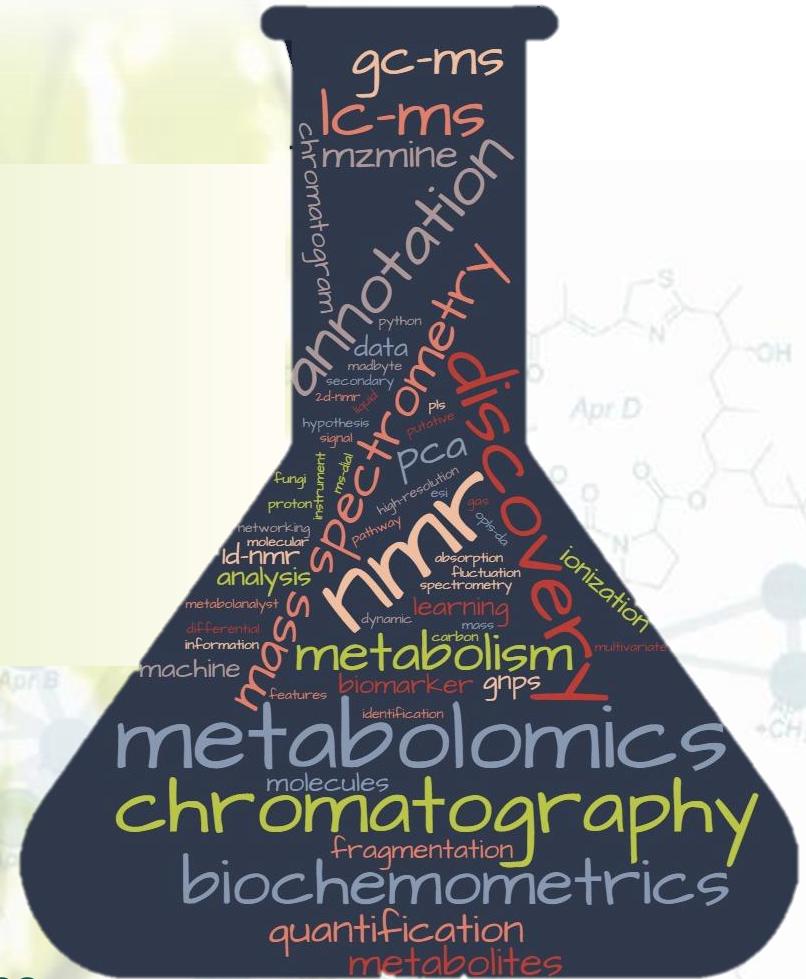


The Mysterious Multiverse of Metabolomics

Josh Kellogg, Ph.D.

Assistant Professor of Metabolomics
Veterinary and Biomedical Sciences
Pennsylvania State University

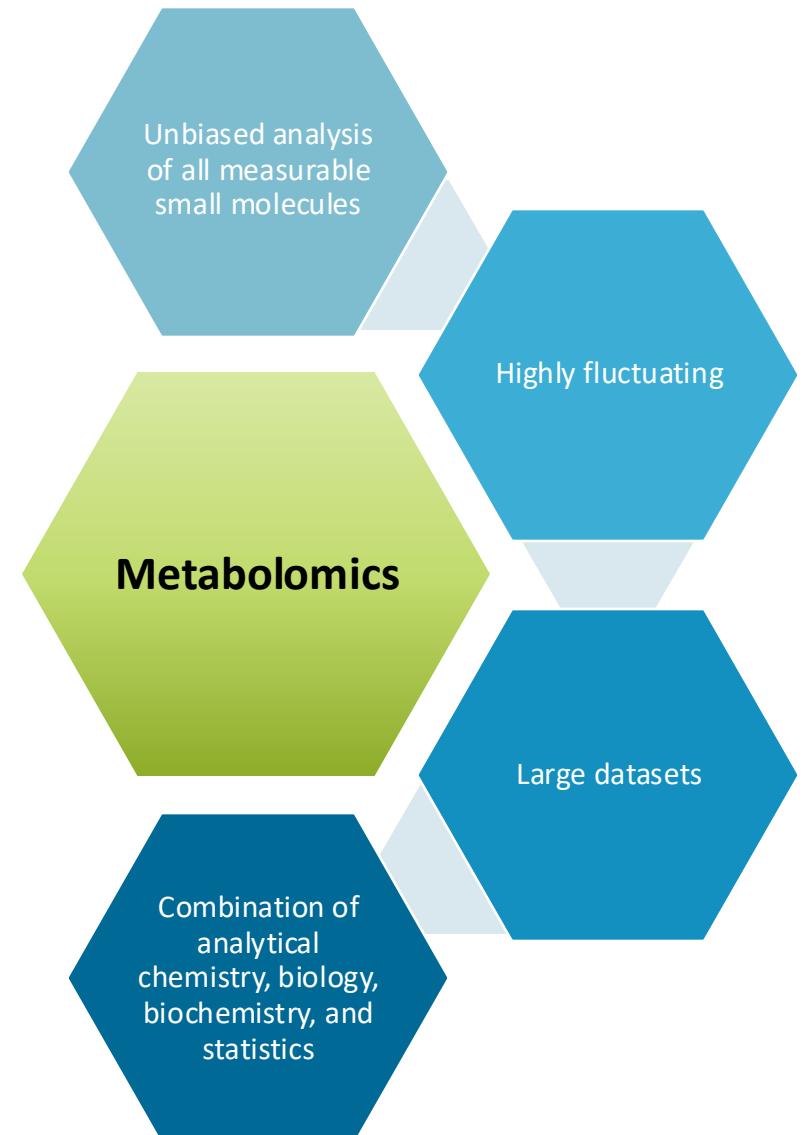


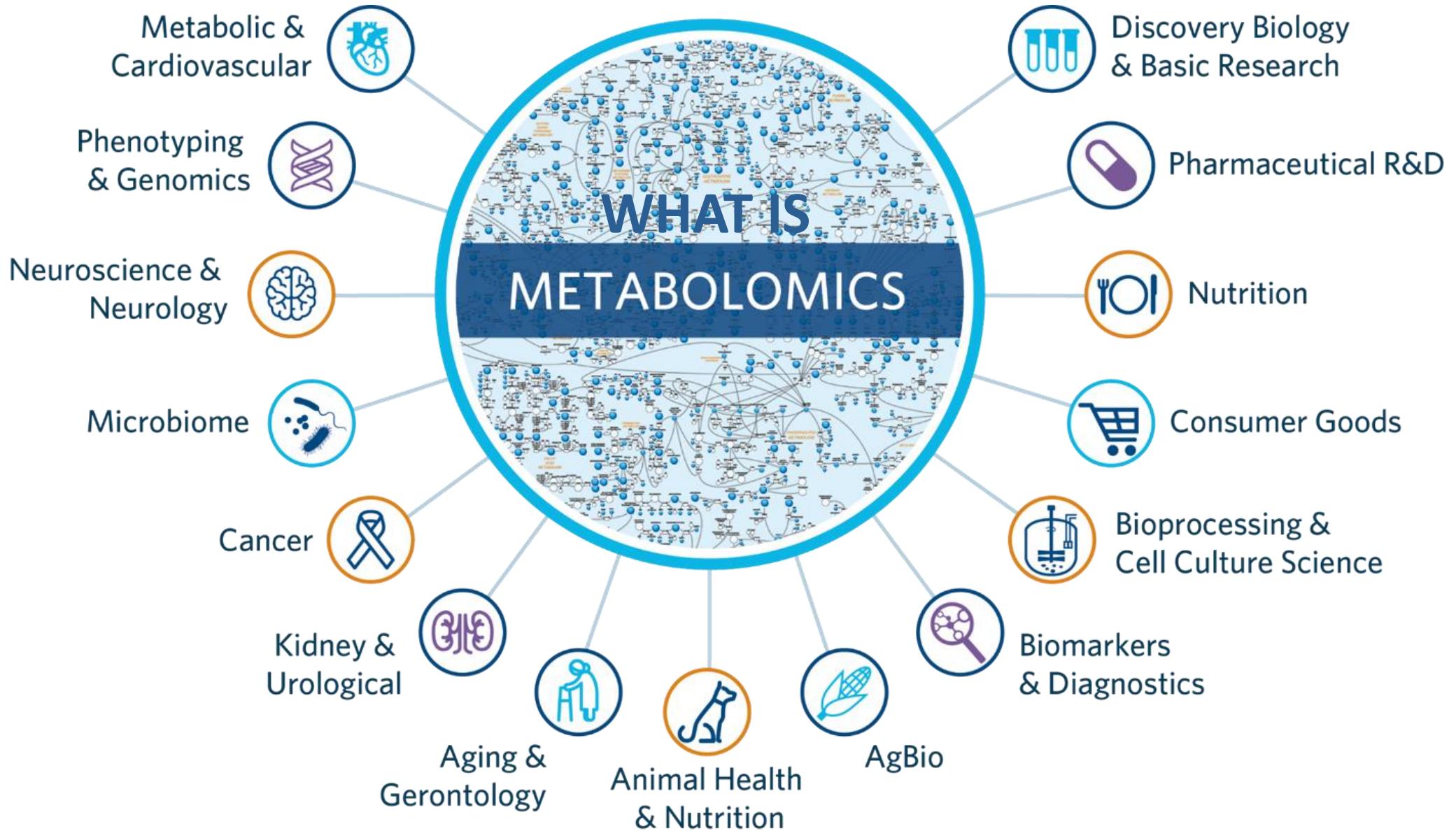
What is metabolomics?



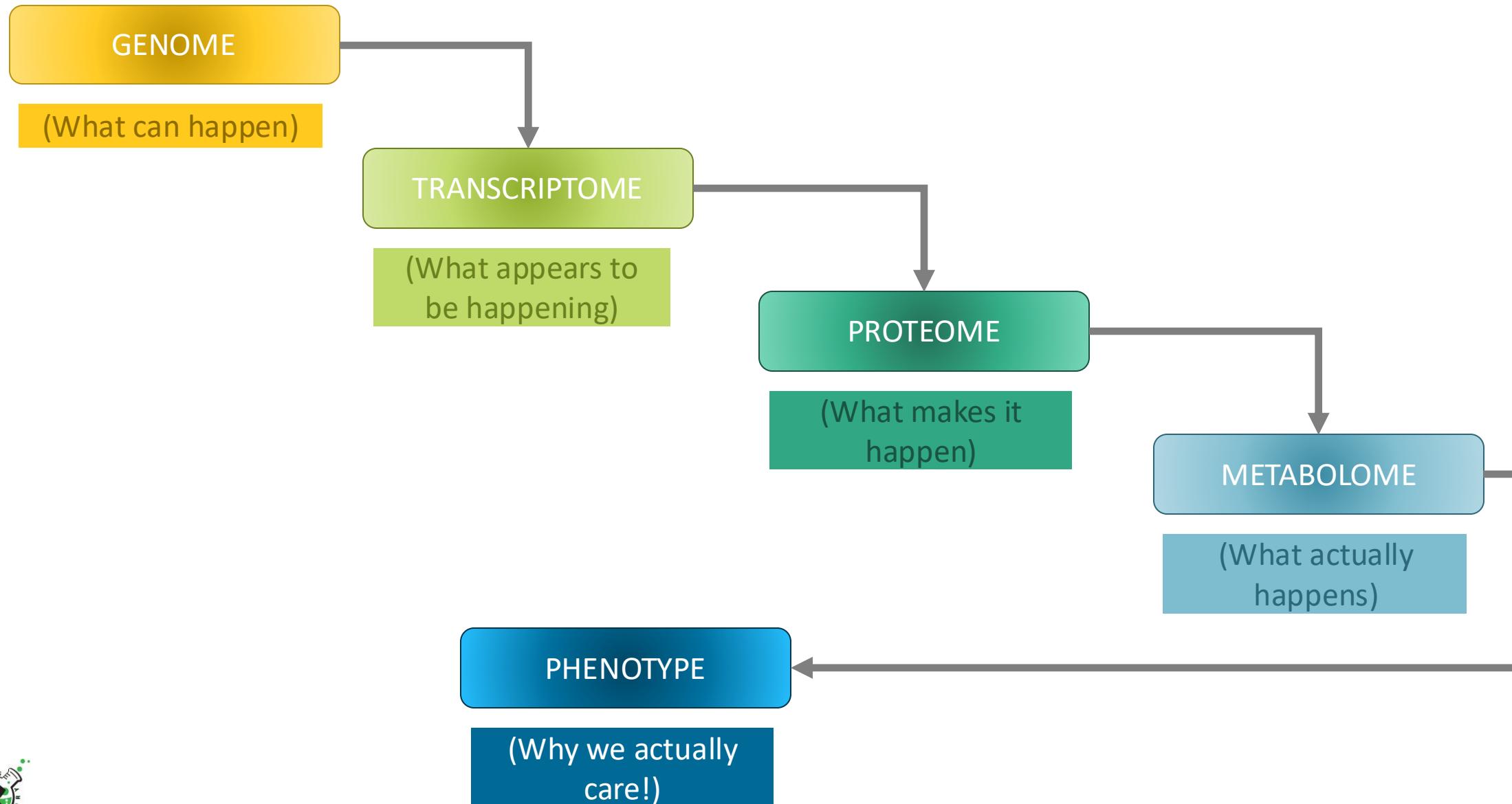
What is Metabolomics?

- **Metabolome:**
 - Total collection of all metabolites produced by an organism
 - Result of the totality of the organism's interactions with its surroundings
- **Metabolite**
 - Small molecule (< 1500 Da)
 - Produced by, metabolized by, or otherwise present in the organism

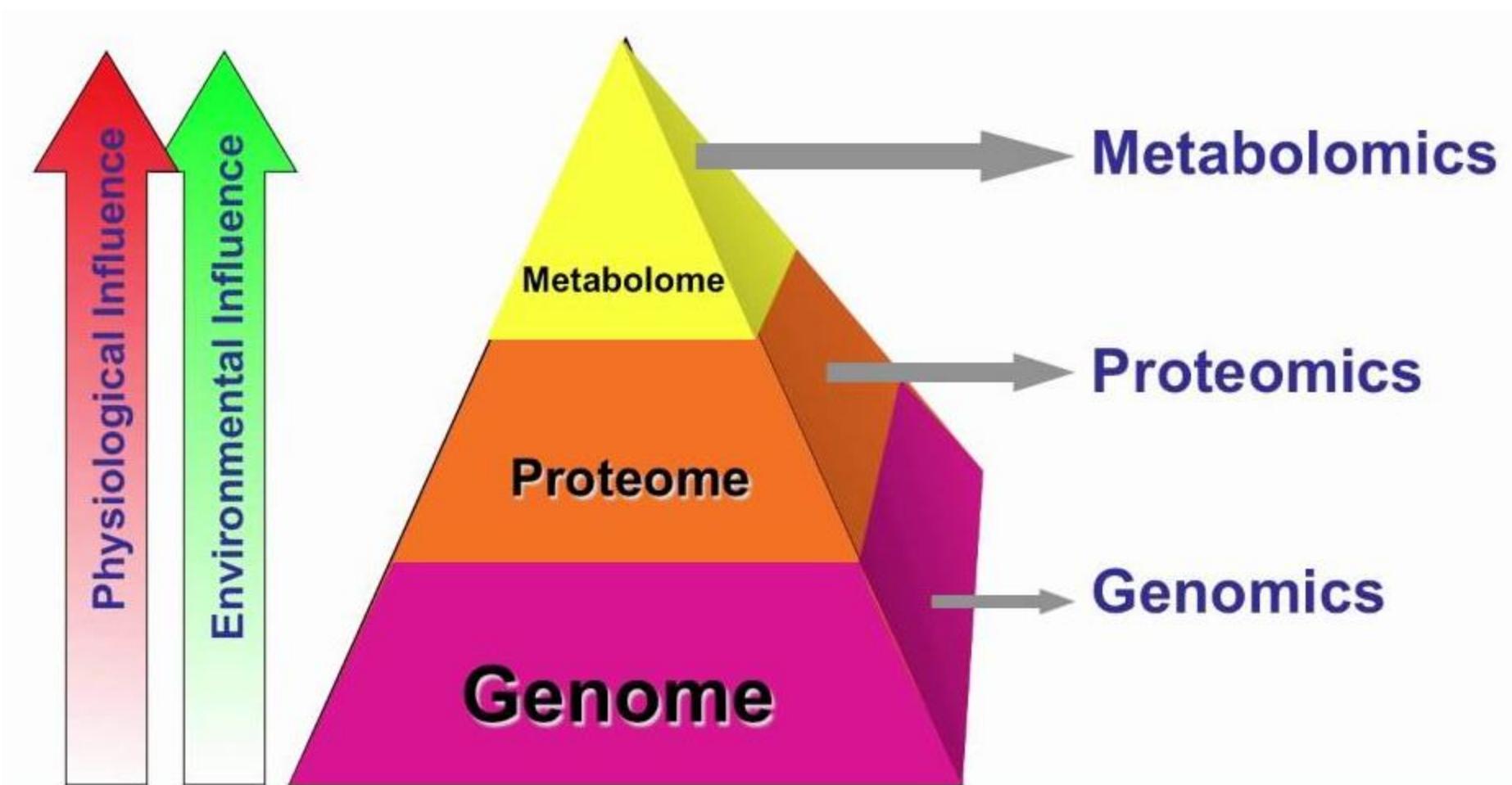




Why the Metabolome is a Joy



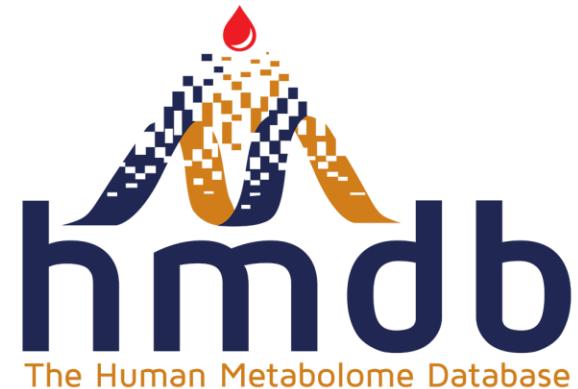
The interface between the organism and the environment



Why the Metabolome is a Challenge

Compound Diversity

- Organisms produce a wide variety of metabolites
- Plants produce vastly higher numbers than humans or fungi (so far...)

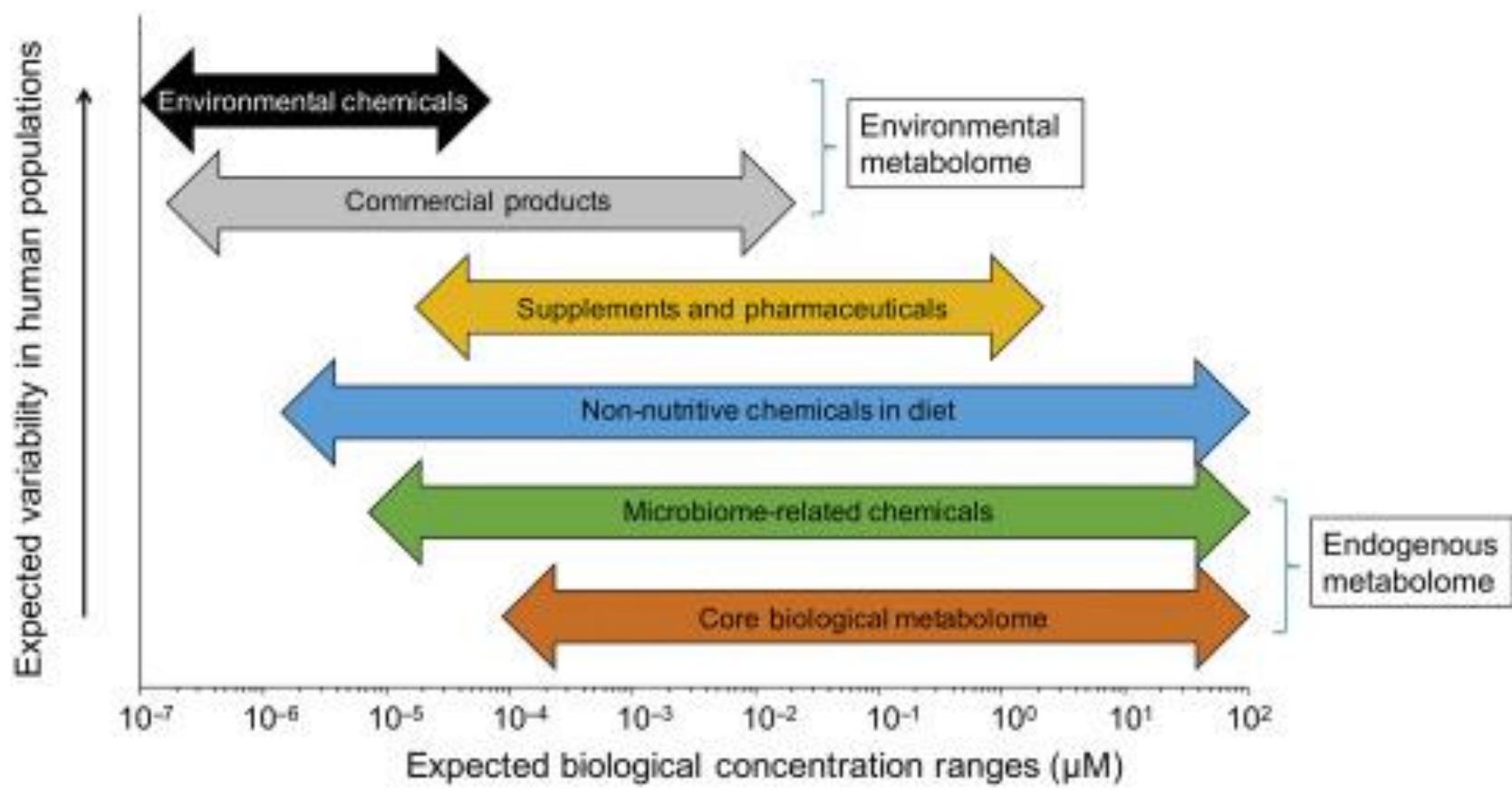


	Commonly circulating	Total possible**
Humans	ca. 3,000-6,000	ca. 115,000
Plants	ca. 7,000-15,000	200,000-1,000,000
Fungi	Variable	ca. 75,000+
Bacterial	Variable	

