MOVISS

Metabolomic Bio & Data 2017

Vorau, Austria

September 20-23

Book of Abstracts

Sponsors

Agilent

BRUKER

ThermoFisher SCIENTIFIC

Scientific committee

Tomáš Adam
Laboratory of metabolomics, Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry; Palacký University in Olomouc; Czech Republic

Peter Filzmoser
Institute of Statistics and Mathematical Methods in Economics; Vienna University of Technology; Austria

Karel Hron
Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science; Palacký University in Olomouc; Czech Republic

Jennifer Kirwan
Berlin Institute of Health / Max Delbrück Center for Molecular Medicine; Germany

Reza Salek
European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Cambridge; United Kingdom

Organizing committee

Peter Filzmoser and Jan Walach
Vienna University of Technology, Austria

Tomáš Adam, Eva Fišerová, David Friedecký, Alžběta Gardlo, Karel Hron, Lukáš Najdekr
Palacký University in Olomouc, Czech Republic

Contacts: Peter: +43 664 60588 1051 | David: +420 604 871 961

Registration in Education Center: Sep 20, 18:30-19:30 and Sep 21, 8:00 - 8:45

Wednesday, Sep 20

19:30 - 21:00	Welcome reception in the „Fuerstenzimmer“ of the Education Center - with buffet
21.00 - 21.30	Organ concert
Afterwards	Get together at „Clubraum“ of the Education Center

Thursday, Sep 21

8:45 - 9:00	Opening of MOVISS: Peter Filzmoser and Tomáš Adam

I. From samples to peaks Chairman: Reza Salek and Jennifer Kirwan

9:00 - 9:40	**I1: Yanes O., From spectrometric data to metabolic networks: An integrated quantitative view of cell metabolism**
9:40 - 10:00	Discussion
10:00 - 10:40	**I2: Burgess K., A communications problem: Context in metabolomics**
10:40 - 11:00	Discussion
11:00 - 11:30	Coffee break and poster viewing
11:30 - 11:40	O1: Mičová K., LC/MS targeted metabolomics based on HILIC aminopropyl stationary phase
11:40 - 11:50	O2: Najdekr L., Impact of high resolution of mass spectrometry in metabolomics
11:50 - 12:00	O3: Narduzzi L., Characterization of the VOCs profile of alpine grass fermentation used in “phytothermotherapy”
12:00 - 12:10	O4: Spicer R., Compliance with MSI Guidelines in Public Metabolomics Repositories
12:10 - 13:00	Discussion
13:00 - 14:00	Lunch

Thursday, Sep 21

II. From peaks to numbers Chairman: David Friedecký and Karel Hron

14:00 - 14:40	**I3: Neuman S., From numbers to patterns and back to structures**
14:40 - 15:00	Discussion
15:00 - 15:40	**I4: Walczak B., Pre-processing of metabolomic data as a challenging enterprise**
15:40 - 16:00	Discussion
16:00 - 16:30	Coffee break and poster viewing
16:30 - 16:40	O5: Guitton Y., Referencing LC-MS, GC-MS, and NMR workflows on the W4M Galaxy infrastructure for reproducible metabolomics data analysis
16:40 - 16:50	O6: Hartler J., An universal flexible tool for automated LC-MS/MS profiling
16:50 - 17:00	O7: Beirnaert Ch., Speaq 2.0: Large scale NMR metabolomics data analysis made easy
17:00 - 17:20	C1: Kamleh A., Complete small molecule research and structure identification in a Next Generation Platform
17:20 - 18:00	Discussion
18:15 - 19:00	Guided tour of the Monastery
19:30 - 22:00	Dinner at Kutscherwirt

Friday, Sep 22

III. From numbers to pictures Chairman: Peter Filzmoser and Lukáš Najdekr

9:00 - 9:40	**I5: Smilde A., Numerical Representations of Metabolic Systems**
9:40 - 10:00	Discussion
10:00 - 10:30	Coffee break and poster viewing
10:30 - 10:40	O8: Gardlo A., Compositional data analysis used in metabolomics
10:40 - 10:50	O9: Hoffmann I., Robust and sparse estimation methods for high dimensional linear and logistic regression
10:50 - 11:00	O10: Schroeder F., On an exact nonparametric test for class separability for the purpose of filter - type model selection
11:00 - 12:00	Discussion
12:00 - 13:00	Lunch

IV. From pictures to understanding Chairman: Tomáš Adam

13:00 - 13:40	**I6: Zamboni N., The journey from data to testable hypotheses**
13:40 - 14:00	Discussion
14:00 - 14:30	Coffee break and poster viewing
14:30 - 14:40	O11: Coene K., Next generation metabolic screening: application of metabolomics for diagnosis of inborn errors of metabolism
14:40 - 14:50	O12: Friedecký D., LC/MS metabolomics in diagnosing and treatment of rare diseases
14:50 - 15:00	O13: Hertel J., Refining risk factors via metabolomics – a general methodology and its application

Friday, Sep 22

15:00 - 15:10	O14: Meoni G., A serum metabolomic analysis of HCV - infected patients successfully treated with IFN - free DAA regimens
15:10 - 15:20	O15: Vignoli A., Serum metabolomic profiles identify ER - positive early breast cancer patients at increased risk of disease recurrence: a multicentre population study
15:20 - 16:00	Discussion
16:00 - 16:30	Coffee break and poster viewing
16:30 - 16:50	C2: Miller Ch., Investigation of pyrazinamide mechanism of action for tuberculosis using metabolomics
16:50 - 17:10	C3: Neuweger H., Using non-targeted high resolution LC-QTOF profiling to characterize metabolic responses of Nicotiana attenuata during infection with Rhizophagus irregularis
17:10 - 18:00	Discussion
18:00 - 18:15	Closing the workshop
18:30 - 19:30	Walk to „Mostschank Kuchlbauer“ (approx. 1 hour) - alternative bus transfer if raining
19:30 - 22:30	Dinner at Kuchlbauer
Afterwards	Walk back or bus transfer

Saturday, Sep 23

9:00-18:00	Hiking tour to Teichalm; in case of rain: cultural tour
19:30-22:30	Dinner at Vorauerhof

Wednesday, Sep. 20: Welcome reception & Organ Concert (Mag. Gerhard Filsegger)

The welcome reception with a complimentary buffet will be held in the impressive Fuerstenzimmer of the Education Center. Afterwards, you will have an opportunity to attend an organ concert by Mag. Gerhard Filsegger, the organist of the monastery. The concert will take place in the monastery church. The organ was completely refurbished in 2013.

Thursday, Sep. 21: Guided tour through the monastery

The tour will start at 18:15. Meeting point will be in front of the church. Besides the church, you will have an opportunity to visit the sacristy, with a very impressive painting by Hackhofer from the beginning of the 18th century. Another part will be the visit of the library, which contains almost 18 000 books (the oldest ones are from the 11th century). The tour guide will be Herr Patrick, the youngest priest in Vorau. After the guided tour, the common dinner will be served at the Kutscherwirt.

Friday, Sep. 22: Local farm dinner at Mostschank Kuchlbauer

After the end of the scientific program, we will walk to Mostschank Kuchlbauer. We will leave at 18:30 from the monastery entrance (people accommodated at Kutscherwirt or Vorauerhof may join us on the way). The walk is approximately 3 km long. Optionally, you may use a bus service to travel on the way there and/or back, if you want to use it, please inform the organizers in advance. "Mostschank" means in German "cider house". You will have the opportunity to taste their homebrewed apple cider and for the cooking their use products which they produce at the farm (a.k.a. the-most-best-super-organic-food which give you superpowers and makes you super-healthy J). Kuchlbauer's apple cider is excellent, famous for its quality and rewarded by many awards.

Saturday, Sep. 23: Teichalm

We meet at 9am in front of the Monastery to take a bus to Teichalm. In case of very bad weather (something like the doomsday weather or when fish-and-chips will be raining), we will take a cultural tour instead. Nevertheless, the use of hiking shoes and warm, waterproof clothes will be an advantage for sure. As they say: "There is no bad weather, just bad clothing". It will take around 1 hour to reach the Sommeralm by bus.

The hike will start at Point 3 on the map, and we will walk 8 km to Point 11. There are several hills in between, like Kulmkogel, Mooskogel, Nestlbodnerkogel, Grubbauerkogel, Heulantsch, which we need to pass. If this seems to be too exhaustive, there is the option of a shorter hike, since the bus can carry you to another starting point.

The total walking time might be 4-5 hours. In the middle of the tour, we will visit a small chalet Stoahandhuette, where drinks and cold food will be served. After arriving at Teichalm, we can take a coffee in one of the restaurants, and then leave to Vorau, where dinner will be served at Vorauerhof at 19:30.

Adam Tomáš	Palacký University Olomouc, Olomouc, Czech Republic
Aksenov Alexander	University of California San Diego, San Diego, United States
Battista Anna	Owlstone Medical Ltd, Cambridge, United Kingdom
Beirnaert Charlie	University of Antwerp, Antwerp, Belgium
Brodinová Šárka	Vienna University of Technology, Vienna, Austria
Burgess Karl	University of Glasgow, Glasgow, United Kingdom
Büscher Jörg	MPI for Immunbiology and Epigenetics, Freiburg, Germany
Čechová Eliška	RECETOX, Brno, Czech Republic
Cechová Monika	Palacký University Olomouc, Olomouc, Czech Republic
Chae Young Kee	Sejong University, Seoul, Korea
Cichon Morgan	The Ohio State University, Columbus, United States
Coene Karlien	Radboud University Medical Centre, Nijmegen, Netherlands
Damrath Alexandra	Agilent, Vienna, Austria
Darnhofer Barbara	Omics Center Graz, Graz, Austria
El Abiead Yasin	Uni Wien, Wien, Austria
Fesl Jan	BC CAS, České Budějovice, Czech Republic
Filzmoser Peter	Vienna University of Technology, Vienna, Austria
Fišerová Eva	Palacký University Olomouc, Olomouc, Czech Republic
Friedecký David	Palacký University Olomouc, Olomouc, Czech Republic
Gardlo Alžběta	Palacký University Olomouc, Olomouc, Czech Republic
Gloaguen Yoann	Max Delbrück Center, Berlin, Germany
Guitton Yann	LABERCA, Nantes, France
Hartler Juergen	TU Graz, Graz, Austria
Hertel Johannes	University Medicine Greifswald, Greifswald, Germany
Hoffmann Irene	Vienna University of Technology, Vienna, Austria
Hron Karel	Palacký University Olomouc, Olomouc, Czech Republic
Jáčová Jaroslava	Palacký University Olomouc, Olomouc, Czech Republic
Jakubec Pavel	Faculty of Pharmacy, Charles University, Prague, Czech Republic
Johnsen Lea	MS-Omics, Frederiksberg, Denmark
Kamleh Anas	Thermo Scientific, Hägersten, Sweden
Karlíková Radana	Palacký University Olomouc, Olomouc, Czech Republic
Kirwan Jennifer	Berlin Institute of Health/Max Delbrück center, Berlin, Germany
Klavins Kristaps	CeMM, Vienna, Austria
Koellensperger Gunda	BOKU Wien, Wien, Austria
Kouřil Štěpán	Palacký University Olomouc, Olomouc, Czech Republic
Krizek Elisabeth	Owlstone Medical Ltd, Cambridge, United Kingdom
Mádrová Lucie	Palacký University Olomouc, Olomouc, Czech Republic
Majek Peter	CeMM Forschungszentrum GmbH, Vienna, Austria

Martínez Avila Jose Carlos	Clinical Pharmacology Department. La Paz, Madrid, Spain
Meoni Gaia	CERM, University of Florence, Prato, Italy
Miller Christine	Agilent, Wien, Austria
Mičová Kateřina	Palacký University Olomouc, Olomouc, Czech Republic
Moos Martin	BC CAS, České Budějovice, Czech Republic
Najdekr Lukáš	Palacký University Olomouc, Olomouc, Czech Republic
Narduzzi Luca	Fondazione Edmund Mach, San Michele all'Adige, Italy
Neumann Steffen	IPB Halle, Halle, Germany
Neuweger Heiko	Bruker Daltonik GmbH, Bremen, Germany
Opialla Tobias	Berlin Institute of Health, Berlin, Germany
Rendlová Julie	Palacký University Olomouc, Olomouc, Czech Republic
Rohmatul Fajriyah	Universitas Islam Indonesia, Yogyakarta , Indonesia
Rosadi Dedi	Universitas Gadjah Mada, Yogyakarta , Indonesia
Salek Reza	EMBL-EBI, Cambridge, United Kingdom
Schafferer Simon	Biocrates Life Sciences AG, Innsbruck, Austria
Schroeder Fabian	Austrian Institute of Technology, Vienna, Austria
Šimek Petr	BC CAS, České Budějovice, Czech Republic
Široká Jitka	Palacký University Olomouc, Olomouc, Czech Republic
Smilde Age	University of Amsterdam, Amsterdam, The Netherlands
Spáčil Zdeněk	RECETOX, Brno, Czech Republic
Spicer Rachel	EMBL-EBI, Cambridge, United Kingdom
Štefelová Nikola	Palacký University Olomouc, Olomouc, Czech Republic
Schwaiger Michaela	Uni Wien, Wien, Austria
Trautwein Christoph	University Hospital Tuebingen, Tuebingen, Germany
Václavík Jan	IMTM - UPOL, Olomouc, Czech Republic
Vignoli Alessia	Magnetic Resonance Center – CERM, Sesto Fiorentino, Italy
Vrobel Ivo	Palacký University Olomouc, Olomouc, Czech Republic
Wachsmuth Christian	Philip Morris International, Schwaz, Austria
Walczak Beata	University of Silesia, Katowice, Poland
Walach Jan	Vienna University of Technology, Vienna, Austria
Wei Chunyue	ETH Zurich, Zurich, Switzerland
Wilkinson Daniel	University of Nottingham, Derby, United Kingdom
Yanes Oscar	Universitat Rovira i Virgili & CIBERDEM, Reus, Spain
Yang Heejung	Kangwon National University, Chuncheon, South Korea
Zamboni Nicola	ETH Zurich, Zurich, Switzerland
Zareie Ashkan	BIOCEV, Prague, Czech Republic
Zhu Dan	ETH Zurich, Zurich, Switzerland

Invited Speakers

Karl Burgess

Oscar Yanes

Steffen Neuman

Beata Walczak

Age K. Smilde

Nicola Zamboni

I1: Oscar Yanes

From spectrometric data to metabolic networks: An integrated quantitative view of cell metabolism

Oscar Yanes received his Ph.D. degree in Biochemistry from the Autonomous University of Barcelona (Spain). In 2007 he became Research Associate in Gary Siuzdak's lab at The Scripps Research Institute (La Jolla, USA). Since 2011 he coordinates the metabolomics platform of the Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders (CIBERDEM), he is affiliated member at the IRB Barcelona and assistant professor at the Universitat Rovira i Virgili where he also leads his own research group (www.yaneslab.com). He has long experience in developing new technologies, methods and applications in mass spectrometry-based metabolomics. His lab now focuses on understanding metabolic dysregulations in disease through integrating MS and NMR-based metabolomics with other omic platforms.

