Applications of Proteomics in Medical Researches

Piriya Wongkongkathep, PhD

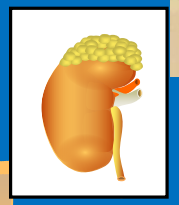
Center of Excellence in Systems Biology
Chulalongkorn University

Outline

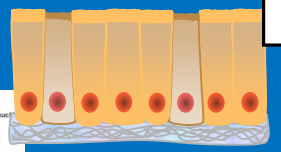
- Introduction to proteomics and systems biology
- Clinical proteomics and identifying biomarkers
- Experimental designs
- Recent advances in medical proteomics
- Proteomics in cancer immunotherapy at CUSB and KCMH

Hierarchies in Systems Biology

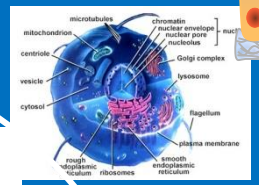
Organism



Organ



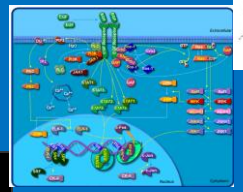
Tissue



Cell

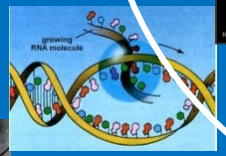
phenomics

Network



Pathway

Protein



RNA

transcriptomics

DNA

genomics

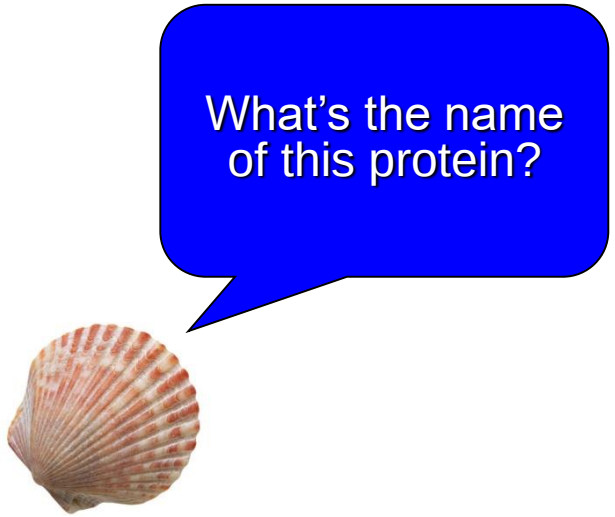
interactomics

proteomics

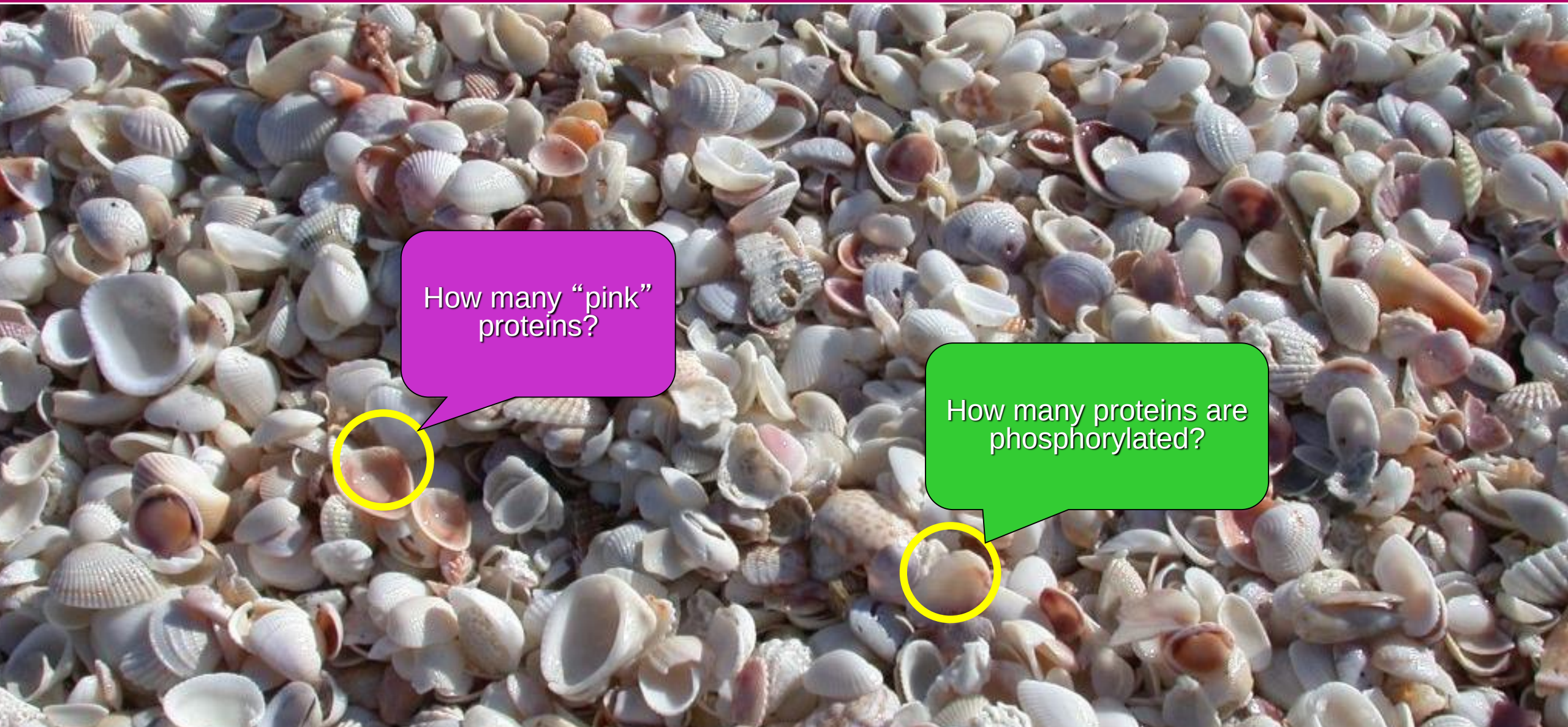
Physical Laws of Nature

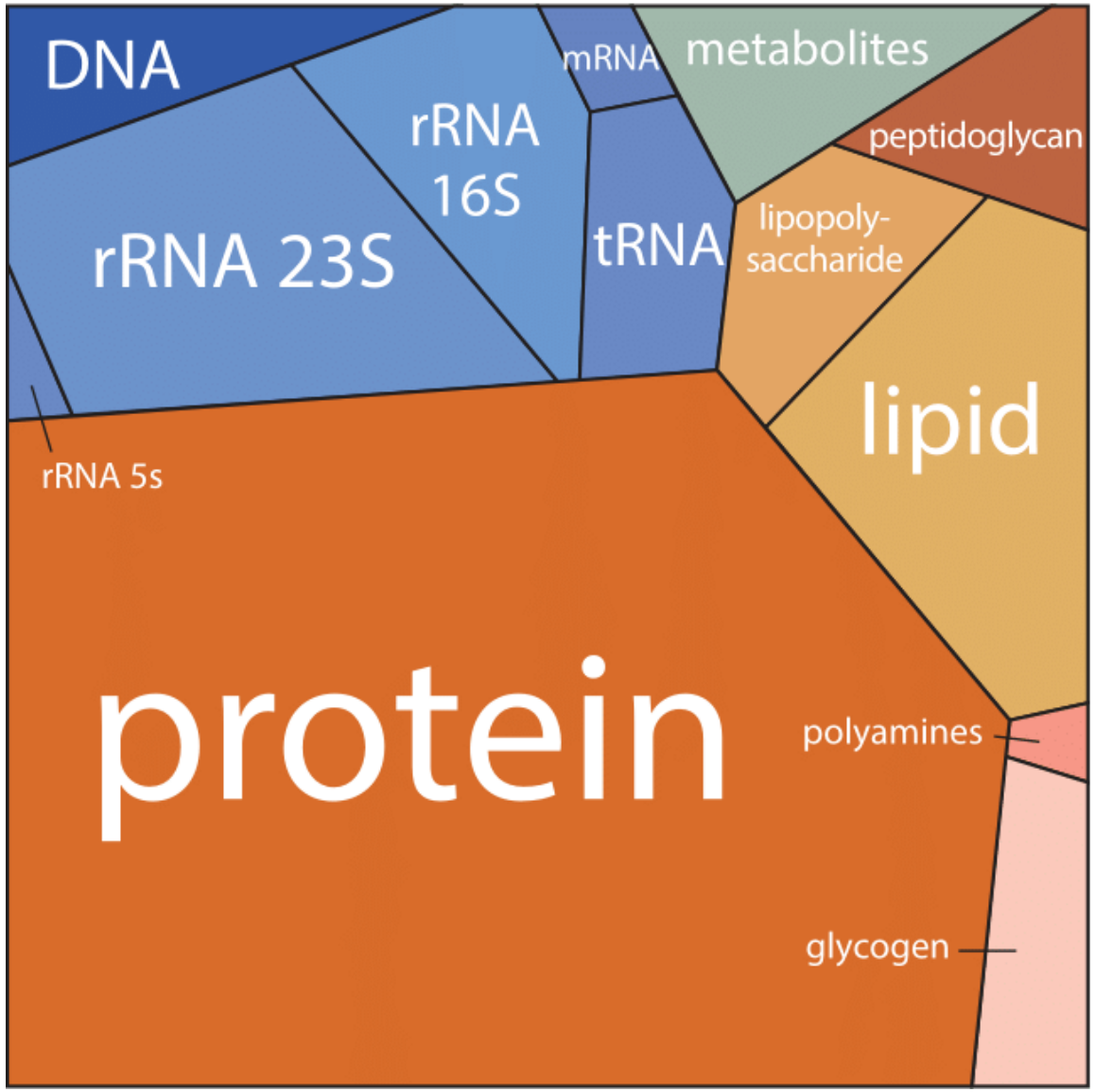
(Slide compliments of Joe Nadeau)

Proteomics



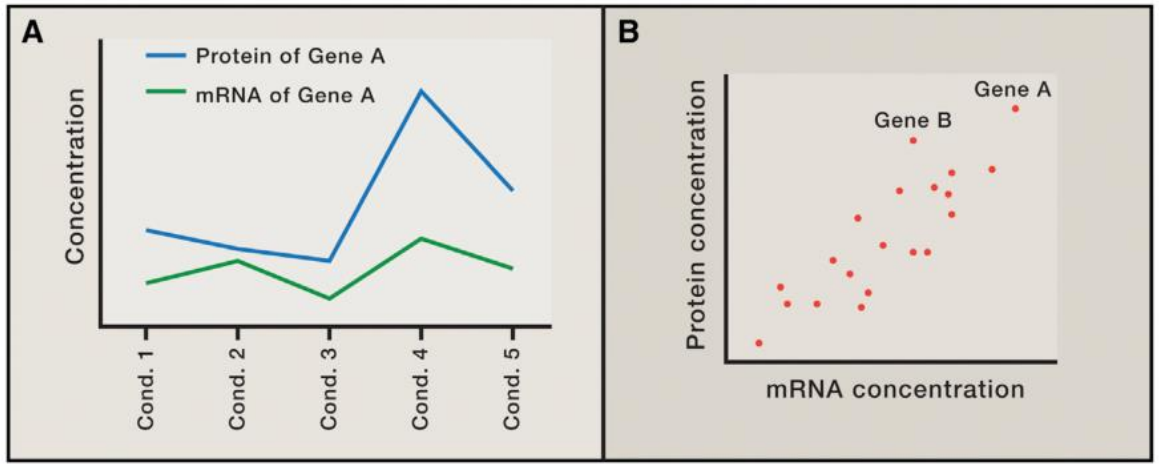
Proteomics – Characterizing Many Proteins

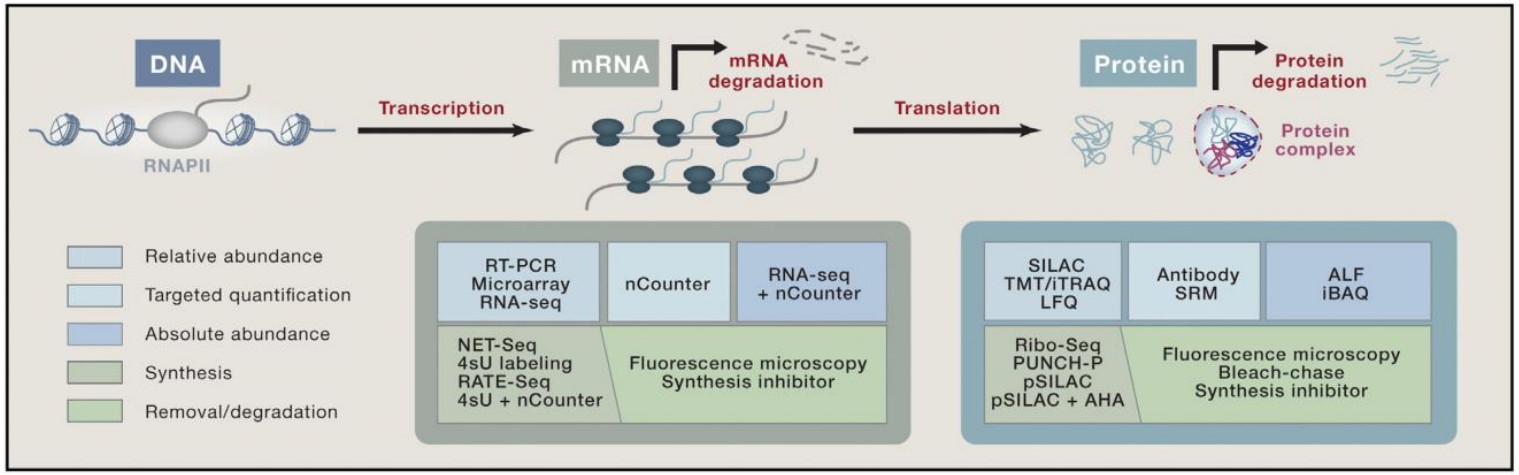


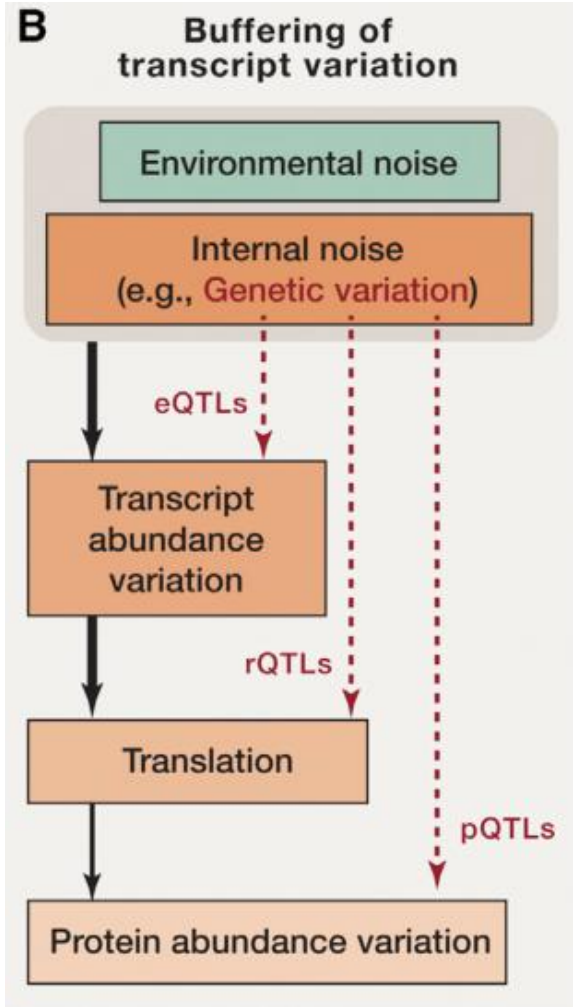


A Voronoi tree diagram of the composition of an E. coli cell growing with a doubling time of 40 min.

Dependency of Cellular Protein Levels on mRNA Abundance

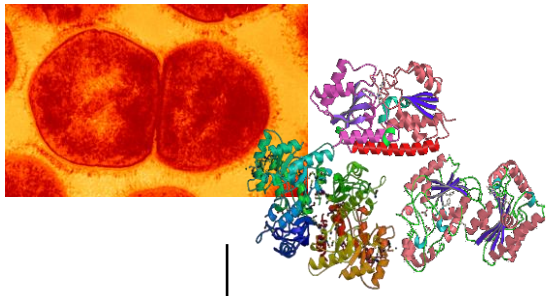




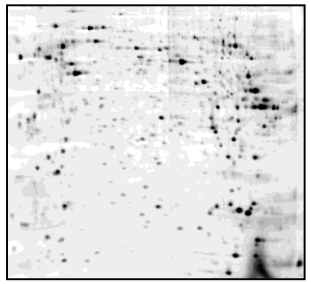


Liu Y, Beyer A, Aebersold R.
Cell. 2016 Apr 21;165(3):535-50.