Complexity

Concentration

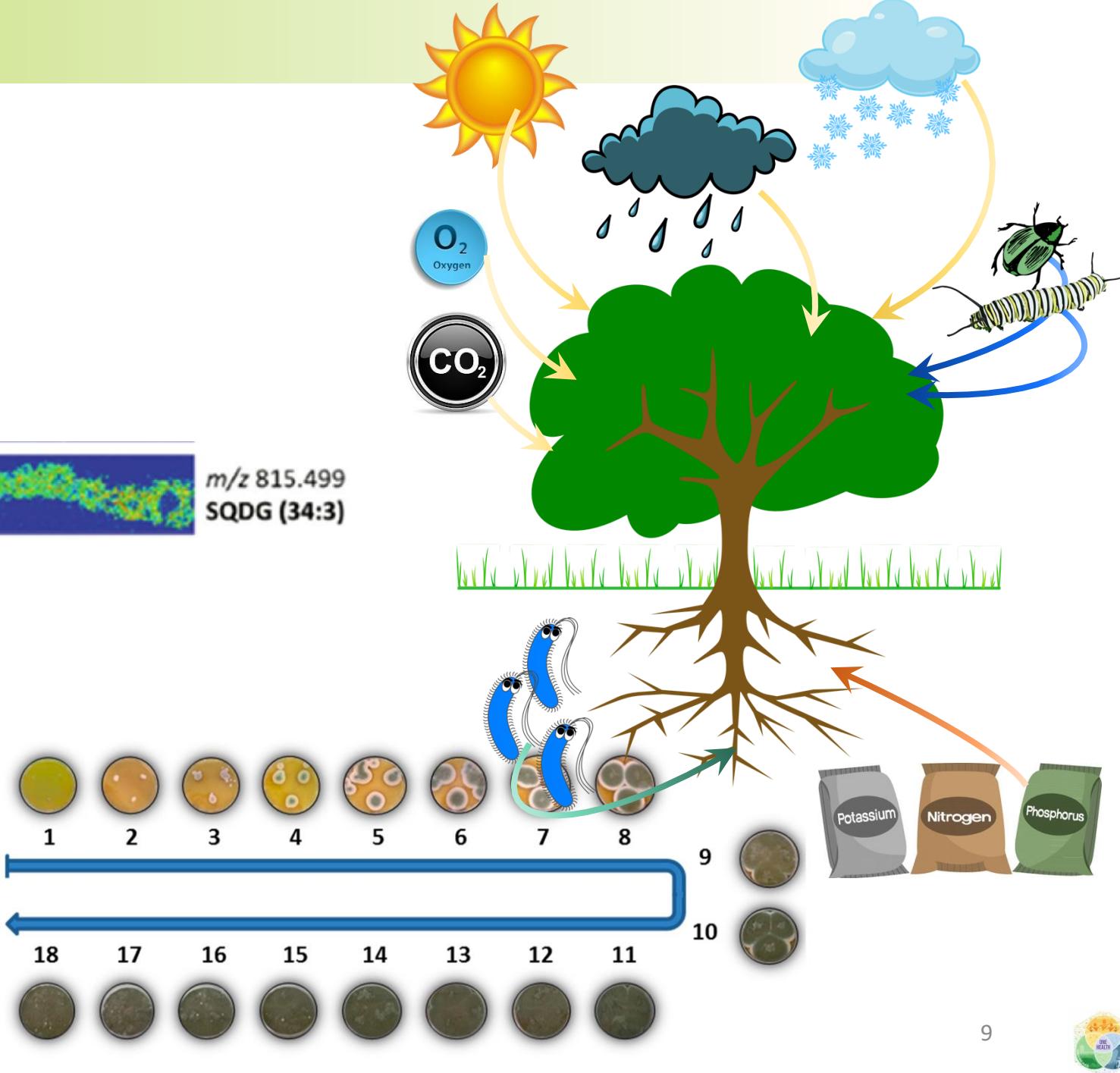
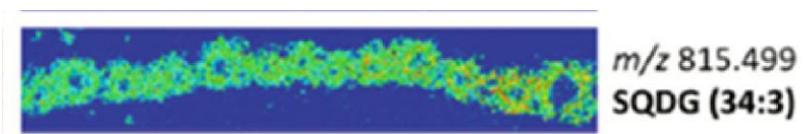
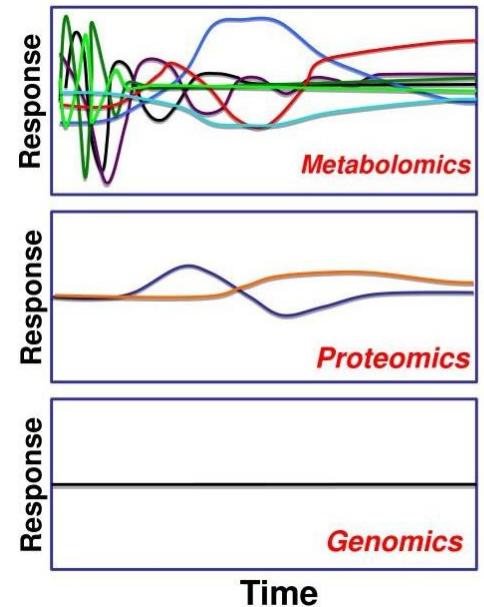
- Often 4-8 orders of magnitude difference in concentration



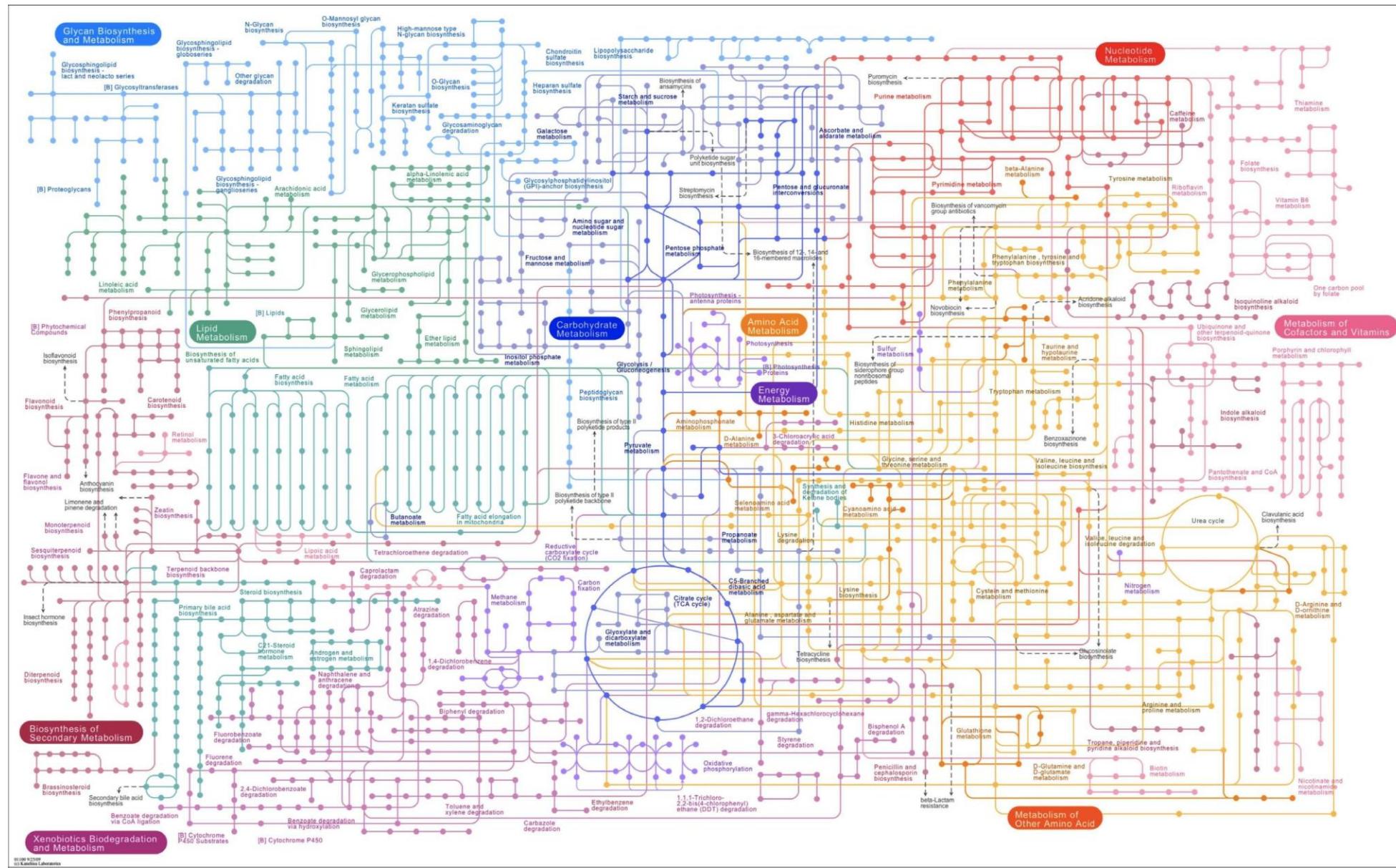
Complexity

Dynamic

- Spatial (tissue-specific)
- Temporal (intra-day, seasonal, annual)
- Environmental (abiotic and biotic stresses)
- Geographical

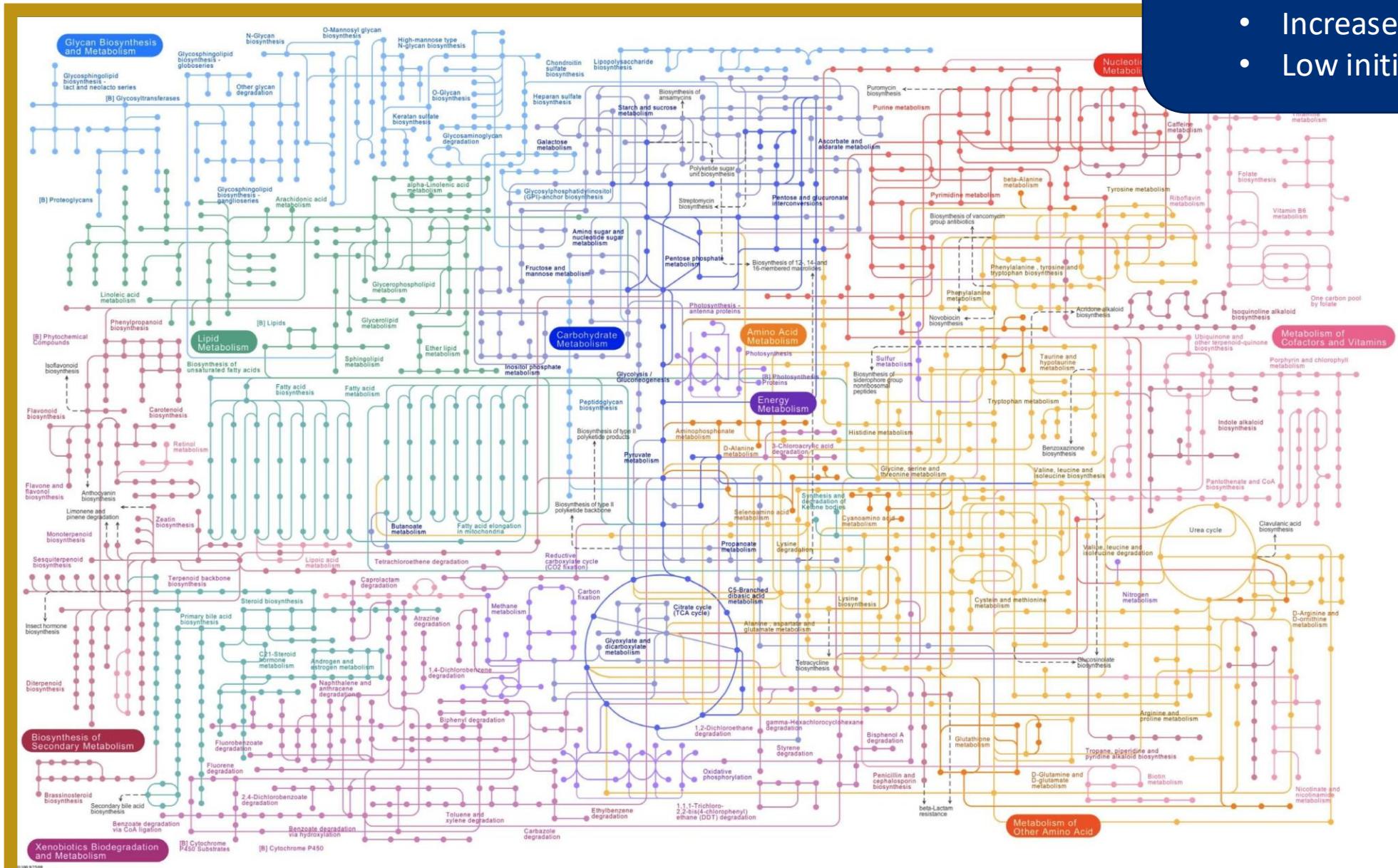


Metabolism



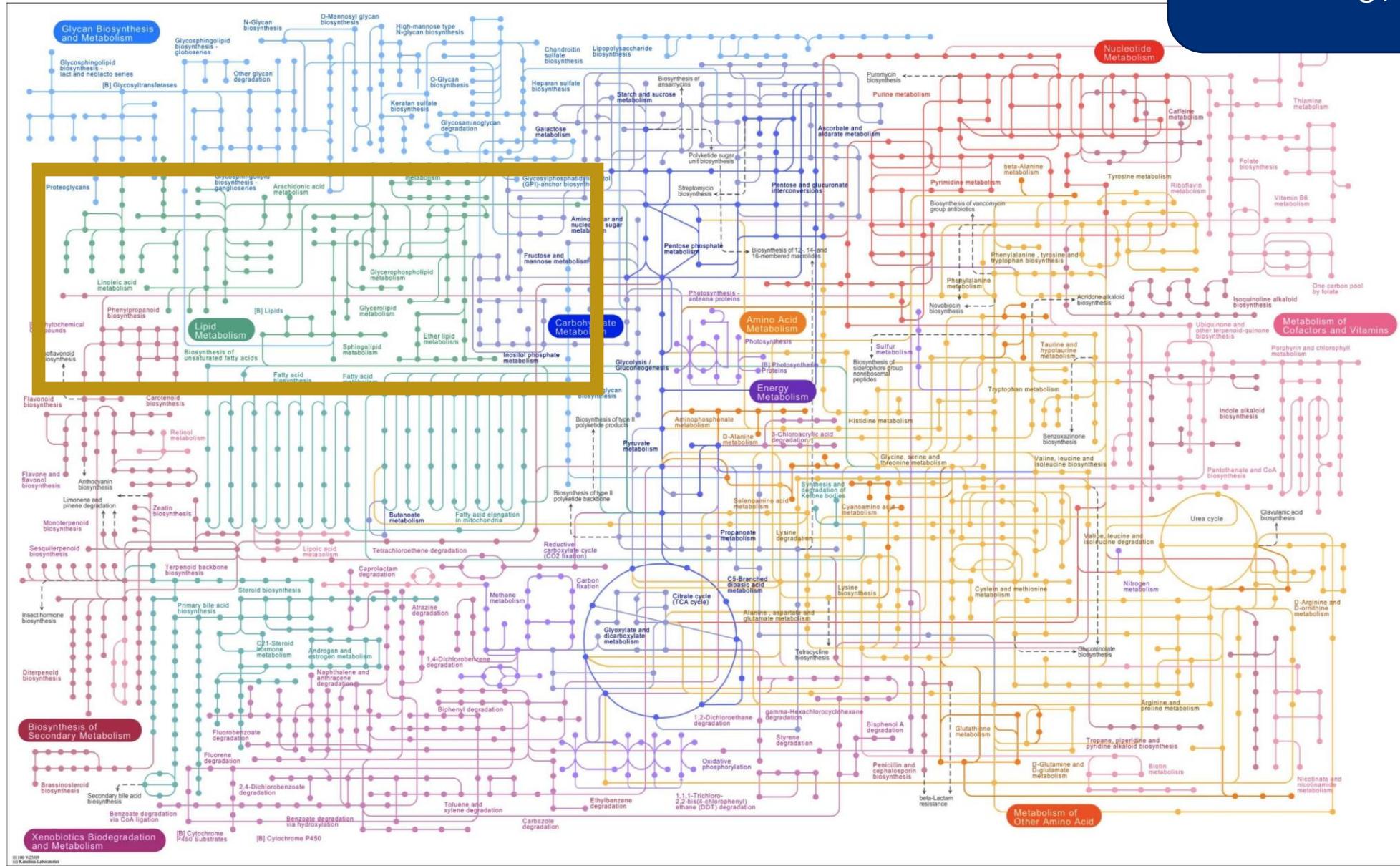
Global Metabolomics

- Untargeted metabolomics
- Profile all measurable metabolites
 - Increased coverage
 - Low initial annotation



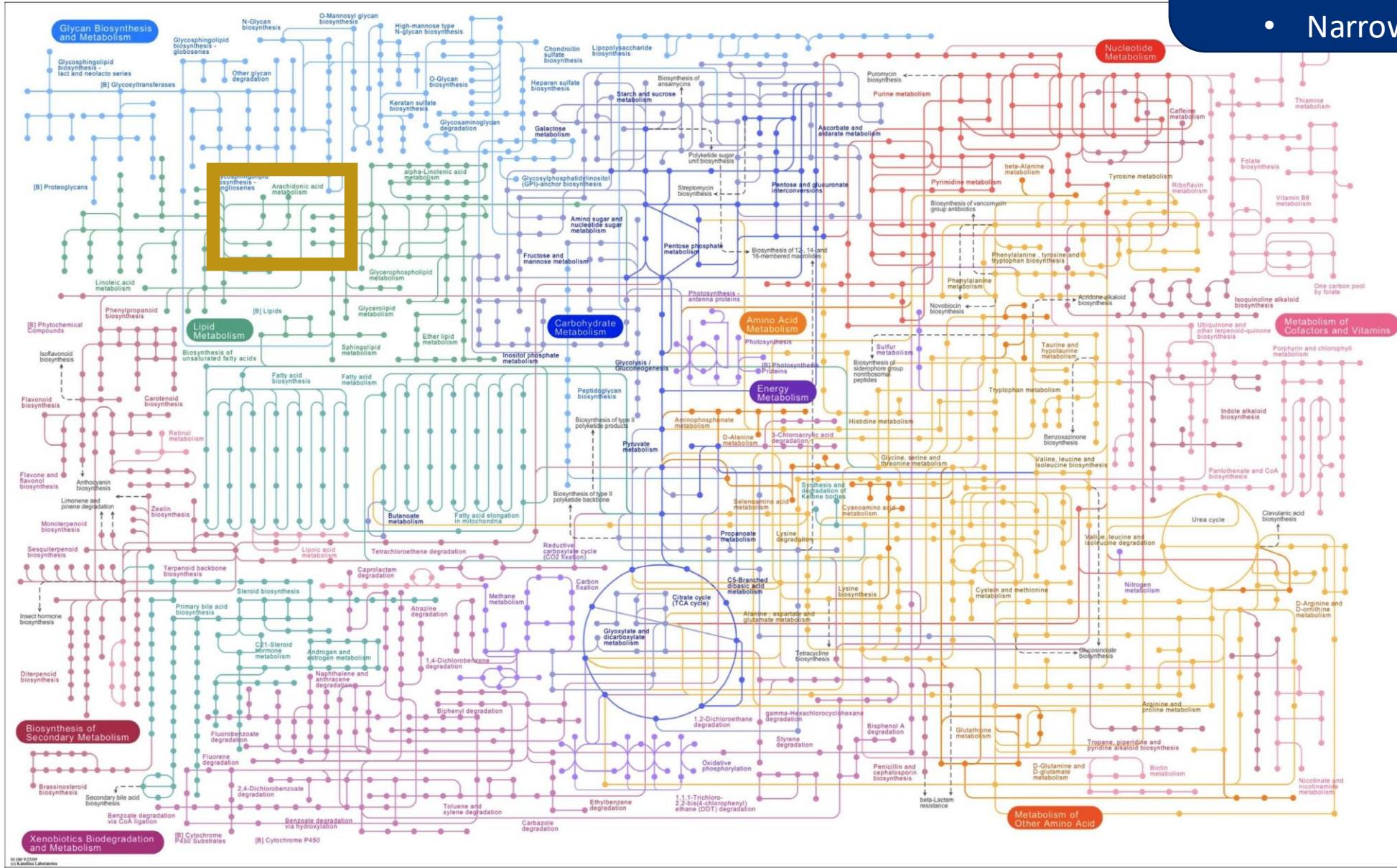
Metabolic Profiling

- Profile a subset of metabolites (usually one class... e.g., lipidomics)

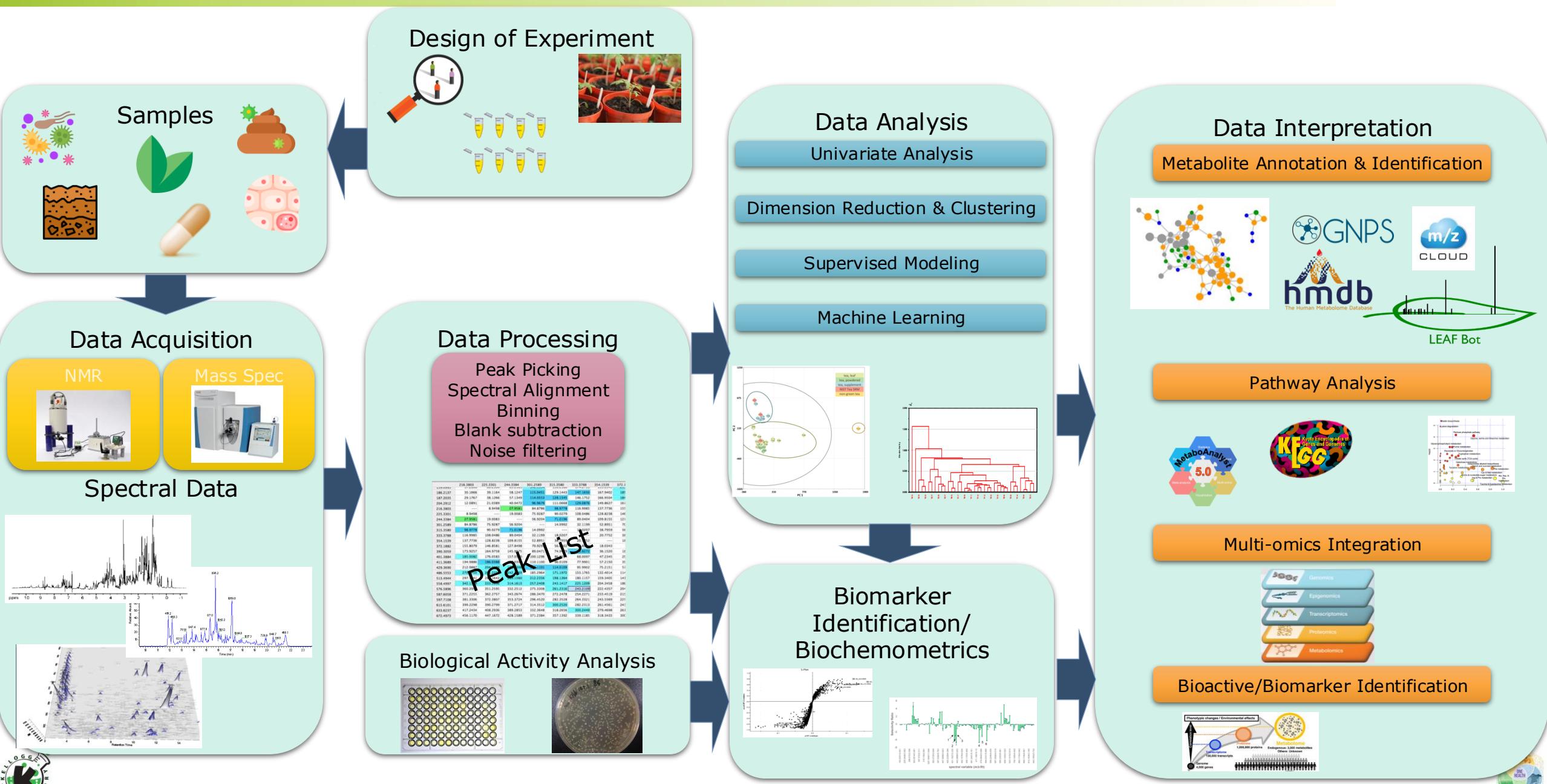


Targeted Metabolomics

- Profile a narrow subset of metabolites (e.g., prostaglandins)
 - More quantifiable
 - Narrower scope

