitute.org (M.A.G.), shaikha@broadinstitute.org (S.S.), scaar@broad.mit.edu (S.A.C.)

In Brief
Comprehensive proteogenomic characterization of lung adenocarcinoma and paired normal adjacent tissues from patients of diverse smoking status and ethnicity on organoids yields insights into cancer biology, offers new drug targets, and provides a common resource for future

2020



2020



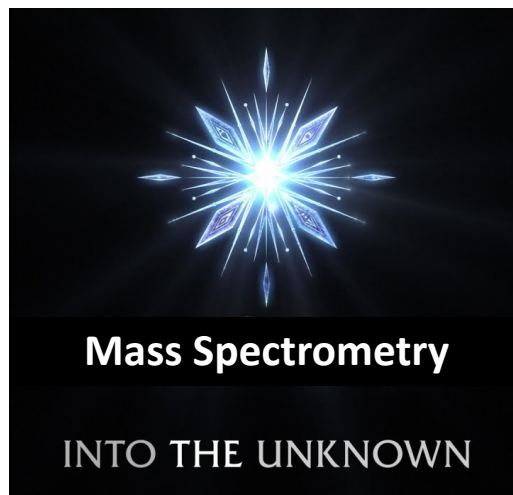
2020

Many more to come...

83

Conclusions

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



Mass Spectrometry

INTO THE UNKNOWN

84

단백체, 멀티오믹스 생명정보학 포스닥, 석/박사과정생 모집 중

(김민식, mkim@dgist.ac.kr)



85

Thank you!

KSBi-BIML 2021

Practice in Mass Spectrometry-based Proteomics

김민식, New Biology, DGIST

본 강의 자료는 한국생명정보학회가 주관하는 KSBi-BIML 2021 워크숍 온라인 수업을 목적으로 제작된것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다. 수업 목적으로 배포 및 전송 받은 경우에도 이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없습니다.

만약 이러한 사항을 위반할 경우 발생하는 모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고합니다.

실습 내용

1. 단백질 데이터 구조
2. 펩티드 MS/MS 스펙트럼 해석
3. Mascot 설명 및 Enolase 단백질 동정
4. MaxQuant 설명 및 HeLa 단백질 동정
5. 결과 파일 해석
6. PTM 분석
7. https://drive.google.com/open?id=1RImRN4R2MZe4ap9Q0_TPczGXS7bGoK

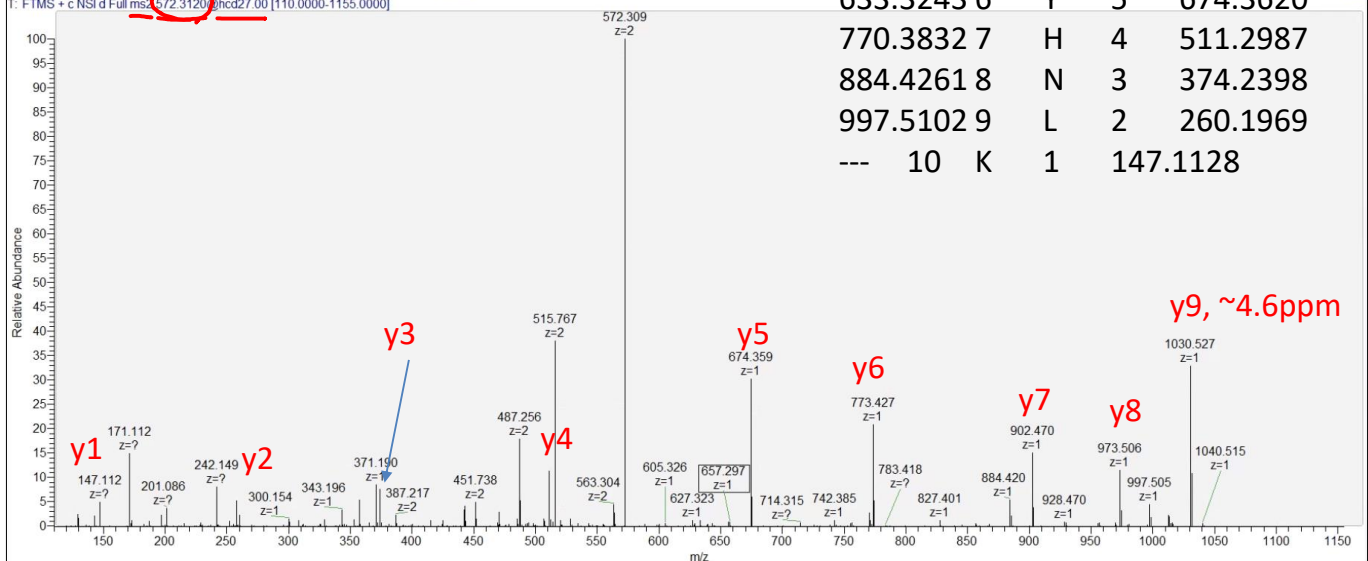
질량스펙트럼 – 펩티드 분석

Human alpha-enolase

I-G-A-E-V-Y-H-N-L-K

y₉

20180502_50cm-HeLa_100-9-6716_305C_01#11439_RT: 32.43_AV: 1_NL: 3.10E6
T: FTMS + c NSI d Full ms4 572.3120 [hcd27.00 [110.0000-1155.0000]]



$$(1030.5316 - 1030.527) / 1030.5316 \times 10^6 = (\sim 4.6) \text{ ppm}$$

큐바이오 정밀의학 연구실

단백체, 멀티오믹스 생명정보학 포스닥, 석/박사과정생 모집 중
(김민식, mkim@dgist.ac.kr)



감사합니다!