Computational Tools for Mass Spectrometry based Proteomic Data

Chris Adams
Stanford University Mass Spectrometry

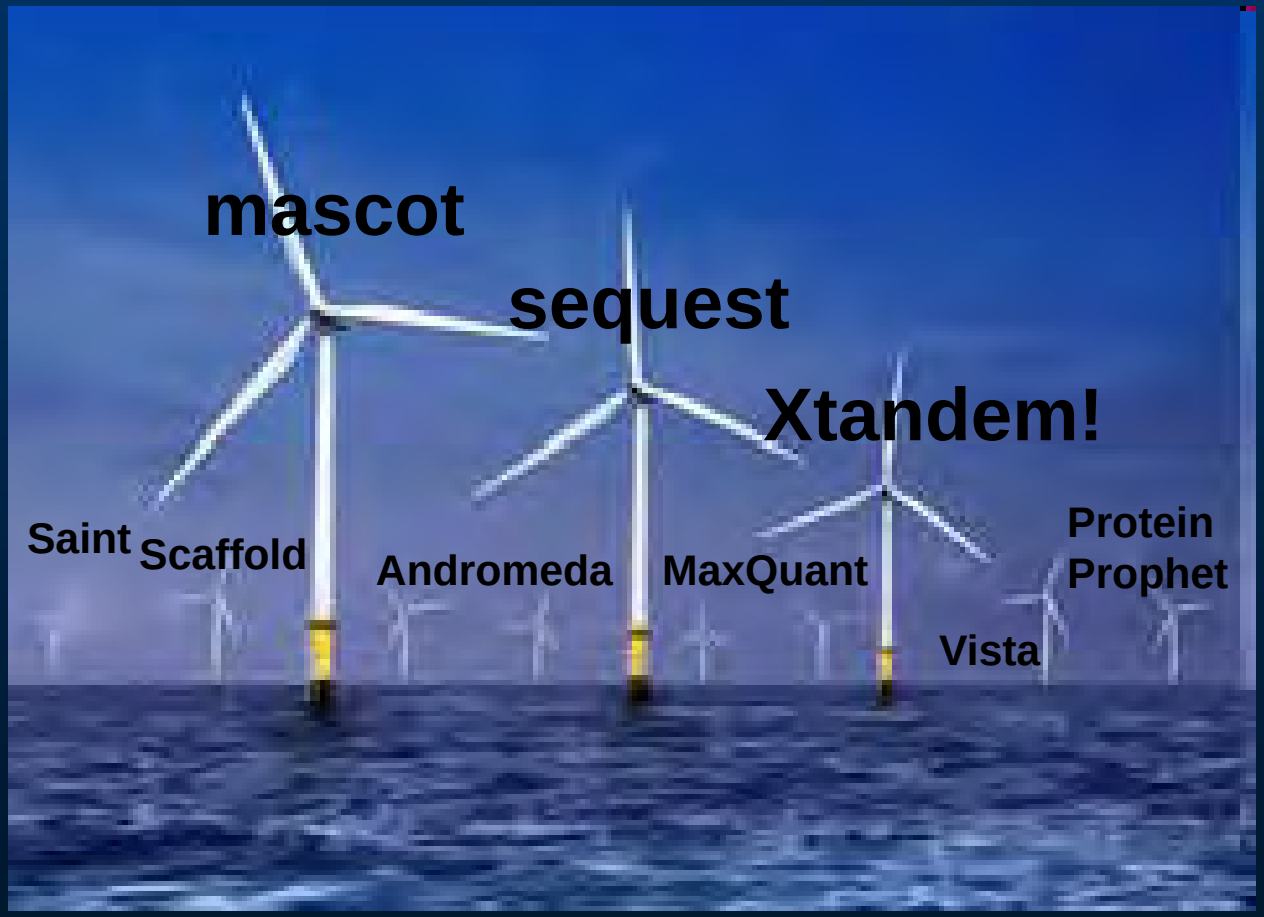
http://mass-spec.stanford.edu/

Presented at the 2011 Stanford Mass Spectrometry Users' Meeting

For personal use only.
Please do not reuse or reproduce without the author's permission.

Proteonancial Advisor???

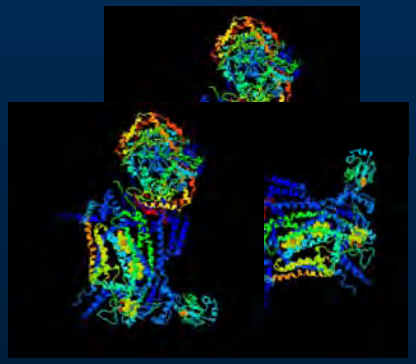
Deciding which tools work for you.....



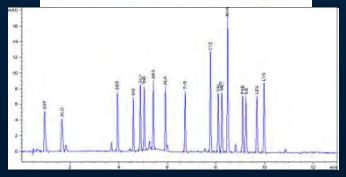
Shotgun Proteomics Works from the "Bottom Up"

Start with a mixture of proteins

Intermediary experiment?



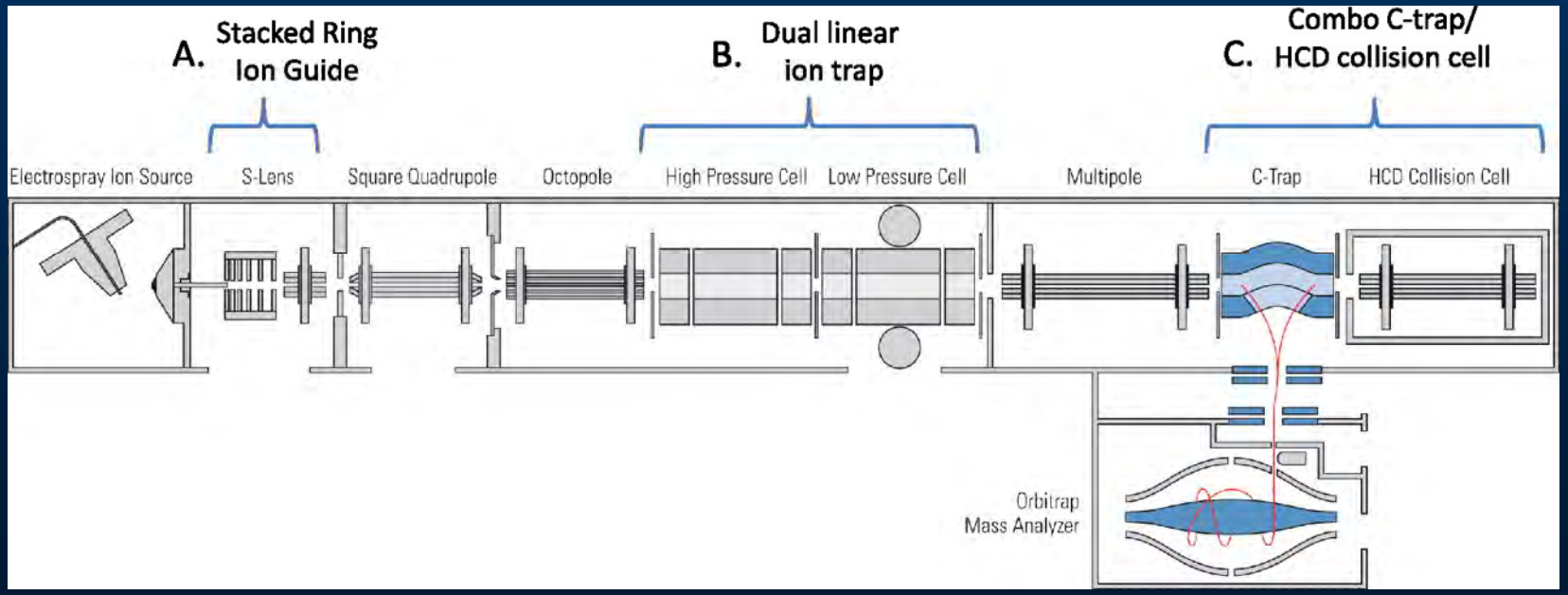
Digestion



ESI

The Guts of the Mass Spectrometer

Schematic of the LTQ Orbitrap Velos MS instrument with three new hardware implementations.

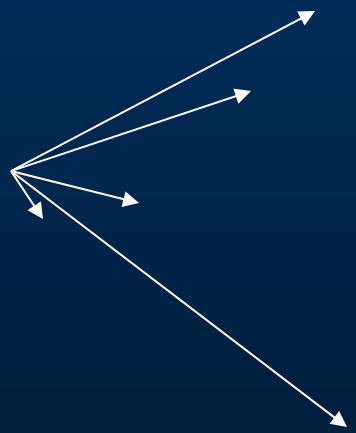


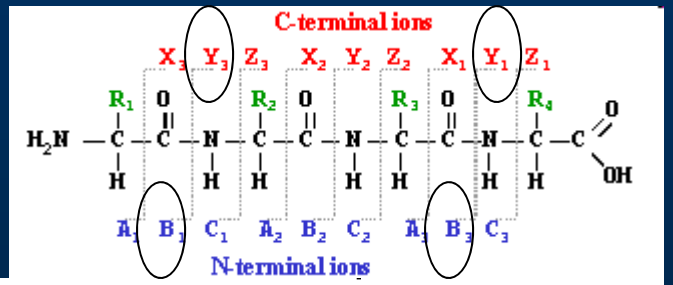
Olsen J V et al. Mol Cell Proteomics 2009;8:2759-2769

Fragmentation via Vibrational Excitation (CAD, CID)

