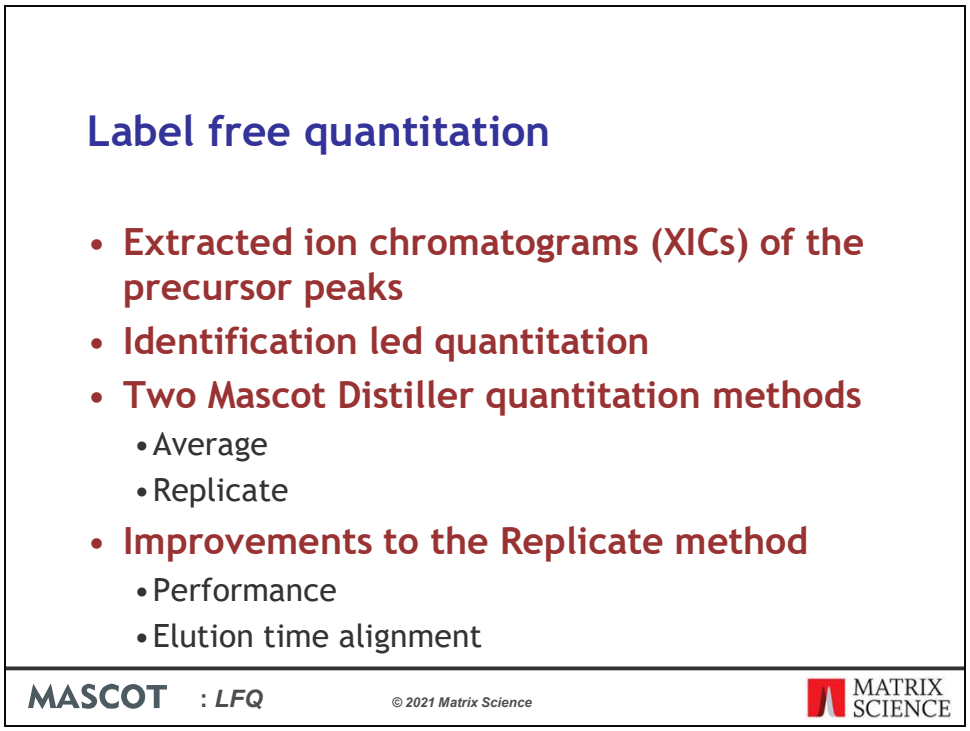
Hello this is Richard Jacob for Matrix Science and my talk is LFQ for the masses. It's a Mascot Distiller label free quantitation tutorial.



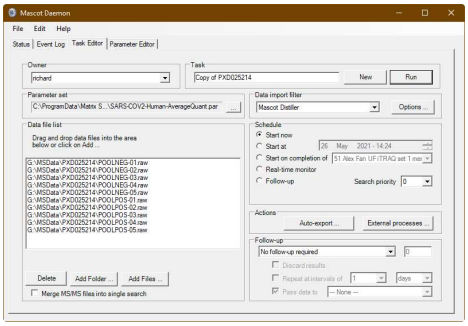
Within the Mascot software ecosystem we can perform Label free quantitation using several different methods. Today I won't be talking about spectra counting or emPAI methods, just those that use extracted ion chromatograms or XIC's of the precursor peaks.

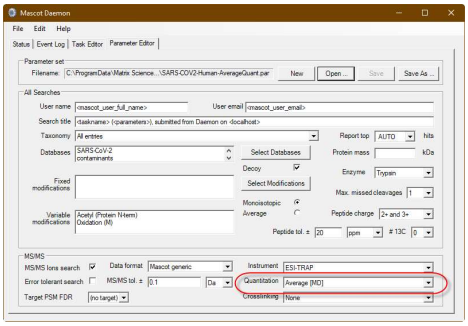
In Mascot we perform identification led quantitation. This means that after peptide identification we go back to the raw data to calculate the XIC's. That way we don't waste time quantitating peaks that might be present in the data but remain unidentified.

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





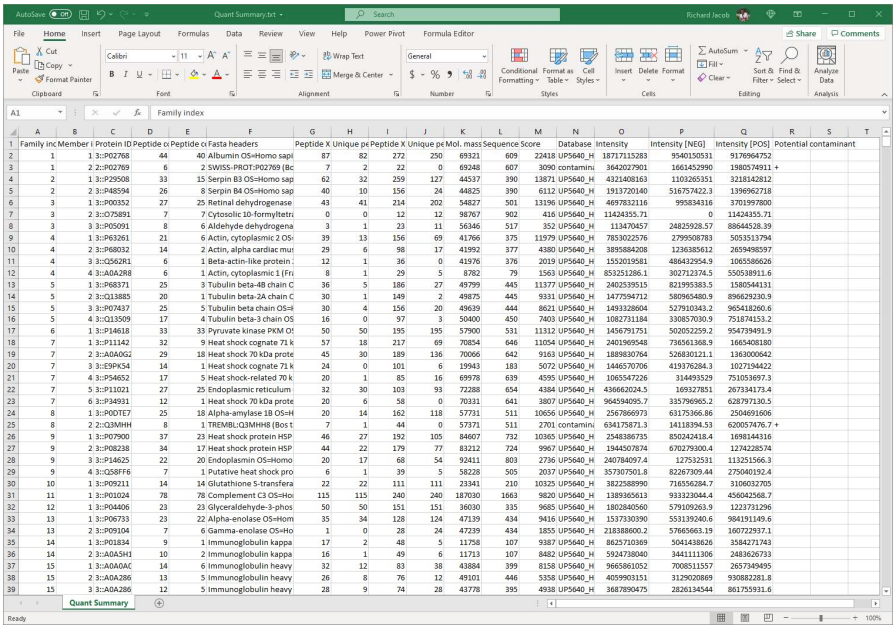
Export of results detailed in last years Quantitation Summary presentation https://www.matrixscience.com/workshop_2020.html

Let's start by processing a data set though Mascot Daemon to create a Quantitation Summary table or report. This feature was introduced in 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.

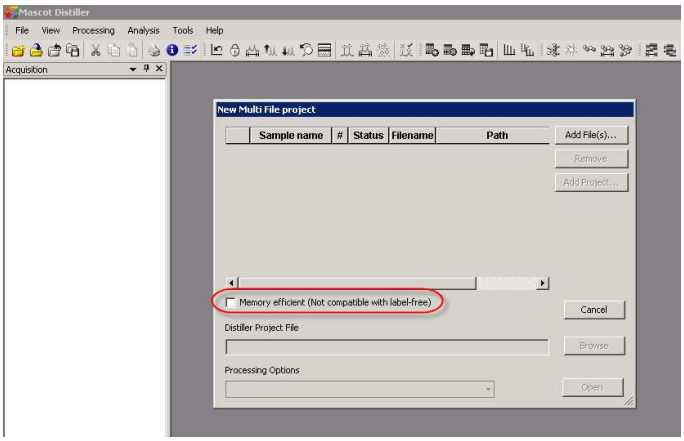
When you start the task Mascot Daemon will processes raw files and search the peak lists serially. The search results are then returned to Distiller and quantified.

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.



