

LFQ for the masses

A Mascot Distiller LFQ tutorial
Richard Jacob

MASCOT : LFQ

© 2021 Matrix Science

 **MATRIX
SCIENCE**

Hello this is Richard Jacob for Matrix Science and my talk is LFQ for the masses. It's a Mascot Distiller label free quantitation tutorial.

Label free quantitation

- **Extracted ion chromatograms (XICs) of the precursor peaks**
- **Identification led quantitation**
- **Two Mascot Distiller quantitation methods**
 - Average
 - Replicate
- **Improvements to the Replicate method**
 - Performance
 - Elution time alignment

MASCOT : Lfq

© 2021 Matrix Science



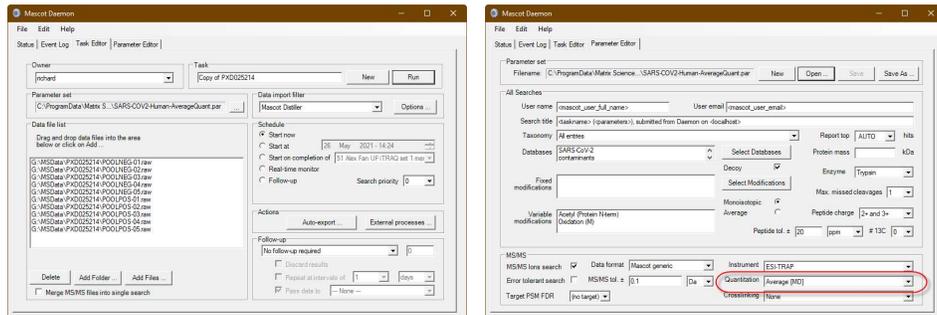
Within the Mascot software ecosystem we can perform Label free quantitation using several different methods. Today I won't be talking about spectra counting or emPAI methods, just those that use extracted ion chromatograms or XIC's of the precursor peaks.

In Mascot we perform identification led quantitation. This means that after peptide identification we go back to the raw data to calculate the XIC's. That way we don't waste time quantitating peaks that might be present in the data but remain unidentified.

There are two Mascot Distiller label free quantitation methods that we can use: the Average method and the Replicate method.

There have been some improvements to label free quantitation in Mascot Distiller release 2.8, particularly for the Replicate method. Overall Mascot Distiller 2.8 is faster than earlier versions and there have been a lot of general improvements to the application. Please go and watch the "New features in Mascot Distiller 2.8" talk to find out more. Updates to Mascot Distiller are currently free so I recommend installing the latest version to take advantage of these improvements.

LFQ through Mascot Daemon and a Quantitation Summary report



Export of results detailed in last years Quantitation Summary presentation https://www.matrixscience.com/workshop_2020.html

MASCOT : LFQ

© 2021 Matrix Science

MATRIX SCIENCE

Let's start by processing a data set through Mascot Daemon to create a Quantitation Summary table or report. This feature was introduced in Mascot Server/Daemon 2.7 so I will be brief.

Set up the Mascot Daemon task as per normal. When configuring the Mascot Server search parameters make sure you set the quantitation method to "Average [MD]". That is very important! We want to quantitate the files independently and summarize the results.

When you start the task Mascot Daemon will process raw files and search the peak lists serially. The search results are then returned to Distiller and quantified.

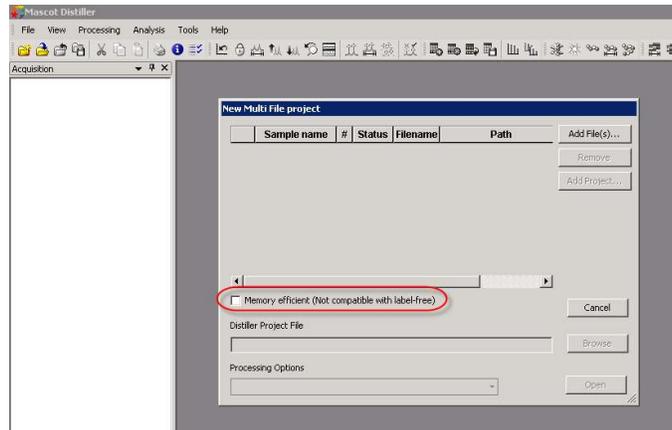
Once the task is complete, we can start the quantitation results export process which is detailed in last years **Quantitation Summary** presentation at <https://www.matrixscience.com/pdf/2020WKSHP4.pdf>.

MASCOT : LFQ © 2021 Matrix Science

MATRIX SCIENCE

You can open the file in a text editor like notepad but it makes more sense to view it as a spreadsheet in Excel. For further statistical analysis you can use Excel, R, Perseus or other software. Again, previous presentations and blog posts have some examples.

