Metabolomics Common Fund Program

Arthur Castle Ph.D. Program Officer (NIDDK)
Co-Coordinator Metabolomics Common Fund Program

National Institutes of Health
Office of Strategic Coordination–The Common Fund

Metabolomics

Definition:
Study of large number of metabolites in living system
Dynamic with thousands to millions of diverse chemicals
Involves Endogenous metabolism, signaling, microbiome, exposures
Represents current state of the biological system

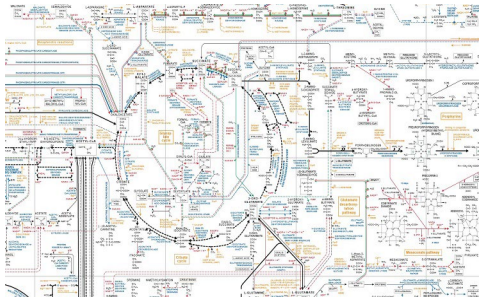
Exposures

Food

Drugs

Toxins

Endogenous Metabolism



Structural (Phospholipids, etc.)
Signaling (Eicosanoids, etc.)

Transformations/Symbiosis

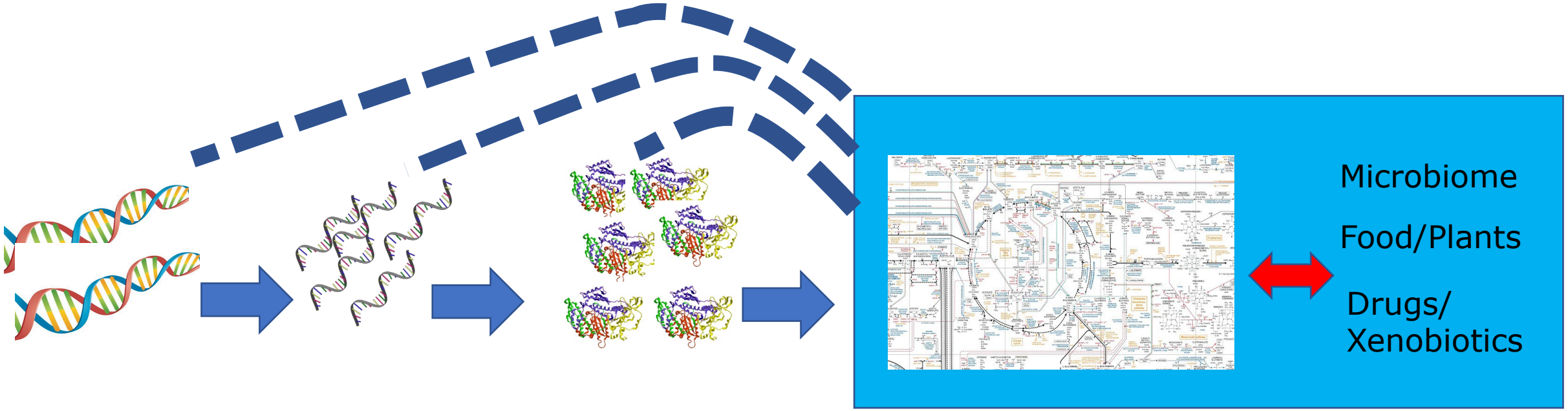
Microbiome

Metabolic by products

Excrements

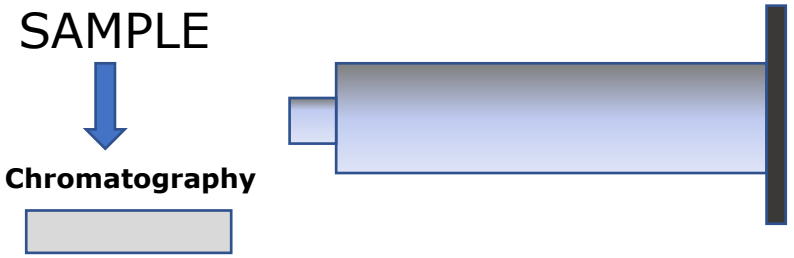
Metabolomics

Function:
Combined with genomics and proteomics for system biology
Metabolites interact at all levels
Often needed to understand other omics



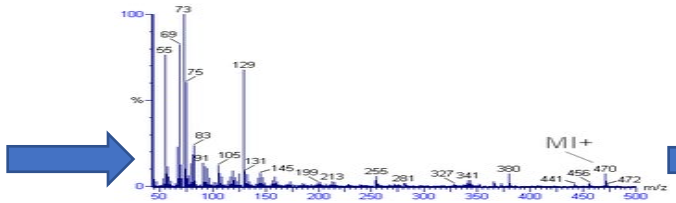
Metabolomics

Collection of data:
Use of mass spectrometry and NMR to collect information
Known standards for hundreds of metabolites and quantification
Many thousands of unknown metabolites with some properties identified
(MW, collision cross section, basic formula, functional groups, fragments)



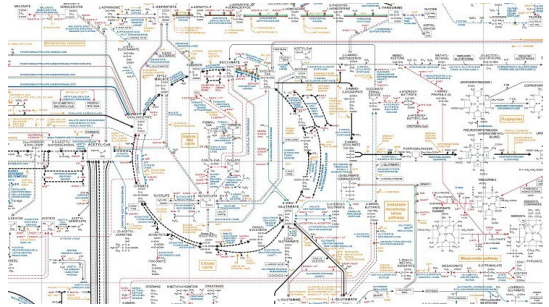
Instrumentation and Techniques
(LC/GC, MS, MS/MS, IMS, NMR Etc.)

Spectrometry Data



Detection, Identification, Analysis Tools

Metabolites / Knowledge



Why Common Fund?

What we learned from discussions and workshops

Metabolomics has potential to advance health related research in many fields

There is much more demand for technology than trained investigators and resources available

The analytical chemistry needed for detection, identification and quantifications is more complex and diverse than genomics and proteomics

Translating metabolite identification into a biological context of pathways requires sophisticated software and expertise

Sharing of experimental data is limited and not easily accessible.

