

Quantitation Summary: exporting protein expression data for complex experiments

John Cottrell
Matrix Science

- **Biological samples:** 6
- **Time points:** 4
- **Technical replicates:** 3
- **Fractions:** 6

$$6 \times 4 \times 3 \times 6 = 432 \text{ raw files}$$

Studies that use mass spectrometry-based quantitation often contain very large numbers of individual analyses: samples from different sources or treatments or time points, possibly fractionated, with replicates and so forth.

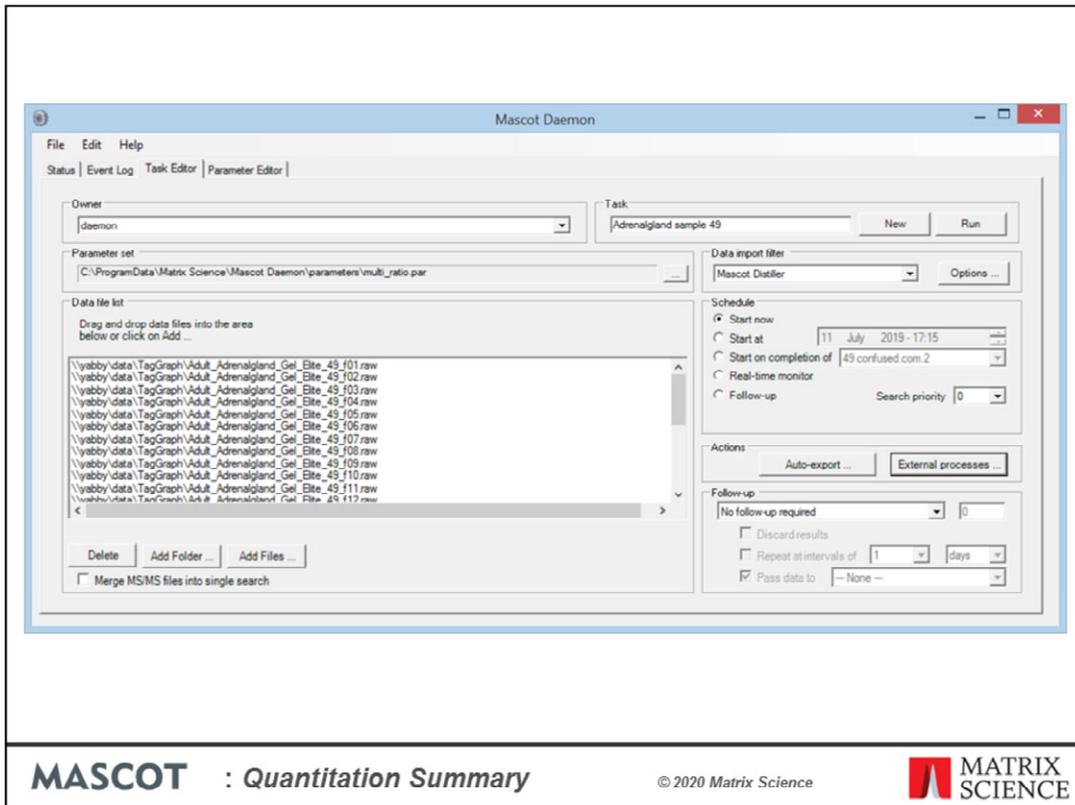
Even a simple study can generate a large number of files. For example, imagine a label-free study of 6 biological samples, 3 control and 3 treated, each of which has been analysed at 4 time points in 3 technical replicates and each replicate has been separated into 6 fractions prior to analysis. This would result in 432 raw files.

	Sample 1 Time 1 Replicate 1	Sample 1 Time 1 Replicate 2	Sample 1 Time 1 Replicate 3	Sample 1 Time 2 Replicate 1	Sample 1 Time 2 Replicate 2	Sample 1 Time 2 Replicate 3	Sample 1 Time 3 Replicate 1
Protein1							
Protein2							
Protein3	Abundance measurements						
Protein4							
Protein5							
Protein6							

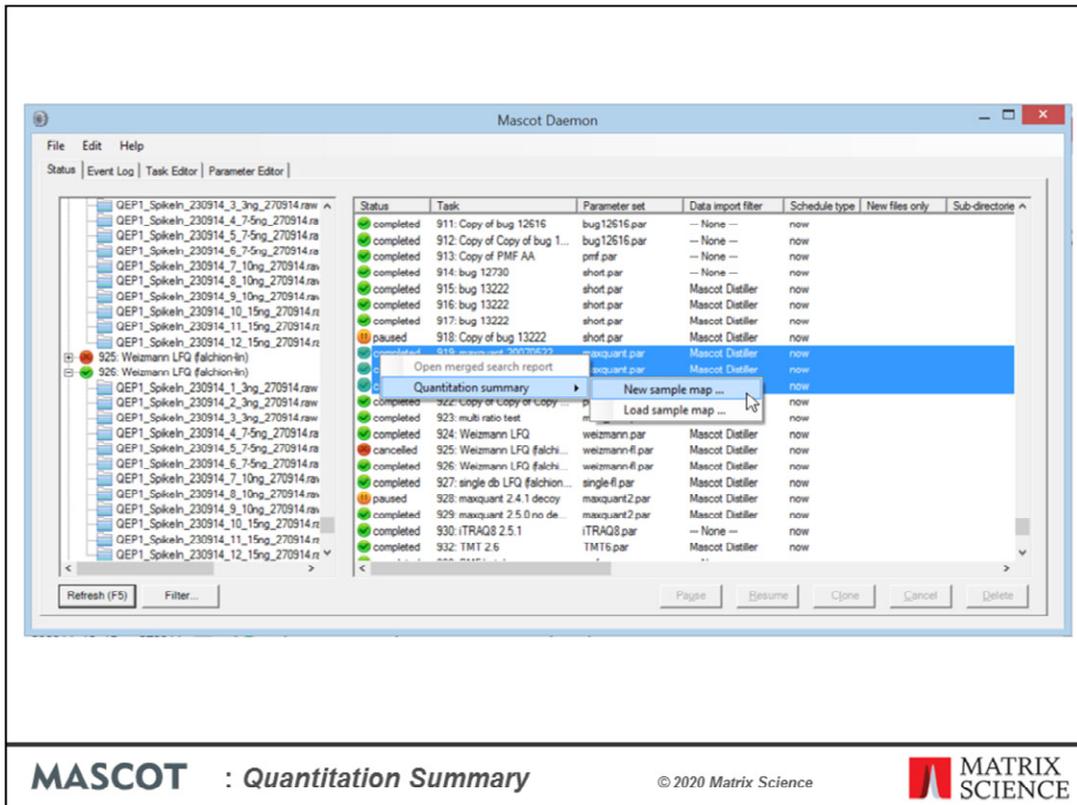
MASCOT : *Quantitation Summary* © 2020 Matrix Science 

Using statistical methods to extract meaningful information, and report it as charts and tables is a complex task that requires custom scripting in a language such as R or specialised software such as Perseus. These take their input in spreadsheet form, called a Quantitation Summary, where the rows correspond to proteins and the columns contain expression data for the various samples in the form of abundances or ratios of abundances.