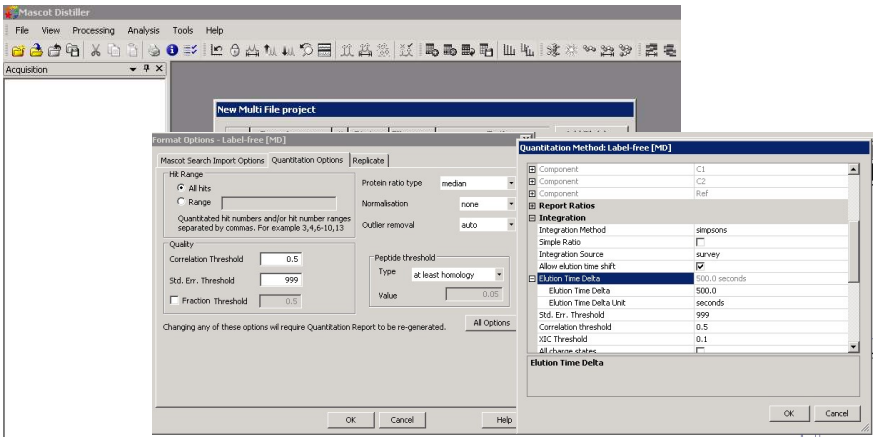
You can open the file in a text editor like notepad but it makes more sense to view it as a spreadsheet in Excel. For further statistical analysis you can use Excel, R, Perseus or other software. Again, previous presentations and blog posts have some examples.

LFQ in Mascot Distiller 2.7 and earlier



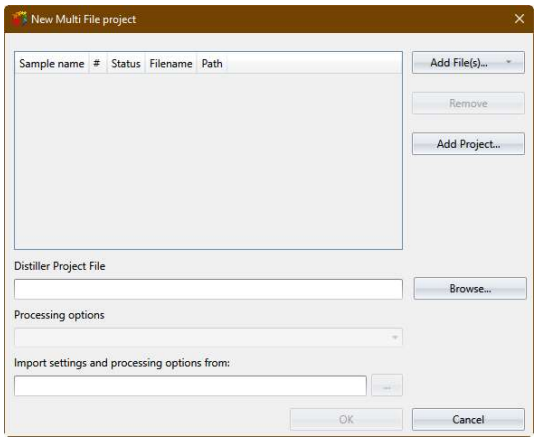
Now we are going to look at processing data in Mascot Distiller directly using multi file projects and the replicate analysis. With Mascot Distiller version 2.7 and earlier a multi-file project was created, and the “Memory Efficient” box had to be left unchecked. Mascot Distiller would open all the files 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



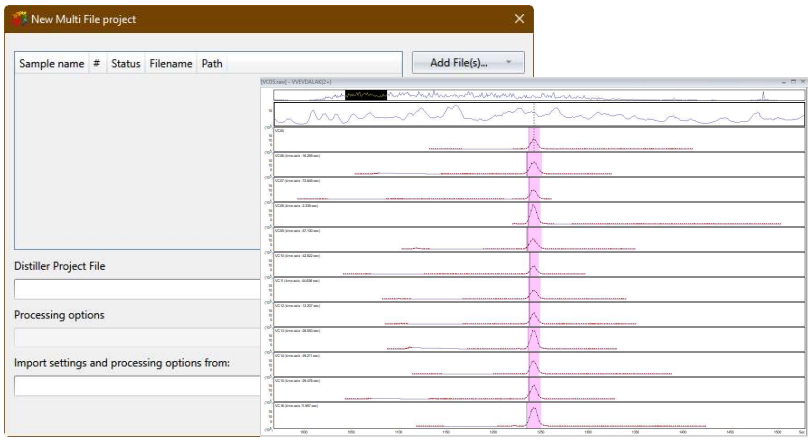
When it came to the quantitation Distiller would use a global elution time delta to look for peptides identified in at least one of the files in a data set but not in others.

LFQ in Mascot Distiller 2.8

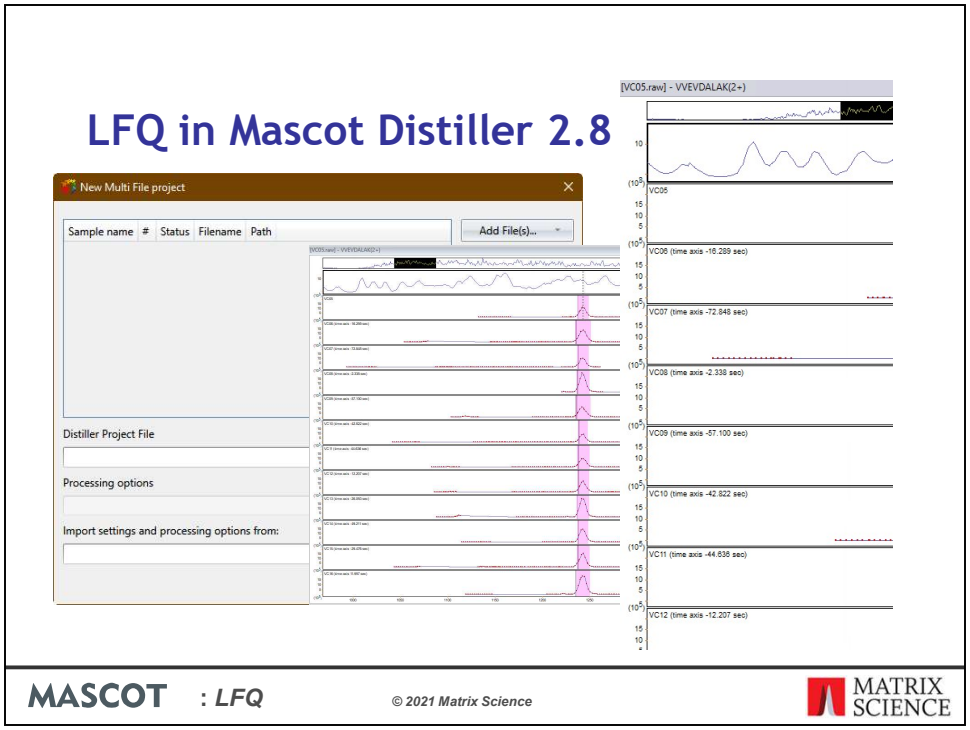


In Mascot Distiller 2.8 we now process all the files independently. The Memory Efficient check box has been removed as well as the Elution Time Delta variable which has been replaced by a new global time-alignment algorithm. This algorithm builds a a consensus dataset by aligning the Total Ion Chromatogram (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 that the older approach.

