Targeted Metabolomics & Proteomics Laboratory

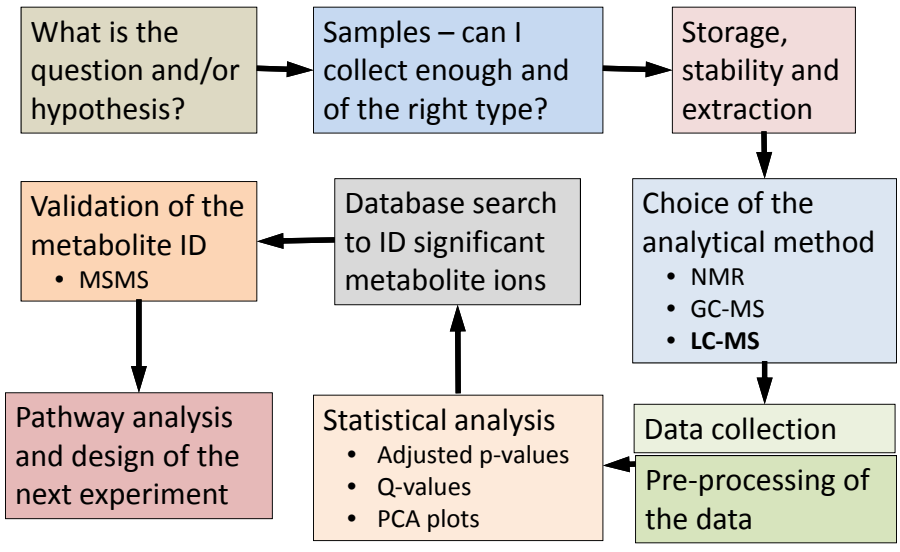
Introduction to metabolomics

Stephen Barnes, PhD

Professor of Pharmacology & Toxicology

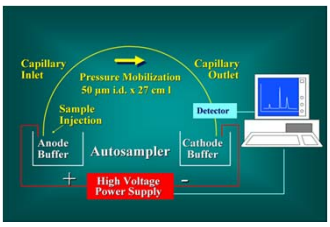
sbarnes@uab.edu; 205 934-7117

Metabolomics workflow



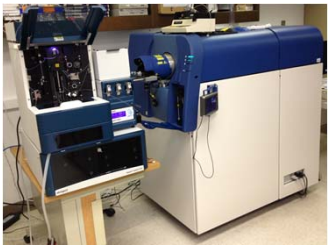
Platforms for metabolomics analysis

NMR



Capillary electrophoresis

GC-MS



LC-MS/MSMS

Experimental Design

- **LC-MS analysis in a metabolomics experiment will generate >1,000 and even as high as 10,000 discernible and reproducible features**
  - **All things being equal and using a p-value cutoff of 5%, under the null hypothesis (H_0) 50-500 of the above will appear to be (falsely) significant.**
  - **Therefore, it is critical to design the experiment to ensure the likelihood of a meaningful outcome.**

Selecting the problem

- Power is hard to pre-estimate in a metabolomics experiment
- Exploring a subtle phenomic difference would require a very large number of samples/patients/animals/cells
- Best problems are ones where there are clear cut differences
- Samples of three per group are not adequate for statistical interpretation
- The best design is where the phenomic event varies across the experimental cohort

Avoiding bias

- Variability is unavoidable, but should not be added to unnecessarily
- Carefully control the biological variation
- All sources of non-biological variation should be both minimized (if possible) and evenly distributed across all groups
- Randomly process the samples from each group and randomize the order in which the samples are analyzed
  - Requires the services of an experienced statistician

Measurement Issues

- Sources of errors at the prep stage
  - Within subject variation
  - Within tissue variation
  - Contamination by cleaning solvents
  - Evaporation of volatiles
  - Calibration uncertainty (LC retention times; masses of ions)

Executing the design

- Make sure that all the samples are collected in the same way
  - Have a standard operating procedure
  - If collecting blood/serum/urine, buy enough sample tubes from the same lot for the entire study
  - Label the samples well and store them in random order in a rack in the freezer

Sample Size and Power Calculation

Often the number of samples to be used for the experiments is dictated by the reality of resources available, not science.

- **How much money is available for the experiment?**
- **What is the cost per sample?**
- **Thus, sample size = $ available through NIH/ cost per sample**

Costs for Metabolomics at UAB

- **Step 1: untargeted LC-MS analysis**
  - Need to run each sample in positive and negative and on reverse phase and normal phase
  - We limit the run times (with re-equilibration) to 30 min – so, 4 x 30 min per sample (2 hr)
  - Basic LC-MS charge is **$175/hr**, so $350 per sample
  - Preliminary run with 3 samples in each group would cost $2,100, **with discount $1,890**
  - *Alternative, 2 groups x 3 samples on reverse-phase and in positive mode only – **$525** – **good for a pilot study***

Costs for Metabolomics at UAB

- **Step 1: now you have a NIH grant**
  - For a clinical study consisting of 50 samples in each of two groups and just reverse-phase and positive mode, cost would be **$8,750**
  - For all four run conditions, cost is **$35,000**
  - In addition, training will be provided to use **XCMS,** a program developed at Scripps that is freely available, to process the LC-MS data
  - This software will determine which of the ions are statistically significant between the groups.

Fundamentals of Metabolism

- **"Metabolites" represent a very wide range of chemical structures**
  - **Volatiles**
    - **Gases (H_2, CO_2)**
    - **Low boiling point (acetone, skatole)**
  - **Ionic**
    - **Negatively charged (organic acids)**
    - **Positively charged (amines, amino acids, oligopeptides)**
  - **Neutrals**
    - **Hydrophilic (Glucose)**
    - **Hydrophobic (vitamins A, D, E K; cholesterol esters)**
- **Mol Wt <1,500 Da**

Fundamentals of LC separation

- The goal in untargeted metabolomics is to collect as much data as possible
- Requires two types of chromatography
  - Reverse-phase columns (C_4, C_8 or C_{18} hydrocarbons attached to silica)
    - Separation on the basis of hydrophobicity
    - Increasing gradient of acetonitrile or methanol in aqueous
  - Normal or HILIC phases
    - Separation on the basis of hydrophilicity
    - Decreasing gradient of acetonitrile or methanol in aqueous

