Mascot Insight

A new application to organise, analyse and report Mascot search and quantitation results

What is Mascot Insight?

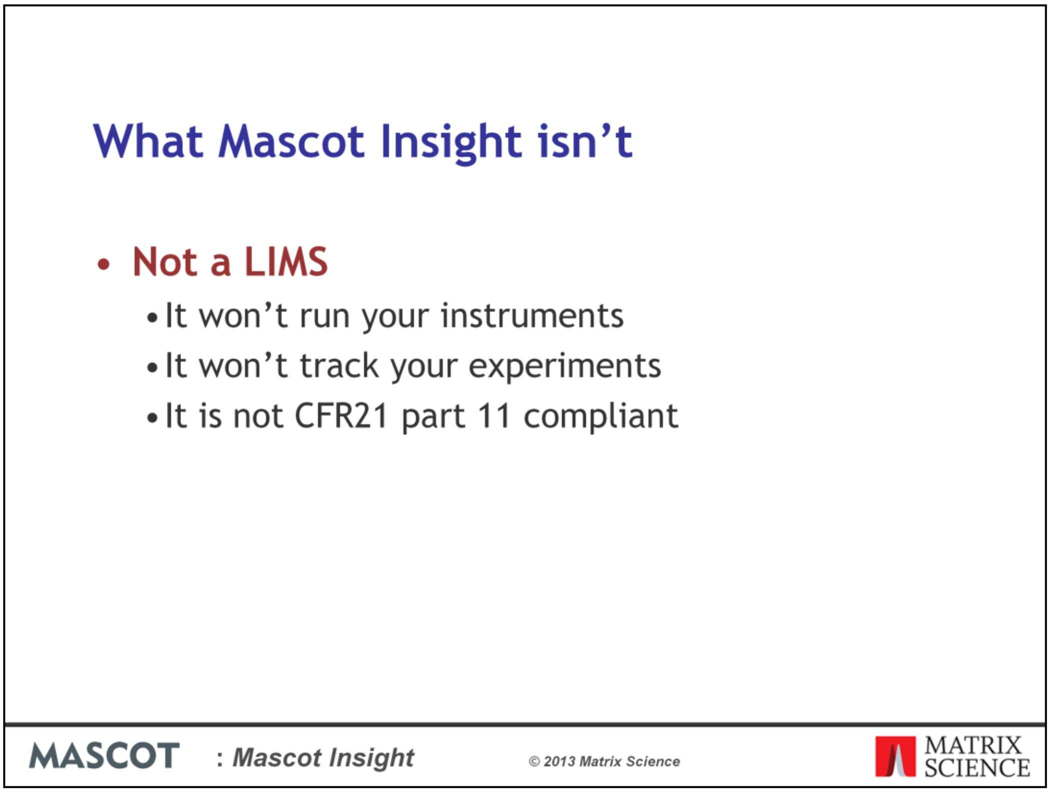
- A new application for:
  - Organising and managing results
    - Mascot, protXML, mzIdentML, Distiller de novo supported
  - Merging Mascot search and quantitation results
  - Annotating results (e.g. GO, Interactions databases, manual annotation and approval)
  - Comparing between datasets
  - Visualising, reporting and exporting results

Mascot Insight is a new application designed to help you to organise and manage your Mascot search and quantitation results.

Mascot Insight provides flexible merging of your Mascot search and quantitation results, including MS-1 based quantitation data from Mascot Distiller, such as SILAC data, and MS2 based quantitation such as iTRAQ and TMT.

Mascot Insight allows you to annotate your results automatically using Gene Ontology and molecular interactions databases and manually, to approve protein hits.

Mascot Insight also provides a wide range of reports covering areas such as dataset comparison, quantitation analysis, plotting charts etc., and provides a variety of exports to allow you to easily share results and export data in machine readable formats

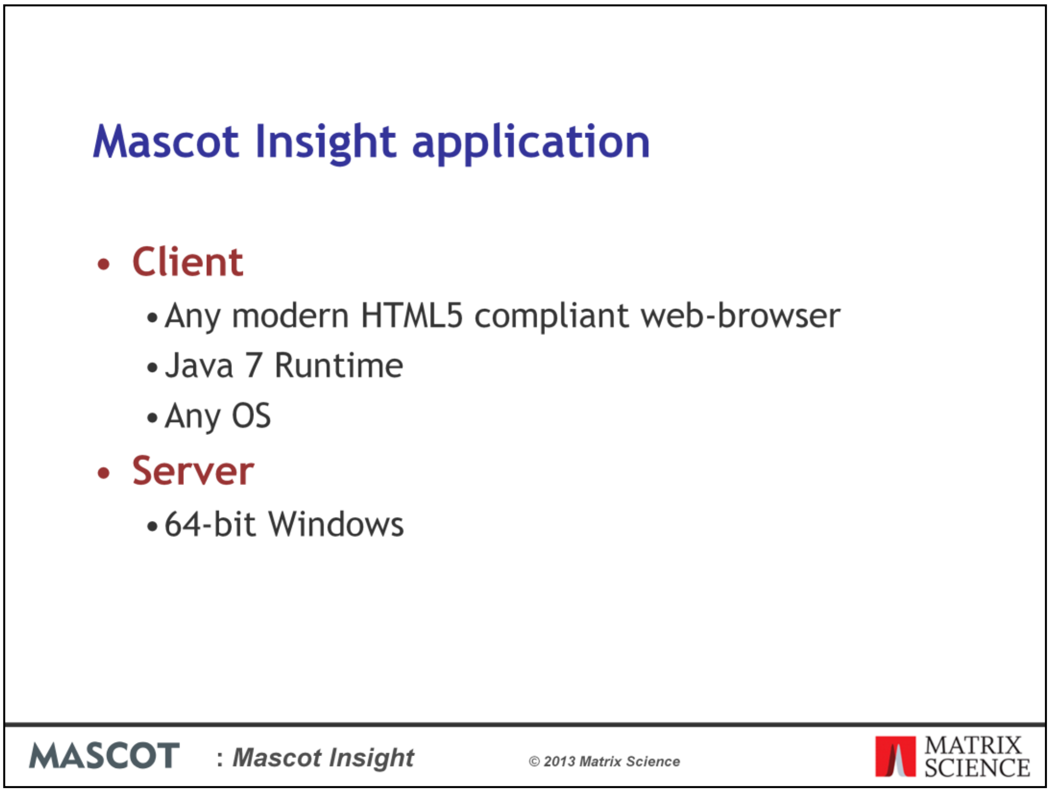


What Mascot Insight isn't

- **Not a LIMS**
  - It won't run your instruments
  - It won't track your experiments
  - It is not CFR21 part 11 compliant

MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

It is also important to mention one thing that Mascot Insight isn't. It is not a LIMS system. Although you can use it to organise your results and report on them, and it also allows you to record details about your experimental workflow, if you wish, it does not implement many of the functions you would associate with a LIMS such as instrument integration, sample tracking and FDA regulatory compliance.



Mascot Insight application

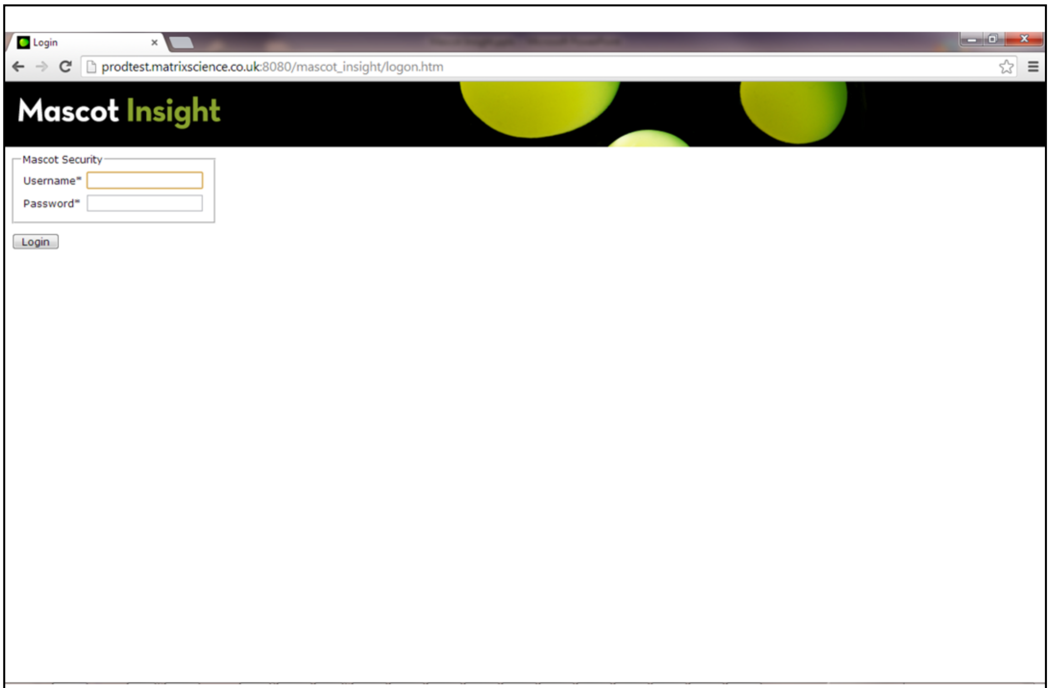
- **Client**
  - Any modern HTML5 compliant web-browser
  - Java 7 Runtime
  - Any OS
- **Server**
  - 64-bit Windows

MASCOT : *Mascot Insight*

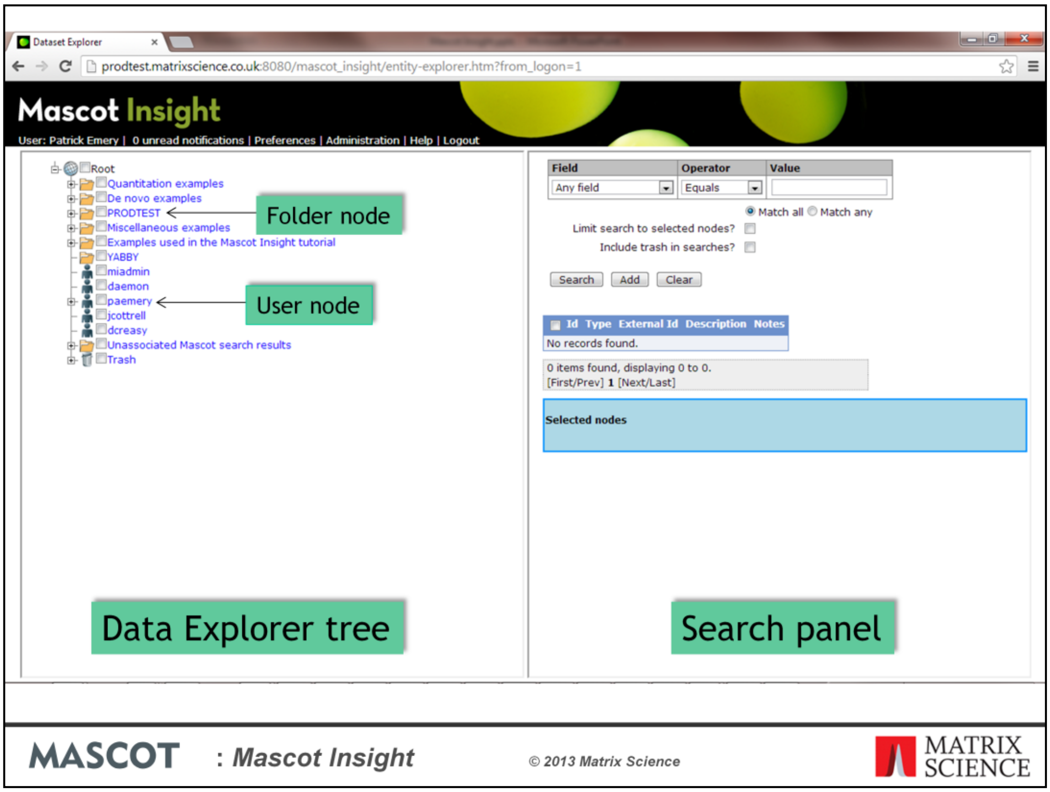
MATRIX SCIENCE

Mascot Insight runs as a web application, and the user interface is a web browser, like Mascot Server. All modern HTML5 compliant web browsers are supported (such as Chrome, Firefox and the more recent versions of Internet Explorer). Much of the functionality of the system is provided by a Java applet, so you will need the latest version of Java runtime installed on the client.

The Mascot Insight server must be 64-bit Microsoft Windows. If your Mascot server is running on a 64-bit Windows OS and you have plenty of spare capacity, you can optionally install Mascot Insight onto the same system.



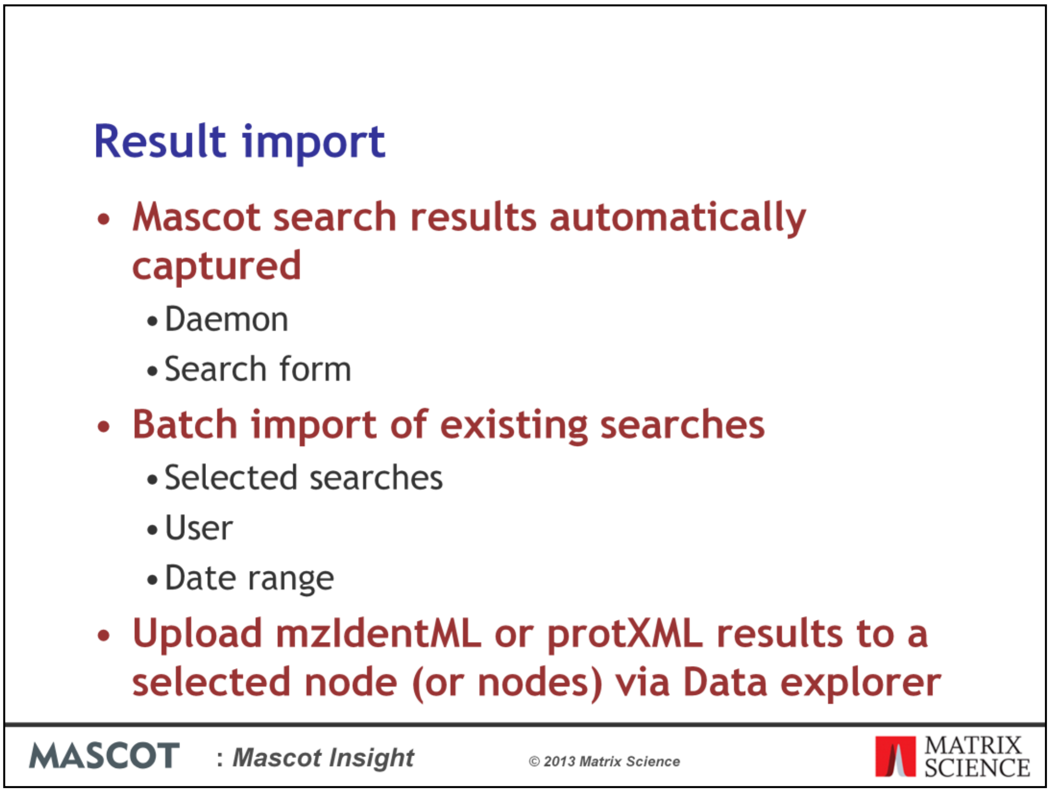
Mascot security provides a single login to both Mascot Server and to the Mascot Insight web application to identify users and to control user access levels.



Once you've logged in, this is the home page of the application. On the left we have the Data Explorer tree. This is where the results are organised and displayed, with different types of node for different types of information. Data can be manually positioned on the tree, or structures can be automatically generated to your own design.

Here, we can see two main types of node – User nodes and Folder nodes. Each user is given a User node automatically, to organise their own data and experiments. The folder node is a generic node, used to group related nodes. Other node types available in the system include Project, Study, Experiment, Sample and Result type nodes. These different node types allow you to record different pieces of information (meta-data). You can modify the meta-data fields associated with a node or create a new node type.

On the right hand side of the home page, we have the Search panel, which allows us to rapidly search the tree to find a particular node or set of nodes. Security is role-based and handled by Mascot security. Exactly what you see on the dataset explorer tree when you log in depends on your users roles, and on what access settings have been given to particular nodes on the tree by their owners.



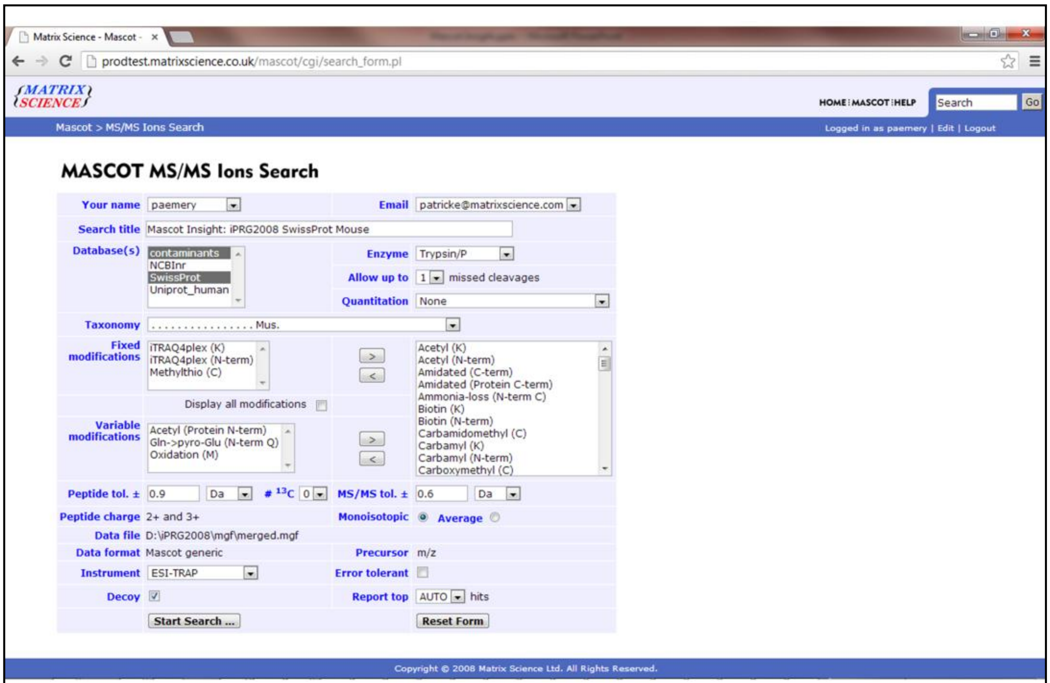
Result import

- Mascot search results automatically captured
  - Daemon
  - Search form
- Batch import of existing searches
  - Selected searches
  - User
  - Date range
- Upload mzIdentML or protXML results to a selected node (or nodes) via Data explorer

Mascot search results are automatically and transparently captured and imported into the system. For the highest level of integration, we’d recommend using Mascot Daemon to submit your searches and Mascot Distiller for your raw data reduction and MS-1 based quantitation. Mascot Daemon comes with your Mascot server, but Mascot Distiller is a separate product. However, neither is required and you can simply submit a search through the standard search form, or third party client software, if you wish.

You can also batch import existing search results from your Mascot server. This can be the entire search history or can be limited to selected searches, e.g. a selected user or a date range.

Other result types, such as results in mzIdentML or protXML formats, can be imported via the Data Explorer. Multiple files can be zipped together and the archive uploaded if you have multiple searches you want to import in a batch.