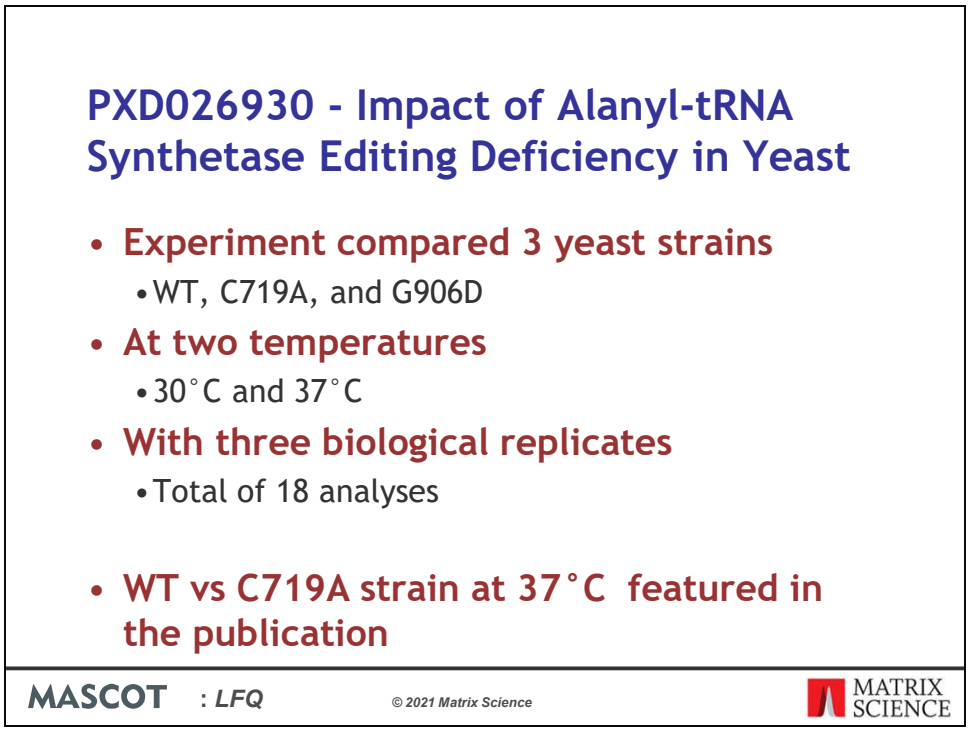
LFQ in Mascot Distiller 2.8



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



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.



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.

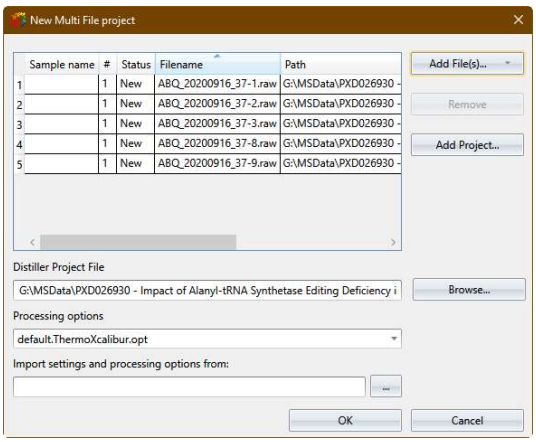
The cultures were grown at 30°C, sampled and then increased to 37°C for 2 hours before sampling again.

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

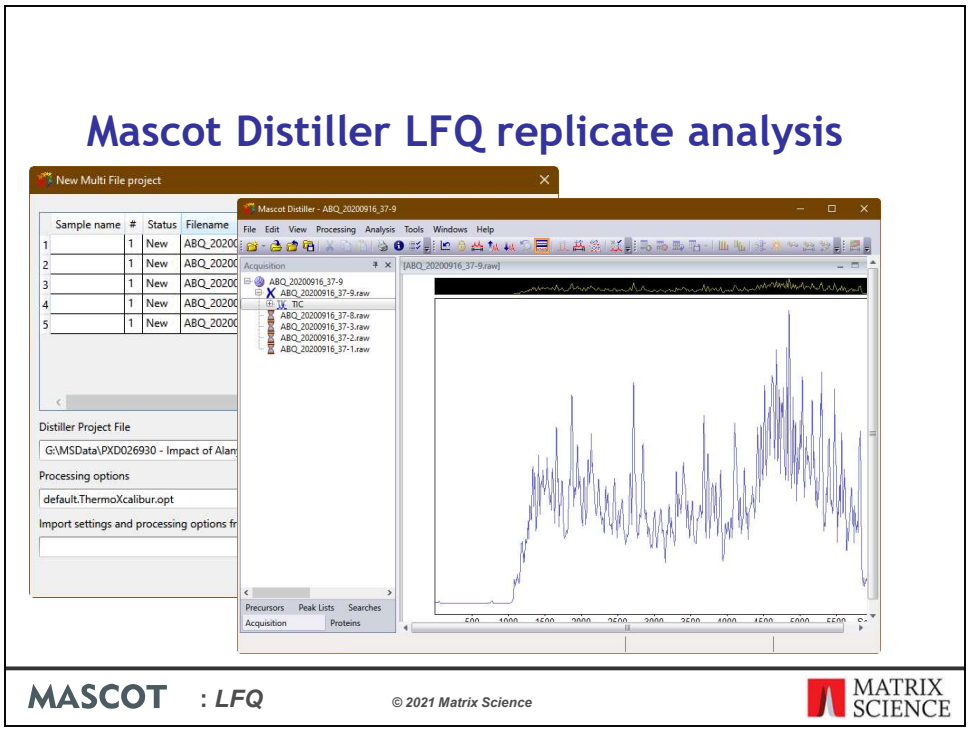
Rather than process the full data set I selected just the files for one a comparison of Wild Type to the C719A mutant strain at 37°C as this was featured in the publication.