Fundamentals of the interface

- Electrospray ionization (ESI)
  - For compounds that are naturally charged at the pH of the mobile phase
    - Positive
    - Negative
- Atmospheric pressure chemical ionization (APCI)
  - Good for compounds that do not naturally carry a charge
    - Positive
    - Negative

Mass spectrometer analyzers

- **Quadrupole**
  - A mass filter with high sensitivity, but low mass accuracy and mass resolution, slow scan speed
- **Time-of-flight (TOF)**
  - Good mass accuracy and mass resolution, highest scan speed
- **Ion motion analyzers**
  - Orbitrap, Fourier Transform ion cyclotron resonance (FT-ICR)
  - Highest mass accuracy and mass resolution, but slow compared to the quadrupoles and TOF detectors

Other parameters to consider

- **pH of the mobile phase**
  - 0.1% formic acid
  - 10 mM NH_4OAc
- **Temperature**
  - Must be kept constant
  - Elevated temperature lowers solvent viscosity
- **Chemical derivative**
  - Reagents for keto- and aldo-groups

Column size, flow rate and sensitivity

- **Regular flow (2.1-4.6 mm ID)**
  - **200 – 1000 μl/min (uPLC)**
- **Microflow (0.5-1.0 mm ID)**
  - **1-100 μl/min (10-200 times more sensitive)**
- **Nanoflow (25-500 μm ID)**
  - **25-500 nl/min (800-1000 times more sensitive)**
- **Column lengths are 10-20 cm**
- **Nanocolumns best in a LC-on-Chip format**
  - **Can be made more reproducibly and easier to maintain at constant temperature**

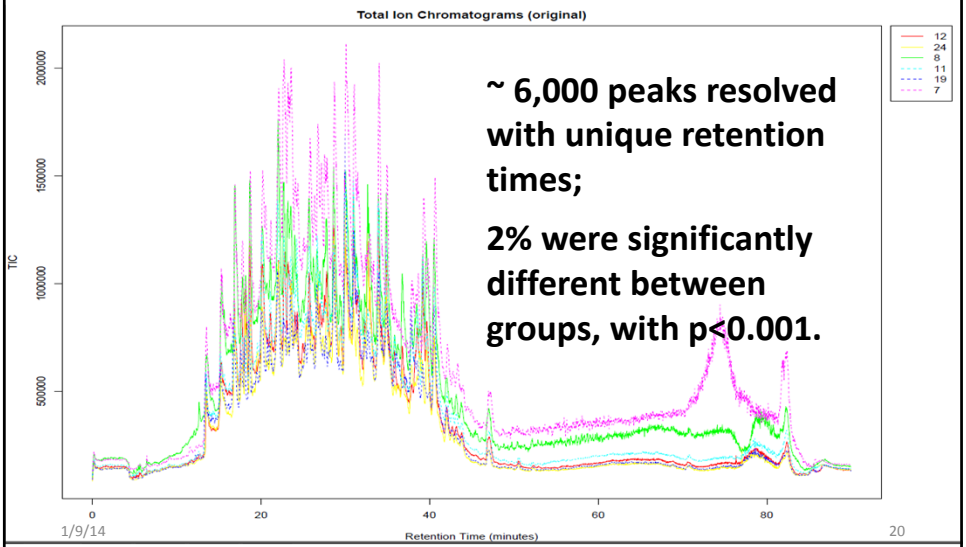
What data are collected in LC-MS?

- **Totally untargeted LC-$(MS)^1$ analysis**
  - **Collect successive high resolution (~40,000)/high mass accuracy (2-3 ppm) mass spectra**
  - **All data (over the specified mass range) are collected**
  - **Acquisition period is 100 msec for Q-TOFs but longer for Orbitraps and FT-ICR instruments**

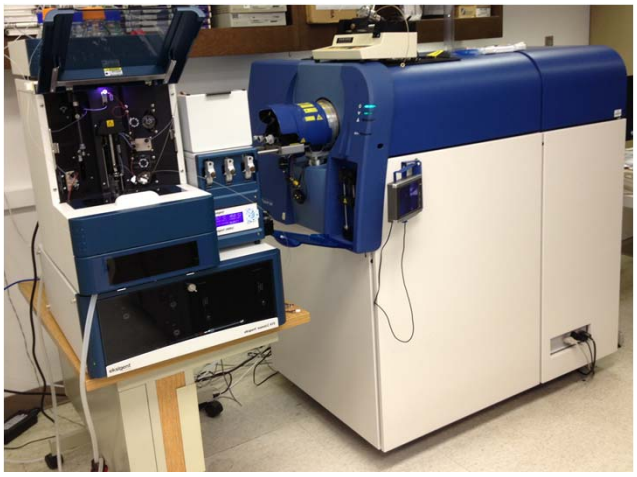
Untargeted, data-dependent analysis

- Think in terms of a 1 sec duty cycle
- For the first 100 msec collect a high resolution (~40,000)/high mass accuracy (2-3 ppm) mass spectrum
  - From the MS1 spectrum, select the most abundant ions: on these MSMS spectra are collected every 50 msec
  - If the MSMS of an ion was collected in the previous 1 sec, it is put on an exclude list for the next 30 sec

LC-MS: total ion currents of ethyl acetate extracts of urines from sham-OVX and OVX

