

SOFTWARE

Open Access



MetHoS: a platform for large-scale processing, storage and analysis of metabolomics data

Konstantinos Tzanakis^{1*} , Tim W. Nattkemper² , Karsten Niehaus³  and Stefan P. Albaum⁴ 

*Correspondence:

ktzan@cebitec.uni-bielefeld.de

¹ International Research Training Group "Computational Methods for the Analysis of the Diversity and Dynamics of Genomes", Faculty of Technology, Bielefeld University, Bielefeld, Germany

² Biodata Mining Group, Center for Biotechnology (CeBiTec), Faculty of Technology, Bielefeld University, Bielefeld, Germany

³ Proteome and Metabolome Research, Center

for Biotechnology (CeBiTec), Faculty of Biology, Bielefeld University, Bielefeld, Germany

⁴ Bioinformatics Resource Facility, Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany

Abstract

Background: Modern mass spectrometry has revolutionized the detection and analysis of metabolites but likewise, let the data skyrocket with repositories for metabolomics data filling up with thousands of datasets. While there are many software tools for the analysis of individual experiments with a few to dozens of chromatograms, we see a demand for a contemporary software solution capable of processing and analyzing hundreds or even thousands of experiments in an integrative manner with standardized workflows.

Results: Here, we introduce MetHoS as an automated web-based software platform for the processing, storage and analysis of great amounts of mass spectrometry-based metabolomics data sets originating from different metabolomics studies. MetHoS is based on Big Data frameworks to enable parallel processing, distributed storage and distributed analysis of even larger data sets across clusters of computers in a highly scalable manner. It has been designed to allow the processing and analysis of any amount of experiments and samples in an integrative manner. In order to demonstrate the capabilities of MetHoS, thousands of experiments were downloaded from the MetaboLights database and used to perform a large-scale processing, storage and statistical analysis in a proof-of-concept study.

Conclusions: MetHoS is suitable for large-scale processing, storage and analysis of metabolomics data aiming at untargeted metabolomic analyses. It is freely available at: <https://methos.cebitec.uni-bielefeld.de/>. Users interested in analyzing their own data are encouraged to apply for an account.

Keywords: Mass spectrometry data, Large-scale metabolomics, Parallel processing, Distributed storage, Distributed analysis

Background

Metabolomics is a unique part of modern life science and molecular biology due to its multidisciplinary requirements: knowledge from biology, chemistry, physics as well as mathematics and statistics needs to be integrated. It deals with the quantification and identification of small molecules called metabolites (< 1500 Da) which are the intermediate and ending products in cellular processes of an organism [1]. The key technology



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

to investigate the metabolites that are abundant in an organism or a tissue is mass spectrometry, where state-of-the-art methods allow gaining tens of thousands of mass spectra within a few minutes which in turn characterize hundreds of potential compounds. The entirety of these metabolites in a cell at a specific moment can be considered as a high-dimensional molecular snapshot of the organism which carries an imprint of all genetic, epigenetic and environmental factors. Thus, one of the main goals of metabolomics research is to bridge the gap between the genotype and phenotype in order to get a complete picture of the internal structure and behavior of a cell [2]. Metabolomics has a wide application in many different fields such as toxicology assessment, nutritional genomics, biomarker discovery and identification, drug development and disease prognosis [3].

High-throughput metabolomics experiments, generally, follow an untargeted approach which is characterized by the simultaneous measurement of a large number of metabolites from each sample, thus analyzing the global metabolomic profile [4]. In this, raw datasets obtained from compound separation and detection techniques, such as Gas Chromatography (GC) or Liquid Chromatography (LC) coupled to Mass spectrometry (MS), are transformed to quantitative metabolite information [5–7]. The general processing strategy includes noise filtering and baseline reduction [8] followed by peak detection [9] and deconvolution [10], chromatographic alignment [11], identification of metabolomic features [12], substitution of missing values [13], normalization [14] and statistical analysis [15, 16].

To support researchers in the complicated and complex analytical workflow, a large number of software tools have been developed. Some of the available tools are focused either on the quantification or identification of metabolites. For instance, iMet-Q [17] and apLCMS [18] deal with the quantification step while Metabolizer [19] with the identification of metabolites. Similarly, MetaBox [20] puts the focus only on the statistical analysis that follows the processing step. Other software packages include both the quantification and identification step like XCMS [21], MetAlign 3.0 [22], MZmine 2 [23], MAVEN [24], mzMatch [25] or MS-Dial [26]. However, they either lack statistical capabilities or are not web-based, which confines them at the usage of computational resources. There are also web-based tools like MeltDB [27], XCMS Online [28] and MetaboAnalyst [29] that not only offer support in data storage and retrieval but also analytical tools for quantification, identification and statistical analysis. However, none of these tools is able to deal with large collections of data sets, referred to as large-scale metabolomics. In the context of this work, we categorize data sets ranging from more than a couple of hundreds to thousands of files to be a large-scale metabolomics data set. For instance, MeltDB is a semi-automated system not prepared for big amounts of data and XCMS Online places the focus on processing with restrictions in the storage and limited statistics, while the recently released MetaboAnalyst 5.0 at least supports up to 200 files.

With repositories giving access to hundreds or even thousands of files from different experiments we see a clear demand for an easy-to-use software solution capable of handling large amounts of metabolomics data in short processing and analysis times. In recent years, first big data frameworks and libraries have been developed which have proven to be capable of dealing with such amounts of data. PhenoMeNal [30] is

a framework that can handle large volumes of data as it utilizes a cluster of computers. However, it follows the Infrastructure-as-a-Service cloud model and requires prior knowledge of cloud computing concepts, which most metabolomics researchers may not be familiar with. In a similar manner, workflow4metabolomics [31] is a repository for Galaxy-based workflows that covers many aspects of metabolomics data processing and analysis. Yet again this requires set up of the technical infrastructure and does not include a scalable storage solution. In this study, we present MetHoS, a ready-to-use web-based platform for large-scale processing, storage and analysis of metabolomics data sets. MetHoS is based on big data frameworks and provides users efficient and user-friendly handling of their own experimental metabolomics data. With Apache Spark [32], Apache Cassandra [33] and KNIME (Konstanz Information Miner) [34] as our fundament we propose a different way of handling large-scale datasets with the prospect of handling even largest amounts of data.