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

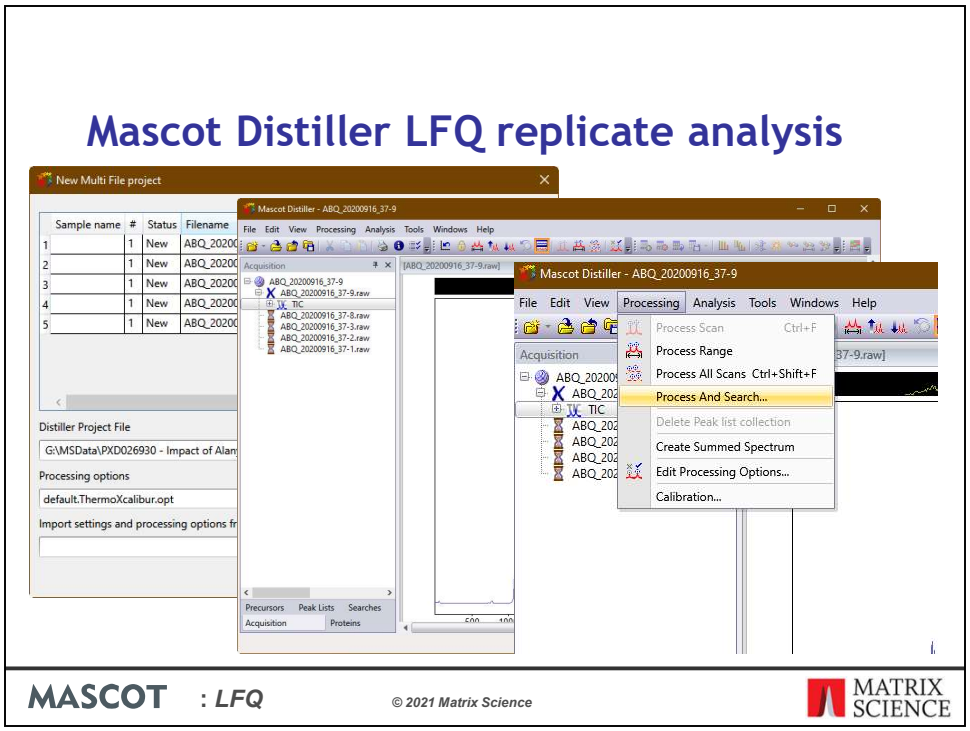
Mascot Distiller LFQ replicate analysis



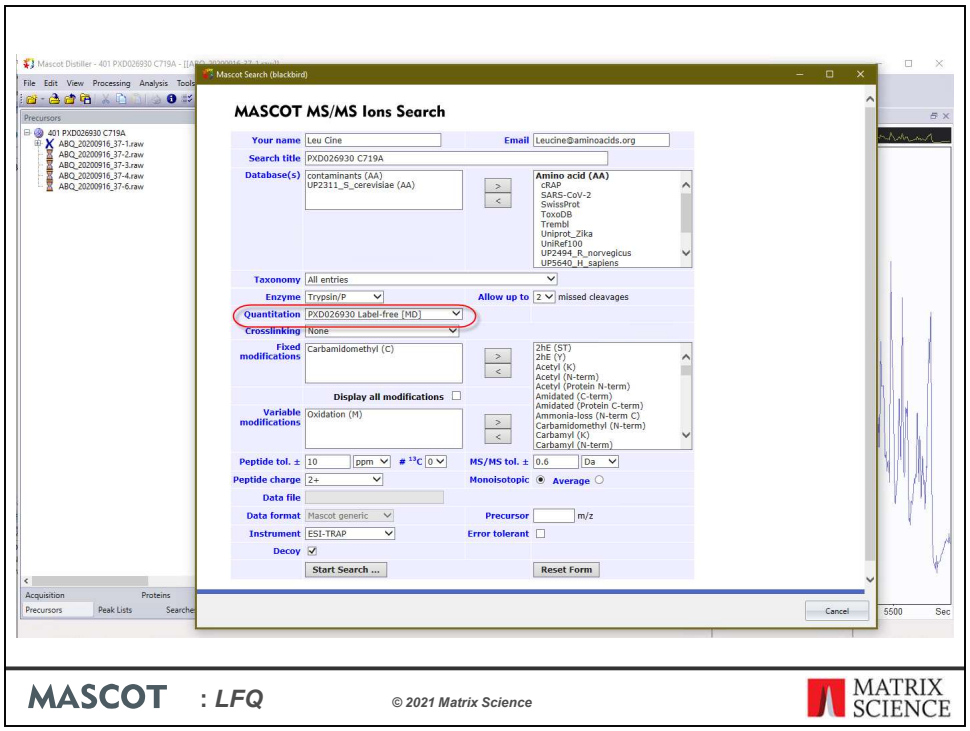
We can set up label free quantitation in Mascot Distiller as a multi file project.



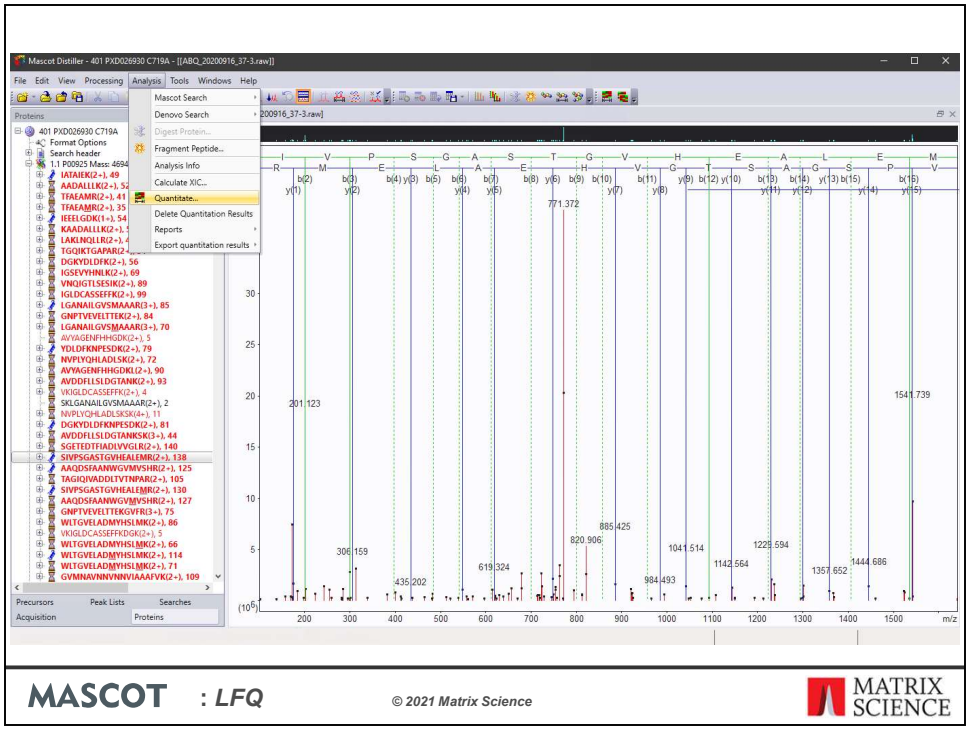
After the data set is open in Mascot Distiller



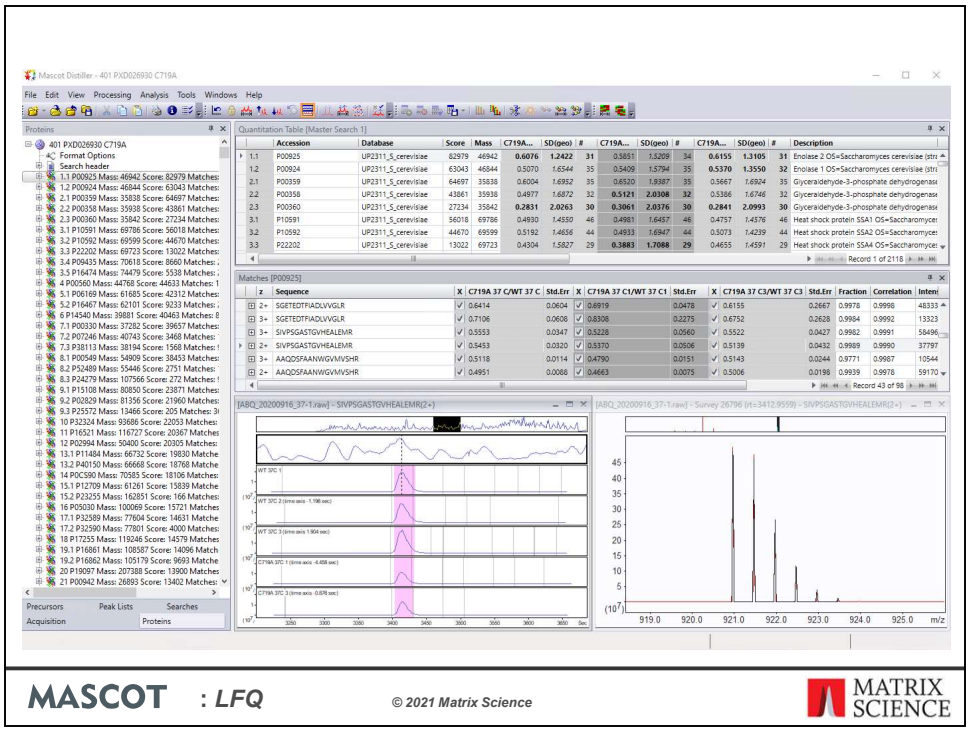
Choose Process and Search to identify the peptides.



