

What is Proteomics?

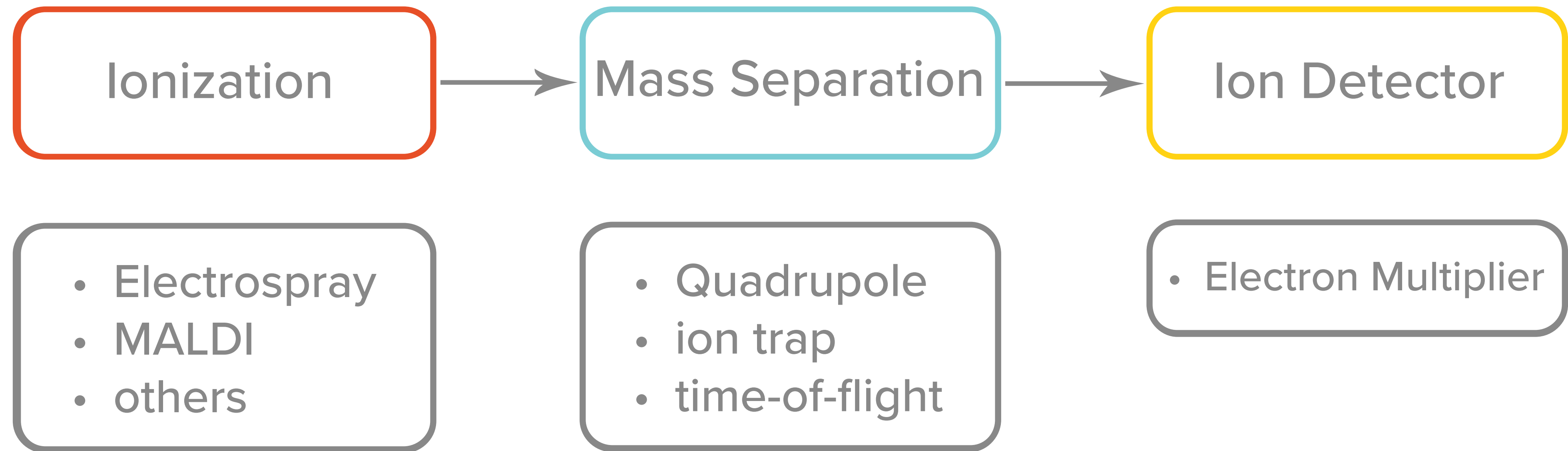
“Proteomics includes not only the identification and quantification of proteins, but also the determination of their localization, modifications, interactions, activities, and, ultimately, their function.”

Stan Fields, *Science*, 2001.

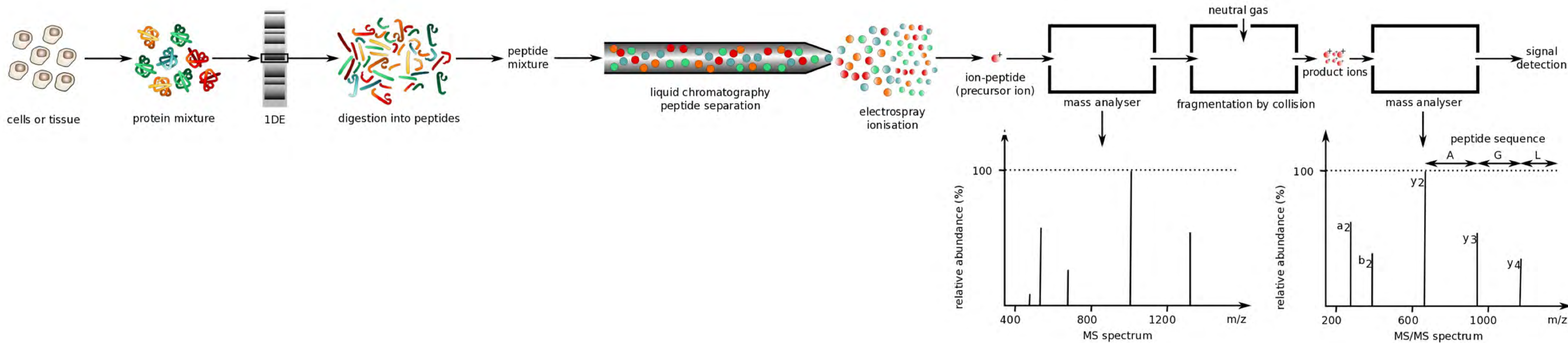
What is Mass Spectrometry?

