

Quantitation: Reporting

MASCOT

: *Quantitation*

© 2007-2023 Matrix Science



This presentation is the last in a series on quantitation. In this presentation we cover reporting, in particular the reporting options that are available in Mascot Distiller and the quantitation toolbox.

Reporting

Supports both MS1 and MS2 quantitation

In Mascot Distiller

Basic tables of peptide and protein information

Quality control plots

Statistical analysis plots

Stats package export

XML export

In Mascot Daemon

Quantitation summaries

MASCOT

: Quantitation

© 2007-2023 Matrix Science



These reports are normally used with precursor based quantitation methods like label free and SILAC, but can also be used with MS2 based quantitation reporter ion methods like iTRAQ and TMT, as long as the peak picking was performed with Mascot Distiller and search results have been saved in the Distiller file.

The Mascot Distiller GUI is a good way to review the quantitation results, to see that everything has worked as expected. The raw results can be used to generate several different reports and tables.

You can then take the next step and analyze the quantitation results to determine the significantly proteins that are up or down regulated using statistical analysis plots.

Or the results can be exported to a table ready for third party statistical analysis or an XML file.

Mascot Daemon can also be used to generate some additional peptide intensity based reports.

Quantitation reports in Mascot Distiller

Reports

Peptides

Proteins

Quality control

Tables

Matches

Peptides

Peptides with intensity info - good for peptide level quantitation
experiments like phosphopeptide quantitation

Proteins

MASCOT

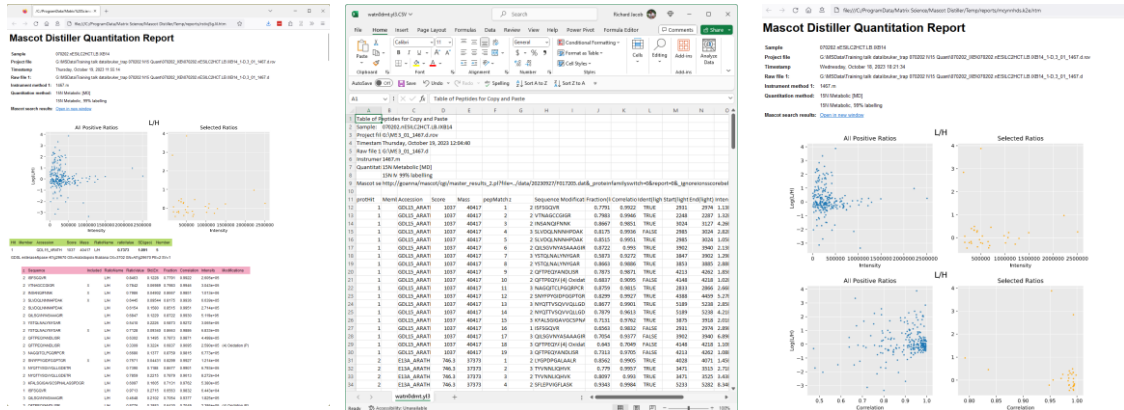
: *Quantitation*

© 2007-2023 Matrix Science



Starting with the straight forward quantitation results you can view them as reports or tables at the match, peptide or protein level.