Until the release of Mascot Server 2.7, we did not have a convenient way to create a Quantitation Summary from individual Mascot Server or Mascot Distiller result files. Mascot Daemon now includes this functionality. Searches run through Daemon that include label or label-free quantitation, including reporter methods such as iTRAQ and TMT, can be combined and annotated to create just such a Quantitation Summary.



The steps are, first, use Daemon to submit the search and initiate quantitation. The analyses can be spread across any number of Daemon tasks, and can include existing results from earlier versions of Daemon. For reporter ion experiments, you can use Distiller for peak picking, but this is not a requirement. For MS1 quantitation methods, such as SILAC and label-free, the raw files must be peak picked and quantified by Mascot Distiller.



MASCOT : Quantitation Summary

© 2020 Matrix Science



Second, you select the relevant tasks and create a Sample Map.

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"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17.2...	1	3ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.0...	2	3ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18.4...	3	3ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_4_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19.2...	1	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_5_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.0...	2	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_6_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20.4...	3	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21.1...	1	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.0...	2	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22.4...	3	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23.2...	1	15ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00.0...	2	15ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\akippy.matroci...	GEP1_SpikeIn_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00.4...	3	15ng

Contaminant DB: Average [MD]:

MASCOT : Quantitation Summary © 2020 Matrix Science 

The Sample Map is used to annotate the list of result files with recognisable sample identifiers. In this case we just need to complete the two right-hand columns. Third, choose ‘Save quantitation summary’

LFQ-merge-yabby-924.txt - Excel

John Cottrell

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
2	1	13:P06733	37	33	Alpha-enolase OS=Homo sapie	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- α	193	138	185	132	184	132	182	127	85006	
5	2	2:3:P08238	54	34	Heat shock protein HSP 90- β	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:Q12931	12	11	Heat shock protein 75 kDa, miti	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	3:2:Q3KNV1	36	2	TREMBL:Q3KNV1;Q596E1 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:Q6NXH9	6	1	TREMBL:Q6NXH9 Tax_Id=10090	15	0	13	1	17	0	16	1	59502	
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:O75369-8	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	

LFQ-merge-yabby-924

MASCOT : Quantitation Summary

© 2020 Matrix Science

MATRIX SCIENCE

And Daemon creates and saves the Quantitation Summary, a tab delimited text file which is the input file for statistical analysis. The best way to explain how this works in detail is with a couple of examples

PRIDE - Proteomics Identification

ebi.ac.uk/pride/archive/projects/PXD001385

Project PXD001385

PRIDE Assigned Tags: **Technical**

Summary

Title
Intensity-based label-free proteomics using a quadrupole orbitrap mass spectrometer

Description
We present a unique data set for benchmarking label free quantitative proteomics using a quadrupole orbitrap mass spectrometer. Soluble Escherichia coli digest was spiked into a HeLa digest in four different concentrations, simulating protein expression differences. The data set, which is available online, provides a unique opportunity to test the proteomic platform (instrumentation and software).

[Read more](#)

Sample Processing Protocol
Four groups of samples, called 3, 7.5, 10 and 15 were prepared in three replicates. The numbers indicate the amount of E. Coli (in nanograms) spiked into 200ng HeLa digestion, which was loaded onto the LC column for each sample. This simulated 5, 2 and 1.5 fold changes relative to the 15ng sample.

Data Processing Protocol
Raw data was imported into the Expressionist data analysis system (GeneSight). Data was filtered, simulated and aligned in retention time. This was followed by feature detection based on peak volume and isotopic clustering. The parameters for all

Properties

Organism
Homo sapiens (human)
Escherichia coli

Organism part
Cell culture

Diseases
Unknown

Modification
monohydroxylated residue
iodoacetamide derivatized residue

Instrument
Q Exactive

Software
Unknown

Experiment Type
Technical

Quantification

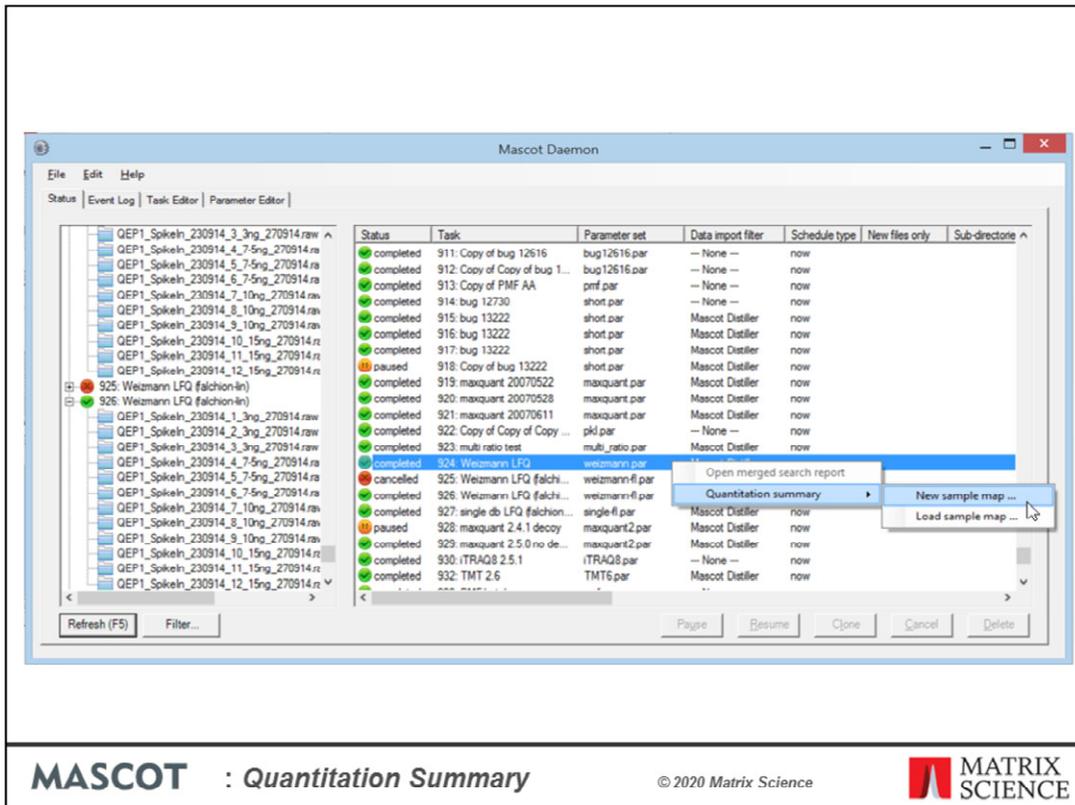
This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#). [I agree, dismiss this banner](#)

