GCTOF Primary metabolite analysis

Sample Preparation:
30µl plasma samples or 5 mg tissue samples are extracted by 1 ml of degassed acetonitrile : water (95:5, v/v) at –20°C, centrifuged and decanted with subsequent evaporation of the solvent to complete dryness. A clean-up step with acetonitrile/water (1:1) removes lipids. The cleaned extract is aliquoted into two equal portions and the supernatant is dried down again. Internal standards C08-C30 FAMEs are added and the sample is derivatized by methoxyamine hydrochloride in pyridine and subsequentlyby N-methyl-N-trimethylsilyltrifluoroacetamide for trimethylsilylation of acidic protons.

GCTOF parameters :
GCTOF analysis are performed by Agilent GC6890/LECO Pegasus III MS (for more details : Fiehn O. et al. Plant J. 53 (2008) 691–704.)
GC conditions:
Column: Restek corporation rtx5Sil-MS (30 m length x 0.25 mm internal diameter with 0.25 µm film made of 95% dimethly/5%diphenylpolysiloxane)
Mobile phase: Helium
Column temperature: 50 to , at 20OC /min , and hold at 330°C for 5 minutes.
Helium Flow-rate: 1 mL min-1
Injection volume: 0.5 µL
Injection: 25 splitless time into a multi-baffled glas liner
Injection temperature: 50°C ramped to 250°C by 12°C s-1
Gradient: 50°C for 1 min, then ramped at 20°C min-1 to 330°C, held constant for 5 min.
Mass Spectrometry conditions:

Scan Rate: 17/specs/sec

Solvent Delay:330sec

Mass Range: 80-Da

ionization energy : -70 eV

transfer line 280°C

ion source: 250°C

HILIC (Amines) Summary

The biogenic amine HILIC analysis workflow involves analysis of polar phase of lipid extraction by ultra high pressure liquid chromatography (UHPLC) on a Waters BEH Amide Column, interfaced to a SCIEX Triple TOF 6600 mass spectrometer (high resolution, accurate mass), with a 16.8 minute total run time. Data are collected in both positive and negative ion mode, and analyzed using MS DIAL, open software for metabolome analysis

Sample Preparation

Extraction is carried out using a bi-phasic solvent system of cold methanol, methyl *tert*-butyl ether (MTBE), and water. In more detail, cold methanol (225 µL is added to a 5mg tissue sample aliquot, which is placed into a 1.5 mL Eppendorf tube. Then, 750 µL of cold MTBE is added, followed by vortexing for 10 s. and shaking for 6 min. at 4ºC. Phase separation is induced by adding 188 µL of mass spec-grade water. After vortexing for 20 s. the sample is centrifuged at 14,000 rpm for 2 min. The upper organic phase is collected in two 300 µL aliquots for lipid analysis polar layer is collected in two 125 µL aliquots for HILIC analysis. One is stored at -20ºC as a backup and the other is evaporated to dryness in a SpeedVac. Dried extracts are resuspended in acetonitrile.

LC/MS parameters

The LC/QTOFMS analyses are performed using an Agilent 1290 Infinity LC system (G4220A binary pump, G4226A autosampler, and G1316C Column Thermostat) coupled to a SCIEX Triple TOF mass spectrometer. Polar compounds are separated on an Acquity UPLC BEH Amide Column, 130Å, 1.7 µm, 2.1 mm X 150 mm maintained at 45°C at a flow-rate of 0.4 mL/min. Solvent pre-heating (Agilent G1316) was used. The mobile phases consist of: Water, 10 mM Ammonium Formate, 0.125% Formic Acid (A) and Acetonitrile: Water (95/5, v/v), 10 mM Ammonium Formate, 0.125% Formic Acid (B)

The gradient is as follows: 0 min 100% (A); 0–2 min 100% (A); 2–7.7 min 30% (A); 7.7–9.5 min 60% (A); 9.5–10.3 min 70% (A); 10.3–12.8 min 0% (A); 12.8–16.8 min 0% (A. A sample volume of 1 µL for positive mode and 3 µL for negative mode is used for the injection. Sample temperature is maintained at 4°C in the autosampler.