Goals of the Program

Stage I 2012-2017

Increase national capacity (6 regional centers)
Provide training and mentoring (collaborative projects, career awards, courses)
Promote data sharing (establish a repository)
Support technology and standards development (R01s, R03s, Contracts)

Stage II 2018-2022

Develop the repository into a national resource
Improve metabolite identification
Overcome analysis and interpretation hurdles
Promote best practices

Building Capacity and Training/Technology (Phase I)

Funded 6 Comprehensive Centers to Provide Services

> 2,000 Services and P&F projects supported
83 Collaborative Supplements to RPGs

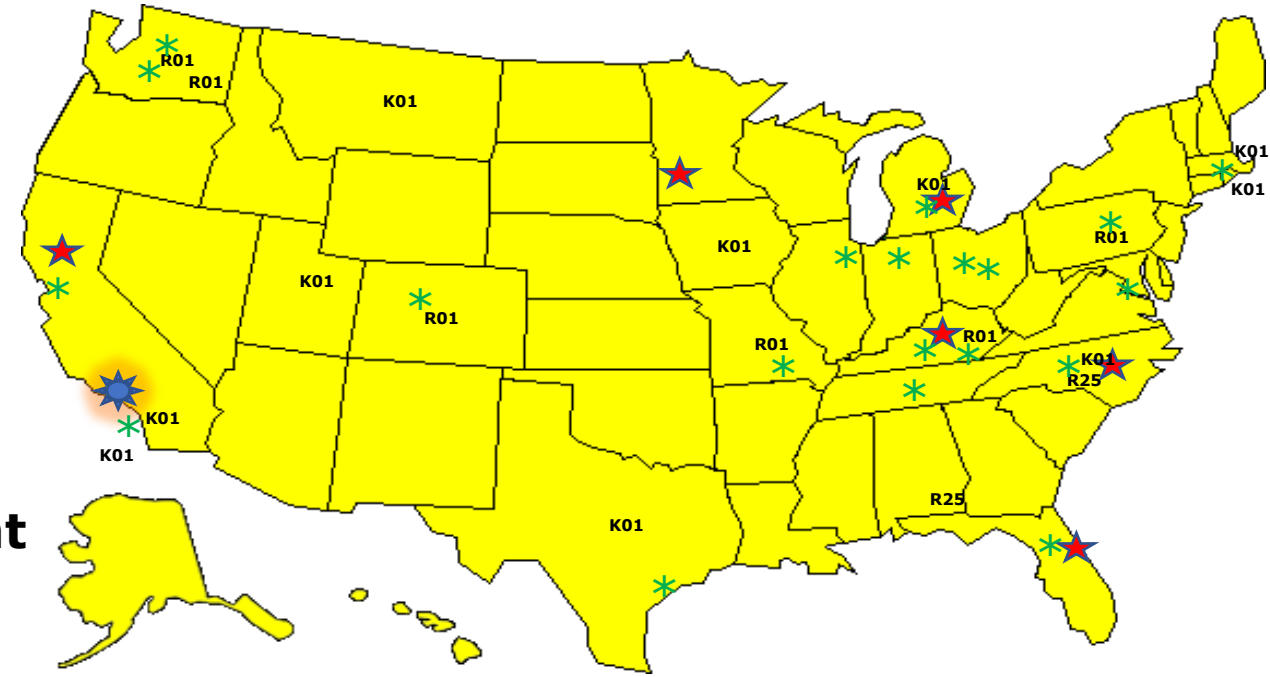
Univ. of Michigan (C. Burant)
Univ. of California Davis (O. Fiehn)
Univ. of North Carolina (S. Sumner)
Univ. of Florida (A. Edison, R. Yost)
Univ. of Kentucky (R. Higashi, T. Fan, A. Lane, H. Moseley)
Mayo Clinic (S. Nair)

Data Repository

UCSD (S. Subramaniam)

Training, Technology and Career Development

10	**Career awards K01**
2	**Educational courses R25**
6	**Technology R01s**
22	**Pilot R03s** *



https://commonfund.nih.gov/metabolomics

Building Capacity and Training/Technology (Phase I Outcomes)

821 Publications directly associate with the centers

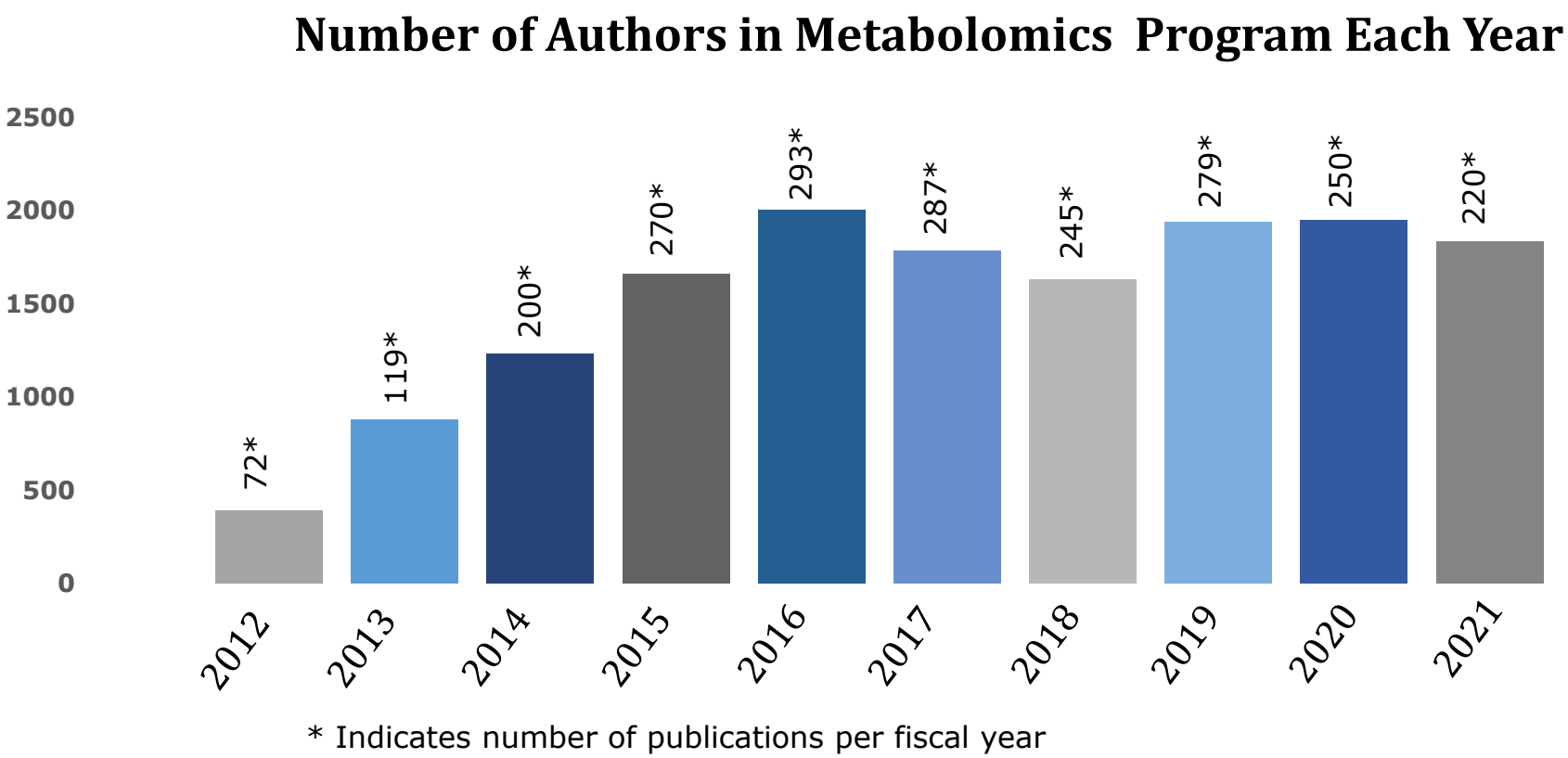
102 Publications associated with K01 career awards (All in science, 2 R35s & 2 R01s awarded)

