KSBi-BIML 2024

Bioinformatics & Machine Learning(BIML) Workshop for Life and Medical Scientists

생명정보학 & 머신러닝 워크샵 (온라인)

Proteogenomic analysis of human cancer

빈진혁 _ 연세대학교

KSBI KOREAN SOCIETY FOR BIOINFORMATICS | 한국생명정보학회

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2024 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 2024

Bioinformatics & Machine Learning(BIML)
Workshop for Life and Medical Scientists

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2024에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 벌써 10년 차를 맞이하게 되었습니다. BIML 워크샵은 국내 생명정보학 분야의 최초이자 최고 수준의 교육프로그램으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되어 있습니다. 올해 인공지능 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 인공지능 기반 자료모델링 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 단일세포오믹스, 공간오믹스, 메타오믹스, 그리고 롱리드염기서열 자료 분석에 대한 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다.

올해 BIML의 가장 큰 변화는 최근 연구 수요가 급증하고 있는 의료정보자료 분석에 대한 현장 강의를 추가하였다는 것입니다. 특히 의료정보자료 분석을 많이 수행하시는 의과학자 및 의료정보 연구자들께서 본 강좌를 통해 많은 도움을 받으실 수 있기를 기대하고 있습니다. 또한 다양한 생명정보학 분야에 대한 온라인 강좌 프로그램도 점차 증가하고 있는 생명정보 분석기술의 다양화에 발맞추기 위해 작년과 비교해 5강좌 이상을 신규로 추가했습니다. 올해는 무료 강좌 5개를 포함하여 35개 이상의 온라인 강좌가 개설되어 제공되며, 연구 주제에 따른 연관된 강좌 추천 및 강연료 할인 프로그램도 제공되며, 온라인을 통한 Q&A 세션도 마련될 예정입니다. BIML-2024는 국내 주요 연구중심 대학의 전임 교원이자 각 분야 최고 전문가들의 강의로 구성되었기에 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것이라 확신합니다.

BIML-2024을 준비하기까지 너무나 많은 수고를 해주신 운영위원회의 정성원, 우현구, 백대현, 김태민, 김준일, 김상우, 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 강사분들께 깊은 감사를 드립니다.

2024년 2월

한국생명정보학회장 이 인 석

강의개요

Proteogenomic analysis of human cancer

최근에 급속하게 발전된 바이오테크놀로지 기술의 발달로 인해 생명 시스템에 대한 다차원적 이해가 가능하게 되었으며, 특히 Next generation sequencing (NGS)을 활용한 유전체 연구와 Mass-spectrometry (MS)를 활용한 단백체 연구가 활발해지고 있다. 암과 같이 복잡한 분자 기전으로 일어나는 질병은 단일 차원의 데이터만으로는 진단 및 치료 타겟 발굴을 위한 생물학적 이해가 어렵다고 알려지면서 유전자 레벨에서 단백질 레벨까지 통합적인 해석을 하기 위한 연구들이 활발해지고 있다. 이러한 접근법 중 대표적인 유전단백체 (Proteogenomics: Proteomics + Genomics) 연구 방법은 동일한 생물학적 샘플에서 유전체, 전사체, 단백체 데이터를 동시에 생산하고 이들 간의 통합분석을 통해 연구대상에 대한 새로운 생물학적 의미를 도출하는 데 유용하게 활용되고 있다. 본 강의를 통해 현대 오믹스 연구의 두 축인 유전체, 단백체 데이터에 대해 이해하고 이 두 데이터가 어떻게 통합되어 암을 진단하고 치료 타겟을 발굴하는데 활용될 수 있는지 배울 수 있다.

강의는 다음의 내용을 포함한다:

- NGS 및 MS 개요
- 유전단백체 연구 개요
- 유전체-단백체 데이터 통합 방법론 소개
- 암 연구를 위한 유전단백체 연구동향

* 참고강의교재: 강의 자료 제공

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

* 강의 난이도: 초급

* 강의: 빈진혁 교수 (연세대학교 의과대학 의생명시스템정보학교실)

Curriculum Vitae

Speaker Name: Jinhyuk Bhin, Ph.D.

▶ **Personal Info**

Name	Jinhyuk Bhin
Title	Assistant Professor
Affiliation	Yonsei University College of Medicine

▶ **Contact Information**

Address	20, Eonju-ro 63-gil, Gangnam-gu, Seoul, Republic of Korea
Email	jbhin@yuhs.ac
Phone Number	02-2019-5470

Research Interest

Oncogenomics, Proteogenomics, single cell and spatial transcriptomics

Educational Experience

2010	B.S. in Chemical Engineering, Sungkyunkwan University, Korea
2016	Ph.D. in Bioinformatics & Systems Biology, POSTECH, Korea

Professional Experience

2016-2016	Postdoc, Institute for Basic Science, DGIST, Korea
2017-2022	Postdoc, Netherlands Cancer Institute, The Netherlands
2023-	Assistant Professor, Yonsei University College of Medicine

Selected Publications (5 maximum)

1. Bhin J*, Yemelyanenko J* et al., MYC is a clinically significant driver of mTOR inhibitor resistance in breast cancer, **Journal of Experimental Medicine**. 2023 Nov;220(1):e20211743
2. Bhin J*, Dias.M* et al., Multi-omics analysis reveals distinct non-reversion mechanisms of PARPi resistance in BRCA1- versus BRCA2-deficient mammary tumors, **Cell Reports**. 2023 May; 42(5):112538.
3. Zingg D*, Bhin J* et al., Truncated FGFR2 is a clinically actionable oncogene in multiple cancers, **Nature**. 2022 Aug;608(7923):609-617.
4. Moon D*, Bhin J* et al., Proteogenomic characterization of human early-onset gastric cancer, **Cancer Cell**. 2019 Jan 14;35(1):111-124.e10.
5. Bin B*, Bhin J* et al, Requirement of zinc transporter ZIP10 for epidermal development: Implication of the ZIP10-p63 axis in epithelial homeostasis, **Proc Natl Acad Sci U S A**. 2017 Nov 14;114(46):12243-12248.

KSBI | 한국생명정보학회

KSBi-BIML 2024

Proteogenomic analysis of human cancer

빈진혁
연세대학교 의과대학
의생명시스템정보학교실

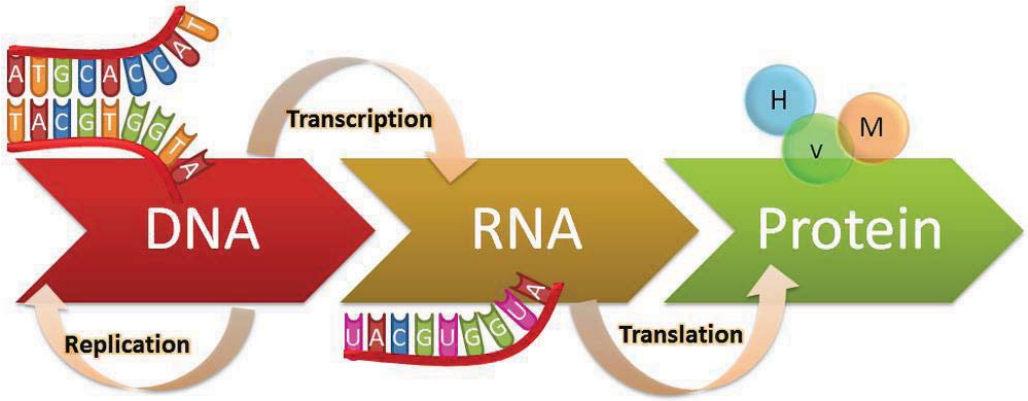
목차

- 유전단백체학 (Proteogenomics) 개념 소개
- 유전단백체학을 이용한 암 연구 동향
- 유전단백체학 분석 방법론/툴 소개

목차

- 유전단백체학 (Proteogenomics) 개념 소개
- 유전단백체학을 이용한 암 연구 동향
- 유전단백체학 분석 방법론/툴 소개

Central Dogma

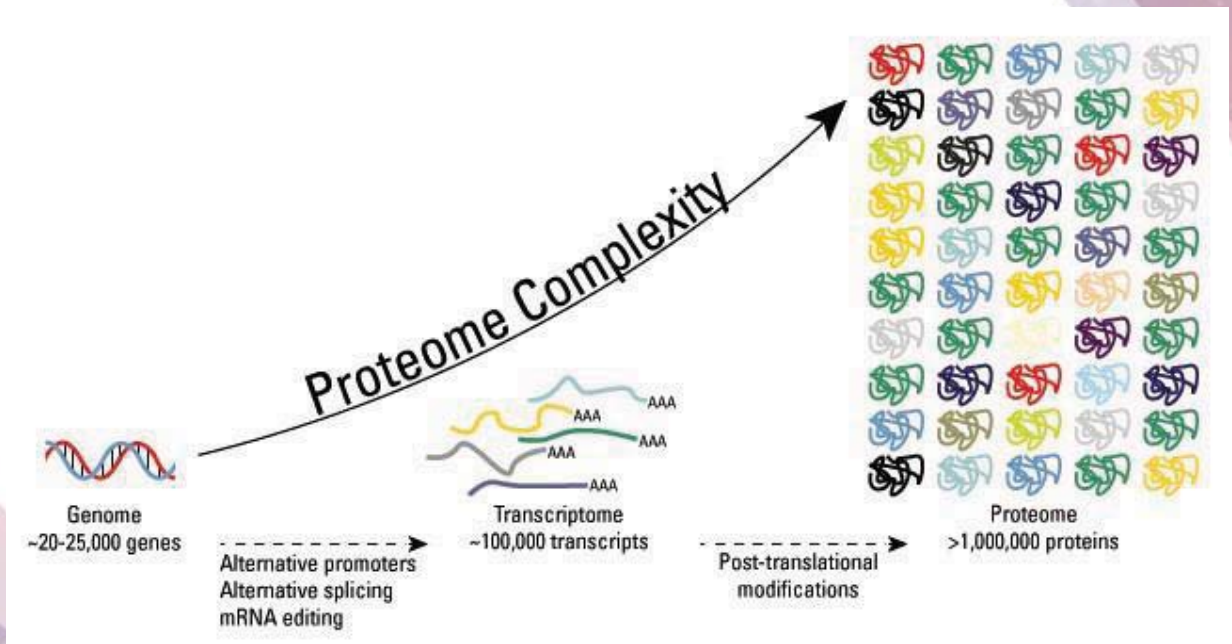


Genome
(유전체)

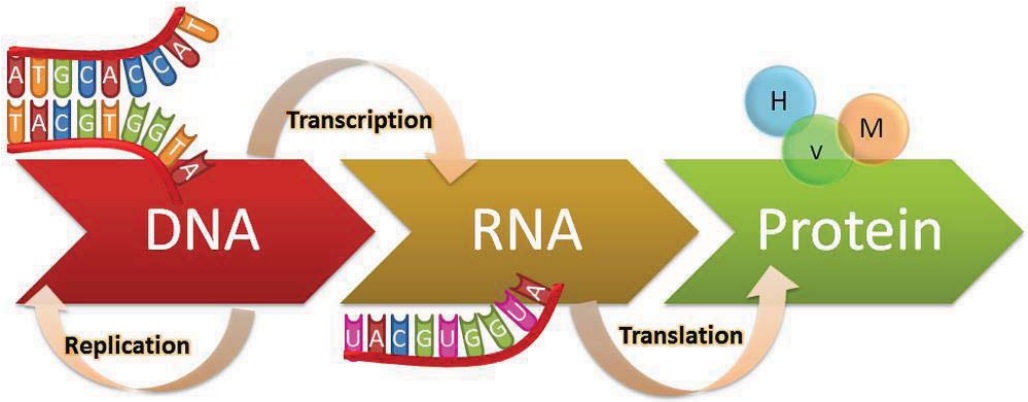
Transcriptome
(전사체)

Proteome
(단백체)

Complexity of molecular entities



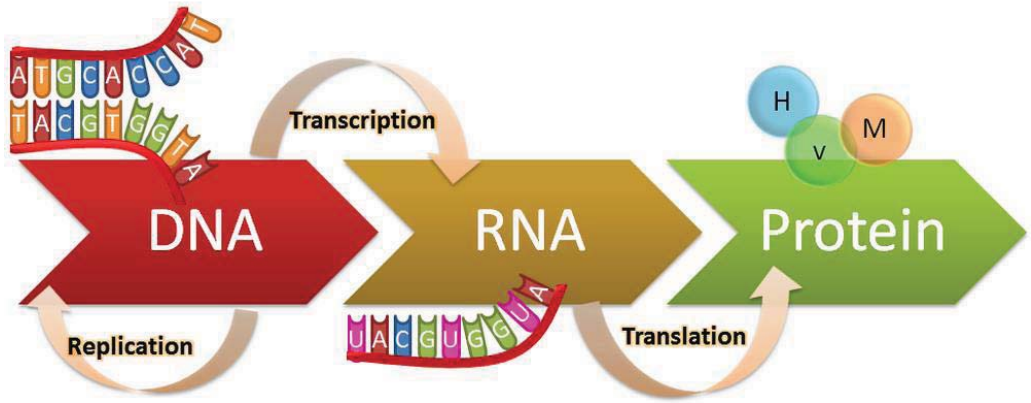
Central Dogma



차세대염기서열분석법
Next generation sequencing
Genomics
(유전체학)

질량분석기
Mass-spectrometry
Proteomics
(단백체학)

유전단백체학 (Proteogenomics)

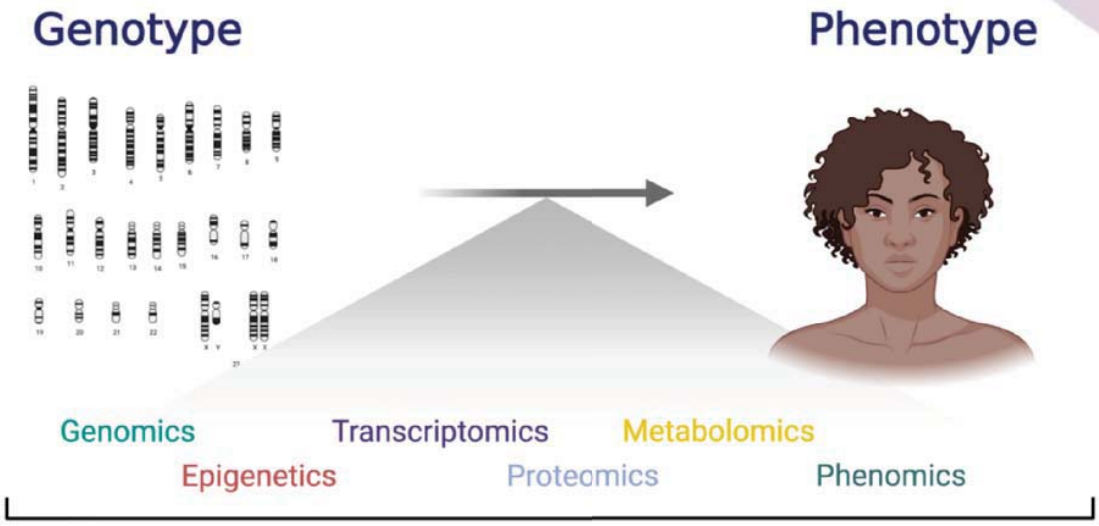


차세대염기서열분석법
Next generation sequencing
Genomics
(유전체학)