Quantitation Reports and Tables



MASCOT

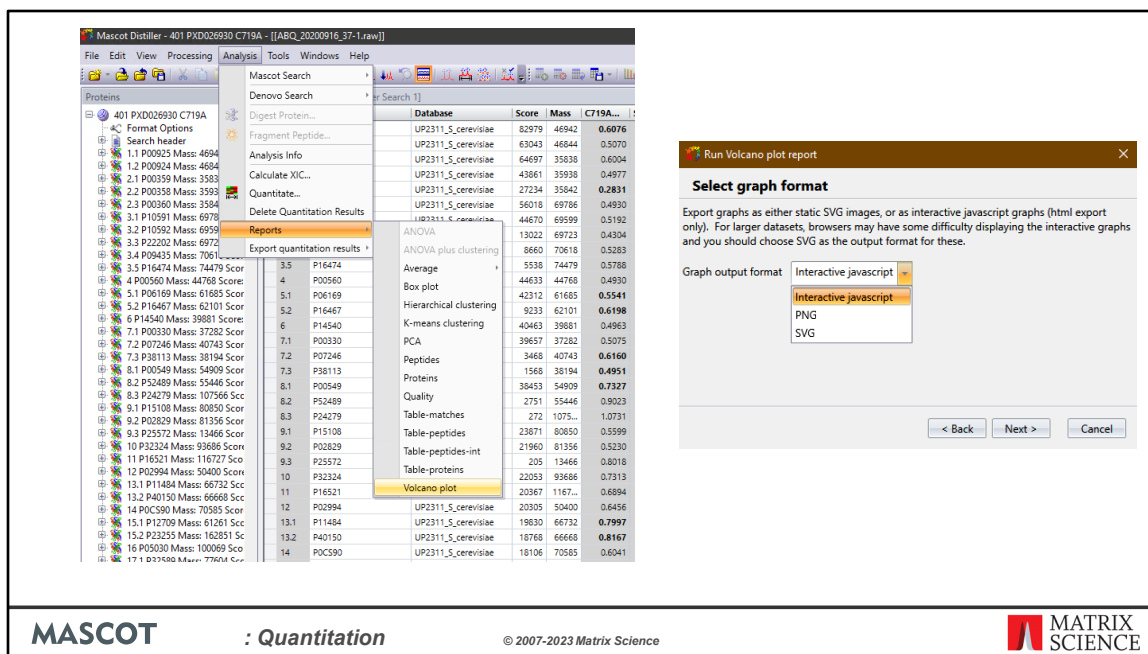
: **Quantitation**

© 2007-2023 Matrix Science



MATRIX
SCIENCE

Choose a report from *Reports...* on the Analysis menu and in the dialog chose a format. Tables can be exported as HTML or CSV (comma-separate values) format and click Finish to generate the report. The CSV file which can be opened directly in Excel or any other spreadsheet package.



Next we have the statistical analysis reports included with Mascot Distiller. These reports are generated by python scripts and cover ANOVA, Hierarchical and K-means Clustering, Principal Component Analysis and Volcano plots.

Let's create a volcano plot. Go to the Analysis->Reports menu and select Volcano plot. Step through the dialog boxes setting the protein ratio significance threshold. I left it at the least stringent default values, p less than 0.05 which is the same as used in the publication the data came from. Filter out hits from the contaminant database, primarily Trypsin in this case.

And chose the output format. SVG is a good choice as it can be scaled without losing resolution. PNG is good for pasting into simple reports or for use online. The interactive Javascript report allows you mouse over points in the graph and see the protein accession numbers. I've selected interactive Javascript.

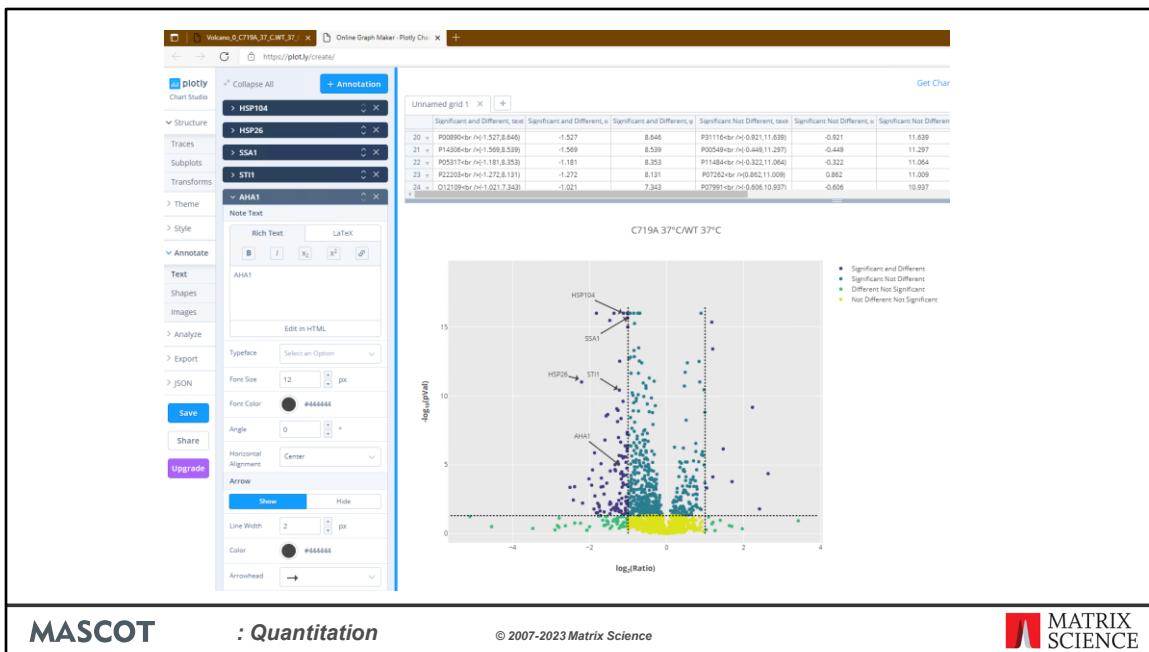
The report is generated and saved to a temporary archive. Move the archive to a project folder, uncompress it and open the HTML page in a browser by double clicking on the file.



