

Software Packages

MetaboAnalyst
XCMS

1

MetaboAnalyst 2.0

- This is an online set of statistical tools
 - <http://www.metaboanalyst.ca/>
- First, create an Excel file for each sample containing three columns, the *m/z* of each metabolite, its retention time and the peak intensity
- Put the files for a group in a folder.
- Then zip the two folders
- On the MetaboAnalyst website, click on “Welcome, click here to start”

mzmed	rtmed	11
397.2229	28.23	185337
201.112	26.02	32332
736.4979	32.65	5807
313.1259	27.99	421955
273.1484	28.24	70922
213.1121	27.34	48047
398.226	28.23	50953
368.2514	32.63	85062
261.0085	17.75	146080
627.3737	27.68	7324
734.4926	32.65	68539
300.131	24.27	179044
347.2216	28.42	393972
731.4695	27.36	5166
620.3867	29.29	3848
305.1023	19.44	41092
301.1655	37.24	336365
733.4899	32.63	156945
177.9917	89.42	4644
253.1425	26.52	77807
403.0859	13.6	33240
285.1845	25.38	13622
310.2012	26.6	116783
793.4364	28.23	28120
322.2092	25.36	13908
384.246	26.48	12131
592.0967	13.6	53698
591.0937	13.58	131973
361.1995	25.38	38533
309.1082	12.55	7465
323.2217	31.45	58926

Structure of Excel file for submission to MetaboAnalyst

MetaboAnalyst data entry

Comma Separated Values (.csv) :

Data type : Concentrations Spectral bins Peak intensity table

Format:

Data file : No file chosen

Zippped Files (.zip) : For WinZip 12.x, choose "Legacy compression (Zip 2.0 Compatible)"


Data type : NMR peak list MS peak list MS spectra

Data : No file chosen


Pairs : No file chosen (required for paired comparison)

Zipped Files (.zip) : For WinZip 12.x, choose "Legacy compression (Zip 2.0 Compatible)"

Data type : NMR peak list MS peak list MS spectra

Data :  Cutts_test.zip

Pairs : No file chosen (required for paired comparison)



Initial parameter setting

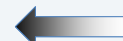
MS peak grouping :

Peaks need to be matched across samples in order to be compared. For two-column format (mass and intensities), peaks are grouped by their m/z values. For three column data (mass, retention time, and intensities), the program will further group peaks based on their retention time. Users need to supply tolerance values in order to proceed. **Here are some suggested values:** mass tolerance - 0.25 (m/z); retention time - 30 (seconds) for LC-MS peak, and 5 (seconds) for GC-MS peaks.

Please note, If a sample has more than one peak in a group, they will be replaced by their sum; some groups will be excluded if none of the classes has at least half its samples represented. Finally, the program create a peak intensity table in which each sample occupies a row and each column represents a peak group identified by the median values of its position (m/z and/or retention time).

Mass tolerance (m/z)

Retention time tolerance (unit)



Initial evaluation

MS peak processing information

The uploaded files are peak lists and intensities data.

A total of 6 samples were found.

These samples contain a total of 12072 peaks.

with an average of 2012 peaks per sample

A total of 925 peak groups were formed.

Peaks of the same group were summed if they are from one sample.

Peaks appear in less than half of samples in each group were ignored.

More assessment of the data

Data processing information:

Checking data content ...passed

The uploaded files are peak lists and intensities data.

A total of 6 samples were found.

These samples contain a total of 12072 peaks.

with an average of 2012 peaks per sample

2 groups were detected in samples.

Samples are not paired.

All data values are numeric.

A total of 0 (0%) missing values were detected.

By default, these values will be replaced by a small value.

Click **Skip** button if you accept the default practice

Or click **Missing value imputation** to use other methods

Data filtering

- Interquartile range (IQR)
- Standard deviation (SD)
- Median absolute deviation (MAD)
- Relative standard deviation (RSD = $SD/mean$)
- Non-parametric relative standard deviation (MAD/median)
- Mean intensity value
- Median intensity value
- None