“Mass spectrometry is the art of measuring atoms and molecules to determine their molecular weight. Such mass or weight information is sometimes sufficient, frequently necessary, and always useful in determining the identity of a species. To practice this art one puts charge on the molecules of interest, i.e., the analyte, then measures how the trajectories of the resulting ions respond in vacuum to various combinations of electric and magnetic fields.”

John B. Fenn – Nobel Prize in Chemistry 2002



A mass spectrometer measures mass to charge ratio or m/z

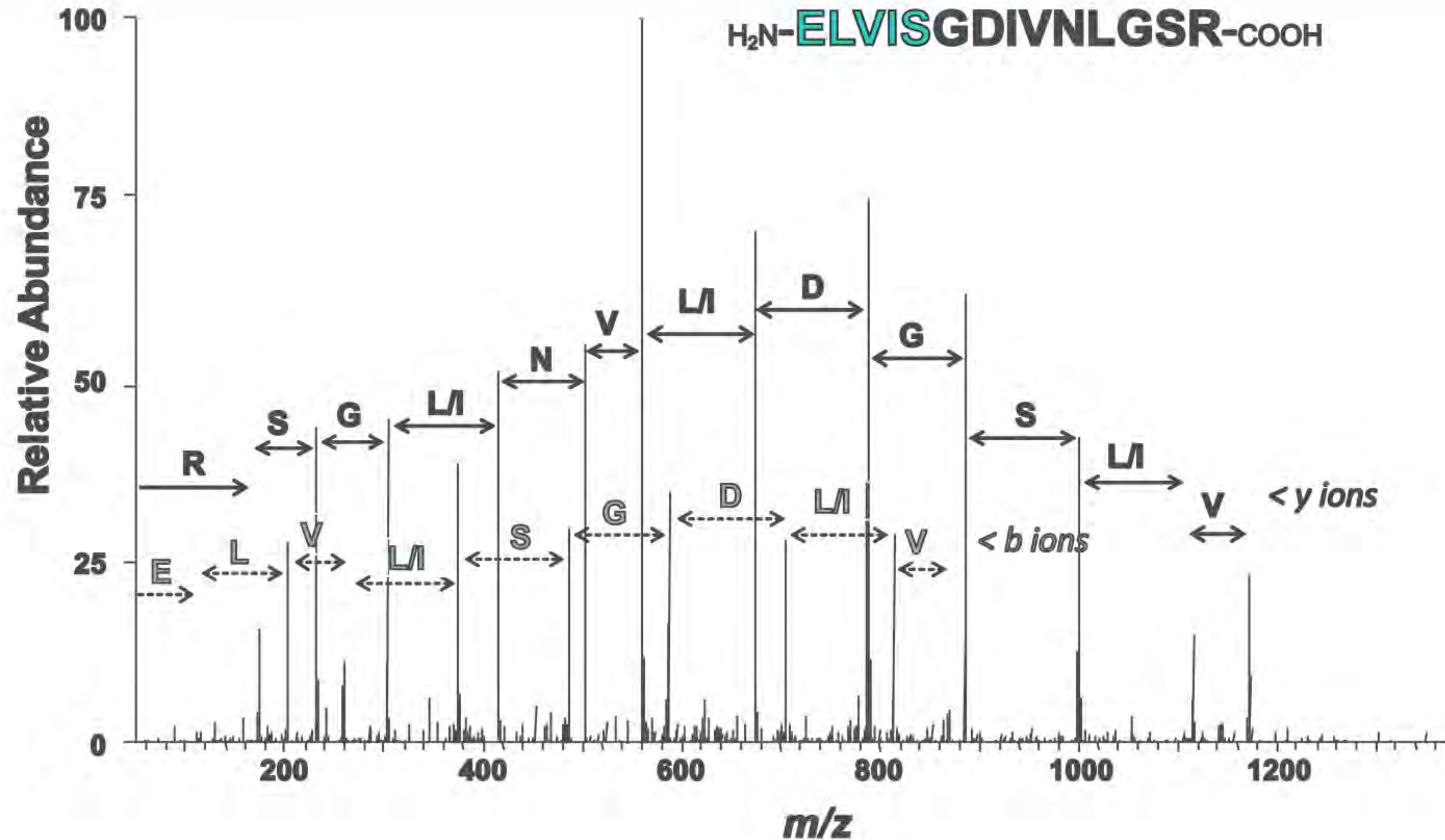


MS/MS fragmentation
along
the peptide backbone
creates
b and *y* ions



| <i>b</i> | <i>y</i> |
|--------------|----------------|
| +E | LVISGDIVNLGSR+ |
| +EL | VISGDIVNLGSR+ |
| +ELV | ISDIVNLGSR+ |
| +ELVI | SIVNLGSR+ |
| +ELVIS | IVNLGSR+ |
| +ELVISG | VNLGSR+ |
| +ELVISGD | NLGSR+ |
| +ELVISGDI | LGSR+ |
| +ELVISGDIV | GSR+ |
| +ELVISGDIVN | GR+ |
| +ELVISGDIVNL | R+ |

The fragment ladder allows assignment



Peptide/Protein identification and is accomplished computationally by searching databases of protein sequences or spectra

Search against a database containing all amino acid sequences assumed to be present in the sample or a spectral library of previously observed spectra for peptide/proteins

If not in the database you will not see it in your results.

Multiple software programs and algorithms exist

Sometimes multiple sets of software are run and combined

End results are Peptide Specific Matches (PSMs) which can then be matched with proteins. All steps come with a variety of statistical parameters and probabilities usually unique to the method...

Thirteen human proteins contain **ELVIS** peptide sequence.

- 1) Short peptides are seldom unique,
- 2) One peptide is seldom enough for protein identification.

| Protein AC | Protein ID | Protein Name | Length | Organism | Match Range |
|---|--|---|--------|----------------------|-----------------------------------|
