18th Annual Conference of the Metabolomics Society

METABOLOMICS 2022

Valencia, Spain | JUNE 19-23

Pre-Conference Workshops

Sunday 12 noon – 2pm

W1: Ion Mobility in Metabolomics: New Technologies and Workflows

Presenters

- Prof. Juan Vicente Sancho Llopis (Universitat Jaume I) (https://www.uji.es/departaments/com/base/estructura/personal?p_departamento=2284&p_profesor=56216)
- Kelly Hines (UGA) (https://www.chem.uga.edu/directory/people/kelly-m-hines). Kelly.Hines@uga.edu
- Christopher Chouinard (https://www.fit.edu/faculty-profiles/c/chouinard-christopher/). cchouinard@fit.edu
- Serge Rudaz (https://ispso.unige.ch/labs/fanal/:en)

Description

Ion mobility (IM) separations coupled to mass spectrometric detection (IM-MS) have seen increasing growth in the area of metabolomics, and shown to be a viable complement to other established techniques. IM-MS requires lower sample concentrations than NMR, shorter analysis times than either NMR or LC and, with sufficient resolving power, can distinguish isomeric species in complex mixtures. IM separations depend on the mobility coefficient (K), which determines the velocity of the gas phase ions in the electric field. K depends on a variety of instrumental parameters, masses of ion and gas molecules, and their rotationally-averaged collision cross section (CCS). The latter parameter can be correlated to the 3D structure of ions and has seen wide applicability in the area of metabolite annotation. In this workshop we will showcase the latest techniques and workflows in the field, with special emphasis to the annotation of lipids and xenobiotics and also computational approaches for predicting CCS values through machine learning and quantum mechanical approaches.

Workshop Objectives

Teach fundamentals of Ion Mobility. Explore new IM Technologies applied to Metabolomics

Learning Outcomes

Become familiar with IM-MS technology strengths and limitations

W2: Spectra processing, functional integration and covariate adjustment of global metabolomics data using MetaboAnalyst 5.0 (Part 1)

*Hands-On – laptop required

Presenters
Jianguo (Jeff) Xia / McGill University

Description
Liquid-chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has become a workhorse in global metabolomics studies with growing applications across biomedical and environmental sciences. However, outstanding bioinformatics challenges in terms of raw data processing, statistical analysis and functional interpretation remain critical barriers to the wider adoption of this technology. To help address growing user requests, we have made major updates to the well-established MetaboAnalyst platform (metaboanalyst.ca). This workshop aims to introduce these features combined with hands-on training on how to use MetaboAnalyst 5.0 to perform: 1) optimized LC-HRMS spectra processing and peak annotation; 2) functional analysis and integration with other omics data; and 3) exploratory statistical analysis with complex metadata. A case study based on recent COVID-19 datasets will be used to demonstrate the key features of each component. A detailed step-by-step protocol (~80 pages) will be made available for workshop attendees to follow during or after the workshop.

Workshop Objectives
1) optimized LC-MS spectra processing and peak annotation;
2) functional analysis and integration with other omics data; and
3) exploratory statistical analysis with complex metadata.

Learning Outcomes
How to use MetaboAnalyst 5.0 web server to perform LC-MS spectra processing, complex statistical analysis and functional integration for global metabolomics data

Sunday 2:15 – 4:15pm

W2 Part 2: Spectra processing, functional integration and covariate adjustment of global metabolomics data using MetaboAnalyst 5.0 (Part 2)

*Hands-On – laptop required

Presenters
Jianguo (Jeff) Xia / McGill University

Description
Liquid-chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has become a workhorse in global metabolomics studies with growing applications across biomedical and environmental sciences. However, outstanding bioinformatics challenges in terms of raw data processing, statistical analysis and functional interpretation remain critical barriers to the wider adoption of this technology. To help address growing user requests, we have made major updates to the well-established MetaboAnalyst platform (metaboanalyst.ca). This workshop aims to introduce these features combined with hands-on training on how to use MetaboAnalyst 5.0 to perform: 1) optimized LC-HRMS spectra processing and peak annotation; 2) functional analysis and integration with other omics data; and 3) exploratory statistical analysis with complex metadata. A case study based on recent COVID-19 datasets will be used to demonstrate the key features of each component. A detailed step-by-step protocol (~80 pages) will be made available for workshop attendees to follow during or after the workshop.