LFQ in Mascot Distiller 2.7 and earlier



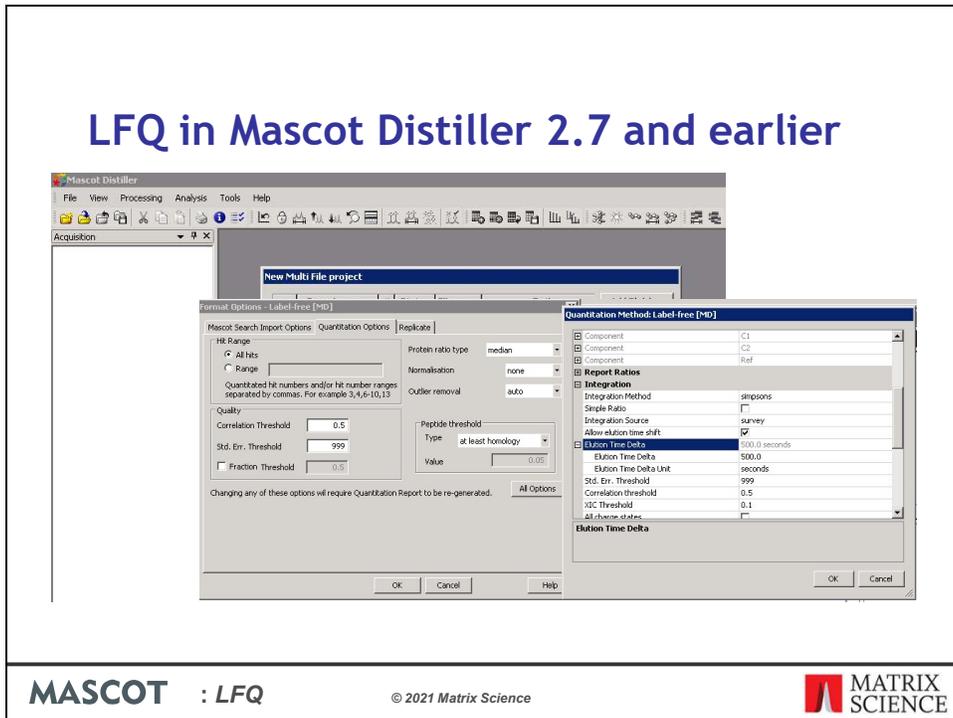
MASCOT : LFQ

© 2021 Matrix Science

MATRIX
SCIENCE

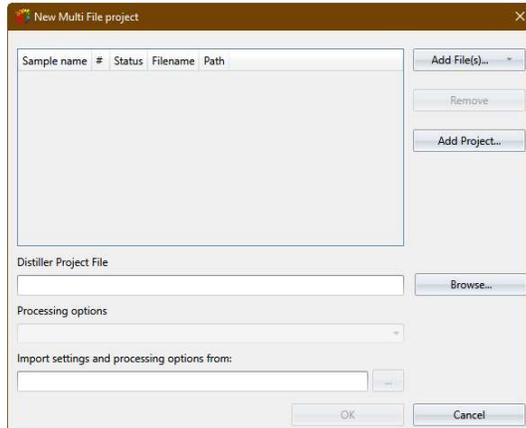
Now we are going to look at processing data in Mascot Distiller directly using multi file projects and the replicate analysis. With Mascot Distiller version 2.7 and earlier a multi-file project was created, and the “Memory Efficient” box had to be left unchecked. Mascot Distiller would open all the files in one big project and process them together. For the search and quantitation steps you must use a quantitation method that of the type “Replicate”. Depending on the number of files in the data set this could use quite a bit of RAM.

LFQ in Mascot Distiller 2.7 and earlier



When it came to the quantitation Distiller would use a global elution time delta to look for peptides identified in at least one of the files in a data set but not in others.

LFQ in Mascot Distiller 2.8



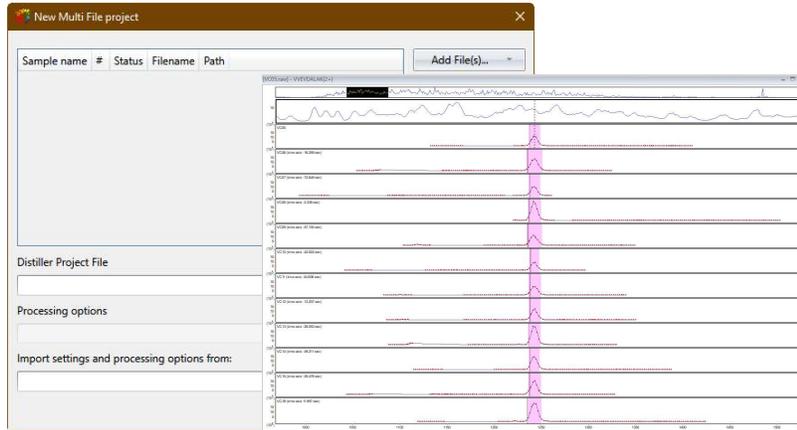
MASCOT : LFQ

© 2021 Matrix Science

MATRIX
SCIENCE

In Mascot Distiller 2.8 we now process all the files independently. The Memory Efficient check box has been removed as well as the Elution Time Delta variable which has been replaced by a new global time-alignment algorithm. This algorithm builds a consensus dataset by aligning the Total Ion Chromatogram (TIC) of each raw file and carrying out a rough alignment. This is then refined with further calculations to create a time shift for the individual file at different retention times and m/z charge states. Files are then compared using the consensus data set as an intermediate. The resulting algorithm can process the data sets in parallel which is faster and more memory efficient than the older approach.