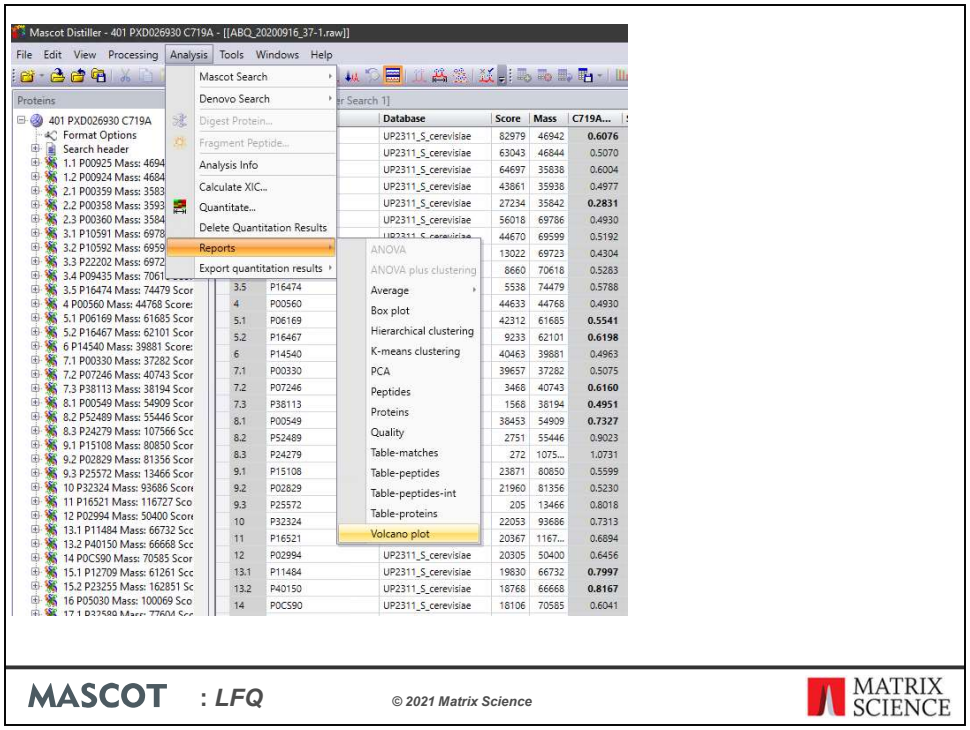
Set up the search parameters, I like to customize the quantitation method on Mascot Server prior to the search.



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



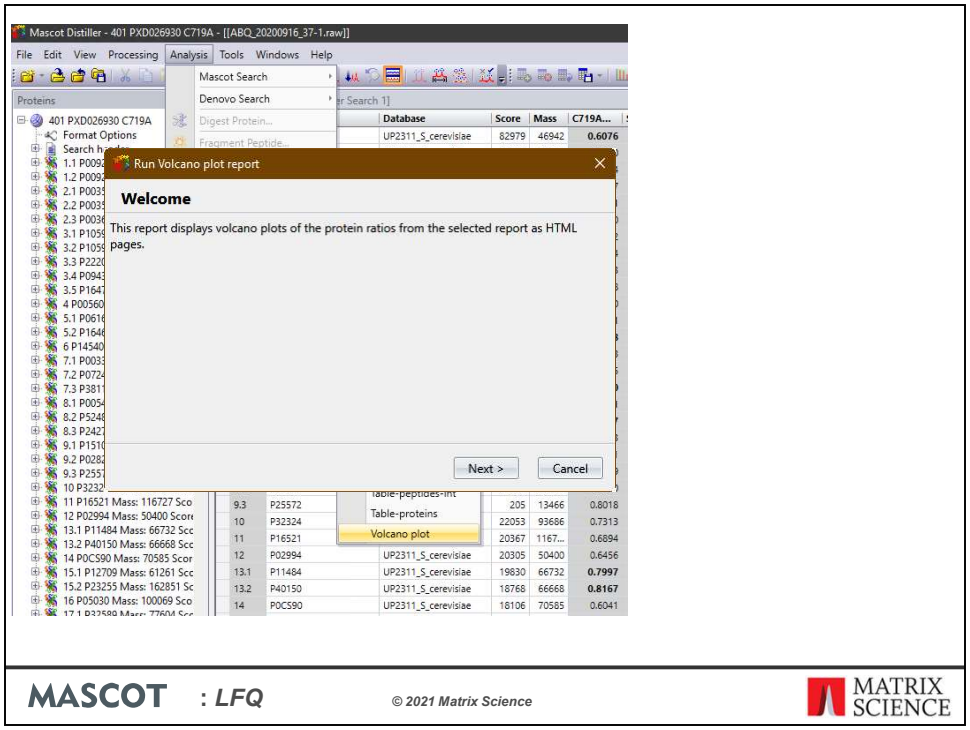
Once quantitation is complete the results are ready for further analysis.



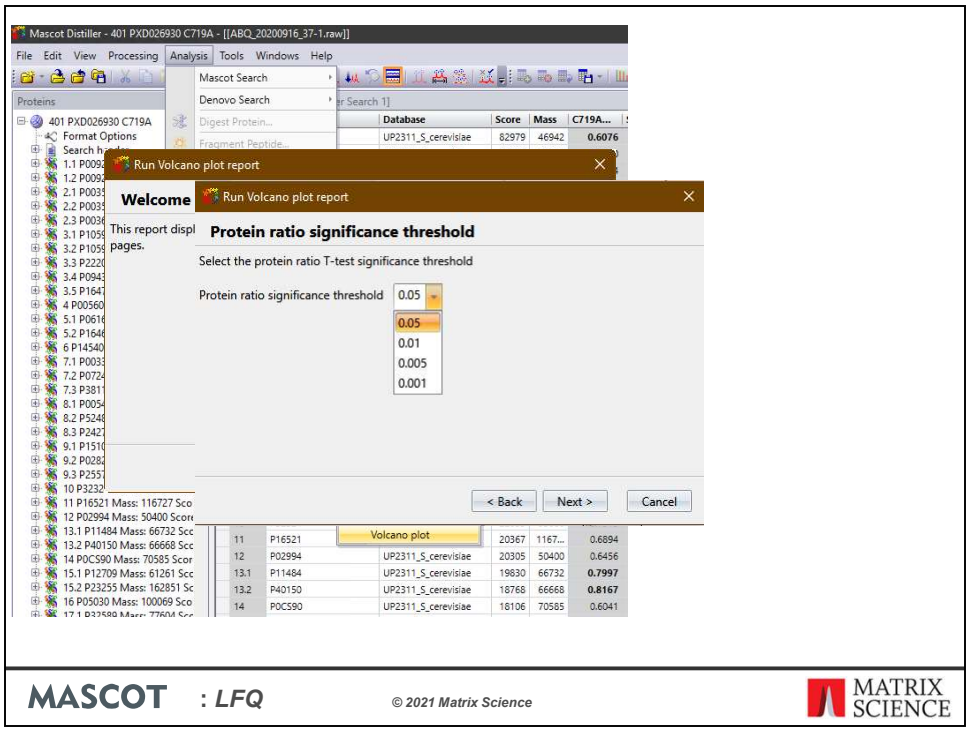
There are five new statistical analysis reports included with Mascot Distiller. These reports are generated by python scripts and cover ANOVA, Hierarchical and K-means Clustering, Principal Component Analysis and Volcano plots. There are also some quality control reports and the preexisting tables and HTML reports.