165 Publications associated with R01s (technology development)

93 Publications associated with R03s (analysis/technique pilots)

2 Publications associated with R25 (educational courses) (196 attended in person course)

Building Community of Researchers

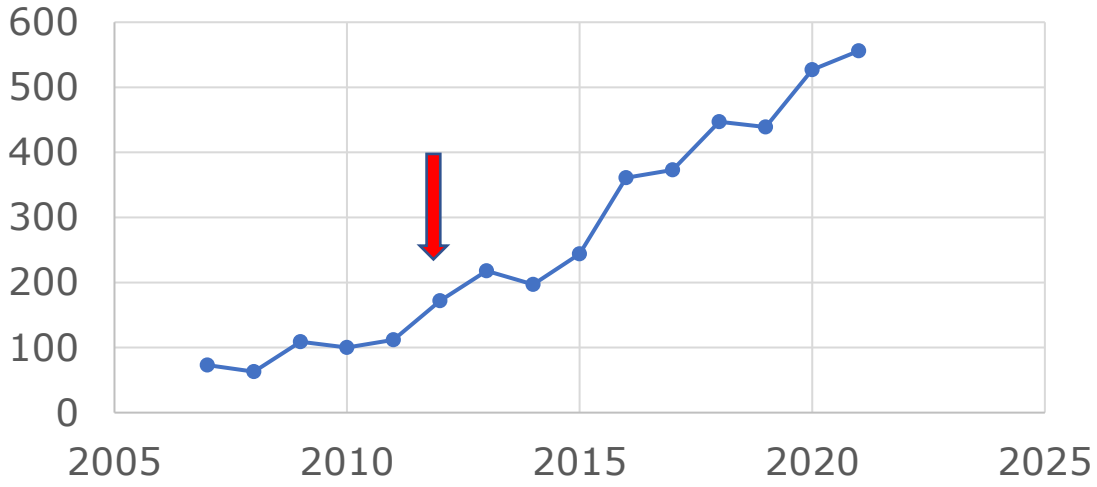


* Indicates number of publications per fiscal year

Methods: Appl IDs were obtained from QVR for all Common Fund grants associated with the metabolomics program. Publications per year associated with these Appl IDs for indicated years were obtained from iSearch. Unique author numbers were generated in Cystoscape.

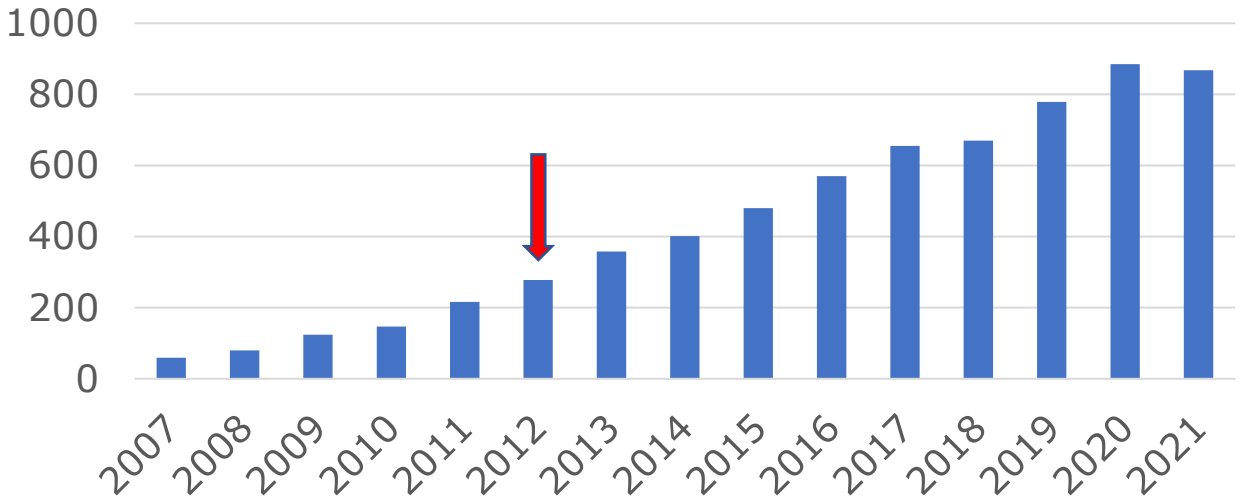
Metabolomic Outcomes for NIH

Number of NIH Grant Awards that Mention Metabolomics by Fiscal Year



Methods: iSearch-Grants was used to search title, abstract, and specific aims for the search term: metabolomic*. Data was restricted to Type 1 applications and NIH as the funding agency. Date of search: 1/6/2022.

Number of NIH-Funded Publications that Mention Metabolomics by Publication Year



Methods: iSearch-Publications was used to search title, abstract, and MeSH keywords for the search term: metabolomic*. Publications were restricted to NIH as the funding agency. Date of search: 1/6/2022.

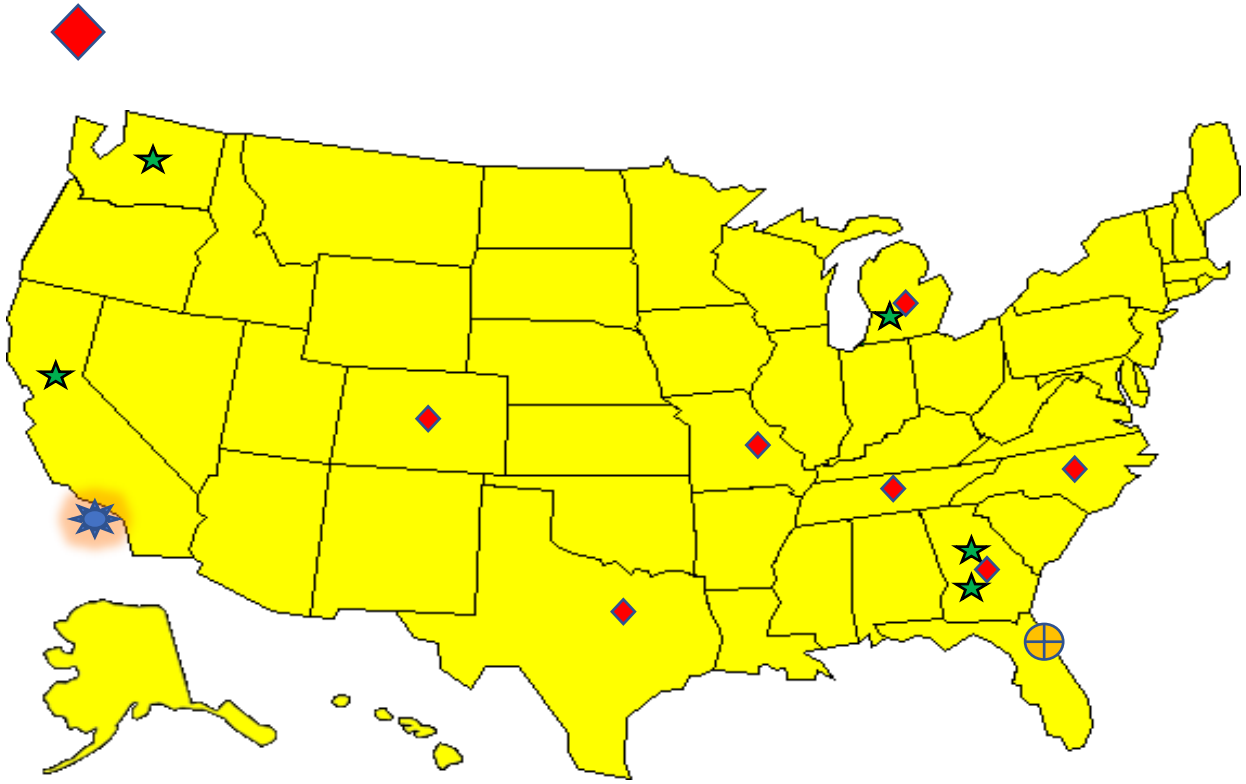
Improving Compound Identifications and Data Analysis and Interpretation (phase II)