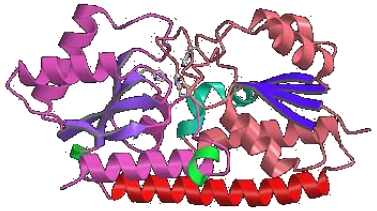
Measuring Proteome



complex protein mixture



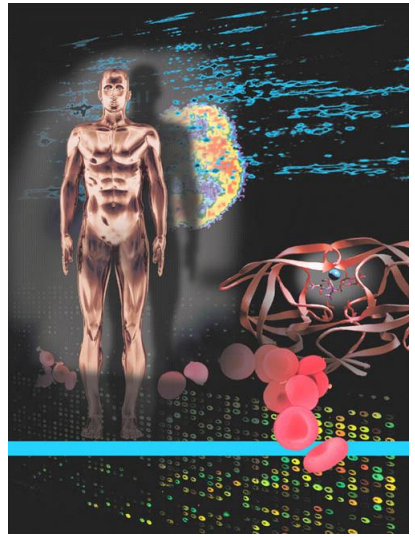
protein separation



protein identification
protein modification
protein abundance

protein detection (mass spectrometry)

Biology



Protein Identification

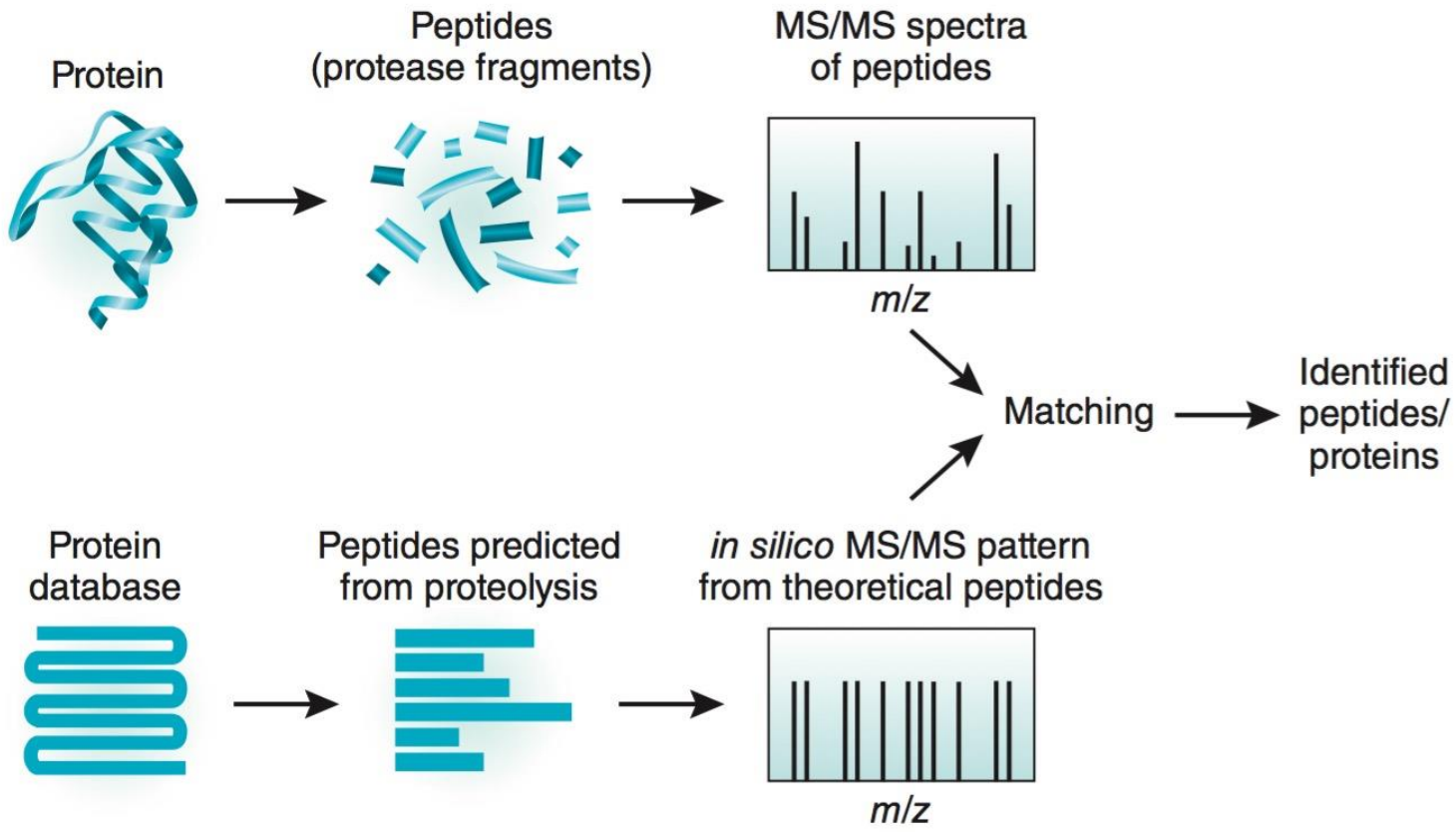
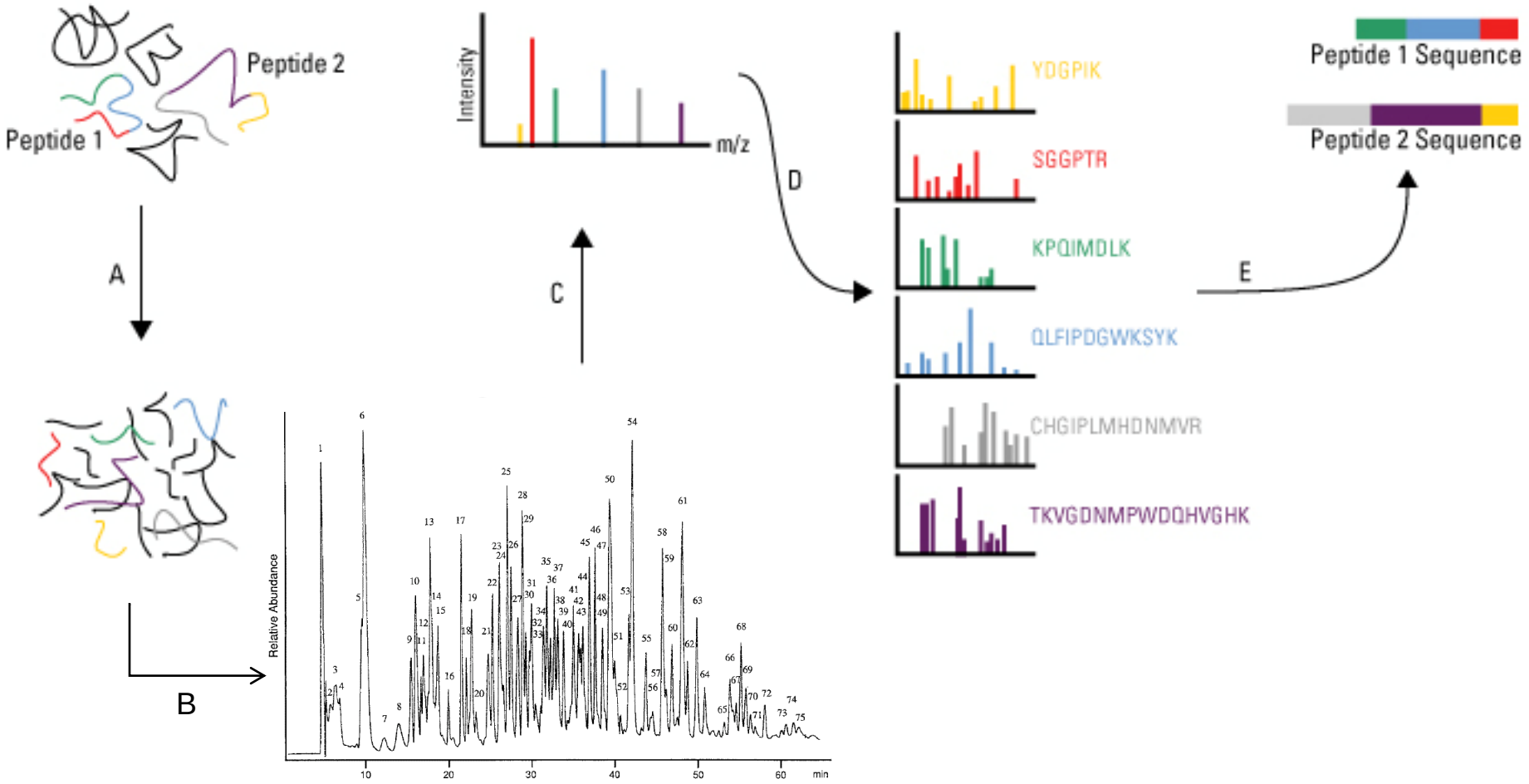
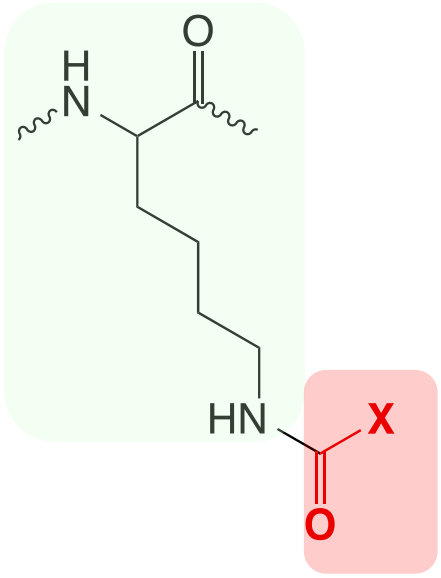


Figure 1 General approach used by peptide-centric MS technologies for the identification of proteins in complex mixtures. After proteolysis of a protein or complex mixture of proteins, the spectra associated with protease fragments are matched with spectra generated *in silico* using information obtained from protein databases.

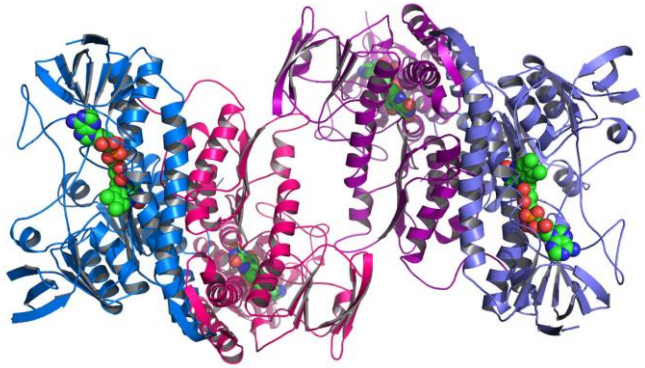
Protein Identification

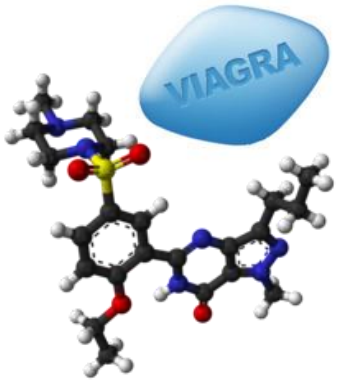


Molecular to Medicine



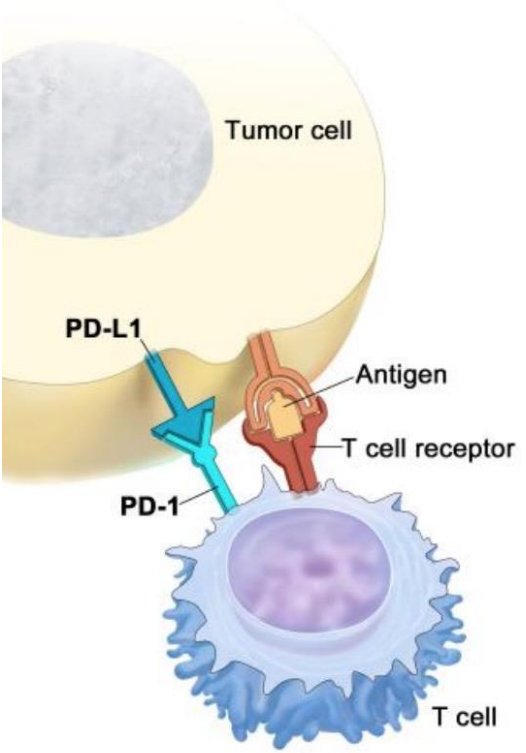
Post-translational Modifications





Drug Binding

Protein Biomarker



Cancer

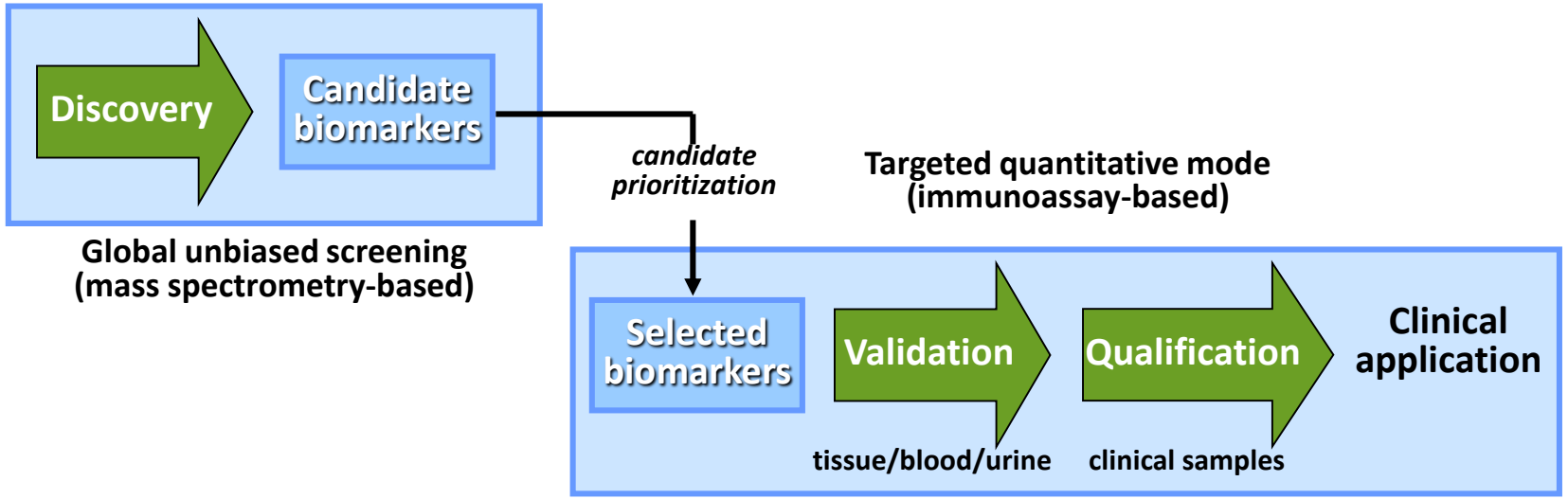
Clinical Proteomics and Discovering Biomarkers

What is a Biomarker and why do we care?

- Any molecule whose presence/abundance is indicative of a diseased state
- Quick and accurate diagnosis allow physicians to manage risks and treat patients more effectively
- Provides insight into the biochemistry of disease progression
- Detecting diseases when they are at their earliest stages could result in better clinical outcomes

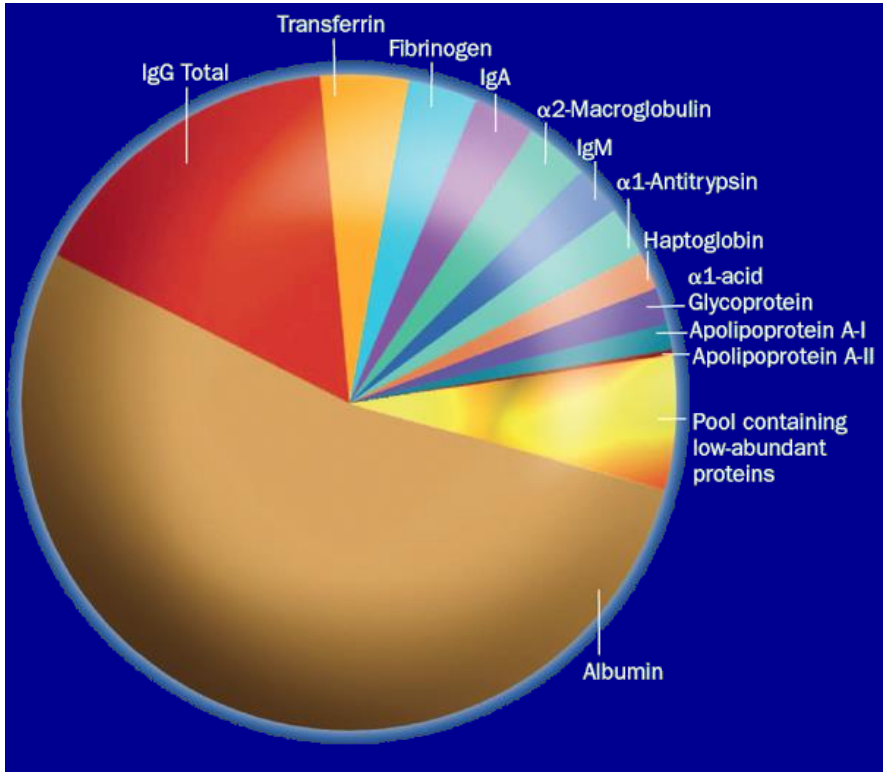
Clinical Proteomics and Discovering Biomarkers

- **Clinical proteomics:** to define proteins that provide clinically useful information about susceptibility to disease, diagnosis, prognosis, and guided therapy
- **Biomarker:** Quantifiable molecules or processes indicative of a certain biological state or condition
  - Detecting diseases when they are at their earliest stages could result in better clinical outcomes



Challenges of Plasma Proteome

- 2.5 L of plasma in an adult contains 250 g of proteins
- 10^{12}-fold concentration range between proteins
- 99% = 22 proteins
- Serum albumin 35-50 mg/mL
- Interleukin 6 (IL-6) 5 pg/mL
- Amyloid-β ~40 pg/mL
  - 10^{9} times less than albumin
- Most potential biomarkers are secreted into blood at very low copy number, especially in the early onset of diseases
- Approaches
  - Depleting high abundant proteins
  - Immunoenrichment