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

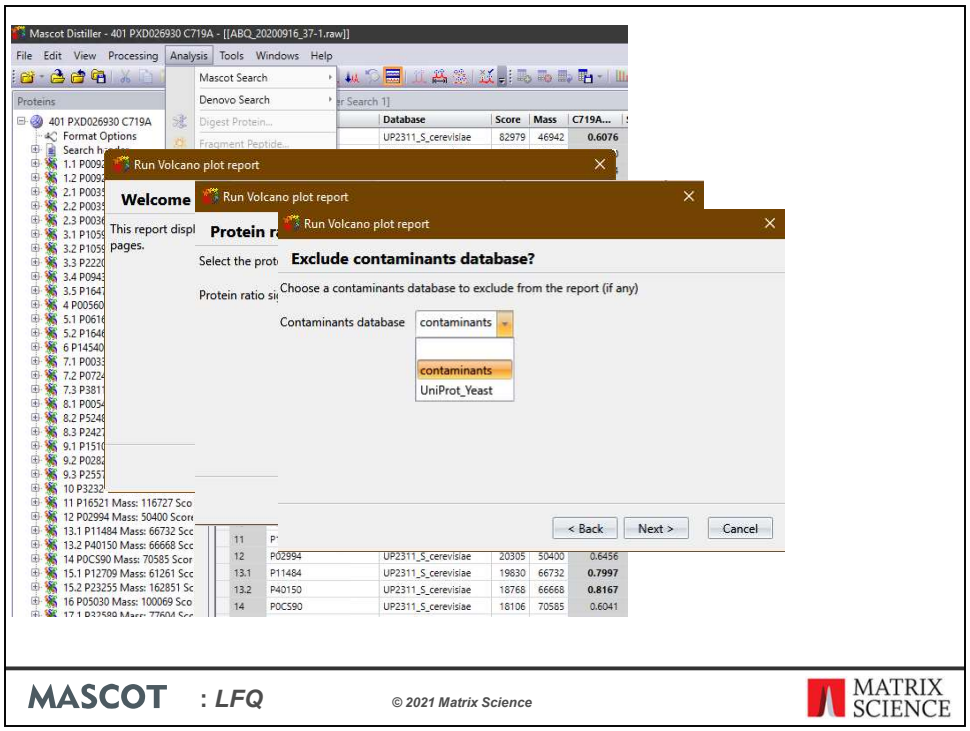
Go to the Analysis->Reports menu and select Volcano plot



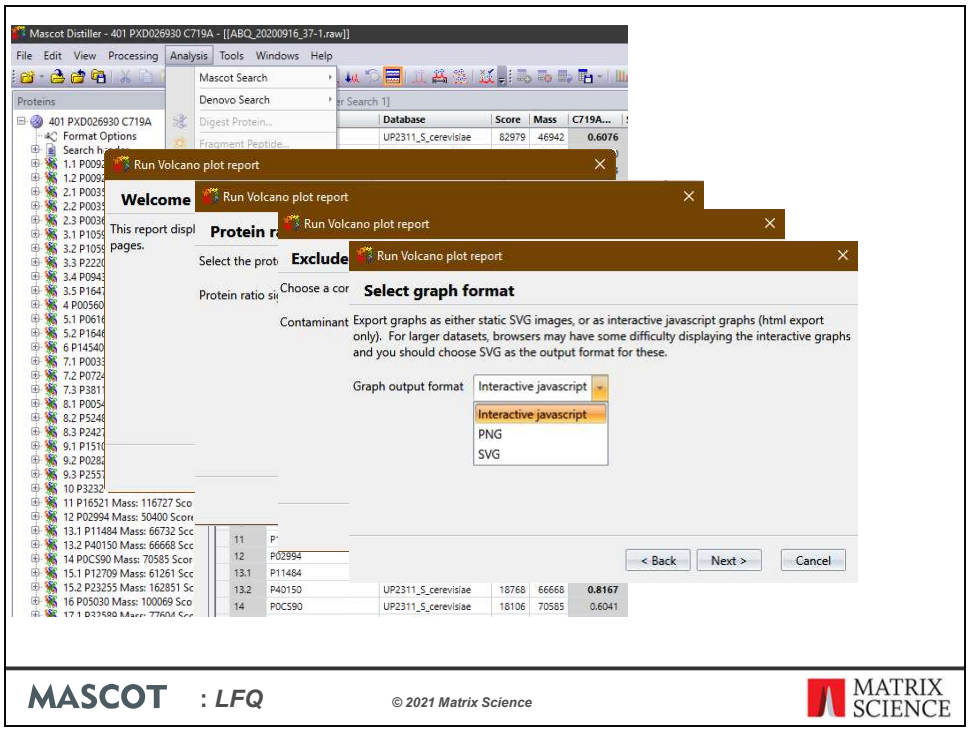
Step through the dialog boxes



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.

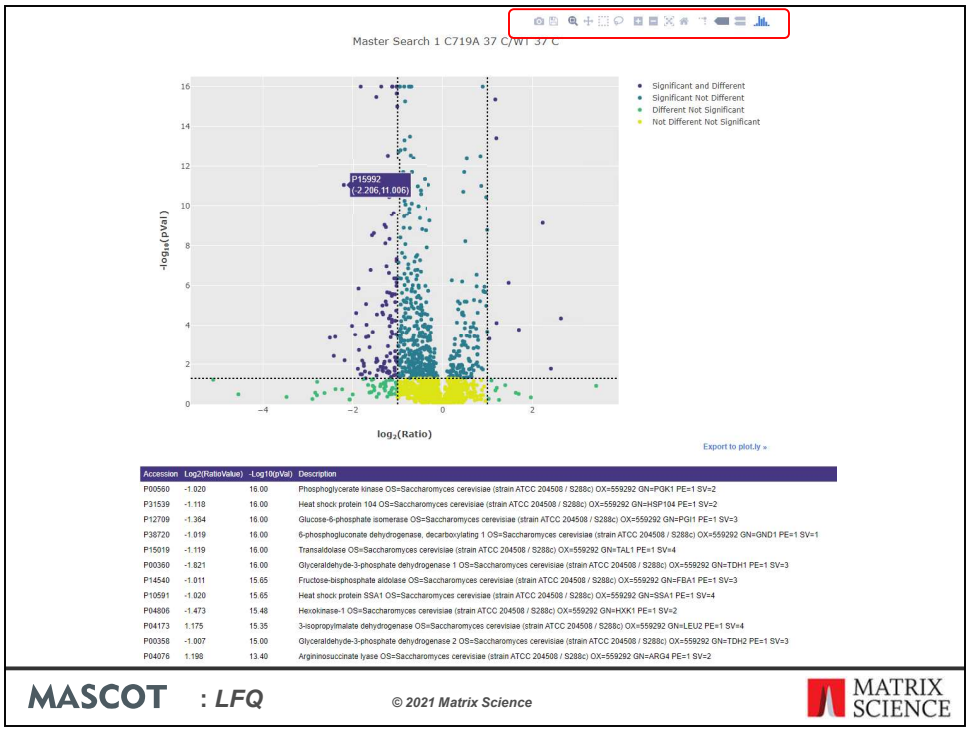


Filter out hits from the contaminant database, primarily Trypsin in this case.