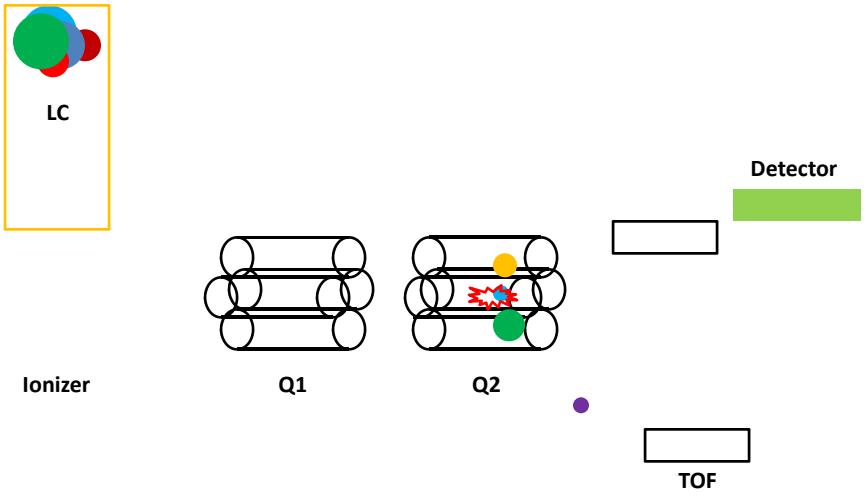


Metabolomics analysis



AB Sciex 5600 TripleTOF using nanofluidics

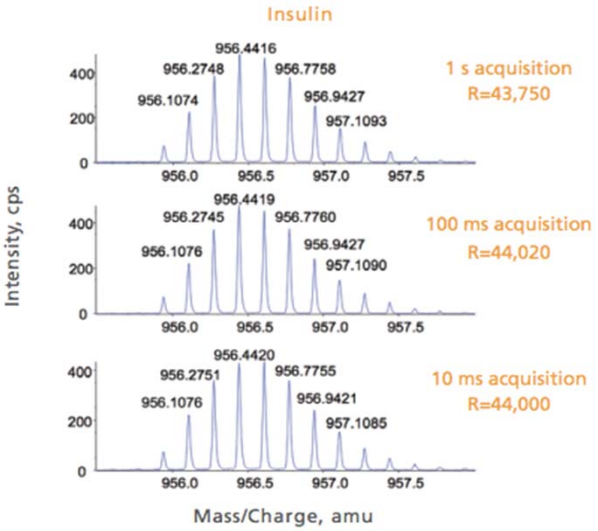
Liquid Chromatography-Triple TOF Mass Spectrometry (LC-QQTOF MS)



LC-MS Conditions

- 15 cm x 75 μm i.d. ChipLC in a temperature-controlled environment
- 0-50% gradient of acetonitrile in 0.1% formic acid over 120 min at 300 nl/min
- Nanoelectrospray ionization in the positive mode
- First 250 msec
  - MS1 scan over the range from 150-1000 *m/z* using the TOF analyzer
- 250-2250 msec
  - MS2 scans for 50 msec on the 20 most abundant ions to obtain product ion spectra – ions selected by the quadrupole filter and analyzed by the TOF

Advantage of the TripleTOF



Unlike FT-ICR and Orbitrap detectors, a TOF analyzer is not dependent on acquisition time. It is ideal for high-speed metabolomics analysis.

Processing of LC-MS metabolomics data

- **Need to align the peaks according to their *m/z* values and their retention times**

https://xcmsonline.scripps.edu/

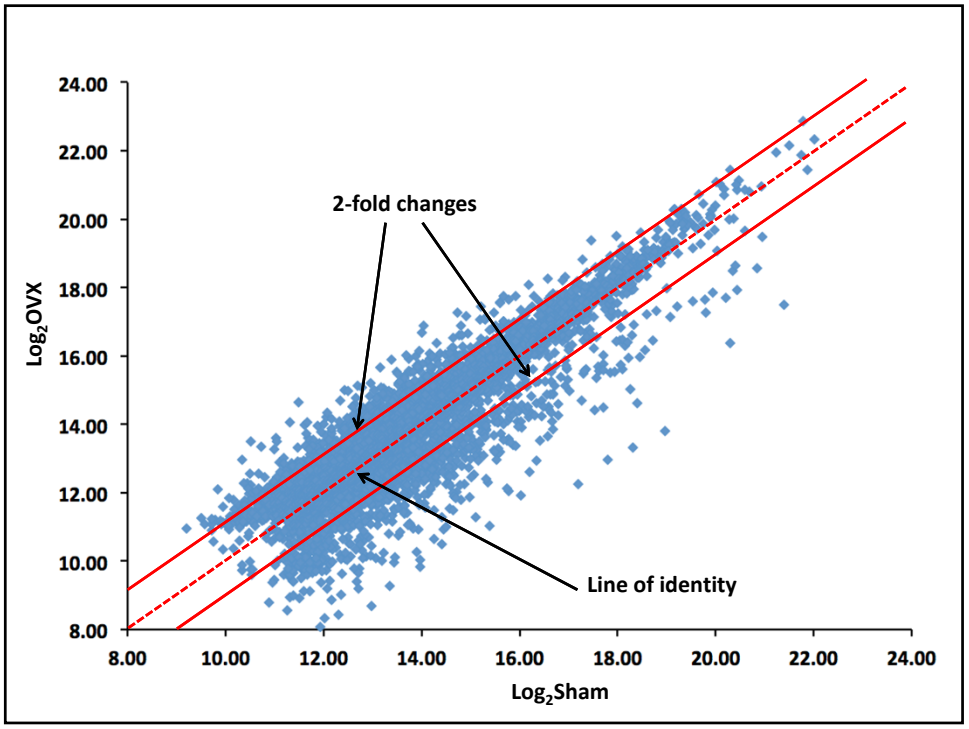
XCMS run locally under R on a PC or a Mac – command line

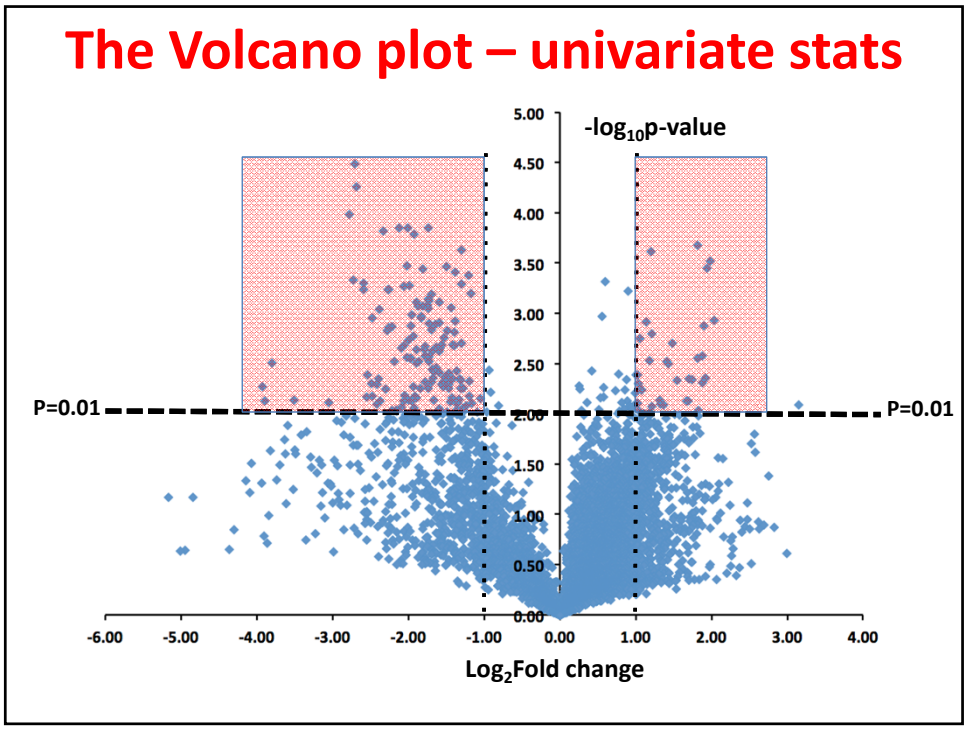
Statistical analysis

Output of XCMS analysis (abridged)

