UAB Workshop on the analysis of LC-MS data
December 17th, 2013

# Introduction to LC-MS metabolomics

**Stephen Barnes, PhD**

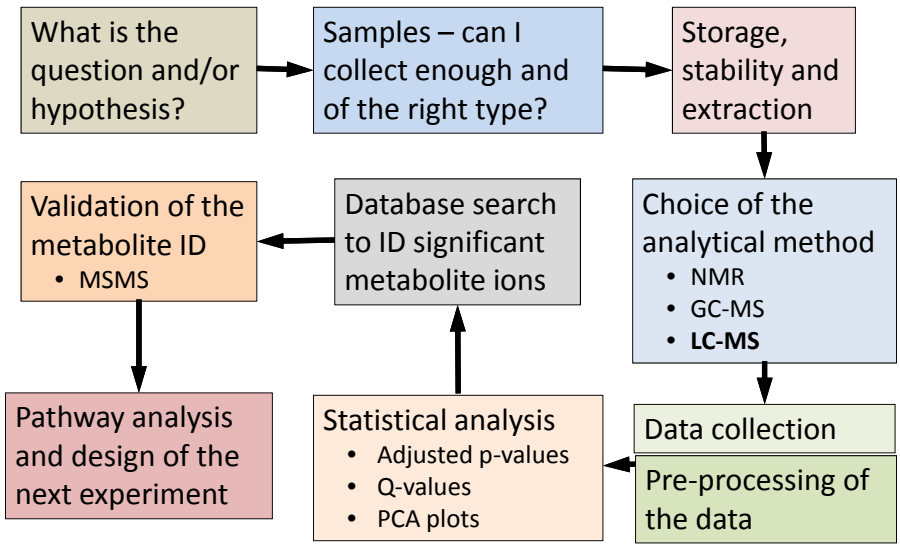
Professor of Pharmacology & Toxicology

sbarnes@uab.edu; 205 934-7117

Targeted Metabolomics & Proteomics Laboratory

CCTS

# Metabolomics workflow



## 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

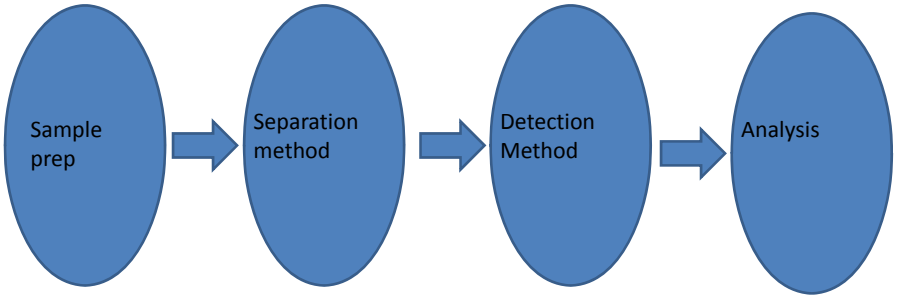
## Measurement Issues

- 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**

# Steps in metabolomics experiment




At each step, the errors can occur
Aim of the experiment is to minimize the errors

