Fundamentals of Mass Spectrometry Based Proteomics and Applications for Quantitation

Part II: Quantification using mass spectrometry

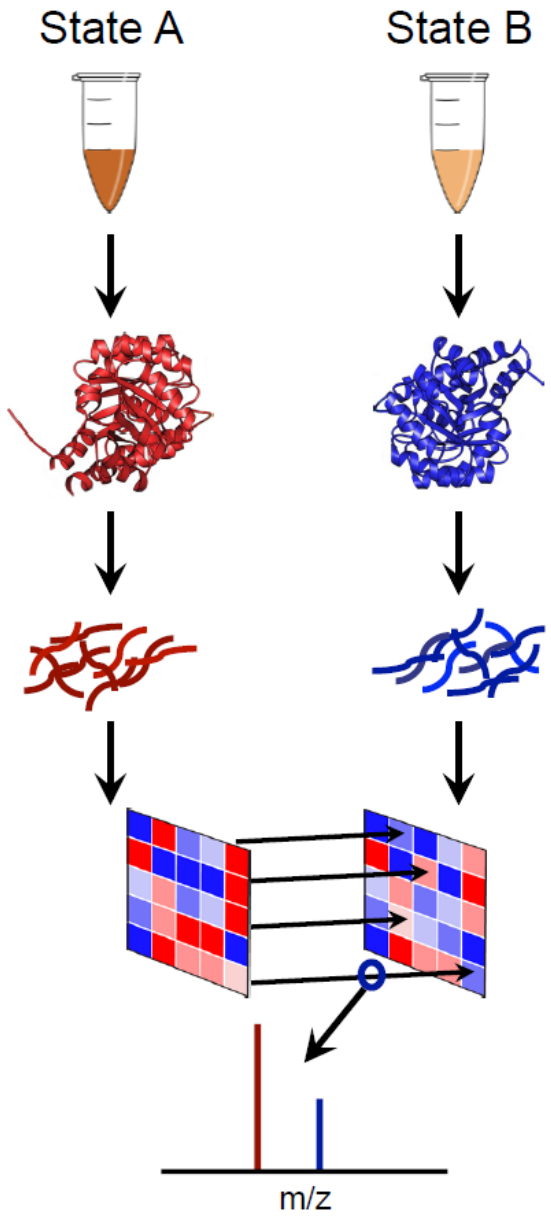
Lisa Jenkins, PhD

CCR Mass Spectrometry Resource

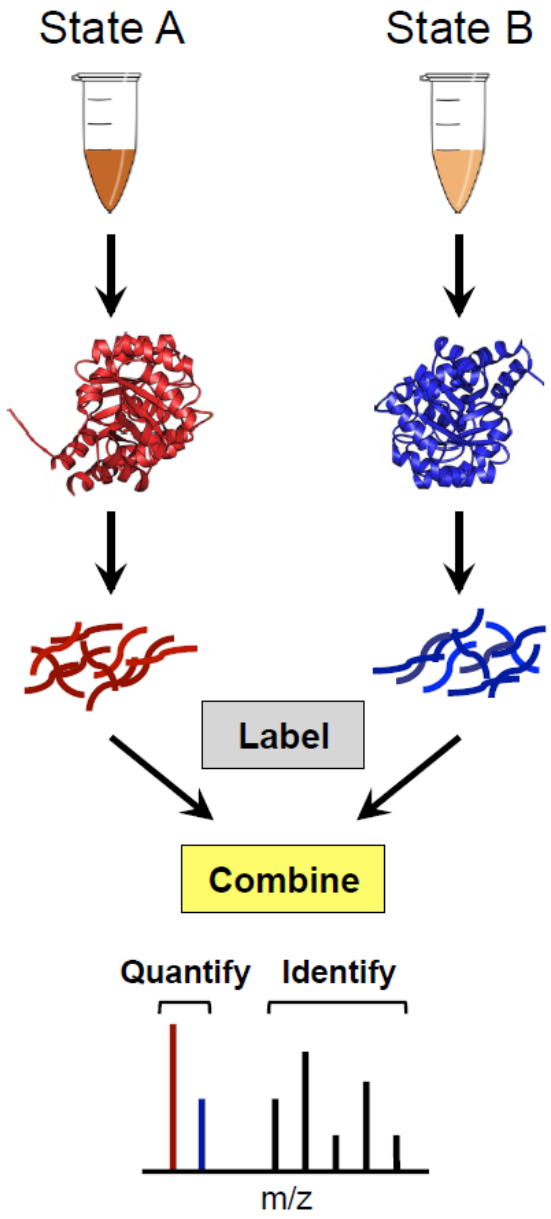
Fundamentals of mass spec-based quantitation

- Like many other quantification techniques, signal intensity is analyte dependent
- Two types of quantification are possible: relative and “absolute”
- In proteomics, most applications are looking at relative quantification – in this case, we are comparing the level of a protein/peptide/modification between different conditions/strains
  - End result of this analysis is always a ratio → WT/mut; Treated/untreated
  - Several methods for performing the quantitation
- For “absolute” (targeted) quantitation, comparison of the signal intensity to that of a standard curve is required

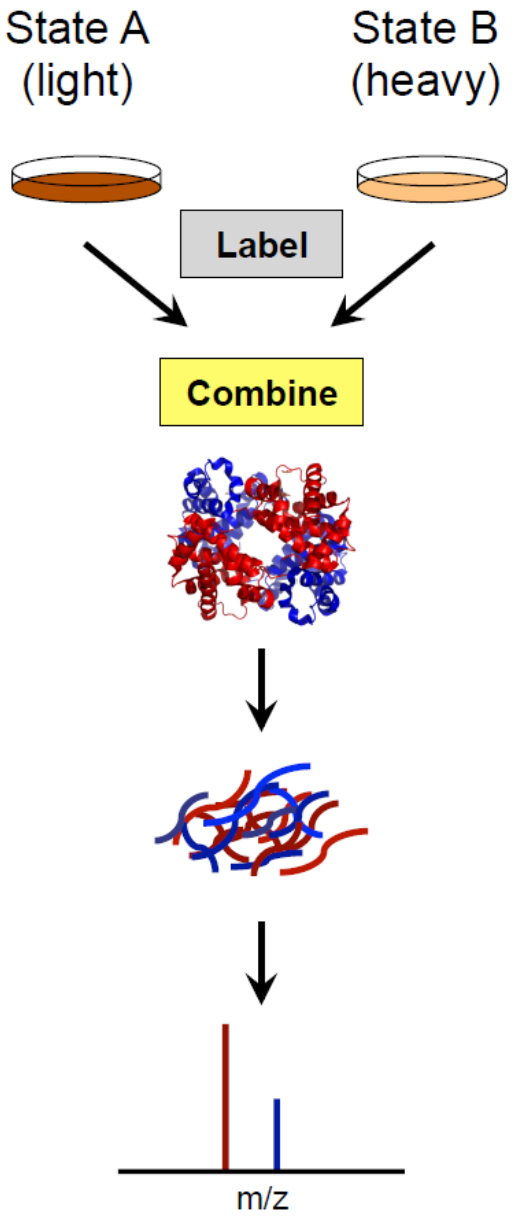
Label-free quantification (1 sample at a time)



Peptide chemical labeling (up to 18 samples at a time)

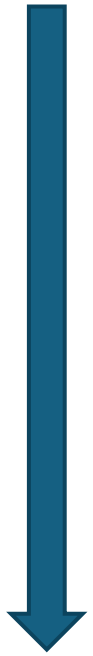


Metabolic labeling (SILAC) (up to 3 samples at a time)



Another way to consider these three approaches

- Label free quantitation
  - Spectral counting
  - Area under curve
- Peptide level labeling
  - iTRAQ up to 8 plex
  - TMT up to 18 plex
- Protein level labeling
  - SILAC – cell only up to 3 plex

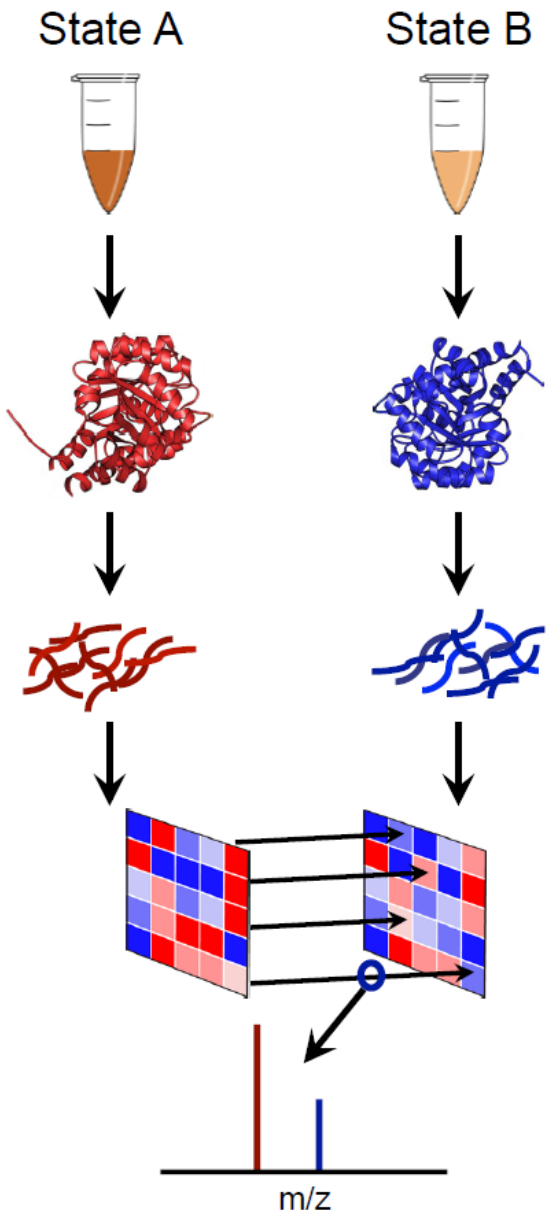


Increasing reliability by reducing experimental variability

Multiplexing improves throughput

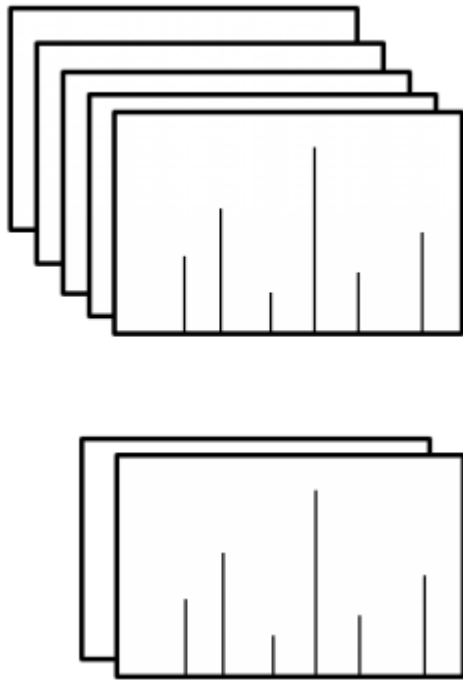
“stitching” data from multiple multiplex experiments can be a challenge if your sample # exceeds that of your reagent

Label-free quantitation (LFQ)



- In LFQ, each sample is processed and analyzed separately
  - Lowest throughput
  - No special reagents are required

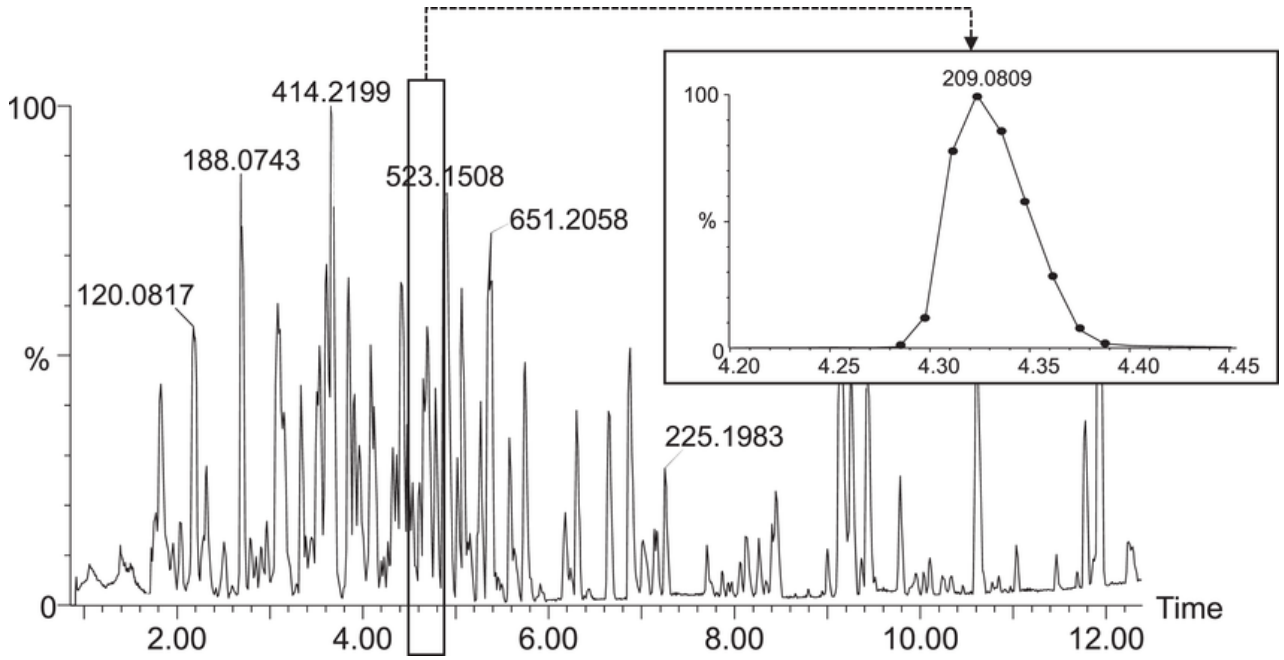
- Quantitation performed either on the number of MS/MS spectra that are linked to the protein (spectral counts) or abundance of the peptide signal