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.

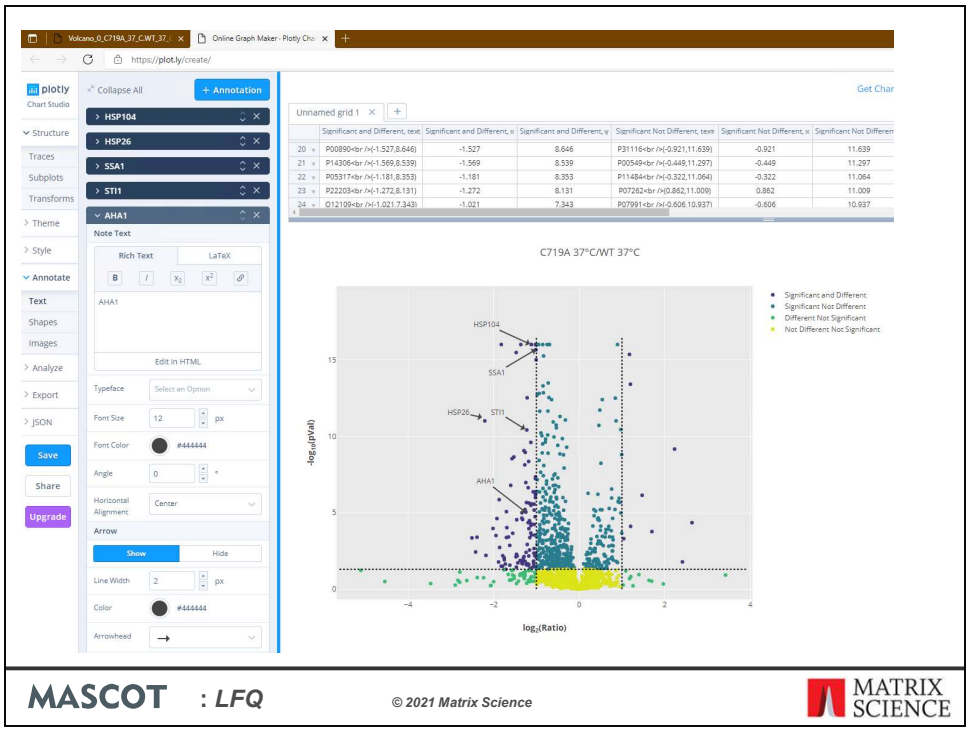
The report is generated and saved to a temporary archive. Move the archive to a project folder, uncompress it and open the html page in a browser by double clicking on the file.



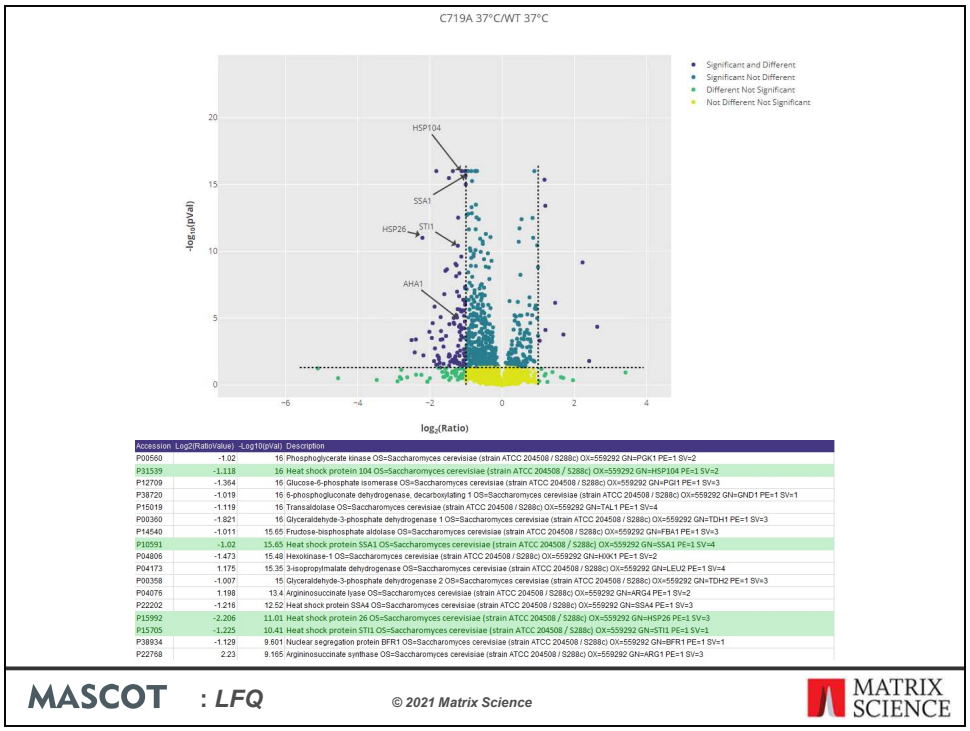
Here is the report from Distiller. The icons at the top right allow you to zoom in and interact with the plot.

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.



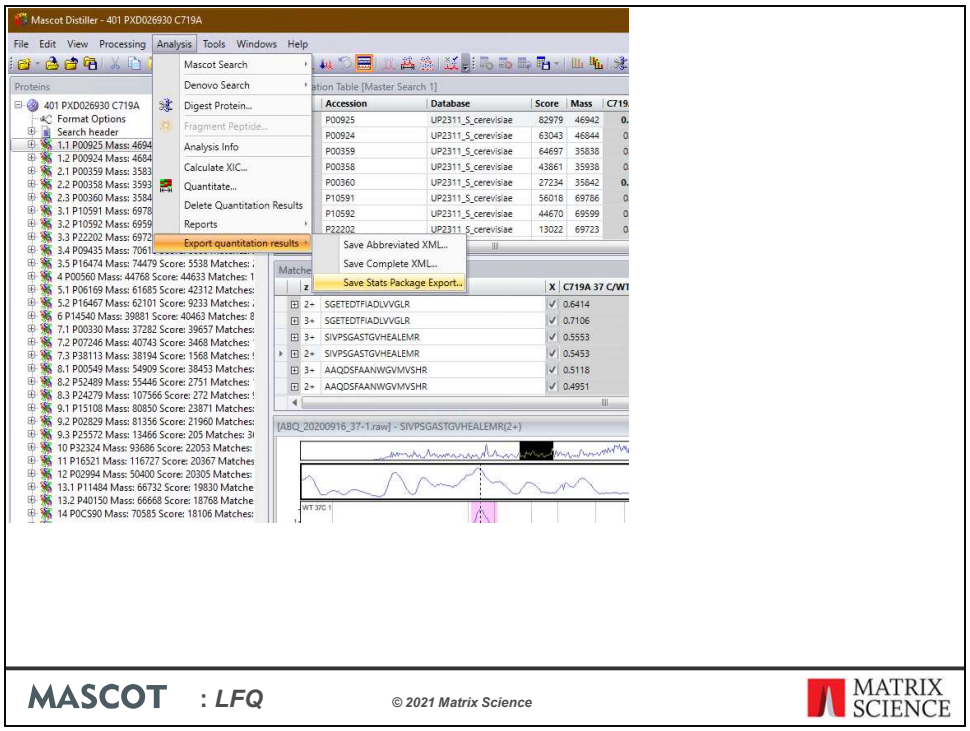
Where you can add annotations. I have annotated four heat shock proteins and a chaperone that are all downregulated and in the same pathway or were highlighted in the study.