## 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.

## Costs for Metabolomics at UAB

- **Validation and bioinformatics**
  - To be discussed later

# 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

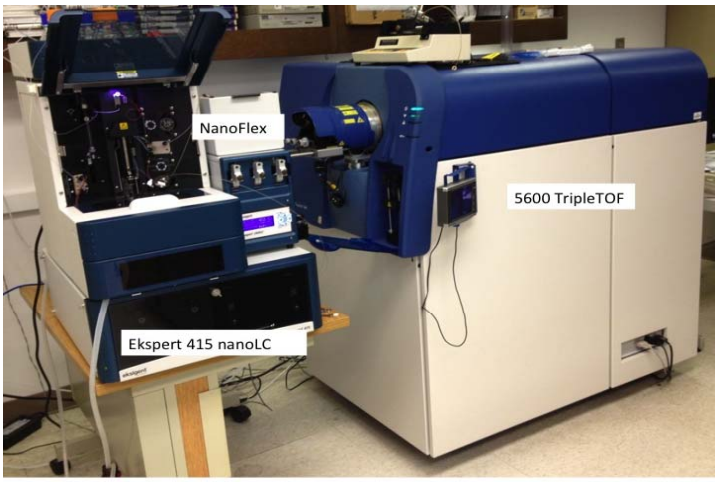
## Other parameters to consider

- **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

# nanoLC-MS instrumentation




# 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**

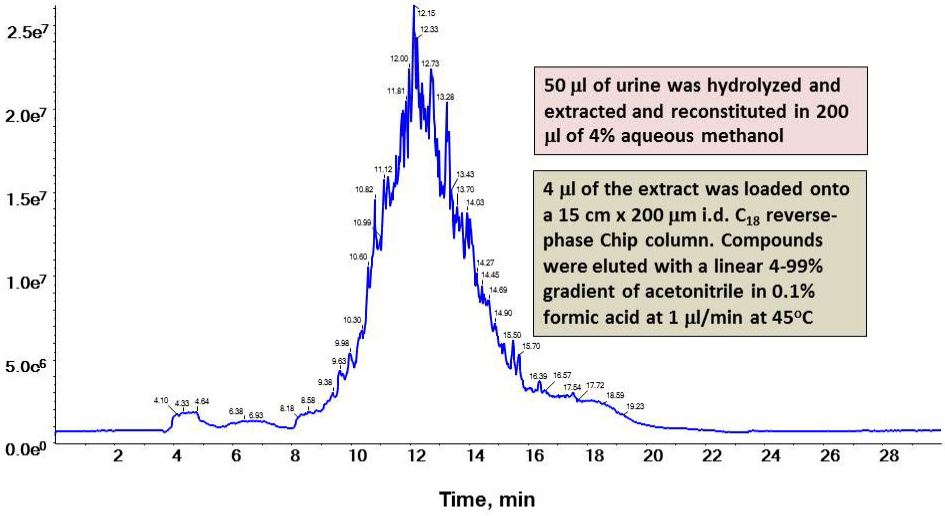
## Untargeted, data-dependent analysis

- 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

## 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

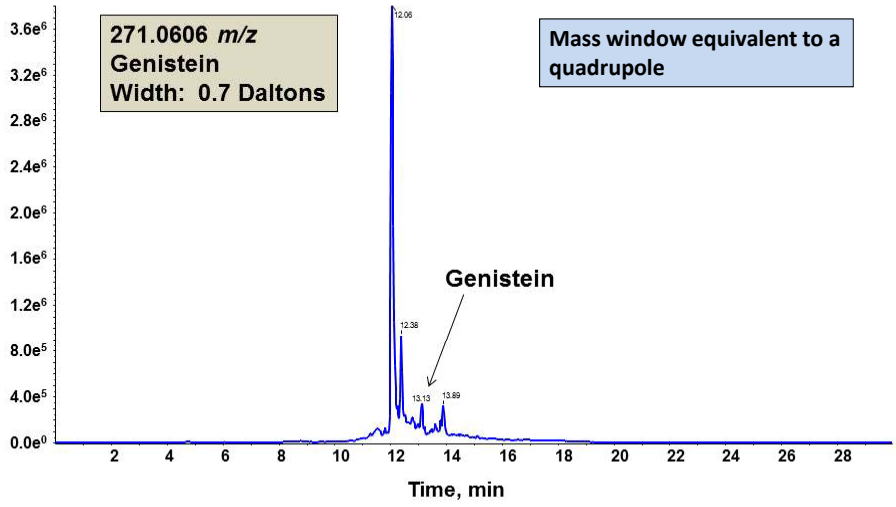



50 µl of urine was hydrolyzed and extracted and reconstituted in 200 µl of 4% aqueous methanol

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

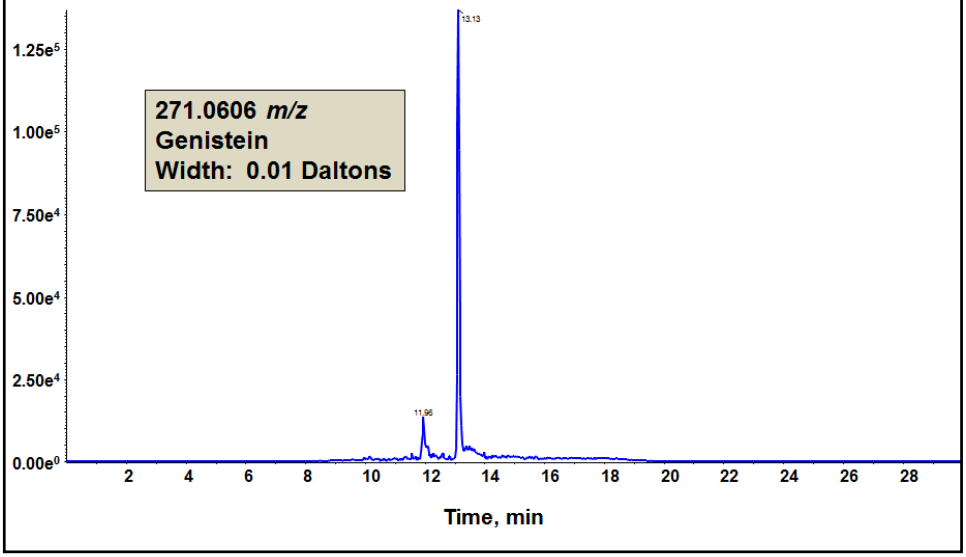
# Many compounds are similar to Genistein in urine