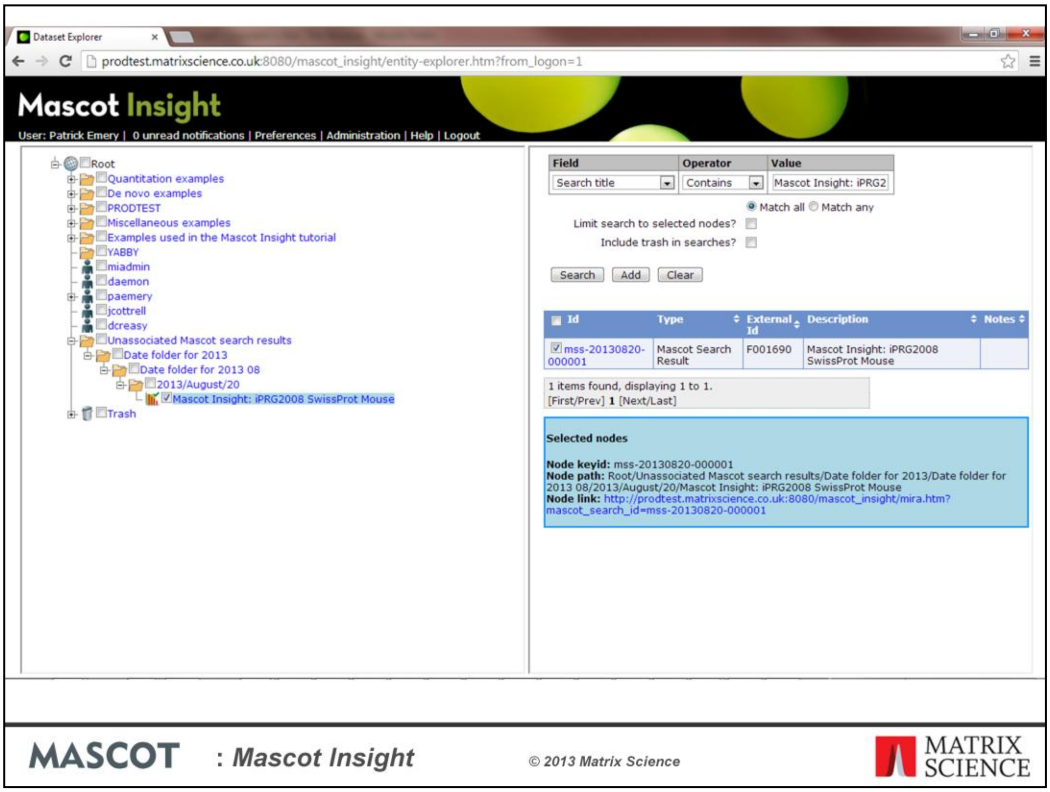


MASCOT MS/MS Ions Search

Your name: paemery
Email: patricke@matrixscience.com
Search title: Mascot Insight: iPRG2008 SwissProt Mouse
Database(s): contaminants, SwissProt (selected from: contaminants, NCBInr, SwissProt, Uniprot_human)
Enzyme: Trypsin/P
Allow up to 1 missed cleavages
Quantitation: None
Taxonomy: . . . . Mus.
Fixed modifications: iTRAQ4plex (K), iTRAQ4plex (N-term), Methylthio (C)
Display all modifications ☐
Variable modifications: Acetyl (Protein N-term), Gln->pyro-Glu (N-term Q), Oxidation (M)

Peptide tol. ± 0.9 Da # ^{13}C 0
MS/MS tol. ± 0.6 Da
Peptide charge: 2+ and 3+
Monoisotopic ● Average ○
Data file: D:\iPRG2008\mgf\merged.mgf
Data format: Mascot generic
Precursor: m/z
Instrument: ESI-TRAP
Error tolerant ☐
Decoy ☑
Report top AUTO hits

To illustrate how a new search result is captured, I'm carrying out a search of the ABRF iPRG 2008 dataset on a Mascot server which is registered with a Mascot Insight server. I just submit the search as normal…



Field	Operator	Value
Search title	Contains	Mascot Insight: iPRG2

● Match all ○ Match any

Limit search to selected nodes? ☐

Include trash in searches? ☐

☐ Id	Type	External Id	Description	Notes
☑ mss-20130820-000001	Mascot Search Result	F001690	Mascot Insight: iPRG2008 SwissProt Mouse	

1 items found, displaying 1 to 1.
[First/Prev] **1** [Next/Last]

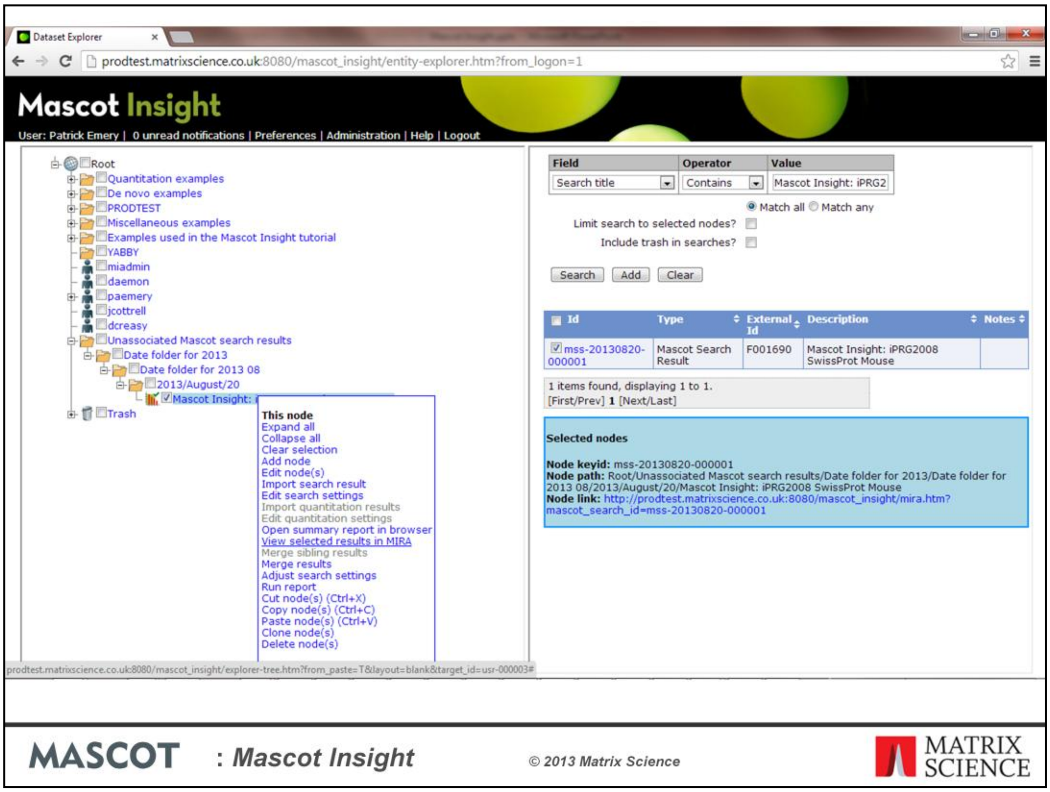
Selected nodes

Node keyid: mss-20130820-000001
Node path: Root/Unassociated Mascot search results/Date folder for 2013/Date folder for 2013 08/2013/August/20/Mascot Insight: iPRG2008 SwissProt Mouse
Node link: http://prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?mascot_search_id=mss-20130820-000001

MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

Shortly after the search has completed, the result will appear in Mascot Insight. In this case the result has been imported to the default location under the ‘Unassociated Mascot search results’ node in a date folder. I could move the result to a different location on the tree by either dragging and dropping it, or by using the cut and paste options from the dataset explorer tree’s context menu. A more efficient way to organise results is to set up filters to import results into a meaningful tree structure. These filters can automatically create new folders on the data explorer tree as required. We’ll take a look at how this works later in the presentation.

If you’re not sure where a search result has gone on the tree, then the tree is fully searchable using any available field – for example, on the search title



Field	Operator	Value
Search title	Contains	Mascot Insight: iPRG2

● Match all ○ Match any

Limit search to selected nodes? ☐

Include trash in searches? ☐

☐ Id	Type	External Id	Description	Notes
☑ mss-20130820-000001	Mascot Search Result	F001690	Mascot Insight: iPRG2008 SwissProt Mouse	

1 items found, displaying 1 to 1.
[First/Prev] **1** [Next/Last]

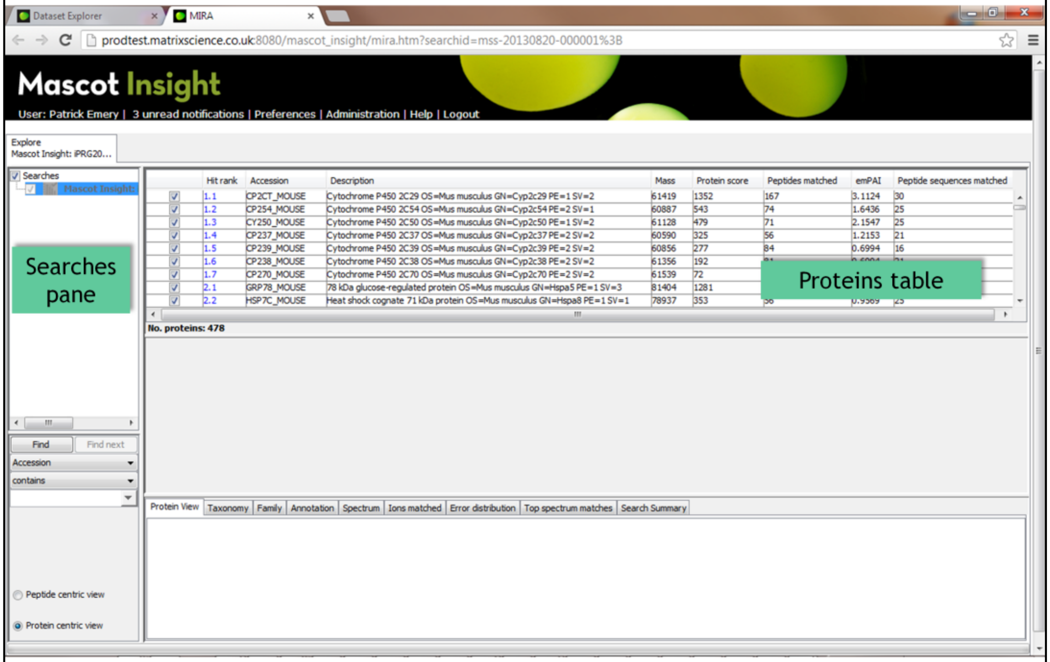
Selected nodes

Node keyid: mss-20130820-000001
Node path: Root/Unassociated Mascot search results/Date folder for 2013/Date folder for 2013 08/2013/August/20/Mascot Insight: iPRG2008 SwissProt Mouse
Node link: http://prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?mascot_search_id=mss-20130820-000001

: *Mascot Insight*

© 2013 Matrix Science

If I right click over a node, or a group of selected nodes, I get a pop up context menu that allows me to carry out a number of operations, including editing the node, editing default search settings, cut, copy, clone, and paste operations. For a search result, I also get the option of viewing the original summary report in the browser, or in MIRA – the Mascot Insight Result Applet

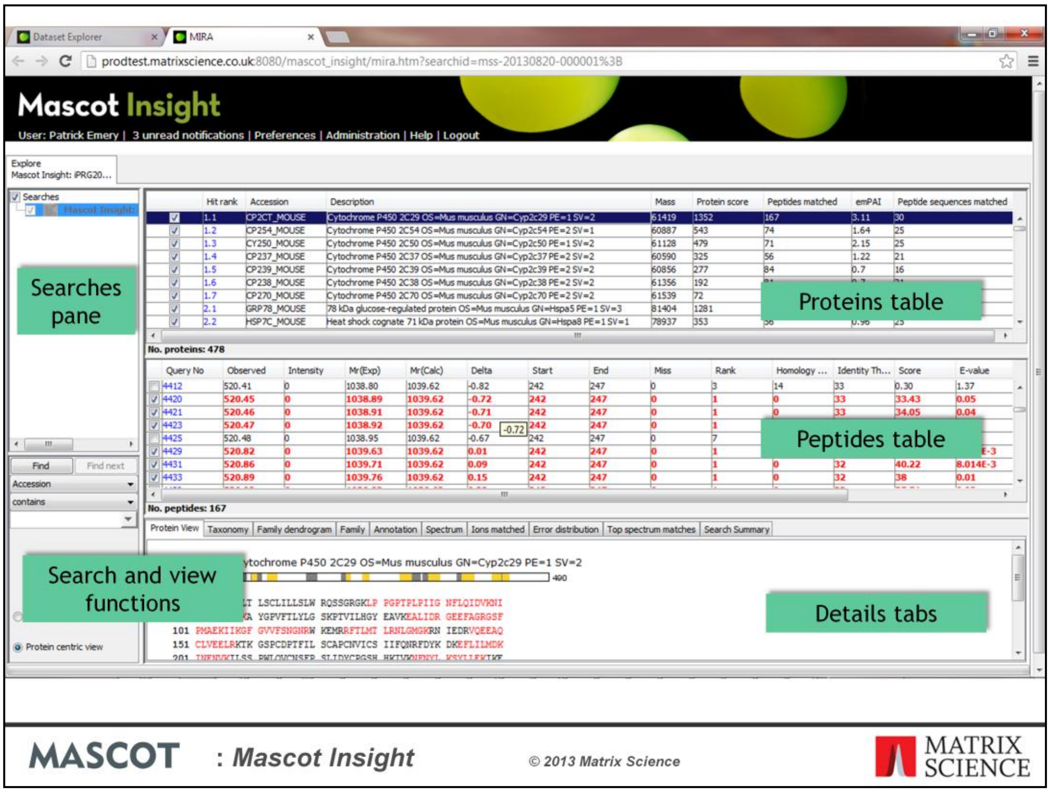


	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	1.1	CP2CT_MOUSE	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2	61419	1352	167	3.1124	30
☑	1.2	CP254_MOUSE	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1	60887	543	74	1.6436	25
☑	1.3	CY250_MOUSE	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=2	61128	479	71	2.1547	25
☑	1.4	CP237_MOUSE	Cytochrome P450 2C37 OS=Mus musculus GN=Cyp2c37 PE=2 SV=2	60590	325	56	1.2153	21
☑	1.5	CP239_MOUSE	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=2	60856	277	84	0.6994	16
☑	1.6	CP238_MOUSE	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=2	61356	192	[illegible]	[illegible]	[illegible]
☑	1.7	CP270_MOUSE	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2	61539	72	[illegible]	[illegible]	[illegible]
☑	2.1	GRP78_MOUSE	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	81404	1281	[illegible]	[illegible]	[illegible]
☑	2.2	HSP7C_MOUSE	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	78937	353	56	0.9569	25

No. proteins: 478

MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

MIRA provides a tabular view of a single result or merged set of results. You can view results in either the Protein centric mode (as we are doing here), or in a peptide centric view. On the upper left we have the searches pane, which is used to move between the selected results, changing which result is displayed in the tables. This pane is also used for running search level reports and for setting formatting parameters across multiple searches. The upper panel on the right contains the proteins table. This is the protein hit list for the currently selected result on the searches pane, and resembles the 'Report builder' view in the standard Mascot protein family report. If we click on a row in the Proteins table:



prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?searchid=mss-20130820-000001%3B

	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	1.1	CP2CT_MOUSE	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2	61419	1352	167	3.11	30
☑	1.2	CP254_MOUSE	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1	60887	543	74	1.64	25
☑	1.3	CY250_MOUSE	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=2	61128	479	71	2.15	25
☑	1.4	CP237_MOUSE	Cytochrome P450 2C37 OS=Mus musculus GN=Cyp2c37 PE=2 SV=2	60590	325	56	1.22	21
☑	1.5	CP239_MOUSE	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=2	60856	277	84	0.7	16
☑	1.6	CP238_MOUSE	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=2	61356	192			
☑	1.7	CP270_MOUSE	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2	61539	72			
☑	2.1	GRP78_MOUSE	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	81404	1281			
☑	2.2	HSP7C_MOUSE	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	78937	353			

No. proteins: 478

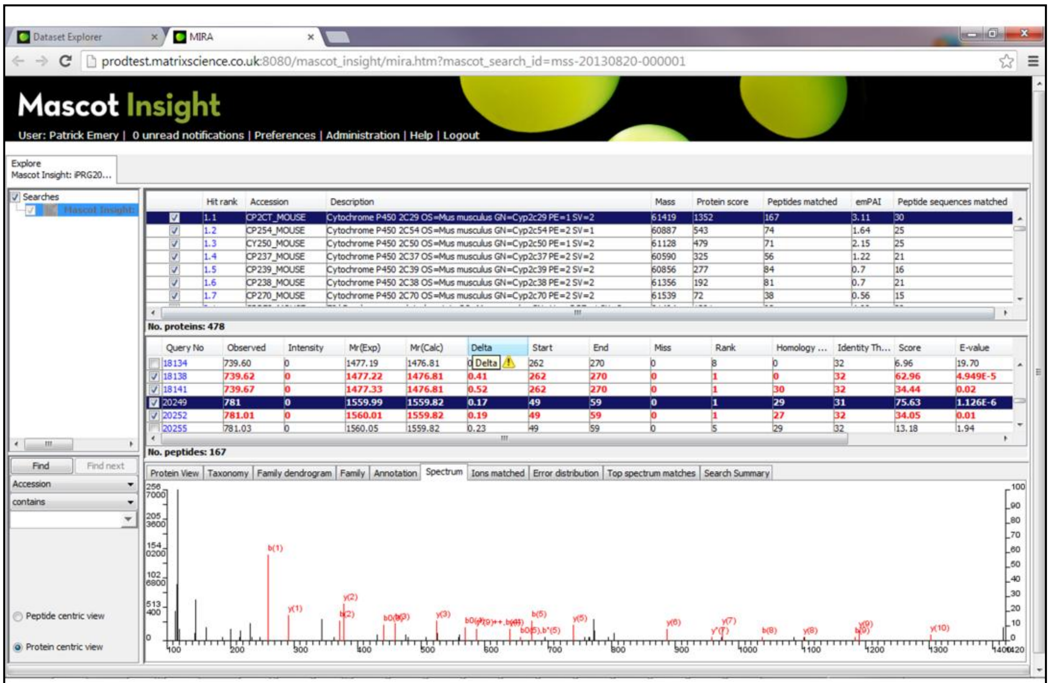
	Query No	Observed	Intensity	Mr(Exp)	Mr(Calc)	Delta	Start	End	Miss	Rank	Homology ...	Identity Th...	Score	E-value
☐	4412	520.41	0	1038.80	1039.62	-0.82	242	247	0	3	14	33	0.30	1.37
☑	4420	520.45	0	1038.89	1039.62	-0.72	242	247	0	1	0	33	33.43	0.05
☑	4421	520.46	0	1038.91	1039.62	-0.71	242	247	0	1	0	33	34.05	0.04
☑	4423	520.47	0	1038.92	1039.62	-0.70 -0.72	242	247	0	1				
☐	4425	520.48	0	1038.95	1039.62	-0.67	242	247	0	7				
☑	4429	520.82	0	1039.63	1039.62	0.01	242	247	0	1				E-3
☑	4431	520.86	0	1039.71	1039.62	0.09	242	247	0	1	0	32	40.22	8.014E-3
☑	4433	520.89	0	1039.76	1039.62	0.15	242	247	0	1	0	32	38	0.01

No. peptides: 167

ytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2

```
    LT LSCLILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
    KA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
```

The peptides table lists details of the peptide matches assigned to the selected protein, and a number of tabs are populated with information about the selected protein, such as coverage. Clicking on a column header in either table sorts on that column. You can re-order the columns by dragging them and show and hide columns as you wish. At the bottom left is a panel with text search functions, and options to switch between protein and peptide-centric views.



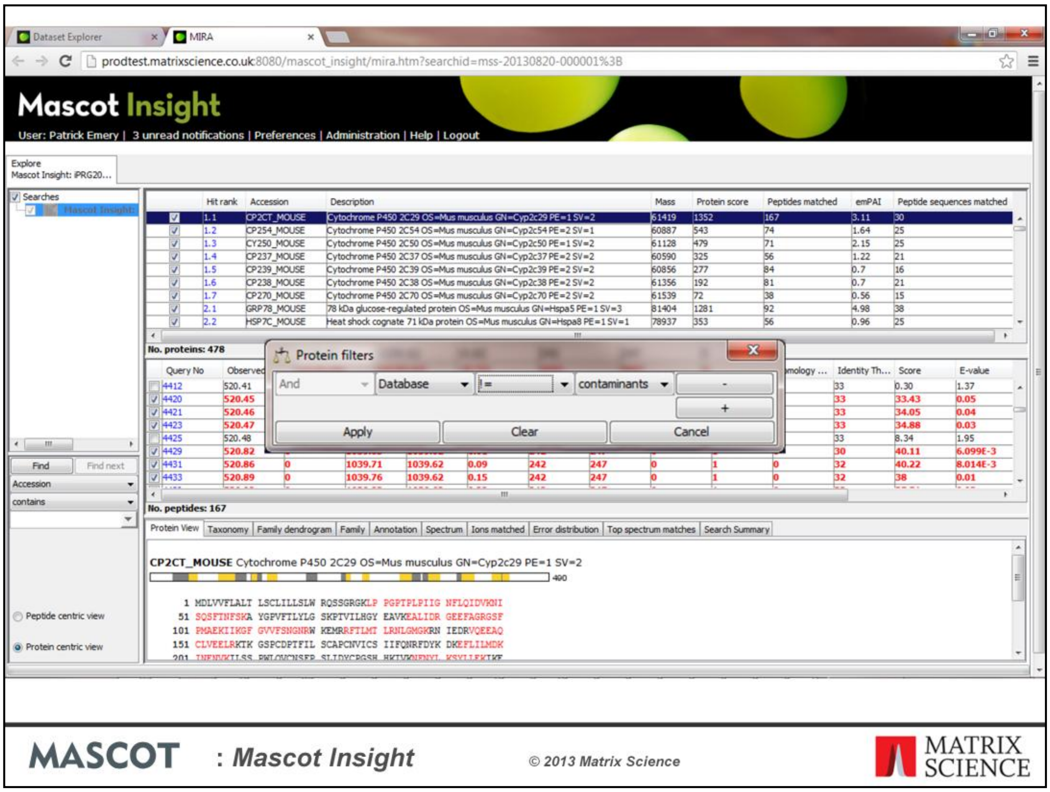
	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	1.1	CP2CT_MOUSE	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2	61419	1352	167	3.11	30
☑	1.2	CP254_MOUSE	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1	60887	543	74	1.64	25
☑	1.3	CY250_MOUSE	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=2	61128	479	71	2.15	25
☑	1.4	CP237_MOUSE	Cytochrome P450 2C37 OS=Mus musculus GN=Cyp2c37 PE=2 SV=2	60590	325	56	1.22	21
☑	1.5	CP239_MOUSE	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=2	60856	277	84	0.7	16
☑	1.6	CP238_MOUSE	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=2	61356	192	81	0.7	21
☑	1.7	CP270_MOUSE	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2	61539	72	38	0.56	15

No. proteins: 478

	Query No	Observed	Intensity	Mr(Exp)	Mr(Calc)	Delta	Start	End	Miss	Rank	Homology ...	Identity Th...	Score	E-value
☐	18134	739.60	0	1477.19	1476.81	0 Delta ⚠	262	270	0	8	0	32	6.96	19.70
☑	18138	**739.62**	**0**	**1477.22**	**1476.81**	**0.41**	**262**	**270**	**0**	**1**	**0**	**32**	**62.96**	**4.949E-5**
☑	18141	**739.67**	**0**	**1477.33**	**1476.81**	**0.52**	**262**	**270**	**0**	**1**	**30**	**32**	**34.44**	**0.02**
☑	20249	**781**	**0**	**1559.99**	**1559.82**	**0.17**	**49**	**59**	**0**	**1**	**29**	**31**	**75.63**	**1.126E-6**
☑	20252	**781.01**	**0**	**1560.01**	**1559.82**	**0.19**	**49**	**59**	**0**	**1**	**27**	**32**	**34.05**	**0.01**
☐	20255	781.03	0	1560.05	1559.82	0.23	49	59	0	5	29	32	13.18	1.94