SCIEX Triple TOF 6600mass spectrometers are operated with electrospray ionization (ESI) performing full scan in the mass range m/z 50–1200. Number of cycles in MS1 is 1667 with cycle time of 500ms and accumulation time 475ms. Mass spectrometer parameters are as follows (positive mode) Gas Temp 300°C, gas pressures in psi units with : GS1 and GS2 50 psi, CUR: 35. ISVF is 4500V and DP and CE are 10V and 100 V.

References Cited

1. Matyash, V., et al., *Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics.* J Lipid Res, 2008. **49**(5): p. 1137-46.

2. Tsugawa, H.,et al., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature Methods 12, p 523–526 (2015)

Lipidomics Summary

The lipidomics workflow involves sample extraction in MTBE with addition of internal standards, followed by ultra high pressure liquid chromatography (UHPLC) on a Waters CSH column, interfaced to a QTOF mass spectrometer (high resolution, accurate mass), with a 15 minute total run time. Data are collected in both positive and negative ion mode, and analyzed using MassHunter (Agilent). Approximately 400 lipids can be identified from plasma, with additional unknowns. The method is highly stable and has been validated on large datasets (>8,000 samples) collected over long time periods (> 1 year).

Sample Preparation

Extraction of blood plasma and tissue lipids is based on the "Maytash" method [1] which was subsequently modified. Extraction is carried out using a bi-phasic solvent system of cold methanol, methyl *tert*-butyl ether (MTBE), and water. In more detail, cold methanol (225 µL) containing a mixture of odd chain and deuterated lipid internal standards [lysoPE(17:1), lysoPC(17:0), PC(12:0/13:0), PE(17:0/17:0), PG(17:0/17:0), sphingosine (d17:1), d_7-cholesterol, SM(17:0), C17 ceramide, d_3-palmitic acid, MG(17:0/0:0/0:0), DG(18:1/2:0/0:0), DG(12:0/12:0/0:0), and d_5-TG(17:0/17:1/17:0)] is added to a 20 µL blood plasma or 5 mg tissue sample aliquot, which is placed into a 1.5 mL Eppendorf tube, and the tube is vortexed for 10 s. Then, 750 µL of cold MTBE containing CE(22:1) (internal standard) are added, followed by vortexing for 10 s. and shaking for 6 min. at 4ºC. Phase separation is induced by adding 188 µL of mass spec-grade water. After vortexing for 20 s. the sample is centrifuged at 14,000 rpm for 2 min. The upper organic phase is collected in two 300 µL aliquots. One is stored at -20ºC as a backup and the other is evaporated to dryness in a SpeedVac. Dried extracts are resuspended using a mixture of methanol/toluene (9:1, v/v) (60 µL) containing an internal standard [12-[[(cyclohexylamino)carbonyl]amino]-dodecanoic acid (CUDA)] used as a quality control.

LC/MS parameters

The LC/QTOFMS analyses are performed using an Agilent 1290 Infinity LC system (G4220A binary pump, G4226A autosampler, and G1316C Column Thermostat) coupled to either an Agilent 6530 (positive ion mode) or an Agilent 6550 mass spectrometer equipped with an ion funnel (iFunnel) (negative ion mode). Lipids are separated on an Acquity UPLC CSH C18 column (100 x 2.1 mm; 1.7 µm) maintained at 65°C at a flow-rate of 0.6 mL/min. Solvent pre-heating (Agilent G1316) was used. The mobile phases consist of 60:40 acetonitrile:water with 10 mM ammonium formate and 0.1% formic acid (A) and 90:10 propan-2-ol:acetonitrile with 10 mM ammonium formate and 0.1% formic acid. The gradient is as follows: 0 min 85% (A); 0–2 min 70% (A); 2–2.5 min 52% (A); 2.5–11 min 18% (A); 11–11.5 min 1% (A); 11.5–12 min 1% (A); 12–12.1 min 85% (A); 12.1–15 min 85% (A). A sample volume of 3 µL is used for the injection. Sample temperature is maintained at 4°C in the autosampler.

