Analyzing metabolomics data using XCMS

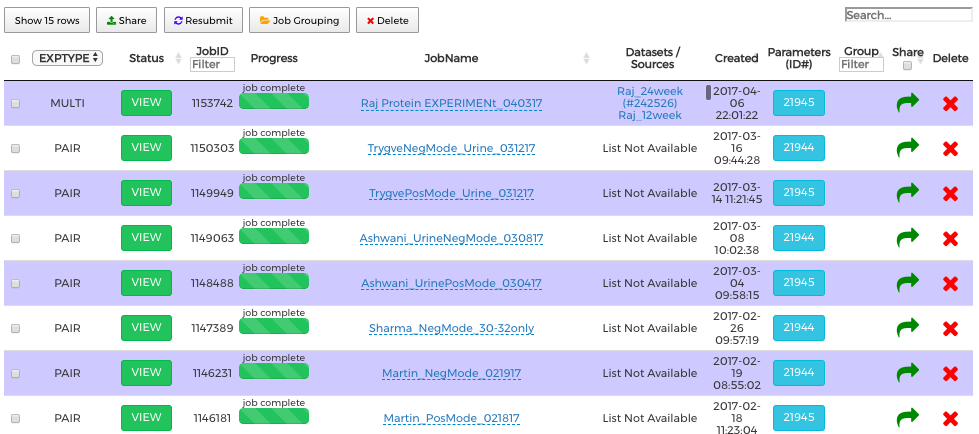
XCMS is a tool to analyze GC-MS and LC-MS metabolomics datasets. It comes in two forms – a standalone version in R. Those of you who are familiar with R should speak to Dr. Xiuxia Du who can demonstrate its use.

XCMSonline (<https://xcmsonline.scripps.edu>) is a web-based version of XCMS. To use it, you must first register to get an account. The account name will be your e-mail address and you will need to select a password. Once the account is activated and you have logged on, you may then upload raw data sets that have been collected from GC-MS and LC-MS analyses. There are 3-5 min short videos explaining the major features of XCMS.

For the tutorial, the datasets (urines from 6 controls and 6 genistein-treated mice) have already been uploaded, the parameters for the analysis have been set and XCMS has processed the data which is available for inspection. Some of the large number of features of XCMS will be demonstrated during the live tutorial, ending with a download of all the information processed by XCMS. One of the downloaded files will be used for the statistical analysis of the data in the Metaboanalyst session.

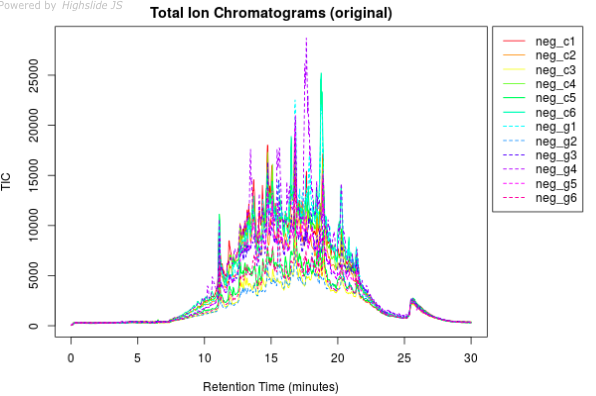
The following represents some of the figures from the XCMS analysis of the experiment. Those wishing to analyze the data from this experiment should contact Dr. Barnes who will share it (the data are not publicly available otherwise).

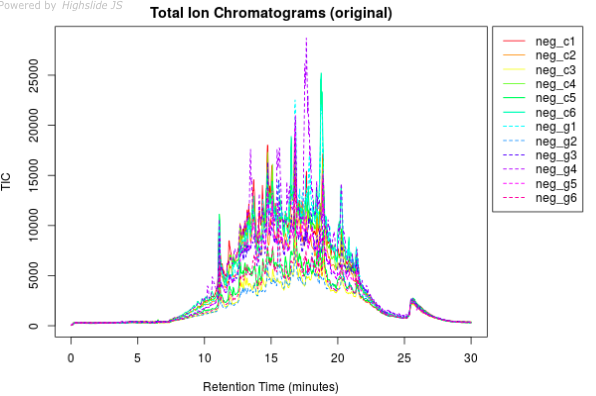
Clicking on the View Results tab takes you to a summary page.