| P42126 <small>/ProClass UniProtKB/Swiss-Prot</small> | ECI1_HUMAN <small>UniProtKB/Swiss-Prot</small> | Enoyl-CoA delta isomerase 1, mitochondrial precursor | 302 | Homo sapiens (Human) | LEFLT ELVIS LEKLE 77-81 |
| A0AUX1 <small>/ProClass UniProtKB/TrEMBL</small> | A0AUX1_HUMAN <small>UniProtKB/TrEMBL</small> | ZNF33A protein | 811 | Homo sapiens (Human) | FNTV SELVIS SKINYL 154-158 |
| A6NCM1 <small>/ProClass UniProtKB/Swiss-Prot</small> | IQCAL_HUMAN <small>UniProtKB/Swiss-Prot</small> | IQ and AAA domain-containing protein 1-like | 817 | Homo sapiens (Human) | ESLY EELVIS GLLRK 492-496 |
| B3KSV8 <small>/ProClass UniProtKB/TrEMBL</small> | B3KSV8_HUMAN <small>UniProtKB/TrEMBL</small> | cDNA FLJ37145 fis, clone BRACE2024781, highly similar to Synapsin-2 | 514 | Homo sapiens (Human) | RQLIT ELVIS SKMNQL 340-344 |
| B4DH52 <small>/ProClass UniProtKB/TrEMBL</small> | B4DH52_HUMAN <small>UniProtKB/TrEMBL</small> | cDNA FLJ55916, highly similar to General transcription factor II-I | 993 | Homo sapiens (Human) | VVKK PELVIS YLPPG 642-646 |
| B4DH82 <small>/ProClass UniProtKB/TrEMBL</small> | B4DH82_HUMAN <small>UniProtKB/TrEMBL</small> | cDNA FLJ54437, highly similar to RRP5 protein homolog (Fragment) | 1299 | Homo sapiens (Human) | EVNE LELVIS LPNGL 97-101 |
| Q05D89 <small>/ProClass UniProtKB/TrEMBL</small> | Q05D89_HUMAN <small>UniProtKB/TrEMBL</small> | ZNF33A protein (Fragment) | 173 | Homo sapiens (Human) | FNTV SELVIS SKINYL 154-158 |
| Q3B799 <small>/ProClass UniProtKB/TrEMBL</small> | Q3B799_HUMAN <small>UniProtKB/TrEMBL</small> | ZNF33B protein (Fragment) | 606 | Homo sapiens (Human) | FNTV SELVIS SKINYL 154-158 |
| Q499G6 <small>/ProClass UniProtKB/TrEMBL</small> | Q499G6_HUMAN <small>UniProtKB/TrEMBL</small> | General transcription factor II, i | 976 | Homo sapiens (Human) | VVKK PELVIS YLPPG 625-629 |
| Q59GM1 <small>/ProClass UniProtKB/TrEMBL</small> | Q59GM1_HUMAN <small>UniProtKB/TrEMBL</small> | Synapsin II isoform IIb variant (Fragment) | 514 | Homo sapiens (Human) | RQLIT ELVIS SKMNQL 444-448 |
| Q66T69 <small>/ProClass UniProtKB/TrEMBL</small> | Q66T69_HUMAN <small>UniProtKB/TrEMBL</small> | Eyes absent homolog | 508 | Homo sapiens (Human) | MV ELVIS PSLTV 3-7 |