LFQ in Mascot Distiller 2.8



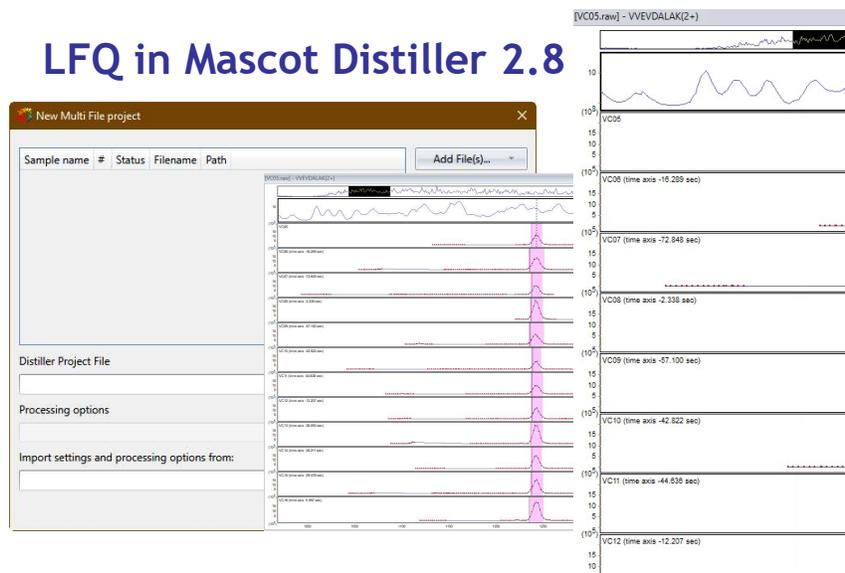
MASCOT : LFQ

© 2021 Matrix Science

MATRIX
SCIENCE

You can see the time shift values in the XIC window for a quantitated peptide.

LFQ in Mascot Distiller 2.8



MASCOT : LFQ

© 2021 Matrix Science

MATRIX
SCIENCE

You can see the time shift values in the XIC window for a quantitated peptide. Zooming in we can see the difference reported in seconds for each file.

PXD026930 - Impact of Alanyl-tRNA Synthetase Editing Deficiency in Yeast

- **Experiment compared 3 yeast strains**
 - WT, C719A, and G906D
- **At two temperatures**
 - 30°C and 37°C
- **With three biological replicates**
 - Total of 18 analyses
- **WT vs C719A strain at 37°C featured in the publication**

MASCOT : Lfq

© 2021 Matrix Science

 MATRIX SCIENCE

The rest of the talk I am going to use a data set that I obtained from the PRIDE repository. The data set I selected is from a publication that looks at the role of alanyl-tRNA synthetases in *S. cerevisiae*. Aminoacyl-tRNA synthetases are essential enzymes linked with neurological disorders in humans. The publication shows that the mutations have more general effects in yeast on the amino acid control pathway and heatshock response.

The initial experiment looked at three yeast strains, two of which have mutations in alanyl-tRNA synthetases.

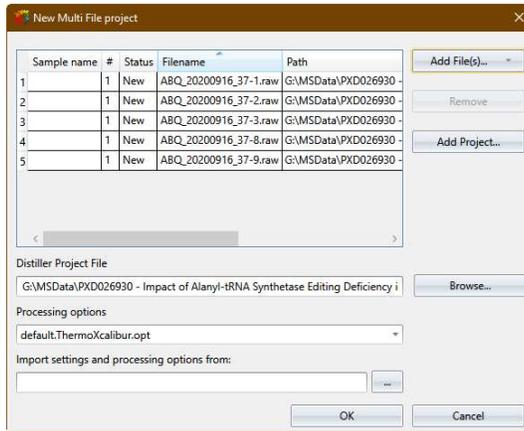
The cultures were grown at 30°C, sampled and then increased to 37°C for 2 hours before sampling again.

Each strain and temperature condition were sampled three times as biological replicates for a total of 18 MS analysis.

Rather than process the full data set I selected just the files for one a comparison of Wild Type to the C719A mutant strain at 37°C as this was featured in the publication.

Note that at this point I created a replicate quantitation method for the analysis on the Mascot Server ready for searching and quantitation.

Mascot Distiller LFQ replicate analysis



MASCOT : LFQ

© 2021 Matrix Science



We can set up label free quantitation in Mascot Distiller as a multi file project.

Mascot Distiller LFQ replicate analysis

The screenshot displays the Mascot Distiller software interface. On the left, a 'New Multi File project' window shows a table with the following data:

Sample name	#	Status	Filename
1	1	New	ABQ_20200916_37-9
2	1	New	ABQ_20200916_37-8
3	1	New	ABQ_20200916_37-3
4	1	New	ABQ_20200916_37-2
5	1	New	ABQ_20200916_37-1

Below the table, the 'Distiller Project File' is set to 'G:\MSData\PX0026930 - Impact of Alan...', and the 'Processing options' are set to 'default.ThermoXcalibur.opt'. The main window shows a 'Total Ion Chromatogram (TIC)' plot with a y-axis labeled 'TIC' and an x-axis labeled 'Time (min)'. The plot shows a noisy baseline that rises significantly starting around 1000 minutes, reaching a plateau of approximately 4000 units by 2000 minutes, and remains relatively stable with high-frequency noise until 6000 minutes. The x-axis has major ticks at 0, 1000, 2000, 3000, 4000, 5000, and 6000. The software interface includes a menu bar (File, Edit, View, Processing, Analysis, Tools, Windows, Help) and a toolbar with various icons for file operations and analysis.