MASCOT : Quantitation Summary

© 2020 Matrix Science

MATRIX SCIENCE

The first example is a label-free experiment. A set of 12 raw files was downloaded from PRIDE project PXD001385. According to the project description: "Four groups of samples, called 3, 7.5, 10 and 15 were prepared in three replicates. The numbers indicate the amount of E. Coli (in nanograms) spiked into 200ng HeLa digestion, which was loaded onto the LC column for each sample. This simulated 5, 2 and 1.5 fold changes relative to the 15ng sample."



The files were processed in a single Mascot Daemon task, using Mascot Distiller for peak picking and quantitation. The data were searched against human and E. coli proteomes plus a contaminants database using typical search settings for Q Exactive data. The quantitation method was 'Average [MD]', which is label-free MS1 quantitation for individual files.

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.

✓	Task	Task name	Raw file path	Raw file name	Parameter set	Fasta(s)	Submitted	Fraction	Intensity LFQ
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17:2...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:0...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:4...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_4_7-5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19:2...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_5_7-5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:0...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_6_7-5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:4...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21:1...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:0...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:4...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23:2...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00:0...		
✓	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00:4...		

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

MASCOT : Quantitation Summary

© 2020 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	Fasta(s)	Submitted	Fraction	Intensity LFQ
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17:2...	1	3ng*
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:0...	2	3ng*
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:4...	3	3ng*
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_4_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19:2...	1	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_5_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:0...	2	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_6_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:4...	3	7.5ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21:1...	1	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:0...	2	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:4...	3	10ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23:2...	1	15ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_Spikeln_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00:0...	2	15ng
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	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 Summary © 2020 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.

The screenshot shows a window titled "Sample map - *". It contains a table with the following columns: Task, Task name, Raw file path, Raw file name, Parameter set, Fasta(s), Submitted, Fraction, and Intensity LFQ. The table lists 15 rows of sample data, each with a checked checkbox in the first column. Below the table, there are controls for "Contaminant DB" (set to "None"), "Average [MD]" (set to "Average [MD]"), and buttons for "Settings ...", "Save sample map ...", "Save quantitation summary ...", and "Close".

<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"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_1_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 17:2...		3ng_rep1
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_2_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:0...		3ng_rep2
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_3_3ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 18:4...		3ng_rep3
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_4_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 19:2...		7.5ng_rep1
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_5_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:0...		7.5ng_rep2
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_6_7.5ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 20:4...		7.5ng_rep3
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_7_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 21:1...		10ng_rep1
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_8_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:0...		10ng_rep2
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_9_10ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 22:4...		10ng_rep3
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_10_15ng_270914.raw	weizmann.par	ecoli_proteome_...	06/08/2019 23:2...		15ng_rep1
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_11_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00:0...		15ng_rep2
<input checked="" type="checkbox"/>	924	Weizmann LFQ	\\skippy.mattsci...	GEP1_SpikeIn_230914_12_15ng_270914.raw	weizmann.par	ecoli_proteome_...	07/08/2019 00:4...		15ng_rep3

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
1	1	1 3::P06733	37	33	Alpha-enolase OS=Homo sapie	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- α	193	138	185	132	184	132	182	127	85006	
5	2	2 3::P08238	54	34	Heat shock protein HSP 90- β	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::Q12931	12	11	Heat shock protein 75 kDa, miti	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 2::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::Q6NXH9	6	1	TREMBL:Q6NXH9 Tax_Id=10090	15	0	13	1	17	0	16	1	59502	
14	3	7 3::K7EP14	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::O75369-8	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 Summary

