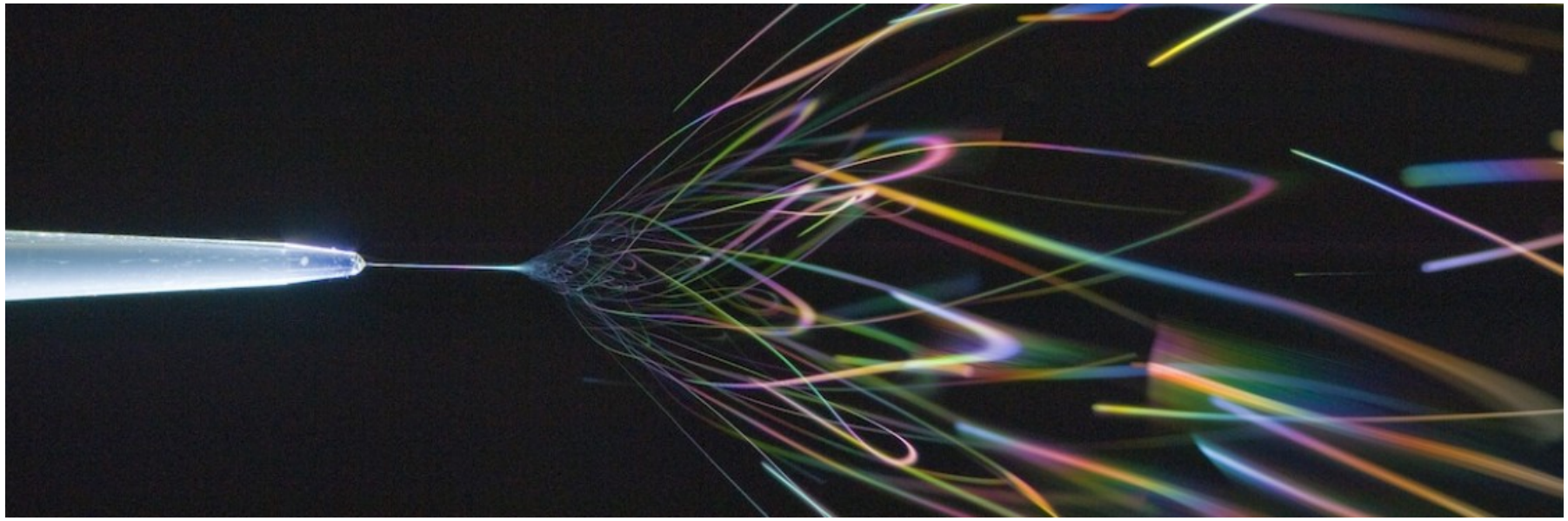
Mass Spectrometry and Proteomics

Andreas Schlosser

Rudolf Virchow Center
University of Würzburg



LC-MS-based Proteomics

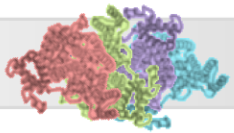
nanoHPLC

Reversed phase (C-18)

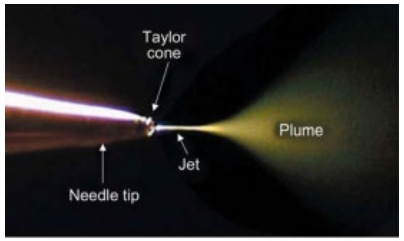
proteins

peptides

Bottom-up proteomics

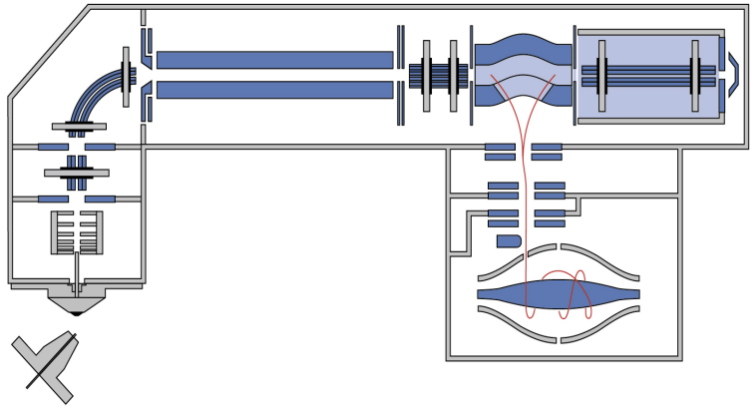


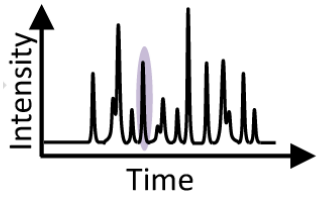
Digestion



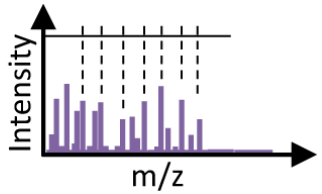
ESI

MS





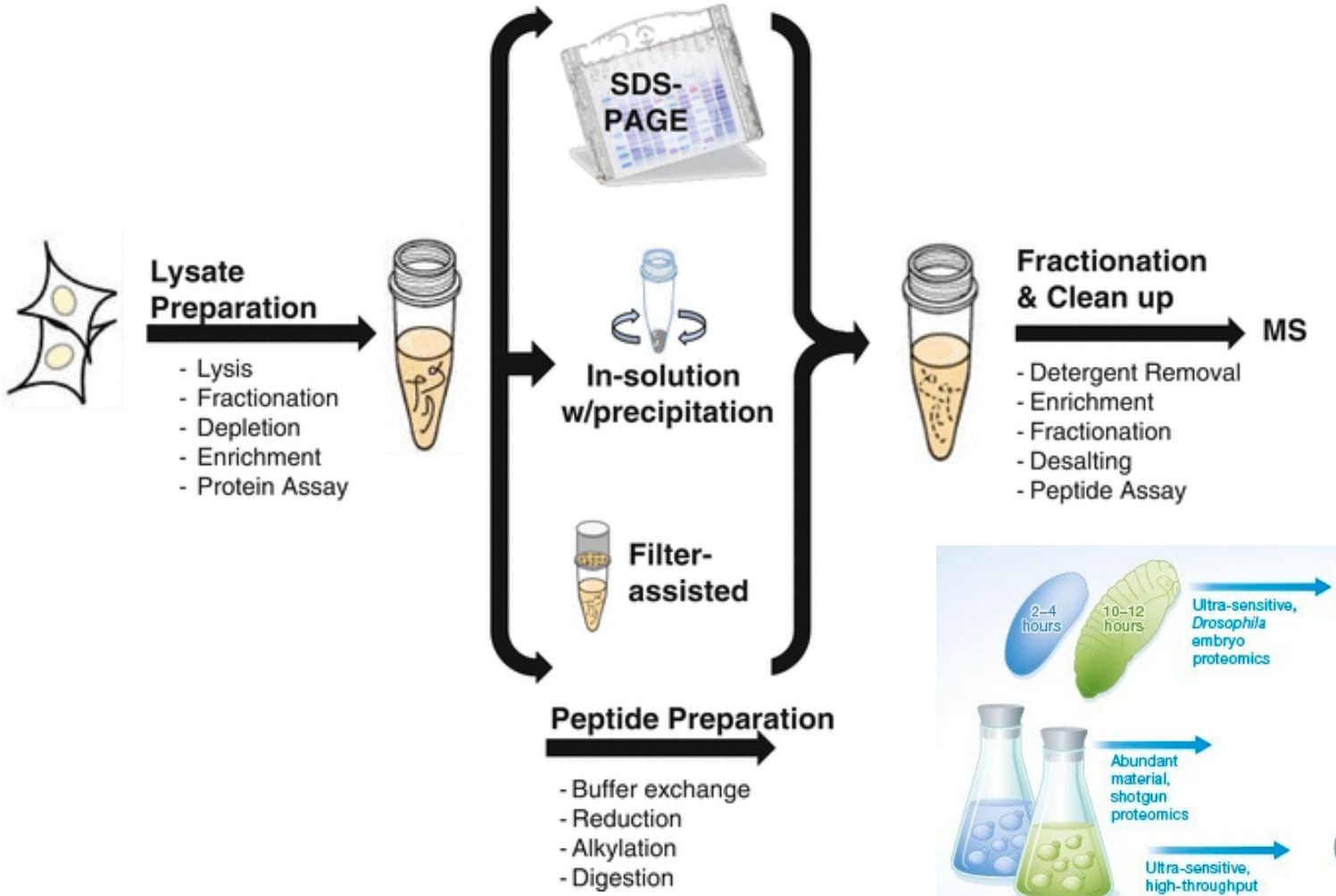
LC-MS

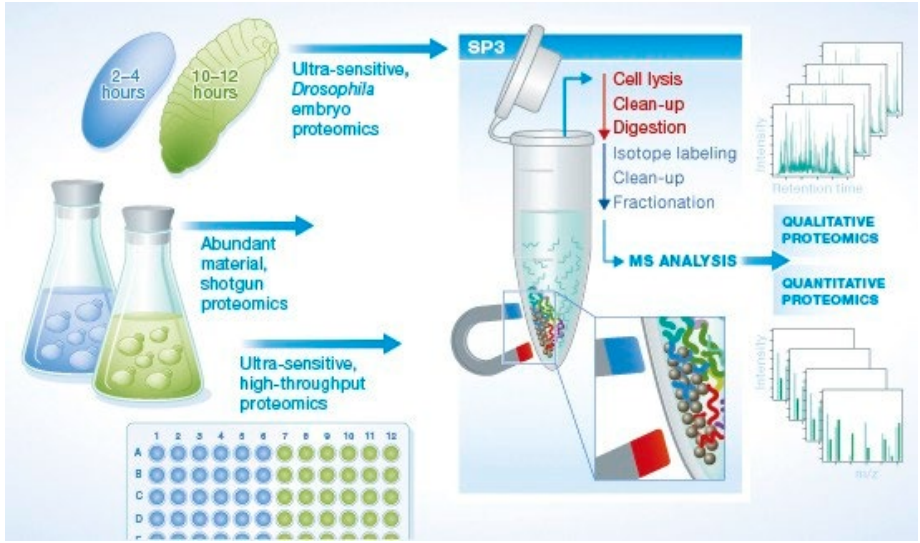


MS/MS

adapted from J. Proteome Res. 2013, 12, 1067−1077

Sample Preparation Overview





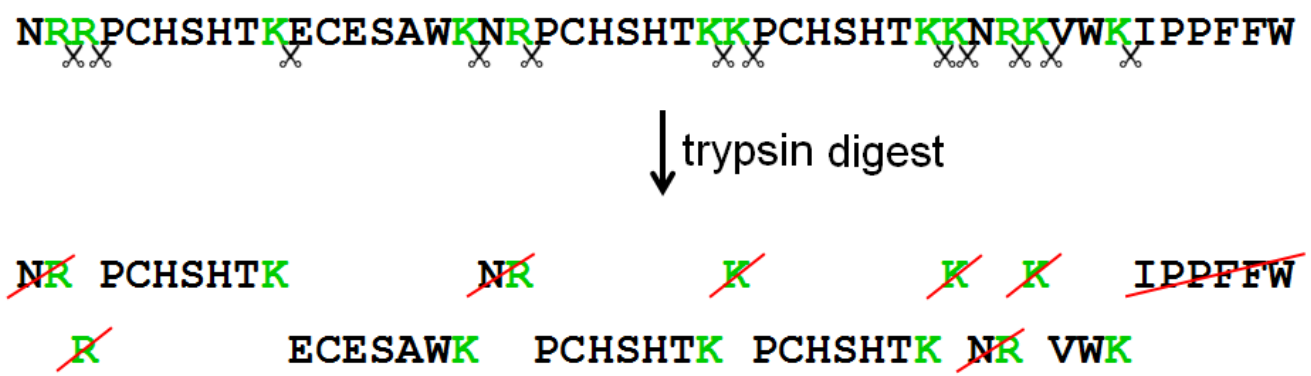
<u>On-bead digest:</u>
Single-Pot Solid-Phase-enhanced Sample Preparation (SP3)

Mirzaei H., Carrasco M. (eds) Modern Proteomics – Sample Preparation, Analysis and Practical Applications. Advances in Experimental Medicine and Biology, vol 919. Springer

Digestion

Trypsin is the most common protease used in proteomics, since trypsin specifically cleaves C-terminal of Arg (R) and Lys (K).

- Sequence specificity greatly improves sensitivity of database searching
- Tryptic peptides typically have an optimal length (10-20 aa) for LC-MS analyses
- C-terminal basic residues greatly improve fragmentation behavior of peptides (improving identification rate)



Limitation of trypsin:

- Complete sequence coverage typically cannot be achieved with trypsin alone

LC-MS-based Proteomics

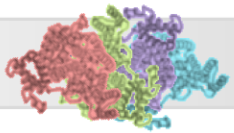
nanoHPLC

Reversed phase (C-18)

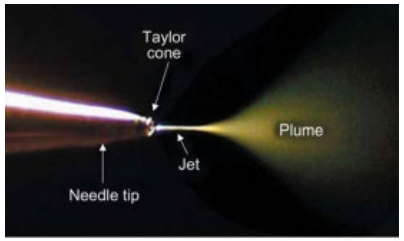
proteins

peptides

Bottom-up proteomics

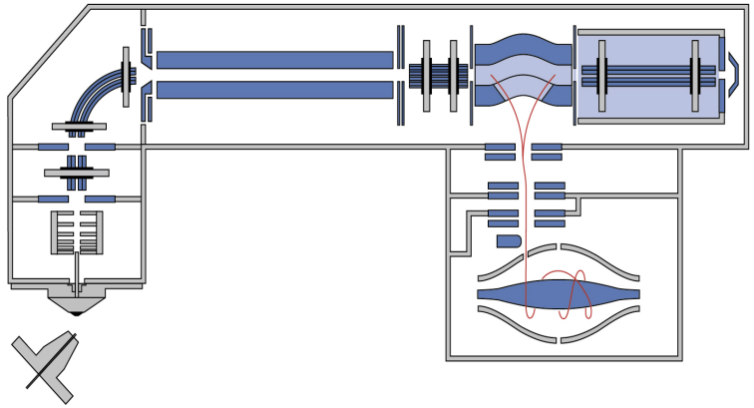


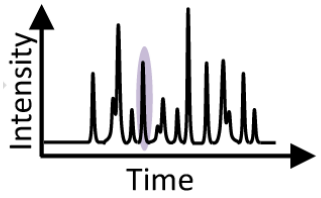
Digestion



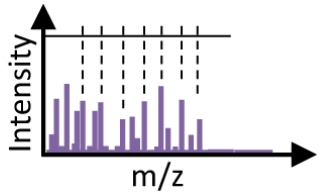
ESI

MS





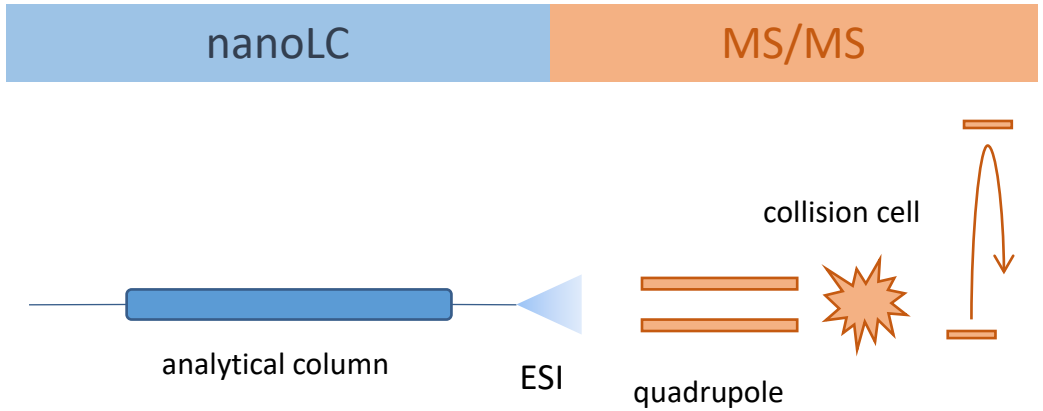
LC-MS



MS/MS

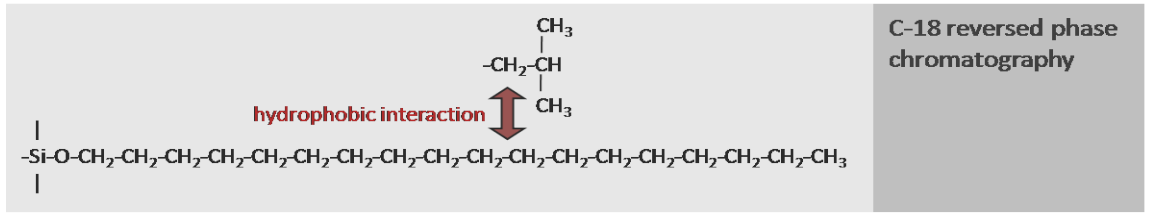
adapted from J. Proteome Res. 2013, 12, 1067−1077

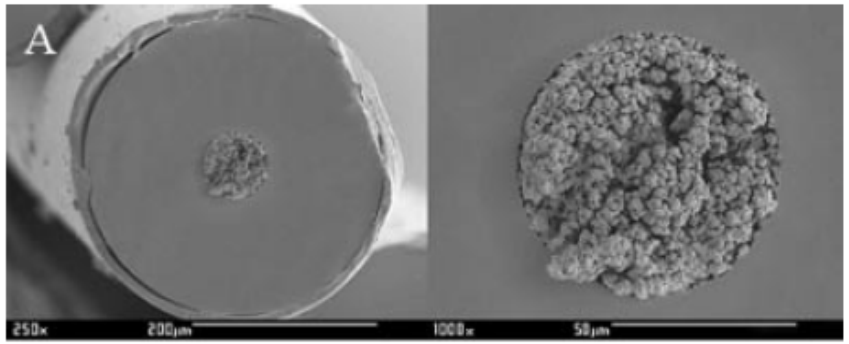
nanoLC-MS

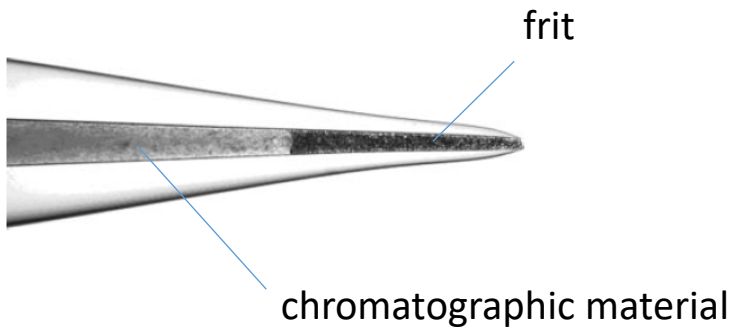


nano HPLC

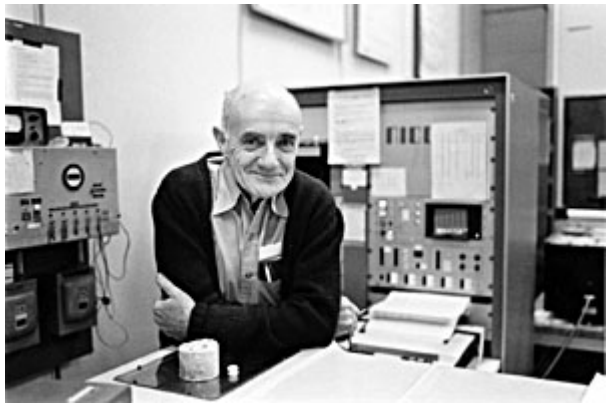
50 - 75 µm i.d.
length: 15 – 50 cm
C-18 (1.7 - 5 µm particles)
flow rate: 200 – 300 nL/min
linear acetonitile gradient





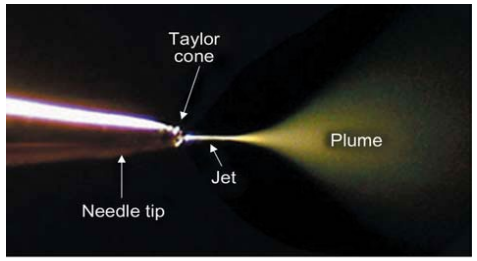


Electrospray Ionization (ESI)

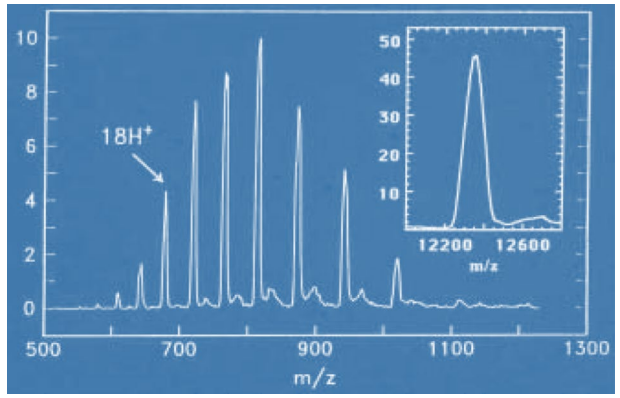


John Fenn

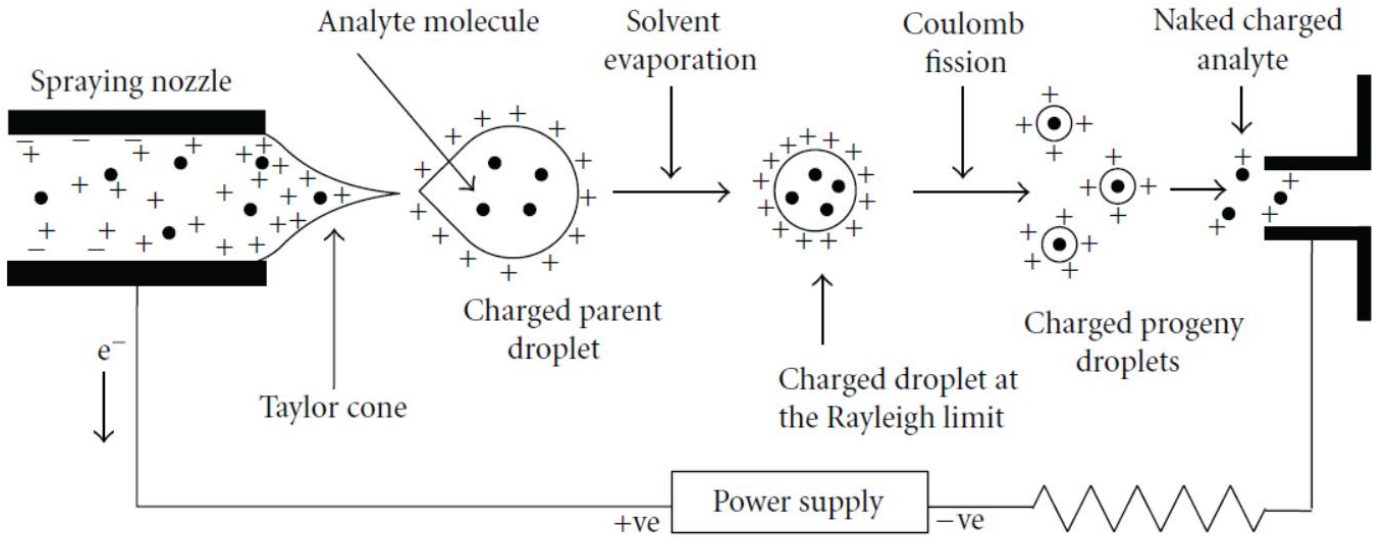
Nobel Price for Chemistry 2002



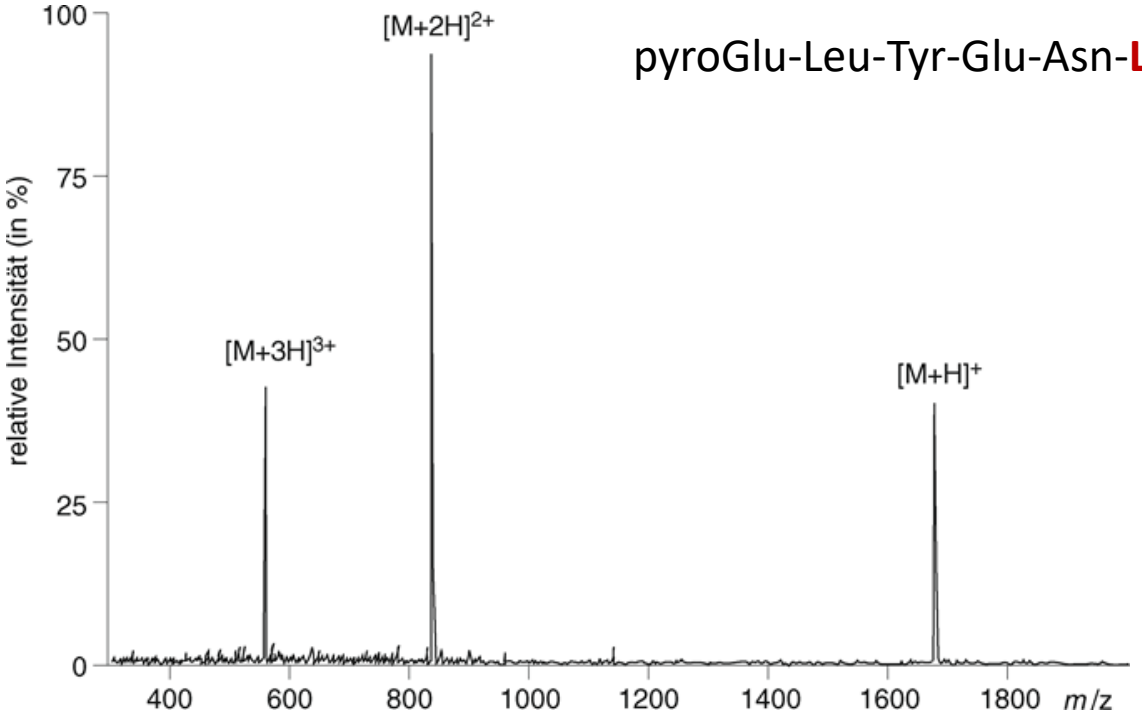
Cone Jet Mode



Cytochrome C (12 kDa), Noble Lecture 2002



Electrospray Ionization (ESI)

