XCMS differentiation

David L. Tabb, Ph.D.

dtabb1973@gmail.com

With many valuable slide contributions from H. Paul Benton and Gary Siuzdak, The Scripps Research Institute

Overview

- Statistics for difference testing in biological mass spectrometry

- Experimental designs paired with statistical tests in XCMS.

“Masters of the x-axis”

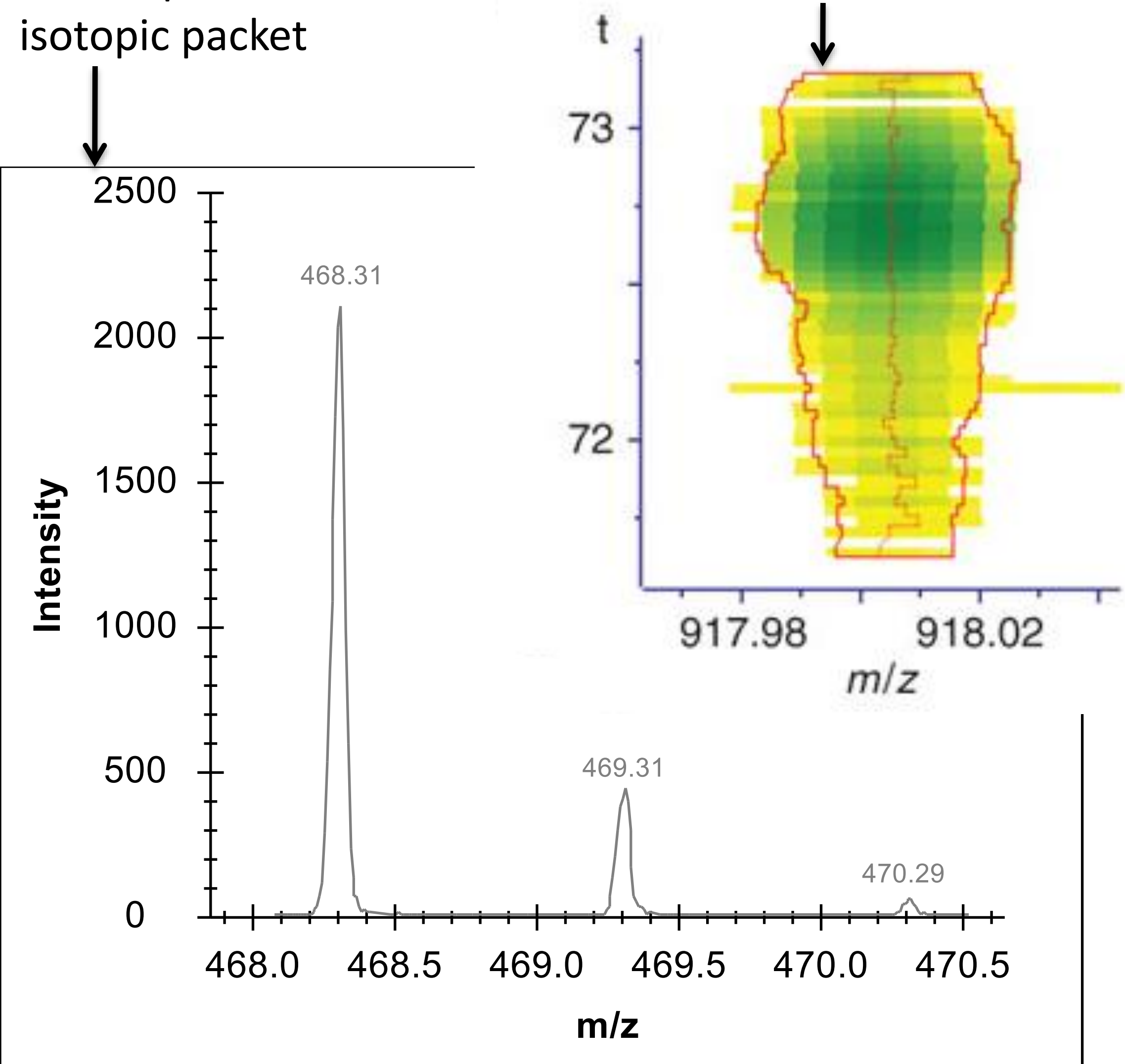
- Mass spectrometers offer very high precision for m/z values, with FT mass analyzers varying well within 10 ppm for mass measurement.
- Intensity measurement, however, is far more variable, especially in low signal-to-noise ions.
- Intralaboratory variation for targeted quantitative experiments generally range up to 10% CV.

