Knowledge that will change your world

Introduction to metabolomics research

Stephen Barnes, PhD
University of Alabama at Birmingham
934-717; sbarnes@uab.edu

What is "Metabolomics"?

- Metabolomics is like other types of –omics analysis (microarray, RNA-Seq, proteomics, etc.)
  - Offers a "comprehensive" view of all detectable chemicals (not just metabolites)
  - Can be applied to body fluids
    - Plasma/sera, urine, saliva, tears, fecal water, etc.
  - Also to tissues
    - Liver, lung, heart, kidney, brain, eyes, etc.
  - And to single cells
    - Human, rodent, yeast, bacteria, etc.

What is the metabolome?

- Not just the intermediates in the described metabolic pathways (glycolysis, Krebs cycle, etc.) in biochemistry textbooks
- It's all the chemicals that are in tissues and biofluids of us, in experimental animals, in cell lines and even in foods we eat.
- Also, the air we breathe/smell

Where does the metabolome come from?

- It starts with what fixes CO_2 and N_2

Trees convert CO_2 to organic compounds

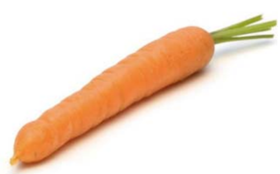
Field of soybeans – they fix N_2 because of nitrogen-fixing bacteria in their root nodules

Plants have more genes than humans

- **Why? Plants can't run away**
- **Instead, they have to practice chemical warfare to prevent attack by aphids and microorganisms**
- **Many plants are poisonous to us**
- **Understanding which plants were safe to eat, or were so if cooked, represented the rise of agriculture and civilization**

Compounds in plants and fruits

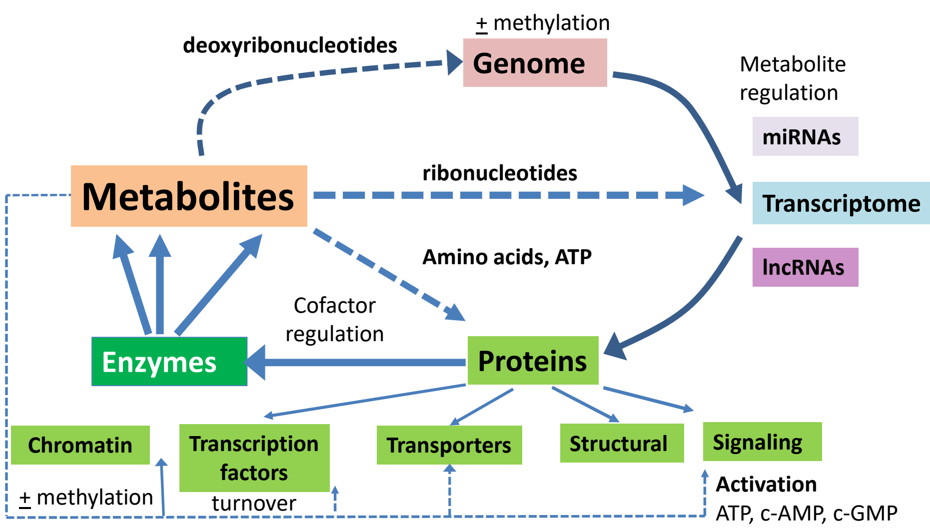
- Carotenoids
- Many vitamins
- Polyphenols and anthocyanins
- Not made by human cells



Other sources of body chemicals

- The microbiomes
  - Humans are not single organisms
  - Instead, we are super-organisms
  - The gut microbiome has 10 times the number of cells found in the rest of the (human) body
  - It makes novel compounds that are absorbed, enter the blood stream and tissues
- Chemicals from the environment
- Interactions between the xenobiotics and the human enzyme systems