Workshop Objectives
1) optimized LC-MS spectra processing and peak annotation;
2) functional analysis and integration with other omics data; and
3) exploratory statistical analysis with complex metadata.

Learning Outcomes
How to use MetaboAnalyst 5.0 web server to perform LC-MS spectra processing, complex statistical analysis and functional integration for global metabolomics data

W3: Mass spectrometry data processing with MZmine 3: from LC-MS to ion-mobility-enabled molecular networking

Presenters

- Tomáš Pluskal, IOCB Prague, Czechia
- Steffen Heuckeroth, University of Münster, Germany
- Robin Schmid, UC San Diego, USA
- Daniel Petras, University of Tübingen, Germany
- Tito Damiani, IOCB Prague, Czechia

Description

MZmine 3 is the third generation of the popular open-source, platform-independent software framework for MS data processing and visualization. Since the introduction of MZmine 2 a decade ago, the project has matured into a community-driven, collaborative platform and its functions have been expanded with a variety of new modules to process LC-MS, GC-MS, LC-ion mobility-MS, and mass spectrometry imaging data. MZmine enables research by spectral preprocessing, feature detection, and various options for metabolite identification. Its modern graphical user interfaces and interactive charts facilitate data exploration and validation of results from every processing step. Recent developments have strengthened the ties to other open source projects, including the GNPS platform with Ion Identity Molecular Networking and the SIRIUS suite for compound annotation. In this workshop, we will introduce new and experienced users into feature finding concepts in MZmine leading up to molecular networking on the GNPS web platform, a new way to map the chemical diversity of samples by spectral fragmentation similarity. While mostly focussing on LC-IMS-MS data analysis, this session will also be interesting for users seeking to process any other type of MS data.

Workshop Objectives

1. Introducing the MS data processing capabilities of MZmine 3, with a particular focus on new features such as ion mobility spectrometry or MS imaging;
2. Demonstrating the workflow for Ion Identity Molecular Networking for LC-(IMS)-MS/MS data together with the GNPS web platform (including the easy entry through the MZmine Wizard);
3. Demonstrating the modules for spectral annotation, including the integration with other tools such as SIRIUS or molecular networking on GNPS

Learning Outcomes

1. Getting a broad overview of the MZmine 3 features for MS data processing and metabolite identification;
2. Learning where to find resources for MZmine usage and integration with other tools for metabolomics;
3. Learning how to connect MZmine processing results to downstream analysis in GNPS or SIRIUS

W4: Frontiers in NMR Metabolomics

Presenters

- NMR metabolomics overview (high fields) (Panteleimon Takis, Imperial College London & Leo Cheng, Massachusetts General Hospital & Harvard Medical School)
- NMR metabolomics at low fields (Stefan Gloeggler, Max Planck Institute)
- NMR metabolomics in medicine (cancer and other diseases) (Dr. Julian Griffin, University of Aberdeen)
- Integration of NMR and MS metabolomics data (Daniel Raftery, University of Washington)
- QA/QC Issues seen in current publications (Robert Powers, University of Nebraska-Lincoln)
- Metabolomics data analyses and presentations (Nathaniel Mercaldo, Massachusetts General Hospital & Harvard Medical School)