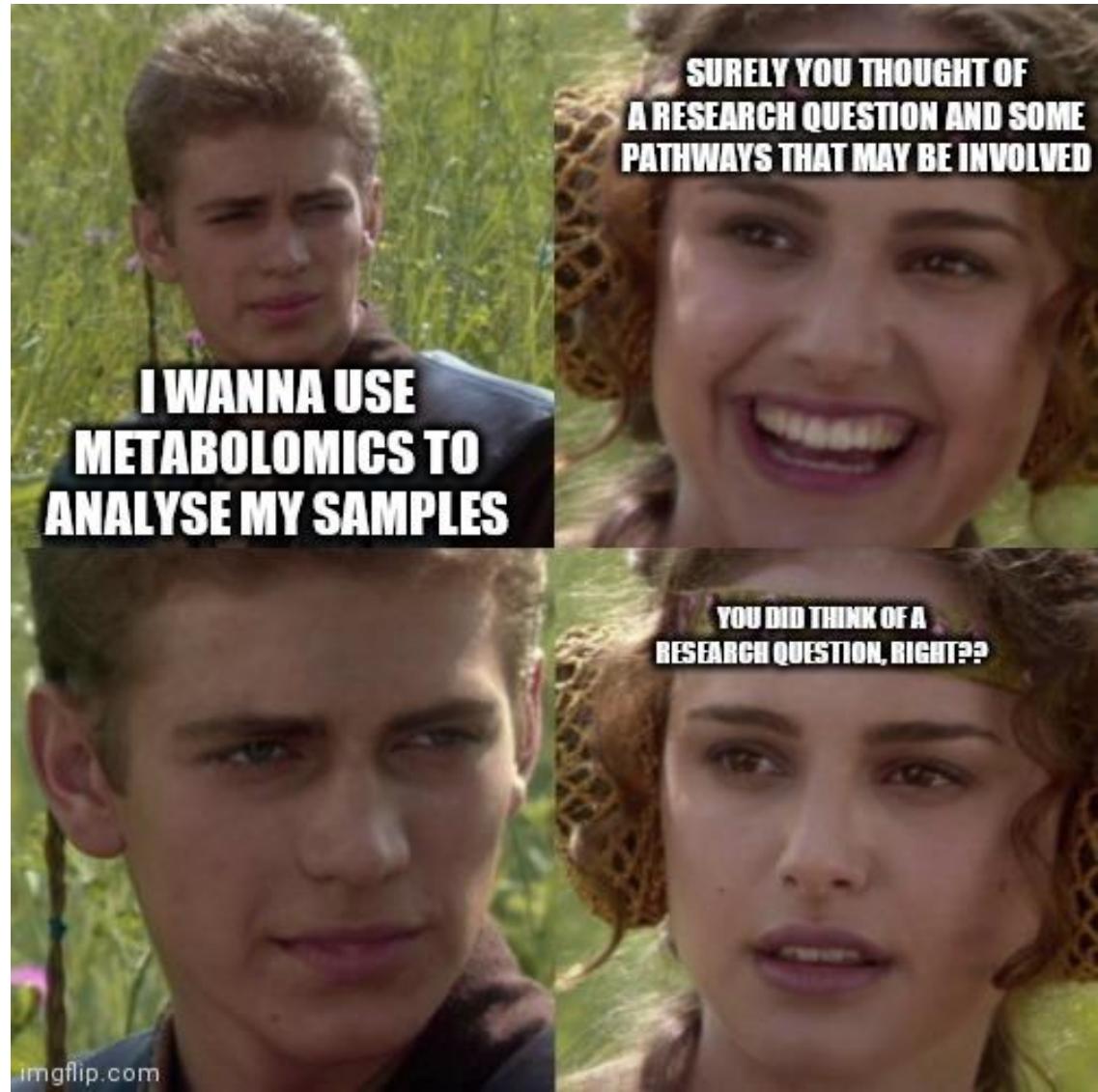


Metabolomics Workflow



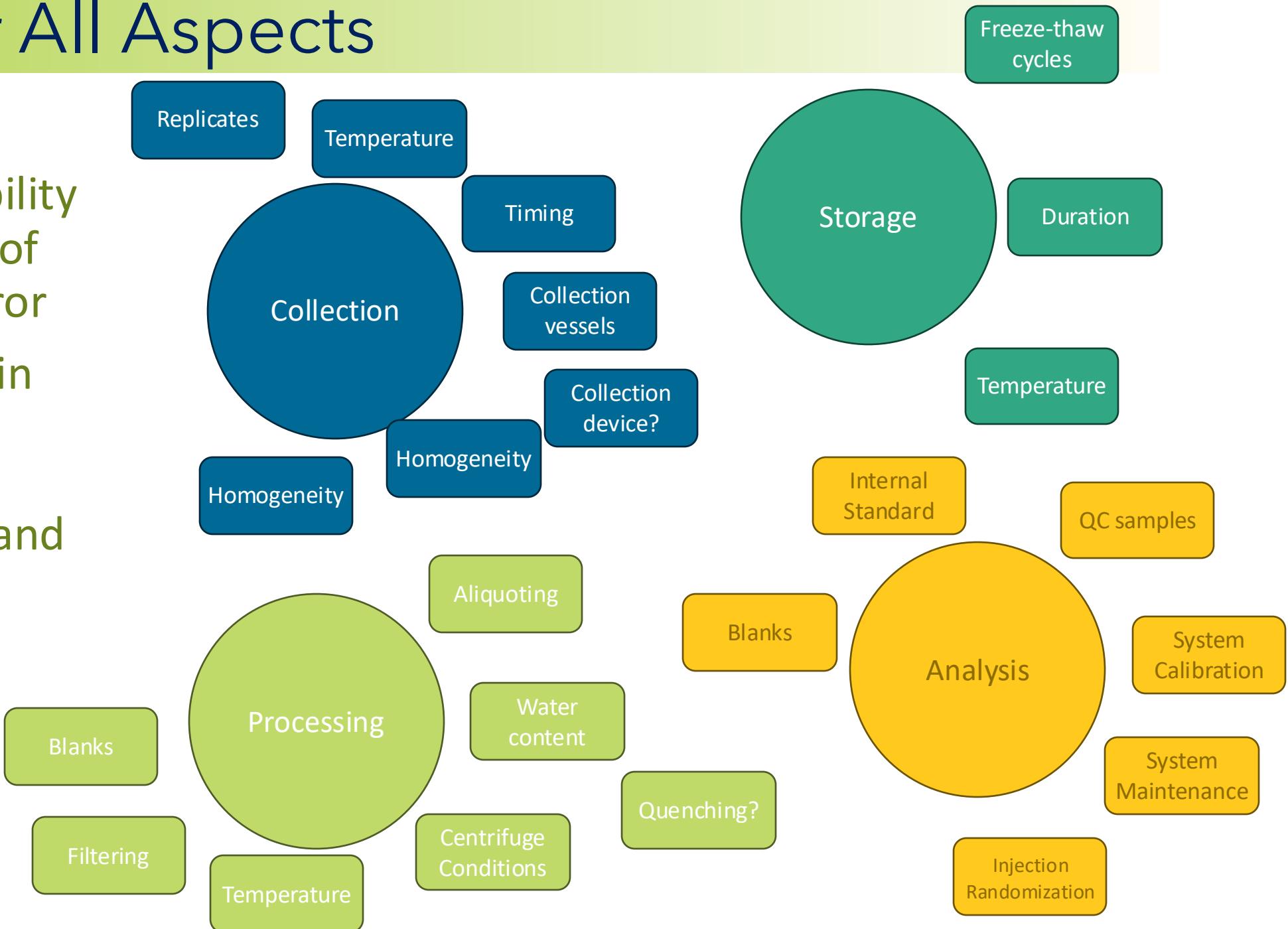
Experimental Design

- What you put in is what you get
- Metabolomics studies, to be effective and robust, need to be planned from the get-go
 - Not as an after thought
- Want statistical analysis/clustering to result from metabolome change not experimental variability
- **EVERYTHING** should be as **CONSTANT** as possible between samples or the study lacks power and rigor



Consider All Aspects

- Lower variability and sources of unknown error
- Consistency in methods, protocols, equipment, and technique



Design of Experiments



Define your objective



Consider all aspects of your study to reduce error



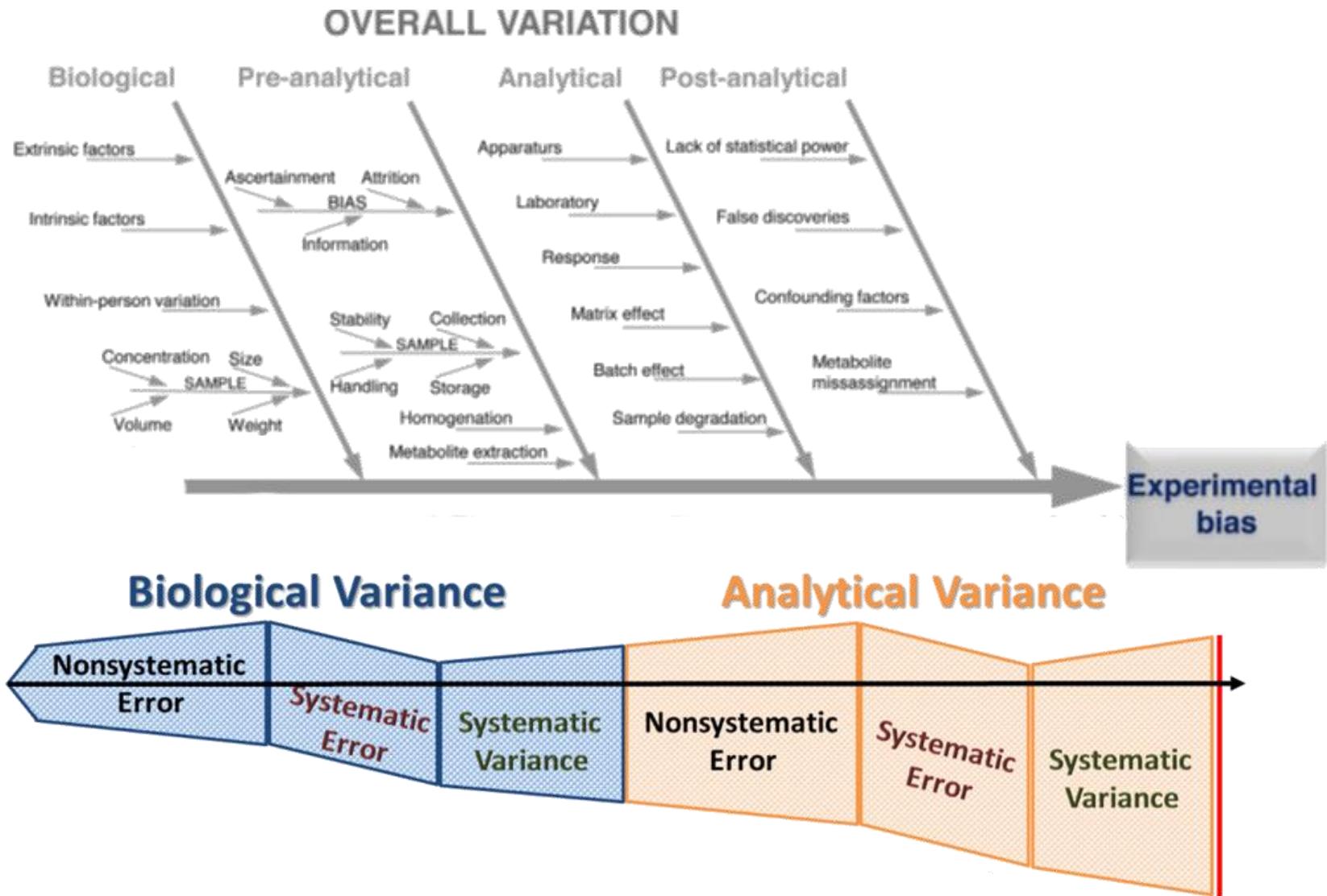
Replication, replication, replication!



Analytical Technique



Have a Data Interpretation Plan in Place



Extraction & Fractionation

Goal: Access as much of the chemistry as possible in the samples



Solid
Phase
Extraction

Distillation

Organic
Solvent
Maceration

Microwave
-assisted
Extraction

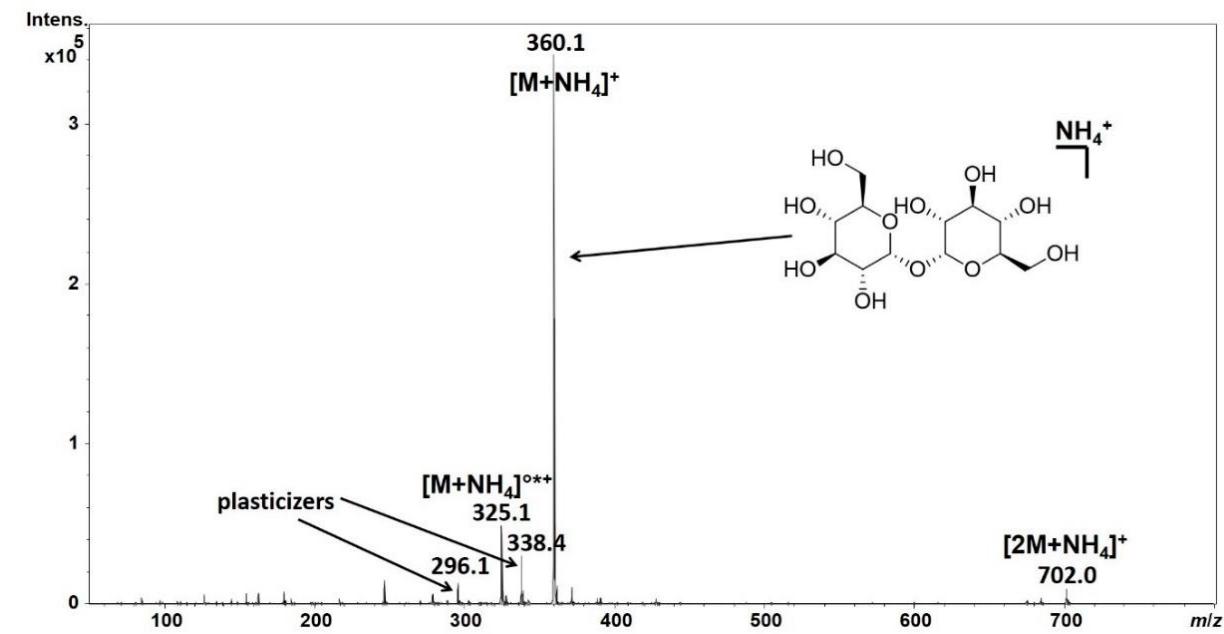
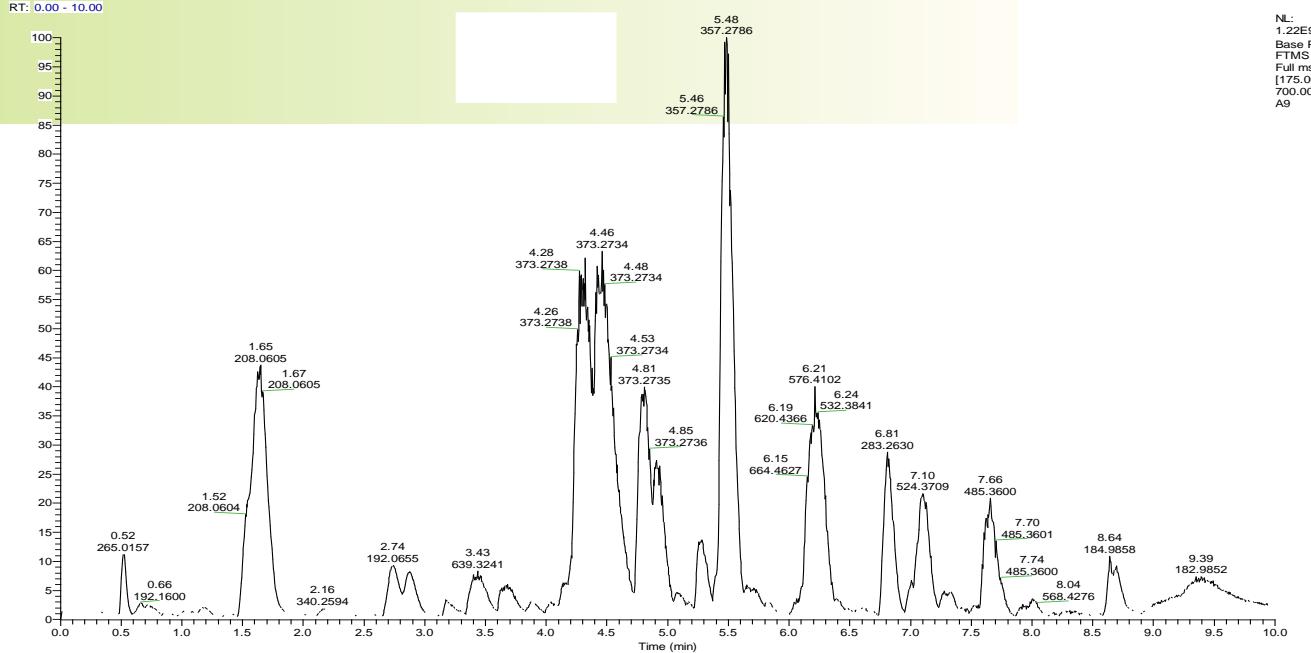
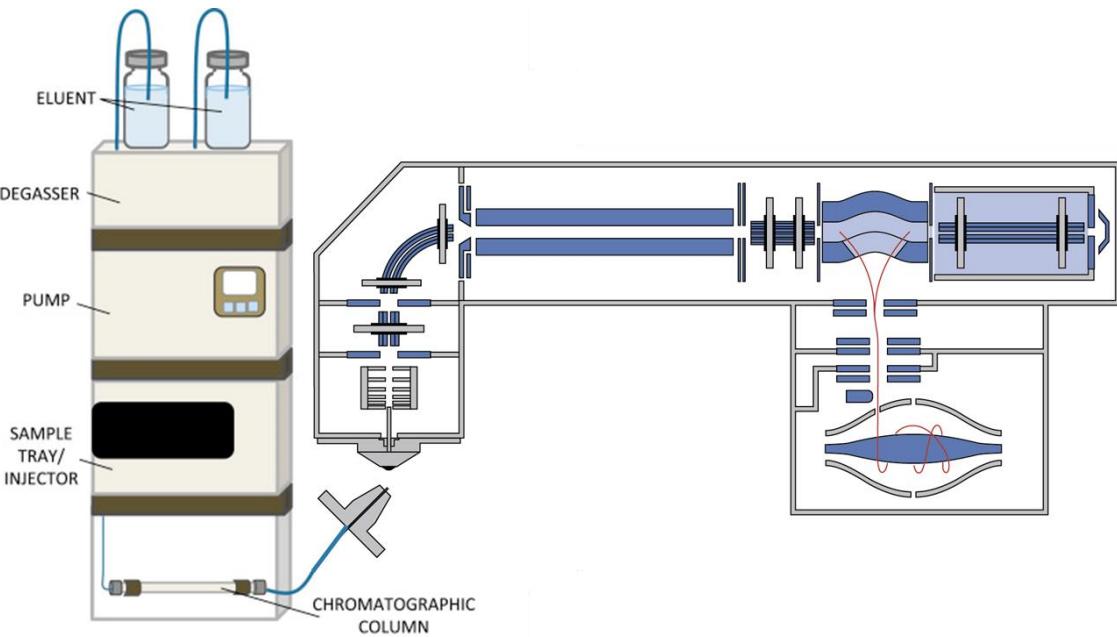
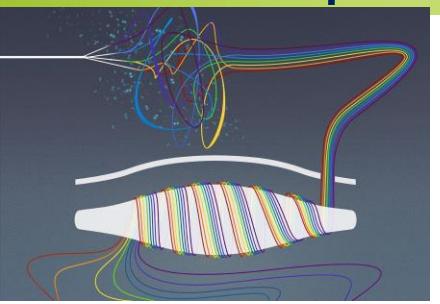
Supercritical
Fluid
Extraction

Ionic
Liquids

Ultrasound
-assisted
extraction

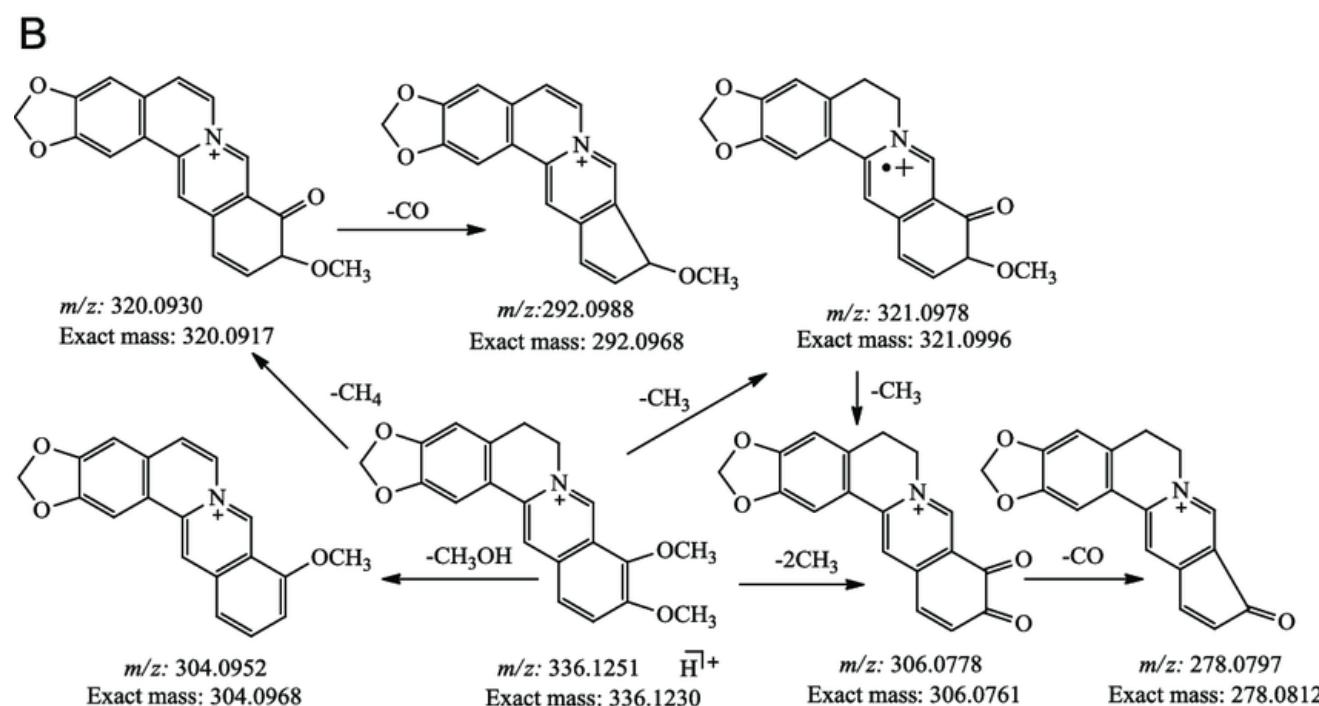
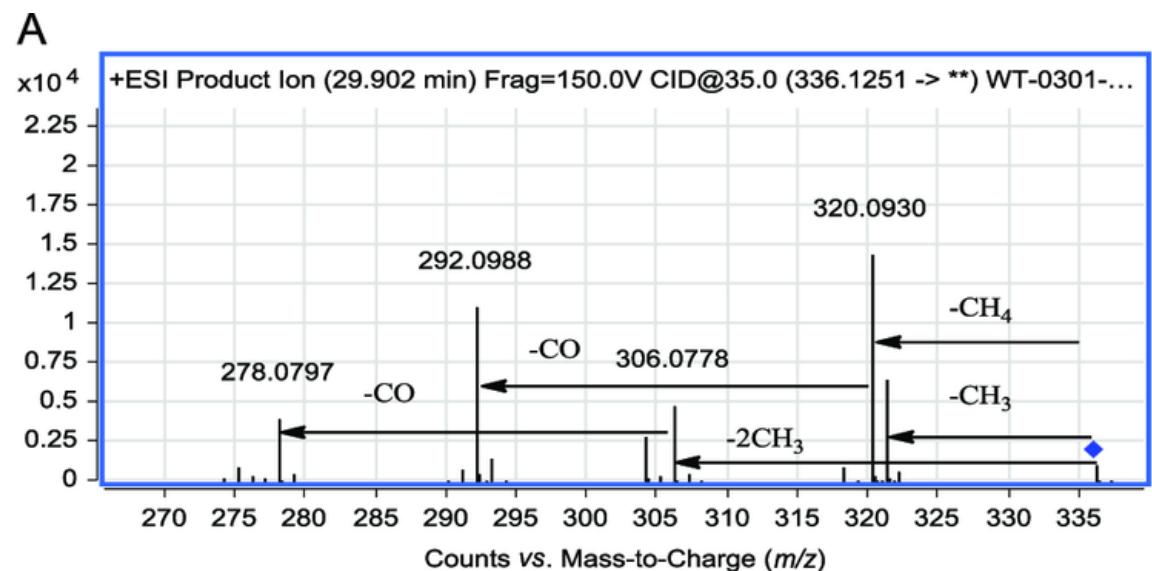
Accelerated
Solvent
Extraction