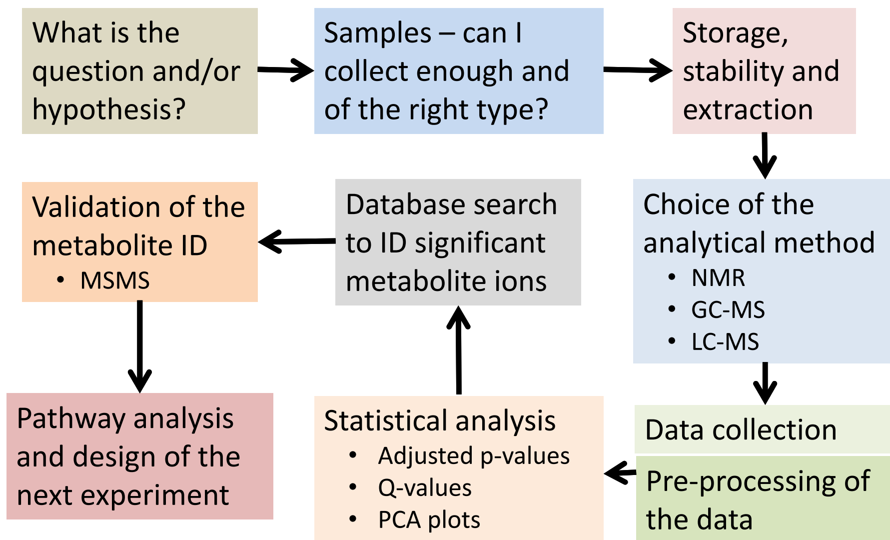
Metabolites are associated with every aspect of cellular events



World without gas!

The metabolome is very complex!