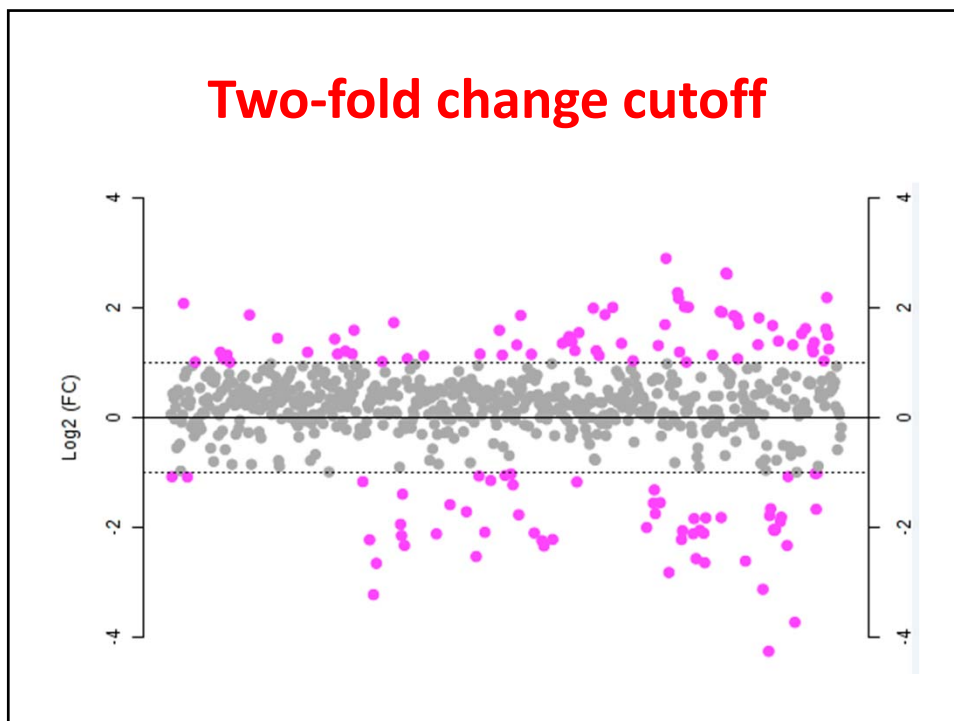
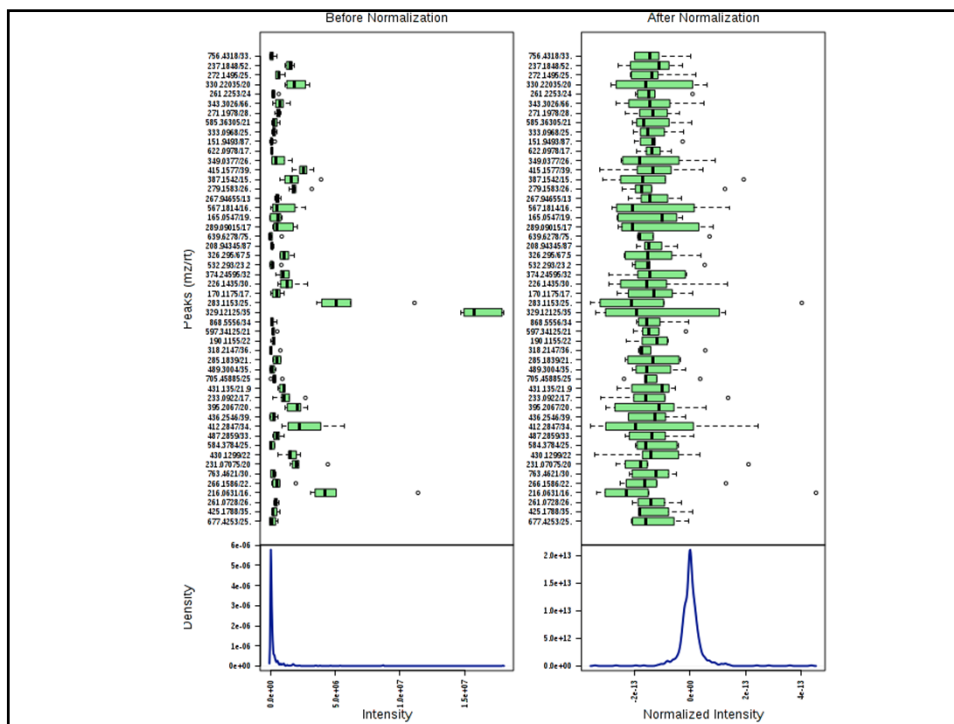
Data transformation and scaling