Mass Spectrometry



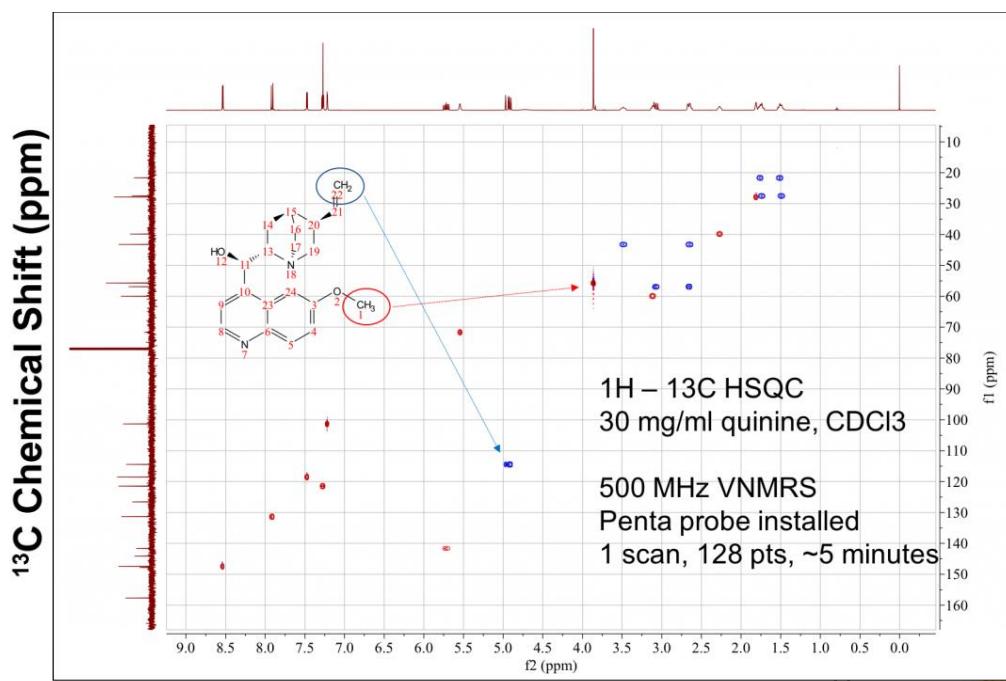
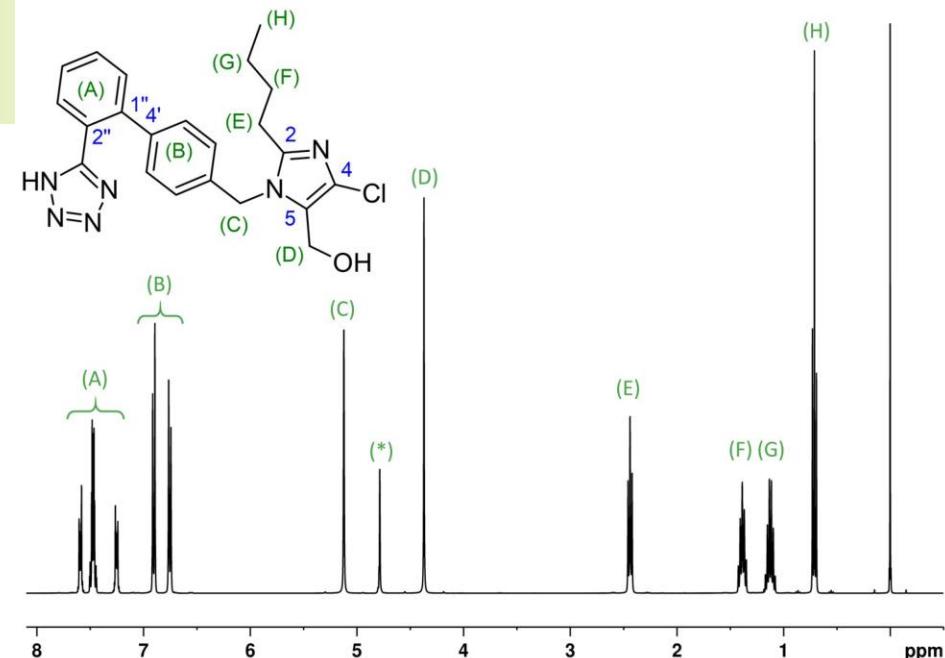
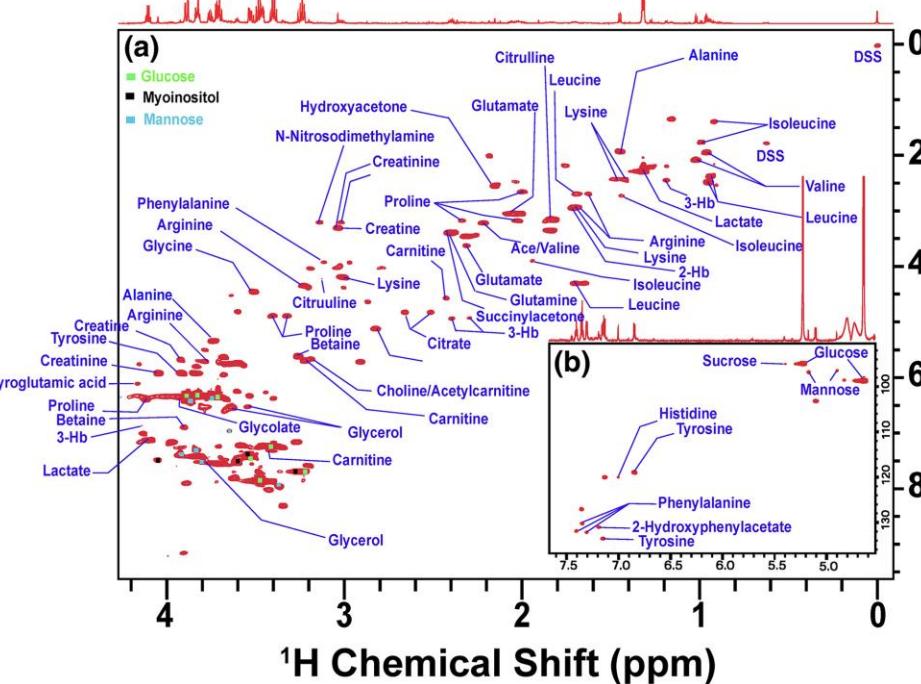
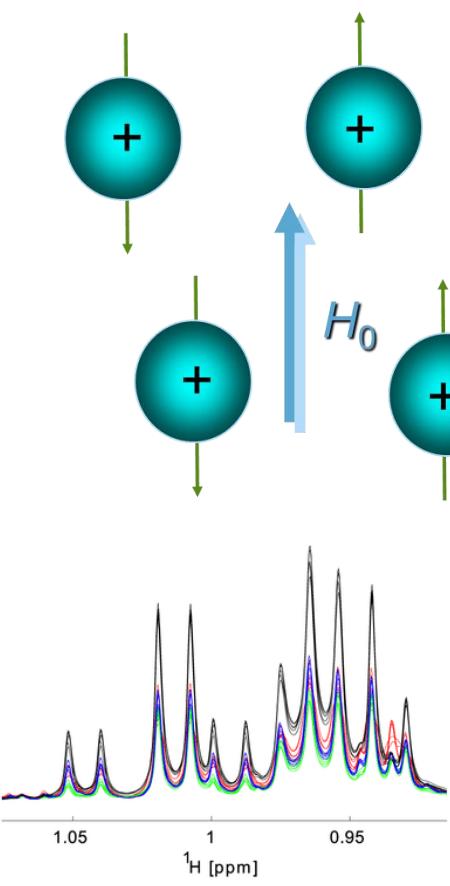
MS Fragmentation

- Applying voltage or forcing collisions within the MS causes molecules to fragment
 - Provide additional structural information
 - Allow for library construction for annotation of unknowns



NMR

An external magnetic field causes nuclear magnetic moments to align parallel and antiparallel to applied field.



PSU Metabolomics & NMR Core



SciEx
ZenoTOF
7600



Thermo
Exploris
240

Thermo
Exploris
120



Bruker AV-
III 850

High resolution
MS

Thermo
Fusion
Lumos

NMR

Bruker AV-
III 600



Bruker NEO
600

Quant MS



Thermo
TSQ



SCIEX
6500+
QTrap

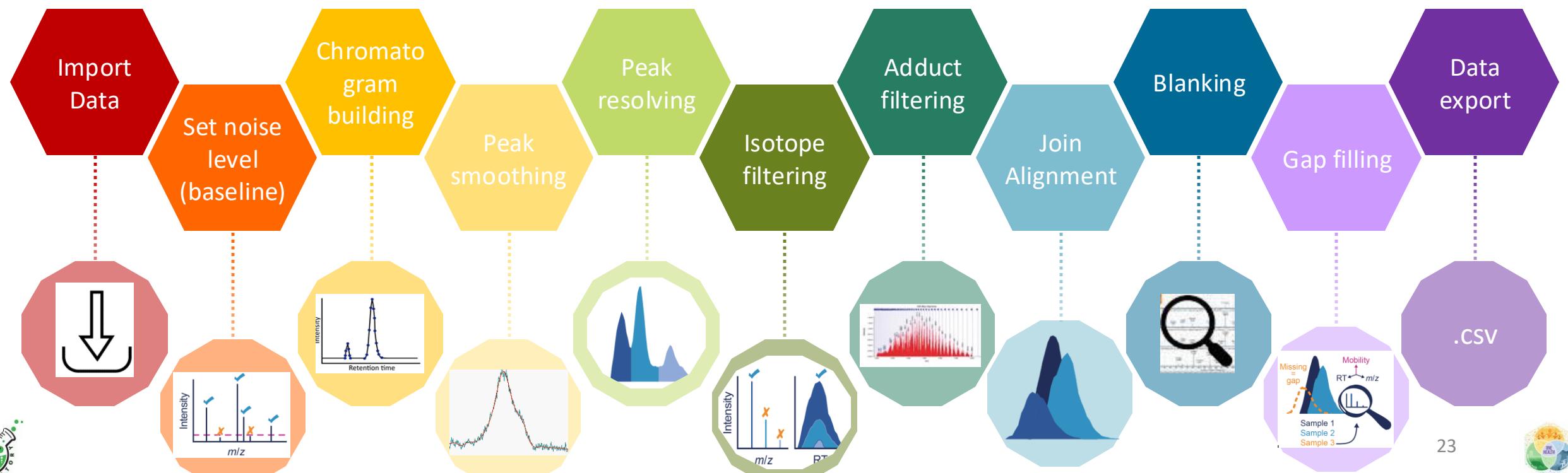


Agilent
5975 GC-
MS



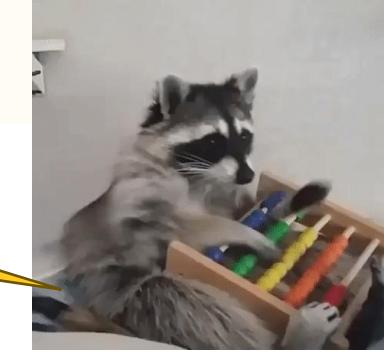
Bruker AV-
III 500

Data Analysis

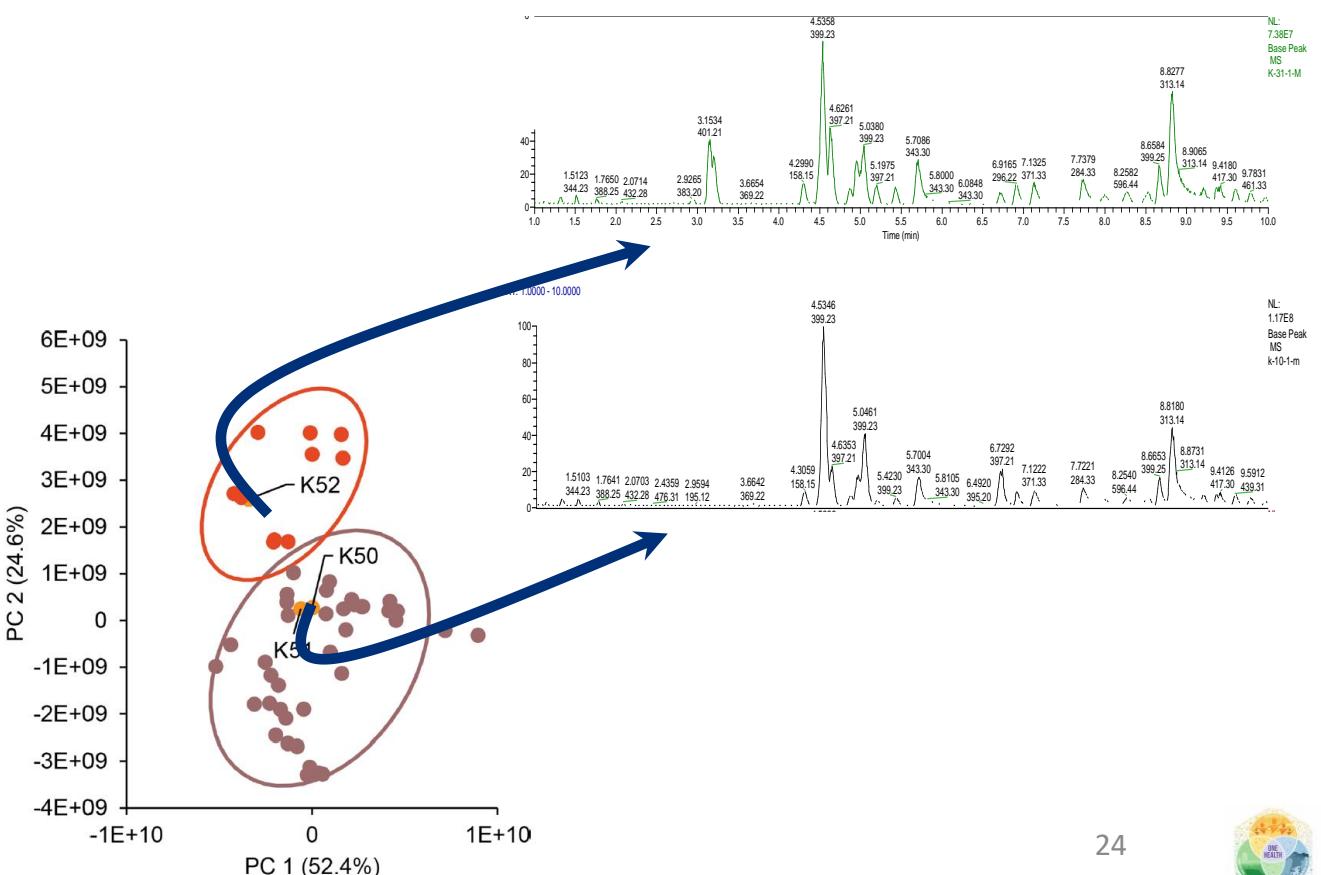
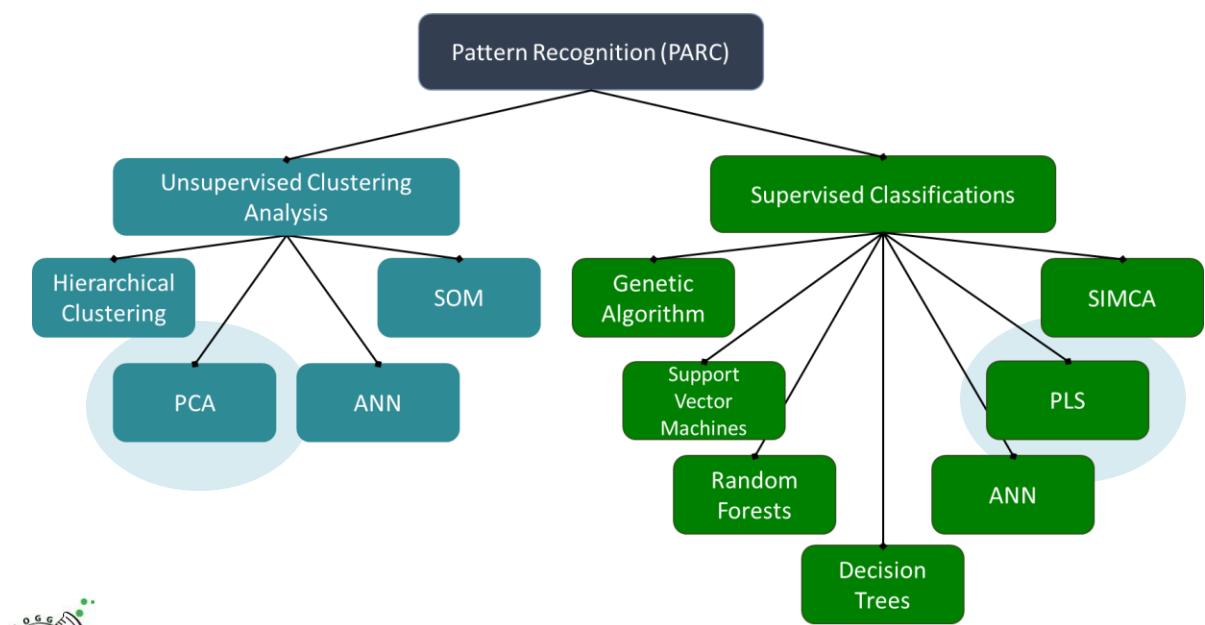


Data Analysis

Computation is key!
We ❤️ stats!!



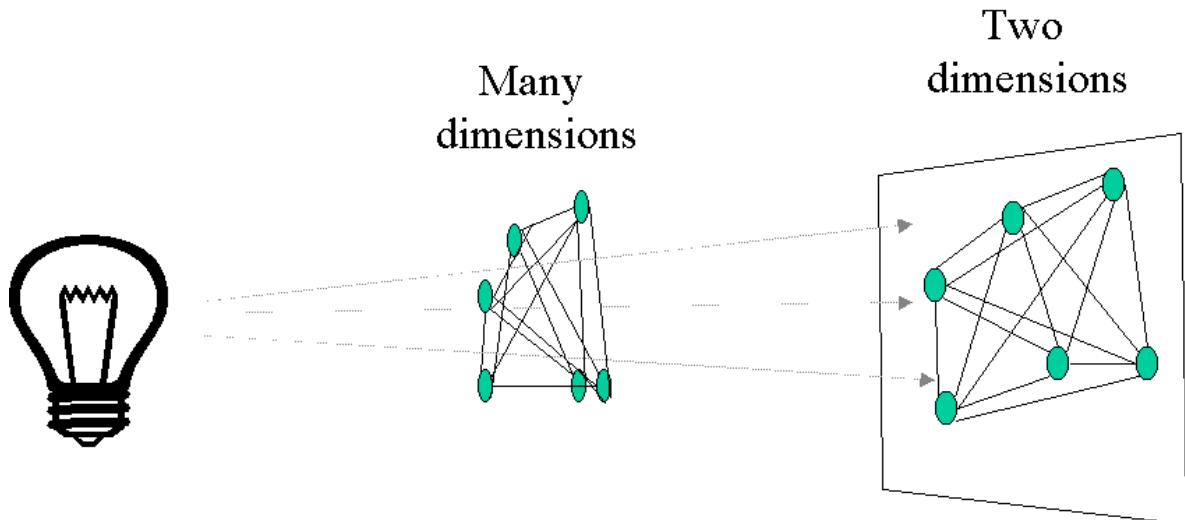
- Data from any metabolomic platform will produce “fat datasets.”
 - datasets with relatively few rows, i.e. samples or biological replicates, and many columns, i.e. variables, in this case, mass features or metabolites.
- Many variables in the dataset are interdependent, e.g., because they are in the same biosynthetic route.
- Most statistical models, such as Multiple Analysis of Variance (MANOVA), are not suitable for dealing with this kind of dataset.