+

질량분석기
Mass-spectrometry
Proteomics
(단백체학)

유전체-단백체 통합 연구의 필요성



A multi-omics approach is needed to understand complex biological systems

https://arimagenomics.com

유전체-단백체 통합 연구의 필요성

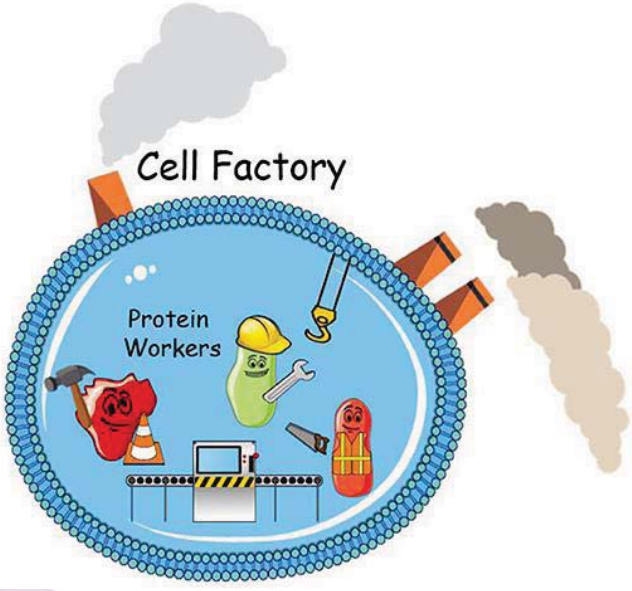
Genome remains the same but proteome changes

DNA - tells what possibly

RNA – tells what probably

Proteins – tells what actually happens

왜 단백체를 이해해야 하는가?



Proteins are the workers inside your cells

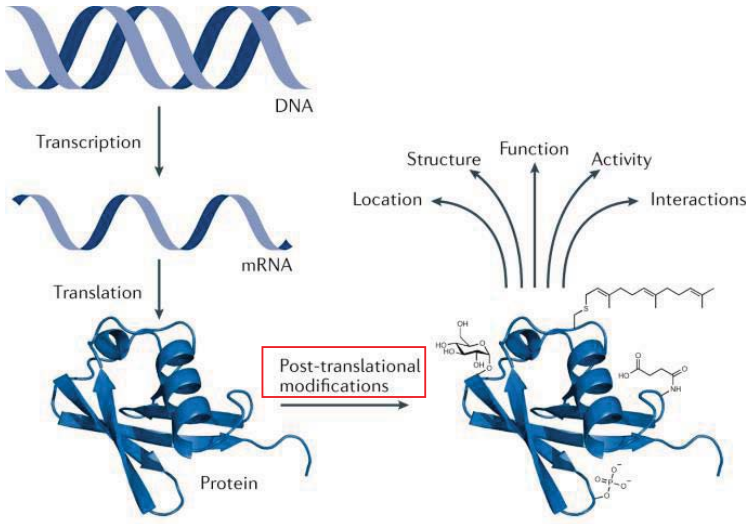
https://kids.frontiersin.org

To be functional in a cell

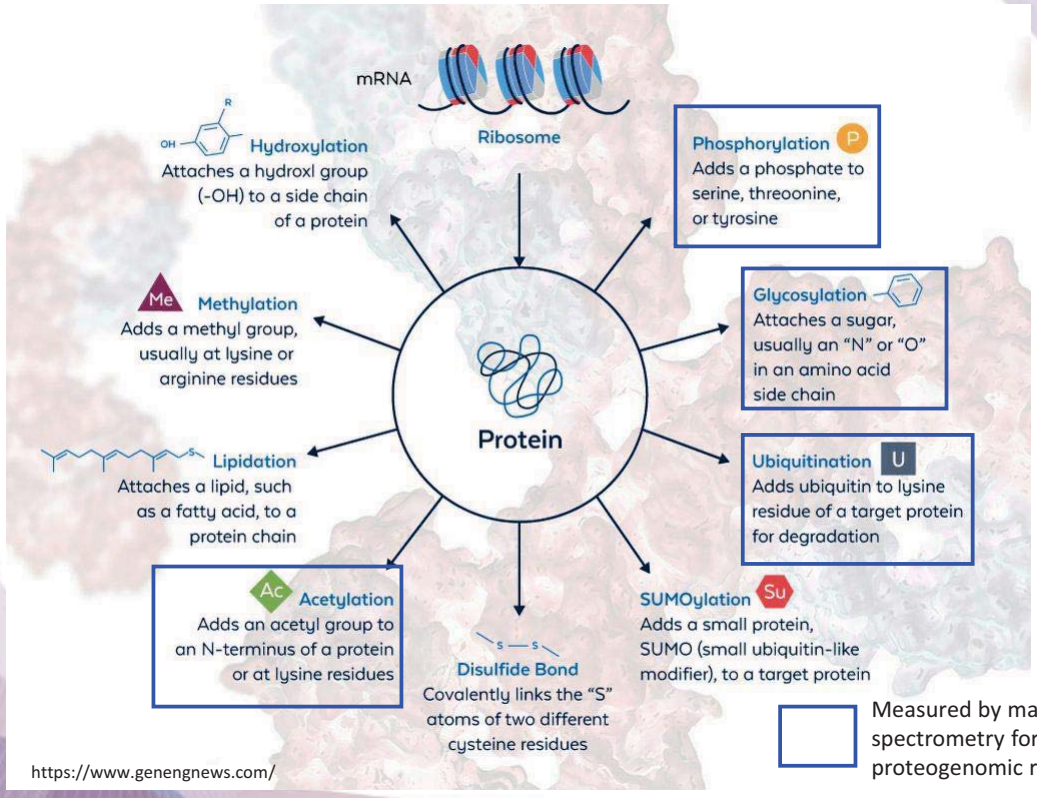
- Folding
- Post-translational modification
- Transport
- Assembly into complexes
- Quality control

단백질 번역 후 변형 – 기능 조절

Post-translational modification (PTM)
: chemical modification of a protein after translation is complete, essential for its functionality



단백질 번역 후 변형 – 기능 조절

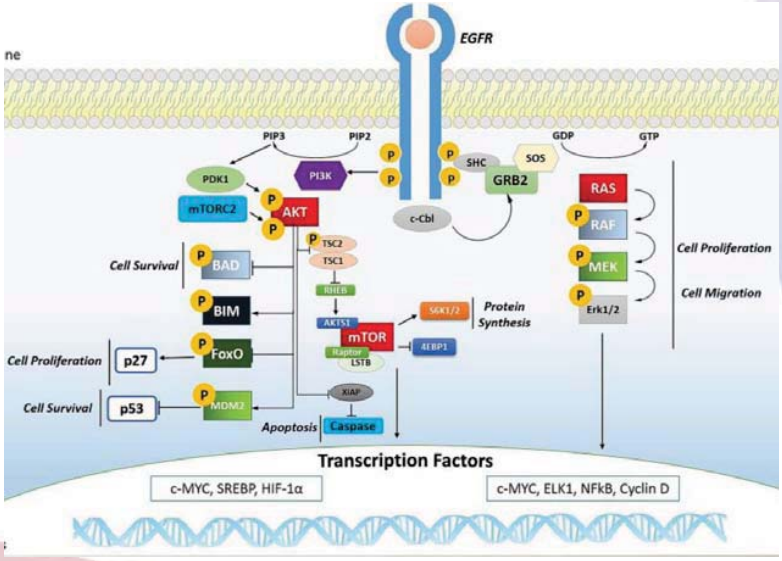


단백질 번역 후 변형 – 기능 조절

- **Phosphorylation**

  - addition of a phosphate group to an amino acid, commonly serine (86.4%), threonine (11.8%), or tyrosin (1.8%)

  - key regulatory mechanism in many cellular processes, including cell cycle, signal transduction pathway, and metabolic regulation



단백질 번역 후 변형 – 기능 조절

- **Glycosylation**

  - attachment of sugar moieties to proteins, typically at asparagine (N-linked) or serine/threonine (O-linked) residue

  - cell-cell recognition, immune response, signaling pathway

- **Acetylation**

  - addition of an acetyl group to the lysine residue of proteins

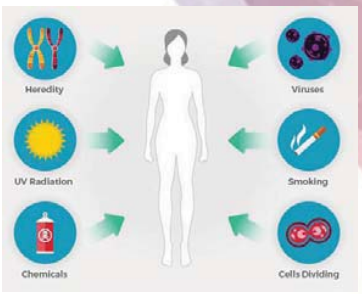
  - gene expression regulation by modifying histone, chromatin structure

- **Ubiquitination**

  - attachment of ubiquitin (a small regulatory protein) to lysine residue

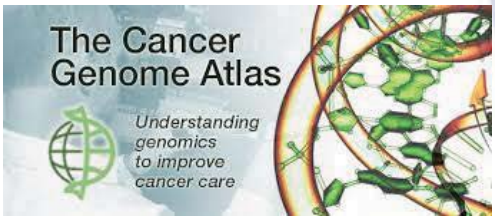
  - cell cycle control, DNA repair, immune response

유전단백체 연구의 필요성 - 암 연구

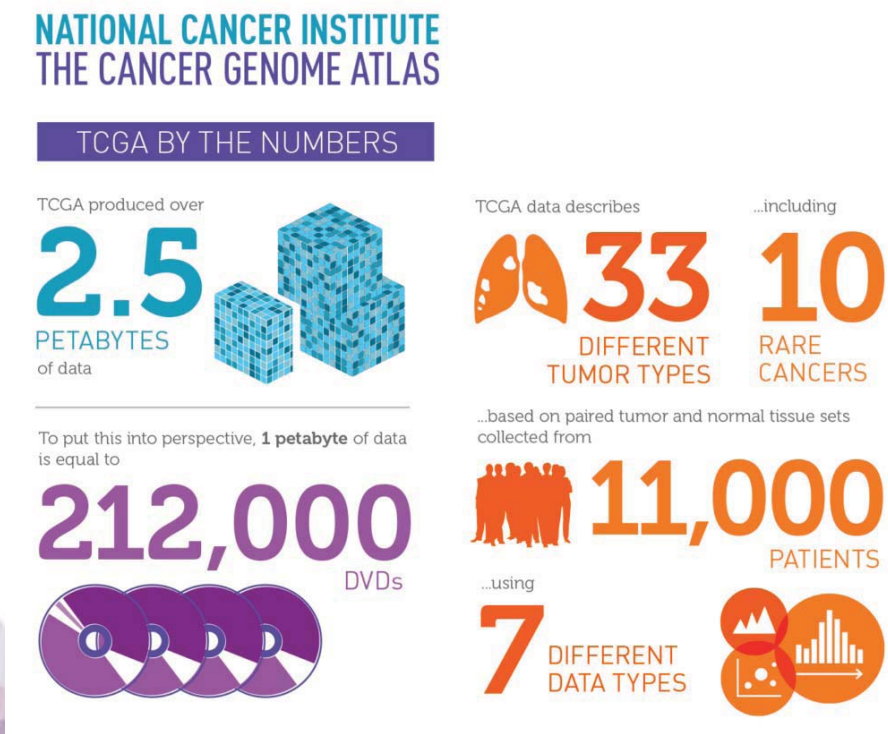


- Cancer is a genetic disease, caused by changes in genes that control the way cells grow and multiply

- **The Cancer Genome Atlas (TCGA)**
  - started in 2006 and ended in 2013, by NCI
  - to catalog genetic mutations responsible for cancer
  - to identify unique and reproducible genomic differences that exist among patients and determine whether these differences could lead to the development of individual treatment

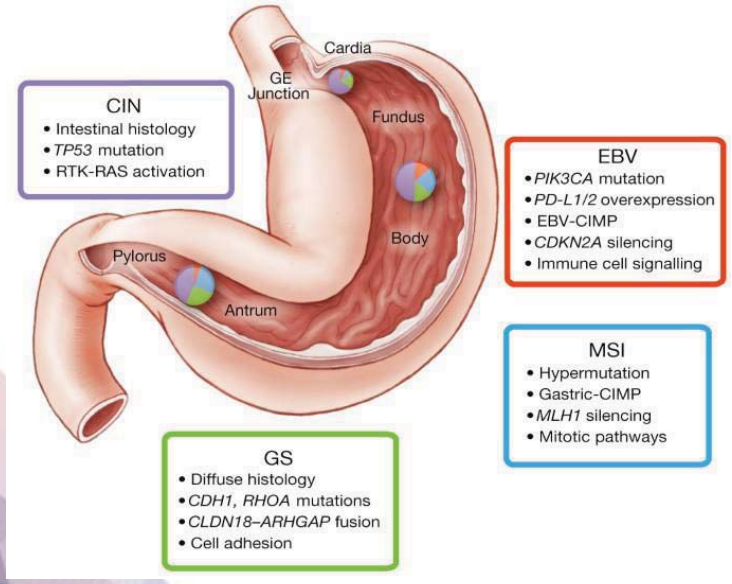


유전단백체 연구의 필요성 - 암 연구



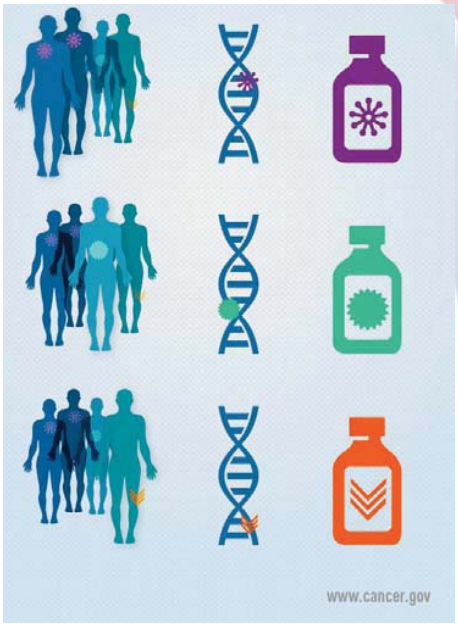
유전단백체 연구의 필요성 - 암 연구

TCGA - Gastric Cancer

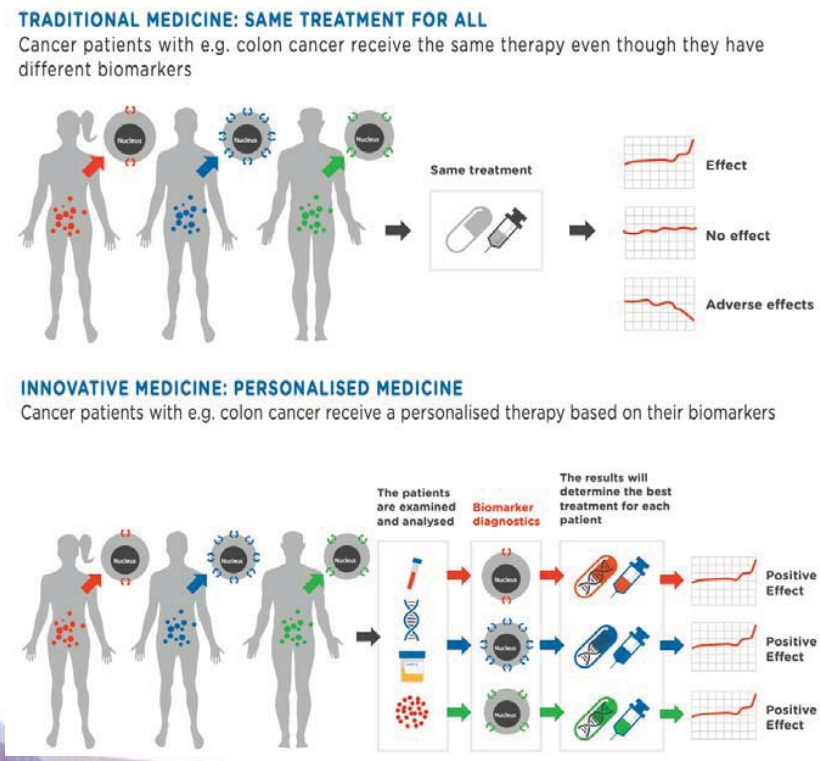


TCGA, 2014, Nature (Gastric cancer)