Presentation will focus on major principles of metabolomics experimental design, strengths and weaknesses of metabolomics analytical approaches, and factors that impact upon subsequent analysis of the data. This will include discussing about current trends and challenges in metabolomics, with particular emphasis on experimental and computational approaches based on LC- and GC-mass spectrometry (MS) and nuclear magnetic resonance (NMR) to enable a comprehensive analysis of cellular metabolites.

I2: Karl Burgess

A communications problem: Context in metabolomics

I am a Senior Research Fellow and Head of Metabolomics at the University of Glasgow, with a focus on methodological and bioinformatic development of technologies for metabolomics and proteomics, especially in the context of bacterial adhesion and biofilm development. With a background in both biology and computing science, I found my niche in biological mass spectrometry, where my expertise in processing and interpreting large datasets in a biological context became invaluable in the rapidly developing field of proteomics, and, later, metabolomics, with several highly cited papers in the field. My personal research focuses primarily on improved quality control, assisted interpretation of complex 'omics data and improved metabolome coverage. Highlights of this research include publication of a combined HILIC/RP method for increasing metabolome coverage and the Polyomics Metabolomics Pipeline (PiMP) software platform for metabolomic analysis. In light of the importance of 'omics to modern biological research, I also have made significant contributions to: regenerative medicine in collaboration with Prof. Matt Dalby; host-pathogen interactions with Dr. Andy Roe and biochemical parasitology with Prof. Mike Barrett. My profile was recently highlighted in 'The Analytical Scientist'.

My presentation will focus on biological interpretation of metabolomics datasets: from the perspective of a core facility, how do you interact with collaborators, how much interpretation can and do we provide, and how that interaction is affected by statistics, data and ambiguity, especially in the context of identification in untargeted analysis? I will focus primarily on LC-MS but with some context from NMR and GCMS.

I3: Steffen Neuman

From numbers to patterns and back to structures

Steffen Neumann studied computer science and bioinformatics at Bielefeld University, and now his group focuses on the development of tools and databases for metabolomics and computational mass spectrometry. The group develops algorithms for data processing of metabolite profiling experiments, which are available in several Open Source Bioconductor packages, and addresses the identification of unknowns in mass spectrometry data. The IPB is member of the MassBank consortium and operated the first MassBank server in Europe. The MetFrag and MetFusion tools allow the identification of compounds where no reference spectra are available.

During the MOVISS meeting, I am looking forward to discussions on large-scale metabolite profiling, and large-scale characterisation of metabolite features from MS data. Last-but-not-least, the value of the data drastically increases if it is shared in standardised formats, so that others can grab it, re-process and make the most of the data.

I4: Beata Walczak

Pre-processing of metabolomic data as a challenging enterprise

In metabolomics studies, we are faced with multiple challenges associated with the properties of the studied samples (complexity, diversity, wide concentration range of metabolites, unknown overall concentration, etc.) as well as with the properties of the applied analytical platforms (NMR, LC-MS, GC-MS, UPLC-MS, FTIR, etc). To focus data analysis on the induced data variation, it is necessary to remove all the remaining undesired effects at the data pre-processing step.

The pre-processing step usually involves an enhancement of individual signals via their de-nosing and background elimination, as well as pre-processing of a signal set such, as alignment, normalization, scaling/transformation, etc. In many studies, data pre-processing is performed based on a trial-and-error approach, whereas it should be based on the data characteristics (determined by the samples and instrumental signals characteristics) and on the final goal of the data analysis. Although construction of exact rules seems impossible, some indications and diagnostics tools can help avoiding evident abuse of the existing methods [1].

Different aspects of data pre-processing and their influence on data analysis are going to be illustrated by the simulated and real data sets. Additionally, different aspects of data organization and presentation, and their influence on data analysis will be discussed.

Reference: 1. P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers?, J. Chromatography A, 1362 (2014) 194-205

I5: Age K. Smilde

Numerical Representations of Metabolic Systems

Age K. Smilde is full professor of Biosystems Data Analysis at the Swammerdam Institute for Life Sciences at the University of Amsterdam and is also affiliated with the Academic Medical Centre of that same university. As of June 1, 2013 he holds a part-time position as professor of Computational Systems Biology at the University of Copenhagen. His research interest focuses on two topics: data fusion and network inference. Data fusion concerns integrating functional genomics data and fusing data with prior biological knowledge. The network topic encompasses network reconstruction, deriving network properties from time-resolved data and computational network analysis. He has published more than 200 peer-reviewed papers and has been the Editor-Europe of the Journal of Chemometrics during the period 1994-2002. He chaired the 1996 Gordon Research Conference on Statistics in Chemistry & Chemical Engineering. In 2006, he received the Achievements in Chemometrics Award of the Eastern Analytical Symposium.

Data is not the same as numbers. Instruments generate numbers but before these become meaningful they have to be transformed to data. That transformation involves theory (both instrumental as well as biological). Hence, data is loaded with theory and this has repercussions for the subsequent data analysis and interpretation of the results. These notions will be illustrated with examples from metabolomics. There are many open questions which will also be discussed.

I6: Nicola Zamboni

The journey from data to testable hypotheses

Nicola Zamboni earned his PhD in Biotechnology in the group of Jay Bailey at ETH Zurich with a thesis on metabolic engineering and 13C metabolic flux analysis. In 2004 he moved as a postdoctoral fellow to the Stanford Genome Technology Center, where he developed and applied metabolomics-based approaches for unraveling metabolic changes in eukaryotic cells. Since late 2005, he's a group leader and independent PI at the Institute of Molecular Systems Biology of ETH Zurich.

The lab focuses on the development of mass spectrometry and computational methods to investigate cellular metabolism from bacteria to human cells in a variety of questions related to systems biology, preclinical research, biotech, and healthcare. It pursues a primarily data-driven approach largely based on mass spectrometry, i.e. high-throughput and non-targeted metabolomics and 13C-metabolic flux analysis. These data are integrated with a variety of computational approaches to support data mining, integration, interpretation, and prediction.

At the MOVISS meeting, the discussion will focus on the interpretation of metabolome data, i.e. on how to test testable hypotheses from typically complex, ill-dimensioned, non-quantitative, sparse and noisy data.

Short oral presentations

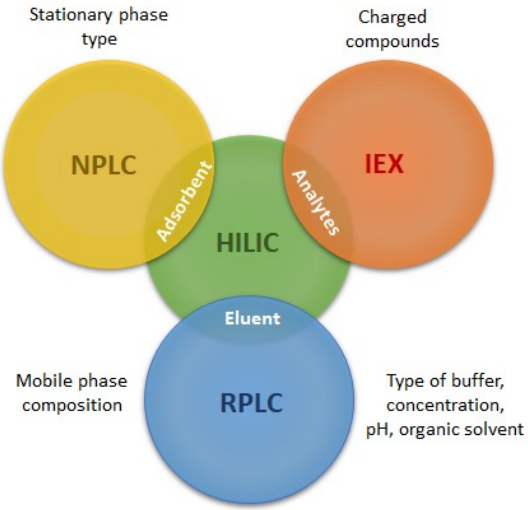
O1: LC/MS targeted metabolomics based on HILIC aminopropyl stationary phase

Kateřina Mičová[1], David Friedecký[2,3], Radana Karlíková[1,2], Alžběta Gardlo[1,2], Hana Janečková[3], Tomáš Adam[1,2,3]

[1]Department of Clinical Biochemistry, University Hospital Olomouc, I. P. Pavlova 6, 775 20 Olomouc, Czech Republic; [2]Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacký University Olomouc, Hněvotínská 5, 779 00 Olomouc, Czech Republic; [3]Laboratory for Inherited Metabolic Disorders, Faculty of Medicine and Dentistry, Palacký University Olomouc, I. P. Pavlova 6, 775 20 Olomouc, Czech Republic

Metabolomic analysis focuses to reveal differences of metabolic profiles between physiological and pathological conditions of an organism. Metabolites are present in biological matrices in wide range of concentrations, therefore sensitive techniques for their identification are needed. High performance liquid chromatography using aminopropyl column (Luna 3 µm NH2, 2x100 mm, Phenomenex) coupled to tandem mass spectrometry (Sciex, Foster City, CA, USA) represents suitable tool for analysis of polar and semi-polar metabolites. Our method enables analysis of 354 intermediary metabolites (acylcarnitines, amino acids, organic acids, nucleotides, etc.) that can be found in various biological matrices. Metabolites, detected by multiple reaction monitoring mode, are quantified using MultiQuant 3.0 software (AB Sciex, Foster City, CA) and statistically evaluated in R program language (version 3.1.2). Targeted metabolomic analysis can be used for study of biochemical changes caused by metabolic diseases (inherited or acquired), poisoning and monitoring drug effect on metabolome. Various clinical application of targeted metabolomics (e.g. tauopathy in transgenic rats, patients with chronic myeloid leukemia) will be presented.

This work was supported by the Ministry of Health of the Czech Republic [AZV 15-31604A].

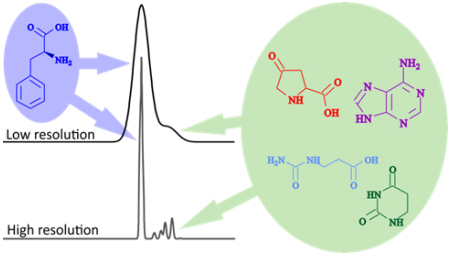


O2: Impact of high resolution of mass spectrometry in metabolomics

Lukáš Najdekr[1], David Friedecký[1], Ralf Tautenhahn[2], Tomáš Pluskal[3], Junhua Wang[2], Yingying Huang[2], Tomáš Adam[1]

[1]Laboratory of Metabolomics, Institute of Molecular and Translational medicine, University Hospital and Palacký University in Olomouc, Hněvotínská 5, 775 15 Olomouc, Czech Republic; [2]Thermo Fisher Scientific, 355 River Oaks Pkwy, San Jose, 95134 CA, USA; [3]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142-1479, USA

Modern separation methods in conjunction with high resolution accurate mass (HRAM) spectrometry can provide an enormous number of features characterized by exact mass and by chromatographic behavior. The aim of this work is to describe influence of mass resolving power on exact mass spectrometry profiling experiments covering the whole picture of a sample. Higher mass resolving power requires usually longer scanning times and thus less data points across the peak are acquired. This could present an issue in quantification, component detection and possible problem with deconvolution. Experiment was addressed by *In silico* calculations and by analyzing human plasma by LC-HRMS at different resolutions.



From metabolic databases (HMDB, LipidMaps, KEGG), a list of compounds (41 474) was compiled and potential adducts and isotopes were calculated (622 110 features). The number of distinguishable masses was calculated for up to 3840k resolution. To evaluate these models, human plasma samples were analyzed by LC-HRMS on an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific, CA, USA) at resolving power settings of 15k (7.8 Hz) up to a maximum of 480k (1.2 Hz). Software XCMS 1.44, MZmine 2.13.1, and Compound Discoverer 2.0.0.303 were used for evaluation. In plasma samples, the number of detected features increased sharply up to 60k in both positive and negative mode. However, beyond these values, it either flattened out or decreased owing to technical limitations. In conclusion, the most effective mass resolving powers for profiling analyses of metabolite rich biofluids on the Orbitrap Elite were around 60 000– 120 000 fwhm to retrieve the highest amount of information. The region between 400–800 m/z was influenced the most by resolution.

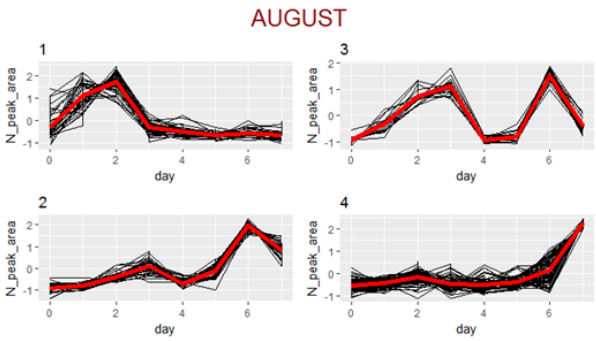
The infrastructural part of this project (Institute of Molecular and Translational Medicine) was supported from NPU I (LO1304) and Czech Science Foundation Grant 15-34613L.

Reference: Najdekr, L. *et al.* Influence of mass resolving power in orbital ion-trap mass spectrometry-based metabolomics. *Anal. Chem.* **88,** acs.analchem.6b02319 (2016).