Ruiz A, *et al*, Plos One. 2013
Plasma Proteome Institute

Identifying HCC Therapeutic Targets

LETTER

https://doi.org/10.1038/s41586-019-0987-8

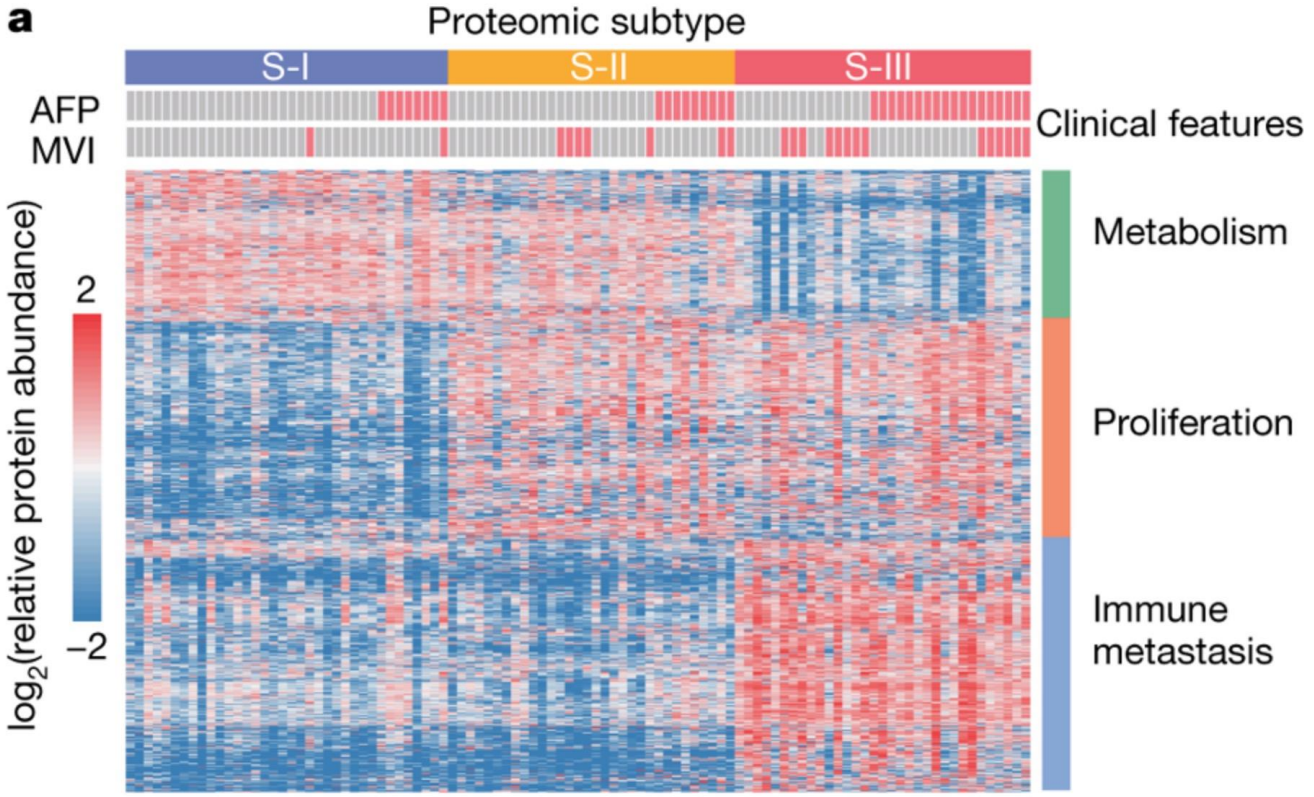
Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma

Ying Jiang[1,8], Aihua Sun[1,8], Yang Zhao[1,2,8], Wantao Ying[1,8], Huichuan Sun[3,8], Xinrong Yang[3,8], Baocai Xing[4,8], Wei Sun[1], Liangliang Ren[1], Bo Hu[3], Chaoying Li[1], Li Zhang[5], Guangrong Qin[6], Menghuan Zhang[6], Ning Chen[1], Manli Zhang[1], Yin Huang[1], Jinan Zhou[1], Yan Zhao[1], Mingwei Liu[1], Xiaodong Zhu[3], Yang Qiu[1], Yanjun Sun[1], Cheng Huang[3], Meng Yan[1], Mingchao Wang[1], Wei Liu[4], Fang Tian[1], Huali Xu[1], Jian Zhou[3], Zhenyu Wu[1], Tieliu Shi[5], Weimin Zhu[1], Jun Qin[1], Lu Xie[6], Chinese Human Proteome Project (CNHPP) Consortium[7], Jia Fan[3]*, Xiaohong Qian[1,2]* & Fuchu He[1]*

Hepatocellular carcinoma is the third leading cause of deaths from cancer worldwide. Infection with the hepatitis B virus is one of the leading risk factors for developing hepatocellular carcinoma, particularly in East Asia[1]. Although surgical treatment may be effective in the early stages, the five-year overall rate of survival after developing this cancer is only 50–70%[2]. Here, using proteomic and phospho-proteomic profiling, we characterize 110 paired tumour and non-tumour tissues of clinical early-stage hepatocellular carcinoma related to hepatitis B virus infection. Our quantitative proteomic data highlight heterogeneity in early-stage hepatocellular carcinoma: we used this to stratify the cohort into the subtypes S-I, S-II and S-III, each of which has a different clinical outcome. S-III, which is characterized by disrupted cholesterol homeostasis, is associated with the lowest overall rate of survival and the greatest risk of a poor prognosis after first-line surgery. The knockdown of sterol O-acyltransferase 1 (SOAT1)—high expression of which is a signature specific to the S-III subtype—alters the distribution of cellular cholesterol, and effectively suppresses the proliferation and migration of hepatocellular carcinoma. Finally, on the basis of a patient-derived tumour xenograft mouse model of hepatocellular

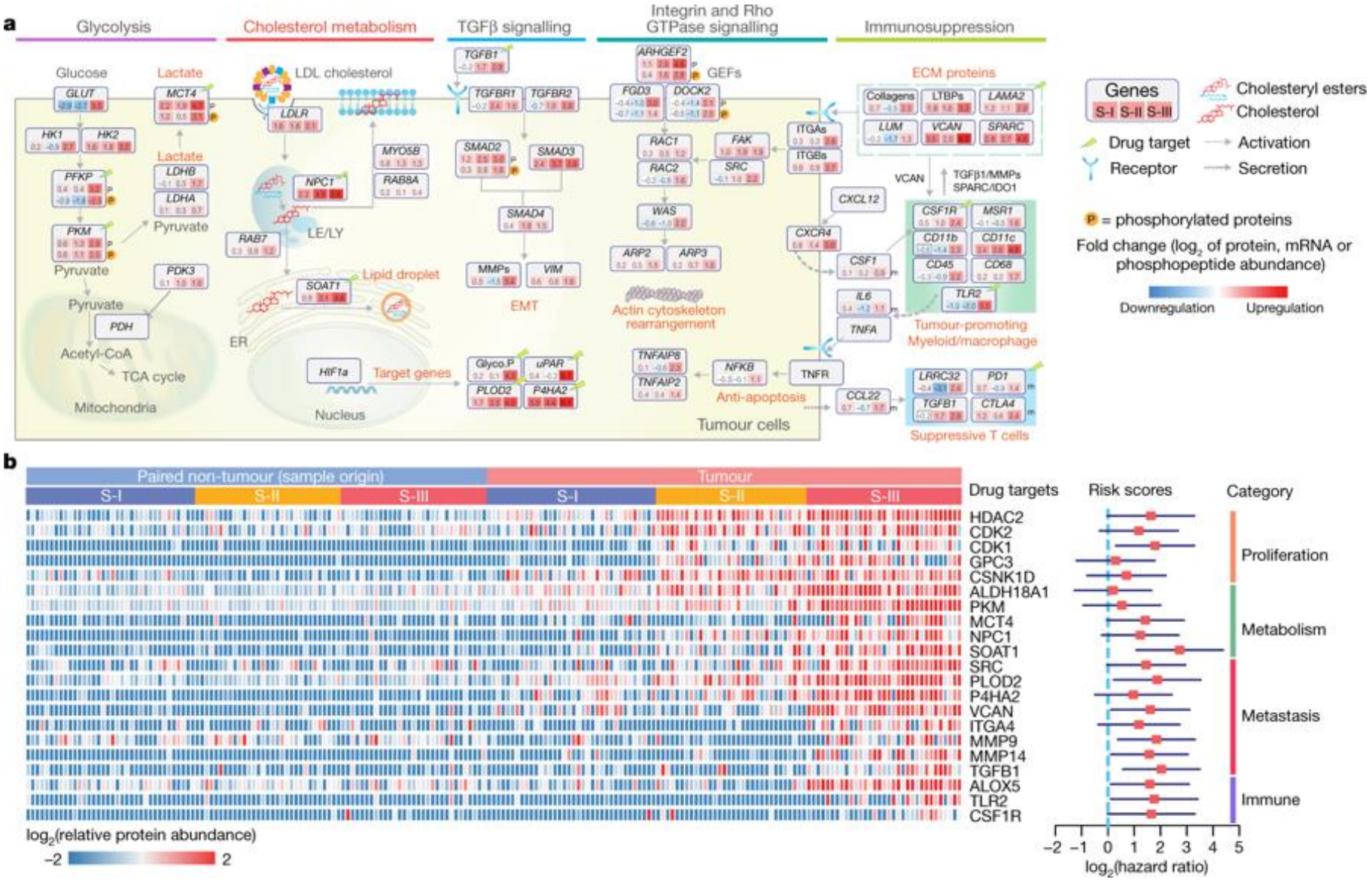
A case-by-case review shows that the number of proteins identified in the tumours is significantly higher than that identified in the paired non-tumour tissues (Fig. 1, Extended Data Fig. 3a). On average, 5,953 proteins per tumour and 5,114 proteins per non-tumour liver tissues were identified (Extended Data Fig. 1a). Furthermore, in the tumour samples, high levels of α-fetoprotein (AFP^{high}; AFP > 200ng ml^{-1}) and microscopic vascular invasion-positive (MVI^{+}) patients have a higher level of protein identification than AFP^{low} and MVI^{-} patients (Extended Data Fig. 3b). This association is also supported by the distribution of the RNA-seq results (Extended Data Fig. 3c–e).

A pathway enrichment analysis reveals that the cell cycle, integrin, PDGF signalling, MAPK, TNF, MET and other pathways are over-represented in the upregulated proteins, and that PPAR signalling and metabolism-related pathways are over-represented in the down-regulated proteins in early-stage HCC (Extended Data Fig. 4a, Supplementary Table 6). Compared with the proteome data, the phospho-proteome data additionally show the hyper-phosphorylation of signalling pathways—including the p38 pathway, RHO regulatory pathway, myosin pathway, RB1 pathway and IL1 signalling pathway (Extended Data Fig. 4b, Supplementary Table 7)—in HCC, which



Jiang Y, et al. Nature 2019

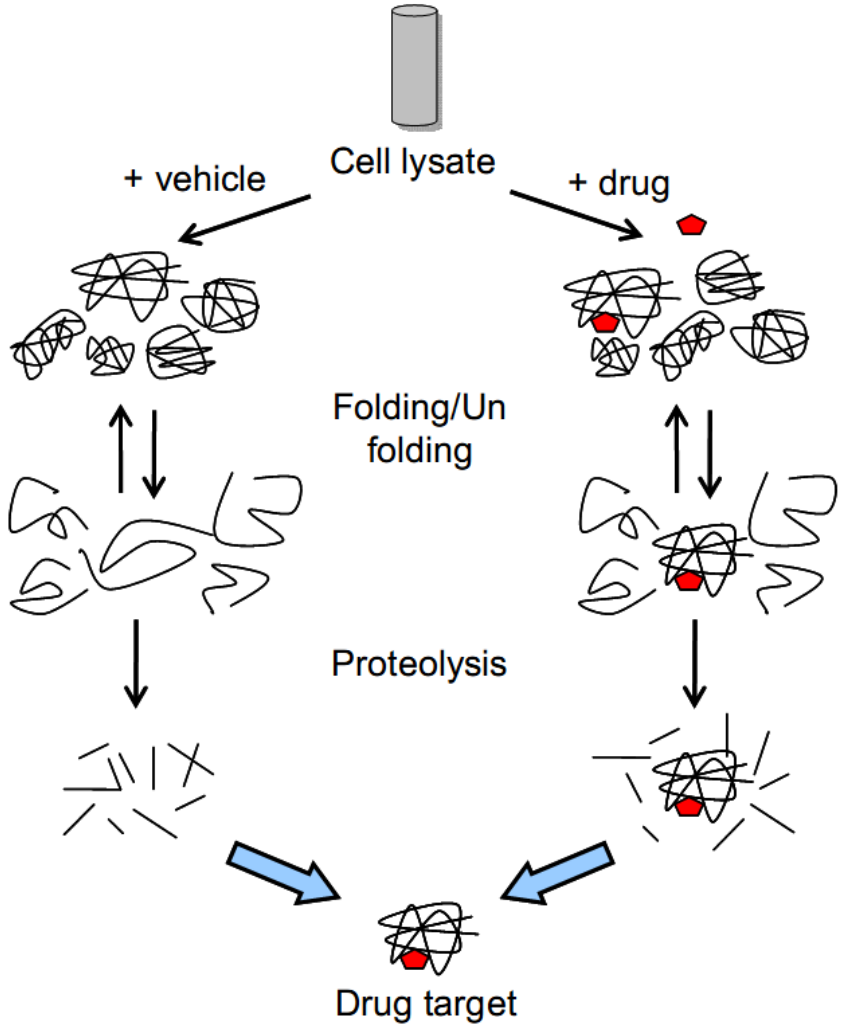
Identifying HCC Therapeutic Targets



Jiang Y, et al. Nature 2019

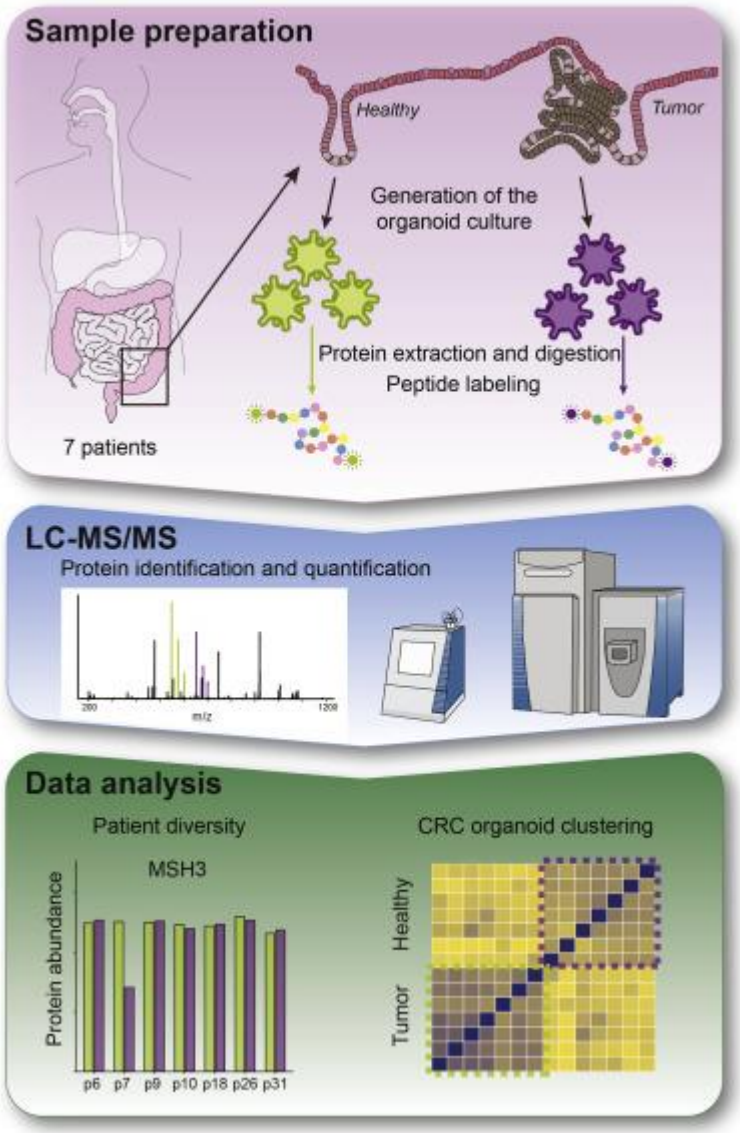
A new global way to detect drug targets - DARTS

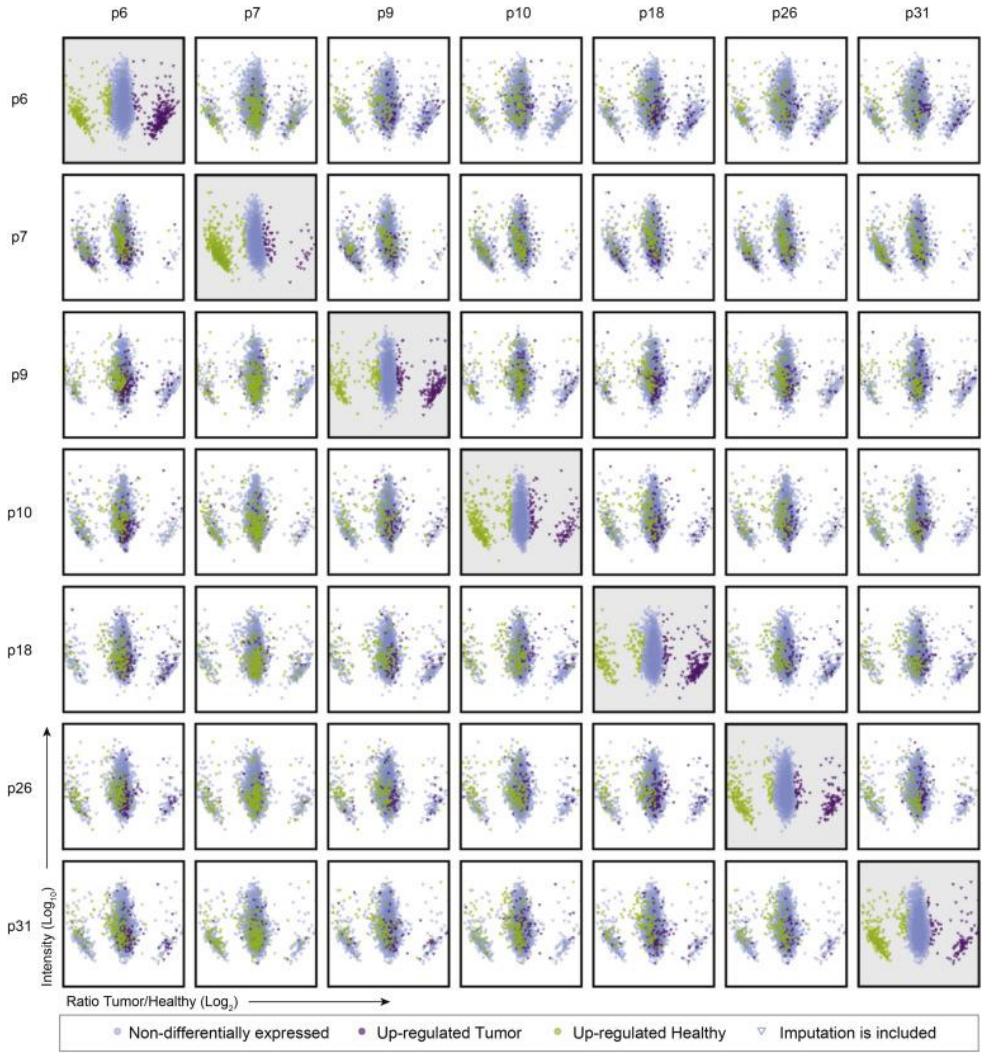
- **Drug Affinity Responsive Target Stability (DARTS)**
- Drug binding **reduces protease susceptibility**
- **Protein target stabilized** either globally or locally
- Does **not require modification** or immobilization of the small molecule