No. peptides: 167

Clicking on a peptide row allows you to drill down and see peptide specific information, such as an annotated spectrum.



prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?searchid=mss-20130820-000001%3B

	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	1.1	CP2CT_MOUSE	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2	61419	1352	167	3.11	30
☑	1.2	CP254_MOUSE	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1	60887	543	74	1.64	25
☑	1.3	CY250_MOUSE	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=2	61128	479	71	2.15	25
☑	1.4	CP237_MOUSE	Cytochrome P450 2C37 OS=Mus musculus GN=Cyp2c37 PE=2 SV=2	60590	325	56	1.22	21
☑	1.5	CP239_MOUSE	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=2	60856	277	84	0.7	16
☑	1.6	CP238_MOUSE	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=2	61356	192	81	0.7	21
☑	1.7	CP270_MOUSE	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2	61539	72	38	0.56	15
☑	2.1	GRP78_MOUSE	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	81404	1281	92	4.98	38
☑	2.2	HSP7C_MOUSE	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	78937	353	56	0.96	25

No. proteins: 478

Protein filters: And | Database | != | contaminants

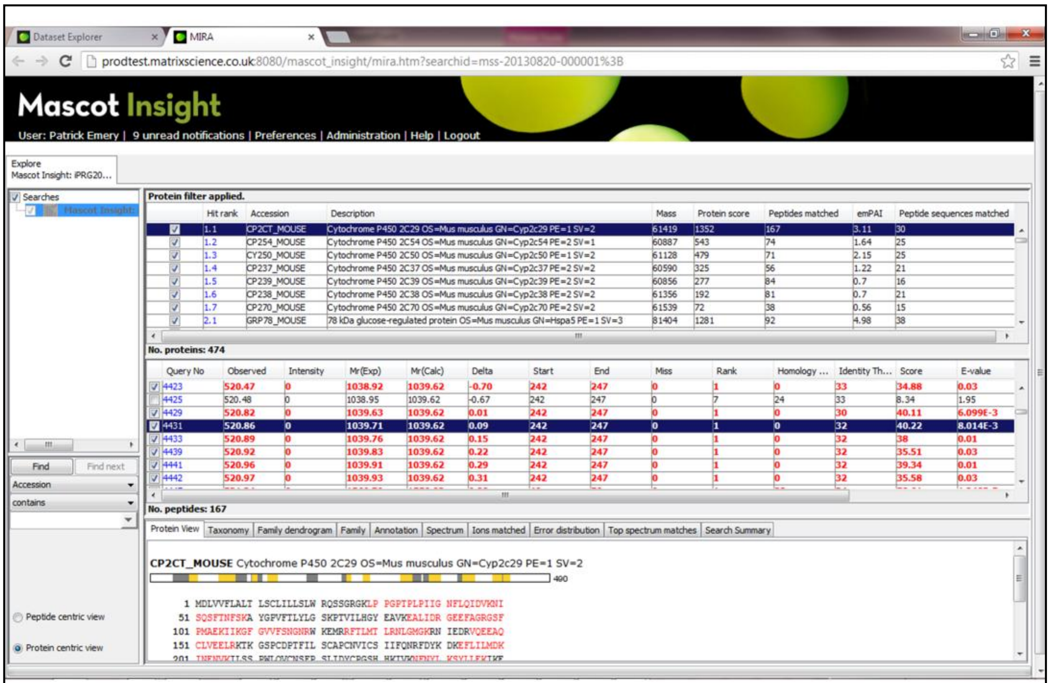
No. peptides: 167

CP2CT_MOUSE Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2

```
  1 MDLVVFLALT LSCLILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
 51 SQSFTNFSKA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
```

MASCOT : *Mascot Insight* © 2013 Matrix Science

You can easily apply protein and peptide level filters in MIRA. Here, we are going to filter out any matches to the contaminants database. Set up the filter and click 'Apply'



Protein filter applied.

	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	1.1	CP2CT_MOUSE	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2	61419	1352	167	3.11	30
☑	1.2	CP254_MOUSE	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1	60887	543	74	1.64	25
☑	1.3	CY250_MOUSE	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=2	61128	479	71	2.15	25
☑	1.4	CP237_MOUSE	Cytochrome P450 2C37 OS=Mus musculus GN=Cyp2c37 PE=2 SV=2	60590	325	56	1.22	21
☑	1.5	CP239_MOUSE	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=2	60856	277	84	0.7	16
☑	1.6	CP238_MOUSE	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=2	61356	192	81	0.7	21
☑	1.7	CP270_MOUSE	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2	61539	72	38	0.56	15
☑	2.1	GRP78_MOUSE	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	81404	1281	92	4.98	38

No. proteins: 474

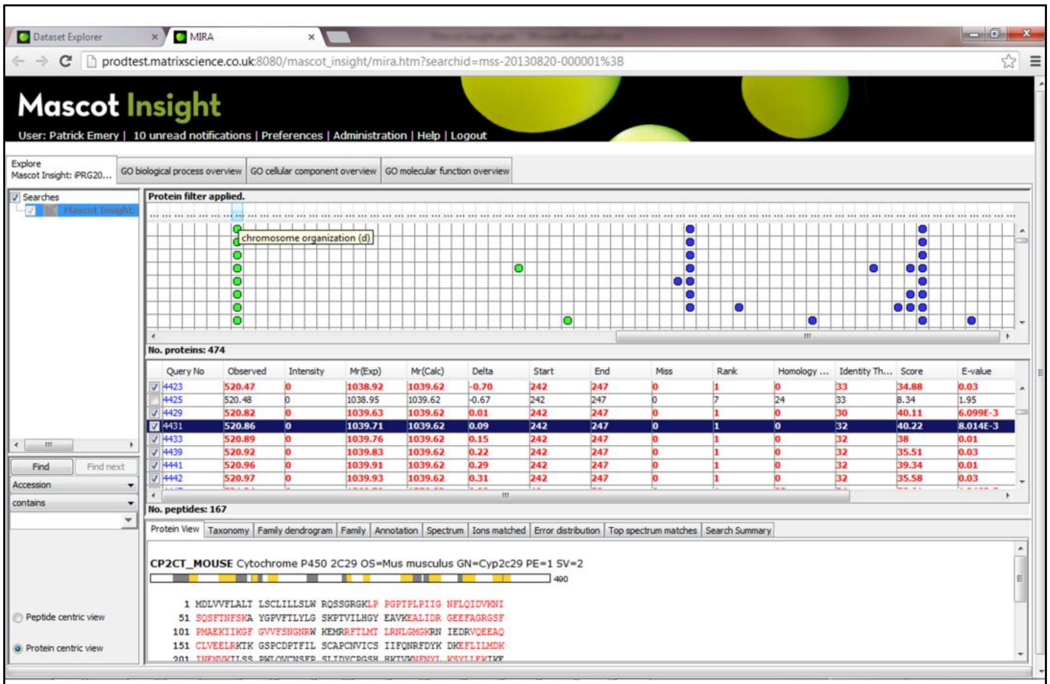
	Query No	Observed	Intensity	Mr(Exp)	Mr(Calc)	Delta	Start	End	Miss	Rank	Homology ...	Identity Th...	Score	E-value
☑	4423	520.47	0	1038.92	1039.62	-0.70	242	247	0	1	0	33	34.88	0.03
☐	4425	520.48	0	1038.95	1039.62	-0.67	242	247	0	7	24	33	8.34	1.95
☑	4429	520.82	0	1039.63	1039.62	0.01	242	247	0	1	0	30	40.11	6.099E-3
☑	4431	520.86	0	1039.71	1039.62	0.09	242	247	0	1	0	32	40.22	8.014E-3
☑	4433	520.89	0	1039.76	1039.62	0.15	242	247	0	1	0	32	38	0.01
☑	4439	520.92	0	1039.83	1039.62	0.22	242	247	0	1	0	32	35.51	0.03
☑	4441	520.96	0	1039.91	1039.62	0.29	242	247	0	1	0	32	39.34	0.01
☑	4442	520.97	0	1039.93	1039.62	0.31	242	247	0	1	0	32	35.58	0.03

No. peptides: 167

CP2CT_MOUSE Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2

```
  1 MDLVVFLALT LSCLILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
 51 SQSFTNFSKA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
```

And we’ve now hidden the 4 matches to the contaminants database for this search result.



No. proteins: 474

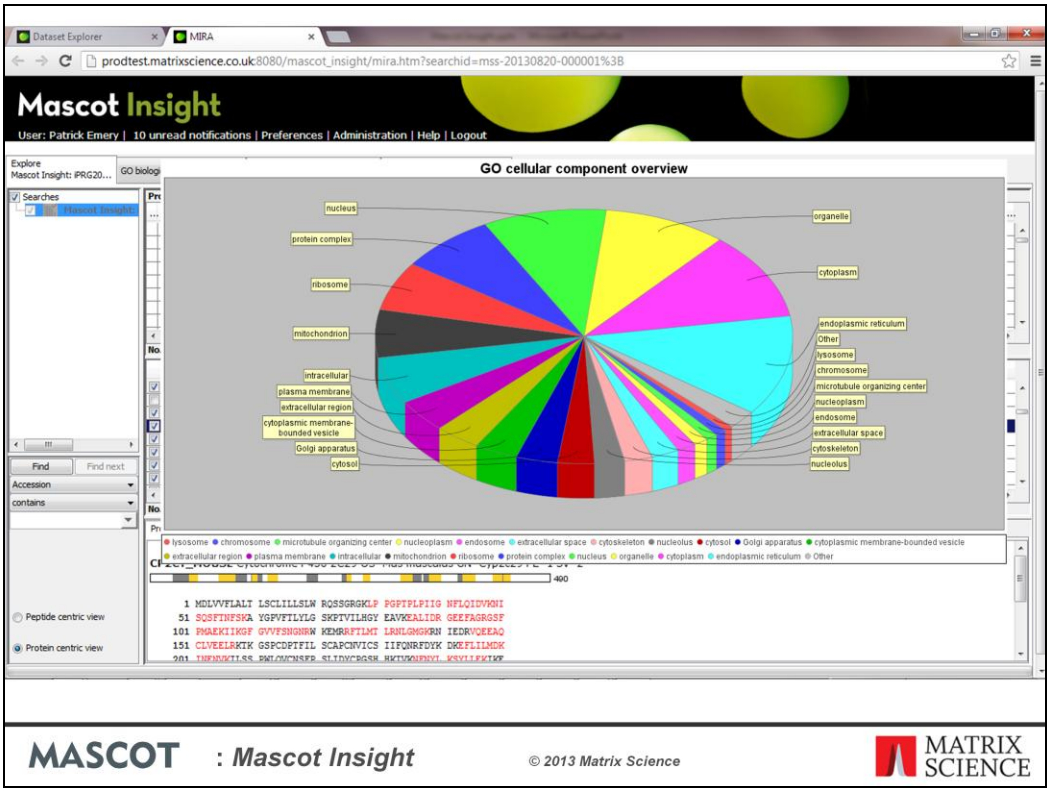
	Query No	Observed	Intensity	Mr(Exp)	Mr(Calc)	Delta	Start	End	Miss	Rank	Homology …	Identity Th…	Score	E-value
☑	4423	520.47	0	1038.92	1039.62	-0.70	242	247	0	1	0	33	34.88	0.03
☐	4425	520.48	0	1038.95	1039.62	-0.67	242	247	0	7	24	33	8.34	1.95
☑	4429	520.82	0	1039.63	1039.62	0.01	242	247	0	1	0	30	40.11	6.099E-3
☑	4431	520.86	0	1039.71	1039.62	0.09	242	247	0	1	0	32	40.22	8.014E-3
☑	4433	520.89	0	1039.76	1039.62	0.15	242	247	0	1	0	32	38	0.01
☑	4439	520.92	0	1039.83	1039.62	0.22	242	247	0	1	0	32	35.51	0.03
☑	4441	520.96	0	1039.91	1039.62	0.29	242	247	0	1	0	32	39.34	0.01
☑	4442	520.97	0	1039.93	1039.62	0.31	242	247	0	1	0	32	35.58	0.03

No. peptides: 167

CP2CT_MOUSE Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=2

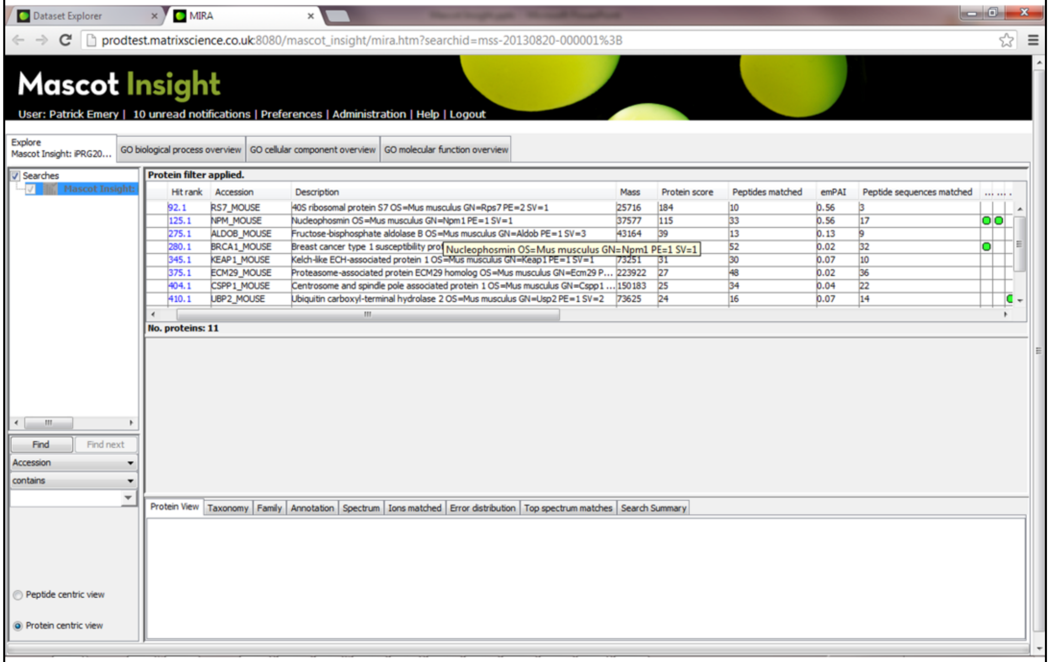
```
  1 MDLVVFLALT LSCLILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
 51 SQSFTNFSKA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
```

You can enable Gene Ontology (GO) Analysis in MIRA. GO annotations are shown in the proteins table, as different coloured dots. Hover the mouse over a cell to display category information in a tooltip.



```
  1 MDLVVFLALT LSCLILLSLW RQSSGRGKLP PGPTPLPIIG NFLQIDVKNI
 51 SQSFTNFSKA YGPVFTLYLG SKPTVILHGY EAVKEALIDR GEEFAGRGSF
101 PMAEKIIKGF GVVFSNGNRW KEMRRFTLMT LRNLGMGKRN IEDRVQEEAQ
151 CLVEELRKTK GSPCDPTFIL SCAPCNVICS IIFQNRFDYK DKEFLILMDK
```

GO assignments are also presented in additional tabs as pie-charts. Click on a wedge of the pie-chart, and the protein table is filtered on that GO assignment



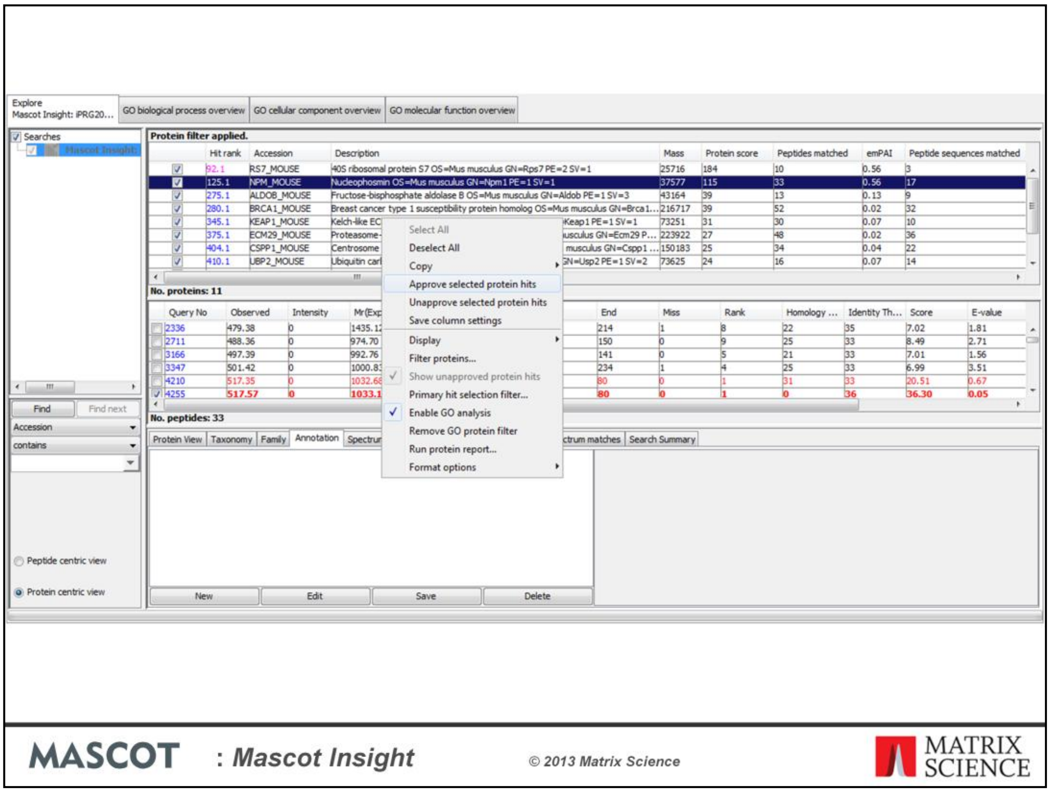
Protein filter applied.

Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
92.1	RS7_MOUSE	40S ribosomal protein S7 OS=Mus musculus GN=Rps7 PE=2 SV=1	25716	184	10	0.56	3
125.1	NPM_MOUSE	Nucleophosmin OS=Mus musculus GN=Npm1 PE=1 SV=1	37577	115	33	0.56	17
275.1	ALDOB_MOUSE	Fructose-bisphosphate aldolase B OS=Mus musculus GN=Aldob PE=1 SV=3	43164	39	13	0.13	9
280.1	BRCA1_MOUSE	Breast cancer type 1 susceptibility pro			52	0.02	32
345.1	KEAP1_MOUSE	Kelch-like ECH-associated protein 1 OS=Mus musculus GN=Keap1 PE=1 SV=1	73251	31	30	0.07	10
375.1	ECM29_MOUSE	Proteasome-associated protein ECM29 homolog OS=Mus musculus GN=Ecm29 P…	223922	27	48	0.02	36
404.1	CSPP1_MOUSE	Centrosome and spindle pole associated protein 1 OS=Mus musculus GN=Cspp1 …	150183	25	34	0.04	22
410.1	UBP2_MOUSE	Ubiquitin carboxyl-terminal hydrolase 2 OS=Mus musculus GN=Usp2 PE=1 SV=2	73625	24	16	0.07	14

Nucleophosmin OS=Mus musculus GN=Npm1 PE=1 SV=1

No. proteins: 11

Now we are just looking at the 11 protein matches with the GO cellular component assignment of ‘Micro Tubule Organising Centre’ (MTOC)



Protein filter applied.