Implementation

Application architecture

Cloud computing allows for the parallel execution of tasks on a large number of virtual machines. Moreover, it allows for scalability: if the size of the problem increases, more machines can easily be added. MetHoS supports horizontal scaling, thus ensuring performance by not being limited to the capacity of a single unit and redundancy with no single point of failure. It is written in Java and utilizes a set of software tools that enable parallel processing, distributed storage and distributed analysis on the cloud (Fig. 1). MetHoS uses a variable number of computer nodes and has been designed for OpenStack, the popular open-source software platform for cloud computing.

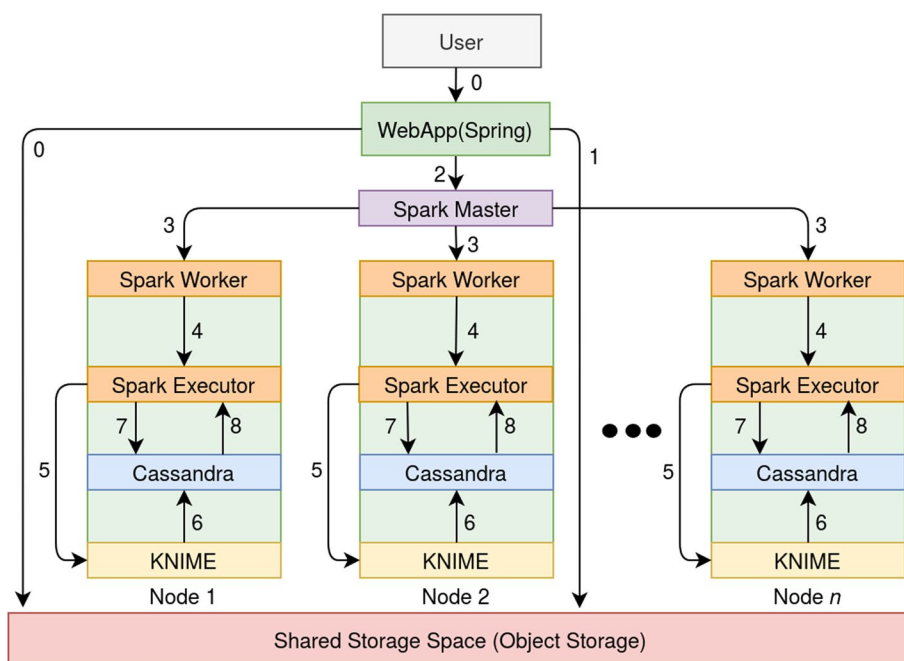


Fig. 1 This figure shows the data-flow during the processing and analysis steps of the combination of Apache Spark, Apache Cassandra and KNIME

The general processing of uploaded metabolomics data is relying on the software KNIME, which is a well-known open-source data workflow engine and analytics platform. Apache Cassandra, a NoSQL database for distributed storage, was selected to store the results of the processing across the cluster. We integrated Apache Spark, a fast in-memory cluster-computing engine, that can perform distributed processing and analysis of big amounts of data. Finally, we combined all three software tools together with the Spring Application framework providing users a friendly and easy-to-understand web-based application.

Project and data management

MetHoS provides sophisticated user and project management, in which uploaded samples can be grouped to create experiments in already pre-defined projects. The owner of a project has the ability to edit it and also manage access rights. MetHoS makes use of OpenStack's object storage through the swift API, where the experiments are stored in a shared storage space (Fig. 1).

An *experiment* in MetHoS refers to a biological experiment that consists of many biological replicates. After uploading, experiments can be processed (quantification and identification step) by selecting one of the automated KNIME workflows. In result, each workflow is also responsible for storing the results straight into the Cassandra database. Apache Spark verifies fair job-scheduling and workload distribution, e.g. in terms of KNIME workflow processing, giving thus the ability to process thousands of experiments in a matter of hours.

Results of the processing are presented in the *View* section. The representation can be either in a form of a table by selecting an experiment and observing all the identified and unidentified metabolomic features or with boxplots by selecting more experiments and observing the metabolomic features among them. Various statistical tests enable further investigation and analysis of the processed metabolomics data within an experiment but, in particular, across a multitude of experiments. The results are presented and visualized with the help of the D3 Javascript library.

Parallel processing

The architecture and design of MetHoS are based on flexibility allowing for an easy integration of workflows implemented in KNIME Analytics Platform. KNIME has integrated OpenMS [35], an open-source software C++ library for management and analyses of metabolomics data, allowing for quantification and identification of metabolites, against spectral databases, in each sample or group of samples.

Once the raw chromatographic data are uploaded to the Openstack object storage, they can be selected for processing (Fig. 1, step 1) with a pre-defined workflow. The Spark master, which is a single coordinator that acquires cluster nodes (Spark Workers), receives the request and distributes the tasks to each node (Fig. 1, steps 2 and 3). Then, each Spark Worker activates a Spark Executor, an agent responsible for carrying out the task and activating KNIME on every node in order to process the selected experiments (Fig. 1, steps 4 and 5).

MetHos supports the well-known and accepted open-source file formats .mzML, .mzData and .mzXML as inputs and produces information about identified and

unidentified metabolites which is automatically being stored in the Cassandra database (Fig. 1, step 6). The workflows currently provided in MetHoS target MS1 and MS2 level data and support identification by exact mass search as well as spectral matching. Figure 2 gives an example of two of these workflows which cover all steps required to analyze metabolomics data: conversion of mass spectra in an appropriate format, the quantification and normalization (Additional file 4: Table S1), the control of ionization mode, corrections of retention time distortions (Additional file 5: Table S2 and Additional file 6: Table S3), the identification of metabolites and the storage of metabolite measurements into the Cassandra database. For the quantification of metabolites both workflows rely on the software tool FeatureFinderMetabo [36] of OpenMS. Concerning the identification of metabolites the first workflow uses the AccurateMassSearch algorithm, which identifies metabolite features by comparing their exact mass to databases like HMDB [37], MassBank [38] and MoNA:Fiehn [39], while the second workflow uses the MetaboliteSpectralMatcher [40], which identifies small molecules from tandem MS spectra using a spectral library such as MassBank.