Description

This workshop will review and examine the current state of NMR-based metabolomics, from its achievements in various research fronts to aspects that require further attention. Expert discussion will cover current areas of NMR metabolomics studies achievable at high field for different scientific disciplines using traditional solution-phase and high-resolution magic angle spinning approaches, as well as metabolomics imaging with localized spectroscopy. Specific examples will be illustrated with various cancer studies. New NMR horizon potential metabolomics measurements at extremely low magnetic field will be presented to introduce the concept of NMR metabolomics imaging beyond chemical shifts, and connections between NMR and mass spectrometry (MS) based metabolomics will be discussed to demonstrate the combined strength of these two methods for metabolomics research. To further improve and standardize research efforts across labs, quality assurance (QA) and quality control (QC) issues will be discussed, together with recommendations on metabolomics data statistical analyses to ensure the accuracy and reproducibility. A discussion on data presentation will help communicate more effectively with the larger scientific community.

Workshop Objectives

1. Understand the current state of NMR-based metabolomics in different scientific disciplines;
2. Learn different NMR methodologies applicable to metabolomics studies;
3. Understand current issues, such as QA/QC, that enable broader metabolomics integration across studies and laboratories; and
4. Learn effective data analysis and presentation tools.

Learning Outcomes

This workshop will provide attendees with a background in NMR with an opportunity to overview the field, to be informed by new NMR approaches, and to consider how to standardize their studies to ensure accuracy and reproducibility. For non-NMR focused researchers, in addition to learning information on the landscape of the current NMR metabolomics fields, attendees may appreciate the enhanced capabilities obtained when MS is assisted by NMR in metabolomics studies. For all attendees, systematic presentations of recommendations on optimal and accurate data statistical analyses and presentations will enhance their ability to perform correct and accurate data mining and to better communicate their results.

Sunday 4:30 – 6:30pm

W5: State of QA/QC Best Practices in LC-MS-Based Untargeted Metabolomics, Informed Through mQACC Community Engagement Initiatives

Presenters

1. Introduction to mQACC and QA/QC principles, Warwick Dunn, University of Liverpool, UK;
2. Pooled and intra-study QC samples, Tracey Schock, National Institutes of Standards and Technology, USA;
3. System suitability evaluation, Dajana Vuckovic, Concordia University, Canada;
4. Use of internal standards, Julia Kuligowski, Health Research Institute La Fe, Spain;
5. Design of the analytical batch, Jonathan Mosley, U.S. Environmental Protection Agency, USA;
6. Open Discussion on Living Guidance Document, Jonathan Mosley, U.S. Environmental Protection Agency, USA

Description

There is a critical need to standardize and implement QA and QC best practices in untargeted metabolomics to ensure high quality data generation and analysis. To address this need, the metabolomics Quality Assurance and quality Control Consortium (mQACC) was established in 2017 and currently has over 90 international members. The mQACC Best Practices Working Group has actively engaged the metabolomics community over the last three years through a combination of eight interactive forums and workshops with a total attendance of 390 participants across all events to identify key practices across relevant technical areas: pooled/intra-study QC samples, system suitability evaluation, use of internal standards, and analytical batch design. The goal of this workshop is to disseminate the key findings from these extensive efforts to the broader metabolomics community. Additionally, it will act as a forum for critical discussions to further fine tune the compiled information to establish an open-access best practices “living guidance” document that reflects the community feedback received and becomes the QA/QC go-to resource for metabolomics scientists.

Workshop Objectives

1. To disseminate findings from the mQACC Best Practices Working Group’s extensive community engagement efforts to establish best practices for LC-MS data collection in untargeted metabolomics. 2. To solicit further feedback from the international metabolomics community on the compiled and summarized findings to establish an open-access best practices “living guidance” document that will be freely accessible to researchers.

Learning Outcomes

1.Attendees will understand the community feedback received by the mQACC Best Practices Working Group and recognize how the feedback will be used to support QA/QC best practices for untargeted LC-MS-based metabolomics.
2.Attendees will be able to identify how to participate in mQACC, including mechanisms to contribute to the best practices community engagement efforts

W6: EMN – Professional Career Development
Post-graduate: Where do I go from there?