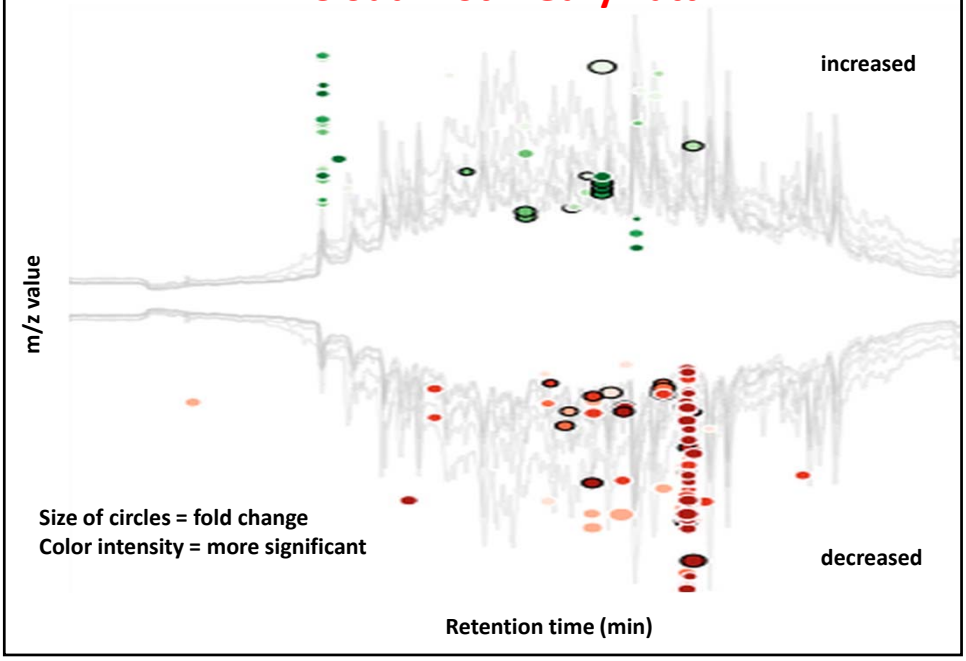
fold	log2fold	pvalue	`-logP	updown	mzmed	rtmed	X_Sham	SD Sham	X_OVX	SD OVX
6.52	-2.704	0.00003	4.483	DOWN	733.7044	32.74	4458	211	684	231
6.43	-2.685	0.00005	4.260	DOWN	733.4420	32.73	26764	742	4162	1230
6.86	-2.778	0.00010	3.986	DOWN	893.5196	33.07	11779	821	1717	761
4.36	-2.125	0.00014	3.853	DOWN	370.8050	32.77	3505	103	804	182
4.01	-2.005	0.00014	3.850	DOWN	367.5347	32.74	3770	242	939	181
3.32	-1.732	0.00014	3.849	DOWN	627.3729	27.75	32883	1969	9898	1475
5.03	-2.329	0.00015	3.819	DOWN	736.9168	32.75	4270	280	850	309
3.80	-1.927	0.00016	3.784	DOWN	371.5100	32.75	3592	210	944	245
3.52	1.816	0.00021	3.675	UP	419.2028	28.28	9583	2230	33753	2355
2.45	-1.293	0.00023	3.630	DOWN	321.6359	32.82	2631	154	1074	151
2.30	1.203	0.00024	3.611	UP	422.0146	13.59	1762	232	4055	219
3.94	1.979	0.00030	3.521	UP	397.2190	28.28	49630	12341	195655	15643
4.04	-2.015	0.00034	3.468	DOWN	751.9241	32.77	3907	310	967	322
2.83	-1.499	0.00034	3.463	DOWN	601.3587	32.79	40835	2854	14446	2089
3.85	1.943	0.00035	3.450	UP	397.1969	28.25	49417	11421	190039	15415
3.51	-1.812	0.00036	3.438	DOWN	480.7295	32.82	3946	312	1124	230
2.60	-1.378	0.00039	3.406	DOWN	349.2509	32.68	394842	16847	151885	6886
2.30	-1.202	0.00042	3.374	DOWN	356.6478	32.78	3798	161	1651	237
6.61	-2.725	0.00047	3.330	DOWN	894.5268	33.04	7444	755	1126	576
1.52	0.600	0.00048	3.315	UP	274.2859	30.11	2995	117	4540	49
6.03	-2.592	0.00050	3.300	DOWN	893.5615	33.05	3374	192	560	308
2.45	-1.293	0.00051	3.289	DOWN	624.3335	32.83	18828	867	7685	1292
3.97	-1.988	0.00053	3.272	DOWN	990.5281	32.73	5873	544	1481	512
4.15	-2.054	0.00054	3.266	DOWN	527.2847	33.09	36935	3291	8896	3484
4.77	-2.255	0.00058	3.237	DOWN	368.2513	32.71	430969	34932	90317	18841

This file has 6201 lines of data


24.00
22.00
20.00
18.00
16.00
14.00
12.00
10.00
8.00
Log2OVX
2-fold changes
Line of identity
8.00
10.00
12.00
14.00
16.00
18.00
20.00
22.00
24.00
Log2Sham


The Volcano plot – univariate stats
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
-log10p-value
P=0.01
P=0.01
-6.00
-5.00
-4.00
-3.00
-2.00
-1.00
0.00
1.00
2.00
3.00
4.00
Log2Fold change

Cloud Plot – Gary Patti



What are the metabolites affected by OVX?