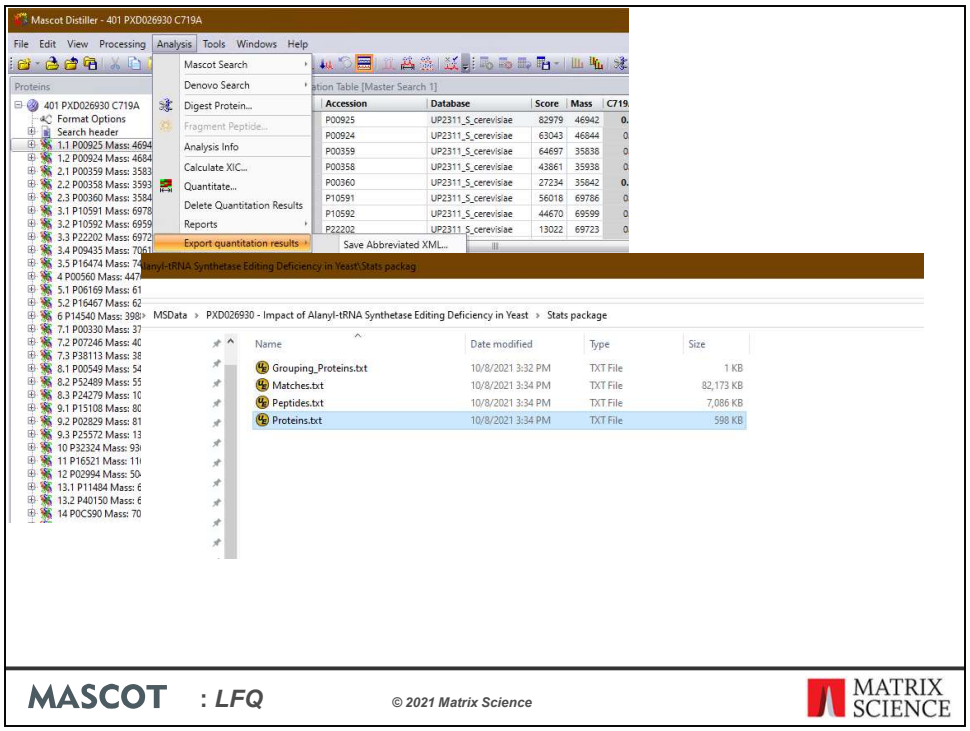
Accession	Log2(RatioValue)	-Log10(pVal)	Description
P00560	-1.02	16	Phosphoglycerate kinase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=PGK1 PE=1 SV=2
P31539	-1.118	16	Heat shock protein 104 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HSP104 PE=1 SV=2
P12709	-1.364	16	Glucose-6-phosphate isomerase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=PGI1 PE=1 SV=3
P38720	-1.019	16	6-phosphogluconate dehydrogenase, decarboxylating 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=GND1 PE=1 SV=1
P15019	-1.119	16	Transaldolase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=TAL1 PE=1 SV=4
P00360	-1.821	16	Glyceraldehyde-3-phosphate dehydrogenase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=TDH1 PE=1 SV=3
P14540	-1.011	15.65	Fructose-bisphosphate aldolase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=FBA1 PE=1 SV=3
P10591	-1.02	15.65	Heat shock protein SSA1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SSA1 PE=1 SV=4
P04806	-1.473	15.48	Hexokinase-1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HXK1 PE=1 SV=2
P04173	1.175	15.35	3-isopropylmalate dehydrogenase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=LEU2 PE=1 SV=4
P00358	-1.007	15	Glyceraldehyde-3-phosphate dehydrogenase 2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=TDH2 PE=1 SV=3
P04076	1.198	13.4	Argininosuccinate lyase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=ARG4 PE=1 SV=2
P22202	-1.216	12.52	Heat shock protein SSA4 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SSA4 PE=1 SV=3
P15992	-2.206	11.01	Heat shock protein 26 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HSP26 PE=1 SV=3
P15705	-1.225	10.41	Heat shock protein STI1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=STI1 PE=1 SV=1
P38934	-1.129	9.601	Nuclear segregation protein BFR1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=BFR1 PE=1 SV=1
P22768	2.23	9.165	Argininosuccinate synthase OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=ARG1 PE=1 SV=3

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.

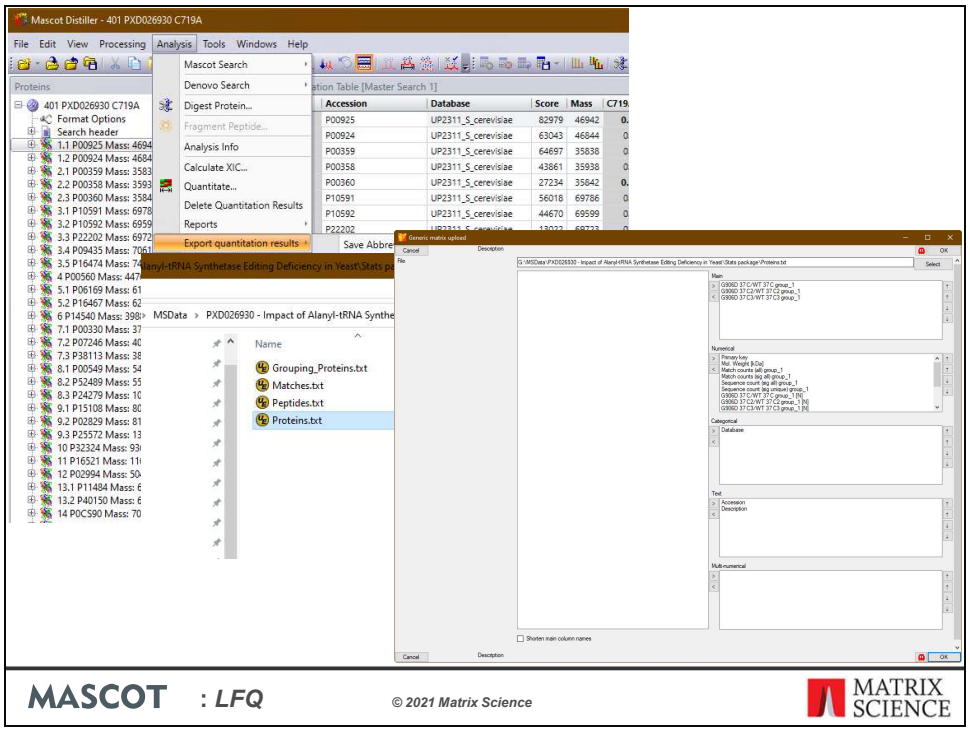
The paper suggests that Alanyl-tRNA Synthetase C719A mutant has an editing deficiency which causes misincorporation of Ser into Ala positions in the proteome leading to a cascade of misfolding and degradation events of key regulatory proteins. Heat stress exacerbates protein misfolding and results in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



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



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 Peruses 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 coverers 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