Precision Medicine



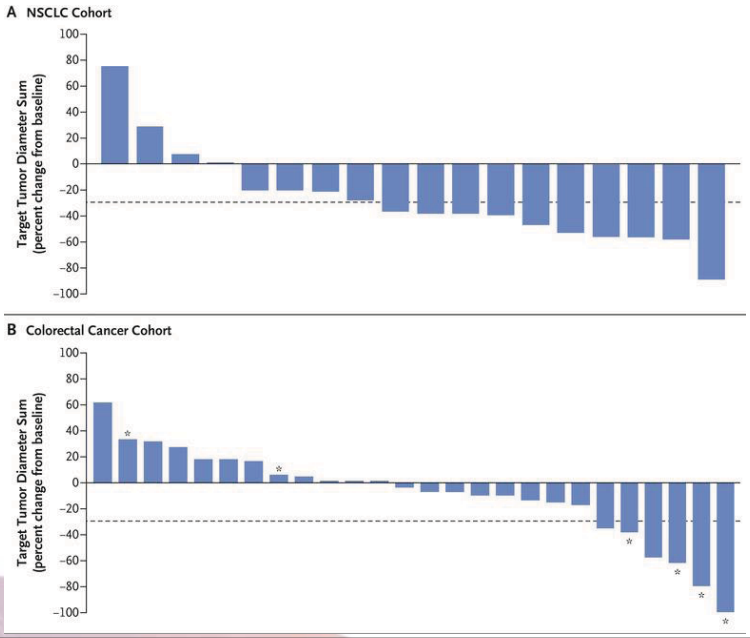
유전단백체 연구의 필요성 - 암 연구



유전단백체 연구의 필요성 - 암 연구

- However, mutational profiling alone cannot guarantee accurate predictions about the quantity or activity of the proteins affected by driver mutations

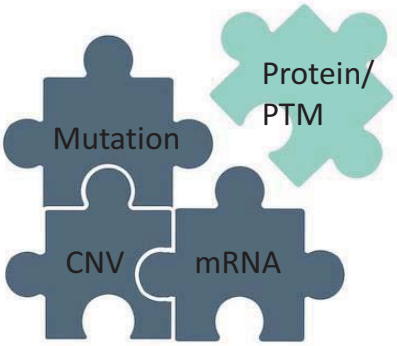
Different responses to BRAF inhibitor Vemurafenib among patients with BRAF V600 mutations



유전단백체 연구의 필요성 - 암 연구

- Need to meaningfully interpret the contribution of genomic features to a specific patient's tumor biology
- Need to have a more complete and precise picture of molecular pathology
- RNA expression levels are often poor predictors of actual protein levels
- Oncogenic/cancer-driving signalings are usually mediated by PTM

Need to incorporate proteomic information



CPTAC – 유전단백체 콘소시움 탄생

CPTAC
CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM

From 2011, by NCI

Flagship papers from CPTAC


nature

Explore content · About the journal · Publish with us · Subscribe

nature > articles > article

Article | Published: 20 July 2014

Proteogenomic characterization of human colon and rectal cancer

Bing Zhang, Jing Wang, Xiaojing Wang, Jing Zhu, Qi Liu, Zhiao Shi, Matthew C. Chambers, Lisa J. Zimmerman, Kent F. Shaddox, Sangtae Kim, Sherri R. Davies, Sean Wang, Pei Wang, Christopher R. Kinsinger, Robert C. Rivers, Henry Rodriguez, R. Reid Townsend, Matthew J. C. Ellis, Steven A. Carr, David L. Tabb, Robert J. Coffey, Robbert J. C. Slebos, Daniel C. Liebler & the NCI CPTAC

Colorectal cancer
(Nature, 2014)


nature

Explore content · About the journal · Publish with us · Subscribe

nature > articles > article

Article | Published: 25 May 2016

Proteogenomics connects somatic mutations to signalling in breast cancer

Philipp Mertins, D. R. Mani, Kelly V. Ruggles, Michael A. Gillette, Karl R. Clauser, Pei Wang, Xianlong Wang, Jana W. Qiao, Song Cao, Francesca Petralia, Emily Kawaler, Filip Mundt, Karsten Krug, Zhidong Tu, Jonathan T. Lei, Michael L. Gatza, Matthew Wilkerson, Charles M. Perou, Venkata Yellapantula, Kuan-lin Huang, Chenwei Lin, Michael D. McLellan, Ping Yan, Sherri R. Davies, NCI CPTAC

Breast cancer
(Nature, 2016)


Cell

Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer

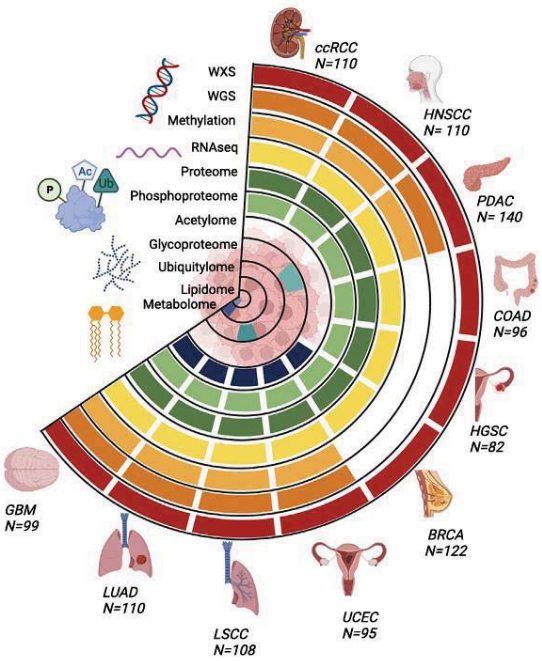
Ovarian cancer
(Cell, 2016)

CPTAC – 유전단백체 콘소시움 탄생

CPTAC
CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM

From 2011, by NCI

~10 cancer types, ~ 1,000 tumors



국내유전단백체 연구 근황

Gastric cancer, 2019

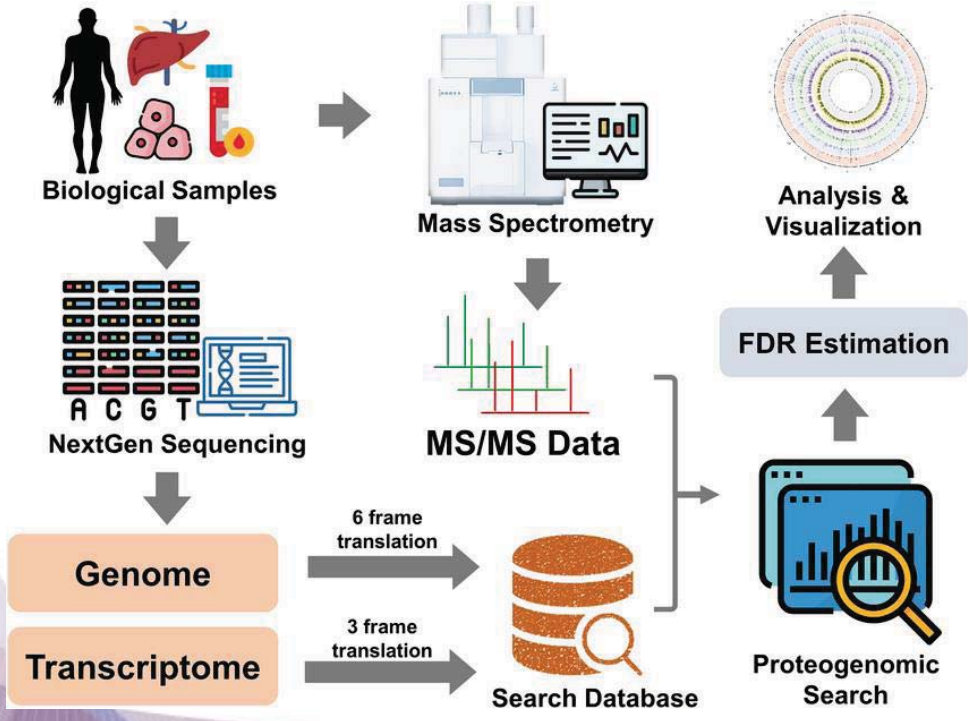
Pancreatic cancer, 2022

Glioblastoma, 2024

목차

- 유전단백체학 (Proteogenomics) 개념 소개
- **유전단백체학을 이용한 암 연구 동향**
- 유전단백체학 분석 방법론/툴 소개

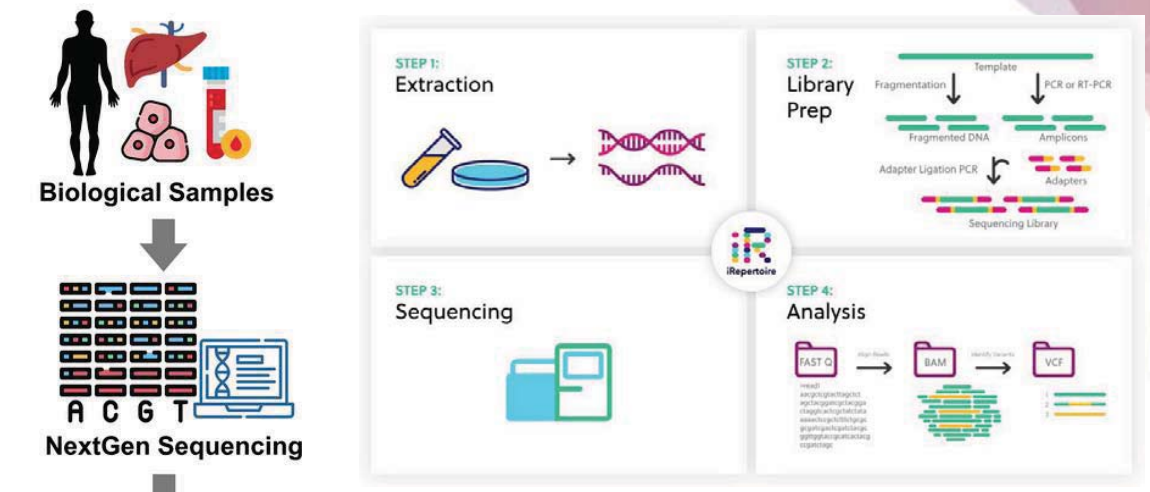
유전단백체 기반 암 연구 workflow



Raj et al., 2023, J.Proteins and Proteomics

유전단백체 기반 암 연구 workflow

1. NGS for genomic and transcriptomic analyses

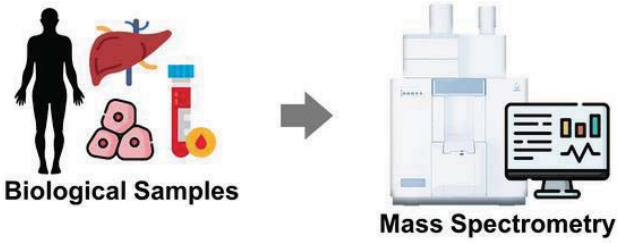


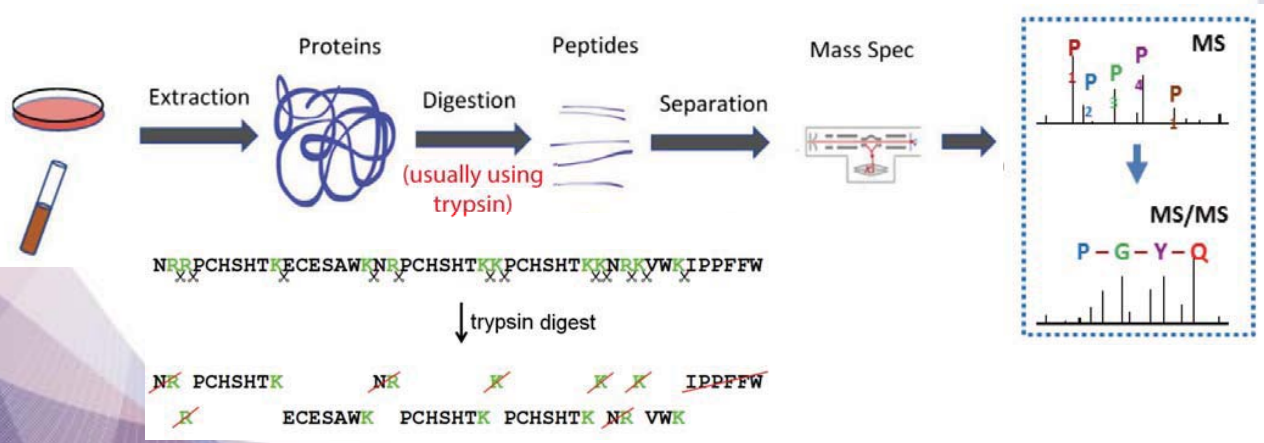
Somatic mutations (SNVs, INDEL – vcf/maf format)