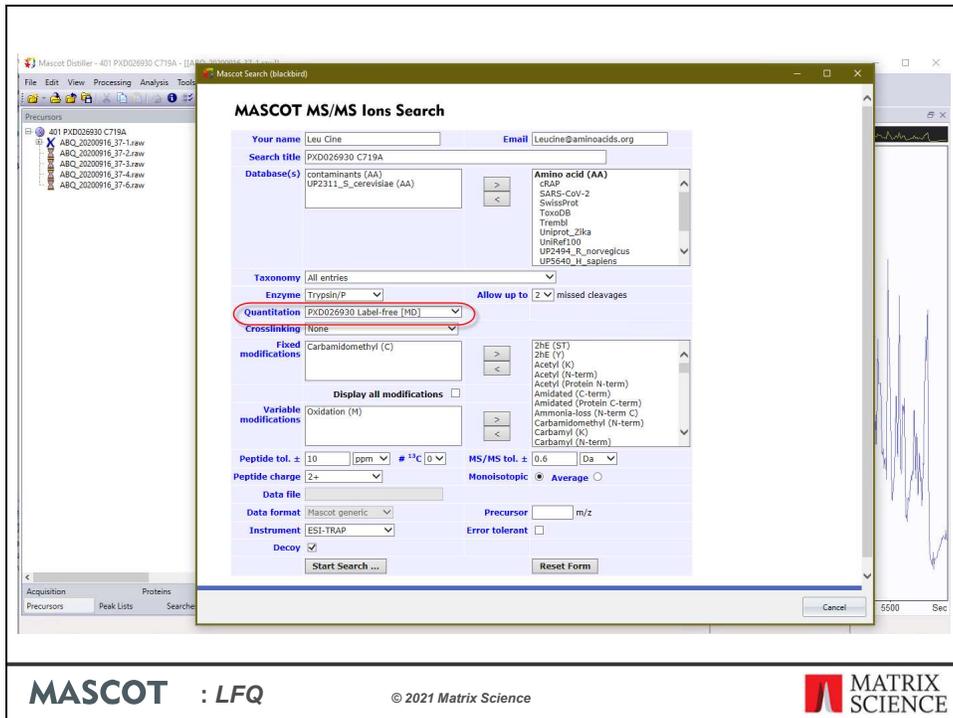
After the data set is open in Mascot Distiller

Mascot Distiller LFQ replicate analysis

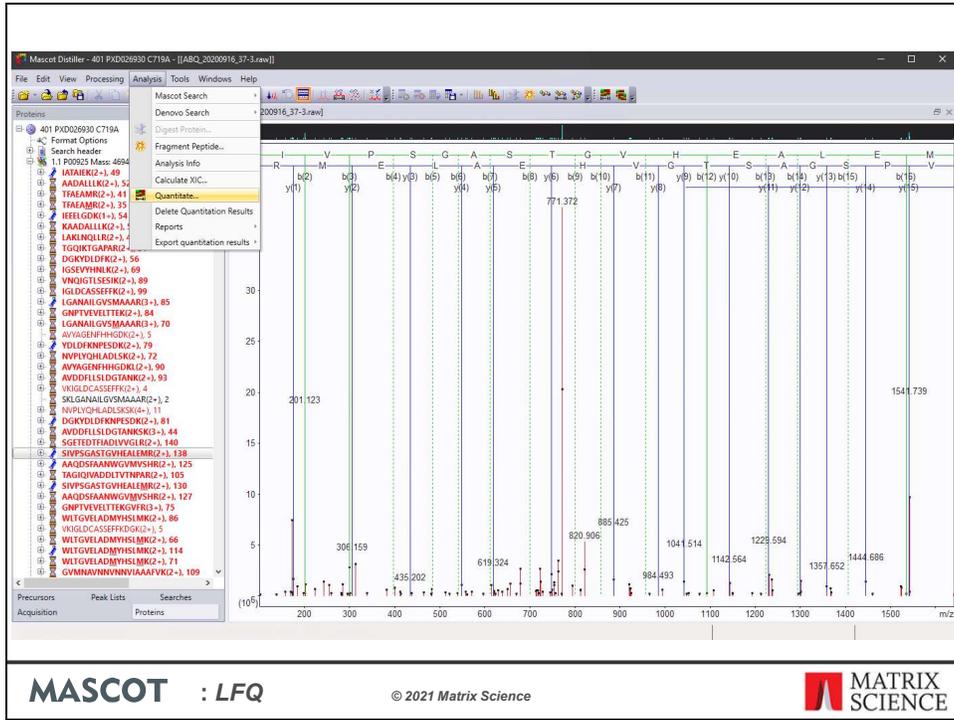
The screenshot displays the Mascot Distiller software interface. On the left, a 'New Multi File project' window shows a table with columns 'Sample name', '#', 'Status', and 'Filename'. The table contains five rows, all with 'Status' set to 'New' and 'Filename' starting with 'ABQ_20200916'. Below the table are sections for 'Distiller Project File', 'Processing options', and 'Import settings and processing options from'. The main window shows a file tree with folders for 'Acquisition' and 'TIC'. A context menu is open over the 'Acquisition' folder, with the 'Process And Search...' option highlighted. The menu also includes options like 'Process Scan', 'Process Range', 'Process All Scans', 'Delete Peak list collection', 'Create Summed Spectrum', 'Edit Processing Options...', and 'Calibration...'. At the bottom of the interface, there are tabs for 'Precursors', 'Peak Lists', and 'Searches', and a status bar with 'MASCOT : LFQ', '© 2021 Matrix Science', and the 'MATRIX SCIENCE' logo.