© 2020 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

LFQ-merge-yabby-924.txt - Excel

John Cottrell

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
	Ratio [7.5ng / 3ng]	Ratio variabilit y [%] [7.5ng / 3ng]	Ratio count [7.5ng / 3ng]	Ratio type [7.5ng / 3ng]	Ratio [10ng / 3ng]	Ratio variabilit y [%] [10ng / 3ng]	Ratio count [10ng / 3ng]	Ratio type [10ng / 3ng]	Ratio [15ng / 3ng]	Ratio variabilit y [%] [15ng / 3ng]	Ratio count [15ng / 3ng]	Ratio type [15ng / 3ng]	Intensity [3ng]	Intensity [7.5ng]	Intensity [10ng]	Intensity [15ng]	Poten contant	
1	0.964398	1.23323	114	median	0.937089	1.166451	111	median	0.904996	1.175257	110	median	1.77E+10	4.46E+09	4.49E+09	4.51E+09	4.25E+09	
2	0.938147	1.651989	17	median	0.933939	1.117947	13	median	0.926565	2.716981	18	median	2.45E+09	6.7E+08	6.01E+08	6.2E+08	5.56E+08	
3	0.954782	1.200489	152	median	0.949287	1.152321	150	median	0.896981	1.128214	153	median	9.86E+09	2.59E+09	2.44E+09	2.49E+09	2.35E+09	
4	0.963448	1.205042	139	median	0.953098	1.120576	145	median	0.910348	1.098493	140	median	1.15E+10	2.97E+09	2.89E+09	2.86E+09	2.72E+09	
5	0.959232	1.12963	68	median	0.958475	1.069606	61	median	0.891961	1.092244	66	median	1.66E+09	4.4E+08	4.23E+08	4.17E+08	3.85E+08	
6	0.993295	1.12617	14	median	0.968	1.103151	14	median	0.91096	1.108296	13	median	1.21E+09	3.19E+08	3.07E+08	3.03E+08	2.84E+08	
7	0.941643	1.28903	121	median	0.943693	1.390515	127	median	0.901383	1.196093	130	median	1.35E+10	3.62E+09	3.4E+09	3.35E+09	3.15E+09 +	
8	0.952834	1.159354	104	median	0.965514	1.104482	105	median	0.901746	1.091268	107	median	4.38E+09	1.13E+09	1.16E+09	1.06E+09	1.03E+09	
9	0.959542	1.182998	61	median	0.935061	1.155986	70	median	0.908992	1.131608	68	median	3E+09	8.02E+08	7.08E+08	7.71E+08	7.21E+08 +	
10	0.955059	1.185466	57	median	0.935715	1.151932	71	median	0.907494	1.133647	65	median	2.96E+09	7.91E+08	6.93E+08	7.66E+08	7.14E+08 +	
11	0.846192	3.984523	16	median	1.093249	3.081166	16	median	0.864921	3.618244	17	median	3.6E+08	92390477	67282502	1.04E+08	97078292	
12	0.859335	2.64959	9	median	0.933901	2.129753	11	median	0.815931	1.130961	9	median	5.58E+08	1.5E+08	1.08E+08	1.35E+08	1.65E+08 +	
13	0.686559	3.005479	7	median	0.985981	2.420498	8	median	0.879415	2.761304	8	median	2.51E+08	70722576	46660789	60733001	72825656	
14	0.896751	1.291621	6	median	0.825243	1.180451	9	median	0.920102	1.189858	10	median	1.24E+09	2.99E+08	2.99E+08	3.46E+08	2.96E+08 +	
15	0.998271	1.204713	6	median	1.028609	1.196913	7	median	0.899796	1.177787	9	median	8.97E+08	2.23E+08	2.32E+08	2.15E+08	2.27E+08 +	
16	0.954127	1.136827	144	median	0.94622	1.134982	142	median	0.899993	1.10823	143	median	2.3E+09	6.1E+08	5.7E+08	5.76E+08	5.46E+08	
17	0.966112	1.184959	112	median	0.956142	1.168124	115	median	0.908647	1.118169	114	median	1.32E+09	3.49E+08	3.31E+08	3.23E+08	3.15E+08	
18	0.953137	1.165342	119	median	0.961408	1.13252	128	median	0.909731	1.114404	125	median	6.42E+09	1.65E+09	1.66E+09	1.6E+09	1.51E+09	

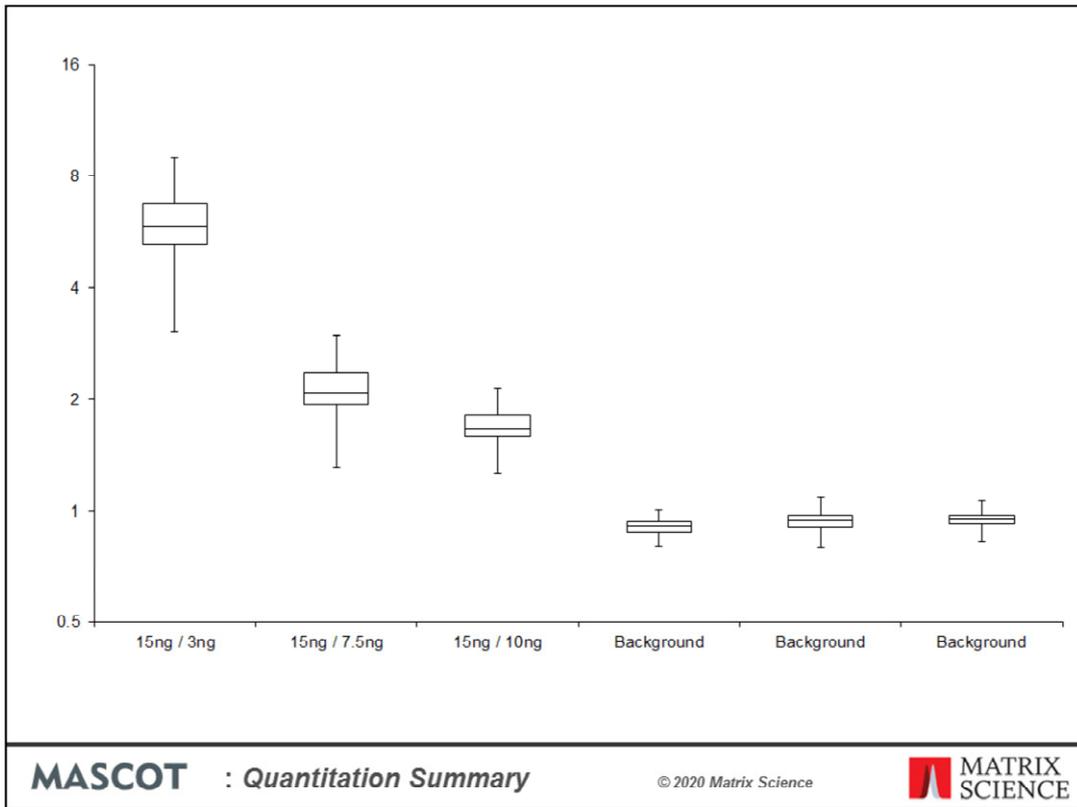
LFQ-merge-yabby-924

MASCOT : Quantitation Summary

© 2020 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.

(s)	Submitted	Fraction	Intensity 113	Intensity 114	Intensity 115	Intensity 116	Intensity 117	Intensity 118	Intensity 119	Intensity 121
rot_2015_...	30/08/2019 10.0...	1	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	2	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	3	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	4	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	5	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	6	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	7	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	8	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	1	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.0...	2	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.0...	3	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.1...	4	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.1...	5	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.1...	6	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.1...	7	A*	I	J	K	L	M	N	O
rot_2015_...	30/08/2019 10.1...	8	A*	I	J	K	L	M	N	O

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.