Huang, Lomenick et al., PNAS 2009

Precision Medicine

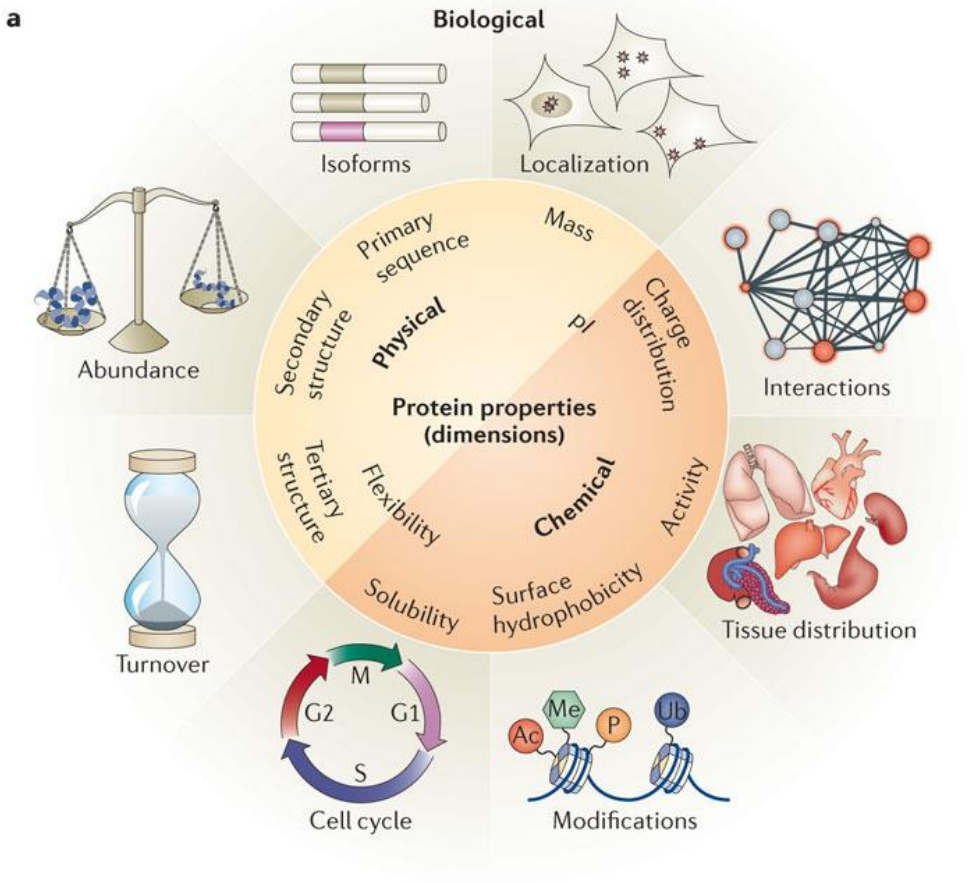


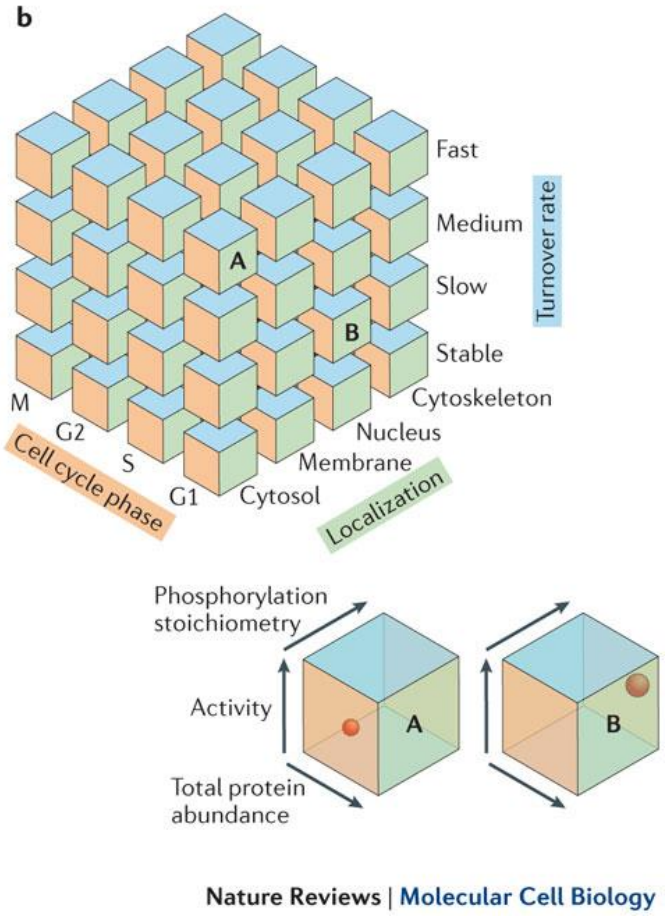


Cristobal A, et al, Cell Reports 2017

Experimental design considerations

Multidimensional proteomics for cell biology





Nature Reviews | Molecular Cell Biology

Dimension	Examples of techniques used	Refs
Abundance (absolute and relative)	Label-free quantitation	5,16–18
	SILAC	19
	^{15}N-labelling	20
	NeuCode SILAC	21
	Dimethyl-labelling	22,23
	TMT	24
	iTRAQ	25
Cell cycle regulation	Centrifugal elutriation	124
	Chemical inhibitors of cell cycle regulators	125
	FACS (for DNA content or phase-specific markers)	126
Tissue distribution	Dissection	95,127
	FACS (for cell-type-specific markers)	126
Interactions	Affinity-enrichment (endogenous immuno-precipitation or tagged fusion protein pull-down)	63–67
	Protein correlation profiling	9,70,71
	Proximity-labelling	39,68
Post-translational modifications	Affinity enrichment: TiO_2	128,129
	Affinity enrichment: IMAC	128,130
	Modification-specific antibodies	90,131–133
	Chromatography: IEX	87
	Chromatography: HILIC	94
	Chromatography: ERLIC	134
Localization	Centrifugation	3,43,135
	Protein correlation profiling	38,44
	Proximity-labelling	39
	Detergent solubility	4
Turnover	Metabolic pulse-labelling	3,5,6,55
	Cycloheximide treatment	4
Isoform expression	• High sequence coverage to identify isoform-specific peptides • Targeted mass spectrometry analysis may be used to detect isoform-specific peptides	136,137
Solubility	Thermal denaturation followed by differential centrifugation	138
Activity	Analogue-sensitive kinases	139
	Activity-dependent binding domains	140
Tertiary Structure	Protease sensitivity	141
	Crosslinking	77,78

ERLIC, electrostatic repulsion hydrophilic interaction chromatography; FACS, fluorescence-associated cell sorting; HILIC, hydrophilic interaction chromatography; IEX, ion-exchange chromatography; IMAC, immobilized metal affinity chromatography; iTRAQ, isobaric tags for relative and absolute quantification; LC-MS/MS, liquid chromatography followed by tandem mass spectrometry; SILAC, stable isotope labelling by amino acids in cell culture; TiO_2, titanium dioxide; TMT, tandem mass tag.

Larance M, Lamond AI.
Nat Rev Mol Cell Biol. 2015 May;16(5):269-80.

Experimental design considerations

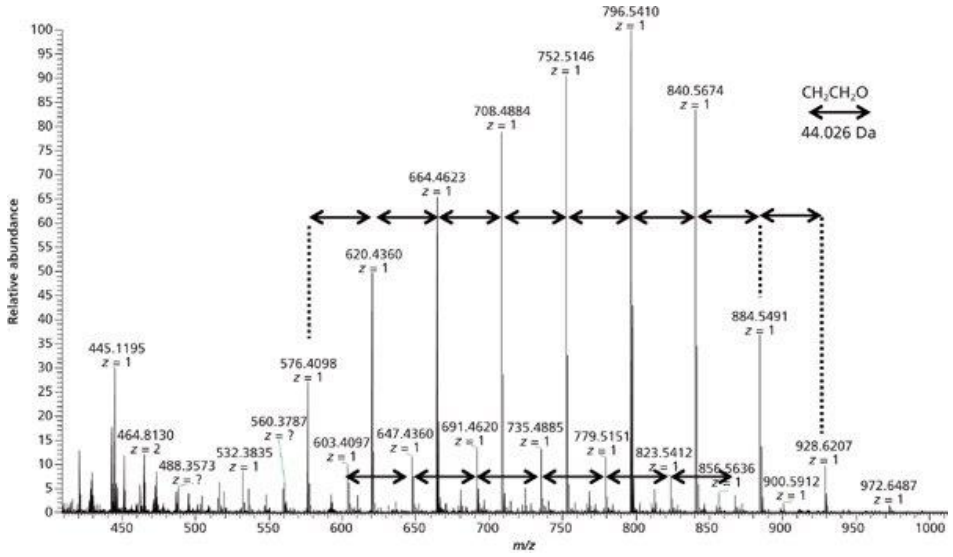
- Identification & Quantification
  - Sample heterogeneity
  - Database availability/completeness
  - Dynamic range of sample source
  - Loading considerations e.g. per cell vs. per protein amount; use ‘per unit of time’ for highly dynamic system such as urine
  - Relative vs. absolute quantification
- Types of control
  - Time control
  - Vehicle control
  - KO control
  - Physical control e.g. pressure, flow rate, temperature

Experimental design considerations

- Normalization
  - Based on assumptions e.g. expecting highly skewed data from AP-MS vs. control bead
- Sample size
  - ‘missing value’ problem
  - Statistical power
- Replication
  - Biological vs. technical
- Statistical analysis
  - Identification level
  - Quantification level
  - Biological interpretation level

What Should We Avoid in Proteomics Sample Preparation

- Detergents
  - Suppress ionization of peptides
  - Bind to the column
- Polymers, PEG, plasticizer
- Involatile solvents
- Salts and buffers (small molecules)
  - Can be removed
- Keratins
  - Human skin
  - Wool
  - Always wear gloves
  - Use a laminar hood



PEG contamination

Recent advances in Medical Proteomics

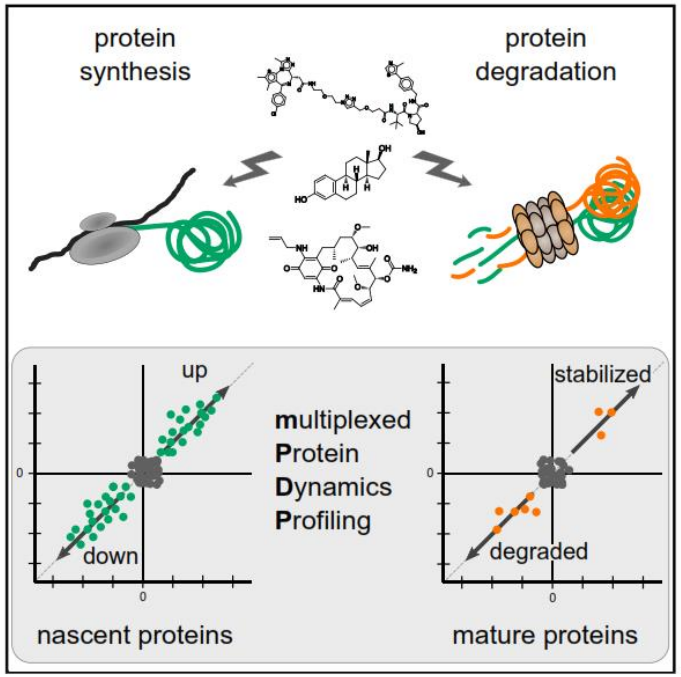
Dynamic SILAC with chemical labeling using neutron-encoded tandem mass tags (TMT)

Resource

Cell

Multiplexed Proteome Dynamics Profiling Reveals Mechanisms Controlling Protein Homeostasis

Graphical Abstract



Authors

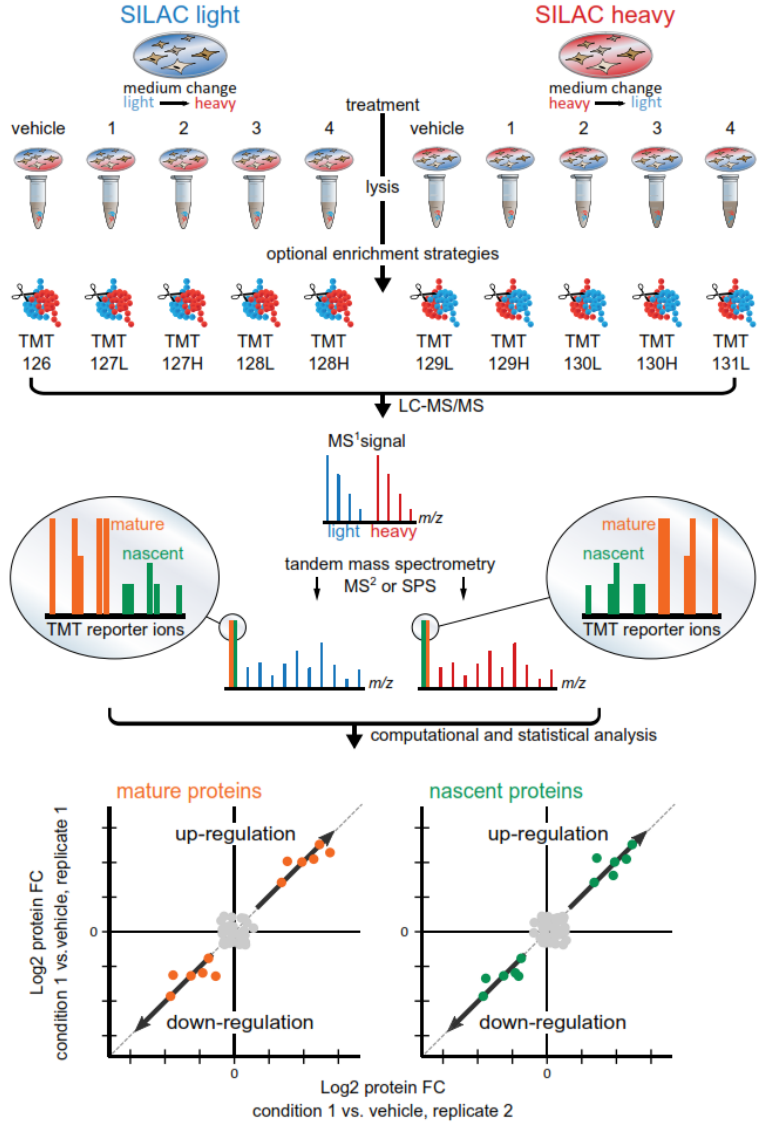
Mikhail M. Savitski, Nico Zinn, Maria Faelth-Savitski, ..., Paola Grandi, Giovanna Bergamini, Marcus Bantscheff

Correspondence

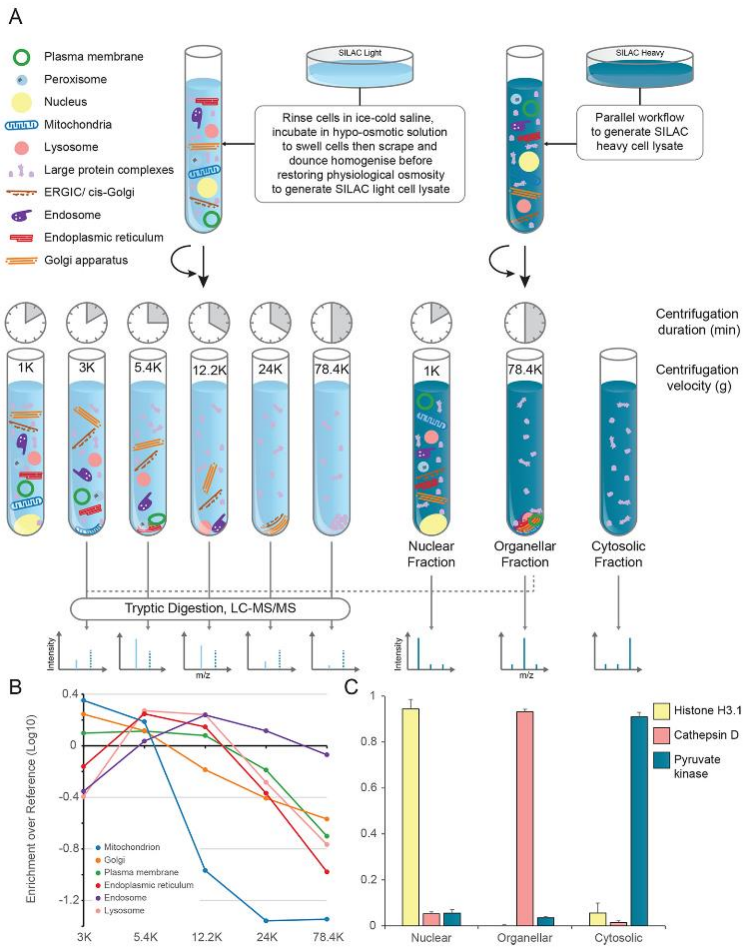
mikhail.savitski@embl.de (M.M.S.), giovanna.2.bergamini@gsk.com (G.B.), marcus.x.bantscheff@gsk.com (M.B.)