Data transformation

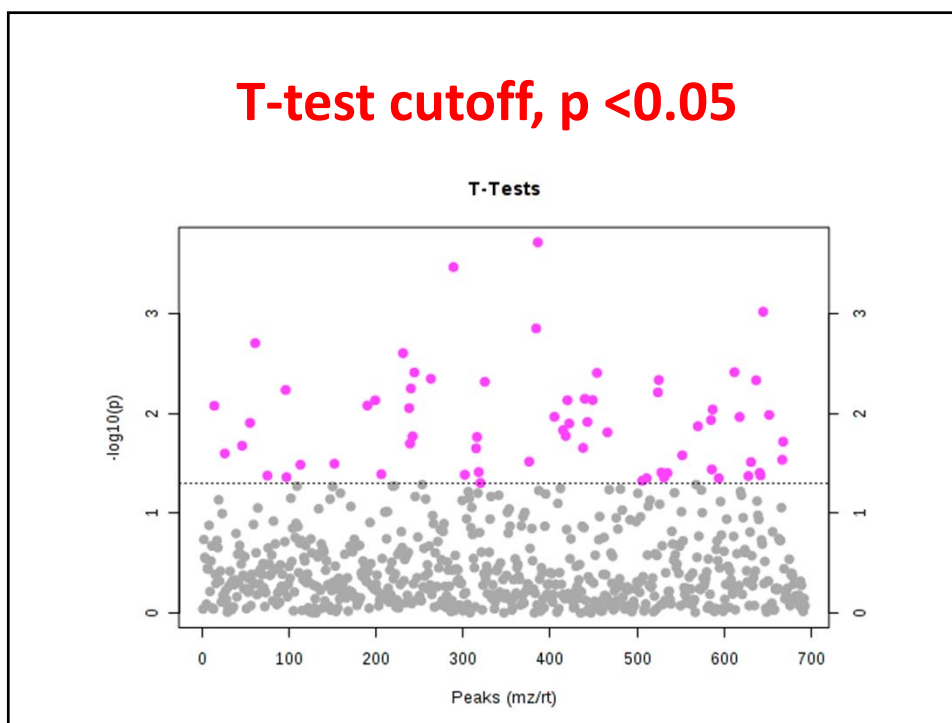
- None
- Log transformation (generalized logarithm transformation or glog)
- Cube root transformation (take cube root of data values)

Data scaling

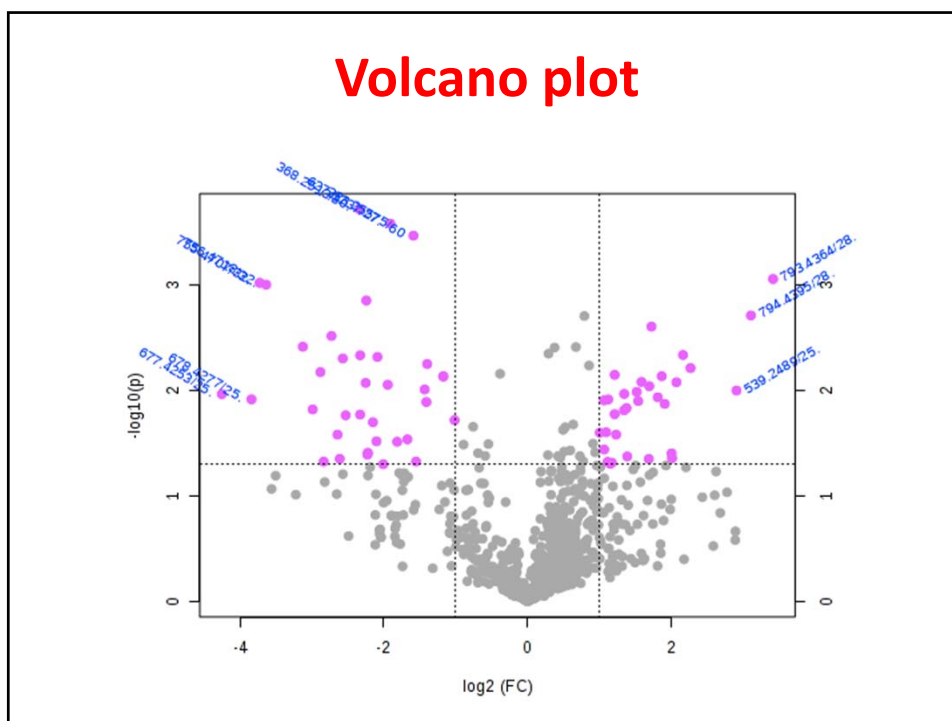
- None
- Autoscaling (mean-centered and divided by the standard deviation of each variable)
- Pareto Scaling (mean-centered and divided by the square root of standard deviation of each variable)
- Range Scaling (mean-centered and divided by the range of each variable)