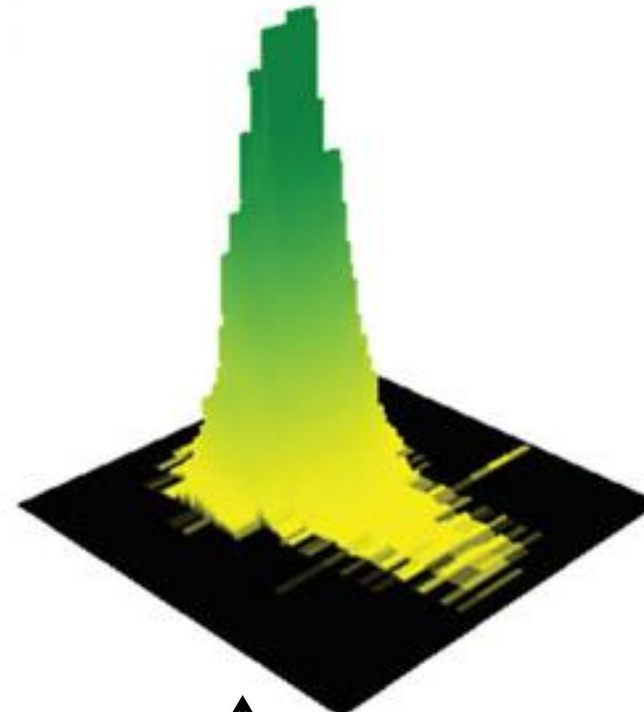
Addona et al. *Nature Biotech*. (2009) 27: 633

MS Intensity as quantitative data

Close-up of isotopic packet

Peak through time





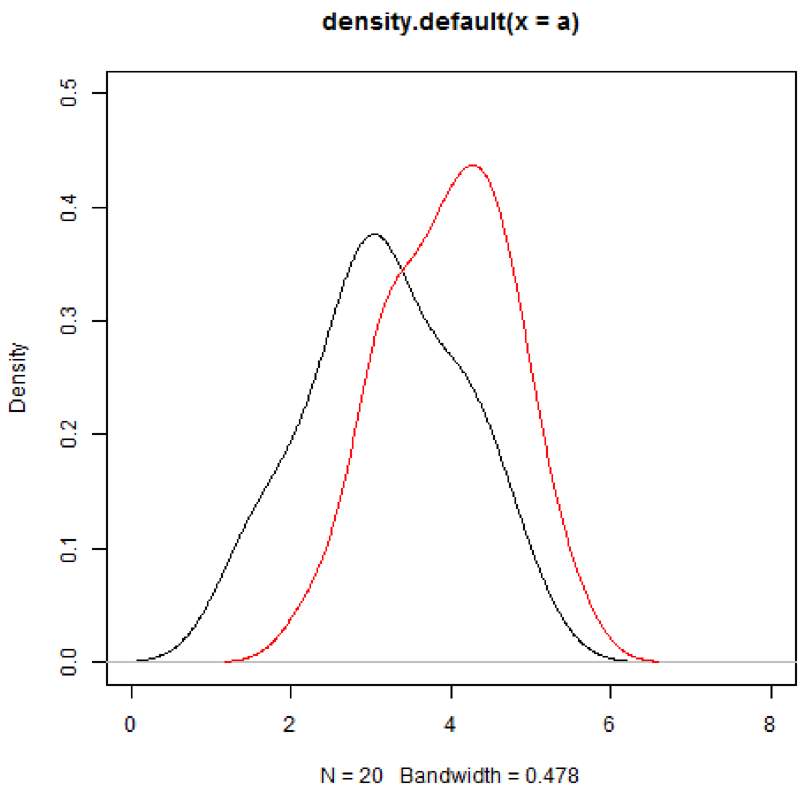
3D volume reconstruction

Upper image from Cox (2008) *Nat. Biotech.* 26: 1367-1372.