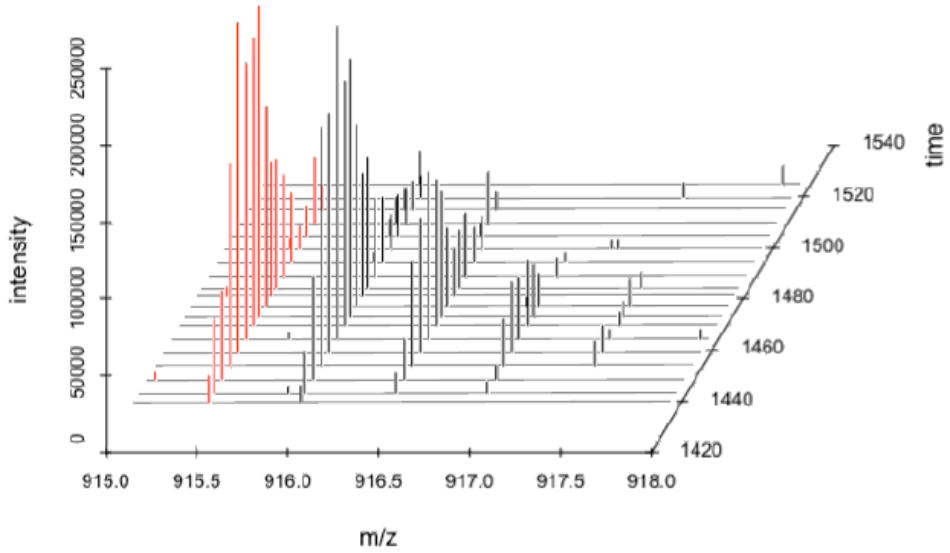


Spectral Counts – MS2

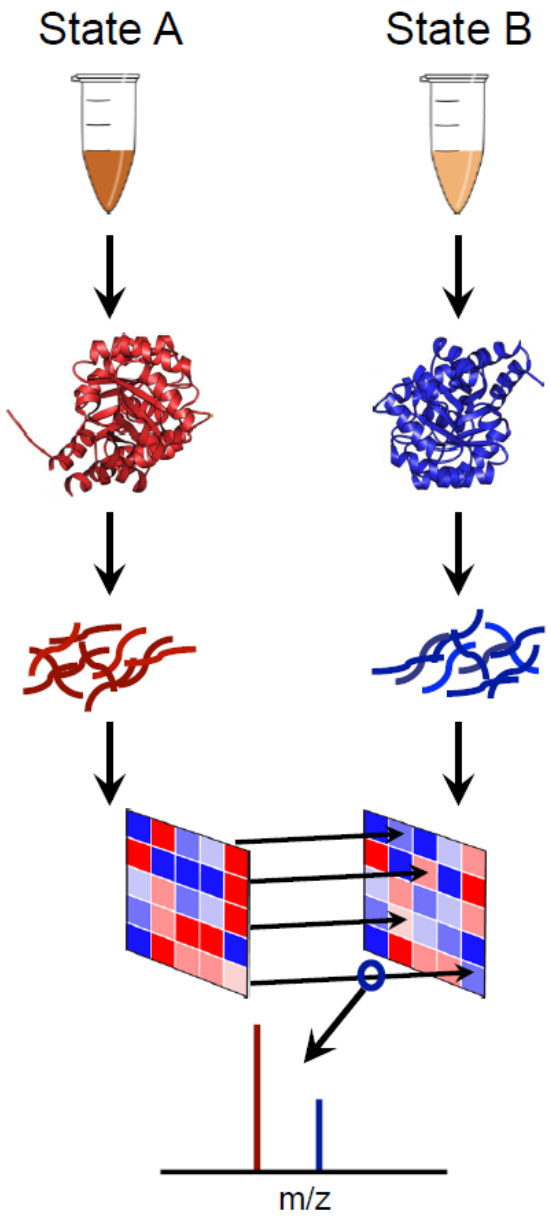
- Spectral counts:
  - Pro: every protein identified will have a count
  - Con: for small numbers, lose discrimination of differences

Quantitation MS1-based Abundance



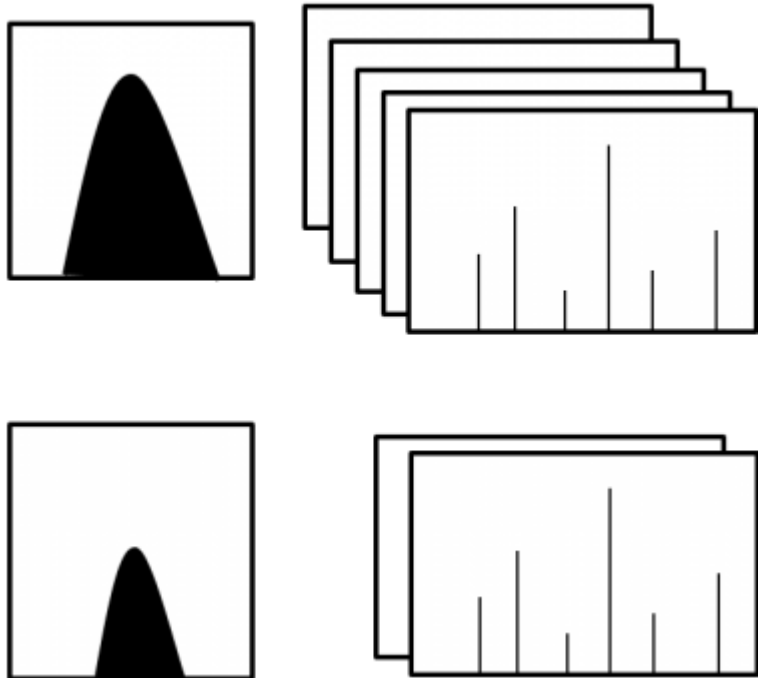


Label-free quantitation (LFQ)



- In LFQ, each sample is processed and analyzed separately
  - Lowest throughput
  - No special reagents are required

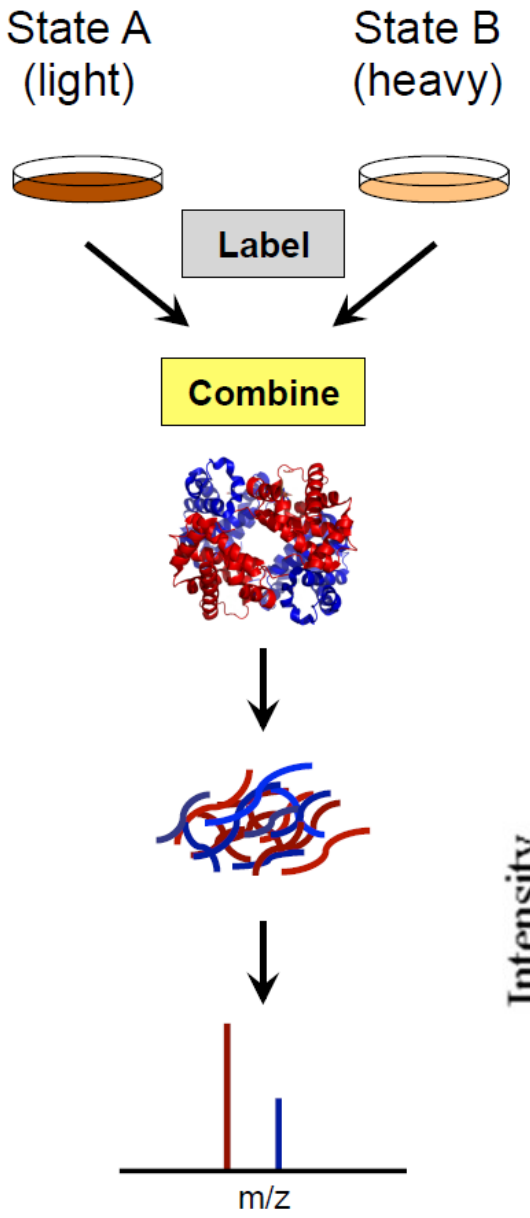
- Quantitation performed either on the number of MS/MS spectra that are linked to the protein (spectral counts) or abundance of the peptide signal



Abundance – MS1 Spectral Counts – MS2

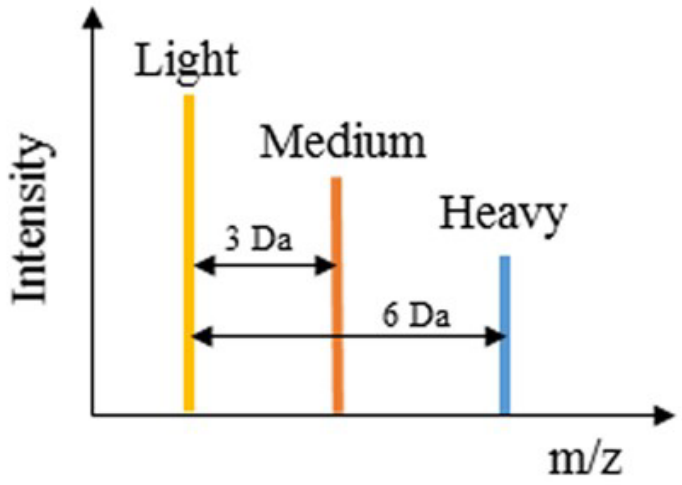
- Abundances
  - Pro: because derived from elution profile peak areas, they are large numbers (e3-e10) so see differences for low abundance proteins
  - Con: some peptides/proteins will not be able be quantified (“missing value problem”)
    - Missing values can be even more problematic in PTM analysis

Stable Isotope Labeling by Amino acids in Cell culture (SILAC)

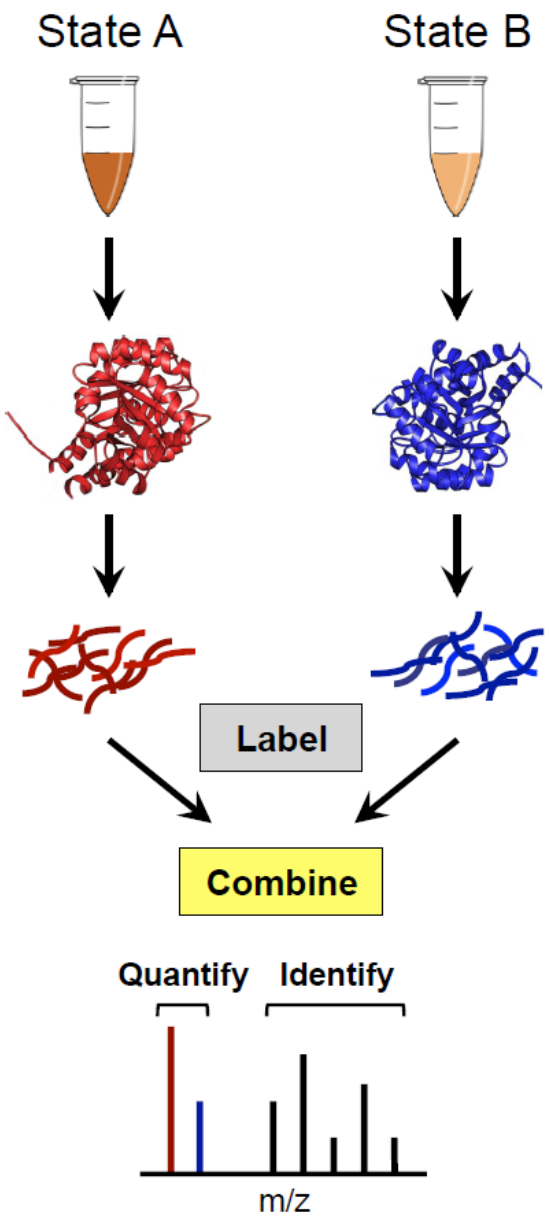


- With SILAC labeling, the cells are grown in isotopically-labeled media and then combined prior to sample processing
  - Requires labeled amino acids, intermediate cost, but allows 3-plexing
  - Is compatible with analysis PTMs and minimizes variation in sample processing
  - Is not compatible with human samples and special chow is required for mice