Sample name	#	Status	Filename
1	1	New	ABQ_20200916_37-9
2	1	New	ABQ_20200916_37-8
3	1	New	ABQ_20200916_37-3
4	1	New	ABQ_20200916_37-2
5	1	New	ABQ_20200916_37-1

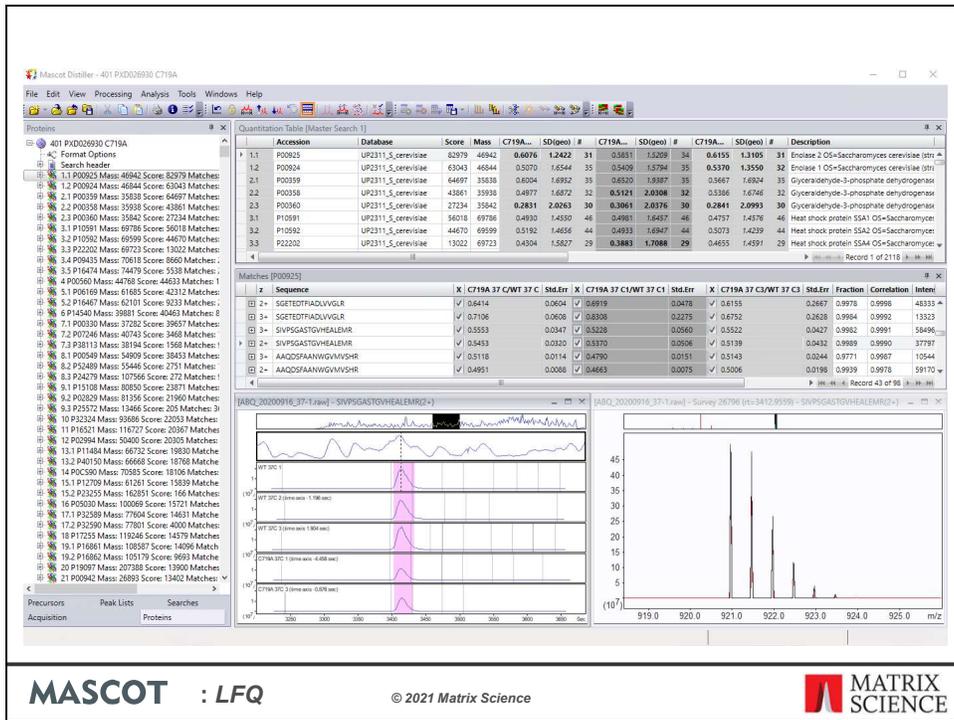
Choose Process and Search to identify the peptides.



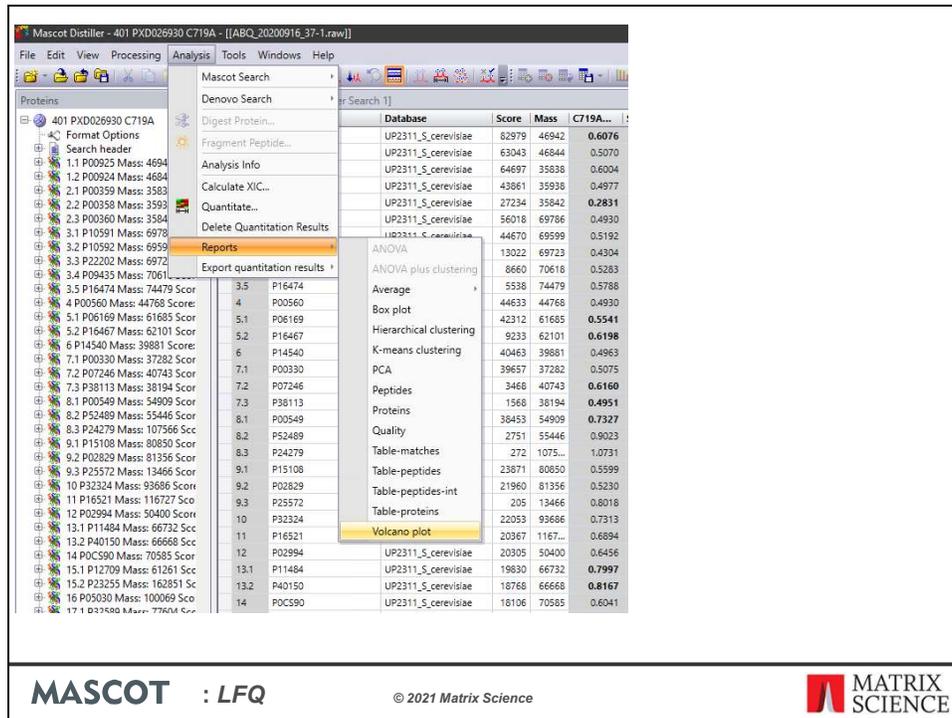
Set up the search parameters, I like to customize the quantitation method on Mascot Server prior to the search.