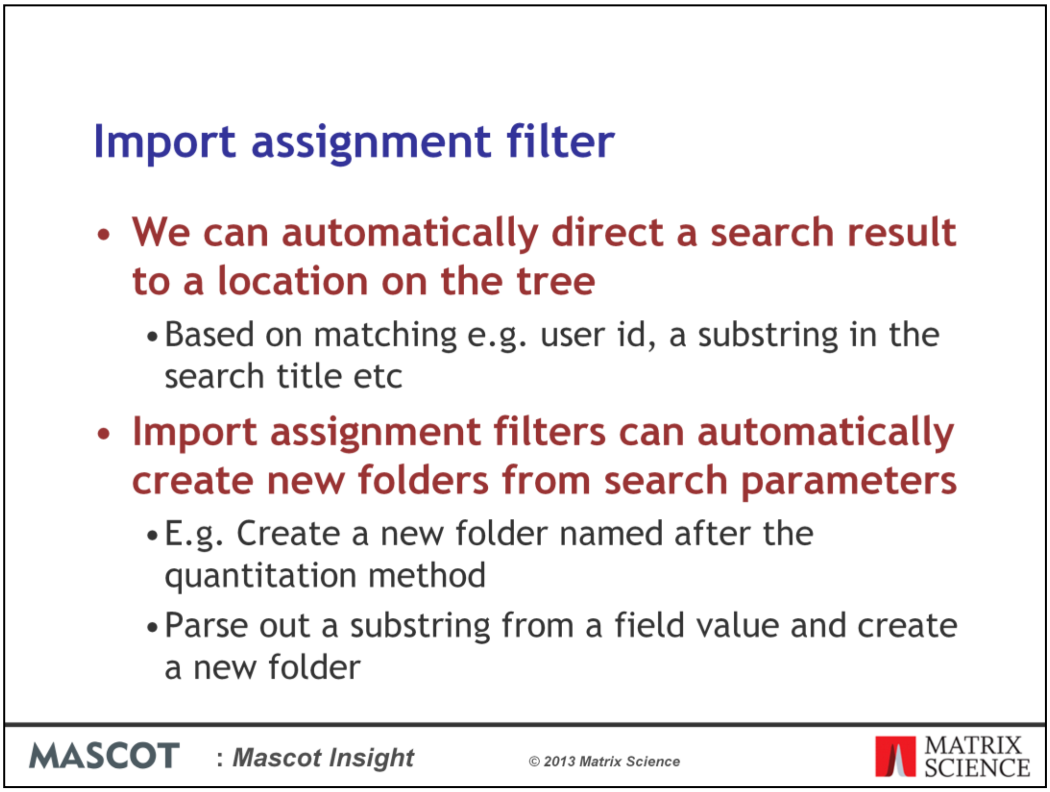
	Hit rank	Accession	Description	Mass	Protein score	Peptides matched	emPAI	Peptide sequences matched
☑	92.1	RS7_MOUSE	40S ribosomal protein S7 OS=Mus musculus GN=Rps7 PE=2 SV=1	25716	184	10	0.56	3
☑	125.1	NPM_MOUSE	Nucleophosmin OS=Mus musculus GN=Npm1 PE=1 SV=1	37577	115	33	0.56	17
☑	275.1	ALDOB_MOUSE	Fructose-bisphosphate aldolase B OS=Mus musculus GN=Aldob PE=1 SV=3	43164	39	13	0.13	9
☑	280.1	BRCA1_MOUSE	Breast cancer type 1 susceptibility protein homolog OS=Mus musculus GN=Brca1...	216717	39	52	0.02	32
☑	345.1	KEAP1_MOUSE	Kelch-like EC Keap1 PE=1 SV=1	73251	31	30	0.07	10
☑	375.1	ECM29_MOUSE	Proteasome- usculus GN=Ecm29 P...	223922	27	48	0.02	36
☑	404.1	CSPP1_MOUSE	Centrosome musculus GN=Cspp1 ...	150183	25	34	0.04	22
☑	410.1	UBP2_MOUSE	Ubiquitin carl GN=Usp2 PE=1 SV=2	73625	24	16	0.07	14

No. proteins: 11

	Query No	Observed	Intensity	Mr(Exp	End	Miss	Rank	Homology ...	Identity Th...	Score	E-value
☐	2336	479.38	0	1435.1	214	1	8	22	35	7.02	1.81
☐	2711	488.36	0	974.70	150	0	9	25	33	8.49	2.71
☐	3166	497.39	0	992.76	141	0	5	21	33	7.01	1.56
☐	3347	501.42	0	1000.8	234	1	4	25	33	6.99	3.51
☐	4210	517.35	0	1032.68	80	0	1	31	33	20.51	0.67
☑	4255	**517.57**	**0**	**1033.1**	**80**	**0**	**1**	**0**	**36**	**36.30**	**0.05**

No. peptides: 33

Once you’re happy with a particular set of protein matches, you can approve the protein hits and make manual annotations if you wish. Approving a protein hit flags the protein in the database. You can use this to highlight and filter out just those protein hits that are of interest.



Import assignment filter

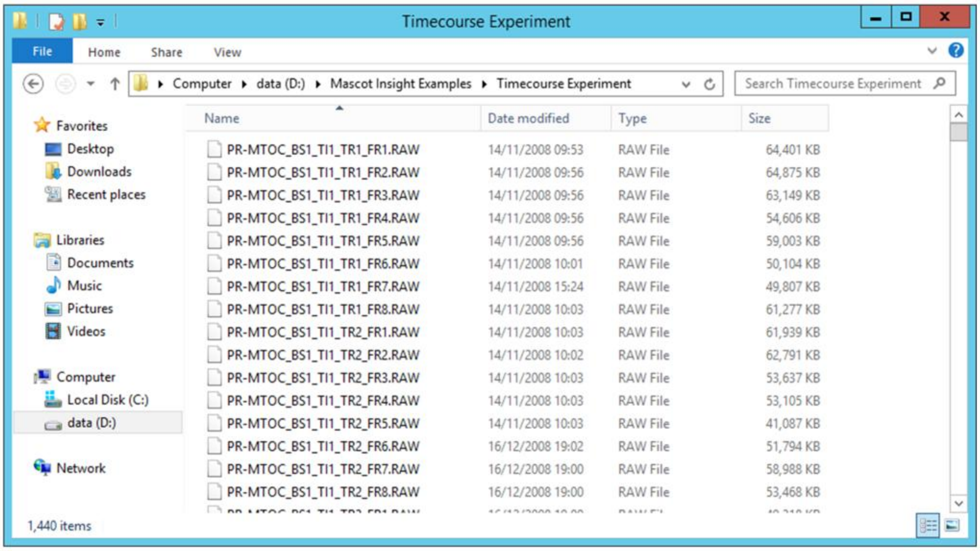
- **We can automatically direct a search result to a location on the tree**
  - Based on matching e.g. user id, a substring in the search title etc
- **Import assignment filters can automatically create new folders from search parameters**
  - E.g. Create a new folder named after the quantitation method
  - Parse out a substring from a field value and create a new folder

MASCOT : *Mascot Insight*

MATRIX SCIENCE

Earlier, I showed an example of a search result being imported into the default location in Dataset Explorer. However, you can override this using an Import assignment filter. This is a filter to direct a search result to a specified node based on just about any property, such as user name or file name. Import assignment filters can automatically create new folders on the tree based on the values of search parameters, or by parsing out a string from a field.

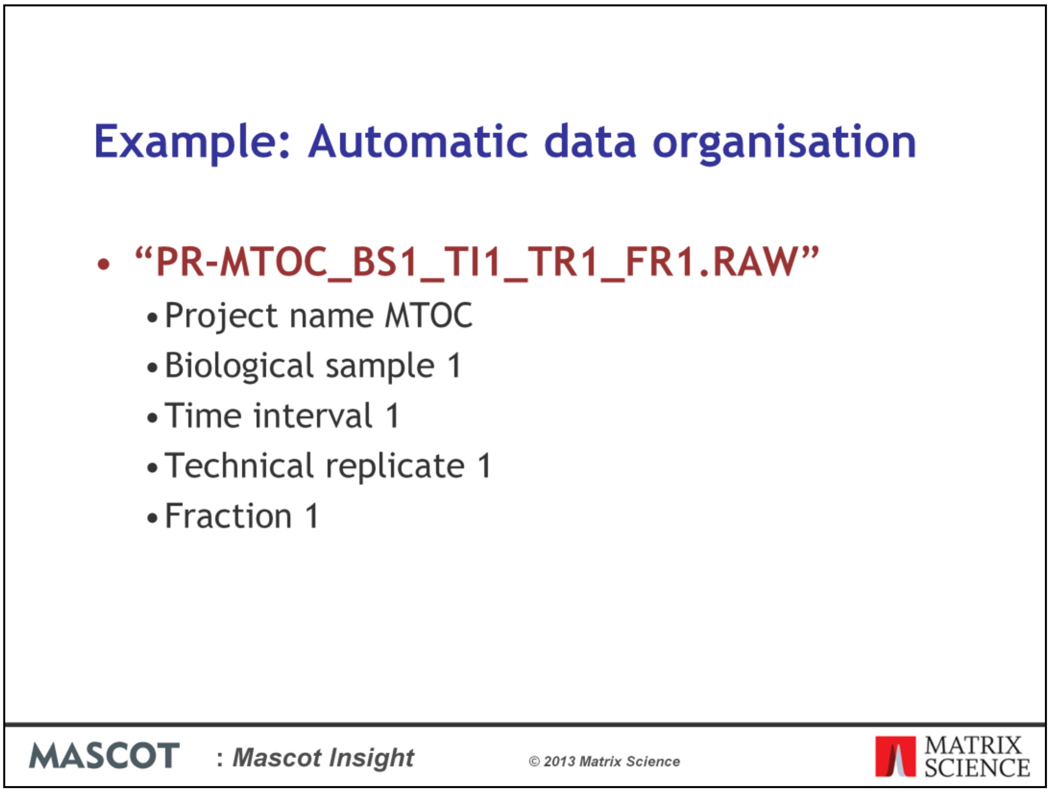
Example: Automatic data organisation



Name	Date modified	Type	Size
PR-MTOC_BS1_TI1_TR1_FR1.RAW	14/11/2008 09:53	RAW File	64,401 KB
PR-MTOC_BS1_TI1_TR1_FR2.RAW	14/11/2008 09:56	RAW File	64,875 KB
PR-MTOC_BS1_TI1_TR1_FR3.RAW	14/11/2008 09:56	RAW File	63,149 KB
PR-MTOC_BS1_TI1_TR1_FR4.RAW	14/11/2008 09:56	RAW File	54,606 KB
PR-MTOC_BS1_TI1_TR1_FR5.RAW	14/11/2008 09:56	RAW File	59,003 KB
PR-MTOC_BS1_TI1_TR1_FR6.RAW	14/11/2008 10:01	RAW File	50,104 KB
PR-MTOC_BS1_TI1_TR1_FR7.RAW	14/11/2008 15:24	RAW File	49,807 KB
PR-MTOC_BS1_TI1_TR1_FR8.RAW	14/11/2008 10:03	RAW File	61,277 KB
PR-MTOC_BS1_TI1_TR2_FR1.RAW	14/11/2008 10:03	RAW File	61,939 KB
PR-MTOC_BS1_TI1_TR2_FR2.RAW	14/11/2008 10:02	RAW File	62,791 KB
PR-MTOC_BS1_TI1_TR2_FR3.RAW	14/11/2008 10:03	RAW File	53,637 KB
PR-MTOC_BS1_TI1_TR2_FR4.RAW	14/11/2008 10:03	RAW File	53,105 KB
PR-MTOC_BS1_TI1_TR2_FR5.RAW	14/11/2008 10:03	RAW File	41,087 KB
PR-MTOC_BS1_TI1_TR2_FR6.RAW	16/12/2008 19:02	RAW File	51,794 KB
PR-MTOC_BS1_TI1_TR2_FR7.RAW	16/12/2008 19:00	RAW File	58,988 KB
PR-MTOC_BS1_TI1_TR2_FR8.RAW	16/12/2008 19:00	RAW File	53,468 KB

Lets take a look at an example of how you might set up and use an Import assignment filter. You can make the Data Explorer tree as simple or as complicated as you wish, but the idea is to organise results logically so as to facilitate merging and reporting. For example, consider a quantitation experiment where 5 biological samples are analysed at 12 time intervals and each analysis is performed in triplicate (to give technical replicates). Each analysis involves 2D chromatography, requiring the mass spectrometry analysis of 8 fractions. This one experiment gives rise to 1440 Mascot search results. By grouping these results into a logical hierarchy of folders, we can minimise the work required to merge results and create meaningful reports.

In this case, the RAW data files are all in a single folder as shown, but the data file names contain structured information that can be used to generate a meaningful tree structure.



Example: Automatic data organisation

- “PR-MTOC_BS1_TI1_TR1_FR1.RAW”
  - Project name MTOC
  - Biological sample 1
  - Time interval 1
  - Technical replicate 1
  - Fraction 1

MASCOT : *Mascot Insight*

MATRIX SCIENCE

So, here we have the filename of the first raw data file. The name structure of the raw data file has been logically put together and breaks down like this:

PR – Project/experiment name

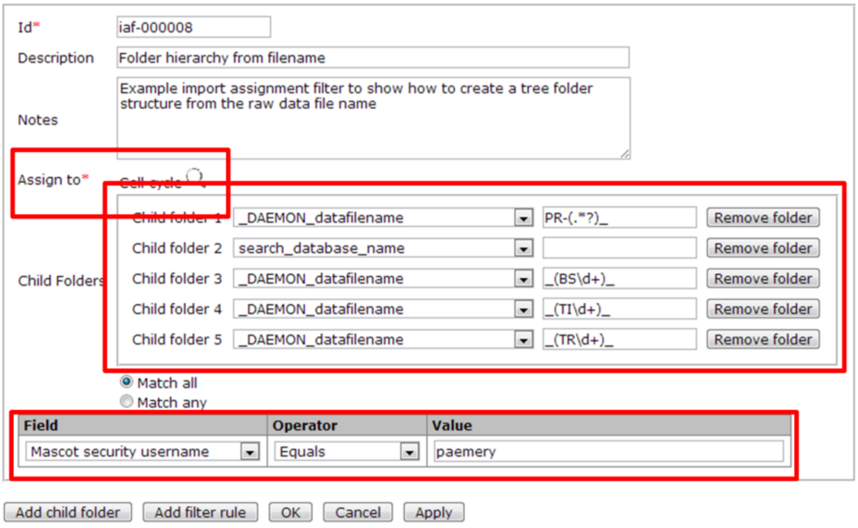
BS – Biological sample number

TI – Time interval

TR – Technical replicate number

FR – Fraction number

Example: Automatic data organisation



Id* iaf-000008

Description Folder hierarchy from filename

Notes Example import assignment filter to show how to create a tree folder structure from the raw data file name

Assign to* Cell-cycle

Child Folders

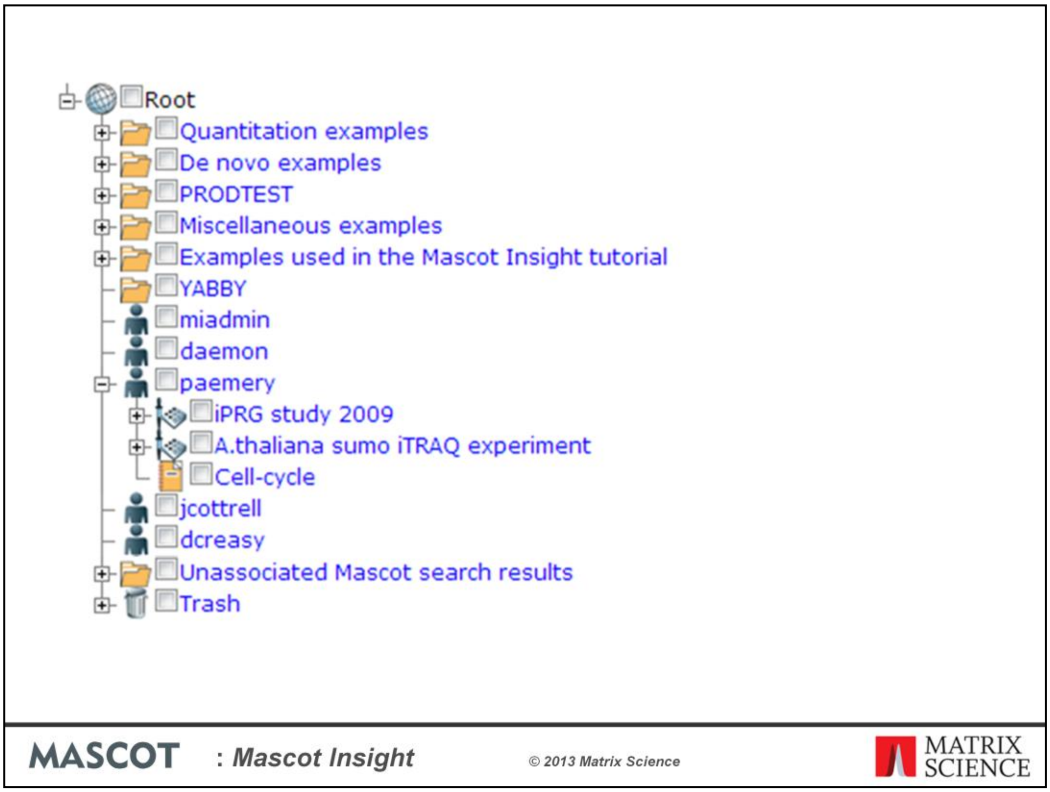
Child folder 1	_DAEMON_datafilename	PR-(.*?)_	Remove folder
Child folder 2	search_database_name		Remove folder
Child folder 3	_DAEMON_datafilename	_(BS\d+)_	Remove folder
Child folder 4	_DAEMON_datafilename	_(TI\d+)_	Remove folder
Child folder 5	_DAEMON_datafilename	_(TR\d+)_	Remove folder

● Match all
○ Match any

Field	Operator	Value
Mascot security username	Equals	paemery

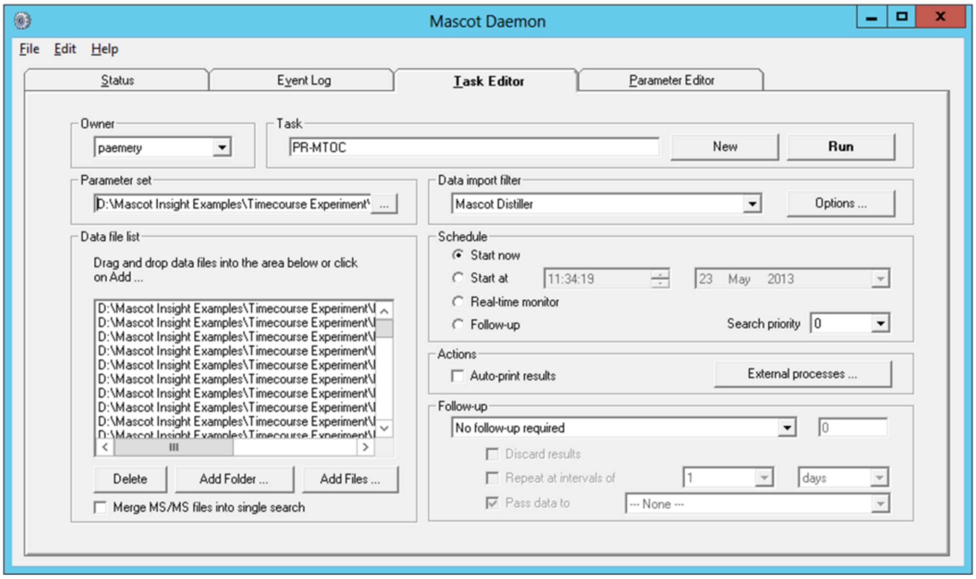
Now we set up the import assignment filter. First, we specify that the search must have been run using my Mascot security username, so that this filter doesn't accidentally grab somebody else's results. We have to select a parent node on the tree under which new folders and results will be added. In this case, I've set up an project node called 'Cell-cycle'. Now we want the system to automatically create folders and group the fractions for each technical replicate at each time point. We can pull the required pieces of information out of the raw data filename using regular expressions, as shown above. So, under the Cell-cycle project node, we'll generate a folder named after the experiment name. Then, a folder named after the sequence database that was searched, followed by a sub-folder for the biological sample, then the time interval and finally the technical replicate.

Although in this example, the required data for the import filter is contained in the raw data filename, it could be extracted from any other available field. For example, if the raw files were grouped together in a meaningful way on your hard drive, we could easily mimic this folder structure in the Data Explorer tree.



☐Root
- ☐Quantitation examples
- ☐De novo examples
- ☐PRODTEST
- ☐Miscellaneous examples
- ☐Examples used in the Mascot Insight tutorial
- ☐YABBY
- ☐miadmin
- ☐daemon
- ☐paemery
  - ☐iPRG study 2009
  - ☐A.thaliana sumo iTRAQ experiment
  - ☐Cell-cycle
- ☐jcottrell
- ☐dcreasy
- ☐Unassociated Mascot search results
- ☐Trash

Before we run the searches, this is what the explorer tree in Mascot Insight looks like



Owner: paemery

Task: PR-MTOC

Parameter set: D:\Mascot Insight Examples\Timecourse Experiment\

Data import filter: Mascot Distiller

Data file list

Drag and drop data files into the area below or click on Add ...

D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\
D:\Mascot Insight Examples\Timecourse Experiment\

☐ Merge MS/MS files into single search

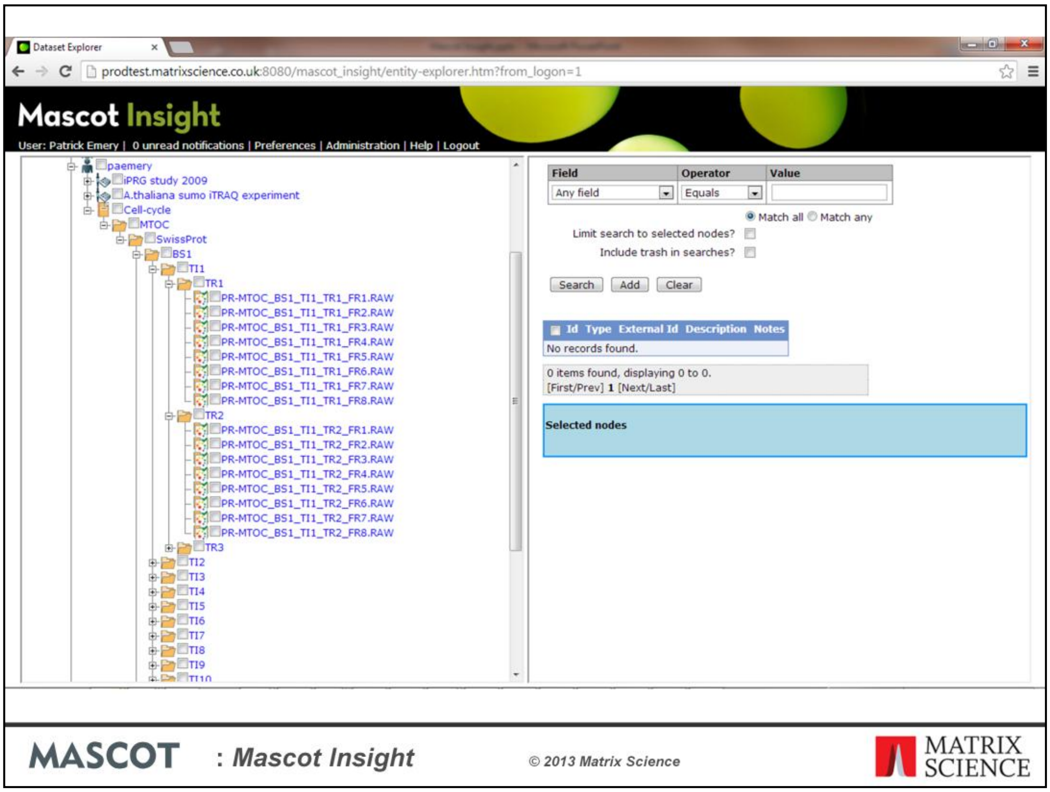
Schedule
● Start now
○ Start at 11:34:19 23 May 2013
○ Real-time monitor
○ Follow-up
Search priority 0

Actions
☐ Auto-print results

Follow-up
No follow-up required 0
☐ Discard results
☐ Repeat at intervals of 1 days
☑ Pass data to --- None ---

MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

We set up the task in Mascot Daemon as normal, making sure we choose Mascot Distiller as the import filter and, since this is an N15 quantitation dataset, we’ll set it to automatically carry out quantitation using Mascot Distiller as well. When a search is completed, the result will automatically be imported into Mascot Insight, and the Import assignment filter we set up will direct it to the desired location on the explorer tree. When Daemon completes quantitation for a file, the quantitation results are also automatically uploaded and imported.