- Like LFQ, quantitation is based on peptide abundances

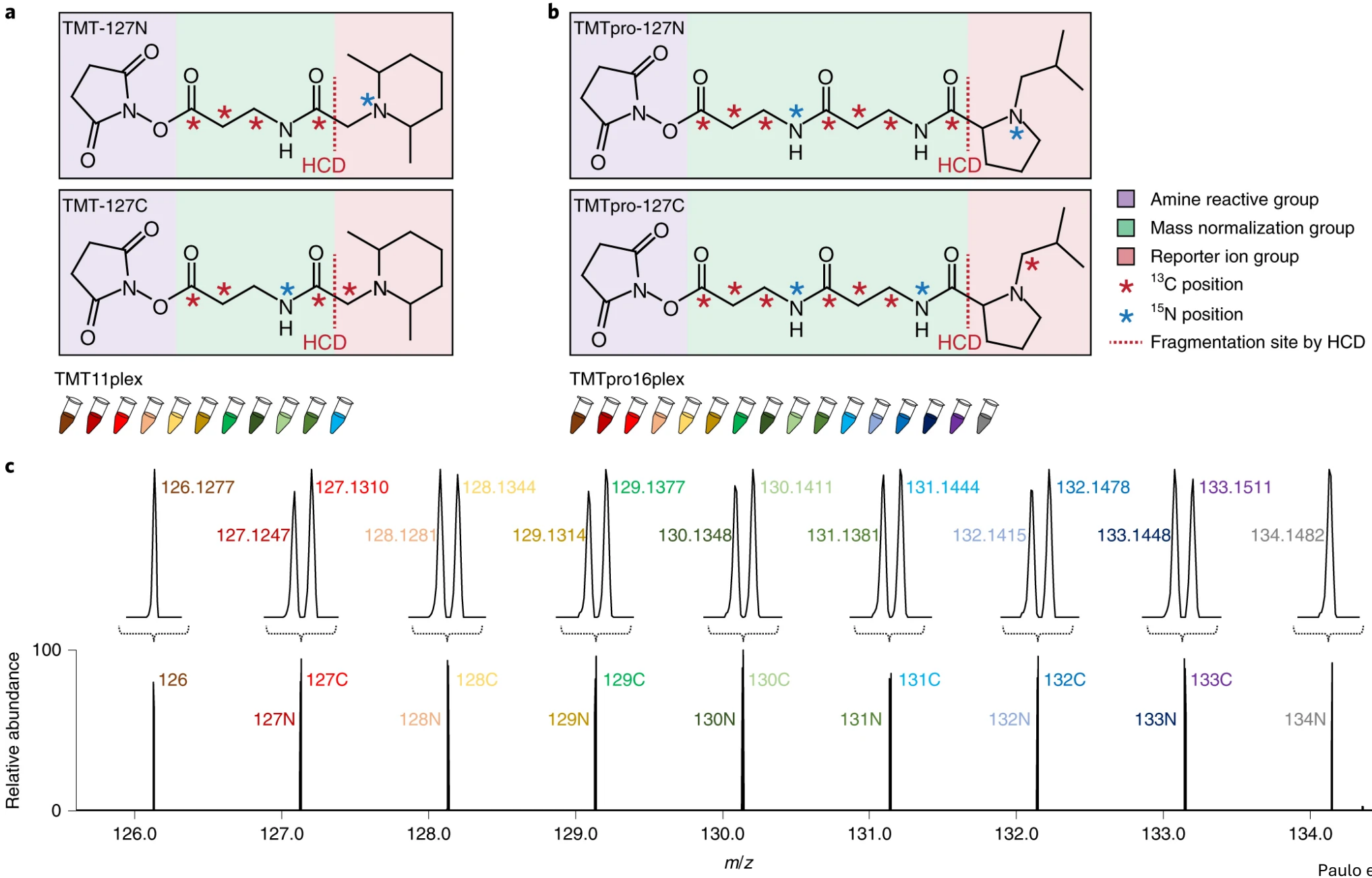


Multiplexed quantitation with isobaric mass tags



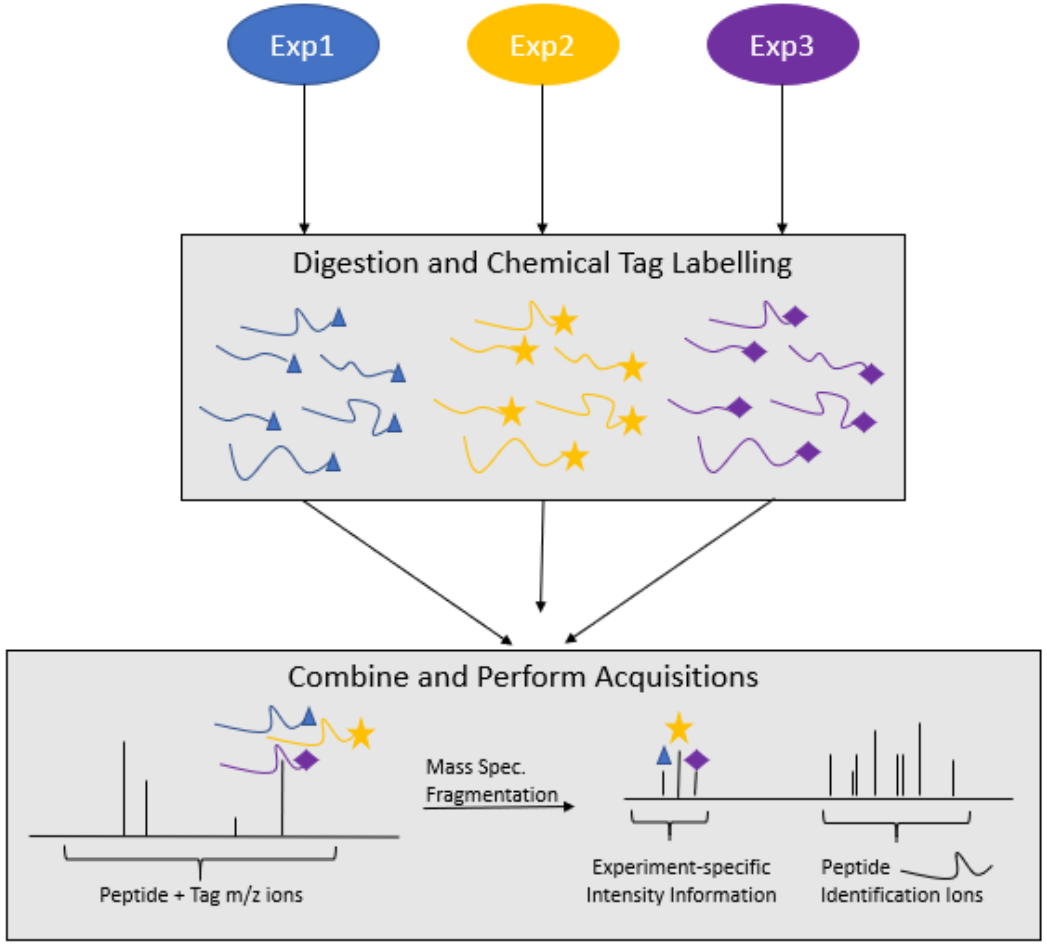
- Peptides are labeled with amine-reactive isobaric tags after proteolytic digestion
  - Requires specific reagent, which can be expensive (~$125/sample), but allows multiplexing of samples up to 18-plex on Orbitrap or 10-plex on TOF
  - Is compatible with analysis PTMs and all sample types
  - As samples are combined after processing, some variation in processing steps can be observed so replication is critical

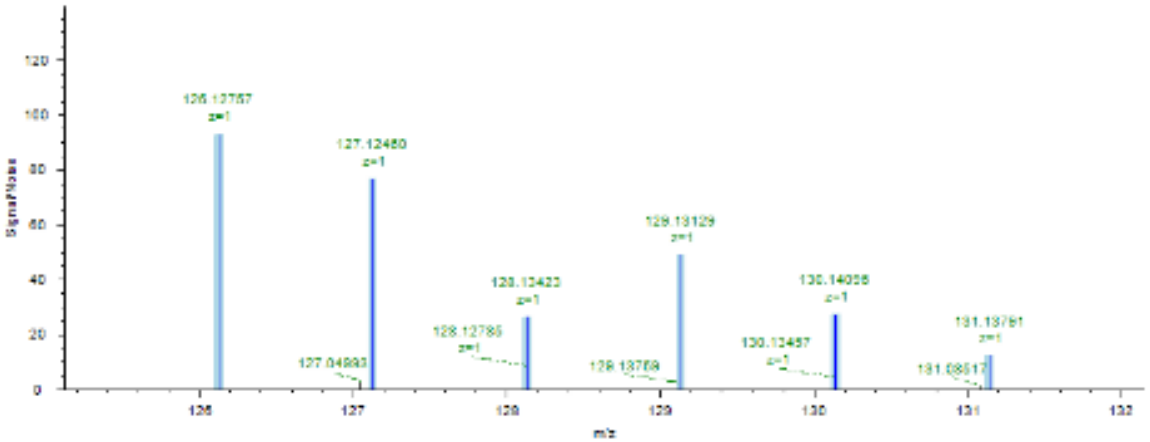
Tandem mass tags (TMT)