PSMs

| Peptide | Charge | Mods | MinFDR | Count | Count | Sample | Protein | Gene |
|------------------------------|--------|-------------|--------|-------|-------|---------------------|-------------------------------|----------|
| ATAFNEQVDKFFPLIEVNK | 3;4;5 | [0:+144.102 | 0 | 50 | 0 | 13-1489-01A:42-2590 | NP_002936.1;P27694 | RPA1 |
| LHDNQNGWSGDSAPVELILSDETLPAPE | 3;4;5 | [0:+144.102 | 0 | 58 | 0 | 13-1489-01A:42-2590 | M0R009;NP_570602.2;P04217;P0 | A1BG |
| SLSLCNMFLDEMAK | 2;3 | [0:+144.102 | 0 | 78 | 0 | 13-1489-01A:42-2590 | NP_038464.1;NP_995314.1;Q9Y2 | NCKAP1 |
| TNPFPLLEDEDDLFTDQK | 3 | [0:+144.102 | 0 | 34 | 0 | 13-1489-01A:42-2590 | E7ESD2;F8W7U3;J3KP36;NP_001 | FAM21A;I |
| LFPHEVADVEPVLDLVTIQNPK | 3;4;5 | [0:+144.102 | 0 | 68 | 0 | 13-1489-01A:42-2590 | J3KR97;NP_005984.3;Q9BTW9;Q9 | TBCD |
| AELGALPDDFIDSLEK | 2;3 | [0:+144.102 | 0 | 36 | 0 | 13-1489-01A:42-2590 | E9PCB6;NP_065777.1;Q9BYT8;XFN | NLN |
| FGDTSLQEVINVESLVR | 2;3 | [0:+144.102 | 0 | 25 | 0 | 13-1489-01A:42-2590 | B8ZZB2;F5H860;NP_001127696.1 | INPP4A |
| GIEELFLDLCKR | 2;3;4 | [0:+144.102 | 0 | 78 | 0 | 13-1489-01A:42-2590 | NP_055814.1;Q9UL25 | RAB21 |
| LMSSNSTDLPLNIECFMNDK | 2;3;4 | [0:+144.102 | 0 | 31 | 0 | 13-1489-01A:42-2590 | B4DY72;B4DYH1;NP_001273432. | HSPH1 |