- paemery
  - iPRG study 2009
  - A.thaliana sumo iTRAQ experiment
  - Cell-cycle
    - MTOC
      - SwissProt
        - BS1
          - TI1
            - TR1
              - PR-MTOC_BS1_TI1_TR1_FR1.RAW
              - PR-MTOC_BS1_TI1_TR1_FR2.RAW
              - PR-MTOC_BS1_TI1_TR1_FR3.RAW
              - PR-MTOC_BS1_TI1_TR1_FR4.RAW
              - PR-MTOC_BS1_TI1_TR1_FR5.RAW
              - PR-MTOC_BS1_TI1_TR1_FR6.RAW
              - PR-MTOC_BS1_TI1_TR1_FR7.RAW
              - PR-MTOC_BS1_TI1_TR1_FR8.RAW
            - TR2
              - PR-MTOC_BS1_TI1_TR2_FR1.RAW
              - PR-MTOC_BS1_TI1_TR2_FR2.RAW
              - PR-MTOC_BS1_TI1_TR2_FR3.RAW
              - PR-MTOC_BS1_TI1_TR2_FR4.RAW
              - PR-MTOC_BS1_TI1_TR2_FR5.RAW
              - PR-MTOC_BS1_TI1_TR2_FR6.RAW
              - PR-MTOC_BS1_TI1_TR2_FR7.RAW
              - PR-MTOC_BS1_TI1_TR2_FR8.RAW
            - TR3
          - TI2
          - TI3
          - TI4
          - TI5
          - TI6
          - TI7
          - TI8
          - TI9
          - TI10

Field	Operator	Value
Any field	Equals	

● Match all ○ Match any

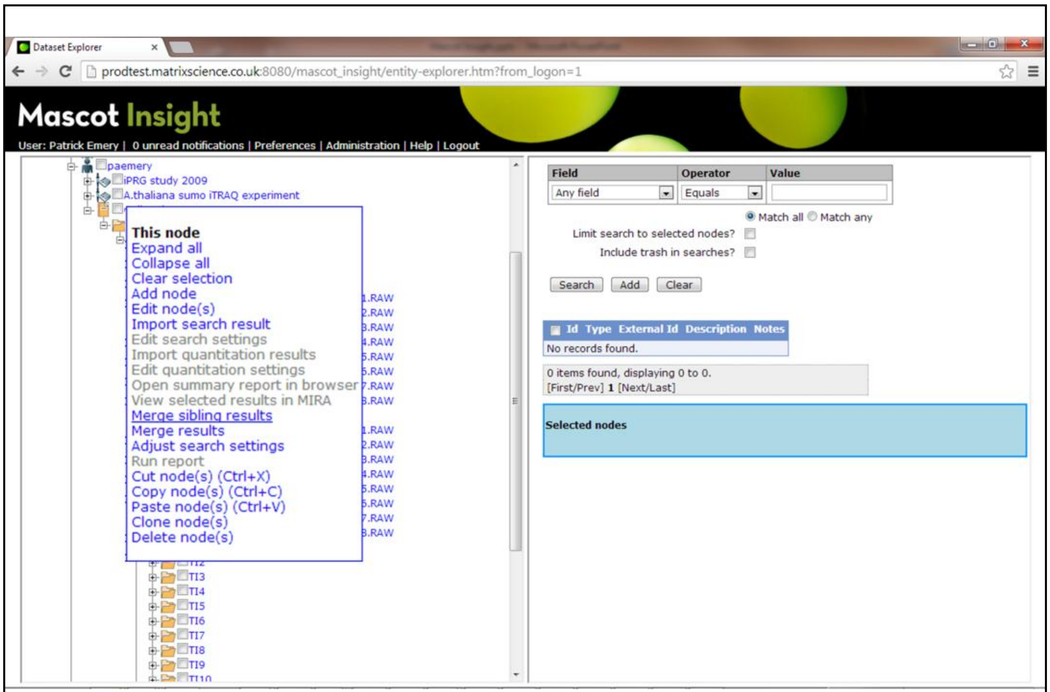
Limit search to selected nodes? ☐

Include trash in searches? ☐

☐	Id	Type	External Id	Description	Notes
No records found.					

0 items found, displaying 0 to 0.
[First/Prev] 1 [Next/Last]

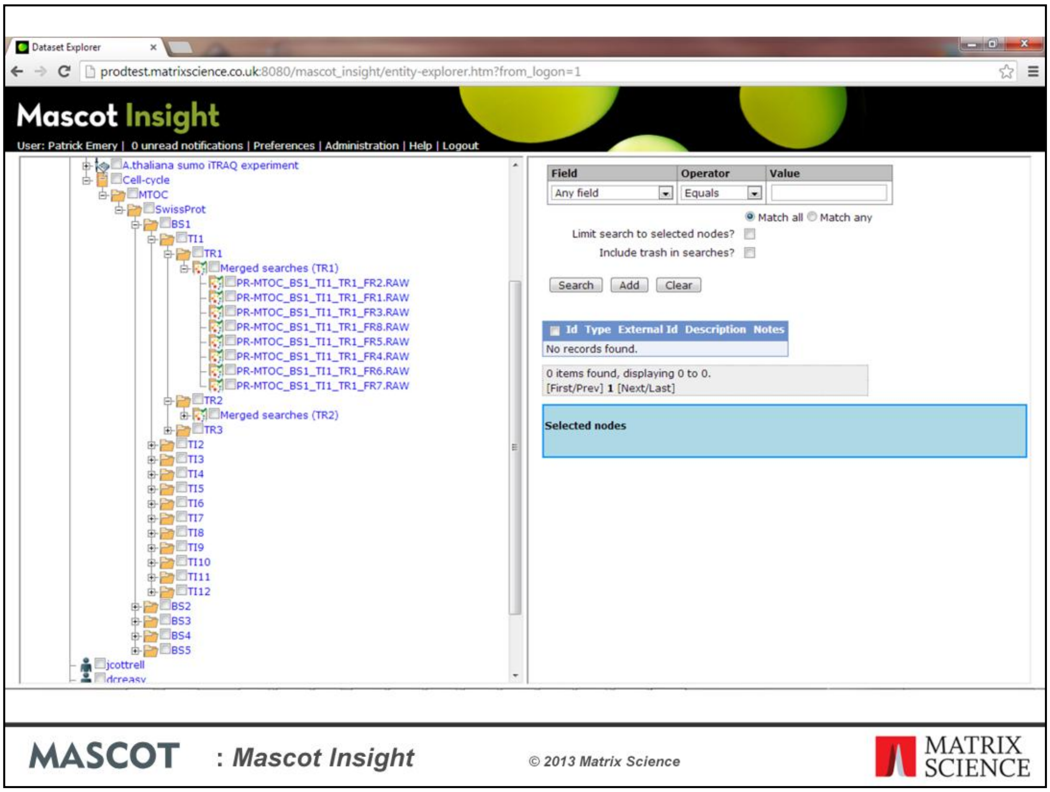
When the searches have completed, the Explorer tree looks like this. A total of 182 folders automatically created under the ‘Cell-cycle’ project node, and under each of those folders the eight fractions for each technical replicate at each time interval, all without any manual intervention after the import assignment filter had been set up.



MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

Now we need to merge the results for each of those groups of eight fractions. Because the results are grouped together on the tree in the way we want to merge them, we just need to right click over the 'Cell-cycle' project node (the parent node for all the results) and select the 'Merge sibling results' option from the context menu.

The system will then run through the dataset, merging together search results at the same level of the tree so, all of the fractions for each technical replicate will be merged together. More advanced, manual merge options are available if you need to do some additional processing when carrying out the merge – for example, if you have rotated the samples around iTRAQ report ions channels between replicates using a Latin square approach



prodtest.matrixscience.co.uk:8080/mascot_insight/entity-explorer.htm?from_logon=1

Mascot Insight

User: Patrick Emery | 0 unread notifications | Preferences | Administration | Help | Logout

- A.thaliana sumo iTRAQ experiment
- Cell-cycle
  - MTOC
    - SwissProt
      - BS1
        - TI1
          - TR1
            - Merged searches (TR1)
              - PR-MTOC_BS1_TI1_TR1_FR2.RAW
              - PR-MTOC_BS1_TI1_TR1_FR1.RAW
              - PR-MTOC_BS1_TI1_TR1_FR3.RAW
              - PR-MTOC_BS1_TI1_TR1_FR8.RAW
              - PR-MTOC_BS1_TI1_TR1_FR5.RAW
              - PR-MTOC_BS1_TI1_TR1_FR4.RAW
              - PR-MTOC_BS1_TI1_TR1_FR6.RAW
              - PR-MTOC_BS1_TI1_TR1_FR7.RAW
          - TR2
            - Merged searches (TR2)
          - TR3
        - TI2
        - TI3
        - TI4
        - TI5
        - TI6
        - TI7
        - TI8
        - TI9
        - TI10
        - TI11
        - TI12
      - BS2
      - BS3
      - BS4
      - BS5
- jcottrell
- dcreasy

Field	Operator	Value
Any field	Equals	

● Match all ○ Match any

Limit search to selected nodes? ☐

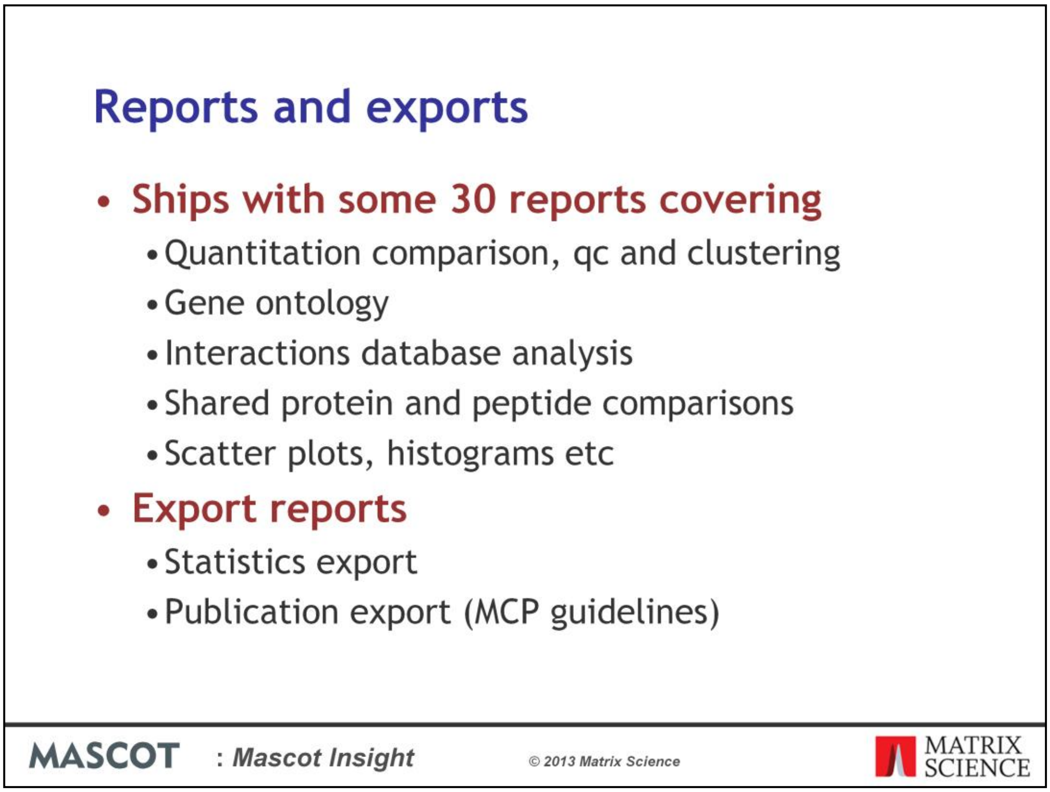
Include trash in searches? ☐

No records found.

0 items found, displaying 0 to 0.
[First/Prev] **1** [Next/Last]

When the merge is completed, for each repliacte, the results for individual fractions become children of a new ‘Merged searches’ result node. The merged search nodes can be treated in every way like any ‘normal’ search result node; viewed in MIRA or used to generate reports.

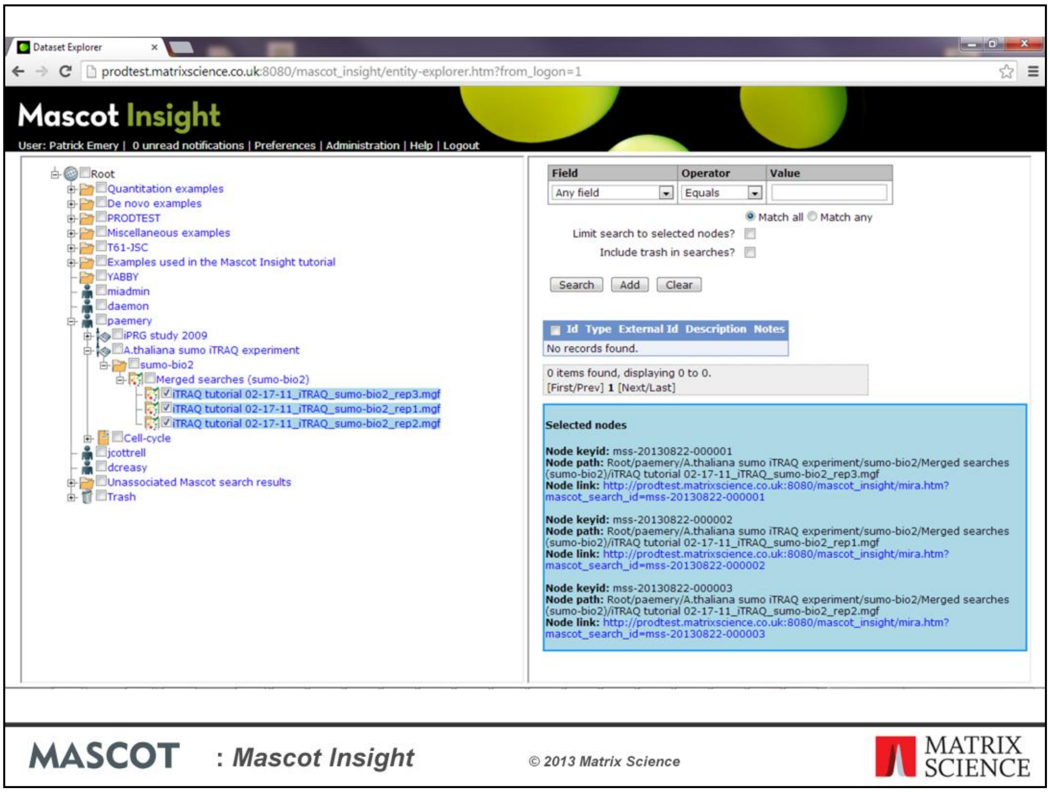
That was a fairly complex example of how you can use an import assignment filter, and often the desired folder structure would be much simpler. However, it does illustrate that you can easily and automatically create a complex, organised folder structure in Mascot Insight.



Reports and exports

- **Ships with some 30 reports covering**
  - Quantitation comparison, qc and clustering
  - Gene ontology
  - Interactions database analysis
  - Shared protein and peptide comparisons
  - Scatter plots, histograms etc
- **Export reports**
  - Statistics export
  - Publication export (MCP guidelines)

The system ships with some 30 reports designed specifically for proteomics data. These cover a wide range of areas such as result comparison, quantitation and quantitation clustering/grouping, Gene Ontology analysis, Interactions database analysis and general graphing reports such as scatter plots. Reports and exports can be run either directly from the Data explorer or via MIRA. Once generated, reports can be exported in a variety of formats, including CSV, SVG, PNG and HTML, or saved to the dataset explorer tree to be easily viewed at a later date. The export reports are used to get data out of the system in various formats - two important examples of the export reports are a ‘Statistic export’ which exports protein, peptide and quantitation data in a tab delineated format for use in statistics packages such as ‘R’ and ‘Perseus’, and a publication export which is based around the MCP guidelines. Let’s look at a few examples of some of the reports that are available.



- ☐ Root
  - ☐ Quantitation examples
  - ☐ De novo examples
  - ☐ PRODTEST
  - ☐ Miscellaneous examples
  - ☐ T61-JSC
  - ☐ Examples used in the Mascot Insight tutorial
  - ☐ YABBY
  - ☐ miadmin
  - ☐ daemon
  - ☐ paemery
    - ☐ iPRG study 2009
    - ☐ A.thaliana sumo iTRAQ experiment
      - ☐ sumo-bio2
        - ☐ Merged searches (sumo-bio2)
          - ☑ iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep3.mgf
          - ☑ iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep1.mgf
          - ☑ iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep2.mgf
    - ☐ Cell-cycle
  - ☐ jcottrell
  - ☐ dcreasy
  - ☐ Unassociated Mascot search results
  - ☐ Trash

Field	Operator	Value
Any field	Equals	

● Match all ○ Match any

Limit search to selected nodes? ☐

Include trash in searches? ☐

☐ Id	Type	External Id	Description	Notes
No records found.				

0 items found, displaying 0 to 0.
[First/Prev] **1** [Next/Last]

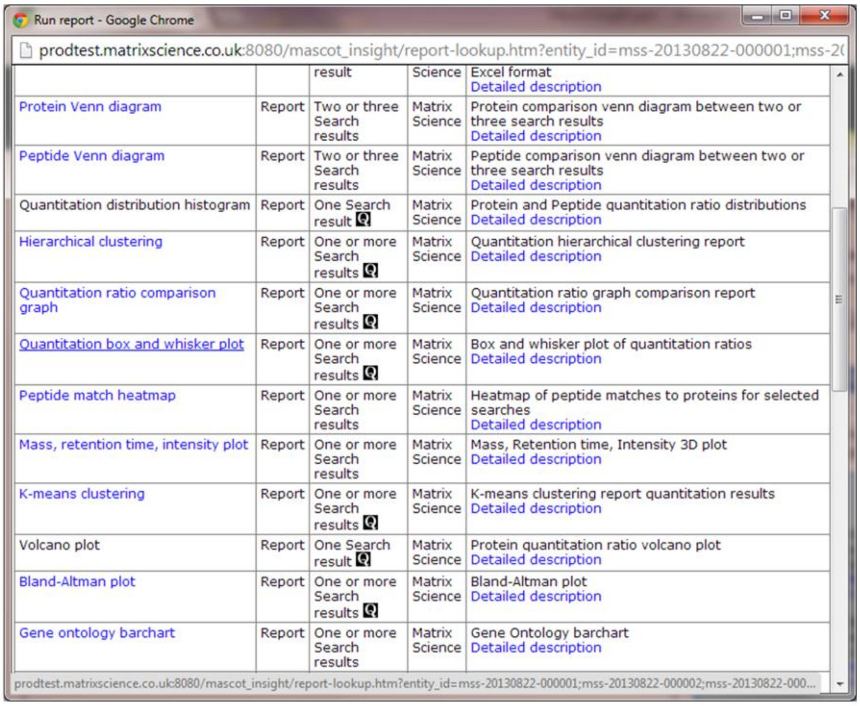
Selected nodes

Node keyid: mss-20130822-000001
Node path: Root/paemery/A.thaliana sumo iTRAQ experiment/sumo-bio2/Merged searches (sumo-bio2)/iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep3.mgf
Node link: http://prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?mascot_search_id=mss-20130822-000001

Node keyid: mss-20130822-000002
Node path: Root/paemery/A.thaliana sumo iTRAQ experiment/sumo-bio2/Merged searches (sumo-bio2)/iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep1.mgf
Node link: http://prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?mascot_search_id=mss-20130822-000002

Node keyid: mss-20130822-000003
Node path: Root/paemery/A.thaliana sumo iTRAQ experiment/sumo-bio2/Merged searches (sumo-bio2)/iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep2.mgf
Node link: http://prodtest.matrixscience.co.uk:8080/mascot_insight/mira.htm?mascot_search_id=mss-20130822-000003

Here we have an iTRAQ dataset with three technical replicates. One way to judge how reproducible the technical replicates have been is a box and whisker plot of the quantitation ratios. This will give us global information for each ratio from each replicate, showing us the global mean and median ratios, the data spread, presence of outliers etc. We can generate the report directly from the data explorer tree. First we select the three searches for the replicates, then right click to bring up the context menu and select the ‘Run report’ option.