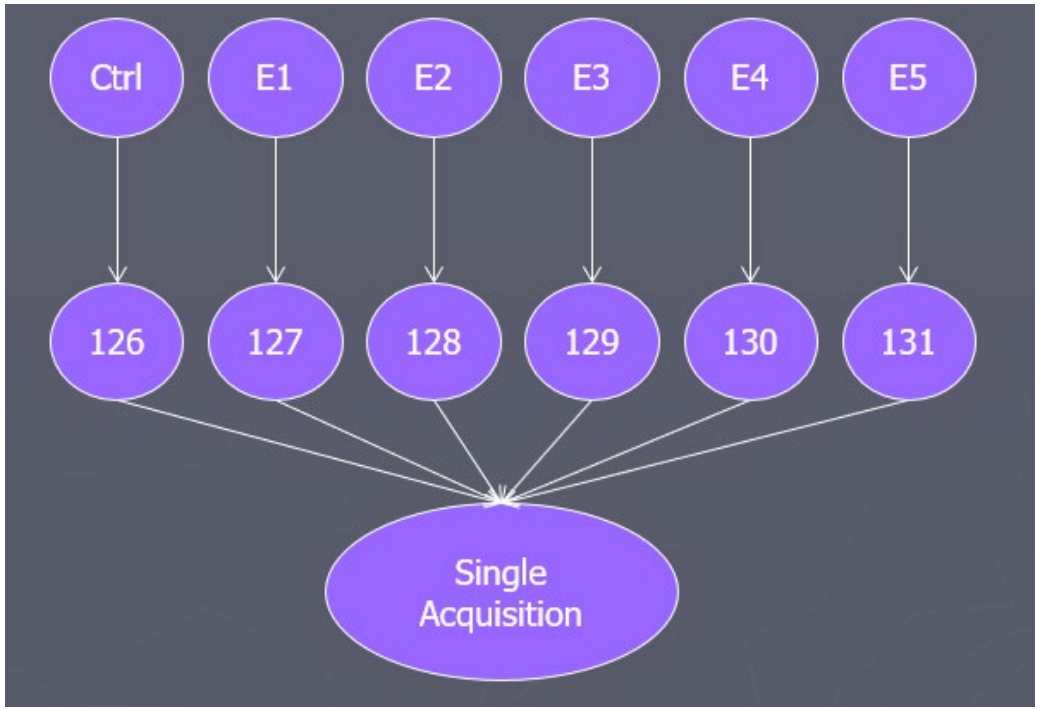


Paulo et al. Nat Methods 2020

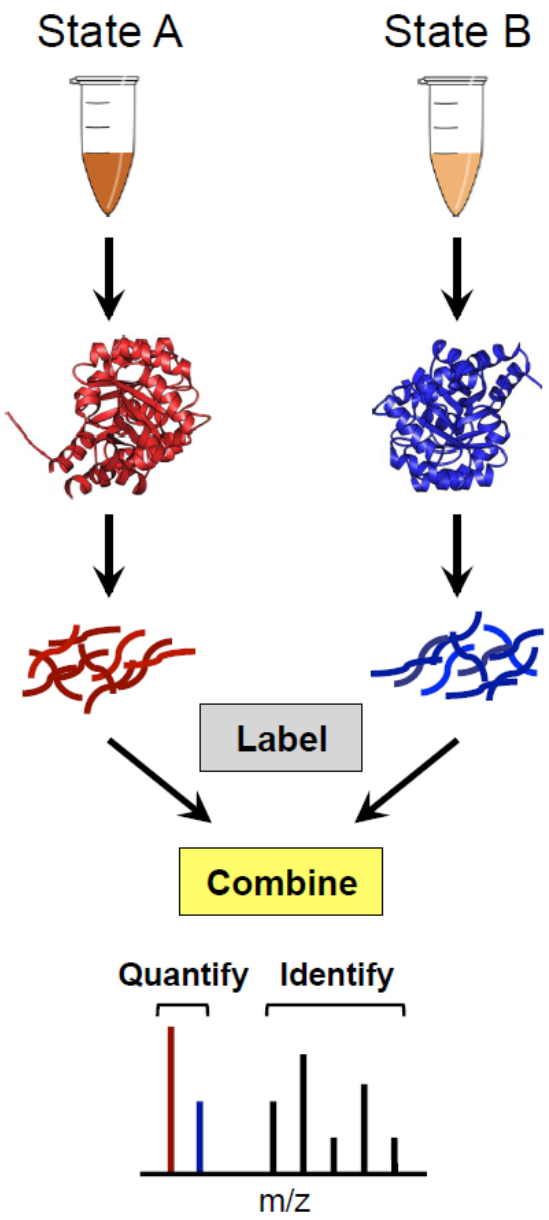
Quantitation MSn-based





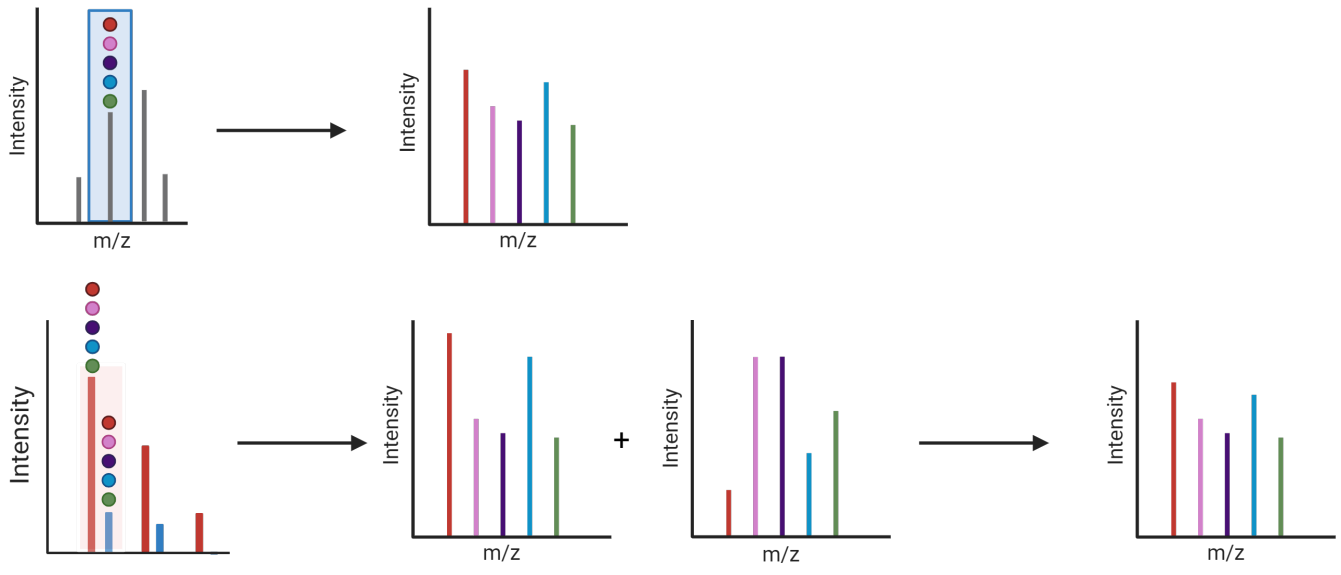


Multiplexed quantitation with isobaric mass tags



- Peptides are labeled with amine-reactive isobaric tags after proteolytic digestion
  - Requires specific reagent, which can be expensive (~$125/sample), but allows multiplexing of samples up to 18-plex on Orbitrap or 9-plex on TOF
  - Is compatible with analysis PTMs and all sample types
  - As samples are combined after processing, some variation in processing steps can be observed so replication is critical

- Quantitation is based on the intensity of reporter ions observed in MS/MS spectrum
  - Co-isolation of peptides of similar mass leads to ratio compression effects
  - Specific instrumental design can be used to minimize this, but that requires increased instrument time which may result in fewer identifications



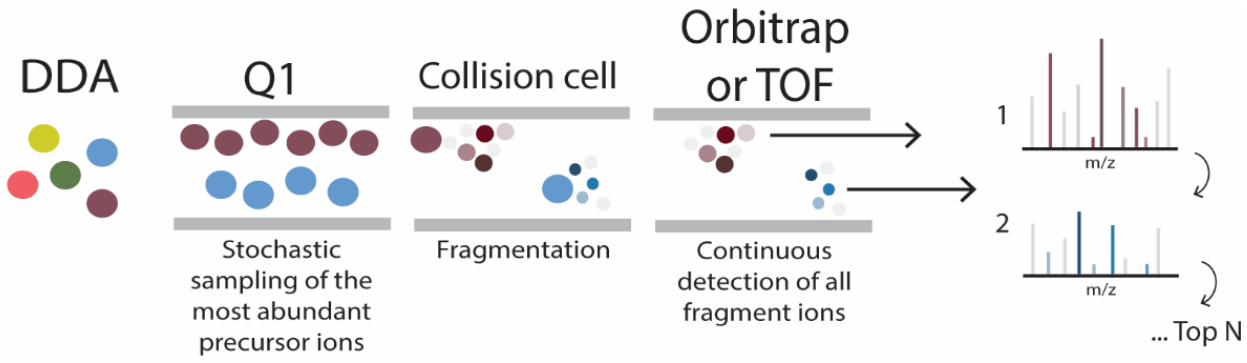
iBAQ for Comparison Across Proteins

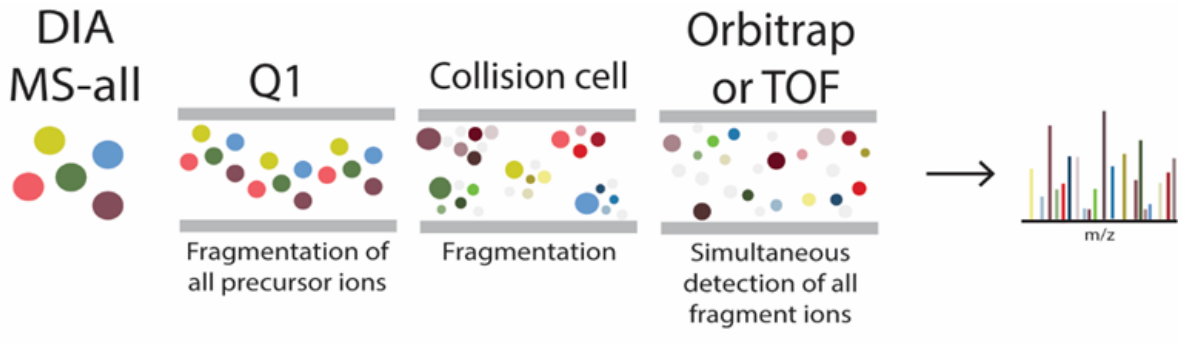
- In general, the relative quantitation is the same protein across conditions
  - Differences in protein sequence (how many peptides, length of peptides, sequence of peptides) determine how it ionizes
  - Assume those effects are consistent for the same protein in different samples or conditions

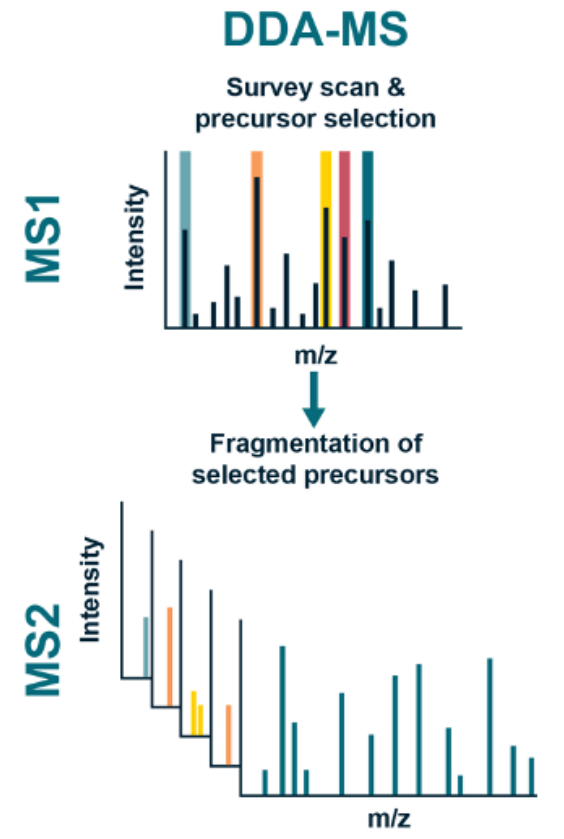
- When experimental need requires comparison of proteins in the same sample, a different algorithm is needed
  - iBAQ - intensity Based Absolute Quantitation
  - iBAQ metric is normalized to the number of identifiable peptides for a given protein to provide a measure of the protein's absolute abundance

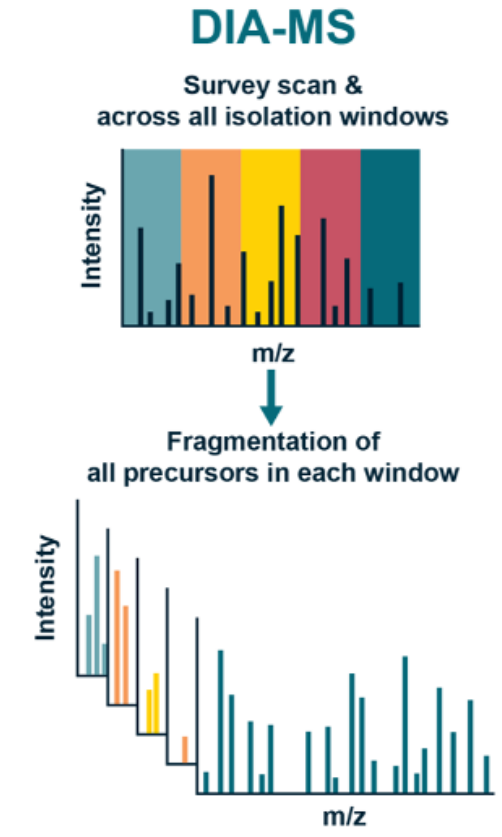
$$iBAQ = \Sigma intensity / \#theoretical\ peptides$$

Data-independent acquisition









Data-independent acquisition

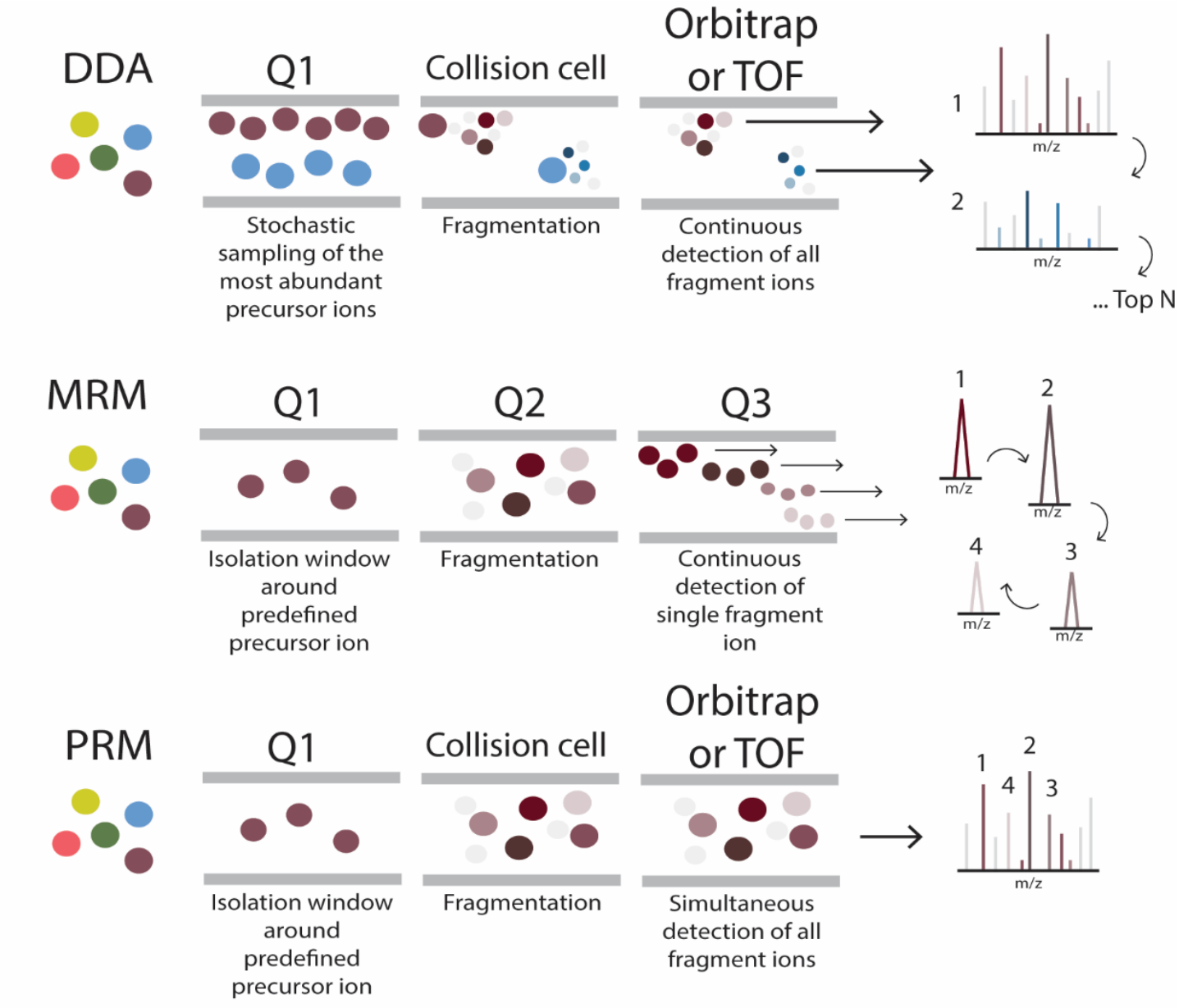
DIA data requires different software for analysis

- originally deconvolute the MS/MS spectra by matching to spectral library

- currently, neural networks are used to predict properties of peptides, such as retention time, ion mobility, or fragmentation, from the sequence so that analysis can be done from sequence file by predicting spectral library

- Quantitation can be performed at the MS1 or MS2 level

Targeted Quantitation



What do relative quantitative proteomic data look like?