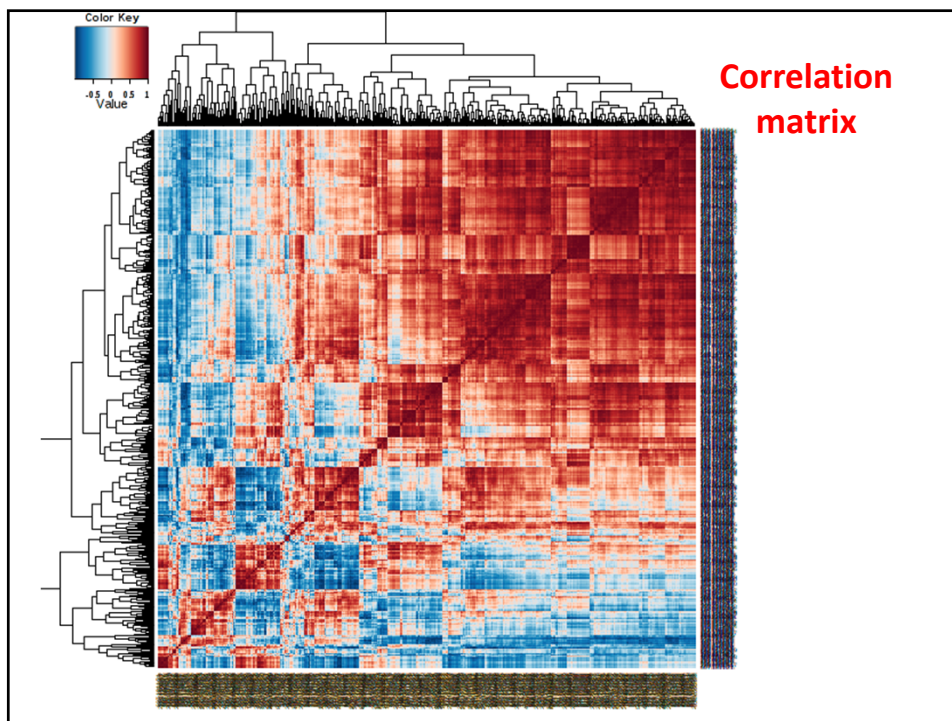


T-test cutoff, $p < 0.05$



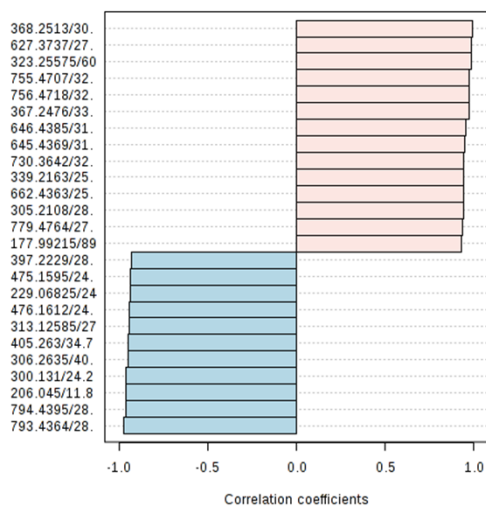
Volcano plot



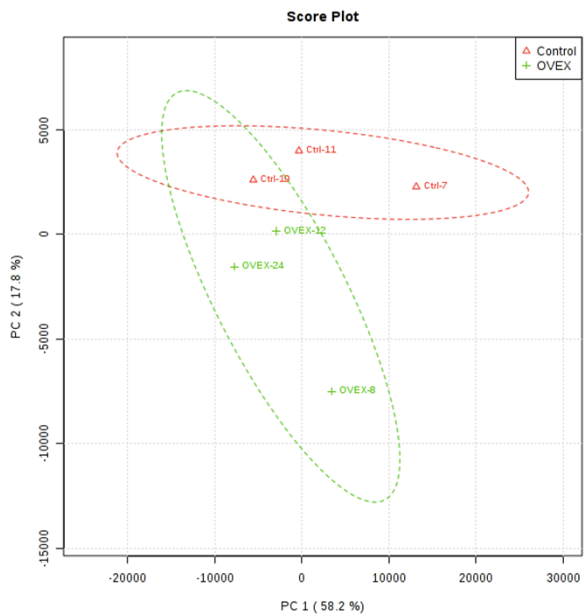