The quadrupole/time-of-flight (QTOF) mass spectrometers are operated with electrospray ionization (ESI) performing full scan in the mass range *m*/*z* 65–1700 in positive (Agilent 6530, equipped with a JetStreamSource) and negative (Agilent 6550, equipped with a dual JetStream Source) modes producing both unique and complementary spectra. Instrument paramaters are

as follows (positive mode) Gas Temp 325°C, Gas Flow 8 l/min, Nebulizer 35 psig, Sheath Gas 350°C, Sheath Gas Flow 11, Capillary Voltage 3500 V, Nozzle Voltage 1000V, Fragmentor 120V, Skimmer 65V. Data (both profile and centroid) are collected at a rate of 2 scans per second. In negative ion mode, Gas Temp 200°C, Gas Flow 14 l/min, Fragmentor 175V, with the other parameters identical to positive ion mode. For the 6530 QTOF, a reference solution generating ions of 121.050 and 922.007 *m/z* in positive mode and 119.036 and 966.0007 m/z in negative mode, and these are used for continuous mass correction. For the 6550, the reference solution is introduced via a dual spray ESI, with the same ions and continuous mass correction.

Samples are injected (1.7 □l in positive mode and 5 □l in negative ion mode) with a needle wash for 20 seconds (wash solvent is isopropanol). The valve is switched back and forth during the run for washing; this has been shown to be essential for reducing carryover of less polar lipids.

Data Analysis

For the data processing the MassHunter software is used, and a unique ID is given to each lipid based on its retention time and exact mass (RT_mz). This allows the report of peak areas/heights or concentration of lipids based on the use of particular internal standards. Lipids are identified based on their unique MS/MS fragmentation patterns using in-house software, Lipidblast. Using complex lipid class-specific internal standards this approach is used to quantify >400 lipid species including: mono-, di- and triacylglycerols, glycerophospholipids, sphingolipids, cholesterol esters, ceramides, and fatty acids.

This approach is highly reproducible, displaying an RSD of 0.1% for the retention time and 1.7% for peak area based on replicate analysis of plasma samples (*n*=10). An average shot-to-shot carryover of less than 0.1% is observed. The described method is specifically developed to deal with high-throughput analyses, and can be used to analyze over 300 samples a week, including both positive and negative mode data acquisition.

References Cited

1. Matyash, V., et al., *Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics.* J Lipid Res, 2008. **49**(5): p. 1137-46.

This methods documentation is output by MetaboAnalyst5.0 after statistical analysis of unprocessed metabolomics data from the West Coast Metabolomics Center. All credit belongs to the respective authors (1-5).

.

1 Data Processing and Normalization

1.1 Reading and Processing the Raw Data

MetaboAnalyst accepts a variety of data types generated in metabolomic studies, including compound concentration data, binned NMR/MS spectra data, NMR/MS peak list data, as well as MS spectra (NetCDF, mzXML, mzDATA). Users need to specify the data types when uploading their data in order for MetaboAnalyst to select the correct algorithm to process them. Table 1 summarizes the result of the data processing steps.

1.1.1 Reading Concentration Data

The concentration data should be uploaded in comma separated values (.csv) format. Samples can be in rows or columns, with class labels immediately following the sample IDs.

1.1.2 Data Integrity Check

Before data analysis, a data integrity check is performed to make sure that all the necessary information has been collected. The class labels must be present and contain only two classes. If samples are paired, the class label must be from -n/2 to -1 for one group, and 1 to n/2 for the other group (n is the sample number and must be an even number). Class labels with same absolute value are assumed to be pairs. Compound concentration or peak intensity values should all be non-negative numbers. By default, all missing values, zeros and negative values will be replaced by the half of the minimum positive value found within the data (see next section).