7 analysis and interpretation tool development projects
Univ. of North Carolina Charlotte (X. Du)
Univ. of Michigan (A. Karnovsky, G. Michalidis)
Univ. of Colorado Denver (K. Kechris-Mays, D. Ghosh)
The Jackson Laboratory (S. Li, G. Suizdak)
Washington Univ. (G. Patti)
Univ. of Texas MD Anderson (J. Weinstein, R. Akbani, B. Broom)
Vanderbilt Univ. (J. Young)

5 Compound Identification Cores
Univ. of Georgia (A. Edison)
Univ. of California at Davis (O. Fiehn)
Emory Univ. (D. Jones, S. Li, G. Miller, E. Morgan)
Univ. Michigan (A. Nesvizhskii, C. Evans)

Data Repository Univ. of San Diego (S. Subramaniam)

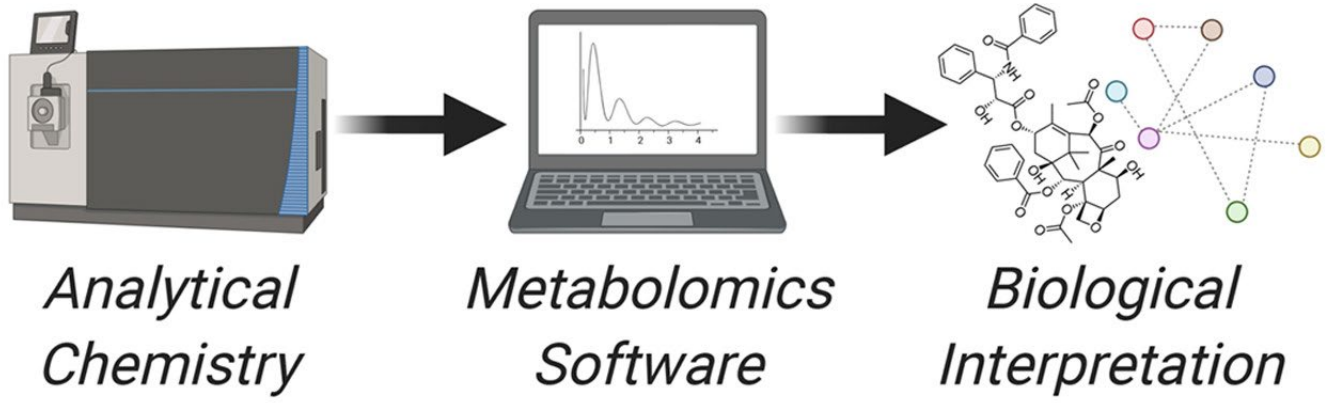
Coordinating center Univ. of Florida (R. Yost, M. Conlon)