Moderator: Susana Palma, PhD, The Francis Crick Institute, UK
Presenters:

- Amani Said, PhD, Success Beyond the Lab, Spain
- Ewy Mathé, PhD, National Center for Advancing Translational Sciences, USA
- Reza Salek, PhD, Bruker BioSpin GmbH, Germany

Description
The phrase “alternative career paths” is no longer appropriate when the majority of PhD scientists are entering roles outside of academia. In this year’s EMN Professional Development Workshop, we’ll discuss prominent career paths outside of academia and lesser-known career paths PhD metabolomics scientists have chosen to pursue. This workshop is targeted to Early Career Researchers (ECRs) to explore career pathways both within and beyond metabolomics, and address the skills needed to transition. The EMN Professional Career Development workshop will support ECRs who are curious about the approach to career transitions and career paths across academia, industry, government, and non-profit corporations among others. The workshop will be separated into four sections: **(1)** assessing skill sets that align with the vision of one's career path, **(2)** career transitions between government and academia, **(3)** career paths in industry and non-profit, and **(4)** a question and answer section for audience members to further engage with the speakers.

Workshop Objectives
1. Enhance awareness of the numerous career pathways for an individual within the metabolomics field;
2. Discuss transitions between different fields of metabolomics within academia and the shift to non-academic career paths;
3. Understand how scientific training fits with many different career possibilities and assess current skills

Learning Outcomes
Attendees will receive insights on the skill-set required to identify their next career path. For ECRs, some skills can still be honed or acquired before completing their current tenure and tips will be provided to aid their next career endeavors. The panelist will discuss how they were able to translate their academic work experiences into “skills” that were desired from hiring managers in different fields.

W7: Towards Spatial Metabolomics

*Hands-On – laptop required

Presenters

- María García-Altares (URV) https://orcid.org/0000-0003-4255-1487
- Pere Ràfols (URV) https://orcid.org/0000-0002-9240-4058 https://github.com/prafols
- Christoph Bookmeyer (URV) https://orcid.org/0000-0002-3479-3841

Description

Spatial Metabolomics explores the metabolism and its changes directly on biological tissues, providing a "chemical map" of each detected metabolite that informs about its location and relative abundance on the tissue. Spatial Metabolomics is, for instance, widely used to understand how tumors exchange metabolites with their environment, or how microbial communities transfer compounds among colonies.
In this workshop, we aim to present how Mass Spectrometry Imaging (MSI) can be used for Spatial Metabolomic studies. In the first part, we will use short videos made in our lab to show how to design and perform MSI experiments, and we will discuss how to adapt already existing metabolomics workflows to include Spatial Metabolomics. In the second part of the workshop, we will demonstrate how to use our in-house developed open-source R package rMSI, able to visualize, process, and analyze MSI data. This demonstration will include test data for the participants of the workshop to gain hands-on experience with rMSI and its integration into the R platfom.

Workshop Objectives

1) To appreciate how Spatial Metabolomics can provide valuable complementary information on Metabolomics studies.
2) To design experiments for MSI that can be integrated into current Metabolomic workflows.
3) To learn how to visualize, process, and approach the data analysis of an MSI experiment.

Learning Outcomes

1) How Mass Spectrometry Imaging is progressing towards Spatial Metabolomics (i.e. strategies in quality control, standardization, annotation, and identification, etc.).
2) What are the steps to conduct an MSI analysis, from experimental design to data acquisition.
3) How to visualize, process, and analyze MSI data with the open-source software family rMSI, developed by Mil@b group specifically for MSI.