Here is the report from Distiller. The icons at the top right allow you to zoom in and interact with the plot.

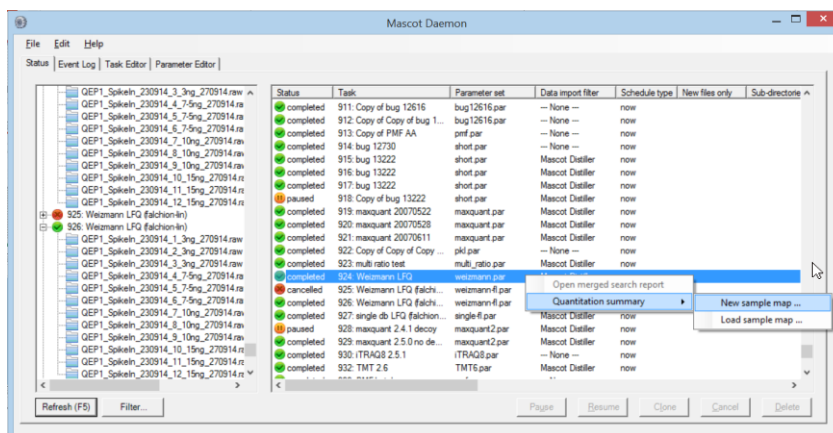
You can mouse over the data points to see the protein accession numbers, ratio and p values.

We use plot.ly as the graphing library to create these plots and there is also an option to open it in plot.ly for further editing.



The interactive report can add annotations. I have annotated four heat shock proteins and a chaperone that are all downregulated and in the same pathway or were highlighted in the original study.

Quantitation summary



MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

Finally, I should mention Quantitation summaries, an alternative method to output the quantitation results in a format ready to use for statistical analysis. You can export the results from any quantitation method where you have used Mascot Daemon to automate the processing.

Here the files were processed in a single Mascot Daemon task, using Mascot Distiller for peak picking and quantitation. The data is a Label Free Quantitation (LFQ) data set and searched with the '**Average [MD]**' quantitation method.

Once processing was complete, the task was selected in the list view on the Mascot Daemon status tab. Right clicking the selection invoked a context menu, from which Quantitation Summary; New sample map ... was chosen.

Sample map - *

<input checked="" type="checkbox"/>	Task	Task name	Raw file path	Raw file name	Parameter set	Fastq(s)	Submitted	Fraction	Intensity LFQ
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 17.2...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 18.0...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 18.4...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_4_7_5ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 19.2...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_5_7_5ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 20.0...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_6_7_5ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 20.4...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 21.1...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 22.0...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 22.4...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome...	06/08/2019 23.2...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome...	07/08/2019 00.0...		
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\akippy\matlaci...	GEP1_SpkeIn_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome...	07/08/2019 00.4...		

Contaminant DB: Average (MD): Settings ... Save sample map ... Save quantitation summary ... Close

MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

We've tried to minimise the amount of typing required to create a Sample Map. As in Excel, columns can be sorted so that a repeating value can be pasted to a range of cells. In this case, sorting on raw file name (by clicking the column header) is all that is required. For more complex data sets, sorting on file path or task name or time of submission may help organise the files in a useful way. This is particularly important when samples have been separated into large numbers of fractions. You don't want to have to type in every fraction number. Just sort appropriately, select the cell range in the fraction column, right click and choose Fill with integer series.