O3: Characterization of the VOCs profile of alpine grass fermentation used in "phytothermotherapy"

Luca Narduzzi[1], Elena Franciosi[1], Kieran Tuohy[1] & Fulvio Mattivi[1]

[1]*Research and innovation center, Fondazione Edmund Mach, Via E. Mach, 1 San Michele all'Adige, 38010, (TN) Italy*



Phytothermotherapy treats patients aching from rheumatic diseases in hot grass baths, obtained through spontaneous natural fermentation. Despite the uncontrolled fermentation process, phytothermotherapy demonstrated to be very effective against multiple rheumatic diseases. The mechanisms of action are unknown: while the fermentation heat (up to 65°) is responsible for the short-term effects, it cannot explain the long term improvement of the patients (up to 12 months). Despite its marketing attractiveness, the "natural" process of the grass baths fermentations has limitations; in particular, the process shows high variability in terms of heat produced and odor quality, undermining the reproducibility and the pleasantness of the cure. VOCs profiling has been widely used to evaluate the quality of the fermentations in many food products (wine, beer, bread, cheese); through this approach, it is possible to establish the metabolism of such "living food organisms". Similarly, in this work, we used it to estimate the quality of the grass baths and to determine whether their metabolism is going toward an undesired direction. Furthermore, as mechanisms of action are unknown, among the VOCs produced there might be some candidate metabolite enhancing the effects of the cure, as reported in other cures for some plant/bacteria volatiles. Using passive diffusion samplers and TD-GCxGC-TOF-MS, we established a method to characterize VOCs profile of the grass baths. Several hundreds of metabolites have been detect in every single run. Through a "fuzzy" clustering approach, we discarded the noisy variables and grouped the remaining volatiles into four time trends. Interestingly, the trend groups were reflecting the main metabolic processes: the plants' stress response (trend 1), the microbial fermentation (trend 2) and start of the putrefaction (trend 4). While the identification of the metabolites gave us a list of several candidates having health-related effects, their trend profiling gave a clear snapshot of the metabolic status of the fermentation, allowing to critically evaluate the grass baths outcome. For example, we obtained a putative list of metabolites responsible for the insurgence of unpleasant odours related to the rapid growth of undesired bacteria (clostridia). In future, this metabolomics approach can be used to evaluate how the setting of the parameters in the fermentation protocol (like: starter microbial cultures, grass maturation, freezing and movement, and others) may improve or worsening the quality of the grass baths.

O4: Compliance with MSI Guidelines in Public Metabolomics Repositories

Rachel Spicer[1], Reza Salek[1], Christoph Steinbeck[1,2]

[1]European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD; [2]Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller-University, Jena, Germany

Since publication of the Metabolomics Standards Initiative (MSI) reporting standards[1] in 2007, the metabolomics community has matured. There is now greater understanding of both environmental and genetic factors with significant effects on the metabolome. The MSI consisted of five working groups, reflecting different areas of metabolomics requiring standards. The MSI biological context metadata working group was divided into four subgroups: mammalian and *in vivo*, microbial and *in vitro*, plant and environmental. The two general-purpose metabolomics data repositories MetaboLights and Metabolomics Workbench were developed to comply with MSI guidelines for minimal metadata reporting. We selected three model organisms, *Homo sapiens*, *Mus musculus* and *Arabidopsis thaliana*, to assess compliance with the MSI biological context guidelines. Studies involving these species cover the mammalian and *in vivo*, microbial and *in vitro,* and plant guidelines. Compliance with the environmental reporting standards was not assessed, due to the low number of environmental studies on the repositories. The plant minimal reporting standards were the most complied with (average compliance = 50.0%) and the microbial/ *in vitro* guidelines were the least (average compliance = 16.5%). Unlike many of the most successful reporting standards[2], the MSI standards were developed using a top-down approach. The lack of compliance with the MSI guidelines shows that they do not reflect the needs of the metabolomics community. We propose that the MSI guidelines should be revisited and revised, consulting with the community, data curators, publishers and funders.

The work was supported by an EMBL PhD studentship.

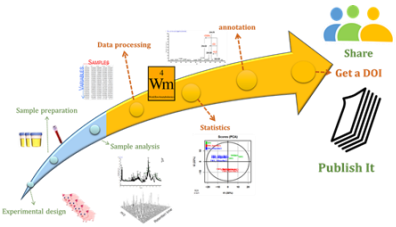
Reference 1. Fiehn, O. *et al.* The metabolomics standards initiative (MSI). *Metabolomics* **3,** 175–178 (2007). 2. Brazma, A., Krestyaninova, M. & Sarkans, U. Standards for systems biology. *Nat. Rev. Genet.* **7,** 593–605 (2006).

O5: Referencing LC-MS, GC-MS, and NMR workflows on the W4M Galaxy infrastructure for reproducible metabolomics data analysis

Yann Guitton[1,§], Marie Tremblay-Franco[2,§], Gildas Le Corguillé[3], Jean-François Martin[2], Mélanie Pétéra[4], Pierrick Roger-Mele[5], Alexis Delabrière[5], Sophie Goulitquer[6], Misharl Monsoor[3], Christophe Duperier[4], Cécile Canlet[2], Rémi Servien[2], Patrick Tardivel[2], Christophe Caron[7], Franck Giacomoni[4], and Etienne A. Thévenot[5]

[1]LUNAM Université, Oniris, LABERCA, Nantes, F-44307, France; [2]Toxalim, Université de Toulouse, INRA, INP-ENVT, INP-EI-Purpan, Univ. Toulouse 3 Paul Sabatier, 31027 Toulouse, France; [3]UPMC, CNRS, FR2424, ABiMS, Station Biologique, 29680, Roscoff, France; [4]INRA, UMR 1019, PFEM, 63122, Saint Genes Champanelle, France; [5]CEA, LIST, Laboratory for Data Analysis and Systems' Intelligence (LADIS), MetaboHUB, F-91191 Gif-sur-Yvette, France; [6]INSERM/UBO, U1078, Brest; [7]INRA, Ingenum, Toulouse, France

Metabolomics data analysis is a complex, multistep process, which is constantly evolving with the development of new analytical technologies, mathematical methods, and bioinformatics tools and databases. The Workflow4Metabolomics Galaxy online infrastructure (W4M, http://workflow4metabolomics.org) provides a unique centralized, user-friendly, and high-performance environment to build, run, and share metabolomics workflows for LC-MS, GC-MS, and NMR technologies[1]. W4M now supports the publication of workflows online: for the first time, users have the opportunity to get a reference ID for their study. The whole workflow (i.e. the modules and the parameter values) and the input/output data can thus be publicly accessed and cited with a simple Digital Object Identifier (DOI). By referencing your history, you get a permanent DOI which you can cite in your publications[2]. Making your workflows and associated data available to the community is essential to demonstrate the value and the reproducibility of your analysis (e.g., to reviewers). This initiative follow the four FAIR principles "Findability, Accessibility, Interoperability, and Reusability" of data, algorithms and tools[3]. As for raw data, journal editors will increasingly require that the process of generating the results (code, parameter values, output data) is made available on reference repositories. Funding agencies such as European Programmes also require that the generated data are made public. Finally, by sharing your analysis, you get the opportunity to receive feedback on your results, be cited, and initiate new collaborations. In addition, it offers new opportunities for training. Workflow4Metabolomics thus not only offers a unified, evolving environment for metabolomics data analysis, but also should become the reference repository for shared workflows. The work was supported by MetaboHub and IFB.



References: 1. Giacomoni et al. (2015) 'Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics', *Bioinformatics*, 31: 1493-95. 2. http://workflow4metabolomics.org/referenced_W4M_histories; 3. Wilkinson et al. (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018.

O6: An universal flexible tool for automated LC-MS/MS profiling

Jürgen Hartler[1,2,3], Alexander Triebl[2], Andreas Ziegl[1], Martin Trötzmüller[2,3], Gerald N Rechberger[3,4], Oana A Zeleznik[5,6], Kathrin A Zierler[4], Federico Torta[7], Amaury Cazenave-Gassiot[7], Markus R Wenk[7], Alexander Fauland[8], Craig E Wheelock[8], Aaron M Armando[9], Oswald Quehenberger[9], Qifeng Zhang[10], Michael J O Wakelam[10], Guenter Haemmerle[4], Friedrich Spener[4,11], Harald C Köfeler[2,3], Gerhard G Thallinger[1,3]

[1]Institute of Molecular Biotechnology, Graz University of Technology, 8010 Graz, Austria; [2]Center for Medical Research, Medical University of Graz, Austria; [3]Omics Center Graz, BioTechMed-Graz, Austria; [4]Department of Molecular Biosciences, NAWI Graz, University of Graz, Austria; [5]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, U.S.A.; [6]Department of Medicine, Harvard Medical School, Boston, Massachusetts, U.S.A.; [7]Singapore Lipidomics Incubator, Department of Biochemistry, National University of Singapore, Singapore; [8]Department of Medical Biochemistry and Biophysics, Division of Physiological Chemistry 2, Karolinska Institutet, Stockholm, Sweden; [9]School of Medicine, University of California San Diego, California, U.S.A.; [10]The Babraham Institute, Babraham Research Campus, Cambridge, U.K.; [11]Department of Molecular Biology and Biochemistry, Medical University of Graz, Austria

Appropriate methods for analyzing metabolites in high-throughput fashion are needed in fundamental and applied research to accelerate biomedical, clinical and nutritional research including intervention studies. Liquid chromatography-linked tandem mass spectrometry (LC-MS) enables quantitative measurements of hundreds to thousands of metabolites in complex mixtures. In addition, MS/MS spectra are required to obtain structural information for more detailed characterization of the metabolite. For lipid profiling, automated lipid annotation relies on spectral libraries up to date. Yet, variables such as the type of mass spectrometer, the collision energy applied, the type of adduct, and the charge state influence heavily the pattern of lipid MS/MS spectra.

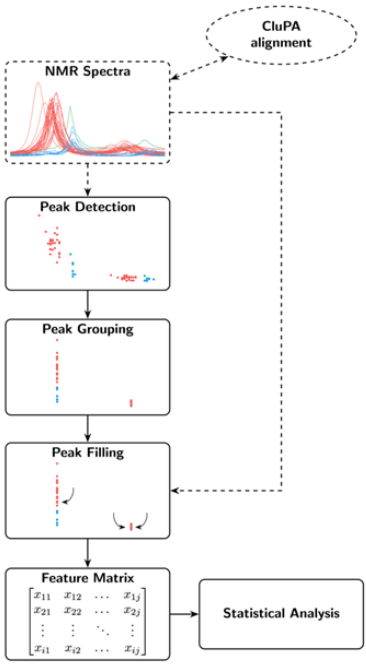
To solve these problems, we have developed Lipid Data Analyzer 2 (LDA 2), enabling automated annotation and quantitation of lipid species and of their molecular structures in high-throughput LC-MS/MS data. The software interprets spectra based on intuitive *decision rule sets*, and flexibly accommodates changes in fragmentation behavior. Platform independence was proven in experiments with eight different mass spectrometric set-ups, comprising low- and high-resolution instruments at various collision energies and use of several adduct ions. With LDA 2, both, the number of correctly identified lipid molecular species and the reliability of the results increased compared to present state-of-the-art. Moreover, this approach allows for *de novo* detection of unanticipated structures, and is easily extensible to other metabolite classes by a graphical user interface providing direct visual feedback on acquired spectra.

O7: Speaq 2.0: Large scale NMR metabolomics data analysis made easy

Charlie Beirnaert[1,2], Pieter Meysman[1,2], Nina Hermans[3], Sandra Apers[3], Luc Pieters[3], Adrian Covaci[4], Kris Laukens[1,2]

[1]Advanced Database Research and Modelling (ADReM), Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium; [2]Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp/Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem; [3]Natural Products & Food Research and Analysis (NatuRA), Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium; [4]Toxicological Center, Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium

Numerous present day metabolomics experiments rely on both LC-MS and NMR spectroscopy data to fully quantify the available information. Since LC-MS data analysis is inherently more complex compared to NMR, most advanced tools are developed for LC-MS. Yet often NMR data analysis takes up a large portion of time, since there is a lot of manual intervention involved in spectral processing. This is a consequence of the default method used to tackle these spectra, the so-called interval methods. Spectra are divided in bins or buckets (the intervals) and the signals within one bucket are summarized to represent a single variable or feature. This approach effectively summarizes the spectra and allows for easier data processing and statistics, but introduces several problems, both in pre-processing and in following analysis, that compromise the automation potential. Large scale metabolomics studies in which hundreds of LC-MS and NMR samples have to be analyzed require automated workflows. We present a new workflow for the analysis of NMR spectra that uses wavelets to fully automate the process of converting raw spectra to peaks, which are then aligned and grouped into features. This involves minimal user interaction and since wavelets are also often used in peak picking for LC-MS data (to convert spectra into peak data with minimal information loss) this method advances the integration of LC-MS and NMR. The framework is validated by replicating the results of two papers demonstrating the speed, ease of use, and improved results. The algorithms will be made available in the speaq [1] R-Package.



Reference: [1] T. N. Vu et al. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC Bioinformatics* 12:405 (2011)

O8: Compositional data analysis used in metabolomics

Alžběta Gardlo[1,2], Karel Hron[3]

[1]Laboratory of Metabolomics, Institute of Molecular and Translational medicine, Palacký University, Olomouc, Hněvotínská 5, 779 00 Olomouc, Czech Republic; [2]Department of Clinical Biochemistry, University Hospital Olomouc, I.P. Pavlova 6, 775 20 Olomouc, Czech Republic; [3]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, 779 00 Olomouc, Czech Republic

Compositional data (or compositions for short) are multivariate observations with positive components, and they can be represented without loss of information as data with a constant sum constraint like proportions or percentages (1,2). In such a case, the sum of the compounds (parts) is not important and the only relevant information is contained in the ratios between the parts. Compositional data occur in a wide range of applications involving geochemistry, analytical chemistry, and its related fields and it is nowadays successfully applied also to the field of metabolomics (3,4).

