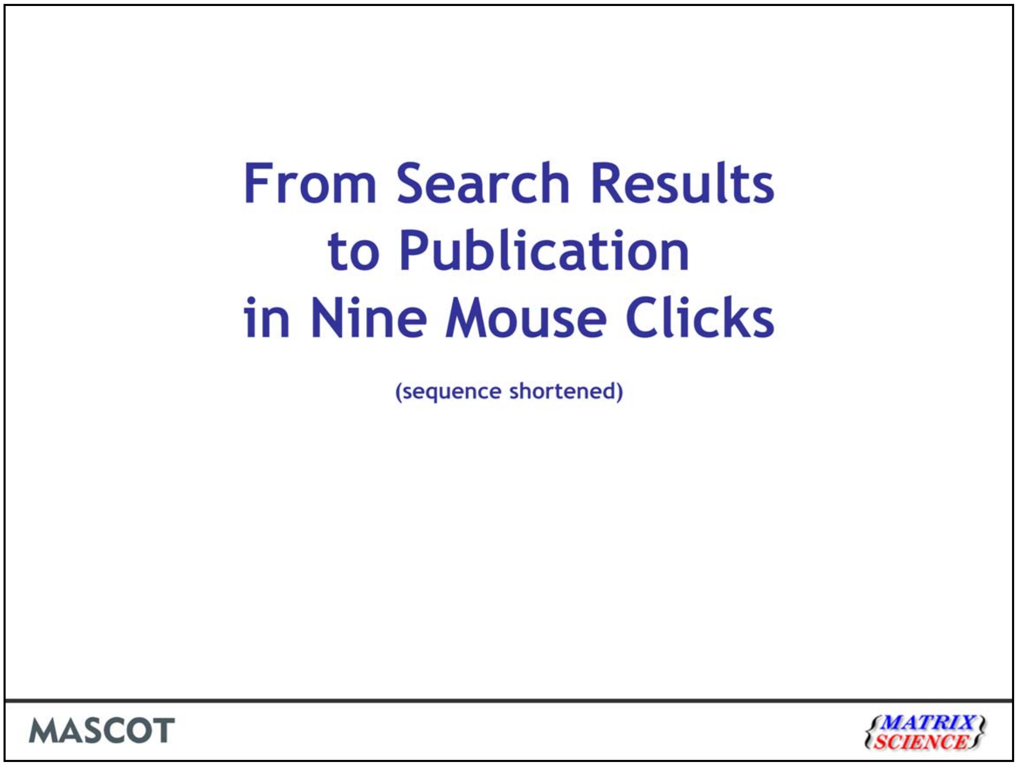


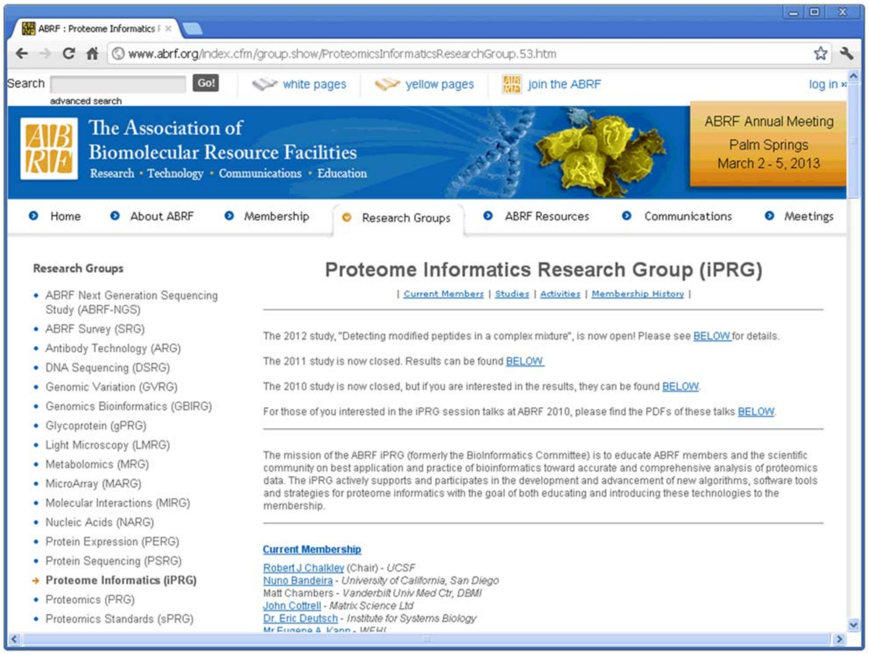
From Search Results to Publication in Nine Mouse Clicks

(sequence shortened)

MASCOT

MATRIX SCIENCE

I'd like to show you how the new features in the protein family report make it easy to generate the figures and tables needed for a publication.



ABRF : Proteome Informatics
www.abrf.org/index.cfm/group.show/ProteomicsInformaticsResearchGroup.53.htm
Search
Go!
advanced search
white pages
yellow pages
join the ABRF
log in
The Association of Biomolecular Resource Facilities
Research • Technology • Communications • Education
ABRF Annual Meeting
Palm Springs
March 2 - 5, 2013
Home
About ABRF
Membership
Research Groups
ABRF Resources
Communications
Meetings

Research Groups

- ABRF Next Generation Sequencing Study (ABRF-NGS)
- ABRF Survey (SRG)
- Antibody Technology (ARG)
- DNA Sequencing (DSRG)
- Genomic Variation (GVRG)
- Genomics Bioinformatics (GBIRG)
- Glycoprotein (gPRG)
- Light Microscopy (LMRG)
- Metabolomics (MRG)
- MicroArray (MARG)
- Molecular Interactions (MIRG)
- Nucleic Acids (NARG)
- Protein Expression (PERG)
- Protein Sequencing (PSRG)
- → **Proteome Informatics (iPRG)**
- Proteomics (PRG)
- Proteomics Standards (sPRG)

Proteome Informatics Research Group (iPRG)

| Current Members | Studies | Activities | Membership History |

The 2012 study, "Detecting modified peptides in a complex mixture", is now open! Please see BELOW for details.

The 2011 study is now closed. Results can be found BELOW

The 2010 study is now closed, but if you are interested in the results, they can be found BELOW.

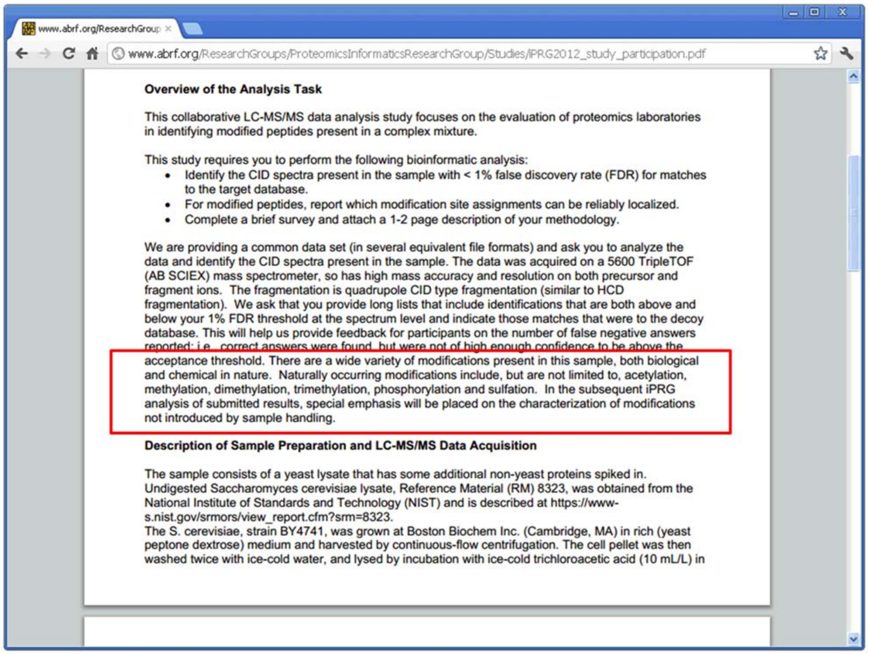
For those of you interested in the iPRG session talks at ABRF 2010, please find the PDFs of these talks BELOW.

The mission of the ABRF iPRG (formerly the BioInformatics Committee) is to educate ABRF members and the scientific community on best application and practice of bioinformatics toward accurate and comprehensive analysis of proteomics data. The iPRG actively supports and participates in the development and advancement of new algorithms, software tools and strategies for proteome informatics with the goal of both educating and introducing these technologies to the membership.

Current Membership

Robert J Chalkley (Chair) - *UCSF*
Nuno Bandeira - *University of California, San Diego*
Matt Chambers - *Vanderbilt Univ Med Ctr, DBMI*
John Cottrell - *Matrix Science Ltd*
Dr. Eric Deutsch - *Institute for Systems Biology*
Mr Eugene A. Kapp - *WEHI*

To illustrate, I'm going to use data from this year's ABRF Proteome Informatics Workgroup study - iPRG2012



www.abrf.org/ResearchGroup
www.abrf.org/ResearchGroups/ProteomicsInformaticsResearchGroup/Studies/iPRG2012_study_participation.pdf

Overview of the Analysis Task

This collaborative LC-MS/MS data analysis study focuses on the evaluation of proteomics laboratories in identifying modified peptides present in a complex mixture.

This study requires you to perform the following bioinformatic analysis:

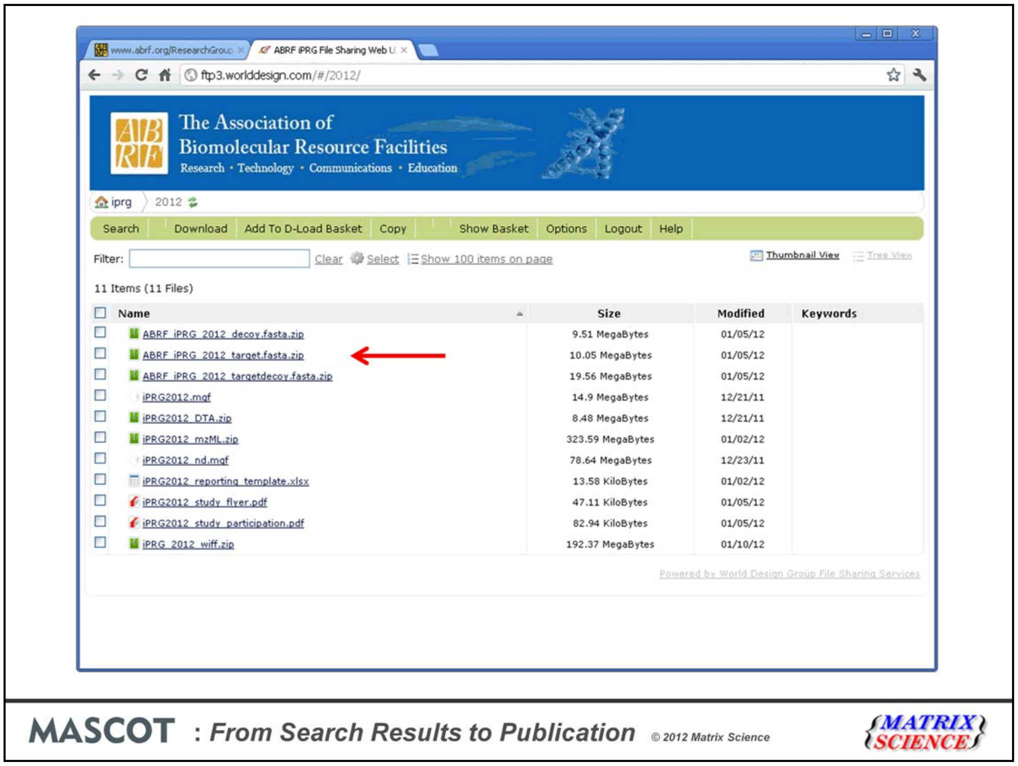
- Identify the CID spectra present in the sample with < 1% false discovery rate (FDR) for matches to the target database.
- For modified peptides, report which modification site assignments can be reliably localized.
- Complete a brief survey and attach a 1-2 page description of your methodology.

We are providing a common data set (in several equivalent file formats) and ask you to analyze the data and identify the CID spectra present in the sample. The data was acquired on a 5600 TripleTOF (AB SCIEX) mass spectrometer, so has high mass accuracy and resolution on both precursor and fragment ions. The fragmentation is quadrupole CID type fragmentation (similar to HCD fragmentation). We ask that you provide long lists that include identifications that are both above and below your 1% FDR threshold at the spectrum level and indicate those matches that were to the decoy database. This will help us provide feedback for participants on the number of false negative answers reported; i.e., correct answers were found, but were not of high enough confidence to be above the acceptance threshold. There are a wide variety of modifications present in this sample, both biological and chemical in nature. Naturally occurring modifications include, but are not limited to, acetylation, methylation, dimethylation, trimethylation, phosphorylation and sulfation. In the subsequent iPRG analysis of submitted results, special emphasis will be placed on the characterization of modifications not introduced by sample handling.

Description of Sample Preparation and LC-MS/MS Data Acquisition

The sample consists of a yeast lysate that has some additional non-yeast proteins spiked in. Undigested Saccharomyces cerevisiae lysate, Reference Material (RM) 8323, was obtained from the National Institute of Standards and Technology (NIST) and is described at https://www-s.nist.gov/srmors/view_report.cfm?srm=8323.
The S. cerevisiae, strain BY4741, was grown at Boston Biochem Inc. (Cambridge, MA) in rich (yeast peptone dextrose) medium and harvested by continuous-flow centrifugation. The cell pellet was then washed twice with ice-cold water, and lysed by incubation with ice-cold trichloroacetic acid (10 mL/L) in

The sample was a yeast lysate that had some additional non-yeast proteins spiked in. It was analysed on an AB Sciex 5600 tripleTOF and both raw data and peak lists were provided. Participants were asked to search a specified database and use target decoy to report peptide matches at 1% FDR. They were also asked to characterize modifications with special emphasis on modifications not introduced by sample handling


www.abrf.org/ResearchGroup ×
ABRF iPRG File Sharing Web U ×
ftp3.worlddesign.com/#/2012/
The Association of
Biomolecular Resource Facilities
Research • Technology • Communications • Education
iprg
2012
Search
Download
Add To D-Load Basket
Copy
Show Basket
Options
Logout
Help
Filter:
Clear
Select
Show 100 items on page
Thumbnail View
Tree View

11 Items (11 Files)

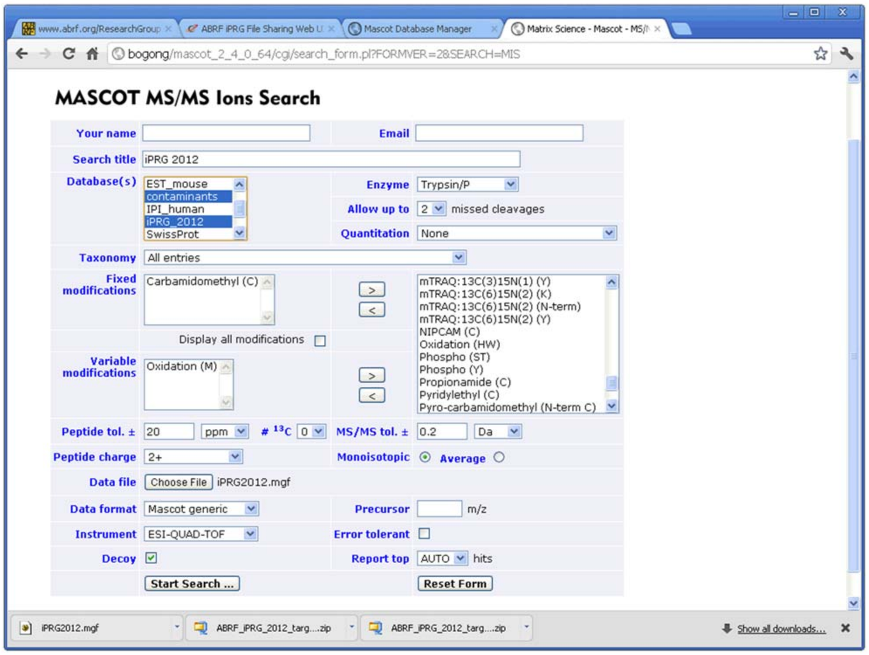
□	Name	Size	Modified	Keywords
□	ABRF_iPRG_2012_decoy.fasta.zip	9.51 MegaBytes	01/05/12	
□	ABRF_iPRG_2012_target.fasta.zip	10.05 MegaBytes	01/05/12	
□	ABRF_iPRG_2012_targetdecoy.fasta.zip	19.56 MegaBytes	01/05/12	
□	iPRG2012.mgf	14.9 MegaBytes	12/21/11	
□	iPRG2012_DTA.zip	8.48 MegaBytes	12/21/11	
□	iPRG2012_mzML.zip	323.59 MegaBytes	01/02/12	
□	iPRG2012_nd.mgf	78.64 MegaBytes	12/23/11	
□	iPRG2012_reporting_template.xlsx	13.58 KiloBytes	01/02/12	
□	iPRG2012_study_flyer.pdf	47.11 KiloBytes	01/05/12	
□	iPRG2012_study_participation.pdf	82.94 KiloBytes	01/05/12	
□	iPRG_2012_wiff.zip	192.37 MegaBytes	01/10/12	

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

I'm not going to show you all the steps required to participate in the study because it was a peptide-centric study and most of the work was in formatting the results to fit the spreadsheet template.