Accession	Gene	Description	Ratio: (C2) / (C1)	Ratio P-Value: (C2) / (C1)	Abundance: 126, C1, 1	Abundance: 127N, C1, 2	Abundance: 127C, C1, 3	Abundance: 128N, C2, 1	Abundance: 128C, C2, 2	Abundance: 129N, C2, 3
O15234	CASC3	Protein CASC3	1.392	7.35E-02	7.42E+01	6.84E+01	6.46E+01	8.71E+01	1.02E+02	1.12E+02
P14635	CCNB1	G2/mitotic-specific cyclin-B1	1.195	8.15E-04	1.53E+03	1.56E+03	1.45E+03	1.90E+03	1.94E+03	1.99E+03
Q8N573	OXR1	Oxidation resistance protein 1	0.953	1.00E+00	2.02E+02	2.07E+02	1.79E+02	1.98E+02	2.12E+02	2.05E+02
Q04206	RELA	Transcription factor p65	1.177	3.68E-02	2.66E+02	2.51E+02	2.56E+02	3.23E+02	3.35E+02	3.51E+02
Q9BYD3	MRPL4	39S ribomal protein L4, mitochondrial	1.001	1.00E+00	1.85E+03	1.87E+03	1.83E+03	1.92E+03	2.05E+03	1.98E+03
P49427	CDC34	Ubiquitin-conjugating enzyme E2 R1	1.206	8.84E-04	2.27E+02	2.29E+02	2.20E+02	2.77E+02	2.98E+02	2.92E+02

Other columns may be included, but frequently:

- Accession number from database
- Gene symbol
- Ratio
- p-value
- Raw abundance values

There are multiple points for sample normalization:

- Biological sample (number of cells, total protein)
- Peptide level, injection size
- Total signal intensity for a sample/channel

What do relative quantitative proteomic data look like?

Accession	Gene	Description	Ratio: (C2) / (C1)	Ratio P-Value: (C2) / (C1)	Abundance: 126, C1, 1	Abundance: 127N, C1, 2	Abundance: 127C, C1, 3	Abundance: 128N, C2, 1	Abundance: 128C, C2, 2	Abundance: 129N, C2, 3
O15234	CASC3	Protein CASC3	1.392	7.35E-02	7.42E+01	6.84E+01	6.46E+01	8.71E+01	1.02E+02	1.12E+02
P14635	CCNB1	G2/mitotic-specific cyclin-B1	1.195	8.15E-04	1.53E+03	1.56E+03	1.45E+03	1.90E+03	1.94E+03	1.99E+03
Q8N573	OXR1	Oxidation resistance protein 1	0.953	1.00E+00	2.02E+02	2.07E+02	1.79E+02	1.98E+02	2.12E+02	2.05E+02
Q04206	RELA	Transcription factor p65	1.177	3.68E-02	2.66E+02	2.51E+02	2.56E+02	3.23E+02	3.35E+02	3.51E+02
Q9BYD3	MRPL4	39S ribomal protein L4, mitochondrial	1.001	1.00E+00	1.85E+03	1.87E+03	1.83E+03	1.92E+03	2.05E+03	1.98E+03
P49427	CDC34	Ubiquitin-conjugating enzyme E2 R1	1.206	8.84E-04	2.27E+02	2.29E+02	2.20E+02	2.77E+02	2.98E+02	2.92E+02

Other columns may be included, but frequently:

- Accession number from database
- Gene symbol
- Ratio
- p-value
- Raw abundance values

There are multiple points for sample normalization:

- Biological sample (number of cells, total protein)
- Peptide level, injection size
- Total signal intensity for a sample/channel

Special case – PTM analysis

Accession	Gene	Positions in Master Proteins	Annotated Sequence	Modifications in Master Proteins	Phospho Abundance Ratio: (S2) / (S1)	Phospho Abundance p-value: (S2) / (S1)	Total Abundance Ratio: (S2) / (S1)	Total Abundance Ratio Adj. P-Value: (S2) / (S1)
Q2M2I8	AAK1	Q2M2I8 [670-680]	[K].SATTTPSGSPR.[T]	Q2M2I8 2xPhospho [T674(100); S678(100)]	0.991	1.00E+00	0.884	9.87E-01
Q8NE71	ABCF1	Q8NE71 [221-245]	[K].AKKAEQGSEEEGEGEEEEEEGGESK.[A]	Q8NE71 1xPhospho [S228(100)]	1.147	1.35E-01	1.587	4.34E-03
Q15057	ACAP2	Q15057 [512-528]	[K].FVDKYSISLSPPEQQKK.[F]	Q15057 1xPhospho [S521(99.6)]	1.922	2.77E-04	1.903	9.93E-04
Q9UKV3	ACIN1	Q9UKV3 [326-336]	[K].TRSQEQEVLER.[G]	Q9UKV3 1xPhospho [S328(100)]	0.285	1.09E-03	2.052	1.60E-03

With PTM analysis, data look a little different because the focus is on the site of modification rather than the protein

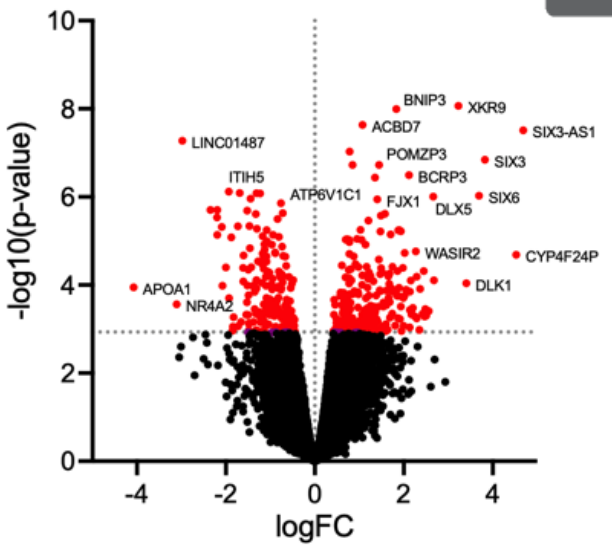
- The output will give the sequence and residue numbers of the modified peptide
- Frequently localization scores are given to indicate confidence in which residue has the modification
- Quantitation of peptide abundance should be compared to total protein abundance to see how the modification site changes relative to the protein overall

Proteomics data will associate with the official gene symbol, which means that pathway- and functional-based approaches can be used for mining of the data.

Ingenuity Pathway Analysis

QLUCORE

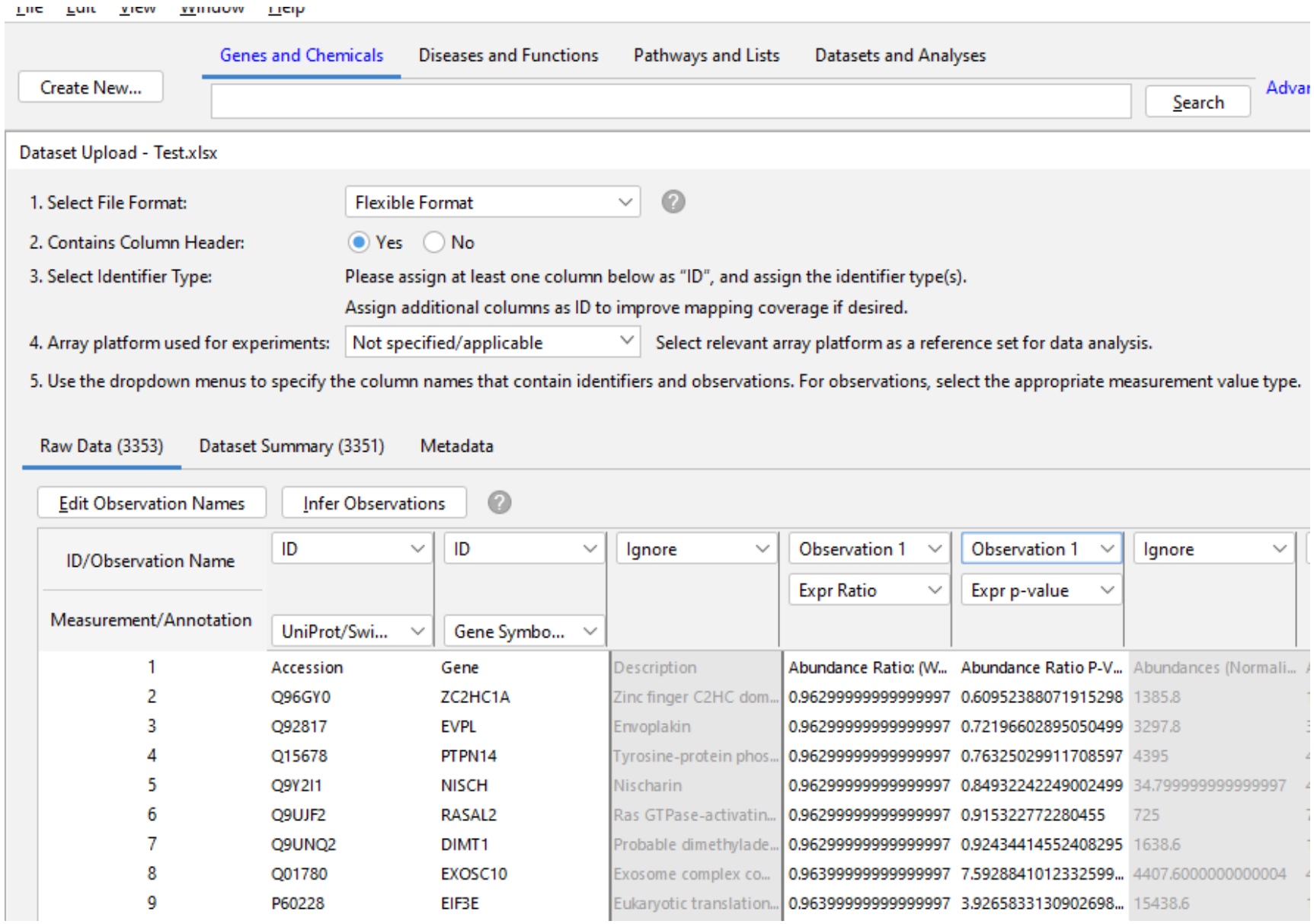
NIH National Institutes of Health
DAVID Bioinformatics