In metabolomics, absolute values of biomarkers compared with reference ranges (data from the healthy population) is the most frequently used approach. Ratios of metabolite data are frequently analyzed in the biochemical diagnostic practice; hence relative changes are more relevant/ informative than absolute values. It suggests that metabolomic data can indeed be considered as observations carrying relative information, i.e. as compositional data.

The specific geometry for compositional data is called the Aitchison geometry, however statistical data analysis is usually carried out in the Euclidean geometry. Thus, the central idea is to express compositions from the simplex (sample space of compositions) in real coordinates and then to apply the standard multivariate methods. Three basic logratio coordinate systems are used for this purpose: additive, centered and isometric logratio coordinates. This contribution focuses on centered logratio (clr) coordinates, which are popularly used in metabolomics, because their use improves the identification of specific biomarkers. On the other hand, this approach has also some limitations which will be also discussed.

The work was supported by NPU I (LO1304), GACR 15-34613L.

Reference: 1. Aitchison, J. *The statistical analysis of compositional data*. Chapman & Hall, London, 1986. 2. Pawlowsky-Glahn, V. et al. *Modeling and analysis of compositional data*. Wiley, Chichester, 2015. 3. Gardlo, A. et al. Normalization techniques for PARAFAC modeling of urine metabolomic data. *Metabolomics* 12 (7), article 117, (2016). 4. Najdekr, L. et al. Oxidized phosphatidylcholines suggest oxidative stress in patients with medium-chain acyl-CoA dehydrogenase deficiency. *Talanta* 139, 62-66, (2015).

O9: Robust and sparse estimation methods for high dimensional linear and logistic regression

Fatma Sevinc Kurnaz[1], Irene Hoffmann[2], Peter Filzmoser[2]

[1]Department of Statistics, Yildiz Technical University, 34220 Istanbul, Turky; [2]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstraße 8 -10, 1040 Vienna, Austria

Elastic net estimators penalize the objective function of a regression problem by adding a term containing the L1 and L2 norm of the coefficient vector. This type of penalization achieves intrinsic variable selection and similar coefficient estimates for highly correlated variables. We propose a robust and sparse estimator for both, linear and logistic regression. The algorithm searches for outlier-free subsets on which the classical elastic net estimators can be applied. With a final reweighting step the statistical efficiency of the methods can be improved. We compare our methods to the classical counterparts as well as to the sparse LTS estimator, a robust regression estimator with L1 penalty. Simulation studies and real data examples demonstrate the superior performance of the proposed method.

The work was supported by grant TUBITAK 2214/A from the Scientific and Technological Research Council of Turkey and by the Austrian Science Fund (FWF), project P 26871-N20.

O10: On an exact nonparametric test for class separability for the purpose of filter-type model selection

Fabian Schroeder

Austrian Institute of Technology, Department of Molecular Diagnostics, Vienna, Austria

Consider the well-known problem of supervised classification with one variable and two classes. This method can be used for the purpose of variable selection on a univariate basis, which is often referred to as filter-type model selection. In most applications it is applied when the dimension of the sample space in high or ultra high to drastically reduce the number of variables which subsequently allows the use of more sophisticated model selection methods.

In this paper I will introduce and discuss a statistical test based on the prediction error of a nonparametric classifier for this purpose. This approach has several advantages. First, it naturally accounts for the operating conditions of the classification task, comprising the misclassification costs of the class distribution in the overall population. Many common filter statistics, e.g. the t-test disregard the operation conditions as well as differences in the variances between the class conditionals, which may lead to false conclusions. Secondly, the nonparametric approach guarantees that the differences in the test statistic do not only reflect the differences in the extent to which the variables satisfy the distributional assumption. Thirdly, it is possible to obtain the exact finite sample distribution of the test statistic under the assumption of equal class conditionals. The exact significance of the classification result, as given by the p-value, is very handy since it gives an absolute criterion for the filter rather than just a relative one, which only allows ranking the variables. Furthermore it also allows adjusting the p-values for the number of tests performed.

O11: Next generation metabolic screening: application of metabolomics for diagnosis of inborn errors of metabolism

Karlien Coene[1], Leo Kluijtmans[1], Udo Engelke[1], Brechtje Hoegen[1], Ed van der Heeft[1], Siebolt de Boer[1], Marleen Huigen[1], Jasper Engel[2], Maartje van de Vorst[3], Christian Gilissen[3], Ron Wevers[1]

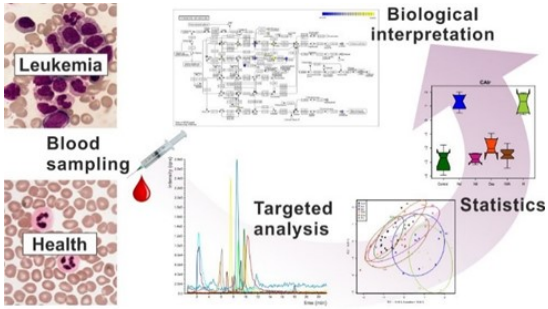
[1]Dept. of Lab. Med., Translational Metabolic Laboratory, Radboud University Medical Centre, Nijmegen, The Netherlands; [2]Dept. of Anal. Chemistry/Chemometrics, Inst. for Molecules and Materials, Radboud University Nijmegen, The Netherlands.; [3]Dept. of Human Genetics, Donders Centre for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands.

The implementation of 'Next Generation Sequencing' (NGS) in clinical routine generates an increasing need for functional data to evaluate the pathogenicity of variants discovered. To provide the required functional information, we have developed a metabolic complement to NGS, which we have termed 'Next Generation Metabolic Screening' (NGMS). Instead of developing dedicated assays for individual classes of metabolites, we have followed an untargeted, mass spectrometry-based metabolomics approach. We now present a proof-of-principle study that demonstrates the feasibility of NGMS for the diagnosis of inborn errors of metabolisms (IEMs). Ultra-high pressure, reversed-phase liquid chromatography (UHPLC)-mass spectrometry was performed on an Agilent 1290 LC-QTOF 6545 system. Plasma samples of patients previously diagnosed with an IEM (11 different IEMs) and control samples were selected for this study and analyzed in duplo in a blinded fashion: i.e. the QTOF operator, bioinformatician and laboratory specialist were unaware of the diagnoses. Raw data was processed by Agilent MassHunter Qual software, and subsequently analyzed by XCMS online data software for retention time alignment, peak detection, and peak matching. The resulting features (i.e., an ion signal with accurate mass (m/z), intensity and retention time) were uploaded to an in-house developed chemometric pipeline, which selects features that are significantly different between patients and controls (*t*-test with Bonferroni-Holm correction), and automatically cross-references exact masses with the human metabolome database (HMDB) for putative metabolite annotation. To reduce data complexity and extract relevant information for diagnosis of IEMs, the automated pipeline included an initial targeted evaluation of ~350 known IEM-associated metabolites using a customized 'IEM panel'. As a second step, analysis of NGMS data could be expanded to include all HMDB annotated- and unknown features, analogous to the stepwise strategy used for interpreting complex NGS data. In each plasma sample, ~10.000 individual features could be detected. After chemometric analysis, approximately 50-500 significantly different features remained in patient samples as compared with controls. Through automated cross-referencing to the HMDB and application of our 'IEM panel', the relevant biomarkers associated with the IEMs studied could be readily identified in each individual patient. Upon extension of the analysis to all HMDB annotated- and unknown features, several potential novel biomarkers were identified for most of the IEMs. This study shows that the IEMs selected for this study, i.e. PKU, MCAD, VLCAD, MCC, amino acylase I, MSUD, MAT I/III, HMG-CoA lyase, prolinemia type II, xanthinuria type II and hyperlysinemia, could all be diagnosed unambiguously through an untargeted LC-QTOF-based metabolomics strategy in combination with an in-house developed chemometric pipeline. We now refer to this workflow as NGMS. This proof-of-principle study demonstrates the feasibility of this holistic metabolomics approach in diagnosing IEMs and illustrates the promises of NGMS for the present-day functional genomics laboratory.

O12: LC/MS metabolomics in diagnosing and treatment of rare diseases

David Friedecký[1,2], Radana Karlíková[1], Tomáš Adam[1,2]

[1]Laboratory of Metabolomics, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Hněvotínská 5, 779 00 Olomouc, Czech Republic; [2]Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry, University Hospital Olomouc, I.P. Pavlova 6, 775 20 Olomouc, Czech Republic



In last decade, metabolomics plays an important role in clinical research. As a picture of whole metabolic status, it brings complex insight into biochemistry and pathobiochemistry of diseases. Two analytical approaches are applied: targeted metabolomics based on analysis of hundreds of selected metabolites, and untargeted metabolomics for complex profiling of thousands features – potential metabolites. Nowadays, it becomes a part of clinical diagnosing and treatment monitoring due to better understanding of mechanisms of diseases and capability to find new biomarkers. Two examples of clinical metabolomics will be presented.

In chronic myeloid leukemia study, targeted LC/MS metabolomics method covering 350 metabolites was applied in order to understand biochemical changes of the disease(1). The metabolic profiles distinguished newly diagnosed patients and patients treated with hydroxyurea or tyrosine kinase inhibitors from healthy controls. The changes were found in glycolysis, citric acid cycle, and amino acid metabolism. Differences between resistant and well responding patients to imatinib treatment in levels of amino acids and acylcarnitines may be potentially used as an additional tool for the assessment of response.

In order to find new potential biomarkers of inborn errors of metabolism, untargeted LC/MS method allowing analysis of polar metabolites including phospholipids was used. Metabolic profile in dry blood spots from patients with medium chain acyl-CoA dehydrogenase deficiency was clearly distinguished from controls. Acylcarnitines (already well-known biomarkers) and oxidized phospholipids were found to be the most discriminating metabolites. Elevated levels of oxidized phospholipids in patients suggest an increased presence of oxidative stress as one of the pathophysiological mechanisms.

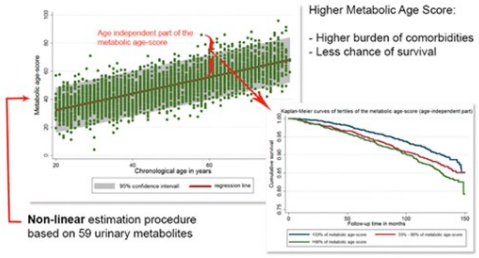
The work was supported by IGA_LF_2016_014, NPU I (LO1304), GACR I 1910-N26.

Reference: 1. Karlíková, R. *et al.* Metabolite Profiling of the Plasma and Leukocytes of Chronic Myeloid Leukemia Patients. *J. Proteome Res.* **15,** 3158–66 (2016).

O13: Refining Risk Factors via Metabolomics - a General Methodology and its Application

Johannes Hertel[1,4], Maik Pietzner[2,3], Nele Friedrich[2,5], Matthias Nauck[2,3], Hans Jörgen Grabe[1,4]

[1]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, 17475 Greifswald, Germany; [2]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany; [3]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, 17475 Greifswald, Germany; [4]German Center for Neurodegenerative Diseases DZNE, Site Rostock/Greifswald, 17487 Greifswald, Germany; [5]Research Centre for Prevention and Health, Glostrup University Hospital, Denmark



By drawing a holistic picture of the metabolic activities at a certain time-point, metabolomic techniques represent promising platforms for the development of personalized diagnostics which may help to establish a more targeted approach to the treatment and prevention of human diseases. Here, we will explicate a statistical framework deriving metabolomic prediction metrics for biological age and metabolic health with the goal of capturing biologically meaningful variance that is not represented by classical risk factors like chronological age or anthropometric measurements. The statistical framework was applied to large NMR measured urinary metabolomic data-sets from the Study of Health in Pomerania (SHIP-0: n=4068; SHIP-TREND-0: n=996) and, for replication in a clinical setting, to a small sample (n=38) of heavily obese individuals who underwent bariatric surgery and had pre- and post-surgery metabolomic measurements. The results demonstrated that the derived prediction scores cover the multimorbidity accompanied by aging and obesity and were even predictive for survival in 13 years of follow-up. The presentation for the workshop focusses on the aspects of validation strategies, data normalization among the three samples and the modeling of the inherent sex-specific non-linear relations of urinary metabolites to age and anthropometric measurements. Finally, a formal justification for the overall methodology is presented, specifying the necessary statistical prerequisites. In conclusion, we developed successfully a metabolomic analyses pipeline aiming at the refinement of classical risk factors which is in its abstract principle transferable to other omics layers. This work was part of the project GANI_MED (Greifswald Approach to Individualized Medicine) which was funded by the Federal Ministry of Education and Research (grant no. 03IS2061A).

References: Hertel, J et al. (2016). Measuring Biological Age via Metabonomics: The Metabolic Age Score. J. Proteome Res. *15*, 400S–410. Hertel, J., et al (2017). Two statistical criteria to choose the method for dilution correction in metabolomic urine measurements. Metabolomics *13*. (In Press). Hertel, J. et al. (2017) The Metabolic Waist Circumference Score – Quantifying the Metabolic Consequences of Obesity Using 59 Urinary Metabolites. Metabolism. (In Revision).