Gene expression (read count matrix)
Gene fusion

유전단백체 기반 암 연구 workflow

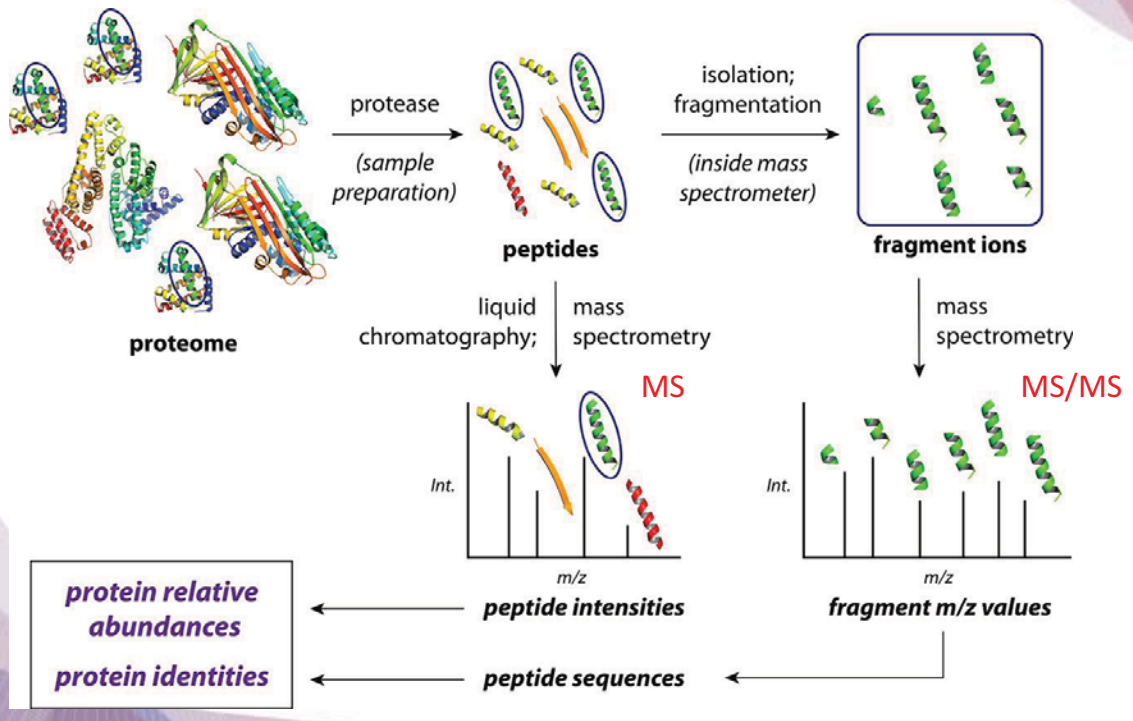
2. Mass-spectrometry for proteomic analyses





유전단백체 기반 암 연구 workflow

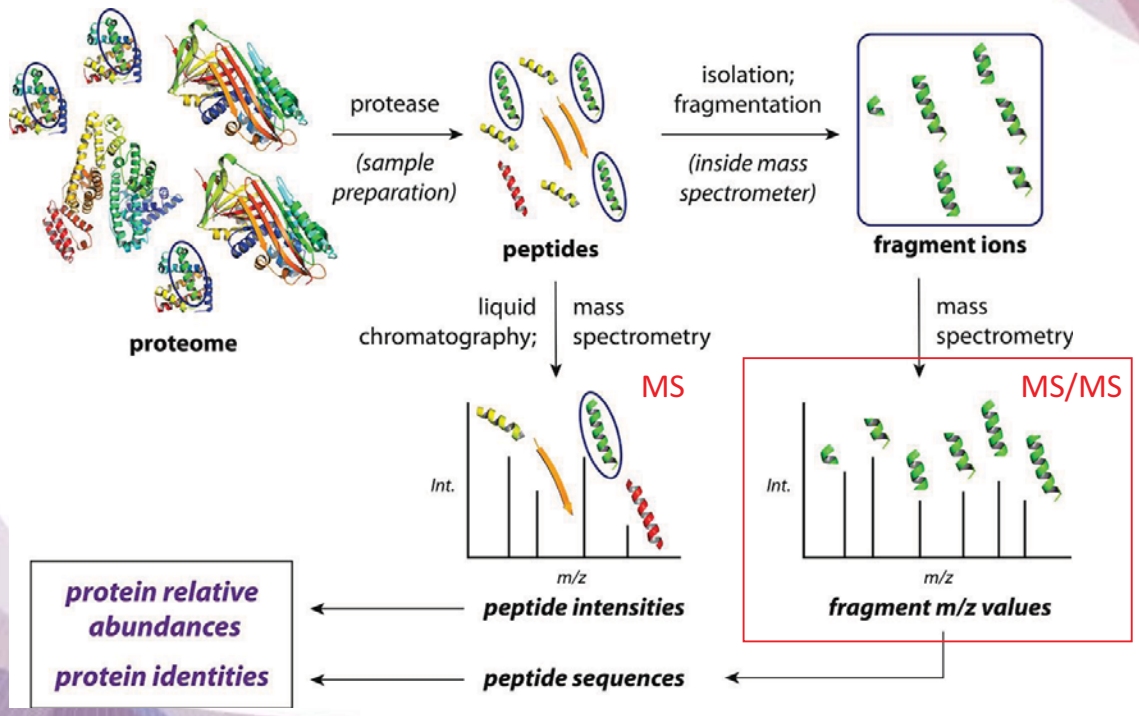
2. Mass-spectrometry for proteomic analyses



Shuken, 2023, J.Proteome Res

유전단백체 기반 암 연구 workflow

2. Mass-spectrometry for proteomic analyses

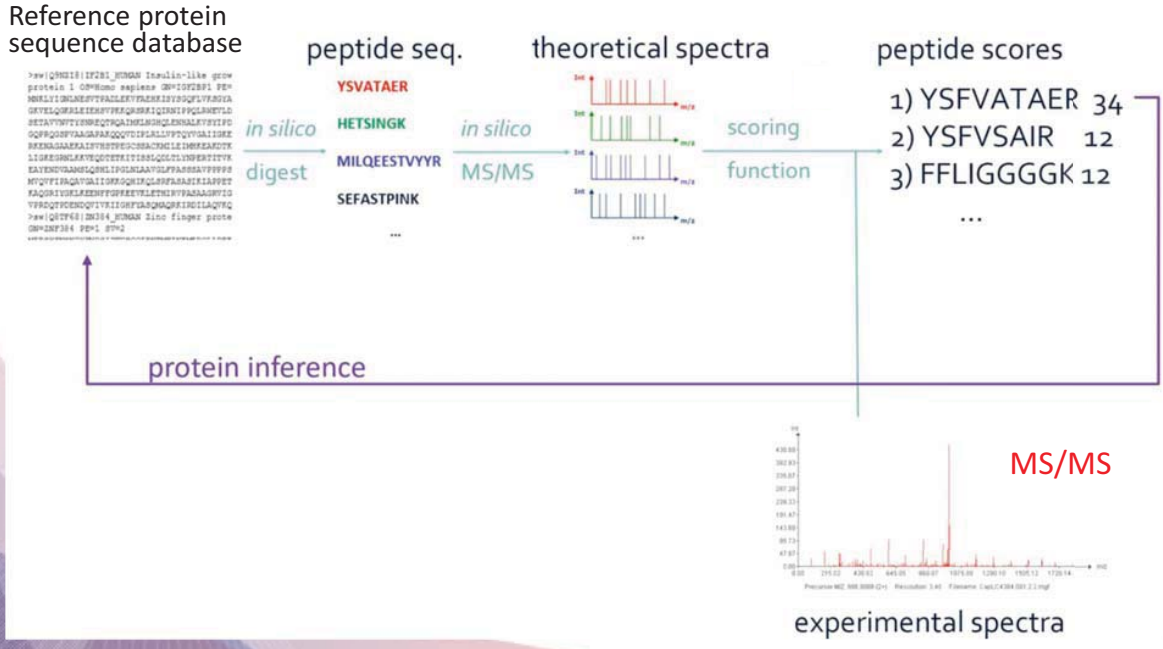


Shuken, 2023, J.Proteome Res

유전단백체 기반 암 연구 workflow

2. Mass-spectrometry for proteomic analyses

Peptide sequence identification



유전단백체 기반 암 연구 workflow

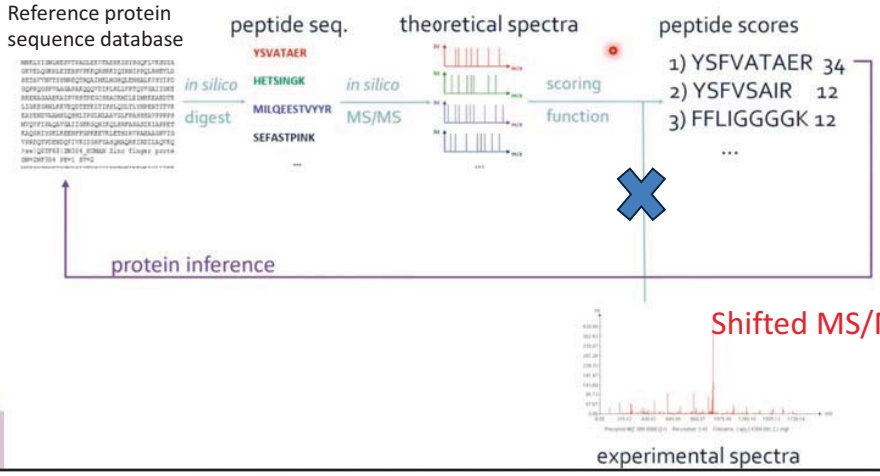
2. Mass-spectrometry for proteomic analyses

- **Peptide identification in tumor samples is incomplete!**

  - A typical method to identify peptide sequences from MS/MS scans relies on known protein sequences (wild-type) in reference proteome database

  - tumor samples contain <u>many mutations, novel splice junctions, and novel protein-coding regions, resulting in novel peptide sequences</u>

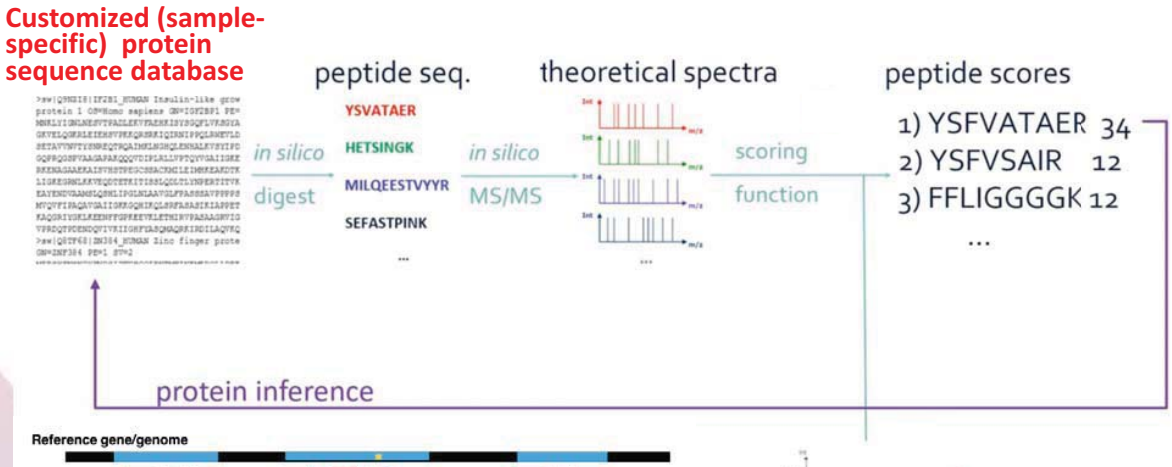
  - experimental MS/MS scans cannot match theoretical spectra due to shifts in scan peaks