Metabolomics workflow



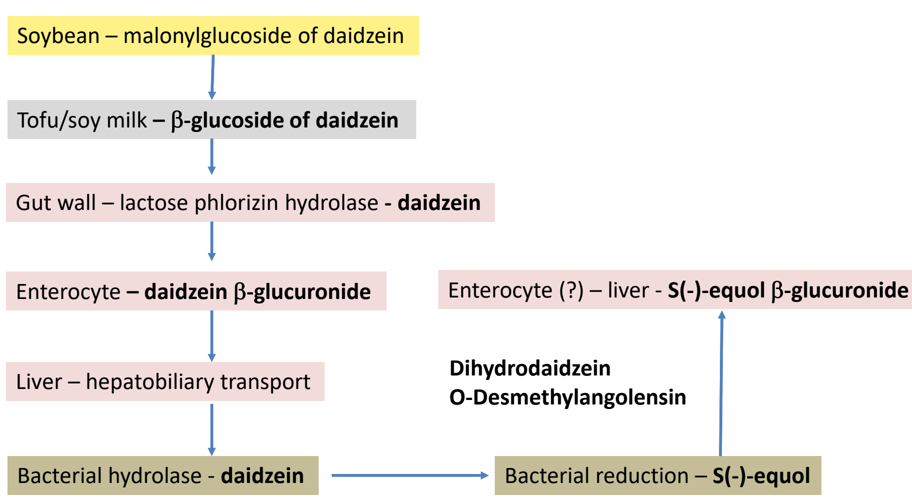
Course goals

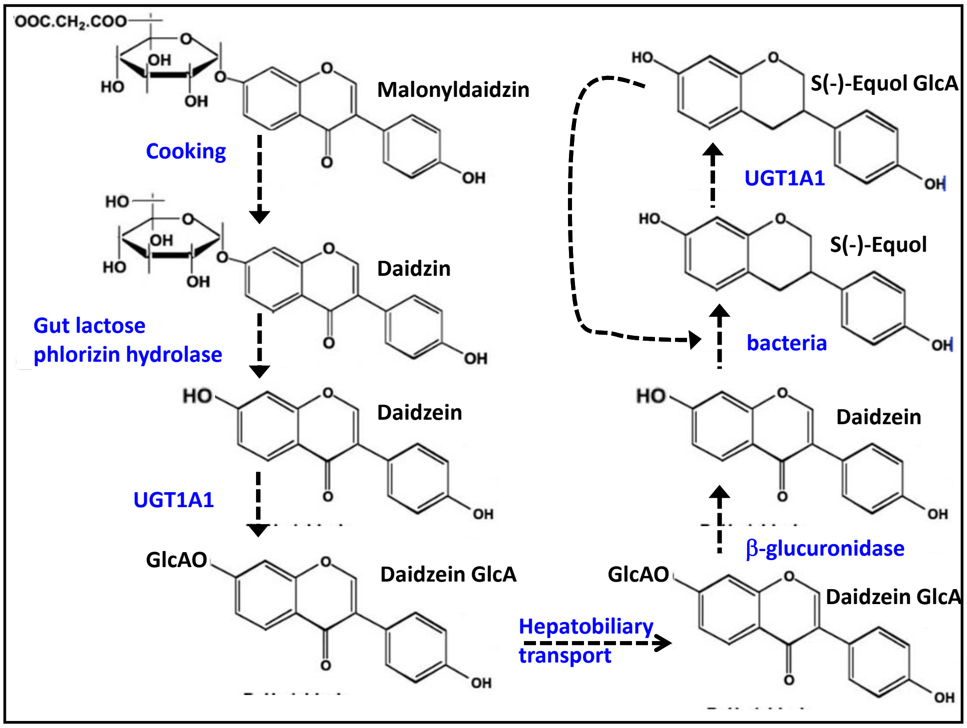
1. **To understand the vital roles of metabolites**
   - **To provide energy for the chemical and enzymatic processes of life**
   - **To provide the building blocks for the macromolecules (DNA, RNA, proteins, carbohydrates, lipids)**
   - **As co-factors**
   - **As signaling molecules**
   - **As biomarkers for disease**

Course goals

2. To understand the origins of metabolites
   - Produced by (human) cells
   - Produced by the things that we eat (the food-ome)
     - Plants (wheat, corn)
     - Fruits (apples, oranges, strawberries)
     - Vegetables (rice, potatoes, broccoli, peas)
     - Dairy products, including fermented forms
     - Meat from other animals
     - Xenobiotics
   - Produced by microorganisms in our bodies

Complex metabolism of daidzein





Course goals

3. To understand that a metabolomics experiment is high dimensional
   - i.e., it is compares the intensities of hundreds, if not thousands, of distinct species
   - Very important statistical consequences
   - Cannot afford to do a robust experiment that fully satisfies theoretical statistical principles
   - Very important to sit down with a statistician <u>prior</u> to executing an experiment

Course goals

4. To select the appropriate method for extracting/recovering metabolites
   - Metabolites encompass an enormous range of chemistries
     - Gaseous (H_2, H_2S)
     - Volatile (butyric acid, acetone, skatole)
     - Hydrophilic (glucose)
     - Charged-positive/negative (amino acids, nucleotides, organic acids, amines)
     - Hydrophobic (lipids, steroids, hydrocarbons)
   - No single method suitable for all metabolites

Course goals

5. Selecting the analytical approach
   - *In situ* analysis
     - Laser ablation of frozen tissue
     - Other desorption methods
     - Magic angle spinning NMR
     - Other spectroscopic methods
   - Extracted samples
     - NMR
     - GC-MS (1- and 2D chromatography and MSMS)
     - LC-MS (1- and 2D chromatography and MSMS)
     - CE-MS
   - Targeted vs untargeted analysis

Course goals

6. Analysis of the data
   - Data visualization
     - Mzmine 2
     - XMCSonline
   - Data alignment
     - NMR methods
     - LC-MS and GC-MS methods (XCMS/Mzmine 2)
     - MS-DIAL
   - Statistical evaluation
     - XCMSonline
     - Univariate and multivariate analysis (MetaboAnalyst)
     - Mummichog

