THE UNIVERSITY OF ALABAMA AT BIRMINGHAM
Knowledge that will change your world

Introduction to metabolomics research

Stephen Barnes, PhD, FASN
University of Alabama at Birmingham
sbarnes@uab.edu
205-934-3462; BBRB 711

Targeted Metabolomics & Proteomics Laboratory

1

Course goals

1. To understand the vital roles of small molecules/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

2

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
   - Synthetic: Therapeutics, smoking, household chemicals

3

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 prior to executing an experiment

Dr. Hemant Tiwari

4

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

Dr. Prasain

5

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 such as Raman CAR
   – 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

Drs. Placzek and Barnes

6

Course goals

6. Analysis of the data
   - Data alignment
     - NMR methods
     - LC-MS and GC-MS methods (XCMS; ADAP; MS-DIAL)
   - Statistical evaluation
     - Univariate and multivariate analysis (MetaboAnalyst)
     - XCMSonline
     - Peaks to Pathways (Metaboanalyst)
   - Data visualization
     - XMCSonline
     - MZmine

Dr. Barnes

7

Course goals

7. Identifying metabolites
   - Use of MS (absolute mass)
     - MS-DIAL
     - METLIN
     - Peaks to Pathways
     - ChemSpider
   - MSMS (fragmentation spectra)
     - METLIN
     - MS-DIAL
   - Metabolite standards (IROA kit)
   - Importance of retention time
     - Multiple column conditions

Dr. Barnes and Prasain

8

Course goals

8. Pathways and applications
   - Peaks to Pathways/Metaboanalyst
   - KEGG pathway mapping
   - Applications to:
     - Adverse cardiovascular risk
     - Diabetes
     - Lens and kidney diseases including COVID-19
     - Cancer
   - Integration with other –Omics (Dr. Hu)
   - Machine learning/Artificial Intelligence (Dr. Marquez-Lago)

9

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.

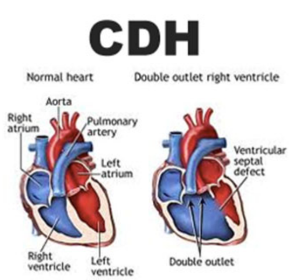
10

Defining who we are chemically

- Does an understanding of the functions of human genes define the chemical make up of our body fluids and tissues?
- How does metabolomics provide information on the circulating chemicals?
- Are the detected chemicals metabolites produced by human enzymes?
- So, what are we really exposed to? And does it make a difference?