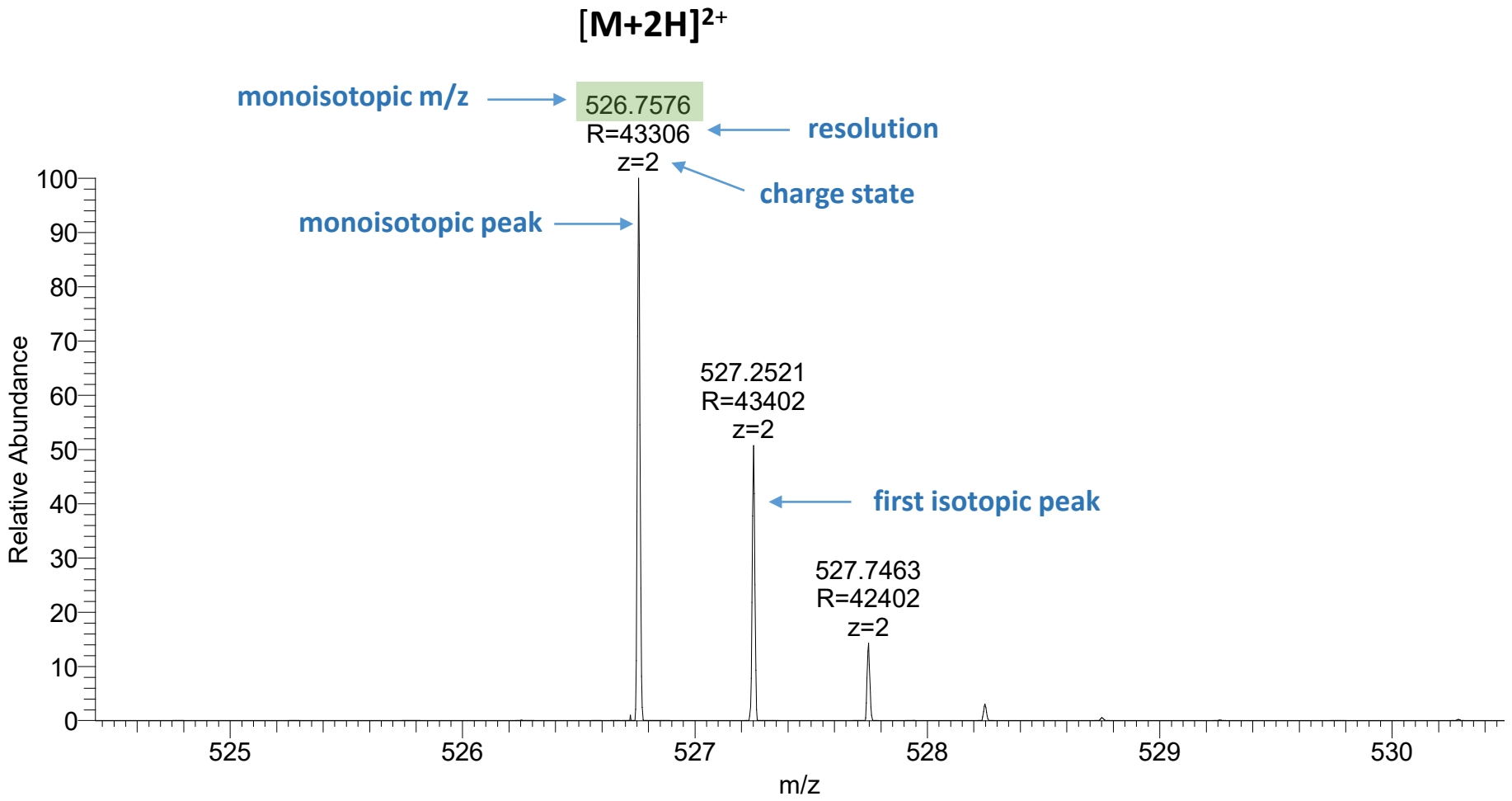


pyroGlu-Leu-Tyr-Glu-Asn-**Lys**-Pro-**Arg**-**Arg**-Pro-Tyr-Ile-Leu

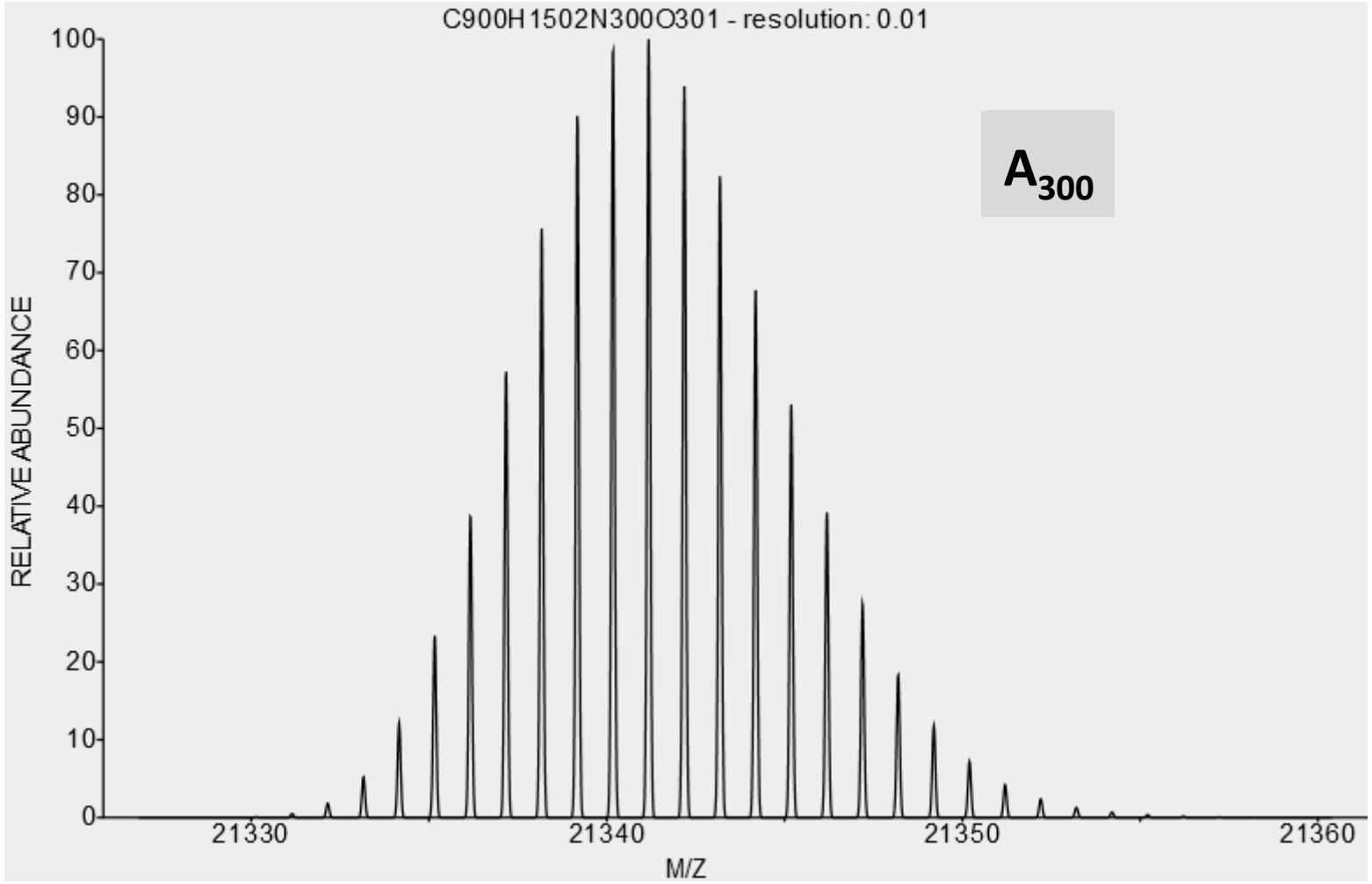
ESI spectrum of the **peptide neurotensin**.

For peptides the observed maximum charge state corresponds to number of basic sites!

Peptide Mass Spectrum

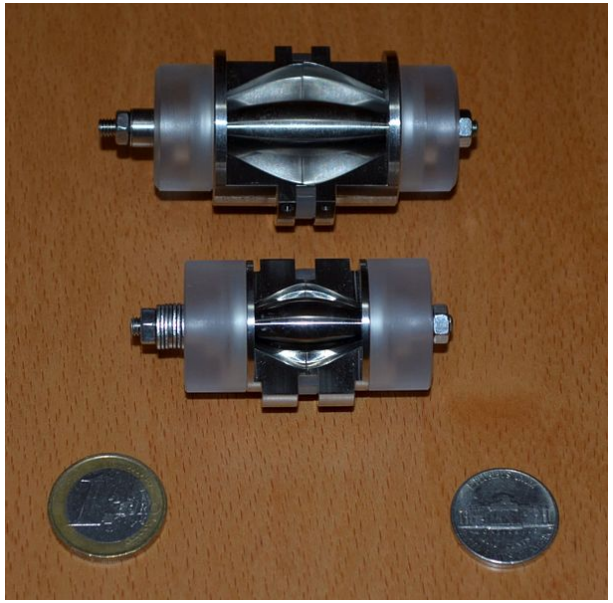


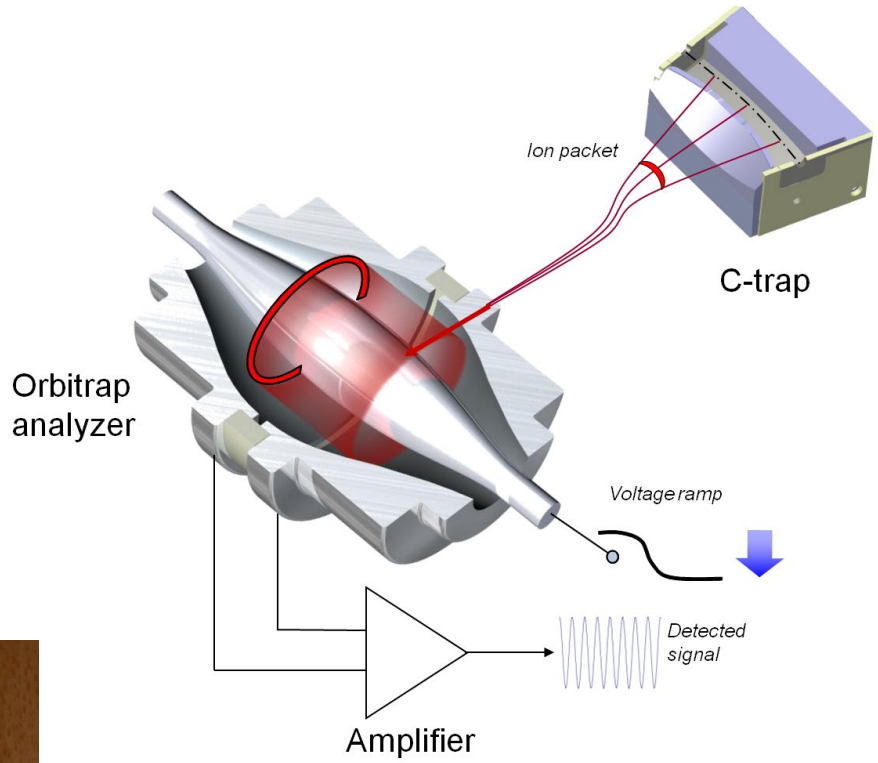
Isotopic Pattern of Peptides and Proteins



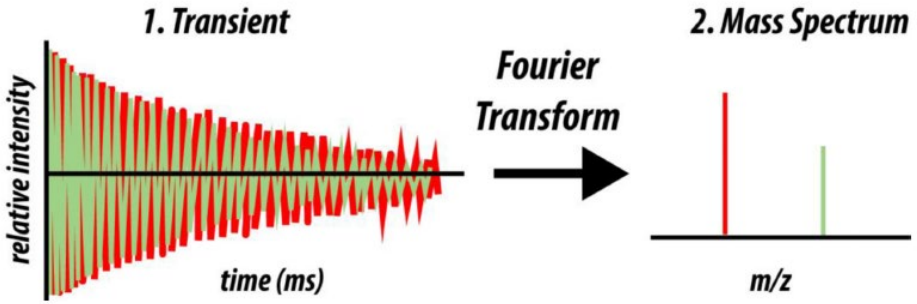
Orbitrap

Alexander Makarov (Thermo, Bremen)





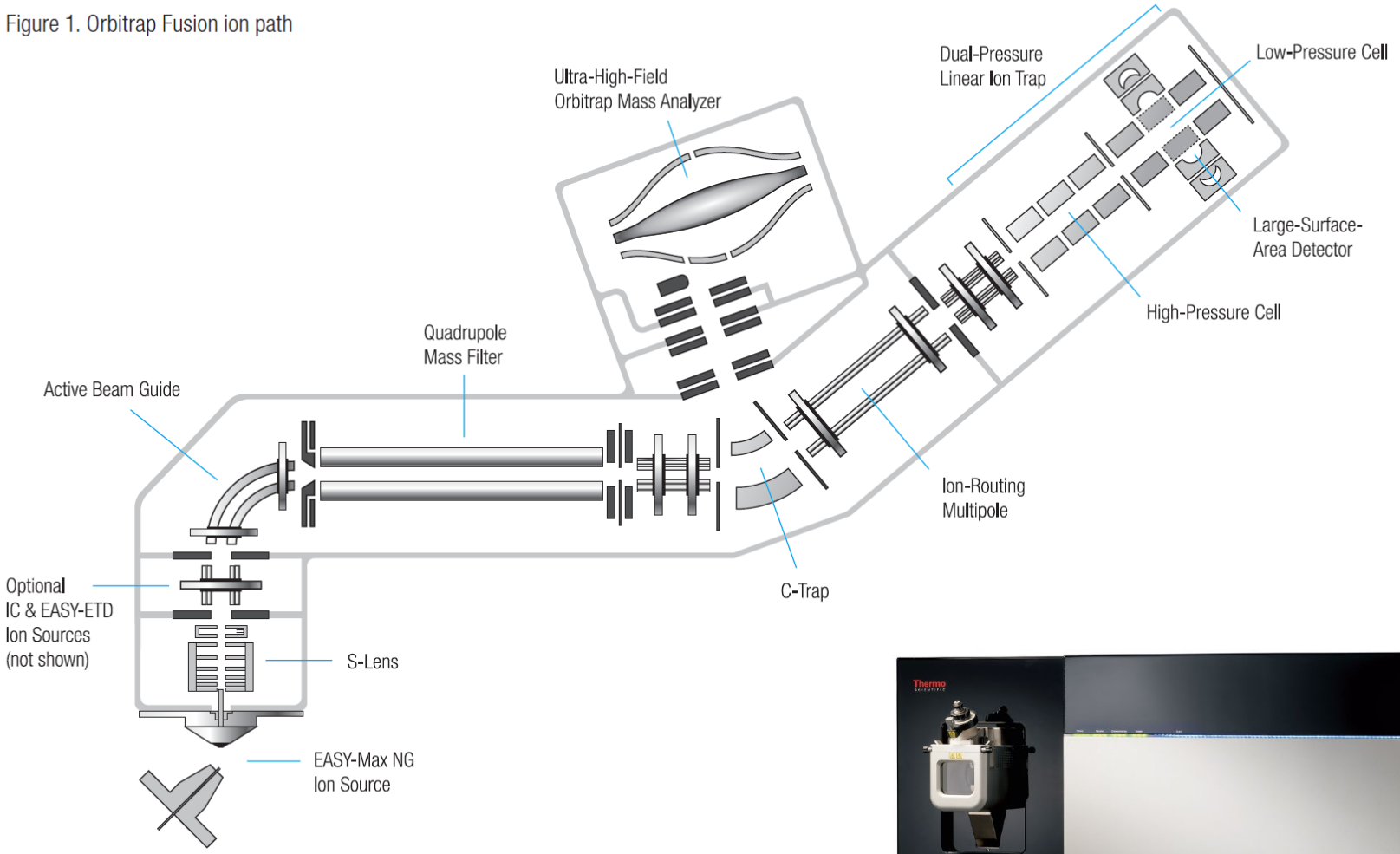
$$\omega_z = \sqrt{\frac{k}{m/z}}$$



- electrostatic ion trap
- very high resolution (up to 450 000)
- very high mass accuracy (1 ppm)

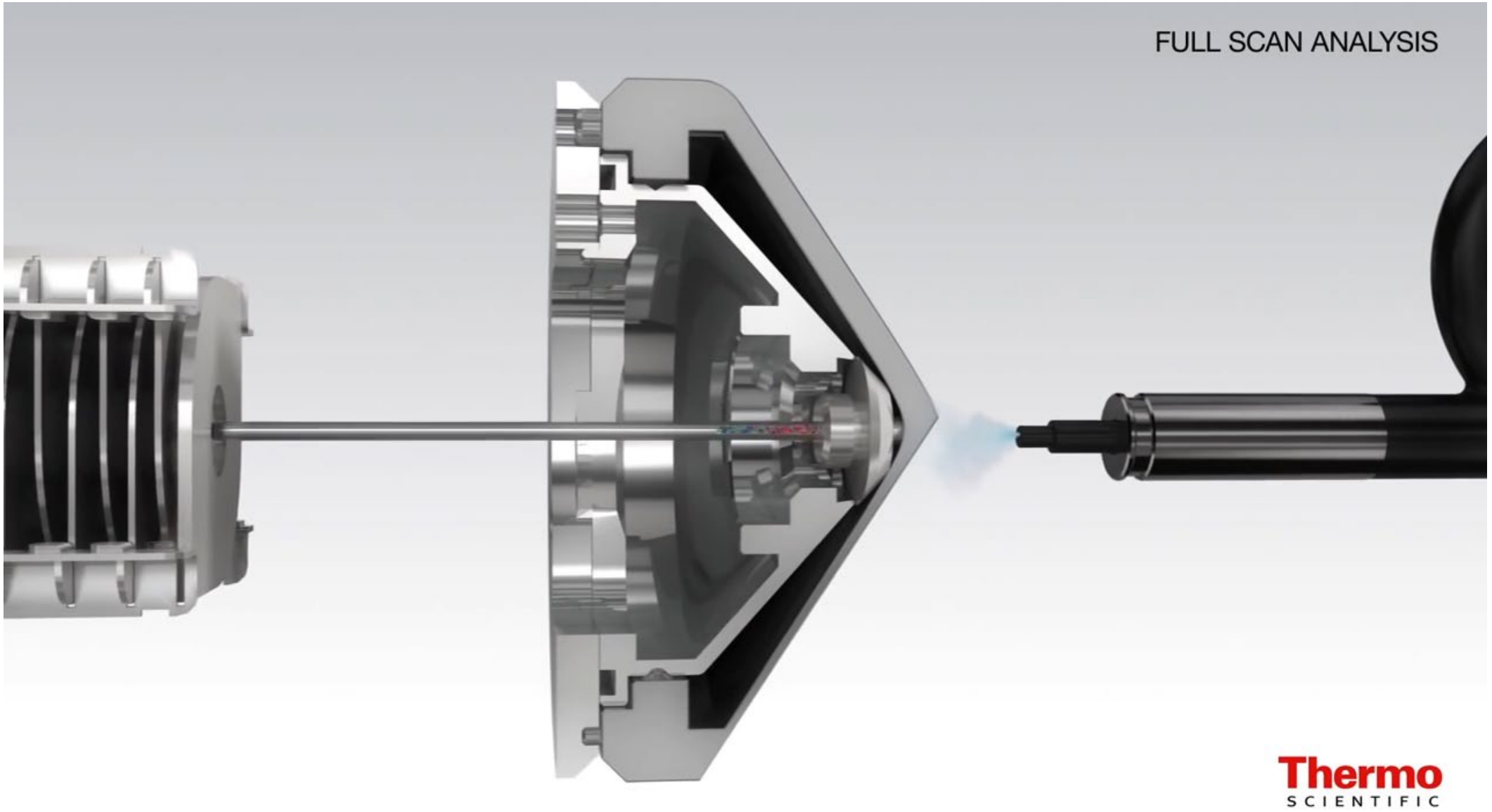
Orbitrap Fusion

Figure 1. Orbitrap Fusion ion path

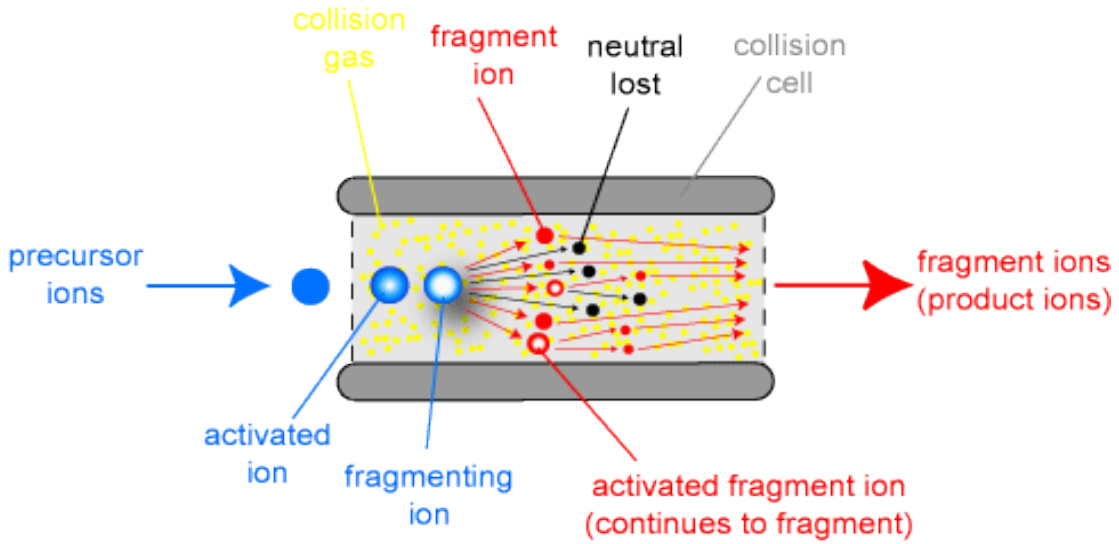


- Tribrid instrument – combination of three mass analyzer
- Ultra-High-Field Orbitrap mass analyzer with resolving power up to 450 000 (FWHM)