Course goals

7. Identifying the “interesting” metabolites
   - Use of Mass Spectrometry (absolute mass)
     - METLIN
     - MetaboSearch
     - Mummichog
     - ChemSpider
   - MSMS (fragmentation spectra)
     - METLIN
   - Metabolite standards
   - Importance of retention time
     - Multiple column conditions

Course goals

8. Pathways and applications
   - Mummichog
   - KEGG pathway mapping
   - Metaboanalyst
   - Applications to:
     - Adverse cardiovascular risk
     - Diabetes
     - Lens and kidney diseases
     - Cancer

Use of stable isotopes

9. Novel pathways and pathways atom by atom
   - The pathways in the text books were built when
     1. First GC-MS, then high field NMR and most recently LC-MS were not available to take advantage of stable isotope labeling of unstable intermediates
     2. We didn't appreciate the role of the microbiomes in the pathway "stories" being built
   - Now specifically labeled ^{13}C, ^{15}N and $^{17}O/^{18}O$ precursors and intermediates are available to trace pathways and their dynamics

A brief history of metabolomics

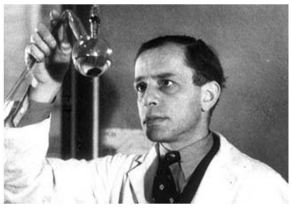
Nuclear physics moves to biology

- 1897 JJ Thomson discovers the electron (cathode rays)

- 1919 Aston using a mass spectrograph shows that Neon with a non-integer MW (20.2 Da) is composed of two isotopes, ^{20}Ne and ^{22}Ne

http://www.asms.org/Publications/Historical/HistoryofMassSpectrometry/tabid/94/Default.aspx

Transition to biology



Ralf Schoenheimer

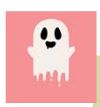
David Rittenberg

- **While the politicians, tyrants, dictators and despots were salivating at the thought of developing nuclear weapons from unstable isotopes in the early part of the 20th Century, two scientists began the pursuit of the peaceful use of stable isotopes, initially deuterium (^{2}H), and later carbon (^{13}C) and nitrogen (^{15}N), to study biochemical pathways**
- **Understanding the pathways of metabolism was born**

Metabolomics and NIH Research 1948-2015

Microbiome

Metabolomics



1950s-60s emphasis on determining metabolic pathways – 20+ Nobel prizes

Bloch Lynen Krebs

1950s-early 1980s Identification and purification of proteins

1980-1988 Sequencing of genes – cDNA libraries – *orthogonal research*

1988-2000 Sequencing of the human genome – *period of non-orthogonal research* – where did all the genes go? junk DNA?

2004 Tiling arrays reveal that most of the genome is expressed

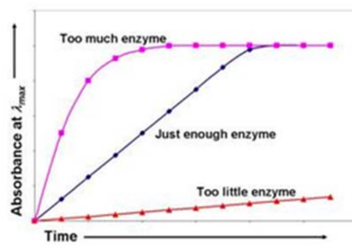
2006 First ENCODE project on 1% of the human genome reveals RNAs coming from more than one gene

2012 Human genome ENCODE project reveals the extent of DNA expression and roles for "junk" DNA, as well as intergenic proteins

2014 –"deep" proteomics reveals the presence of 400+ proteins that **are not encoded by the genome**

Metabolism to metabolomics

- Measured with enzymes – changes in NAD(P)H absorbance/fluorescence
  - Studies of glycolytic and the TCA cycle intermediates one at a time



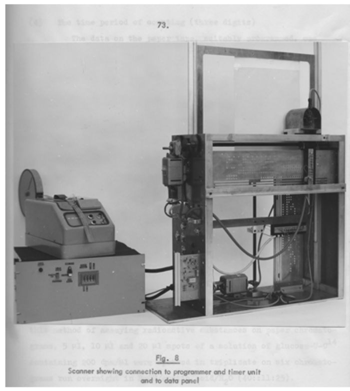
Gas chromatography and GC-MS

- Organic acids, fatty acids and amino acids by GC
  - Short chain fatty acids
  - Volatile derivatives, Flame Ionization Detection
- GC-MS started in mid-70s
- 1970s also saw introduction of capillary GC which gave far higher chromatographic resolution than the packed ¼” ID columns (1975/6)