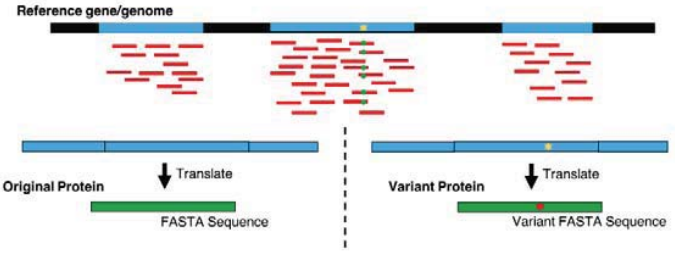


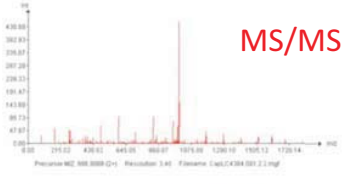
유전단백체 기반 암 연구 workflow

2. Mass-spectrometry for proteomic analyses

Peptide sequence identification



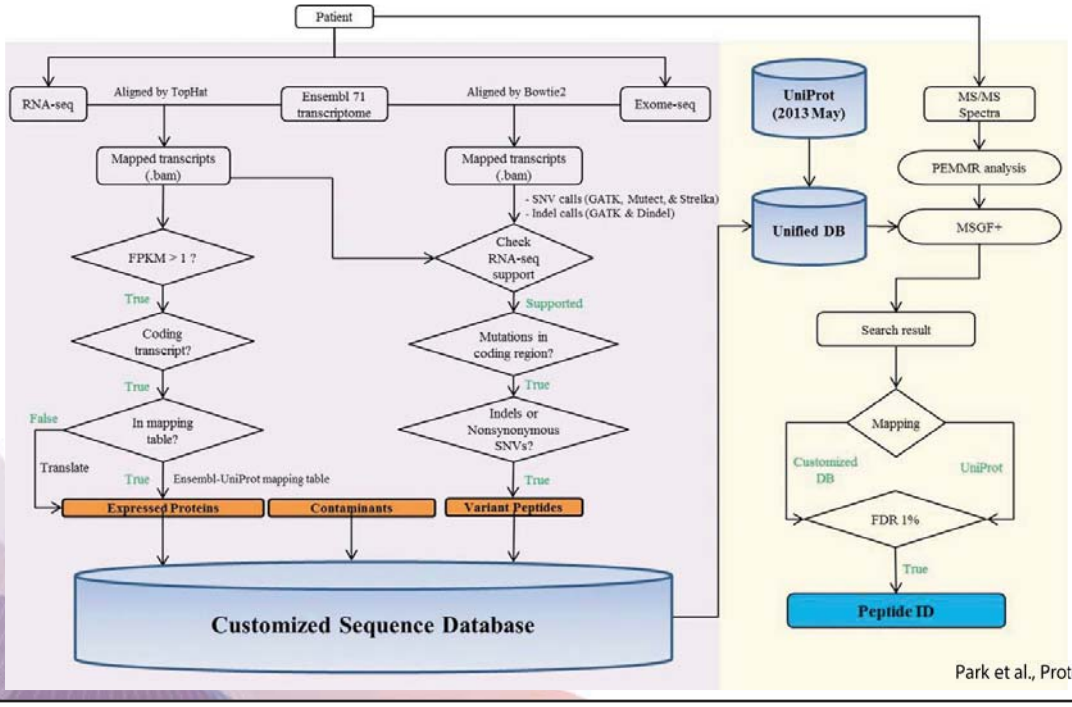




유전단백체 기반 암 연구 workflow

2. Mass-spectrometry for proteomic analyses

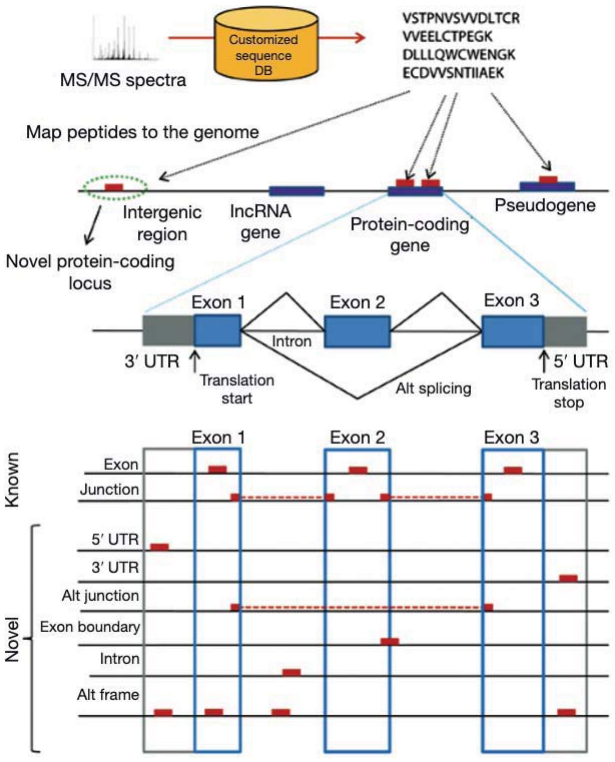
Building customized protein sequence database



Park et al., Proteomics, 2014

유전단백체 기반 암 연구 workflow

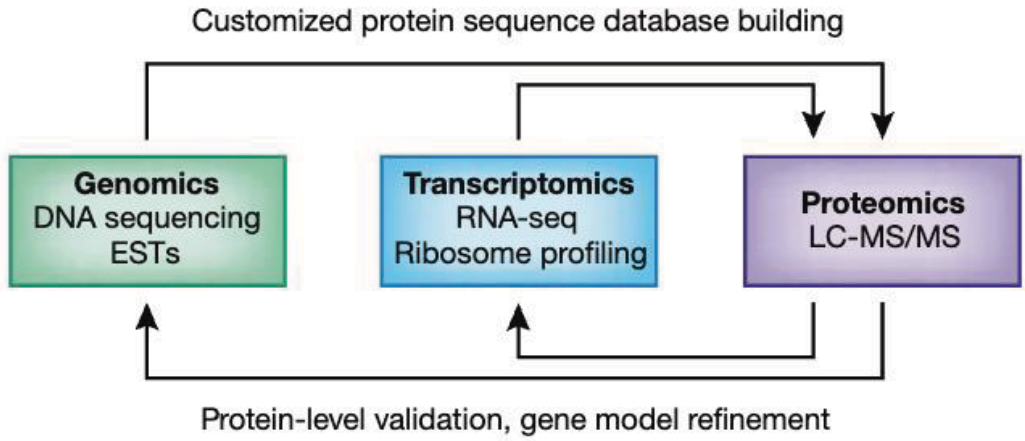
2. Mass-spectrometry for proteomic analyses



Nesvizhskii, Nature Methods, 2014

유전단백체 기반 암 연구 workflow

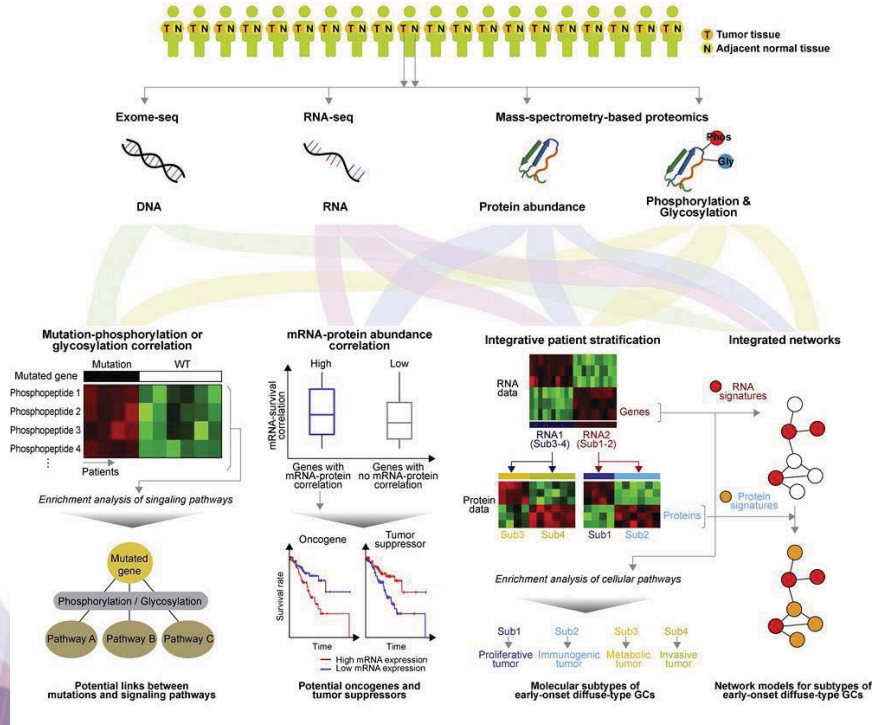
2. Mass-spectrometry for proteomic analyses



Nesvizhskii, Nature Methods, 2014

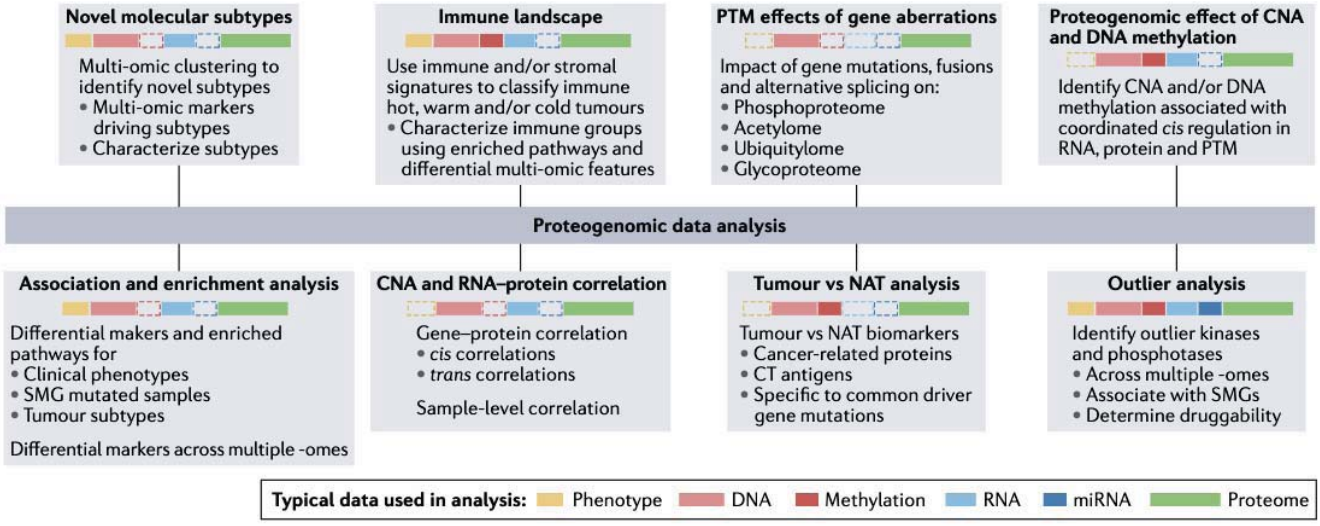
유전단백체 기반 암 연구 workflow

3. integrative analyses of genomic and proteomic data



Bhin and Mun et al., Cancer Cell, 2019

유전단백체 기반 암 연구 동향

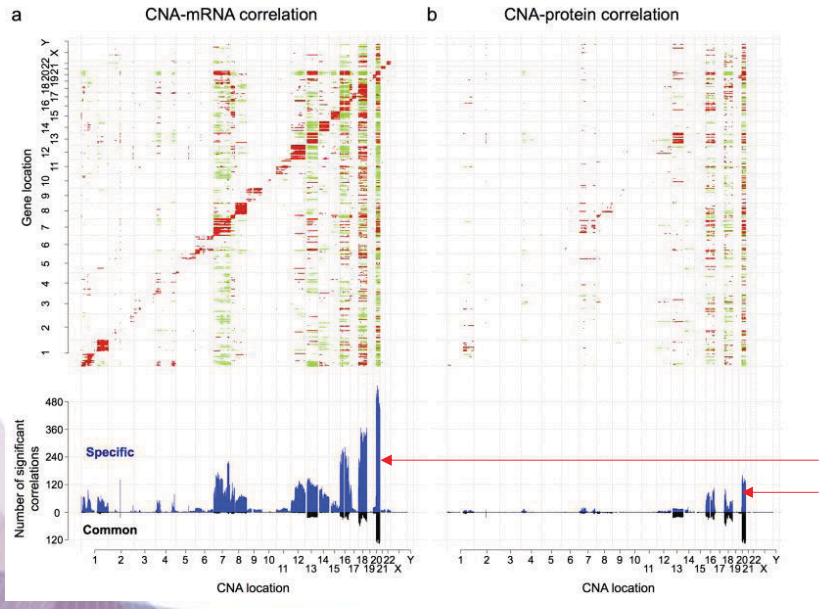


Carr, Nature Review Cancer, 2022

유전단백체 기반 암 연구 동향

1. Deciphering proteomic effects of genomic alterations

- measuring the correlation between different entities



Proteogenomic effect of CNA and DNA methylation

Identify CNA and/or DNA methylation associated with coordinated *cis* regulation in RNA, protein and PTM

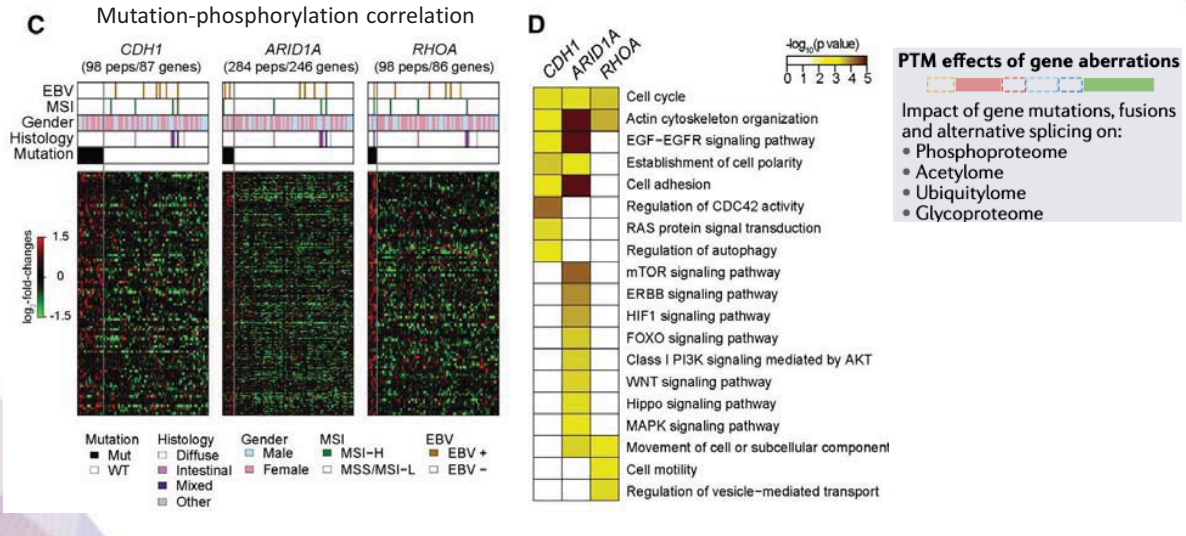
Impact of chr20q amplification on mRNA and protein (trans-acting CNN hotspot)
→ potential cancer driver

Zhang, Nature, 2014 (Colorectal cancer)

유전단백체 기반 암 연구 동향

1. Deciphering proteomic effects of genomic alterations

- measuring the correlation between different entities

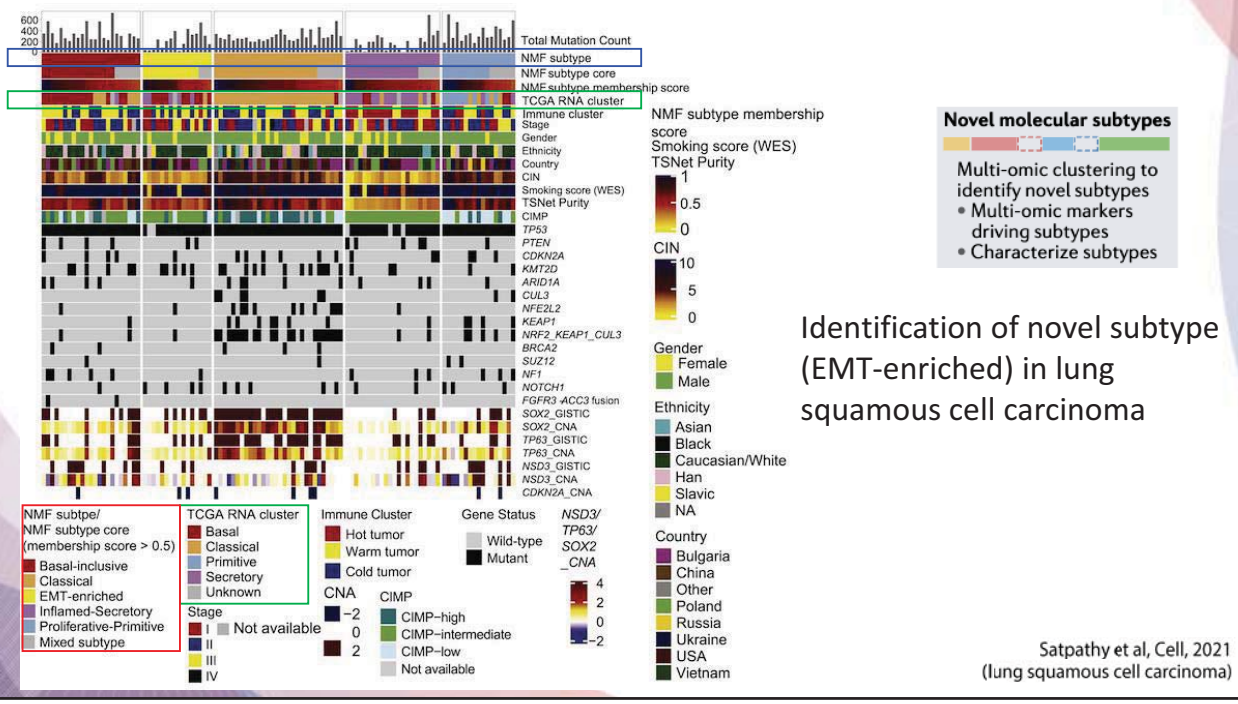


Bhin and Mun et al, Cancer Cell, 2019 (Gastric cancer)