11

A great deal of emphasis has been placed on the importance of DNA sequencing

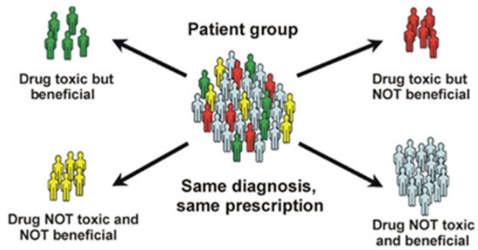


This model works for congenital diseases

Biliary atresia

https://loveyabeckett.files.wordpress.com

This has evolved into precision medicine and optimization of therapy



http://personalizedmedicineproject.weebly.com/

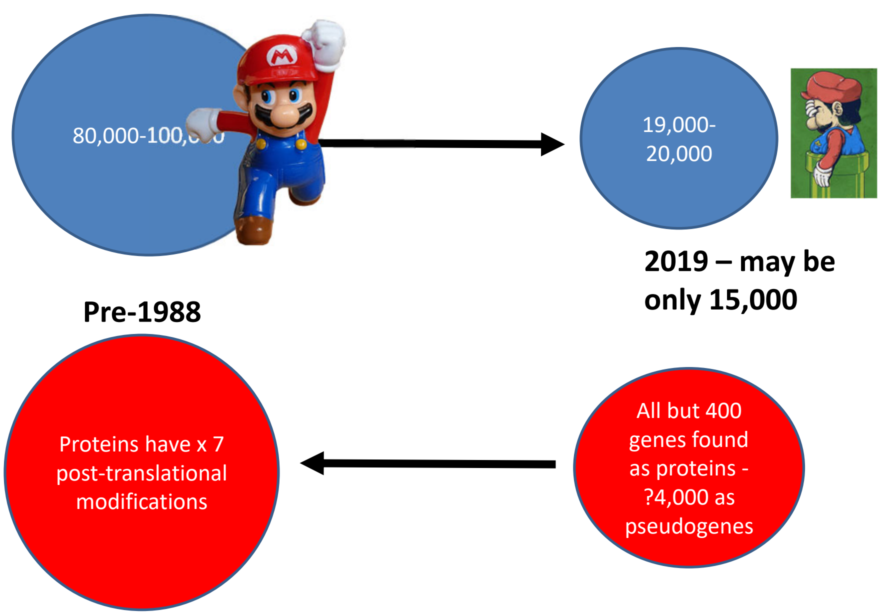
12

Metabolomics in the newborn

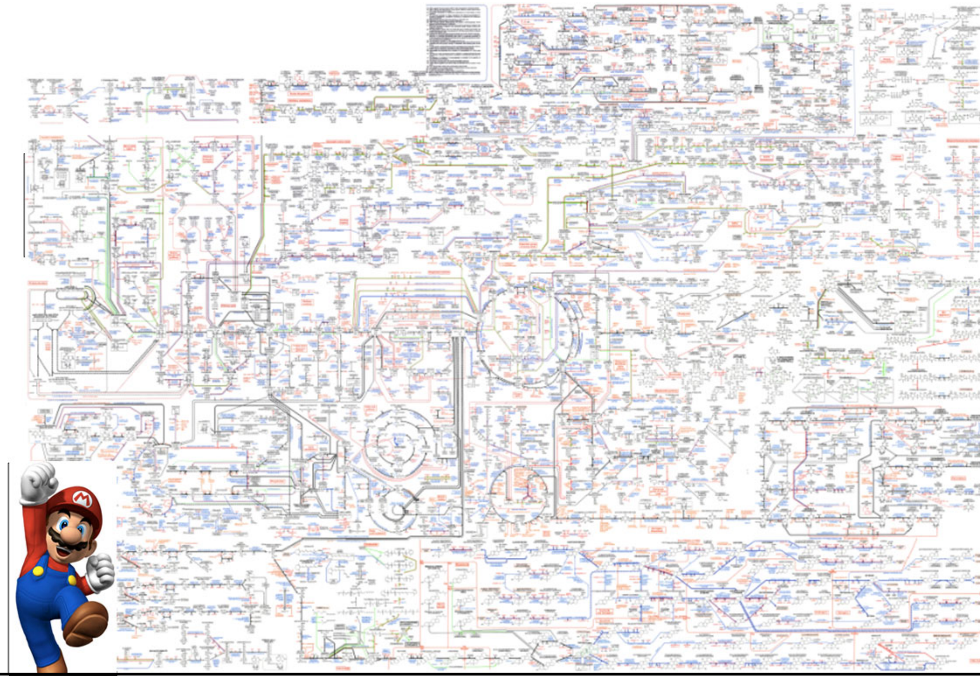
Dr. Dan Sharer

13

However, genes failed to meet full expectations



14



15

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