1.1.3 Missing value imputations

Too many zeroes or missing values will cause difficulties for downstream analysis. MetaboAnalyst offers several different methods for this purpose. The default method replaces all the missing and zero values with a small value (the half of the minimum positive values in the original data) assuming to be the detection limit. The assumption of this approach is that most missing values are caused by low abundance metabolites (i.e. below the detection limit). In addition, since zero values may cause problem for data normalization (i.e. log), they are also replaced with this small value. User can also specify other methods, such as replace by mean/median, or use K-Nearest Neighbors, Probabilistic PCA (PPCA), Bayesian PCA (BPCA) method, Singular Value Decomposition (SVD) method to impute the missing values [1]. Please choose the one that is the most appropriate for your data. Zero or missing variables were replaced with a small value: 5.5

1.1.4 Data Filtering

The purpose of the data filtering is to identify and remove variables that are unlikely to be of use when modeling the data. No phenotype information is used in the filtering process, so the result can be used with any downstream analysis. This step can usually improves the results. Data filter is strongly recommended for datasets with large number of variables (> 250) datasets contain much noise (i.e. chemometrics data). Filtering can usually improve your results[2].
For data with number of variables < 250, this step will reduce 5% of variables; For variable number between 250 and 500, 10% of variables will be removed; For variable number between 500 and 1000, 25% of variables will be removed; And 40% of variables will be removed for data with over 1000 variables.

1.2 Data Normalization

The data is stored as a table with one sample per row and one variable (bin/peak/metabolite) per column. The normalization procedures implemented below are grouped into four categories. Sample specific normalization allows users to manually adjust concentrations based on biological inputs (i.e. volume, mass); row-wise normalization allows general-purpose adjustment for differences among samples; data transformation and scaling are two different approaches to make features more comparable. You can use one or combine both to achieve better results.

The normalization consists of the following options:
1. Sample specific normalization (i.e. normalize by dry weight, volume) 2. Row-wise procedures:
    Normalization by the sum
    Normalization by the sample median

Normalization by a reference sample (probabilistic quotient normalization)[3] Normalization by a reference feature (i.e. creatinine, internal control)

3. Data transformation:
   Generalized log transformation (glog 2) Cube root transformation
4. Data scaling:
   Unit scaling (mean-centered and divided by standard deviation of each variable)
   Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)
   Range scaling (mean-centered and divided by the value range of each variable)

2 Statistical and Machine Learning Data Analysis

MetaboAnalyst offers a variety of methods commonly used in metabolomic data analyses. They include:

1. Univariate analysis methods:
   Fold Change Analysis
   T-tests
   Volcano Plot
   One-way ANOVA and post-hoc analysis Correlation analysis
2. Multivariate analysis methods:
   Principal Component Analysis (PCA)
   Partial Least Squares - Discriminant Analysis (PLS-DA)
3. Robust Feature Selection Methods in microarray studies
   Significance Analysis of Microarray (SAM)
   Empirical Bayesian Analysis of Microarray (EBAM)
4. Clustering Analysis
   Hierarchical Clustering – Dendrogram
   – Heatmap
   Partitional Clustering
   – K-means Clustering
   – Self-Organizing Map (SOM)
5. Supervised Classification and Feature Selection methods
   Random Forest
   Support Vector Machine (SVM)
   Please note: some advanced methods are available only for two-group sample analysis.

2.1 Univariate Analysis

Univariate analysis methods are the most common methods used for exploratory data analysis. For two-group data, MetaboAnalyst provides Fold Change (FC) analysis, t-tests, and volcano plot which is a combination of the first two methods. All three these methods support both unpaired and paired analyses. For multi-group analysis, MetaboAnalyst provides two types of analysis - one-way analysis of variance (ANOVA) with associated post-hoc analyses, and correlation analysis to identify significant compounds that follow a given pattern. The univariate analyses provide a preliminary overview about features that are potentially significant in discriminating the conditions under study.

For paired fold change analysis, the algorithm first counts the total number of pairs with fold changes that are consistently above/below the specified FC threshold for each variable. A variable will be reported as significant if this number is above a given count threshold (default > 75% of pairs/variable).

2.2 Principal Component Analysis (PCA)

PCA is an unsupervised method aiming to find the directions that best explain the variance in a data set (X) without referring to class labels (Y). The data are summarized into much fewer variables called scores which are weighted average of the original variables. The weighting profiles are called loadings. The PCA analysis is performed using the prcomp package. The calculation is based on singular value decomposition. The Rscript chemometrics.R is required.

2.3 Partial Least Squares - Discriminant Analysis (PLS-DA)

PLS is a supervised method that uses multivariate regression techniques to extract via linear combination of original variables (X) the information that can predict the class membership (Y). The PLS regression is performed using the plsr function provided by R pls package[4]. The classification and cross-validation are performed using the corresponding wrapper function offered by the caret package[5].