- Unlike proteomics, you cannot predict the metabolites from another domain
- Databases are being built
  - METLIN (Scripps) is attached to XCMS-online and is supplemented by the Human Metabolome Database (David Wishart)
  - ChemSpider is a comprehensive small molecule database maintained by the Royal Institute of Chemistry
- The largest number of unique metabolites detected in human biofluids come from what we eat

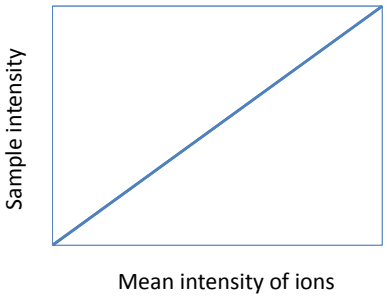
METLIN (at Scripps)

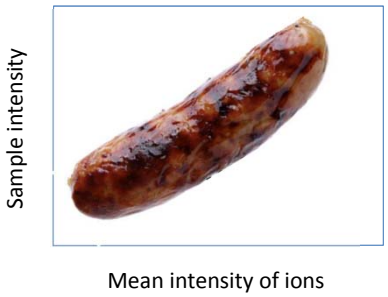
- Primary identification is based on MS1 data
  - If you can measure the mass of an ion to 1-2 ppm, you can usually write down its empirical formula
  - This excludes many compounds having the same nominal mass, e.g., 76 Da
- METLIN will assign an "identification" based on metabolites having a mass in its database within a user defined mass window (say 5 ppm) – if it's not in the database, no assignment can be made
- To improve the identifications, METLIN is adding MSMS spectral data (~5% of the database so far)

-Omics requires multivariate statistics

- Principal Components Analysis
  - 2D- and 3D-analysis
- Partial Least Squares Discriminant Analysis
- MetaboAnalyst (online free software)
  - http://www.metaboanalyst.ca/MetaboAnalyst/faces/Home.jsp

What are Principal Components?

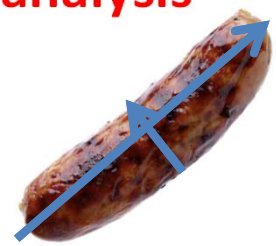




- **Think of the data as a short fat sausage**
- **The dimensions of the sausage represent the variation in the data**

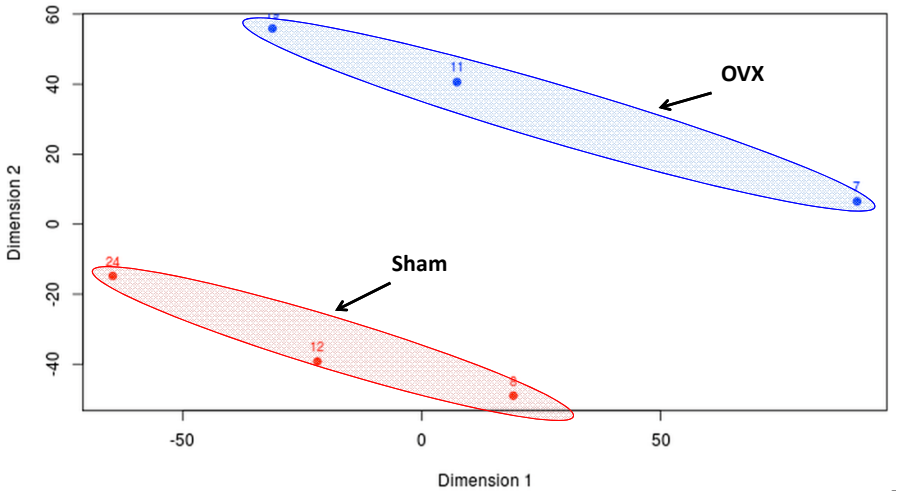
Principal components analysis

- **We look for the largest component of variation – that's the long axis of the sausage – PC1**



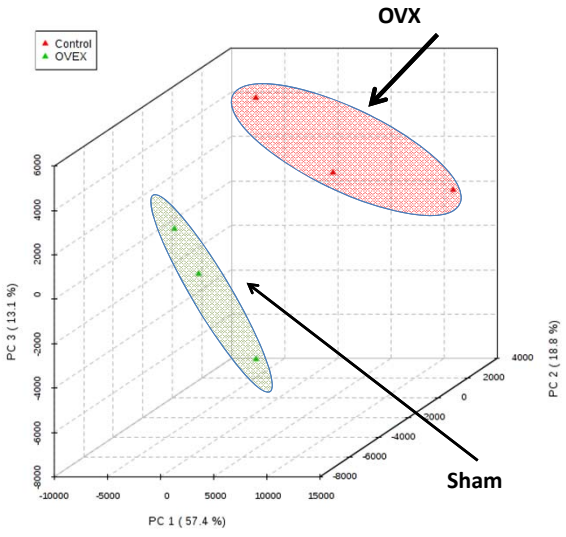
- **The second source of variation is the diameter of the sausage and is typically orthogonal to PC1 – it's PC2**
- **We then determine how much each ion contributes to the variation of PC1 and PC2 – weightings for each ion**
- **Then for each sample there is a PC1, PC2 (same as x,y data pair) that can be plotted in a 2D-plot**

2D-PCA analysis

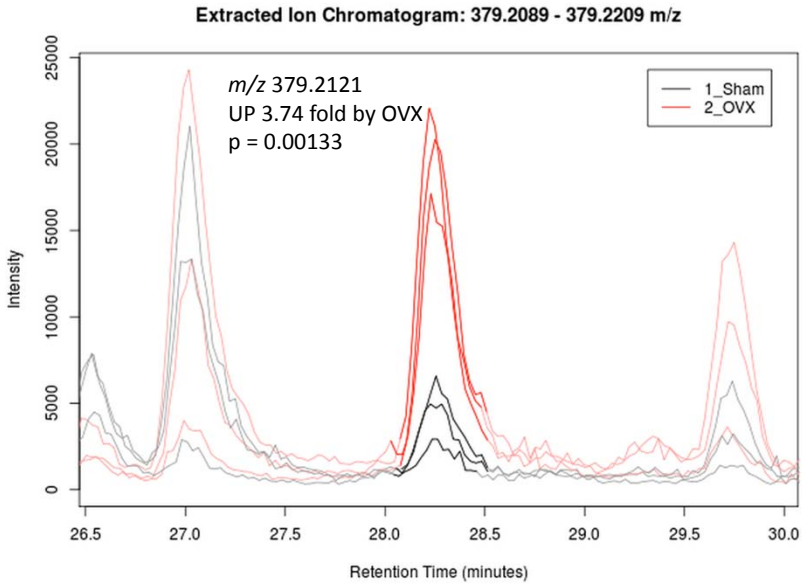


3D-PCA analysis

Two principal components did not explain all the variation, so a third principal component was introduced giving rise to a 3D-plot