16

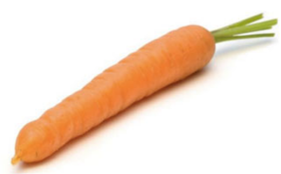
Plants have more genes than humans

- **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**

17

Compounds in plants and fruits

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



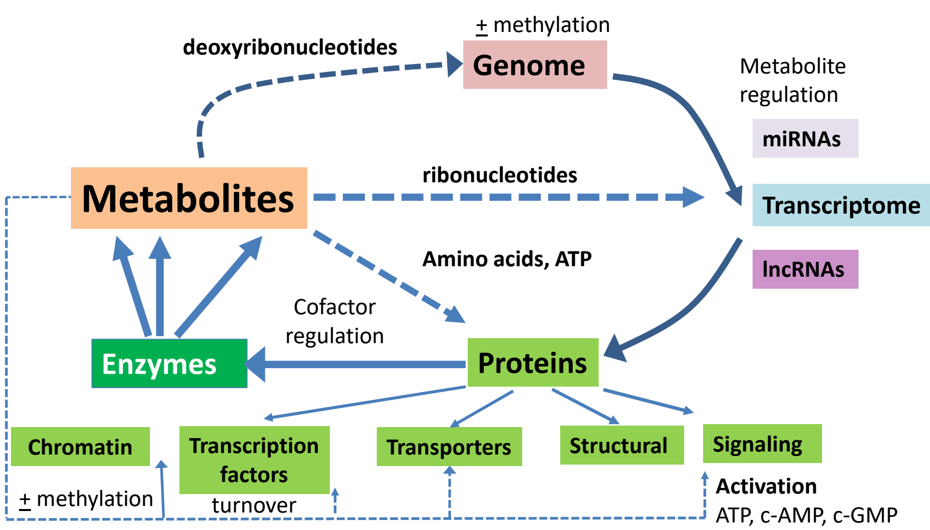
18

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
  - industrial contaminants, therapeutics, supplements
- Interactions between the xenobiotics and the human enzyme systems

19

Metabolites are associated with every aspect of cellular events



20

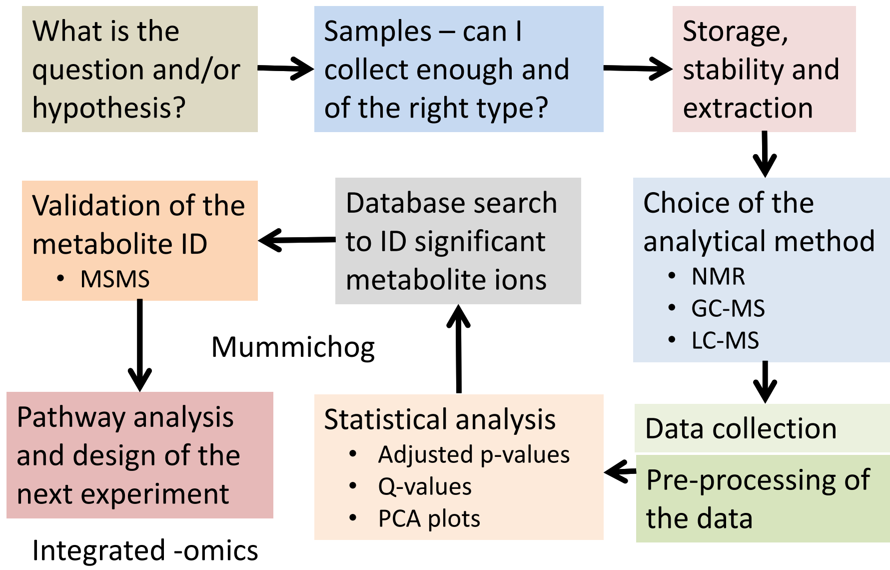
World without gas!