Parametric vs. Nonparametric

- Parametric tests (such as Student's t-test) expect that data come from a specified distribution; if this is untrue, the p-value is not accurately estimated.
- Non-parametric tests (such as the Mann-Whitney/Wilcoxon U-test) do not rely upon distributional assumptions. For a given data set size, non-parametric tests are slightly less likely to achieve significant p-values.

Are these distributions different?



- Welch two-sample t-test yields p-value of 0.006355.
- Wilcoxon rank sum test yields p-value of 0.009484.
- Model samples 20 replicates from normal distribution with mean=3 (A) or mean=4 (B), std. dev.=1.

Student’s t-test

- Created in 1908 by Gosset for monitoring beer quality at Guinness brewery in Dublin, Ireland.
- Compares difference between sample means to variance of samples to produce t statistic.
- Yields p-value that the difference between means would be at least this large by chance.
- Assumes Gaussian (normal) distribution of continuous values; counts are problematic.

Welch later added a correction for data in which variances were unequal.

Mann-Whitney-Wilcoxon U-test

- Imagine the values from A are blue marbles and the values from B are red marbles, with sizes representing their values.
- Sort the marbles by size. Do the colors separate? If so, the p-value is lower.
- This non-parametric alternative to t-test works best when many samples are available. Don't use this in a 3 vs. 3 test!

The curse of dimensionality

- Many tests evaluate differences for each ion independently. Using them on 1,000 ions is almost sure to yield "significance."
- Multiple testing correction by Bonferroni or Benjamini-Hochberg procedures adjusts these p-values to reflect many applications of test.
- Permuting column labels can help to estimate expected p-values from random data.

Bland, Altman (1995) *British Medical J*. 310: 170.
Benjamini, Hochberg (1995) J Royal Stat. Soc. 57: 289-300.

Test types in XCMS

- Two-cohort comparisons
- Paired comparisons
- Meta-analysis of disparate studies
- Multiple group comparisons

Table 1. Statistical Tests and Interactive Visualization Tools Implemented within the XCMS Online Platform for Univariate Analysis (One Variable at a Time)

univariate statistical analysis			
experimental design	parametric test (data follow normal distribution)	nonparametric alternative test (no assumption about data distribution)	visualization tools
comparison of two independent groups	independent *t* test (Welch *t* test)	Mann–Whitney	interactive two-group cloud plot, box–whisker plots, extracted-ion chromatograms
comparison of two dependent (paired or matched) groups	paired *t* test	Wilcoxon signed-rank	
comparison of ≥ three independent groups	one-way ANOVA with post-HOC	Kruskal–Wallis	interactive multigroup cloud plot, box–whisker plots, extracted-ion chromatograms
comparison of ≥ three dependent (matched or related) groups	repeated measures ANOVA	Friedman	