Protein/Gene Identification

| A | B | C | D | CH | CI | CJ | CK | CM | CN | CO | CP | CQ | CR |
|-------------|--|--|--|-----------------|-------------------|-------------------|---------------------------------------|-----------|----------------|--|----|----|----|
| Gene | 13-1489-01A:42-2590-01A:36-2529-01A:POOL Spectral Counts | 13-1489-01A:42-2590-01A:36-2529-01A:POOL Distinct Peptides | 13-1489-01A:42-2590-01A:36-2529-01A:POOL Unshared Peptides | Spectral Counts | Distinct Peptides | Unshared Peptides | Description | Chr# | Locus | Proteins | | | |
| RP9 | 1 | 1 | 1 | 8 | 3 | 3 | retinitis pigmentosa 9 (autosomal don | 7 | 7p14.3 | C9J6V2;NP_976033.1;Q8TA86 | | | |
| RPA1 | 36 | 21 | 21 | 929 | 42 | 41 | replication protein A1, 70kDa | 17 | 17p13.3 | I3L2M5;I3L4R8;I3L524;NP_002936.1;P27694 | | | |
| RPA2 | 21 | 11 | 11 | 486 | 16 | 16 | replication protein A2, 32kDa | 1 | 1p35 | NP_001273005.1;NP_002937.1;P15927;P15927 | | | |
| RPA3 | 21 | 7 | 7 | 424 | 9 | 9 | replication protein A3, 14kDa | 7 | 7p22 | B5MC59;NP_002938.1;P35244 | | | |
| RPAP1 | 1 | 1 | 1 | 31 | 12 | 12 | RNA polymerase II associated proteir | 15 | 15q15.1 | H3BPM3;H3BPY8;H3BRE8;H3BTJ6;NP_056355 | | | |
| RPAP2 | 1 | 1 | 1 | 8 | 5 | 5 | RNA polymerase II associated proteir | 1 | 1p22.1 | NP_079089.2;Q8IXW5;Q8IXW5-2;XP_00527128 | | | |
| RPAP3 | 7 | 5 | 5 | 193 | 31 | 31 | RNA polymerase II associated proteir | 12 | 12q13.11 | NP_001139547.1;NP_001139548.1;NP_078880. | | | |
| RPE | 9 | 6 | 6 | 196 | 15 | 15 | ribulose-5-phosphate-3-epimerase | 2 | 2q32-q33.3 | B4E016;C9IYE8;C9IZU8;C9J6A7;C9J8S0;C9J9T | | | |
| RPF1 | 0 | 0 | 0 | 9 | 5 | 5 | ribosome production factor 1 homolog | 1 | 1p22.3 | NP_079341.2;Q9H9Y2 | | | |

Other Mass Spec Methods

Isotope labeled methods for improved quantitation

- SILAC- Stable Isotope Labeling with Amino acids in Cell culture (labeling in cell culture)
- ICAT - Isotope Coded Affinity Tags (label proteins in lysate)
- iTRAQ - isotope Tags for Relative and Absolute Quantitation (label peptides)

Multiple Reaction Monitoring (MRM)

- To identify particular peptides/proteins (biomarkers) with highest sensitivity
- Typically 2 to 3 peptides per protein, synthetic standards used, instruments tuned to best sensitivity for peptide/proteins of interest.

Sources of Mass Spec Data

Mass Spec Protein Data Repositories

PRIDE Proteomics IDentifications Archive (www.ebi.ac.uk/pride)

MassIVE Mass spec Interactive Virtual Environment (massive.ucsd.edu)

The Global Proteome Machine Database (gpmdb.thegpm.org/)

Peptide Atlas (www.peptideatlas.org)