First of all, we need to make the Fasta database available for searching in Mascot. As you’ve seen in the earlier presentation, Database Manager makes this very easy. If you wanted to use Mascot's automatic target/decoy function, you would download the target only database, which contains SwissProt entries.



www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
Matrix Science - Mascot - MS/I
bogong/mascot_2_4_0_64/cgi/search_form.pl?FORMVER=2&SEARCH=MIS
iPRG2012.mgf
ABRF_iPRG_2012_targ....zip
ABRF_iPRG_2012_targ....zip
Show all downloads...

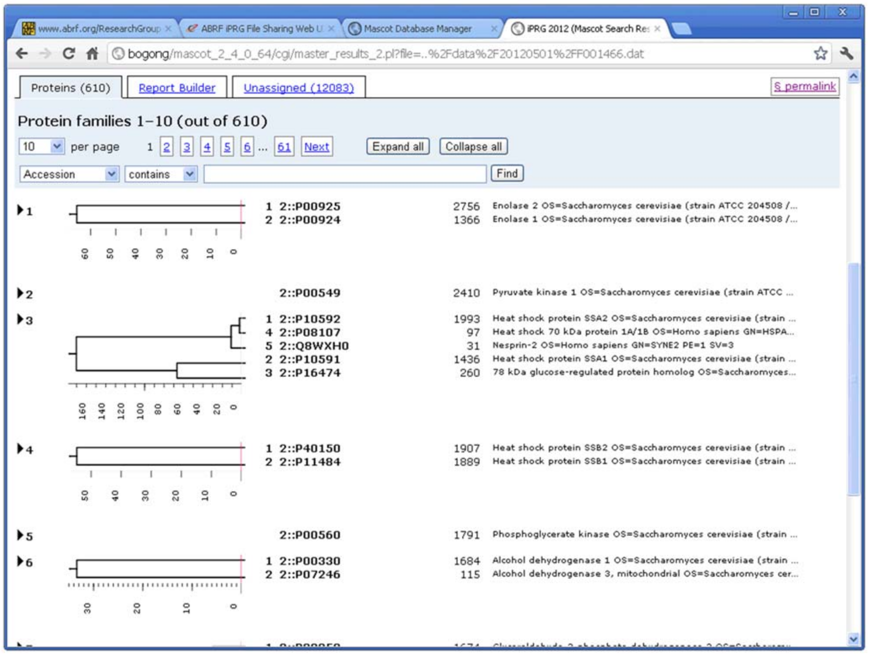
MASCOT MS/MS Ions Search

Your name		Email	
Search title	iPRG 2012		
Database(s)	EST_mouse contaminants IPI_human iPRG_2012 SwissProt	Enzyme	Trypsin/P
		Allow up to	2 missed cleavages
		Quantitation	None
Taxonomy	All entries		
Fixed modifications	Carbamidomethyl (C)	> <	mTRAQ:13C(3)15N(1) (Y) mTRAQ:13C(6)15N(2) (K) mTRAQ:13C(6)15N(2) (N-term) mTRAQ:13C(6)15N(2) (Y) NIPCAM (C) Oxidation (HW) Phospho (ST) Phospho (Y) Propionamide (C) Pyridylethyl (C) Pyro-carbamidomethyl (N-term C)
	Display all modifications □		
Variable modifications	Oxidation (M)	> <	
Peptide tol. ±	20 ppm # ^{13}C 0	MS/MS tol. ±	0.2 Da
Peptide charge	2+	Monoisotopic	● Average ○
Data file	Choose File iPRG2012.mgf		
Data format	Mascot generic	Precursor	m/z
Instrument	ESI-QUAD-TOF	Error tolerant	□
Decoy	☑	Report top	AUTO hits
	Start Search ...		Reset Form

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

We know from the sample description that the cysteine alkylation is carbamidomethyl. Usually, the only other modification I would select for a first, trial search is Met-Ox. The other settings are guesses which we will refine by looking at the results.

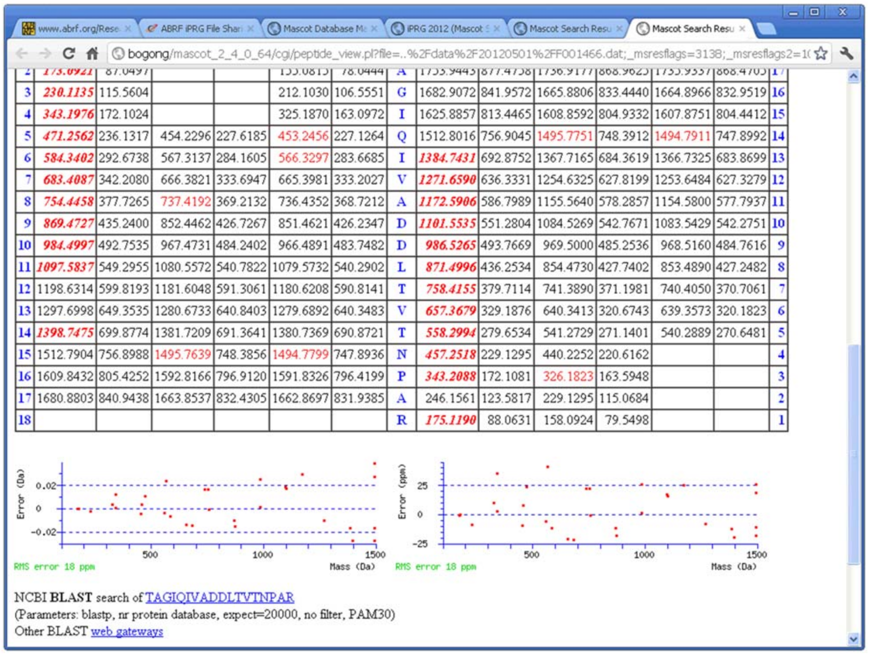

www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Res
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001466.dat
Proteins (610)
Report Builder
Unassigned (12083)
§ permalink

Protein families 1–10 (out of 610)

10
per page
1 2 3 4 5 6 ... 61 Next
Expand all
Collapse all
Accession
contains
Find

▸1	1 2::P00925	2756	Enolase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...
	2 2::P00924	1366	Enolase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...
▸2	2::P00549	2410	Pyruvate kinase 1 OS=Saccharomyces cerevisiae (strain ATCC ...
▸3	1 2::P10592	1993	Heat shock protein SSA2 OS=Saccharomyces cerevisiae (strain ...
	4 2::P08107	97	Heat shock 70 kDa protein 1A/1B OS=Homo sapiens GN=HSPA...
	5 2::Q8WXH0	31	Nesprin-2 OS=Homo sapiens GN=SYNE2 PE=1 SV=3
	2 2::P10591	1436	Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strain ...
	3 2::P16474	260	78 kDa glucose-regulated protein homolog OS=Saccharomyces...
▸4	1 2::P40150	1907	Heat shock protein SSB2 OS=Saccharomyces cerevisiae (strain ...
	2 2::P11484	1889	Heat shock protein SSB1 OS=Saccharomyces cerevisiae (strain ...
▸5	2::P00560	1791	Phosphoglycerate kinase OS=Saccharomyces cerevisiae (strain ...
▸6	1 2::P00330	1684	Alcohol dehydrogenase 1 OS=Saccharomyces cerevisiae (strain ...
	2 2::P07246	115	Alcohol dehydrogenase 3, mitochondrial OS=Saccharomyces cer...

In particular, we can use Peptide View and Protein View to estimate mass accuracy.



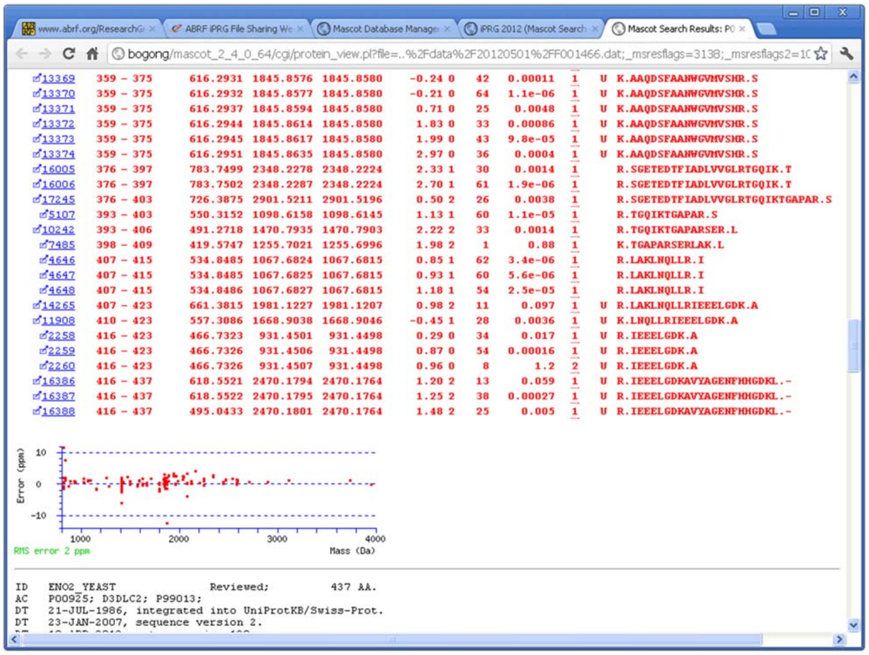
www.abrf.org/Rese
ABRF iPRG File Shar
Mascot Database Ma
iPRG 2012 (Mascot S
Mascot Search Resu
Mascot Search Resu
bogong/mascot_2_4_0_64/cgi/peptide_view.pl?file=..%2Fdata%2F20120501%2FF001466.dat;_msresflags=3138;_msresflags2=1(

2	***173.0921***	87.0497			155.0815	78.0444	A	1753.9443	877.4758	1736.9177	868.9625	1735.9337	868.4705	17
3	***230.1135***	115.5604			212.1030	106.5551	G	1682.9072	841.9572	1665.8806	833.4440	1664.8966	832.9519	16
4	***343.1976***	172.1024			325.1870	163.0972	I	1625.8857	813.4465	1608.8592	804.9332	1607.8751	804.4412	15
5	***471.2562***	236.1317	454.2296	227.6185	453.2456	227.1264	Q	1512.8016	756.9045	1495.7751	748.3912	1494.7911	747.8992	14
6	***584.3402***	292.6738	567.3137	284.1605	566.3297	283.6685	I	***1384.7431***	692.8752	1367.7165	684.3619	1366.7325	683.8699	13
7	***683.4087***	342.2080	666.3821	333.6947	665.3981	333.2027	V	***1271.6590***	636.3331	1254.6325	627.8199	1253.6484	627.3279	12
8	***754.4458***	377.7265	737.4192	369.2132	736.4352	368.7212	A	***1172.5906***	586.7989	1155.5640	578.2857	1154.5800	577.7937	11
9	***869.4727***	435.2400	852.4462	426.7267	851.4621	426.2347	D	***1101.5535***	551.2804	1084.5269	542.7671	1083.5429	542.2751	10
10	***984.4997***	492.7535	967.4731	484.2402	966.4891	483.7482	D	***986.5265***	493.7669	969.5000	485.2536	968.5160	484.7616	9
11	***1097.5837***	549.2955	1080.5572	540.7822	1079.5732	540.2902	L	***871.4996***	436.2534	854.4730	427.7402	853.4890	427.2482	8
12	1198.6314	599.8193	1181.6048	591.3061	1180.6208	590.8141	T	***758.4155***	379.7114	741.3890	371.1981	740.4050	370.7061	7
13	1297.6998	649.3535	1280.6733	640.8403	1279.6892	640.3483	V	***657.3679***	329.1876	640.3413	320.6743	639.3573	320.1823	6
14	***1398.7475***	699.8774	1381.7209	691.3641	1380.7369	690.8721	T	***558.2994***	279.6534	541.2729	271.1401	540.2889	270.6481	5
15	1512.7904	756.8988	1495.7639	748.3856	1494.7799	747.8936	N	***457.2518***	229.1295	440.2252	220.6162			4
16	1609.8432	805.4252	1592.8166	796.9120	1591.8326	796.4199	P	***343.2088***	172.1081	326.1823	163.5948			3
17	1680.8803	840.9438	1663.8537	832.4305	1662.8697	831.9385	A	246.1561	123.5817	229.1295	115.0684			2
18							R	***175.1190***	88.0631	158.0924	79.5498			1

NCBI **BLAST** search of TAGIQIVADDLTVTNPAR
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

MASCOT : *From Search Results to Publication* © 2012 Matrix Science MATRIX SCIENCE

Looks like 0.05 Da is safe for MS/MS.



www.abrf.org/ResearchG
ABRF iPRG File Sharing We
Mascot Database Manage
iPRG 2012 (Mascot Search
Mascot Search Results: P0
bogong/mascot_2_4_0_64/cgi/protein_view.pl?file=..%2Fdata%2F20120501%2FF001466.dat;_msresflags=3138;_msresflags2=10