Peaks differentiating Ctrl/OVEX

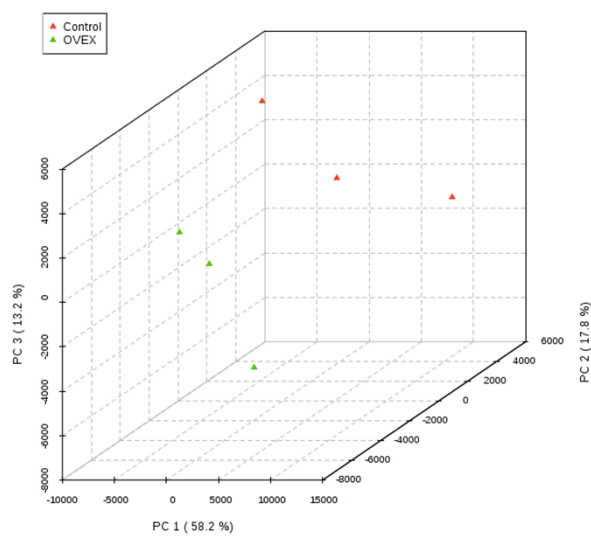
Top 25 peaks (mz/rt) correlated with the 1-2

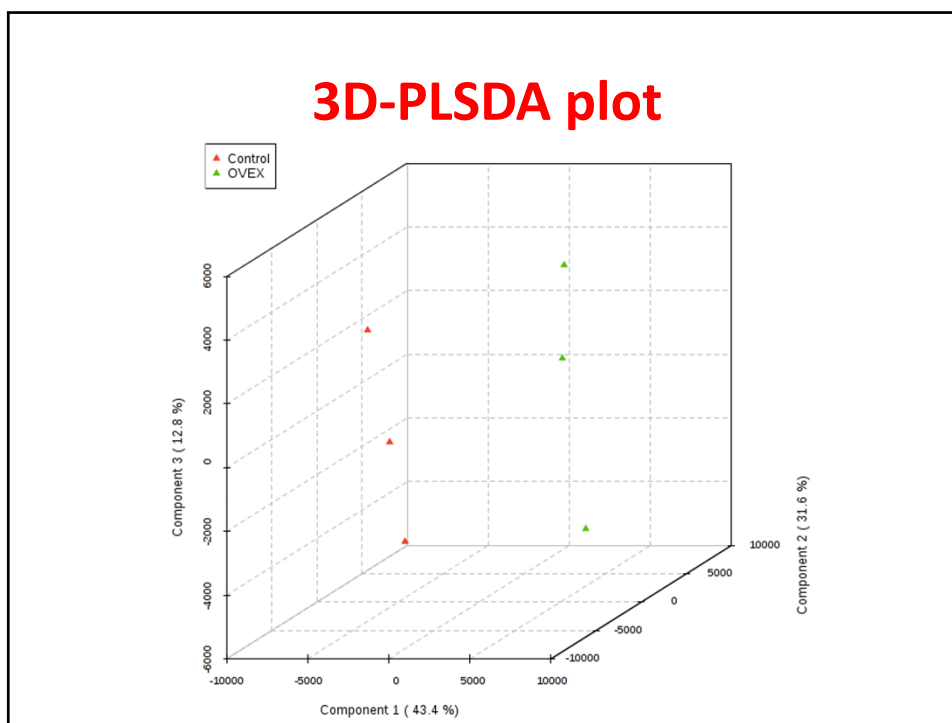
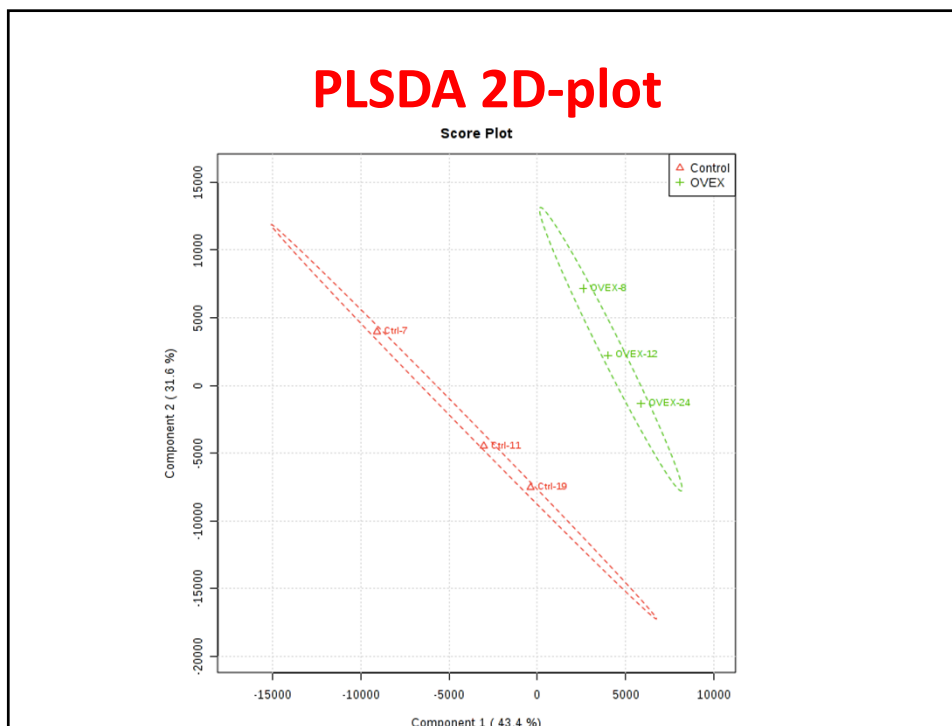