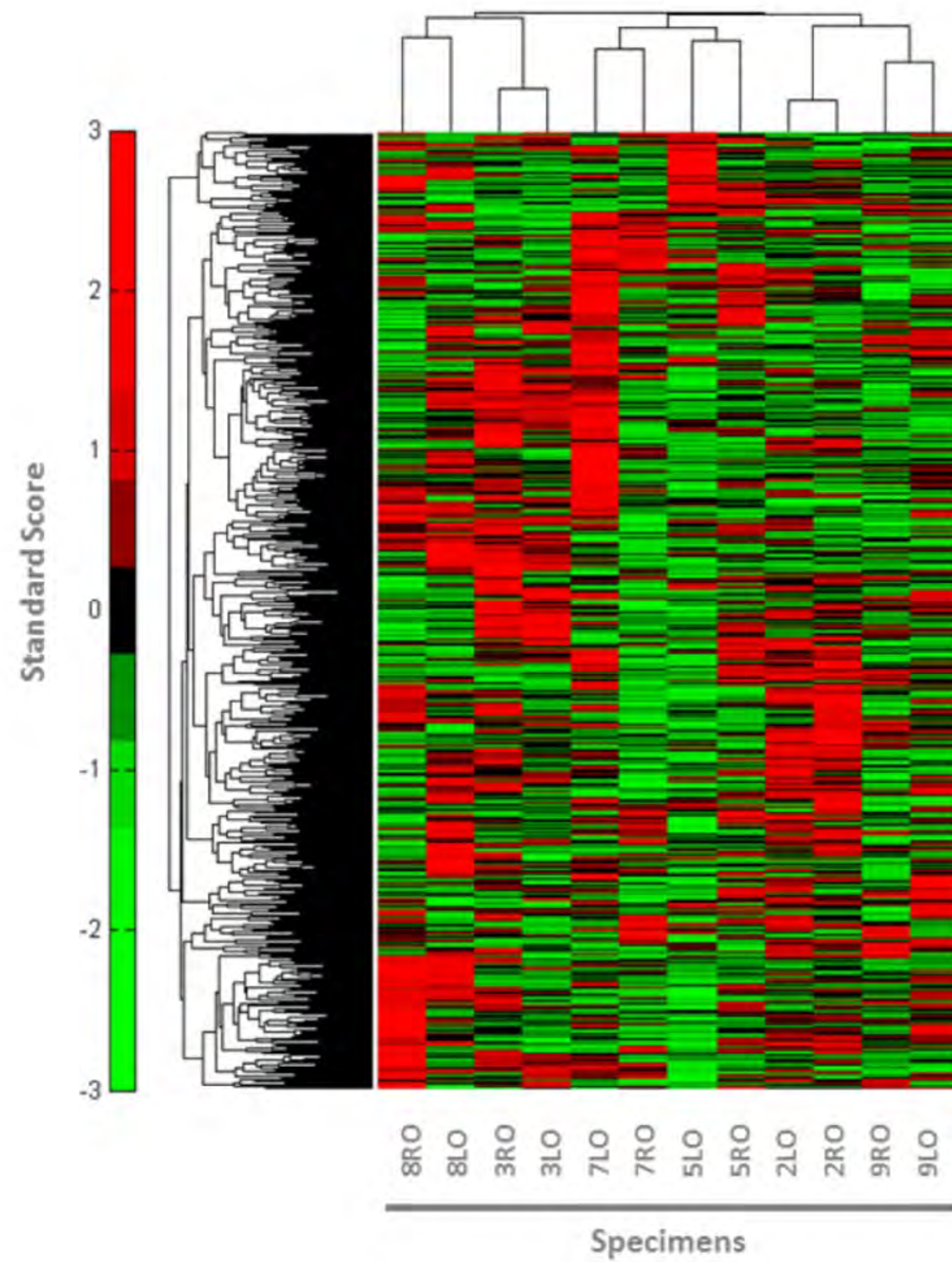
Clinical Proteomics Tumor Analysis Consortium (CPTAC)

- (cptac-data-portal.georgetown.edu/cptacPublic/) < data sets
- (<https://assays.cancer.gov>) < Peptides and assays for targeted methods like MRM