When the search is complete the next step is to quantitate the data set.



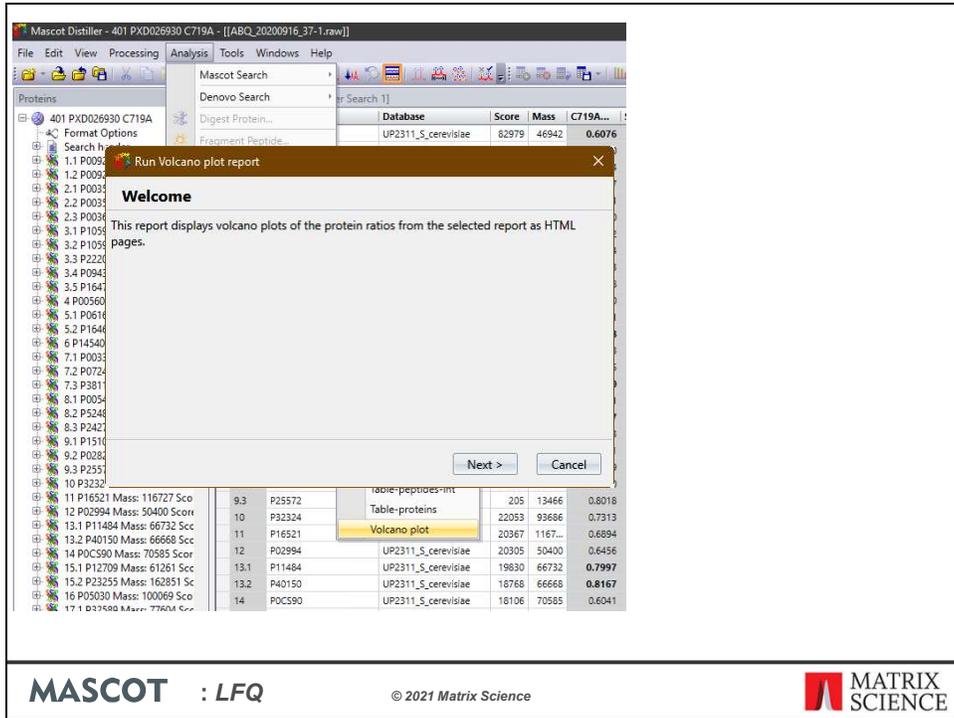
Once quantitation is complete the results are ready for further analysis.



There are five new 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. There are also some quality control reports and the preexisting tables and HTML reports.

We are going to follow along with the publication and create a volcano plot.

Go to the Analysis->Reports menu and select Volcano plot



Step through the dialog boxes

Mascot Distiller - 401 PXD026930 C719A - [ABQ_20200916_37-1.raw]

File Edit View Processing Analysis Tools Windows Help

Mascot Search
Denovo Search
Digest Protein...
Fragment Peptide...

Search []

Database	Score	Mass	C719A...
UP2311_S_cerevisiae	62979	46942	0.6076

Proteins

401 PXD026930 C719A

Format Options

Search by

Run Volcano plot report

Welcome Run Volcano plot report

This report displays 17 pages.

Protein ratio significance threshold

Select the protein ratio T-test significance threshold

Protein ratio significance threshold 0.05

0.05
0.01
0.005
0.001

< Back Next > Cancel

11	P16521	Volcano plot	20367	1167...	0.6894
12	P02994	UP2311_S_cerevisiae	20305	50400	0.6456
13.1	P11484	UP2311_S_cerevisiae	19830	66732	0.7997
13.2	P40150	UP2311_S_cerevisiae	18768	66668	0.8167
14	P0CS90	UP2311_S_cerevisiae	18106	70585	0.6041

MASCOT : LFQ

© 2021 Matrix Science

MATRIX SCIENCE

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 screenshot shows the Mascot Distiller interface with a search results table and a dialog box for excluding a contaminant database.

Search Results Table:

Database	Score	Mass	C719A...
UP2311_S_cerevisiae	62979	46942	0.6076

Exclude contaminant database? Dialog Box:

Select the protein ratio site:

Choose a contaminant database to exclude from the report (if any):