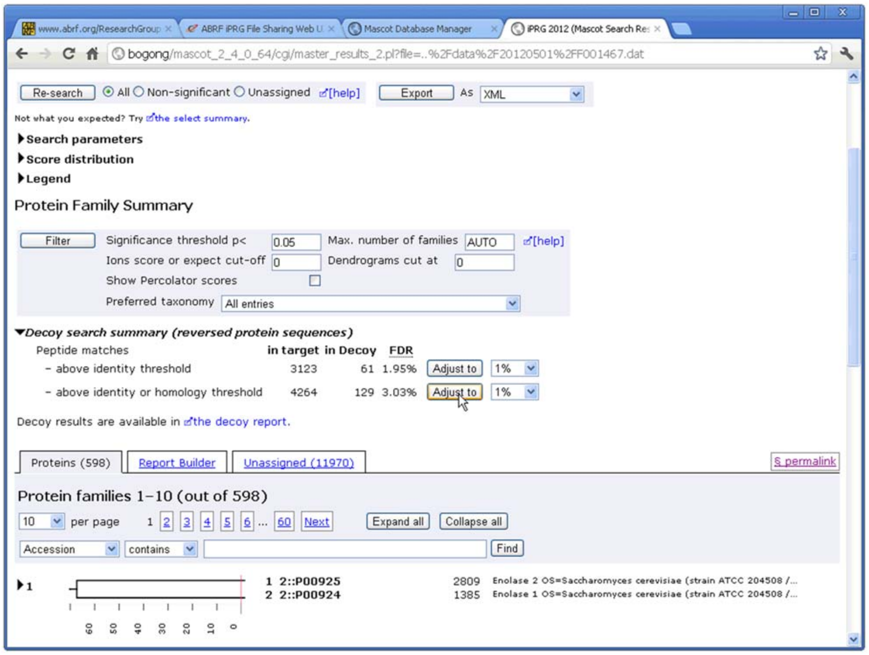
```
13369   359 - 375   616.2931 1845.8576 1845.8580  -0.24 0  42  0.00011  1  U K.AAQDSFAANWGVMVSHR.S
13370   359 - 375   616.2932 1845.8577 1845.8580  -0.21 0  64  1.1e-06  1  U K.AAQDSFAANWGVMVSHR.S
13371   359 - 375   616.2937 1845.8594 1845.8580   0.71 0  25   0.0048  1  U K.AAQDSFAANWGVMVSHR.S
13372   359 - 375   616.2944 1845.8614 1845.8580   1.83 0  33  0.00086  1  U K.AAQDSFAANWGVMVSHR.S
13373   359 - 375   616.2945 1845.8617 1845.8580   1.99 0  43  9.8e-05  1  U K.AAQDSFAANWGVMVSHR.S
13374   359 - 375   616.2951 1845.8635 1845.8580   2.97 0  36   0.0004  1  U K.AAQDSFAANWGVMVSHR.S
16005   376 - 397   783.7499 2348.2278 2348.2224   2.33 1  30   0.0014  1    R.SGETEDTFIADLVVGLRTGQIK.T
16006   376 - 397   783.7502 2348.2287 2348.2224   2.70 1  61  1.9e-06  1    R.SGETEDTFIADLVVGLRTGQIK.T
17245   376 - 403   726.3875 2901.5211 2901.5196   0.50 2  26   0.0038  1    R.SGETEDTFIADLVVGLRTGQIKTGAPAR.S
 5107   393 - 403   550.3152 1098.6158 1098.6145   1.13 1  60  1.1e-05  1    R.TGQIKTGAPAR.S
10242   393 - 406   491.2718 1470.7935 1470.7903   2.22 2  33   0.0014  1    R.TGQIKTGAPARSER.L
 7485   398 - 409   419.5747 1255.7021 1255.6996   1.98 2   1     0.88  1    K.TGAPARSERLAK.L
 4646   407 - 415   534.8485 1067.6824 1067.6815   0.85 1  62  3.4e-06  1    R.LAKLNQLLR.I
 4647   407 - 415   534.8485 1067.6825 1067.6815   0.93 1  60  5.6e-06  1    R.LAKLNQLLR.I
 4648   407 - 415   534.8486 1067.6827 1067.6815   1.18 1  54  2.5e-05  1    R.LAKLNQLLR.I
14265   407 - 423   661.3815 1981.1227 1981.1207   0.98 2  11    0.097  1  U R.LAKLNQLLRIEEELGDK.A
11908   410 - 423   557.3086 1668.9038 1668.9046  -0.45 1  28   0.0036  1  U K.LNQLLRIEEELGDK.A
 2258   416 - 423   466.7323  931.4501  931.4498   0.29 0  34    0.017  1  U R.IEEELGDK.A
 2259   416 - 423   466.7326  931.4506  931.4498   0.87 0  54  0.00016  1  U R.IEEELGDK.A
 2260   416 - 423   466.7326  931.4507  931.4498   0.96 0   8      1.2  2  U R.IEEELGDK.A
16386   416 - 437   618.5521 2470.1794 2470.1764   1.20 2  13    0.059  1  U R.IEEELGDKAVYAGENFHHGDKL.-
16387   416 - 437   618.5522 2470.1795 2470.1764   1.25 2  38  0.00027  1  U R.IEEELGDKAVYAGENFHHGDKL.-
16388   416 - 437   495.0433 2470.1801 2470.1764   1.48 2  25    0.005  1  U R.IEEELGDKAVYAGENFHHGDKL.-
```

RMS error 2 ppm

```
ID   ENO2_YEAST              Reviewed;         437 AA.
AC   P00925; D3DLC2; P99013;
DT   21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT   23-JAN-2007, sequence version 2.
```

For the MS errors, looks like 10 ppm is safe. We might get away with 5 ppm but, with a small database, this is going to limit the number of candidate sequences available for matching to each spectrum, so 10 ppm is a better choice.

We also see quite a few high scoring matches with 2 missed cleavages, so maybe push this up to 3.


www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Res
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat
Re-search
Export
[help]

● All ○ Non-significant ○ Unassigned [help]

Export As XML

Not what you expected? Try the select summary.

▸ Search parameters

▸ Score distribution

▸ Legend

Protein Family Summary

Filter

Significance threshold p< 0.05

Max. number of families AUTO [help]

Ions score or expect cut-off 0

Dendrograms cut at 0

Show Percolator scores □

Preferred taxonomy All entries

▾ *Decoy search summary (reversed protein sequences)*

Peptide matches	in target	in Decoy	FDR		
– above identity threshold	3123	61	1.95%	Adjust to	1%
– above identity or homology threshold	4264	129	3.03%	Adjust to	1%

Decoy results are available in the decoy report.

Proteins (598) | Report Builder | Unassigned (11970) | § permalink

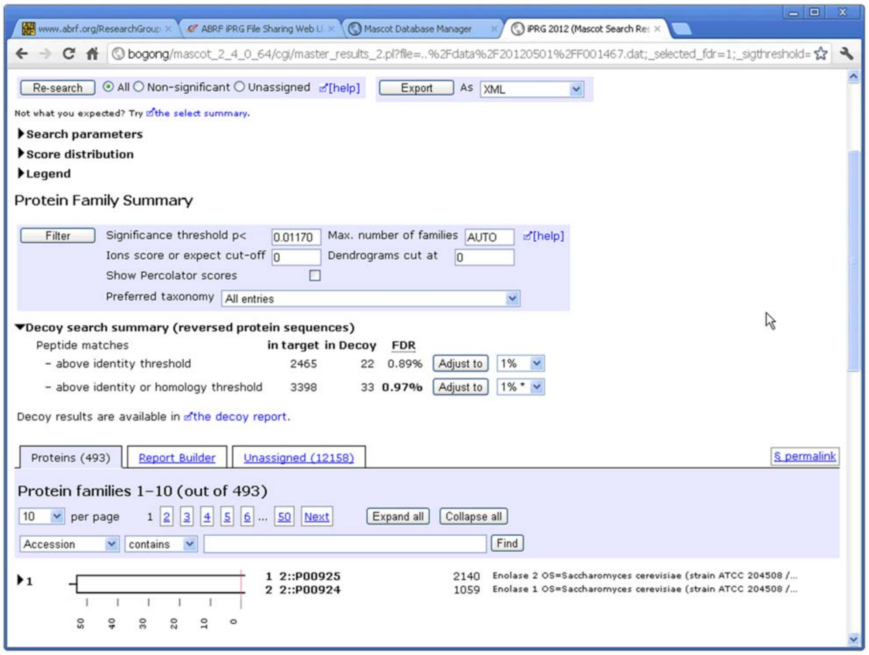
Protein families 1–10 (out of 598)

10 per page 1 2 3 4 5 6 ... 60 Next Expand all Collapse all

Accession contains Find

▸1	1 2::P00925	2809	Enolase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...
	2 2::P00924	1385	Enolase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...

Repeating with the new settings, we can see that the FDR for the default setting of 5% significance threshold is approximately 3%. The iPRG study requested matches to be reported with an FDR of 1%. This is where another of the new features in Mascot 2.4 comes in useful. The 'adjust to FDR' button. Getting back to the title of the talk, let's use the first of our nine mouse clicks to obtain the required 1% FDR


www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Re:
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=

Re-search ● All ○ Non-significant ○ Unassigned [help] Export As XML

Not what you expected? Try the select summary.

- Search parameters
- Score distribution
- Legend

Protein Family Summary

Filter
Significance threshold p< 0.01170 Max. number of families AUTO [help]
Ions score or expect cut-off 0 Dendrograms cut at 0
Show Percolator scores □
Preferred taxonomy All entries

Decoy search summary (reversed protein sequences)

Peptide matches	in target	in Decoy	FDR		
– above identity threshold	2465	22	0.89%	Adjust to	1%
– above identity or homology threshold	3398	33	**0.97%**	Adjust to	1% *

Decoy results are available in the decoy report.

Protein families 1–10 (out of 493)

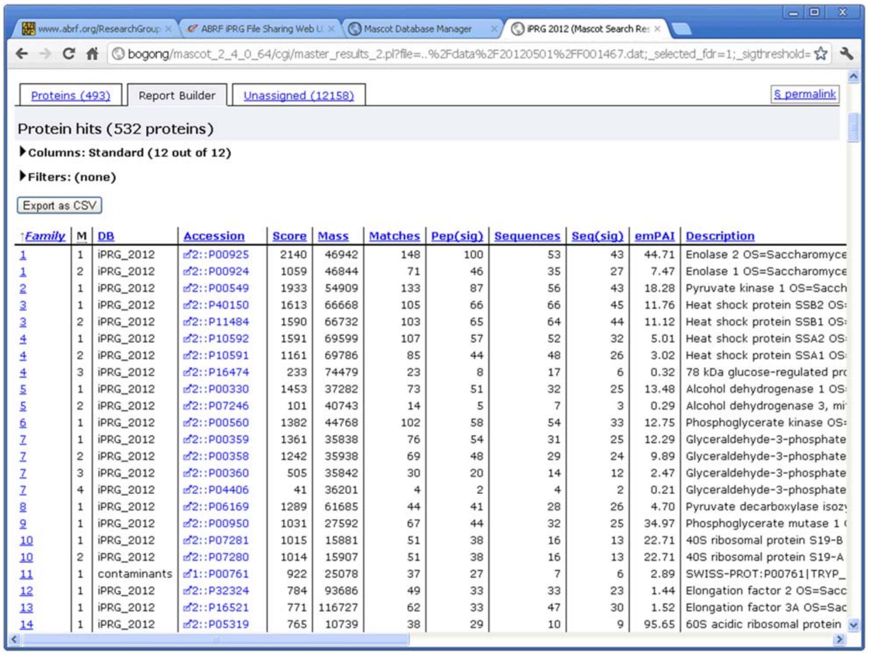
10 per page 1 2 3 4 5 6 ... 50 Next Expand all Collapse all

Accession contains Find

1	1 2::P00925	2140	Enolase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...
	2 2::P00924	1059	Enolase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 /...

In Mascot 2.3 and earlier, you had to use trial and error to adjust the FDR to a specific value, so this button is a time saver. You may also notice that the decoy sequences are reversed and not randomised. This is another new feature in Mascot 2.4. The default is reversed for MS/MS searches with enzyme specificity and randomised for no enzyme searches, but you can change these defaults if you wish.

To get a table of proteins suitable for publication, we use a second mouse click to switch to the Report Builder tab.


www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Re:
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=
Proteins (493)
Report Builder
Unassigned (12158)
§ permalink
Export as CSV

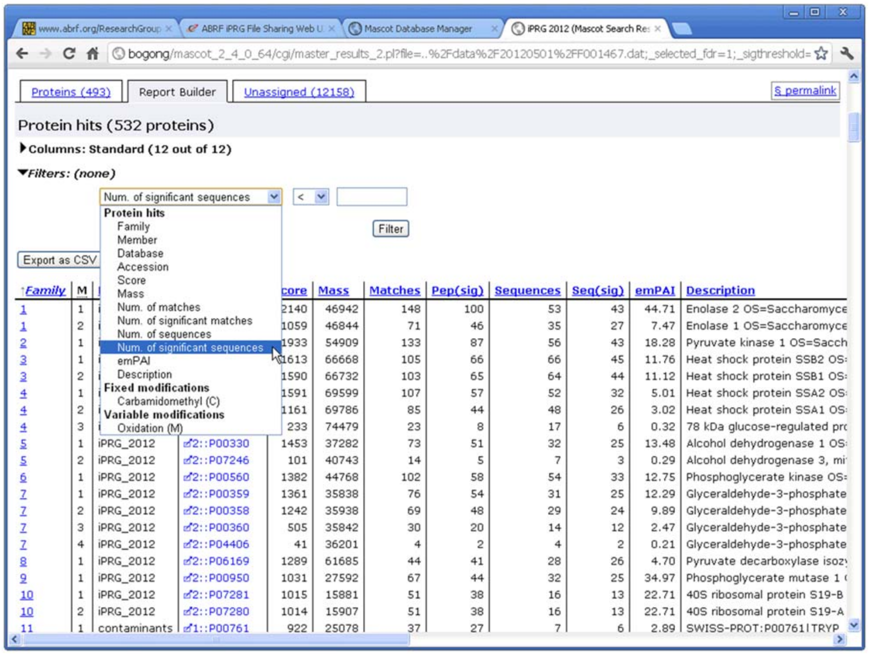
Protein hits (532 proteins)

▶Columns: Standard (12 out of 12)

▶Filters: (none)

↑Family	M	DB	Accession	Score	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	iPRG_2012	2::P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyce
1	2	iPRG_2012	2::P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyce
2	1	iPRG_2012	2::P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Sacch
3	1	iPRG_2012	2::P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS
3	2	iPRG_2012	2::P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS
4	1	iPRG_2012	2::P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS
4	2	iPRG_2012	2::P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS
4	3	iPRG_2012	2::P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated pr
5	1	iPRG_2012	2::P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS
5	2	iPRG_2012	2::P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mi
6	1	iPRG_2012	2::P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS
7	1	iPRG_2012	2::P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate
7	2	iPRG_2012	2::P00358	1242	35938	69	48	29	24	9.89	Glyceraldehyde-3-phosphate
7	3	iPRG_2012	2::P00360	505	35842	30	20	14	12	2.47	Glyceraldehyde-3-phosphate
7	4	iPRG_2012	2::P04406	41	36201	4	2	4	2	0.21	Glyceraldehyde-3-phosphate
8	1	iPRG_2012	2::P06169	1289	61685	44	41	28	26	4.70	Pyruvate decarboxylase isoz
9	1	iPRG_2012	2::P00950	1031	27592	67	44	32	25	34.97	Phosphoglycerate mutase 1
10	1	iPRG_2012	2::P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B
10	2	iPRG_2012	2::P07280	1014	15907	51	38	16	13	22.71	40S ribosomal protein S19-A
11	1	contaminants	1::P00761	922	25078	37	27	7	6	2.89	SWISS-PROT:P00761|TRYP_
12	1	iPRG_2012	2::P32324	784	93686	49	33	33	23	1.44	Elongation factor 2 OS=Sacc
13	1	iPRG_2012	2::P16521	771	116727	62	33	47	30	1.52	Elongation factor 3A OS=Sac
14	1	iPRG_2012	2::P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein

Lets assume we want to drop the ‘one hit wonders’ and only report proteins that have significant matches to at least 2 different peptide sequences



www.abrf.org/ResearchGroup ×
ABRF iPRG File Sharing Web U ×
Mascot Database Manager ×
iPRG 2012 (Mascot Search Re ×

bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=

Protein hits (532 proteins)

▶Columns: Standard (12 out of 12)

▼Filters: (none)

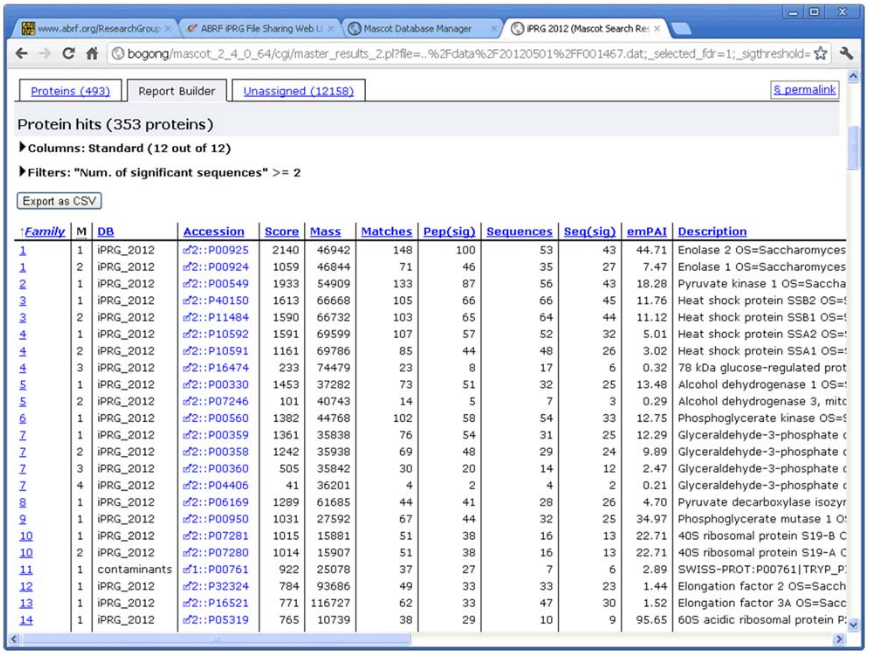
Num. of significant sequences ▾ < ▾

↑Family	M	I		core	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	i		2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyce
1	2	i		1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyce
2	1			1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Sacch
3	1	i		1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS:
3	2	i		1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS:
4	1	i		1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS:
4	2			1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS:
4	3	i		233	74479	23	8	17	6	0.32	78 kDa glucose-regulated prc
5	1	iPRG_2012	2::P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS:
5	2	iPRG_2012	2::P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mi
6	1	iPRG_2012	2::P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=
7	1	iPRG_2012	2::P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate
7	2	iPRG_2012	2::P00358	1242	35938	69	48	29	24	9.89	Glyceraldehyde-3-phosphate
7	3	iPRG_2012	2::P00360	505	35842	30	20	14	12	2.47	Glyceraldehyde-3-phosphate
7	4	iPRG_2012	2::P04406	41	36201	4	2	4	2	0.21	Glyceraldehyde-3-phosphate
8	1	iPRG_2012	2::P06169	1289	61685	44	41	28	26	4.70	Pyruvate decarboxylase isoz\
9	1	iPRG_2012	2::P00950	1031	27592	67	44	32	25	34.97	Phosphoglycerate mutase 1 (
10	1	iPRG_2012	2::P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B
10	2	iPRG_2012	2::P07280	1014	15907	51	38	16	13	22.71	40S ribosomal protein S19-A
11	1	contaminants	1::P00761	922	25078	37	27	7	6	2.89	SWISS-PROT:P00761|TRYP