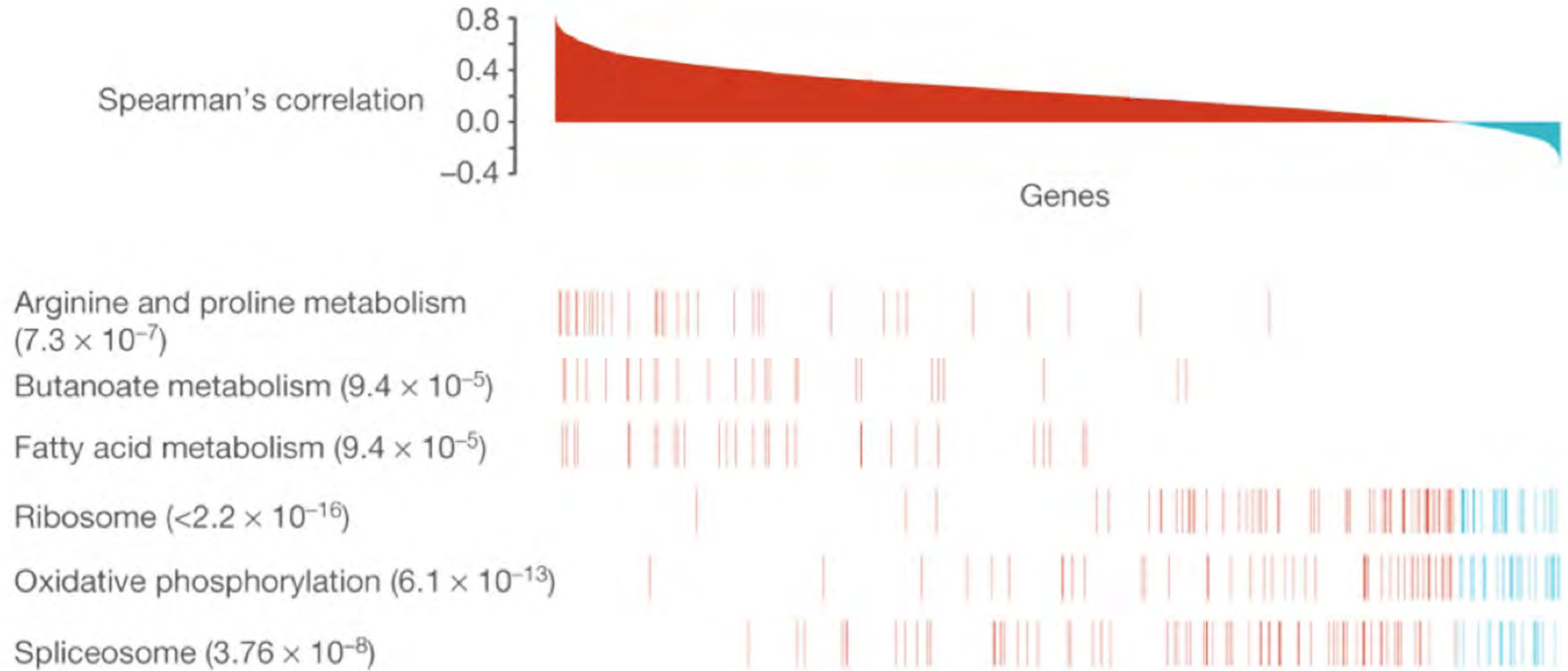
Many others ...