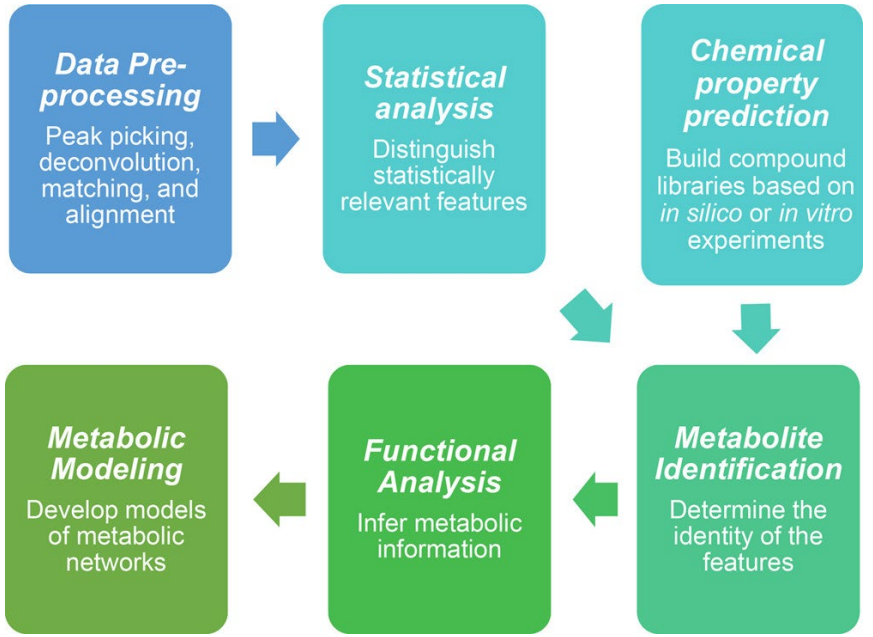
https://commonfund.nih.gov/metabolomics

Improving Data Analysis and interpretation tools

7 projects, 11 tools covering most of the pipeline from data to knowledge

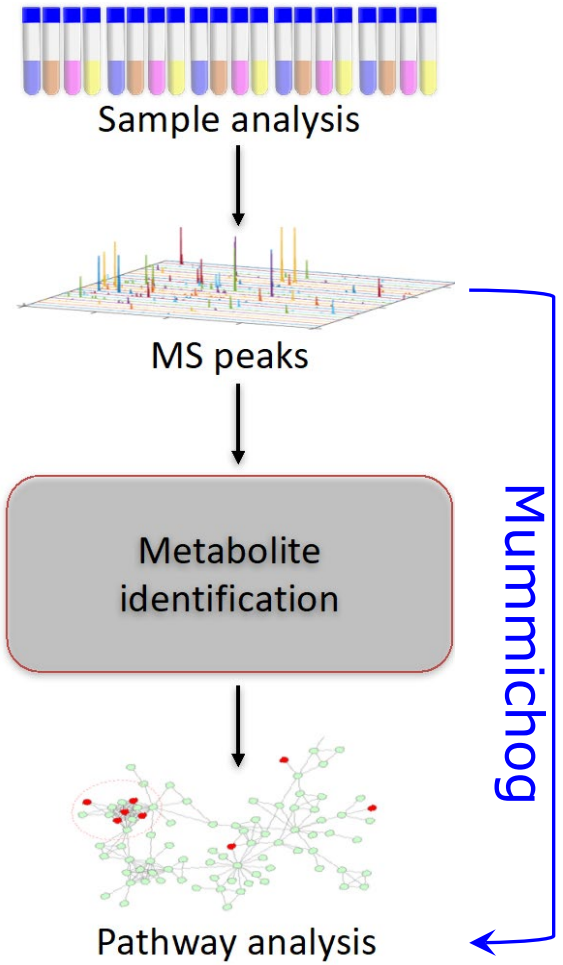


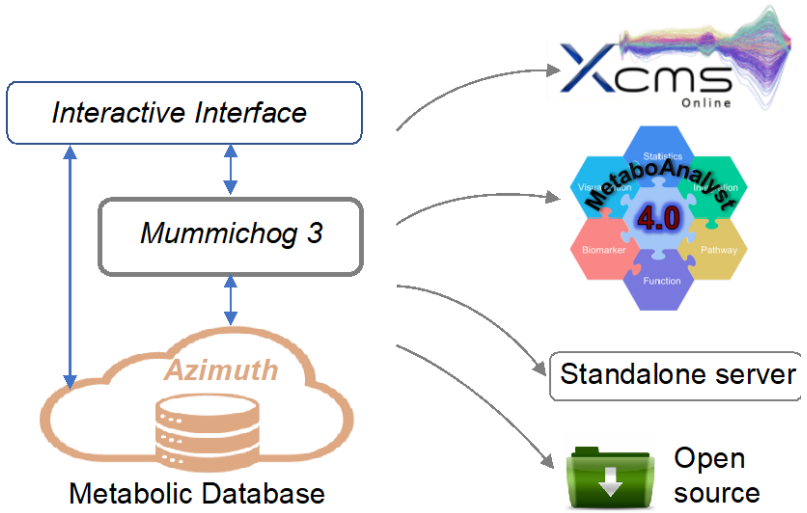
Overview of a typical metabolomics workflow for analysis of MS datasets.



Chang, et al. A Practical Guide to Metabolomics Software Development. Anal Chem. 2021, 93, 1912-23

Mummichog – pathway analysis





- Coupling visual analytics
- Azimuth, the chemical database of biology
- Metabolic network reconstruction using mass spectrometry data
- Data integration

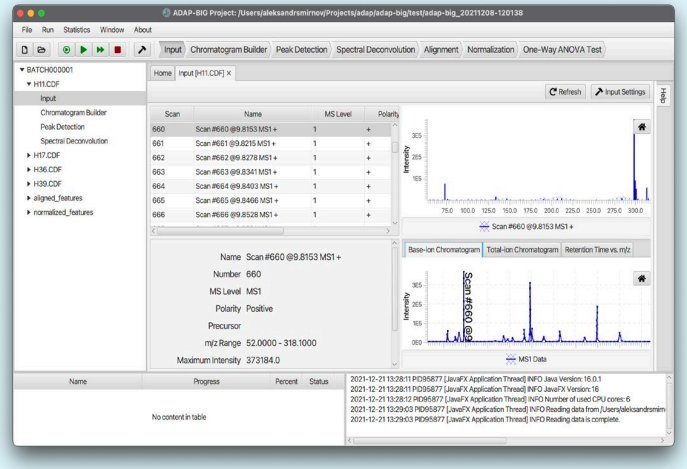
Impact:

- Original publication
  - 500 citations
- Mummichog v2 server at Jax Labs
  - >6,000 user jobs completed
- Used as integrated component via MetaboAnalyst and XCMS Online
  - MetaboAnalyst has >500,000 users

ADAP – data preprocessing

ADAP-BIG

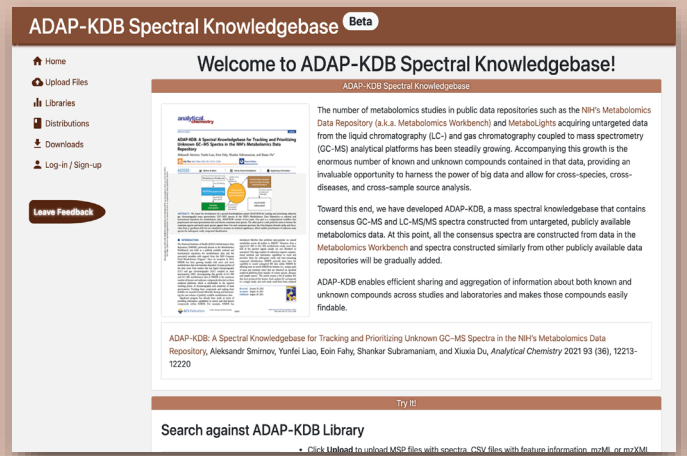
https://adap-big.github.io



- Processes big data from large multi-batch studies
- Provides visualization of every data processing step
- Performs statistical analysis (ANOVA, PCA)
- Processes raw data for ADAP-KDB

ADAP-KDB

https://adap.cloud



- Contains ~48k spectra from 113 studies from the Metabolomics Workbench
- Harmonizes data by constructing consensus spectra
- Calculates distributions of species, sample sources, diseases, etc.

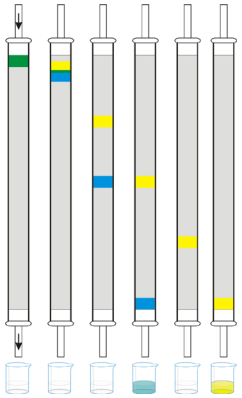
Impact:

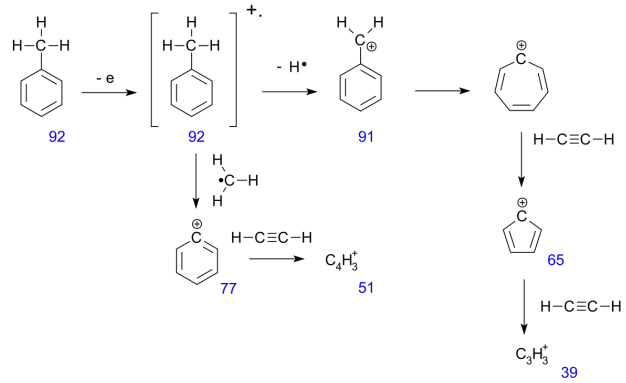
- ADAP-BIG and ADAP-KDB
  Data processing and library matching in the NIEHS HHEAR Initiative and Nutrition for Precision Health Common Fund Program.
- ADAP-BIG
  Processing metabolomics data from the NHLBI CARDIA Study.