Distributed storage

The main advantage of a NoSQL database over an SQL database is horizontal scalability and distributed storage, giving the possibility to store any amount of data, just limited by the number of utilized storage nodes. Apache Cassandra is such a NoSQL distributed database. Its masterless architecture appoints it unique among other NoSQL databases as it ensures high availability of data at all times. In combination with adjusting

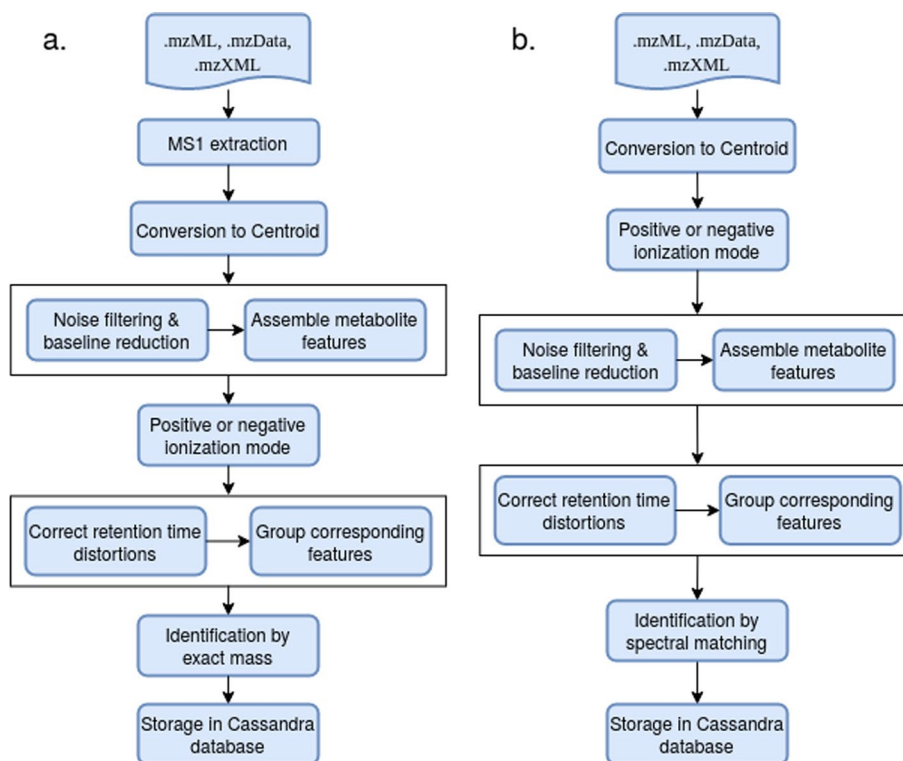


Fig. 2 This figure shows the steps of two of the workflows currently implemented in MetHoS. **a** Identification by exact mass, **b** Identification by spectral matching

the replication factor to three in our Cassandra model, we are able to provide a no single point of failure model.

Cassandra-specific, our data model (Fig. 3) was modeled around queries, with the *experiment* column family being the most important where the processing results of millions of metabolomic features are stored with the KNIME workflows. Furthermore, with an appropriate partitioning within Cassandra, it is made sure that data is equally distributed and stored efficiently.

Distributed statistical analysis

Apache Spark is a fast, distributed in-memory large-scale data processing engine providing powerful Machine Learning algorithms, with their own library, which perform statistical tests in a distributed manner. MetHoS uses the Apache Spark Machine Learning library (Spark MLlib) in combination with the recently developed data structure called *Dataset*, which provides the convenience of an RDD, less memory consumption and automatic optimization. For every statistical test of our application, Apache Spark creates *jobs* which are comprised from *tasks* that are distributed from the Spark master and executed from the Spark executors on data *partitions*. Each executor of a node has been assigned a number of cores and the more cores can be used the more tasks can be performed in parallel.

The statistical methods for analysis available in MetHoS are:

- Basic statistics (mean and standard deviation)
- Metabolite filtering (mean and standard deviation between two defined groups of experiments)
- Pearson correlation
- Spearman correlation
- Principal component analysis (PCA)
- Clustering (K-means, Bisecting k-means)

In our analyses, we provide a set of choices depending on the desired depth of analysis. Specific metabolites can be selected to set the focus of the analysis. If required, handling of missing values can be incorporated. In the default settings, analyses are conducted on any common metabolite that exists in every selected experiment. Respectively, for the replicate level, the list of the metabolites that exist at least once in every replicate of the selected experiments will be selected for analysis. Missing values strategies include replacement with zero, mean or median. Results of an analysis, e.g. basic statistics such

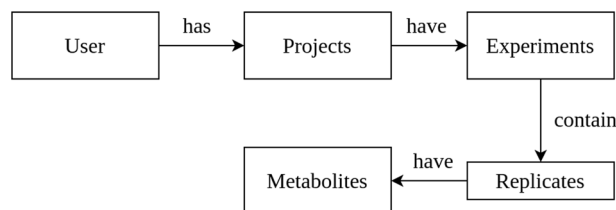


Fig. 3 This figure shows the conceptual data model of the Cassandra database

as mean and standard deviation values of all or common metabolites in experiments can be exported in .csv format for further analysis and observation.

Results

The capabilities of MetHoS are presented with more detail in the following evaluation and a use case in which we performed a large-scale processing, storage and statistical analysis in thousands of experiments downloaded from the MetaboLights database belonging to 38 different studies [36, 41–73].

Evaluation of scalability

Processing thousands of experiments in a linear manner with the traditional methods would take several days or weeks depending on the complexity and size of the files. In order to prove the efficiency of parallel processing and the horizontal scalability of MetHoS, we processed 200 experiments originating from a study of MetaboLights repository, under the study identifier MTBLS28, with a variable number of Spark workers (Fig. 4). The results indicated that the more worker nodes are added in the cluster, the faster the processing is completed.

Use case data

In our comprehensive use case, we extended to all the samples of studies from 2012 to 2020 which were downloaded and grouped in experiments, after excluding problematic and corrupted files. They originate from mass spectrometry experiments of the human organism which contain .mzData, .mzML and .mzXML files (Additional file 7: Table S4). Uploading the experiments in MetHoS resulted in 4827 experiments occupying approximately 1.1 Tb of disk space in the Openstack object storage space.