21

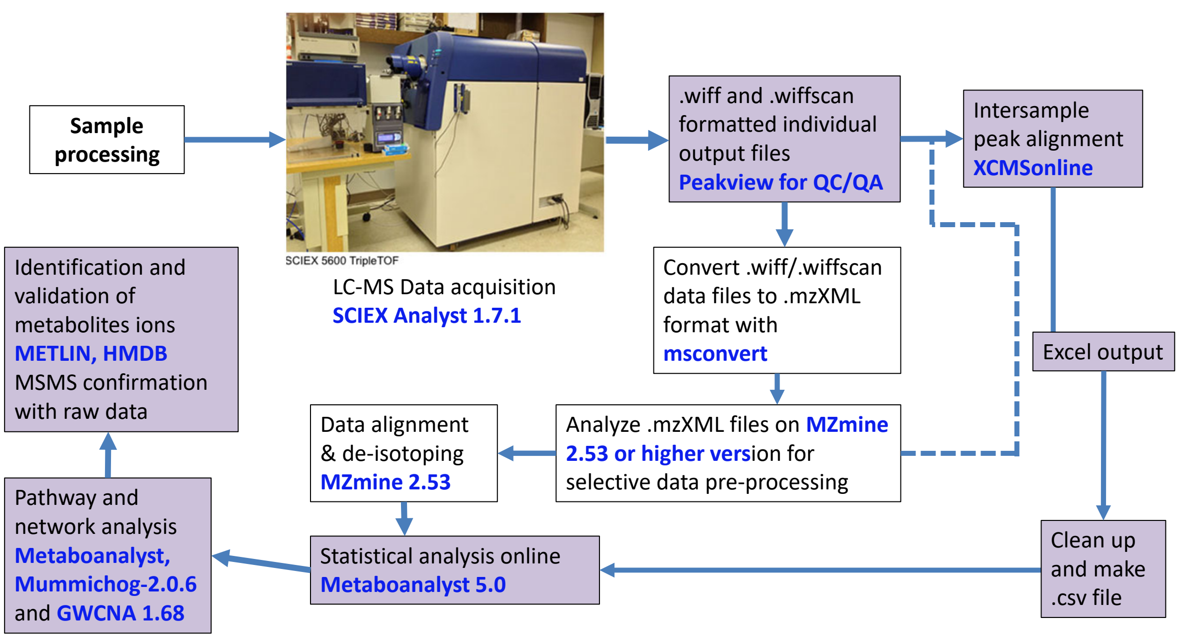
The metabolome is very complex!

22

Metabolomics workflow



23



24

Sample processing →

SCIEX 5600 TripleTOF

LC-MS SWATH data acquisition#
SCIEX Analyst 1.7.1

→ .wiff and .wiffscan formatted individual output files → **Upload to MS-DIAL**

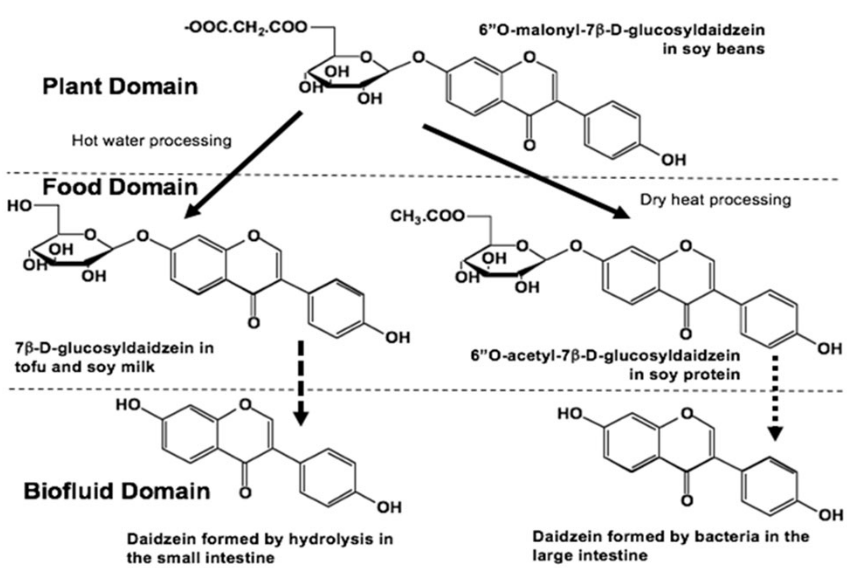
→ **Analysis for files by MS-DIAL 4.70**

→ **Identification using MSMS of drugs, their metabolites, physiological intermediates, food intermediates, food additives and domestic and industrial chemicals**

→ **Alignment allows comparison of subjects to see which drug is in which subject and how much**

25

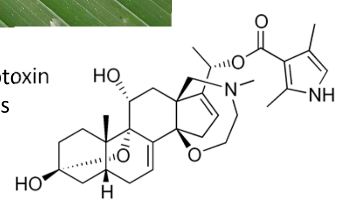
Complexity in metabolism beyond cell culture
Inter-genome events



26

The Amazonian poison dart frog

Their skin contains molecules like batrachotoxin which irreversibly poisons the Na^+-channels



27

Two questions

Why isn't the batrachotoxin a poison to the frog?