Monday 8:15 – 10:15am

W8: Clinical Lipidomics

Presenters

- Margrét Thorsteinsdóttir; University of Iceland, Iceland
- Tuulia Hyötyläinen; Örebro University, Sweden
- Dajana Vuckovic; Concordia University, Canada
- Markus Wenk; National University of Singapore
- Maria Fedorova; Technische Universität Dresden, Germany
- Matej Oresic; Örebro University, Sweden

Details from Organizers

https://www.nordicmetsoc.org/CLIPflier-MetSoc2022.pdf

Description

Within the metabolomics community, there is growing interest in the comprehensive analysis of lipids (lipidomics), especially in the clinical domain, because lipid-related disturbances play important roles in the pathogenesis of most of the common diseases. However, there are many (pre)analytical and biological factors that impact human lipid levels as measured by lipidomics, which must be considered when planning and conducting lipidomic analyses. This workshop will introduce (1) the key factors that impact human lipid levels as measured by lipidomics, including biological (e.g., diet, circadian rhythm) and preanalytical (e.g., sample collection and storage), and (2) the analytical techniques for lipidomic analysis of human biofluids and tissues. The specific topics covered will be: (1) Clinical lipidomics – an overview, (2) Preanalytical factors affecting lipidomes, (3) Microsampling techniques, (4) Analytical techniques for lipidomics, including inter-laboratory comparability and reference ranges, (5) Tissue lipidomics – analytical aspects, and (6) Pipeline for clinical lipidomics.
The workshop will conclude with panel discussion. The workshop is intended for metabolomics researchers entering or already pursuing research in clinical lipidomics. They are expected to learn valuable lessons about planning and conducting lipidomic studies.
This workshop was initiated as part of the European COST action “Pan-European Network in Lipidomics and EpiLipidomics”.

Workshop Objectives

(1) To introduce the key factors that impact human lipid levels as measured by lipidomics, including biological (e.g., diet, circadian rhythm) and preanalytical (e.g., sample collection and storage),
(2) To introduce the analytical techniques, and related challenges, for lipidomic analysis of human biofluids and tissues, including discussion of inter-laboratory comparability and reference ranges.

Learning Outcomes

(1) Understanding of key (pre)analytical and biological factors that may affect the lipid levels in human studies.
(2) Knowledge of key methods for sample collection, preparation and analysis for lipidomics of human biofluids and tissues.

W9: Mining the Metabolome Using the Mass Spectrometry Query Language and MS2LDA

*Hands-On – laptop required

Presenters

- Mingxun Wang
- Allegra Aron
- Justin van der Hooft

Description

This workshop will introduce both the use of Mass Spectrometry Query Language (MassQL) to describe and search for specific patterns in mass spectrometry data and the use of MS2LDA for unsupervised substructure discovery.

Participants will learn how to:
i) Express mass spectrometry patterns in MassQL
ii) Query data using MassQL in GNPS
iii) Browse through the results on the GNPS interface
iv) Run an MS2LDA job using the GNPS-MS2LDA interface and which parameters are important to consider
v) Browse through the results and annotate Mass2Motifs on the ms2lda.org interface
vi) Store annotated Mass2Motifs in MotifDB and use them in future metabolome mining workflows

Finally, the participants will get a perspective on how MS2LDA-MotifDB and MassQL can be used in tandem to enhance the metabolome mining experience.

Workshop Objectives

1. Teach participants how to use MS2LDA for substructure discovery
2. Teach participants how to write simple MS1 and MS2 queries using MassQL to find substructure in dataset and in public data
3. Teach participants to explore and reuse MassQL queries written by the community

Learning Outcomes

1. Accelerate researcher's ability to mine metabolomics data, especially when searching for specific chemical classes or scaffolds
2. Enable researchers to find substructures within metabolomics data
3. Enable researchers to codify and share mass spectrometry knowledge to enable data reproducibility