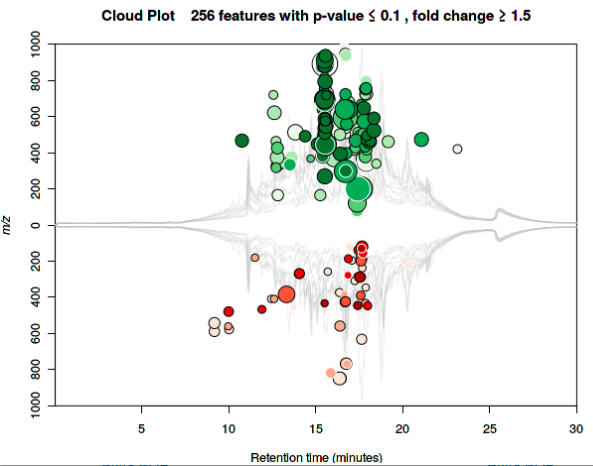
By clicking VIEW, several summary figures appear. The first is an overlay of the total ion chromatograms (TICs) from each urine. At first glance, the ratio of the most intense to least intense TIC is 2-3 fold, a reflection of the variation in fluid intake/urine output.

In XCMS, sentinel ions are used to build a polynomial function that corrects for small variations in retention times between individual sample runs.



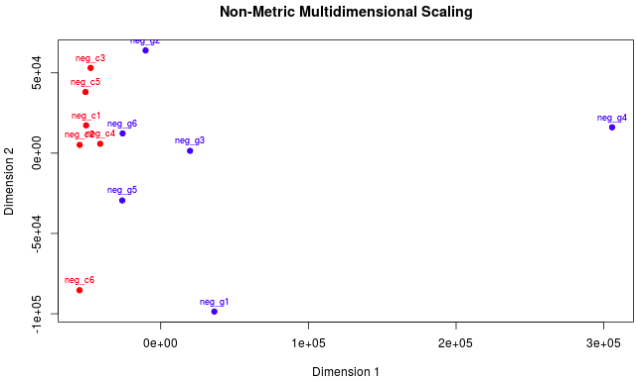


A useful feature of XCMSonline when getting an overall picture of the data is the Cloud plot.



The Cloud plot maps onto the average TIC the ions that exhibit fold changes greater than 1.5 (green ones are increased in the urines of genistein-treated mice, whereas red ones are decreased). The sizes of the circles reflect the fold change and the depth of color the statistical significance (darker circles have smaller p-values). Besides the default Cloud plot, there is also interactive Cloud plot.

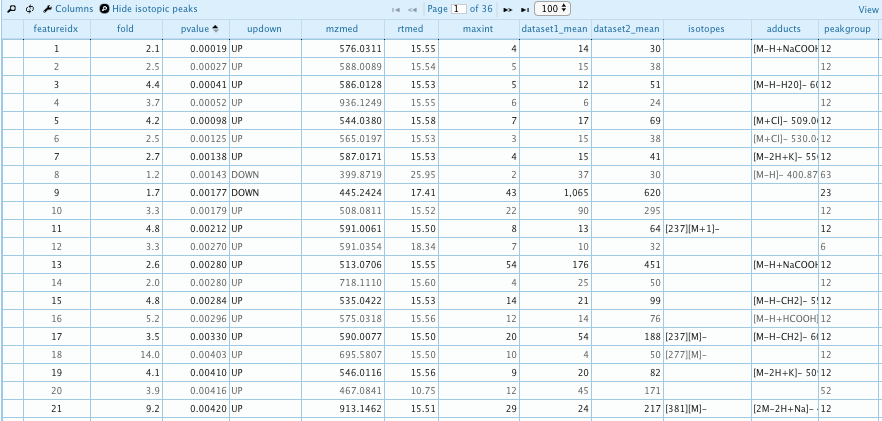
As a default, XCMSonline also provides multivariate analysis of the data. Note that XCMS does not take into account the need for data normalization. Urine volumes are dependent on fluid intake that can be variable from animal to animal.

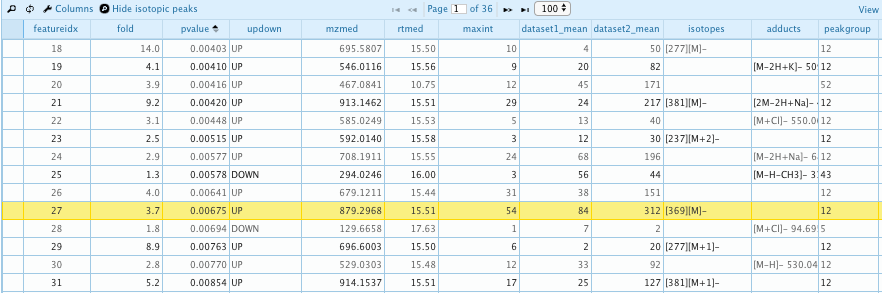
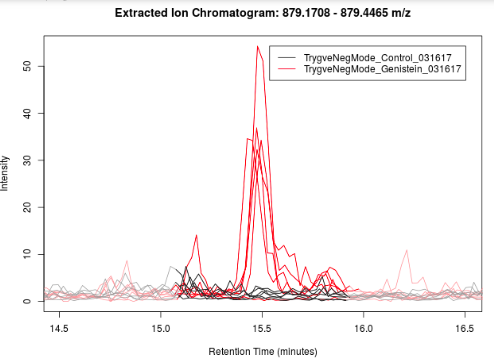