ANSWER: The frog has mutations of three residues in the Na^+-channel protein that prevent binding of the batrachotoxin

Does the frog synthesize the toxin?

ANSWER: It doesn't, it gets the toxin from what it eats – ants, beetles, etc.

So, it all depends on what you eat.

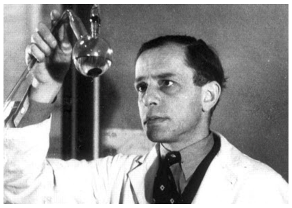
Dart frogs bred in captivity and fed a non-insect diet don't make batrachotoxin

28

Where did metabolomics came from?

29

Transition of mass spectrometry to biology



Ralf Schoenheimer

David Rittenberg

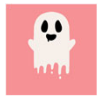
- **While the politicians, tyrants, dictators and despots were salivating at the thought of developing nuclear weapons from unstable isotopes in the early part of the 20th Century, two scientists began the pursuit of the peaceful use of stable isotopes, initially deuterium (^{2}H), and later carbon (^{13}C)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

30

Metabolomics and NIH Research 1948-2016

Microbiome

Metabolomics



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

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

31

Structure of NIH Research

Individual Research institutes (27)
NIDDK, NHLBI, NIAID, etc.

National Centers
CIT, CSR, FIC, NCCIH, NCRR, NIHCC

Office of the Director

2005 Review by Congress

Division of Program Coordination, Planning, and Strategic Initiatives, DPCPSI

NIH Common Fund

NIH Council of Councils

32

Structure of NIH Research

Individual Research institutes (27)
NIDDK, NHLBI, NIAID, etc.