NIDAP

STRING

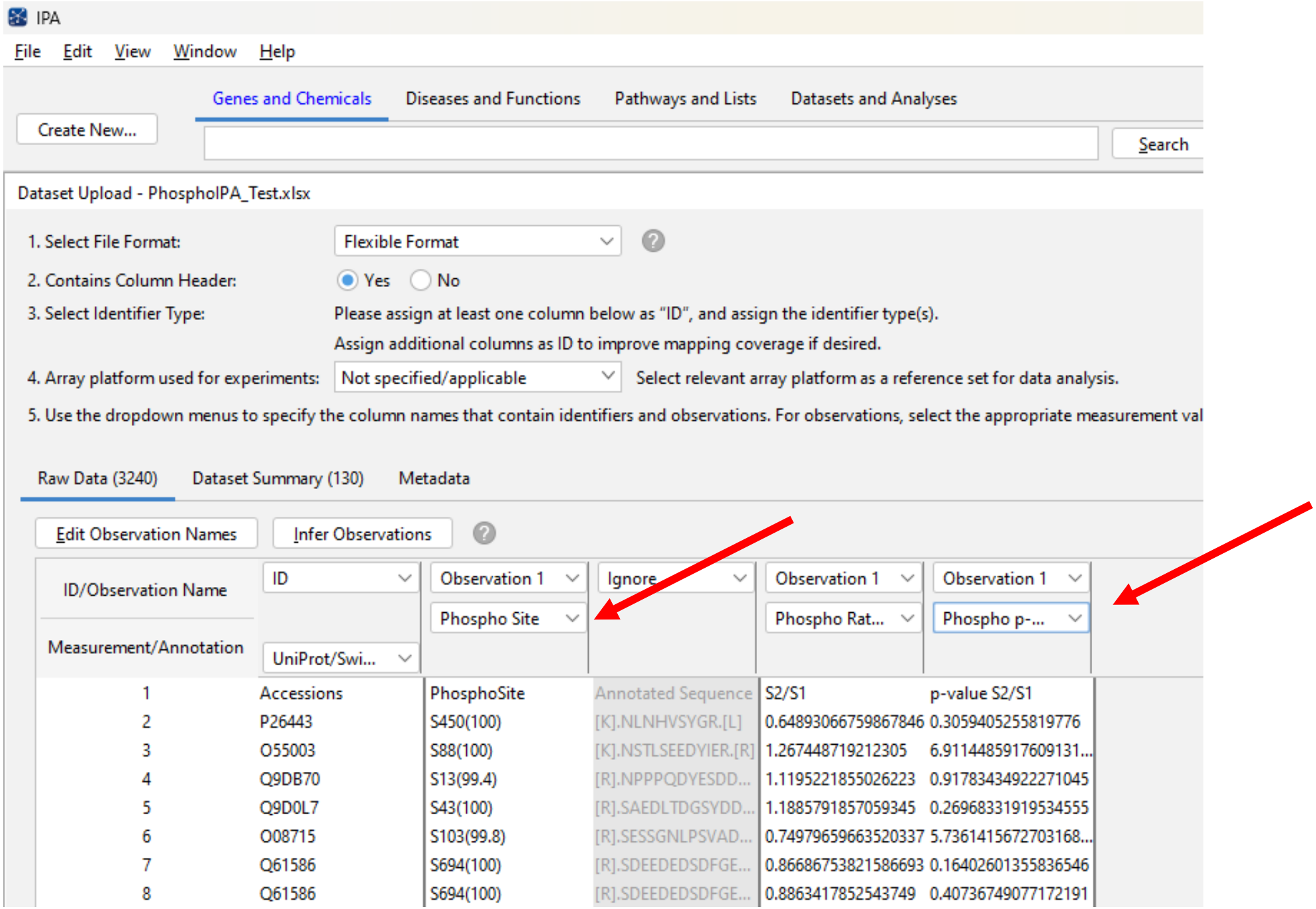
Ingenuity Pathway Analysis

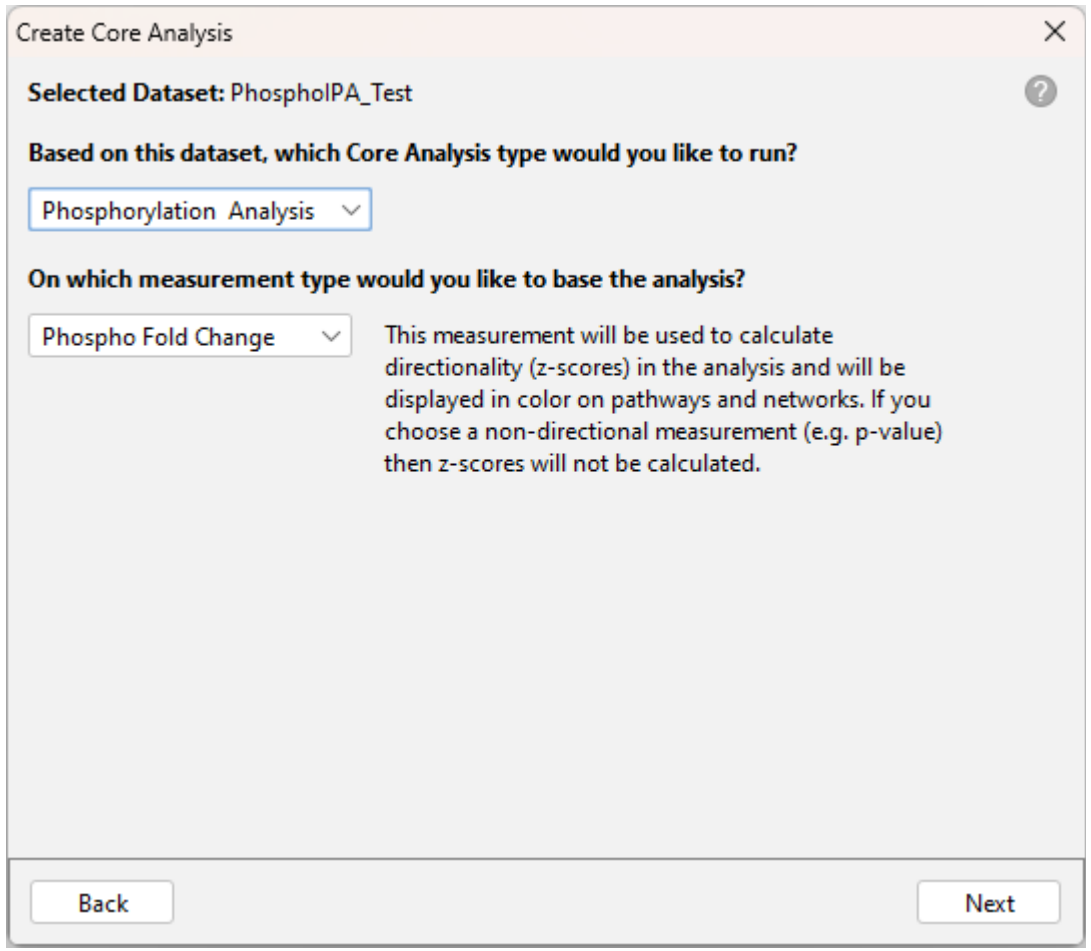


Ingenuity has added new modules for analysis of phosphoproteomic data

Accessions	PhosphoSite	Annotated Sequence	S2/S1	p-value S2/S1
P26443	S450(100)	[K].NLNHVSYGR.[L]	0.65	0.31
O55003	S88(100)	[K].NSTLSEEDYIER.[R]	1.27	0.07
Q9DB70	S13(99.4)	[R].NPPPQDYESDDESYEVLDLTEYAR.[R]	1.12	0.92
Q9D0L7	S43(100)	[R].SAEDLTDGSYDDILNAEQLKK.[L]	1.19	0.27
O08715	S103(99.8)	[R].SESSGNLPSVADTR.[S]	0.75	0.01
Q61586	S694(100)	[R].SDEEDEDSDFGEEQR.[D]	0.87	0.16
Q61586	S694(100)	[R].SDEEDEDSDFGEEQRDCYLK.[V]	0.89	0.41
Q61586	S687(100); S694(100)	[R].SDEEDEDSDFGEEQRDCYLK.[V]	0.71	0.15

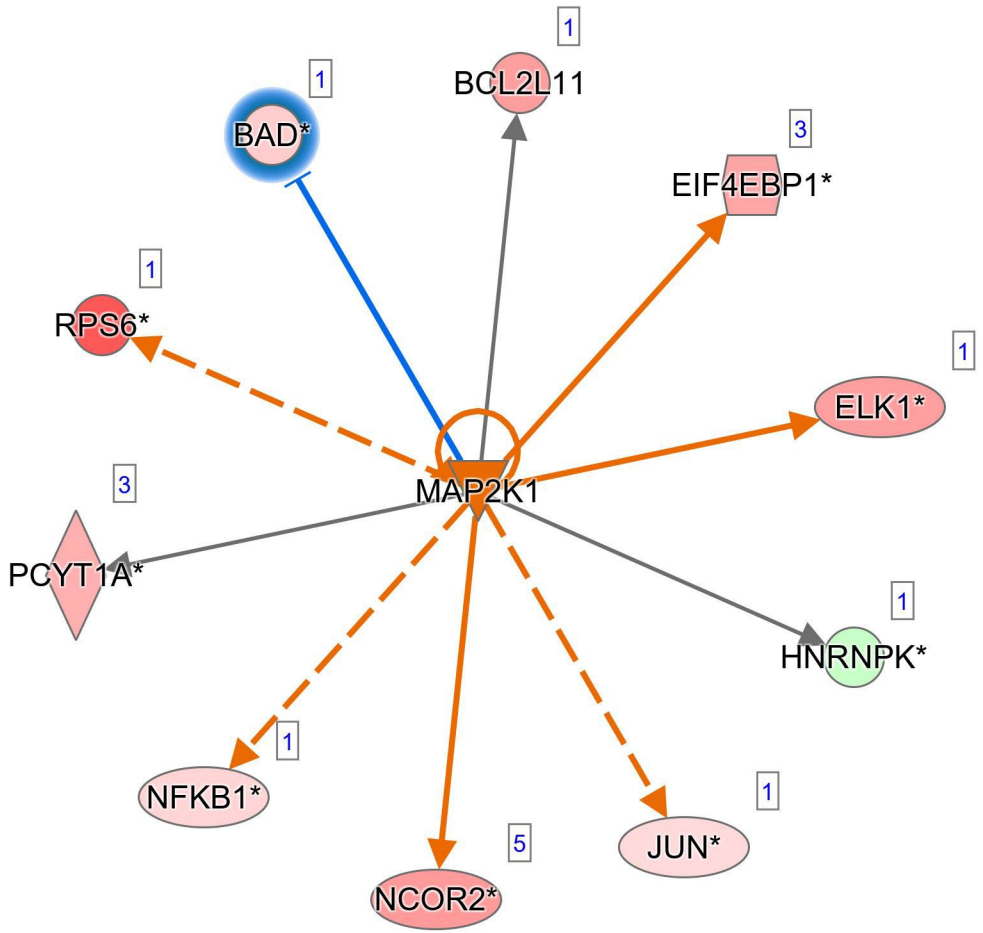
Requires different upload format that has a unique phosphosite identifier for each phosphopeptide.


IPA
File Edit View Window Help
Genes and Chemicals
Diseases and Functions
Pathways and Lists
Datasets and Analyses
Create New...
Search
Dataset Upload - PhosphoIPA_Test.xlsx
1. Select File Format:
Flexible Format
2. Contains Column Header:
Yes
No
3. Select Identifier Type:
Please assign at least one column below as "ID", and assign the identifier type(s).
Assign additional columns as ID to improve mapping coverage if desired.
4. Array platform used for experiments:
Not specified/applicable
Select relevant array platform as a reference set for data analysis.
5. Use the dropdown menus to specify the column names that contain identifiers and observations. For observations, select the appropriate measurement val
Raw Data (3240)
Dataset Summary (130)
Metadata
Edit Observation Names
Infer Observations
ID/Observation Name
Measurement/Annotation
ID
Observation 1
Ignore
Observation 1
Observation 1
Phospho Site
Phospho Rat...
Phospho p-...
UniProt/Swi...
1 Accessions PhosphoSite Annotated Sequence S2/S1 p-value S2/S1
2 P26443 S450(100) [K].NLNHVSYGR.[L] 0.64893066759867846 0.3059405255819776
3 O55003 S88(100) [K].NSTLSEEDYIER.[R] 1.267448719212305 6.9114485917609131...
4 Q9DB70 S13(99.4) [R].NPPPQDYESDD... 1.1195221855026223 0.91783434922271045
5 Q9D0L7 S43(100) [R].SAEDLTDGSYDD... 1.1885791857059345 0.26968331919534555
6 O08715 S103(99.8) [R].SESSGNLPSVAD... 0.74979659663520337 5.7361415672703168...
7 Q61586 S694(100) [R].SDEEDEDSDFGE... 0.86686753821586693 0.16402601355836546
8 Q61586 S694(100) [R].SDEEDEDSDFGE... 0.8863417852543749 0.40736749077172191


Create Core Analysis
Selected Dataset: PhosphoIPA_Test
Based on this dataset, which Core Analysis type would you like to run?
Phosphorylation Analysis
On which measurement type would you like to base the analysis?
Phospho Fold Change
This measurement will be used to calculate directionality (z-scores) in the analysis and will be displayed in color on pathways and networks. If you choose a non-directional measurement (e.g. p-value) then z-scores will not be calculated.
Back
Next