To assess the significance of class discrimination, a permutation test was performed. In each permutation, a PLS-DA model was built between the data (X) and the permuted class labels (Y) using the optimal number of components determined by cross validation for the model based on the original class assignment. MetaboAnalyst supports two types of test statistics for measuring the class discrimination. The first one is based on prediction accuracy during training. The second one is separation distance based on the ratio of the between group sum of the squares and the within group sum of squares (B/W- ratio). If the observed test statistic is part of the distribution based on the permuted class assignments, the class discrimination cannot be considered significant from a statistical point of view.[6]

There are two variable importance measures in PLS-DA. The first, Variable Importance in Projection (VIP) is a weighted sum of squares of the PLS loadings taking into account the amount of explained Y-variation in each dimension. Please note, VIP scores are calculated for each components. When more than components are used to calculate the feature importance, the average of the VIP scores are used. The other importance measure is based on the weighted sum of PLS-regression. The weights are a function of the reduction of the sums of squares across the number of PLS components. Please note, for multiple-group (more than two) analysis, the same number of predictors will be built for each group. Therefore, the coefficient of each feature will be different depending on which group you want to predict. The average of the feature coefficient are used to indicate the overall coefficient-based importance.

2.4 Hierarchical Clustering

In (agglomerative) hierarchical cluster analysis, each sample begins as a separate cluster and the algorithm proceeds to combine them until all samples belong to one cluster. Two parameters need to be considered when performing hierarchical clustering. The first one is similarity measure - Euclidean distance, Pearson's correlation, Spearman's rank correlation. The other parameter is clustering algorithms, including average linkage (clustering uses the centroids of the observations), complete linkage (clustering uses the farthest pair of observations between the two groups), single linkage (clustering uses the closest pair of observations) and Ward's linkage (clustering to minimize the sum of squares of any two clusters). Heatmap is often presented as a visual aid in addition to the dendrogram.

Hierarchical clustering is performed with the hclust function in package stat.

2.5 Random Forest (RF)

Random Forest is a supervised learning algorithm suitable for high dimensional data analysis. It uses an ensemble of classification trees, each of which is grown by random feature selection from a bootstrap sample at each branch. Class prediction is based on the majority vote of the ensemble. RF also provides other useful information such as OOB (out-of-bag) error, variable importance measure, and outlier measures. During tree construction, about one-third of the instances are left out of the bootstrap sample. This OOB data is then used as test sample to obtain an unbiased estimate of the classification error (OOB error). Variable importance is evaluated by measuring the increase of the OOB error when it is permuted. The outlier measures are based on the proximities during tree construction.

RF analysis is performed using the randomForest package[7].

2.6 Support Vector Machine (SVM)

SVM aims to find a nonlinear decision function in the input space by mapping the data into a higher dimensional feature space and separating it there by means of a maximum margin hyperplane. The SVM- based recursive feature selection and classification is performed using the R-SVM script[8]. The process is performed recursively using decreasing series of feature subsets (ladder) so that different classification models can be calculated. Feature importance is evaluated based on its frequencies being selected in the best classifier identified by recursive classification and cross-validation. Please note, R-SVM is very computationally intensive. Only the top 50 features (ranked by their p values from t-tests) will be evaluated.

3 Data Annotation

Please be advised that MetaboAnalyst also supports metabolomic data annotation. For NMR, MS, or GC-MS peak list data, users can perform peak identification by searching the corresponding libraries. For compound concentration data, users can perform metabolite set enrichment analysis and metabolic pathway analysis.

[1]Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. pcaMethods: a bioconductor package, providing PCA methods for incomplete data., Bioinformatics 2007 23(9):1164-1167

[2]Hackstadt AJ, Hess AM.Filtering for increased power for microarray data analysis, BMC Bioinformatics. 2009; 10: 11.