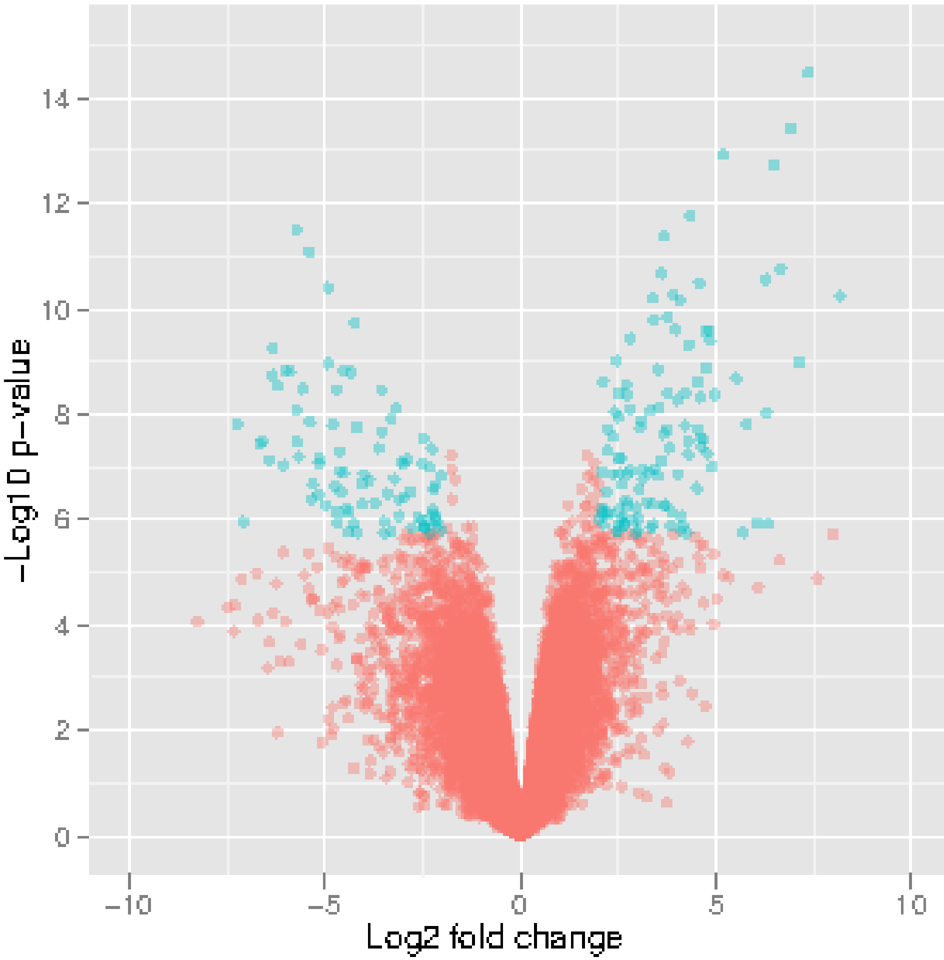
Gowda et al. *Anal. Chem*. (2014) 86: 6931

Paired comparisons

Label	Before	After
Patient A	5.3	5.5
Patient B	4.7	4.9
Patient C	5.7	5.9
Patient D	4.9	5.1
Patient E	5.0	5.2
Patient F	4.8	5.0
Patient G	3.6	3.8
Patient H	4.5	4.7
Patient I	6.1	6.3
Patient J	5.6	5.8

- Do not compare the “before” value for patient A to the “after” value for patient B.
- These are *paired* data, and using an unpaired test will miss the obvious difference between the two columns.