Nonetheless, it can be seen that the controls tend to cluster together (except for animal #6). The GEN group are much more scattered.

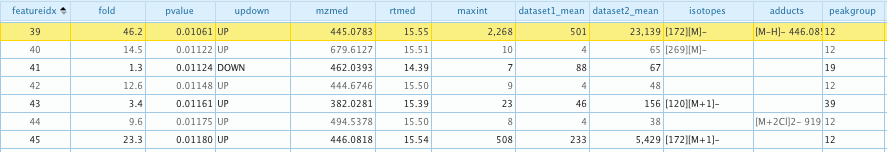
**Results table**: This option presents information on the fold change (increased or decreased) for each ion feature, its median mass-to-charge ratio (*m/z*), median retention time, maximum intensity, mean peak areas of the control and treatment groups and the nature of the ion (adduct/isotope). The default table is sorted based on the lowest p-values (below). Clicking on the headings allows alternative sorting, e.g., on median *m/z* or median retention time values.

Selecting a particular feature in the table (where the p-value is less than 0.05) yields an extracted ion chromatogram (EIC) centered around the peak for that feature.

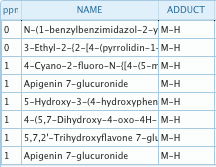


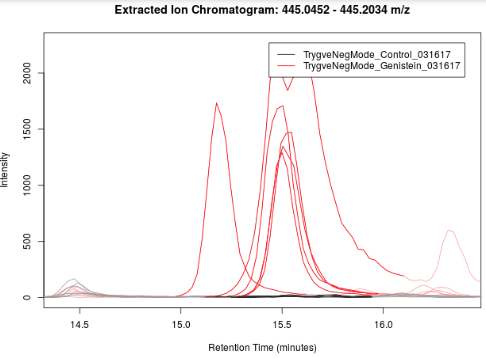


This ion feature is associated with the GEN treatment (red line) and not the controls (black line)

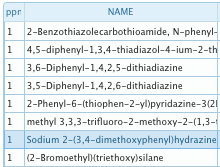


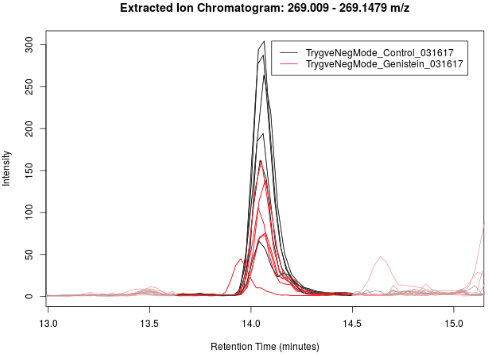
13C-isotope of *m/z* 445.0783

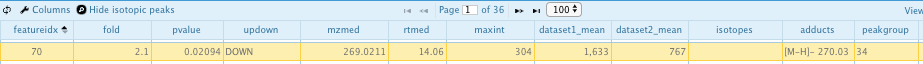




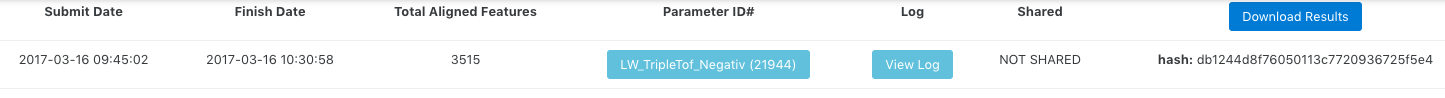
Not all peaks are increased in the treatment group. The ion feature at *m/z* 269.0211 is higher in the control group.







To do further statistical analysis, it is necessary to go back a page and download the XCMS data.



This downloads a zip file that contains many files including an Excel file containing *m/z*, retention time and peak area values that we will use to further process the data and carry out statistical tests and perform pathway analysis. Other files include box and whisker plots and EICs.