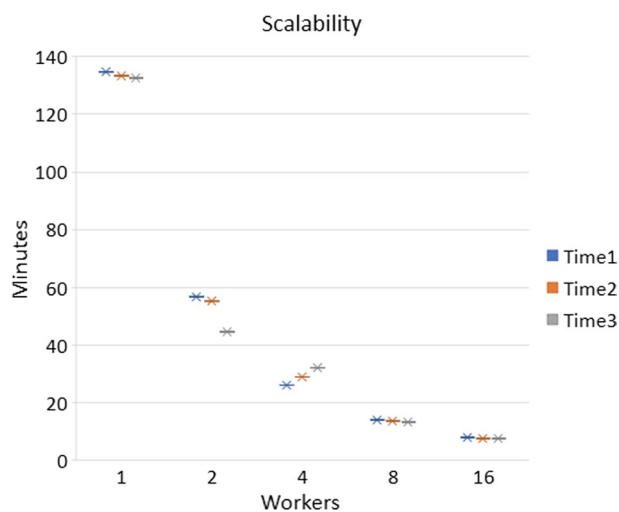


Fig. 4 This figure shows the scalability of MetHoS with 1, 2, 4, 8 and 16 Spark workers compare to the time it takes to process 200 experiments. The processing was performed 3 times on the same 200 experiments for every number of workers

Processing

The processing with MetHoS lasted approximately 12 hours and ended up quantifying and identifying more than 2 billion metabolite features. The results were automatically stored in MetHoS (Fig. 5a).

Analysis results

For our analysis, we performed a k-means clustering in all 4.827 experiments using 15 clusters, with all metabolites, on experiment level and replacing missing values with the mean (Fig. 5b, Additional file 1: Fig. S1). Results show that two big clusters were formed, clusters 1 and 14 (Additional file 8: Table S5, Additional file 9: Table S6). Although there are similarities between several urine samples and blood or blood plasma or blood serum samples, the majority of samples in cluster 1 belong to blood samples and for cluster 14 to urine samples. Clusters 3, 6, 7, 8, 11, 12, 13 and 15 consist exclusively of urine samples, while clusters 2 and 9 only of solvent samples. Moreover, cluster 4 consists only of blood serum samples and cluster 10 only of blood plasma samples. Samples originating from lung, feces, renal tubule or cerebrospinal fluid have been split almost equally in clusters 1 and 14 indicating the formation of two distinct groups possibly originating from two different conditions in each group of experiments. Furthermore, all samples originating from THP-1-cell and Breath have been clustered together in cluster 1 and all samples from MCF-10A-cell and umbilical vein endothelial cell line in cluster 14. Finally, cluster 5 suggests similarities between some solvent and urine samples.

We selected 201 experiments (Additional file 7: Table S4) of blood and urine and performed a Principal Component Analysis on experiment level, selecting all metabolites, replacing missing values with zero and using the z-score normalization. The analysis ended up differentiating the urine samples from the blood samples successfully in an interactive PCA plot (Fig. 5c). It is shown that the intensities of 40.101 metabolites from 201 experiments took part in the calculation of the PCA, while metabolites that were not present in all the experiments are depicted in the table next to the PCA plot.

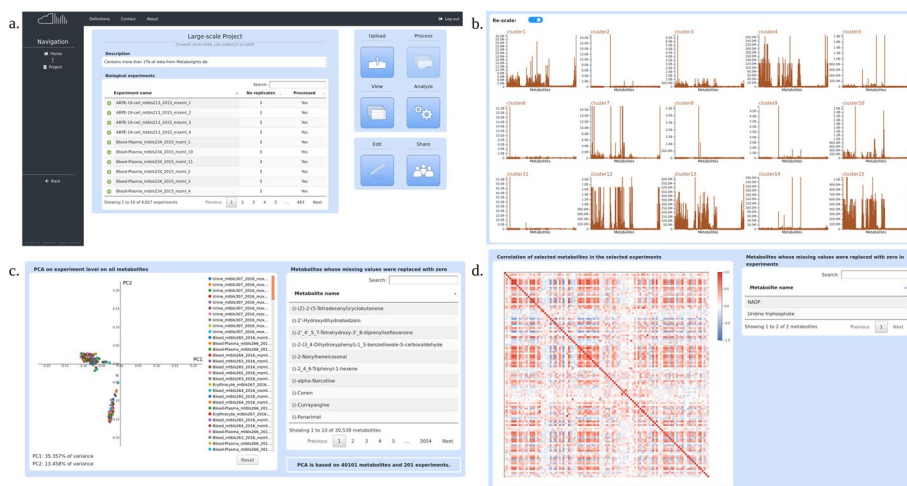


Fig. 5 **a** Web interface of a project. **b** K-means clustering on all 4827 experiments (re-scaled). **c** PCA analysis of 144 experiments originating from whole blood, blood plasma and erythrocyte samples and 57 experiments originating from urine samples. **d** Pearson Correlation of 90 experiments on 112 compounds

Afterwards, 90 experiments (Additional file 7: Table S4) were selected, containing blood samples from 30 individuals, 15 young and 15 elderly (30 whole blood, 30 erythrocyte and 30 blood plasma samples). Pearson Correlation was implemented on 112 compounds (Additional file 10: Table S7) that according to literature [47], show age-related increases or decreases while replacing the missing values with zero (Fig. 5d).

The results indicated that Fructose 6-phosphate and Glucose 6-phosphate are highly positively correlated. The same stands for 2-Phosphoglyceric acid and 2,3-Diphosphoglyceric acid which have a very strong correlation. Closely correlated are also Nicotinamide Adenine Dinucleotide (NAD) and Nicotinamide Adenine Dinucleotide Phosphate (NADP). Last but not least, there were missing values of NADP and Uridine Triphosphate which indicates that they were not present in all 90 experiments.

Thereafter, we selected 45 of the 90 aforementioned experiments that contain samples originating from the 15 young individuals and performed a k-means clustering with 10 clusters, on metabolite level and replacing missing values with zero. The same was implemented for the rest 45 experiments of the 15 elder individuals (Additional file 2: Fig. S2, Additional file 3: Fig. S3).

The results suggest that metabolites like Leucine and Isoleucine, which may play a distinct role in supporting skeletal muscle activity, are clustered together in both cases (Additional file 11: Table S8, Additional file 12: Table S9). Ergothioneine is clustered alone in both cases showing more fluctuations in the elder individuals. In both young and elder clusters, Adenosine Diphosphate (ADP) and NAD are clustered together while Adenosine Triphosphate (ATP) is clustered separately in both cases. L-Acetylcarnitine is clustered separately for the young people while for the elder people it is clustered together with metabolites that are involved in the glucose metabolism (2-phosphoglyceric acid, 3-phosphoglyceric acid, Guanosine Triphosphate (GTP), NADP and Uridine Diphosphate Glucose) and shows a decrease of its abundance in the elder people.

Cluster setup

MetHoS is using Apache Spark 3.0.1 in standalone mode and currently uses 16 worker nodes and 1 master node. Each Spark worker has one executor with two cores making it possible to parallelize two tasks per worker or 32 tasks in total. We provide 23 Gb of RAM to every executor and 115 Gb to the master node. Spark is able to access the Cassandra database through the Spark-Cassandra-Connector 3.0.0 while it is also authorized to access Openstack Object Storage for downloading the experiments to be processed every time.