Upstream Regulator	Phospho Log Ratio	Phospho p-value	Molecule Type	Predicted Activation State	Activation z-score	p-value of overlap	Target Molecules in Dataset	Mechanistic Network
MASTL	↓-0.390	2.80E-02	kinase			9.20E-03	↑ARPP19, ↑EGFR, ↑ENSA, ↓MASTL	...all 4	
CLK1			kinase		0.218	2.28E-02	↑PTPN1, ↓SRSF2, ↓SRSF6, ↑U2AF2	...all 4	
NME7			kinase			4.72E-02	↑CTNNB1, ↓GSK3B	...all 2	
CDKL5	↑0.170	2.89E-01	kinase		-0.283	4.42E-02	↑ARHGEF2, ↑ELOA, ↓MAP1S, ↓MECP2	...all 4	
PRKDC	↑0.740	5.30E-03	kinase		1.724	1.87E-02	↑ATM, ↓CBX5, ↑CHEK1, ↓EIF2S2, ↑...	...all 14	
MAP2K4			kinase		1.976	1.00E00	↑ATF2, ↑JUN, ↑MAP1B, ↓MYC	...all 4	
MAP4K1			kinase		-1.980	3.08E-01	↓CRKL, ↓MAP3K11, ↑PLCG1, ↓PSMD2	...all 4	
MAPK13			kinase		1.025	1.70E-02	↑ATF2, ↓CCND3, ↑EEF2K, ↑EIF4EBP1,	...all 9	
PRKCE	↓-0.230	4.59E-03	kinase		1.797	2.23E-01	↓ADAM17, ↑AFF4, ↓AKT1S1, ↑BAD, ↓...	...all 19	
DAPK1	↑0.370	2.08E-02	kinase		1.342	4.56E-02	↑DAPK1, ↓MAPT, ↑MCM3, ↑PRKD1, ↓...	...all 5	
LATS1	↓-0.310	1.78E-02	kinase		-1.987	4.42E-02	↑BRCA2, ↓MAP3K11, ↓PPP1R12A, ↓R...	...all 4	
ULK1	↓-0.500	8.47E-02	kinase		-0.816	4.36E-02	↓AMBRA1, ↓CDC37, ↑EPHA2, ↓PSM...	...all 6	
CDK1	↑1.270	6.13E-03	kinase	Inhibited	-2.475	3.51E-05	↑AJUBA, ↓AMBRA1, ↓ANAPC1, ↓APC...	...all 50	
MAP2K1	↑0.050	4.12E-01	kinase	Activated	2.621	1.00E00	↑BAD, ↑BCL2L11, ↑EIF4EBP1, ↑ELK1,	...all 10	
MAPK8			kinase		0.491	1.37E-02	↑ATF2, ↑BAD, ↑BCL2L11, ↓CCND3, ↓...	...all 26	
CHEK1	↑1.990	8.18E-05	kinase		0.304	1.02E-02	↑CDC25A, ↓CDC25B, ↓CDC25C, ↓C...	...all 17	
MTOR	↑1.280	2.97E-02	kinase		0.606	2.97E-02	↓AKT1S1, ↓AMBRA1, ↑CAD, ↑CLIP1,	...all 20	
PRKG1	↑0.390	7.46E-03	kinase		0.804	3.89E-02	↑BAD, ↑CALD1, ↓CREB1, ↑CTNNB1,	...all 12	
CHUK			kinase	Activated	2.191	1.00E00	↑CTNNB1, ↑HTT, ↑IRS1, ↑NCOR2, ↑...	...all 5	
FES			kinase		-0.181	1.80E-02	↓BCR, ↓DPYSL2, ↑HDGF, ↓HDGFL2,	...all 12	
CAMK4	↑1.580	1.67E-03	kinase		-0.941	3.32E-02	↑BAD, ↓CABIN1, ↑CAMK4, ↓CREB1, ↓...	...all 9	
PIM1			kinase		0.366	4.25E-02	↑BAD, ↓CBX3, ↑CDC25A, ↓CDC25C,	...all 13	
MAPK11			kinase		0.363	3.72E-02	↑ATF2, ↑ATF7, ↓CCND3, ↓CDC25B, ↓...	...all 14	
AKT1	↓-0.050	1.63E-01	kinase		-0.392	4.53E-03	↓AKT1S1, ↑AMOTL1, ↑BAD, ↑BCL2L...	...all 55	103 (5)
PLK1	↑0.060	9.90E-01	kinase		-1.778	7.47E-02	↓ANAPC1, ↑ATM, ↑BRCA2, ↓CDC20,	...all 24	
CHEK2	↑0.270	1.81E-02	kinase		0.161	8.34E-04	↑BRCA1, ↑CDC25A, ↓CDC25C, ↑CD...	...all 13	23 (3)
MARK4			kinase		-1.091	2.28E-02	↑MAP2, ↓MAP4, ↓MAPT, ↓RPTOR	...all 4	
CDC7	↑1.040	1.12E-02	kinase		0.770	9.71E-03	↑CLSPN, ↑MCM2, ↑MCM3, ↓MCM4, ↑...	...all 6	
ATM	↑0.620	4.89E-03	kinase	Activated	3.156	2.83E-06	↑ATF2, ↑ATM, ↑BCL11A, ↑BRCA1, ↑...	...all 44	58 (6)
EGFR	↑0.560	2.82E-02	kinase		1.844	4.95E-01	↑AGO2, ↑ATF2, ↓CAV1, ↑CBL, ↓CLTC	...all 27	
AURKB			kinase		-0.850	1.39E-05	↑ATM, ↓CDCA8, ↓CENPC, ↓DES, ↑...	...all 19	
MAPK9			kinase		1.352	1.45E-02	↑ATF2, ↑BAD, ↓CCND3, ↑CDC25A, ↓...	...all 18	
HUNK			kinase			3.44E-02	↑ARHGEF2, ↑EGFR, ↓LIMK1	...all 3	
PKN2	↓-0.330	3.77E-02	kinase		-0.577	1.75E-02	↑BAD, ↓CDC25B, ↑CTNNB1, ↓CTTN, ↑...	...all 6	
YES1			kinase		1.914	4.47E-01	↑CBL, ↑CTNNB1, ↑EGFR, ↑GIT2, ↑PR...	...all 5	
RPS6KB2			kinase		0.557	2.37E-03	↑BAD, ↓EIF4B, ↑IRS1, ↓RNF168, ↑RP...	...all 5	67 (4)
RPS6KB1			kinase		0.618	3.65E-05	↓AKT1S1, ↑BAD, ↓CREB1, ↑DDX20, ↓...	...all 18	73 (4)
CDKN1A	↓-1.420	4.96E-02	kinase		-0.503	1.46E-02	↑ATM, ↑CDC6, ↑CDK1, ↑CHEK1, ↑E...	...all 11	
PRKCQ	↓-0.110	5.02E-01	kinase		0.480	6.89E-03	↑ARHGEF6, ↑BAD, ↑CBL, ↓CCDC88A	...all 12	72 (4)
DBF4	↑0.640	1.18E-02	kinase		0.444	2.37E-03	↑EPHA2, ↑MCM2, ↓MCM4, ↑MCM6,	...all 5	
SPHK1			kinase		-1.941	1.00E00	↑EIF4EBP1, ↓FLNA, ↓GSK3B, ↑JUN	...all 4	
ERBB3			kinase		1.941	3.69E-01	↑EGFR, ↑MTOR, ↑NDRG1, ↓PDCD4, ↑...	...all 7	
PIK3C3			kinase			4.72E-02	↑EGFR, ↑EIF4EBP1	...all 2	
CSF1R			kinase	Activated	2.219	4.94E-01	↑CBL, ↑CTNNB1, ↑CTNND1, ↑GAB2, ↑...	...all 5	
PIP5K1C			kinase			4.72E-02	↑EGFR, ↓WNK1	...all 2	
MAP4K4	↓-0.950	7.58E-03	kinase	Inhibited	-2.408	4.36E-02	↑EIF4EBP1, ↓MAP3K11, ↑MTOR, ↓PS...	...all 6	
CSNK2A2			kinase	Activated	2.415	3.83E-01	↑CTNNB1, ↑MAP1B, ↑MRE11, ↑NASP,	...all 6	
CDK6			kinase		1.705	4.83E-07	↑ABI1, ↑BCL11A, ↑CASC3, ↑CBY1, ↑...	...all 40	50 (3)
IKBKG			kinase		1.767	4.69E-01	↑HTT, ↑IRS1, ↑JUN, ↑NCOR2, ↑NFKB1	...all 6	
PIM2			kinase		0.369	1.76E-03	↓API5, ↑BAD, ↓CDKN1A, ↓EIF4B, ↑E...	...all 10	

MAP2K1 1



Analysis for Protein-Protein Interactions

Analysis of interactome data is unique

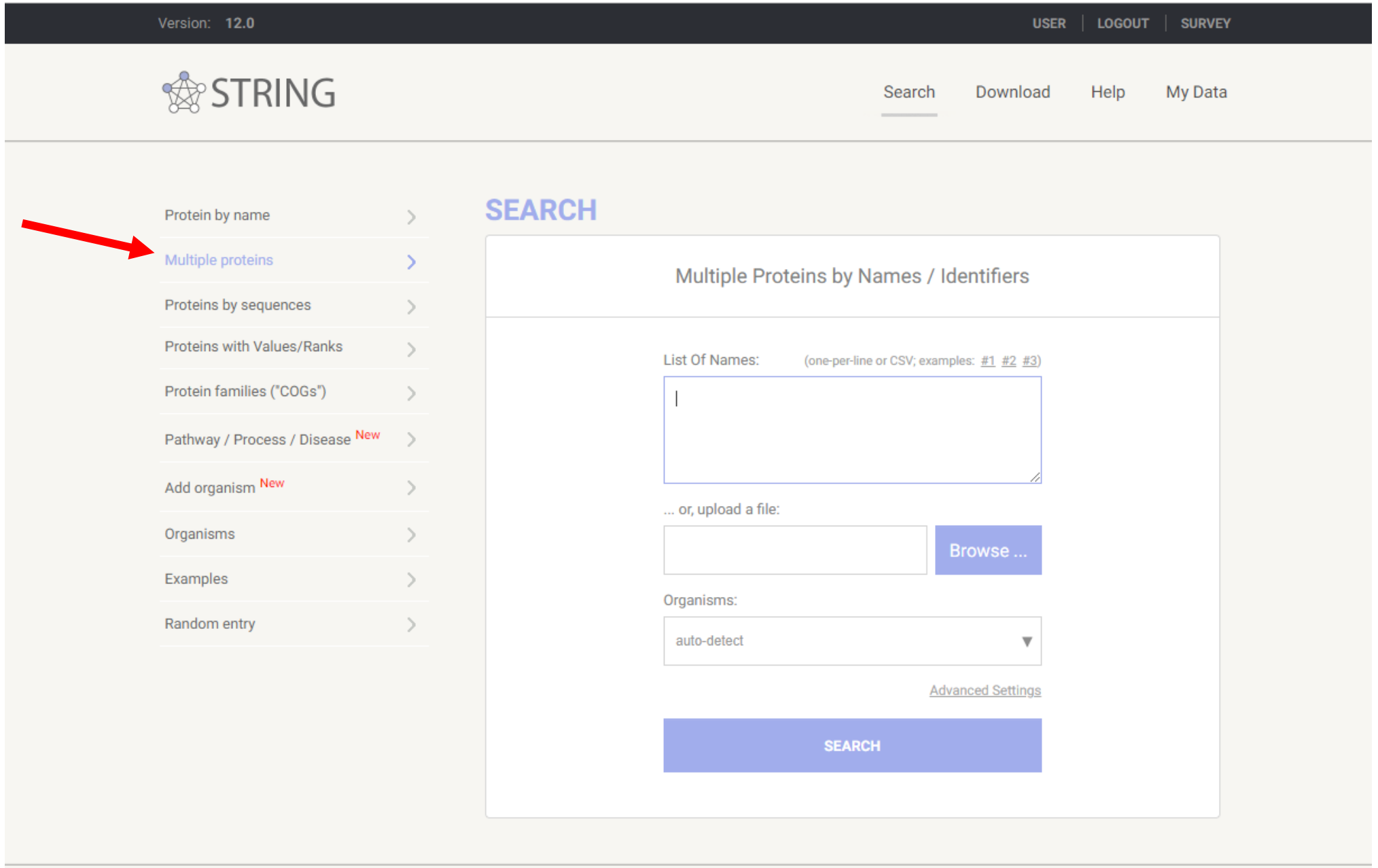
- by nature of the experiment, working with subset of proteome
- can require different normalization approaches, for example normalize to the bait protein
- generally focus on only a single direction of change (increase with bait, not decrease with bait)
- it can be helpful to identify PPI to characterize when a complex is coming down