[3]Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics, 2006, Anal Chem 78 (13);4281 - 4290

[4]Ron Wehrens and Bjorn-Helge Mevik.pls: Partial Least Squares Regression (PLSR) and Principal Component Regres- sion (PCR), 2007, R package version 2.1-0

[5]Max Kuhn. Contributions from Jed Wing and Steve Weston and Andre Williams.caret: Classification and Regression Training, 2008, R package version 3.45

[6]Bijlsma et al.Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation, Anal Chem. 2006, 78 567 - 574

[7]Andy Liaw and Matthew Wiener. Classification and Regression by randomForest, 2002, R News

[8] http://www.hsph.harvard.edu/bioinfocore/RSVMhome/R-SVM.html

MetaboAnalyst documentation authors:

1. Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P. E., Li, S., and Xia, J. (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* **49**, W388-W396
2. Chong, J., and Xia, J. (2020) Using MetaboAnalyst 4.0 for Metabolomics Data Analysis, Interpretation, and Integration with Other Omics Data. *Methods Mol Biol* **2104**, 337-360
3. Chong, J., Wishart, D. S., and Xia, J. (2019) Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr Protoc Bioinformatics* **68**, e86
4. Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D. S., and Xia, J. (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **46**, W486-W494
5. Xia, J., Psychogios, N., Young, N., and Wishart, D. S. (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* **37**, W652-660

Plasma Metabolite Extraction. Metabolites were extracted from 50 uL of plasma obtained from Nadk2 mutant and B6 control mouse models according to standard operating procedures in the Mass Spectrometry and Protein Chemistry Service at The Jackson Laboratory. Plasma metabolites were extracted using an ice cold 2:2:1 methanol:acetonitrile:water (MeOH:ACN:H2O) extraction buffer. Caffeine, 1-napthylamine ($C_{10}H_9N$), and 9-anthracene carboxylic acid ($C_{15}H_{10}O_2$) were all added at 0.5 ng/uL in the extraction buffer as internal standards. After adding the extraction buffer, samples were vortexed for 30 seconds on the highest setting, vigorously mixed for one minute with the Tissue Lyser II in pre-chilled cassettes, and placed in the -20^{o}C freezer overnight (16 hours) for extraction. Samples were then centrifuged at 21,000 x g at 4^{o}C and each metabolite extract (supernatant) was equally divided into five 2 mL microcentrifuge tubes. Each sample supernatant was divided into five equal volume aliquots, one for each of the four modes and the rest to create equal representation pools of all samples, one for each mode. Each aliquot was then dried down using a vacuum centrifuge for storage at -80^{o}C until further use.

Discovery Metabolomics Analysis. Four mode discovery metabolomics analysis was performed in the Mass Spectrometry and Protein Chemistry Service at The Jackson Laboratory using a Thermo Q-Exactive Orbitrap mass spectrometer coupled to a dual-channel Vanquish Ultra-Performance Liquid Chromatography system. The samples were analyzed with a 25 minute gradient over a hydrophobic C18 column (Agilent InfinityLab Poroshell 120 EC-C18, #699775-902T) and a hydrophilic HILIC column (Agilent InfinityLab Poroshell 120 HILIC-Z, #689775-924) column under both positive and negative polarity to maximize identification of metabolites. For C18 positive and negative modes, solvent A was 99.8% H_2O with 0.2% acetic acid and solvent B was 99.8% ACN with 0.2% acetic acid. All runs for HILIC negative use 10 mM ammonium formate in H_2O with 0.1% formic acid as solvent A and 90% ACN with 10 mM ammonium formate in H_2O with 0.1% formic acid as solvent B. HILIC positive runs utilize 10 mM ammonium acetate in H_2O, pH 9.0 with 0.1% AffinityLab Deactivator Inhibitor (Agilent, #5191-3940) for solvent A and 85% ACN with 10 mM ammonium acetate in H_2O with 0.1% AffinityLab Deactivator Inhibitor as solvent B. The 25 minute gradient for the C18 modes consisted of the following steps: 0-1 minutes at 98% A/2% B, 1-13 minutes from 98% A/2% B to 10% A/90% B, 13-15 minutes at 10% A/90% B, 15-16 minutes from 10% A/90% B to 98% A/2% B, and was re-equilibrated from 16-25 minutes at 98% A/2% B. The 25 minute gradient for the HILIC modes consisted of the following steps: 0-1 minutes at 2% A/98% B, 1-11 minutes from 2% A/98% B to 30% A/70% B, 11-12 minutes from 30% A/70% B to 40% A/60% B, 12-16 minutes from 40% A/60% B to 95% A/5% B, was held at 95% A/5% B from 16-18 minutes, 18-20 minutes from 95% A/5% B to 2% A/98% B, and was re-equilibrated from 20-25 minutes at 2% A/98% B.