SPAFDSIMAETLK

He



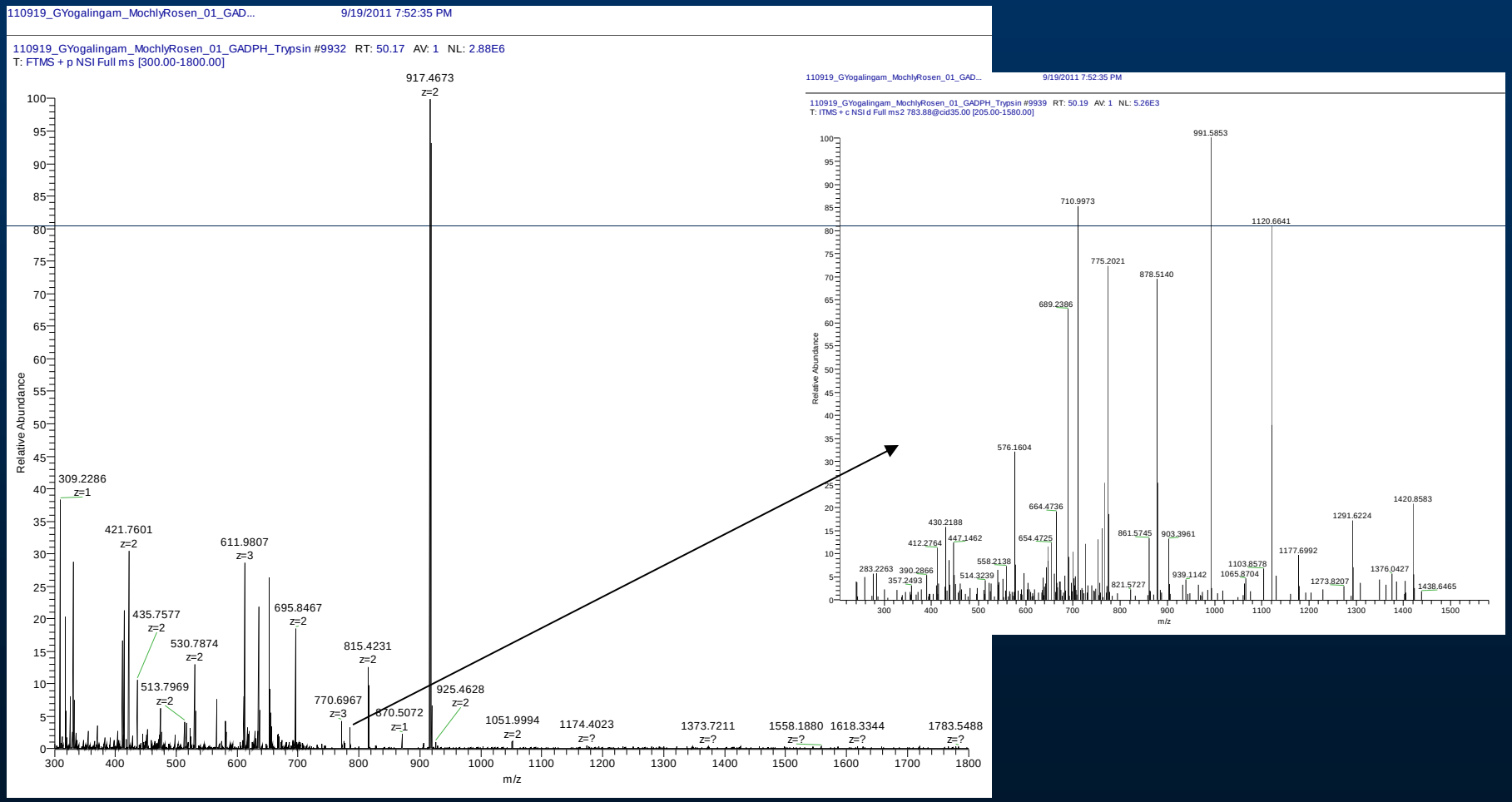


S-P-A-F-D-S-I-M-A-E-T-L-K
(protonated mass 1410.6)

mass+	b-ions	y-ions	mass+
88.1	S	PAFDSIMAETLK	1323.6
185.2	SP	AFDSIMAETLK	1226.4
256.3	SPA	FDSIMAETLK	1155.4
403.5	SPAF	DSIMAETLK	1008.2
518.5	SPAFD	SIMAETLK	893.1
605.6	SPAFDS	IMAETLK	806.0
718.8	SPAFDSI	MAETLK	692.3
850.0	SPAFDSIM	AETLK	561.7
921.1	SPAFDSIMA	ETLK	490.6
1050.2	SPAFDSIMAE	TLK	361.5
1151.3	SPAFDSIMAET	LK	260.4
1264.4	SPAFDSIMAETL	K	147.2

(ECD, ETD)

Pieces of Information



Stanford University
MASS SPECTROMETRY

MANY tools- a small briefing

PI

Aebersold (ISB) — Mallick, Nesvizhskii

Gygi

Mann

Smith

Yates — MacCoss

http://www.nesvilab.org/software.html
https://gygi.med.harvard.edu/index.html/software
http://www.systemsbiology.org/Public_Resources/Downloadable_Software
http://proteowizard.sourceforge.net/downloads.shtml
http://www.biochem.mpg.de/mann/publications/index.html
http://omics.pnl.gov/
http://fields.scripps.edu/downloads.php

Tool(s)

SuperHirn
msConvert/proteowizard
Peptide Prophet
Protein Prophet
Saint

A-score
RDDBS

MaxQuant
Andromeda
Perseus

DanteR
DeconMSn
Viper

Sequest
Skyline
ProLucid

Nature Methods
Nesvizhskii et. al 2007

35 programs!

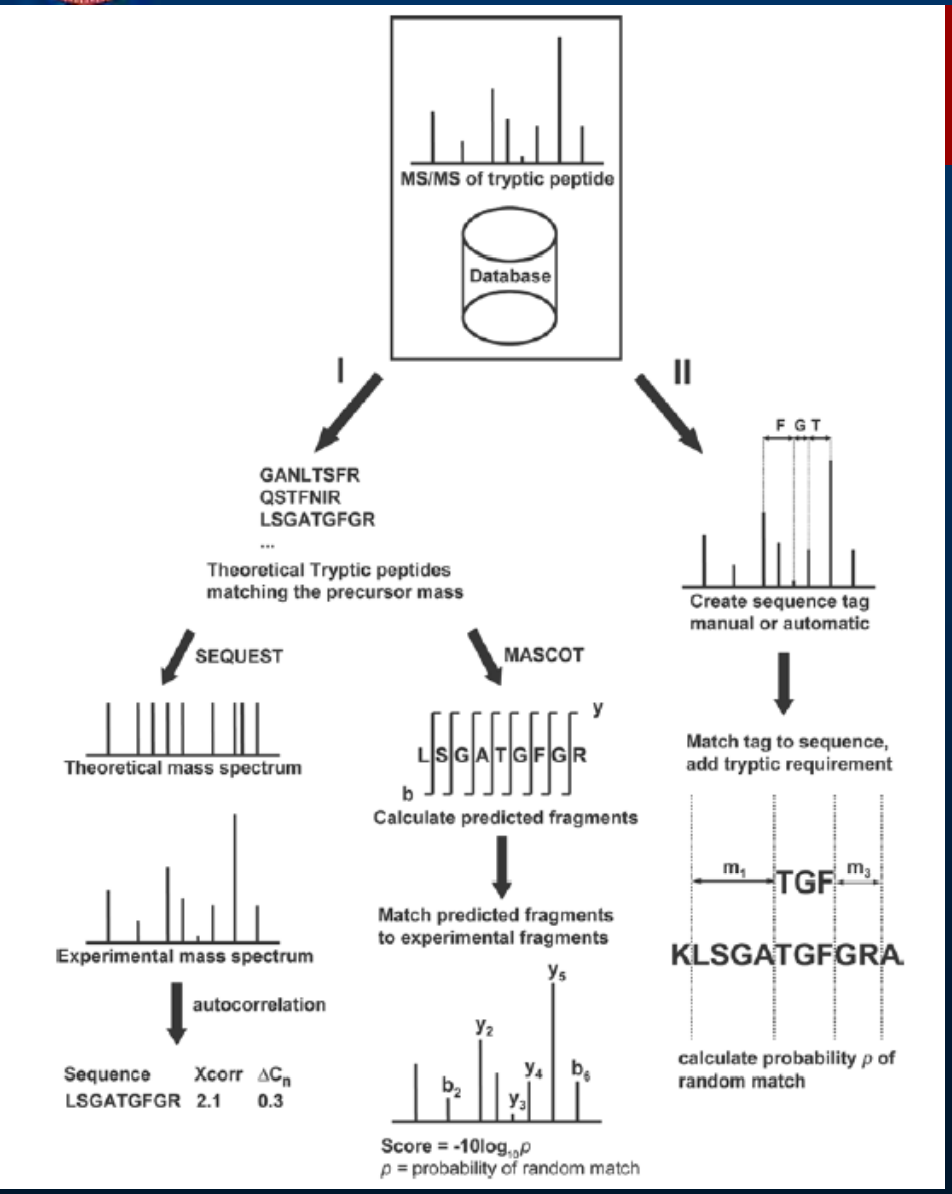
Table 1 | A list of publicly available tools for MS/MS-based proteomics

Program	Reference	Website
Database search tools		
SEQUEST	84	http://www.thermo.com
MASCOT	85	http://matrixscience.com[a]
ProteinProspector	86	http://prospector.ucsf.edu[a]
ProbID	87	http://tools.proteomecenter.org/wiki/index.php?title=Software:ProbID[b]
TANDEM	88	http://www.thegpm.org[a,b]
SpectrumMill		http://www.chem.agilent.com
Phenyx	89	http://www.phenyx-ms.com
OMSSA	4	http://pubchem.ncbi.nlm.nih.gov/omssa[a,b]
VEMS	90	http://personal.cicbiogune.es/rmatthiesen[b]
MyriMatch	91	http://www.mc.vanderbilt.edu/msrc/bioinformatics[b]
Spectral matching tools		
SpectraST	12	http://www.peptideatlas.org/spectrast
X! P3	92	http://p3.thegpm.org/tandem/ppp.html
Biblispec	11	http://proteome.gs.washington.edu/bibliospec
De novo sequencing tools		
Lutefisk	93	http://www.hairyfatguy.com/lutefisk[b]
PepNovo	94	http://peptide.ucsd.edu/pepnovo.py[a,b]
PEAKS	95	http://www.bioinformaticssolutions.com
Sequit		http://www.proteomefactory.com
Sequence tag/hybrid approaches		
GutenTag	16	http://fields.scripps.edu/GutenTag
Inspect	17	http://peptide.ucsd.edu/inspect.html[a,b]
Popitam	96	http://www.expasy.org/tools/popitam
Statistical validation of peptide and protein identifications		
PeptideProphet	21	http://www.proteomecenter.org/software.php[b]
ProteinProphet	56	http://www.proteomecenter.org/software.php[b]
Scaffold		http://www.proteomesoftware.com
Databases for storing and mining of mass spectrometry data		
PeptideAtlas	97	http://www.peptideatlas.org
Proteios		http://www.proteios.org
SBEAMS		http://sbeams.org
CPAS	98	https://www.labkey.org
PRIDE	99	http://www.ebi.ac.uk/pride
Data sharing		
Tranche		http://www.proteomecommons.org/dev/dfs[b]
Tools for protein quantification		
PEPPER (label free)	74	http://www.broad.mit.edu/cancer/software/genepattern[b]
EXPRES (isotopic labeling)		http://www.proteomecenter.org/software.php[b]
Libra (iTRAQ isobaric labeling)		http://www.proteomecenter.org/software.php[b]
ASAPRatio (label free)	100	http://www.proteomecenter.org/software.php[b]
MSQuant (isotopic labeling)		http://msquant.sourceforge.net[b]
RelEx (isotopic labeling)	101	http://fields.scripps.edu/relex