National Centers
CIT, CSR, FIC, NCCIH, NCATS, NIHCC

Office of the Director

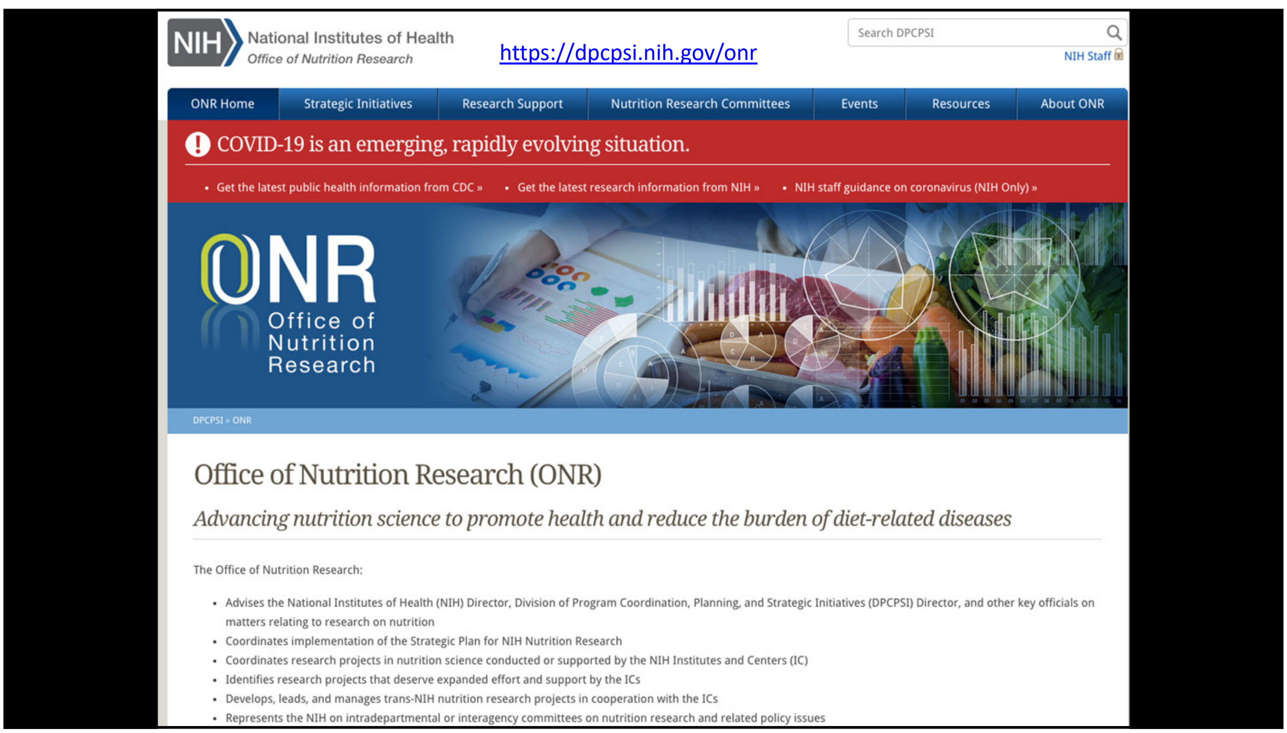
2005 Review by Congress

Division of Program Coordination, Planning, and Strategic Initiatives, DPCPSI

NIH Common Fund

NIH Council of Councils

33



34

Technologies in Metabolomics

- **Gas-chromatography-mass spectrometry (GC-MS)**
- **Liquid chromatography-mass spectrometry (LC-MS)**
- **Capillary electrophoresis-mass spectrometry (CE-MS)**
- **Nuclear magnetic resonance (NMR)**

35

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**

36

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

37

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)

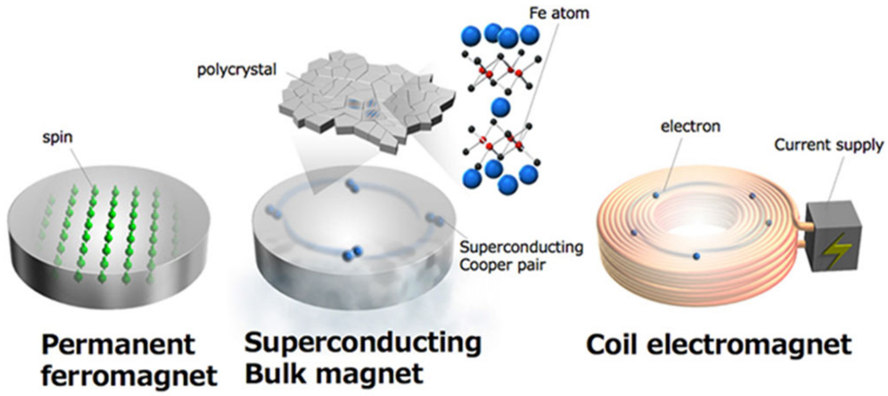
Q-TOF

TripleTOF

Ion Mobility

38

NMR spectroscopy and metabolomics



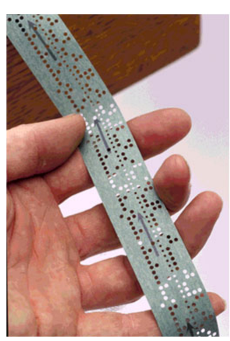
https://nationalmaglab.org/images/news_events/news/2015/october/pnictide_magnetism_1oct2015.jpg

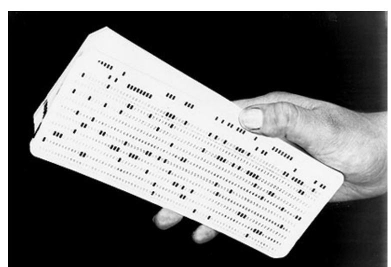
NMR has had several critical development steps – Fourier Transform analysis of collected data, increase in field strength with superconducting magnets, micro-coil, cryogenic analysis, and hyperpolarization.