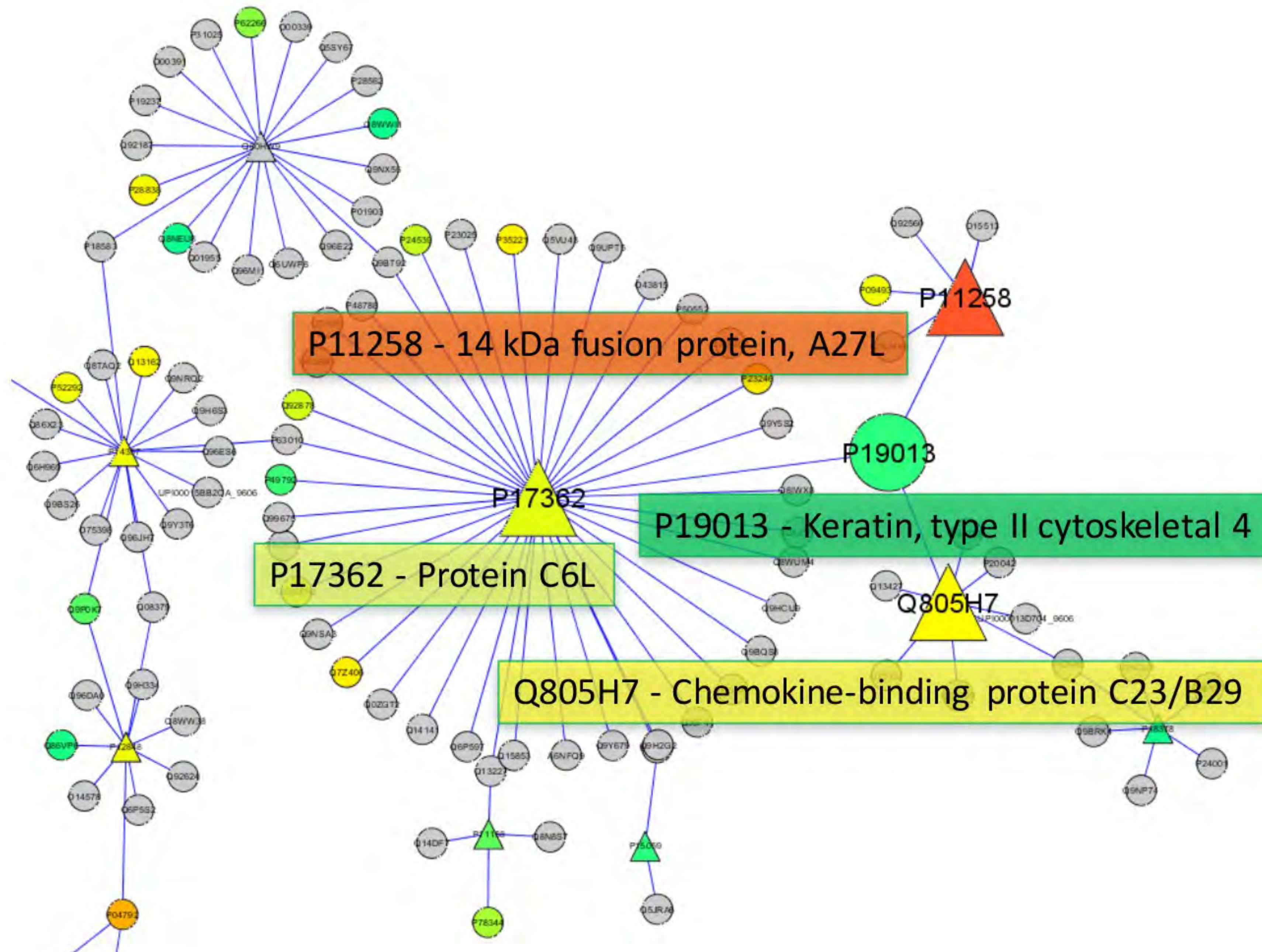
How to analyze MS proteomics data?

- Raw MS data files are manufacturer specific Need expertise and software to reanalyze.
- PSMs and protein/gene reports are more approachable though multiple vendors and search algorithms have different outputs.
- The end goal is list of protein identifications with quantitation under experimental conditions.



Proteomes 2013, 1(3), 240-253; doi:[10.3390/proteomes1030240](https://doi.org/10.3390/proteomes1030240)





Summary of MS proteomics

- A highly sophisticated powerful tool for protein identification and quantification.
- Complementary to other technologies such as microarray, RNASeq and other methods.
- Produces Terabytes of data you need to manage.
- Not a Simple technology
- Not Cheap. Machines and supporting infrastructure can cost over \$100,000 and will be rapidly replaced by better ones in a few year.