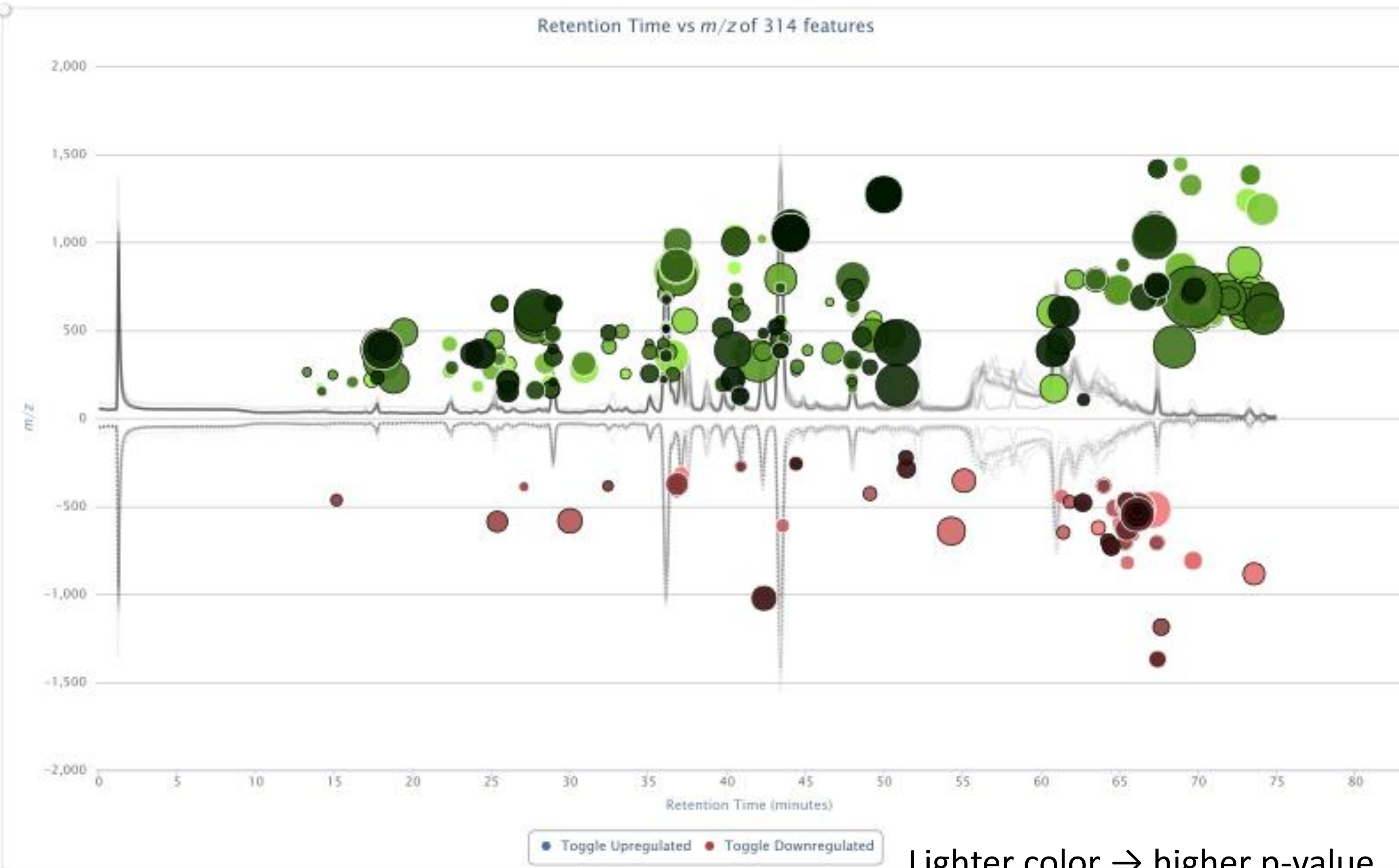
Volcano plots



- Standard visualization for significance and magnitude of change
- Items too close to the center don't change enough to matter.
- Items too close to the bottom do not give significant p-values.

http://bioinformatics.knowledgeblog.org/2011/06/21/volcano-plots-of-microarray-data/