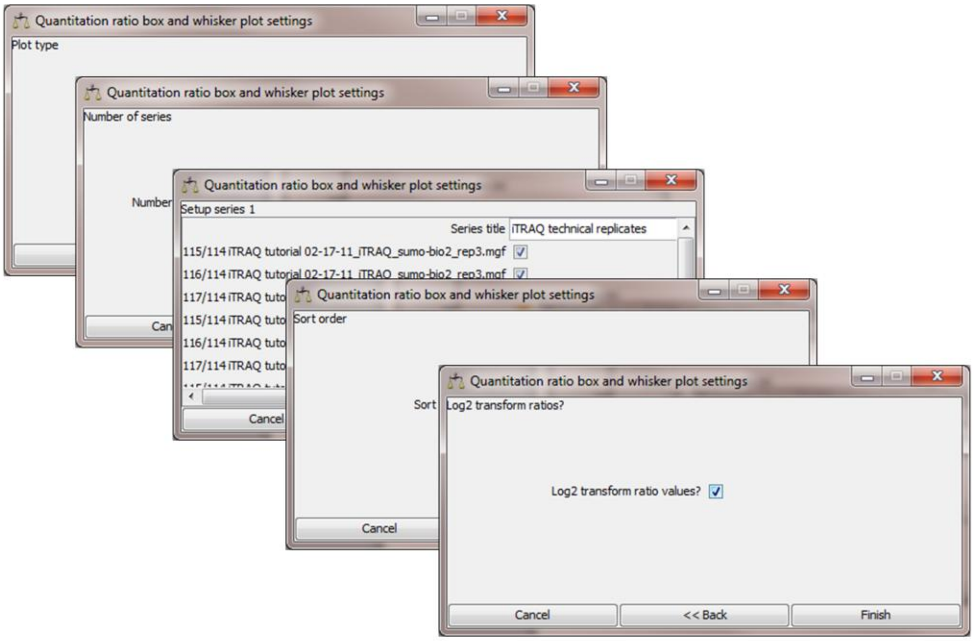
Run report - Google Chrome

prodtest.matrixscience.co.uk:8080/mascot_insight/report-lookup.htm?entity_id=mss-20130822-000001;mss-2(

		result	Science	Excel format Detailed description
Protein Venn diagram	Report	Two or three Search results	Matrix Science	Protein comparison venn diagram between two or three search results Detailed description
Peptide Venn diagram	Report	Two or three Search results	Matrix Science	Peptide comparison venn diagram between two or three search results Detailed description
Quantitation distribution histogram	Report	One Search result Q	Matrix Science	Protein and Peptide quantitation ratio distributions Detailed description
Hierarchical clustering	Report	One or more Search results Q	Matrix Science	Quantitation hierarchical clustering report Detailed description
Quantitation ratio comparison graph	Report	One or more Search results Q	Matrix Science	Quantitation ratio graph comparison report Detailed description
Quantitation box and whisker plot	Report	One or more Search results Q	Matrix Science	Box and whisker plot of quantitation ratios Detailed description
Peptide match heatmap	Report	One or more Search results	Matrix Science	Heatmap of peptide matches to proteins for selected searches Detailed description
Mass, retention time, intensity plot	Report	One or more Search results	Matrix Science	Mass, Retention time, Intensity 3D plot Detailed description
K-means clustering	Report	One or more Search results Q	Matrix Science	K-means clustering report quantitation results Detailed description
Volcano plot	Report	One Search result Q	Matrix Science	Protein quantitation ratio volcano plot Detailed description
Bland-Altman plot	Report	One or more Search results Q	Matrix Science	Bland-Altman plot Detailed description
Gene ontology barchart	Report	One or more Search results	Matrix Science	Gene Ontology barchart Detailed description

prodtest.matrixscience.co.uk:8080/mascot_insight/report-lookup.htm?entity_id=mss-20130822-000001;mss-20130822-000002;mss-20130822-000...

A list selection page of reports will pop open – only the reports we can run on the selected searches will be enabled. Click on the link for the ‘Quantitation box and whisker’ report



Quantitation ratio box and whisker plot settings

Plot type

Quantitation ratio box and whisker plot settings

Number of series

Quantitation ratio box and whisker plot settings

Setup series 1

Series title: iTRAQ technical replicates

115/114 iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep3.mgf ☑

116/114 iTRAQ tutorial 02-17-11_iTRAQ_sumo-bio2_rep3.mgf ☑

Quantitation ratio box and whisker plot settings

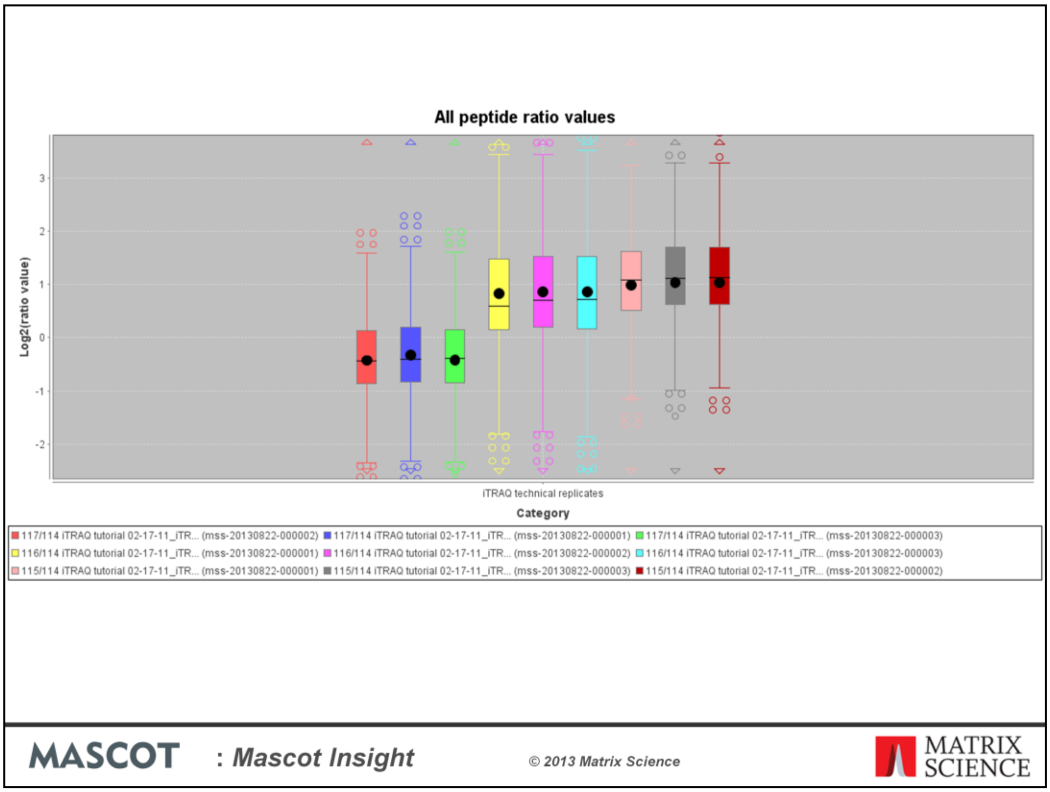
Sort order

Quantitation ratio box and whisker plot settings

Log2 transform ratios?

Log2 transform ratio values? ☑

This will cause the system to launch MIRA and initiate the Wizard used to set up the report

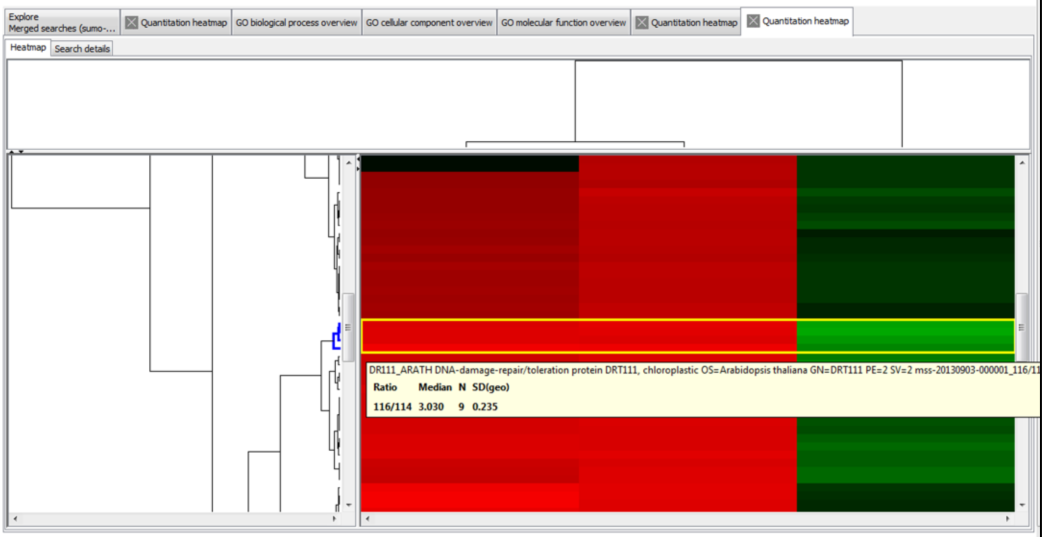


MASCOT : *Mascot Insight* *© 2013 Matrix Science* MATRIX SCIENCE

And here is the generated box and whisker plot. As you can see, the distributions of peptide ratios for each technical replicate are very similar.

Once a report is generated, you can copy and paste it to other applications as a graphic or export it as a CSV file, in case you want to get the data into another package such as Excel or R.

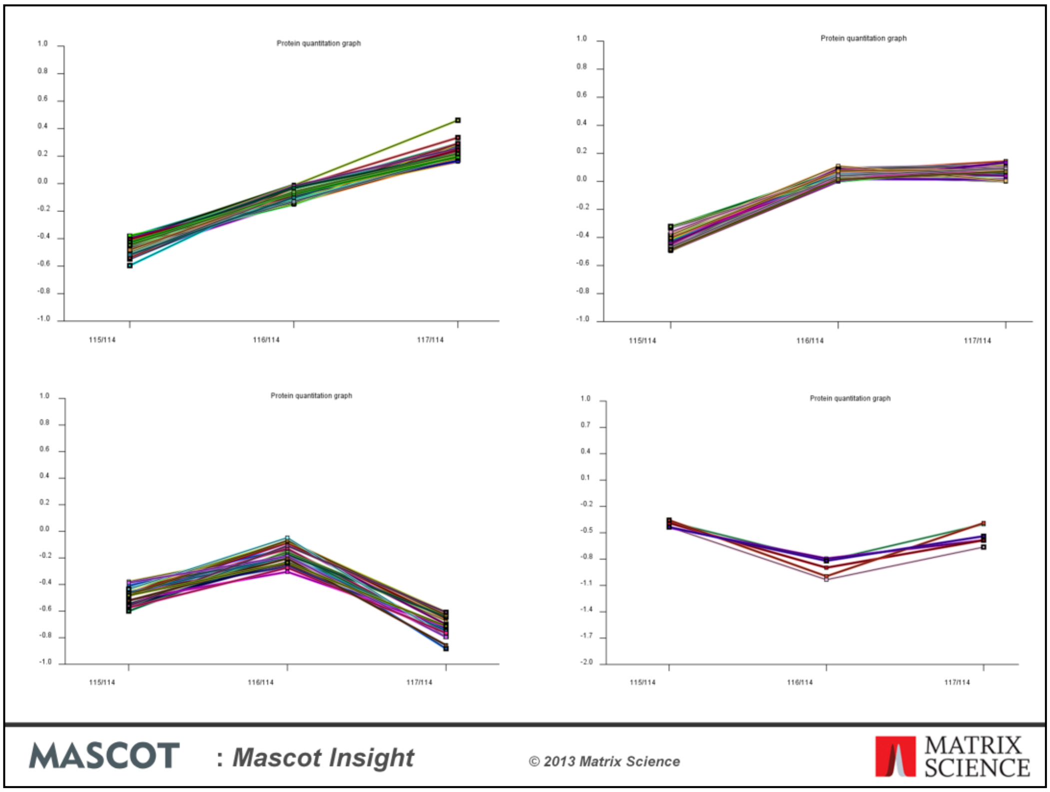
Lets take a look at a few more examples of the types of report you can generate



DR111_ARATH DNA-damage-repair/toleration protein DRT111, chloroplastic OS=Arabidopsis thaliana GN=DRT111 PE=2 SV=2 mss-20130903-000001_116/11

Ratio	Median	N	SD(geo)
116/114	3.030	9	0.235

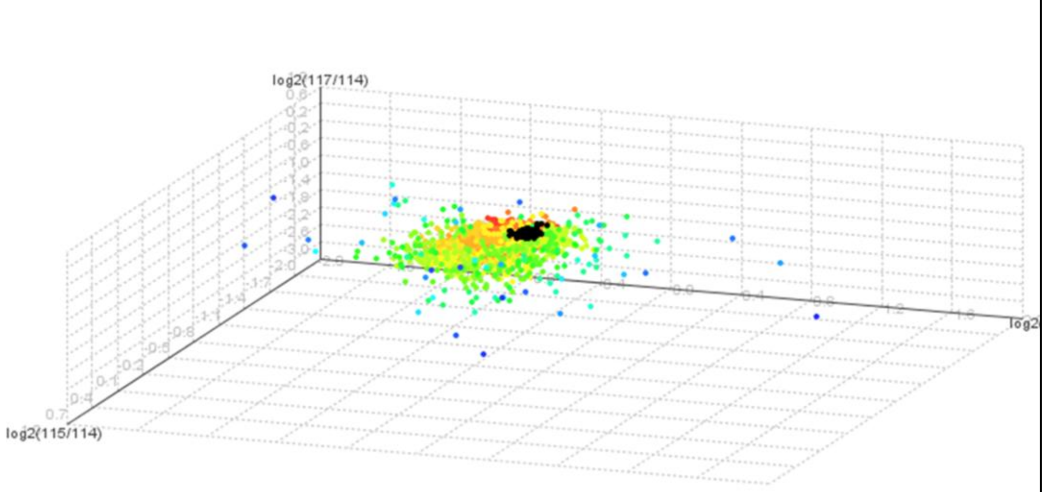
There are several reports available for clustering quantitation ratio values across one or more searches, implementing two common clustering methods. The first of these is the Hierarchical clustering report, which uses hierarchical clustering to group both the proteins and the quantitation ratios into clusters and then generates a heatmap of the results. Here we are looking at the heatmap generated for an iTRAQ 4-plex dataset where Arabidopsis was exposed to heat stress, where red is up-regulated and green down. I’ve applied a Gene Ontology filter to the data, so we’re only looking at the ratios for those proteins that have ‘response to stress’ GO annotation. The rows are individual proteins, and the columns are the quantitation ratios, and we have dendrograms to show the distance between proteins and quantitation ratios. Hover the mouse over a cell on the heatmap to display a tooltip with detailed information about that protein and ratio value.



MASCOT : *Mascot Insight*

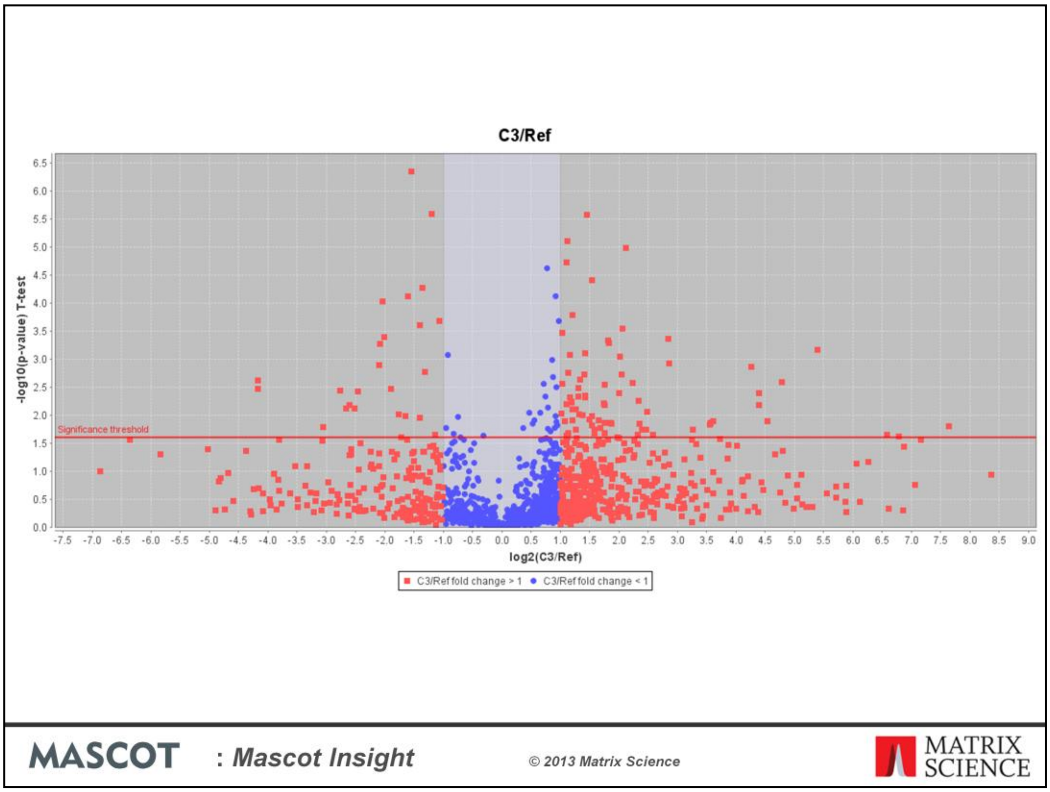
MATRIX SCIENCE

The other type of clustering report that is available is the K-means clustering report. The main difference between K-means clustering and hierarchical clustering is that K-means is a supervised clustering algorithm, which means that you have to specify the number of clusters the algorithm should generate. Here we have four different clusters from an iTRAQ dataset, where each ratio represents a different point on the cell cycle, and the four different clusters are showing different behaviour across those points.



MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

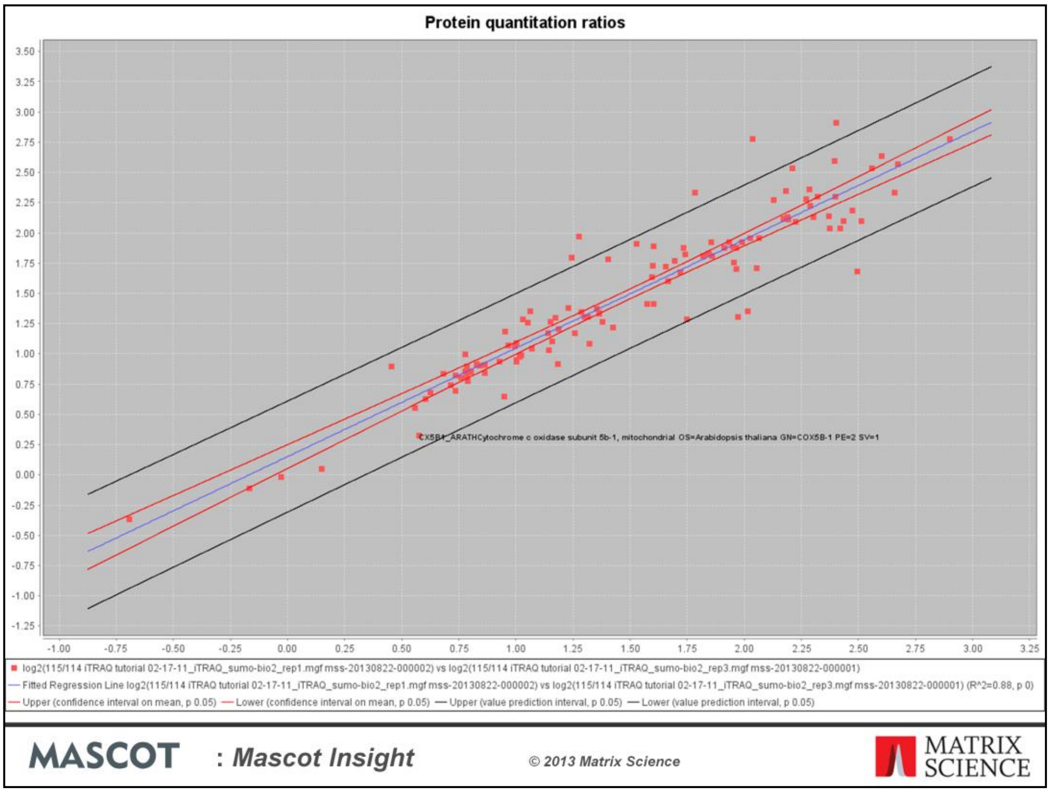
We can also get the report to plot a scatter graph of the clusters – here we have the data for that iTRAQ dataset plotted as a 3D scatter plot. The clusters generated are colour coded. You can rotate 3D scatter plots to get a better view of a particular cluster, hover over a data point and you get a tooltip giving you the protein name. You can also select individual clusters in order to highlight them. Here, I’ve selected the cluster which we’re showing in the top left hand graph.