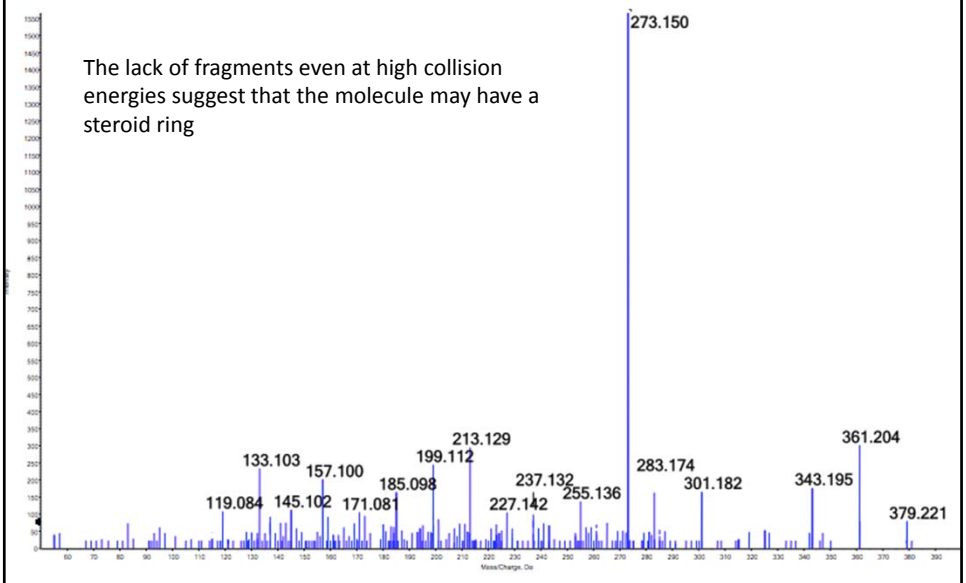


What significant ions are affected by OVX?



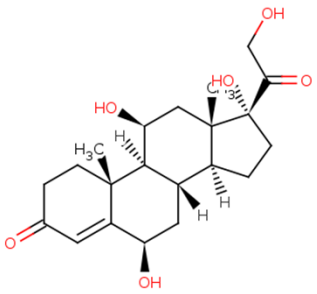
MSMS of m/z 379.2121

The lack of fragments even at high collision energies suggest that the molecule may have a steroid ring



METLIN ID of *m/z* 379.2121

6β-Hydroxycortisol	M+H
18-Hydroxycortisol	M+H
Nigakilactone A	M+H
Isohumulinone A	M+H
Humulinone	M+H



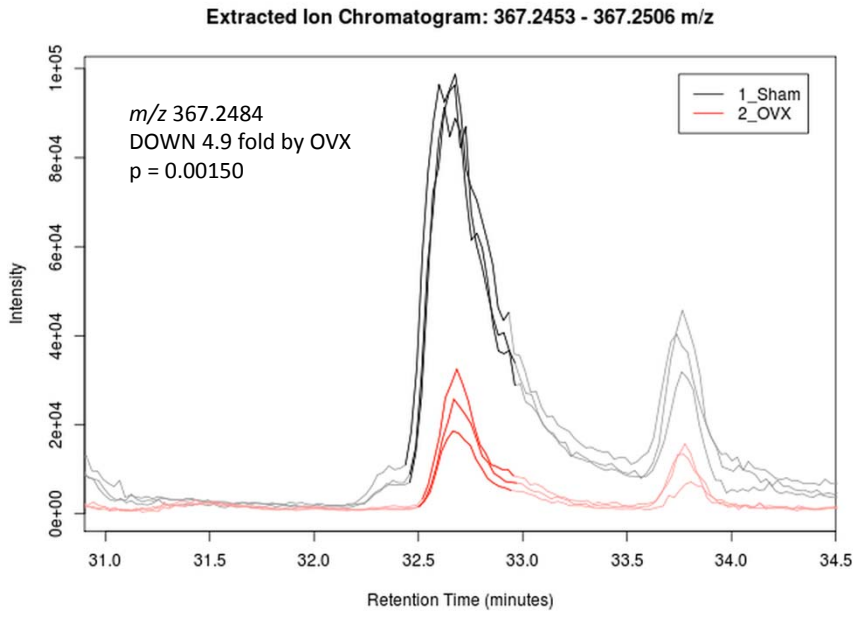
Clin Pharmacol Ther. 1998 Jun;63(6):617-22.

Urinary excretion of 6 beta-hydroxycortisol as an in vivo marker for CYP3A induction: applications and recommendations.

Kovacs SJ, Martin DE, Everitt DE, Patterson SD, Jorkasky DK.

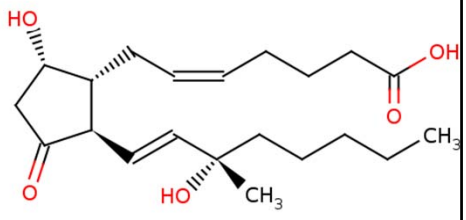
CONCLUSIONS: Urinary excretion of 6 beta-hydroxycortisol may be useful as a screening tool in early-phase development to assess the potential for an investigational drug to induce CYP3A.

Ion down-regulated by OVX



Possible ID of *m/z* 367.2484

15-methyl-15S-PGD2
15-methyl-15R-PGE2
15-methyl-15S-PGE2
15R-PGE2 methyl este
15-methyl-15R-PGD2
Urocortisol

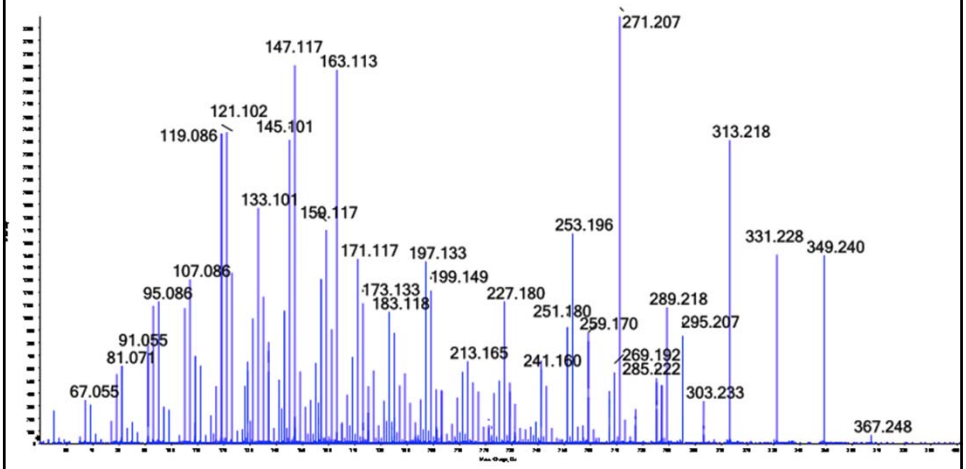


Endocrinology. 1985 Aug;117(2):656-61.