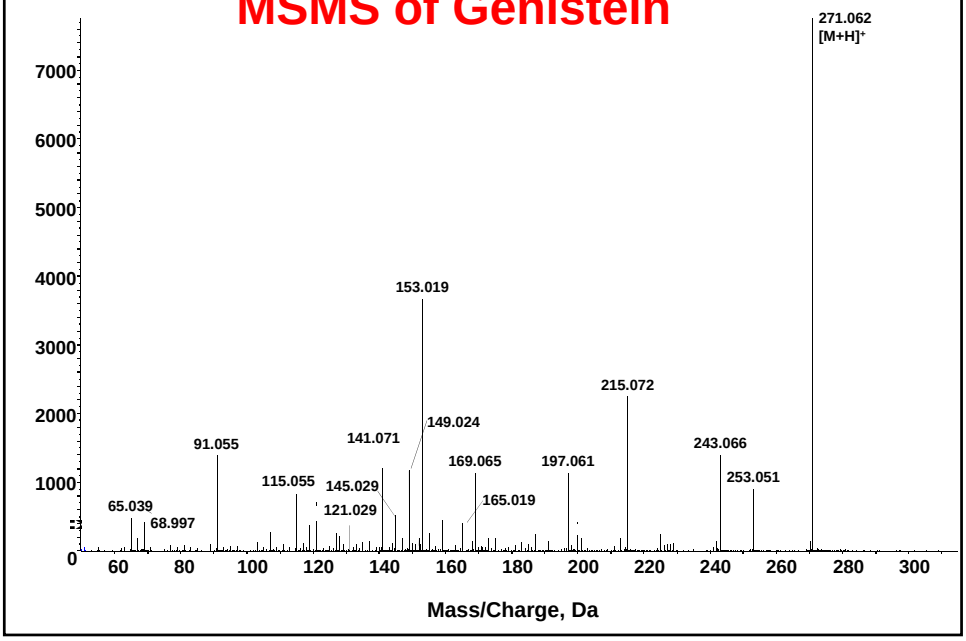

271.0606 *m/z*
Genistein
Width: 0.7 Daltons

Mass window equivalent to a quadrupole

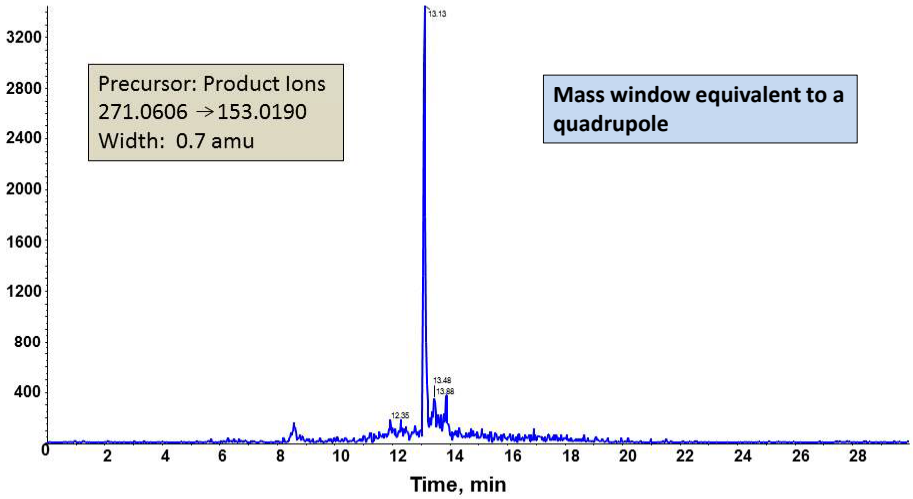
## Shrinking the mass window allows MS detection of Genistein directly in urine




## MSMS of Genistein



# Post data collection MRM for Genistein




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