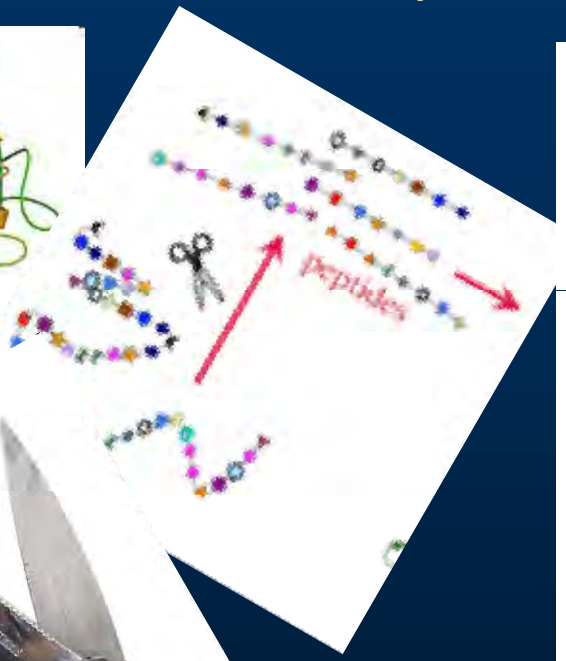
[a]Free access through the web interface (functionality might be limited). [b]Free distribution.

© 2007 Nature Publishing Group http://www.nature.com/naturemethods

Identifying Peptides



Critical Nature of Proteases Combined with High Mass Accuracy for DB Searching



Benefits of Accurate Mass Measurements

Distinguishing K/Q (36.4 mDa) :

With 1 ppm accuracy, K/Q can be distinguished in peptides up to 3.8 kDa.

Identification of post-translational modifications :

Tryptic peptide, m = 997.514 +/- 0.001 Da (1 ppm accuracy)

Did not match any peptide sequences predicted for methionine-tRNA ligase

Assumption: it is a N-terminal peptide with Met_1 removed and Ser_2 acetylated.

Testing: calculated mass m=997.5126 Da, agrees with the measured value

(Gibson, B.W.; Biemann, K. Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 1956)

Data base search :

Tryptic peptide, m = 1513.794 +/- 0.001 (1 ppm accuracy)

Unique sequence, GAAFICAIHSPTLR, is found in the yeast genome.

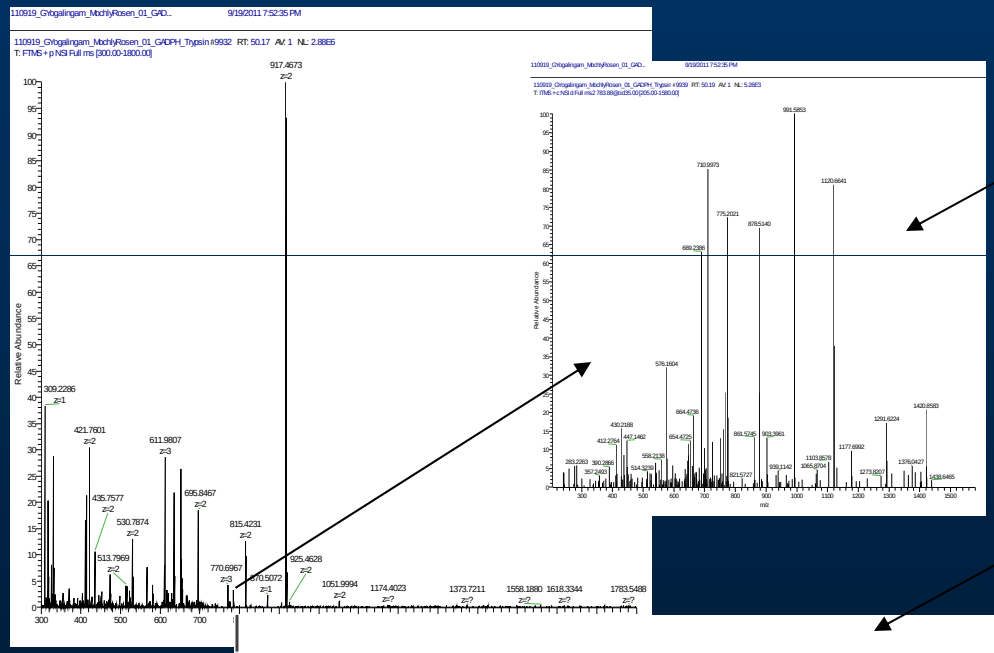
Four sequences are found in a non-redundant database of 203 000 entries; including TFHRIQQMLPDK with the same elemental composition.

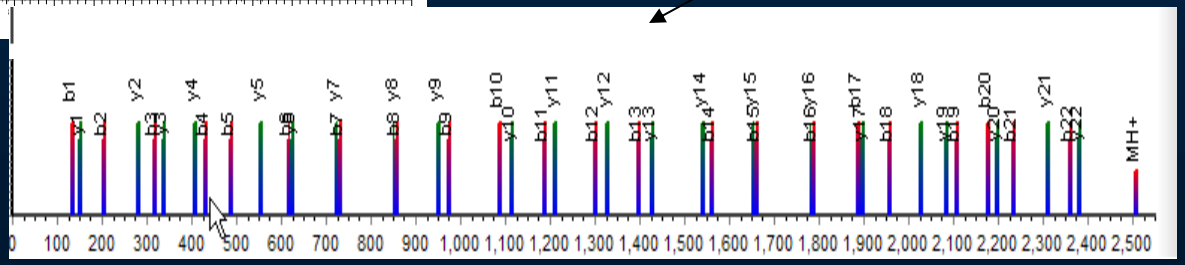
With 1 ppm accuracy, tryptic peptides may be unique for a small genome

(Jensen, O.N.; Podtelejnikov, A.; Mann, M. Rapid Commun. Mass Spectrom. 10 (1996) 1371)

From precursor mass measurement we can provide suggestions

Heuristic – ie Sequest





Similitude

Experimental

vs.

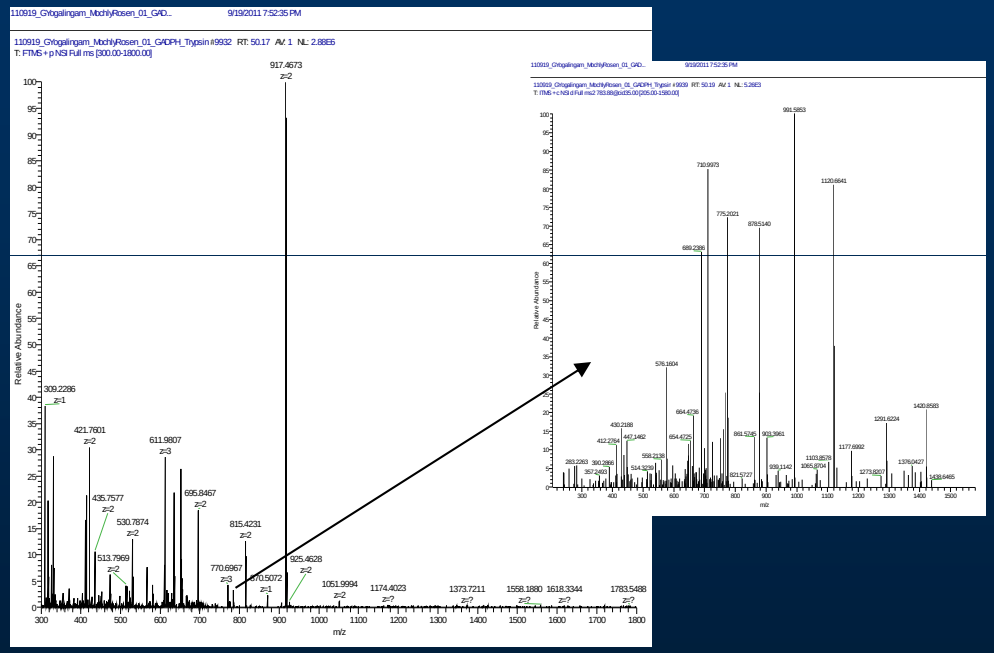
Theoretical

Xcorr- "Fit"

ΔCN
How sure in relation to other peptides (best v. next best)