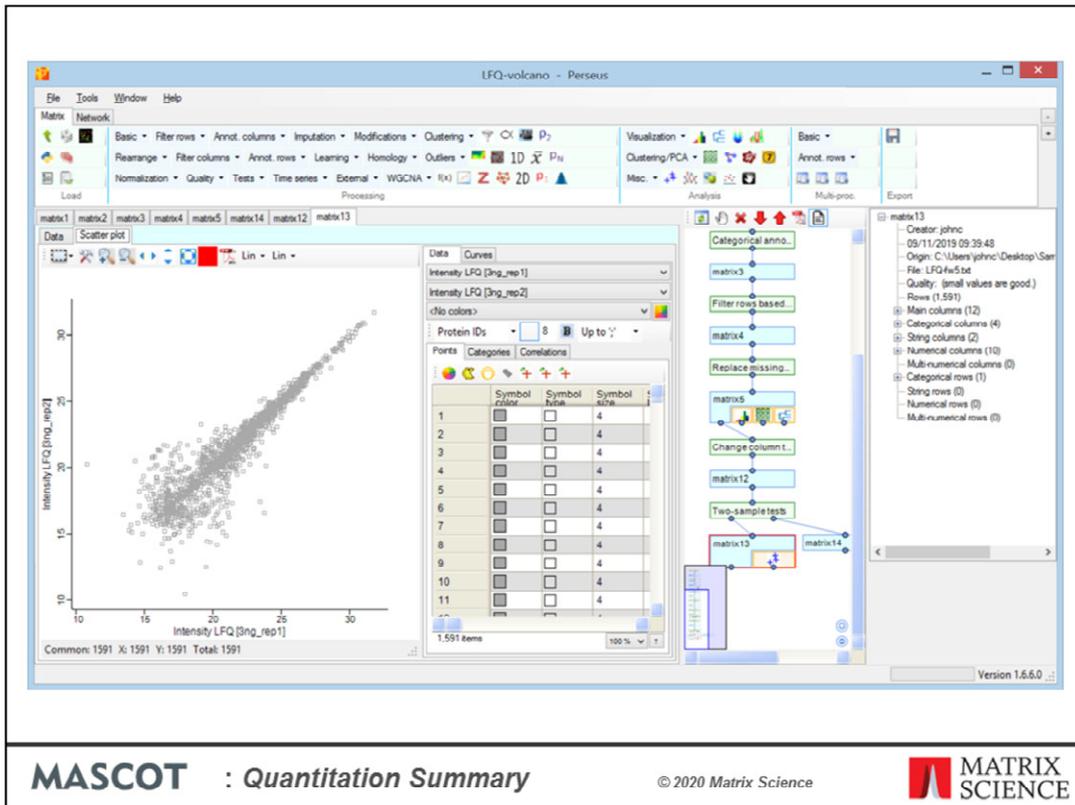
Sample map - *

(s)	Submitted	Fraction	Intensity 113	Intensity 114	Intensity 115	Intensity 116	Intensity 117	Intensity 118	Intensity 119	Intensity 121
rot_2015_...	30/08/2019 10.0...	1	A*	B	C	D	E	F	G	H
rot_2015_...	30/08/2019 10.0...	2	H	A*	B	C	D	E	F	G
rot_2015_...	30/08/2019 10.0...	3	G	H	A*	B	C	D	E	F
rot_2015_...	30/08/2019 10.0...	4	F	G	H	A*	B	C	D	E
rot_2015_...	30/08/2019 10.0...	5	E	F	G	H	A*	B	C	D
rot_2015_...	30/08/2019 10.0...	6	D	E	F	G	H	A*	B	C
rot_2015_...	30/08/2019 10.0...	7	C	D	E	F	G	H	A*	B
rot_2015_...	30/08/2019 10.0...	8	B	C	D	E	F	G	H	A*
rot_2015_...	30/08/2019 10.0...									
rot_2015_...	30/08/2019 10.0...									
rot_2015_...	30/08/2019 10.0...									
rot_2015_...	30/08/2019 10.1...									
rot_2015_...	30/08/2019 10.1...									
rot_2015_...	30/08/2019 10.1...									
rot_2015_...	30/08/2019 10.1...									

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

MASCOT : Quantitation Summary © 2020 Matrix Science 

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 Summary

© 2020 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

Bioconductor - BioViews
 bioconductor.org/packages/devel/BiocViews.html#_Proteomics

Bioconductor
 OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About

Home » BioViews

All Packages

Bioconductor version 3.11 (Development) **Packages found under Proteomics:**

Rank based on number of downloads: lower numbers are more frequently downloaded.

Developers: check this box to toggle the visibility of childless biocViews. Show entries Search table:

Autocomplete biocViews search:

Package	Maintainer	Title	Rank
limma	Gordon Smyth	Linear Models for Microarray Data	12
ProtGenerics	Laurent Gatto	S4 generic functions for Bioconductor proteomics infrastructure	44
pathview	Weijun Luo	a tool set for pathway based data integration and visualization	66
mzB	Steffen Neumann, Laurent Gatto, Qiang Kou	parser for netCDF, mzXML, mzData and mzML and mzIdentML files (mass spectrometry data)	97
MSnbase	Laurent Gatto	Base Functions and Classes for Mass Spectrometry and Proteomics	101
mixOmics	Kim-Anh Le Cao	Omics Data Integration Project	102
mzID	Laurent Gatto	An mzIdentML parser for R	107
MassConeq	Ben Du	Mass spectrum processing by	135

MASCOT : Quantitation Summary © 2020 Matrix Science **MATRIX SCIENCE**

If you are willing to do a bit of scripting, the R language provides access to a huge range of statistical and graphical tools. Bioconductor is a collection of packages for genomic and proteomic applications. Currently, 135 packages are indexed under proteomics and 91 under mass spectrometry.

Home » Bioconductor 3.10 » Software Packages » DEP

DEP

platforms all rank 294 / 1823 posts 1 / 0 / 2 / 0 In Bioc 2.5 years
 build ok updated before release dependencies 150

DOI: [10.18129/B9.bioc.DEP](https://doi.org/10.18129/B9.bioc.DEP)  

Differential Enrichment analysis of Proteomics data

Bioconductor version: Release (3.10)

This package provides an integrated analysis workflow for robust and reproducible analysis of mass spectrometry proteomics data for differential protein expression or differential enrichment. It requires tabular input (e.g. txt files) as generated by quantitative analysis softwares of raw mass spectrometry data, such as MaxQuant or IsobarQuant. Functions are provided for data preparation, filtering, variance normalization and imputation of missing values, as well as statistical testing of differentially enriched / expressed proteins. It also includes tools to check intermediate steps in the workflow, such as normalization and missing values imputation. Finally, visualization tools are provided to explore the results, including heatmap, volcano plot and barplot representations. For scientists with limited experience in R, the package also contains wrapper functions that entail the complete analysis workflow and generate a report. Even easier to use are the interactive Shiny apps that are provided by the package.