유전단백체 기반 암 연구 동향

2. Redefining and characterizing cancer subtypes

- integrative clustering of transcriptomic and proteomic data



Identification of novel subtype (EMT-enriched) in lung squamous cell carcinoma

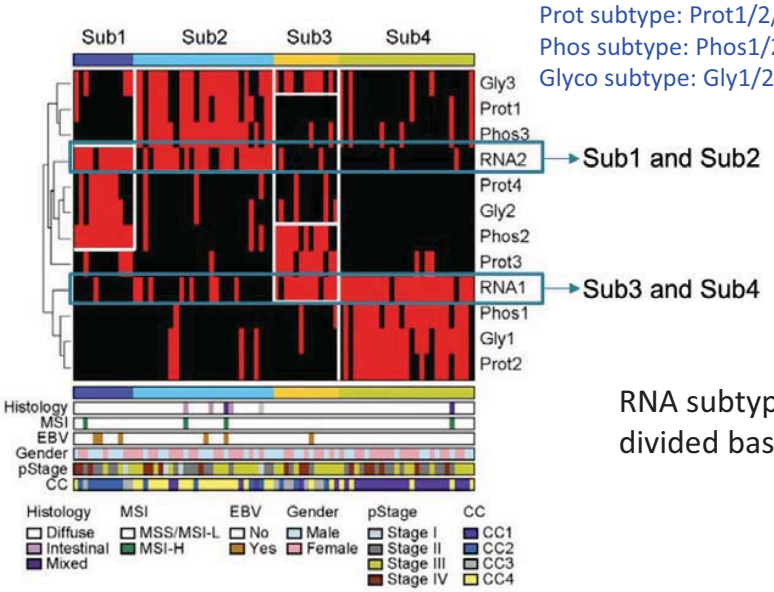
Satpathy et al, Cell, 2021
(lung squamous cell carcinoma)

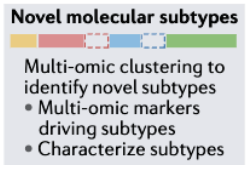
유전단백체 기반 암 연구 동향

2. Redefining and characterizing cancer subtypes

- integrative clustering of transcriptomic and proteomic data

RNA subtype: RNA1/2
Prot subtype: Prot1/2/3/4
Phos subtype: Phos1/2/3
Glyco subtype: Gly1/2/3





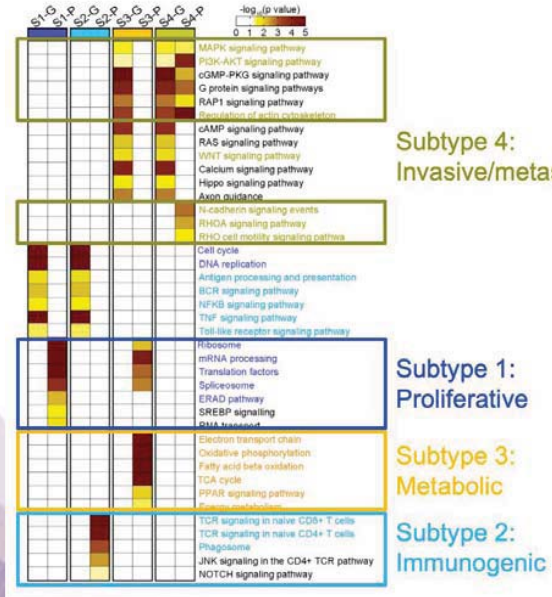
RNA subtypes can be further divided based on proteomics data

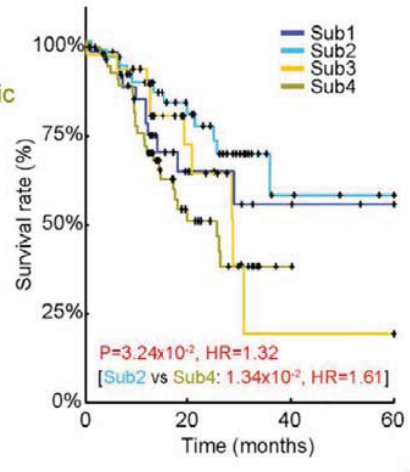
Bhin and Mun et al, Cancer Cell, 2019 (Gastric cancer)

유전단백체 기반 암 연구 동향

2. Redefining and characterizing cancer subtypes

- integrative clustering of transcriptomic and proteomic data



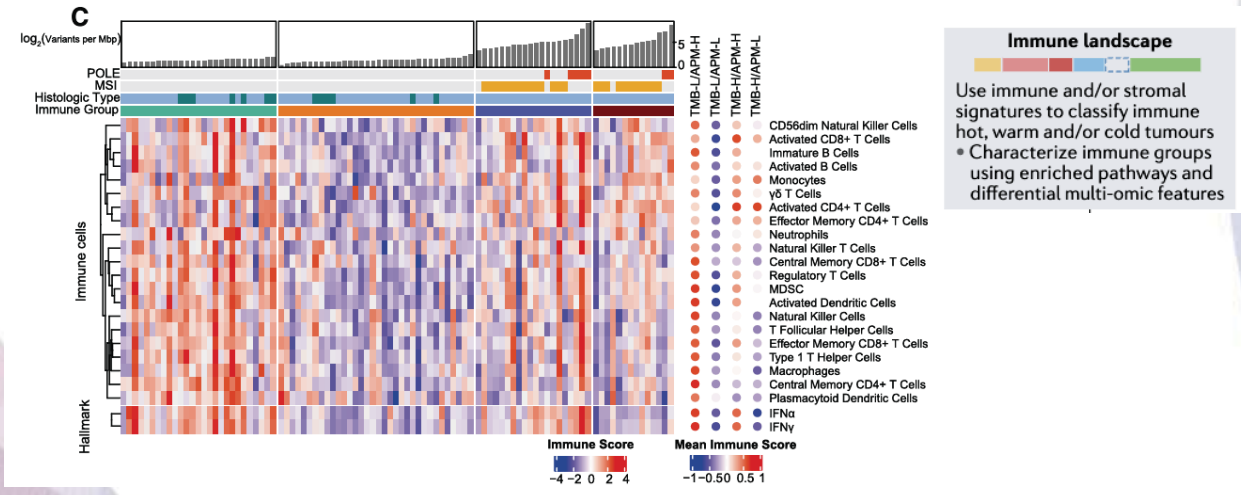


Bhin and Mun et al, Cancer Cell, 2019 (Gastric cancer)

유전단백체 기반 암 연구 동향

3. Assessing tumor immunogenicity

- proteomic evidence for immune status and neoantigen prediction



Immune landscape

Use immune and/or stromal signatures to classify immune hot, warm and/or cold tumours

- Characterize immune groups using enriched pathways and differential multi-omic features

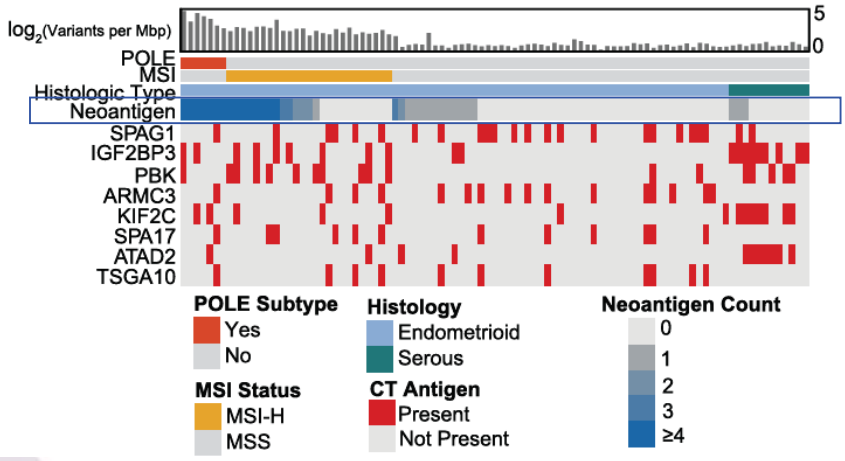
4 Immune subtypes in Endometrial carcinoma supported by proteomics data

Dou et al, Cell, 2020 (Endometrial carcinoma)

유전단백체 기반 암 연구 동향

3. Assessing tumor immunogenicity

- proteomic evidence for immune status and neoantigen prediction



Immune landscape

Use immune and/or stromal signatures to classify immune hot, warm and/or cold tumours

- Characterize immune groups using enriched pathways and differential multi-omic features

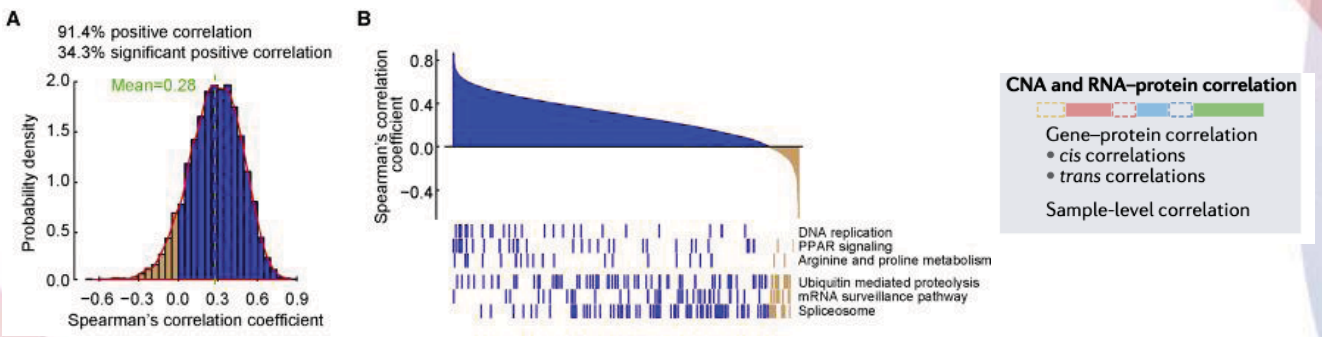
~50% of samples have evidence for putative neoantigens in endometrial carcinoma

Dou et al, Cell, 2020 (Endometrial carcinoma)

유전단백체 기반 암 연구 동향

4. mRNA-protein correlation

- identifying pathways showing high or low mRNA-protein correlation



- In general, Pearson's correlation coefficient between mRNA and protein is 0.3-0.45 (~1/3 of genes show significant correlations)
- Genes in amino acid and fatty acid metabolism show good correlations while genes in house keeping pathways show poor correlations

Bhin and Mun et al, Cancer Cell, 2019 (Gastric cancer)

유전단백체 기반 암 연구 동향

5. Outlier analysis

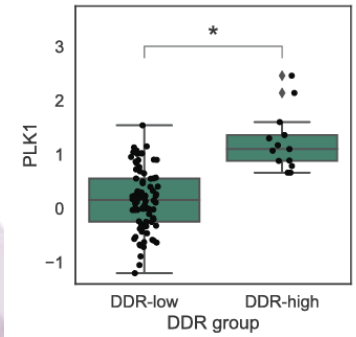
- identifying new potential therapeutic targets

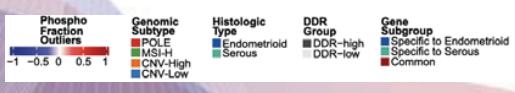
Outlier analysis

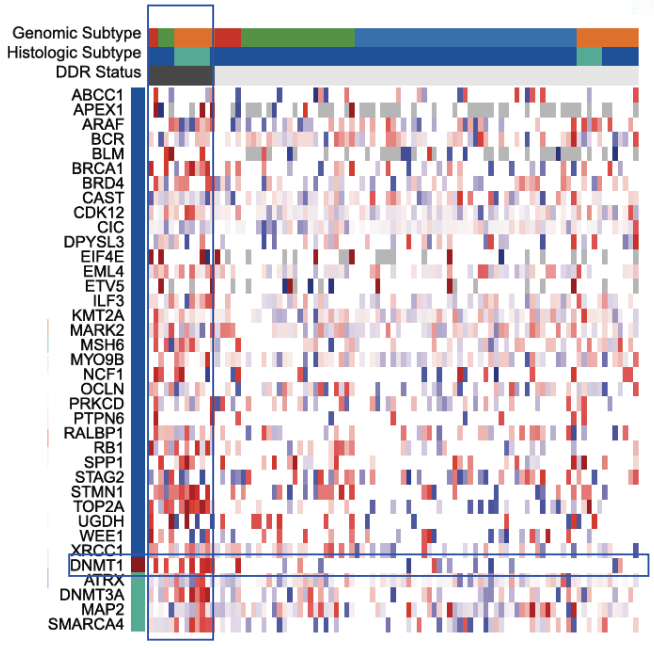
Identify outlier kinases and phosphotases
- Across multiple -omes
- Associate with SMGs
- Determine druggability

DNMT1: FDA-approved drug target

Phosphorylation levels of known DNA damage response (DDR) markers
→ DDR-low and DDR-high







Dou et al, Cell, 2020 (Endometrial carcinoma)

목차

- 유전단백체학 (Proteogenomics) 개념 소개
- 유전단백체학을 이용한 암 연구 동향
- **유전단백체학 분석 방법론/툴 소개**

유전단백체 연구를 위한 분석툴

1. Data access
2. Generation of customized protein sequence database
3. Multi-omics-based patient subtyping
4. Correlation analysis between different molecular entities
5. Neoantigen prediction