XCMS Interactive Cloud Plot

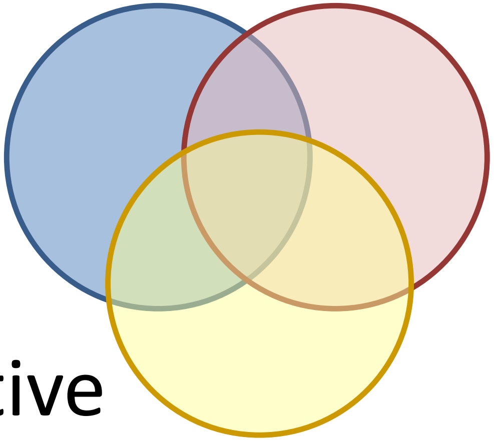


https://xcmsonline.scripps.edu/

Lighter color → higher p-value

Larger radius → bigger fold change

Meta-analysis



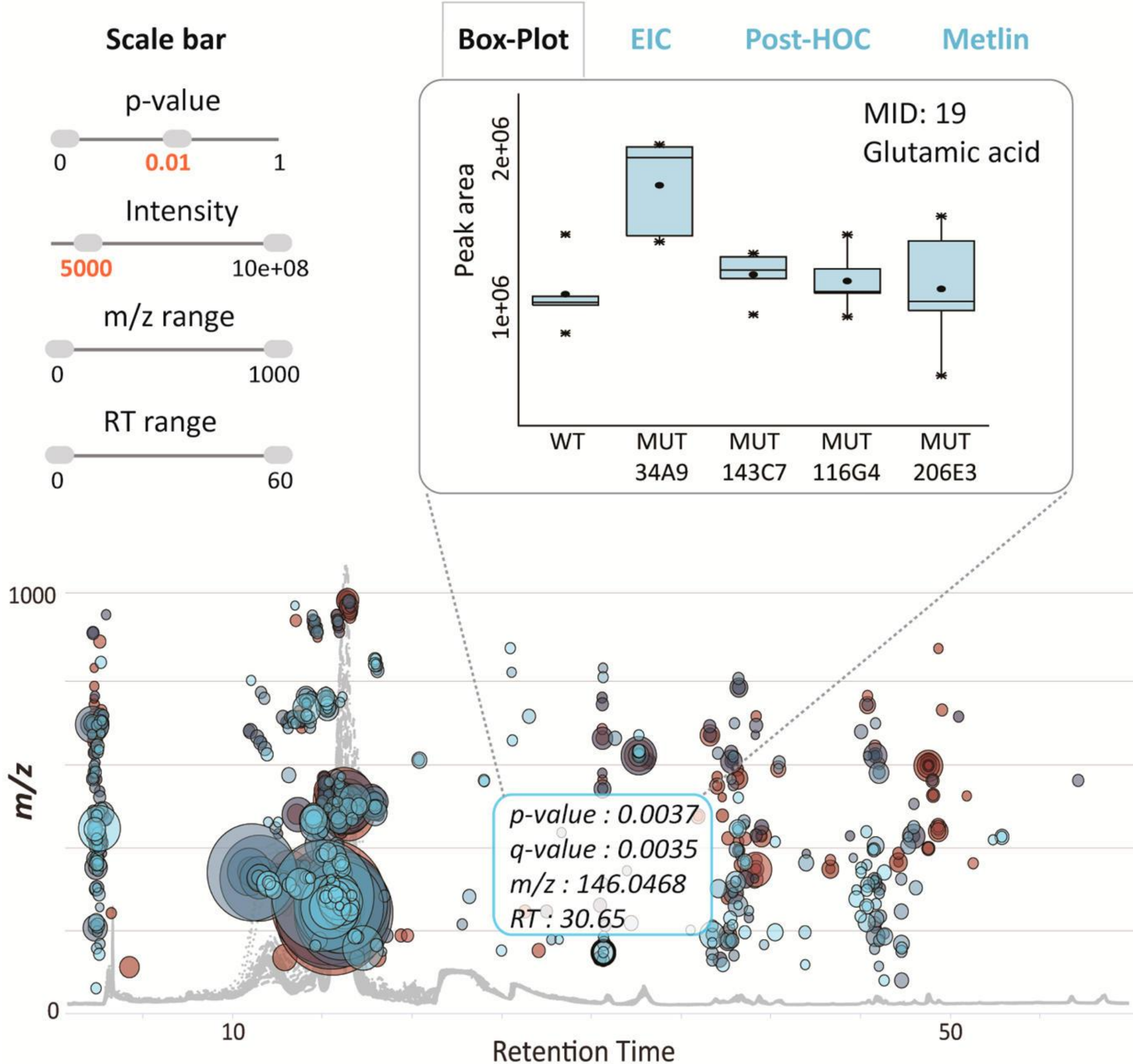
- My experiment produced 147 putative differences, and prior experiments yielded 124 and 78 differences, respectively.
- How do these differential ions intersect among these three studies?
- Venn Diagrams are a standard way to show these overlaps, and XCMS can for visualize overlaps for more than three groups.

This type of meta-analysis reanalyzes the data sets, mapping LC-MS data among experiments.