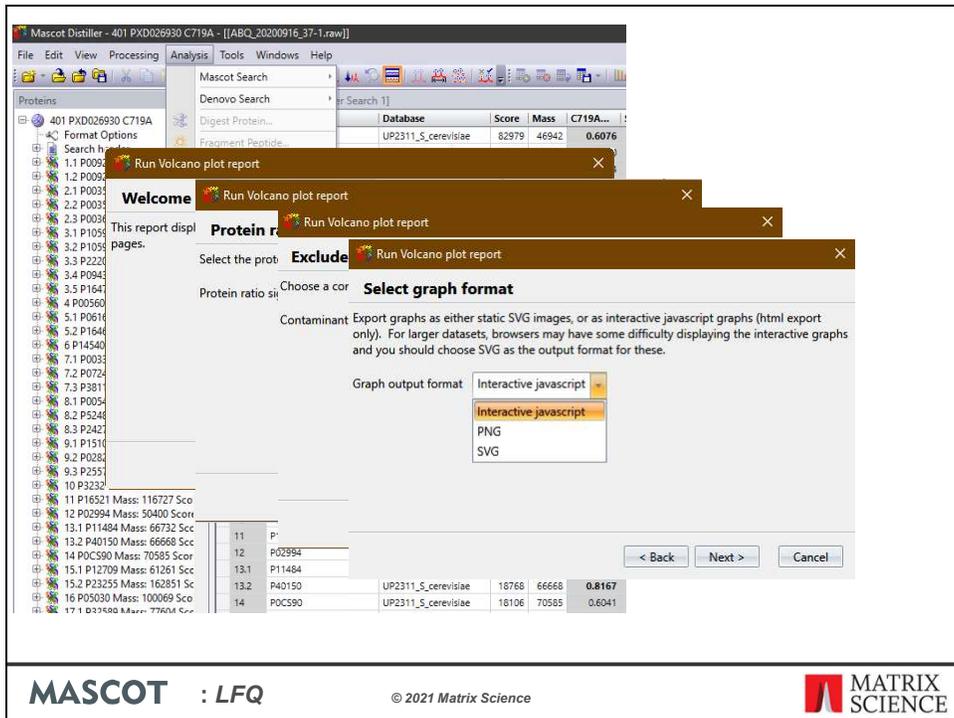
Contaminant database: **contaminants** (selected)

Other options: UniProt_Yeast

Buttons: < Back, Next >, Cancel

MASCOT : LFQ © 2021 Matrix Science **MATRIX SCIENCE**

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.



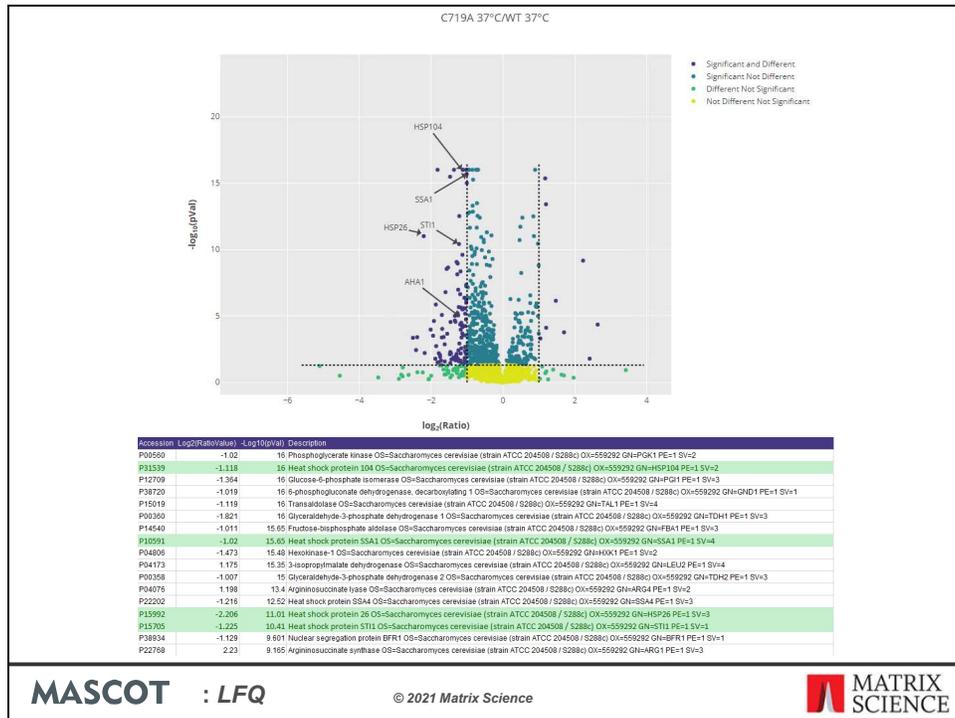
Here is the report from Distiller. The icons at the top right allow you to zoom in and interact with the plot.

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.



Where you 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 study.



Four of these proteins are shown in this table, the fifth is further down the table with a lower p value. They have some of the most significant changes in expression when the Alanyl-tRNA Synthetase C719A mutant *S. cerevisiae* strain is exposed to higher growth temperatures for two hours.

The paper suggests that Alanyl-tRNA Synthetase C719A mutant has an editing deficiency which causes misincorporation of Ser into Ala positions in the proteome leading to a cascade of misfolding and degradation events of key regulatory proteins. Heat stress exacerbates protein misfolding and results in loss of function for key factors in the carbon metabolism, heatshock response, highlighted here, and protein synthesis pathways.

The screenshot shows the Mascot Distiller software interface for a search of PXD026930 C719A. The main window displays a list of protein matches with columns for Accession, Database, Score, Mass, and C719. A context menu is open over the list, with 'Save Stats Package Export...' highlighted. Below the list, a 'Matches' table shows details for selected peptides, including their sequence, score, and match count. At the bottom, a chromatogram plot is visible with a peak labeled 'WT 37C 1'.