Author: Arne Smits [cre, aut], Wolfgang Huber [aut]

Maintainer: Arne Smits <smits.arne@gmail.com>

Citation (from within R, enter `citation("DEP")`):

Zhang X, Smits A, van Tilburg G, Ovaa H, Huber W, Vermeulen M (2018). "Proteome-wide identification of ubiquitin interactions using UbIA-MS." *Nature Protocols*, 13, 530-550.

Documentation »

Bioconductor

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

R / [CRAN](#) packages and [documentation](#)

Support »

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers

MASCOT : Quantitation Summary

© 2020 Matrix Science



I'll use a package called DEP to illustrate the types of analysis that can be achieved with a few lines of scripting.

The image shows a screenshot of a web browser displaying a Nature Oncogene article. The browser's address bar shows the URL [nature.com/articles/onc2016242](https://www.nature.com/articles/onc2016242). The article title is "microRNAs with AAGUGC seed motif constitute an integral part of an oncogenic signaling network". The authors listed are Y Zhou, O Frings, R M Branca, J Boekel, C le Sage, E Fredlund, R Agami & L M Orre. The article was published on 01 August 2016. It has 413 accesses, 6 citations, and 2 altmetric metrics. The abstract begins with "microRNA (miRNA) dysregulation is a common feature of cancer cells, but the complex roles of miRNAs in cancer are not". The page includes a "Download PDF" button and a table of contents with sections like Abstract, Introduction, Results, Discussion, Materials and methods, Accession codes, References, Acknowledgements, Author information, and Ethics declarations. At the bottom of the page, there is a MASCOT logo with the text "Quantitation Summary", a copyright notice "© 2020 Matrix Science", and the Matrix Science logo.

The data comes from this study to identify oncogenic microRNAs in non-small cell lung cancer. Quantitation used 10plex TMT

PRIDE - Proteomics Identification

ebi.ac.uk/pride/archive/projects/PXD004163

Project PXD004163

PRIDE Assigned Tags: Biological Biomedical

Summary

Title
Proteomics of U1810 cells upon treatment with microRNAs with an AAGUGC seed motif.

Description
microRNA dysregulation is a common feature of cancer cells, but the complex roles of microRNAs in cancer are not fully elucidated. Here we used functional genomics to identify oncogenic microRNAs in non-small cell lung cancer and to evaluate their impact on response to EGFR targeting therapy. Our data demonstrate that microRNAs with an AAGUGC-motif in their seed-sequence increase both cancer cell proliferation and sensitivity to EGFR inhibitors. Global transcriptomics, proteomics and target prediction resulted in the identification of several tumor suppressors involved in the G1/S transition as targets of AAGUGC-microRNAs. The clinical implications of our findings were evaluated by analysis of public domain data supporting the link between this microRNA seed-family, their tumor suppressor targets and cancer cell proliferation. In conclusion we propose that AAGUGC-microRNAs are an integral part of an oncogenic signaling network, and that these findings have potential therapeutic implications, especially in selecting patients for

Properties

Organism
Homo sapiens (human)

Organism part
Epithelial cell
Lung

Diseases
Non-small cell lung carcinoma

Modification
monohydroxylated residue
Iodoacetamide derivatized residue

Instrument
Q Exactive

Software
Unknown

Experiment Type
Proteomics

MASCOT : *Quantitation Summary* © 2020 Matrix Science **MATRIX SCIENCE**

72 files downloaded from PRIDE project PXD004163 were processed and searched using Mascot Daemon, as described earlier.

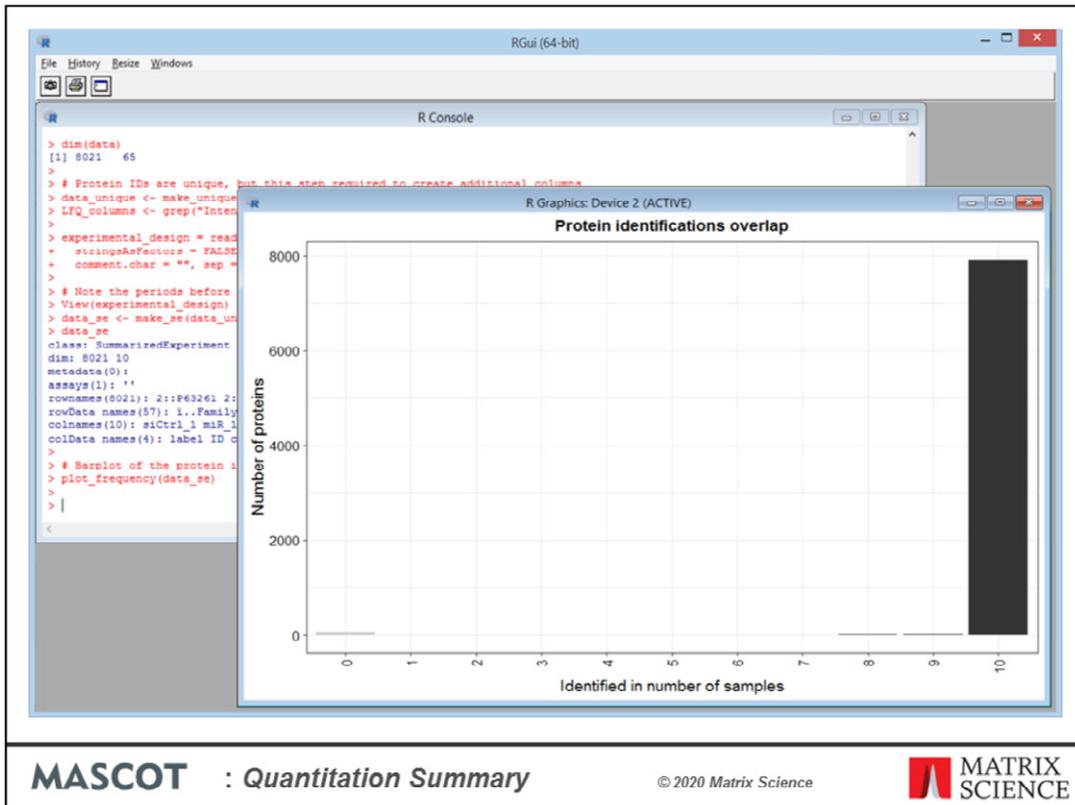
Sample map - PXD004163

Fraction	Intensity 126	Intensity 127N	Intensity 127C	Intensity 128N	Intensity 128C	Intensity 129N	Intensity 129C	Intensity 130N	Intensity 130C	Intensity 131
1.2. 1	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.3. 2	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.3. 3	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.4. 4	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.1. 5	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.2. 6	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.4. 7	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.5. 8	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.0. 9	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.1. 10	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.3. 11	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.4. 12	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.5. 13	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.0. 14	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.2. 15	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C
1.3. 16	sCtrl_A'	mR_191_B	mR_372_A	mR_519c_A	sCtrl_B	mR_372_B	mR_519c_B	sCtrl_C	mR_191_A	mR_372_C

Contaminant DB: contaminants | TMT 10plex | Settings... | Save sample map... | Save quantitation summary... | Close

MASCOT : Quantitation Summary © 2020 Matrix Science 

The Sample Map looks like this. 3 replicates of the control and one of the microRNA treatments, 2 replicates of the other two treatments. Peptide FDR was set to 1% by target/decoy.