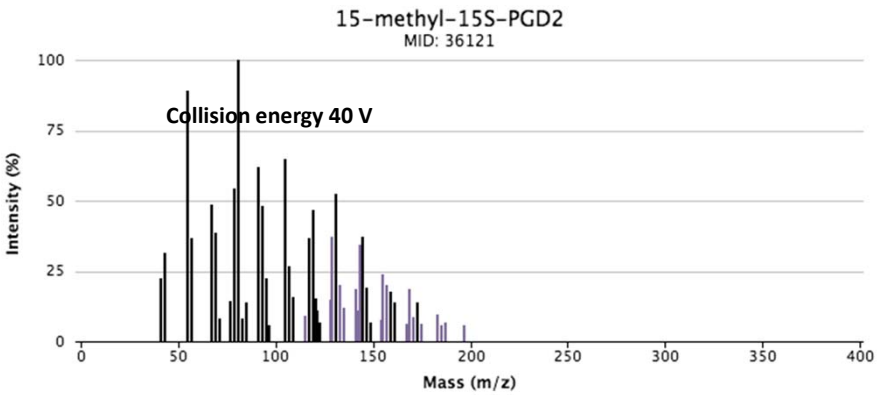
Estradiol stimulates rat renopapillary prostaglandin E2 (PGE2), but not PGF2 alpha biosynthesis.

Katayama S, Lee JB.

MSMS of *m/z* 367.2484



MSMS spectra at METLIN

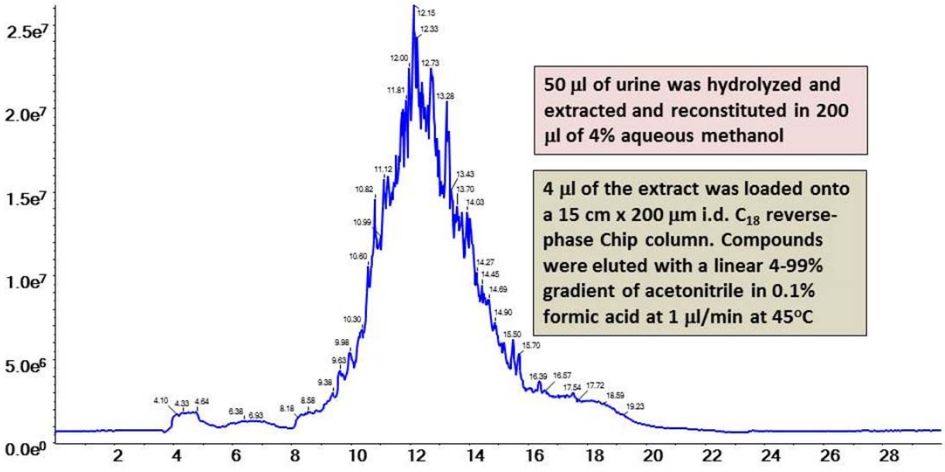


On the 5600 TripleTOF we use a rolling collision energy between 20 and 40 V. The observed ions are within 1 mDa of the library ions.

Best of both worlds analysis

- **Untargeted and targeted analysis performed simultaneously**
  - **As before, collect high mass resolution/high mass accurate MS1 data for 100 msec (untargeted)**
  - **Then collect MSMS data on eighteen pre-selected precursor ions for 50 msec (targeted)**
  - **Repeat data collection in the next second and following second periods**
  - **This technique is called pseudoMRM**

Total ion current of extracted urine

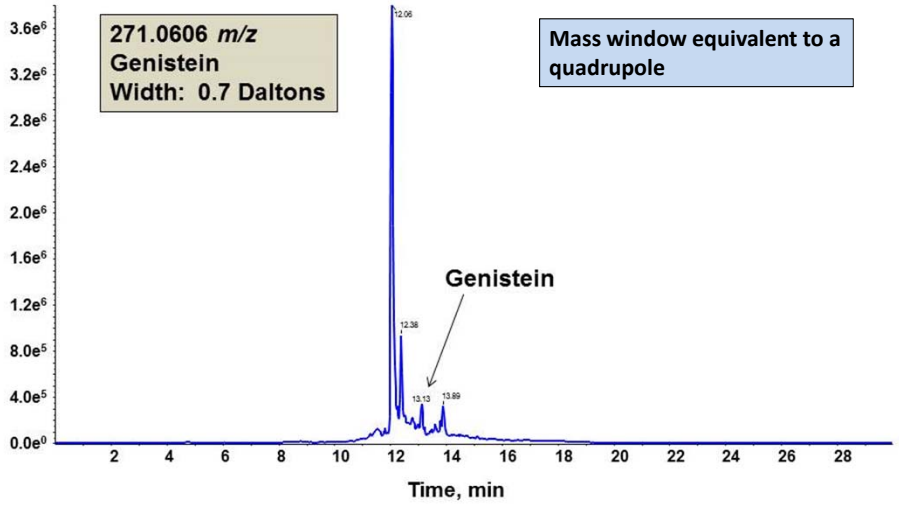


50 µl of urine was hydrolyzed and extracted and reconstituted in 200 µl of 4% aqueous methanol

4 µl of the extract was loaded onto a 15 cm x 200 µm i.d. C_{18} reverse-phase Chip column. Compounds were eluted with a linear 4-99% gradient of acetonitrile in 0.1% formic acid at 1 µl/min at 45°C

The collected data represent a data library that can be searched and experiments conducted AFTER data collection