유전단백체 연구를 위한 분석툴

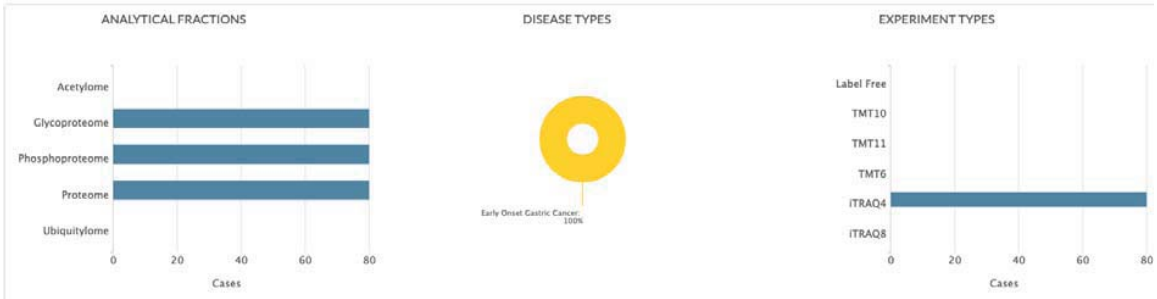
1. Data access

Proteomics data from CPTAC and its partners: Proteomic Data Commons
https://proteomic.datacommons.cancer.gov/pdc/

유전단백체 연구를 위한 분석툴

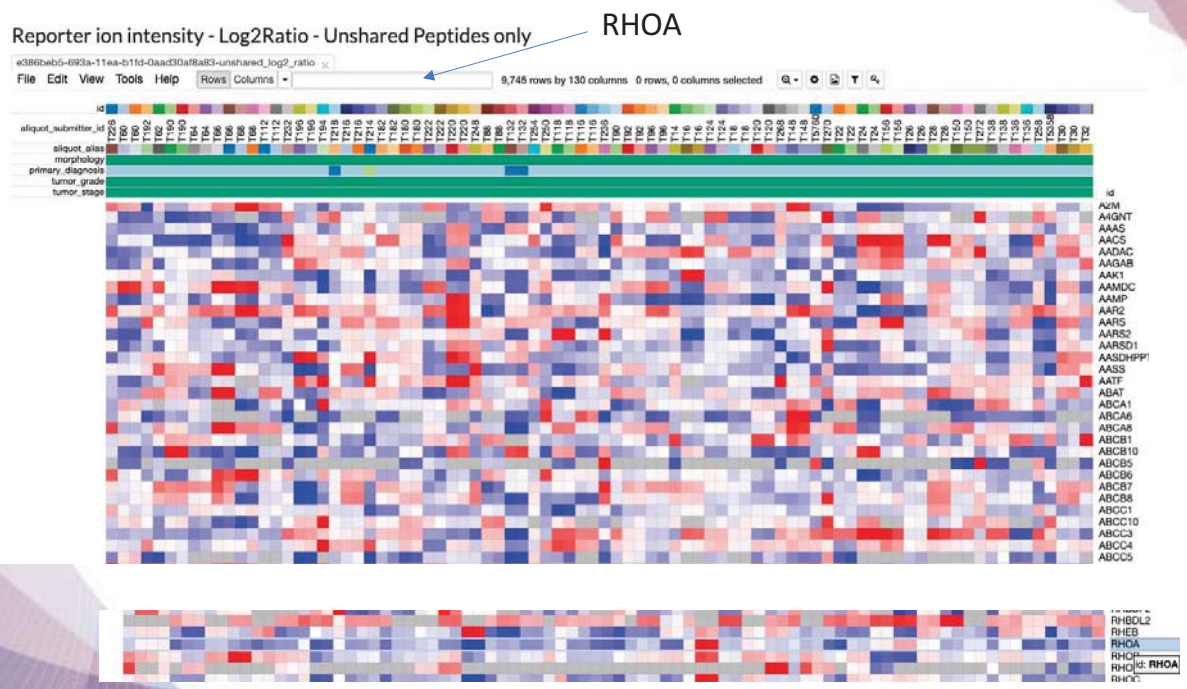
1. Data access

Proteomics data from CPTAC and its partners: **Proteomic Data Commons**
https://proteomic.datacommons.cancer.gov/pdc/



PDC Study ID	Study	Project	Program	Disease Type	Primary Site	Analytical Fraction	Experiment Type	Cases #	Available files for data category					
									Raw	Processed Mass Spectra	Metadata	PSM	Protein Assembly	Quality Metrics
PDC000216	Proteogenomics of Gastric Cancer - Glycoproteome	Human Early-Onset Gastric Cancer - Korea University	International Cancer Proteogenome Consortium	Early Onset Gastric Cancer	Stomach	Glycoproteome	iTRAQ4	80	779	779	4	0	0	0
PDC000215	Proteogenomics of Gastric Cancer - Phosphoproteome	Human Early-Onset Gastric Cancer - Korea University	International Cancer Proteogenome Consortium	Early Onset Gastric Cancer	Stomach	Phosphoproteome	iTRAQ4	80	780	780	4	1560	5	2
PDC000214	Proteogenomics of Gastric Cancer - Proteome	Human Early-Onset Gastric Cancer - Korea University	International Cancer Proteogenome Consortium	Early Onset Gastric Cancer	Stomach	Proteome	iTRAQ4	80	1560	1560	4	3120	4	2

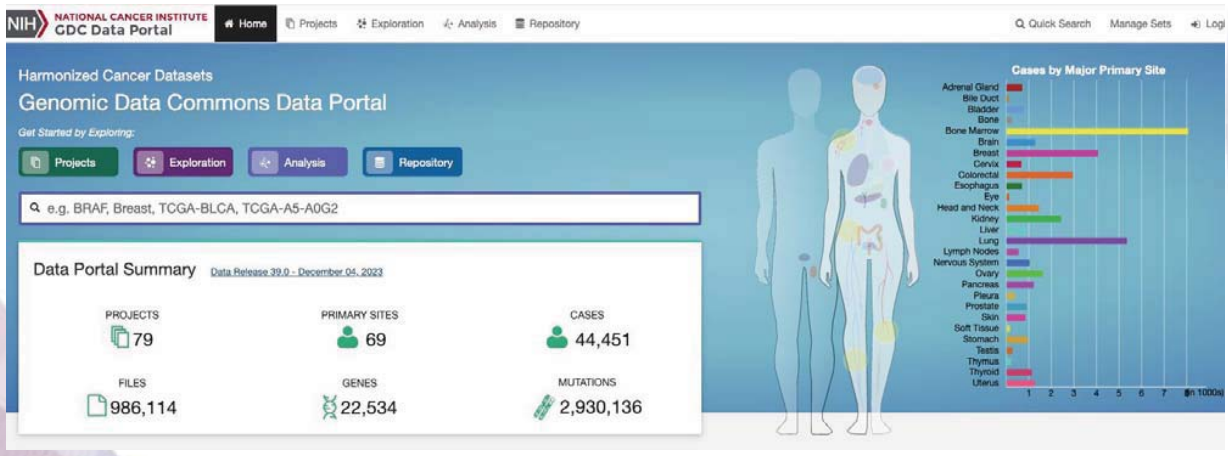
유전단백체 연구를 위한 분석툴



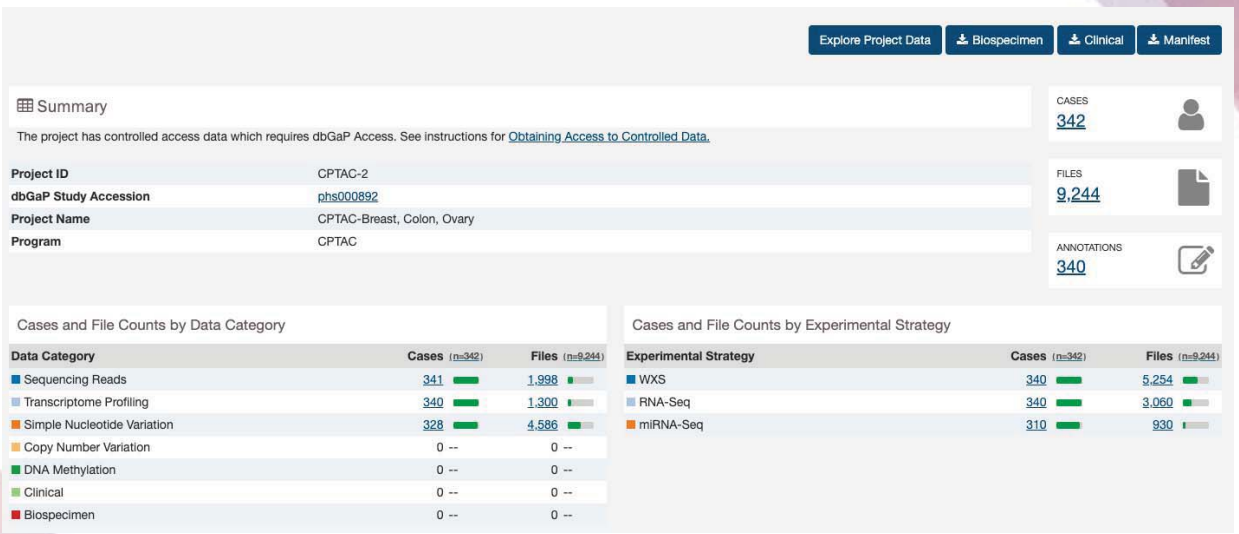
유전단백체 연구를 위한 분석툴

1. Data access

Genomics data from CPTAC: **Genomic data commons**
https://portal.gdc.cancer.gov/



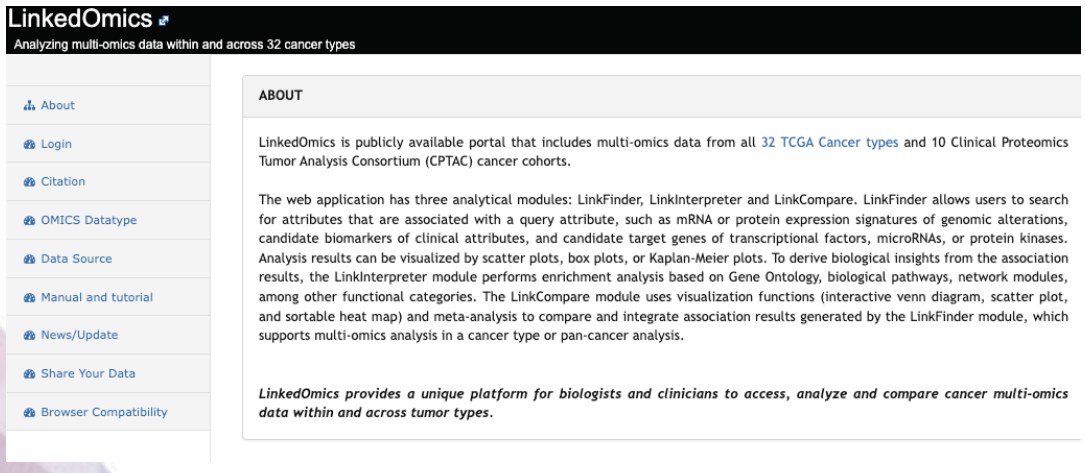
유전단백체 연구를 위한 분석툴



유전단백체 연구를 위한 분석툴

1. Data access

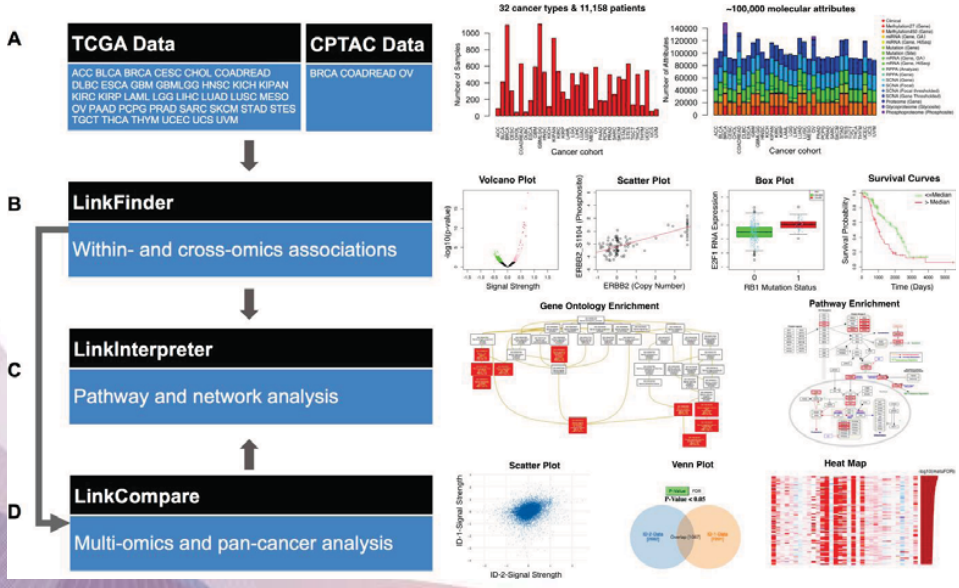
Standardized multi-omic data matrices (TCGA, CPTAC)
https://www.linkedomics.org/



유전단백체 연구를 위한 분석툴

1. Data access

Standardized multi-omic data matrices (TCGA, CPTAC)
https://www.linkedomics.org/



유전단백체 연구를 위한 분석툴

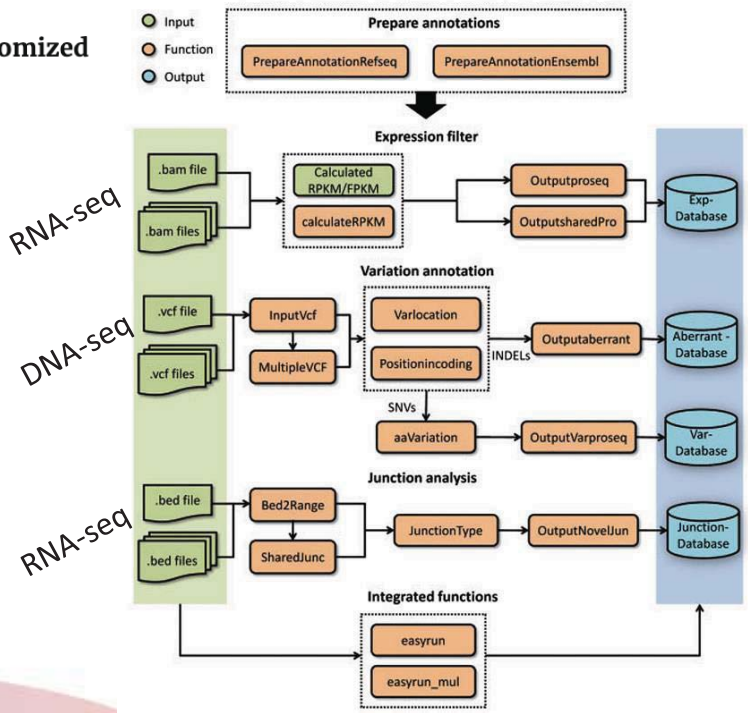
2. Customized protein sequence database

JOURNAL ARTICLE

***customProDB*: an R package to generate customized protein databases from RNA-Seq data for proteomics search**

Xiaojing Wang, Bing Zhang ✉ Author Notes

Bioinformatics, Volume 29, Issue 24, December 2013, Pages 3235–3237,
https://doi.org/10.1093/bioinformatics/btt543
Published: 20 September 2013 **Article history** ▾