In the same 16 computer nodes, we have also installed KNIME providing it with 2 Gb of RAM. Spark downloads an experiment locally on the node and activates KNIME so that it can be processed on the same node. Consequently, a number of 32 experiments can be processed in parallel on the cluster. For our Cassandra setup, we are using the same 16 nodes used as Spark workers.

Conclusions

Here we introduced MetHoS as a flexible and easy-to-use web-based platform, based on big data frameworks, that provides automated processing, distributed storage and distributed analysis in short processing and analysis times (Additional file 13). Our aim was to provide users a bioinformatics platform for the efficient and user-friendly handling

of experimental data originating from different metabolomics studies allowing in that way the integration of metabolomics data. MetHoS is built on Apache Spark to enable metabolomics data processing and analysis using KNIME and SparkML but also to constitute the basis for future analysis functionalities. We evaluated the scalability of the platform using a variable number of Spark workers and demonstrated its capabilities by handling 1.1 Tb of data which were processed in only 12 h ending up to more than 2 billion metabolite features. MetHoS allows for automated processing of large numbers of chromatographic datasets in terms of untargeted metabolite profiling, quantification and de novo identification and by that reaches a time-efficient and target-oriented integration and interpretation of metabolomics data.

Abbreviations

MS	Mass spectrometry
GC	Gas chromatography
LC	Liquid chromatography
KNIME	Konstanz Information Miner
MMLib	Machine Learning library
PCA	Principal component analysis
NAD	Nicotinamide Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
GTP	Guanosine Triphosphate

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-022-04793-w>.

Additional file 1: Fig. S1. K-means clustering on all 4827 experiments (not re-scaled).

Additional file 2: Fig. S2. K-means clustering of 45 experiments originating of young individuals on 112 compounds, on metabolite level and replacing missing values with zero.

Additional file 3: Fig. S3. K-means clustering of 45 experiments originating of elder individuals on 112 compounds, on metabolite level and replacing missing values with zero.

Additional file 4: Table S1. List of the parameters and their values that are used in the KNIME workflow for assembling the metabolite features.

Additional file 5: Table S2. List of the parameters and their values that are used in the KNIME workflow for the map alignment (correcting retention time distortions between maps).

Additional file 6: Table S3. List of the parameters and their values that are used in the KNIME workflow for grouping corresponding features from multiple maps.

Additional file 7: Table S4. The study identifiers from 38 studies of the MetaboLights repository that were used for processing and analysis.

Additional file 8: Table S5. List of clusters and number of experiments in each one.

Additional file 9: Table S6. List of clusters and number of experiments in each one.

Additional file 10: Table S7. List of the metabolites as they appear in Pearson Correlation heatmap starting from the top left corner.

Additional file 11: Table S8. Clusters of the metabolites of 45 experiments originating of 15 young individuals.

Additional file 12: Table S9. Clusters of the metabolites of 45 experiments originating of 15 elder individuals.

Additional file 13. A walk through MetHoS.

Acknowledgements

The authors wish to thank the de.NBI cloud administrators for the expert technical support. This work was supported by grants of the German Federal Ministry of Education and Research (BMBF) of the project 'Bielefeld-Gießen Center for Microbial Bioinformatics—BiGi' (Grant number 031A533) within the German Network for Bioinformatics Infrastructure (de.NBI). We acknowledge support for the publication costs by the Open Access Publication Fund of Bielefeld University and the Deutsche Forschungsgemeinschaft (DFG).

Authors contributions

KT designed and implemented the software. TWN, KN and SPA initiated, supervised, and directed the project. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. KT received funding from the German-Canadian DFG international research training group 'Computational Methods for the Analysis of the Diversity and Dynamics of Genomes' (DiDy) GRK 1906/1. The funding body did not play any role in the design of the study, writing of the manuscript nor did they have any influence on the data collection, analysis or interpretation of the data and results. Contents are the authors' sole responsibility and do not necessarily represent official DFG views.

Availability of data and materials

The datasets used and analysed during the current study are publicly available in the MetaboLights repository at <https://www.ebi.ac.uk/metabolights/> and their study identifiers can be found in the Additional file 7: Table S4. Source code is available at <https://gitlab.uni-bielefeld.de>. Project name: MetHoS; Project home page: <https://methos.cebitec.uni-bielefeld.de/>; Operating system(s): Platform independent; Programming language: Java, Javascript, HTML, CQL; Other requirements: Apache Spark, Apache Cassandra, Knime, Spring Framework, OpenStack; License: GPLv3; Any restrictions to use by non-academics: None.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 21 June 2021 Accepted: 14 June 2022

Published online: 08 July 2022

References

1. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW. Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 2004;9(9):418–25. <https://doi.org/10.1016/j.tplants.2004.07.004>.
2. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol.* 2002;48(1):155–71. <https://doi.org/10.1023/A:1013713905833>.
3. Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: the future of metabolomics in a personalized world. *New Horiz Transl Med.* 2017;3(4):294–305. <https://doi.org/10.1016/j.nhtm.2017.06.001>.
4. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol.* 2015;3:23. <https://doi.org/10.3389/fbioe.2015.00023>.
5. Villas-Bôas SG, Højer-Pedersen J, Akesson M, Smedsgaard J, Nielsen J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast.* 2005;22(14):1155–69. <https://doi.org/10.1002/yea.1308>.
6. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev.* 2007;26(1):51–78. <https://doi.org/10.1002/mas.20108>.
7. El-Anead A, Cohen A, Banoub J. Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Appl Spectrosc Rev.* 2009;44(3):210–30. <https://doi.org/10.1080/05704920902717872>.
8. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem.* 2003;75(18):4818–26. <https://doi.org/10.1021/ac026468x>.
9. Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform.* 2008;9(1):504. <https://doi.org/10.1186/1471-2105-9-504>.
10. Du X, Zeisel SH. Spectral deconvolution for gas chromatography mass spectrometry-based metabolomics: current status and future perspectives. *Comput Struct Biotechnol J.* 2013;4(5):201301013. <https://doi.org/10.5936/csbi.201301013>.
11. Zhou B, Xiao JF, Tuli L, Ransom HW. LC–MS-based metabolomics. *Mol BioSyst.* 2012;8:470–81. <https://doi.org/10.1039/C1MB05350G>.
12. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G. METLIN: a metabolite mass spectral database. *Ther Drug Monit.* 2005;27(6):747–51. <https://doi.org/10.1097/01.ftd.0000179845.53213.39>.
13. Steinfath M, Groth D, Liseč J, Selbig J. Metabolite profile analysis: from raw data to regression and classification. *Physiol Plant.* 2008;132(2):150–61. <https://doi.org/10.1111/j.1399-3054.2007.01006.x>.
14. Xi B, Gu H, Baniyasadi H, Raftery D. Statistical analysis and modeling of mass spectrometry-based metabolomics data. *Methods Mol Biol (Clifton, NJ).* 2014;1198:333–53. https://doi.org/10.1007/978-1-4939-1258-2_22.
15. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. *Metabolites.* 2012;2(4):775–95. <https://doi.org/10.3390/metabo2040775>.