XTandem! Considers Ion intensities

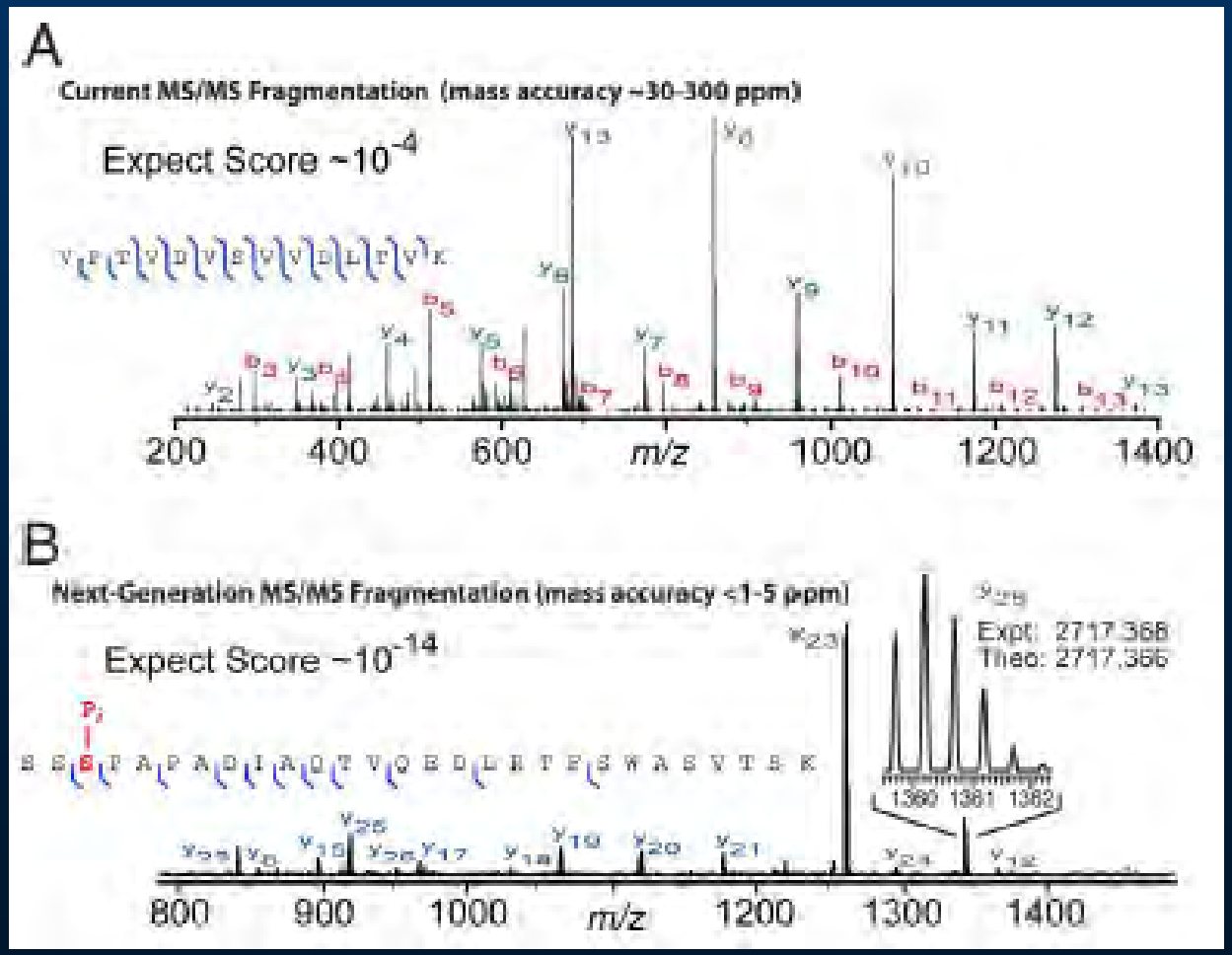
Probability Based – ie Mascot/Andromeda



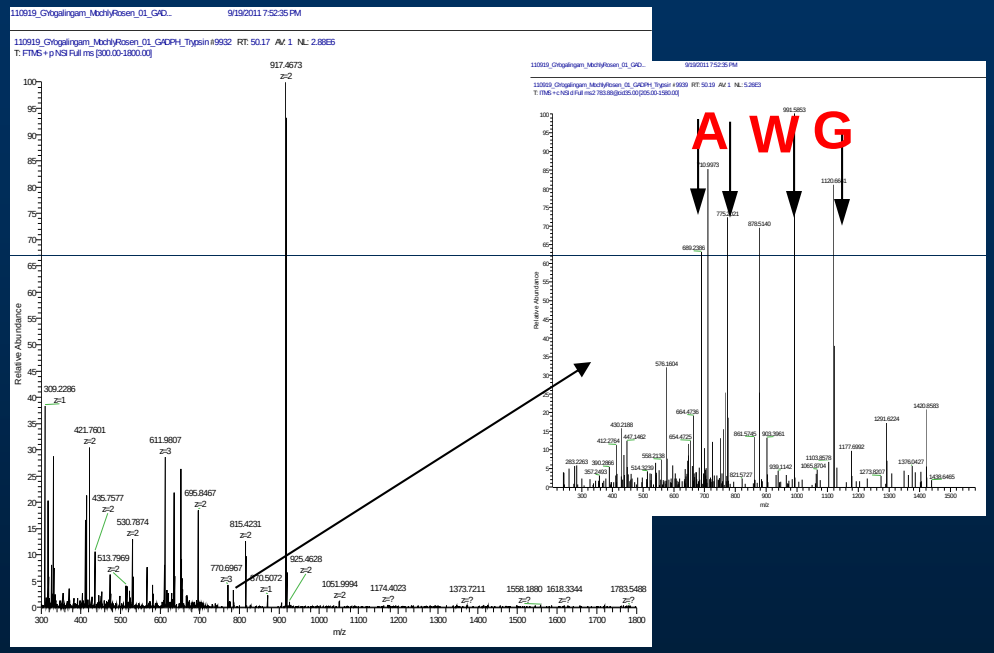
Calculating a probability that the observed number of matches between calculated and experimental fragments could have occurred by chance.

Cox et. Al, J. Proteome Res. 2010

Mass accuracy helps p-values in a probability based model



Database Independent – sequence tag approach/ DeNovo



ID a "sequence tag" from the MS/MS spectra and subsequently perform a hybrid search or attempt to DeNovo

Determining a FDR

Target Decoy Database Approach

MAGPRPSPWARLLLAALISVSLSGTLANRCKKAPVKSCTECVRVDKDCAYCTDEMFRDRRCNTQAELLAAGCQRESIVVMESSFQITEETQIDTTLRRSQMSP
QGLRVRLRPGEERHFELEVFEPLESPVDLYILMDFSNSMSDDLDNLKKMGQNLARVLSQLTSDYTIGFGKFVDKVSVPQTDMRPEKLKEPWPNSDPPFSFKNV
ISLTEDVDEFRNKLQGERISGNLDAPEGGFDAILQTAVCTRDIGWRPDSTHLLVFSTESAFHYEADGANVLAGIMSRNDERCHLDTTGTYTQYRTQDYPSVPTLV
RLLAKHNIIPIFAVTNYSYSYYEKLHTYFPVSSLGVLQEDSSNIVELLEEAFNRIRSNLDIRALDSPRGLRTEVTSKMFQKTRTGSFHIRRGEVGIYQVQLRALEHV
DGTHVCQLPEDQKGNIHLKPSFSDGLKMDAGIICDVCTCELQKEVRSARCSFNGDFVCGQCVCSEGWSGQTCNCSTGSLSDIQPCLREGEDKPCSGRGECQC
GHCVCYGEGRYEGQFCEYDNFQCPRTSGFLCNDRGRCSMGQCVCEPGWTGPSCDCPLSNATCIDSNGGICNGRGHCECGRCHCHQQSLYTDTICEINYSAIH
PGLCEDLRSCVQCQAWGTGEKKGRTCEECNFKVKMVDELKRAEEVVVRCSFRDEDDDCTYSYTMEGDGAPGPNSTVLVHKKKDCPPGSFWWLIPLLLLLLP
LLALLLLLCWKYCACCKACLALLPCCNRGHMVGFKEDHYMLRENLMASDHLDTPMLRSGNLKGRDVVRWKVTNNMQRPGFATHAASINPTELVPYGLSLRL
ARLCTENLLKPDTRECAQLRQEVEENLNEVYRQISGVHKLQQTKFRQQPNAGKKQDHTIVDTVLMAPRSAKPALLKLTEKQVEQRAFHDLKVAPGYYTLTAD
QDARGMVEFQEGVELVDVRVPLFIRPEDDDEKQLLVEAIDVPAGTATLGRRLVNITIIKEQARDVVSFEQPEFSVSRGDQVARIPVIRRVLDGGKSQVSYRTQDG
TAQGNRDYIPVEGELLFQPGEAWKELQVKLLELQEVDSLLRGRQVRRFHVQLSNPKFGAHLGQPHSTTIIIRDPDELDRSFTSQMLSSQPPPHGDLGAPQNPN
AKAAGSRKIHFNWLPPSGKPMGYRVKYWIQGDSESEAHLLDSKVPSVELTNLYPYCDYEMKVCAYGAQGEGPYSSLVSCRTHQEVPSEPGRLAFNVVSSTVT
QLSWAEPAETNGEITAYEVCYGLVNDDNRPIGPMKKVLVDNPKNRMLLIENLRESQPYRYTVKARNGAGWGPEREAIINLATQPKRPMSIPIIPDIPIVDAQSGED
YDSFLMYSDDVLRSPSGSQRPSVSDDTEHLVNGRMDFAFPGSTNSLHRMTTTSAAAYGTHLSPHVPHRVLSTSSTLTRDYNSLTRSEHSHSTTLPRDYSTLTS
VSSHDSRLTAGVPDTPTRLVFSALGPTSLRVSWQEPRCERPLQGYSVEYQLLNGGELHRLNIPNPAQTSVVVEDLLPNHSYVFRVRAQSQEGWGREREGVITI
ESQVHPQSPLCPLPGSAFTLSTPSAPGPLVFTALSPDSLQLSWERPRRPNGDIVGYLVTCEMAQGGGPATAFRVDGDSPESRLTVPGLSENVPYKFKVQARTT
EGFGPEREGIITIESQDGGPFPQLGSRAGLFQHPLQSEYSSITTTHTSATEPFLVDGPTLGAQHLEAGGSLTRHVTQEFVSRTLTTSGTLSTHMDQQFFQT

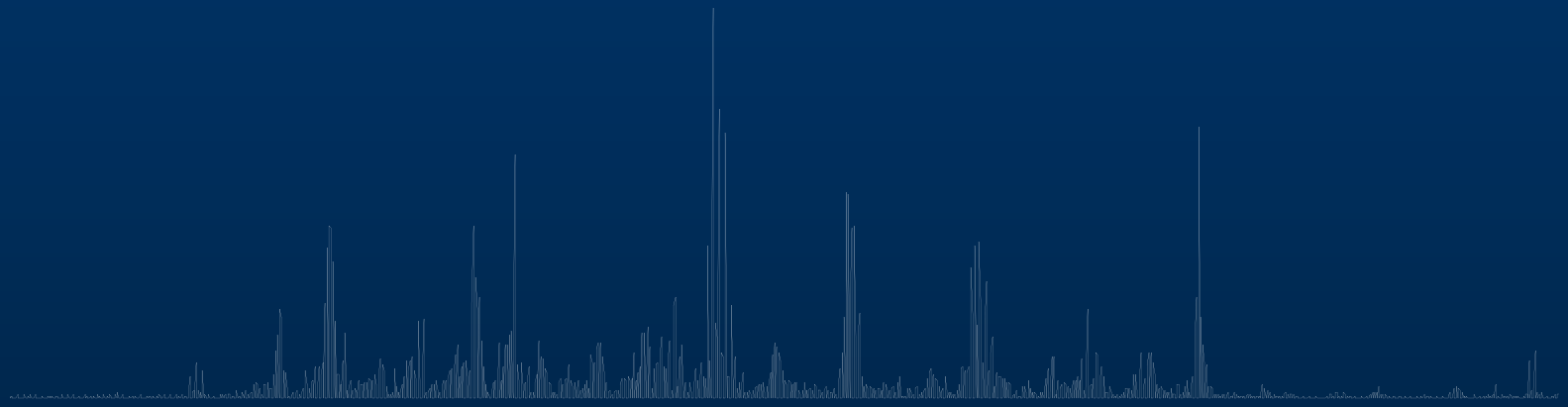
Elias et. al. Nat. Meth. 2007

FDR offers transparency

Independent of the search of choice, instrument, ect. data can be filtered

Global peptide FDR known

Protein Inference:



A typical LC MSMS run of a complex mixture Produces 15000 fragment ion spectra- of which 50-60+ % are ID'd-

Now decide which proteins they belong to……..