O14: A serum metabolomic analysis of HCV-infected patients successfully treated with IFN-free DAA regimens

Gaia Meoni[1], Giorgia Ceccotti [2], Leonardo Tenori [1], Laura Gragnani [2], Elisa Fognani [2], Elena Gianni [2], Claudio Luchinat [1] and Anna Linda Zignego [2]

[1]CIRMMP, University of Florence, Italy; [2]Center for Systemic Manifestations of Hepatitis Viruses, Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy

HCV infects about 170 million of subjects worldwide. The virus has a high propensity to persist in the host, leading to cirrhosis and liver cancer. Metabolomics is the study of metabolic changes in biological systems and may identify specific profiles associated with subtle alterations induced by diseases. Few studies are available on metabolic changes in liver injuries, and since none of them was focused on HCV-infected patients before and after reaching a sustained virological response (SVR) following treatment with direct acting antivirals (DAAs), the aim of this study was to perform a serum metabolomics analysis in this setting. Sera were collected from 67 HCV patients (36 men, mean age 63 ±9,5) successfully undergoing different IFN-free DAA regimens, before therapy (baseline), at post-treatment week 12 (SVR12) and at post treatment week 24 (SVR24). HCV genotype was 1a/1b in 70%, 2a/2c in 23%, 3 in 4.7% and 4 in 2.3%. METAVIR score indicated F3-F4 score in 55% of patients, the remaining 45% had F0-F2. We also analysed a group of 43 sera from healthy subjects (Hs) in order to localize them in the PLS plot with respect to baseline and SVR12 as a preliminary negative control for both groups. Samples were analysed using proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) operating at 600.13MHz. Multilevel Partial Least Squares (MPLS) analysis was applied to study the within-subject changes introduced in the individual metabolic profile by the DAAs therapy, demonstrating a significant pair-wise discrimination comparing both SVR12 and SVR24 samples with the corresponding baseline subjects (90% and 82% of discrimination accuracy respectively), indicating that the efficacy of the treatment is already visible after 12 weeks from the SVR. Several metabolites with unequivocal assignment (i.e. 3-hydroxybutyrate, 2-oxoglutarate, creatine, choline, valine, alanine, acetate, tyrosine, phenylalanine and formate) differed when comparing baseline with SVR12 and SVR24. Analyses revealed also a distinct metabolic profile of the heathy subjects compared with the two groups of patients after viral eradication. Both after viral clearance (SVR12 and SVR24) and in Hs our study shows a progressive increase of BCAA (branched chain amino acids) levels and diminished levels of AAAs (aromatic amino acids). This data suggests an increase in Fisher's ratio (ratio BCAAs to AAAs). Lower levels of essential amino acids in SVR serum samples imply the attenuation of the hypermetabolic status (decreased uptake of essential AAs from diet), and on the other hand decreased levels of non-essential amino acids indicate deceleration of the TCA cycle (1).

References: Sugiyama K. *et al.*, Prominent steatosis with hypermetabolism of the cell line permissive for years of infection with hepatitis C virus. PLoS One 9:e94460 (2014).

O15: Serum metabolomic profiles identify ER - positive early breast cancer patients at increased risk of disease recurrence: a multicentre population study

Alessia Vignoli[1], Christopher D Hart[2] Leonardo Tenori[1], Laura Biganzoli[2], Emanuela Risi[2], Richard Love[3], Angelo Di Leo[2], Claudio Luchinat[1]

[1]Magnetic Resonance Center (CERM), University of Florence, Sesto Fiorentino, Italy; [2]"Sandro Pitigliani" Medical Oncology Department, Hospital of Prato, Prato, Italy; [3]The International Breast Cancer Research Foundation, USA

Detecting signals of micrometastatic disease in early breast cancer (EBC) patients could improve risk stratification and enable better tailoring of adjuvant therapies. Has been shown that postoperative serum metabolomic profiles are predictive of relapse in a single-centre cohort of ER-negative EBC patients (1). Here, we investigated this further using pre-operative serum samples from ER-positive, premenopausal women with EBC who were enrolled in an international phase III trial (2,3). Proton nuclear magnetic resonance (NMR) spectroscopy of 590 EBC samples (319 with relapse or ≥ 6 years clinical follow up) and 109 metastatic breast cancer (MBC) samples was performed. A Random Forest (RF) classification model was built using a training set of 85 EBC and all MBC samples. The model was then applied to a test set of 234 EBC samples, and a risk of recurrence score was generated based on the likelihood of the sample being misclassified as metastatic. In the training set, the RF model separated EBC from MBC with discrimination accuracy of 84.9%. In the test set, the RF recurrence risk score correlated with relapse, with an area under the curve of 0.747 in receiver operator characteristics analysis. Accuracy was maximized at 71.3% (sensitivity 70.8%, specificity 71.4%). The model performed independently of age, tumor size, grade, HER2 status and nodal status, and also of AdjuvantOnline risk of relapse score. In conclusion, in a multicenter group of EBC patients, we developed a model based on preoperative serum metabolomic profiles that was prognostic for disease recurrence, independent of traditional clinicopathological risk factors.

References: 1. Tenori, L. *et al.* Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. *Mol. Oncol.* **9,** 128–139 (2015). 2. Love, R. R. *et al.* Timing of adjuvant surgical oophorectomy in the menstrual cycle and disease-free and overall survival in premenopausal women with operable breast cancer. *J. Natl. Cancer Inst.* **107,** (2015). 3. Hart, C. D. *et al.* Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population. *Clin. Cancer Res.* (2016).

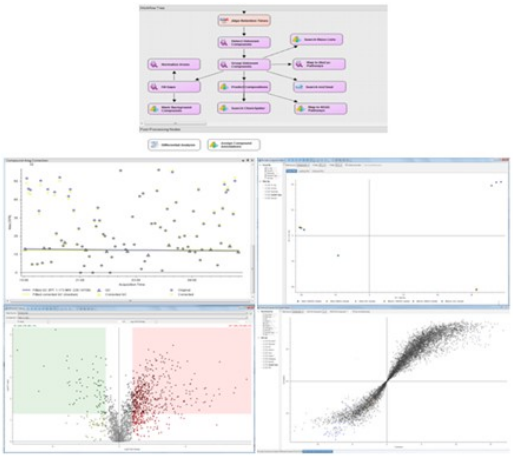
Company presentations

C1: Complete small molecule research and structure identification in a Next Generation Platform

Anas Kamleh[1], Claire Dauly[2], Ralf Tautenhahn[3] and Mark Sanders[3]

[1]Thermo Fisher Scientific, Hägersten, Sweden; [2]Thermo Fisher Scientific, Courtaboeuf, France; [3]Thermo Fisher Scientific, San Jose, CA, USA.

Metabolomics is the closest of Omics to the phenotype. It provides essential information about the mechanistic event in biological systems at the cellular level, as well as enriching our knowledge of the biosystem-environment interaction at the organ and organism levels. Metabolomics is a relatively new field of research and, despite great strides in the last decade, is still lacking powerful tools for data handling, processing and statistical analyses, especially when compared to the older, more established analyses of proteomics, transcriptomics and genomics data. More specifically, unknown compounds identification has been a bottleneck for the discovery of insightful mechanistic information and new markers for health and disease. Additionally, the advent of high resolution mass spectrometry with resolving power of 500K, albeit essential for correct discoveries, adds to the complexity of information and burden on the processing tools. Compound Discoverer™ (CD) is a new research tool released by Thermo Fisher Scientific that aims at providing a complete solution for mass spectrometry based metabolomics data mining. With more scientists, especially in clinical research, realising the importance of metabolomics and embracing the field, the software package was made user friendly to facilitate the transition and help them focus on their discovery aims rather than testing/coding tools for data handling. In fact, establishing an experiment in CD environment is only a few steps away thanks to a flexible, node based experimental setup. The user can choose which pipelines and workflows are most relevant to their work. For advanced users, the platform is accommodating new ideas and programmers can plug-in their own codes to customize the analysis. The software excels in the identification capabilities, with both *in-silico* fragmentation of suspected structures and a direct query of mzCloud™, a highly curated spectral library, and/or query of user's own spectral libraries. Compound Discoverer possesses multiple statistical tools for differential analysis spanning uni-variate, unsupervised and supervised multi-variate comparisons. Tracking information related to detected compounds such as moving from chromatograms to spectra (in MS1 and MS2) was also made easy and interactive.vIn my presentation, I will introduce the newest release Compound Discoverer™2.1, highlighting its capabilities for extracting the most out of high resolution, high complexity data, on the road to accurate discoveries.



C2: Investigation of Pyrazinamide Mechanism of Action for Tuberculosis Using Metabolomics

Christine A. Miller[1], Yuqin Dai[1], Travis E. Hartman[2], Steven M. Fischer[1] and Kyu Y. Rhee[2,3]

[1]Agilent Technologies, 5301 Stevens Creek Blvd, Santa Clara, CA 95051, [2]Division of Infectious Diseases, Department of Medicine, and [3]Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY 10065, USA

Tuberculosis (TB) is both the leading cause of deaths due to an infectious disease and the leading cause of deaths due to a curable disease. However, drug resistance is increasing while the pipeline of new drugs stagnates, and knowledge of existing drugs remains incomplete. Pyrazinamide (PZA) is a frontline TB drug whose mechanism of action remains among the most poorly understood.

Here, we present a high-performance ion-pairing reversed-phase (IP-RP) Q-TOF LC/MS method that has enabled the biologically unbiased study of the impact of PZA on the Mycobacterium tuberculosis metabolome. Coupled with batch feature extraction and multivariate statistical analysis software, this workflow enabled the discovery of activity-specific metabolic changes that may help explain PZA's unique metabolic effects.

C3: Using non-targeted high resolution LC-QTOF profiling to characterize metabolic responses of Nicotiana attenuata during infection with Rhizophagus irregularis

Heiko Neuweger[1], Sven Heiling[2], Ming Wang[2], Ulrike Schweiger-Hufnagel[1], Philipp Hennersdorf[1], Nikolas Kessler[1], Aiko Barsch[1], Emmanuel Gaquerel[3], Ian T. Baldwin[2]

[1] Bruker Daltonik GmbH, Bremen, Germany; [2] Max Planck Institute for Chemical Ecology, Molecular Ecology Department, Jena, Germany; [3] Centre for Organismal Studies Heidelberg, Plant Defense Metabolism Research Group, Heidelberg, Germany

Exploring the highly dynamic composition of specialized (secondary) metabolism in plants is challenging. Understanding the cause for this modulation is therefore important to disentangle the underlying mechanisms. One example is the mutualistic interaction between plants and root colonizing arbuscular mycorrhizal fungi (AMF) that has been shown to result in massive reprogramming of plant metabolism. AMF are ancient and phylogenetically spreading more than 80% of land plant species. They are known to facilitate the uptake of water and nutrients (plant root endosymbiosis) which influences plant growth or enhances the plant resistance against pathogens and abiotic stresses. Additionally AMF exchange bidirectional substances and influence phytohormone signalling. However the effects of AMF on plant defense mechanism has only been investigated in a few studies. Here we are using a software-based metabolite identification approach to assess the metabolic responses in *N. attenuata*'s during infection with the AMF *Rhizophagus irregularis.* The analysis was conducted using MetaboScape 3.0 providing a fully integrated workflow for discovery metabolomics. The novel T-ReX 3D processing of MetaboScape provides automated mass calibration, parameter free retention time alignment and region complete data extraction across the analytical measurements. The result of the processing workflow are robust and reproducible molecular feature tables. During processing, adducts, common fragments and different charge states of the same compound are automatically combined to enable meaningful statistics and increase confidence in compound identification. For compound ID and de-replication tasks, four orthogonal indicators are employed in MetaboScape: accurate precursor mass, retention time, true isotopic pattern and MS/MS fragmentation information. All four contribute to a unique and transparent annotation quality concept. The integrated SmartFormula, CompoundCrawler, MetFrag and SpectralLibrary tools provided a seamless identification workflow for unknown compounds. The interactive statistics allows focusing on the significantly regulated metabolites and enables to detect and link diverse effects in the phytometabolome of *N. attenuata* during *R. irregularis* infection. Putting the final results into biological context was possible with the integrated metabolic pathway mapping functionality. Using MetaboScape, it could be shown that AMF infection strongly influences malonylation of HGL -DTGs. Furthermore a novel phenolamide was strongly increased by AMF infection and could be identified by comparing MS/MS spectral similarities. The described combination of ecological screenings and unbiased metabolite profiling allows for the identification of potential new compounds and metabolic pathways important for the interactions between AMF and *N. attenuata*.

Posters

P1: Finding groups in large and high-dimensional data using k-means-based algorithm

Šárka Brodinová[1], Peter Filzmoser[1], Tomas Ortner[1], Maia Zaharieva[2] , Christian Breiteneder[2]

[1]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria; [2]Institute of Software Technology and Interactive Systems, Vienna University of Technology, Austria

Rapid developments in technologies enable to easily collect and store huge amounts of information resulting in large datasets of high-dimension. In any such kind of datasets, collected observations typically tend to form groups and in addition, both outliers (observations that do not fit to groups) and noise variables (variables that do not hold any useful information for the group separation) are very likely to appear. Therefore, identifying a group structure of large and high-dimensional datasets reveals potentially challenges due to both disturbing factors – outliers and noise variables. Moreover, the task of finding groups becomes even more problematic when no information about data, e.g. the number of groups, is available. A clustering algorithm aiming at discovering the data structure for such an application scenario will be presented.

The proposed k-means-based algorithm incorporates a weighting function using the Local Outlier Factor [1], which automatically assigns a weight to each observation. The resulting observation weights reflect the degree of outlyingness based on which the outliers are identified. In order to cope with noise variables, the contribution of each variable to the cluster separation is evaluated by employing a lasso-type penalty leading to sparsity in a variable vector [2]. In case that no prior knowledge about the data is available, the optimal number of clusters and the sparsity parameter are selected based on a gap statistic [2] adjusted by the observation weights. The performed experiments on simulated and real-world data indicate a great ability of the proposed clustering procedure to identify groups, outliers, and informative variables.

This work has been partly funded by the Vienna Science and Technology Fund (WWTF) through project ICT12-010 and by the K-project DEXHELPP through COMET - Competence Centers for Excellent Technologies, supported by BMVIT, BMWFW and the province Vienna. The COMET program is administrated by FFG.

References: 1. Breuning, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. LOF: identifying density-based local outliers. *In ACM sigmod record.* 29, 93–104 (2000). 2. Witten, D. M. and Tibshirani, R. A framework for feature selection in clustering. *Journal of the American Statistical Association.* 105 (490), 713–726 (2010).