Sample map - *

<input checked="" type="checkbox"/>	Task	Task name	Raw file path	Raw file name	Parameter set	Fastq(s)	Submitted	Fraction	Intensity LFQ
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17.2	1	3ng*
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.0	2	3ng*
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.4	3	3ng*
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_4_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19.2	1	7.5ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_5_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.0	2	7.5ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_6_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.4	3	7.5ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21.1	1	10ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.0	2	10ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.4	3	10ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23.2	1	15ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00.0	2	15ng
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\\kippymatbio\...	GEP1_Spikeln_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00.4	3	15ng

Contaminant DB: ecoli_proteome | Average [MD] | Settings | Save sample map ... | Save quantitation summary ... | Close

MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

If a sample is not fractionated, the fraction cells can be left empty, unless you wish to merge replicates by treating them as fractions. In this example, there are two useful ways to fill in the two columns on the right that are used to identify the samples. Like this, to merge replicates as if they were fractions. An asterisk indicates the reference or control sample, and the Quantitation Summary will include ratios relative to this sample.

Sample map - *

<input checked="" type="checkbox"/>	Task	Task name	Raw file path	Raw file name	Parameter set	Fasta(s)	Submitted	Fraction	Intensity LFQ
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17.2...		3ng_rep1
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.0...		3ng_rep2
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.4...		3ng_rep3
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_4_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19.2...		7.5ng_rep1
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_5_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.0...		7.5ng_rep2
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_6_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.4...		7.5ng_rep3
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21.1...		10ng_rep1
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.0...		10ng_rep2
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.4...		10ng_rep3
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23.2...		15ng_rep1
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 00.0...		15ng_rep2
<input checked="" type="checkbox"/>	324	Weizmann LFQ	\ukippy\matbio\...	GEP1_Spikeln_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00.4...		15ng_rep3

Contaminant DB: [None] [Average (MD)] Settings ... Save sample map ... Save quantitation summary ... Close

MASCOT

: Quantitation

© 2007-2023 Matrix Science

**MATRIX
SCIENCE**

Or, like this, to create separate columns in the Quantitation Summary for each replicate; useful if you want statistics for variation across replicates. Sample identifiers can be anything you like as long as the combination of identifier and fraction number for each file is unique.

The Sample Map can be saved to a disk file, even if not complete, and reloaded as required. When Save quantitation summary ... is chosen, some validation is performed.

Family index	Member index	Protein IDs	Peptide counts (all)	Peptide counts (unique)	Fasta headers	Peptide XICs [3ng]	Unique peptide XICs [3ng]	Peptide XICs [7.5ng]	Unique peptide XICs [7.5ng]	Peptide XICs [10ng]	Unique peptide XICs [10ng]	Peptide XICs [15ng]	Unique peptide XICs [15ng]	Mol. weight [kDa]	Sec. structure
1	1	1 3:P06733	37	33	Alpha-enolase OS=Homo sapien	139	119	147	129	141	123	136	117	47481	
3	1	2 3:P13929	7	3	Beta-enolase OS=Homo sapien	21	1	20	2	19	1	22	3	47299	
4	2	1 3:P07900	61	43	Heat shock protein HSP 90-alpha	193	138	185	132	184	132	182	127	85006	
5	2	2 3:P06238	54	34	Heat shock protein HSP 90-beta	182	118	175	113	173	112	172	108	83554	
6	2	3 3:P14625	35	33	Endoplasmic OS=Homo sapien	85	76	84	75	80	71	83	74	92696	
7	2	4 3:Q12911	12	11	Heat shock protein 75 kDa, mit	19	16	19	16	23	19	18	13	80345	
8	3	1 2:P05787	52	44	SWISS-PROT:P05787 Tax_id=96	163	137	159	137	157	132	162	137	53671	
9	3	2 3:P08670	46	42	Vimentin OS=Homo sapiens O	133	121	124	113	125	114	132	120	53676	
10	3	3 3:Q3KNV1	36	2	TREMBL:Q3KNV1;Q96GE1 Tax_i	94	5	87	5	93	2	86	3	51411	
11	3	4 2:P08729	35	1	SWISS-PROT:P08729 Tax_id=96	92	3	82	0	94	3	83	0	51443	
12	3	5 3:K7EPT8	7	4	Glial fibrillary acidic protein (Fr	18	7	19	8	20	9	19	7	8373	
13	3	6 2:C6NKH9	6	1	TREMBL:C6NKH9 Tax_id=10090	15	0	13	1	17	0	16	1	55902	
14	3	7 3:K7EPH4	3	1	Glial fibrillary acidic protein (Fr	10	2	10	2	9	0	9	0	14086	
15	3	8 2:Q5XKE5	6	2	SWISS-PROT:Q5XKE5 Tax_id=96	13	1	10	2	13	2	11	1	58059	
16	3	9 2:Q01546	5	1	SWISS-PROT:Q01546 Tax_id=96	12	0	10	1	10	0	10	0	66400	
17	4	1 3:P21333	78	72	Filamin-A OS=Homo sapiens O	189	177	187	174	190	180	192	183	283301	
18	4	2 3:Q75369-8	83	77	isoform 8 of Filamin-B OS=Hom	158	146	158	145	167	157	157	148	283626	
19	5	1 3:P13639	61	60	Elongation factor 2 OS=Homo si	163	163	170	169	169	168	167	166	96246	