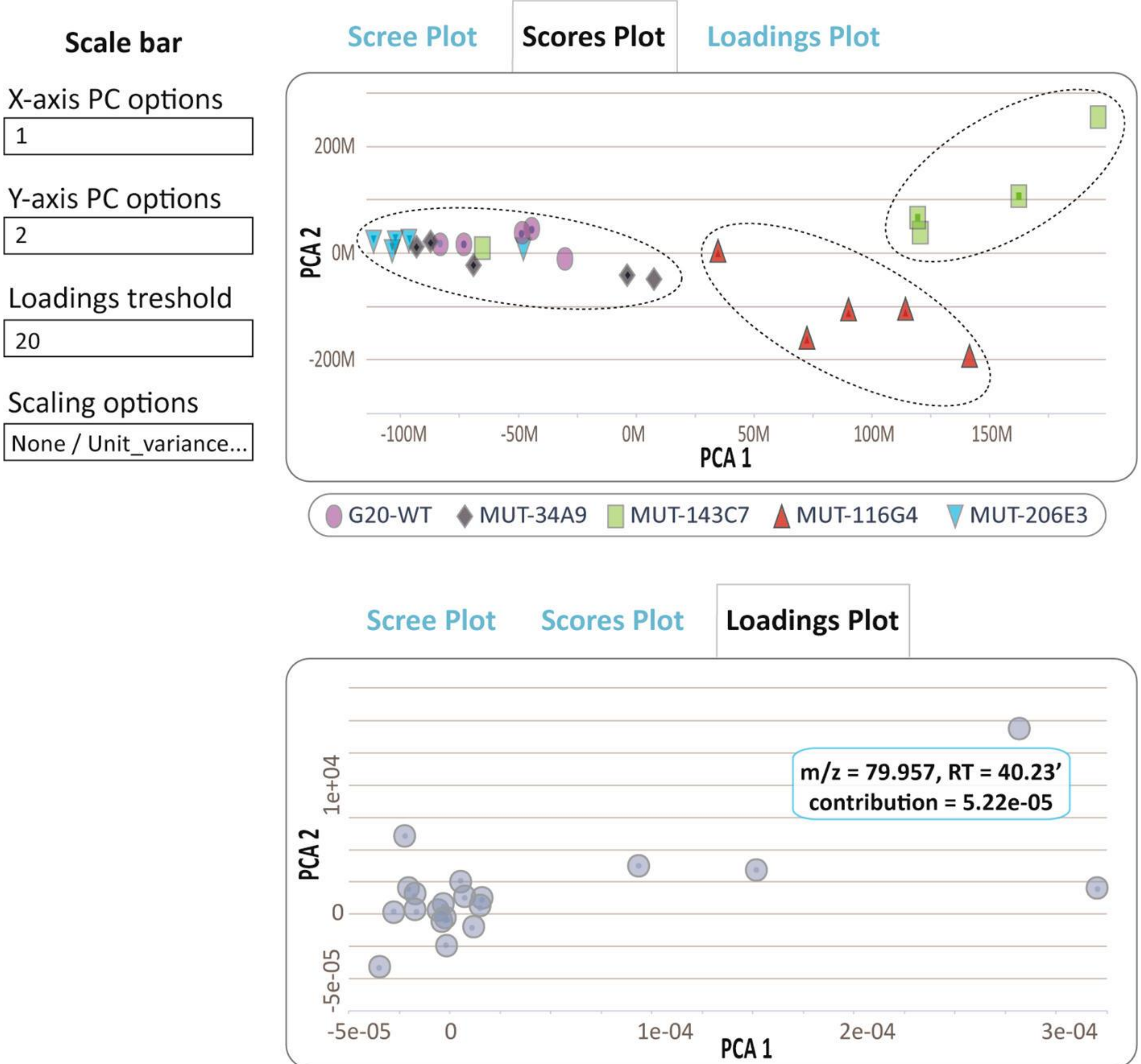
Multi-group comparison

- If you have three groups, it's not best to compare A to B, B to C, and then A to C.
- XCMS uses the univariate ANOVA (parametric) and Kruskal Wallis (nonparametric) tests for multi-group comparison (see next slide).
- To account for covariance among metabolites, XCMS can employ PCA to recognize significant metabolites among groups (two slides ahead).

Gowda et al. *Anal. Chem*. (2014) 86: 6931



Gowda et al. *Anal. Chem.* (2014) 86: 6931



Gowda et al. *Anal. Chem.* (2014) 86: 6931

Takeaway messages

- XCMS is capable of more complex designs than cohort-A-vs-cohort-B comparisons.
- Keep in mind that when many statistical tests are performed, one must correct for it. Q-values from XCMS take multiple testing into account.
- Untargeted biomarker discovery experiments suffer more from multiple comparisons than do targeted experiments on sets of candidates.