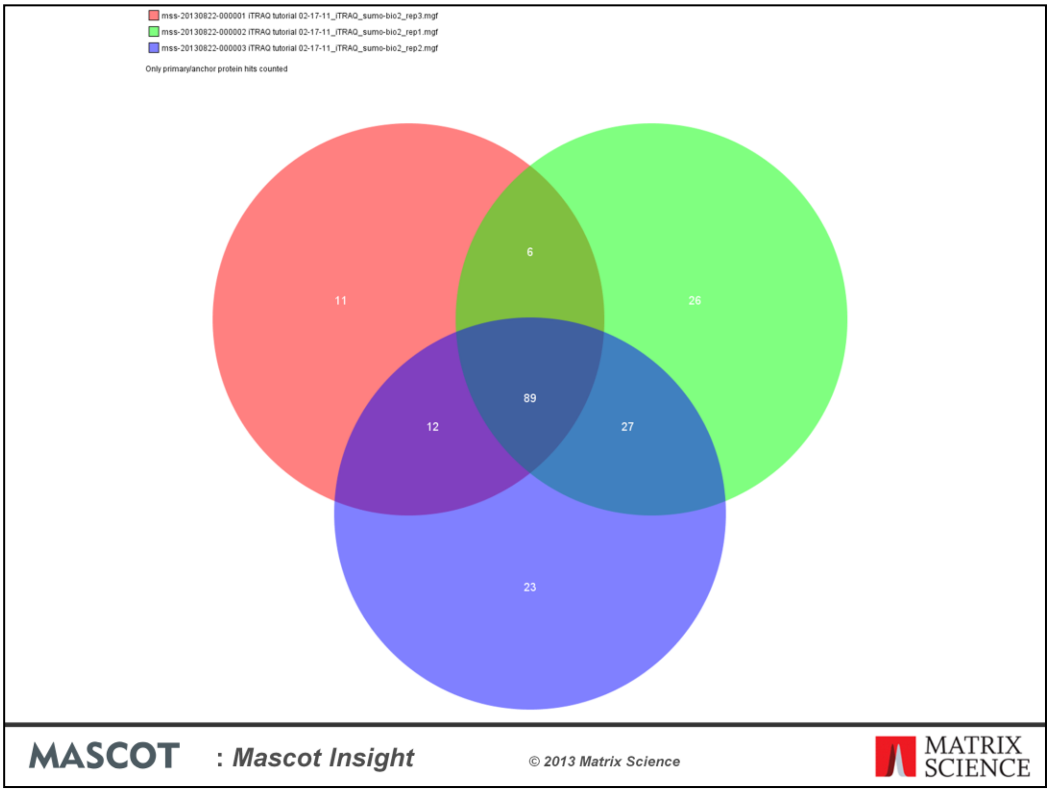
MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

There are various other reports for examining quantitation results. This is a Volcano plot, showing a ratio from a label free dataset

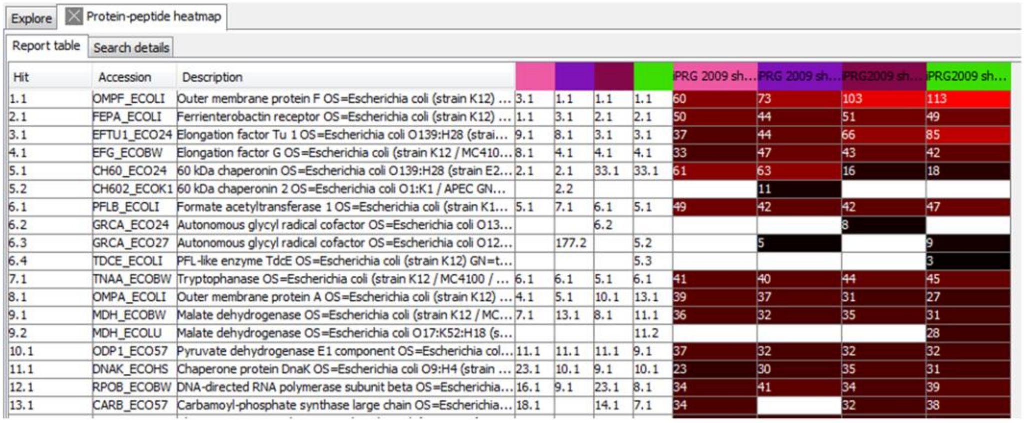
Other quantitation specific plots include a Bland-Altman plot, a QQ plot for peptide ratios matching a protein hit and scatter and histogram plots for comparing ratios (or between searches) and for showing overall ratio distributions.



You can plot scatter plots and histograms of protein and peptide values. In this example, I've plotted the 115/114 iTRAQ ratios from two technical replicates against each other, after excluding matches to the contaminants database, and added a regression line.

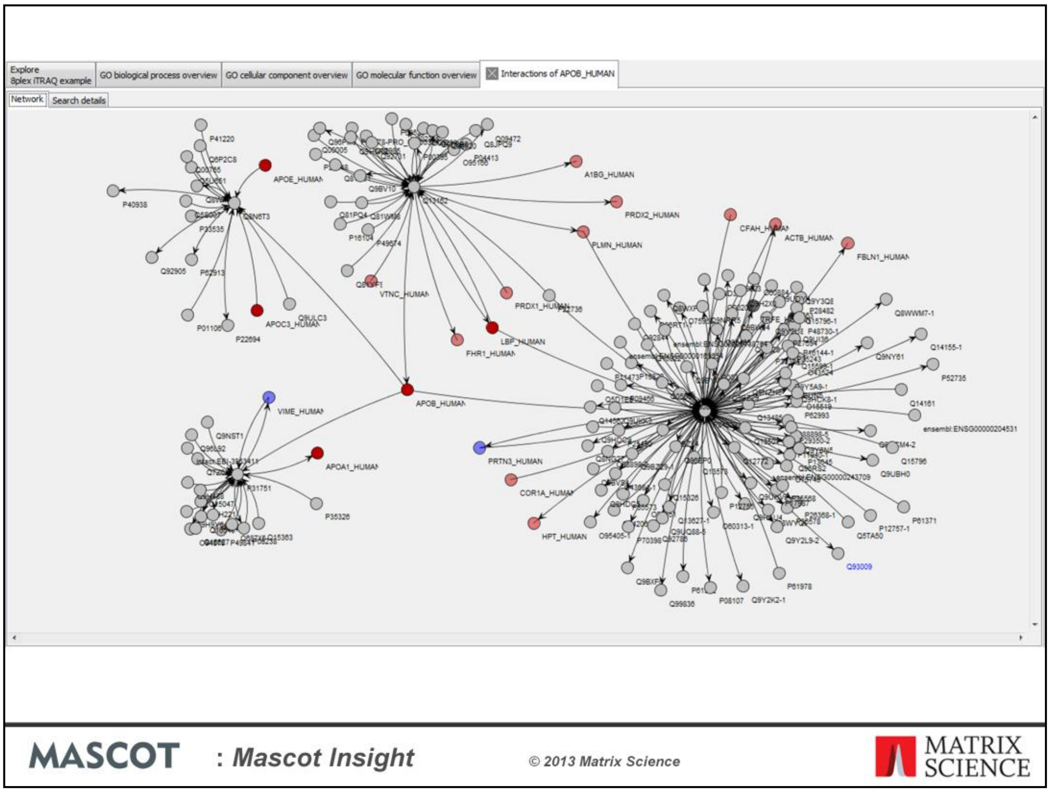


There are also various reports for comparing the protein and peptide matches between searches. For example, you can plot Venn-diagrams for shared protein and peptide matches between searches



Hit	Accession	Description					iPRG 2009 sh…	iPRG 2009 sh…	iPRG2009 sh…	iPRG2009 sh…
1.1	OMPF_ECOLI	Outer membrane protein F OS=Escherichia coli (strain K12) …	3.1	1.1	1.1	1.1	60	73	103	113
2.1	FEPA_ECOLI	Ferrienterobactin receptor OS=Escherichia coli (strain K12) …	1.1	3.1	2.1	2.1	50	44	51	49
3.1	EFTU1_ECO24	Elongation factor Tu 1 OS=Escherichia coli O139:H28 (strai…	9.1	8.1	3.1	3.1	37	44	66	85
4.1	EFG_ECOBW	Elongation factor G OS=Escherichia coli (strain K12 / MC410…	8.1	4.1	4.1	4.1	33	47	43	42
5.1	CH60_ECO24	60 kDa chaperonin OS=Escherichia coli O139:H28 (strain E2…	2.1	2.1	33.1	33.1	61	63	16	18
5.2	CH602_ECOK1	60 kDa chaperonin 2 OS=Escherichia coli O1:K1 / APEC GN…		2.2				11		
6.1	PFLB_ECOLI	Formate acetyltransferase 1 OS=Escherichia coli (strain K1…	5.1	7.1	6.1	5.1	49	42	42	47
6.2	GRCA_ECO24	Autonomous glycyl radical cofactor OS=Escherichia coli O13…			6.2				8	
6.3	GRCA_ECO27	Autonomous glycyl radical cofactor OS=Escherichia coli O12…		177.2		5.2		5		9
6.4	TDCE_ECOLI	PFL-like enzyme TdcE OS=Escherichia coli (strain K12) GN=t…				5.3				3
7.1	TNAA_ECOBW	Tryptophanase OS=Escherichia coli (strain K12 / MC4100 / …	6.1	6.1	5.1	6.1	41	40	44	45
8.1	OMPA_ECOLI	Outer membrane protein A OS=Escherichia coli (strain K12) …	4.1	5.1	10.1	13.1	39	37	31	27
9.1	MDH_ECOBW	Malate dehydrogenase OS=Escherichia coli (strain K12 / MC…	7.1	13.1	8.1	11.1	36	32	35	31
9.2	MDH_ECOLU	Malate dehydrogenase OS=Escherichia coli O17:K52:H18 (s…				11.2				28
10.1	ODP1_ECO57	Pyruvate dehydrogenase E1 component OS=Escherichia col…	11.1	11.1	11.1	9.1	37	32	32	32
11.1	DNAK_ECOHS	Chaperone protein DnaK OS=Escherichia coli O9:H4 (strain …	23.1	10.1	9.1	10.1	23	30	35	31
12.1	RPOB_ECOBW	DNA-directed RNA polymerase subunit beta OS=Escherichia…	16.1	9.1	23.1	8.1	34	41	34	39
13.1	CARB_ECO57	Carbamoyl-phosphate synthase large chain OS=Escherichia…	18.1		14.1	7.1	34		32	38

This is a table based report which shows protein matches for a batch of selected searches with a count of peptide matches for each protein in each search result, with a heat map overlayed over the top of the count values. In this example we're showing counts of significant peptide matches to 4 of the searches from the iPRG2009 study, which are comprised of 2 samples. The two columns on the left are from the 'Red' sample, and the two on the right are from the 'Yellow' sample.

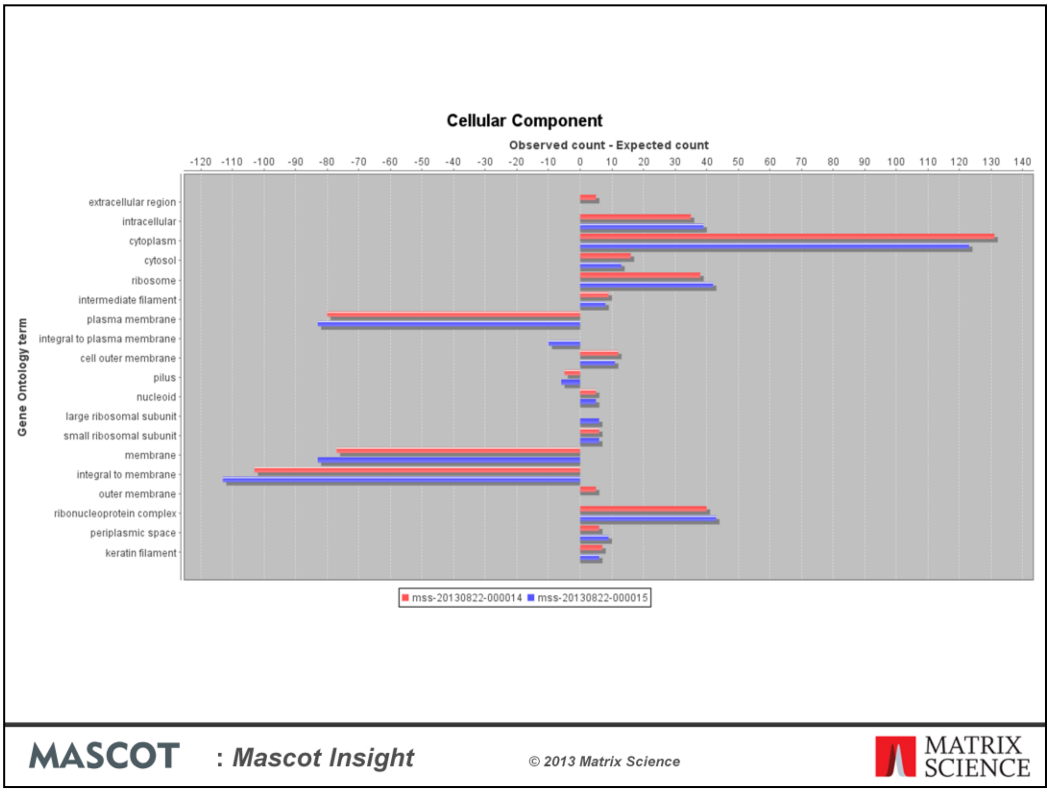


MASCOT : *Mascot Insight*

© 2013 Matrix Science

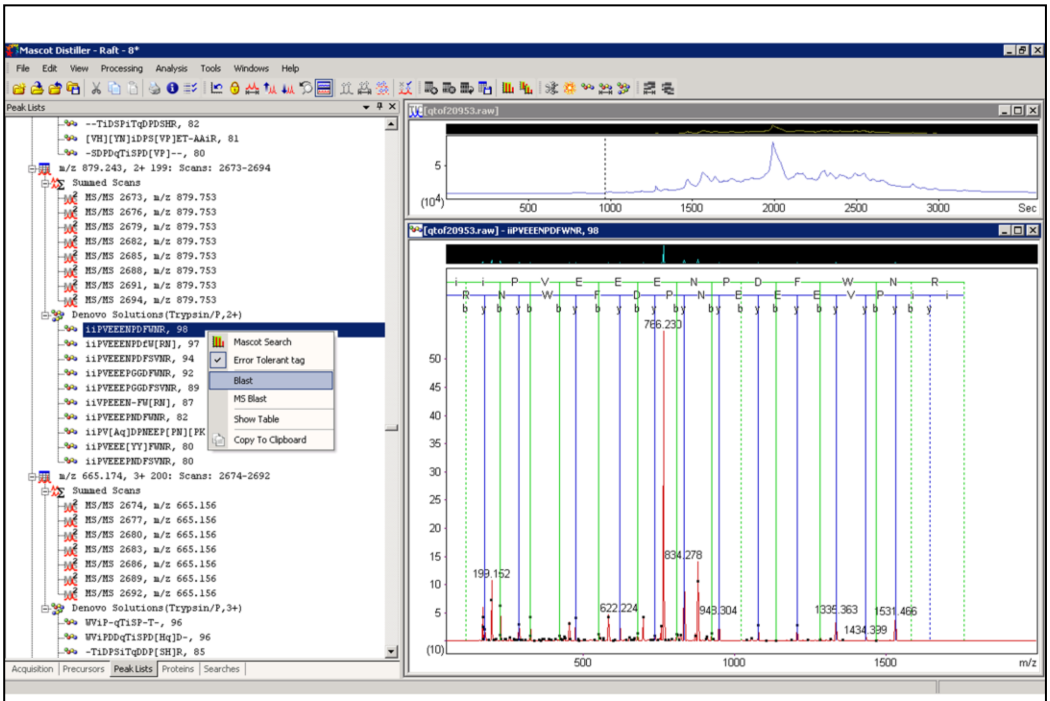
MATRIX SCIENCE

Mascot Insight can generate a molecular interaction network report from any database in the HUPO PSI-MITAB standard format. Here we are looking at a report generated from the IntAct molecular interactions database from the European Bioinformatics Institute for APOB_HUMAN Apolipoprotein B from an iTRAQ 8-plex search. The circles represent proteins – with grey circles representing proteins which were not identified. Proteins which were identified are overlayed by a heatmap, in this case for the iTRAQ 118/113 ratio – with red being up regulated – or shown in blue if the protein was identified by not quantified.



MASCOT : *Mascot Insight* *© 2013 Matrix Science* MATRIX SCIENCE

This report shows enrichment and depletion for GO terms found in selected searches relative to the frequencies in a reference proteome. Proteome GO term frequencies are calculated from the Uniprot reference proteomes, and can be generated for any organism provided the reference proteome is available and you know the NCBI taxonomy id. In the example shown, we’re looking at the results for two merged searches generated from the technical replicates of the ABRF iPRG 2009 study dataset against the E. coli reference proteome – you can see we have a strong enrichment for cytoplasmic proteins, and a depletion of membrane proteins when compared with the proteome frequencies for both searches. You can generate this report using just the GO terms identified for proteins in the selected searches, or against all the GO terms found in the proteome.

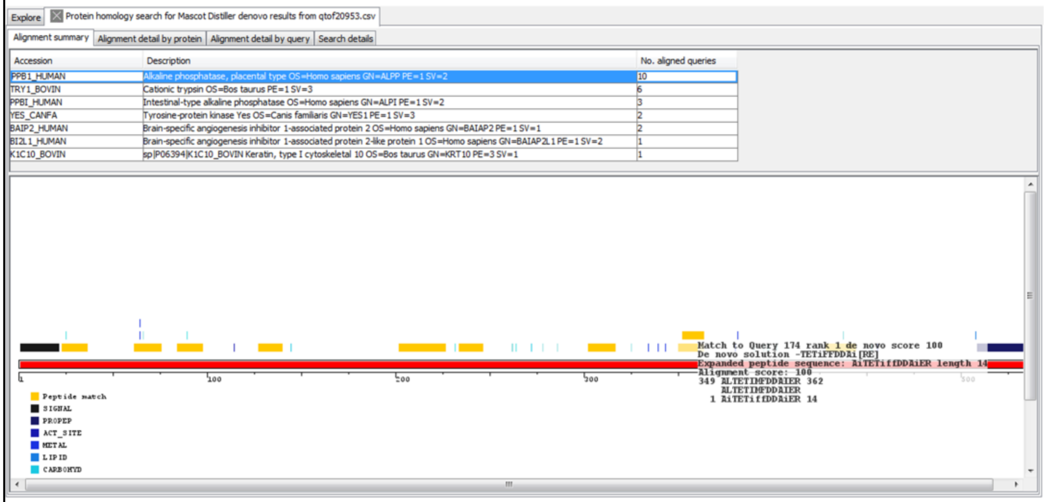


```
--TiDSPiTqDPDSHR, 82
[VH][YN]iDPS[VP]ET-AAiR, 81
-SDPDqTiSPD[VP]--, 80
m/z 879.243, 2+ 199: Scans: 2673-2694
Summed Scans
MS/MS 2673, m/z 879.753
MS/MS 2676, m/z 879.753
MS/MS 2679, m/z 879.753
MS/MS 2682, m/z 879.753
MS/MS 2685, m/z 879.753
MS/MS 2688, m/z 879.753
MS/MS 2691, m/z 879.753
MS/MS 2694, m/z 879.753
Denovo Solutions(Trypsin/P,2+)
iiPVEEENPDFWNR, 98
iiPVEEENPDfW[RN], 97
iiPVEEENPDFSVNR, 94
iiPVEEEPGGDFWNR, 92
iiPVEEEPGGDFSVNR, 89
iiVPEEEN-FW[RN], 87
iiPVEEEPNDFWNR, 82
iiPV[Aq]DPNEEP[PN][PK
iiPVEEE[YY]FWNR, 80
iiPVEEEPNDFSVNR, 80
m/z 665.174, 3+ 200: Scans: 2674-2692
Summed Scans
MS/MS 2674, m/z 665.156
MS/MS 2677, m/z 665.156
MS/MS 2680, m/z 665.156
MS/MS 2683, m/z 665.156
MS/MS 2686, m/z 665.156
MS/MS 2689, m/z 665.156
MS/MS 2692, m/z 665.156
Denovo Solutions(Trypsin/P,3+)
WViP-qTiSP-T-, 96
WViPDDqTiSPD[Hq]D-, 96
-TiDPSiTqDDP[SH]R, 85
```

The Mascot Distiller search toolbox includes a powerful de novo sequencing algorithm. A common reason for carrying out de novo sequencing is to try to find matches to MS/MS spectra which fail to give matches in a database search.

After database search results have been imported into Mascot Distiller, you can choose to de novo all the unassigned spectra. This gives a number of de novo solutions of varying quality. The challenge is how to verify these results. Unless one is working with an organism that is very poorly represented in the databases, truly novel sequences are rare. More usually, there was no match from the database search because of a SNP or unexpected modification or non-specific cleavage. In which case, confidence in a de novo solution comes from aligning it to one of the proteins found in the database search.

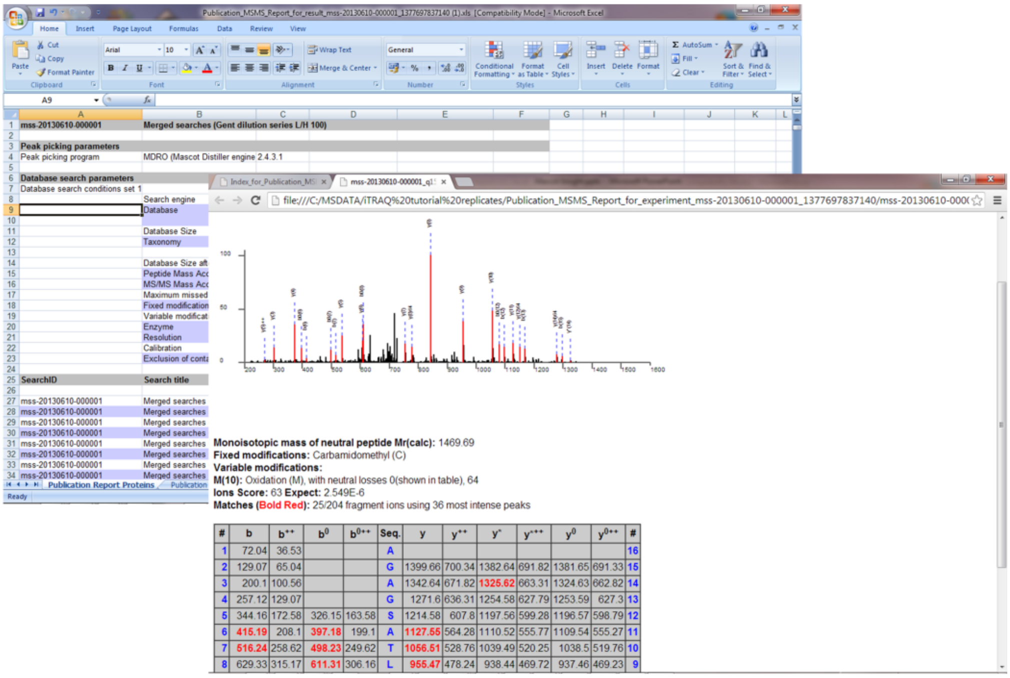
One possibility is to submit the high scoring de novo solutions for a sequence homology search, such as BLAST or MS-BLAST, or for an error tolerant sequence tag search. This is time consuming as the results from these searches have to be individually inspected.



