Bioinformatics Hub (BioHub)

The Bioinformatics Hub is a space designated to bioinformaticians working in metabolomics. The space consists of two round tables with ~20 chairs and a screen for presentations.

Please feel free to use this space anytime during the conference to discuss problems, issues, or ideas regarding metabolomics data processing, tool development, programming, or related topics.

The Hub will be hosting (5) scheduled talks to showcase new developments in computational metabolomics tools, find the complete schedule below.

Tuesday, June 24 9:40 a.m. – 10:10 a.m.

BioHub 1

Profiling 1D1H NMR metabolomics spectra with the CcpNmr Analysis Metabolomics Software via interactive and semi-automated curve-fitting of curated simulated metabolite standards

Instructors

- **Morgan Hayward,** CCPN University of Leicester

Learn how to use the tools in the CcpNmr Analysis Metabolomics software to confidently identify and quantify metabolites in 1D1H NMR spectra. Here we will demonstrate the Profile by Reference module, where users can interactively and accurately quantify metabolites in mixture spectra by overlaying and fitting representative simulations of metabolite standards spectra. Additionally, we will demonstrate the Create Database Reference module, where users can develop interactive in-house simulated libraries for enabling and streamlining specialised analyses in the Profile by Reference module. After demonstration, participants can try out examples or with their own data.

Who would benefit?

- Anyone who currently uses NMR or is interested in using NMR for metabolomics analyses.
- All levels of familiarity with NMR and/or metabolomics.
- Anyone interested in Python coding for NMR and/or metabolomics software.

BioHub 2

Metabonaut: A Collection of Tutorials for Learning Metabolomics Data Analysis in R

Instructors

- **Philippine Louail,** Eurac Research, Bolzano, Italy; Chair for Bioinformatics, Friedrich-Schiller-University Jena, Jena, Germany
- **Johannes Rainer,** Eurac Research, Bolzano, Italy

This session introduces Metabonaut, an educational resource comprising a series of reproducible tutorials for untargeted LC-MS/MS metabolomics data analysis using R. Built around a representative LC-MS/MS dataset, the tutorials demonstrate how to construct an end-to-end analysis workflow using tools such as xcms and other packages from the RforMassSpectrometry ecosystem. We briefly demonstrate how to seamlessly incorporate a variety of R packages, as well as external tools and resources such as public annotation databases, Sirius and Python libraries. The session will conclude with a short discussion on future directions, emphasizing the adoption of standardized data classes, structures, and file formats to streamline the integration of diverse software tools.

Who would benefit?

- Users, data analysts: get to know where to find up-to-date tutorials for various aspects in LC-MS/MS data analysis.
- Software developers: get an overview of the existing software ecosystem and data structures to find connection points for potential future collaboration and contribution.

BioHub 3
Empowering GC-MS metabolomics with PARADISe

Instructors

- **Prof. Rasmus Bro,** University of Copenhagen
- **PostDoc Beatriz Quintanilla,** University of Copenhagen

The interpretation of untargeted Gas Chromatography-Mass Spectrometry (GC-MS) data presents significant challenges due to the complexity of the datasets. PARADISe, a freely available and user-friendly software, offers a robust solution based on tensor decomposition and PARAFAC2 models. In this session, we will showcase the performance of PARADISe in metabolomics GC-MS data, particularly in addressing well-known and otherwise unresolved issues inherent to the analytical technique, e.g. batch analysis, baseline correction, limits of detection and quantification, highly co-eluted compounds, as well as saturated peaks.

Who would benefit?

- Researchers who work with GC-MS metabolomics data and would like to enhance the data analysis and interpretation steps.

Wednesday, June 25 3:00 p.m. – 3:25 p.m.

BioHub 4

Metabolomics data processing, analysis and metabolite identification using Metabox 2.0 and SiMD

Instructors

- **Assoc. Prof. Sakda Khoomrung,** Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand
- **Dr. Kwanjeera Wanichthanarak,** Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand

We developed Metabox 2.0 (https://github.com/kwanjeeraw/metabox2) as a comprehensive tool for data processing, statistical analysis, data integration and biological interpretation. It offers both a user-friendly GUI and ready-to- use R functions, allowing users to assess the quality and characteristics of their data and to select appropriate methods for processing and analysis. In parallel, we introduced the Siriraj Metabolomics Data Warehouse (SiMD) (https://metsysbio.com/simd/) - a four-dimensional metabolite reference library that incorporates mass-to-charge ratio (m/z), retention time (RT), fragment ions (MS/MS), and collision cross-section (CCS) values. These parameters are derived from high-quality reference standards characterized by ion mobility-mass spectrometry (IM-MS). SiMD supports various levels of metabolite identification through robust scoring algorithms based on these orthogonal dimensions.

In this session, we will demonstrate essential features of Metabox 2.0 for data processing and statistical analysis, as well as key functionalities of SiMD for metabolite identification. Participants will gain hands-on experience with critical workflows using our curated example datasets.

Who would benefit?

- Postgraduate students
- Early career researchers

MSB METABOLOMICS & SYSTEMS BIOLOGY

BioHub 5

MSlineaR: a new R package to assess linearity, improving statistical robustness and quality assurance in untargeted metabolomics

Instructors

- **Janine Wiebach,** Berlin Institute of Health at Charite Universitätsmedizin, BIH Metabolomics, Berlin, Germany
- **Ulrike Brüning,** Berlin Institute of Health at Charite Universitätsmedizin, BIH Metabolomics, Berlin, Germany
- **Jennifer Kirwan,** Berlin Institute of Health at Charite Universitätsmedizin, BIH Metabolomics, Berlin, Germany

Mass spectrometry based untargeted metabolomics involves the analysis of complex biological matrices, often containing thousands of individual features. Cleaning up datasets early on in the data processing pipeline reduces data size, reduces risk of false positives, and lowers the burden on false discovery tests. Reduction of features to (1) only those that show linearity in the measured range in the mass spectrometer, and (2) where the samples are measured within the linear range, is thus a major advantage prior to statistical analysis.

MSlineaR uses a serial diluted pooled quality control sample to estimate the linear range per feature. It is designed to perform to a standard similar to a human assessing linearity in an objective way and outperforms standard linear fitting tools. MSlineaR uses a five-step process to assess linearity, detect outliers in the dilution curve, trim curves to remove inconsistencies in the linear shape and reassess the linearity to keep the maximum number of true linear peaks. Graphical output enables the user to inspect the distribution of study samples and quality controls along the linear and nonlinear parts of the dilution curve. The package is designed to be easily implemented in any standard mass spectrometry workflow in R.

Based on an untargeted data set of three batches MSlineaR rejected 3326 (~20%) non-linear features which had not been identified by all our common filter step, which includes signal-to-noise ratio, missing value and RSD filtering. In the same data set MSlineaR was able to rescue 421 features which a standard linear fitting software would have rejected due to outliers.
In this BioHub session we explain the basic requirements how to use and incorporate MSlineaR in an existing pipeline in R. We demonstrate the use on a work example, or the attendees can also bring their own data, which we can discuss.

Who would benefit?

- Researchers with basic understanding in how to use R and which are interested in how to improve quality of untargeted data in a robust way.