Orbitrap Fusion

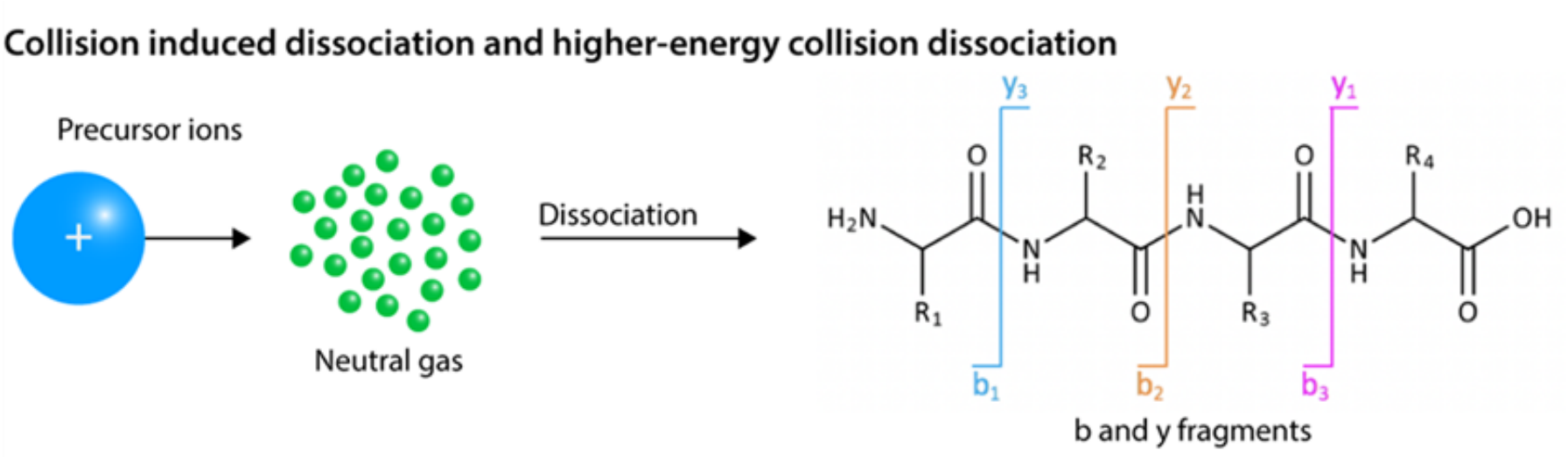


Collision-induced Dissociation (CID)



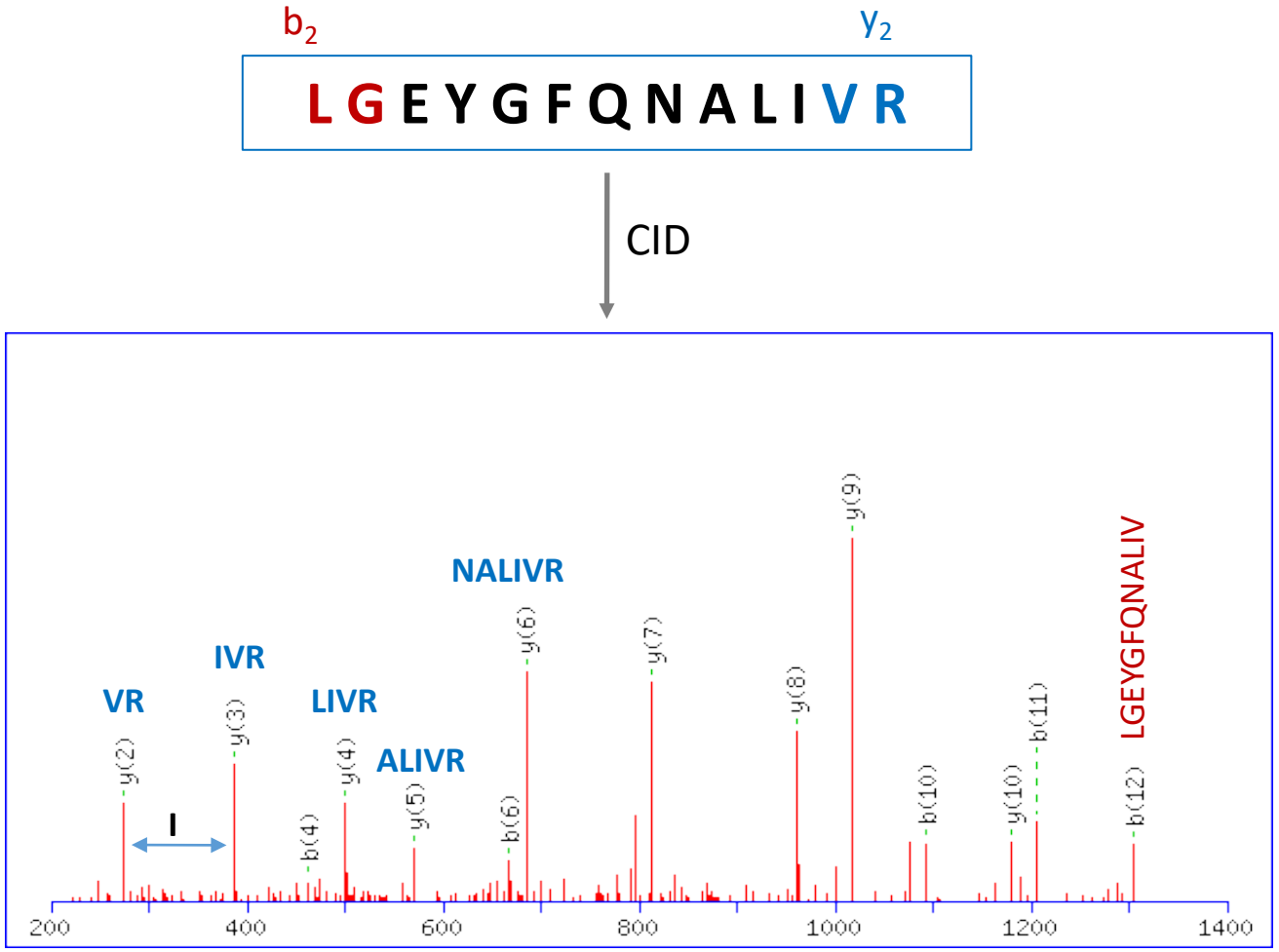
- collision cell is filled with N_2
- a voltage is applied in front of the collision cell (collision energy, typically 10-30 V)
- optimal collision energy depends on peptide mass and charge

Collision-induced Dissociation (CID)



https://www.creative-proteomics.com/blog/

Collision-induced Dissociation (CID)



MS/MS spectrum of the peptide LGEYGFQNALIVR

LC-MS-based Proteomics

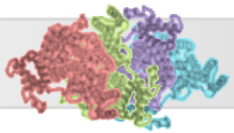
nanoHPLC

Reversed phase (C-18)

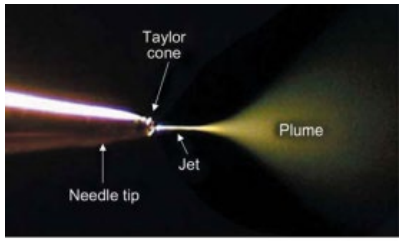
proteins

peptides

Bottom-up proteomics

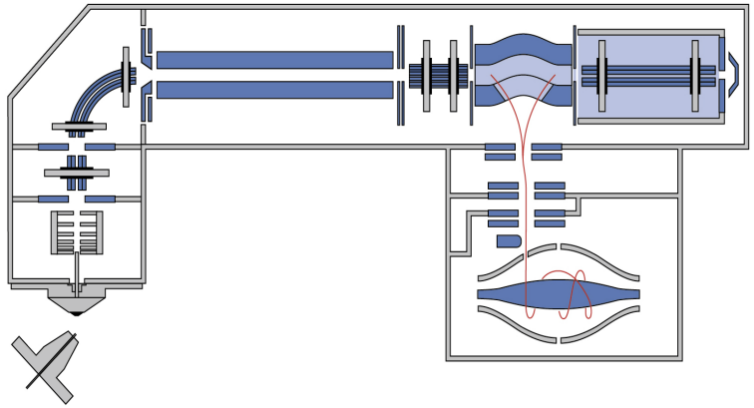


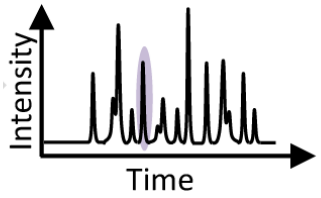
Digestion



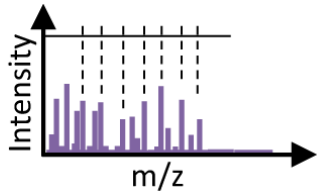
ESI

MS





LC-MS

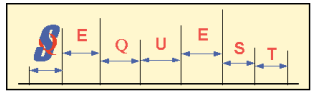


MS/MS

adapted from J. Proteome Res. 2013, 12, 1067−1077

Popular Database Search Engines

Mascot



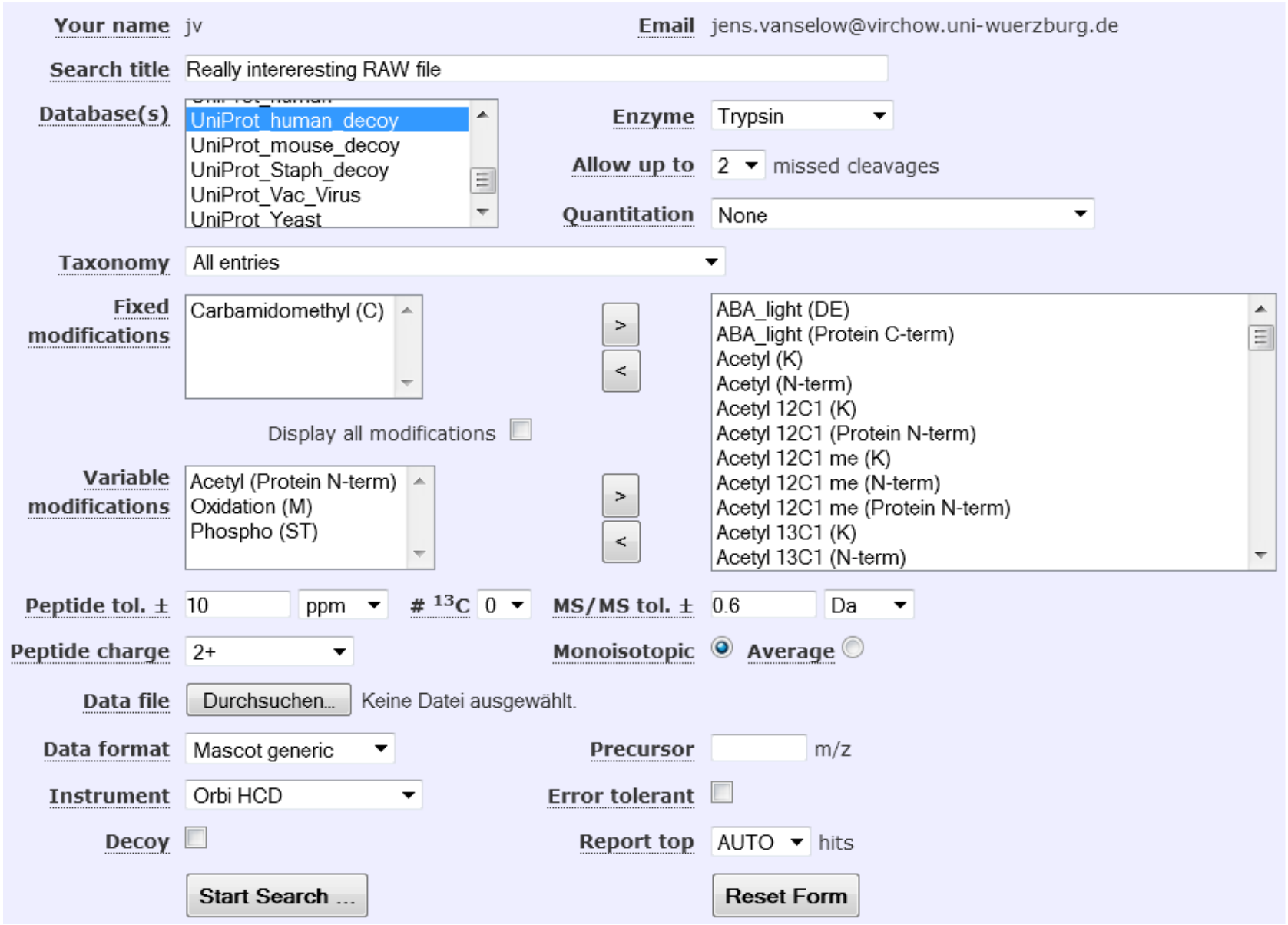
Sequest

MaxQuant (Andromeda)

X!TANDEM

PEAKS

Database Searching for Peptide and Protein ID



Database Searching with Fragment Ion Spectra

Step 1: Generate candidate peptide list

```
1    MSKPHSEAGT  AFIQTQQLHA  AMADTFLEHM  CRLDIDSPPI  TARNTGIICT
51   IGPASRSVET  LKEMIKSGMN  VARLNFSHGT  HEYHAETIKN  VRTATESFAS
101  DPILYRPVAV  ALDTKGPEIR  TGLIKGSGTA  EVELKKGATL  KITLDNAYME
151  KCDENILWLD  YKNICKVVEV  GSKIYVDDGL  ISLQVKQKGA  DFLVTEVENG
201  GSLGSKKGVN  LPGAAVDLPA  VSEKDIQDLK  FGVEQDVDMV  FASFIRKASD
251  VHEVRKVLGE  KGKNIKIISK  IENHEGVRRF  DEILEASDGI  MVARGDLGIE
301  IPAEKVFLAQ  KMMIGRCNRA  GKPVICATQM  LESMIKKPRP  TRAEGSDVAN
351  AVLDGADCIM  LSGETAKGDY  PLEAVRMQHL  IAREAEAAIY  HLQLFEELRR
401  LAPITSDPTE  ATAVGAVEAS  FKCCSGAIIV  LTKSGRSAHQ  VARYRPRAPI
451  IAVTRNPQTA  RQAHLYRGIF  PVLCKDPVQE  AWAEDVDLRV  NFAMNVGKAR
501  GFFKKGDVVI  VLTGWRPGSG  FTNTMRVVPV  P
```

calculate mass list from all tryptic peptides of all proteins

```
MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCR
LDIDSPPITAR
NTGIICTIGPASR
SVETLK
EMIK
…
```

list of calculated masses

```
3643.6793 Da
1196.6401 Da
1358.6976 Da
 675.3803 Da
 519.2727 Da
…
```

± mass accuracy

list of experimental masses

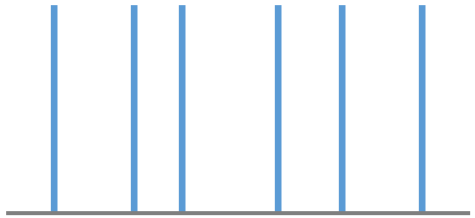
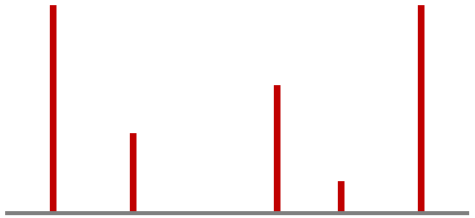
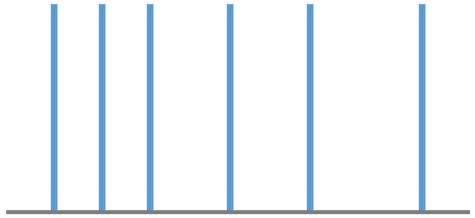
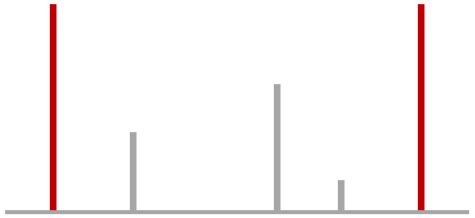
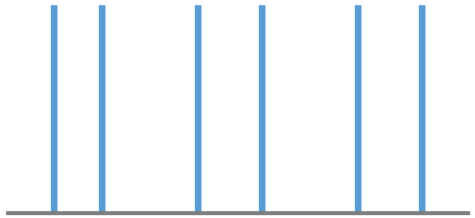
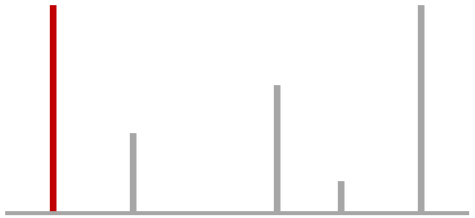
```
2493.1230 Da
1875.9597 Da
1140.6016 Da
1820.9139 Da
1635.8861 Da
1778.8701 Da
1196.6399 Da
1220.6222 Da
1358.6961 Da
1882.8945 Da
2464.2823 Da
 989.5024 Da
1196.5747 Da
1467.6714 Da
1461.8068 Da
1778.8704 Da
1763.9760 Da
…
```

candidate peptides

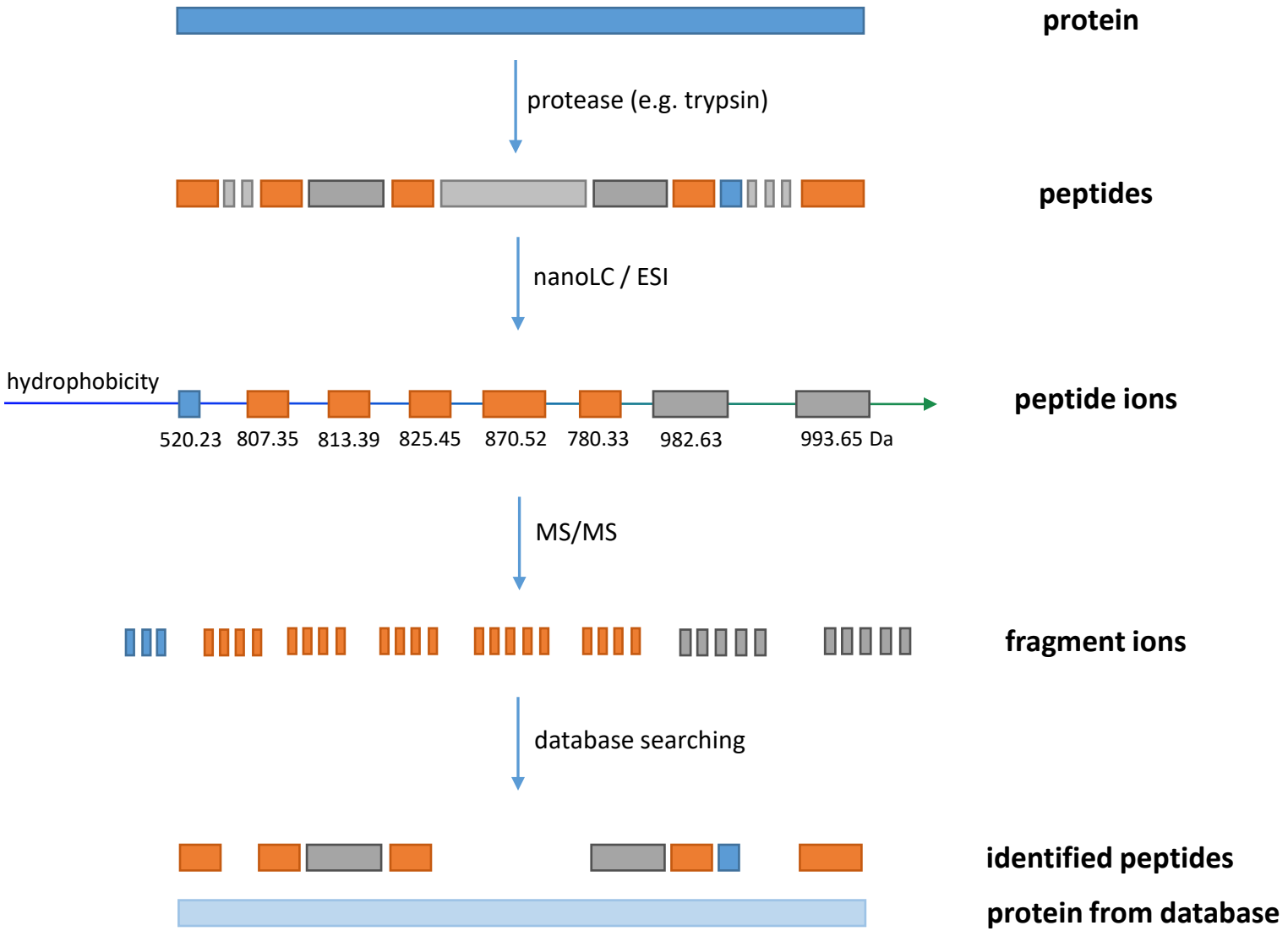
```
LDIDSPPITAR      1196.6401 Da
NTGIICTIGPASR    1358.6976 Da
```

Database Searching with Fragment Ion Spectra

Step 2: Score fragment ion spectra

candidate peptide	theoretical spectrum	experimental spectrum	
839.5229 Da **APIIAVTR**			high score
839.5229 Da **LVGVPSLR**			low score
839.5229 Da **LLAASLPR**			low score

Identification of Peptides from MS/MS Data



False Discovery Rate

False Discovery Rate (FDR) = FP / (FP + TP)

If TP is true positive matches and FP is false positive matches, the number of matches in the target database is TP + FP and the number of matches in the decoy database is FP.

Protein false discovery rate (FDR): fraction of the identified proteins that are in fact wrongly identified

Peptide false discovery rate (FDR): fraction of the identified peptides that are in fact wrongly identified

Evaluation of FDR by using a decoy database

target database	decoy database

example: 6037 identified peptides 224 identified peptides

peptide FDR = 224/6037 = 3.71%

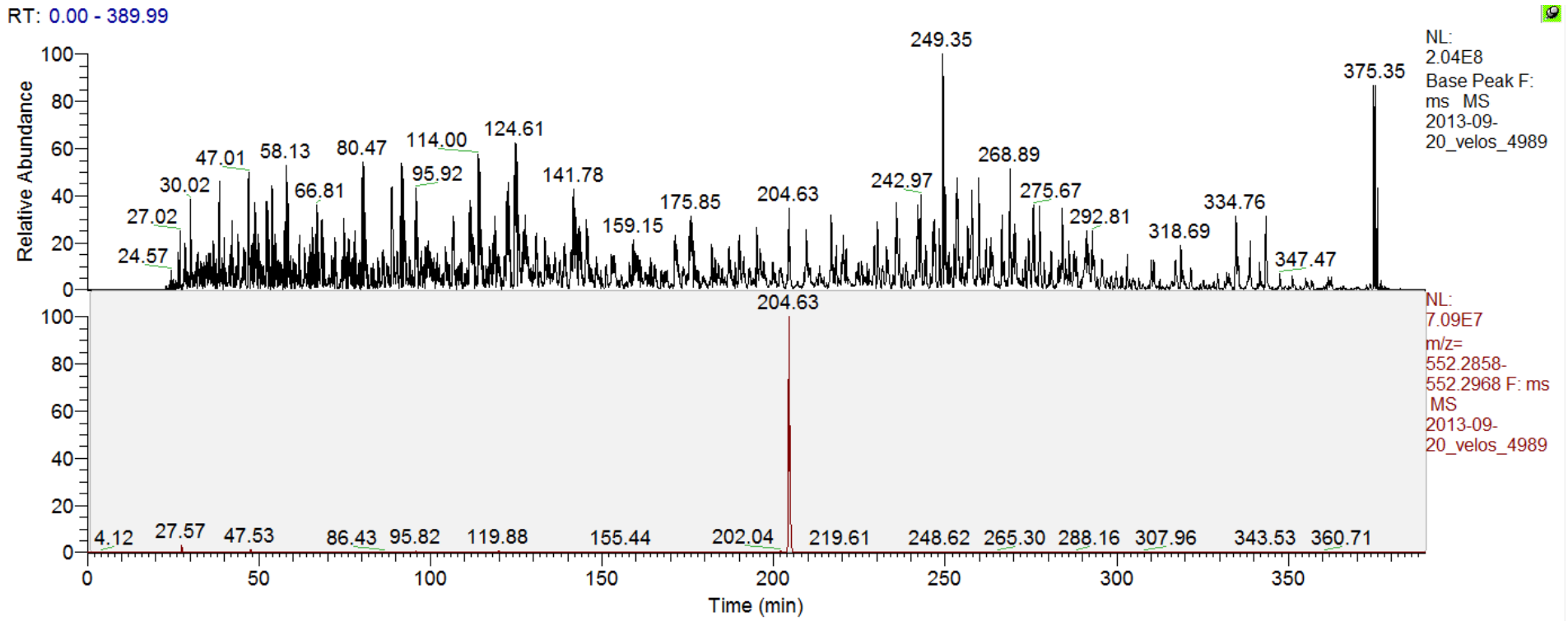
There are different approaches estimating FDR:

- Separate searches in target and decoy database (applied by Mascot; gives higher FDR rates)
- One search using a concatenated target-decoy database
- Decoy database can be random or reversed

Ultra High Pressure High Performance Liquid Chromatography (UHPLC)

yeast tryptic digest of total proteins (5 µg)

6 hour-gradient



pressure: up to 1000 bar

particle size: 1.9 µm

flow: 200 nL/min

column length: 50 cm

column ID: 75 µm

Data-dependent nanoLC-MS/MS

Table 1. Statistics of data.

