A New Integrated Bioinformatics Tool

for Metabolomic Data Handling

Libiseller Gunnar [1], *Sinner Frank*[1], *Pieber Thomas*[1], *Reichelt Wieland*[2], *Kleb Ulrike*[1], *Dvorzak Michaela*[1], *Magnes Christoph*[1]

CONTACT

1
JOANNEUM RESEARCH
Forschungsgesellschaft mbH

HEALTH
Institute for
Biomedicine and
Health Sciences

Christoph Magnes

Elisabethstraße 5
8010 Graz, Austria

Phone +43 316 876-40 00
Fax +43 316 8769-40 00

christoph.magnes@joanneum.at

health@joanneum.at
www.joanneum.at/health

2
University of Graz

Institute of
Molecular Biosciences

Introduction

Until now there has been no tool that can handle everything from data conversion to compound identification let alone individualized storage of results, statistical analysis or management of projects, studies and experiments within one graphical user interface – i.e., an integrated bioinformatic tool.

Aim

We designed and construct an integrated bioinformatic tool for

1) **processing of LC/FTMS raw data** generated by metabolomics studies, including peak picking, peak alignment and grouping across study samples by integrating the XCMS package[1].
2) **organizing and archiving the processed raw data** for further statistical analysis and compound identification
3) **identification of compounds in the processed raw data** by accurate mass and retention time using system-specific databases and public compound DBs

Methods

- Programming language: **Java**
- Database: **MySQL**
- Statistical programming language: **R**

Results

We created JOANNEUM RESEARCH Metabolite Database – JRMDb (Fig. 1). The GUI is divided into three parts. In the left one tree shows projects, studies and experiments. The lower part shows a history of actions within a session. In the right part tables and plots can be displayed within tabs. JRMDb features:

- **Semi-automated data conversion** using ReadW[1] (from vendor-file format to standardized mzXML files)
- **Automated peak picking**, peak alignment and grouping of multiple sample batches utilizing the XCMS Package[2]
- **Data management system** to hierarchically organize the processed data from projects, studies and experiments
- **Data filtering tools** (e.g. Blank-, QC-, System-filter) (Fig. 2)
- **Compound identification** by
  - accurate mass and retention time using a system specific data base updated and maintained by the institute
  - accurate mass using public metabolite data bases (e.g. HMDB[3], MetaCyc[4], ChEBI[5], LipidMaps[6])
- **Data visualization tools** (e.g. EIC diagrams (Fig. 3), trend analysis, density plots, feature correlation plots, heatmaps, PCAs).

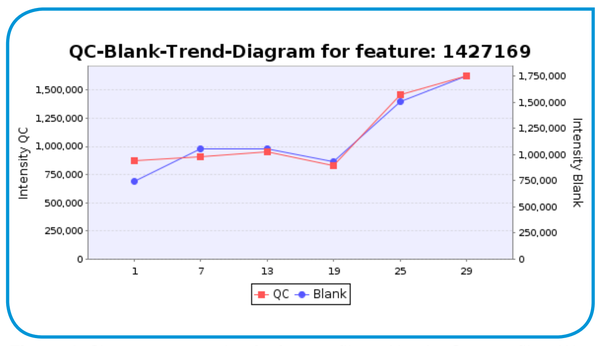

Figure 2*: Shows an example of a feature filtered by the System-filter. We periodically mix in pairs of blank- and QC-samples (mix of all samples) during an analysis. If the intensities behave in the same way we assume the feature is just a peak generated by impurities in the system.*