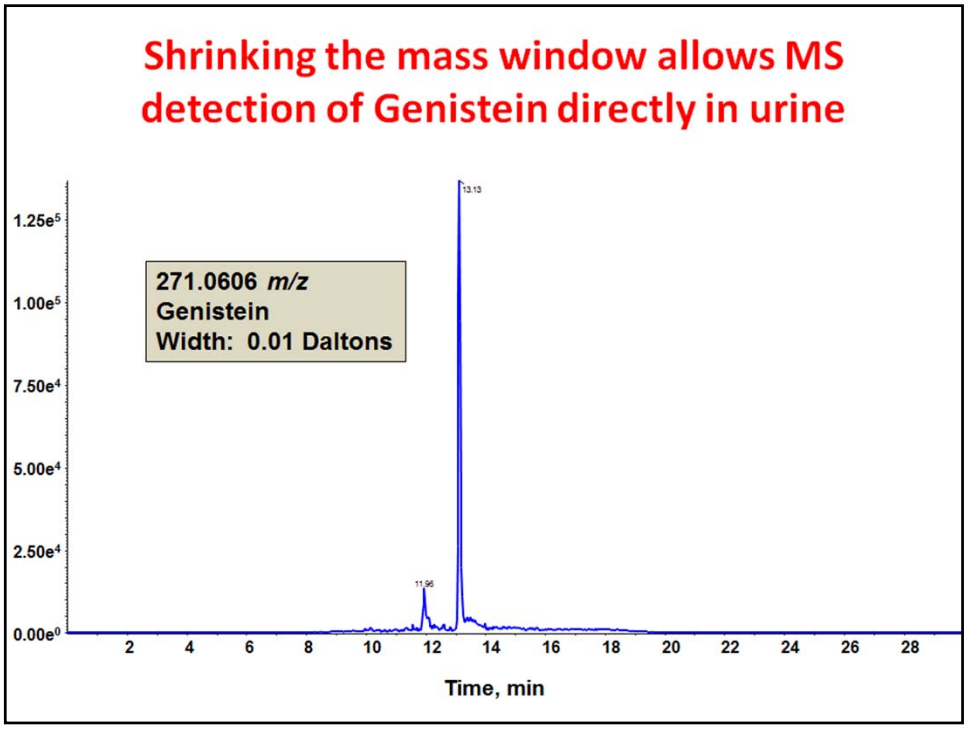
Many compounds are similar to Genistein in urine



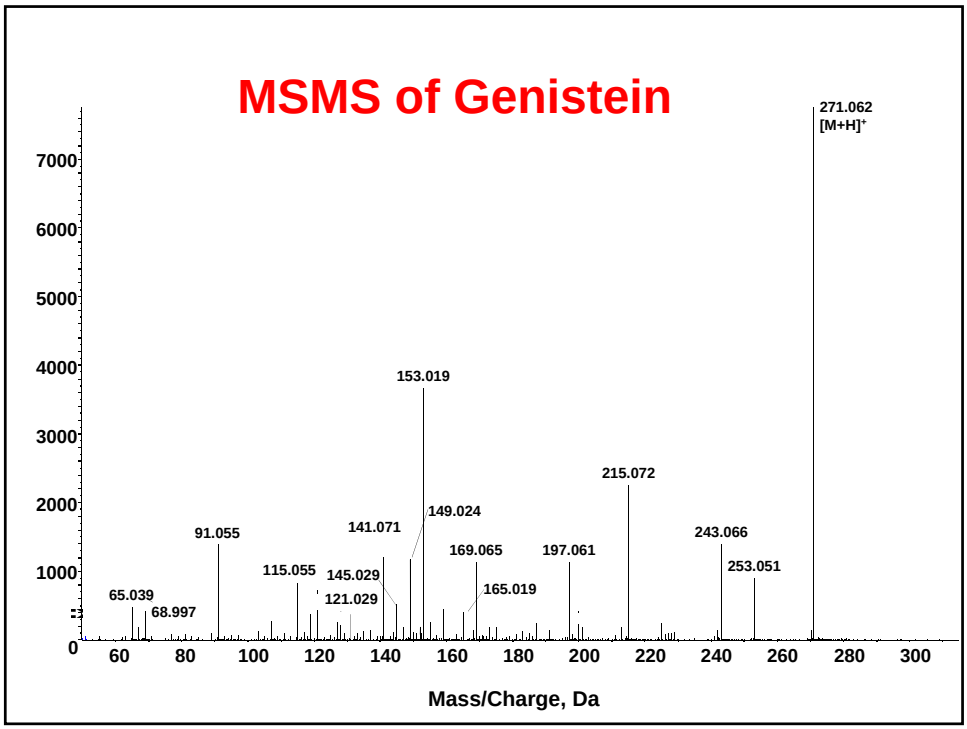
271.0606 *m/z*
Genistein
Width: 0.7 Daltons

Mass window equivalent to a quadrupole

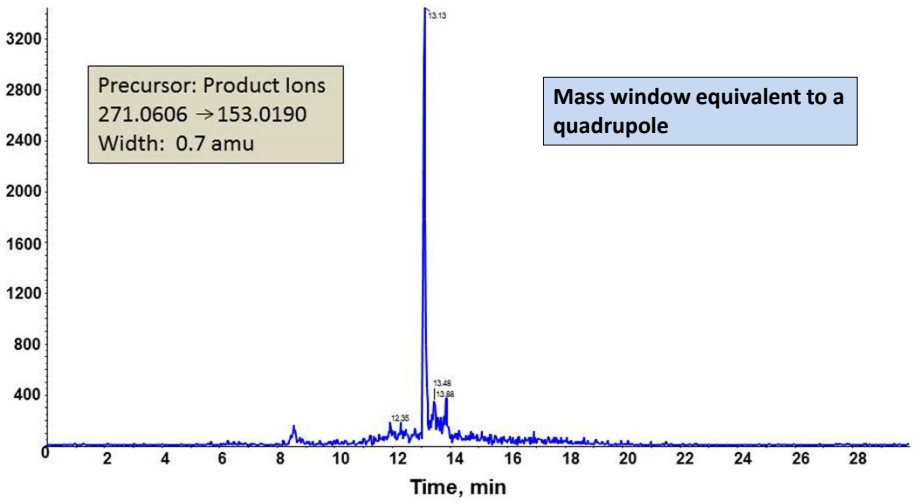
Shrinking the mass window allows MS detection of Genistein directly in urine



MSMS of Genistein



Post data collection MRM for Genistein



MRM for Genistein isn't improved by a narrow mass window