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

We open up the filters section and add a suitable filter. This uses another 5 mouse clicks, so 7 in total



www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Re
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=
Proteins (493)
Report Builder
Unassigned (12158)
§ permalink

Protein hits (353 proteins)

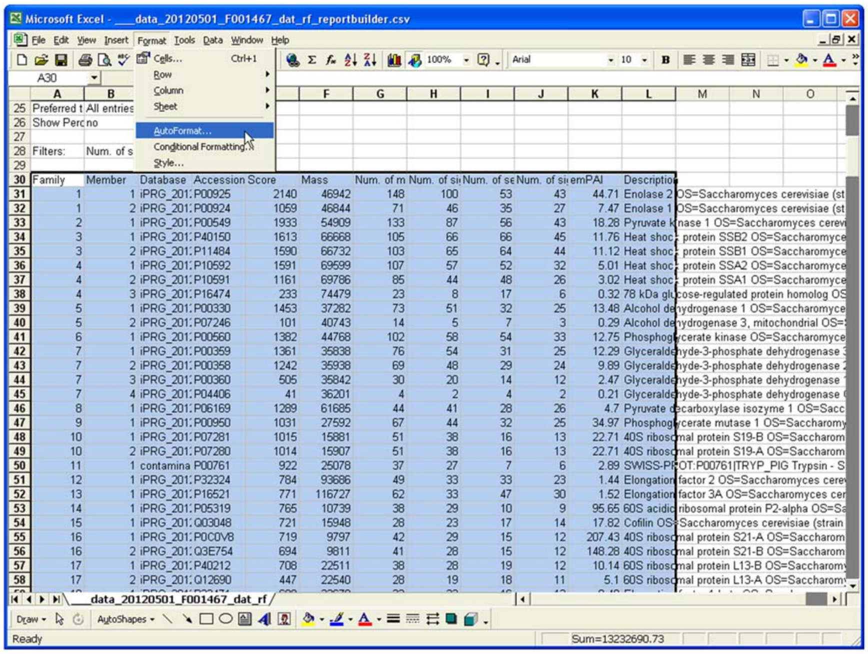
▸Columns: Standard (12 out of 12)

▸Filters: "Num. of significant sequences" >= 2

Export as CSV

↑Family	M	DB	Accession	Score	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	iPRG_2012	2::P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyces
1	2	iPRG_2012	2::P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyces
2	1	iPRG_2012	2::P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Saccha
3	1	iPRG_2012	2::P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS=
3	2	iPRG_2012	2::P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS=
4	1	iPRG_2012	2::P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS=
4	2	iPRG_2012	2::P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS=
4	3	iPRG_2012	2::P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated prot
5	1	iPRG_2012	2::P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS=
5	2	iPRG_2012	2::P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mitc
6	1	iPRG_2012	2::P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=
7	1	iPRG_2012	2::P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate
7	2	iPRG_2012	2::P00358	1242	35938	69	48	29	24	9.89	Glyceraldehyde-3-phosphate
7	3	iPRG_2012	2::P00360	505	35842	30	20	14	12	2.47	Glyceraldehyde-3-phosphate
7	4	iPRG_2012	2::P04406	41	36201	4	2	4	2	0.21	Glyceraldehyde-3-phosphate
8	1	iPRG_2012	2::P06169	1289	61685	44	41	28	26	4.70	Pyruvate decarboxylase isozyr
9	1	iPRG_2012	2::P00950	1031	27592	67	44	32	25	34.97	Phosphoglycerate mutase 1 O
10	1	iPRG_2012	2::P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B C
10	2	iPRG_2012	2::P07280	1014	15907	51	38	16	13	22.71	40S ribosomal protein S19-A C
11	1	contaminants	1::P00761	922	25078	37	27	7	6	2.89	SWISS-PROT:P00761|TRYP_P
12	1	iPRG_2012	2::P32324	784	93686	49	33	33	23	1.44	Elongation factor 2 OS=Sacch
13	1	iPRG_2012	2::P16521	771	116727	62	33	47	30	1.52	Elongation factor 3A OS=Sacc
14	1	iPRG_2012	2::P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein P

Click number 8 is to export as CSV and click number 9 (actually double click) is to open the CSV in excel



Microsoft Excel - ___data_20120501_F001467_dat_rf_reportbuilder.csv
File Edit View Insert Format Tools Data Window Help
Cells... Ctrl+1
Row
Column
Sheet
AutoFormat...
Conditional Formatting...
Style...
100%
Arial
10
A30
Draw
AutoShapes
___data_20120501_F001467_dat_rf
Ready

	A	B	C	D	E	F	G	H	I	J	K	L
25	Preferred t	All entries										
26	Show Perc	no										
27												
28	Filters:	Num. of s										
29												
30	Family	Member	Database	Accession	Score	Mass	Num. of m	Num. of si	Num. of se	Num. of si	emPAI	Description
31	1	1	iPRG_201	P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyces cerevisiae (st
32	1	2	iPRG_201	P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyces cerevisiae (st
33	2	1	iPRG_201	P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Saccharomyces cerevi
34	3	1	iPRG_201	P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS=Saccharomyce
35	3	2	iPRG_201	P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS=Saccharomyce
36	4	1	iPRG_201	P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS=Saccharomyce
37	4	2	iPRG_201	P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS=Saccharomyce
38	4	3	iPRG_201	P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated protein homolog OS
39	5	1	iPRG_201	P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS=Saccharomyce
40	5	2	iPRG_201	P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mitochondrial OS=
41	6	1	iPRG_201	P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=Saccharomyce
42	7	1	iPRG_201	P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate dehydrogenase 3
43	7	2	iPRG_201	P00358	1242	35938	69	48	29	24	9.89	Glyceraldehyde-3-phosphate dehydrogenase 2
44	7	3	iPRG_201	P00360	505	35842	30	20	14	12	2.47	Glyceraldehyde-3-phosphate dehydrogenase 1
45	7	4	iPRG_201	P04406	41	36201	4	2	4	2	0.21	Glyceraldehyde-3-phosphate dehydrogenase
46	8	1	iPRG_201	P06169	1289	61685	44	41	28	26	4.7	Pyruvate decarboxylase isozyme 1 OS=Sacc
47	9	1	iPRG_201	P00950	1031	27592	67	44	32	25	34.97	Phosphoglycerate mutase 1 OS=Saccharomy
48	10	1	iPRG_201	P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B OS=Saccharom
49	10	2	iPRG_201	P07280	1014	15907	51	38	16	13	22.71	40S ribosomal protein S19-A OS=Saccharom
50	11	1	contamina	P00761	922	25078	37	27	7	6	2.89	SWISS-PROT:P00761|TRYP_PIG Trypsin - S
51	12	1	iPRG_201	P32324	784	93686	49	33	33	23	1.44	Elongation factor 2 OS=Saccharomyces cerev
52	13	1	iPRG_201	P16521	771	116727	62	33	47	30	1.52	Elongation factor 3A OS=Saccharomyces cer
53	14	1	iPRG_201	P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein P2-alpha OS=Sa
54	15	1	iPRG_201	Q03048	721	15948	28	23	17	14	17.82	Cofilin OS=Saccharomyces cerevisiae (strain
55	16	1	iPRG_201	P0C0V8	719	9797	42	29	15	12	207.43	40S ribosomal protein S21-A OS=Saccharom
56	16	2	iPRG_201	Q3E754	694	9811	41	28	15	12	148.28	40S ribosomal protein S21-B OS=Saccharom
57	17	1	iPRG_201	P40212	708	22511	38	28	19	12	10.14	60S ribosomal protein L13-B OS=Saccharomy
58	17	2	iPRG_201	Q12690	447	22540	28	19	18	11	5.1	60S ribosomal protein L13-A OS=Saccharomy

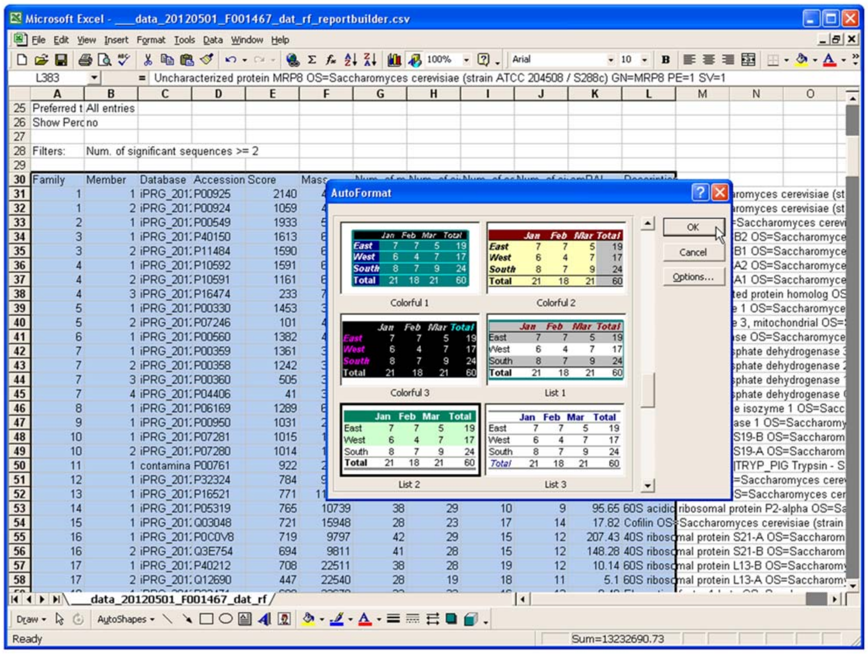
Sum=13232690.73

Now, I’m going to cheat a bit, and ignore all the keystrokes we need to use in Excel to add some formatting to the table.

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

This is where the last bit of the title comes in. You may have noticed the weasel words ‘sequence shortened’ in technology ads. Particularly for a certain cellphone

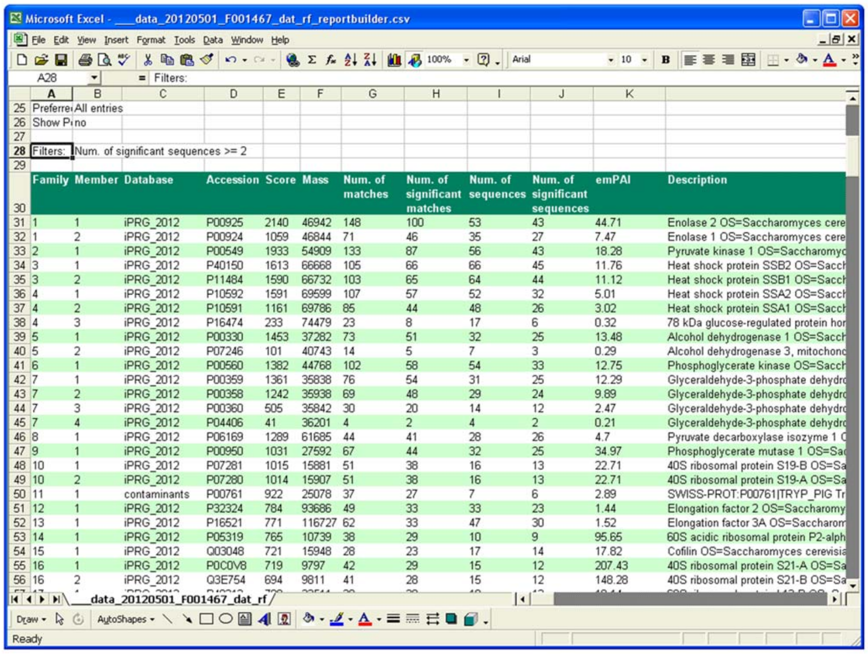

Microsoft Excel - ____data_20120501_F001467_dat_rf_reportbuilder.csv
File Edit View Insert Format Tools Data Window Help
100%
Arial
10
AutoFormat
Colorful 1
Colorful 2
Colorful 3
List 1
List 2
List 3
OK
Cancel
Options...
____data_20120501_F001467_dat_rf
Draw
AutoShapes
Ready

L383 = Uncharacterized protein MRP8 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=MRP8 PE=1 SV=1

	A	B	C	D	E	F	G	H	I	J	K	L	M
25	Preferred t	All entries											
26	Show Perc	no											
27													
28	Filters:	Num. of significant sequences >= 2											
29													
30	Family	Member	Database	Accession	Score	Mass	Num. of m	Num. of si	Num. of se	Num. of si	emPAI	Descriptio	
31	1	1	iPRG_201:	P00925	2140	4							aromyces cerevisiae (st
32	1	2	iPRG_201:	P00924	1059	4							aromyces cerevisiae (st
33	2	1	iPRG_201:	P00549	1933	5							=Saccharomyces cerevi
34	3	1	iPRG_201:	P40150	1613	6							B2 OS=Saccharomyce
35	3	2	iPRG_201:	P11484	1590	6							B1 OS=Saccharomyce
36	4	1	iPRG_201:	P10592	1591	6							A2 OS=Saccharomyce
37	4	2	iPRG_201:	P10591	1161	6							A1 OS=Saccharomyce
38	4	3	iPRG_201:	P16474	233	7							ted protein homolog OS
39	5	1	iPRG_201:	P00330	1453	3							e 1 OS=Saccharomyce
40	5	2	iPRG_201:	P07246	101	4							e 3, mitochondrial OS=
41	6	1	iPRG_201:	P00560	1382	4							se OS=Saccharomyce
42	7	1	iPRG_201:	P00359	1361	3							sphate dehydrogenase 3
43	7	2	iPRG_201:	P00358	1242	3							sphate dehydrogenase 2
44	7	3	iPRG_201:	P00360	505	3							sphate dehydrogenase 1
45	7	4	iPRG_201:	P04406	41	3							sphate dehydrogenase C
46	8	1	iPRG_201:	P06169	1289	6							e isozyme 1 OS=Sacc
47	9	1	iPRG_201:	P00950	1031	2							ase 1 OS=Saccharomy
48	10	1	iPRG_201:	P07281	1015	1							S19-B OS=Saccharom
49	10	2	iPRG_201:	P07280	1014	1							S19-A OS=Saccharom
50	11	1	contamina	P00761	922	2							|TRYP_PIG Trypsin - S
51	12	1	iPRG_201:	P32324	784	9							=Saccharomyces cerev
52	13	1	iPRG_201:	P16521	771	11							S=Saccharomyces cer
53	14	1	iPRG_201:	P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein P2-alpha OS=Sa	
54	15	1	iPRG_201:	Q03048	721	15948	28	23	17	14	17.82	Cofilin OS=Saccharomyces cerevisiae (strain	
55	16	1	iPRG_201:	P0C0V8	719	9797	42	29	15	12	207.43	40S ribosomal protein S21-A OS=Saccharom	
56	16	2	iPRG_201:	Q3E754	694	9811	41	28	15	12	148.28	40S ribosomal protein S21-B OS=Saccharom	
57	17	1	iPRG_201:	P40212	708	22511	38	28	19	12	10.14	60S ribosomal protein L13-B OS=Saccharomy	
58	17	2	iPRG_201:	Q12690	447	22540	28	19	18	11	5.1	60S ribosomal protein L13-A OS=Saccharomy	