W10: Hitchhikers' Guide to Networks in Metabolomics

Presenters

- Steffen Neumann (Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany)
- Elva Novoa (INRAE, Toulouse, France)
- Karl Burgess (University of Edinburgh, Scotland, United Kingdom)
- Thomas Naake (European Molecular Biology Laboratory (EMBL), Heidelberg, Germany)
- Ricardo Roberto da Silva (University of São Paulo, São Paulo, Brazil)
- Horst Joachim Schirra (Griffith University, Queensland, Australia)

Details from Organizers

https://docs.google.com/document/d/1u-P9jDj31OnKYGadcdxrNe0M6O2HUsiuk7IBpJaGqek/edit

Description

Biological interpretation and metabolite annotation in large and complex metabolomics datasets can benefit from the organization as rich graphs or networks with informative relationships as connections. There are two main types of networks in metabolomics studies: knowledge-based and experimental networks. Knowledge-based networks contain entities and connections previously derived, experimentally or theoretically (e.g. metabolites and biochemical reactions) that we know occur in a given organism; whereas experimental networks are derived from data, for example, by applying some statistical measures (such as correlation) directly to the metabolomics data. Each network (knowledge-based or experimental) gives us a different (and complementary) perspective of metabolism.

Workshop Objectives

1) Understand how networks are built and used within the context of metabolomics research
2) Get an overview of different types of networks and how they can contribute to metabolomics and systems biology research
3) Understand the differences between experimental (e.g. molecular, correlation networks) and knowledge-based networks (e.g. genome scale metabolic networks)

Learning Outcomes

1) Understand the different networks which can be produced and used in metabolomics
2) Identify the benefits of using these networks in metabolomics research
3) References to literature, courses and tutorials for future learning

Monday 10:30am – 12:30pm

W11: The 3 R's of Effective Data Sharing in Metabolomic Epidemiology: Rigor, Reproducibility, and Repositories

Presenters

Moderator: Krista Zanetti, PhD, MPH, RD, National Cancer Institute, USA
Speakers and panelists:

- Current state of the field: Jessica Lasky-Su, PhD, Brigham and Women's Hospital, USA
- The epidemiology perspective: Bo Chawes, MD, PhD, DMSc, Copenhagen University Hospital, Denmark
- The data science perspective: Tim Ebbels, PhD, Imperial College London, UK
- Alternative approach to public repositories: Rachel Kelly, PhD, Brigham and Women's Hospital, USA

Description

While the sharing of data is key to the furthering of all science, this is particularly true for metabolomic epidemiology. Access to data is vital to enable this growing field to deliver on its potential. The two primary goals of data sharing in metabolomic epidemiology are to: 1) provide a data resource to facilitate original research involving epidemiological studies; and 2) ensure reproducibility, accuracy, and transparency in data analysis.

Multiple repositories have been established for metabolomic data. However, data deposition requirements specific to metabolomic epidemiology studies are challenging due to the complexities of large-scale demographic, covariate, and phenotypic data required for epidemiological modelling. This workshop will: 1) discuss the current state of the field with respect to data sharing; 2) review key data requirements for metabolomic epidemiologic studies to ensure study rigor and reproducibility; and 3) consider the implications of these requirements for repository design.

From the epidemiology and data science perspective, speakers will present strategies and challenges for facilitating rigor, quality, transparency, and reproducibility in data sharing. Additionally, alternatives to public data repositories will be highlighted, including sharing aggregate data, as demonstrated by the Consortium of Metabolomic Studies (COMETS - https://epi.grants.cancer.gov/comets/). A panel Q&A will end the workshop.

Workshop Objectives

1) To understand critical data needs for metabolomic epidemiological studies to promote data reproducibility, highlighting the important considerations for clinical data.
2) To identify requirements for epidemiological data deposition to public repositories.
3) To describe alternate approaches implemented in metabolomic epidemiological studies to address the need for improved rigor and reproducibility in the absence of individual-level data access.

Learning Outcomes

1) Attendees will understand the unique requirements for data deposition for metabolomic epidemiological studies compared with other metabolomic studies.
2) Attendees will be able to describe established approaches to address data sharing for metabolomic epidemiological studies.