P2: The metabolome analysis of amniotic fluid reveals microbial colonization

Eliška Čechová[1], Tereza Pavlová[1,2], Julie Bienertová-Vašků[1,2], Ivana Musilová[3], Marian Kacerovský[3,4], Jana Klánová[1], Zdeněk Spáčil[1]

[1]Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Brno, Czech Republic; [2]Department of Pathological Physiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic; [3]Department of Obstetrics and Gynecology, Faculty of Medicine, Charles University, Hradec Kralove, Czech Republic; [4]Biomedical Research Center, University Hospital Hradec Kralove, Hradec Kralove, Czech Republic

Preterm births occur prior the 37th week of gestation in 9-11% of pregnancies and can significantly influence children's future development and health [1]. A common cause of preterm birth is a microbial infection as a result of easier access of bacteria into the amniotic fluid due to preterm premature rupture of membrane (PPROM) condition [2]. In this case metabolomics was used to explore microbial colonization in amniotic fluid and potentially provide clinical data appropriate to determine diagnosis. Discovery approach based on ultra-high performance liquid chromatography (UHPLC) and high resolution/accurate mass (HRAM) mass spectrometry was used to analyse the metabolome of amniotic fluid. Samples of amniotic fluid (n=58) were collected from women diagnosed with PPROM and sorted in groups depending on presence or absence of microbial colonization [3]. Sample processing consisted of freeze-drying and extraction of dried samples with isopropanol. The UHPLC/HRAM analysis was performed on Nexera X2 system (Shimadzu corp., Japan) coupled to orbital ion-trap mass analyser (Orbitrap Fusion, Thermo Scientific, San Jose, USA). Data processing algorithm was designed in Compound Discoverer software (v2.0TM, Thermo Scientific). Metabolic profiles of the two groups (± microbial colonization) were compared for differences with a particular attention to the tryptophan metabolism pathways, which is known to be significantly affected by microbiota [4]. Metabolites identified in amniotic fluid as markers of microbial colonization were correlated with metagenomics data (16S rRNA gene sequencing). This study utilized HRAM-MS analysis to investigate microbiota-modulated host metabolome as well as exclusively microbial metabolome of amniotic fluid. The results may provide novel insights to comprehend the interaction between host and microbial metabolism. The research was supported by the Grant Agency of the Czech Republic (GAČR no 17-24592Y) and by the Czech Ministry of Education, Youth and Sports (CETOCOEN PLUS CZ.02.1.01/0.0/0.0/15_003/0000469 and LM2015051).

References: [1] H. Blencowe, et al., Born Too Soon: The global epidemiology of 15 million preterm births, Reprod. Health. 10 (2013) S2. [2] T.M. Medina, D.A. Hill, Preterm premature rupture of membranes: diagnosis and management., Am. Fam. Physician. 73 (2006) 659–664. [3] I. Musilova, et al., Intraamniotic Inflammation in Women with Preterm Prelabor Rupture of Membranes, PLoS One. 10 (2015) e0133929. [4] L. Etienne-Mesmin, B. Chassaing, A.T. Gewirtz, Tryptophan: A gut microbiota-derived metabolites regulating inflammation, World J. Gastrointest. Pharmacol. Ther. 8 (2017) 7–9.

P3: Study of seed dormancy by mass spectrometry combined with PCA and PLS-DA

Monika Cechová[1], Markéta Válková[1], Lukáš Kučera[1], Petr Barták[1], Petr Smýkal[2], Iveta Hradilová[2], Aleš Soukup[3], Anna Janská[3], Karel Hron[4], Kamila Fačevicová[5] and Petr Bednář[1]

[1]Regional Centre of Advanced Technologies and Materials, Department of Analytical Chemistry, Faculty of Science, Palacký University, 17. Listopadu 12, 77146 Olomouc; [2]Department of Botany, Faculty of Science, Palacký University, Šlechtitelů 11, 78371 Olomouc; [3]Department of Experimental Plant Biology, Faculty of Science, Charles University, Viničná 5, 128 44 Prague 2; [4]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic; [5]Department of Mathematics, Faculty of Education, Palacký University, Žižkovo nám. 5, 771 40 Olomouc, Czech Republic

Timing and efficiency of seed germination are key factors for plant life. In a population of wild seeds (unlike crop one) only a part of individuals starts to germinate in proper conditions. Fraction of seeds remaining inactive is expressed by dormancy. The study deals with development of methods distinguishing chemical differences of wild (dormant) pea genotypes from crop (non-dormant) ones in dry state for better understanding the process of imbibition and germination on molecular level. Liquid chromatography combined with high resolution tandem mass spectrometry (LC/HRTMS) was used for the analysis of seed coat extracts. Several polyphenols appeared among markers of dormancy, i.e. dimer and trimer of gallocatechine and rhamnosides of quercetin and myricetin. Laser desorption/ionization mass spectrometry was utilized for direct analysis of seed coat outer layers. Seed coats were fixed on the MALDI plate by double sided tape and immediately measured. Hydrophobic components of the outermost seed coat layers (cuticle), especially hydroxylated long chain fatty acids, appeared to be also significant markers of dormancy (1). The effect of sugar composition on seed imbibition and first phases of germination is now in progress. Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA), especially performed in centered logratio coefficients, respecting relative nature of samples, appeared to be extremely useful tools for data interpretation. In this contribution, the important aspects of data preparation and application of those multivariate methods will be discussed.

This research was funded by Ministry of Education, Youth and Sports of the Czech Republic (LO1305), Grant Agency of Czech Republic (14-11782S) and Palacký University Olomouc (IGA_PrF_2017_020).

Reference: 1. Hradilová I., Trněný O., Válková M., Cechová M., Janská A., Prokešová L., Aamir K., Krezdorn N., Rotter B., Winter P., Varshney R.K., Soukup A., Bednář P., Hanáček P., Smýkal P., A Combined Comparative Transcriptomic, Metabolomic, and Anatomical Analyses of Two Key Domestication Traits: Pod Dehiscence and Seed Dormancy in Pea (Pisum sp.). *Front. Plant Sci*. **8**, 1-25 (2017).

P4: Addressing stability of urine and plasma in large-scale LC-MS metabolomics experiments

Morgan J. Cichon[1,2] and Ken M. Riedl[2,3]

[1]Personalized Food and Nutritional Metabolomics Discover Theme, The Ohio State University, Columbus, Ohio, United States. [2]Department of Food Science & Technology, The Ohio State University, Columbus, Ohio, United States. [3]Nutrient and Phytochemical Analytic Shared Resource, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio, United States

The human metabolome is large and complex and many genetic and environmental factors can contribute to the observed diversity in the metabolome. For this reason, large-scale metabolomics experiments are often necessary in order to adequately address the biological variation that exists in the human population. Of course, with larger sample numbers, there is more room for variability due to sample preparation and handling. In order to reduce batch effects in LC-MS metabolomics experiments, it is desirable to run as many samples continuously in as few batches as possible and to limit the number of different analysts. However, balancing variability and feasibility is important and often challenging. Since previous metabolomics stability studies have focused on handling, storage, or freeze-thaw cycles[1], here we evaluate stability during analysis. Our objectives were to determine the largest analytical batch that could be prepared at one time for LC-QTOF-MS metabolomics studies of urine and plasma and to evaluate instrument stability over a large continuous batch of samples with intermittent pooled QC samples. To determine sample stability in the autosampler pooled samples of urine and blood were extracted and aliquots of the prepared extracts placed in the autosampler and injected sequentially over a 12 hour period (60 injections). The resulting LC-QTOF-MS data were used to determine the maximum number of samples that could be prepared and set in the autosampler at a time. For the second objective, pooled QC samples were evaluated for reproducibility over the course of a 300 sample study run in a continuous batch. From our analysis, there was a moderate separation of samples based on length of time in the autosampler prior to injection and this varied for urine and plasma. In addition, a pre-equilibration period in the autosampler was necessary to further reduce variability. Using the autosampler results we found it feasible to run a large continuous batch of 300 samples as QCs were highly reproducible over the course of the study. Results from this work will allow us to design and conduct large-scale urine and plasma metabolomics studies more efficiently and reproducibly.

The work was supported by NCI P30CA016058.

Reference: 1. Gika, H.G. *et al.* Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry fingerprinting of human urine: Sample stability under different handling and storage conditions for metabonomics studies. *J. Chrom. A.* 1189, 314-322 (2008).

P5: Role of Adipose Triglyceride Lipase in cancer

Barbara Darnhofer[1,2,3], Tamara Tomin[1,2], Katarina Fritz[1,2] , Matthias Schittmayer[1,2], Ruth Birner-Gruenberger[1,2,3]

[1]Functional Proteomics and Metabolic Pathways Research Unit, Institute of Pathology, Medical University of Graz, Austria; [2]Omics Center Graz, BioTechMed-Graz, Austria; [3]Austrian Center of Industrial Biotechnology, Graz, Austria.

Metabolic rewiring is a dominant factor within tumor initiation and progression. Defining the pathways that are limiting for cancer progression can be exploited as therapeutic targets. Fatty acid metabolism is a key process in initiation of metastasis (via import of external fatty acids) and also in cancer cell proliferation (via de novo fatty acid synthesis). Lipases (responsible for mobilization of fatty acids from lipids) play a central role in providing fatty acids and oncogenic signaling lipid molecules thereby affecting cancer aggressiveness. The rate limiting lipase in intracellular triacylglycerol mobilization Adipose Triglyceride Lipase (ATGL) has been found to be down regulated in many human tumors and its depletion induces mouse pulmonary neoplasia. Consistently, we report here that CRISPR/Cas9 deletion of ATGL in A549 lung adenocarcinoma cells leads to increased intracellular lipid accumulation and in addition increases cell proliferation and cell mobility. Our label free quantitative proteomics and transcriptomics data show that protein and mRNA expression levels of the proto-oncogene Src and enzymes of one-carbon metabolism are significantly altered. Increased cell motility of ATGL knock out A549 lung adenocarcinoma cells is reversed by pharmacologic inhibition of Src. One-carbon metabolism is a central biosynthetic hub of cellular proliferation producing building blocks for DNA synthesis, providing methylation equivalents for transcriptional regulation and is a major source for cellular NADPH, a biosynthetic bottleneck in rapidly proliferating cells required for e.g. de novo fatty acid synthesis and oxidative stress response systems. We currently perform metabolomics analyses to better understand the complex interplay of fatty acid metabolism with central carbon and one carbon metabolism to drive cell proliferation directly on the metabolite level.

P6: An introduction of DEG2 package for metabolomics data

Rohmatul Fajriyah[1], Dedi Rosadi[2]

[1]Department of Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia; [2]Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia

Statistics is considered as a gentle guide starting point for further analysis in metabolomics data analysis to achieve the goals such as biomarker discovery and disease diagnosis. Statistical methods of univariate and multivariate are used extensively in metabolomics studies (Xi et. al, 2014). We present a new *R* package, namely **DEG2** (Differentially Expressed Genes 2). For its current version, this package can be used, for instances, to do the cross-variance statistical test for Microarray/Sequencing/metabolomics data which has been proposed and introduced in Fajriyah ((Fajriyah 2014) and (Fajriyah 2017)). In metabolomics studies, the test can be used as an alternative test to show which metabolites have the power to differentiate the two different groups in the data set when the sample size is small, the situation which is usually appeared in the bioinformatics research. Based on its statistical distribution, the p-value can be also provided. The test is built under a homogeneity of variance between samples.

The work was supported by Directorate of Research, Technology and Community Service, Ministry of Research and Technology, Indonesia.

References: 1. Xi, B., Gu, H., Baniasadi, H., and Raftery, D. 2014. "Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data". Methods in Molecular Biology, 1198: 333-353. 2. Fajriyah, R. 2014. "Microarray Data Analysis: Background Correction and Differentially Expressed Genes." PhD thesis, Institute of Statistics, TU Graz, Austria. 3. Fajriyah, R. 2017. "Introducing a Cross-Variance Concept and Its Application." In *12th Icmsa Proceeding*.

P7: Weighted pivot coordinates and their use in metabolomics

Alžběta Gardlo[1,2], Karel Hron[3]

[1]Laboratory of Metabolomics, Institute of Molecular and Translational medicine, Palacký University, Olomouc, Hněvotínská 5, Olomouc, Czech Republic; [2]Department of Clinical Biochemistry, University Hospital Olomouc, I.P. Pavlova 6, Olomouc, Czech Republic; [3]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University Olomouc, 17. listopadu 12, Olomouc, Czech Republic

The logratio methodology for statistical analysis of so called compositional data becomes more popular and it is nowadays successfully applied also to the field of metabolomics (1,2). The logratio methodology converts compositional data from their original geometry (Aitchison geometry) to interpretable real orthonormal coordinates and allows to perform a reasonable statistical processing and graphical outputs of compositional data. However, compositional data (especially in metabolomics) and their logratio coordinates can be influenced by artifacts resulting from processing (imputed) data below detection limit of measurement devices. Weighted pivot coordinates are new orthonormal logratio coordinates (3) which capture the relevant relative information about an original component and also suppress the redundant information from the data in a controlled way. The system of weights is used that accounts for different roles of parts in a compositional data set to find a more reasonable way to extract relative information about single components within orthonormal logratio coordinates. The only limitation is that the information about a part of interest within one orthonormal coordinate system is split into two coordinates, one capturing the relevant information of that part and the other containing its redundant residue (typically the last coordinate).

The practical part of the contribution is focused on real data from metabolomics comparing biological samples of healthy controls and patients suffered from a specific disease. Using weighted pivot coordinates markers of this disease (by proper choice of weights) can be found much easier comparing to other alternatives.

The work was supported by NPU I (LO1304), GACR 15-34613L.

References: 1. Gardlo, A. et al. Normalization techniques for PARAFAC modeling of urine metabolomic data. *Metabolomics* 12 (7), article 117, (2016). 2. Najdekr, L. et al. Oxidized phosphatidylcholines suggest oxidative stress in patients with medium-chain acyl-CoA dehydrogenase deficiency. *Talanta* 139, 62-66, (2015). 3. Filzmoser, P., Hron, K. Robust coordinates for compositional data using weighted balances. In K. Nordhausen and S. Taskinen (Eds.), *Modern Nonparametric, Robust and Multivariate Methods*, 167-184, Springer, Heidelberg, (2015).