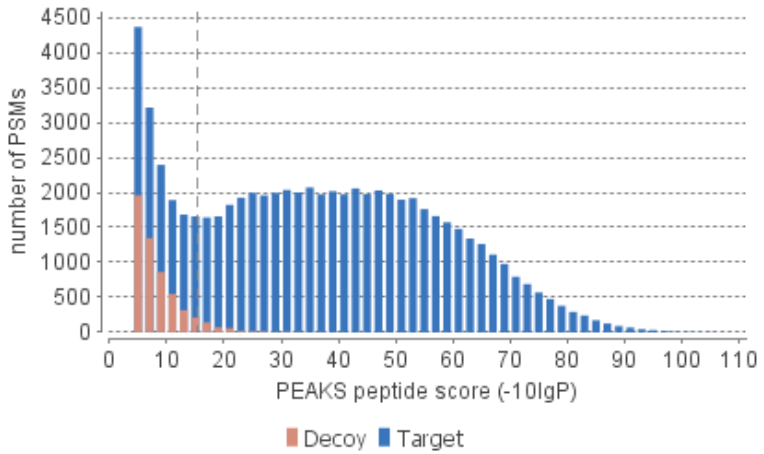
# of MS Scans	14325
# of MS/MS Scans	86519

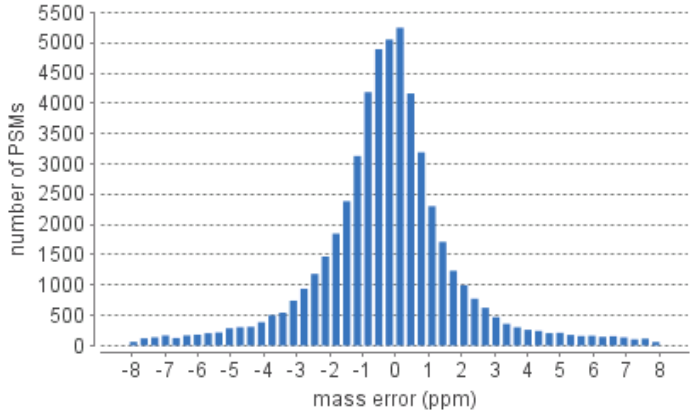
Table 3. Statistics of filtered result.

Peptide-Spectrum Matches	53063
Peptide Sequences	36737
Protein Groups	4463
Proteins	4538
Proteins (#Unique Peptides)	3036 (>2); 494 (=2); 604 (=1);
FDR (Peptide-Spectrum Matches)	1.0%
FDR (Peptide Sequences)	1.4%

Table 5. Number of identified peptides in each sample by the number of missed cleavages

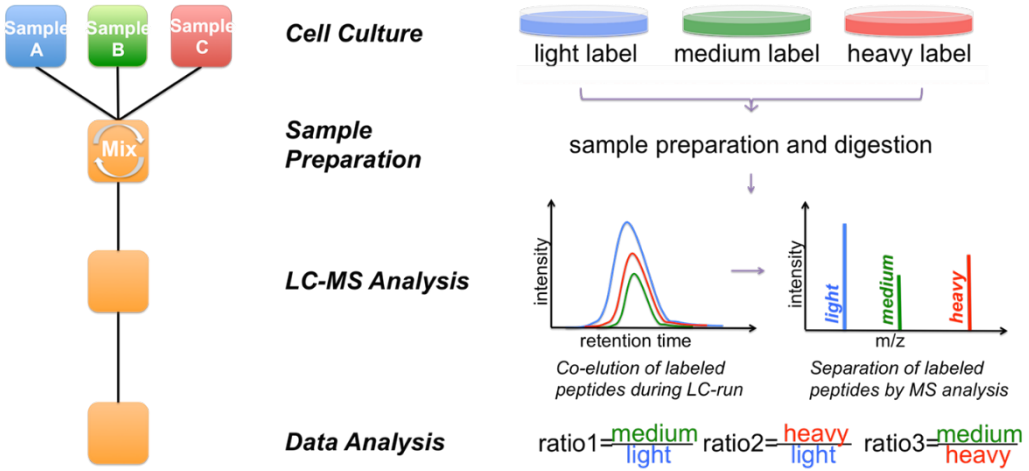
Missed Cleavages	0	1	2	3	4+
single run	35569	1155	13	0	0





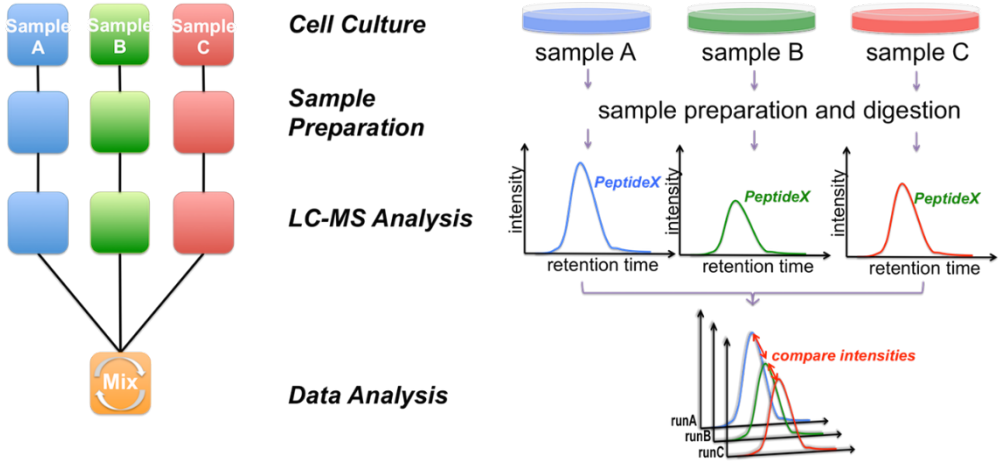
Quantitative Proteomics

Stable isotope labeling, e.g. SILAC



- most accurate quantification
- less instrument time
- higher peak complexity (might result in less identified peptides)
- requires isotopically labeled reagents (expensive)
- labeling strategy depends on sample type

Label-free Quantification



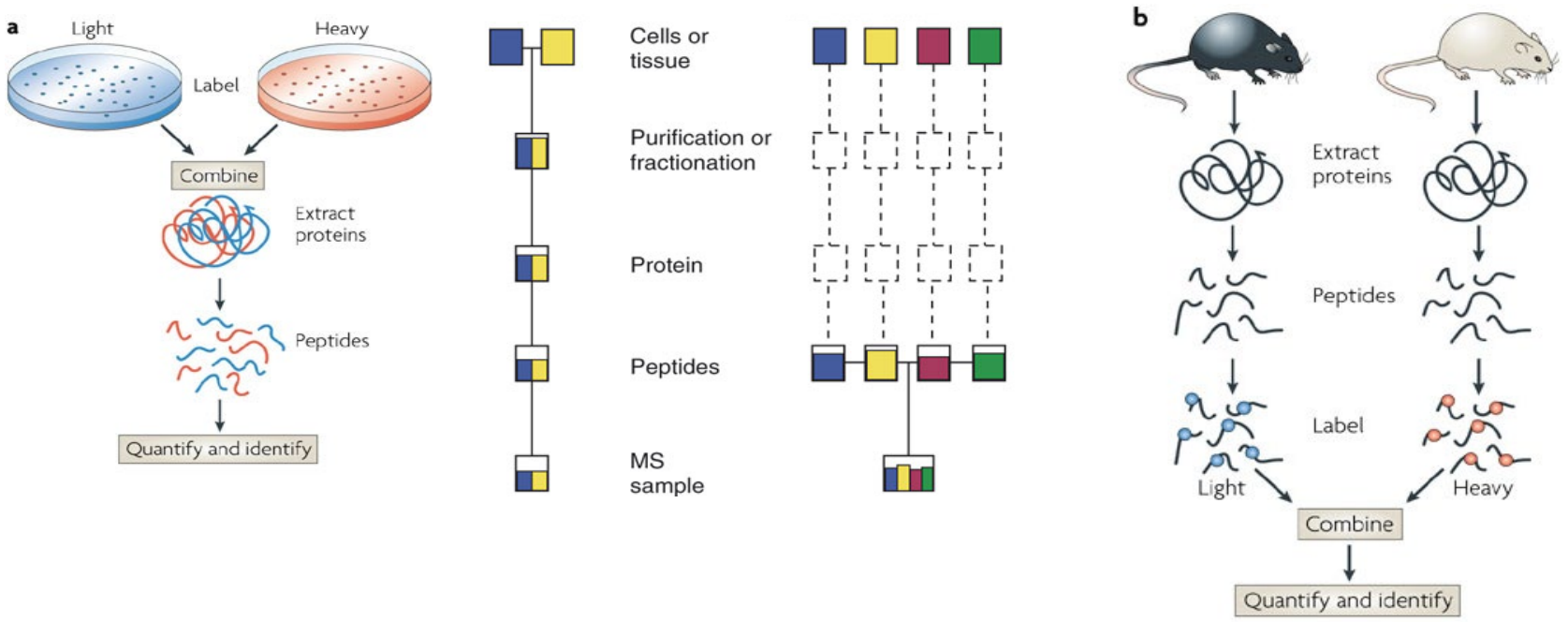
- applicable to any type of sample
- no additional labeling step
- no additional costs
- peak complexity is not increased
- less accurate quantification
- more replicates, more instrument time

adapted from https://www.zmbh.uni-heidelberg.de/Central_Services/Mass_Spectrometry/quant_proteomics.html

Metabolic vs. Peptide-Labeling

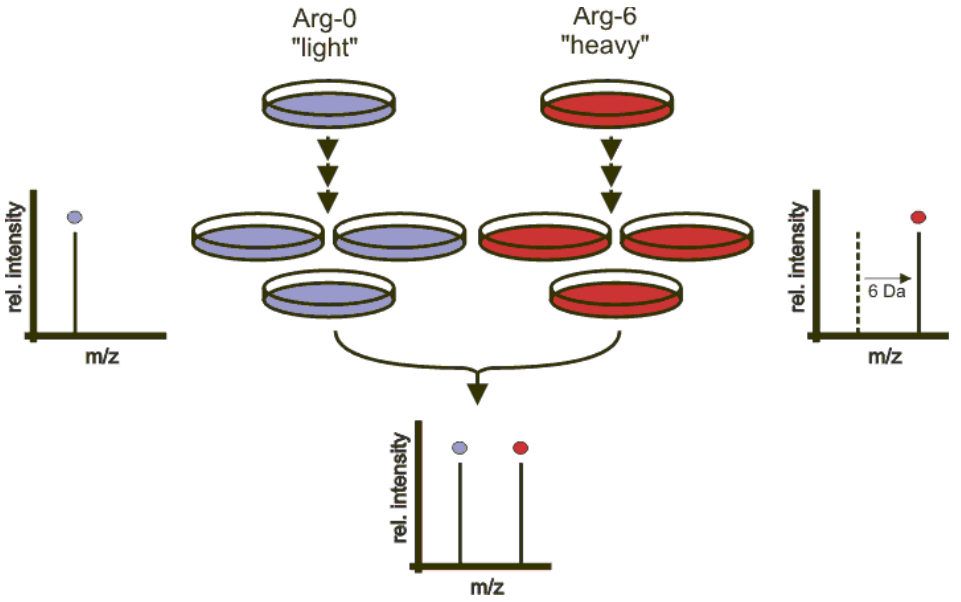
metabolic labeling

chemical labeling



Stable Isotope Labeling by Amino Acids in Cell Culture - SILAC

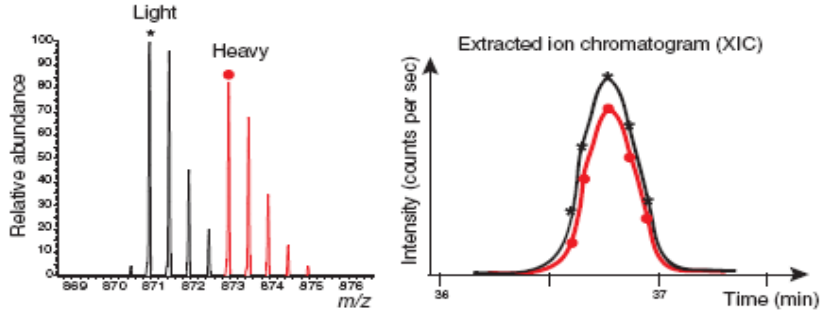
Matthias Mann,
MPI Martinsried



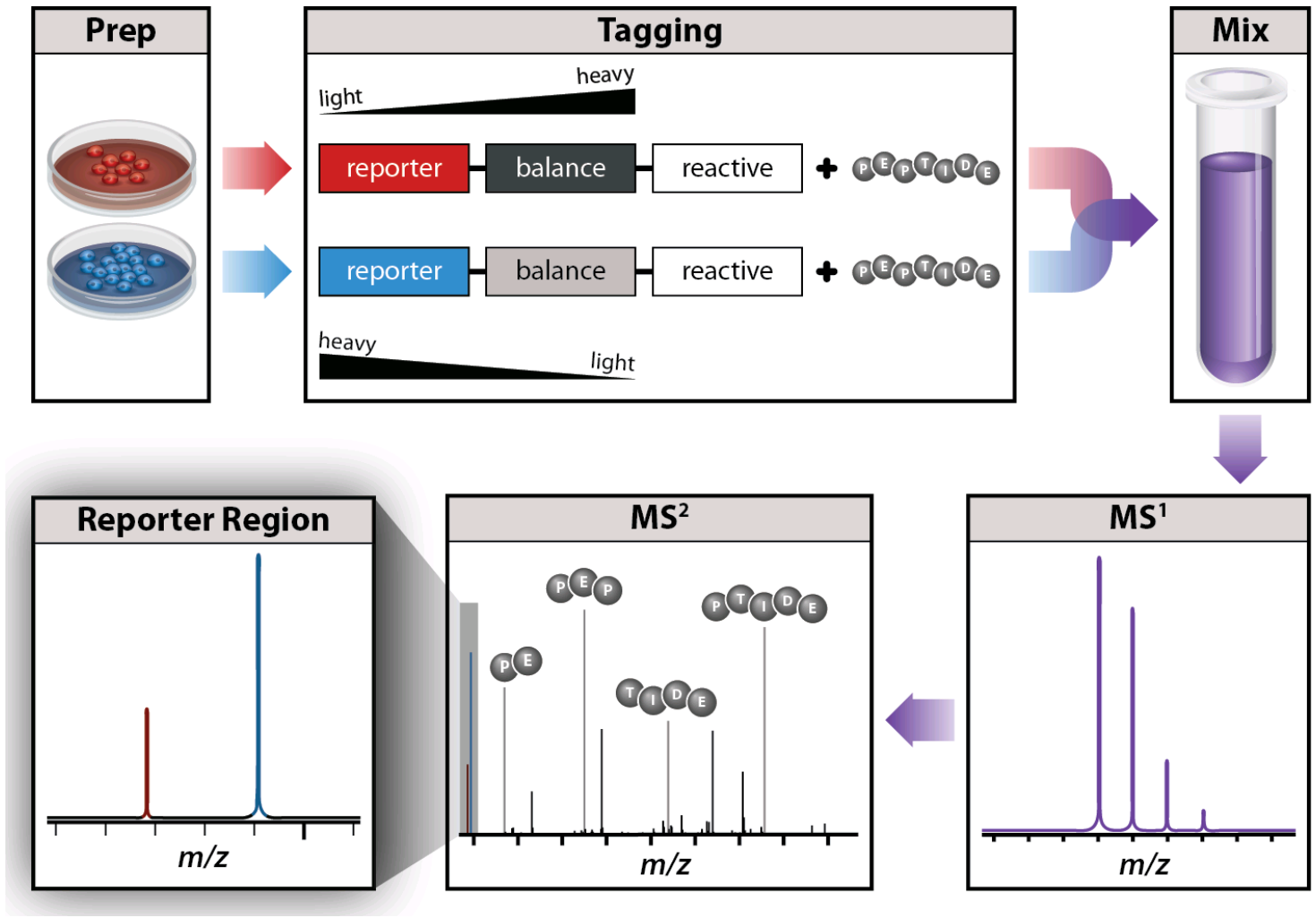
<u>tryptic peptides:</u>

AVPSLTR

AVPSLTK

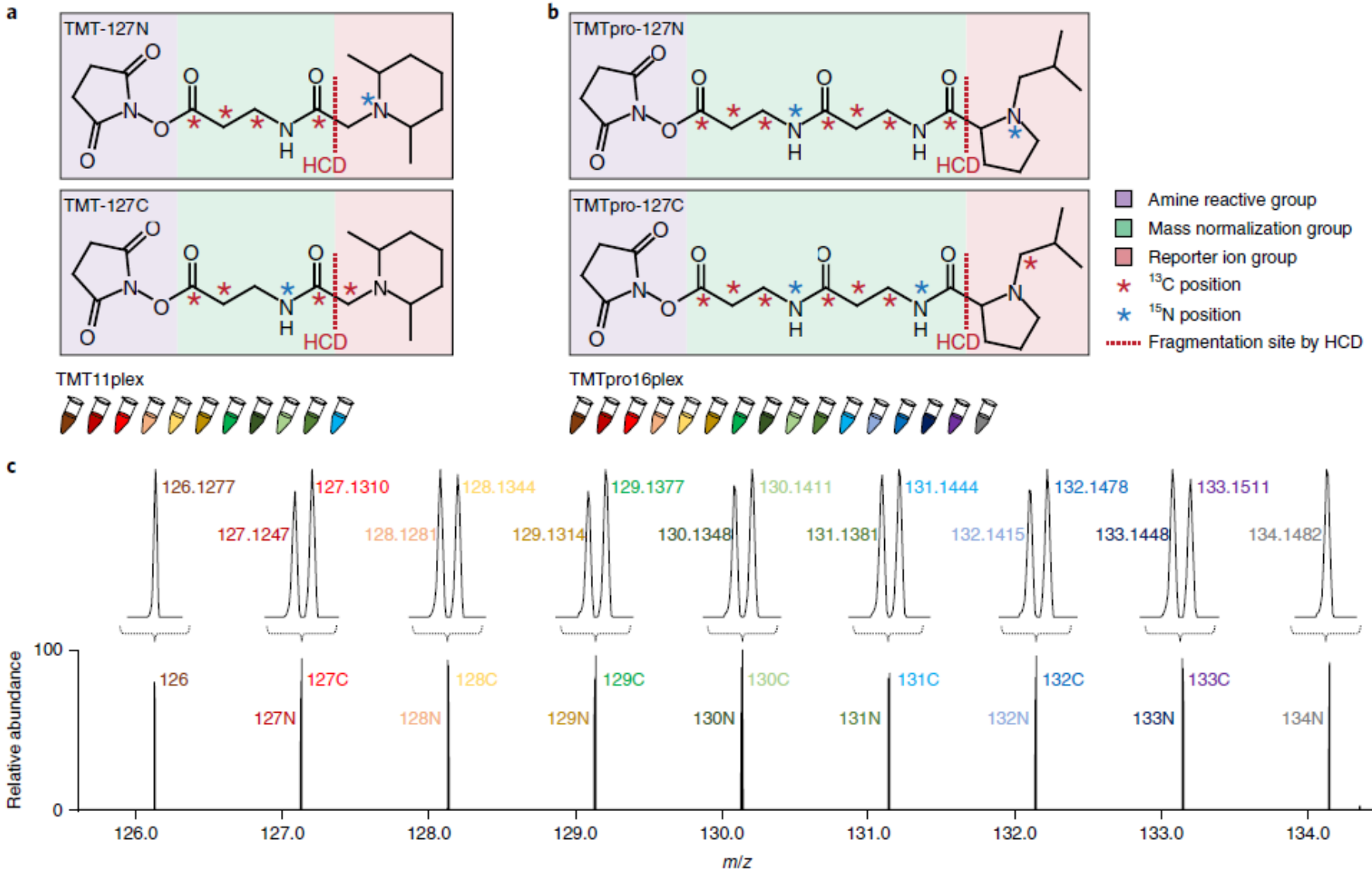


Isobaric Labeling



TMT Labeling

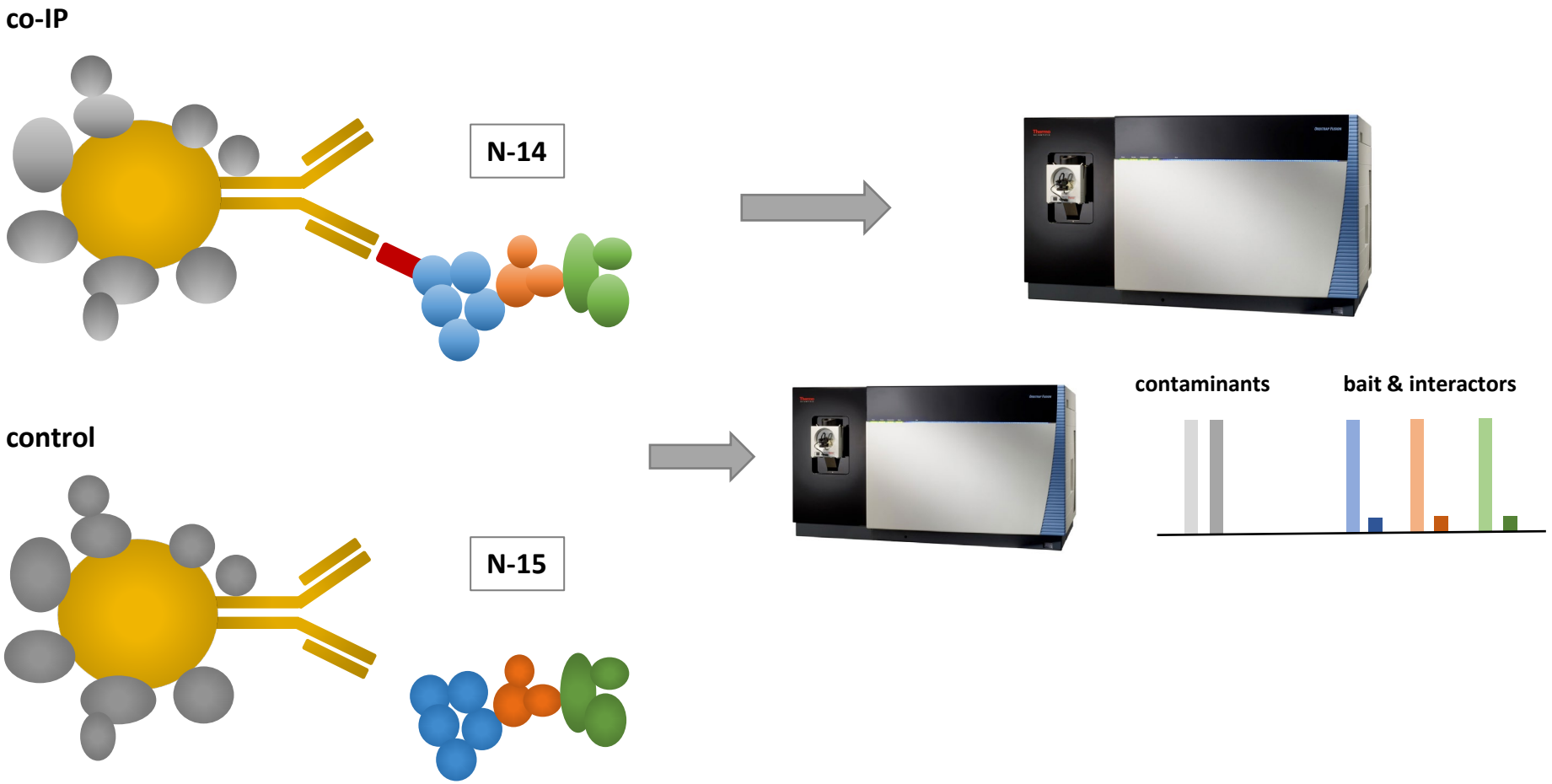
Tandem **M**ass **T**ags



Nature Methods volume 17, 399–404 (2020)