AT (Tony) James

Origins of practical metabolomics

Imperial College 1967-1970

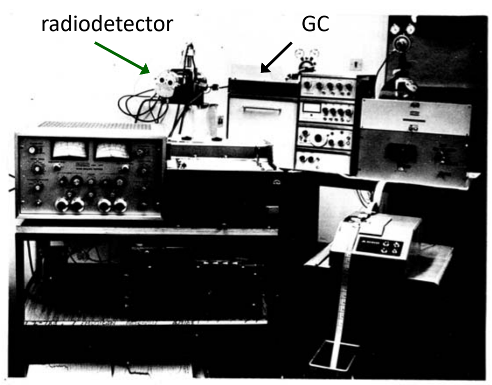


Radio 2D-paper chromatography scanner with digitization of collected data

The room had 10 of these scanners – data analyzed by a central computer (in 1968)

Courtesy of K.R. Mansford, PhD

Radio-GC analysis – metabolomics in its infancy

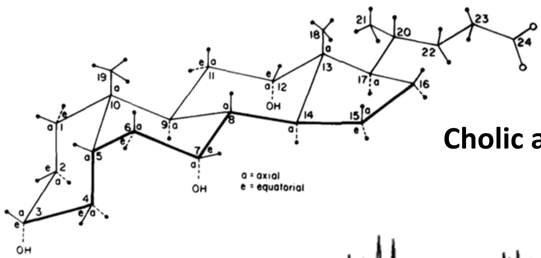


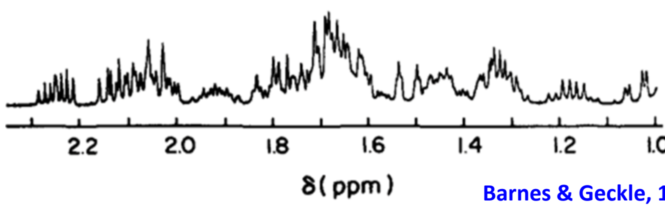
Radio gas-liquid chromatography with digitization of collected data

Developed this for my PhD work (1967-1970) to study glucose metabolism in acellular slime moulds

How Nuclear Magnetic Resonance (NMR) became a player

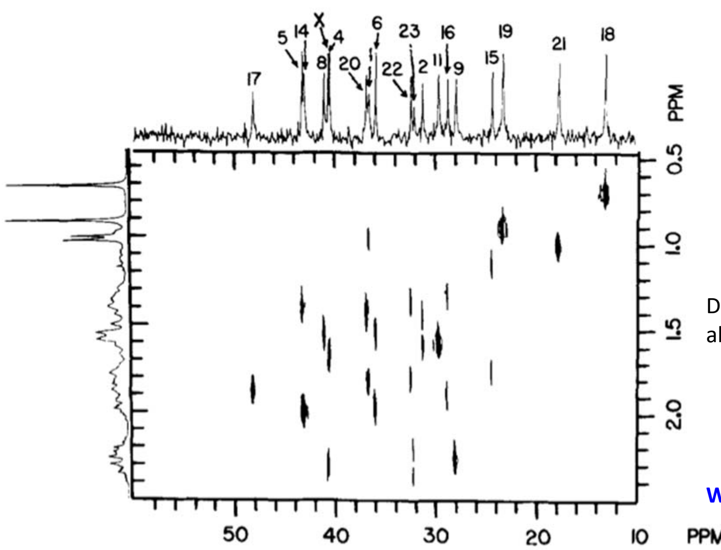
- Mid 60s – introduction of Fourier transform analysis
- Late 70s – introduction of superconducting magnets
- Early 80s - pulse sequences