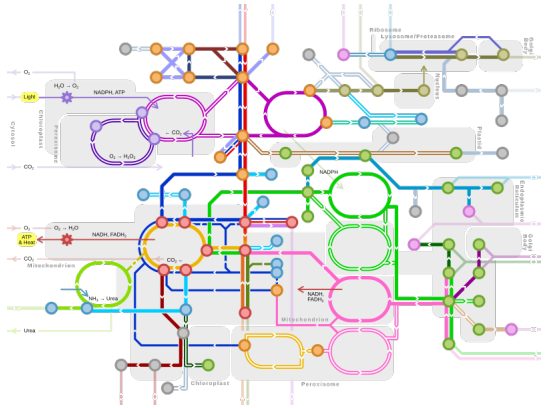
Improving Compound Identification

Problem: Compound Identification, historically, has relied on an authenticated standard… With an infinite number of potential metabolites and other small molecules it is impossible to 'brute force' Compound Identification.

Solution: Link Experimental and Computational approaches to enable predictive identification based on orthogonal data.







Separation Science:

Fragmentation:

Metabolic Transformation:

UC DAVIS
UNIVERSITY OF CALIFORNIA

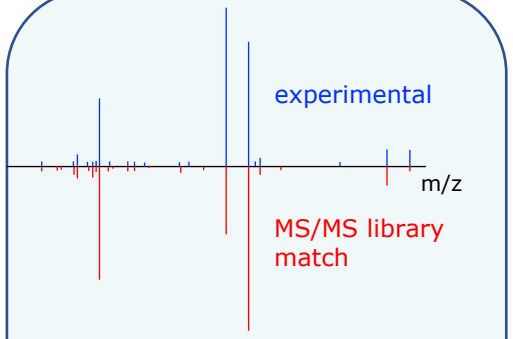
West Coast Metabolomics Center for Compound Identification

Spectral Entropy: Using the complexity of Fragmentation Pattern to improve confidence in identification

Experimental Core: Generate of chromatography and MS/MS libraries from standards
Computational Core: In silico modeling of chromatography and MS/MS
Experimental Core: Validate with a broader library

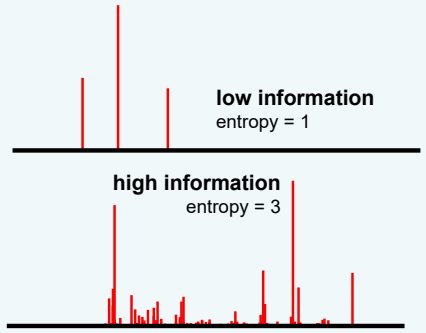
Impact: Reduce False Discovery Rate by 50%; outperform 42 current algorithms

Problem



(a) **How similar is similar enough?**
(b) **What about spectra with few ions?**

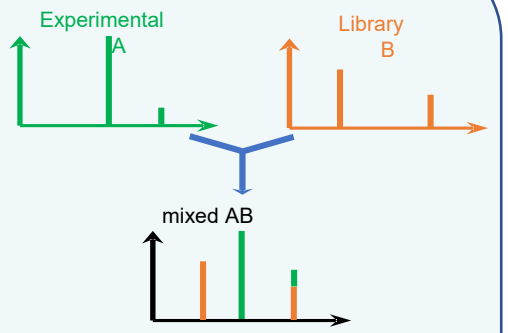
Idea



Convert mass spectra into information content ('entropy')

$$S = -\sum_p I_p \ln I_p$$

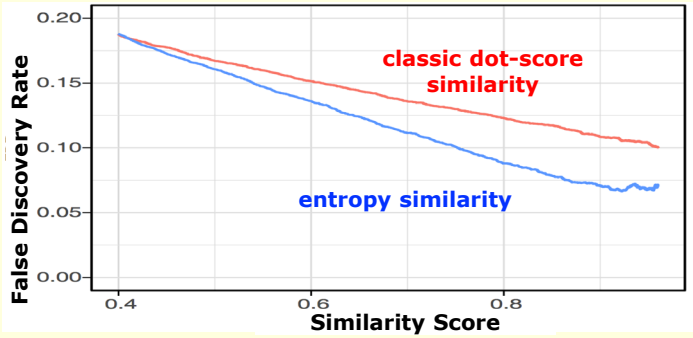
Implementation



entropy similarity =

$$1 - \frac{(S_{AB} - S_A) + (S_{AB} - S_B)}{\ln 4}$$

Result



- **testing by > 100,000 NIST20 spectra**
- **validation by MassBank.us spectra**
- **applied to GI tract metabolomics**

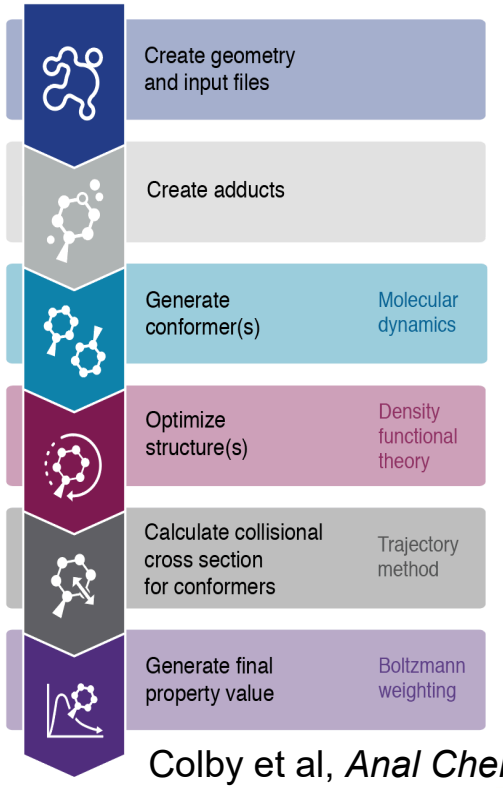
Nature Methods 2021, 18:1524
Fiehn O. *et al.* UC Davis