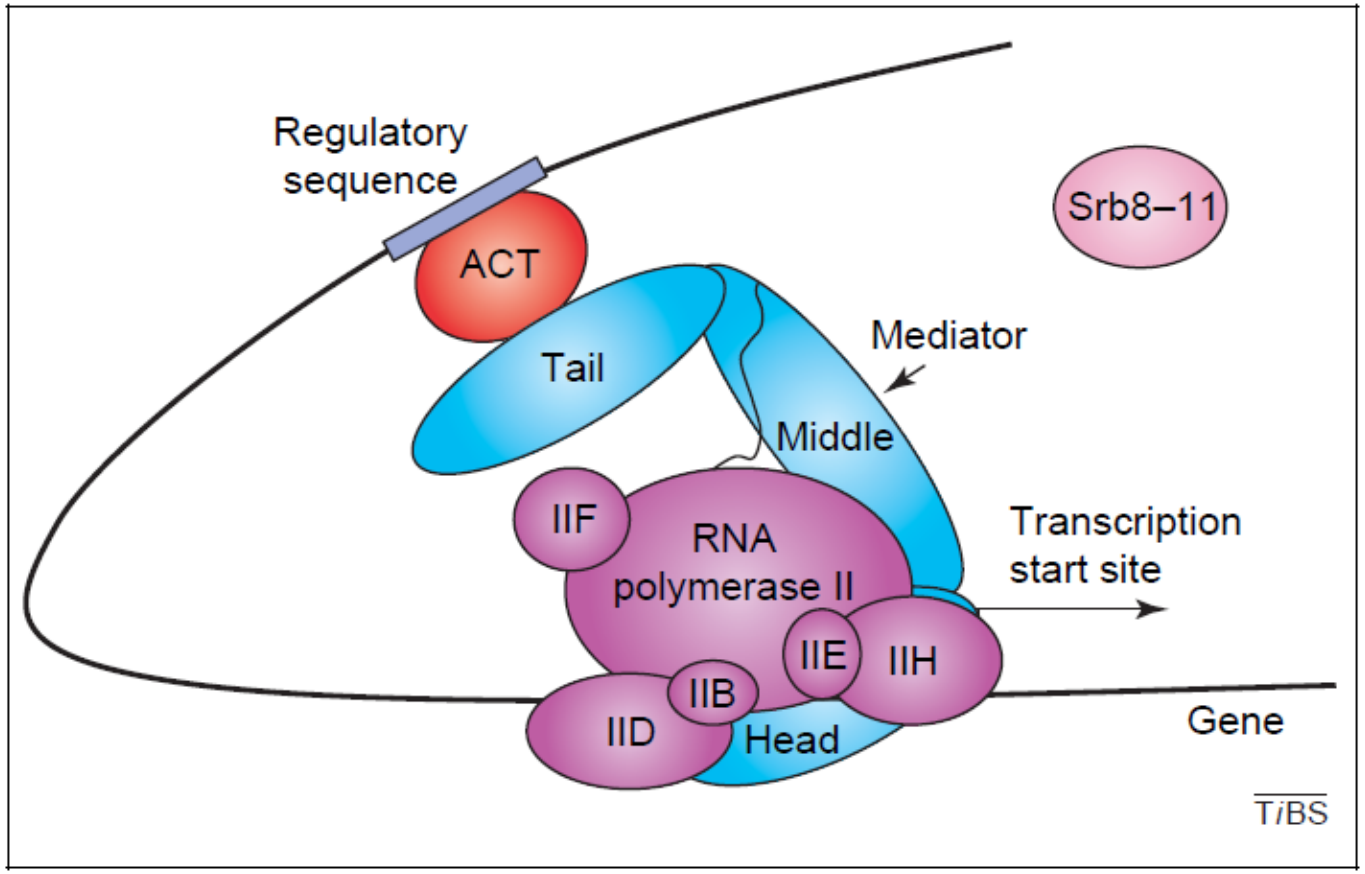
ID of Protein Interaction Partners

Mass Spectrometric Analysis of Co-IPs



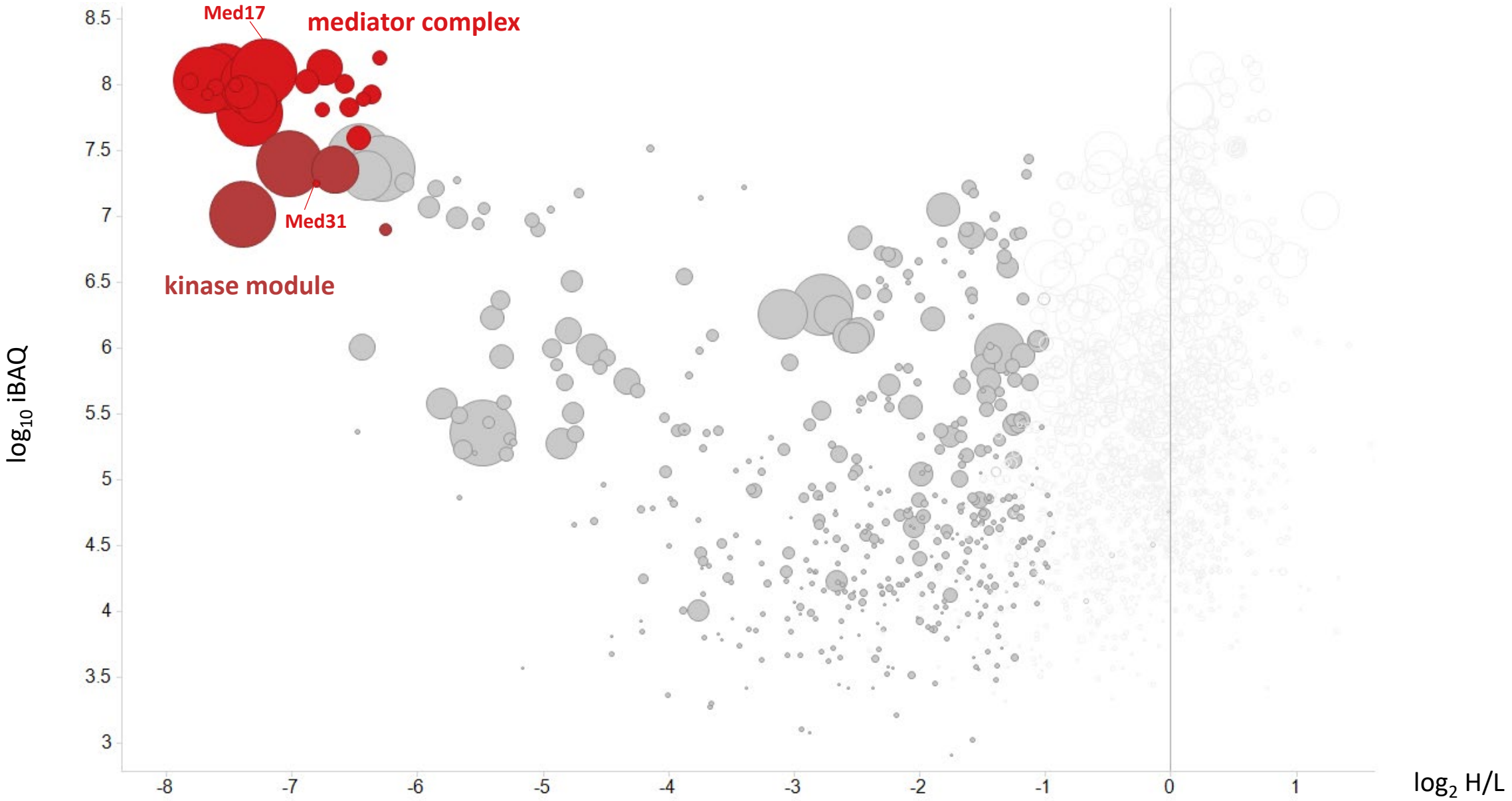
The Mediator Complex

- essential for regulated transcription
- communicates signals from activators to Pol II
- stimulates PIC formation
- structure and function is conserved in higher cells

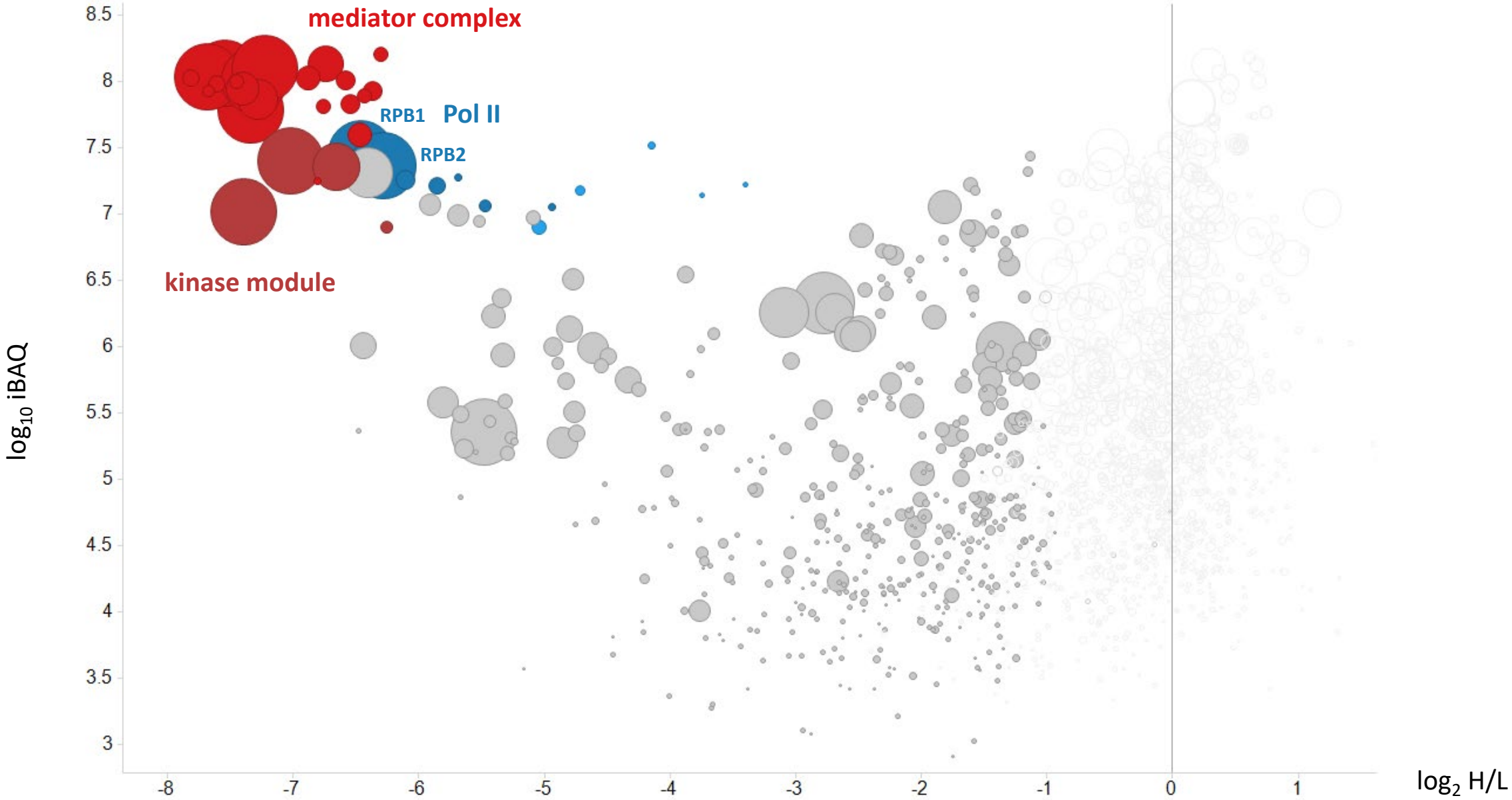


from S. Björklund & C.M. Gustafsson, *TIBS* **2005**, 30:240.

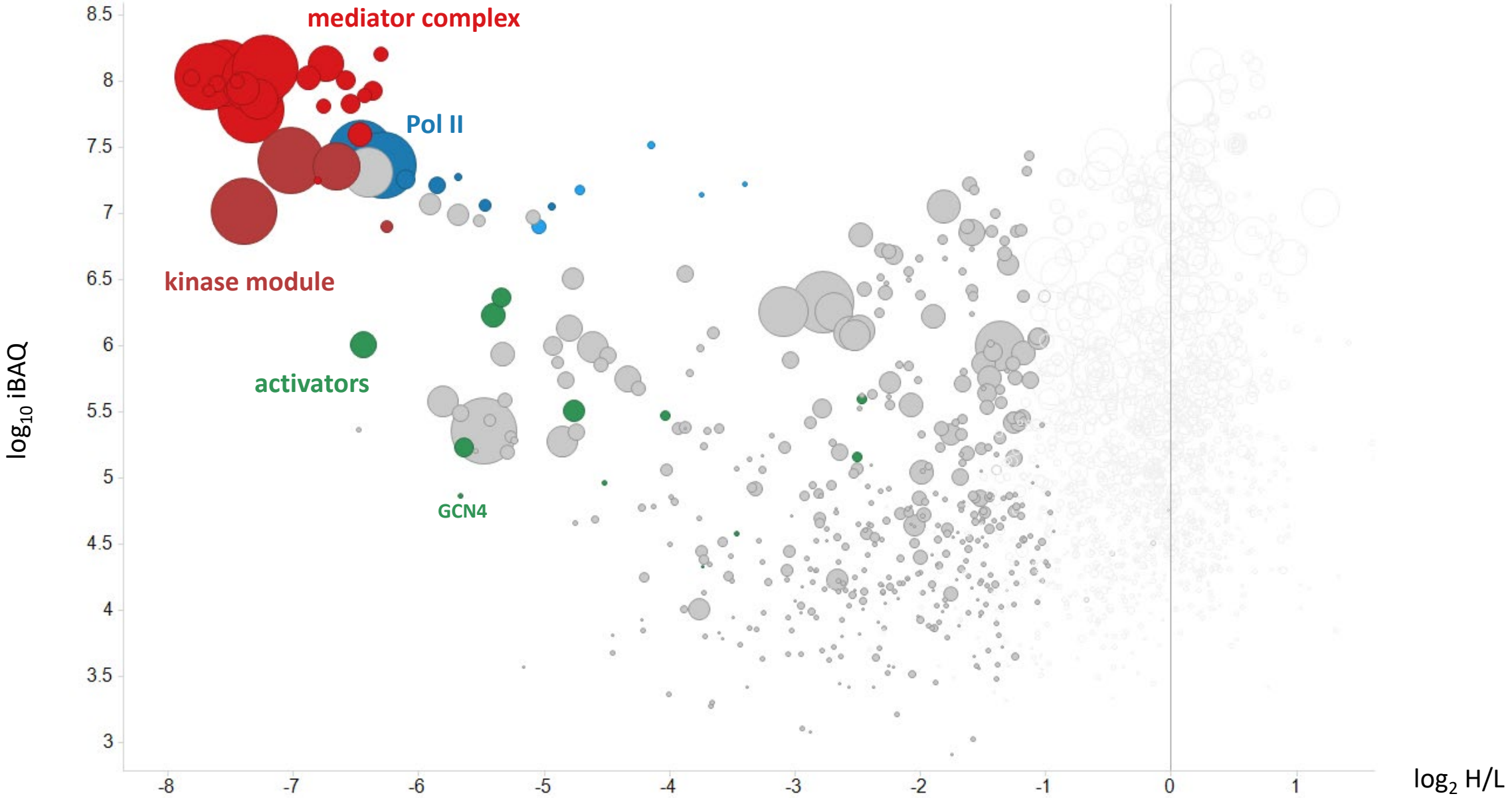
The Yeast Mediator Interactome



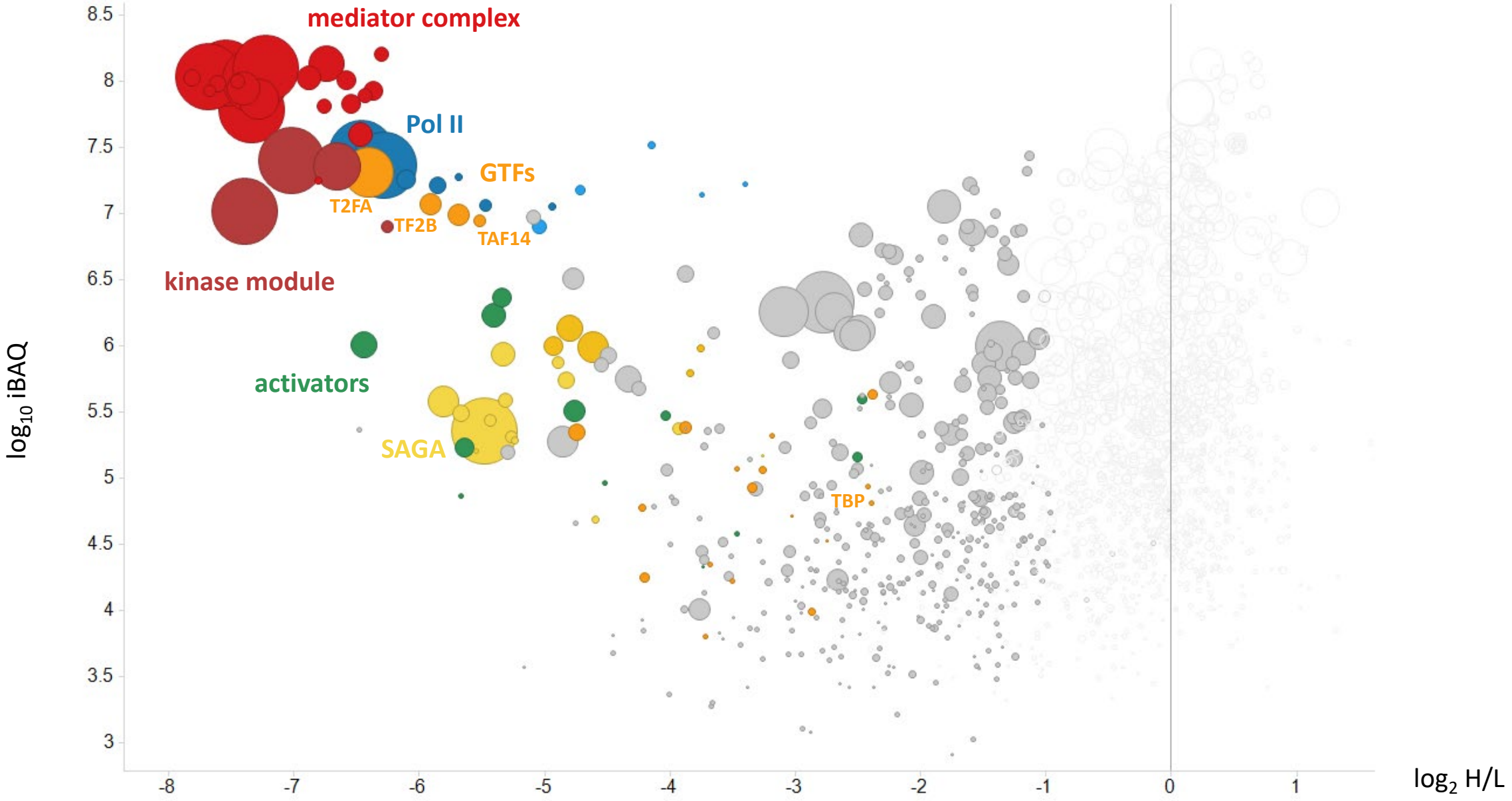
The Yeast Mediator Interactome



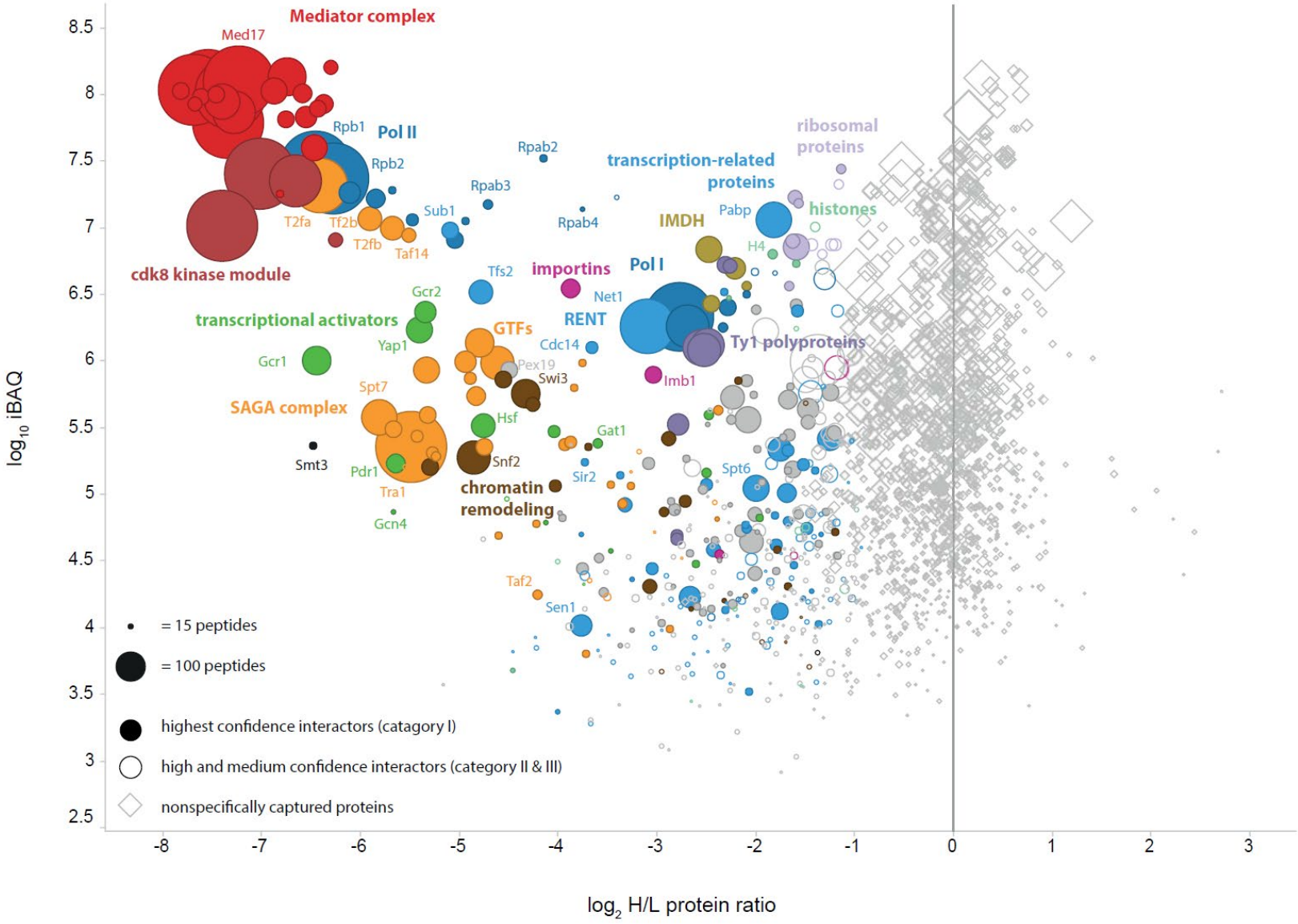
The Yeast Mediator Interactome



The Yeast Mediator Interactome



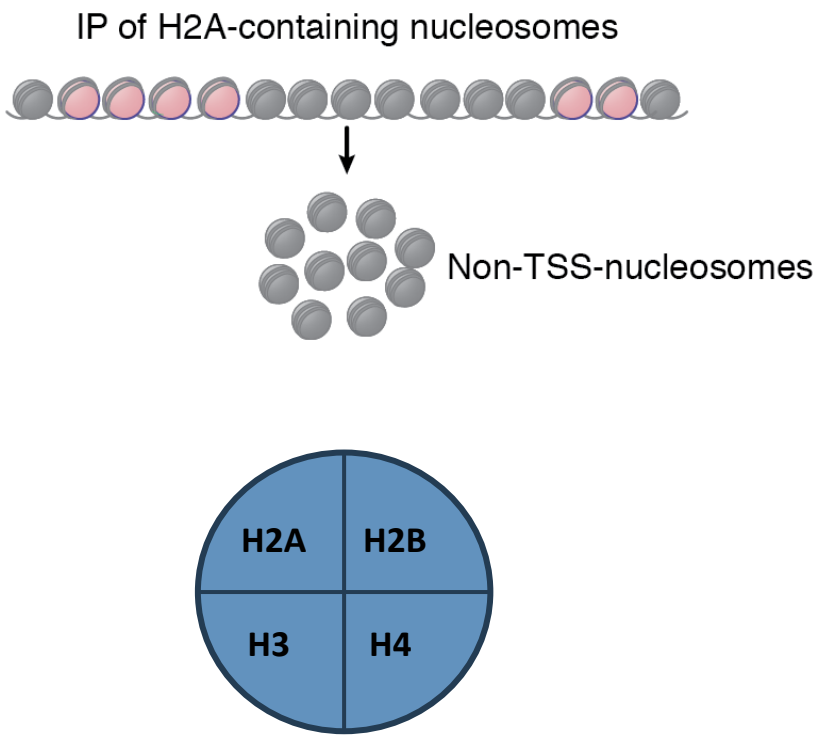
The Yeast Mediator Interactome

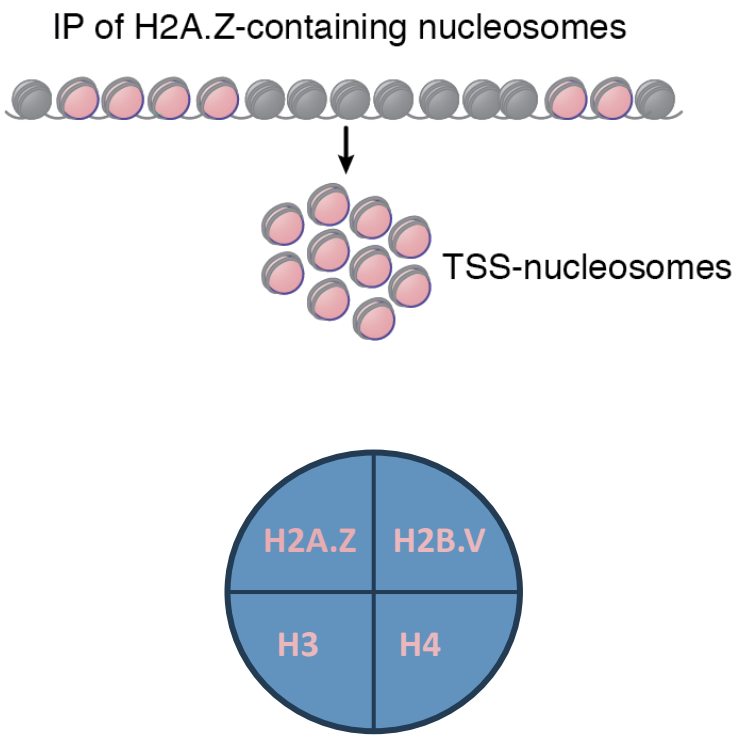


Uthe H, Vanselow JT, Schlosser A. Proteomic Analysis of the Mediator Complex Interactome in Saccharomyces cerevisiae. Sci Rep. 2017;7:43584.