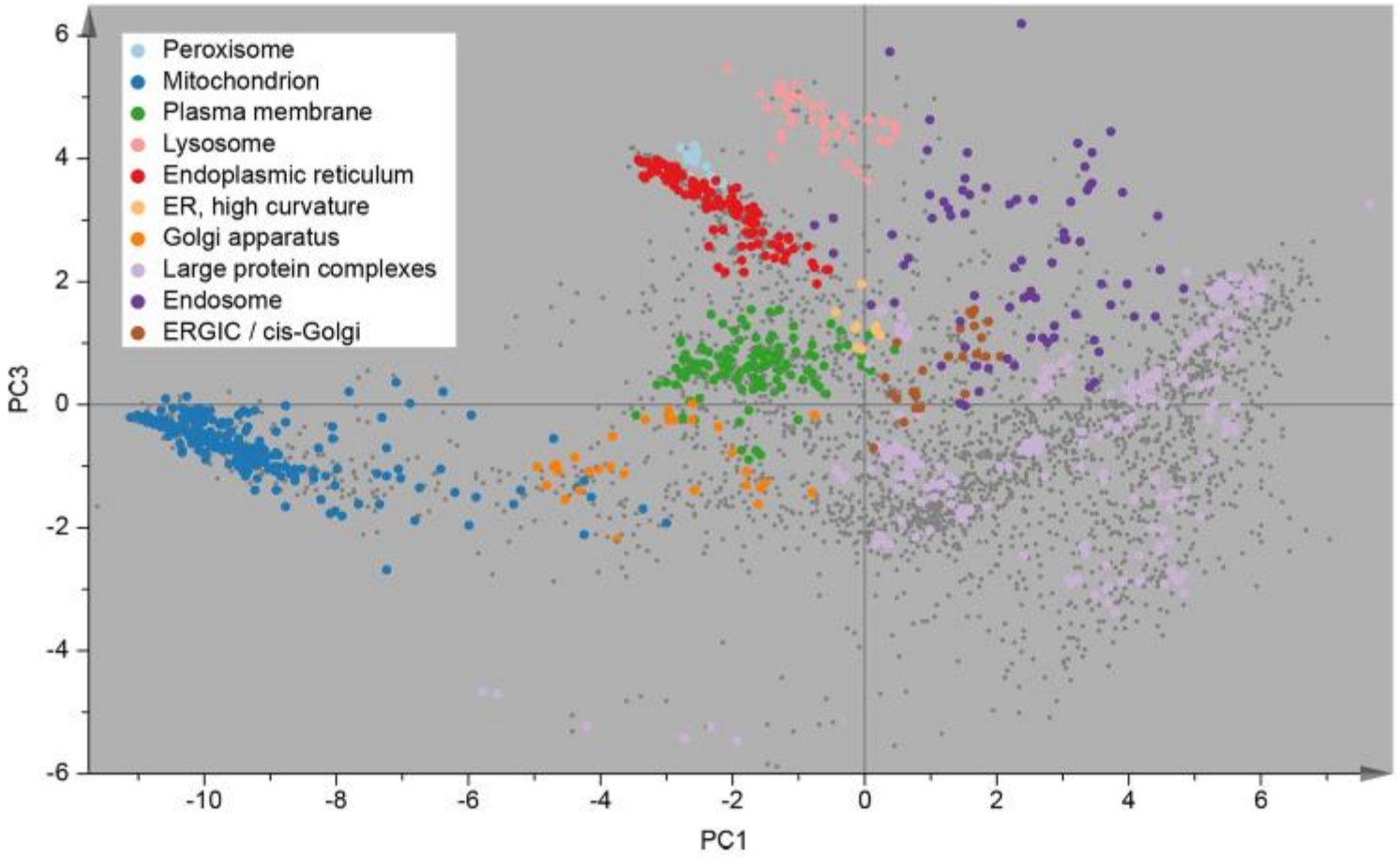
In Brief

Tracking both protein synthesis and degradation across thousands of proteins yields insights into functional regulation by protein degradation.



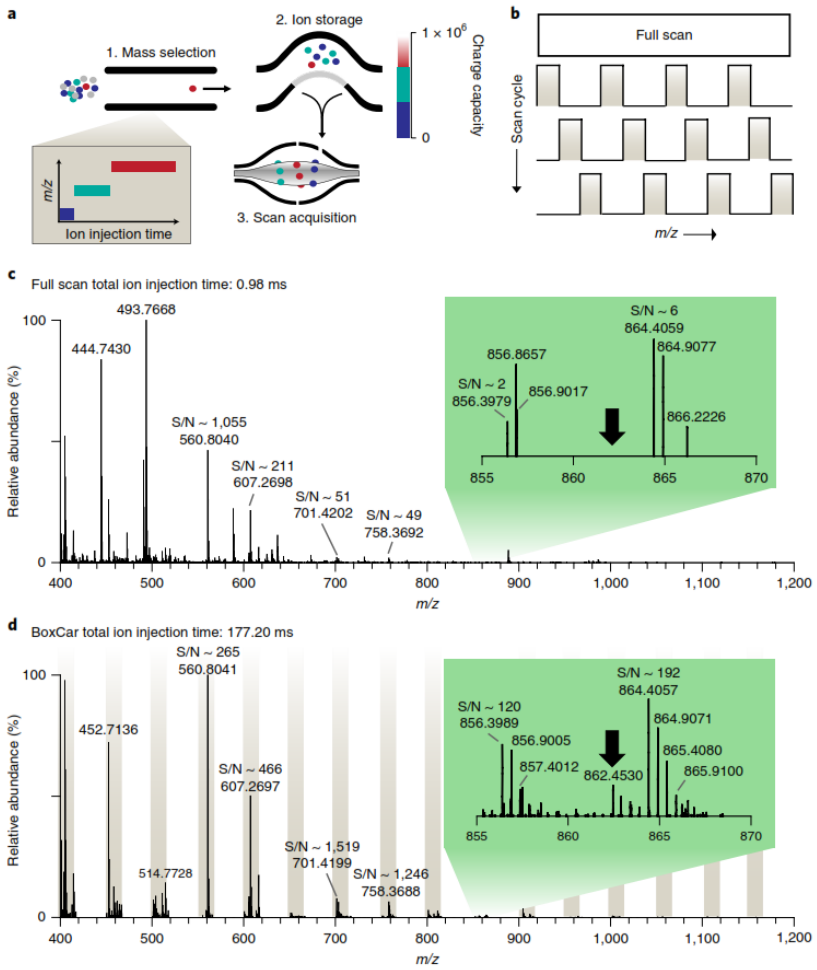
Global, quantitative and dynamic mapping of protein subcellular localization

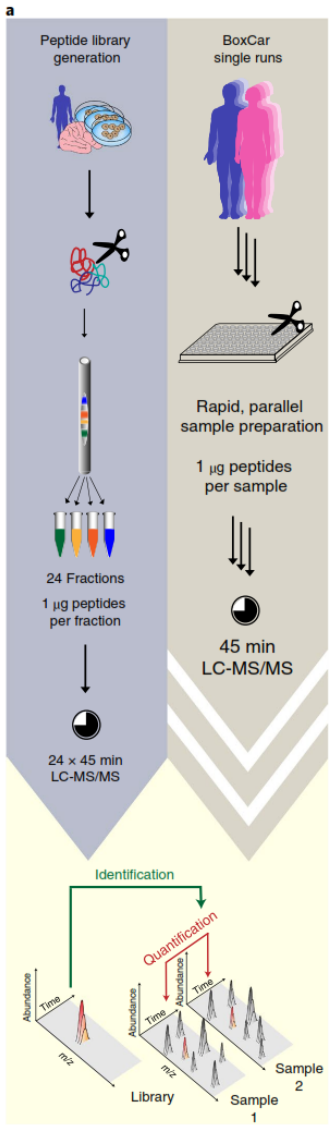


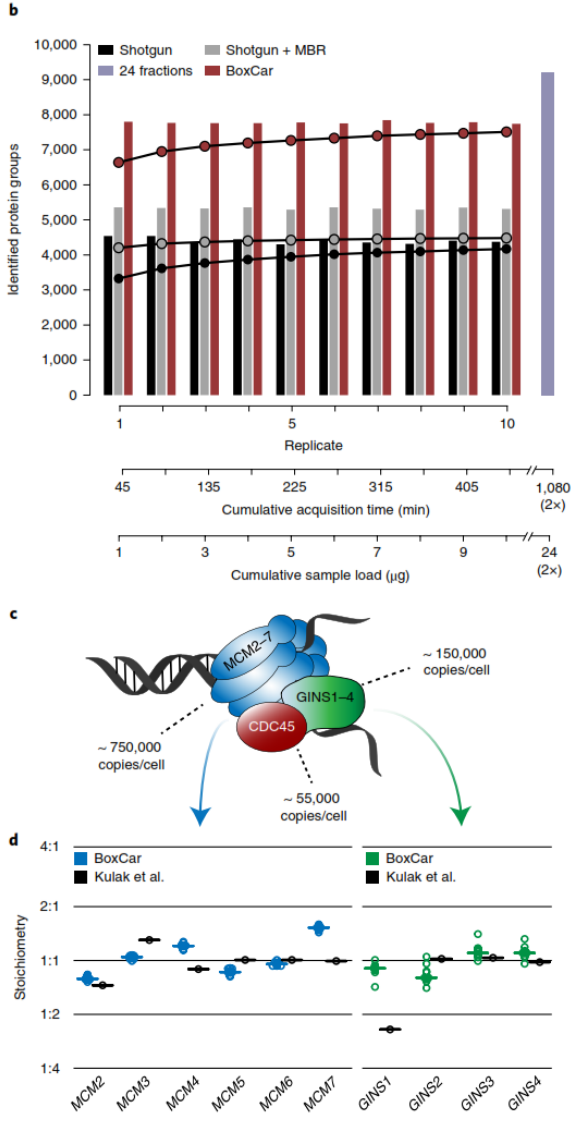


Itzhak DN, Tyanova S, Cox J, Borner GH.
Elife. 2016 Jun 9;5.

BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes



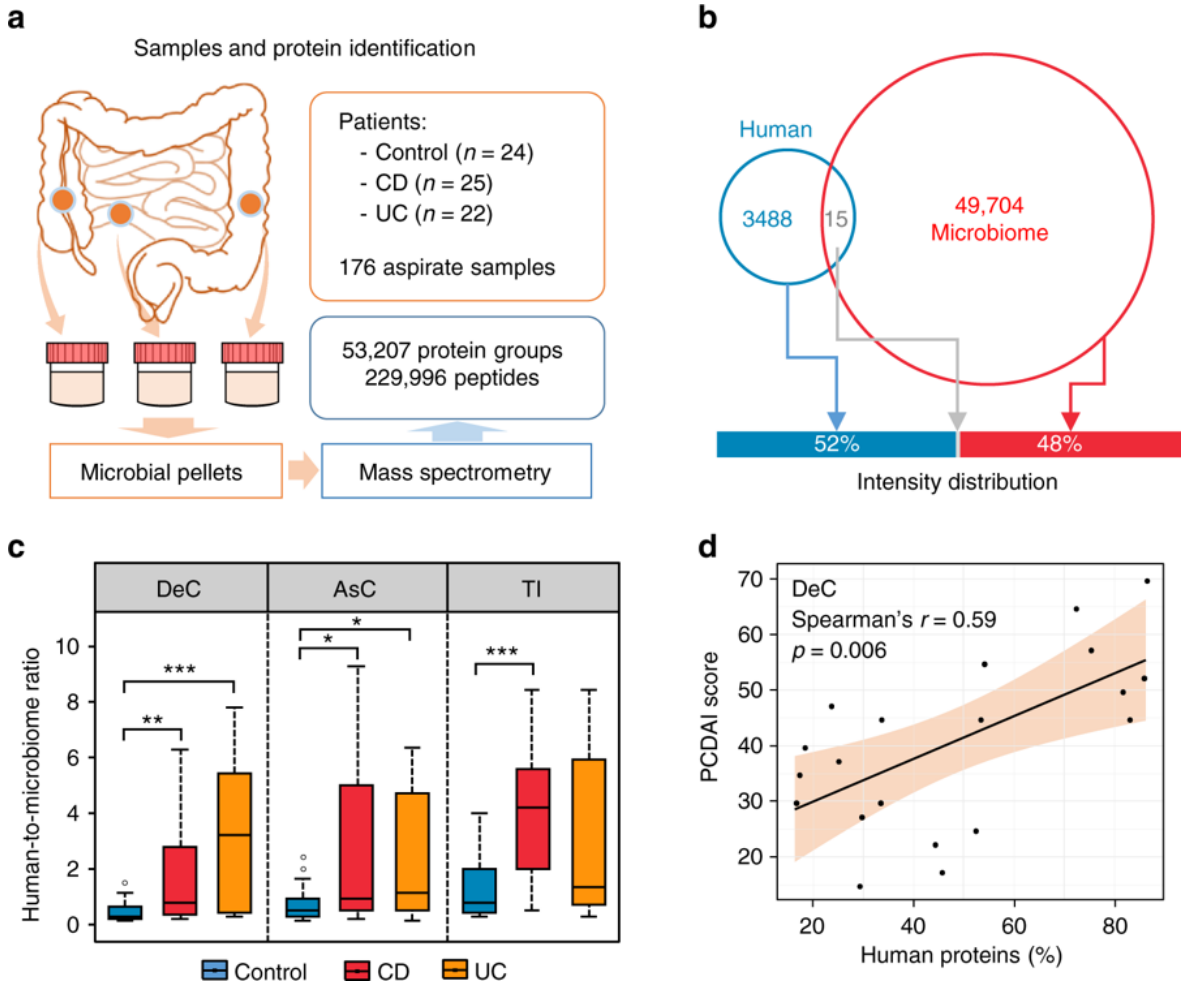




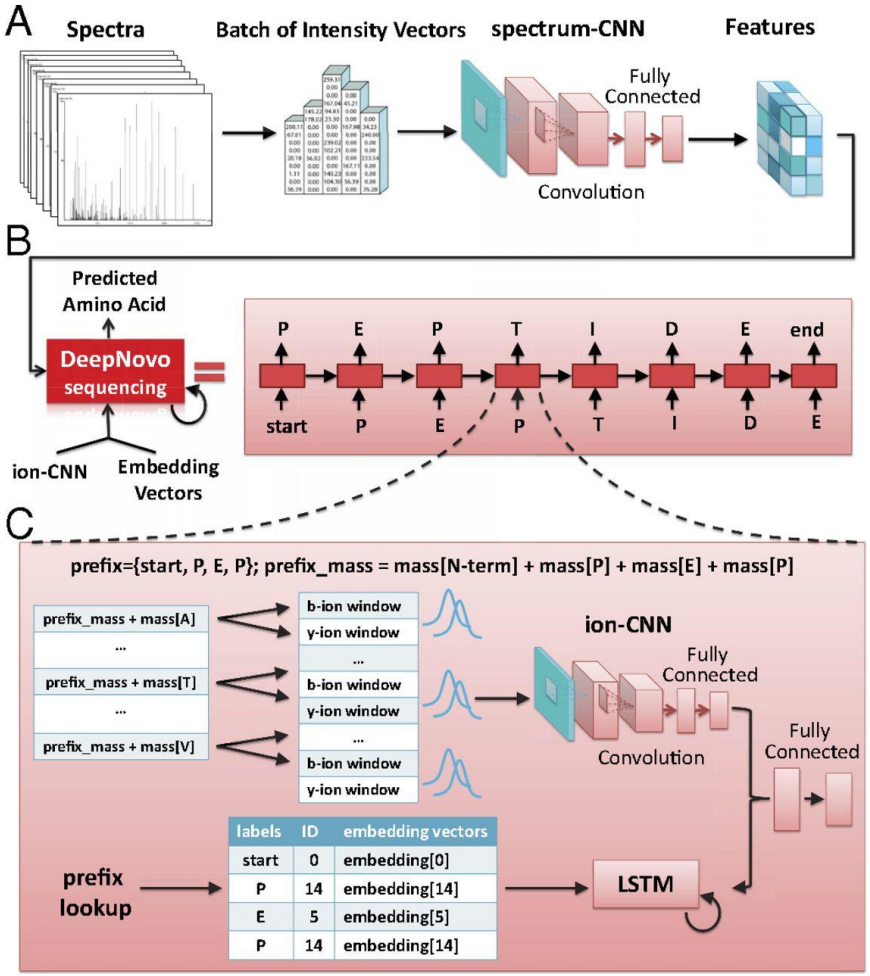
Meier F, Geyer PE, Virreira Winter S, Cox J, Mann M.
Nat Methods. 2018 May 7. doi: 10.1038/s41592-018-0003-5.