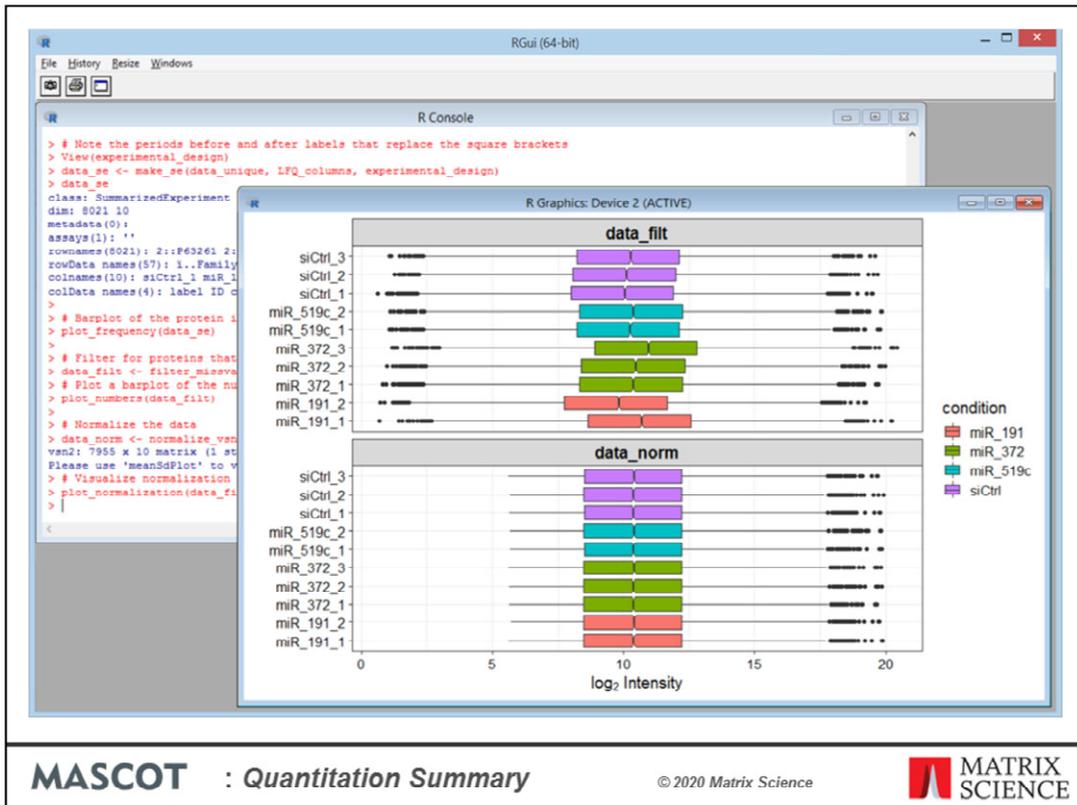


MASCOT : Quantitation Summary

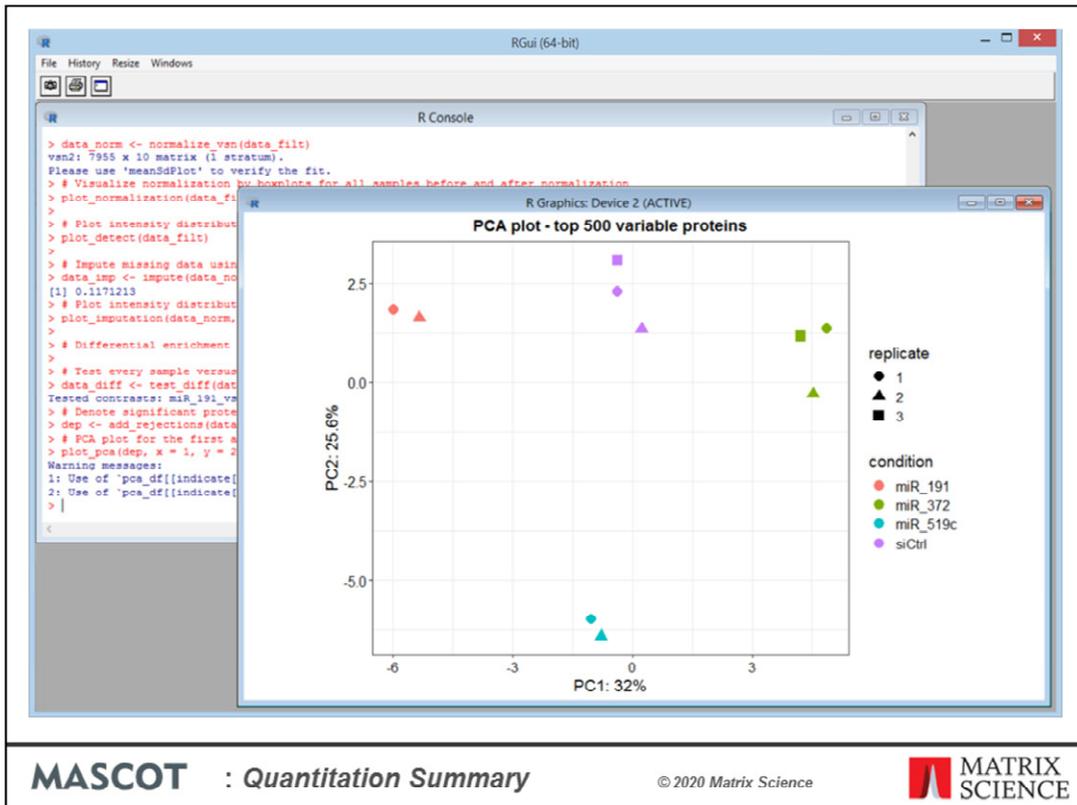
© 2020 Matrix Science



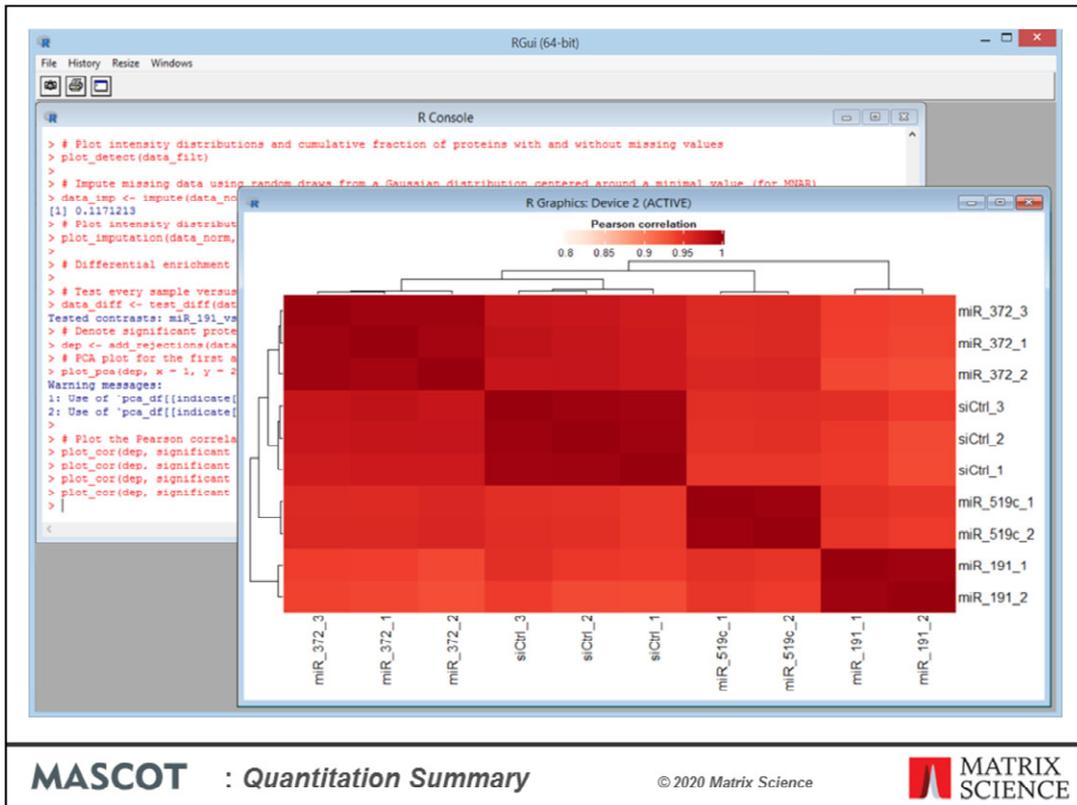
Using DEP, we can very easily create a number of informative charts. Some are for QC, such as this one, which shows we have data for almost all 8021 proteins across all 10 channels – very few missing values.



A box plot showing the intensities before and after normalisation



PCA shows the replicates cluster nicely



MASCOT : Quantitation Summary

© 2020 Matrix Science



A heat map for sample to sample similarity

Mascot Daemon Help

Contents | Index | Search

- Introduction
- Getting Started
- Tutorial
- Getting Help
- Troubleshooting
- In Depth
 - Quantitation Summary
 - Quantitation Summary f
 - Quantitation Summary f
 - Data Import Filters
 - Running Daemon as a sen
 - Stop Masses
- Reference
 - Contacting Matrix Science

Creating or loading a Sample Map

On the [Status tab](#) in Mascot Daemon, select the root node of the tree (Task Database) so that the list view on the right shows a list of tasks.

To open an existing Sample Map, right click any task in the list view and choose **Quantitation summary: Load sample map ...**

To create a new Sample Map for a study, select all the tasks used to process the data in the study. If there are many tasks in the database, you may wish to filter the list by keyword or owner or range so as to narrow down the possibilities. One or more tasks can be selected in the usual way: left mouse click to select one task, shift-left click to extend the selection to a range, control+left click to add or remove individual tasks. When all the tasks have been selected, right click and choose **Quantitation summary: New sample map ...**

Note: You must select the root node of the tree (on the left) and then select tasks in the list view (on the right). You cannot select tasks by clicking on tasks in the tree.

Ideally, all the tasks selected for a sample map will use the same set of search parameters which reference the same quantitation method. This ensures that there will be no problems with merging results to obtain a master list of proteins or with creating a uniform set of quantitation components and ratios. This is not strictly enforced at the time the sample map is created because the parameter set or the quantitation method could have been edited between tasks or maybe different tasks used different sets of parameters, but the important settings were the same. If you create a map from tasks that are not compatible, the error may only come to light during creation of the Quantitation Summary.