Analysis of Posttranslational Modifications (PTMs)

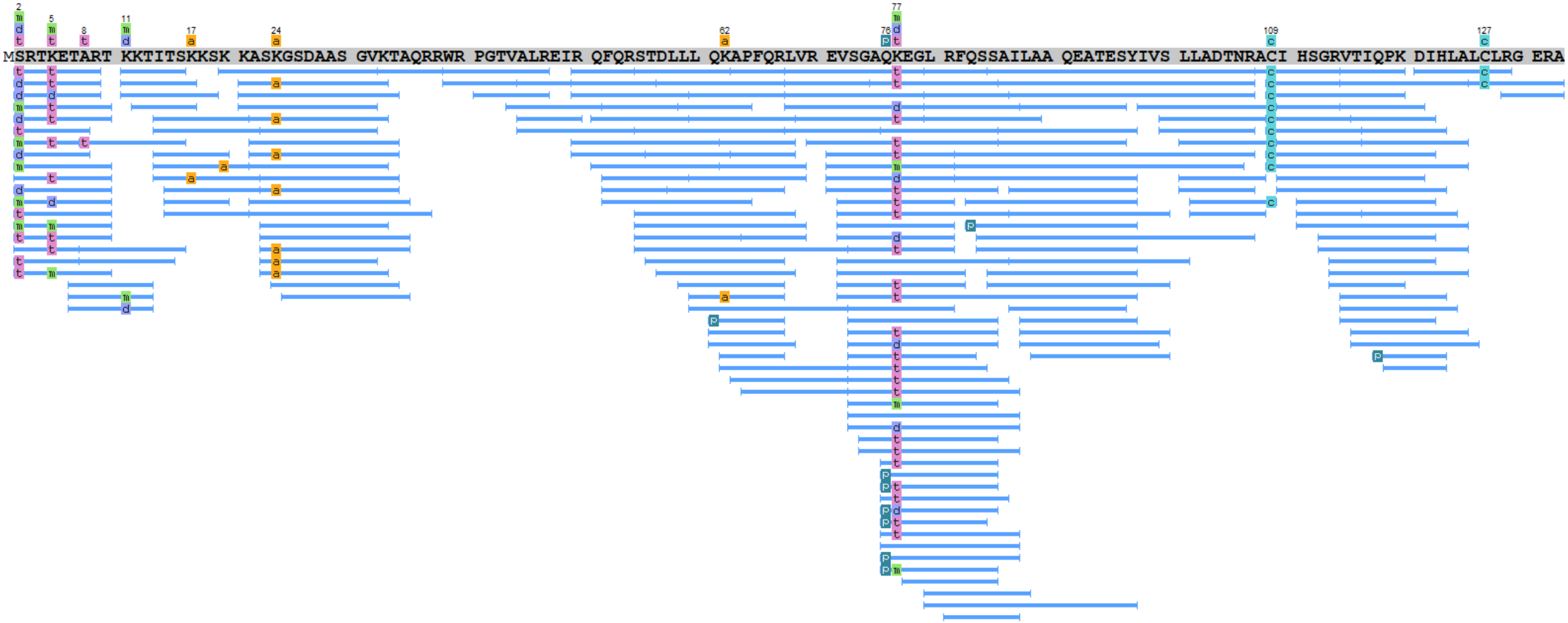
Nucleosomes and Histones in *Trypanosoma brucei*





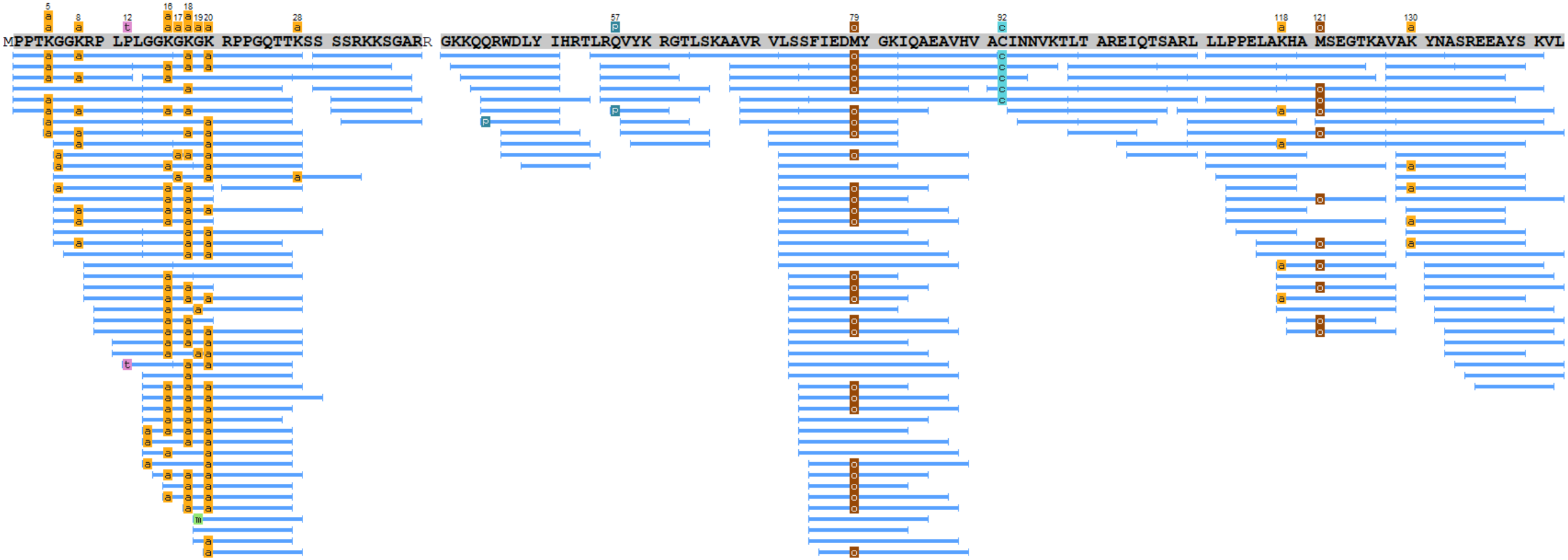
PTM Analysis of Start Site Histones

H3

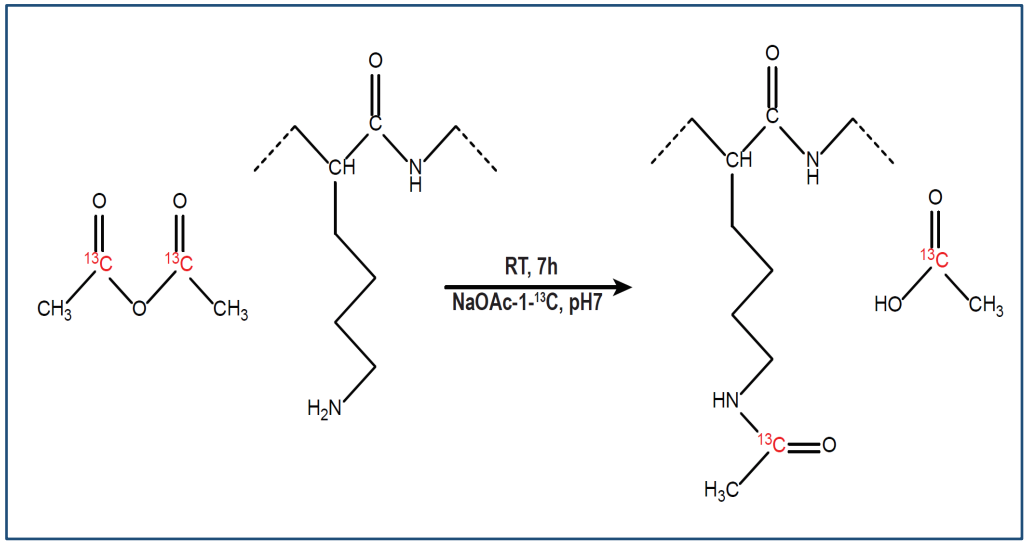


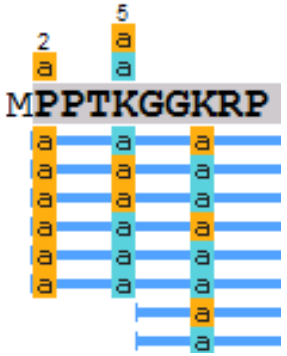
PTM Analysis of Start Site Histones

H2B.V

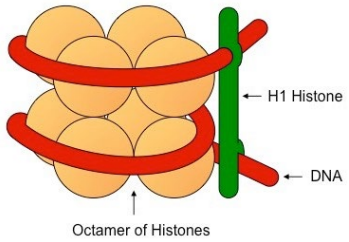


Measuring Site-specific Acetylation Degrees

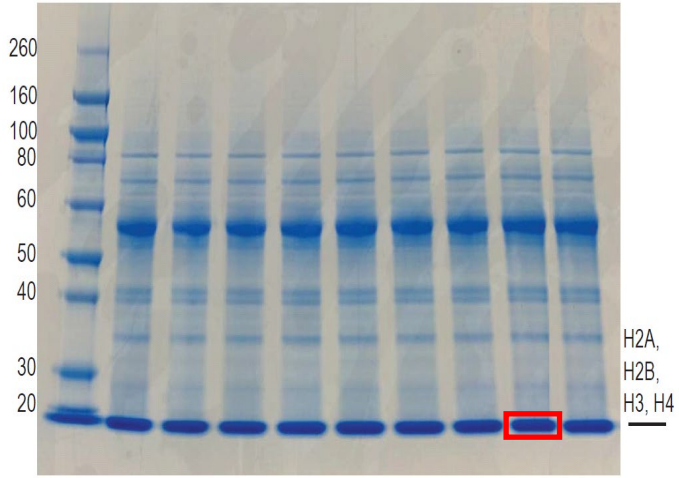




a Acetylation 13C1 (+43.01)
a Acetylation (K) (+42.01)



histone enriched fraction

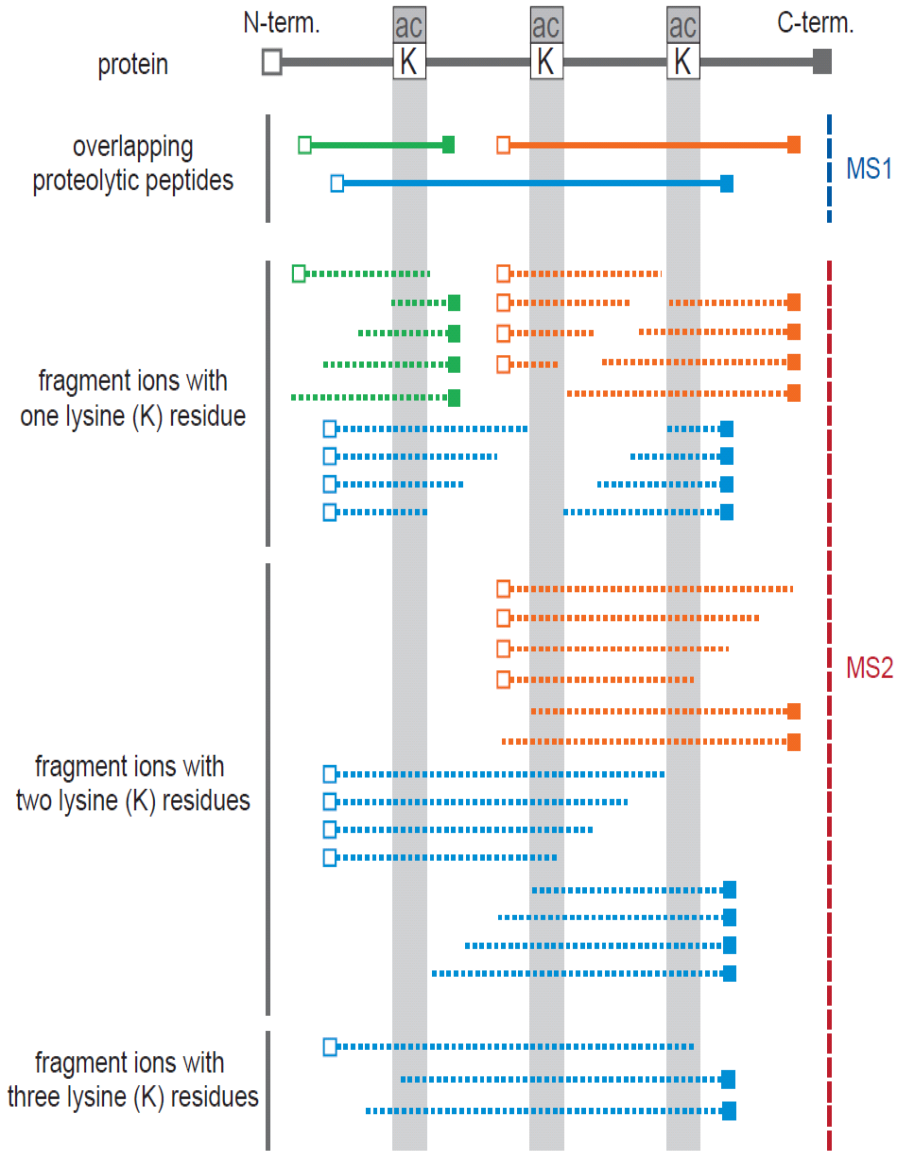


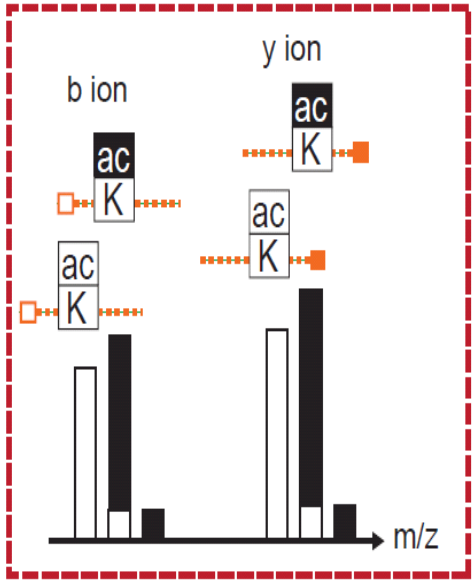
in-gel acetylation with ^{13}C-Ac_2O

elastase thermolysin papain

ElBashir R, Vanselow JT, Kraus A, Janzen CJ, Siegel NT, Schlosser A. (2015) *Anal Chem* 87:9939–9945.

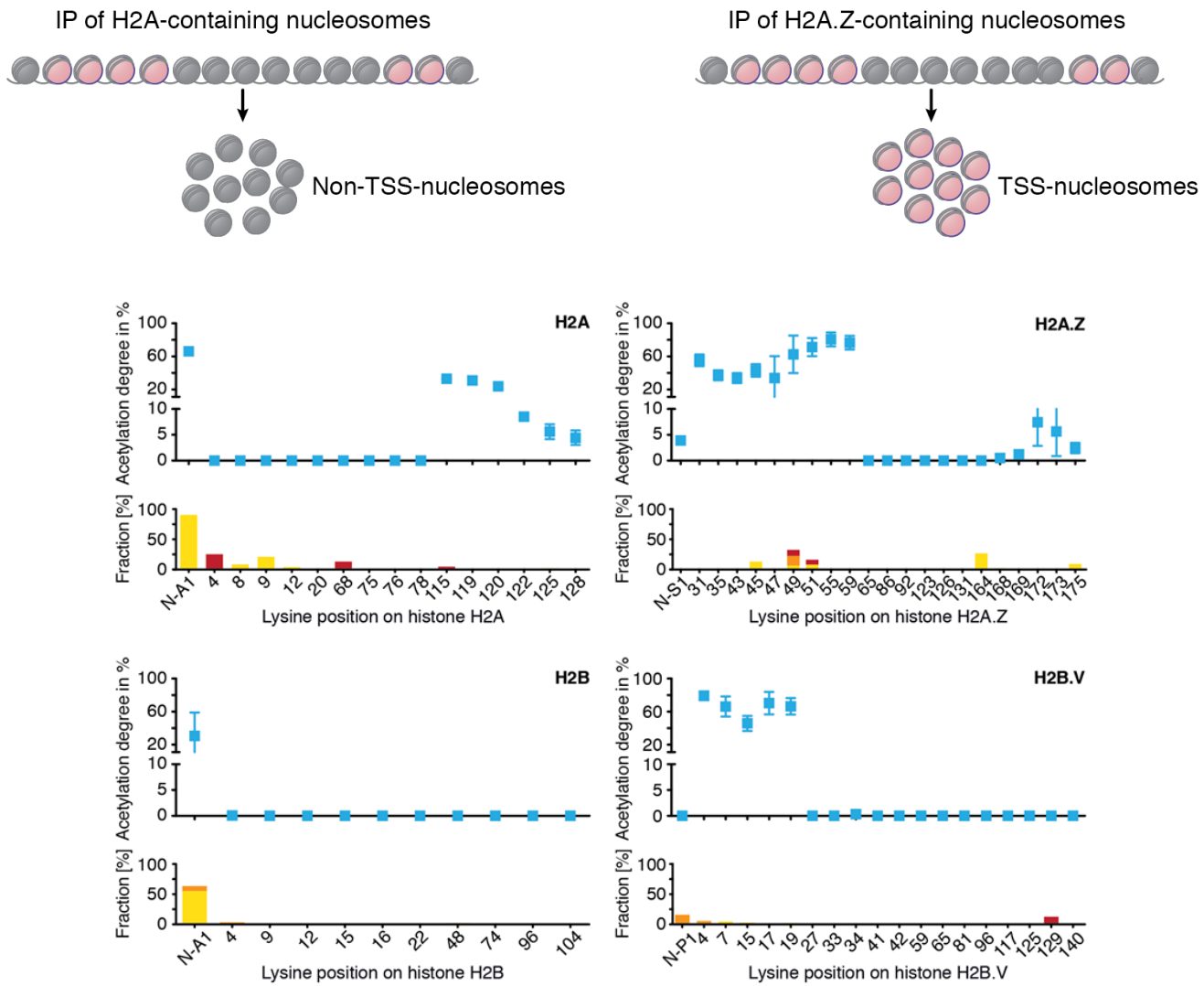
Fragment Ion Patchwork Quantification





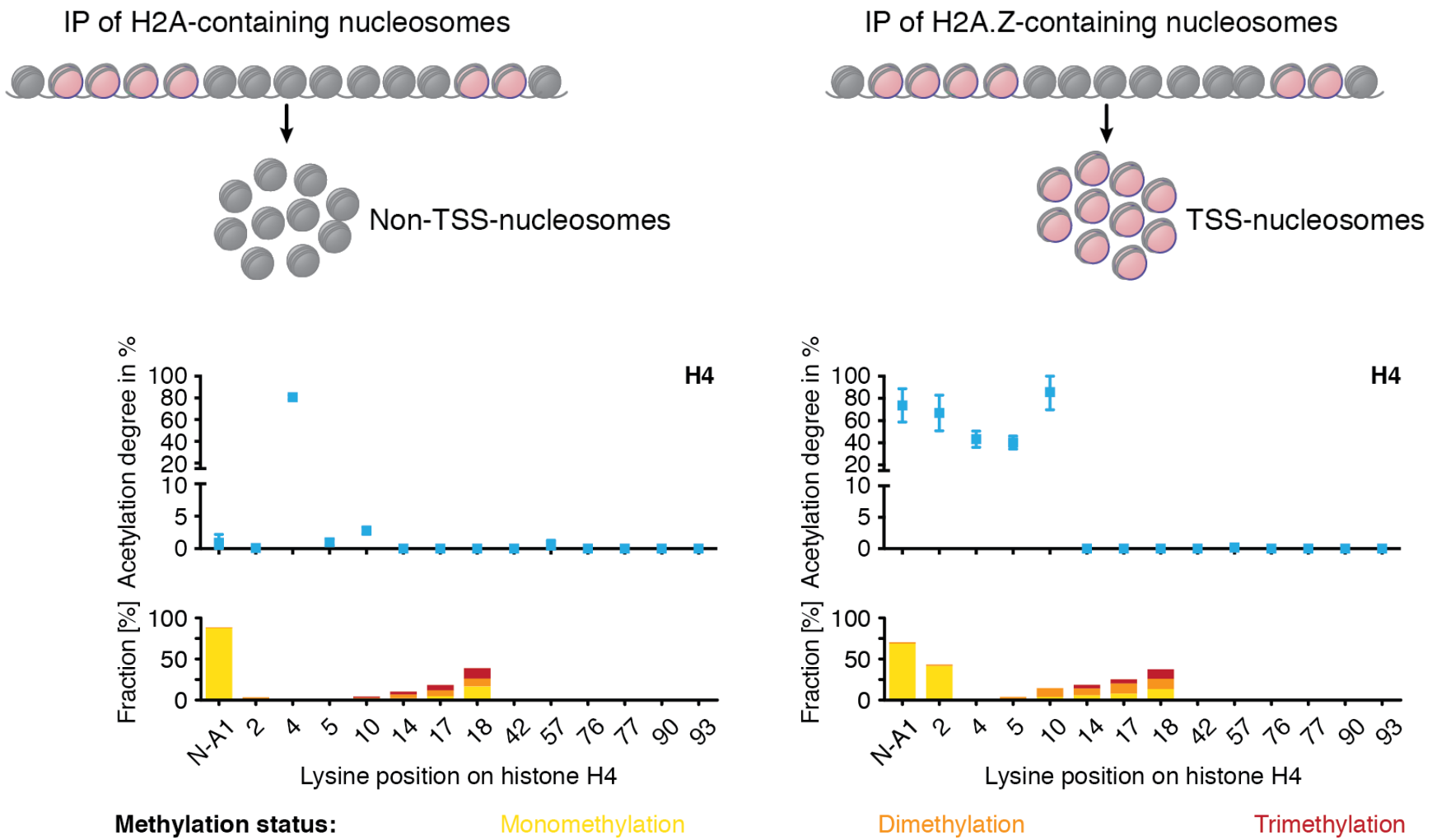
ElBashir R, Vanselow JT, Kraus A, Janzen CJ, Siegel NT, Schlosser A. (2015) *Anal Chem* 87:9939–9945.