Metaproteomics

- Proteome of multi-organism, complex communities
- Gut microbiome

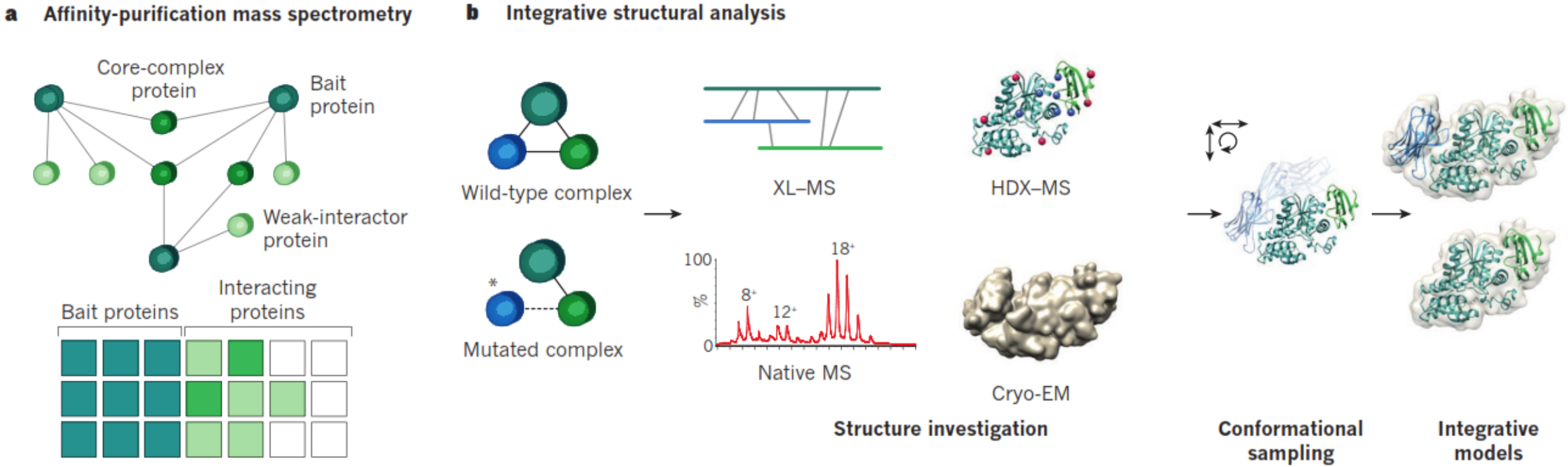


De novo peptide sequencing by deep learning



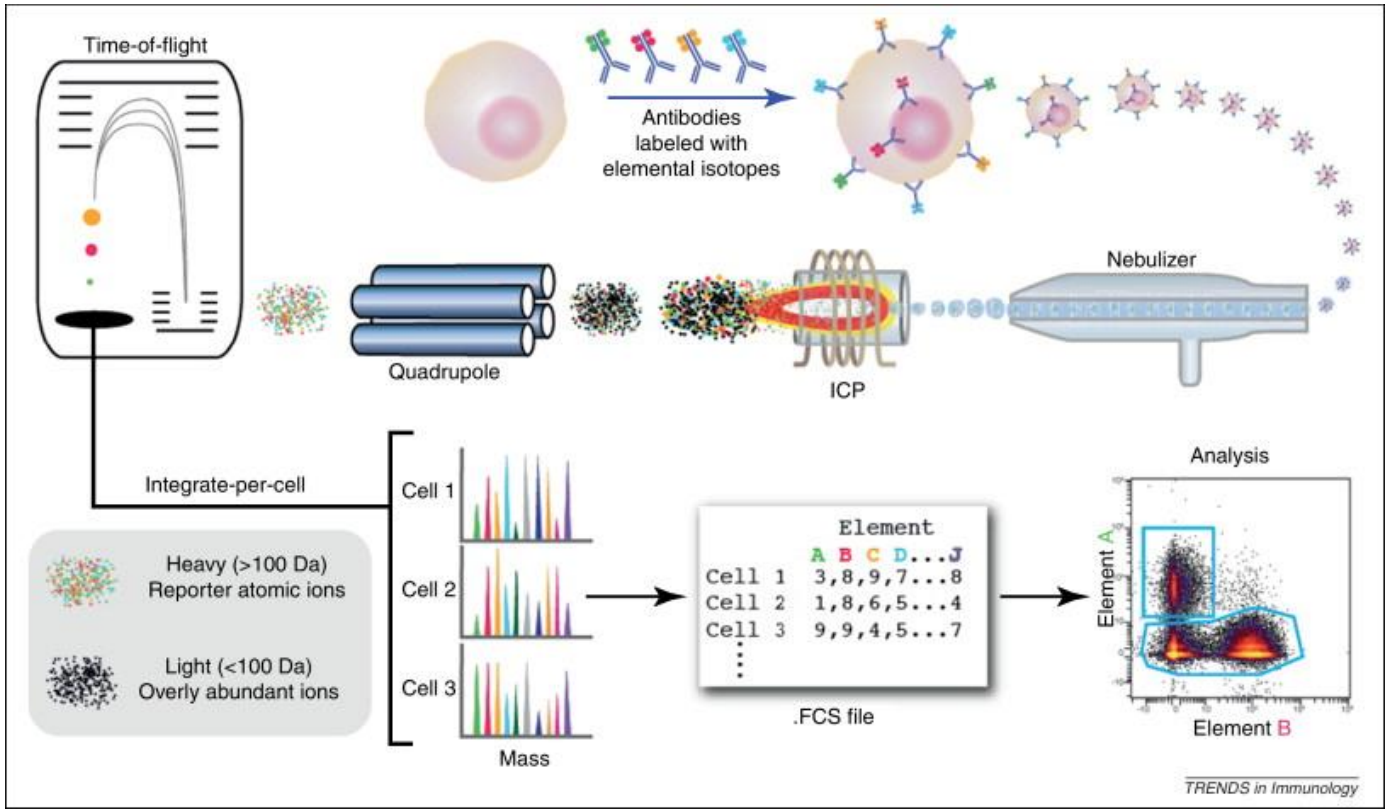
Tran NH, Zhang X, Xin L, Shan B, Li M.
Proc Natl Acad Sci U S A. 2017 Jul 18.

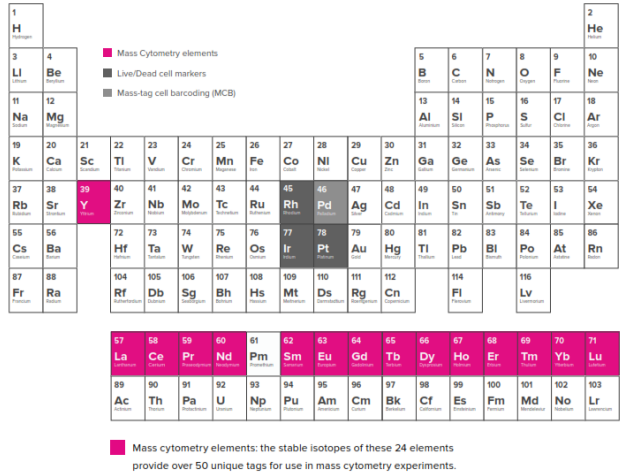
Interaction Proteomics and Structural Proteomics



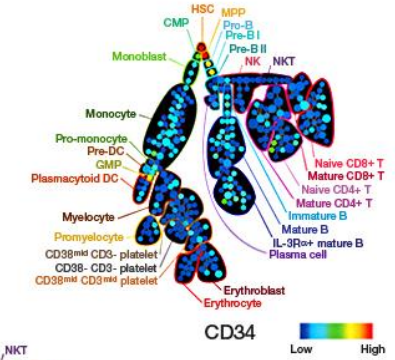
Aebersold R, Mann M.
Nature. 2016 Sep 14;537(7620):347-55.

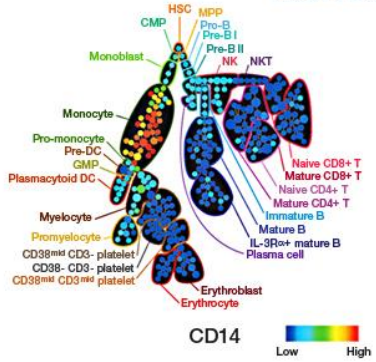
Single-cell Analysis by Mass Cytometry





CyTOF data analyzed using the SPADE algorithm to organize cells into hierarchies of related phenotypes

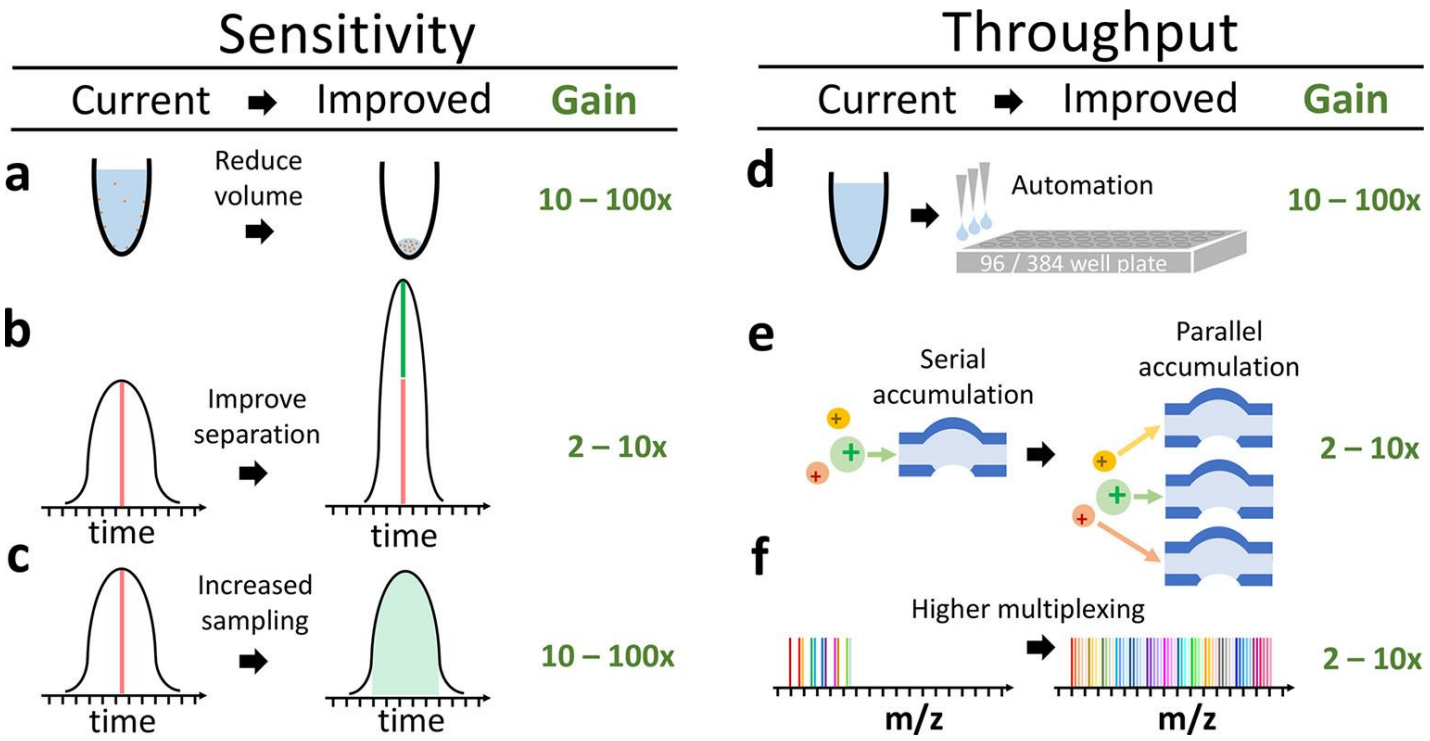




SPADE tree annotated with CD34 and CD14 expression in healthy human bone marrow cells

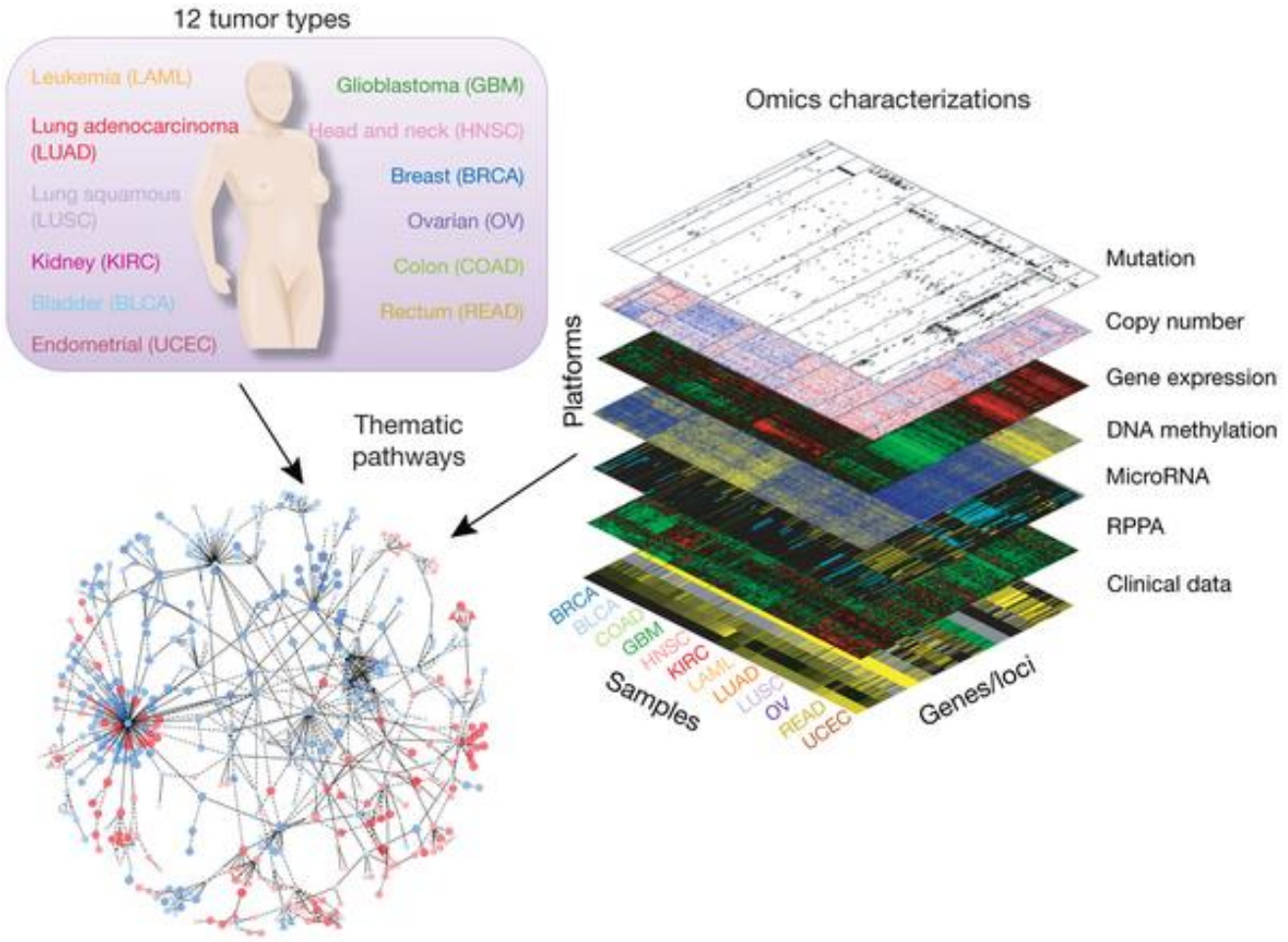
Single Cell Proteomics – Are We There Yet?

- Typical experiments need >10,000 cells for complete proteome
- Only top abundant proteins were identified from a single mammalian cell
- Alternatives?