16. Worley B, Powers R. Multivariate analysis in metabolomics. *Metabolomics*. 2013;1(1):92–107.
17. Chang H-Y, Chen C-T, Lih TM, Lynn K-S, Juo C-G, Hsu W-L, Sung T-Y. iMet-Q: a user-friendly tool for label-free metabolomics quantitation using dynamic peak-width determination. *PLoS ONE*. 2016;11(1):1–18. <https://doi.org/10.1371/journal.pone.0146112>.
18. Yu T, Park Y, Johnson JM, Jones DP. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* (Oxford, England). 2009;25(15):1930–6. <https://doi.org/10.1093/bioinformatics/btp291>.
19. Mak TD, Laiakis EC, Goudarzi M, Fornace AJ. Metabolizer: a novel statistical workflow for analyzing postprocessed LC–MS metabolomics data. *Anal Chem*. 2014;86(1):506–13. <https://doi.org/10.1021/ac402477z>.
20. Wanichthanarak K, Fan S, Grapov D, Barupal DK, Fiehn O. Metabox: a toolbox for metabolomic data analysis, interpretation and integrative exploration. *PLoS ONE*. 2017;12(1):1–14. <https://doi.org/10.1371/journal.pone.0171046>.
21. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78(3):779–87. <https://doi.org/10.1021/ac051437y>.
22. Lommen A, Kools HJ. Metalign 3.0: performance enhancement by efficient use of advances in computer hardware. *Metabolomics Off J Metabolomic Soc*. 2012;8(4):719–26. <https://doi.org/10.1007/s11306-011-0369-1>.
23. Pluskal T, Castillo S, Villar-Briones A, Orešič M. Mzmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinform*. 2010;11(1):395. <https://doi.org/10.1186/1471-2105-11-395>.
24. Clasquin MF, Melamud E, Rabinowitz JD. LC–MS data processing with maven: a metabolomic analysis and visualization engine. *Curr Protoc Bioinform*. 2012;Chapter 14:14–11. <https://doi.org/10.1002/0471250953.bi1411s37>.
25. Scheltema RA, Jankevics A, Jansen RC, Swertz MA, Breitling R. PeakML/mzMatch: a file format, java library, r library, and tool-chain for mass spectrometry data analysis. *Anal Chem*. 2011;83(7):2786–93. <https://doi.org/10.1021/ac2000994>.
26. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;12(6):523–6. <https://doi.org/10.1038/nmeth.3393>.
27. Kessler N, Neuweiger H, Bonte A, Langenkämper G, Niehaus K, Nattkemper TW, Goesmann A. MeltDB 2.0—advances of the metabolomics software system. *Bioinformatics*. 2013;29(19):2452–9. <https://doi.org/10.1093/bioinformatics/btt414>.
28. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS online: a web-based platform to process untargeted metabolomic data. *Anal Chem*. 2012;84(11):5035–9. <https://doi.org/10.1021/ac300698c>.
29. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-E, Li S, Xia J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021. <https://doi.org/10.1093/nar/gkab382>.
30. Peters K, Bradbury J, Bergmann S, Capuccini M, Cascante M, de Atauri P, Ebbels TMD, Foguet C, Glen R, Gonzalez-Beltran A, Günther UL, Handakas E, Hankemeier T, Haug K, Herman S, Holub P, Izzo M, Jacob D, Johnson D, Jourdan F, Kale N, Karaman I, Khalili B, Emami Khonsari P, Kultima K, Lampa S, Larsson A, Ludwig C, Moreno P, Neumann S, Novella JA, O'Donovan C, Pearce JTM, Peluso A, Piras ME, Pireddu L, Reed MAC, Rocca-Serra P, Roger P, Rosato A, Rueedi R, Ruttkies C, Sadawi N, Salek RM, Sansone S-A, Selivanov V, Spjuth O, Schober D, Thévenot EA, Tomasoni M, van Rijswijk M, van Vliet M, Viant MR, Weber RJM, Zanetti G, Steinbeck C. Phenomenal: processing and analysis of metabolomics data in the cloud. *GigaScience*. 2018;8(2):giy149. <https://doi.org/10.1093/gigascience/giy149>.
31. Giacomoni F, Le Corguillé G, Monsoor M, Landi M, Pericard P, Pétéra M, Duperier C, Tremblay-Franco M, Martin J-F, Jacob D, Goullitquer S, Thévenot EA, Caron C. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics*. 2014;31(9):1493–5. <https://doi.org/10.1093/bioinformatics/btu813>.
32. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M. A view of cloud computing. *Commun ACM*. 2010;53(4):50–8. <https://doi.org/10.1145/1721654.1721672>.
33. Lakshman A, Malik P. Cassandra: a decentralized structured storage system. *SIGOPS Oper Syst Rev*. 2010;44(2):35–40. <https://doi.org/10.1145/1773912.1773922>.
34. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B. KNIME: the konstantz information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors. *Data analysis, machine learning and applications*. Berlin: Springer; 2008. p. 319–26.
35. Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich H-C, Gutenbrunner P, Kenar E, Liang X, Nahnsen S, Nilse L, Pfeuffer J, Rosenberger G, Rurik M, Schmitt U, Veit J, Walzer M, Wojnar D, Wolski WE, Schilling O, Choudhary JS, Malmström L, Aebersold R, Reinert K, Kohlbacher O. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods*. 2016;13(9):741–8. <https://doi.org/10.1038/nmeth.3959>.
36. Kenar E, Franken H, Forcisi S, Wörmann K, Häring H-U, Lehmann R, Schmitt-Kopplin P, Zell A, Kohlbacher O. Automated label-free quantification of metabolites from liquid chromatography–mass spectrometry data. *Mol Cell Proteomics MCP*. 2014;13(1):348–59. <https://doi.org/10.1074/mcp.M113.031278>.
37. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorn Dahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Res*. 2013;41(Database issue):801–7. <https://doi.org/10.1093/nar/gks1065>.
38. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom*. 2010;45(7):703–14. <https://doi.org/10.1002/jms.1777>.
39. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem*. 2009;81(24):10038–48. <https://doi.org/10.1021/ac901952z>.

