

3rd International Summer School on Non-Targeted Metabolomics Data Mining for Biomedical Research

Statens Serum Institut, 21th - 25th August 2023

Monday, August 21st 2023, Introduction

8.30-9.00	Registration and Coffee	
9.00-9.30	Welcome and introduction	Madeleine Ernst, Statens Serum Institut, Denmark
9.30-11.00	Mass spectrometry basics	Daniel Petras, University of Tübingen, Germany and Corinna Brungs, IOCB Prague, Czechia
11.00-11.30	Break	
11.30-12.30	Introduction to exercise and dataset	Alan Jarmusch, National Institute of Environmental Health Sciences, United States
12.30-13.30	Lunch break	
13.30-14.30	Guided tours in the Danish National Biobank and SSI's mass spectrometry facilities	Sanne Grundvad Boelt, Karina Meden Sørensen, and Bartłomiej Wilkowski, Statens Serum Institut, Denmark
14:30-14:50	Quality assurance and quality control (QA/QC) best practices: Introduction and overview	Martin Hansen, Aarhus University, Denmark
14.50-15.20	QA/QC from a company perspective	Morten Danielsen, MS-Omics, Denmark
15.20-15.30	Break	
15.30-15.50	QA/QC best reporting practices and QA/QC for neonatal dried blood spot samples	Filip Ljung, Statens Serum Institut, Denmark
15.50-16.20	Data management: security, FAIR principles and data repositories	Justin van der Hooft, Wageningen University & Research, the Netherlands
16.20-16.40	Break	
16.40-17.00	<i>Keynote:</i> Unraveling the metabolomic architecture of autism in a large Danish population-based cohort	Filip Ottosson, Statens Serum Institut, Denmark
17.00-19.00	Ice breaker	

Tuesday, August 22nd 2023, Data preprocessing

8.30-9.00	Coffee	
9.00-09.30	Introduction to data preprocessing	Steffen Heuckeroth, University of Münster, Germany Robin Schmid and Corinna Brungs, IOCB Prague, Czechia
9.30-11.00	Data preprocessing in MZmine 3 I*	Steffen Heuckeroth, University of Münster, Germany Robin Schmid and Corinna Brungs, IOCB Prague, Czechia
11.00-11.30	Break	
11.30-12.30	Data preprocessing in MZmine 3 II	Steffen Heuckeroth, University of Münster, Germany Robin Schmid and Corinna Brungs, IOCB Prague, Czechia
12.30-13.30	Lunch break	
13.30-14.00	<i>Keynote:</i> Collaborative Metabolomics Research in the Intramural Research Program of the National Institutes of Health	Alan K. Jarmusch, National Institute of Environmental Health Sciences, United States
14.00-15.00	Working on data in groups	
15.00-15.30	Break	
15.30-17.00	Working on data in groups	

* Schmid R., Heuckeroth S. et al., Integrative analysis of multimodal mass spectrometry data in MZmine 3, *Nature Biotechnology* (2023).

Wednesday, August 23rd 2023, Metabolite annotation

8.30-9.00	Coffee	
9.00-9.30	Introduction to metabolite identification	Alan Jarmusch, National Institute of Environmental Health Sciences, United States Scott Jarmusch, Technical University of Denmark Daniel Petras, University of Tübingen, Germany
9.30-10.30	Introduction to GNPS**	Alan Jarmusch, National Institute of Environmental Health Sciences, United States Scott Jarmusch, Technical University of Denmark Daniel Petras, University of Tübingen, Germany
10.30-11.00	Break	
11.00-12.30	Sirius+CSI:FingerID***	Markus Fleischauer and Fleming Kretschmer, Friedrich-Schiller University Jena, Germany
12.30-13.30	Lunch break	
13.30-14.00	Introduction to Cytoscape	Camilla Lemser, PREDICT Center of Excellence, Aalborg University and Statens Serum Institut, Denmark and Daniel Petras, University of Tübingen, Germany
14.00-14.30	Integrating Sirius+CSI:FingerID into GNPS Cytoscape network	Alan Jarmusch, National Institute of Environmental Health Sciences, United States Scott Jarmusch, Technical University of Denmark Daniel Petras, University of Tübingen, Germany
14.30-16.00	Working on data in groups	
16.00-16.30	Break	
16.30-17.30	Keynote: Advanced metabolite annotation tools and newest developments within GNPS2	Mingxun Wang, University of California - Riverside, United States (online)
17.30-19.30	Happy hour at Reffen https://refffen.dk/ (Refshalevej 167A, 1432 Copenhagen)	

** Wang M., et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, *Nature Biotechnology* **34**, 828-837 (2016); Nothias, L.F. et al., Feature-based molecular networking in the GNPS analysis environment, *Nature Methods* **17**, 905-908 (2020).

*** Dührkop K., Fleischauer M., et al., SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information, *Nature Methods* **16**, 299-302 (2020).

Thursday, August 24th 2023, Substructure discovery and advanced metabolite annotation

8.30-9.00	Coffee	
9.00-9.30	Intro to Jupyter notebooks	Abzer K. Pakkir Shah, University of Tübingen, Germany
9.30-11.00	Matchms*	Florian Huber, Düsseldorf University of Applied Sciences, Germany
11.00-11.30	Break	
11.30-12.30	MS2Query**	Florian Huber, Düsseldorf University of Applied Sciences, Germany Justin van der Hooft, Wageningen University & Research, the Netherlands
12.30-13.30	Lunch break	
13.30-14.00	Keynote: Spatial and single-cell metabolomics: technology, applications, and computational challenges	Theodore Alexandrov, European Molecular Biology Laboratory, Heidelberg, Germany & BiolInnovation Institute, Copenhagen, Denmark (online)
14.00-15.00	MS2LDA***	Justin van der Hooft, Wageningen University & Research, the Netherlands
15.00-15.30	Break	
15.30-17.00	Working on data in groups	

* Huber F., et al. matchms - processing and similarity evaluation of mass spectrometry data, *Journal of Open Source Software* **52**, 2411 (2020).

** De Jonge N., et al. MS2Query: reliable and scalable MS2 mass spectra-based analogue search, *Nature Communications* **14**, 1752 (2023).

*** Van der Hooft, J.J.J., et al., Topic modeling for untargeted substructure exploration in metabolomics, *PNAS* **113**, 13738-13743 (2016).

Friday, August 25th 2023, Statistical analysis and data visualization

8.30-9.00	Coffee	
9.00-10.00	Data normalization	Abzer K. Pakkir Shah, University of Tübingen, Germany
10.00-11.00	Multivariate and univariate statistical analyses	Filip Ottosson and Madeleine Ernst, Statens Serum Institut, Denmark
11.00-11.30	Break	
11.30-12.30	Working on data in groups	
12.30-13.30	Lunch break	
13.30-14.30	Biological interpretation and conclusions	Alan Jarmusch, National Institute of Environmental Health Sciences, United States
14.30-15.30	Working on data in groups	
15.30-16.00	Break	
16.00-17.00	Project presentations in groups	