MASCOT

: Quantitation

© 2007-2023 Matrix Science

MATRIX SCIENCE

If all is present and correct, a progress dialog is displayed, because calculations can take some time for large numbers of files. The stages are:

- Create a merged report of all search results
- Export the master list of proteins
- For each file, export the expression data for all peptides
- Assign the peptide data to proteins according to the master list
- Calculate protein abundances and ratios as required, including outlier detection
- Write everything to a disk file in TSV format

This is the Quantitation Summary for the label-free data when we choose to merge replicates. Most columns are self-explanatory.

</

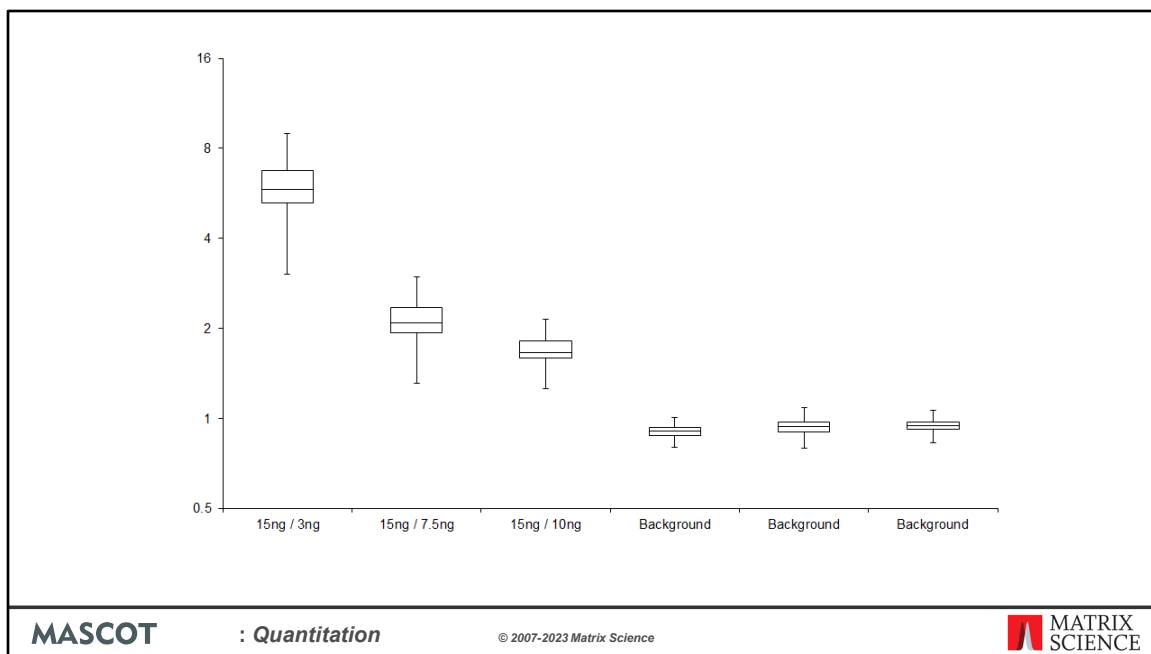
MASCOT

: Quantitation

© 2007-2023 Matrix Science

MATRIX
SCIENCE

Over to the right are columns containing median ratios and total intensity values for each protein. One of the strengths of the Quantitation Summary is that it uses the same rigorous protein inference as the Mascot Protein Family Summary report.



You may be able to get the report you want direct from Excel. For example, this box and whisker plot of the data was produced in Excel.