Sum=13232690.73

You get the idea



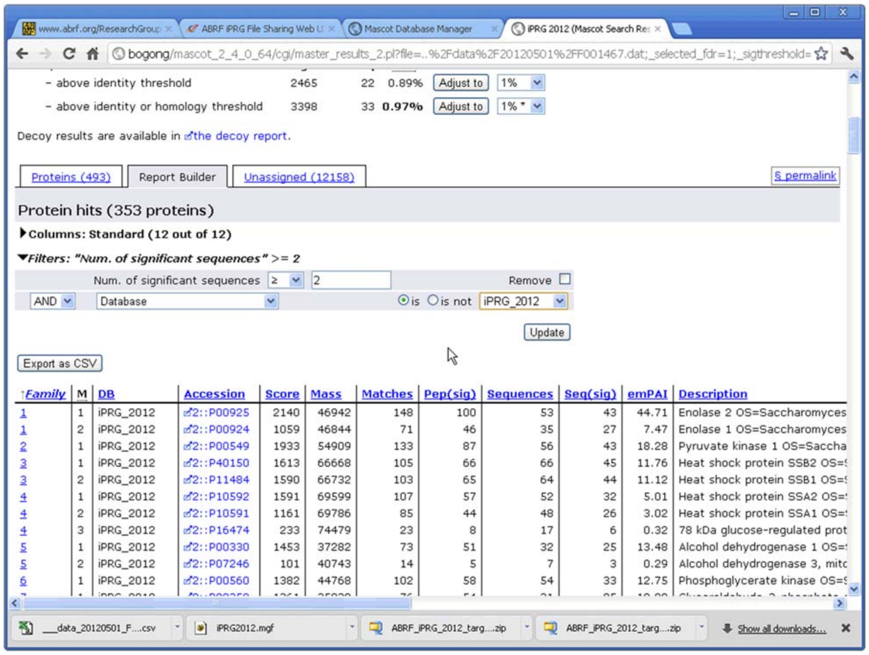
Microsoft Excel - ____data_20120501_F001467_dat_rf_reportbuilder.csv

A28 = Filters:

	A	B	C	D	E	F	G	H	I	J	K	
25	Preferre	All entries										
26	Show P	no										
27												
28	Filters:	Num. of significant sequences >= 2										
29												
30	Family	Member	Database	Accession	Score	Mass	Num. of matches	Num. of significant matches	Num. of sequences	Num. of significant sequences	emPAI	Description
31	1	1	iPRG_2012	P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyces cere
32	1	2	iPRG_2012	P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyces cere
33	2	1	iPRG_2012	P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Saccharomyc
34	3	1	iPRG_2012	P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS=Sacch
35	3	2	iPRG_2012	P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS=Sacch
36	4	1	iPRG_2012	P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS=Sacch
37	4	2	iPRG_2012	P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS=Sacch
38	4	3	iPRG_2012	P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated protein hor
39	5	1	iPRG_2012	P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS=Sacch
40	5	2	iPRG_2012	P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mitochonc
41	6	1	iPRG_2012	P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=Sacch
42	7	1	iPRG_2012	P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate dehydro
43	7	2	iPRG_2012	P00358	1242	35938	69	48	29	24	9.89	Glyceraldehyde-3-phosphate dehydro
44	7	3	iPRG_2012	P00360	505	35842	30	20	14	12	2.47	Glyceraldehyde-3-phosphate dehydro
45	7	4	iPRG_2012	P04406	41	36201	4	2	4	2	0.21	Glyceraldehyde-3-phosphate dehydro
46	8	1	iPRG_2012	P06169	1289	61685	44	41	28	26	4.7	Pyruvate decarboxylase isozyme 1 C
47	9	1	iPRG_2012	P00950	1031	27592	67	44	32	25	34.97	Phosphoglycerate mutase 1 OS=Sac
48	10	1	iPRG_2012	P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B OS=Sa
49	10	2	iPRG_2012	P07280	1014	15907	51	38	16	13	22.71	40S ribosomal protein S19-A OS=Sa
50	11	1	contaminants	P00761	922	25078	37	27	7	6	2.89	SWISS-PROT:P00761|TRYP_PIG Tr
51	12	1	iPRG_2012	P32324	784	93686	49	33	33	23	1.44	Elongation factor 2 OS=Saccharomy
52	13	1	iPRG_2012	P16521	771	116727	62	33	47	30	1.52	Elongation factor 3A OS=Saccharom
53	14	1	iPRG_2012	P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein P2-alph
54	15	1	iPRG_2012	Q03048	721	15948	28	23	17	14	17.82	Cofilin OS=Saccharomyces cerevisia
55	16	1	iPRG_2012	P0C0V8	719	9797	42	29	15	12	207.43	40S ribosomal protein S21-A OS=Sa
56	16	2	iPRG_2012	Q3E754	694	9811	41	28	15	12	148.28	40S ribosomal protein S21-B OS=Sa

____data_20120501_F001467_dat_rf

And there we have it, a table of the reliably identified proteins, suitable for pasting into a publication, in just 9-ish mouse clicks


www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Res
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=

- above identity threshold 2465 22 0.89% Adjust to 1%
- above identity or homology threshold 3398 33 **0.97%** Adjust to 1% *

Decoy results are available in the decoy report.

Proteins (493)
Report Builder
Unassigned (12158)
§ permalink

Protein hits (353 proteins)

Columns: Standard (12 out of 12)

Filters: "Num. of significant sequences" >= 2

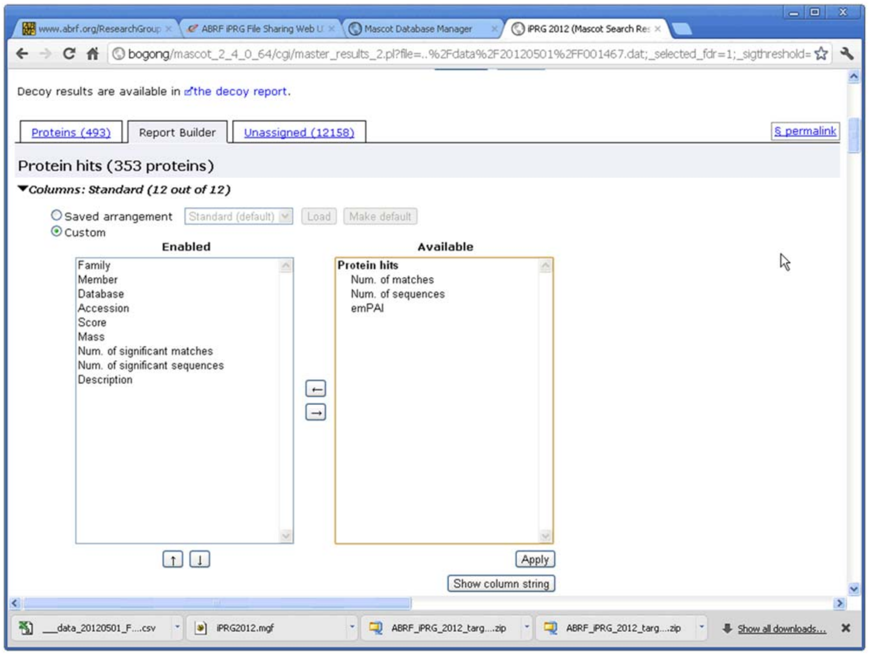
Num. of significant sequences ≥ 2 Remove □

AND Database ● is ○ is not iPRG_2012

Family	M	DB	Accession	Score	Mass	Matches	Pep(sig)	Sequences	Seq(sig)	emPAI	Description
1	1	iPRG_2012	2::P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyces
1	2	iPRG_2012	2::P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyces
2	1	iPRG_2012	2::P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Saccha
3	1	iPRG_2012	2::P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS=
3	2	iPRG_2012	2::P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS=
4	1	iPRG_2012	2::P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS=
4	2	iPRG_2012	2::P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS=
4	3	iPRG_2012	2::P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated prot
5	1	iPRG_2012	2::P00330	1453	37282	73	51	32	25	13.48	Alcohol dehydrogenase 1 OS=
5	2	iPRG_2012	2::P07246	101	40743	14	5	7	3	0.29	Alcohol dehydrogenase 3, mitc
6	1	iPRG_2012	2::P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=

___data_20120501_F....csv
iPRG2012.mgf
ABRF_iPRG_2012_targ....zip
ABRF_iPRG_2012_targ....zip
Show all downloads...

By the way, the filtering is very flexible, with lots of useful terms. Another thing that you could easily do would be to exclude proteins from the contaminants database



bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001467.dat;_selected_fdr=1;_sigthreshold=

Decoy results are available in the decoy report.

Proteins (493) | Report Builder | Unassigned (12158) | § permalink

Protein hits (353 proteins)

▼*Columns: Standard (12 out of 12)*

○ Saved arrangement Standard (default) Load Make default
● Custom

Enabled

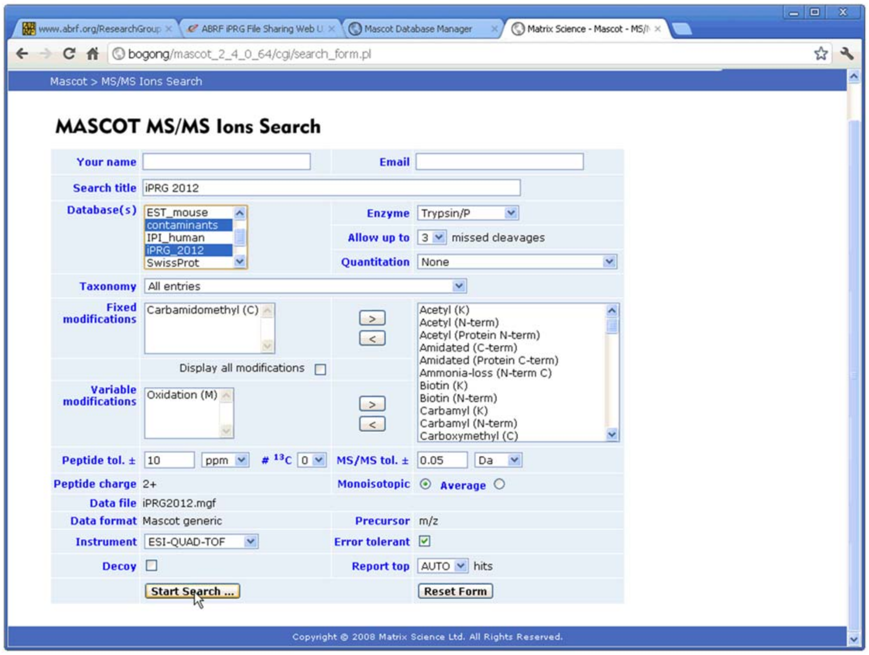
Family
Member
Database
Accession
Score
Mass
Num. of significant matches
Num. of significant sequences
Description

Available

Protein hits
Num. of matches
Num. of sequences
emPAI

www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Res
←
→
↑
↓
Apply
Show column string
___data_20120501_F....csv
iPRG2012.mgf
ABRF_iPRG_2012_targ....zip
ABRF_iPRG_2012_targ....zip
Show all downloads...

The columns section of Report Manager allows you to choose which columns to include and, if required, change their order



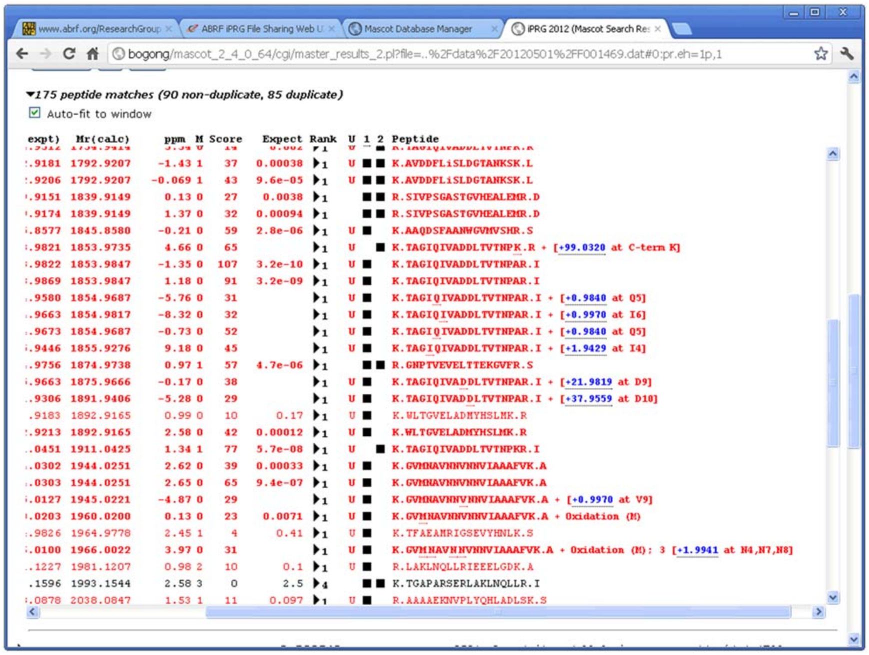
www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
Matrix Science - Mascot - MS/I
bogong/mascot_2_4_0_64/cgi/search_form.pl
Mascot > MS/MS Ions Search

MASCOT MS/MS Ions Search

Your name		Email	
Search title	iPRG 2012		
Database(s)	EST_mouse contaminants IPI_human iPRG_2012 SwissProt	Enzyme	Trypsin/P
		Allow up to	3 missed cleavages
		Quantitation	None
Taxonomy	All entries		
Fixed modifications	Carbamidomethyl (C)		Acetyl (K) Acetyl (N-term) Acetyl (Protein N-term) Amidated (C-term) Amidated (Protein C-term) Ammonia-loss (N-term C) Biotin (K) Biotin (N-term) Carbamyl (K) Carbamyl (N-term) Carboxymethyl (C)
	Display all modifications □		
Variable modifications	Oxidation (M)		
Peptide tol. ±	10 ppm # ^{13}C 0	MS/MS tol. ±	0.05 Da
Peptide charge	2+	Monoisotopic	● Monoisotopic ○ Average
Data file	iPRG2012.mgf		
Data format	Mascot generic	Precursor	m/z
Instrument	ESI-QUAD-TOF	Error tolerant	☑
Decoy	□	Report top	AUTO hits

MASCOT : *From Search Results to Publication* © 2012 Matrix Science MATRIX SCIENCE

Now, the main goal of the iPRG2012 study was to characterise modifications. Quickest way to find out what modifications might be present is an error tolerant search.


www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
iPRG 2012 (Mascot Search Re

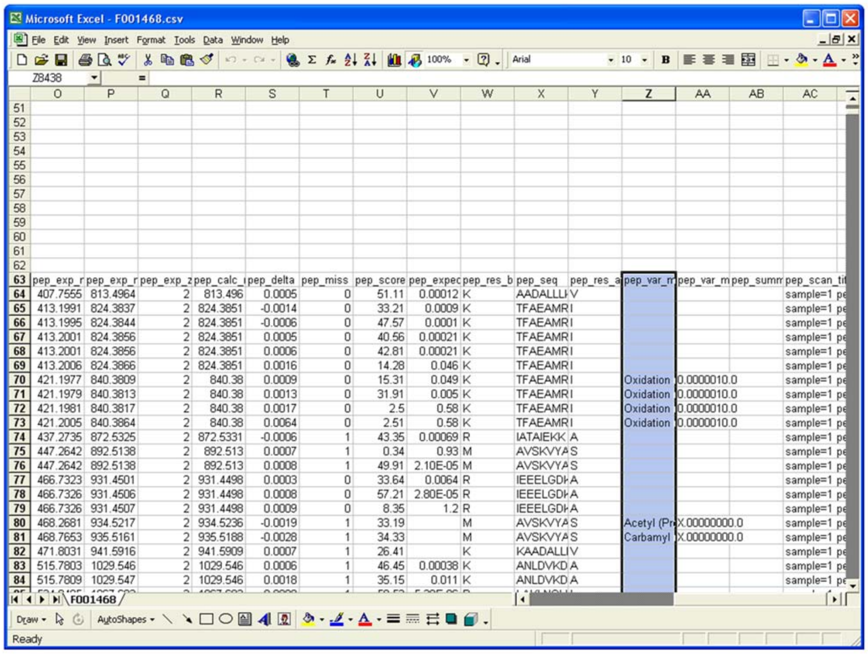
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120501%2FF001469.dat#0:pr.eh=1p,1

▾175 peptide matches (90 non-duplicate, 85 duplicate)