Barnes & Geckle, 1982

Pulse sequences in NMR (HetCor)



Depends on natural abundance of ^{13}C

Waterhous et al. (1985)

Gas chromatography

- Built on critical steps
  - 1908 Twsett introduces the concept of chromatographic separation (of plant pigments)
  - 1941 Martin and Consden conceptualize the rules of partition chromatography (get the Nobel Prize)
  - 1950 James and Martin describe gas chromatography of volatile fatty acids
    - A boon to the oil industry
  - 1975 (Finally) open tubular, capillary gas chromatography becomes commercially available

Progress in LC-MS

- Commercial HPLC appeared in the early 1970s to separate thermally stable and unstable molecules
- The challenge remained to find a way to get the unstable compounds into the gas phase
  - Applied to macromolecules (peptides, proteins) as well as metabolites
- Thermospray had some initial success
- Electrospray ionization and chemical ionization radically changed analysis, allowing compounds to go into the gas phase at atmospheric pressure and room temperature

LC-MS

- Suddenly, there were what appeared to be no limits (or very few) to what could be analyzed
- Unheard of, robust mass spectrometers came into play
  - "A reliable mass spectrometer" was considered in 1990 to be an oxymoron

Types of LC-MS analysis

Single quadrupole LC-MS analysis

LC-time-of-flight (TOF)-MS

FT-ICR MS

Orbi-trap

Triple quadrupole LC-MS analysis

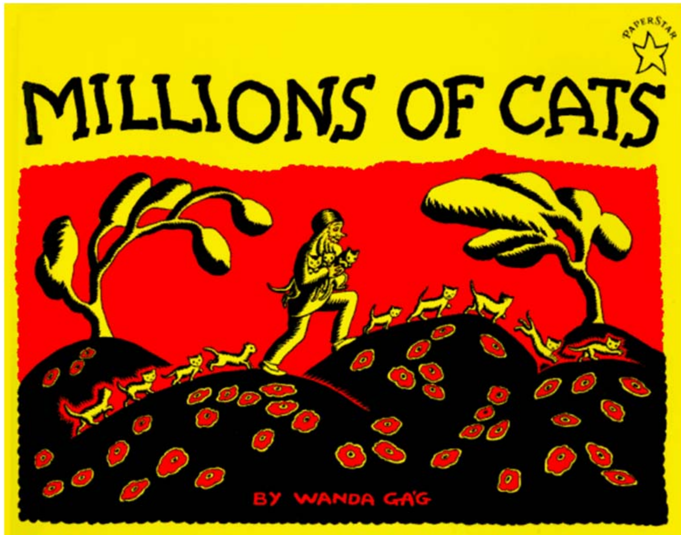
Multiple reaction monitoring (MRM)

Q-TOF

TripleTOF

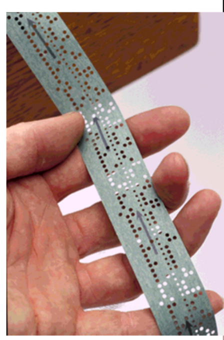
Ion Mobility

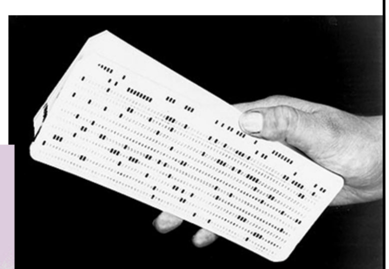
Data explosion



Changing times in Computing

- **1950 The Cambridge colleagues of Watson and Crick calculated the structure of DNA by putting data onto punched cards and taking them by train to London for analysis – and to the fog – the "cloud" in 1950s**
- **1964 Seymour Cray develops the CDC 6600 (1 Mflops)**
- **1967 I used paper tape to collect data from a radio gas chromatograph and then submitted them via a terminal reader to the CDC 6600 at the University of London**