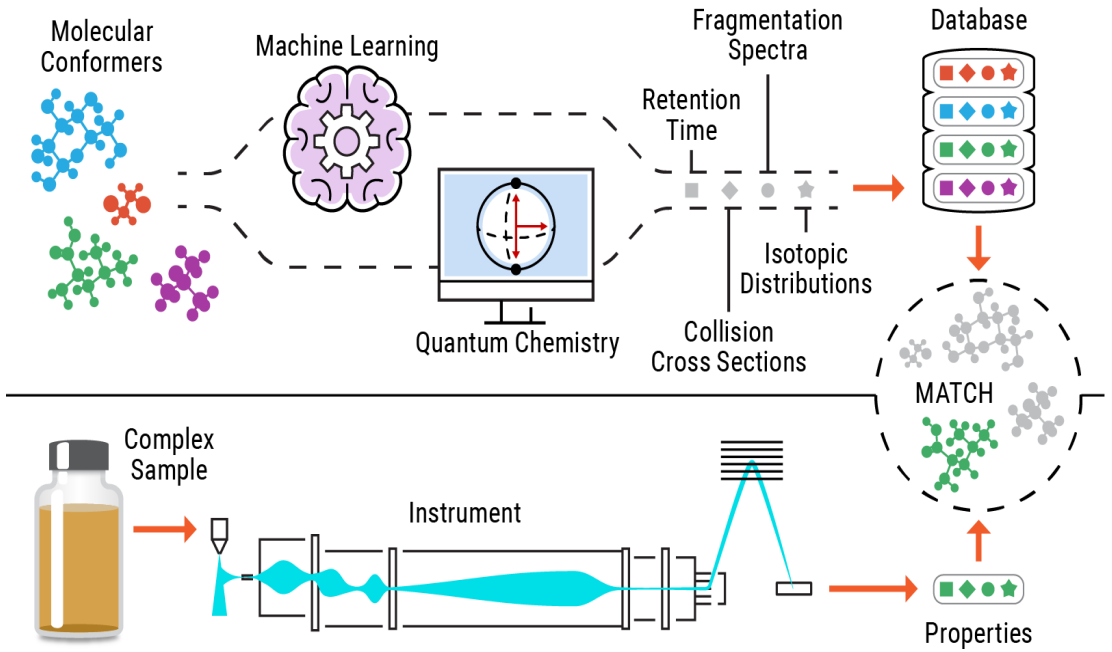
Pacific Northwest Advanced Compound Identification Core

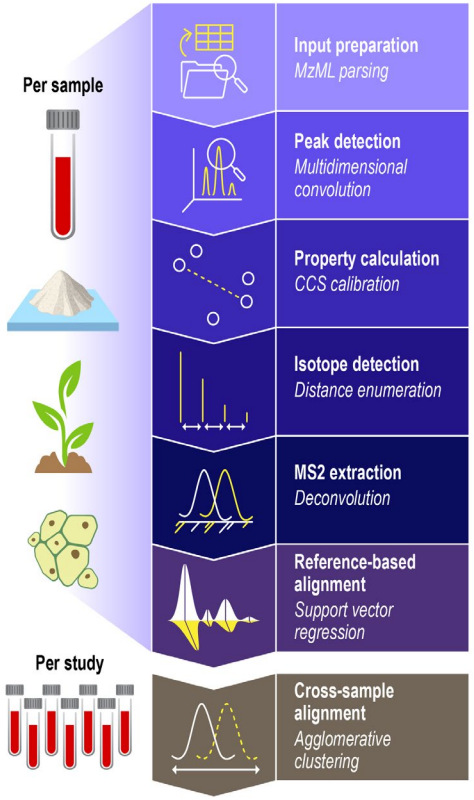
Framework for a Future compound ID paradigm

ISiCLE – Experimentation and Quantum Chemistry to inform Machine Learning-based prediction of chemical shape (Collisional Cross Section in ion mobility)

DEIMoS – high-dimensional Mass Spec data processing







Colby et al, *Anal Chem*, 2019

National Metabolomics Data Repository

Established in 2014 and expanded in 2018

SAMPLES

75,000 samples
From over 2000 deposited studies

METABOLITES

32,000 known metabolites detected (with ~ 3.5m unknow features)
320,000 names cross reference to 22,000 metabolites (RefMet)
143,000 Metabolites structures

ANALYSIS

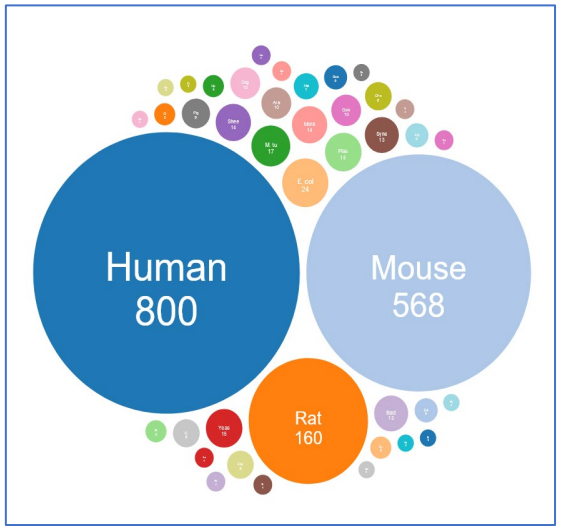
Ability to launch dozens of analysis tools in cloud workspace
25 terabytes of raw spectral data