P8: De-replication study of Natural Products using Mass Spectral Pattern

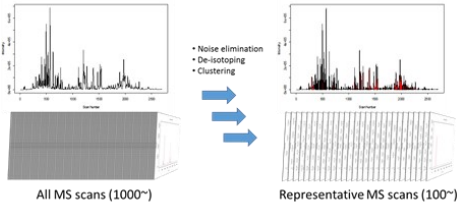
Heejung Yang

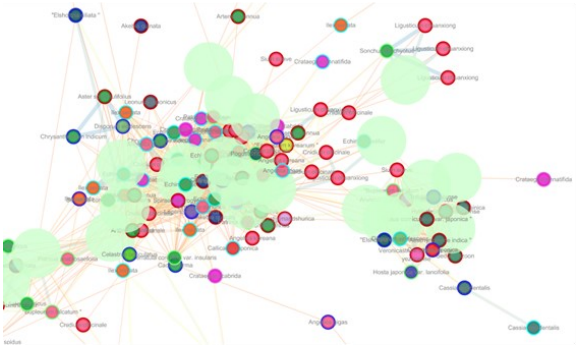
Laboratory of Natural Products Chemistry, College of Pharmacy, Kangwon National University, Chuncheon, 24341, Republic of Korea

Single natural product extract or the combinational formula have been used for coping with illness and treating diseases for a long time. Since even single natural product produces tons of metabolites via the complex metabolism, it is one of the biggest obstacles in natural products research. The complexity and diversity of secondary metabolites are not only frustrating natural chemists but also hindering the discovery of new therapeutic agents.

Recently, mass spectrometry (MS) has been spotlighted as a state-of-art method providing of invaluable information for discovery of novel components in natural products. However, despite such significant advances of MS technology with high-throughput power, too much information derived from MS might be not rather useful in the study of natural products. In the present study, we tried to develop a data-processing pipeline for processing of huge MS spectral information for small molecules which are generally covered in the research field of natural products chemistry using R statistical program. These processes are including of noisy signal elimination, de-isotopoing, clustering, visualization and identification. The network based on the representative scans derived from the MS spectral data of 368 Korean medicinal plants were visualized and analyzed for the dereplication of known components and the discovery of novel ones. We expect this process will aid the inspection of the complexity of natural products.

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI15C0075) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2015R1C1A1A01053892).





P9: Determination of Quorum sensing molecules and other Candida Albicans metabolites in human vaginal lavage using tandem mass spectrometry

Jakubec P.[1], Vlčková H. [1], Pilařová V. [1], Buchta V. [2], Nováková L.[1]

[1]Department of analytical chemistry, Faculty of Pharmacy in Hradec Králové, Charles University in Prague, Czech Republic; [2]Department of clinical microbiology, University Hospital in Hradec Králové, Czech Republic

Candida albicans (CA) is a common part of human microflora, typically present in oral cavity, gastrointestinal tract or vagina. The quorum sensing is a system which regulates population density of the microorganism and enables them to respond to changing physiological parameters of environment including human body. In case of CA quorum sensing molecules are farnesol and tyrosol. Farnesol blocks the morphogenic switch from yeast to hyphae and tyrosol supports it.

Vaginal swabs were collected and eluted in phosphate buffer saline solution from 68 patients and 14 healthy controls with and without vulvovaginal discomfort. Presence of CA and its phenotype was also defined.

The study had two major aims, firstly the determination of the farnesol and tyrosol concentrations in vaginal lavage using two different ultra-high performance liquid chromatography mass spectrometry systems (UHPLC-QqQ and UHPLC-Q-ToF) and secondly the assessment of the other CA metabolites using high resolution mass spectrometry. Three sample preparation methods were evaluated for the quantification of farnesol and tyrosol, the simple protein precipitation with acetonitrile, micro-extraction packed sorbent (using PEP sorbent) and µ-SPE in pipette tips (3 layers of C8 and 3 layers of styrene divinyl benzene sorbent). Analytes were separated on Acquity BEH C18 column using gradient elution with composition of 0.075% formic acid and 0.075% acetonitrile mobile phases. Flow rate of 0.2 ml/min and 4.6 minutes analysis time was used for UHPLC-QqQ and 0.5ml/min, 2.6 minutes analysis for UHPLC-Q-ToF. Quantification was done using deuterium labeled internal standards.

This work was supported by the STARSS project Reg.No.CZ.02.1.01/0.0/0.0/15_003/0000465) co-funded by ERDF, SVV 260 412/2017 and Ministry of Health of the Czech Republic grant nr. 15-29225A.

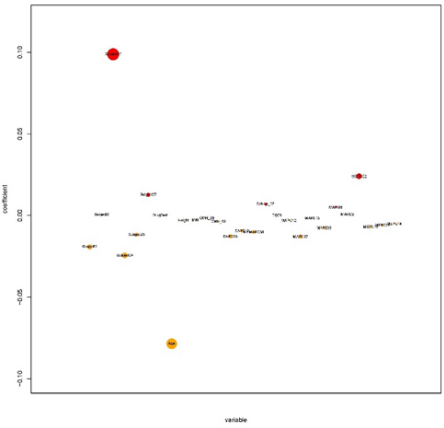
P10: Zonisamide: Plasma Metabolomic profiling in a crossover clinical trial

Jose Carlos Martínez Ávila[1], Alberto Borobia[1,2], Alberto García Bartolomé[1], Antonio Carcás Sanjuan[1,2]

[1]Fundación Investigación Hospital La Paz. IdiPaz Department of Clinical Pharmacology. Paseo de la Castellana 261. 28046 Madrid, Spain.; [2]Department of Pharmacology. Faculty of Medicine. Universidad Autonoma de Madrid. Avda Obispo Morcillo s/n. 28029. Madrid, Spain.

Sequence	Period 1	Washout	Period 2
1	R		T
2	T		R

Bioequivalence trials are designed to find if two pharmaceutical products can be considered equivalent. In order to achieve this aim bioequivalence trials follow a two period, two sequence crossover design. Subjects are randomly allocated in one of the two sequences, receiving reference drug or test as correspond with a washout between periods. After to apply preprocessing methods to the untargeted metabolite profiles such as missing values imputation [1], log2 transformation, normalization and scaling[2] and multivariate outlier detection [3], to look for differences between periods and sequence is a straightforward analysis of a high dimensional problem. 28 volunteers, two samples per volunteer. No remarkable metabolite difference was found between periods. Comparison between sequences gives a difference in taurine conjugated bile acids. A different approach is to estimate the influence of metabolites into pharmacokinetical parameters such as C_{max}, maximum concentration, t_{max}, time when maximum concentration appears, resulting in a repeated measures structure. Penalized regression via Lasso [4] is one of the methods used to find most important metabolites. However, it is necessary to apply a penalized regression accounting for a subject variability, here we included the subject as a fixed effect but a Lasso with random effects is needed to allocate properly variability. Another approach could be to analyze separately with a penalized regression period one and period two and later compare if the same metabolites are selected. As we see variability due to subject is a key topic, how to account for it? A measure of the percent of variability in metabolites profile attributable to subjects is needed in order to assess which methods should be used to characterize their influence in pharmacokinetical variables.



References: [1] D. J. Stekhoven. "Missforest-Non-parametric missing value imputation for mixed-type data," *Bioinformatics*, 28, 112–118, 2012.; [2] R. A. van den Berg, "Centering, scaling, and transformations: improving the biological information content of metabolomics data," *BMC Genomics*, 7, 142, 2006.; [3] P. Filzmoser, "Outlier identification in high dimensions," *Comput. Stat. Data Anal.*, 52, 1694–1711, 2008.; [4] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. (Series B)*, 58, 267–288, 1996.

P11: Bayesian counterpart to *t*-tests in compositional analysis of metabolomic data

Julie Rendlová[1], Karel Hron[1], Ondřej Vencálek[1], Alžběta Gardlo[2,3]

[1]*Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University, Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic;* [2]*Department of Clinical Biochemistry, University Hospital Olomouc, I.P. Pavlova 6, 775 20 Olomouc, Czech Republic;* [3]*Laboratory of Metabolomics, Institute of Molecular and Translational Medicine, Palacký University Olomouc, Hněvotínská 5, 779 00 Olomouc, Czech Republic*

Both targeted and untargeted metabolomic methods aim to find statistically significant differences in chemical fingerprints of patients with some disease and a control group, and to identify biological markers allowing then the prediction of the disease. After the raw measurements are pre-processed and interpreted as intensities of chromatographic peaks, the differences between controls and patients are often evaluated by *t*-tests or their nonparametric version (Wilcoxon test). The aim of this contribution is to propose a Bayesian counterpart to this traditional approach.

Methods of Bayesian inference reallocate some prior credibility across the space of all possible hypothesis or values of parameters consistently with the data evidence until posterior distribution is obtained [1]. Bayesian *t*-test, assuming *t* distribution with higher tails instead of Gaussian, is therefore a robust method from its nature [2]. Moreover, unlike classical statistic methods, it does not run into troubles when multiple tests are done simultaneously, e.g. for each metabolite. To evaluate markers, Maximum Aposteriori (MAP) criterion is available together with more complex information considering the entire posterior distributions.

Furthermore, since each metabolome is a collection of some small-molecule metabolites in a biological tissue, metabolomic data should be always treated as compositions, e.g. relative structure of metabolites should be of interest rather than absolute values of spectroscopic measurements. Compositional data follow so-called Aitchison geometry instead of Euclidian one [3], therefore understanding of some basic principles of this geometry and work in centered logratio coefficients are required for any further analysis including the Bayesian one.

Theoretical part of this contribution is accompanied by analysis of data set containing dry blood spots of patients suffered from medium chain acyl-CoA dehydrogenase deficiency (MCADD).

Reference: 1. Kruschke, J. (2011), *Doing Bayesian Data Analysis*, New York: Academic Press.; 2. Kruschke, J. K. (2012), Bayesian Estimation Supersedes the T Test, *Journal of Experimental Psychology: General,* **142** (2), 573-603.; 3. Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R. (2015), *Modeling and Analysis of Compositional Data*, Wiley, Chichester.

P12: Towards standardized data analysis workflows in targeted metabolomics – challenges and best practices in biomarker discovery

Simon Schafferer, Guido Dallmann, Judith Wahrheit, Therese Koal

Biocrates Life Sciences AG, Innsbruck, AT

Being a relatively new player in the –omics technologies, metabolomics has now developed into a widespread and highly interdisciplinary area of research, confirmed by an exponential growth of publications since 2000. High-throughput targeted metabolomics has opened the doors for large scale projects in conjunction with other omics areas e.g. proteomics or genomics to shed light into biological pathways of various diseases and contribute in biomarker discovery to ultimately provide a diagnostic screening platform.

Despite its rapid growth and large amount of data produced, there is still a lack of awareness on the importance of targeted metabolomics data pre-processing and pre-treatment, i.e. the very first steps of data analysis. Correct data handling at this early stage may significantly enhance later statistical data analysis and finally biological interpretation and biomarker identification. Here we address the cornerstones of standardized data analysis in targeted metabolomics ranging from common pitfalls in study design, to rules for harmonization of data pre-processing, as well as their impact on statistical analysis and biomarker discovery.

Correct treatment of concentration values below the limit of detection (LOD) for instance leads to more accurate biological interpretation and thereby accelerates research by proper planning of follow up experiments. An initial data quality check on sample generation (e.g. time to liquid sample centrifugation and sample storage at room temperature) increases the statistical power of tests at later stages of data analysis. In addition, study design plays an important role, by overcoming technical issues that may occur during measurement.

Raising the awareness of best practices in data pre-processing and pre-treatment will promote targeted metabolomics and will give researchers the opportunity to spend less time on the computational workflow, with guarantee of standardized data pre-processing and data analysis and more time with biological interpretation.

P13: Investigation of diabetes mellitus type II in relation to body weight by LC-MS based metabolic fingerprinting

Jitka Široká[1], Joanna Godzien[2], Michal Ciborowski[3], Coral Barbas[2]

[1]Laboratory of Growth Regulators, Centre of the Region Haná for Biotechnological and Agricultural Research, Palacký University & Institute of Experimental Botany ASCR, Šlechtitelů 27, CZ-783 71, Olomouc, Czech Republic; [2]CEMBIO (Center for Metabolomics and Bioanalysis) Facultad de Farmacia, Universidad San Pablo CEU, Campus Monteprincipe, Madrid, Spain; [3]Centre for Clinical Research, Medical University of Bialystok, Bialystok, Poland

Overweight and obesity are counted as major risk factors of diabetes mellitus type II progression. The exact pathophysiological aspects of this relation have not been fully elucidated, yet (1). Therefore, 203 of plasma samples from subjects of different body weight (normal weight, overweight, obese) with diagnosis of diabetes, prediabetes, insulin resistance and healthy individuals were analysed by liquid chromatography coupled to high resolution mass spectrometry operated in positive and negative ionization mode. The obtained metabolic fingerprints were statistically evaluated and the discriminating signals were subjected to mass spectrometry fragmentation for structure elucidation. From about 200 fragmentation spectra about 100 signals were annotated by manual reviewing of spectra and comparison with databases (Metlin, LipidMaps, HMDB).

In unsupervised principal component analysis score plots constructed for each body weight group in both ionization mode separately the overlap of healthy, insulin resistant and prediabetic was observed, suggesting small differences in metabolic fingerprints of these groups. The cluster of diabetic partly overlapping healthy, prediabetic and insulin resistant in all body weight groups indicate deeper changes in metabolic fingerprints. So far, the overall obtained data have to be carefully interpreted. The work was supported by European Social Fund Postdoc Project No. CZ.1.07/2.3.00/30.0004.