W12: Revisiting CASMI: showcase your success in compound identification for 500 new unknowns, using raw LC-MS/MS data.

Presenters – TBD

Details from Organizers
https://fiehnlab.ucdavis.edu/casmi

Description
In untargeted assays, approximately 80% of all MS/MS spectra remain unidentified. Yet, the last CASMI contest was performed five years ago. As molecules of former CASMI contests are now well known, new datasets must be provided for testing compound ID strategies in 2022. For the first time, we will give full chromatograms as input into a CASMI contest. Workshop organizers from UC Davis will provide raw LC-accurate mass MS/MS data in +/- ESI mode, with a list of 500 compounds and their retention times and nominal m/z values. Compound spectra are not yet included in public libraries, but comprise both metabolites and exposome chemicals. Data will be available for downloads by March 15, including chromatography and MS conditions. While we will encourage specific scientific teams to participate, anyone can utilize the data and submit results. UC Davis will open a portal for submission until two weeks prior to the conference. Submitters will be notified, and we will select six submitters to showcase their results and approaches. Results will be discussed based on four different categories of correct annotations: 1) adducts 2) elemental formula 3) structure class and 4) full chemical structures. All submitters may participate in manuscript writing afterwards.

Workshop Objectives
Enable comparison of the current performance in compound ID approaches from raw data – generate new public datasets and write manuscript; Understand advantages and pitfalls of different compound identification approaches; Advance expertise across disciplines towards compound identification

Learning Outcomes
Learn about current tools, software, and compound identification approaches. Ask panel experts how to use tools, including pitfalls or barriers of applying tools for your own research. Understand the different levels and hierarchies in compound identification, from determining adducts to elemental formulas, structure classes to structure isomers.

W13: Big Data Machine Learning Methods for Metabolomics

*Hands-On – laptop required

Presenters

- Kangni Alemdjrodo (Purdue University)
- Daniel Raftery (University of Washington)
- Dabao Zhang (Purdue University)
- Min Zhang (Purdue University)

Description

The ability to measure a large number of metabolites provides a unique opportunity to systematically reveal the relationship between metabolites and diseases, facilitating biomarker identification and disease prediction. However, the combination of a large number of metabolites and a small number of samples challenges the data analysis, especially when evaluating the nonlinear relationship between diseases and metabolites or gene-metabolite interactions.

First, an overview is given on popular multivariate methods used in metabolomics, such as logistic regression and PLS-DA. We will then introduce methods that can handle a huge number of variables, such as XGBoost, and our newly developed method, gPOCRE-based screening (GPS), which can improve metabolomics data analysis by selecting a small subset from a large number of variables. The workshop will feature hands-on exercises developed by an NIH/NCI-funded research education program, “Big Data Training for Cancer Research”. The workshop will introduce data pre-processing procedures, guide participants to run through hands-on exercises, focus on different data types, including NMR, LC-MS, and genomic data, visualize the results, interpret the outputs, and discuss follow-up analysis. Designed primarily for metabolomics researchers with limited statistical backgrounds and computational skills, this workshop aims to assist participants in analyzing metabolomics data using advanced machine learning methods.

Workshop Objectives

This workshop aims to help biomedical researchers learn how to use popular and advanced machine learning tools to analyze metabolomics data efficiently, especially when the number of metabolites is large but the total number of biological samples is small.
(1)use popular multivariate statistical methods for metabolomics data analysis;
(2)use advanced machine learning tools for efficient analysis of metabolomics data;
(3)evaluate the relationship between disease status and gene-metabolite interactions.

Learning Outcomes

The participants will recognize the value and limitations of popular and advanced machine learning methods in metabolomics research and analyze metabolomics data using these tools.
(1)Analyze metabolomics data using popular multivariate statistical methods;
(2)Analyze metabolomics data using advanced machine learning tools.