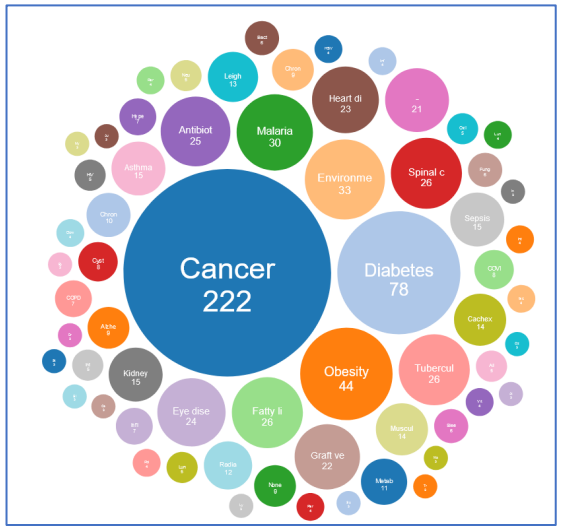
Compare Studies

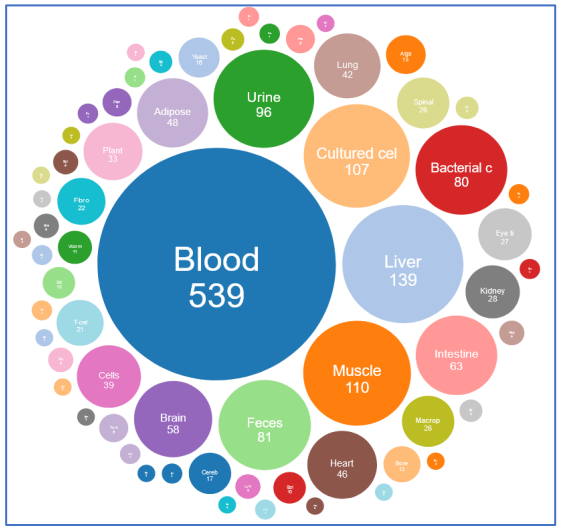
Identify Metabolite Functions

Improve Technology

National Metabolomics Data Repository



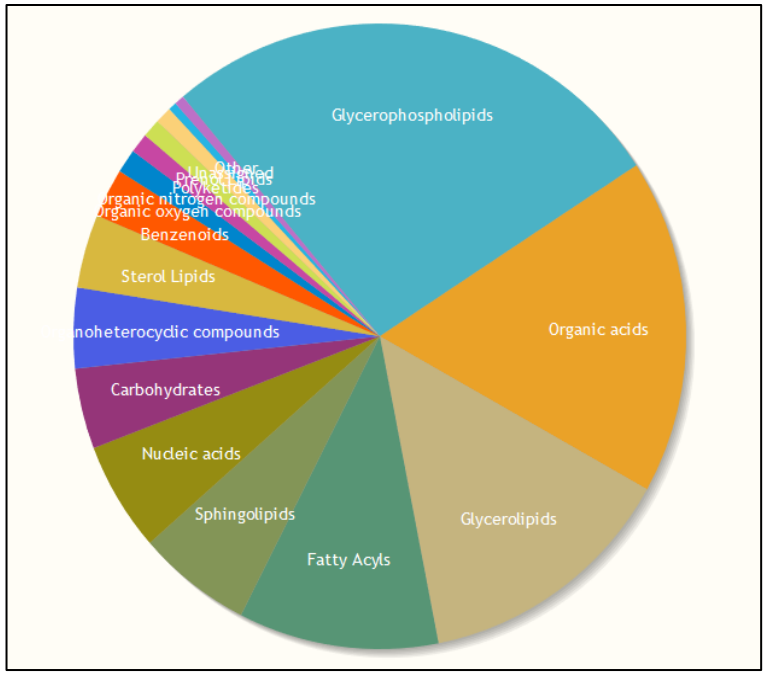




134 Different Species

133 Different Conditions

151 Different Tissue/Sample types



National Metabolomics Data Repository (Data Analysis)

Bringing Analysis Capability to the Data (CLOUD COMPUTING)

Provides and extensive study/sample/metabolite/protocol searches

Extensive R based data mining, statistical analysis and visualization tools, Jupyter workflows

Integration of the metabolomics specific tools being developed

Cross reference metabolite names (RefMet) and structures

Links to many other relevant resources and database of available analysis tools

Promoting Best Practices – Consortium Working Groups

Quantum Mechanics (QM) Working Group

Goal: Define/share best practices/pipelines in QM for metabolomics
Publication: Quantum Chemistry Calculations for Metabolomics, *Chem Rev.* 2021 May 26; 121(10): 5633–5670. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8161423

Software Standards for Metabolomics Tool Development

Goal: Provide recommendations and SOPs for software development within the metabolomics consortium
Publication: A Practical Guide to Metabolomics Software Development, Anal Chem. 2021 Feb 2;93(4):1912-1923. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7859930

Internal Standards Working Group

Goal: Develop and evaluate the use of internal standard kits to improve cross-lab comparability of LC-MS metabolomics data

Unknown Lipids Data Exchange Working Group

Goal: Facilitate cross-laboratory unknown feature alignment and compound identification

Promoting Best Practices – Collaborative Suppl. Projects

17 Supplements in total were awarded in FY19, FY20 and FY21 to support collaboration and best practices in:

Advancing Compound Identification

- Internal standards
- Exposome compound identification
- Tracking unknows in NMDR (ADAP-KDB https://www.adap.cloud)
- Data exchange on unknown features in lipidomics and polar metabolites

Promoting Software Development

- Moving tools to the cloud
- MSCAT: Metabolomics software CATalog (https://mscat.metabolomicsworkbench.org)
- Common datasets for tool and software benchmarking
- Portal for Open Computational Metabolomics Tools (POMICS)

Accomplishments

Access and training to hundreds of PIs in many fields
Helped build a community of collaborating investigator

Repository to share study data
Extensive ability to bring analysis to data (cloud computing)

Metabolomics specific tools at each stage of the analysis pipeline
Standards and guidelines for developing metabolomics software

Experimental and in silico data/tools for compound identification
Standards and guidelines for compound identification/quantification

Future Directions

Supporting Large NIH projects Dependent on Metabolomics Data

Ability to Improve the Integration of Controlled Access Data with Metabolomics

Ability to Combine Multi-Omics Data with Metabolomics Data in Cloud Environment

Expand the use of in silico compound identification

Acknowledgement

Working Group Chairs

Griffin Rodgers, M.D., M.A.C.P., NIDDK; Dinah Singer, Ph.D., NCI; Rick Woychik, Ph.D., NIEHS; Linda Birnbaum, Ph.D., D.A.B.T, A.T.S.; Philip F. Smith, Ph.D., NIDDK

Program Officers and Coordinators

Art Castle, Ph.D., NIDDK; Krista Zanetti, Ph.D., M.P.H., R.D., NCI, and David Balshaw, Ph.D., NIEHS; Yuxia Cui, Ph.D., NIEHS

Common Fund Program Leaders

Douglas Sheeley, Sc.D. , Leslie K. Derr, Ph.D.

Project Scientists

Yuxia Cui, Ph.D., NIEHS, Padma Maruvada, Ph.D., NIDDK, Jill Morris, Ph.D., NINDS, Dan Shaughnessy, Ph.D., NIEHS, Pothur Srinivas, Ph.D., NCI, Mukesh Verma, Ph.D., NCI, Keren Witkin, Ph.D., NCI

Acknowledgement

Full Working Group List

LeeAnna Arrowchis, M.P.A (NIDDK)
David Balshaw, Ph.D. (NIEHS)
Paul J. Barrett, Ph.D. (OD)
Arthur L. Castle, Ph.D. (NIDDK)
Ricardo R. Cibotti, Ph.D. (NIAMS)
Yuxia Cui, Ph.D. (NIEHS)
David Dzamashvili (OD)
Yih-Woei Fridell, Ph.D. (NIA)
Brionna Hair, Ph.D. (OD)
Jennifer Lewis, M.B.A. (OD)
Amy Lossie, Ph.D. (NIDA)
Conrad Mallia, Ph.D. (NIAID)
Padma Maruvada, Ph.D. (NIDDK)
Jill Annette Morris, Ph.D. (NINDS)
Gary J. Murray, Ph.D. (NIAAA)
Laurie S. Nadler, Ph.D. (NIMH)
Griffin Rodgers, M.D., M.A.C.P. (NIDDK)
Sheena Scruggs, Ph.D. (NIEHS)
Dan Shaughnessy, Ph.D. (NIEHS)
Douglas Mark Sheeley, SC.D. (OD)
Lillian Shum, Ph.D. (NIDCR)
Dinah S. Singer, Ph.D. (NCI)
Pothur R. Srinivas, Ph.D., M.P.H. (NHLBI)
Danilo Tagle, Ph.D. (NCATS)
Hung Tseng, Ph.D. (NIAMS)
Mukesh Verma, Ph.D. (NCI)
Keren Lisa Witkin, Ph.D. (NCI)
Rick Woychik, Ph.D. (NIEHS)
Susan Wright, Ph.D. (NIDA)
Ashley Xia, M.D., Ph.D. (NIDDK)
Krista Zanetti, Ph.D. (NCI)