40. Fenyő D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem*. 2003;75(4):768–74. <https://doi.org/10.1021/ac0258709>.
41. Haug K, Cochran K, Nainala VC, Williams M, Chang J, Jayaseelan KV, O'Donovan C. MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res*. 2019;48(D1):440–4. <https://doi.org/10.1093/nar/gkz1019>.
42. Capellades J, Navarro M, Samino S, Garcia-Ramirez M, Hernandez C, Simo R, Vinaixa M, Yanes O. geoRge: a computational tool to detect the presence of stable isotope labeling in LC/MS-based untargeted metabolomics. *Anal Chem*. 2016;88(1):621–8. <https://doi.org/10.1021/acs.analchem.5b03628>.
43. Cai X, Dong J, Liu J, Zheng H, Kaweeteerawat C, Wang F, Ji Z, Li R. Multi-hierarchical profiling the structure–activity relationships of engineered nanomaterials at nano–bio interfaces. *Nat Commun*. 2018;9(1):4416. <https://doi.org/10.1038/s41467-018-06869-9>.
44. Singh KD, Tancev G, Decrue F, Usemann J, Appenzeller R, Barreiro P, Jaumà G, Santiago MM, de Miguel GV, Frey U, Sinues P. Standardization procedures for real-time breath analysis by secondary electrospray ionization high-resolution mass spectrometry. *Anal Bioanal Chem*. 2019;411(19):4883–98. <https://doi.org/10.1007/s00216-019-01764-8>.
45. Qiong Z, Xiaofeng Y, Haifang W, Xing W, Xin L, Yao L, Xiaohe Z, Chen F, Haixia L, Yurong Q. Fecal metabolomics and potential biomarkers for systemic lupus erythematosus. *Front Immunol*. 2019;10:976. <https://doi.org/10.3389/fimmu.2019.00976>.
46. [dataset] Christina R, Rurik M, Kohlbacher O, Huber CG. Multi-omics toxicity profiling of engineered nanomaterials. *metabolights_dataset*, V1. 2006. <https://www.ebi.ac.uk/metabolights/MTBLS277>
47. Chaleckis R, Murakami I, Takada J, Kondoh H, Yanagida M. Individual variability in human blood metabolites identifies age-related differences. *Proc Natl Acad Sci*. 2016;113(16):4252–9. <https://doi.org/10.1073/pnas.1603023113>.
48. Ranninger C, Schmidt LE, Rurik M, Limonciel A, Jennings P, Kohlbacher O, Huber CG. Improving global feature detectabilities through scan range splitting for untargeted metabolomics by high-performance liquid chromatography–orbitrap mass spectrometry. *Anal Chim Acta*. 2016;930:13–22. <https://doi.org/10.1016/j.aca.2016.05.017>.
49. Ranninger C, Rurik M, Limonciel A, Jennings P, Kohlbacher O, Huber CG. Nephron toxicity profiling via untargeted metabolome analysis employing a high performance liquid chromatography–mass spectrometry-based experimental and computational pipeline. *J Biol Chem*. 2015;290(31):19121–32. <https://doi.org/10.1074/jbc.M115.644146>.
50. Herman S, Khoonsari PE, Aftab O, Krishnan S, Strömbom E, Larsson R, Hammerling U, Spjuth O, Kultima K, Gustafsson M. Mass spectrometry based metabolomics for in vitro systems pharmacology: pitfalls, challenges, and computational solutions. *Metabolomics*. 2017;13(7):79. <https://doi.org/10.1007/s11306-017-1213-z>.
51. Vincent IM, Daly R, Courtioux B, Cattanach AM, Biéler S, Ndung'u JM, Bisser S, Barrett MP. Metabolomics identifies multiple candidate biomarkers to diagnose and stage human African trypanosomiasis. *PLOS Negl Trop Dis*. 2016;10(12):1–20. <https://doi.org/10.1371/journal.pntd.0005140>.
52. Venturini G, Malagrino PA, Padilha K, Tanaka LY, Laurindo FR, Dariolli R, Carvalho VM, Cardozo KHM, Krieger JE, Pereira ADC. Integrated proteomics and metabolomics analysis reveals differential lipid metabolism in human umbilical vein endothelial cells under high and low shear stress. *Am J Physiol Cell Physiol*. 2019;317(2):326–38. <https://doi.org/10.1152/ajpcell.00128.2018>.
53. Murakami I, Chaleckis R, Pluskal T, Ito K, Hori K, Ebe M, Yanagida M, Kondoh H. Metabolism of skin-absorbed resveratrol into its glucuronized form in mouse skin. *PLoS ONE*. 2014;9(12):1–20. <https://doi.org/10.1371/journal.pone.0115359>.
54. Herman S, Khoonsari PE, Tolf A, Steinmetz J, Zetterberg H, Åkerfeldt T, Jakobsson P-J, Larsson A, Spjuth O, Burman J, Kultima K. Integration of magnetic resonance imaging and protein and metabolite CSF measurements to enable early diagnosis of secondary progressive multiple sclerosis. *Theranostics*. 2018;8:4477–90. <https://doi.org/10.7150/thno.26249>.
55. Herman S, Niemelä V, Khoonsari PE, Sundblom J, Burman J, Landtblom A-M, Spjuth O, Nyholm D, Kultima K. Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington's disease subjects. *Sci Rep*. 2019;9(1):4129. <https://doi.org/10.1038/s41598-019-40186-5>.
56. Ulaszewska MM, Trost K, Stanstrup J, Tuohy KM, Franceschi P, Chong MF-F, George T, Minihane AM, Lovegrove JA, Mattivi F. Urinary metabolomic profiling to identify biomarkers of a flavonoid-rich and flavonoid-poor fruits and vegetables diet in adults: the flavurs trial. *Metabolomics*. 2016;12(2):32. <https://doi.org/10.1007/s11306-015-0935-z>.
57. Mathé EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, Shields PG, Idle JR, Smith PB, Anami K, Kazandjian DG, Hatzakis E, Gonzalez FJ, Harris CC. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res*. 2014;74(12):3259–70. <https://doi.org/10.1158/0008-5472.CAN-14-0109>.
58. van der Hoof JJJ, Padmanabhan S, Burgess KEV, Barrett MP. Urinary antihypertensive drug metabolite screening using molecular networking coupled to high-resolution mass spectrometry fragmentation. *Metabolomics*. 2016;12:125. <https://doi.org/10.1007/s11306-016-1064-z>.
59. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res*. 2015;14(8):3322–35. <https://doi.org/10.1021/acs.jproteome.5b00354>.
60. Trošt K, Ulaszewska MM, Stanstrup J, Albanese D, De Filippo C, Tuohy KM, Natella F, Scaccini C, Mattivi F. Host: microbiome co-metabolic processing of dietary polyphenols—an acute, single blinded, cross-over study with different doses of apple polyphenols in healthy subjects. *Food Res Int*. 2018;112:108–28. <https://doi.org/10.1016/j.foodres.2018.06.016>.
61. Samino S, Vinaixa M, Díaz M, Beltran A, Rodríguez MA, Mallol R, Heras M, Cabre A, Garcia L, Canela N, de Zegher F, Correig X, Ibáñez L, Yanes O. Metabolomics reveals impaired maturation of HDL particles in adolescents with hyperinsulinaemic androgen excess. *Sci Rep*. 2015;5:11496. <https://doi.org/10.1038/srep11496>.
62. Chen G, Walmsley S, Cheung GCM, Chen L, Cheng C-Y, Beuerman RW, Wong TY, Zhou L, Choi H. Customized consensus spectral library building for untargeted quantitative metabolomics analysis with data independent