Editing the Sample Map

The cells of the Sample Map can be edited, copied and pasted in much the same way as a spreadsheet. That is, you can select a single cell or a range of cells, copy it to the Windows clipboard, and paste it. If you select a range of cells for a paste operation, it must be some multiple of the data on the clipboard. For example, if you copied a 2 x 2 range, you could paste it to 2 x 4 or 4 x 2 but not 2 x 3. There is a context menu for copy, paste and delete, or you can use the usual shortcut keys (CTRL+C, CTRL+V, DEL)

The first column contains checkboxes that can be used to de-select individual files, so that they will be excluded from the Quantitation Summary. The checkbox in the header functions to select all / select none.

The next 7 columns contain information that can be used to sort and group files to facilitate annotation. Click on the column header to sort on the column and click a second time to reverse the sort order. The contents of these columns do not appear in the Quantitation Summary, they are simply to aid annotation. If you wish to use the task name or raw file name as the basis for the sample identifier, copy the relevant cells and paste into the expression columns.

If a sample is separated into fractions, each fraction gets the same identifier(s) except for the number in the fraction column. If a sample is not fractionated, the fraction column is left empty, but in this case, the identifier(s) must be unique. To avoid typing individual fraction numbers, select a range of cells in the fraction column, right click in the range, and choose **Fill with integer series**

Tip: If you want to merge technical replicates, you can pretend they are fractions of a single sample

You can add columns to document the map or to help organise the data. Right click anywhere in the table and choose **Add column**. To edit the

MASCOT : Quantitation Summary © 2020 Matrix Science

As always, detailed help and reference material for the Sample Map and Quantitation Summary can be found in the Mascot Daemon help file