☑ Auto-fit to window

expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	1	2	Peptide
.9181	1792.9207	-1.43	1	37	0.00038	▶1	U	■	■	K.AVDDFLiSLDGTANKSK.L
.9206	1792.9207	-0.069	1	43	9.6e-05	▶1	U	■	■	K.AVDDFLiSLDGTANKSK.L
.9151	1839.9149	0.13	0	27	0.0038	▶1		■	■	R.SIVPSGASTGVHEALEMR.D
.9174	1839.9149	1.37	0	32	0.00094	▶1		■	■	R.SIVPSGASTGVHEALEMR.D
.8577	1845.8580	-0.21	0	59	2.8e-06	▶1	U	■		K.AAQDSFAANWGVMVSHR.S
.9821	1853.9735	4.66	0	65		▶1	U		■	K.TAGIQIVADDLTVTNPK.R + [+99.0320 at C-term K]
.9822	1853.9847	-1.35	0	107	3.2e-10	▶1	U	■		K.TAGIQIVADDLTVTNPAR.I
.9869	1853.9847	1.18	0	91	3.2e-09	▶1	U	■		K.TAGIQIVADDLTVTNPAR.I
.9580	1854.9687	-5.76	0	31		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+0.9840 at Q5]
.9663	1854.9817	-8.32	0	32		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+0.9970 at I6]
.9673	1854.9687	-0.73	0	52		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+0.9840 at Q5]
.9446	1855.9276	9.18	0	45		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+1.9429 at I4]
.9756	1874.9738	0.97	1	57	4.7e-06	▶1		■	■	R.GNPTVEVELTTEKGVFR.S
.9663	1875.9666	-0.17	0	38		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+21.9819 at D9]
.9306	1891.9406	-5.28	0	29		▶1	U	■		K.TAGIQIVADDLTVTNPAR.I + [+37.9559 at D10]
.9183	1892.9165	0.99	0	10	0.17	▶1	U	■		K.WLTGVELADMYHSLMK.R
.9213	1892.9165	2.58	0	42	0.00012	▶1	U	■		K.WLTGVELADMYHSLMK.R
.0451	1911.0425	1.34	1	77	5.7e-08	▶1	U		■	K.TAGIQIVADDLTVTNPKR.I
.0302	1944.0251	2.62	0	39	0.00033	▶1	U	■		K.GVMNAVNNVNNVIAAAFVK.A
.0303	1944.0251	2.65	0	65	9.4e-07	▶1	U	■		K.GVMNAVNNVNNVIAAAFVK.A
.0127	1945.0221	-4.87	0	29		▶1	U	■		K.GVMNAVNNVNNVIAAAFVK.A + [+0.9970 at V9]
.0203	1960.0200	0.13	0	23	0.0071	▶1	U	■		K.GVMNAVNNVNNVIAAAFVK.A + Oxidation (M)
.9826	1964.9778	2.45	1	4	0.41	▶1	U	■		K.TFAEAMRIGSEVYHNLK.S
.0100	1966.0022	3.97	0	31		▶1	U	■		K.GVMNAVNNVNNVIAAAFVK.A + Oxidation (M); 3 [+1.9941 at N4,N7,N8]
.1227	1981.1207	0.98	2	10	0.1	▶1	U	■		R.LAKLNQLLRIEEELGDK.A
.1596	1993.1544	2.58	3	0	2.5	▶4		■	■	K.TGAPARSERLAKLNQLLR.I
.0878	2038.0847	1.53	1	11	0.097	▶1	U	■		R.AAAAEKNVPLYQHLADLSK.S

The error tolerant search discovers lots of modifications, but which ones are interesting? It would be helpful if the report included a table of the modifications that had been found together with their frequency of occurrence. I can assure you that this is on the wish list. Meanwhile, the work around is to export the results as CSV and open in Excel

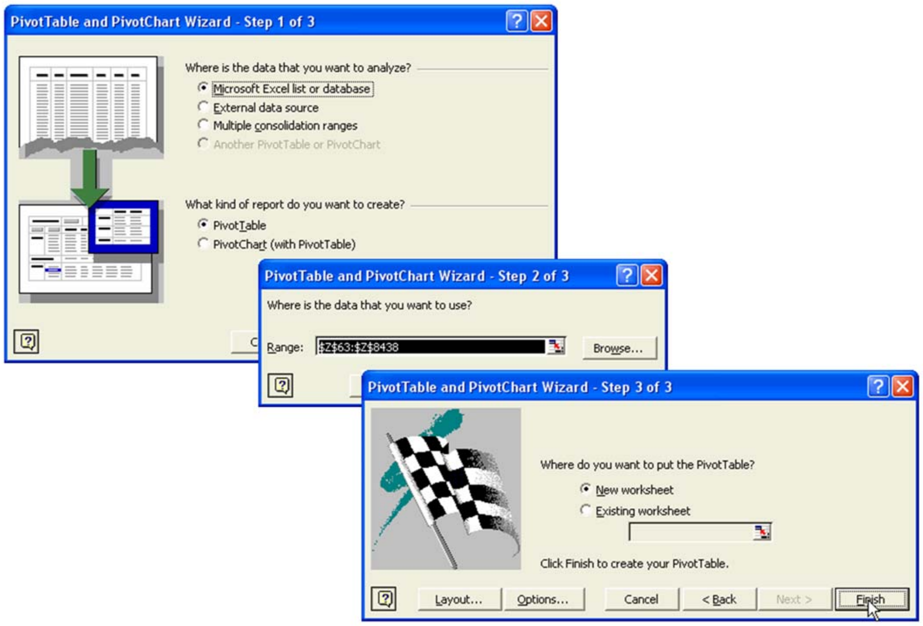

Microsoft Excel - F001468.csv
File Edit View Insert Format Tools Data Window Help
100%
Arial
10
Z8438
Draw
AutoShapes
Ready

	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
51															
52															
53															
54															
55															
56															
57															
58															
59															
60															
61															
62															
63	pep_exp_r	pep_exp_r	pep_exp_z	pep_calc_	pep_delta	pep_miss	pep_score	pep_expec	pep_res_b	pep_seq	pep_res_a	pep_var_m	pep_var_m	pep_summ	pep_scan_ti
64	407.7555	813.4964	2	813.496	0.0005	0	51.11	0.00012	K	AADALLLI	V				sample=1 pe
65	413.1991	824.3837	2	824.3851	-0.0014	0	33.21	0.0009	K	TFAEAMR	I				sample=1 pe
66	413.1995	824.3844	2	824.3851	-0.0006	0	47.57	0.0001	K	TFAEAMR	I				sample=1 pe
67	413.2001	824.3856	2	824.3851	0.0005	0	40.56	0.00021	K	TFAEAMR	I				sample=1 pe
68	413.2001	824.3856	2	824.3851	0.0006	0	42.81	0.00021	K	TFAEAMR	I				sample=1 pe
69	413.2006	824.3866	2	824.3851	0.0016	0	14.28	0.046	K	TFAEAMR	I				sample=1 pe
70	421.1977	840.3809	2	840.38	0.0009	0	15.31	0.049	K	TFAEAMR	I	Oxidation	0.0000010.0		sample=1 pe
71	421.1979	840.3813	2	840.38	0.0013	0	31.91	0.005	K	TFAEAMR	I	Oxidation	0.0000010.0		sample=1 pe
72	421.1981	840.3817	2	840.38	0.0017	0	2.5	0.58	K	TFAEAMR	I	Oxidation	0.0000010.0		sample=1 pe
73	421.2005	840.3864	2	840.38	0.0064	0	2.51	0.58	K	TFAEAMR	I	Oxidation	0.0000010.0		sample=1 pe
74	437.2735	872.5325	2	872.5331	-0.0006	1	43.35	0.00069	R	IATAIEKK	A				sample=1 pe
75	447.2642	892.5138	2	892.513	0.0007	1	0.34	0.93	M	AVSKVYA	S				sample=1 pe
76	447.2642	892.5138	2	892.513	0.0008	1	49.91	2.10E-05	M	AVSKVYA	S				sample=1 pe
77	466.7323	931.4501	2	931.4498	0.0003	0	33.64	0.0064	R	IEEELGDI	A				sample=1 pe
78	466.7326	931.4506	2	931.4498	0.0008	0	57.21	2.80E-05	R	IEEELGDI	A				sample=1 pe
79	466.7326	931.4507	2	931.4498	0.0009	0	8.35	1.2	R	IEEELGDI	A				sample=1 pe
80	468.2681	934.5217	2	934.5236	-0.0019	1	33.19		M	AVSKVYA	S	Acetyl (Pr	X.00000000.0		sample=1 pe
81	468.7653	935.5161	2	935.5188	-0.0028	1	34.33		M	AVSKVYA	S	Carbamyl	X.00000000.0		sample=1 pe
82	471.8031	941.5916	2	941.5909	0.0007	1	26.41		K	KAADALLI	V				sample=1 pe
83	515.7803	1029.546	2	1029.546	0.0006	1	46.45	0.00038	K	ANLDVKD	A				sample=1 pe
84	515.7809	1029.547	2	1029.546	0.0018	1	35.15	0.011	K	ANLDVKD	A				sample=1 pe

F001468

MASCOT : *From Search Results to Publication* © 2012 Matrix Science MATRIX SCIENCE

Select the pep_var_mod column, containing the modifications, and choose Pivot table from the Data menu



PivotTable and PivotChart Wizard - Step 1 of 3

Where is the data that you want to analyze?

● Microsoft Excel list or database
○ External data source
○ Multiple consolidation ranges
○ Another PivotTable or PivotChart

What kind of report do you want to create?

● PivotTable
○ PivotChart (with PivotTable)

PivotTable and PivotChart Wizard - Step 2 of 3

Where is the data that you want to use?

Range: $Z$63:$Z$8438

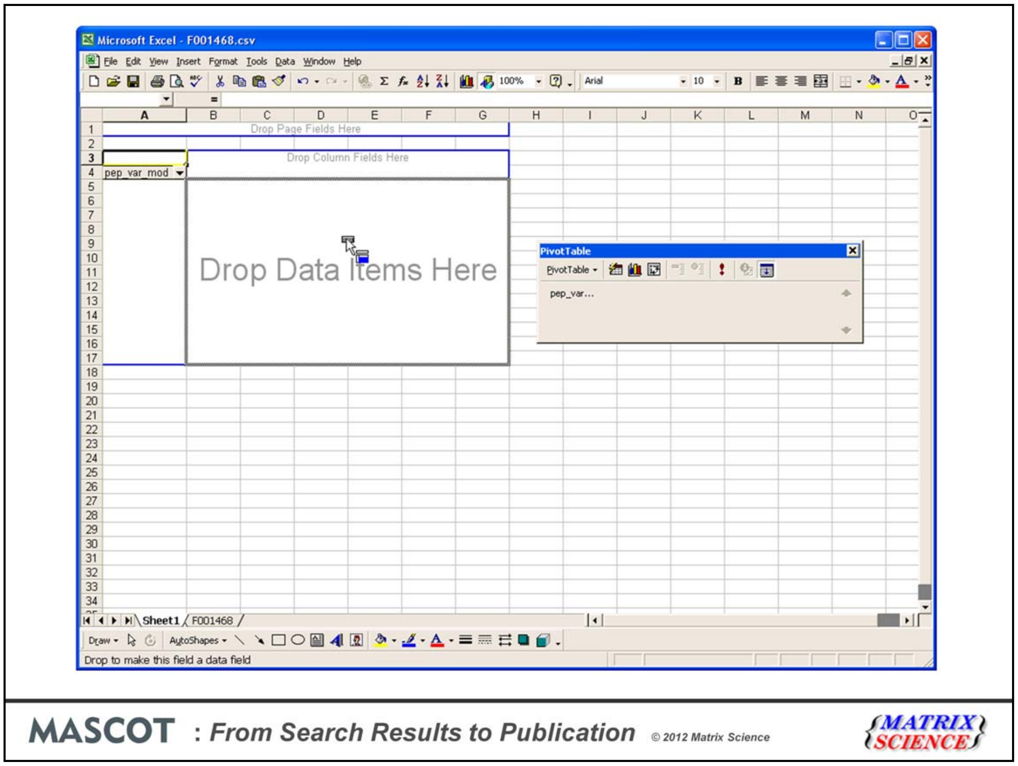
PivotTable and PivotChart Wizard - Step 3 of 3

Where do you want to put the PivotTable?

● New worksheet
○ Existing worksheet

Click Finish to create your PivotTable.

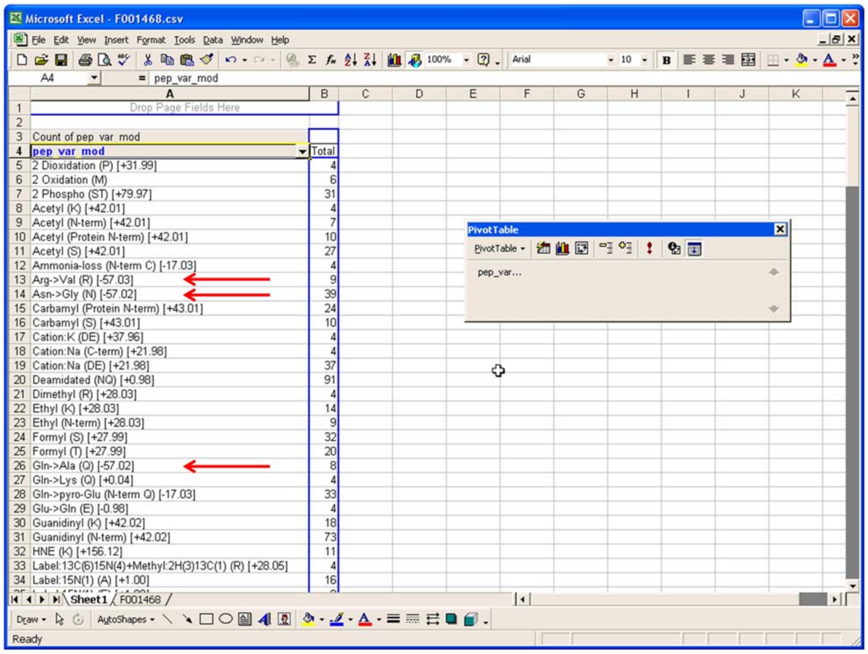
In the pivot table wizard, the defaults are OK



Drop to make this field a data field

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

Drag and drop the pep_var_mods button to both the row fields and data items area


Microsoft Excel - F001468.csv
File Edit View Insert Format Tools Data Window Help
PivotTable
pep_var...
Sheet1
F001468
Draw
AutoShapes
Ready

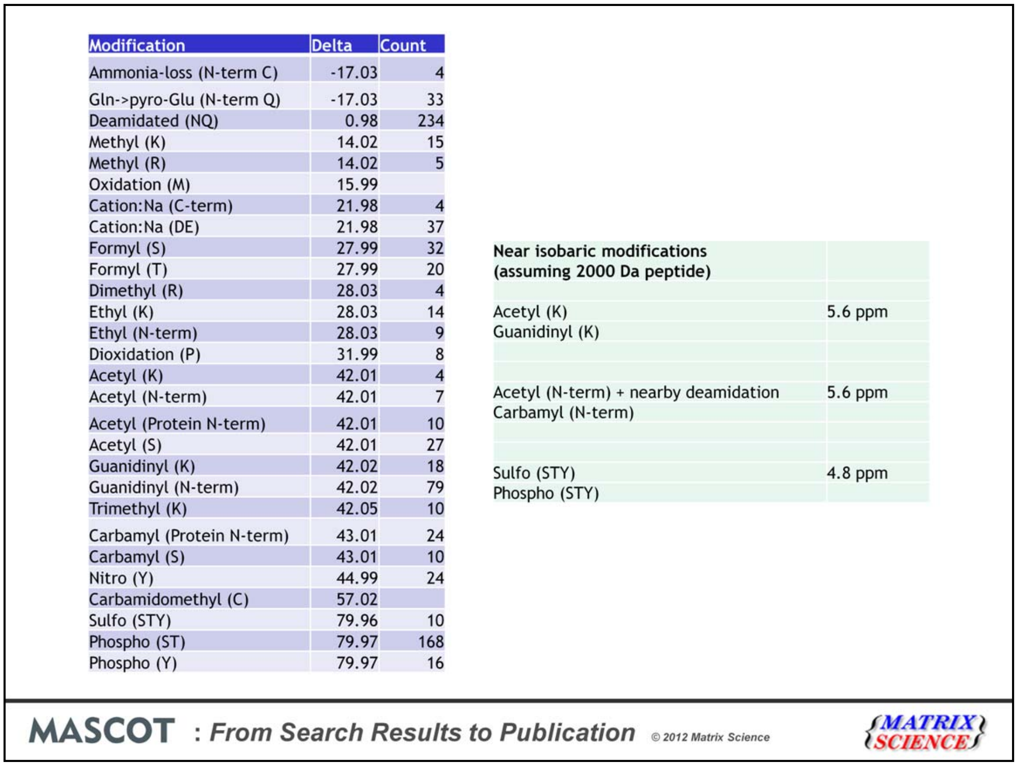
A4 = pep_var_mod

Drop Page Fields Here

	A	B
3	Count of pep_var_mod	
4	pep_var_mod	Total
5	2 Dioxidation (P) [+31.99]	4
6	2 Oxidation (M)	6
7	2 Phospho (ST) [+79.97]	31
8	Acetyl (K) [+42.01]	4
9	Acetyl (N-term) [+42.01]	7
10	Acetyl (Protein N-term) [+42.01]	10
11	Acetyl (S) [+42.01]	27
12	Ammonia-loss (N-term C) [-17.03]	4
13	Arg->Val (R) [-57.03]	9
14	Asn->Gly (N) [-57.02]	39
15	Carbamyl (Protein N-term) [+43.01]	24
16	Carbamyl (S) [+43.01]	10
17	Cation:K (DE) [+37.96]	4
18	Cation:Na (C-term) [+21.98]	4
19	Cation:Na (DE) [+21.98]	37
20	Deamidated (NQ) [+0.98]	91
21	Dimethyl (R) [+28.03]	4
22	Ethyl (K) [+28.03]	14
23	Ethyl (N-term) [+28.03]	9
24	Formyl (S) [+27.99]	32
25	Formyl (T) [+27.99]	20
26	Gln->Ala (Q) [-57.02]	8
27	Gln->Lys (Q) [+0.04]	4
28	Gln->pyro-Glu (N-term Q) [-17.03]	33
29	Glu->Gln (E) [-0.98]	4
30	Guanidinyl (K) [+42.02]	18
31	Guanidinyl (N-term) [+42.02]	73
32	HNE (K) [+156.12]	11
33	Label:13C(6)15N(4)+Methyl:2H(3)13C(1) (R) [+28.05]	4
34	Label:15N(1) (A) [+1.00]	16

And we get a table of the distinct variable modifications with a count for each. For a large search, you might want to restrict the table to just the top 50 most frequent modifications, and this is easily done (Pivot table wizard menu; Field settings; count; custom; advanced; show top 50).