All About Dimension Reduction (Unsupervised)

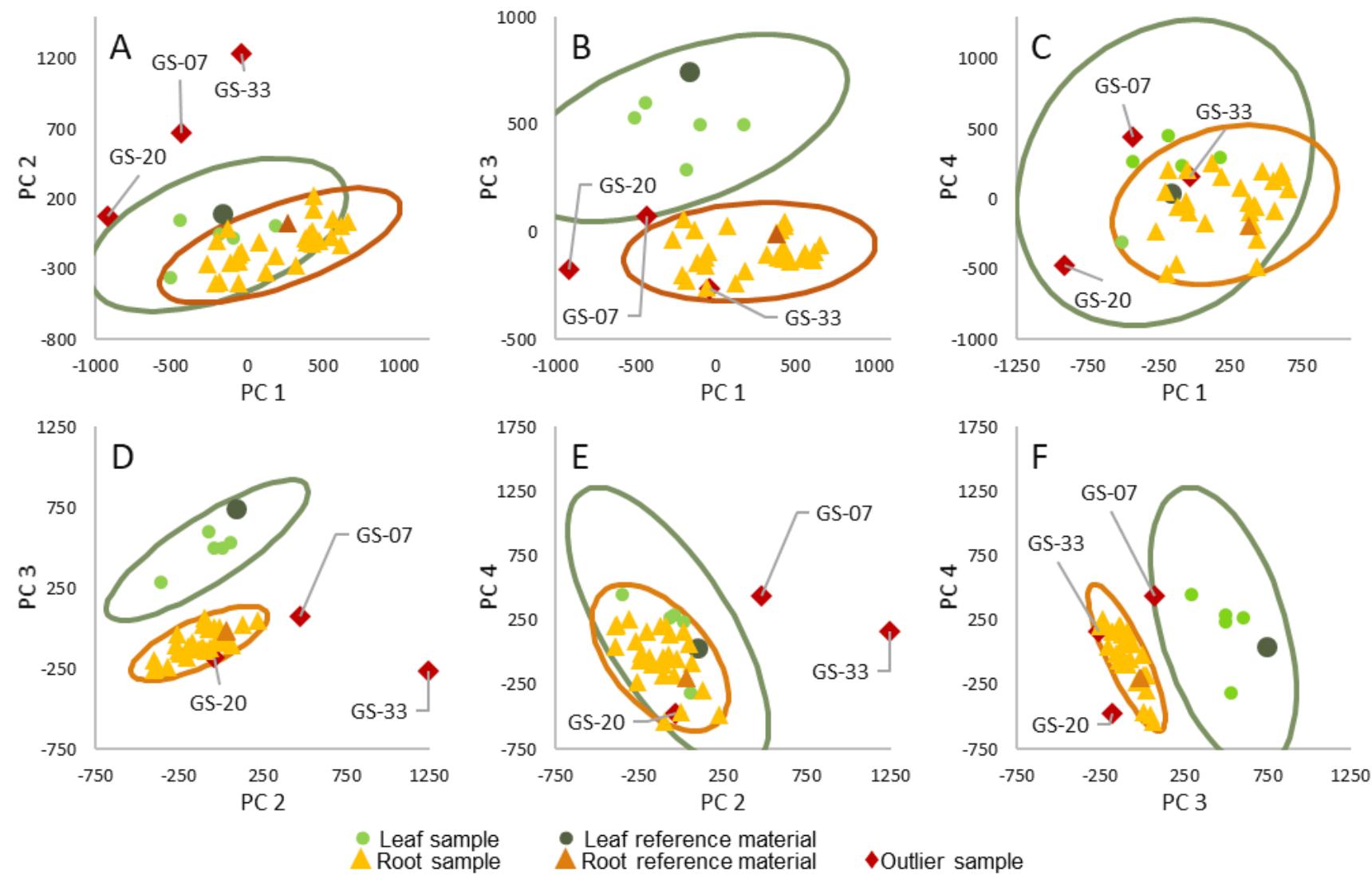
- Uses just the metabolomic/chemistry data
 - Simplifies complex high-dimensional data
 - Summarize data with a lower dimensional model

- Given data points in d dimensions
- Convert them to data points in $r < d-1$ dimensions
 - “principal components”
- With minimal loss of information

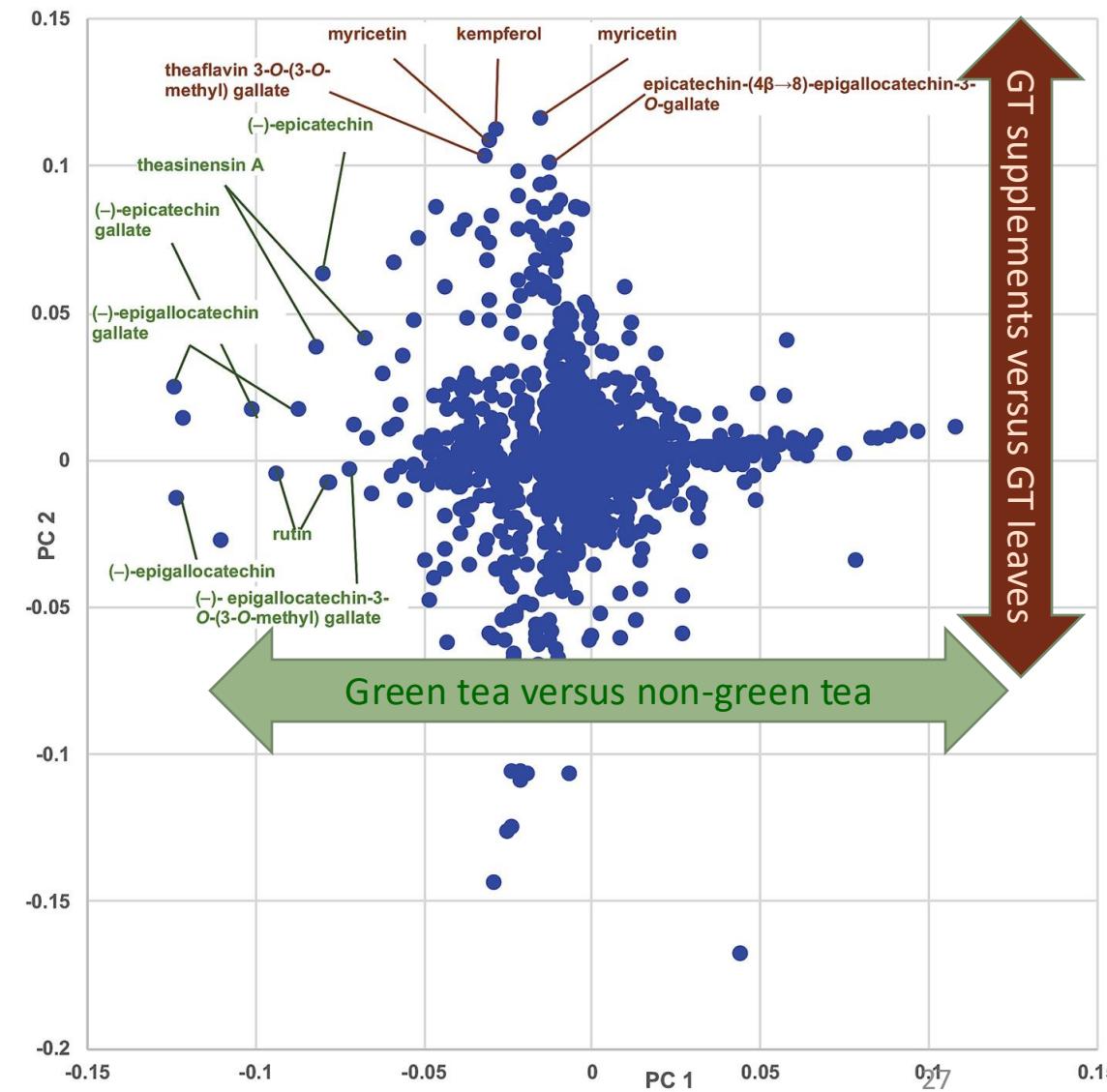
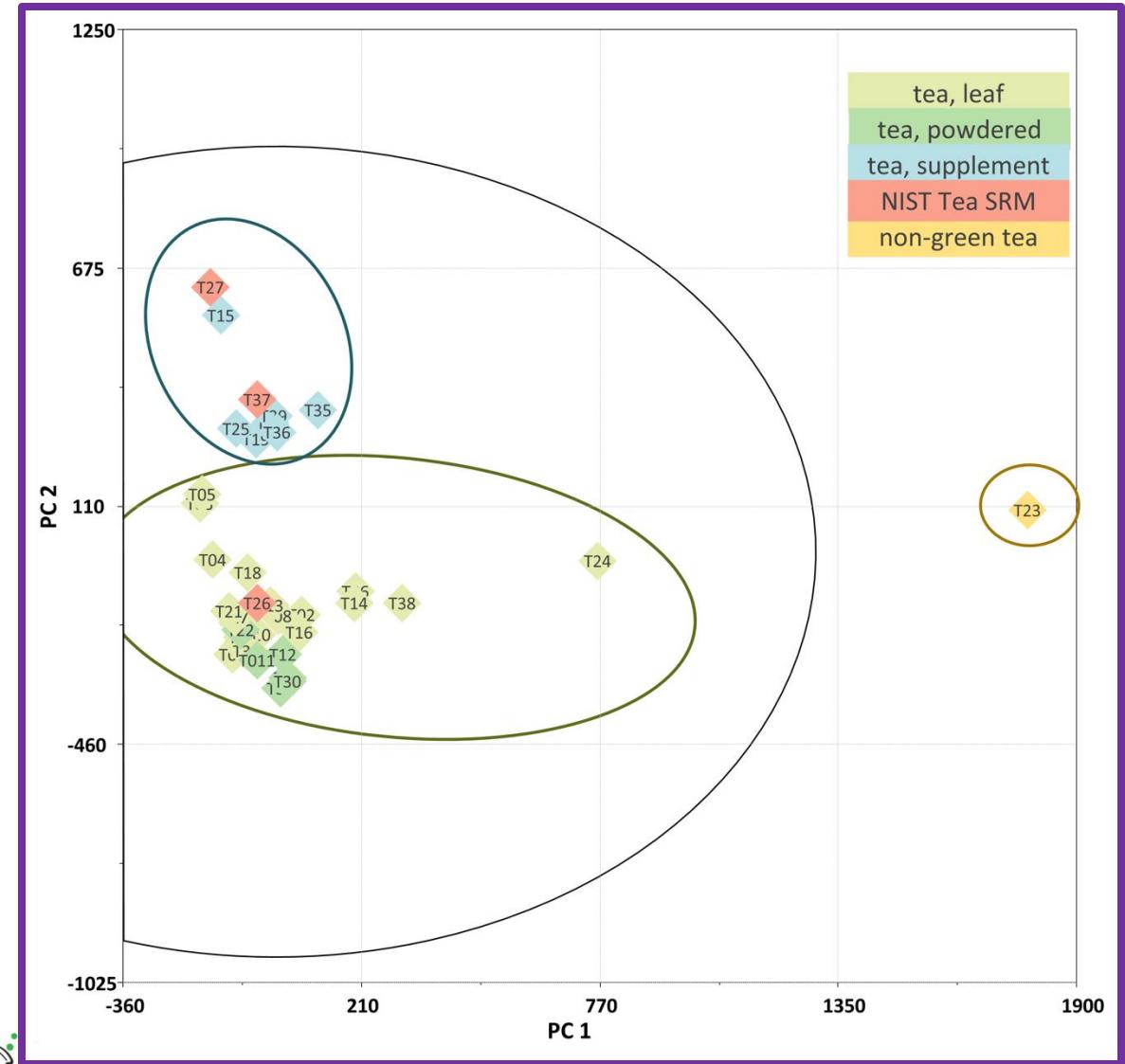


PCA Information

- Scores plots shows the samples plotted using 2 of the principal components.
- Spatial representation of the samples is used to infer similarity/dissimilarity of samples.

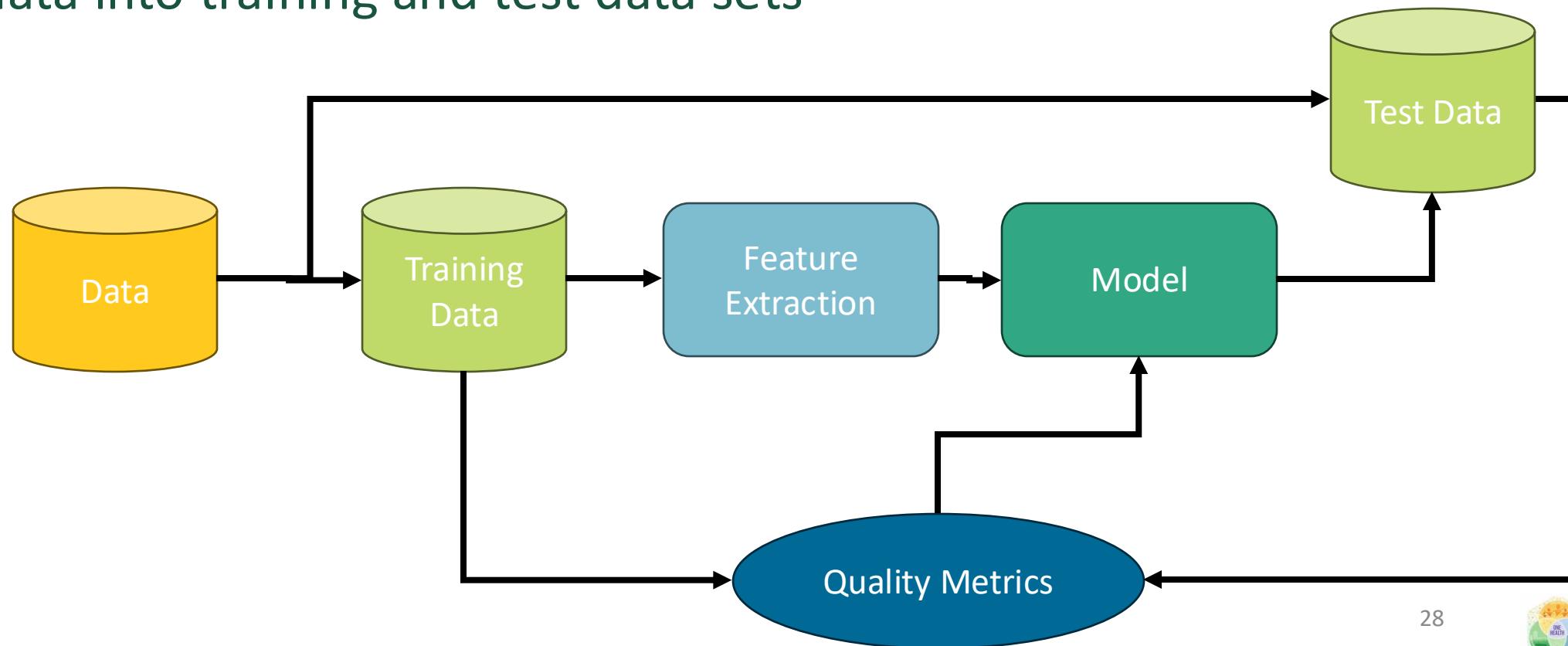


Scores and Loadings



Supervised Approaches (PLS, etc.)

- Use some sort of non-metabolomic (metadata) to group/cluster the samples.
 - Categories
 - Bioactivity
- Split the data into training and test data sets

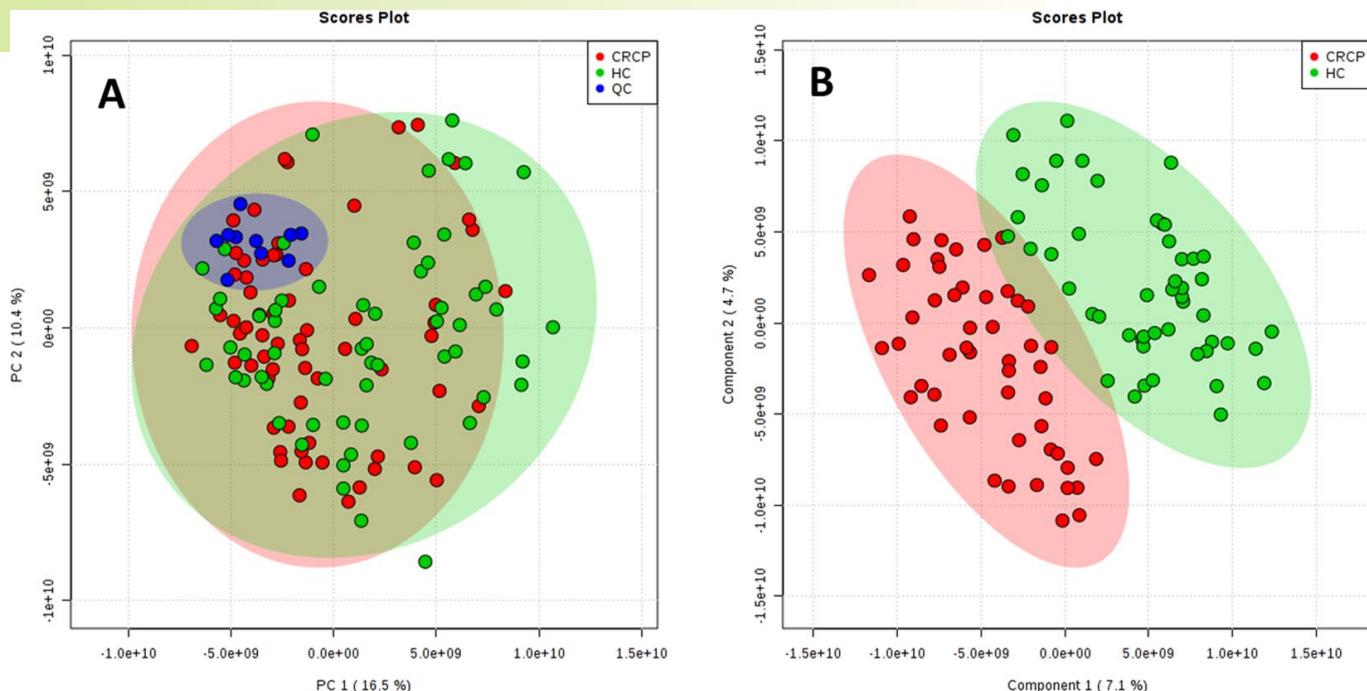
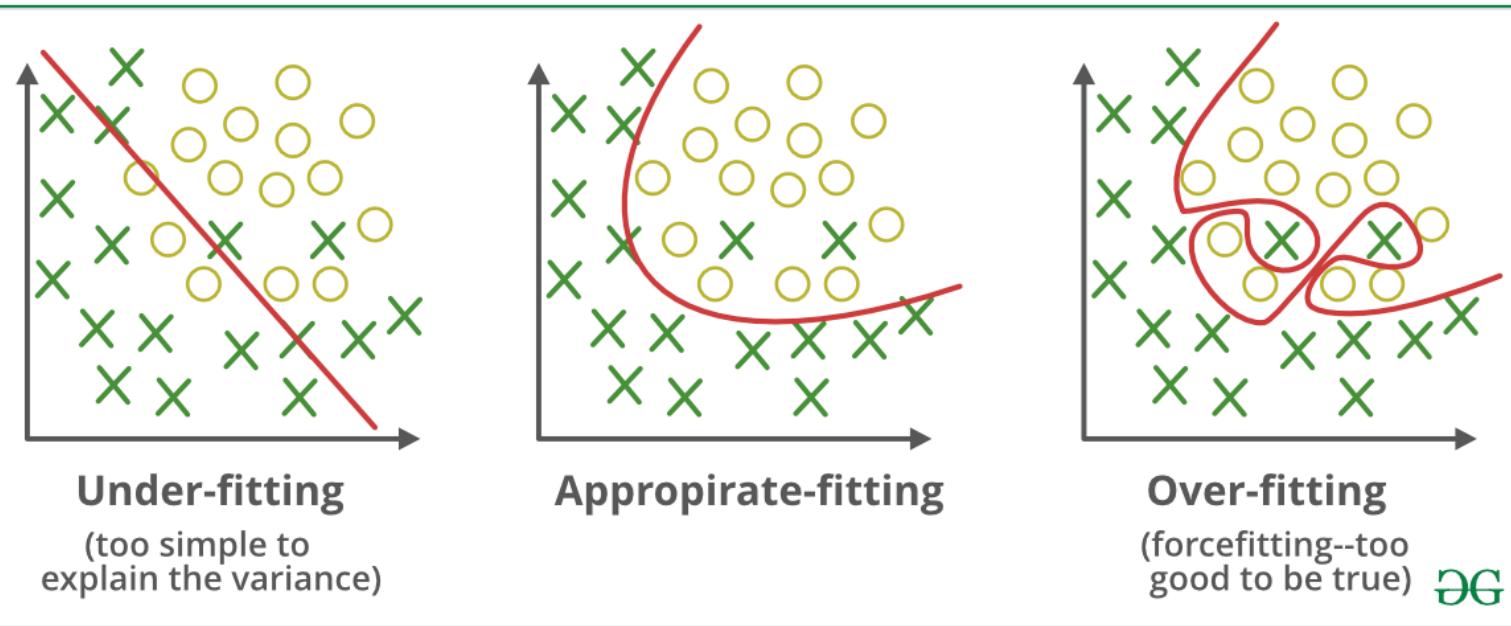


Caution!

- Supervised methods are designed to find differences between groups



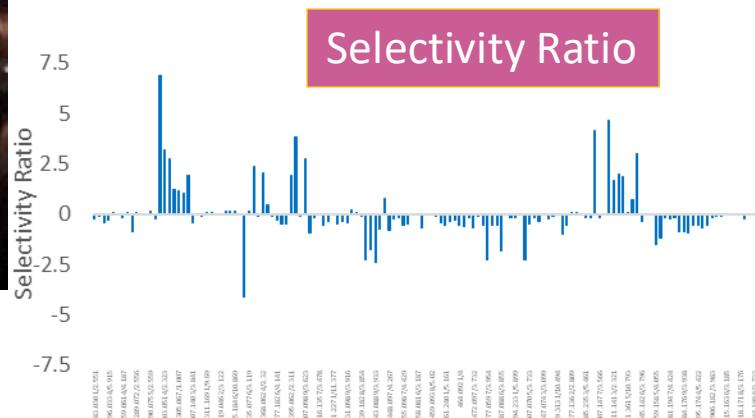
Corporate needs you to find the differences between this picture and this picture.



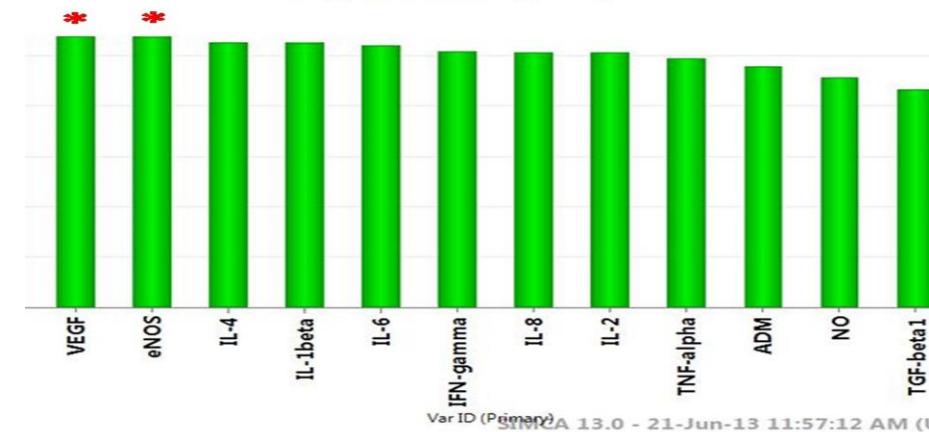
Show Me The Chemistry

Combining bioactivity with metabolomics allows for the modeling of correlative or predictive features (metabolites)

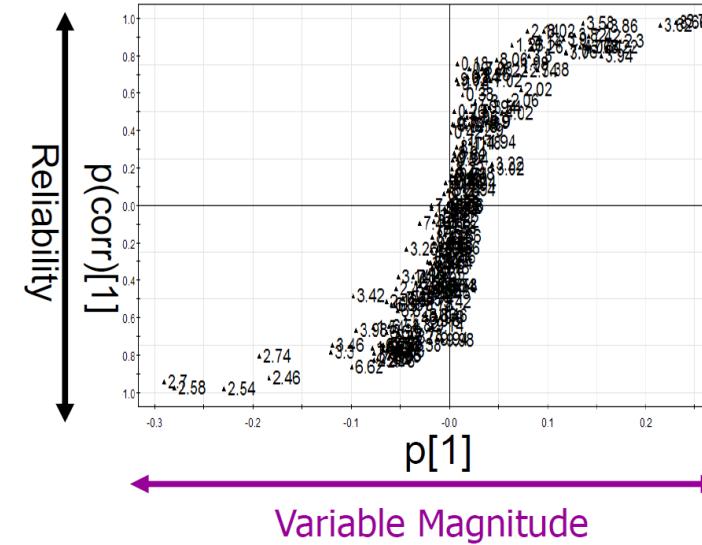
“Biochemometrics”