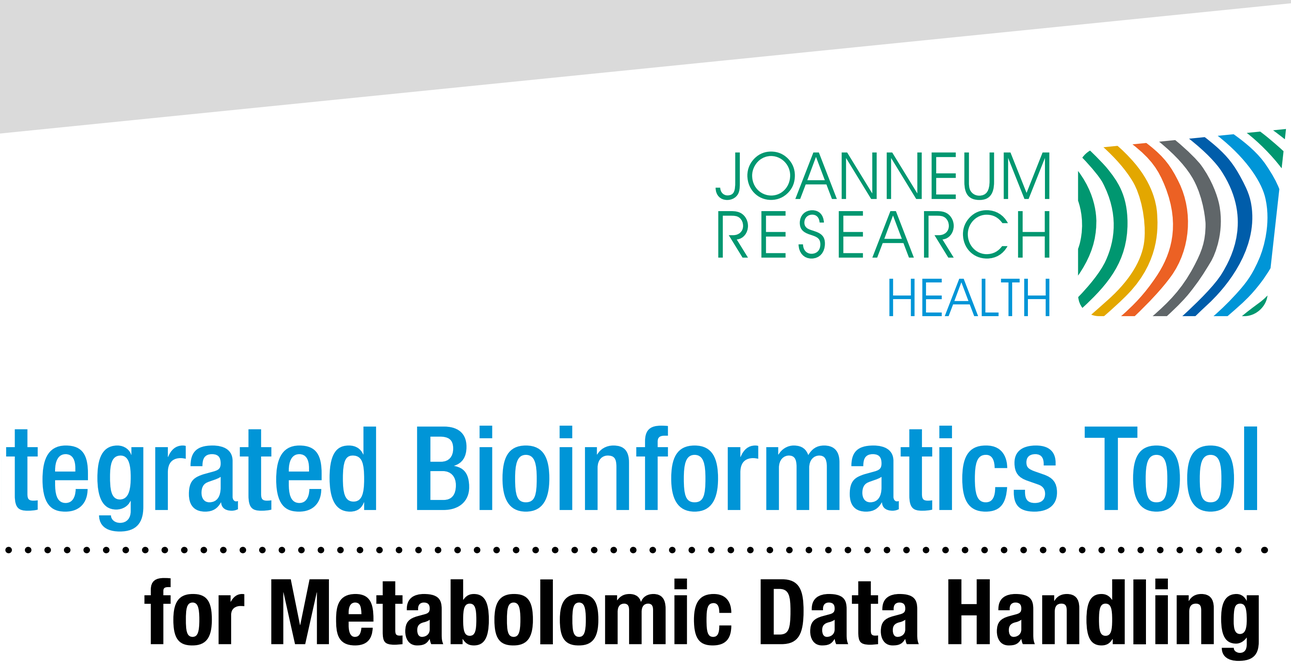

Figure 1*: JRMDb with identified compounds. Green marked ones are within a specified retention time window. The popup-menu shows possible actions.*

Case Study

In order to challenge the software with a rather scientifically problem we compared typical inducers of autophagy in *Saccharomyces cerevesiae*. The aim was to identify key metabolites which distinguish induction of autophagy by chemical treatment (Rapamycin) and induction through nutrient starvation (HBSS/SN-D/SD-N) (Fig. 4).

The workflow was as following:

- Sample extraction (n = 4 per sample group) using trichloroacetic acid
- QC-sample-preparation by mixing equal amounts of each sample in one vial
- Automated analysis-sequence-generation and sample randomization
- LC/FTMS measurement according to[7]
- Data conversion from vendor file format to mzXML
- Automated XCMS-preprocessing
- XCMS-processing detecting 1203 features
- Storing data to MySQL database
- Data filtration reducing the features to 631
- Compound-identification within the remaining features by m/z and retention time
- PCA-creation (Fig. 5)
- Creation of a heat map (Fig. 6)

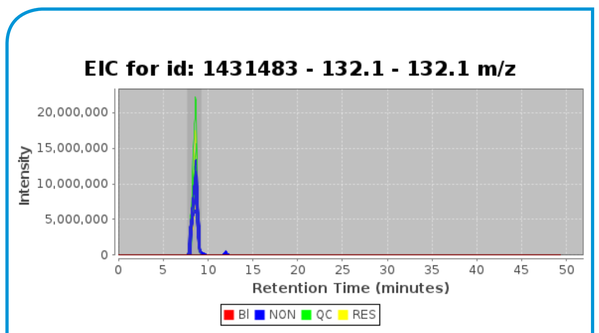

Figure 3*: Extracted ion chromatogram. Samples belonging to the same class get automatically the same color.*

Conclusion

With JRMDb we have consolidated the metabolomic workflow from data conversion and processing as well as study management to statistical and visual interpretation of the metabolomics data within one program. Due to the modular design further tools like normalization and additional plots can be added easily.