The list still needs some interpretation. For example, note the presence of several mods with mass delta -57. These almost certainly indicate that carbamidomethylation is not 100% quantitative. For peptides where Cys is not modified, putting a -57 mod close by cancels out the mass difference well enough to get a decent match.

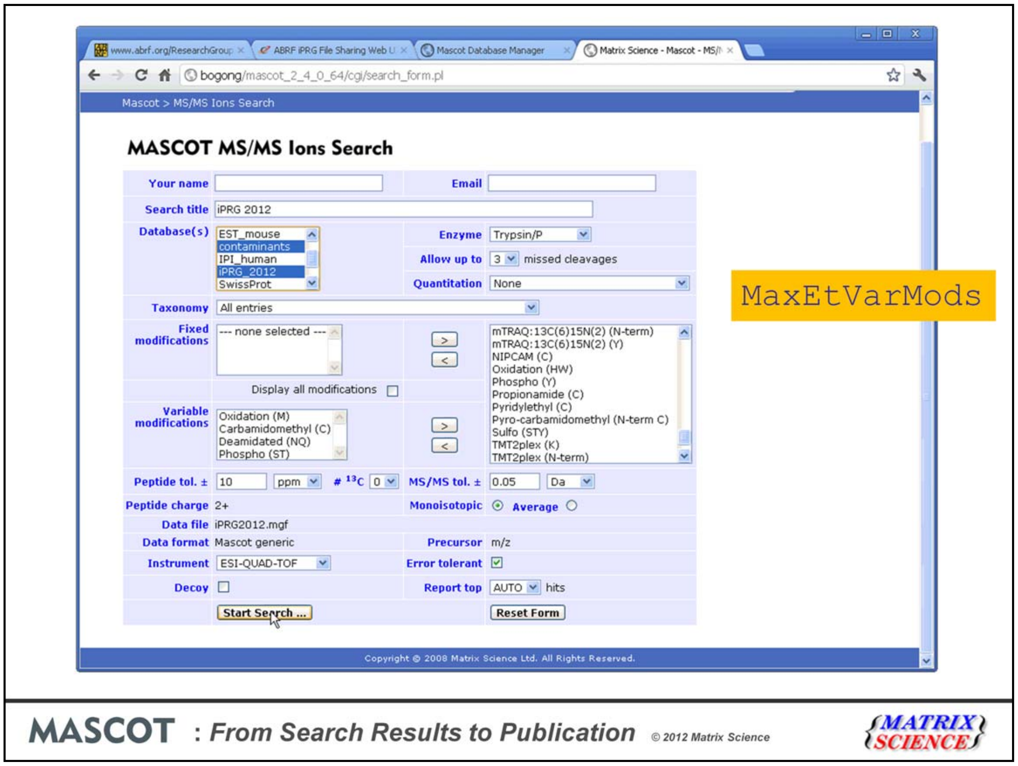


Modification	Delta	Count
Ammonia-loss (N-term C)	-17.03	4
Gln->pyro-Glu (N-term Q)	-17.03	33
Deamidated (NQ)	0.98	234
Methyl (K)	14.02	15
Methyl (R)	14.02	5
Oxidation (M)	15.99	
Cation:Na (C-term)	21.98	4
Cation:Na (DE)	21.98	37
Formyl (S)	27.99	32
Formyl (T)	27.99	20
Dimethyl (R)	28.03	4
Ethyl (K)	28.03	14
Ethyl (N-term)	28.03	9
Dioxidation (P)	31.99	8
Acetyl (K)	42.01	4
Acetyl (N-term)	42.01	7
Acetyl (Protein N-term)	42.01	10
Acetyl (S)	42.01	27
Guanidinyl (K)	42.02	18
Guanidinyl (N-term)	42.02	79
Trimethyl (K)	42.05	10
Carbamyl (Protein N-term)	43.01	24
Carbamyl (S)	43.01	10
Nitro (Y)	44.99	24
Carbamidomethyl (C)	57.02	
Sulfo (STY)	79.96	10
Phospho (ST)	79.97	168
Phospho (Y)	79.97	16

Near isobaric modifications (assuming 2000 Da peptide)	
Acetyl (K)	5.6 ppm
Guanidinyl (K)	
Acetyl (N-term) + nearby deamidation	5.6 ppm
Carbamyl (N-term)	
Sulfo (STY)	4.8 ppm
Phospho (STY)	

After further scrutiny, we end up with these as the believable modifications that occur 4 or more times. Although the mass accuracy of the data is excellent, there can still be ambiguities, such as whether we have acetyl or guanidinyl. In the case of sulfo and phospho, we can often decide which we have from differences in neutral loss behaviour. I'll come back to this later.

Where we go next depends on the goal of the experiment. In the case of the iPRG2012 study, it was to report as many matches as possible. Clearly, this is a slightly artificial case. In real life, we are more likely to be interested in a specific modification or a specific protein. But, how would one search for all of these modifications? You can't simply select them all as variable modifications; the combinatorial explosion would mean that all specificity was lost. However, it is highly unlikely that we will see two rare modifications on the same peptide. As long as we have Oxidation (M), Deamidated (NQ), Phospho (ST), and Carbamidomethyl (C) specified in the search as variable modifications, we shouldn't miss very much when the error tolerant search looks serially through all of the modifications in Unimod.



www.abrf.org/ResearchGroup
ABRF iPRG File Sharing Web U
Mascot Database Manager
Matrix Science - Mascot - MS/I
bogong/mascot_2_4_0_64/cgi/search_form.pl
Mascot > MS/MS Ions Search

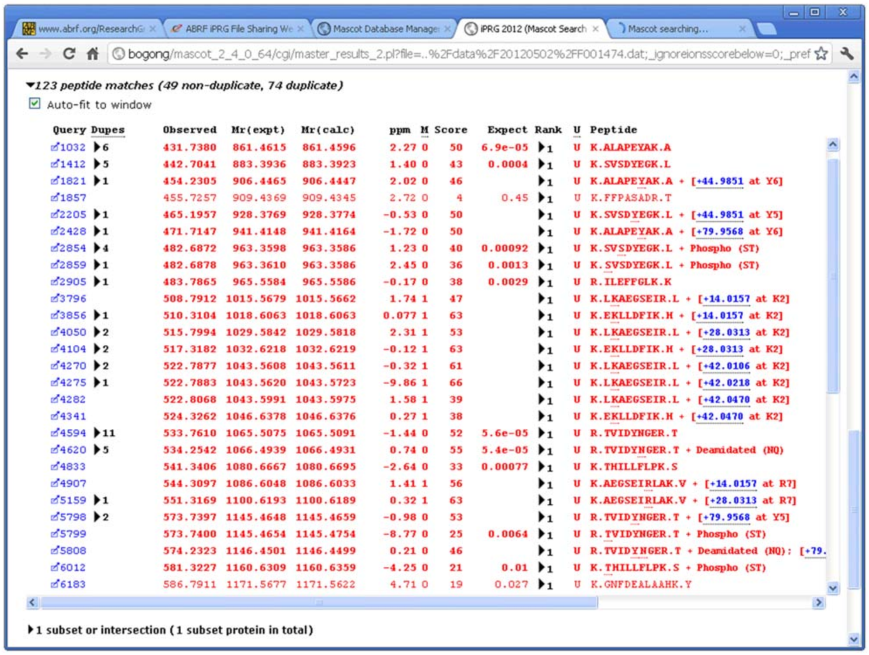
MASCOT MS/MS Ions Search

Your name		Email	
Search title	iPRG 2012		
Database(s)	EST_mouse contaminants IPI_human iPRG_2012 SwissProt	Enzyme	Trypsin/P
		Allow up to	3 missed cleavages
		Quantitation	None
Taxonomy	All entries		
Fixed modifications	--- none selected ---	> <	mTRAQ:13C(6)15N(2) (N-term) mTRAQ:13C(6)15N(2) (Y) NIPCAM (C) Oxidation (HW) Phospho (Y) Propionamide (C) Pyridylethyl (C) Pyro-carbamidomethyl (N-term C) Sulfo (STY) TMT2plex (K) TMT2plex (N-term)
	Display all modifications □		
Variable modifications	Oxidation (M) Carbamidomethyl (C) Deamidated (NQ) Phospho (ST)	> <	
Peptide tol. ±	10 ppm # ^{13}C 0	MS/MS tol. ±	0.05 Da
Peptide charge	2+	Monoisotopic	● Average ○
Data file	iPRG2012.mgf		
Data format	Mascot generic	Precursor	m/z
Instrument	ESI-QUAD-TOF	Error tolerant	☑
Decoy	□	Report top	AUTO hits
	Start Search ...		Reset Form

MaxEtVarMods

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

Note that the default Mascot configuration only allows 2 variable mods in an error tolerant search. You’ll need to change the value of the MaxEtVarMods option to 4 or more to perform such a search.


www.abrf.org/ResearchG
ABRF iPRG File Sharing We
Mascot Database Manage
iPRG 2012 (Mascot Search
Mascot searching...
bogong/mascot_2_4_0_64/cgi/master_results_2.pl?file=..%2Fdata%2F20120502%2FF001474.dat;_ignoreionsscorebelow=0;_pref

▾123 peptide matches (49 non-duplicate, 74 duplicate)

☑ Auto-fit to window

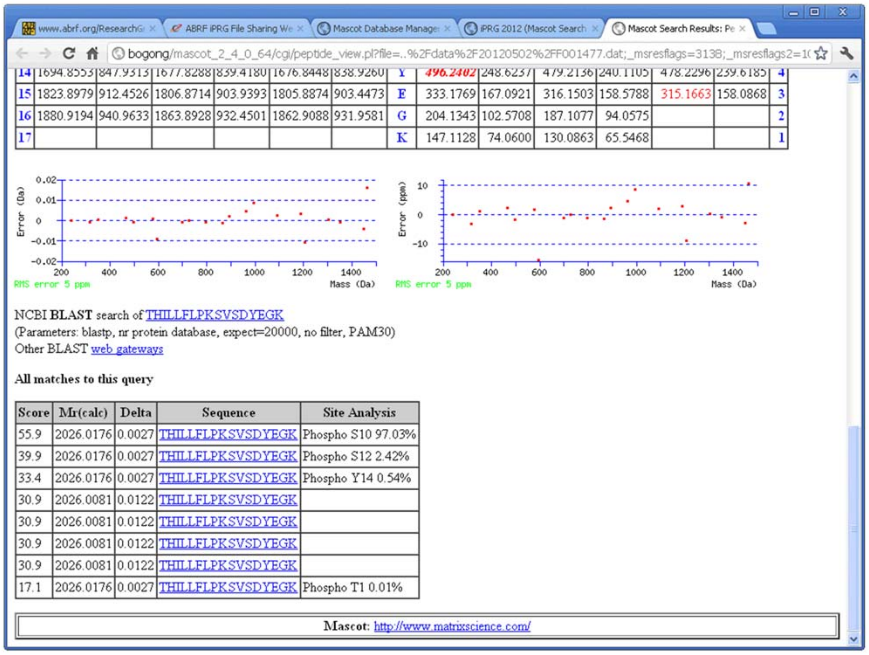
Query	Dupes	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
1032	▶6	431.7380	861.4615	861.4596	2.27	0	50	6.9e-05	▶1	U	K.ALAPEYAK.A
1412	▶5	442.7041	883.3936	883.3923	1.40	0	43	0.0004	▶1	U	K.SVSDYEGK.L
1821	▶1	454.2305	906.4465	906.4447	2.02	0	46		▶1	U	K.ALAPEYAK.A + [+44.9851 at Y6]
1857		455.7257	909.4369	909.4345	2.72	0	4	0.45	▶1	U	K.FFPASADR.T
2205	▶1	465.1957	928.3769	928.3774	-0.53	0	50		▶1	U	K.SVSDYEGK.L + [+44.9851 at Y5]
2428	▶1	471.7147	941.4148	941.4164	-1.72	0	50		▶1	U	K.ALAPEYAK.A + [+79.9568 at Y6]
2854	▶4	482.6872	963.3598	963.3586	1.23	0	40	0.00092	▶1	U	K.SVSDYEGK.L + Phospho (ST)
2859	▶1	482.6878	963.3610	963.3586	2.45	0	36	0.0013	▶1	U	K.SVSDYEGK.L + Phospho (ST)
2905	▶1	483.7865	965.5584	965.5586	-0.17	0	38	0.0029	▶1	U	R.ILEFFGLK.K
3796		508.7912	1015.5679	1015.5662	1.74	1	47		▶1	U	K.LKAEGSEIR.L + [+14.0157 at K2]
3856	▶1	510.3104	1018.6063	1018.6063	0.077	1	63		▶1	U	K.EKLLDFIK.H + [+14.0157 at K2]
4050	▶2	515.7994	1029.5842	1029.5818	2.31	1	53		▶1	U	K.LKAEGSEIR.L + [+28.0313 at K2]
4104	▶2	517.3182	1032.6218	1032.6219	-0.12	1	63		▶1	U	K.EKLLDFIK.H + [+28.0313 at K2]
4270	▶2	522.7877	1043.5608	1043.5611	-0.32	1	61		▶1	U	K.LKAEGSEIR.L + [+42.0106 at K2]
4275	▶1	522.7883	1043.5620	1043.5723	-9.86	1	66		▶1	U	K.LKAEGSEIR.L + [+42.0218 at K2]
4282		522.8068	1043.5991	1043.5975	1.58	1	39		▶1	U	K.LKAEGSEIR.L + [+42.0470 at K2]
4341		524.3262	1046.6378	1046.6376	0.27	1	38		▶1	U	K.EKLLDFIK.H + [+42.0470 at K2]
4594	▶11	533.7610	1065.5075	1065.5091	-1.44	0	52	5.6e-05	▶1	U	R.TVIDYNGER.T
4620	▶5	534.2542	1066.4939	1066.4931	0.74	0	55	5.4e-05	▶1	U	R.TVIDYNGER.T + Deamidated (NQ)
4833		541.3406	1080.6667	1080.6695	-2.64	0	33	0.00077	▶1	U	K.THILLFLPK.S
4907		544.3097	1086.6048	1086.6033	1.41	1	56		▶1	U	K.AEGSEIRLAK.V + [+14.0157 at R7]
5159	▶1	551.3169	1100.6193	1100.6189	0.32	1	63		▶1	U	K.AEGSEIRLAK.V + [+28.0313 at R7]
5798	▶2	573.7397	1145.4648	1145.4659	-0.98	0	53		▶1	U	R.TVIDYNGER.T + [+79.9568 at Y5]
5799		573.7400	1145.4654	1145.4754	-8.77	0	25	0.0064	▶1	U	R.TVIDYNGER.T + Phospho (ST)
5808		574.2323	1146.4501	1146.4499	0.21	0	46		▶1	U	R.TVIDYNGER.T + Deamidated (NQ): [+79.
6012		581.3227	1160.6309	1160.6359	-4.25	0	21	0.01	▶1	U	K.THILLFLPK.S + Phospho (ST)
6183		586.7911	1171.5677	1171.5622	4.71	0	19	0.027	▶1	U	K.GNFDEALAAHK.Y

▸1 subset or intersection (1 subset protein in total)

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

For the iPRG study, the next step would be to export the results to Excel. I don’t want to go into a lot of detail … there isn’t time … so I’ll just highlight a couple of points relating to modification characterisation.