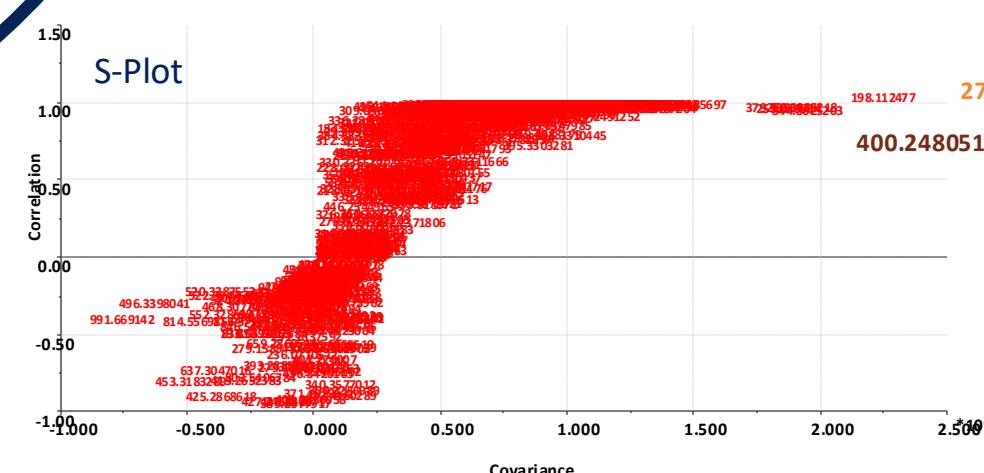
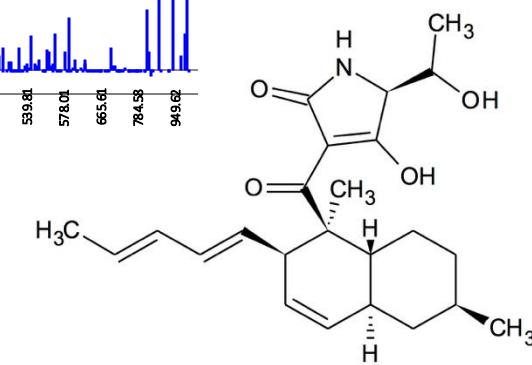
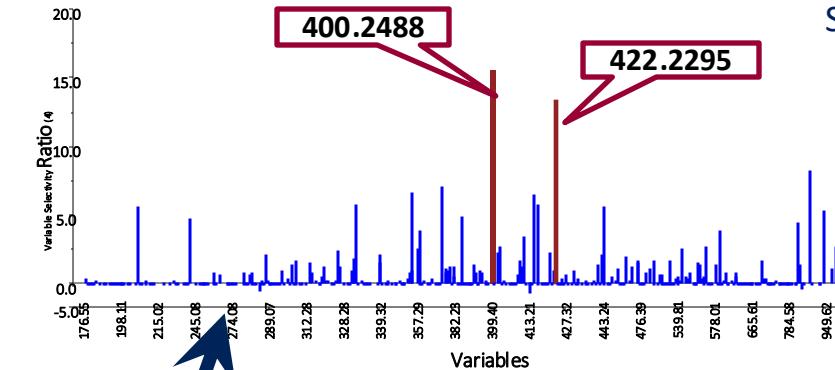
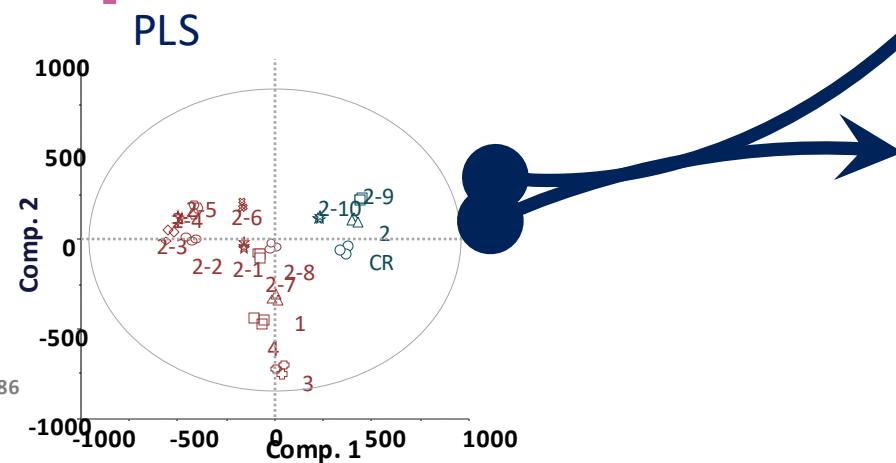
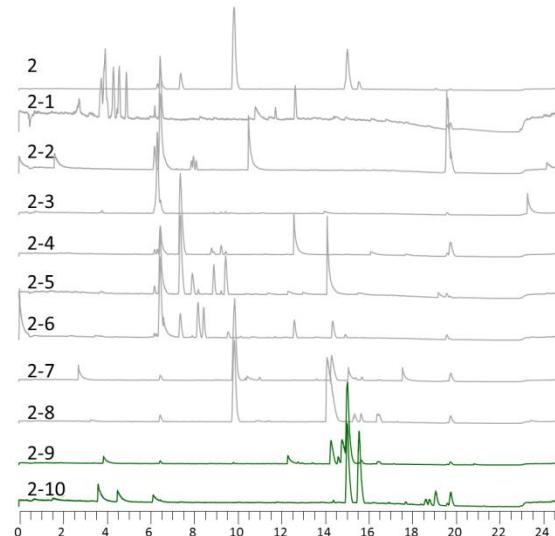
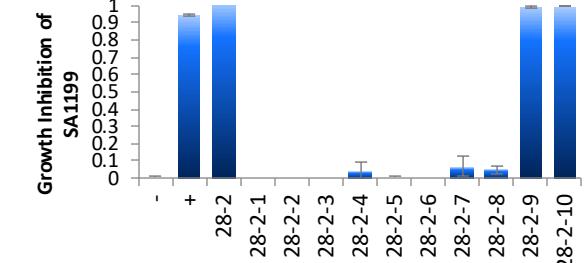
VIP



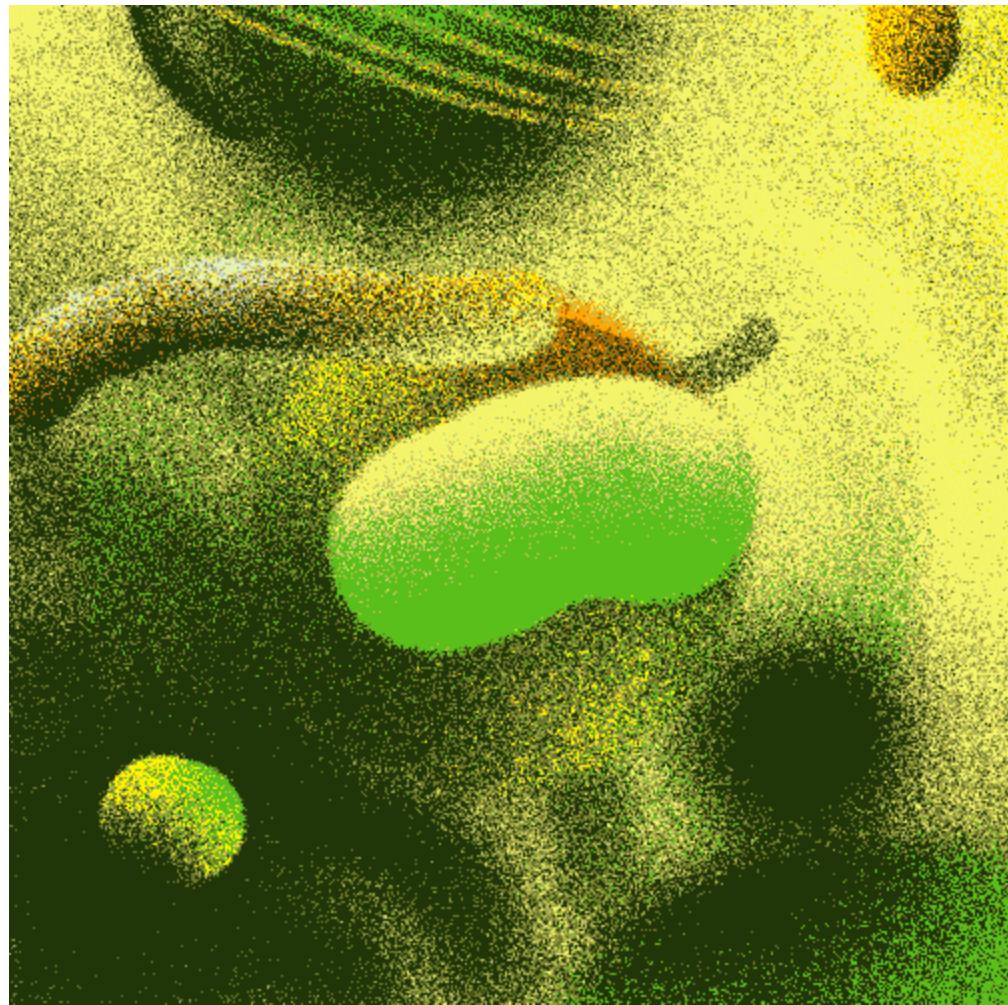
SIMCA 13.0 - 21-Jun-13 11:57:12 AM (1)



Endophytic Fungi Producing Antimicrobials

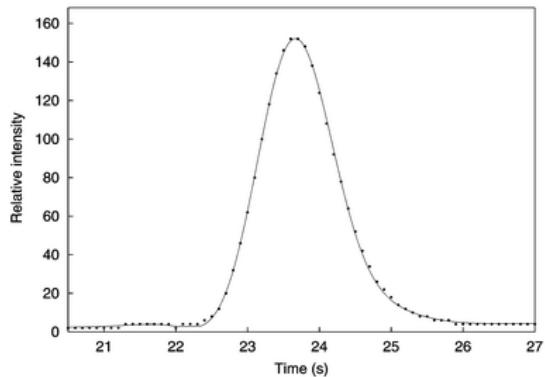


Back to the Chemistry! And Biology!



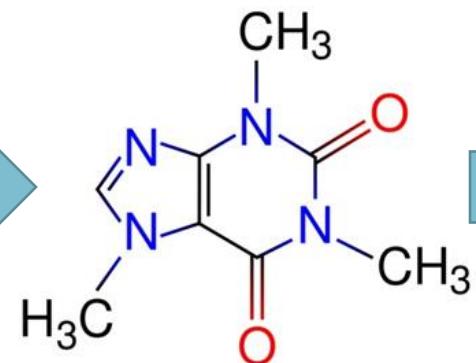
The Chemistry Informs the Biology

Mass and RT

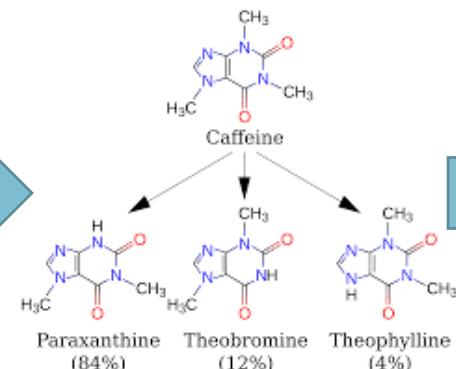


Current
Bottlenecks

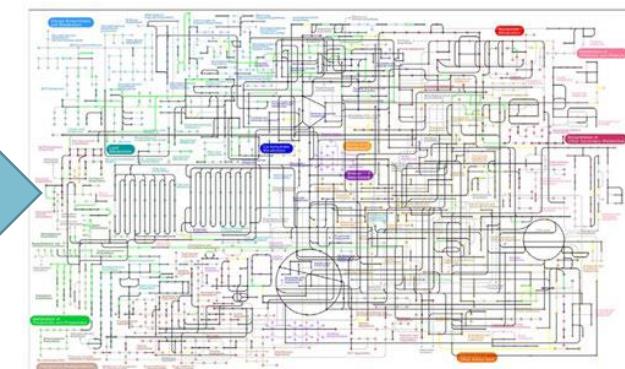
Metabolite
structure/name



Metabolic pathway



Network analysis



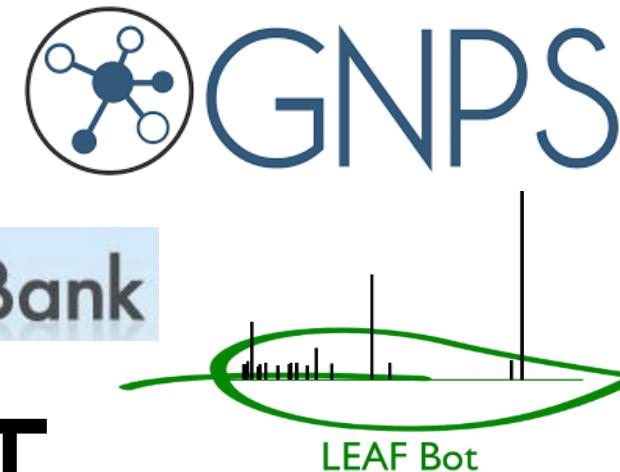
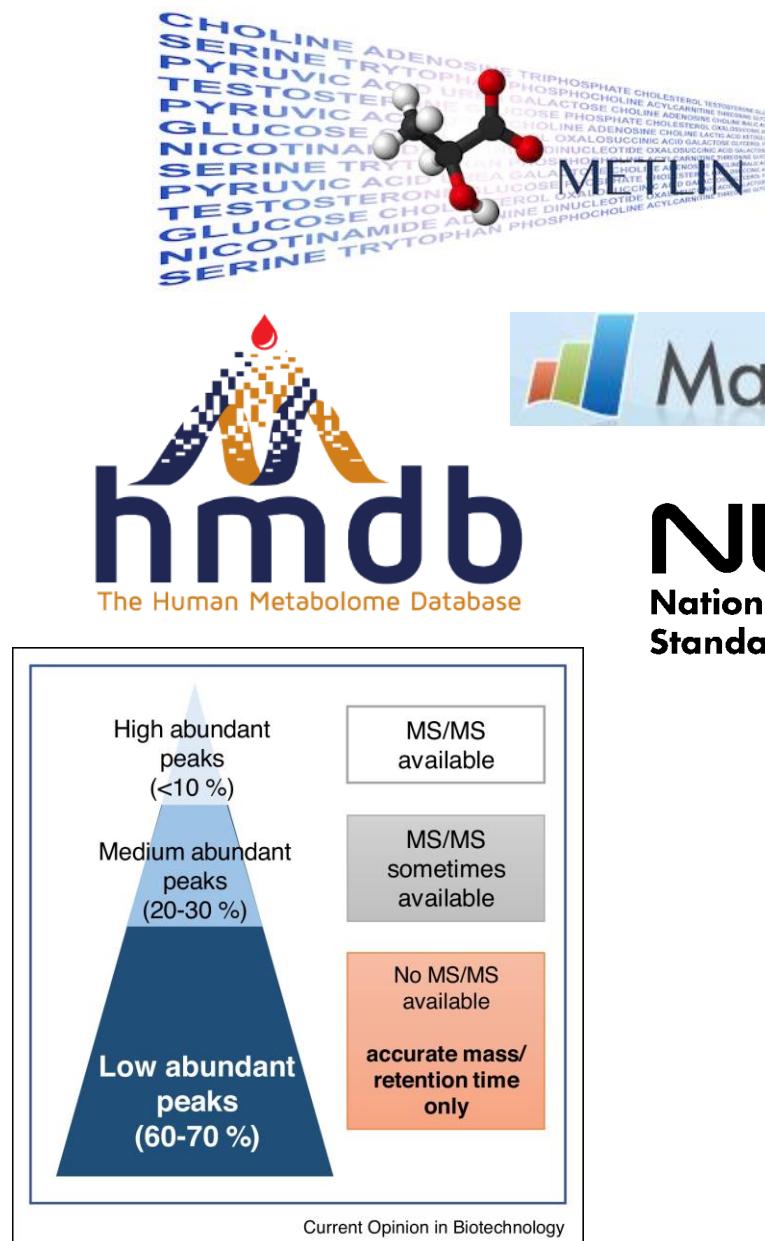
Biological Information

Data Annotation - What is it?



Identification is based on a combination of:

- Retention time in LC
- Accurate mass
- Isotope distribution
- MS/MS product ions of a precursor ion
- Isolation/structural elucidation

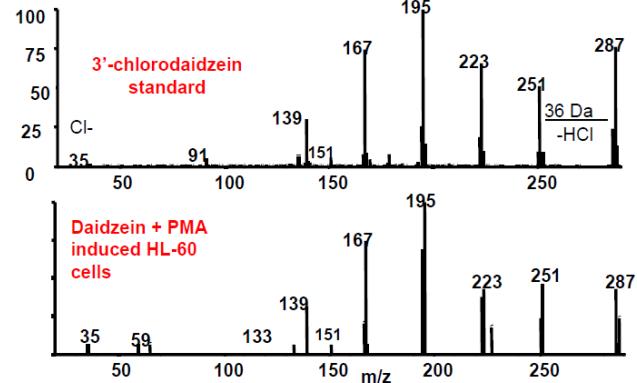


NIST
National Institute of
Standards and Technology

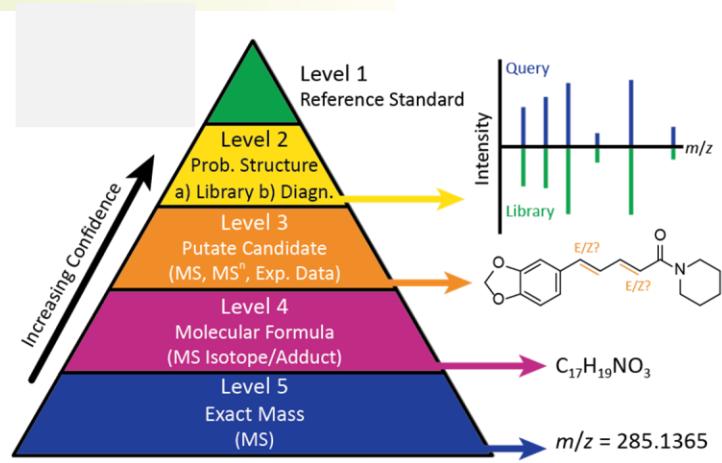
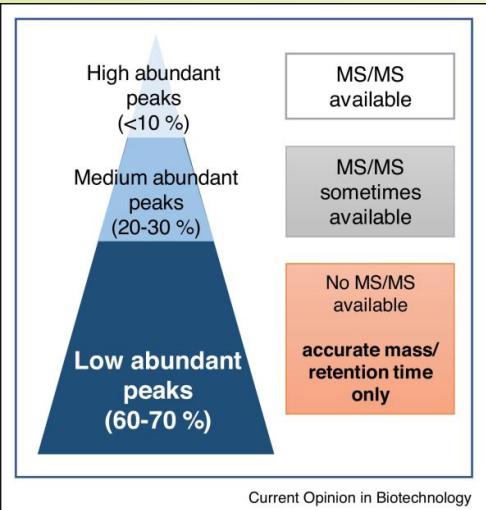


Annotation and Identification

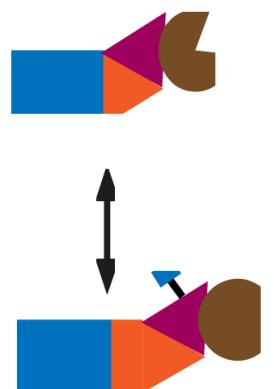
Analytical Standards



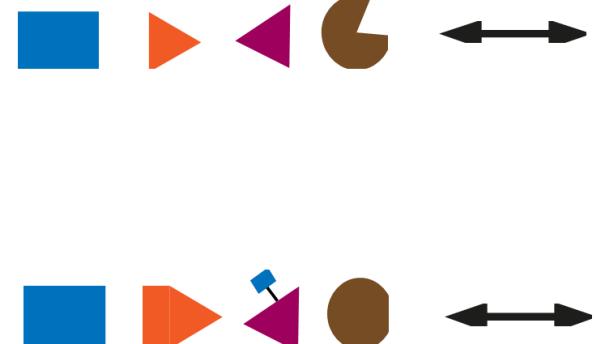
Library Matching



Molecular Structures

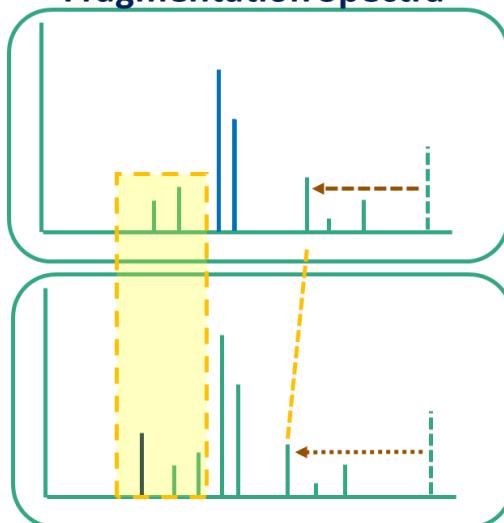


Molecular Fragments



- Best case scenario
- Requires a priori knowledge

Fragmentation Spectra

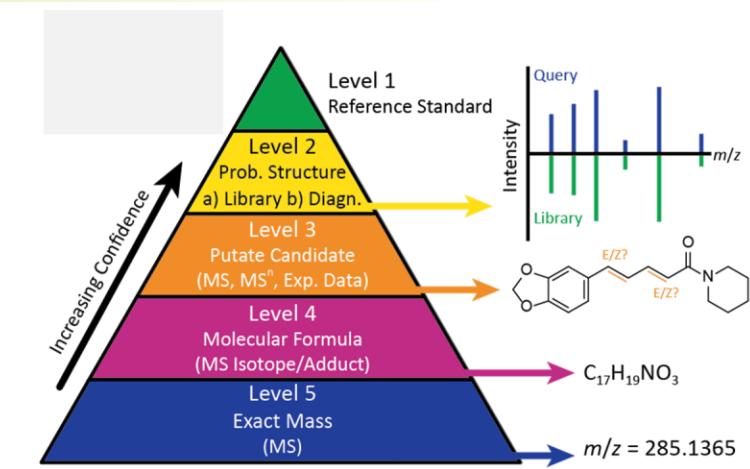
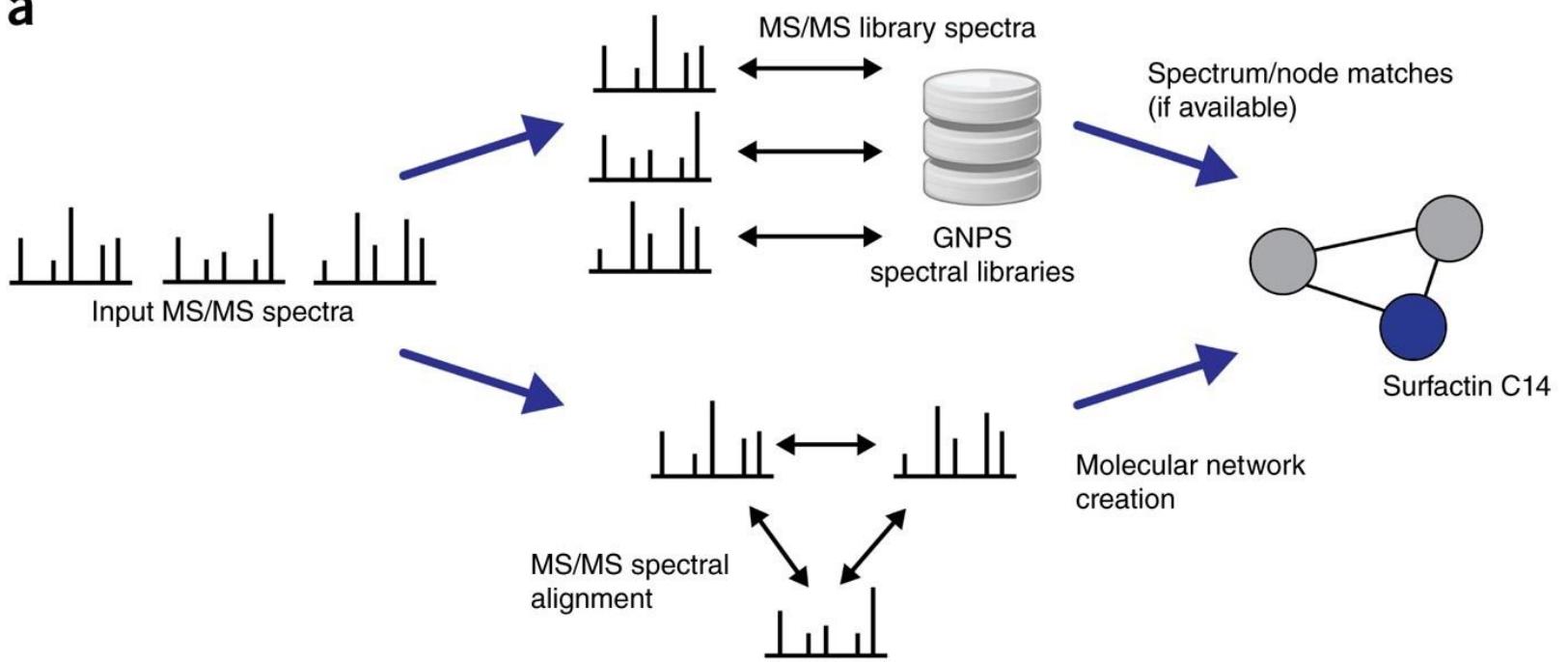