Analyzing Acetylation Patterns in Start Site Nucleosomes



Kraus, A.J., Vanselow, J.T., Lamer, S. *et al.* Distinct roles for H4 and H2A.Z acetylation in RNA transcription in African trypanosomes. *Nat Commun* **11,** 1498 (2020)

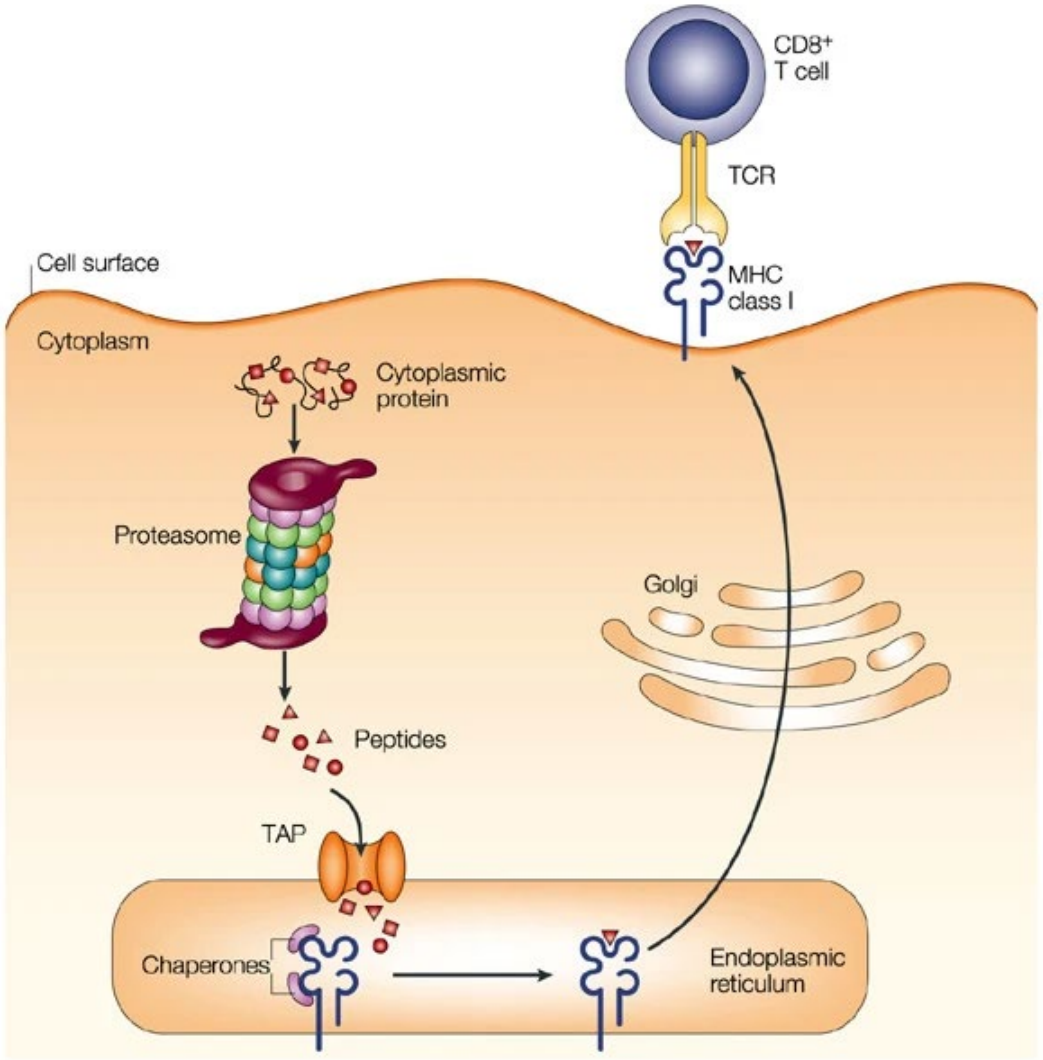
Analyzing Acetylation Patterns in Start Site Nucleosomes



Kraus, A.J., Vanselow, J.T., Lamer, S. *et al.* Distinct roles for H4 and H2A.Z acetylation in RNA transcription in African trypanosomes. *Nat Commun* **11,** 1498 (2020)

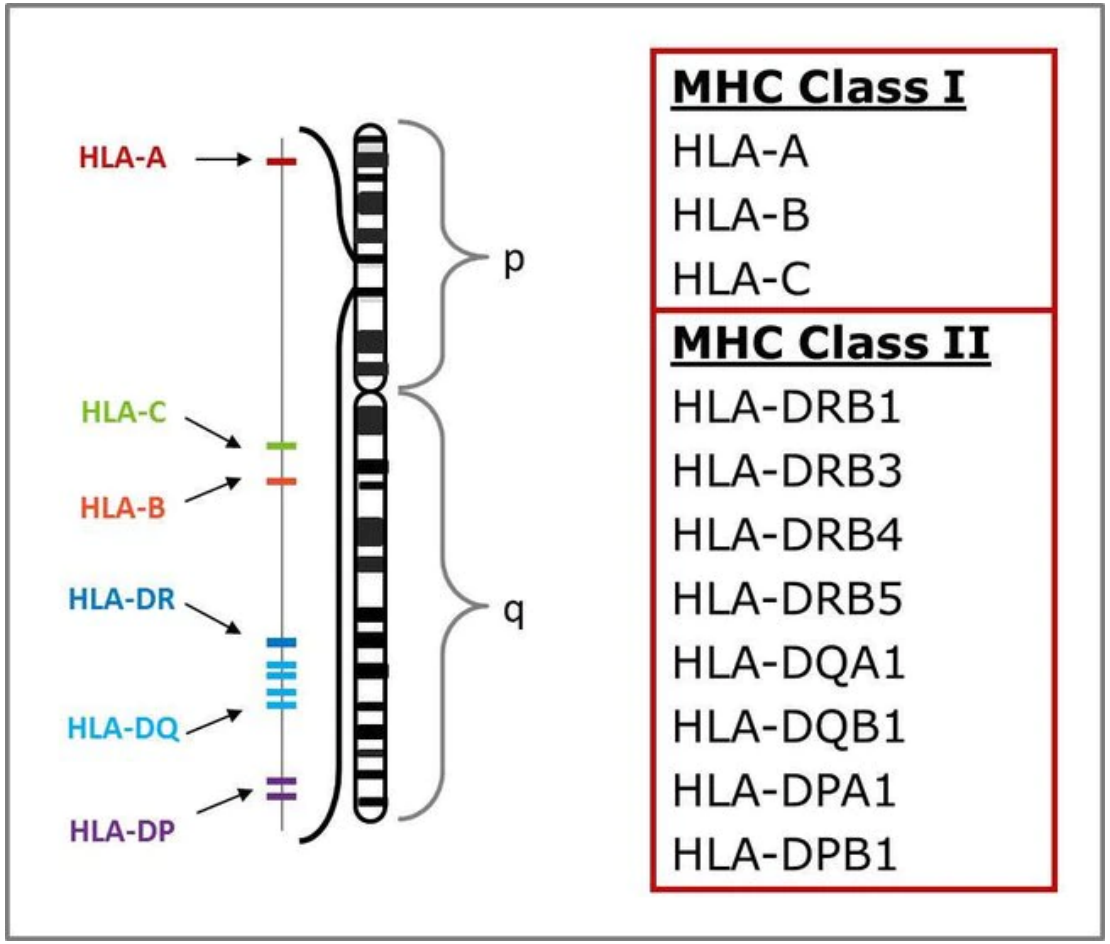
Immunopeptidomics

MHC class I antigen presentation



Nature Reviews Immunology volume 3, pages 952–961 (2003)

HLA Polymorphism



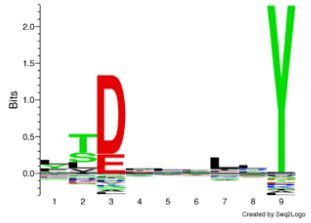
Example: HeLa cell line

HLA-A*68:02
HLA-A*03:19
HLA-B*15:03
HLA-B*15:03
HLA-C*12:03
HLA-C*12:03

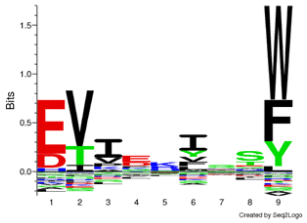
HLA Peptide Motifs

HLA-A1 Supertype

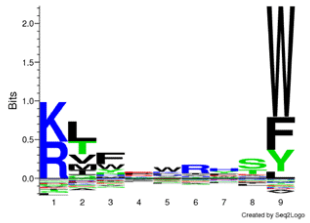
A*01:01



A*25:01

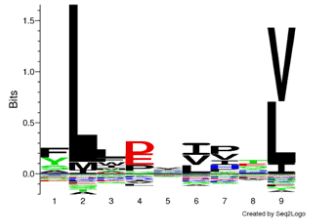


A*32:01

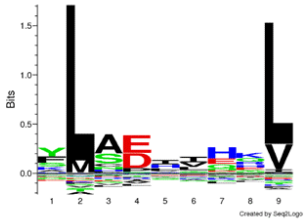


HLA-A2 Supertype

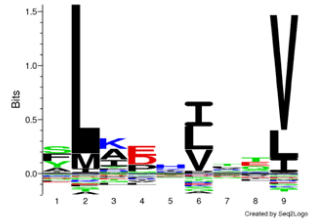
A*02:01



A*02:02

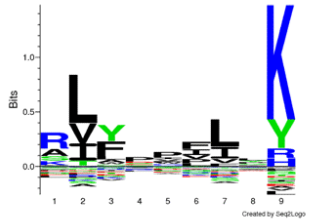


A*02:03

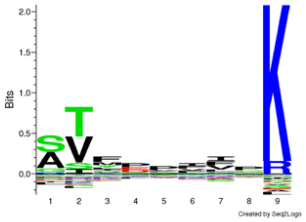


HLA-A3 Supertype

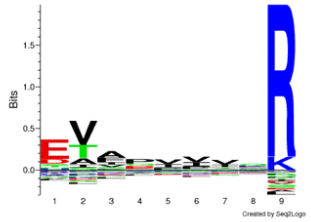
A*03:01



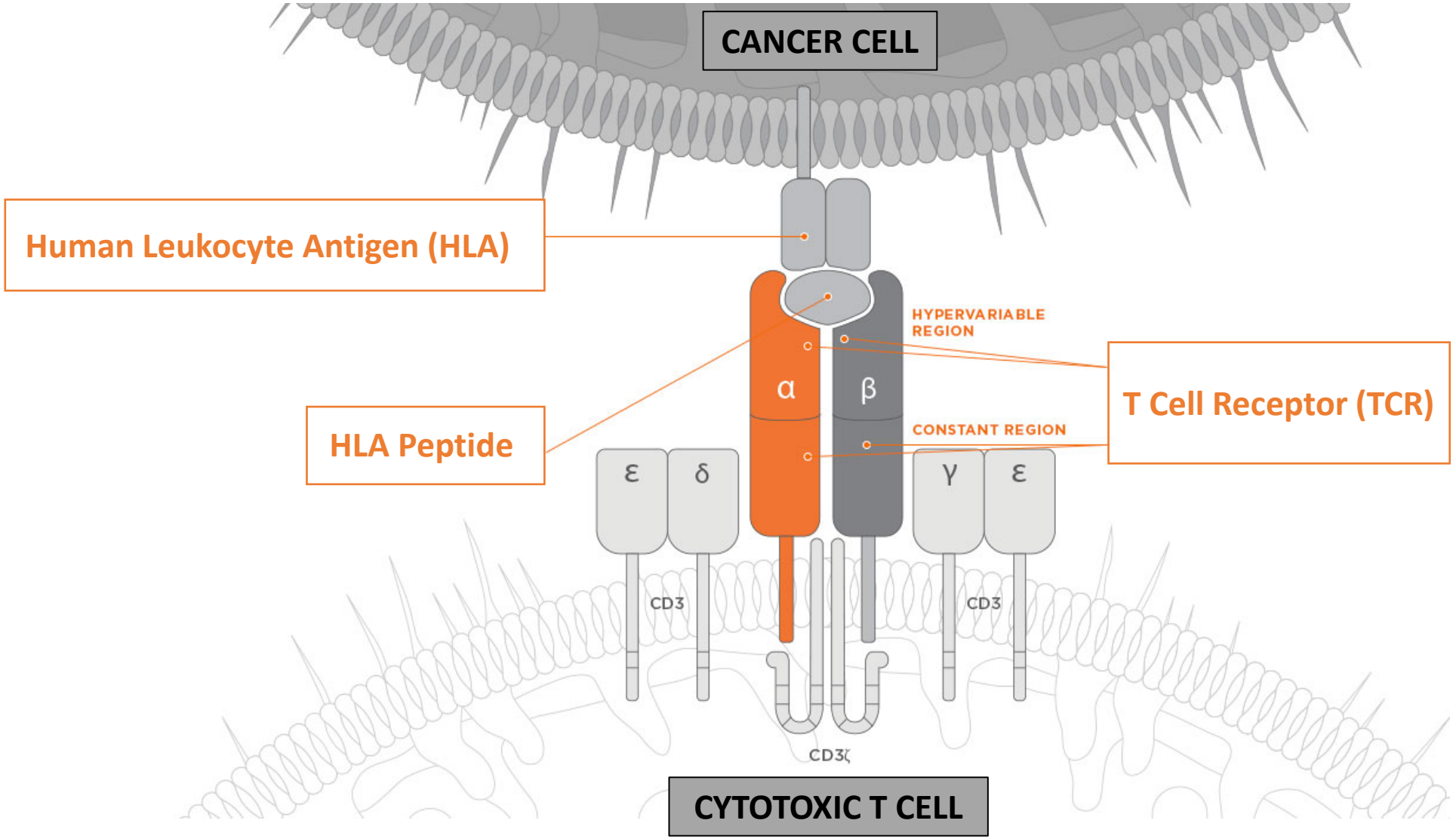
A*11:02



A*68:01

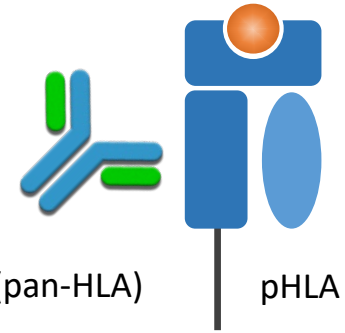


HLA-I Peptides are the Key Players for T Cell Response



modified from junotherapeutics.com

Mass Spectrometry-based Immunopeptidomics



W6/32 (pan-HLA)

pHLA complex

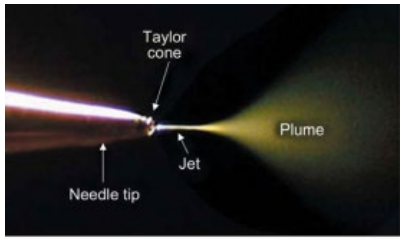
SPE (C-18) or
ultrafiltration

HLA peptides

HLA peptides

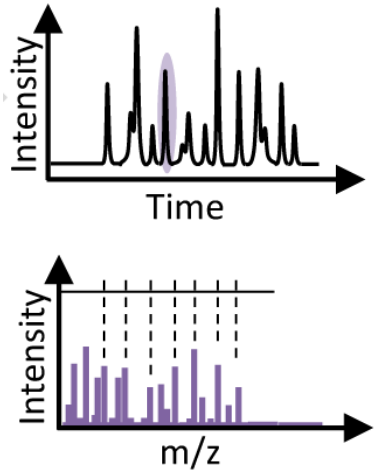
Reversed phase (C-18)

nanoHPLC



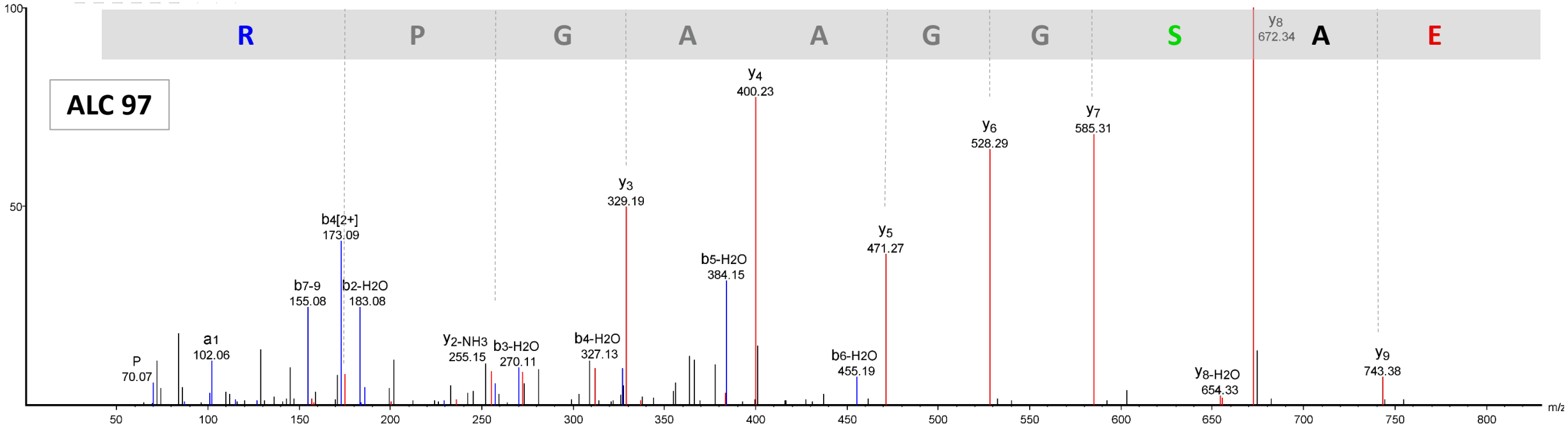
ESI

MS (Orbitrap)



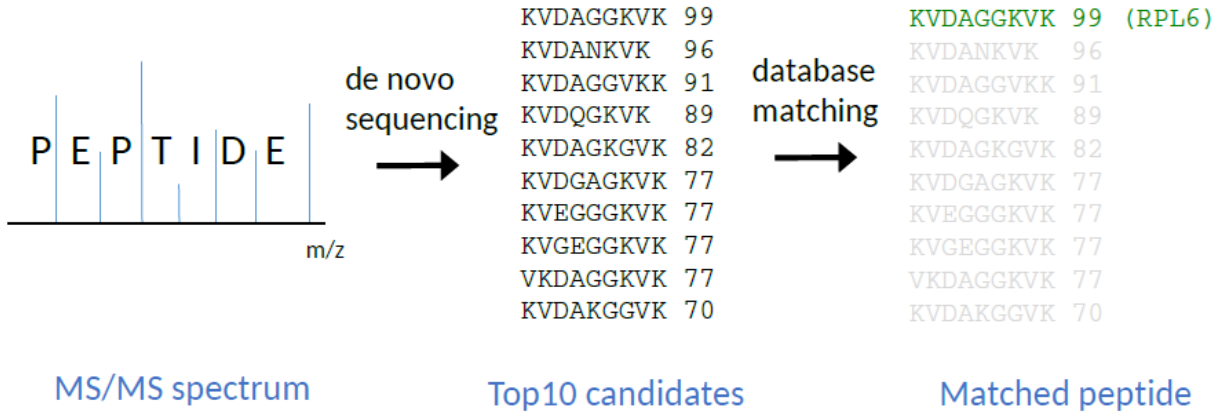
MS/MS data

Fragment Ion Spectrum of HLA-I Peptide



De novo Sequencing-based Approach – Peptide-PRISM

Database matching of *de novo* sequencing candidates

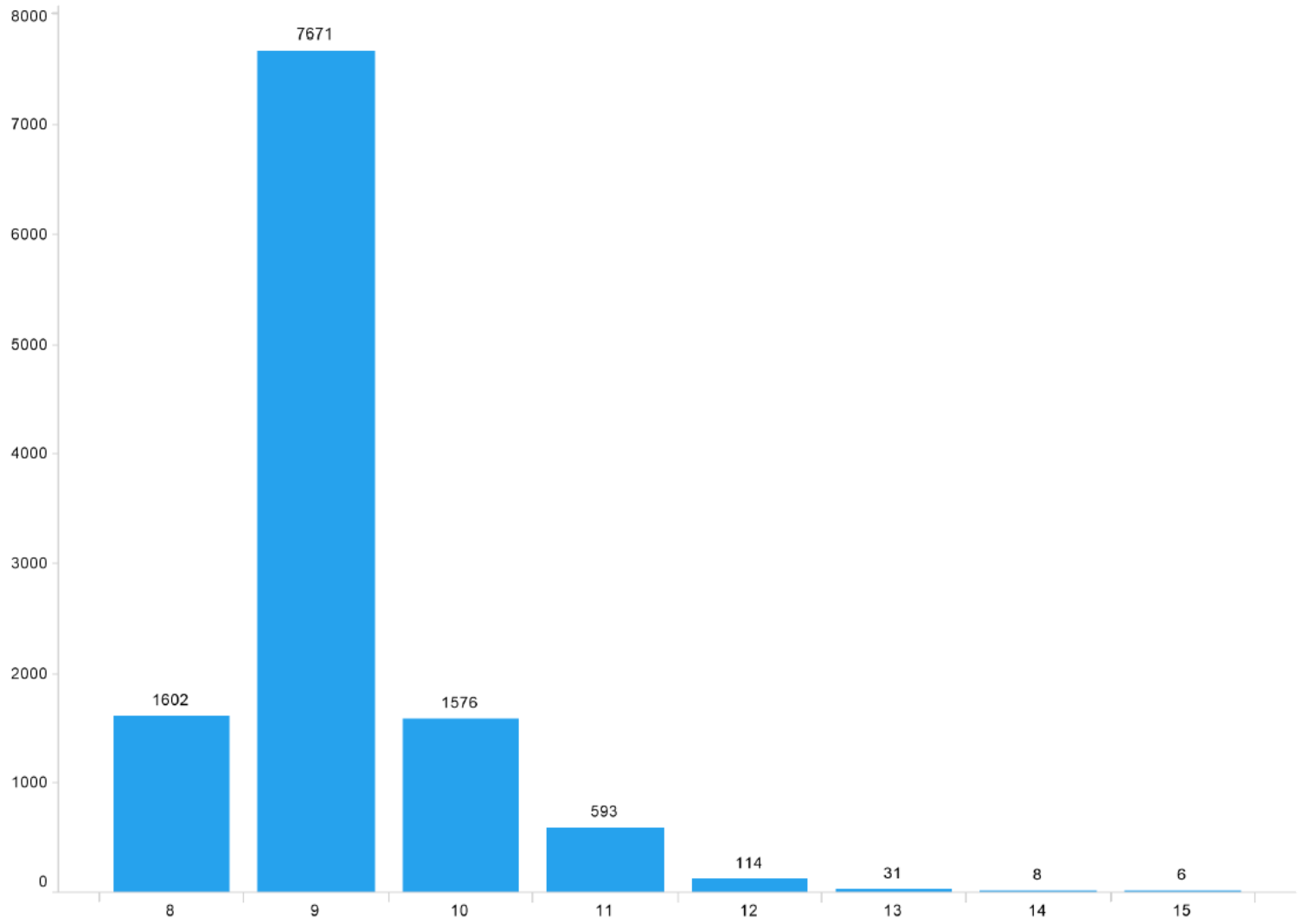


MS/MS spectrum | Top10 candidates | Matched peptide

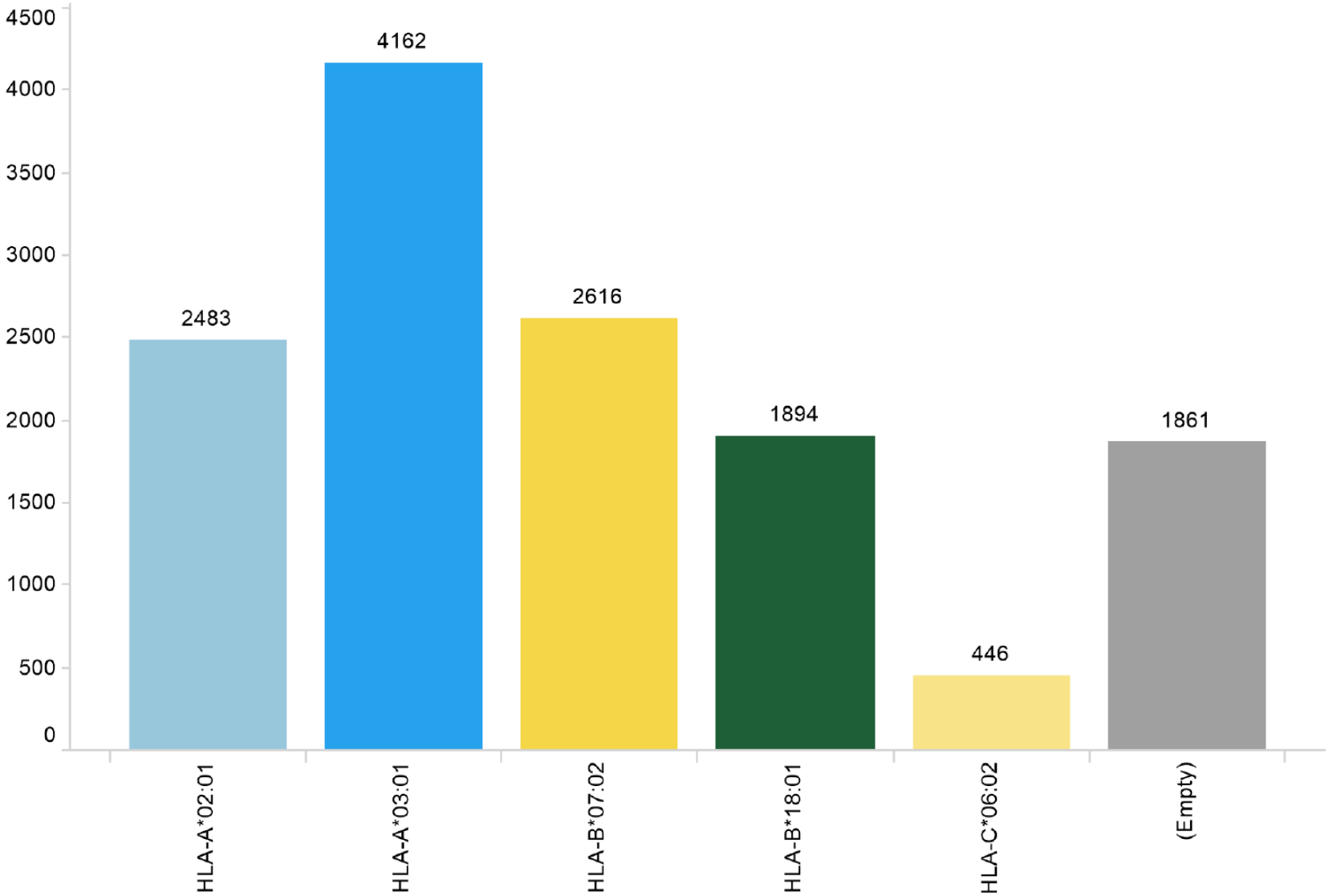
FDR filtering using stratified mixture models