Accession	Database	Score	Mass	C719
P00925	UP2311_S_cerevisiae	82979	46942	0
P00924	UP2311_S_cerevisiae	63043	46944	0
P00359	UP2311_S_cerevisiae	64697	35930	0
P00358	UP2311_S_cerevisiae	43861	35930	0
P00360	UP2311_S_cerevisiae	27234	35642	0
P10591	UP2311_S_cerevisiae	56018	69786	0
P10592	UP2311_S_cerevisiae	44670	69599	0
P22202	UP2311_S_cerevisiae	13022	69723	0

z	Match	X	C719A 37 CWT
2+	SGETDTFIADLVVGLR	✓	0.6414
3+	SGETDTFIADLVVGLR	✓	0.7106
3+	SIVPSGASTGVHEALEMR	✓	0.5533
2+	SIVPSGASTGVHEALEMR	✓	0.5453
3+	AAQDSFAANWGVVMSHR	✓	0.5118
2+	AAQDSFAANWGVVMSHR	✓	0.4951

MASCOT : LFQ © 2021 Matrix Science **MATRIX SCIENCE**

If you want to do further analysis beyond the reports included with Mascot Distiller the data can be exported as a stats package.

The screenshot displays the Mascot Distiller software interface. The main window shows a list of search results with columns for Accession, Database, Score, Mass, and C19. Below the results, there are several menu options including 'Export quantitation results' and 'Save Abbreviated XML...'. The bottom of the window shows a file explorer view of the 'MSData' folder, listing files such as 'Grouping_Proteins.txt', 'Matches.txt', 'Peptides.txt', and 'Proteins.txt'.

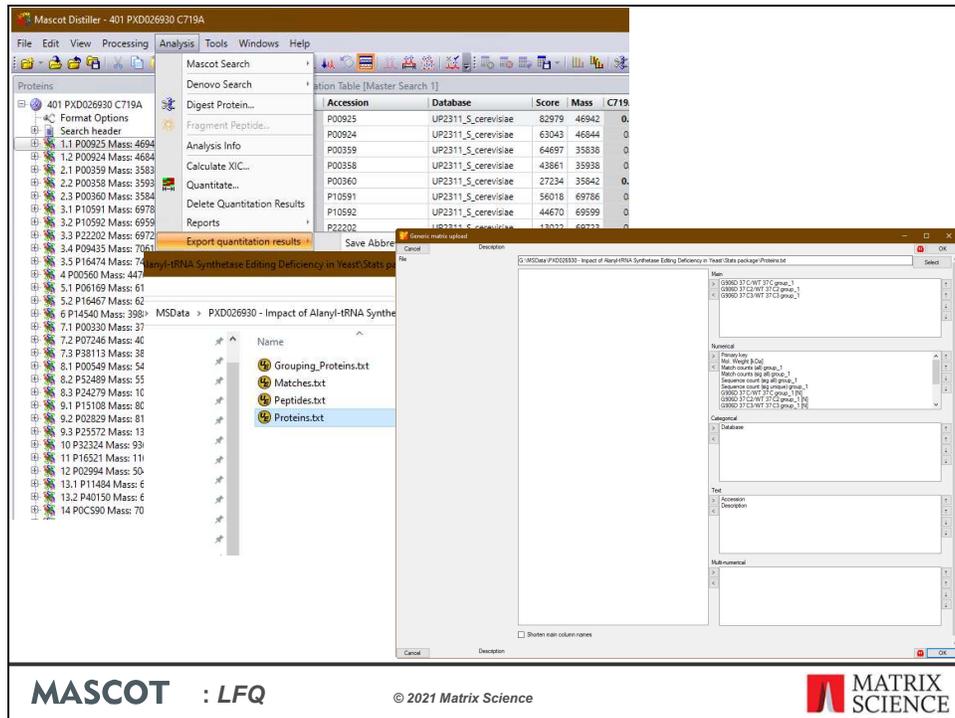
Accession	Database	Score	Mass	C19
P00925	UP2311_S_cerevisiae	82979	46942	0
P00924	UP2311_S_cerevisiae	63043	46944	0
P00359	UP2311_S_cerevisiae	64697	35930	0
P00358	UP2311_S_cerevisiae	43861	35930	0
P00360	UP2311_S_cerevisiae	27234	35642	0
P10591	UP2311_S_cerevisiae	56018	69786	0
P10592	UP2311_S_cerevisiae	44670	69599	0
P22202	UP2311_S_cerevisiae	13022	69723	0

MSData > PXD026930 - Impact of Alanyl-tRNA Synthetase Editing Deficiency in Yeast > Stats package

Name	Date modified	Type	Size
Grouping_Proteins.txt	10/8/2021 3:32 PM	TXT File	1 KB
Matches.txt	10/8/2021 3:34 PM	TXT File	82,173 KB
Peptides.txt	10/8/2021 3:34 PM	TXT File	7,086 KB
Proteins.txt	10/8/2021 3:34 PM	TXT File	598 KB

MASCOT : L_FQ © 2021 Matrix Science **MATRIX SCIENCE**

When the export is complete unzip the archive.



And open the protein file in a third-party analysis program like Perseus.

The files exported from Mascot Distiller include a column type row which Perseus uses to correctly assign the columns during the import making it a one click affair. You could then use Perseus to annotate the results with Gene ontology and KEGG pathway information or carry out other statistical analysis.

This is the end of the tutorial which covers the basics of Label free quantitation in Mascot Distiller. If you have questions on how analyze your data with Mascot Server and Distiller please contact support@matrixscience.com