Annotation and Identification

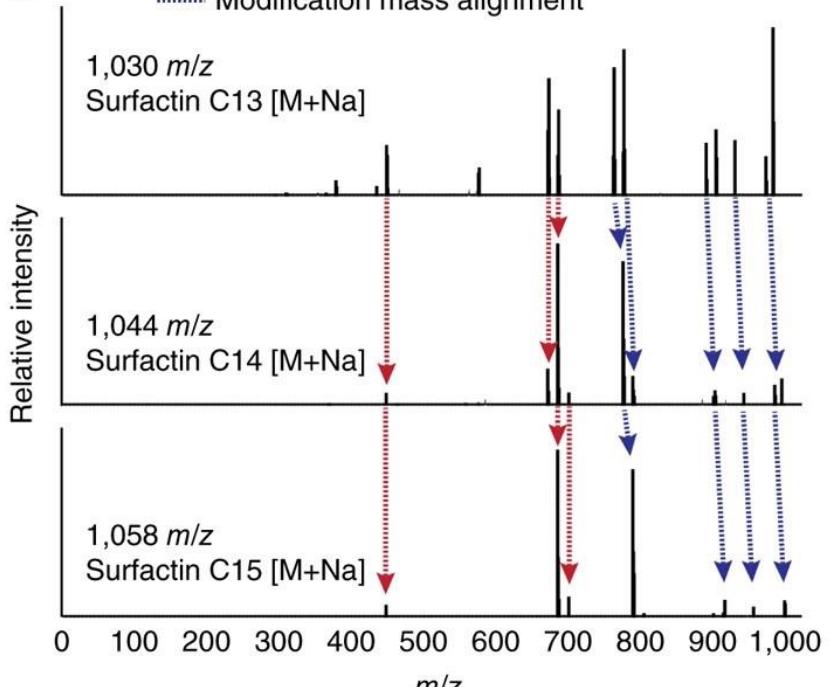


Molecular Networking

a



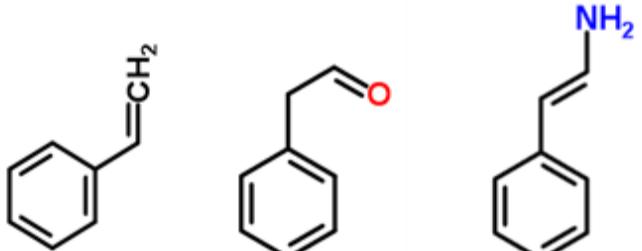
b



Annotation and Identification

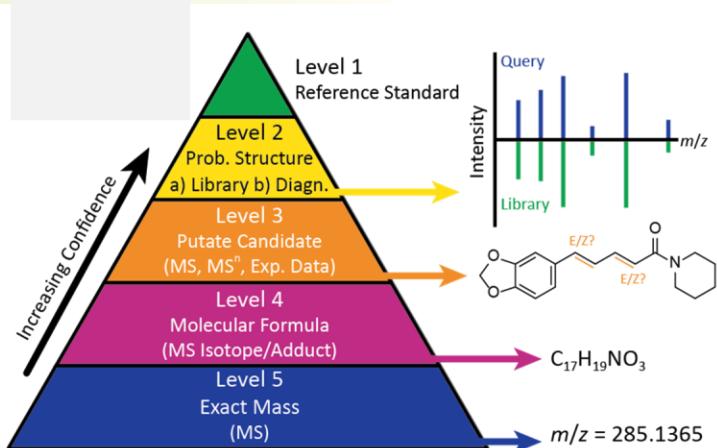
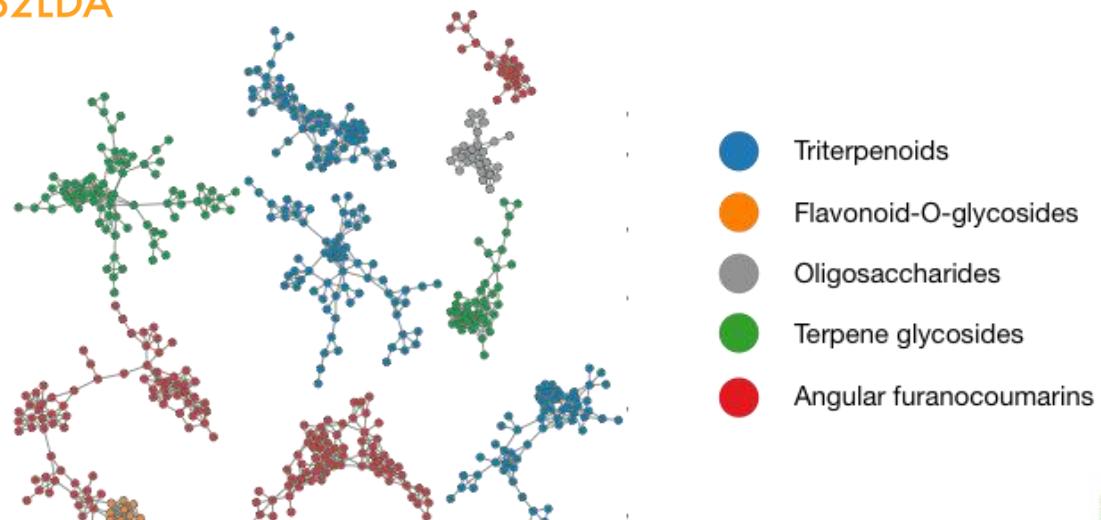
Substructure Hunting

- Chemical classes are related to each other structurally



MOLNET
ENHANCER

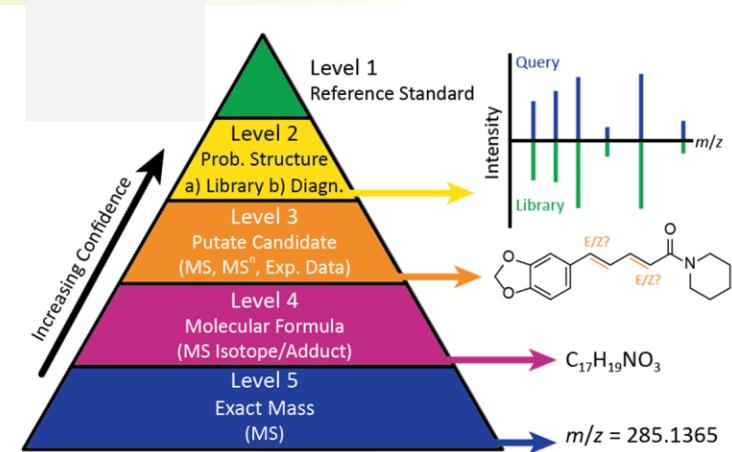
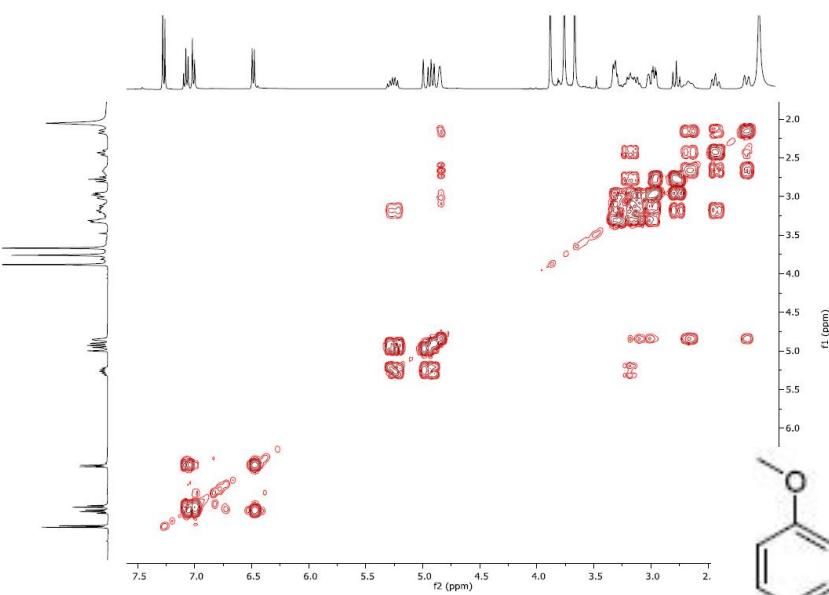
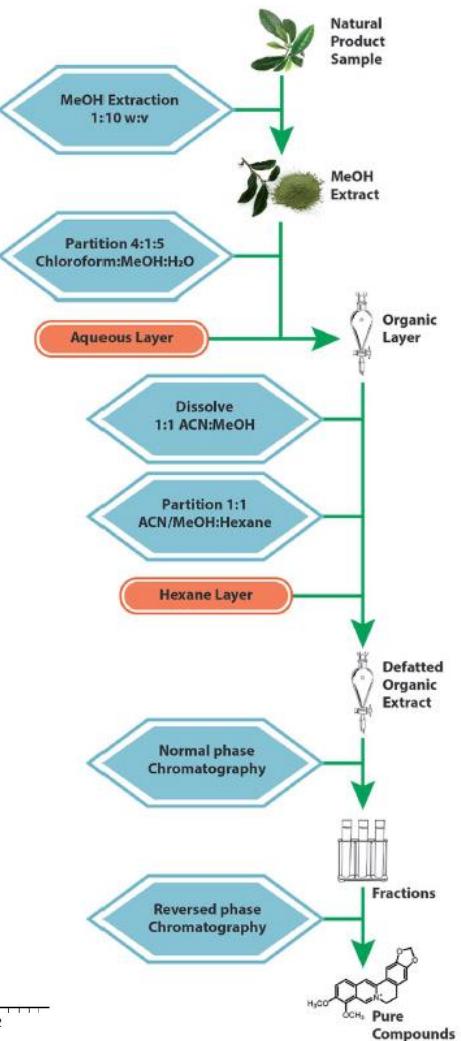
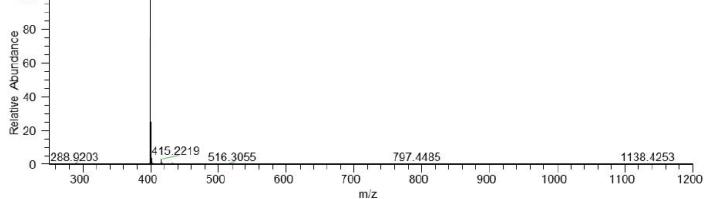
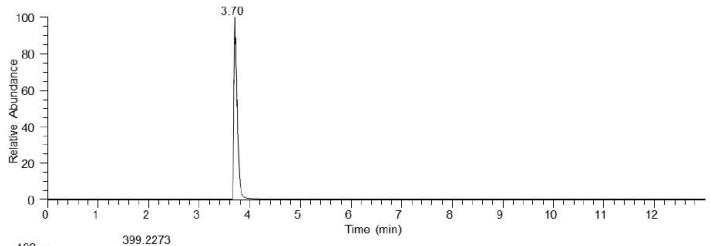
GNPS
MS2LDA



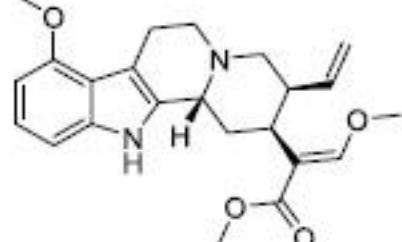
Annotation and Identification

Isolation

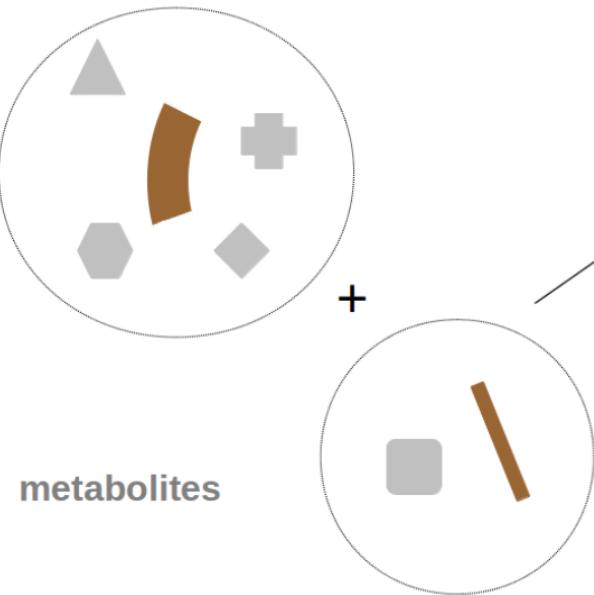
- Rounds of chromatography
- Follow-up with MS and NMR for structure elucidation



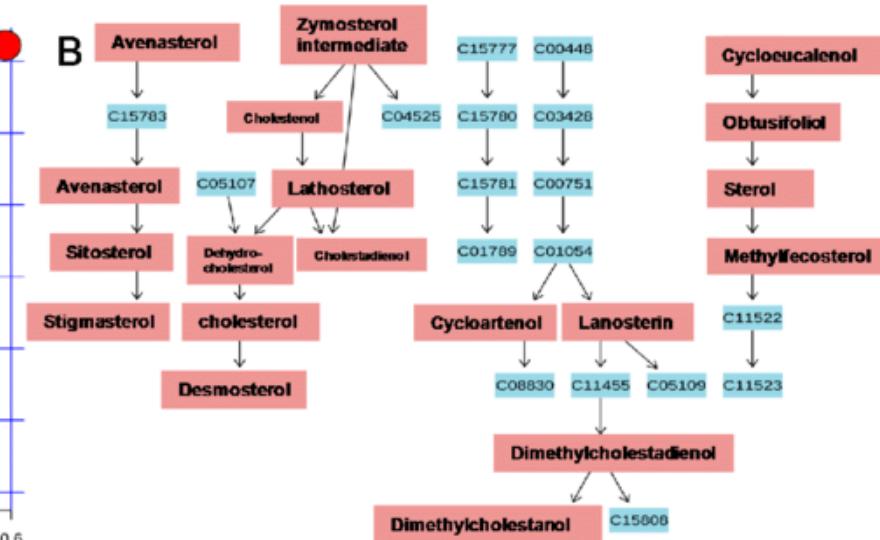
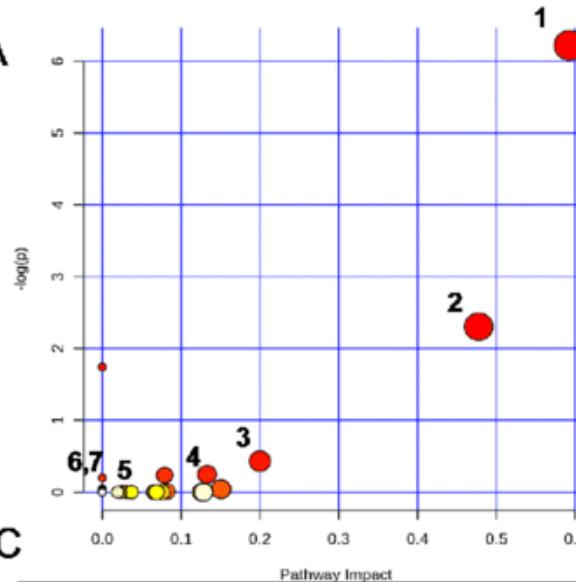
Epiallo-isopaynantheine (7)



Pathway Analysis



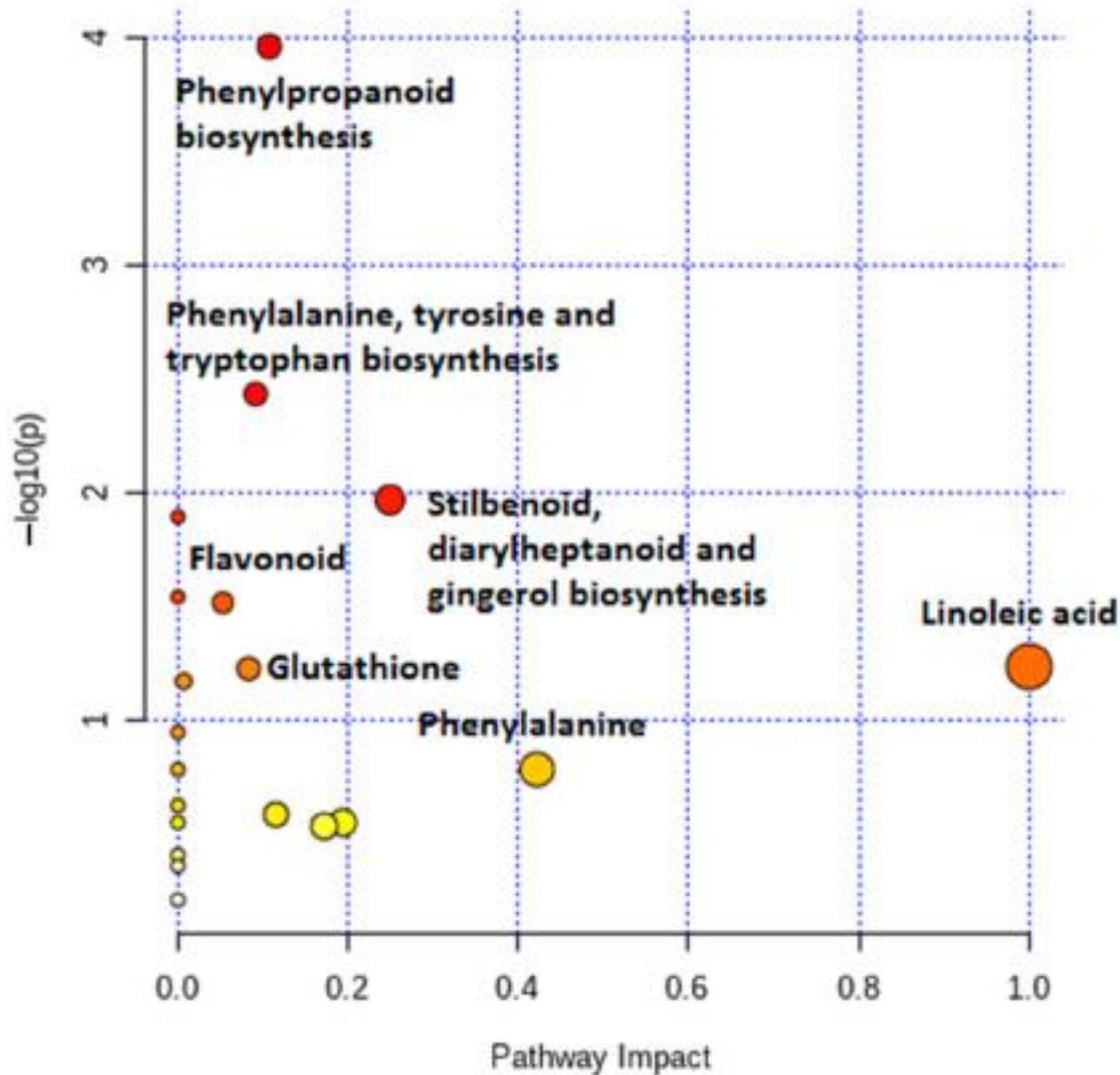
A



Connecting the dots...

#	Pathway name	p-value*	FDR*	p-value#	Q-value#	Impact Score
1	Steroid biosynthesis	1.11E-16	2.85E-15	8.14E-16	1.45E-12	0.5926
2	Diterpenoid biosynthesis	1.16E-06	1.67E-05	0.152	1	0.47745
3	Indole alkaloid biosynthesis/Tryptophan metabolism	1.53E-01	2.33E-06	4.76E-09	3.61E-07	0.2
4	Porphyrin and chlorophyll metabolism	7.25E-17	2.85E-15	0.000952	0.0305	0.13305
5	Carotenoid biosynthesis/Retinol metabolism	4.11E-07	7.06E-06	0.000834	0.0277	0.07913
6	Arachidonic acid metabolism	2.00E-39	1.72E-37	1.09E-19	3.89E-16	0
7	Biosynthesis of unsaturated fatty acids	1.70E-06	2.04E-05	0.0106	0.248	0

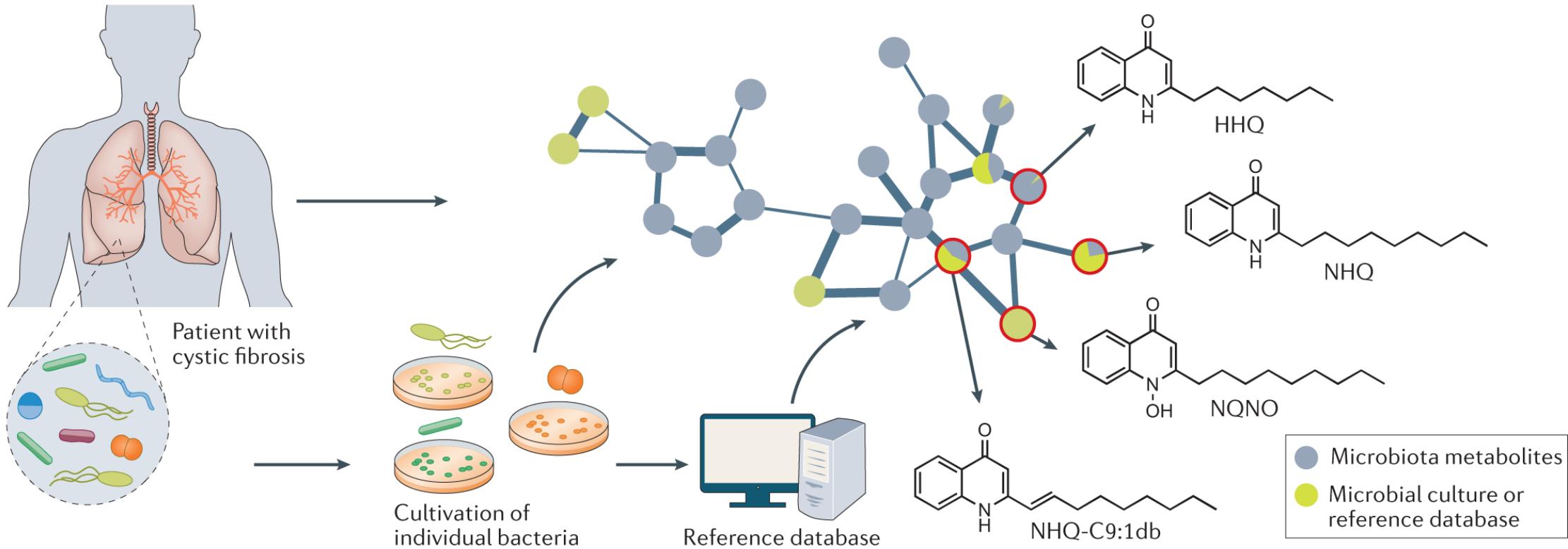
Pathway Analysis



- p-values (y-axis; pathway enrichment analysis)
- pathway impact values (x-axis; pathway topology analysis).