Protein Inference: highly homologous proteins and redundant databases

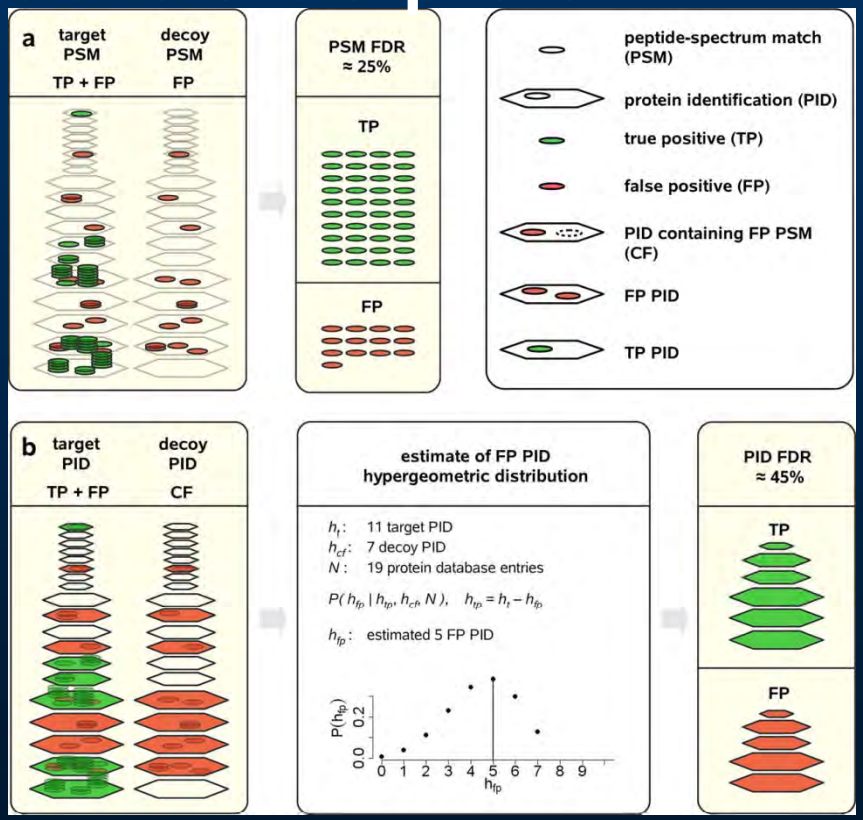
Proteins with high degree of sequence homology

Alternative splice forms

Differentially processed proteins

Protein Inference: some strategies available

Generate a protein FDR



Reiter L et al. Mol Cell Proteomics 2009;8:2405-2417

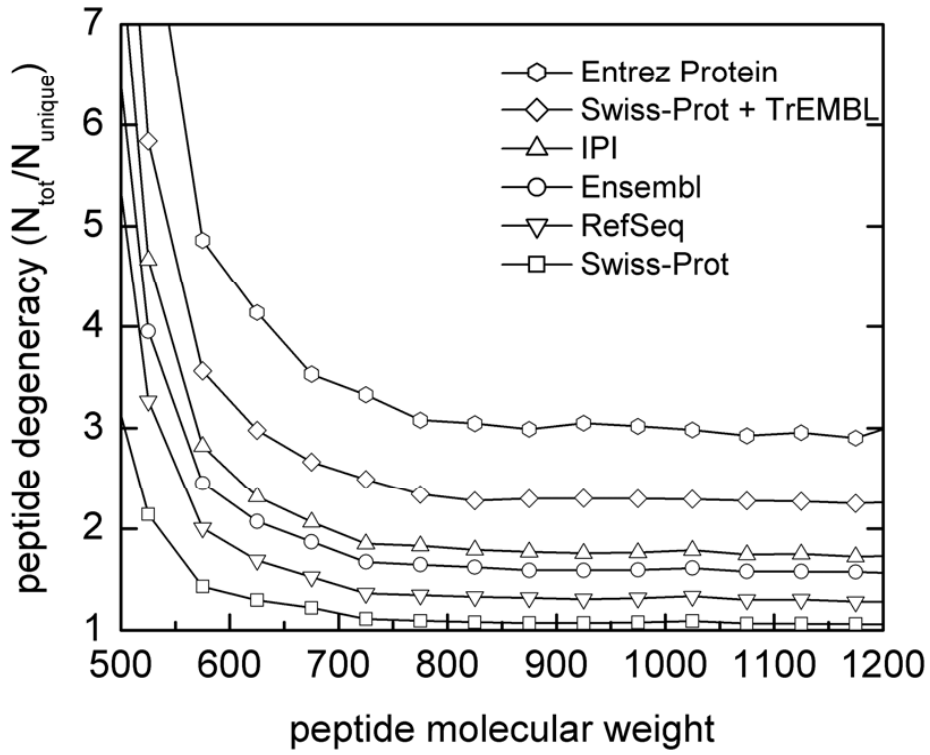
Hierarchical grouping

Group by family

But…..

Kall et al. Nat. Meth. 2007

Databases



Unique peptide sequence

Or

Degenerate

Visualizing Your Data- Scaffold

Scaffold Viewer - Samples - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% Min # Peptides: 1 Min Peptide: 95%

Load Data | Samples | Proteins | Similarity | Quantify | Publish | Statistics

Display Options: Number of Unique Peptides Req Mods: No Filter Search:

Probability Legend:
- over 95%
- 80% to 94%
- 50% to 79%
- 20% to 49%
- 0% to 19%

#	Visible?	Starred?	Bio View: Identified Proteins (276)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	B5_23C 0729_B5_23C	0729_B5_23C_02	0729_B5_23C_03	0729_B5_23C_04	0729_B5_23C_05	0729_B5_23C_06	B5_37C 0729_B5_37C	0729_B5_37C_02	0729_B5_37C_03	0729_B5_37C_04	0729_B5_37C_05	0729_B5_37C_06
1	✓	☆	IPI:IPI00783987.2 Tax_Id=9606 Gene_Symbol=C3 Complement C3 (Fragm...	IPI00783987	187 kDa	★	52	66	61	58	63	51	70	75	62	60	59	49
2	✓	☆	IPI:IPI00021369.1 Tax_Id=9606 Gene_Symbol=CRYAB Alpha-crystallin B c...	IPI00021369	20 kDa		13	18	18	17	19	15	17	19	16	21	18	17
3	✓	☆	IPI:IPI00298497.3 Tax_Id=9606 Gene_Symbol=FGB Fibrinogen beta chain	IPI00298497	56 kDa		16	20	20	18	20	18	16	21	20	19	22	18
4	✓	☆	IPI:IPI00022229.1 Tax_Id=9606 Gene_Symbol=APOB Apolipoprotein B-100	IPI00022229	516 kDa	★	54	73	65	25	28	24	23	34	29	23	28	20
5	✓	☆	IPI:IPI00745872.2 Tax_Id=9606 Gene_Symbol=ALB Isoform 1 of Serum al...	IPI00745872	69 kDa	★	23	36	31	30	39	25	30	31	26	28	26	21
6	✓	☆	IPI:IPI00032291.2 Tax_Id=9606 Gene_Symbol=C5 Complement C5	IPI00032291	188 kDa	★	12	16	18	17	17	17	21	23	24	35	38	29
7	✓	☆	IPI:IPI00478003.1 Tax_Id=9606 Gene_Symbol=A2M Alpha-2-macroglobulin	IPI00478003	163 kDa	★	5	12	12	24	24	22	17	21	21	27	30	22
8	✓	☆	IPI:IPI00021885.1 Tax_Id=9606 Gene_Symbol=FGA Isoform 1 of Fibrinoge...	IPI00021885	95 kDa		9	19	16	17	16	14	14	16	12	15	13	9
9	✓	☆	IPI:IPI00021891.5 Tax_Id=9606 Gene_Symbol=FGG Isoform Gamma-B of F...	IPI0002189...	52 kDa		8	14	14	13	13	10	12	16	11	15	12	8
10	✓	☆	IPI:IPI00941961.1 Tax_Id=9606 Gene_Symbol=- 66 kDa protein	IPI00941961	66 kDa	★	13	16	13	13	15	12	14	14	12	15	13	10
11	✓	☆	IPI:IPI00423463.1 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00423463	53 kDa	★	8	13	11	10	13	13	11	10	11	13	14	10
12	✓	☆	IPI:IPI00302592.2 Tax_Id=9606 Gene_Symbol=FLNA Isoform 2 of Filamin-A	IPI0030259...	280 kDa		19	35	31	20	28	32	1	1	1	17	34	16
13	✓	☆	IPI:IPI00892604.1 Tax_Id=9606 Gene_Symbol=C4A;C4B Complement com...	IPI00892604	193 kDa	★	5	11	8	7	11	9	12	13	14	7	10	10
14	✓	☆	IPI:IPI00291262.3 Tax_Id=9606 Gene_Symbol=CLU Isoform 1 of Clusterin	IPI0029126...	52 kDa		6	13	10	11	13	9	11	11	10	13	10	7
15	✓	☆	IPI:IPI00021439.1 Tax_Id=9606 Gene_Symbol=ACTB Actin, cytoplasmic 1	IPI0002143...	42 kDa	★	10	12	11	11	12	10	6	9	8	11	11	9
16	✓	☆	IPI:IPI00654755.3 Tax_Id=9606 Gene_Symbol=HBB Hemoglobin subunit b...	IPI00654755	16 kDa	★	10	12	10	11	11	7	10	11	8	11	9	7
17	✓		IPI:IPI00298994.6 Tax_Id=9606 Gene_Symbol=TLN1 Talin-1	IPI00298994	270 kDa	★	19	29	24	27	28	24	1	1	1	23	25	17
18	✓		IPI:IPI00019580.1 Tax_Id=9606 Gene_Symbol=PLG Plasminogen	IPI00019580	91 kDa	★	9	13	9	8	13	7	4	10	6	13	12	10

Protein Information:

Lookup Accession Number In: NCBI (ie:gi|1351907,ALBU_BOVIN,P0...