14	1694.8553	847.9313	1677.8288	839.4180	1676.8448	838.9260	Y	***496.2402***	248.6237	479.2136	240.1105	478.2296	239.6185	4
15	1823.8979	912.4526	1806.8714	903.9393	1805.8874	903.4473	E	333.1769	167.0921	316.1503	158.5788	315.1663	158.0868	3
16	1880.9194	940.9633	1863.8928	932.4501	1862.9088	931.9581	G	204.1343	102.5708	187.1077	94.0575			2
17							K	147.1128	74.0600	130.0863	65.5468			1

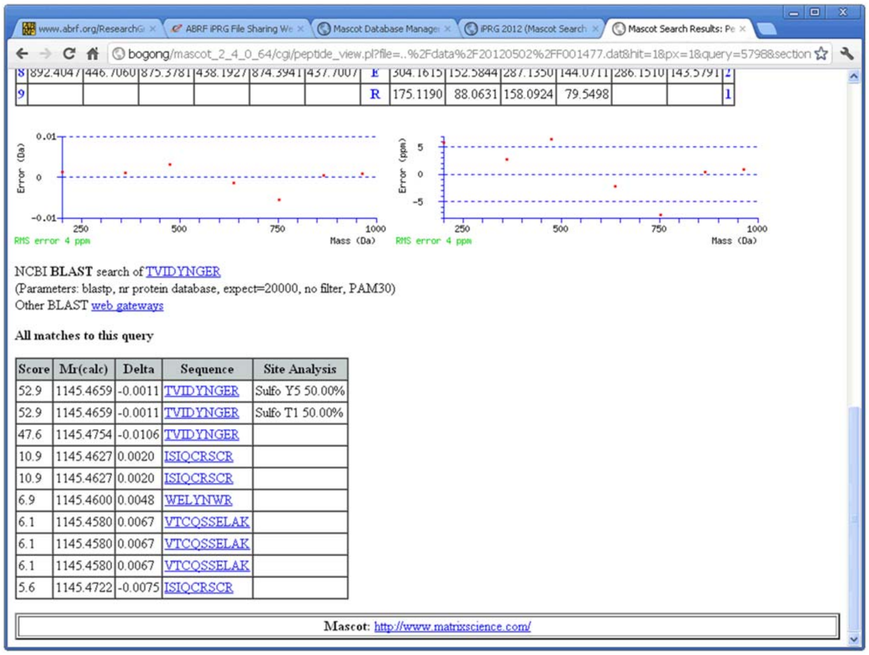
NCBI **BLAST** search of THILLFLPKSVSDYEGK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
55.9	2026.0176	0.0027	THILLFLPKSVSDYEGK	Phospho S10 97.03%
39.9	2026.0176	0.0027	THILLFLPKSVSDYEGK	Phospho S12 2.42%
33.4	2026.0176	0.0027	THILLFLPKSVSDYEGK	Phospho Y14 0.54%
30.9	2026.0081	0.0122	THILLFLPKSVSDYEGK	
30.9	2026.0081	0.0122	THILLFLPKSVSDYEGK	
30.9	2026.0081	0.0122	THILLFLPKSVSDYEGK	
30.9	2026.0081	0.0122	THILLFLPKSVSDYEGK	
17.1	2026.0176	0.0027	THILLFLPKSVSDYEGK	Phospho T1 0.01%

Mascot: http://www.matrixscience.com/

Mascot 2.4 reports site localisation probabilities using the delta score method published in MCP by Bernard Kuster’s group. Here, for example, there are 4 potential phosphorylation sites but, based on the score differences between the matches, it looks fairly clear that the site is S10. The four matches with scores of 30.9 are for Sulfation on each of these four sites. Because the Sulfo modification is lost quantitatively on MS/MS fragmentation, there is no preference for any particular site; the MS/MS is identical in all cases. For this peptide, we can be confident that the modification is phospho because we see extensive loss of 98 from the fragments, and matching these gives the higher score.


www.abrf.org/ResearchG
ABRF iPRG File Sharing We
Mascot Database Manage
iPRG 2012 (Mascot Search
Mascot Search Results: Pe
bogong/mascot_2_4_0_64/cgi/peptide_view.pl?file=..%2Fdata%2F20120502%2FF001477.dat&hit=1&px=1&query=5798&section

8	892.4047	446.7060	875.3781	438.1927	874.3941	437.7007	E	304.1615	152.5844	287.1350	144.0711	286.1510	143.5791	2
9							R	175.1190	88.0631	158.0924	79.5498			1

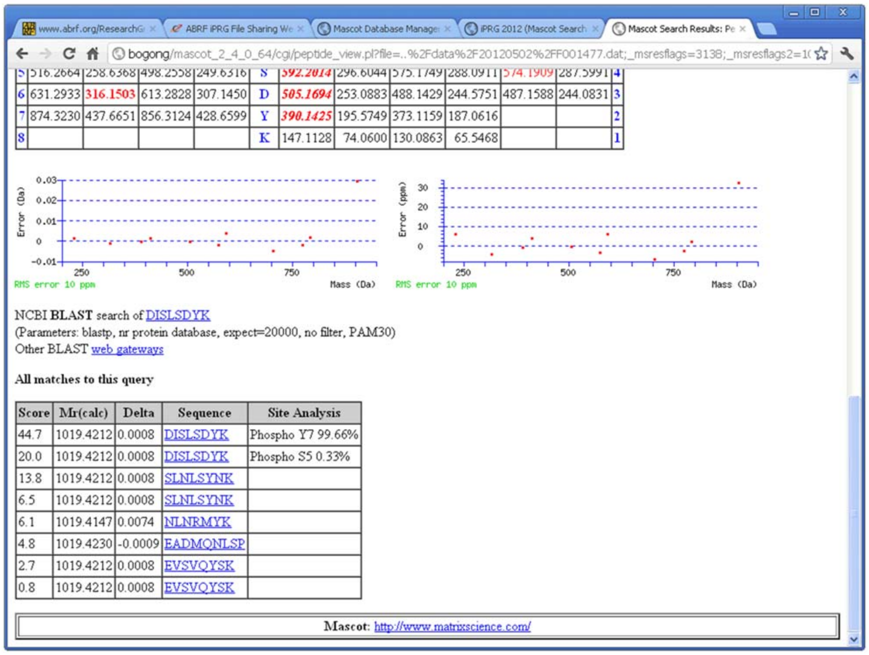
NCBI **BLAST** search of TVIDYNGER
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
52.9	1145.4659	-0.0011	TVIDYNGER	Sulfo Y5 50.00%
52.9	1145.4659	-0.0011	TVIDYNGER	Sulfo T1 50.00%
47.6	1145.4754	-0.0106	TVIDYNGER	
10.9	1145.4627	0.0020	ISIQCRSCR	
10.9	1145.4627	0.0020	ISIQCRSCR	
6.9	1145.4600	0.0048	WELYNWR	
6.1	1145.4580	0.0067	VTCQSSELAK	
6.1	1145.4580	0.0067	VTCQSSELAK	
6.1	1145.4580	0.0067	VTCQSSELAK	
5.6	1145.4722	-0.0075	ISIQCRSCR	

Mascot: http://www.matrixscience.com/

Here is a peptide that has sulfo as the top scoring match. There is simply nothing in the MS/MS to distinguish modification at T1 and Y5. The third match with the greater mass error is for Phospho on T1. Phospho on Y gets a very poor score, not even in the top 10, because it takes out most of the matching y ions



5	516.2664	258.6368	498.2558	249.6316	S	592.2014	296.6044	575.1749	288.0911	574.1909	287.5991	4
6	631.2933	316.1503	613.2828	307.1450	D	505.1694	253.0883	488.1429	244.5751	487.1588	244.0831	3
7	874.3230	437.6651	856.3124	428.6599	Y	390.1425	195.5749	373.1159	187.0616			2
8					K	147.1128	74.0600	130.0863	65.5468			1

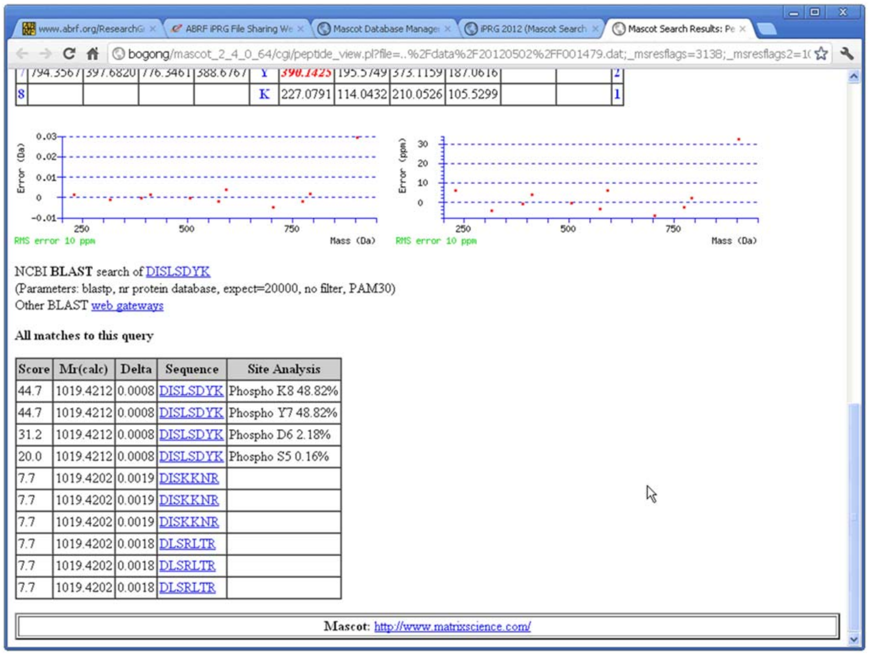
NCBI **BLAST** search of DISLSDYK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
44.7	1019.4212	0.0008	DISLSDYK	Phospho Y7 99.66%
20.0	1019.4212	0.0008	DISLSDYK	Phospho S5 0.33%
13.8	1019.4212	0.0008	SLNLSYNK	
6.5	1019.4212	0.0008	SLNLSYNK	
6.1	1019.4147	0.0074	NLNRMYK	
4.8	1019.4230	-0.0009	EADMQNLSP	
2.7	1019.4212	0.0008	EVSVQYSK	
0.8	1019.4212	0.0008	EVSVQYSK	

Mascot: http://www.matrixscience.com/

A word of warning. Site localisation is often a function of the modifications selected for the search. Here, for example, is another peptide where the localisation looks excellent when we search with Phospho on S, T, and Y. But, in rare cases, other residues can be phosphorylated. Post translational modification of C, R, D, K and H are all documented in RESID and Unimod. If we were to perform a search where these unusual specificities were included …



7	794.3567	397.6820	776.3461	388.6767	Y	390.1425	195.5749	373.1159	187.0616			2
8					K	227.0791	114.0432	210.0526	105.5299			1

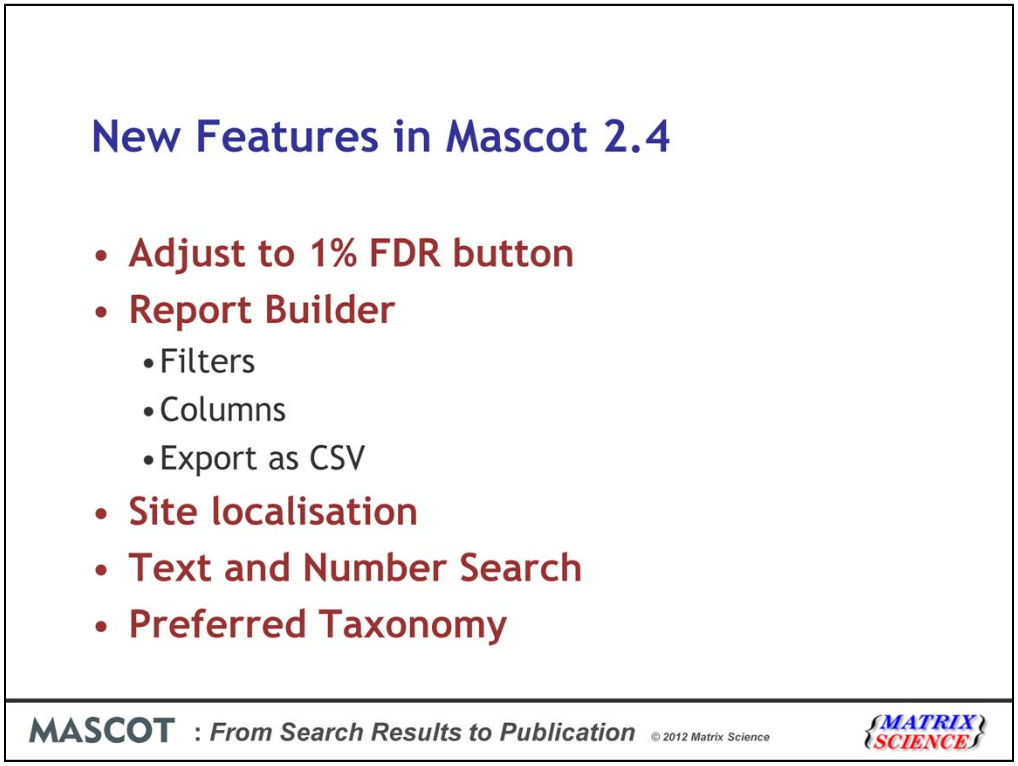
NCBI **BLAST** search of DISLSDYK
(Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
Other BLAST web gateways

All matches to this query

Score	Mr(calc)	Delta	Sequence	Site Analysis
44.7	1019.4212	0.0008	DISLSDYK	Phospho K8 48.82%
44.7	1019.4212	0.0008	DISLSDYK	Phospho Y7 48.82%
31.2	1019.4212	0.0008	DISLSDYK	Phospho D6 2.18%
20.0	1019.4212	0.0008	DISLSDYK	Phospho S5 0.16%
7.7	1019.4202	0.0019	DISKKNR	
7.7	1019.4202	0.0019	DISKKNR	
7.7	1019.4202	0.0019	DISKKNR	
7.7	1019.4202	0.0018	DLSRLTR	
7.7	1019.4202	0.0018	DLSRLTR	
7.7	1019.4202	0.0018	DLSRLTR	

Mascot: http://www.matrixscience.com/

Things are no longer so clear cut. In reality, this is highly likely to be Phospho on Y7 because Phospho K is very unusual. But, when we say we are confident that the phosphate is on Y7, we should really add “assuming the only possibilities are S, T, and Y”



New Features in Mascot 2.4

- Adjust to 1% FDR button
- Report Builder
  - Filters
  - Columns
  - Export as CSV
- Site localisation
- Text and Number Search
- Preferred Taxonomy

MASCOT : *From Search Results to Publication* © 2012 Matrix Science

MATRIX SCIENCE

To summarise, we’ve seen practical examples of several of the new features in the Mascot 2.4 reports. The two that I didn’t mention are the much enhanced text and number search facility. For example, you can search the protein family report for a modification or a mass value. Finally, the facility to set a preferred taxonomy. This wasn’t relevant here, because the database was essentially yeast proteins, but in other searches, you might want to search a wide taxonomy, e.g. green plants, and where there are two proteins with equal scores, always choose the protein from (say) maize, because that is the particular subject of your research.