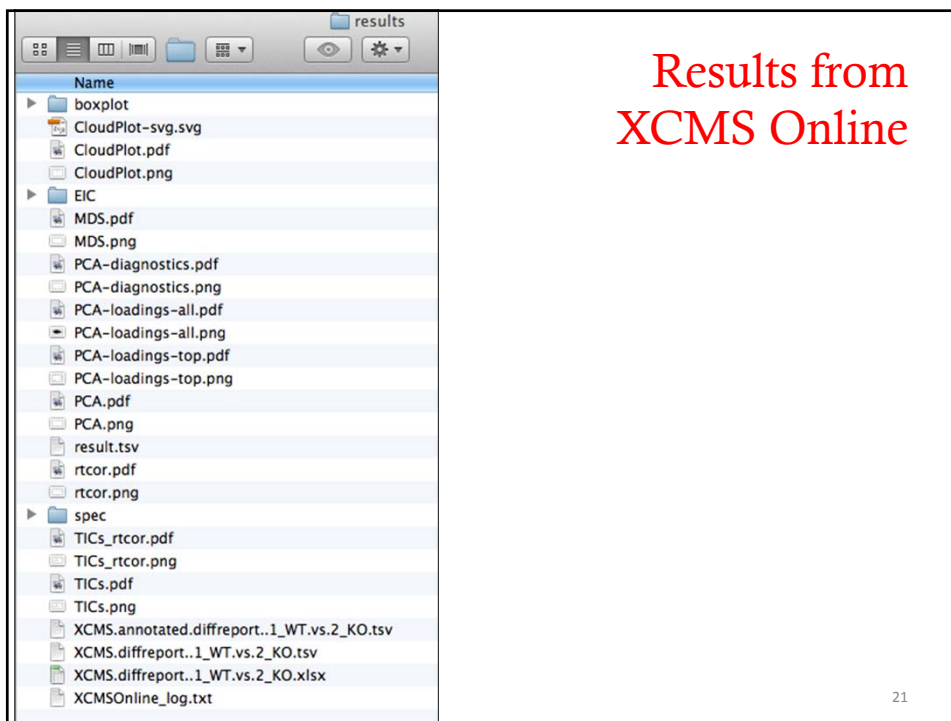
PCA 2D-plot



3D PCA plot







Results from
XCMS Online

21

For in-depth statistical analysis and data interpretation,
please make an appointment with a biostatistician.

Thank you!

23