Specht H, Slavov, N, J Proteom Res, 2018

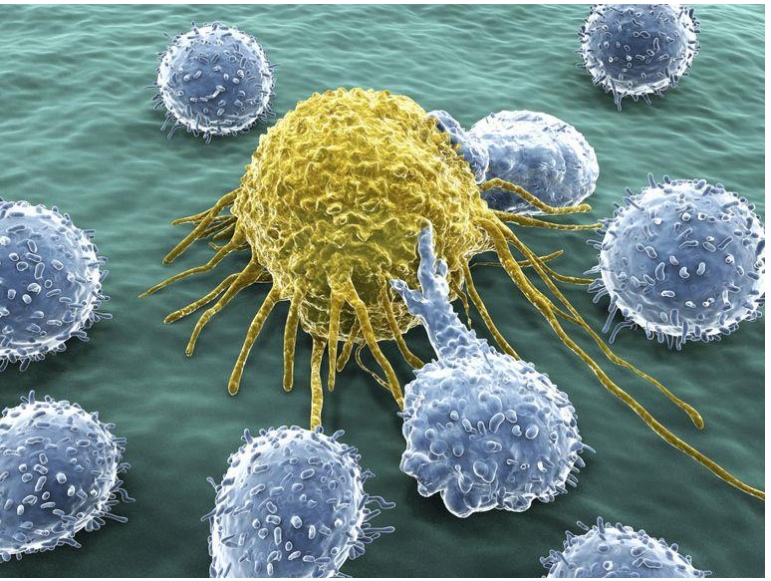
Integrated data set of multi-omics analyses



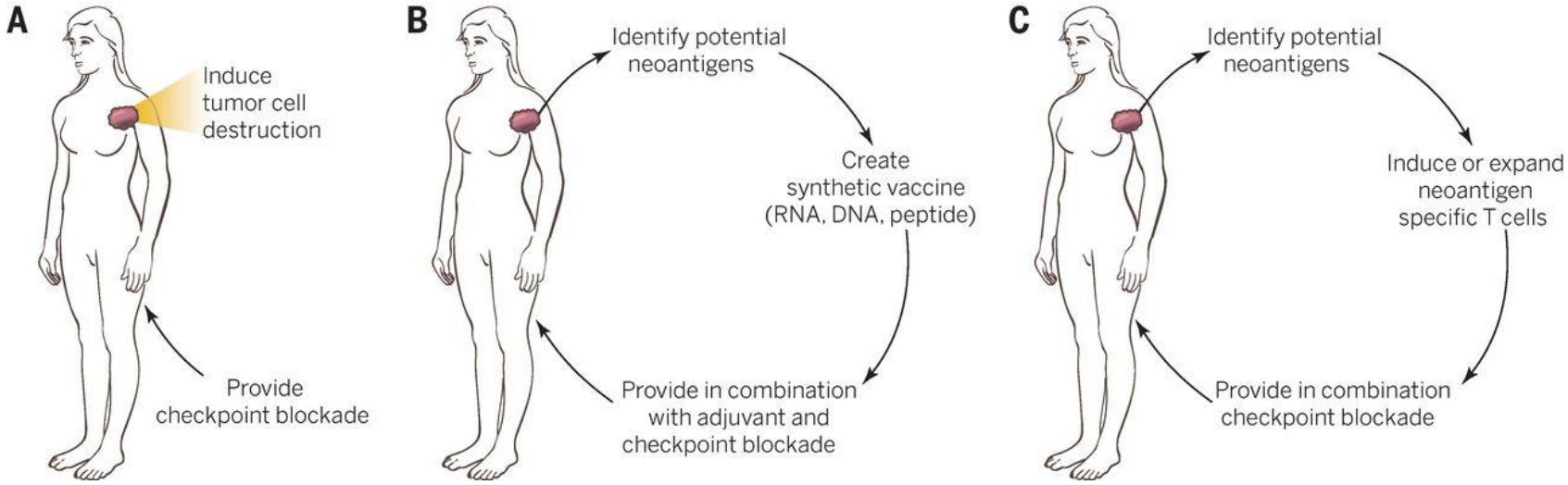
Proteomics in Cancer Immunotherapy

Precision Medicine

Checkpoint Inhibitor

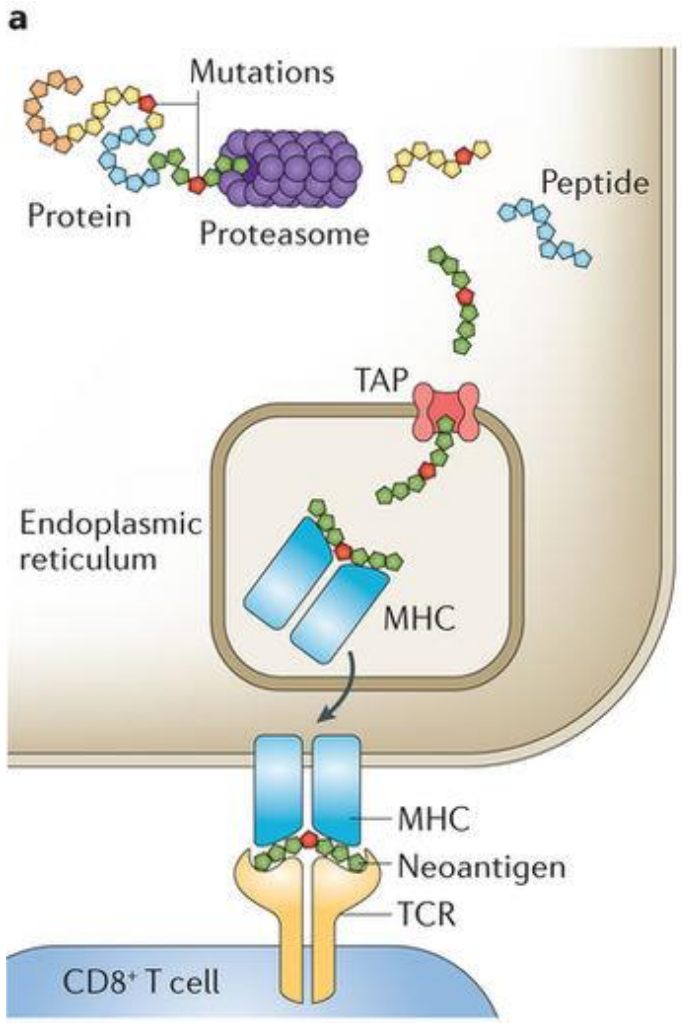


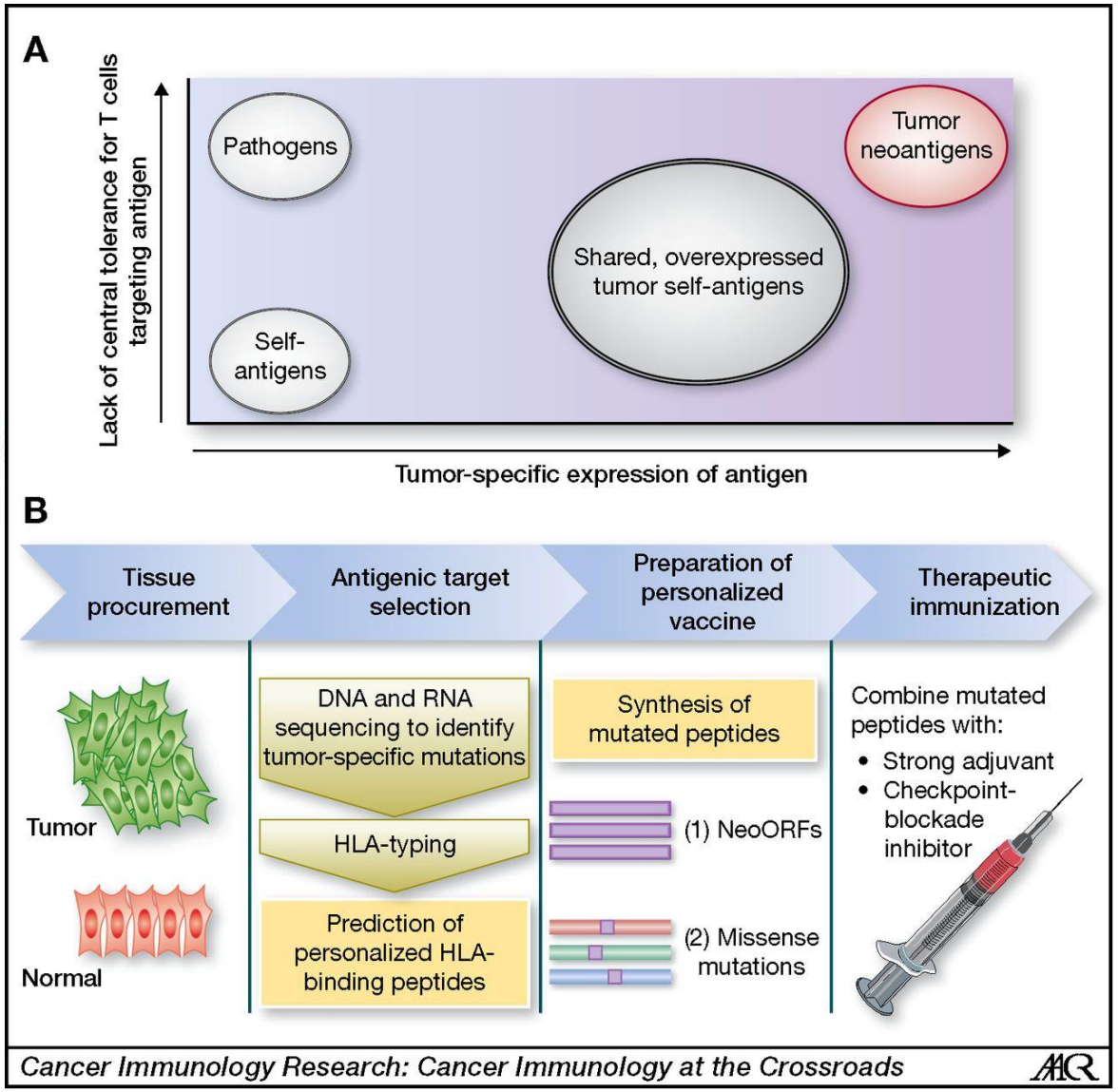
CAR T-cell



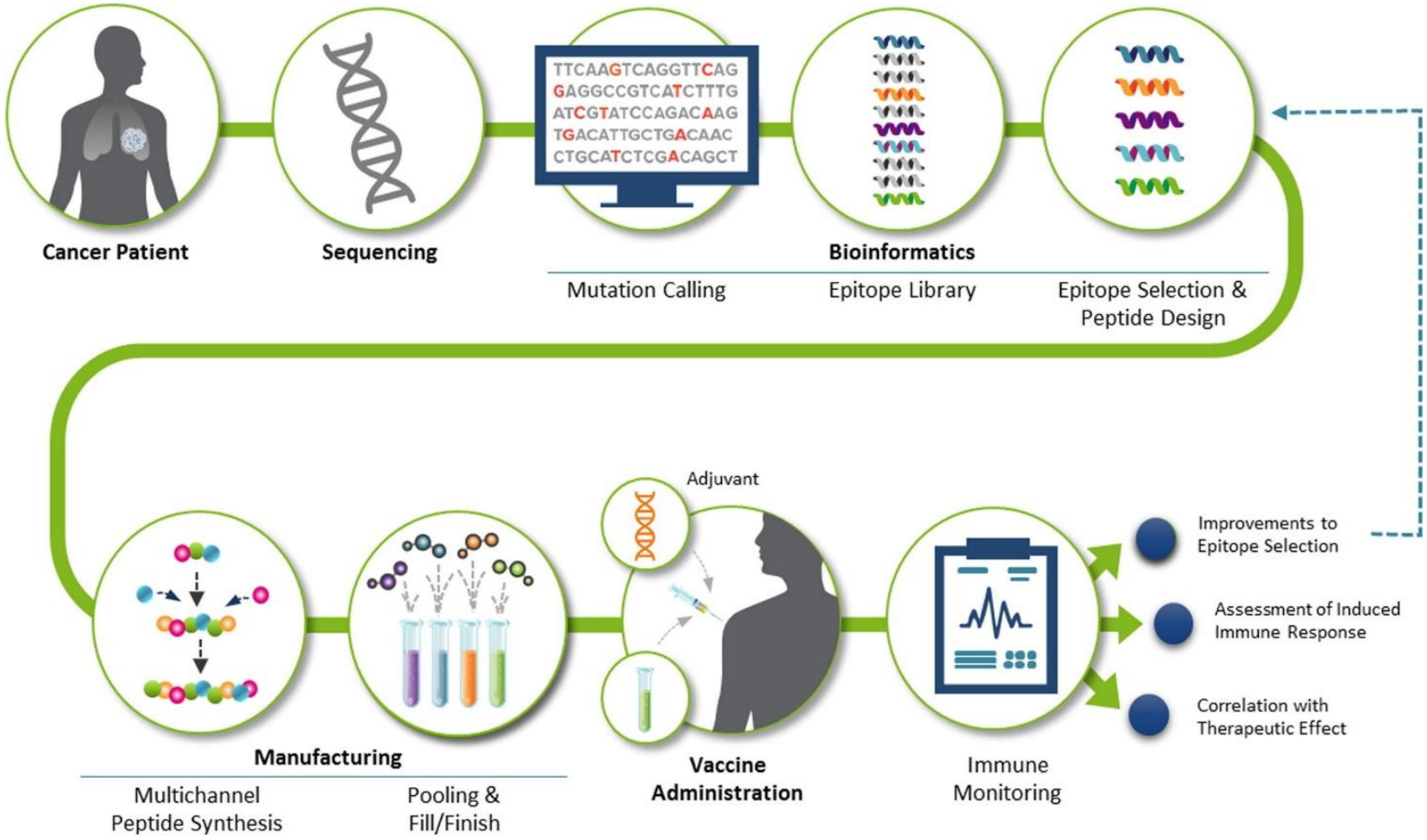
Cancer Vaccine

What Are Neoantigens?

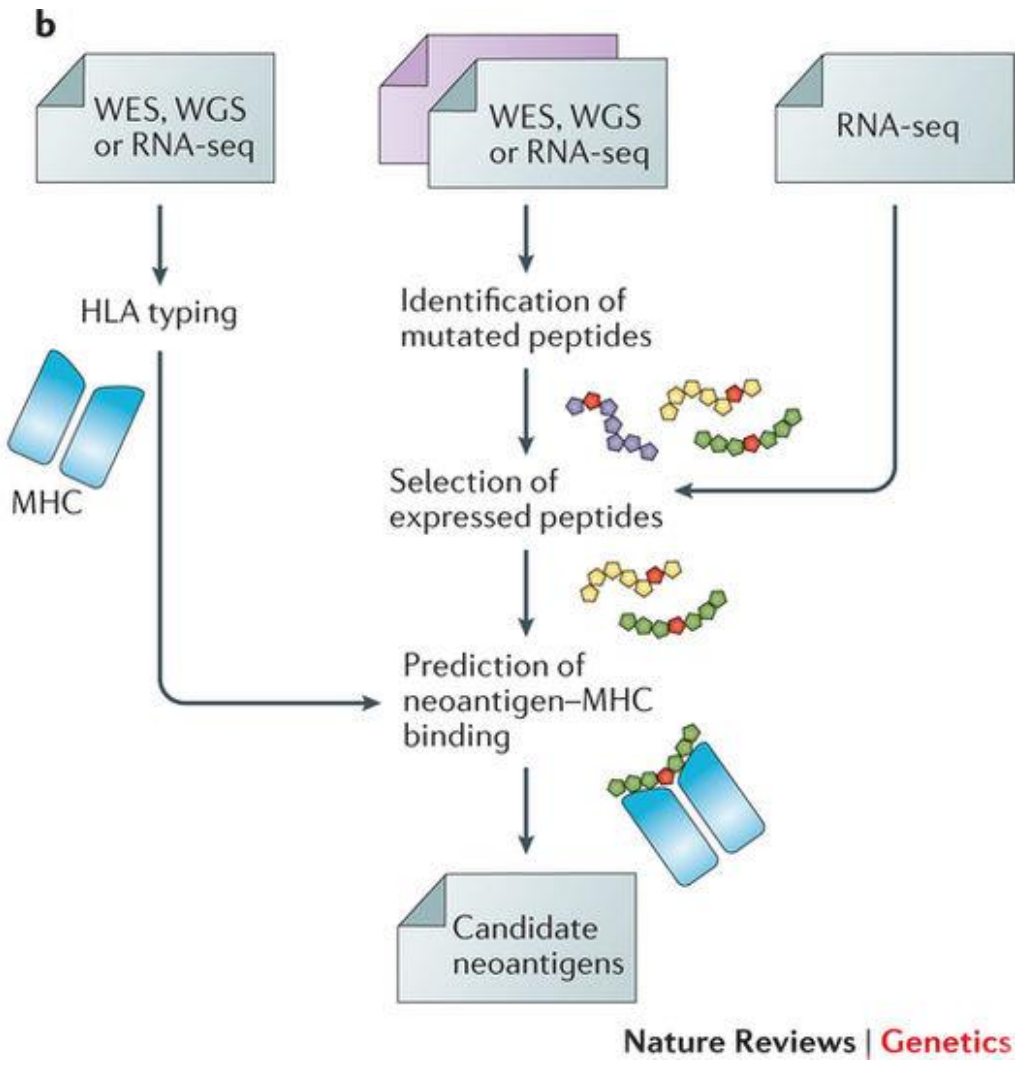




Computational prediction and genomic sequencing approach tumor neoantigen identification

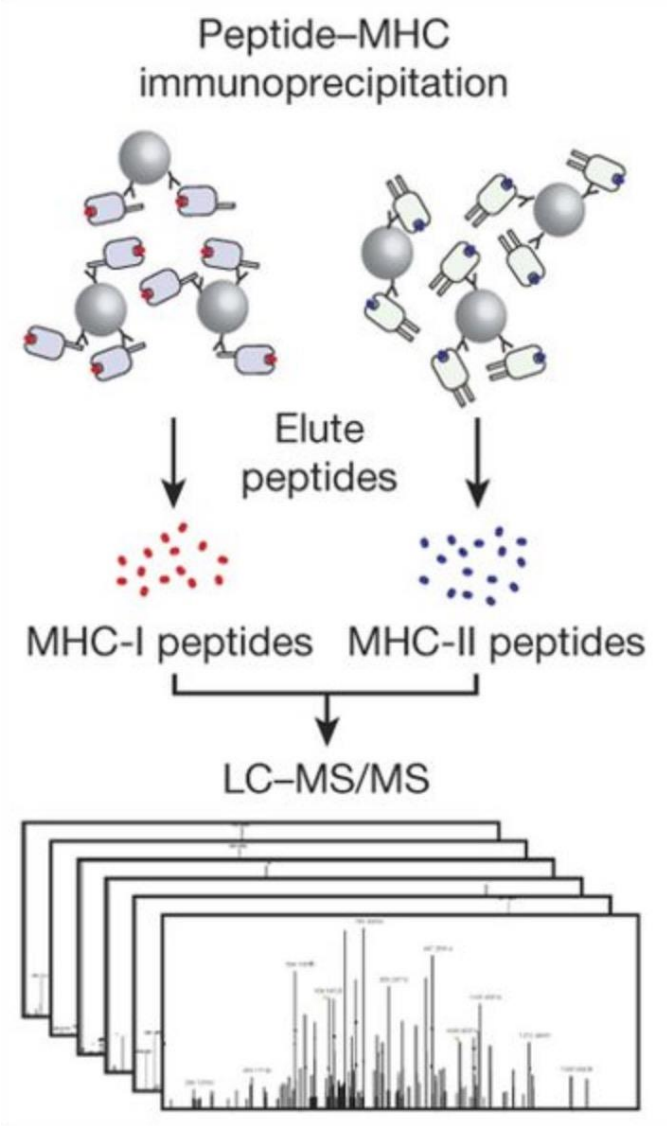


Neoantigen Prediction

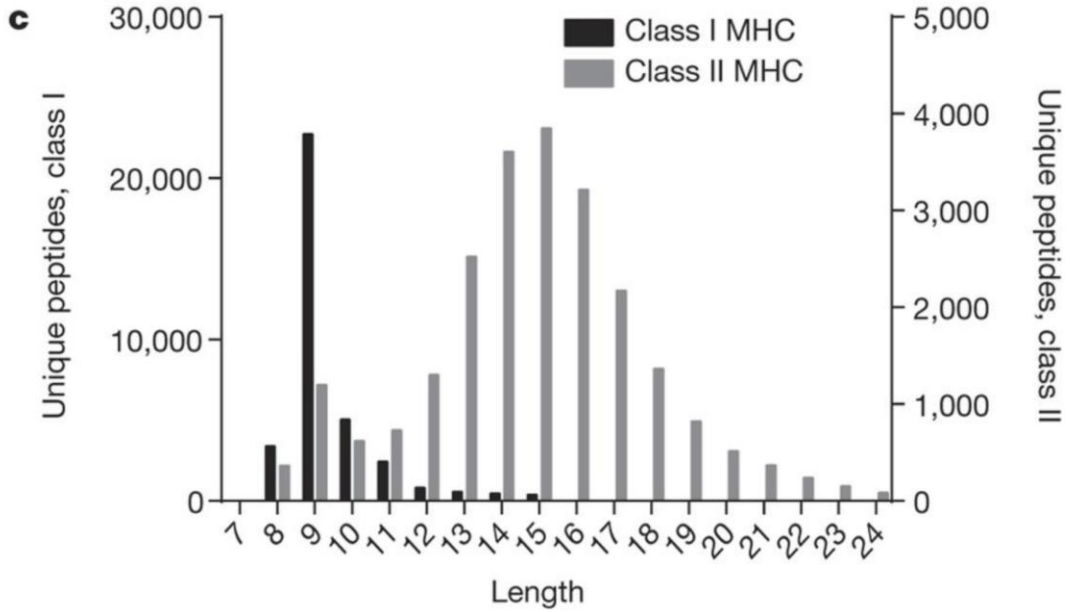


- Somatic mutation calling
- RNA expression
- HLA typing
- Binding prediction by NetMHC and MuPeXI
- Deep learning
- List of candidates