IPI00783987

Preferred Accession Number: IPI00783987

Protein Name: IPI:IPI00783987.2 Tax_Id=9606 Gene_Symbol=C3 Com

Sample Information:

Biological Sample:

Sample Category:

Sample Description:

MS/MS Sample:

MS/MS Sample Notes:

276 Proteins
0.1% Prot %FDR
50008 Spectra
5.3% Pept %FDR

8:53 PM
9/24/2011

Visualizing Your Data- Scaffold Tools and Caveats

Scaffold Viewer - Samples - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% | Min # Peptides: 1 | Min Peptide: 95%

Load Data | Samples | Proteins | Similarity | Quantify | Publish | Statistics

Display Options: Unweighted Spectrum Count | Req Mods: No Filter | Search:

Probability Legend:
- over 95%
- 80% to 94%
- 50% to 79%
- 20% to 49%
- 0% to 19%

Bio View: Identified Proteins (276)

#	Visible?	Starred?	Identified Proteins	Accession Number	Molecular Weight	Protein Grouping Ambiguity	0729_85_23C											
1	✓	☆	IPI:IPI00021369.1 Tax_Id=9606 Gene_Symbol=CRYAB Alpha-crystallin B c...	IPI00021369	20 kDa		86											
2	✓	☆	IPI:IPI00783987.2 Tax_Id=9606 Gene_Symbol=C3 Complement C3 (Fragm...	IPI00783987	187 kDa	★	74											
3	✓	☆	IPI:IPI00022229.1 Tax_Id=9606 Gene_Symbol=APOB Apolipoprotein B-100	IPI00022229	516 kDa	★	57											
4	✓	☆	IPI:IPI00745872.2 Tax_Id=9606 Gene_Symbol=ALB Isoform 1 of Serum al...	IPI00745872	69 kDa	★	32											
5	✓	☆	IPI:IPI00941961.1 Tax_Id=9606 Gene_Symbol=- 66 kDa protein	IPI00941961	66 kDa		23											
6	✓	☆	IPI:IPI00298497.3 Tax_Id=9606 Gene_Symbol=FGB Fibrinogen beta chain	IPI00298497	56 kDa		22											
7	✓	☆	IPI:IPI00302592.2 Tax_Id=9606 Gene_Symbol=FLNA Isoform 2 of Filamin-A	IPI00030259...	280 kDa		20											
8	✓	☆	IPI:IPI00298994.6 Tax_Id=9606 Gene_Symbol=TLN1 Talin-1	IPI00298994	270 kDa	★	20											
9	✓	☆	IPI:IPI00654755.3 Tax_Id=9606 Gene_Symbol=HBB Hemoglobin subunit b...	IPI00654755	16 kDa	★	19	22	17	17	17	13	16	16	14	17	16	13
10	✓	☆	IPI:IPI00021439.1 Tax_Id=9606 Gene_Symbol=ACTB Actin, cytoplasmic 1	IPI0002143...	42 kDa	★	18	24	23	18	24	17	7	12	11	17	20	17
11	✓	☆	IPI:IPI00784842.1 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00784842	52 kDa	★	15	22	15	14	19	21	16	16	19	18	21	20
12	✓	☆	IPI:IPI00032291.2 Tax_Id=9606 Gene_Symbol=C5 Complement C5	IPI00032291	188 kDa	★	14	17	19	18	21	18	24	27	28	41	45	39
13	✓	☆	IPI:IPI00021891.5 Tax_Id=9606 Gene_Symbol=FGG Isoform Gamma-B of F...	IPI0002189...	52 kDa		14	24	23	20	22	17	17	26	18	25	18	18
14	✓	☆	IPI:IPI00423463.1 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00423463	53 kDa	★	14	23	16	17	20	21	16	17	19	22	23	19
15	✓	☆	IPI:IPI00785084.2 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00785084	51 kDa	★	14	22	15	13	19	22		16		19	22	
16	✓	☆	IPI:IPI00021885.1 Tax_Id=9606 Gene_Symbol=FGA Isoform 1 of Fibrinoge...	IPI00021885	95 kDa		13	28	25	24	24	20	17	22	18	22	26	20
17	✓	☆	IPI:IPI00029168.1 Tax_Id=9606 Gene_Symbol=LPA Apolipoprotein(a)	IPI00029168	501 kDa	★	13	17	17	4	4	4	6	6	8	4	5	2
18	✓	☆	IPI:IPI00784985.1 Tax_Id=9606 Gene_Symbol=IGK@ IGK@ protein	IPI00784985	26 kDa	★	13	17	17	15	19	15	14	17	13	19	15	12

276 Proteins
0.1% Prot %FDR
50008 Spectra
5.3% Pept %FDR

Protein Information:

Lookup Accession Number In: NCBI (ie:gi|1351907,ALBU_BOVIN,P0...

Preferred Accession Number:

Protein Name:

Sample Information:

Biological Sample:

Sample Category:

Sample Description:

MS/MS Sample:

MS/MS Sample Notes:

8:56 PM
9/24/2011

Scaffold cont.

Scaffold Viewer - Quantify - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% | Min # Peptides: 1 | Min Peptide: 95%

Load Data
Samples
Proteins
Similarity
Quantify
Publish
Statistics

276 Proteins
0.1% Prot %FDR
50008 Spectra
5.3% Pept %FDR

Quantitative Value

IPI:IPI00783987.2 Tax_Id=9606 ...

Normalized Spectrum Count

Biological Sample

Protein Venn Diagram

Proteins | Unique Peptides | Unique Spectra

1: B5_23C 2: B5_37C 3: B5_42C

B5_23C B5_37C B5_42C
12 3 2 161 22 19 47

Accession Protein N...

Quantitative Scatterplots

Q-Q Scatterplot | Mean/Deviation Scatterplot

X: B5_23C Y: B5_37C

B5_23C/B5_37C Q-Q Scatterplot

B5_37C

B5_23C

Significant Insignificant Optimal 45° Lower Error Bar High Error Bar

Gene Ontology Terms

There is no Gene Ontology Term data to display

9:03 PM
9/24/2011

Scaffold and Spectral Counting

Spectral Counting

- Spectral count correlates well with protein abundance
- Fold change can be calculated and statistically evaluated
- Simple and straightforward implementation
- Sensitive to protein abundance changes – for abundant proteins 2 fold change easily detected with high confidence

Limitations

- The response to increasing protein amount is saturable
- Noisy data at low spectral counts – **large** difference in spectral count necessary to determine significant change

Scaffold and Degenerate Peptides

Scaffold Viewer - Similarity - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% | Min # Peptides: 1 | Min Peptide: 95%

Load Data | Samples | Proteins | Similarity | Quantify | Publish | Statistics

276 Proteins
0.1% Prot %FDR
50008 Spectra
5.3% Pept %FDR

IPI:IPI00783987.2 Tax_Id=9606 Gene_Symbol=C3 Complement C3 (Fragment)

Index	Peptide	Exclusive To	Valid	IPI:IPI00783987.2... IPI00783987	IPI:IPI00887739.3 ... IPI00887739	No Group IPI00942927	IPI00816741	IPI00739237	IPI00419215	IPI00448322	IPI00871685	IPI00916318	IPI00936567
1	AAVYHHFISDGVR		✓	95%	95%	95%							
2	AAVYHHFISDGVRK		✓	95%	95%	95%							
3	ACEPGVDYVYK	IPI:IPI00783987.2 T...	✓	95%		95%		95%					
4	ADIGCTPGSGK		✓	95%	95%	95%							
5	ADIGCTPGSGKDYAGVFSDAG...		✓	95%	95%	95%							
6	AEDLVGKSLYVSATVILHSGSD...		✓	95%	95%	95%							
7	AELQCPQPAAR		✓	95%	95%	95%							
8	AFSDRNTLIIYLDK	IPI:IPI00783987.2 T...	✓	95%				95%					
9	AFTECCVVASQLR		✓				95%						
10	AGDFLEANYMNLQR		✓	95%	95%								
11	AIAYSSLSQSYLYIDWTDNHK		✓				95%						
12	ALLVGEHLNIIVTPK		✓				95%						
13	APSTWLTAYVVK		✓	95%	95%	95%							
14	AVLYNYR		✓	95%	95%								
15	AYYENSPQQVFSTEFEVK		✓	95%	95%	95%							
16	CAEENCFIQK	IPI:IPI00783987.2 T...	✓	95%				95%					
17	CCYDGACVNNDETCEQR		✓				95%						
18	DAPDHQELNLDVSLQLPSR	IPI:IPI00783987.2 T...	✓	95%				95%					
19	DFDFVPPVVR		✓	95%	95%								
20	DICEEQVNSLPGSITK		✓	95%	95%								
21	DMALTAFVLISLQEAK		✓	95%	95%								
22	DSCVGSLVVK		✓	95%	95%	95%							
23	DSITTWEILAVSMSDK		✓	95%	95%								
24	DSITTWEILAVSMSDKK		✓	95%	95%								
25	DSLDQLVGGVPVTLNAQTIDV...		✓				95%						

Identifications | Spectrum | Spectrum/Model Error | Fragmentation Table

Sequence	Prob	SEQU...	SEQU...	NTT	Modifications	Observed	Actual Mass	Charge	Delta ...	Delta ...	Spectrum ID	Category	Bio Sample	MS/MS Sa...
(K)AAVYHHFISDGVR(K)	95%	2.02	0.35	2		491.25	1,470.74	3	0.0037	2.5	100729_Rothbar...	G120_23C	0729_G120...	100729_Rot...
(K)AAVYHHFISDGVR(K)	95%	1.75	0.44	2		736.38	1,470.74	2	0.0039	2.7	100729_Rothbar...	G120_23C	0729_G120...	100729_Rot...
(K)AAVYHHFISDGVR(K)	95%	2.86	0.57	2		491.25	1,470.74	3	-0.0015	-1.1	100729_Rothbar...	G120_42C	0729_G120...	100729_Rot...
(K)AAVYHHFISDGVR(K)	95%	2.67	0.50	2		491.26	1,470.75	3	0.013	9.0	100729_Rothbar...	G120_42C	0729_G120...	100729_Rot...
(K)AAVYHHFISDGVR(K)	95%	3.34	0.45	2		736.37	1,470.73	2	-0.0075	-5.1	100729_Rothbar...	G120_42C	0729_G120...	100729_Rot...