- acquisition mass spectrometry and metabodia workflow. *Anal Chem.* 2017;89(9):4897–906. <https://doi.org/10.1021/acs.analchem.6b05006>.
63. Leal-Witt MJ, Ramon-Krauel M, Samino S, Llobet M, Cuadras D, Jimenez-Chillaron JC, Yanes O, Lerin C. Untargeted metabolomics identifies a plasma sphingolipid-related signature associated with lifestyle intervention in prepubertal children with obesity. *Int J Obesity.* 2018;42(1):72–8. <https://doi.org/10.1038/ijo.2017.201>.
 64. Rochat B, Mohamed R, Sottas P-E. LC-HRMS metabolomics for untargeted diagnostic screening in clinical laboratories: a feasibility study. *Metabolites.* 2018;8(2):39. <https://doi.org/10.3390/metabo8020039>.
 65. Teruya T, Chaleckis R, Takada J, Yanagida M, Kondoh H. Diverse metabolic reactions activated during 58-hr fasting are revealed by non-targeted metabolomic analysis of human blood. *Sci Rep.* 2019;9(1):854. <https://doi.org/10.1038/s41598-018-36674-9>.
 66. Chaleckis R, Ebe M, Pluskal T, Murakami I, Kondoh H, Yanagida M. Unexpected similarities between the schizosaccharomyces and human blood metabolomes, and novel human metabolites. *Mol BioSyst.* 2014;10:2538–51. <https://doi.org/10.1039/C4MB00346B>.
 67. Decuyper S, Maltha J, Deborggraave S, Rattray NJW, Issa G, Bérenger K, Lompo P, Tahita MC, Ruspasinghe T, McConville M, Goodacre R, Tinto H, Jacobs J, Carapetis JR. Towards improving point-of-care diagnosis of non-malaria febrile illness: a metabolomics approach. *PLOS Negl Trop Dis.* 2016;10(3):0004480. <https://doi.org/10.1371/journal.pntd.0004480>.
 68. Domingo-Almenara X, Brezmes J, Vinaixa M, Samino S, Ramirez N, Ramon-Krauel M, Lerin C, Díaz M, Ibáñez L, Correig X, Perera-Lluna A, Yanes O. eRah: A computational tool integrating spectral deconvolution and alignment with quantification and identification of metabolites in GC/MS-based metabolomics. *Anal Chem.* 2016;88(19):9821–9. <https://doi.org/10.1021/acs.analchem.6b02927>.
 69. To KKW, Lee K-C, Wong SSY, Sze K-H, Ke Y-H, Lui Y-M, Tang BSF, Li IWS, Lau SKP, Hung IFN, Law C-Y, Lam C-W, Yuen K-Y. Lipid metabolites as potential diagnostic and prognostic biomarkers for acute community acquired pneumonia. *Diagn Microbiol Infect Dis.* 2016;85(2):249–54. <https://doi.org/10.1016/j.diagmicrobio.2016.03.012>.
 70. Laursen MR, Hansen J, Elkjær C, Stavnager N, Nielsen CB, Pryds K, Johnsen J, Nielsen JM, Bøtker HE, Johannsen M. Untargeted metabolomics reveals a mild impact of remote ischemic conditioning on the plasma metabolome and α -hydroxybutyrate as a possible cardioprotective factor and biomarker of tissue ischemia. *Metabolomics.* 2017;13(6):67. <https://doi.org/10.1007/s11306-017-1202-2>.
 71. Schoeman JC, Moutloatse GP, Harms AC, Vreeken RJ, Scherpbier HJ, Van Leeuwen L, Kuijpers TW, Reinecke CJ, Berger R, Hankemeier T, Bunders MJ. Fetal metabolic stress disrupts immune homeostasis and induces proinflammatory responses in human immunodeficiency virus type 1- and combination antiretroviral therapy-exposed infants. *J Infect Dis.* 2017;216(4):436–46. <https://doi.org/10.1093/infdis/jix291>.
 72. Lamichhane S, Ahonen L, Dyrlund TS, Kempainen E, Siljander H, Hyöty H, Ilonen J, Toppari J, Veijola R, Hyötyläinen T, Knip M, Oresic M. Dynamics of plasma lipidome in progression to islet autoimmunity and type 1 diabetes—type 1 diabetes prediction and prevention study (DIPP). *Sci Rep.* 2018;8(1):10635. <https://doi.org/10.1038/s41598-018-28907-8>.
 73. Sen P, Carlsson C, Virtanen SM, Simell S, Hyöty H, Ilonen J, Toppari J, Veijola R, Hyötyläinen T, Knip M, Orešič M. Persistent alterations in plasma lipid profiles before introduction of gluten in the diet associated with progression to celiac disease. *Clin Transl Gastroenterol.* 2019;10(5):00044. <https://doi.org/10.14309/ctg.0000000000000044>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