39

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**





40

My first computers

At Imperial College (1968)
Digital PDP-9 computer with 8 k of memory
Data points were collected at 1 Hz on paper tape
Fed into the computer with an optical reader
The program to digitize collected GC data was 6.5 k
The model I had had a screen to display the data

At UAB (1979)
Apple II+ computer with 48 k of memory
Programs and collected data were uploaded from magnetic tape
(disk drives came a little later)
With a monitor, the cost (1979 dollars) was $1,999
1980 – "hacked" into the UAB IBM 370 mainframe which had a memory of 12 MB

41

Today in Computing

On my desk in 2022

- **The Apple MacBook Pro with 10 core M1 pro processors and 16-core GPUs**
  - **Operates at >3.2 GHz**
  - **Random access memory 16 GB**
    - **Access 200 GB/s**
  - **1 TB Flash memory storage**
  - **40 GB/s Thunderbolt I/O**
- **Also cost ~$2,000 (2021 dollars)**

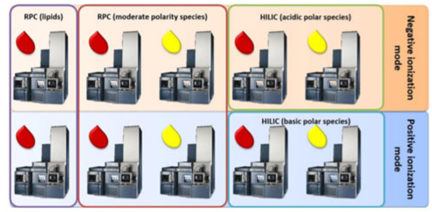
Cheaha high-performance computing

- UAB's first supercomputer (IBM Blue-Gene) operated at 4.733 Tflop/s
- **Replaced by Cheaha - in its current configuration it has 3500+ conventional CPU cores and 6.6 PB raw data storage**
- **It operates at 528 Tflop/s (max)**

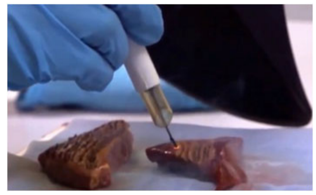
42

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

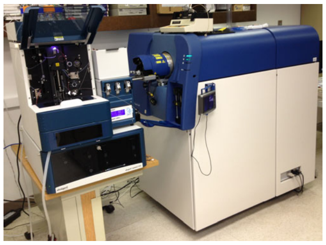
This is Next-GEN precise medicine

43

The UK National Phenome Center, LC-MS labs

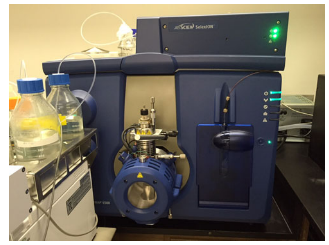
44

UAB capabilities in metabolomics

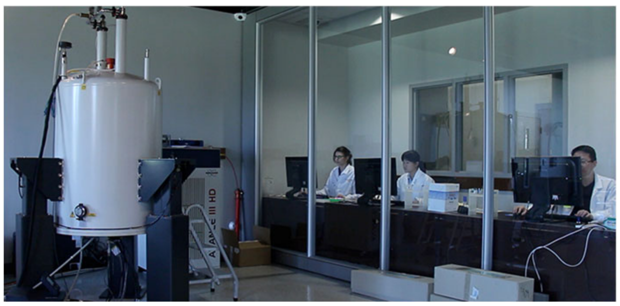


SCIEX 5600 TripleTOF
with Eksigent nanoLC

TMPL mass spec lab
BBRB 709/715
Stephen Barnes, Director
205-934-3462



SCIEX 6500 Qtrap with SelexION



Central Alabama NMR facility
Chemistry Bdg
William Placzek, Director
205-934-2465

45

Great challenges in metabolomics

- **The extent of the metabolome**
  - From gaseous hydrogen to earwax
- **Having complete databases**
  - METLIN has over 1 million 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**

46

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/
  - UNC-CH: http://www.uncnri.org/wp-content/uploads/2016/12/NIHERCMRC.pdf
  - 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
- **Other resources**
  - See this link

47