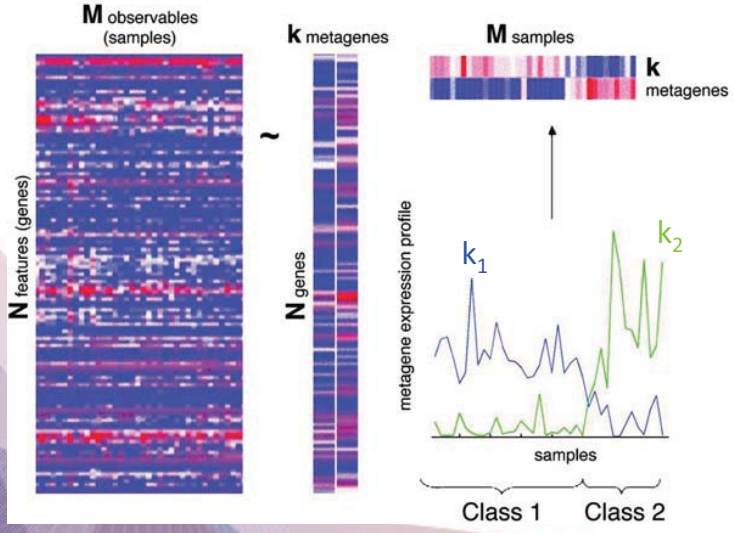


유전단백체 연구를 위한 분석툴

3. Multi-omics-based patient subtyping

- Non-negative matrix factorization (NMF) : for single omics data

A (rank M) ~ W H (rank k=2)



A: non-negative values

A = W x H

Objective function:

$$\sum(\boldsymbol{A} - WH)^2;$$

Assign each sample m to class k that has the highest score

유전단백체 연구를 위한 분석툴

3. Multi-omics-based patient subtyping

- Non-negative matrix factorization (NMF) : for single omics data

- Consensus clustering - multiple iterations of the chosen clustering method on subsamples of the dataset
- By introducing sampling variability with sub-sampling, this provides metrics to assess the stability of the clusters

K = 2 (0: not sampled)

	iter1	iter2	iter3	iter4
Alice	2	0	2	1
Brian	1	1	1	2
Sally	1	2	0	2
James	0	2	1	0

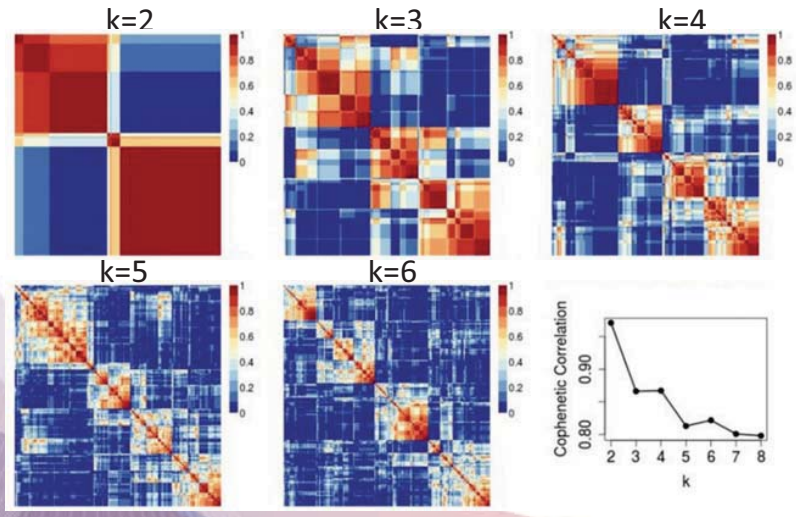
	Alice	Brian	Sally	James
Alice	1	0.00	0.00	0.0
Brian	0	1.00	0.67	0.5
Sally	0	0.67	1.00	1.0
James	0	0.50	1.00	1.0

유전단백체 연구를 위한 분석툴

3. Multi-omics-based patient subtyping

- Non-negative matrix factorization (NMF) : for single omics data

Consensus NMF clustering

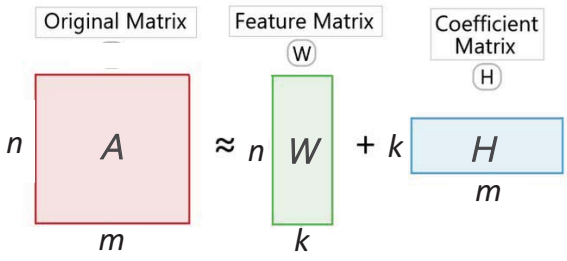


유전단백체 연구를 위한 분석툴

3. Multi-omics-based patient subtyping

- Integrative Non-negative matrix factorization (NMF)

$$A_{n\times m} \approx W_{n\times k} H_{k\times p}$$



Modification for integrative clustering

$$A^{i}_{n\times m} \approx W^{i}_{n\times k} \times H_{k\times m} \quad i : \text{data type } (i = 1, 2, \ldots p)$$

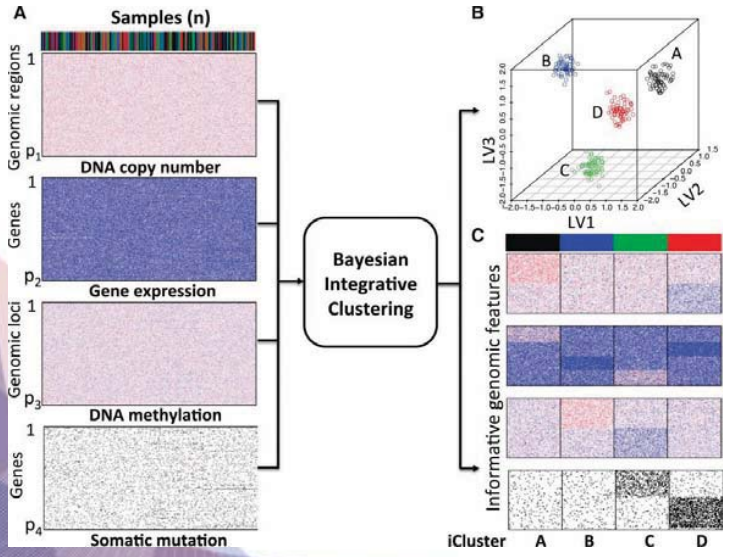
$$Q = \min_{W,H} \sum_{i=1}^{p} \theta^{i} ||Xi - WiH||_2$$

Chalise et al., PLOS ONE (2017)

유전단백체 연구를 위한 분석툴

3. Multi-omics-based patient subtyping

- iCluster → iClusterPlus → iClusterBayes



- Joint integrative clustering to model continuous and discrete omics data
- Reduce multi-high dimensional space to a low dimensional subspace that will collectively capture the major variations of the multiple datasets
  → low dimensional subspace can be used to cluster the samples
- Simultaneous identification of features that contribute to sample clustering

Mo et al, Biostatistics, 2018

유전단백체 연구를 위한 분석툴

iClusterPlus: integrative clustering of multiple genomic data sets

Qianxing Mo[1] and Ronglai Shen[2]

October 24, 2023

[1]Department of Biostatistics & Bioinformatics
H. Lee Moffitt Cancer Center & Research Institute
qianxing.mo@moffitt.org

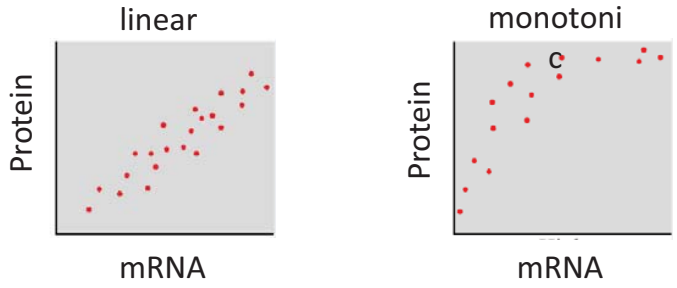
[2]Deparment of Epidemiology and Biostatistics
Memorial Sloan-Kettering Cancer Center
shenr@mskcc.org

R package for iClusterPlus and iClusterBayes

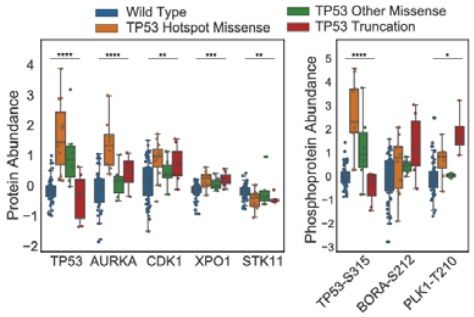
유전단백체 연구를 위한 분석툴

4. Correlation analysis between different molecular entities

- Use Pearson (linear relationship) or Spearman (monotonic relationship)



- t-test (parametric) or Mann-Whitney U test (Non-parametric)

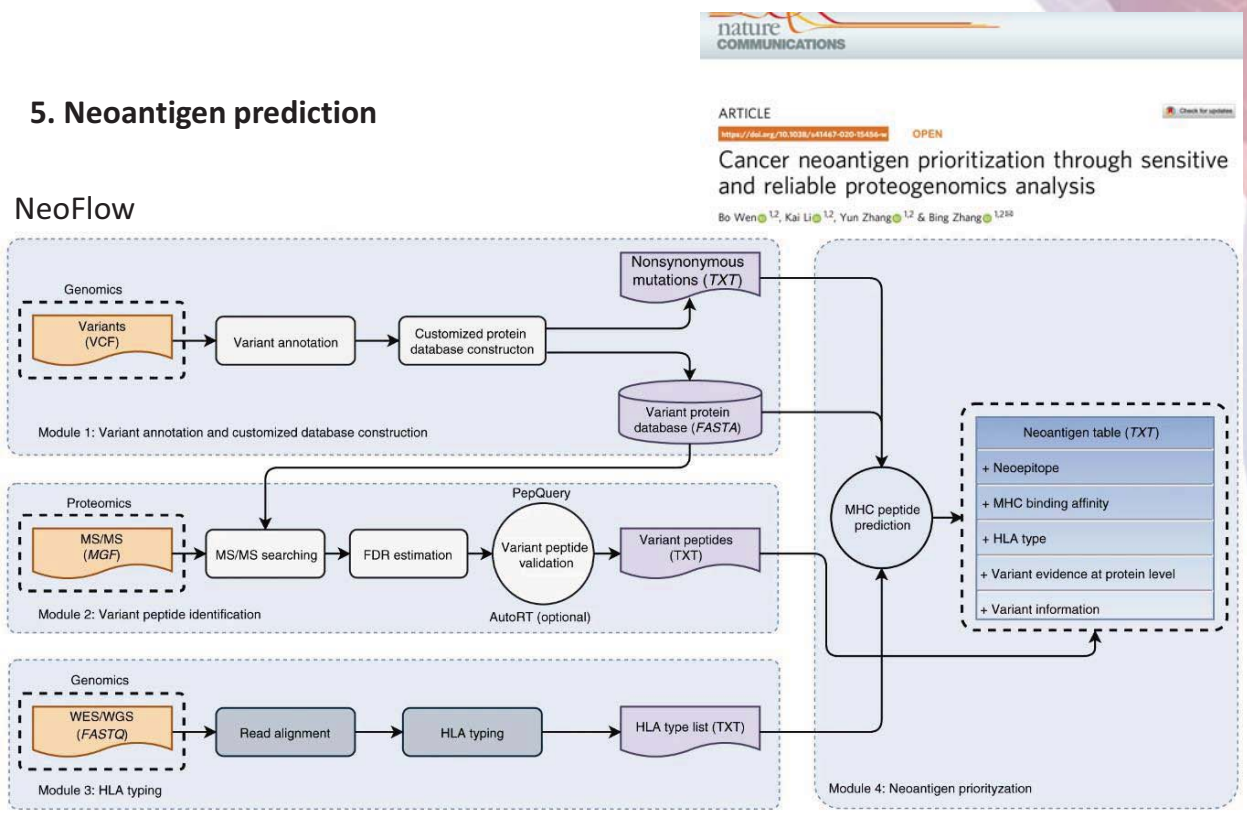


Dou et al, Cell, 2020 (Endometrial carcinoma)

유전단백체 연구를 위한 분석툴

5. Neoantigen prediction

NeoFlow



유전단백체 연구를 위한 툴

Data analysis — association discovery and visualization

- LinkedOmics: Interactive visualization and analysis of multi-omics data from TCGA and CPTAC within and across cancer types[155]
- multiOmicsViz: Visualizing the effect of one omics data type on other omics data along the chromosome[27]
- iProFun: Characterization of functional consequences of copy number alterations (CNAs) and methylation alterations in tumours[99]
- ProNetView: Network visualization inferred from proteogenomic data of the CPTAC-CCRCC project
- NetSAM: Network seriation and modularization[164]
- NetGestalt: Visualization and analysis of biological networks from multi-omics data[164]
- Panoptes (1), Panoptes (2): Analysing pathological images using deep learning
- TSNet: A statistical model for cell type-specific inference based on bulk tumour profiling data[165]
- Perseus: Bioinformatics platform for integrative analysis of proteomic data[166,167]

Data analysis — PTM and pathway analysis

- PTMsigDB: Collection of PTM site-specific signatures[102]
- PTM-SEA: Modified version of ssGSEA for PTMsigDB[102]
- WebGestalt: GSEA toolkit with support for phosphosite enrichment analysis[170]
- PTMcosmos: Database of PTMs and cancer mutations in humans
- PHOTON: Delineating signalling pathways from large-scale phosphoproteomic data[171]
- HotPho: Identification of spatially interacting phosphosites and mutations[172]
- Black Sheep: Identification of samples with aberrant gene, protein or PTM site abundances[173]
- WikiPathways: Database of molecular pathway diagrams contributed and refined by the research community[174]

Mani et al, Nat Rev Cancer, 2022

Contact

연세대학교 의과대학 의생명시스템정보학교실

E-mail: jbhin@yuhs.ac

Lab homepage: https://sites.google.com/view/cpm-yucm/

Office: 연세대학교 강남세브란스 미래의학연구센터 107호

Tel: 02-2019-5470

감사합니다