Peptide sequence

9:05 PM
9/24/2011

Statistics display experimental outliers

Scaffold Viewer - Statistics - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% | Min # Peptides: 1 | Min Peptide: 95%

Load Data | Samples | Proteins | Similarity | Quantify | Publish | Statistics

276 Proteins
0.1% Prot %FDR
50008 Spectra
5.3% Pept %FDR

Category	Bio Sample	MS/MS Sample	#Prot	#IDs	#Spec	%IDs
G120_23C	0729_G120_23C	100729_Rothbard_MS_G12...	156	1126	3307	34%
G120_23C	0729_G120_23C_2	100729_Rothbard_MS_G12...	159	1039	3106	33%
G120_23C	0729_G120_23C_3	100729_Rothbard_MS_G12...	188	1541	5003	31%
G120_23C	0729_G120_23C_4	100729_Rothbard_MS_G12...	154	1026	2758	37%
G120_23C	0729_G120_23C_5	100729_Rothbard_MS_G12...	165	1066	3085	35%
G120_23C	0729_G120_23C_6	100729_Rothbard_MS_G12...	175	1202	3588	34%
G120_37C	0729_G120_37	100729_Rothbard_MS_G12...	161	1040	2786	37%
G120_37C	0729_G120_37_2	100729_Rothbard_MS_G12...	193	1557	5141	30%
G120_37C	0729_G120_37_3	100729_Rothbard_MS_G12...	175	1161	3457	34%
G120_37C	0729_G120_37_4	100729_Rothbard_MS_G12...	153	850	2469	34%
G120_37C	0729_G120_37_5	100729_Rothbard_MS_G12...	184	1234	3922	31%
G120_37C	0729_G120_37_6	100729_Rothbard_MS_G12...	183	1259	4061	31%
G120_42C	0729_G120_42C	100729_Rothbard_MS_G12...	227	2610	8230	32%
G120_42C	0729_G120_42_02	100729_Rothbard_MS_G12...	172	1180	3121	38%
G120_42C	0729_G120_42_03	100729_Rothbard_MS_G12...	210	2299	7229	32%
G120_42C	0729_G120_42_04	100729_Rothbard_MS_G12...	215	2498	7073	35%
G120_42C	0729_G120_42_05	100729_Rothbard_MS_G12...	154	998	2259	44%
G120_42C	0729_G120_42_06	100729_Rothbard_MS_G12...	213	2156	5992	36%
B5_23C	0729_B5_23C	100729_Rothbard_MS_B5_...	144	819	2742	30%
B5_23C	0729_B5_23C_02	100729_Rothbard_MS_B5_...	168	1241	4056	31%
B5_23C	0729_B5_23C_03	100729_Rothbard_MS_B5_...	158	1120	3639	31%
B5_23C	0729_B5_23C_04	100729_Rothbard_MS_B5_...	138	949	3217	29%
B5_23C	0729_B5_23C_05	100729_Rothbard_MS_B5_...	153	1152	4054	28%
B5_23C	0729_B5_23C_06	100729_Rothbard_MS_B5_...	149	979	3162	31%
B5_37C	0729_B5_37C	100729_Rothbard_MS_B5_...	110	759	2486	31%
B5_37C	0729_B5_37C_02	100729_Rothbard_MS_B5_...	127	1008	3237	31%
B5_37C	0729_B5_37C_03	100729_Rothbard_MS_B5_...	117	872	2752	32%
B5_37C	0729_B5_37C_04	100729_Rothbard_MS_B5_...	146	1111	3519	32%
B5_37C	0729_B5_37C_05	100729_Rothbard_MS_B5_...	160	1200	3819	31%
B5_37C	0729_B5_37C_06	100729_Rothbard_MS_B5_...	151	971	3066	32%
B5_42C	0729_B5_42C	100729_Rothbard_MS_B5_...	190	1976	6164	32%
B5_42C	0729_B5_42C_02	100729_Rothbard_MS_B5_...	198	2110	6641	32%
B5_42C	0729_B5_42C_03	100729_Rothbard_MS_B5_...	188	1783	5392	33%
B5_42C	0729_B5_42C_04	100729_Rothbard_MS_B5_...	208	2203	6145	36%

Peptide ROC Plots | Protein Probability Calculation

Number of Identified Spectra

Peptide False Positive Rate

Sequest

Need More Data

At least two analysis programs are required to display scatterplots

Sequest

Charge +2 | Charge +3 | Charge +4 | Charge +5

9:11 PM
9/24/2011

Number of peptide ID's/protein and protein FDR

Scaffold Viewer - Samples - 100729_MS_All_JRothbard_ipi_human

File Edit View Experiment Export Quant Window Help

Min Protein: 95.0% | Min # Peptides: 3 | Min Peptide: 95%

Display Options: Quantitative Value | Req Mods: No Filter | Search:

Probability Legend:
- over 95%
- 80% to 94%
- 50% to 79%
- 20% to 49%
- 0% to 19%

#	Visible?	Starred?	MS/MS View: Identified Proteins (193)	Accession Number	Molecular Weight	Protein Grouping Ambiguity	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_23C 0729... 100729_Rothbard_MS_B5_23C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...	B5_37C 0729... 100729_Rothbard_MS_B5_37C_0...
1	☑	☆	IPI:IPI00021369.1 Tax_Id=9606 Gene_Symbol=CRYAB Alpha-crystallin B c...	IPI00021369	20 kDa		146	168	197	186	196	168	193	217	187	218	166	185
2	☑	☆	IPI:IPI00783987.2 Tax_Id=9606 Gene_Symbol=C3 Complement C3 (Fragm...	IPI00783987	187 kDa	★	125	105	112	116	110	112	203	178	182	95	94	110
3	☑	☆	IPI:IPI00022229.1 Tax_Id=9606 Gene_Symbol=APOB Apolipoprotein B-100	IPI00022229	516 kDa	★	97	87	81	37	34	36	45	49	49	30	33	29
4	☑	☆	IPI:IPI00745872.2 Tax_Id=9606 Gene_Symbol=ALB Isoform 1 of Serum al...	IPI00745872	69 kDa	★	54	52	52	65	68	58	83	64	68	48	40	51
5	☑	☆	IPI:IPI00941961.1 Tax_Id=9606 Gene_Symbol=- 66 kDa protein	IPI00941961	66 kDa	★	39	30	27	32	31	41	32	29	29	31	29	29
6	☑	☆	IPI:IPI00298497.3 Tax_Id=9606 Gene_Symbol=FGB Fibrinogen beta chain	IPI00298497	56 kDa		37	34	36	41	45	54	49	51	55	37	50	62
7	☑	☆	IPI:IPI00302592.2 Tax_Id=9606 Gene_Symbol=FLNA Isoform 2 of Filamin-A	IPI00302592	280 kDa		34	41	39	29	34	48	2	1	2	21	41	23
8	☑	☆	IPI:IPI00298994.6 Tax_Id=9606 Gene_Symbol=TLN1 Talin-1	IPI00298994	270 kDa	★	34	34	33	40	37	35	2	1	2	28	31	25
9	☑	☆	IPI:IPI00654755.3 Tax_Id=9606 Gene_Symbol=HBB Hemoglobin subunit b...	IPI00654755	16 kDa	★	32	24	21	25	21	19	30	22	23	21	18	19
10	☑	☆	IPI:IPI00021439.1 Tax_Id=9606 Gene_Symbol=ACTB Actin, cytoplasmic 1	IPI0002143...	42 kDa	★	31	26	28	27	29	25	13	17	18	21	23	25
11	☑	☆	IPI:IPI00032291.2 Tax_Id=9606 Gene_Symbol=C5 Complement C5	IPI00032291	188 kDa	★	24	18	23	27	25	26	45	38	46	51	52	57
12	☑	☆	IPI:IPI00021891.5 Tax_Id=9606 Gene_Symbol=FGG Isoform Gamma-B of F...	IPI0002189...	52 kDa		24	26	28	29	27	25	32	36	29	31	21	26
13	☑	☆	IPI:IPI00423463.1 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00423463	53 kDa	★	24	25	10	25	24	30	30	24	31	27	26	28
14	☑	☆	IPI:IPI00785084.2 Tax_Id=9606 Gene_Symbol=LOC100290320;IGHV4-31;L...	IPI00785084	51 kDa	★	24	24	18	19	23	32		22		23	25	

Protein Information:

Lookup Accession Number In: NCBI (ie:gi|1351907,ALBU_BOVIN,P0...

Preferred Accession Number:

Protein Name:

Sample Information:

Biological Sample:

Sample Category:

Sample Description:

MS/MS Sample:

MS/MS Sample Notes:

Load Data | Samples | Proteins | Similarity | Quantify | Publish | Statistics

193 Proteins
0.0% Prot %FDR
48796 Spectra
5.3% Pept %FDR

9:19 PM
9/24/2011

Other Tools for Post Processing re-analyzing .RAW data

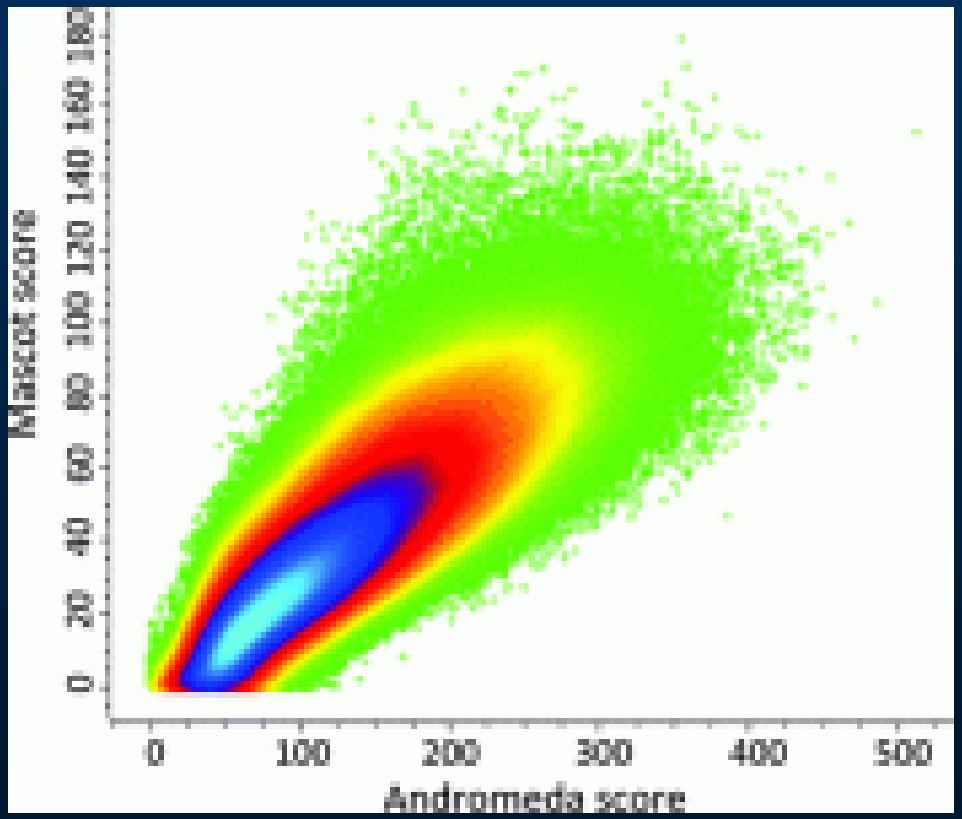
MaxQuant/Andromeda/SILAC

- **Free!**
- **Powerful**
- **Works on your PC**
- **Proficient with SILAC**

http://maxquant.org/

Andromeda

MaxQuant's- Mascot

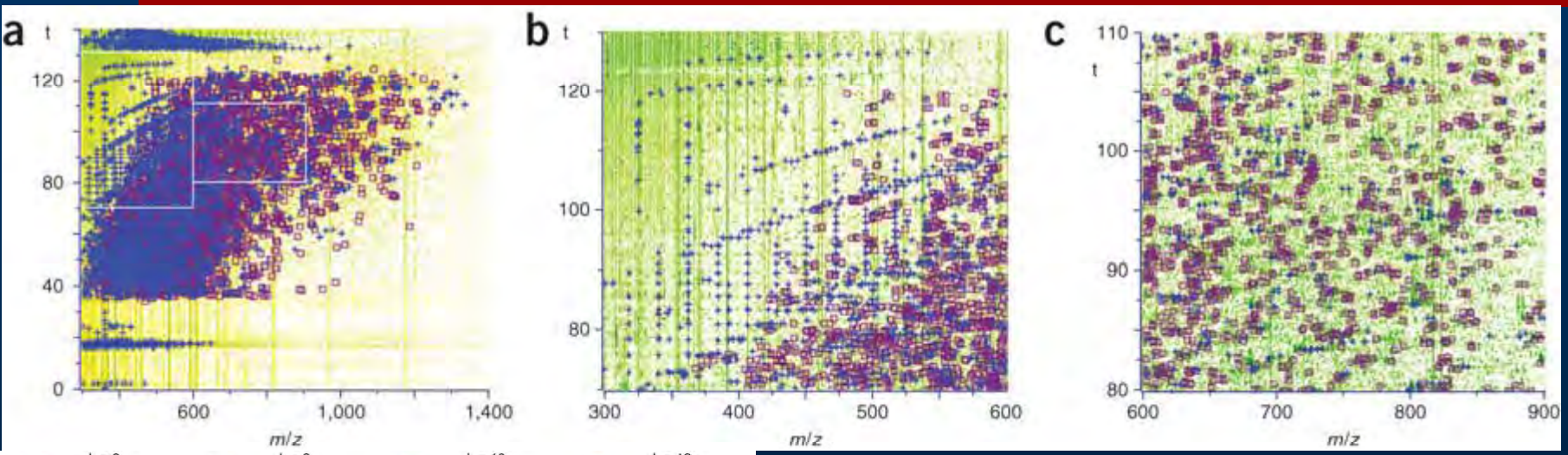


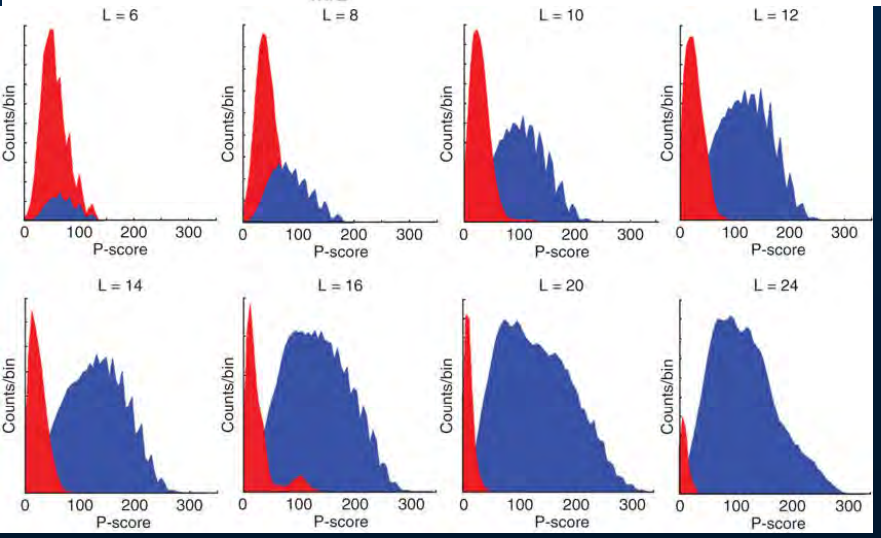
Caveat-

high mass accuracy instruments < 20ppm

Cox et. al JPR 2011

MaxQuant





The panels show the distributions of scores in the forward (blue) and reverse (red) database with peptide length (L) as the parameter. MaxQuant filters potential hits by a priori information, which moves the reverse hit distribution far to the left. These distributions are used to calculate the false-positive rate for peptide identification as a function of peptide length.

Cox and Mann Nat. Biot. 2008

Saint- significance analysis of interactome

AP/IP

Free!

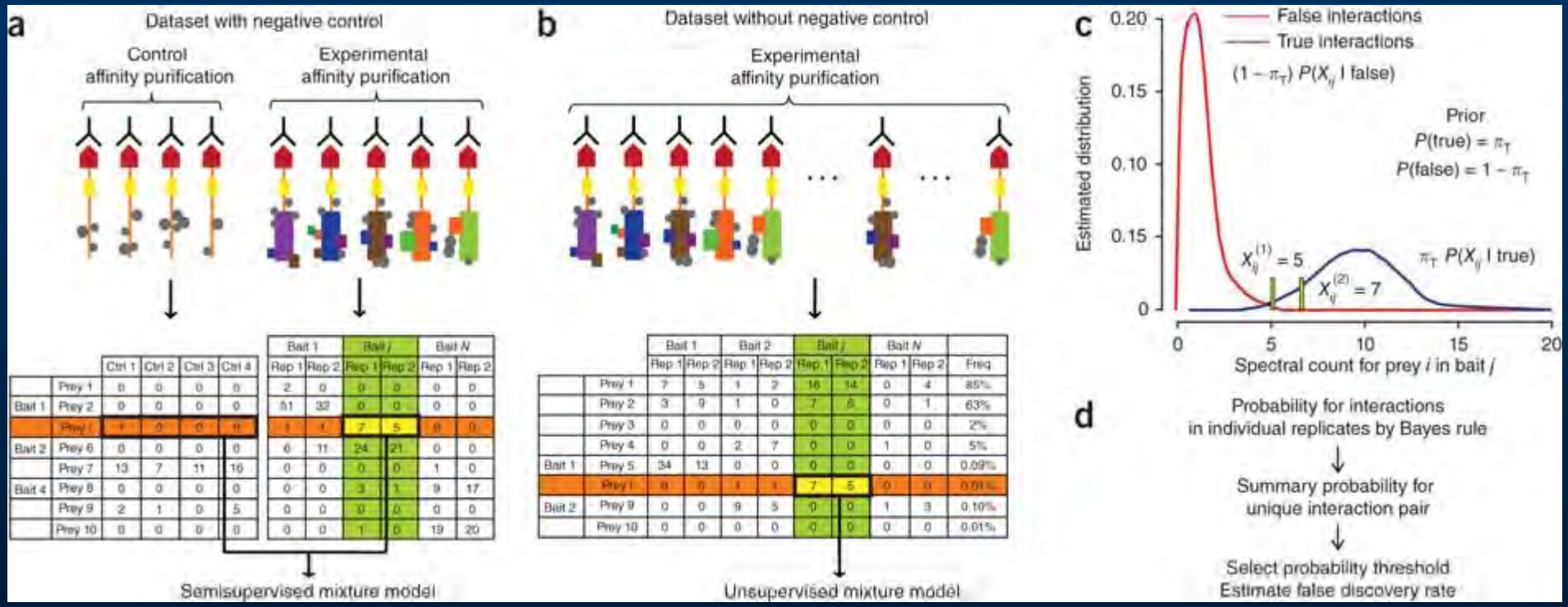
probabilistic scoring of affinity purification mass spectrometry data

True interactors vs. non-specific binders

http://saint-apms.sourceforge.net/Main.html

Choi et al. Nat. Meth. 2010

Saint relies on spectral count data



(a,b) Interaction data in the presence (a) and absence (b) of control purifications. Schematic of the experimental AP-MS procedure is shown at the top and a spectral count interaction table is illustrated at the bottom. Ctrl, control; rep, replicate; freq, frequency. (c) Modeling spectral count distributions for true and false interactions. For the interaction between prey *i* and bait *j*, SAINT uses all relevant data for the two proteins, as shown in the column of the bait (green) and the data in the row of the prey (orange) in a and b. (d) Probability is calculated for each replicate by application of Bayes rule, and a summary probability is calculated for the interaction pair (*i*,*j*).

So many tools, so little time, so much data

Ask questions-

Be proactive-

No one understands your research as well as you do

Acknowledgements

SUMS

- Anna Okumu
- Allis Chien
- Karolina Krasinska
- Pavel Aronov
- Theresa McLaughlin
- Ludmilla Alexandrova

Levitt Lab

- Nir Kalisman

Steinman Lab

- Jonathan Rothbard
- Randy Mann

- Elias Lab
- Snyder Lab
- Vincent and Stella Coates Foundation