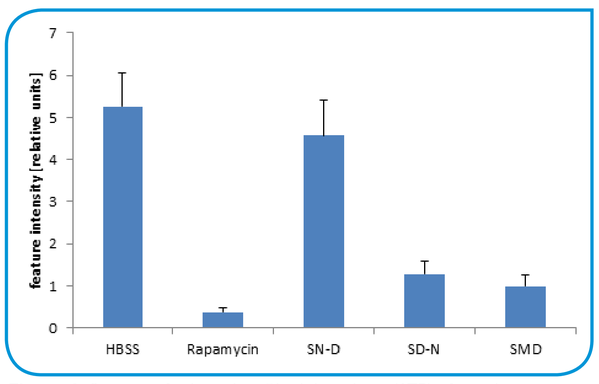

Figure 4*: Content of adenosine-5'-triphosphate (ATP) of saccharomyces cerevesiae normalized on the amount of cells and relative to ATP content of cells grown in SMD (synthetic media containing dextrose) in different autophagy inducing growth conditions. HBSS (Hank's balanced salt solution) buffer containing 0,1% glucose; SD-N synthetic media lacking a nitrogen source; SN-D synthetic media lacking dextrose; rapamycin SMD media containing rapamycin.*

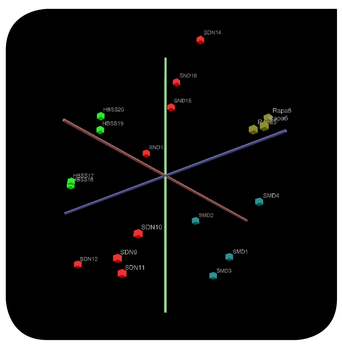

Figure 5*: The 3D-PCA clearly clusters the treatment groups from the case study. See also (Fig. 4).*

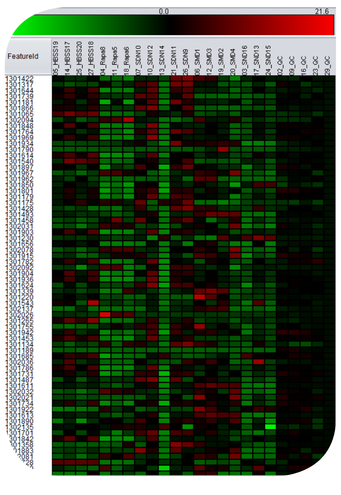

Figure 6*: Shows a heat map generated by JRMDb. The intensities of every feature have been normalized to an average of 0 and a standard deviation of 1. Low intensities are represented by a green and high ones by a red color.*

Acknowledgements

This work was supported financially by the

Austrian Federal Ministry of Transportation, Innovation and Technology (bmvit), Project Met4CAD

and the

Jubiläumsfond of the Austrian Nationalbank (Proj. 13699 to H.S.).

Literature

[1] ReadW: http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW#Current_Version

[2] XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification; Colin A. Smith, et. al; Analytical Chemistry 2006 78 (3), 779–787

[3] Wishart DS, Knox C, Guo AC, et. al., HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009 37 (Database issue): D603–610.

[4] Karp, P.D., Riley, M., Paley, S., and Pellegrini-Toole, A., The MetaCyc Database,Nucleic Acids Research, 30(1):59–61 2002.

[5] Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M. and Ashburner, M. (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 36, D344–D350.

[6] Cotter D, Maer A, Guda C, Saunders B, Subramaniam S. (2006) LMPD: LIPID MAPS proteome database, Nucleic Acids Res. 2006 Jan 1; 34 (Database issue): D507–10

[7] Lu, W., Clasquin, M. F., Melamud, E., Amador-Noguez, D., Caudy, A. A., & Rabinowitz, J. D. (2010). Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Analytical chemistry, 82 (8), 3212–21. doi:10.1021/ac902837x