Today in Computing

On my desk in 2016

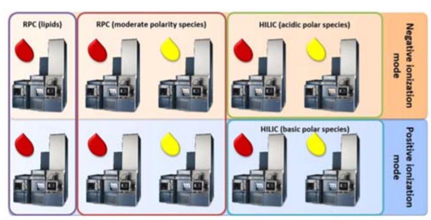
- The Apple MacBook Air with 2 quad core Intel i7 processors
  - Operates at 2.0 GHz
  - Memory of 8 GB
    - Access 1.333 GHz
  - 512 GB Flash memory storage
  - 10 Gbs Thunderbolt I/O
- Also cost ~$2,000

IBM Blue-Gene

- Parallel processing with 2,048 700 MHz computers operating at 4.733 Tflops
- Replaced by Cheaha, in its current configuration it has 48 compute nodes with two 2.66GHz 6-core Intel CPUs per node (576 cores total)
- It operates at 6.125 Tflops

MRC-NIHR National Phenome Centre

600 MHz NMR instruments in surgical suite

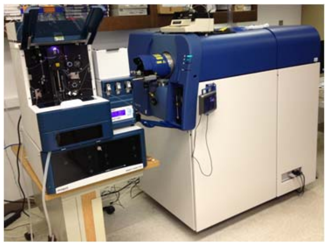


Mass spectrometers (10 Q-TOFs) each dedicated to one assay format

Iknife - revolutionizing surgery

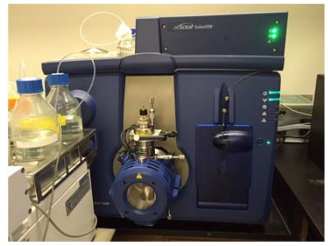
This is Next-GEN precise medicine

UAB capabilities in metabolomics

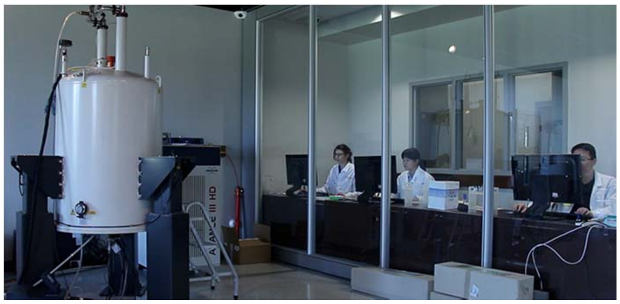


SCIEX 5600 TripleTOF
with Eksigent nanoLC

TMPL mass spec lab
MCLM 459/427
Stephen Barnes, Director
934-7117/3462



SCIEX 6500 Qtrap with SelexION



Central Alabama NMR facility
Chemistry Bdg
N. Rama Krishna, Director
934-5695

Great challenges in metabolomics

- The extent of the metabolome
  - From gaseous hydrogen to earwax
- Having complete databases
  - METLIN has 60,000+ metabolite records, but your problem always creates a need to have more
  - Improvement in the size of a MSMS database
- Storing and processing TBs of data
- Standards and standard operating procedures
- Being able to do the analyses in real time

NIH Common Fund Metabolomics Program

- **Metabolomics Workbench**: http://www.metabolomicsworkbench.org/
- **Regional Comprehensive Metabolomics Research Centers**
  - University of Michigan: http://mrc2.umich.edu/index.php
  - UC Davis Metabolomics Center: http://metabolomics.ucdavis.edu/
  - Eastern Regional Metabolomics Research Center: http://www.UNCNRI.org/metabolomics
  - SE Center for Integrated Metabolomics: http://secim.ufl.edu/
  - Resource Center for Stable Isotope Metabolomics: http://bioinformatics.cesb.uky.edu/bin/view/RCSIRM/
  - Mayo Clinic Metabolomics Resource: http://www.mayo.edu/research/core-resources/metabolomics-resource-core/overview