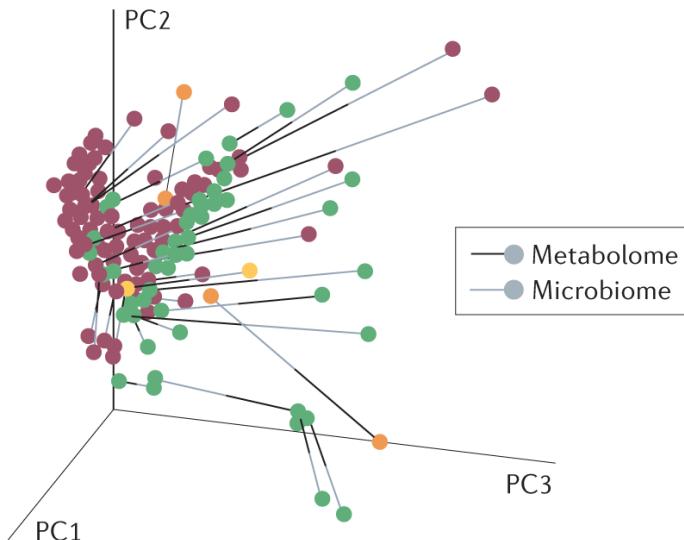
Thus we can determine pathways with high impact: linoleic acid pathway, phenylalanine and stilbenoid biosynthesis, and the pathways with high statistical significance: phenylpropanoid, phenylalanine, tyrosine and tryptophan biosynthesis.

Integration

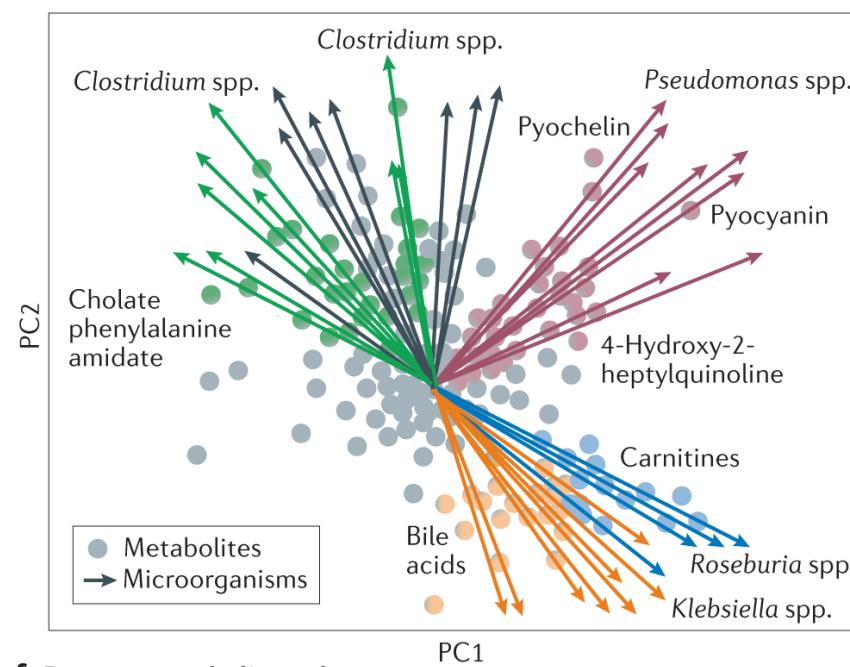


Integration

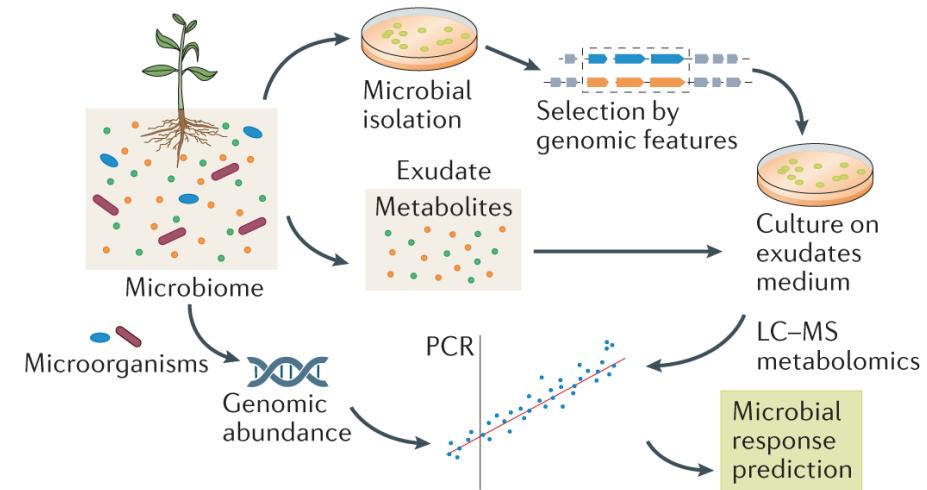
b Procrustes analysis



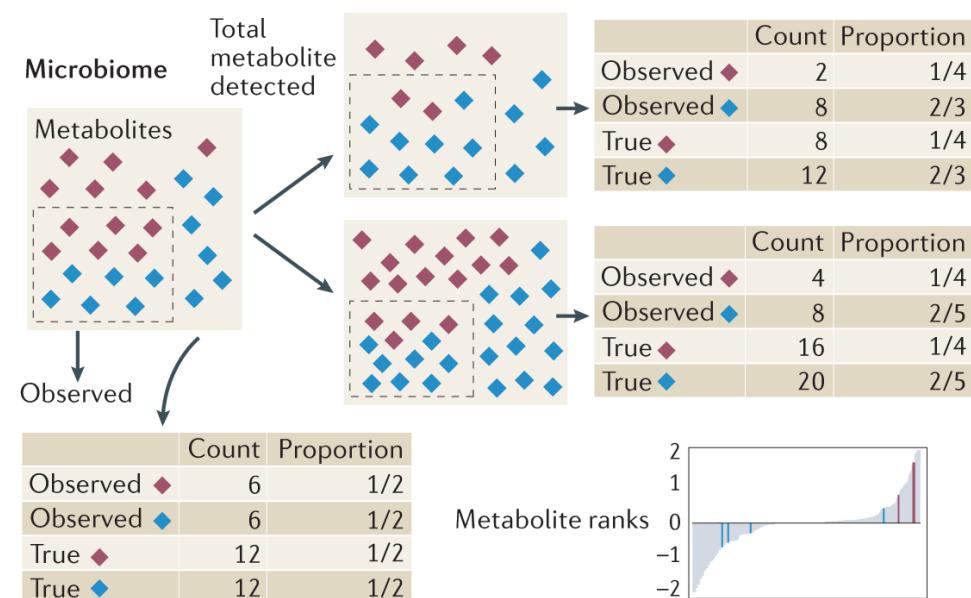
d mmvec



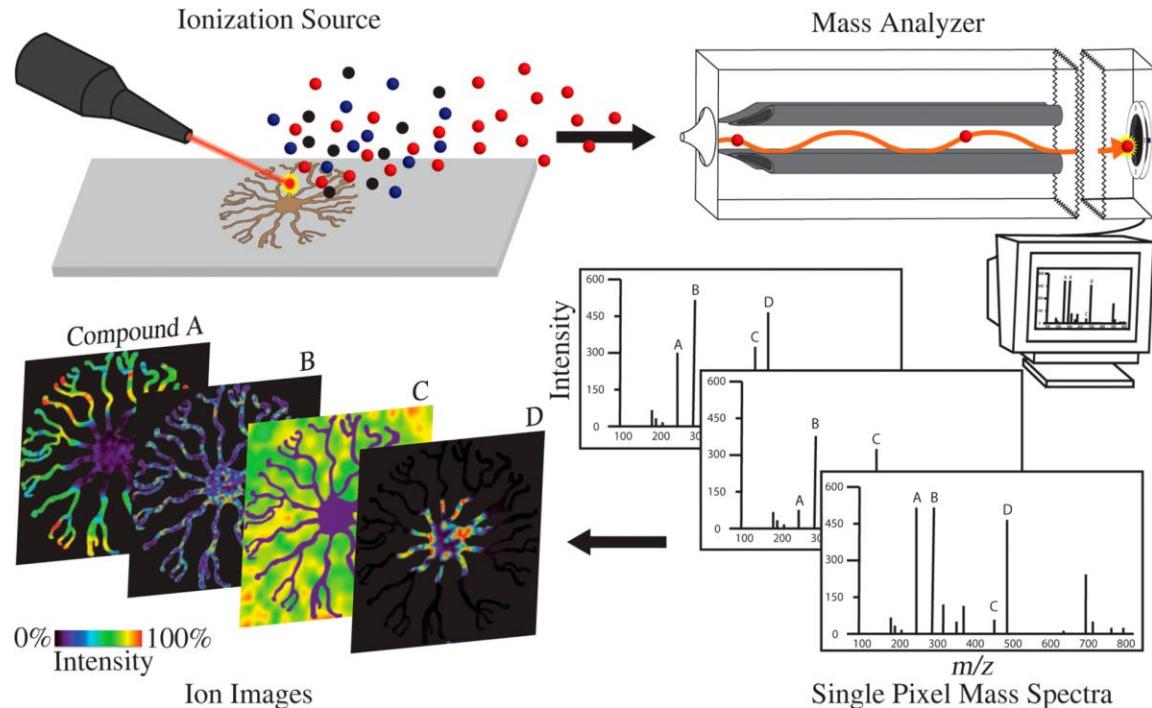
c PCR



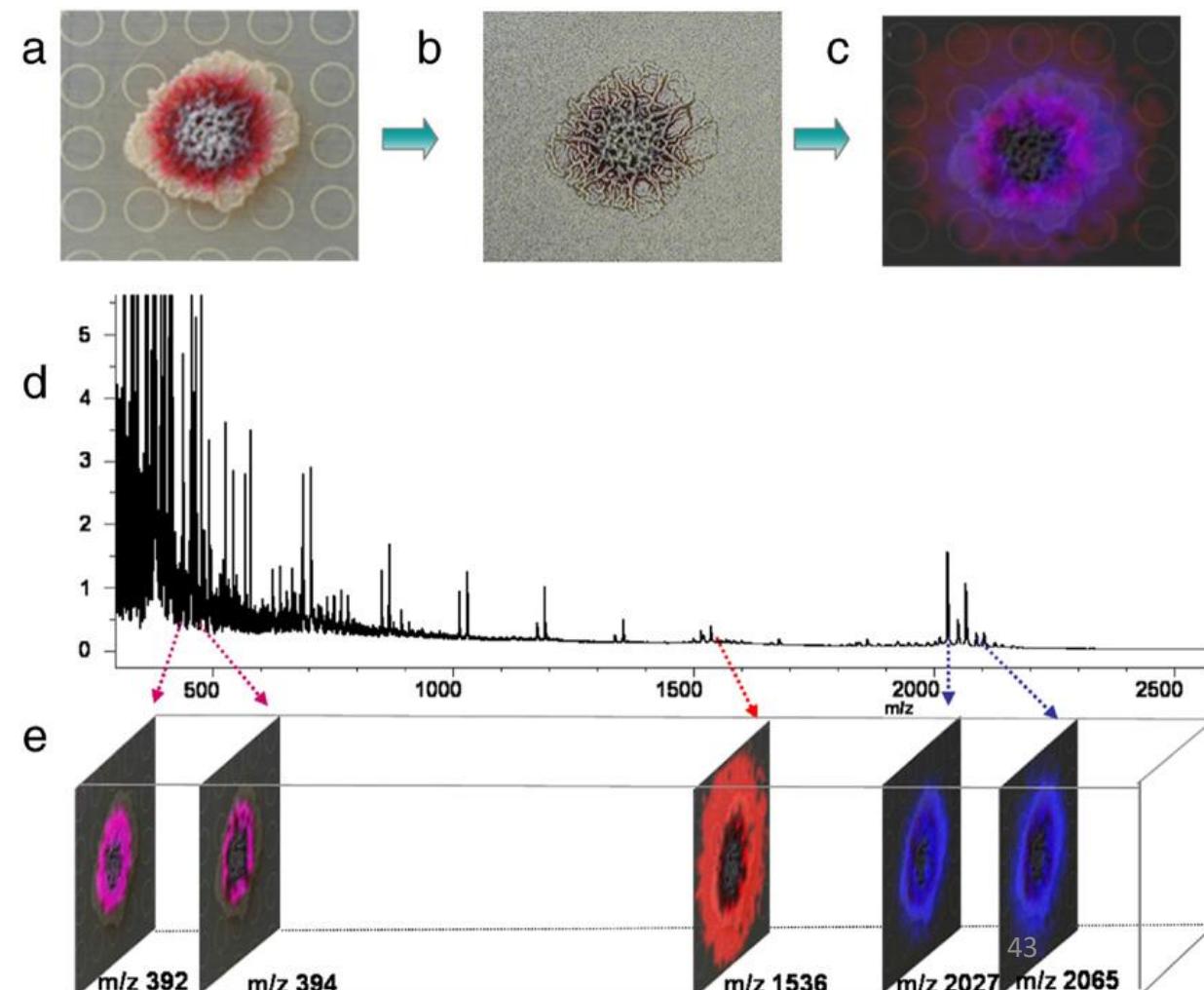
e Songbird



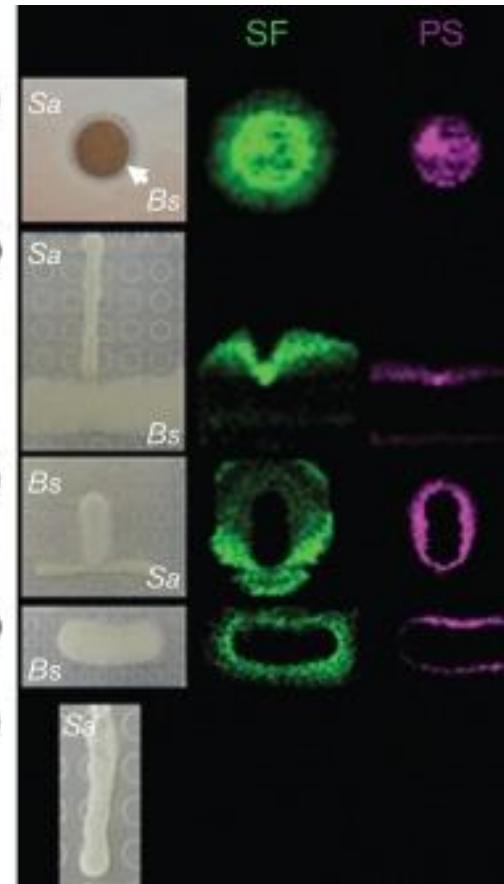
Taking it to 2D (and 3D)



Mass spectrometry imaging (IMS or MSI)



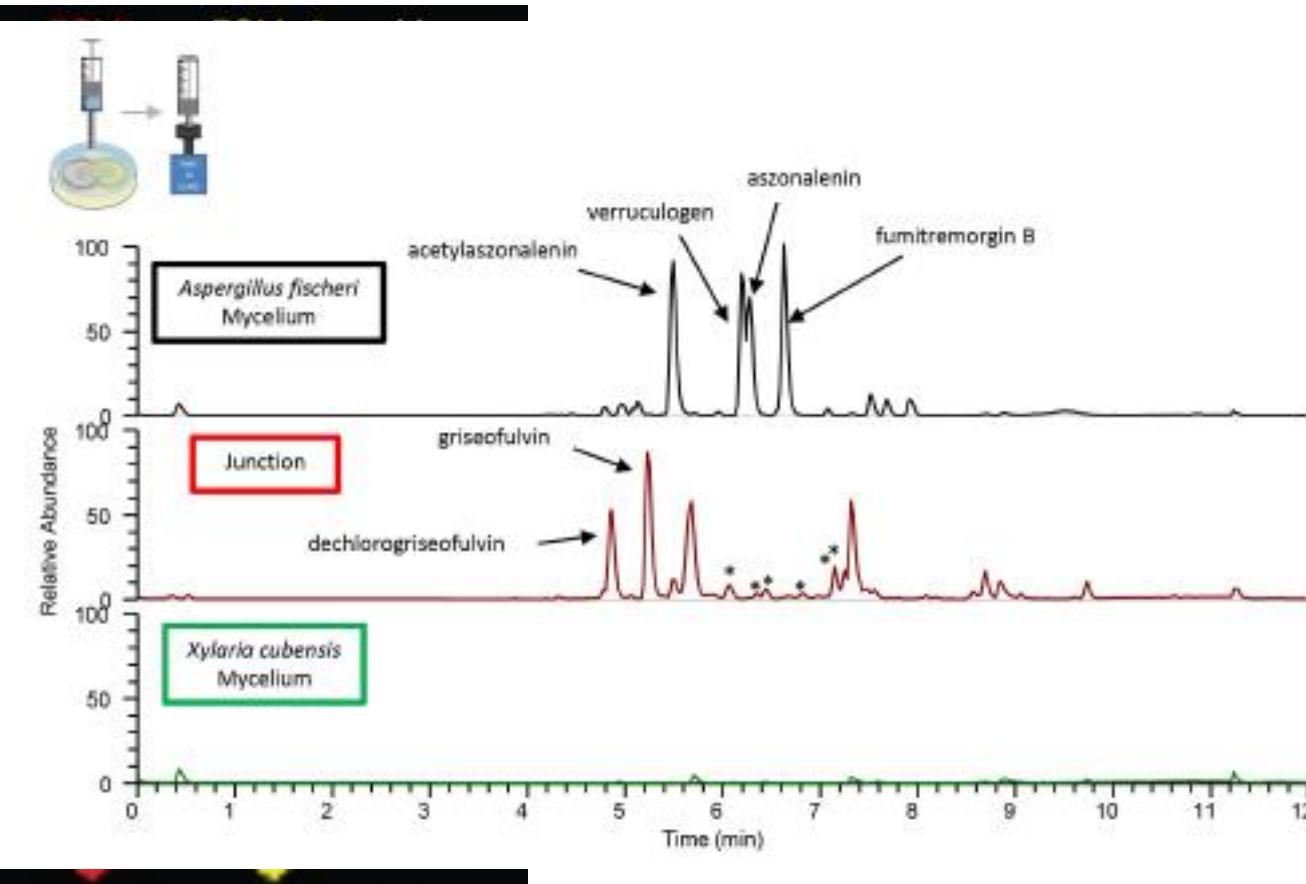
Mixed Microbial Martial Arts (MMMA)



Surfactin Sa toxin production
Plipastatin

interaction between *B. subtilis* (Bs) and *S. aureus* (Sa).

10.1099/mic.0.048736-0



10.1039/c9np00019d
10.1021/ac100954p
10.1021/acs.jnatprod.5b00268



Upregulation of defensive compounds at the junction between *A. fischeri* and *X. cubensis* co-culture

Metabolomics: What Can It Do For You?

Provide high-flux readout of biochemical interactions in the system

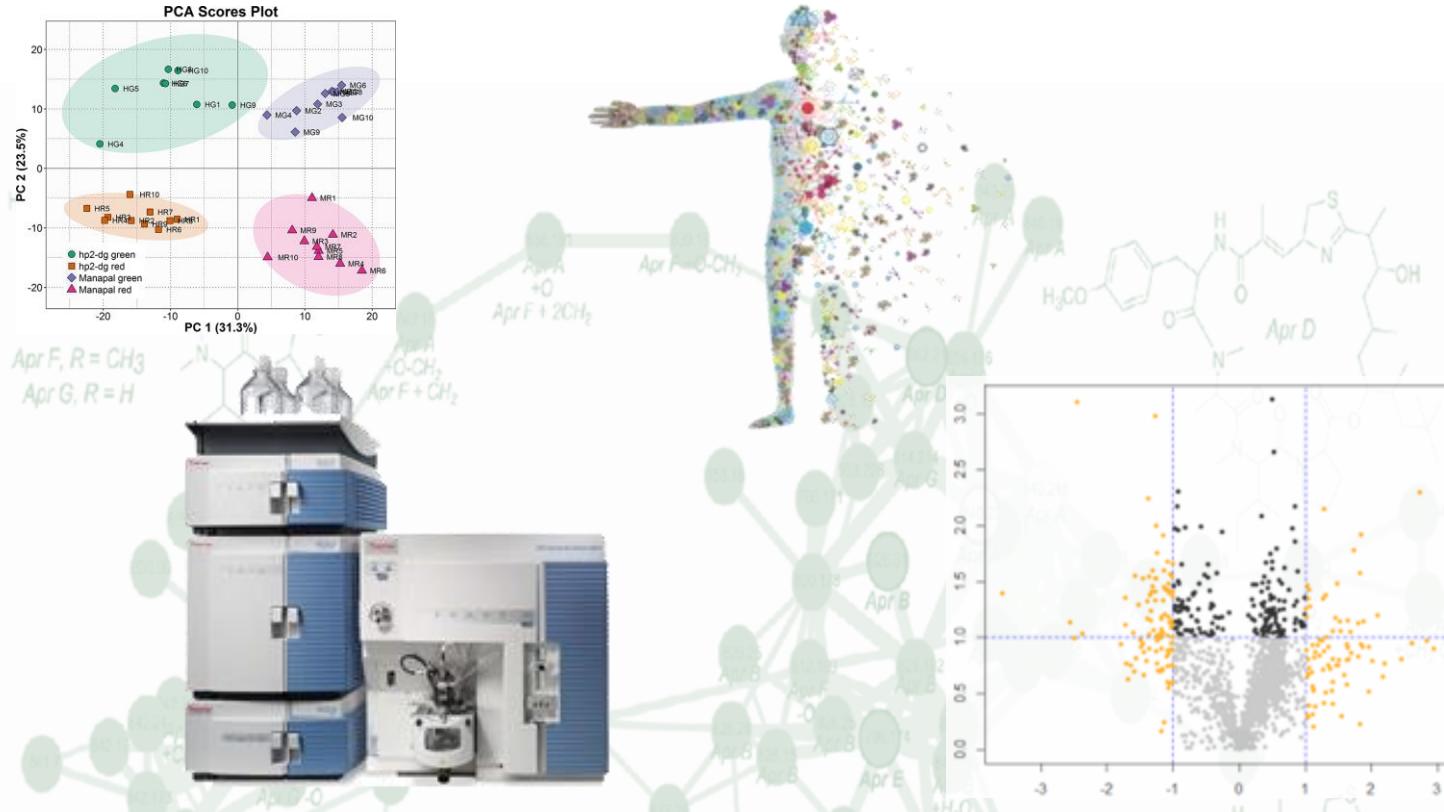
Describe phenotypic differences based upon broad chemical differences

Identify putative biomarkers and bioactive compounds from complex mixtures

Analyze pathways up or down regulated without genomics or proteomics or transcriptomics data

Integrate with other –omics techniques to provide more powerful insight into cellular or organismal systems

Come join me for more fun!!!



Spring 2025 VBSC 555 Principles of Metabolomics



Tu/Th 1:35-2:50pm

Instructor: Dr. Josh Kellogg,
jjk6146@psu.edu

Metabolomics is one of the newer -omics technologies available to researchers and has great potential to impact many aspects of research. This course explores how metabolomics studies are conducted, from experimental design to data analysis. We will draw from current examples in biomedical, ecological, and agricultural research, as well as perform our own data analyses to gain a deeper understanding of the potential for metabolomics in research.



Thank you!

Questions?

jjk6146@psu.edu
325 Huck Life Sciences Bldg.
www.sites.psu.edu/kellogglab
 @kelloggngroup.bsky.social
 @jk kellogg 916
 @kelloggnatprodgrp