Sample map - *

(s)	Submitted	Fraction	Intensity 113	Intensity 114	Intensity 115	Intensity 116	Intensity 117	Intensity 118	Intensity 119	Intensity 121
ht_2015...	30/08/2019 10.0	1	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	2	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	3	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	4	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	5	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	6	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	7	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	8	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0	1	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.0	2	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.0	3	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.1	4	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.1	5	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.1	6	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.1	7	A*	I	J	K	L	M	N	O
ht_2015...	30/08/2019 10.1	8	A*	I	J	K	L	M	N	O

Contaminant DB: None | iTRAQ 8plex | Settings ... | Save sample map ... | Save quantitation summary ... | Close

MASCOT

: Quantitation

© 2007-2023 Matrix Science



For a label-free experiment, there is a single column for the sample identifier. For experiments that use isotopic labels, there will be a column for each component specified in the quantitation method. If it was a typical SILAC experiment with two components, light for unlabelled and heavy for labelled, there would be two columns labelled *Intensity light* and *Intensity heavy*. An experiment that uses isobaric tags might have eight or more components.

This is a sample map for 8plex iTRAQ data. There are many ways of conducting such an study. This shows a case where there are 8 fractions for each sample, so the first 8 rows shows the same arrangement of samples, A to H. These fractions will be merged in the Quantitation summary, and A has an asterisk, so there will be columns for ratios to sample A as well as the total intensities for each channel. The second set of rows contains 7 new samples, plus reference sample A.

If the rows were replicates, and not fractions, then using the same channel for a sample across multiple replicates would be missing a trick.

Sample map - *										
(s)	Submitted	Fraction	Intensity 113	Intensity 114	Intensity 115	Intensity 116	Intensity 117	Intensity 118	Intensity 119	Intensity 121
ht_2015...	30/08/2019 10.0...	1	A*	B	C	D	E	F	G	H
ht_2015...	30/08/2019 10.0...	2	H	A*	B	C	D	E	F	G
ht_2015...	30/08/2019 10.0...	3	G	H	A*	B	C	D	E	F
ht_2015...	30/08/2019 10.0...	4	F	G	H	A*	B	C	D	E
ht_2015...	30/08/2019 10.0...	5	E	F	G	H	A*	B	C	D
ht_2015...	30/08/2019 10.0...	6	D	E	F	G	H	A*	B	C
ht_2015...	30/08/2019 10.0...	7	C	D	E	F	G	H	A*	B
ht_2015...	30/08/2019 10.0...	8	B	C	D	E	F	G	H	A*
ht_2015...	30/08/2019 10.0...									
ht_2015...	30/08/2019 10.0...									
ht_2015...	30/08/2019 10.1...									
ht_2015...	30/08/2019 10.1...									
ht_2015...	30/08/2019 10.1...									
ht_2015...	30/08/2019 10.1...									
ht_2015...	30/08/2019 10.1...									
ht_2015...	30/08/2019 10.1...									

Better to rotate the labels, so as to reduce or eliminate systematic errors. Ideally, a so-called Latin Square, where each sample is rotated through all possible tags, as shown here for the first 8 rows. Rows are merged by sample identifier, so that the Quantitation Summary contains the correct ratio and intensity information.

MASCOT

: Quantitation

© 2007-2023 Matrix Science



Having created a Quantitation Summary, what can you do with it? One option is to open it in Perseus, from the Max Planck Institute. This is a good choice if you prefer to manipulate the data using a spreadsheet type of approach. If you are willing to do a bit of scripting, the R language provides access to a huge range of statistical and graphical tools.

Miscellaneous

- Arg-Pro Conversion of SILAC label is supported
- Quantitation can be limited to 'unique' peptides
- Quantitation results, including all intermediate values, can be exported as XML and CSV
- Reporting uses Python 3.x scripts.
- Editable and extendable by Python programmers
- Exporting to CSV will open the report in Excel
- All steps from peak picking to quantitation can be automated using Mascot Daemon

MASCOT

: *Quantitation*

© 2007-2023 Matrix Science



This is a summary of the reporting and quantitation functionality.

Quantitation talks

- K1. Quantitation - Introduction
- K2. Quantitation - MS2 based methods
- K3. Quantitation - MS1 based methods
- K4. Quantitation - Reporting

MASCOT

: *Quantitation*

© 2007-2023 Matrix Science



Please see the other quantitation presentations to learn about reporter ions, SILAC and label free quantitation as well as reporting formats for the results.

- K1. Quantitation – Introduction.
- K2. Quantitation - MS2 based methods. Quantitation methods that only require information available in the MS/MS peak list are supported in Mascot Server.
- K3. Quantitation - MS1 based methods. Methods that require additional information from the raw data file require Mascot Distiller + Quantitation Toolbox.
- K4. Quantitation – Reporting.