Reference: 1. Park, S. *et al.* A Metabolomic Approach to Understanding the Metabolic Link between Obesity and Diabetes. *Molecules and Cells* **38,** 587-596 (2015).

P14: Regression analysis with compositional covariates in the presence of cellwise outliers

Nikola Štefelová[1], Andreas Alfons[2], Javier Palarea-Albaladejo[3], Peter Filzmose[4], Karel Hron[1]

[1]Department of Mathematical Analysis and Applications of Mathematics, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic; [2]Econometric Institute, Erasmus School of Economics, Erasmus Universiteit Rotterdam, 3000 DR Rotterdam, The Netherlands; [3]Biomathematics and Statistics Scotland, JCMB, The King's Buildings, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, Scotland, UK; [4]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstrasse 8-10, Wien, Austria

Multivariate data are commonly arranged as a rectangular matrix with observations (cases) in the rows and variables in the columns. Ordinary robust estimators are designed to deal with case-wise outliers, i.e. with entire contaminated rows, assuming that there is a majority of non-contaminated observations in the dataset. However, this may not be realistic in many situations where contamination occurs at the cell level. That is, where only a small number of variables is affected per case, but contamination typically propagates throughout many observations. In this case, suppressing entire rows can lead to unacceptable and unnecessary loss of information, particularly in high-dimensional settings. Additional problems arise when data of compositional nature are involved, because then all the relative information about a certain cell representing a compositional part is contained in ratios of such a part to other parts (1). The aim of this work is to present our method for robust compositional regression that can deal with both case-wise and cell -wise outliers. The method is designed for regression with real response variable and composition alongside non-compositional variables as regressors. The proposed procedure involves the following steps. We start with a detection of cell-wise outliers based on a consistent univariate filter presented in (2), taking into account special treatment needed for compositional data. We set all the flagged cells as missing values. Then we impute those missing values through an iterative model-based imputation. We adapt the algorithm from (3) to be able to impute also non-compositional variables. We proceed with a regression carried out according to the approach from (4). As we want to eliminate possible row-wise contamination, we use robust MM-estimator for regression coefficients. Finally, we check a significance of the coefficients via bootstrap. The performance of the presented procedure is illustrated with real-world biological data.

References: 1. Pawlowsky-Glahn, V. et al. *Modeling and analysis of compositional data,* Chichester: Wiley (2015).; 2. Leung, A. et al. *Robust regression estimation and inference in the presence of cellwise and casewise contamination*. Computational Statistics & Data Analysis 99, 1-11 (2016).; 3. Hron, K. et al. *Imputation of missing values for compositional data using classical and robust methods*. Computational Statistics & Data Analysis 54 (12), 3095-3107 (2009).; 4. Hron, K. et al. Journal of Applied Statistics 39 (5), 1115-1128 (2012).

P15: LC-HRMS untargeted metabolomics in drug metabolism studies

Ivo Vrobel[1], David Friedecký[1,2], Edgar Faber[3], Lukáš Najdekr[1], Kateřina Mičová[1], Radana Karlíková[1], Tomáš Adam[1,2]

[1]Laboratory of Metabolomics, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Hněvotínská 5, 779 00 Olomouc, Czech Republic; [2]Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry, University Hospital Olomouc, I.P. Pavlova 6, 775 20 Olomouc, Czech Republic; [3]Department of Hemato-Oncology, Faculty of Medicine and Dentistry, Palacký University Olomouc, I.P. Pavlova 6, 775 20 Olomouc, Czech Republic

Biotransformation of xenobiotics results in range of metabolites. These are sequentially eliminated from the organism. Today we know a number of biotransformation reactions and therefore, it is possible to predict majority of formed metabolites. However, metabolization of various substances including drugs can also proceed through various rearrangements that lead to formation of compounds with significantly modified structure. Prediction of formation of such metabolites is not straightforward and so they might be easily missed.

One possibility how to detect these structurally modified metabolites is untargeted analysis based on liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) followed by statistical treatment. Two groups of samples, one from patients using the surveyed drug and one from control healthy volunteers, are analyzed in one batch, in the same fashion as in the standard metabolomic experiment. The main differences between these groups that are uncovered by statistical comparison are the drug and its metabolites occurring only in patient group.

We tested this approach on samples from patients suffering from chronic myeloid leukemia treated by specific a tyrosine kinase inhibitor imatinib (Glivec, Novartis). The metabolization of imatinib has been studied extensively and large number of metabolites has been reported. Even though, using the untargeted approach mentioned above we were able to detect 24 novel metabolites. The Compound Discoverer 2.0™ software was used for statistical evaluation of data. Taking advantage of multi stage fragmentation and high resolution measurements enabled by Orbitrap technology we also proposed chemical structures of majority of these metabolites. Among those were cysteine and cystine adducts of imatinib formed after *in vivo* rearrangements. The inception of reactive intermediates presumably precedes the formation of these sulphur-containing metabolites. This could shed light on the mechanism of imatinib adverse effects, especially hepatotoxicity and nephrotoxicity.

This work was supported by the Czech Science Foundation grant 15-34613L and grant IGA MZ ČR NT 12218-4/2011 and UP LF 2016-001. The infrastructural part (Institute of Molecular and Translational Medicine) was supported from NPU I (LO1304).

P16: Cell-wise outlier diagnostics and its use for biomarker identification

J. Walach[1], P. Filzmoser[1], K. Hron[2], B. Walczak[3] and L. Najdekr[4,5]

[1]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria; [2]Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University in Olomouc, Czech Republic; [3]Department of Analytical Chemistry, Institute of Chemistry, University of Silesia, Katowice, Poland; [4]Laboratory of Metabolomics, Institute of Molecular and Translational medicine, Palacký University in Olomouc, Czech Republic; [5]University Hospital Olomouc, Czech Republic

In data analysis, outlier diagnostics usually refers to an analysis of rows (observations) of the data matrix. However, it is common that most data cells in a row are regular and only few are deviating. Thus, cell-wise outlier diagnostics can bring more insight into the data. Even though the statistical field provides some possibilities for cell-wise outlier diagnostics, these methods usually fail in the case when the so-called "size-effect" is present. The size-effect occurs due to the different sample volume or concentration of observations. In this case, the absolute information might not be useful anymore and only relative information might lead to adequate results. Here we propose a method that makes use of the log-ratio approach (Pawlowsky-Glahn and others, 2015). We use the elements of the variation matrix defined as the variance of $\log(x_i/x_j)$, for all pairs of variables x_i and x_j. The advantage of log-ratios is that the absolute concentration is irrelevant, which is appropriate in the context. The variation matrix is computed for the joint data as well as for the single groups separately. Instead of using "classical" estimation of variation (square of standard deviation) one can use robust estimation of variance. Here, the τ estimator is used which has two major advantages. Firstly, the variation matrix will be estimated robustly. More importantly, the τ estimator provides robust weights for each value, in our case for each log-ratio. These weights are aggregated and together with applying filtering based on a selected critical value results in cell-wise outlier diagnostics. Since, at least from a statistical point of view, the biomarkers are such variables (metabolites) which differ between the groups (e.g. control and disease), the outlyingness information can be used for biomarker identification. The method has been tested on real as well as on simulated datasets and it shows promising results. This work is supported by the Austrian Science Fund (FWF) and Czech Science Fund (GACR), project number I 1910-N26.

References: 1. Pawlowsky-Glahn, V., Egozcue, R. Tolosana-Delgado, J.J. (2015) Modeling and Analysis of Compositional Data, *Chichester: Wiley*; 2. Walach, J., Filzmoser, P., Hron, K., Walczak, B., Najdekr, L. (2016) Robust biomarker identication in a two-class problem based on pairwise log-ratios, *Manuscript submitted for publication.*

P17: LC-MS based metabolite profiling of oat varieties cultivated in Switzerland

Dan Zhu, Laura Nyström

Laboratory of Food Biochemistry, Institute of Food, Nutrition and Health, ETH Zurich, Schmelzbergstrasse 9, CH 8092 Zurich, Switzerland

Oat is currently sixth in world cereal production, but only 10% of the harvest is used for human food [1]. It is a rich source of dietary fibers and bioactive phytochemicals. Nowadays more and more evidence shows that regular consumption of wholegrain products can prevent several chronic diseases, such as cardiovascular disease. Oat is nutritionally a valuable cereal food because it is almost exclusively consumed as wholegrain products. In comparison with other cereals such as wheat and rice, bioactive phytochemicals of oat are much less studied. It is known that oat is rich in phenolic compounds with antioxidant potential. Moreover, the other minor components with health benefits, such as sterols, tocols, folates, avenanthramides are also in significant amount in oat [2]. Furthermore, the contents of phytochemicals also vary among different oat varieties. In order to evaluate the profile of phytochemicals in oat, untargeted metabolomics based investigation was performed with 25 oat varieties cultivated in Switzerland in this study.

Phytochemicals of oat with different varieties were extracted with methanol, and further analyzed with a UPLC-HR-Q-TOF-MS-based untargeted metabolite profiling approach. Compound spectra were collected in both positive and negative ionization modes using an electrospray ionization source. Multivariate statistical analysis was utilized to find the characteristic markers, enabling us to find the discrimination among the varieties. Moreover, the characteristic markers were tentatively identified by in-house and public databases. Our preliminary results indicate that the metabolite profile is significantly different between oat varieties. Further studies will be done in the future to validate the compound identification. In addition to the new information and understanding about the phytochemical profiles of oat, the results of this study will provide insights for better selection of oat varieties in agriculture as well as improvement of oat food products in the future.

The work was financially supported by Swiss Federal Office of Agriculture (NAP-PGREL Program), and ETH Zurich. We thank Dr. Fabio Mascher and Beate Schierscher-Viret from Agroscope (Changins (VD), Switzerland) for providing oat samples.

References: [1]. Welch, R.W. 2011. Nutrient composition and nutritional quality of oats and comparisons with other cereals. In *Oats: Chemistry and Technology*, chapter 6, pp.95-107. [2]. Shewry, P.R., *et al*. 2008. Phytochemical and Fiber Components in Oat Varieties in the HEALTHGRAIN Diversity Screen. *Journal of Agricultural and Food Chemistry*, 56 (21), 9777-9784.

P18: Issues with tissues: sample stability in NMR-based metabolomics with HR-MAS spectroscopy

C. Trautwein[1], P. Wenk[1], M. A. Krueger[1], A. Schmid[1], B. J. Pichler[1]

[1] Werner Siemens Imaging Center, University Hospital Tuebingen, Germany

Objective: The benefit of *ex-vivo* HR-MAS spectroscopy of intact biopsies alongside *in-vivo* MRI and MRS is unquestioned. However, fast spinning in order to suppress water sidebands destroys the sample and ongoing enzymatic reactions are able to heavily alter the metabolomic profile. Recent studies found that most spectral information of human tumor samples in HR-MAS derives not from the tissue itself but the surrounding liquid phase [1]. A short review summarizes current bottlenecks in HR-MAS [2]. We wanted to investigate these processes in more detail for two common tissue types in our lab: rat liver and murine breast cancer tissue. **Methods**: The current state-of-the-art protocol from Beckonert *et al*. [3] claims that at 10° C maximum, spectra can be recorded without significant changes within one hour of rotation at f = 4-6 kHz. We prepared under the guidelines of [3] from each tissue five technical replicates (~ 12-16 mg) and recorded ten consecutive *noesy* (64 scans) spectra over a total time of ~ 80 min at f = 5 kHz (5 min tuning, matching, shimming, pulse setting and ~ 75 min acquisition). **Results and Discussion**: An overlay of the processed *noesy* spectra of one liver sample shows that peaks in the aliphatic region strongly increase already within the first hour of measurement (Fig. 1 left), leading to integral enhancements by almost a factor of two. The PCA (Fig. 1 right) of the two tissue data sets compared to five "ideal" samples (copies of ten identical *noesy* rat liver extract spectra) show that both sample types strongly alter due to rotational and enzymatic effects. Slow-speed spinning [4] can be an alternative to avoid the effects of physical decomposition but cannot hamper enzymatic degradation. Therefore new methods and protocols for reliable tissue analysis by HR-MAS need to be developed.

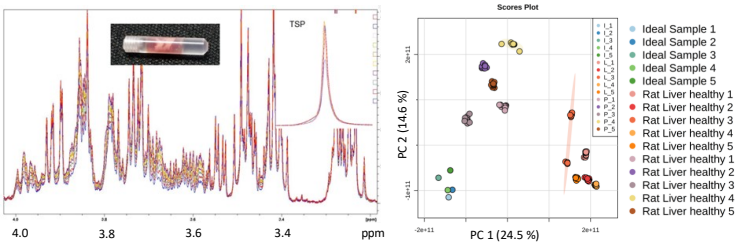


Fig. 1: Overlay of 10 HR-MAS (5 kHz) liver noesy spectra (Δ=7 min) and PCA analysis of ideal vs real samples.

References: [1] Martinez-Bisbal, M.C., et al., *Magnetic Resonance Microscopy Contribution to Interpret High-Resolution Magic Angle Spinning Metabolomic Data of Human Tumor Tissue.* Journal of Biomedicine and Biotechnology, 2011: p. 8.; [2] Esteve, V., B. Martinez-Granados, and M.C. Martinez-Bisbal, *Pitfalls to be considered on the metabolomic analysis of biological samples by HR-MAS.* Frontiers in Chemistry, 2014. **2**: p. 3.; [3] Beckonert, O., et al., *High-resolution magic-angle-spinning NMR spectroscopy for metabolic profiling of intact tissues.* Nature Protocols, 2010. **5**(6): p. 1019-1032.; [4] Andre, M., et al., *Complete Protocol for Slow-Spinning High-Resolution Magic-Angle Spinning NMR Analysis of Fragile Tissues.* Analytical Chemistry, 2014. **86**(21): p. 10749-10754.

Sponsors

MOVISS is supported by