CRAPome – Contaminant Repository for Affinity Purification

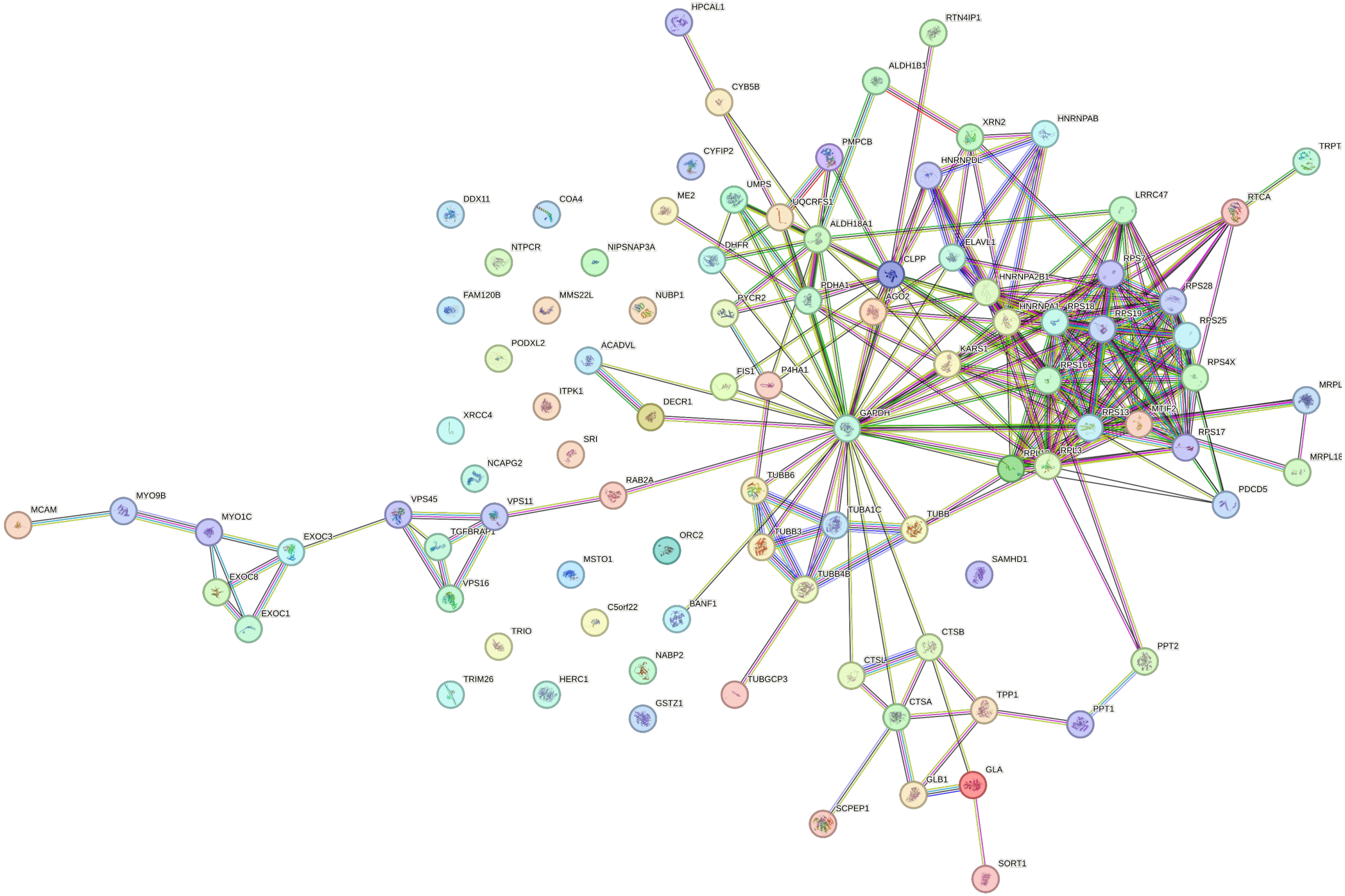
- large database of standardized negative controls, aggregated from several leading labs
- provides a qualitative and semiquantitative indication of how likely a given protein is to be identified as a "nonspecific" interactor
- for example, chaperone proteins frequently co-precipitate with overexpressed proteins
- bear to mind even those proteins that are frequently pulled out could have a specific interaction

STRING is a publicly-available tool for mapping protein-protein interaction networks and performing functional enrichment analysis

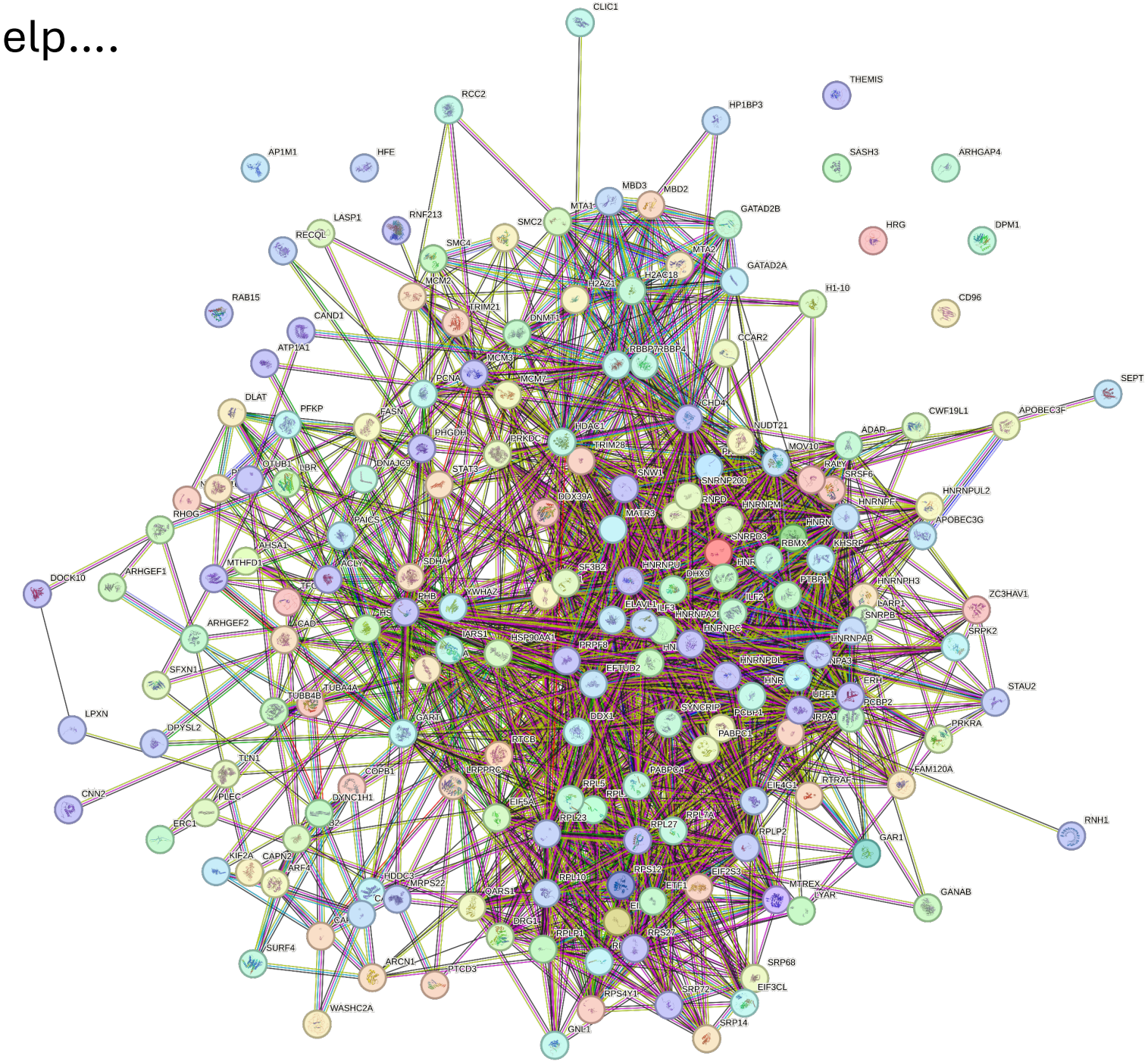
- allows visualize interaction networks and perform functional enrichment



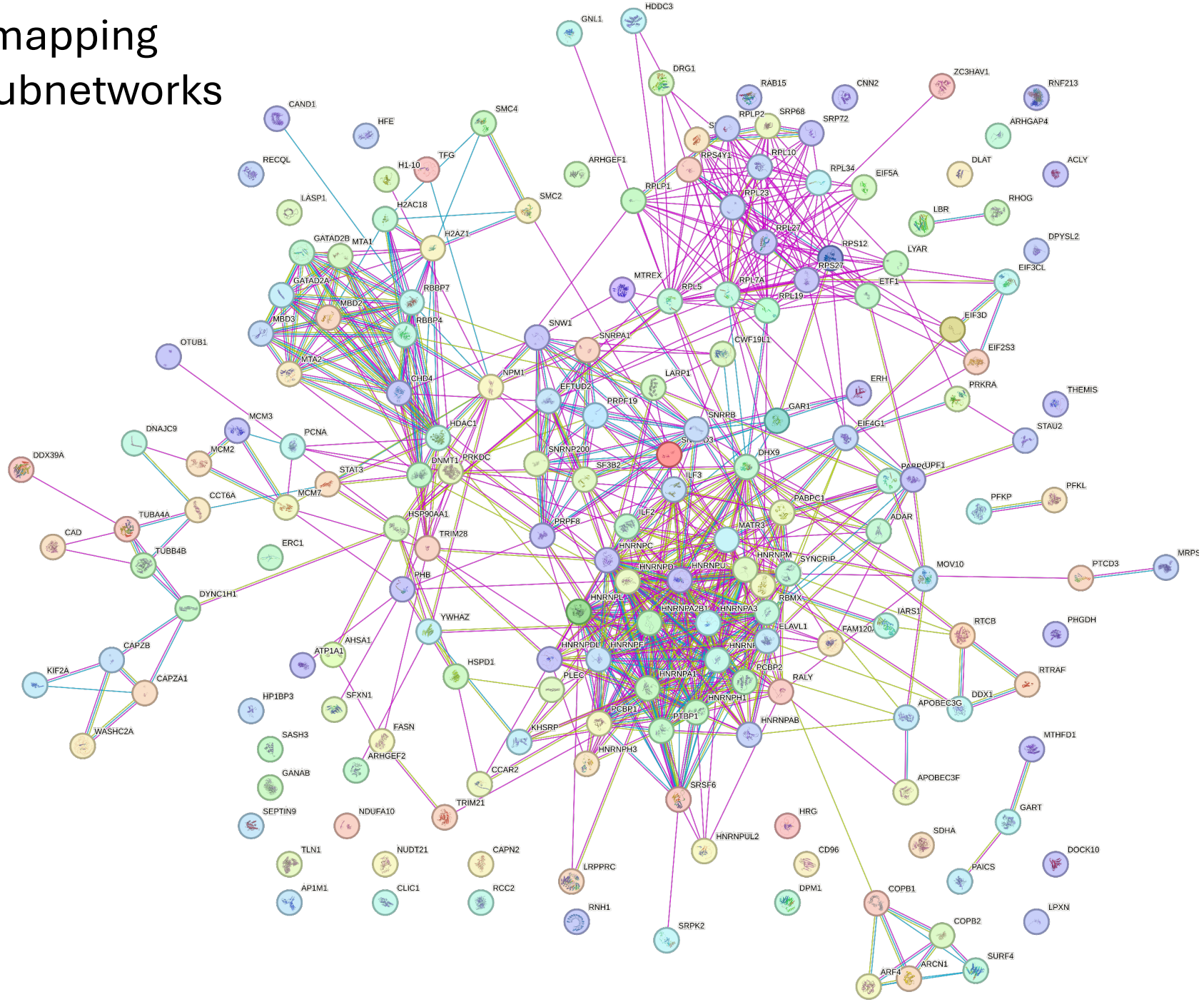
String-db.org


HPCAL1
RTN4IP1
ALDH1B1
CYB5B
XRN2
HNRNPAB
PMPCB
CYFIP2
HNRNPDL
TRPT
UMPS
LRRC47
RTCA
DDX11
COA4
ME2
UQCRFS1
ALDH18A1
ELAVL1
NTPCR
NIPSNAP3A
DHFR
CLPP
RPS7
HNRNPA2B1
RPS28
FAM120B
MMS22L
NUBP1
PDHA1
PYCR2
AGO2
HNRNPA1
RPS18
RPS19
RPS25
PODXL2
ACADVL
KARS1
RPS16
RPS4X
FIS1
P4HA1
MRPL
ITPK1
DECR1
GAPDH
RPS13
MTIF2
XRCC4
RPS17
SRI
RPL3
NCAPG2
MRPL18
TUBB6
RAB2A
PDCD5
MCAM
MYO9B
VPS45
VPS11
TUBA1C
TUBB
MYO1C
TGFBRAP1
TUBB3
ORC2
EXOC3
SAMHD1
MSTO1
TUBB4B
EXOC8
VPS16
BANF1
C5orf22
EXOC1
CTSB
TRIO
PPT2
CTSL
NABP2
TUBGCP3
TRIM26
HERC1
CTSA
TPP1
PPT1
GSTZ1
GLB1
GLA
SCPEP1
SORT1

Too big does not help....



Same dataset mapping
only physical subnetworks



STRING also provides functional enrichment of the dataset

Functional enrichments in your network

explain columns

Biological Process (Gene Ontology)

GO-term	*description*	*count in network*	*strength*	*false discovery rate*
GO:1905696	Regulation of polysome binding	2 of 2	2.04	0.0333
GO:1905663	Positive regulation of telomerase RNA reverse transcriptase...	2 of 2	2.04	0.0333
GO:0006346	DNA methylation-dependent heterochromatin assembly	4 of 7	1.79	0.00026
GO:0070934	CRD-mediated mRNA stabilization	5 of 11	1.69	3.63e-05
GO:1900152	Negative regulation of nuclear-transcribed mRNA catabolic ...	5 of 12	1.66	4.91e-05

(more ...)

Molecular Function (Gene Ontology)

GO-term	*description*	*count in network*	*strength*	*false discovery rate*
GO:0035851	Krueppel-associated box domain binding	2 of 2	2.04	0.0317
GO:0030623	U5 snRNA binding	2 of 2	2.04	0.0317
GO:0016418	S-acetyltransferase activity	2 of 2	2.04	0.0317
GO:0030942	Endoplasmic reticulum signal peptide binding	2 of 3	1.86	0.0475
GO:0003872	6-phosphofructokinase activity	2 of 3	1.86	0.0475

(more ...)

Summary mass spectrometry-based quantification

- Multiple strategies with advantages and disadvantages, right approach depends on experimental question
- Relative quantitation of protein across conditions most common, comparison different proteins in the same sample requires special methods
- Replication is important
- Data analysis will link quantitation to gene symbol so that downstream tools developed for genomics also useful
- Experimental design in proteomics is variable, so it is important to consider how the experiment was performed when determining downstream analysis approach

Questions about a specific project?

Please feel free to reach out:

Ron Holewinski: ronald.holewinski@nih.gov

Lisa Jenkins: lisa.jenkins@nih.gov