Direct Identification of Neoantigens



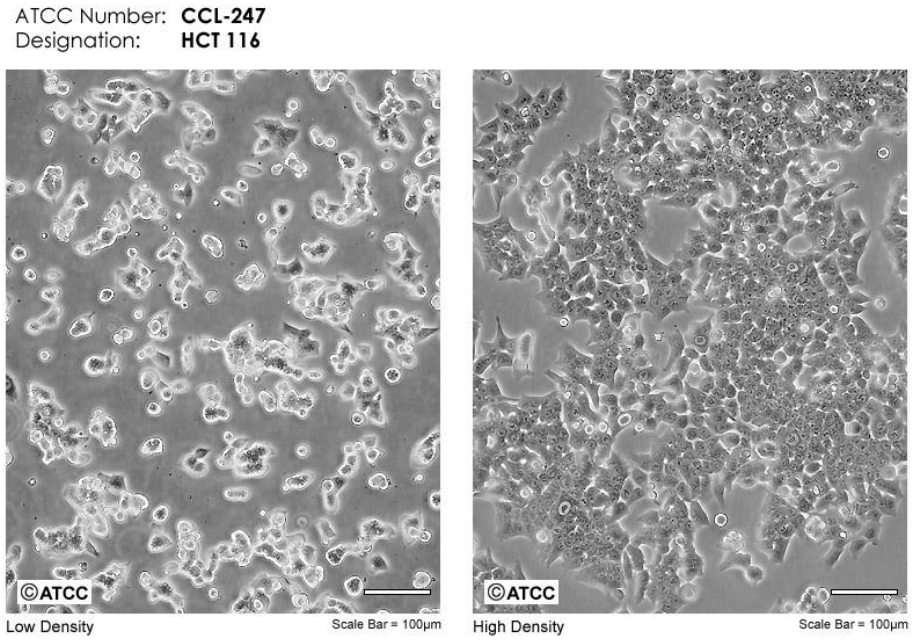
- Immunoprecipitation
  - MHC Class I: W6/32
  - MHC Class II: L243, IVA12
- Mass spectrometry
- Immunopeptidomics



Khodadoust MS, et al. Nature 2017

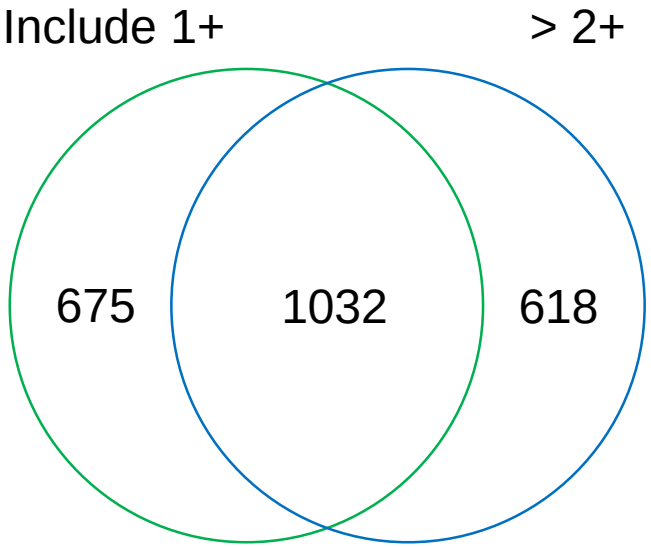
Cell Line Model: HCT116

- Human colorectal carcinoma cell line
- Starting with 10^8 cells, approximately 10 plates, >90% confluence
- HLA genotyping
  - A*01:01, A*02:01
  - B*45:01, B*18:01
  - C*05:01, C*07:01



HLA Peptidomes

- 2325 peptides were identified at 1% FDR
  - Corresponding to 1725 proteins
- 8 phosphopeptides
- 4 neoantigen peptides



CUSB

Gene	Peptide	Mutation
CHMP7	QTDQMVFNTY	A324T
RBBP7	EERVIDEEY	N17D
RNPEP	ALFEVPDGFTA	I195F
UQCRB	EEEKFYLEP	N88K

Mann MCP 2015

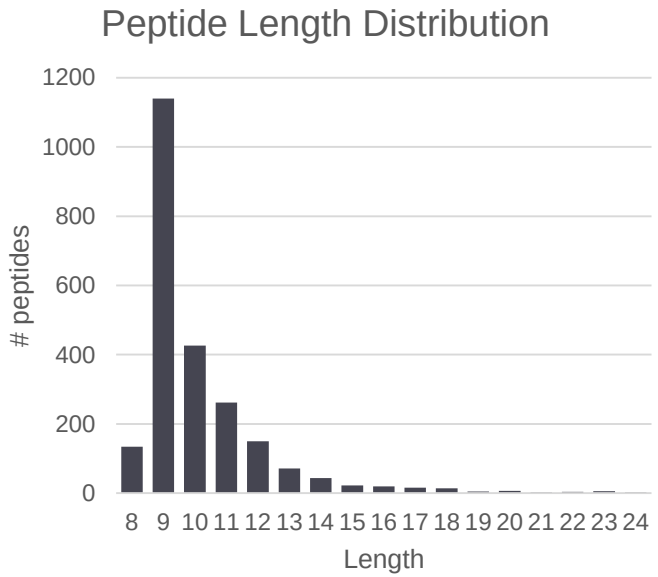
TABLE II

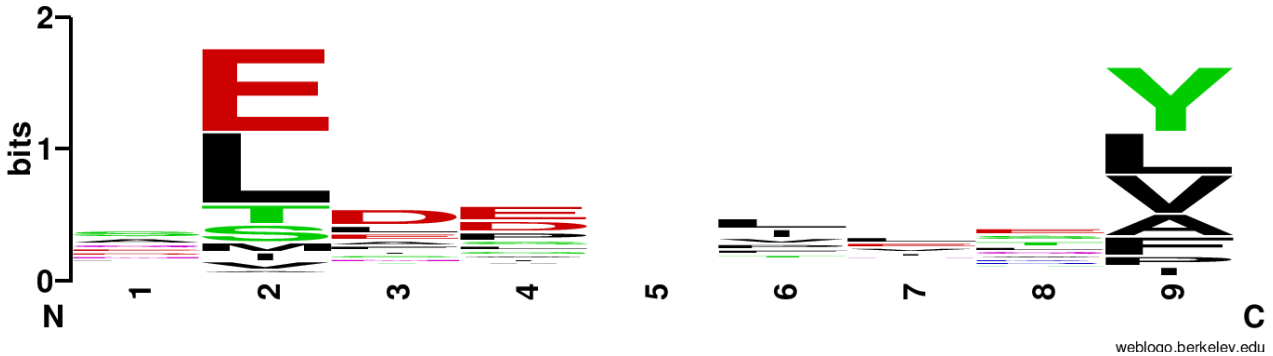
List of MS-identified mutated HLA-I peptides purified from HCT116 cell line. Potential mutation bearing peptide sequences were obtained from (60) and they were added to the UniProt database for MaxQuant search. The mutated positions are marked in bold

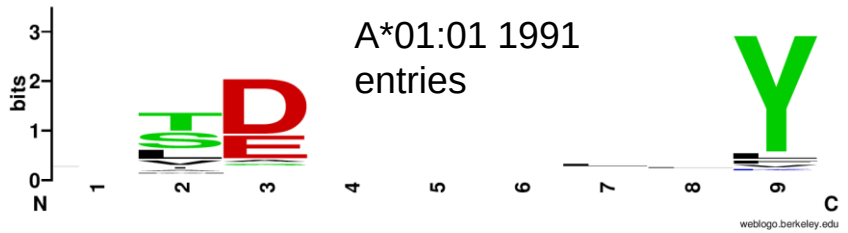
Protein	Peptide	AA change	Length	Identification score	HLA Allele best fit (NetMHC 3.4 predicted affinity, K_d values in nM)
CHMP7	QTDQMVFN**T**Y	p.A324T	10	166	A*01:01 (39), Cw*05:01 (23)
BCL2L13	EEEYPG**V**TA	p.I216V	9	148	B*45:01 (20)
NR1D1	YSDNSN**D**SF	p.G39D	9	134	A*01:01 (36), Cw*05:01 (2)
RBBP7	EERVI**D**EEY	p.N17D	9	116	B*18:01 (149)
UQCRB	EEE**K**FYLEP	p.N88K	9	117	B*45:01 (54)

Comparing Motifs with IEDB

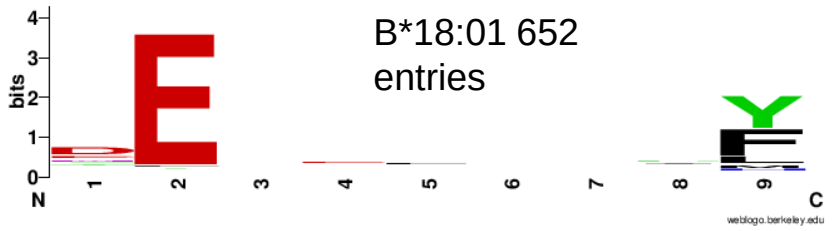
HLA-IP
9-AA
1140 entries



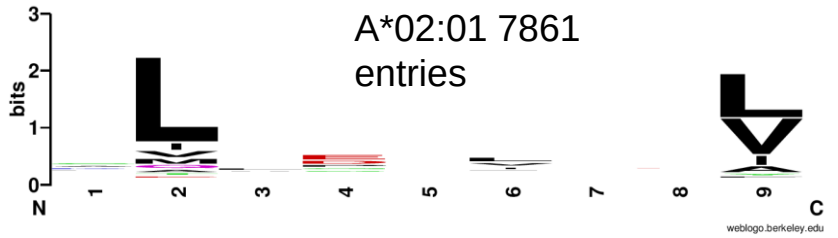




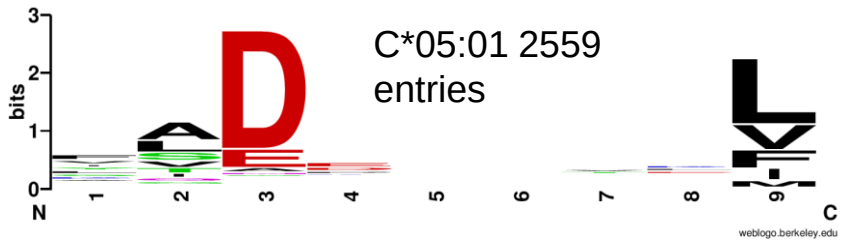
A*01:01 1991 entries



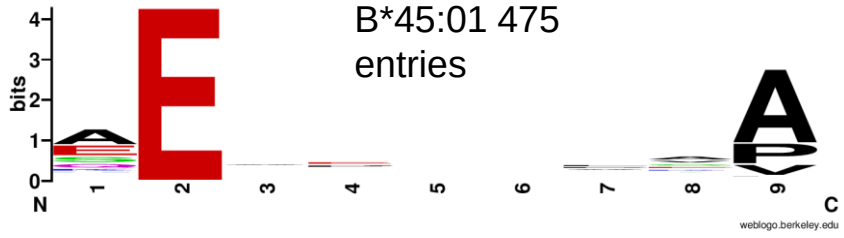
B*18:01 652 entries



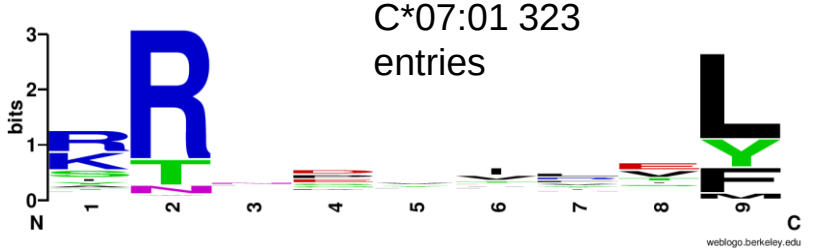
A*02:01 7861 entries



C*05:01 2559 entries

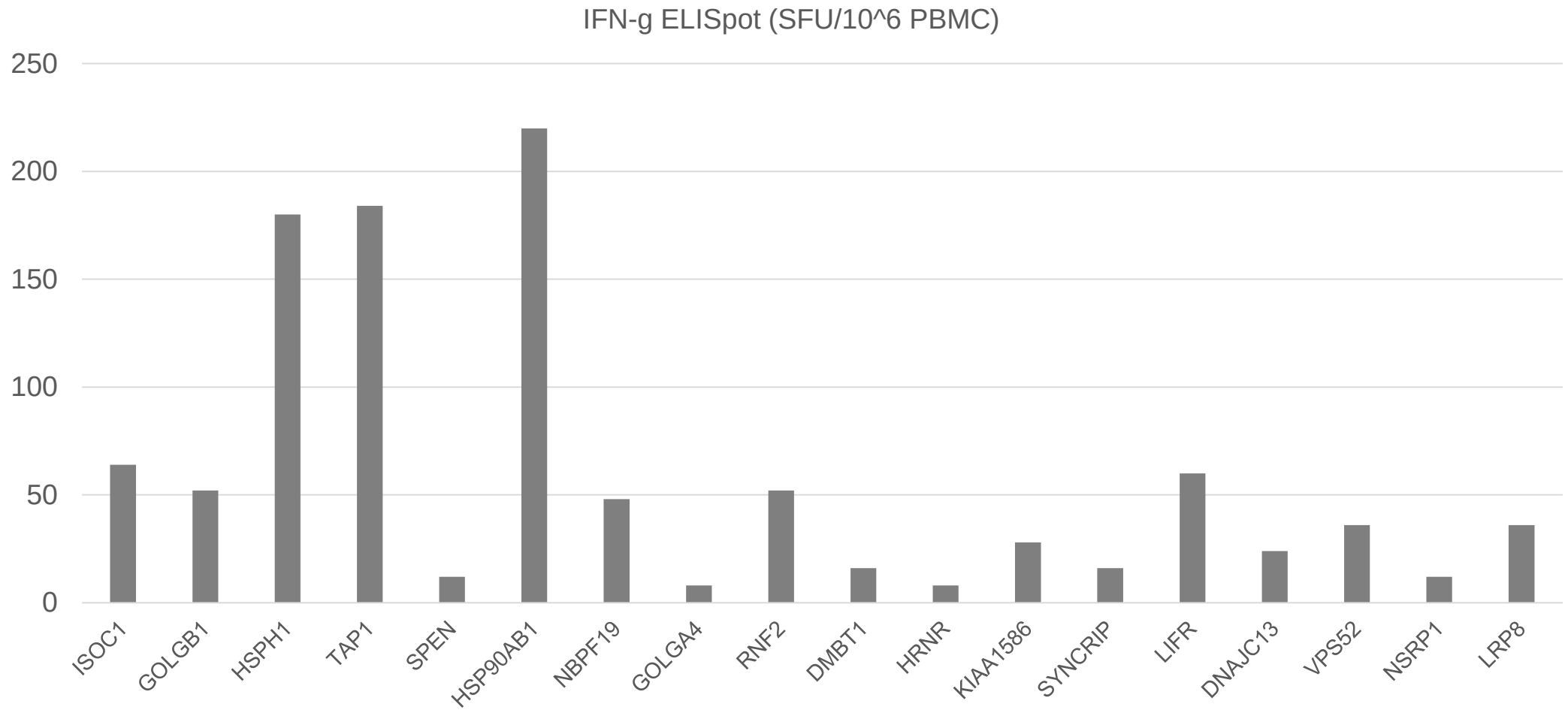


B*45:01 475 entries



C*07:01 323 entries

In Vitro Testing - ELISpot



Summary

- Proteomics is not just an identification tool
- Multi-dimensional proteomics provide more in-depth information about cells, disease stages, multicellular organisms
- Quantitative studies need to be carefully planned and appropriately controlled
- Bioinformatics is an important step
- Multi-omics data are important for precision medicine and cancer immunotherapy

Chulalongkorn University
Systems Biology Center
ศูนย์ชีววิทยาเชิงระบบ คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

More info:
http://sysbio.chula.ac.th/
https://www.facebook.com/CUSBcenter/