Accession	Description	No. aligned queries
PPB1_HUMAN	Alkaline phosphatase, placental type OS=Homo sapiens GN=ALPP PE=1 SV=2	10
TRY1_BOVIN	Cationic trypsin OS=Bos taurus PE=1 SV=3	6
PPBI_HUMAN	Intestinal-type alkaline phosphatase OS=Homo sapiens GN=ALPI PE=1 SV=2	3
YES_CANFA	Tyrosine-protein kinase Yes OS=Canis familiaris GN=YES1 PE=1 SV=3	2
BAIP2_HUMAN	Brain-specific angiogenesis inhibitor 1-associated protein 2 OS=Homo sapiens GN=BAIAP2 PE=1 SV=1	2
BI2L1_HUMAN	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1 OS=Homo sapiens GN=BAIAP2L1 PE=1 SV=2	1
K1C10_BOVIN	sp|P06394|K1C10_BOVIN Keratin, type I cytoskeletal 10 OS=Bos taurus GN=KRT10 PE=3 SV=1	1

```
Match to Query 174 rank 1 de novo score 100
De novo solution -TETiFFDDAi[RE]
Expanded peptide sequence: AiTETiffDDAiER length 14
Alignment score: 100
349 ALTETIMFDDAIER 362
    ALTETIMFDDAIER
  1 AiTETiffDDAiER 14
```

In Mascot Insight, we have implemented a report which allows you to use the protein hit sequences from a selected search result to carry out a BLAST like sequence homology search using de novo solutions, in order to try to find additional possible matches to spectra from your dataset, drilling down into the unassigned MS/MS spectra. The sequence homology search has been tailored to de novo solution data, and allows for Q->K and I->L in the alignment without penalty for example. This allows us to see if we have any convincing de novo sequence matches to peptides from the protein hits from the initial Mascot database search – for example, this looks like a good match where the differences at the n-terminus will have prevented us from getting a sequence database match.



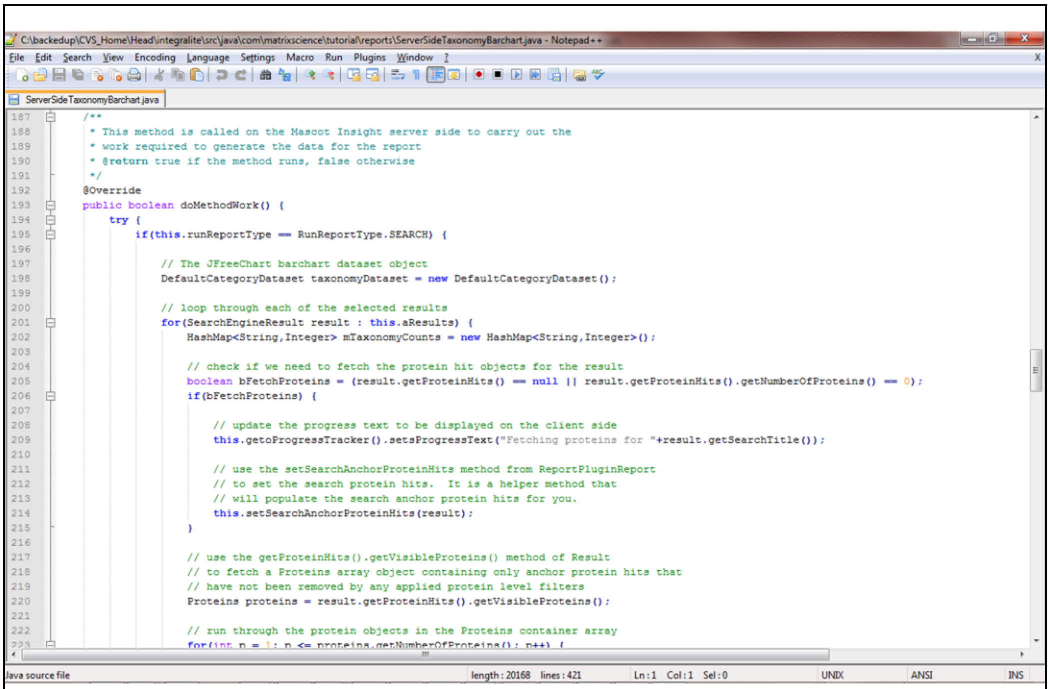
file:///C:/MSDATA/iTRAQ%20tutorial%20replicates/Publication_MSMS_Report_for_experiment_mss-20130610-000001_1377697837140/mss-20130610-000(

	A	B
1	**mss-20130610-000001**	**Merged searches (Gent dilution series L/H 100)**
2		
3	**Peak picking parameters**	
4	Peak picking program	MDRO (Mascot Distiller engine 2.4.3.1
5		
6	**Database search parameters**	
7	Database search conditions set 1	
8		Search engine
9		Database
10		
11		Database Size
12		Taxonomy
13		
14		Database Size aft
15		Peptide Mass Acc
16		MS/MS Mass Acc
17		Maximum missed
18		Fixed modification
19		Variable modificat
20		Enzyme
21		Resolution
22		Calibration
23		Exclusion of conta
24		
25	**SearchID**	**Search title**
26		
27	mss-20130610-000001	Merged searches
28	mss-20130610-000001	Merged searches
29	mss-20130610-000001	Merged searches
30	mss-20130610-000001	Merged searches
31	mss-20130610-000001	Merged searches
32	mss-20130610-000001	Merged searches
33	mss-20130610-000001	Merged searches
34	mss-20130610-000001	Merged searches

Monoisotopic mass of neutral peptide Mr(calc): 1469.69
Fixed modifications: Carbamidomethyl (C)
Variable modifications:
M(10): Oxidation (M), with neutral losses 0(shown in table), 64
Ions Score: 63 **Expect:** 2.549E-6
Matches (Bold Red): 25/204 fragment ions using 36 most intense peaks

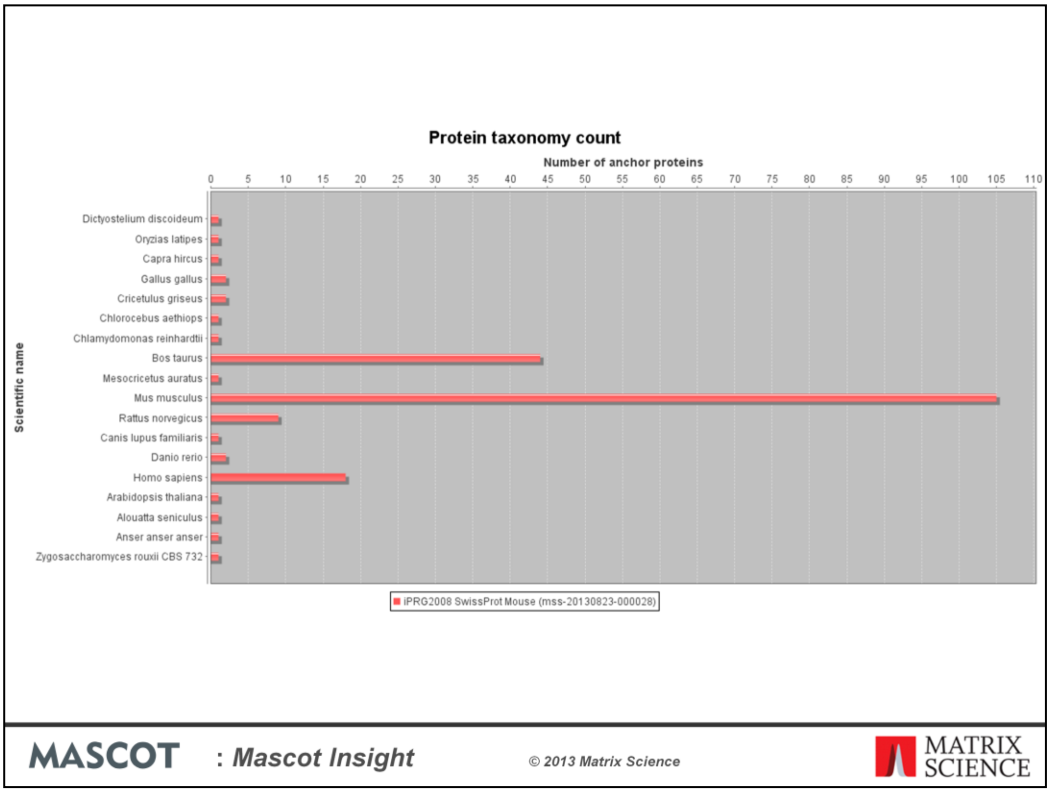
#	b	b++	b0	b0++	Seq.	y	y++	y*	y*++	y0	y0++	#
1	72.04	36.53			A							16
2	129.07	65.04			G	1399.66	700.34	1382.64	691.82	1381.65	691.33	15
3	200.1	100.56			A	1342.64	671.82	**1325.62**	663.31	1324.63	662.82	14
4	257.12	129.07			G	1271.6	636.31	1254.58	627.79	1253.59	627.3	13
5	344.16	172.58	326.15	163.58	S	1214.58	607.8	1197.56	599.28	1196.57	598.79	12
6	**415.19**	208.1	**397.18**	199.1	A	**1127.55**	564.28	1110.52	555.77	1109.54	555.27	11
7	**516.24**	258.62	**498.23**	249.62	T	**1056.51**	528.76	1039.49	520.25	1038.5	519.76	10
8	629.33	315.17	**611.31**	306.16	L	**955.47**	478.24	938.44	469.72	937.46	469.23	9

The Publication export report exports data for selected search results in a report designed to meet the guidelines for publication in the journal Molecular and Cellular Proteomics. Running this report generates a zip archive containing an index HTML page, with links to the main Excel report file and links to static HTML pages which contain the spectrum views for MS/MS peptide matches (or MS protein matches for PMF data). A separate archive is generated for each search result, or you can generate a single report for a merged dataset.

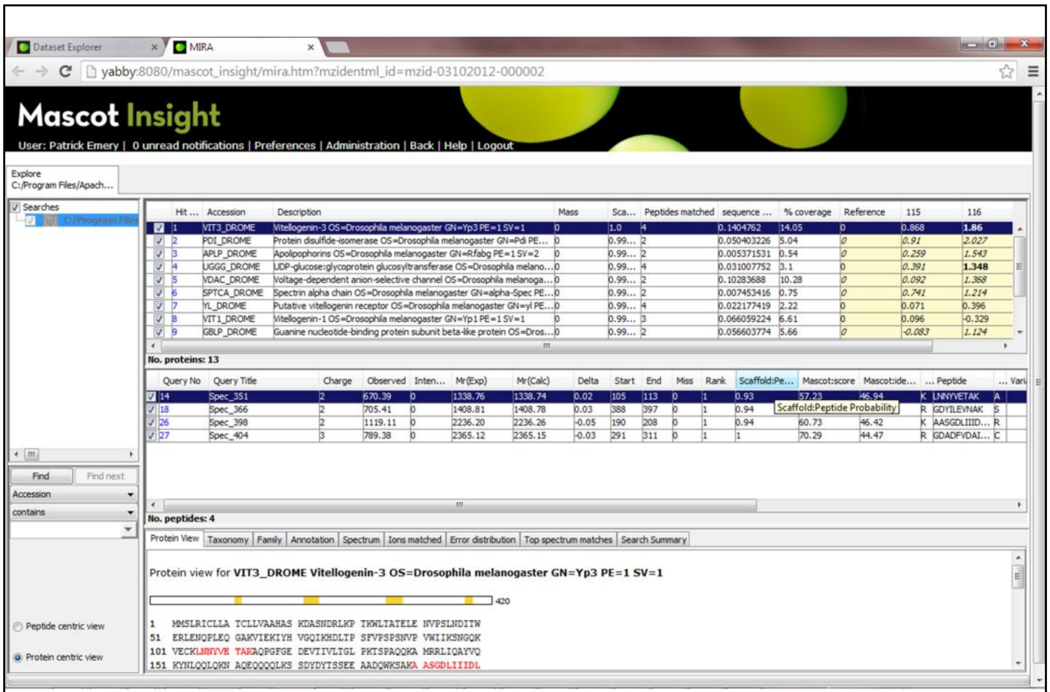


```
    /**
     * This method is called on the Mascot Insight server side to carry out the
     * work required to generate the data for the report
     * @return true if the method runs, false otherwise
     */
    @Override
    public boolean doMethodWork() {
        try {
            if(this.runReportType == RunReportType.SEARCH) {

                // The JFreeChart barchart dataset object
                DefaultCategoryDataset taxonomyDataset = new DefaultCategoryDataset();

                // loop through each of the selected results
                for(SearchEngineResult result : this.aResults) {
                    HashMap<String,Integer> mTaxonomyCounts = new HashMap<String,Integer>();

                    // check if we need to fetch the protein hit objects for the result
                    boolean bFetchProteins = (result.getProteinHits() == null || result.getProteinHits().getNumberOfProteins() == 0);
                    if(bFetchProteins) {

                        // update the progress text to be displayed on the client side
                        this.getoProgressTracker().setsProgressText("Fetching proteins for "+result.getSearchTitle());

                        // use the setSearchAnchorProteinHits method from ReportPluginReport
                        // to set the search protein hits.  It is a helper method that
                        // will populate the search anchor protein hits for you.
                        this.setSearchAnchorProteinHits(result);
                    }

                    // use the getProteinHits().getVisibleProteins() method of Result
                    // to fetch a Proteins array object containing only anchor protein hits that
                    // have not been removed by any applied protein level filters
                    Proteins proteins = result.getProteinHits().getVisibleProteins();

                    // run through the protein objects in the Proteins container array
                    for(int n = 1; n <= proteins.getNumberOfProteins(); n++) {
```

In addition to these reports and others, you can write your own reports. You will need to have a good working knowledge of the Java programming language. You can write both chart-type reports for use in MIRA, and export-type reports to create files in any format you choose. The example shown here will generate a barchart of taxonomy ids for a selected search result. The Mascot Insight API is a high level API, so that the code works as well for searches imported in protXML and mzIdentML formats as for Mascot results.



And this is the output from the report run against the iPRG2008 dataset searched against the whole of SwissProt, with the largest number of protein hits being to mouse (as might be expected since this is a mouse dataset).



	Hit ...	Accession	Description	Mass	Sca...	Peptides matched	sequence ...	% coverage	Reference	115	116
☑	1	VIT3_DROME	Vitellogenin-3 OS=Drosophila melanogaster GN=Yp3 PE=1 SV=1	0	1.0	4	0.1404762	14.05	0	0.868	**1.86**
☑	2	PDI_DROME	Protein disulfide-isomerase OS=Drosophila melanogaster GN=Pdi PE...	0	0.99...	2	0.050403226	5.04	*0*	*0.91*	*2.027*
☑	3	APLP_DROME	Apolipophorins OS=Drosophila melanogaster GN=Rfabg PE=1 SV=2	0	0.99...	2	0.005371531	0.54	*0*	*0.259*	*1.543*
☑	4	UGGG_DROME	UDP-glucose:glycoprotein glucosyltransferase OS=Drosophila melano...	0	0.99...	4	0.031007752	3.1	0	*0.391*	**1.348**
☑	5	VDAC_DROME	Voltage-dependent anion-selective channel OS=Drosophila melanoga...	0	0.99...	2	0.10283688	10.28	*0*	*0.092*	*1.368*
☑	6	SPTCA_DROME	Spectrin alpha chain OS=Drosophila melanogaster GN=alpha-Spec PE...	0	0.99...	2	0.007453416	0.75	*0*	*0.741*	*1.214*
☑	7	YL_DROME	Putative vitellogenin receptor OS=Drosophila melanogaster GN=yl PE...	0	0.99...	4	0.022177419	2.22	0	0.071	0.396
☑	8	VIT1_DROME	Vitellogenin-1 OS=Drosophila melanogaster GN=Yp1 PE=1 SV=1	0	0.99...	3	0.066059224	6.61	0	0.096	-0.329
☑	9	GBLP_DROME	Guanine nucleotide-binding protein subunit beta-like protein OS=Dros...	0	0.99...	2	0.056603774	5.66	*0*	*-0.083*	*1.124*

No. proteins: 13

	Query No	Query Title	Charge	Observed	Inten...	Mr(Exp)	Mr(Calc)	Delta	Start	End	Miss	Rank	Scaffold:Pe...	Mascot:score	Mascot:ide...	...	Peptide	...	Vari
☑	14	Spec_351	2	670.39	0	1338.76	1338.74	0.02	105	113	0	1	0.93	57.23	46.94	K	LNNYVETAK	A	
☑	18	Spec_366	2	705.41	0	1408.81	1408.78	0.03	388	397	0	1	0.94	[illegible]	[illegible]	R	GDYILEVNAK	S	
☑	26	Spec_398	2	1119.11	0	2236.20	2236.26	-0.05	190	208	0	1	0.94	60.73	46.42	K	AASGDLIIID...	R	
☑	27	Spec_404	3	789.38	0	2365.12	2365.15	-0.03	291	311	0	1	1	70.29	44.47	R	GDADFVDAI...	C	

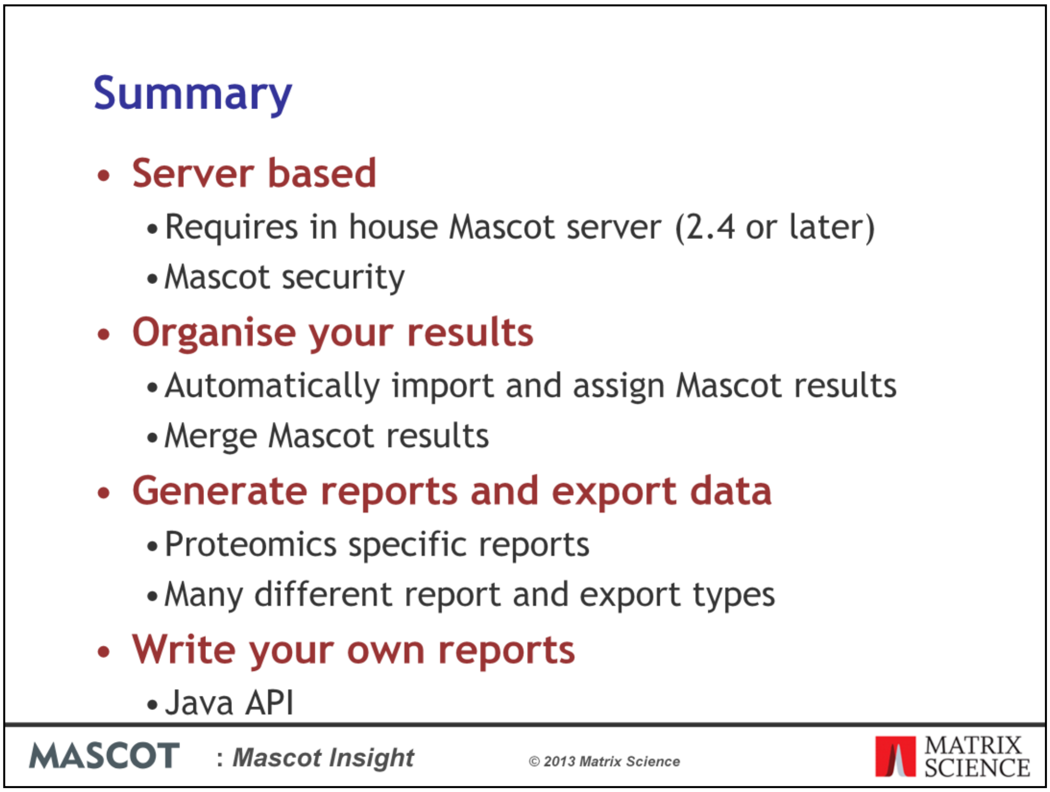
Scaffold:Peptide Probability

No. peptides: 4

Protein view for **VIT3_DROME Vitellogenin-3 OS=Drosophila melanogaster GN=Yp3 PE=1 SV=1**

```
1   MMSLRICLLA TCLLVAAHAS KDASNDRLKP TKWLTATELE NVPSLNDITW
51  ERLENQPLEQ GAKVIEKIYH VGQIKHDLTP SFVPSPSNVP VWIIKSNGQK
101 VECKLNNYVE TAKAQPGFGE DEVTIVLTGL PKTSPAQQKA MRRLIQAYVQ
151 KYNLQQLQKN AQEQQQQLKS SDYDYTSSEE AADQWKSAKA ASGDLIIIDL
```

As mentioned earlier, you can import search results from other search engines in either protXML or mzIdentML standard formats. Here we are looking at an mzIdentML format result exported from Scaffold Q+, and we’ve also imported the Scaffold quantML quantitation result export file as well. We can use these results to generate reports, just as for Mascot search results.



Summary

- **Server based**
  - Requires in house Mascot server (2.4 or later)
  - Mascot security
- **Organise your results**
  - Automatically import and assign Mascot results
  - Merge Mascot results
- **Generate reports and export data**
  - Proteomics specific reports
  - Many different report and export types
- **Write your own reports**
  - Java API

MASCOT : *Mascot Insight* © 2013 Matrix Science MATRIX SCIENCE

In summary, Mascot Insight is a powerful new application that allows you to organise your search results, automatically importing Mascot search results into folder structures of your design. It allows you to merge Mascot search results, including all types of quantitation data.

The system ships with a large number of reports, covering areas such as protein and peptide level comparisons between datasets, quantitation based reports for comparison, ratio clustering and quality control. You can carry out Gene Ontology analysis of your data, including display of enriched and depleted GO terms compared with a reference proteome. Other reports are based on molecular interactions databases in the PSI-MITAB format.

The results of these reports are easily exported from the system, and there are many export options.

You can also write your own reports in Java using a high level API.