All matched peptides | Target & decoy distributions | Identified peptides

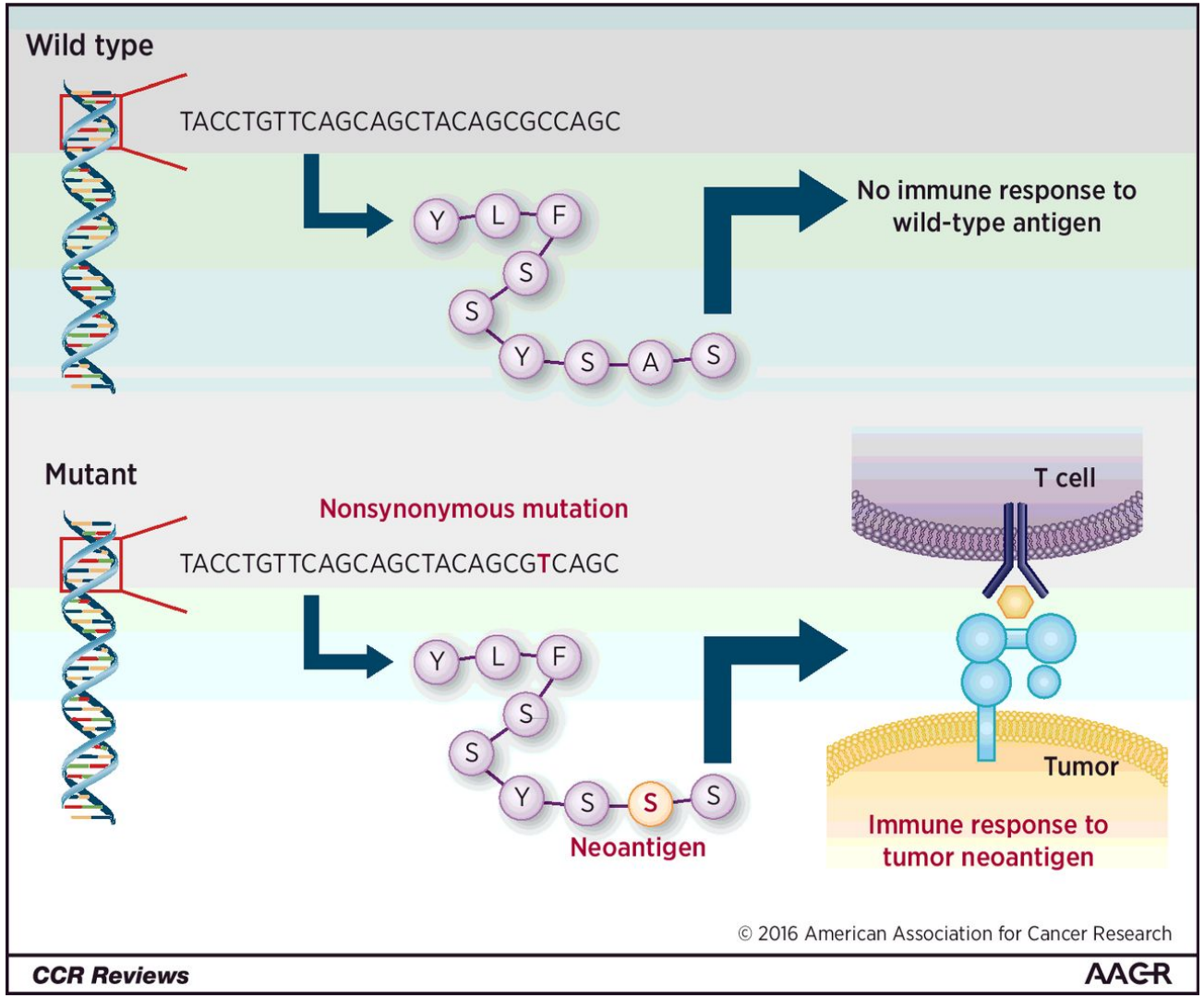
HLA Class I Length Distribution



HLA Class I Binding Prediction (NetMHC)

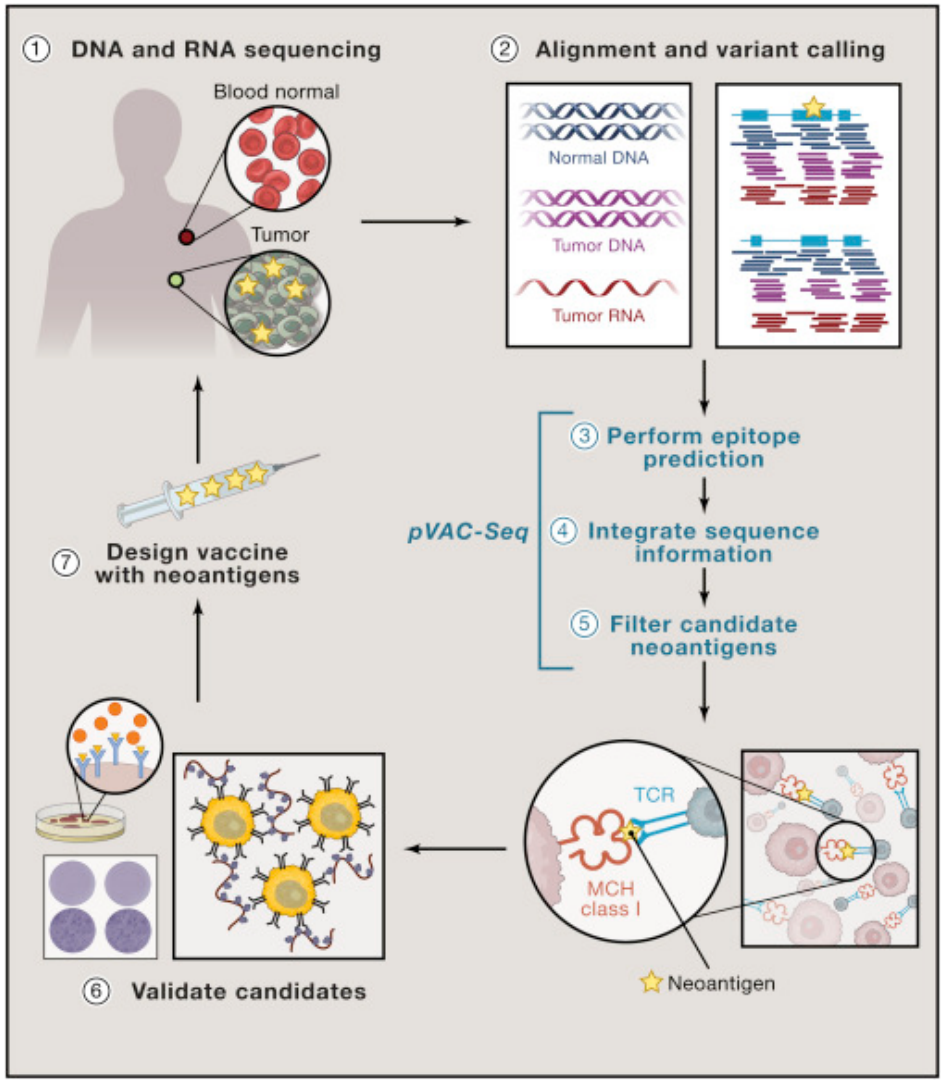


Tumor-specific Mutations can Generate Neoantigens



Braun et al. **2016** *Clin Cancer Res* 22:5642-5650

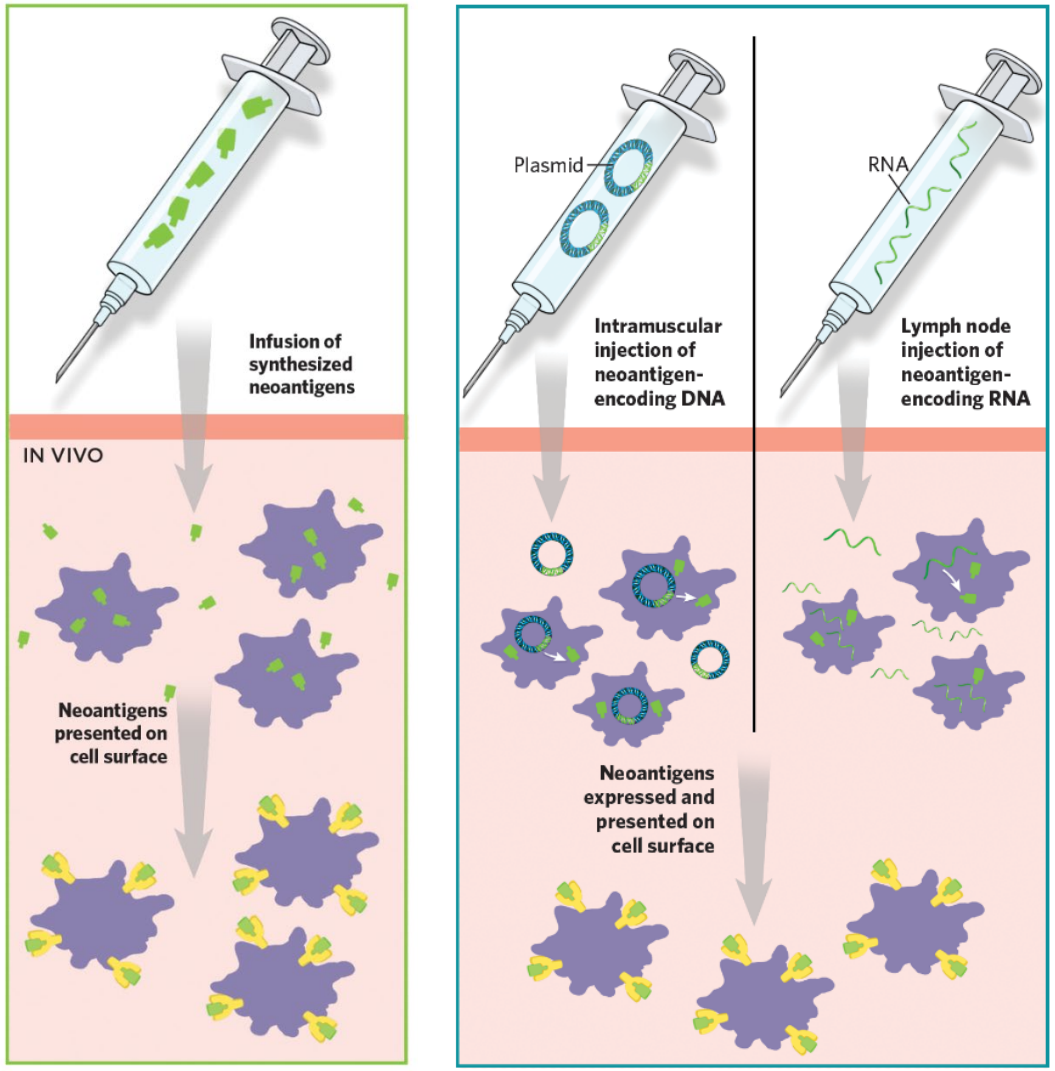
Strategy for Personalized Cancer Immunotherapy with Neoantigens

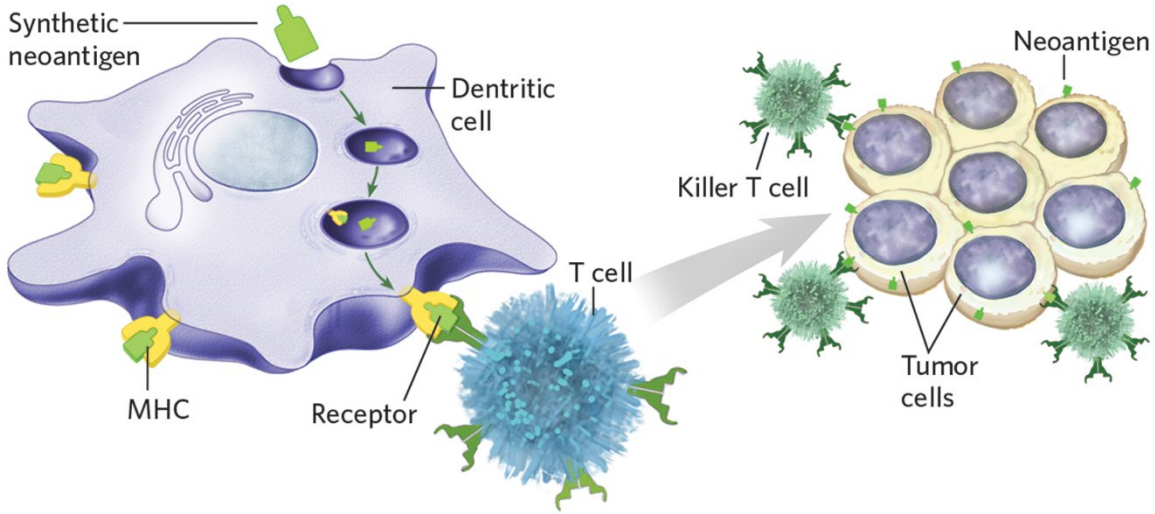


Liu et al. **2017** *Cell* 168:600-612

Strategies for HLA Peptide-based Immunotherapies

Vaccination with peptides or DNA/RNA (BioNTec)

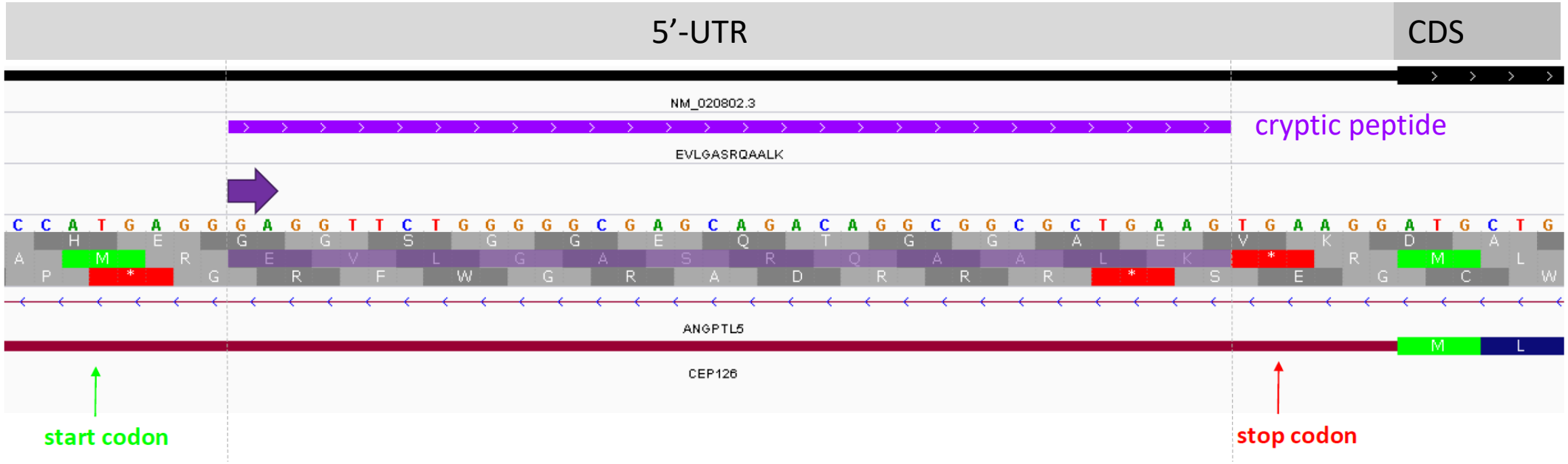




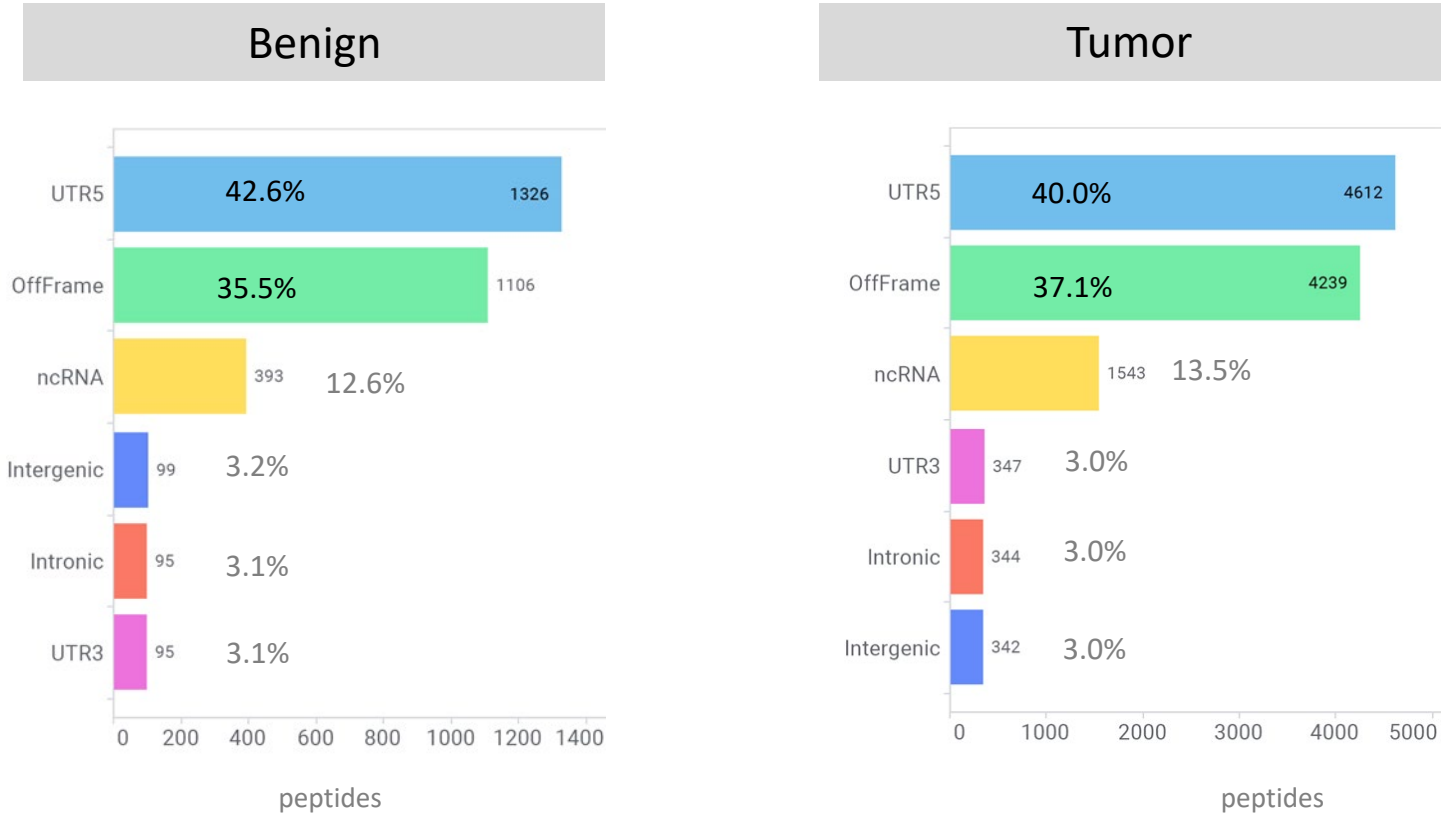
Source: https://www.the-scientist.com/features/personalized-cancer-vaccines-in-clinical-trials-66075

Cryptic HLA-I peptides

Cryptic HLA peptides derive from allegedly non-coding regions, such as 5’- and 3’-UTRs or introns, or from coding sequences (CDS) in non-canonical reading frames.



Origin of cryptic peptides

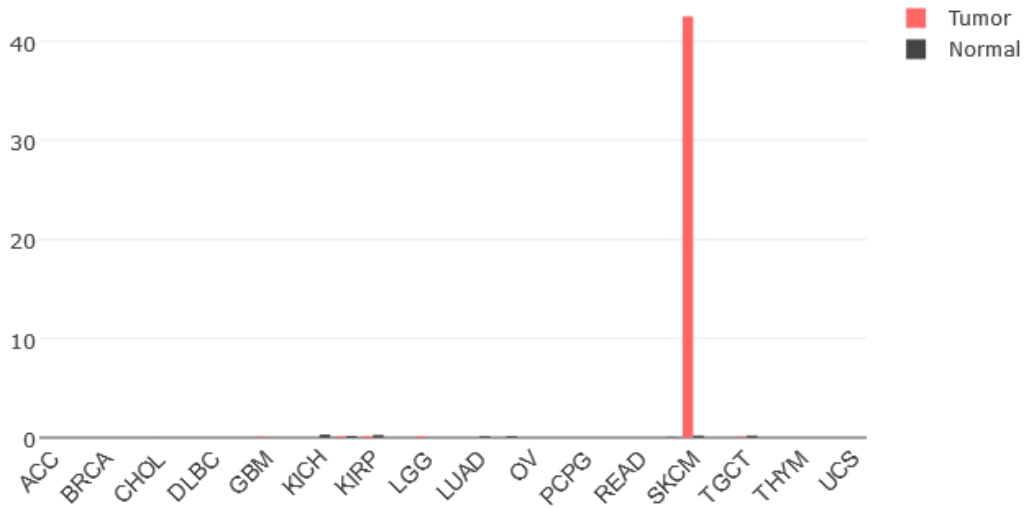


Tumor-specific, recurrent, cryptic peptides (TURCs)

- 40% of the tumor-specific cryptic peptides are identified in 2 or more patients
- One source for TURCs is aberrant, tumor-specific transcription

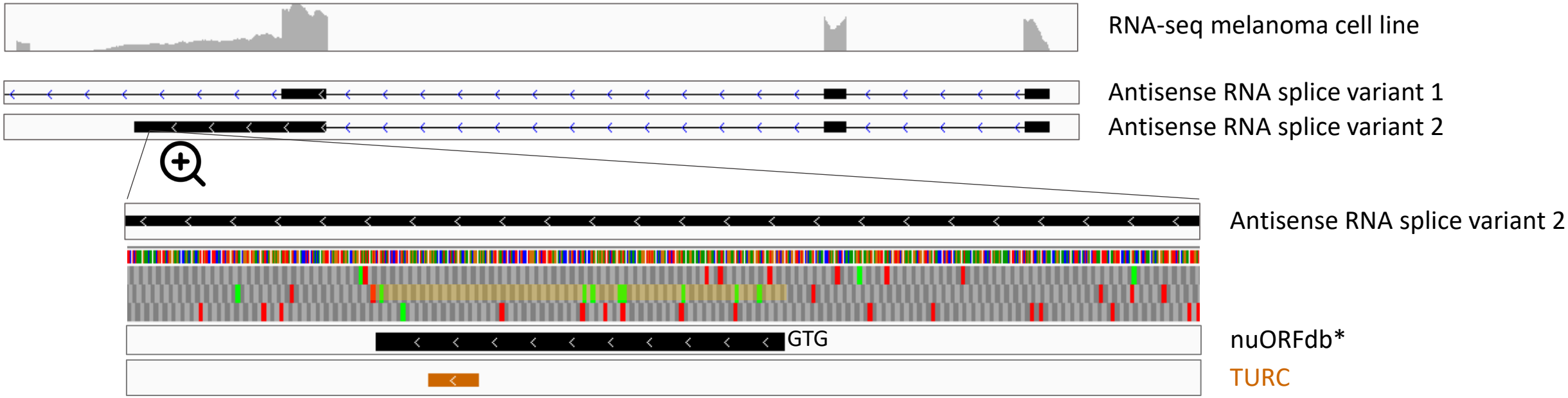
Example: antisense RNA, aberrantly expressed in melanoma

- 5 different cryptic peptides
- in 11 melanoma patients/patient cell lines



From GEPIA (http://gepia.cancer-pku.cn/)

TURC from an antisense RNA



- Cryptic HLA peptides from ncRNAs are recurrently presented on tumor cells, allowing the immune system to recognize aberrant expression of ncRNAs.
- These cryptic peptides might represent a new class of non-private neoantigens exploitable for cancer immunotherapy.