Each sample was reconstituted in 25 uL of 95% H_2O/5% ACN for C18 modes and 95% ACN/5% H_2O for HILIC modes; two technical replicates for each sample were injected at 10 uL (equivalent of 4 uL starting plasma extracted). A quality control metabolite pool representing all plasma samples was run at the beginning and end of the run set at concentrations equivalent to the samples and was used for a normalized quality control batch correction of the runs over time to account for technical variance. The sample sequence run order was randomized using the list randomizer in random.org. For the C18 modes, MS1 precursor settings included a default charge = 1, resolution = 35,000, AGC target = 3e6, maximum IT = 100 ms, and scan range = 80-1200 m/z (C18) or 65-975 m/z (HILIC). Data dependent MS2 fragmentation (dd-MS2) settings included a resolution = 17,500, AGC target = 1e5, maximum IT = 50 ms, top N = 5, isolation window = 2.0 (C18) and 1.2 (HILIC) m/z, and a stepped NCE = 30, 50. General dd settings included a minimum AGC = 8e3, intensity threshold = 1.6e5, and a dynamic exclusion

of 20 seconds.

Metabolomics Data Analysis. All .raw files were exported and analyzed using Thermo Compound Discoverer (v3.2.0.421). Peak filtering, detect compounds, and compound matching using was performed using the settings listed below related to data analysis. General settings include a mass tolerance = 10 ppm, intensity tolerance = 30%, S/N threshold = 2, minimum peak intensity = 500,000, base ions = [M+H]; [M-H]-1, a maximum peak width = 0.5 minutes. All peaks were subject to a blank background subtraction to remove contaminant peaks (minimum S/N threshold = 2) and then subject to a quality control pooled sample correction selecting for peaks only consistently detected in the pool for normalization. All MS1 and MS2 data was then searched against the Thermo mzCloud and ChemSpider databases, along with Metabolika Pathways and mzLogic predicted composition. Further filters were applied to the data including a coefficient of variation $\leq$ 35% in either group and an MS2 FISH coverage filter > 10 in Compound Discoverer to filter for high confidence matches based on consistency of detection and MS2 fragmentation matching. Differential targets were then filtered by adding an abundance ratio filters for fold change >2 or <0.5 and a p-value $\leq$ 0.05 for significance. Overlapping metabolite comparisons for the venn diagrams were performed using Venny 2.1 (bioinfogp.cnb.csic.es/tools/venny/), principal component analysis was completed in Compound Discoverer, and volcano plot comparisons were completed using custom R analysis. For enrichment analysis, compound lists for each mode were filtered in Compound Discoverer for FISH>=10, CV%<=35 in either sample group, and a p-value <=0.05. These compounds were then selected for in the mzCloud analysis matches, followed by a filter for a KEGG ID being present (mainly keeps the endogenous or essential dietary metabolites that are well annotated). The significantly altered metabolite (indicators of pathway flux) KEGG IDs from mzCloud were then extracted from each of the four modes into one list (126 metabolites total and analyzed using the Enrichment Analysis feature in MetaboAnalyst 5.0. Three types of analyses were run on this dataset: 1) SMDB pathway analysis (compared to 99 human metabolic pathway datasets for pathway enrichment), 2) blood disease signature analysis (compared to 344 human blood disease metabolite sets), and 3) predicted dysfunctional enzyme pathways analysis (compared for enrichment in 912 metabolic sets to predict the change in the case of dysfunctional enzymes). All enrichment ratios were computed by the hits in the dataset provided / expected number of hits based on a random distribution.

Compound Discoverer Search Settings:

Search description: Untargeted Metabolomics workflow: Find and identify the differences between samples.

Function: Performs retention time alignment, unknown compound detection, and compound grouping across all samples. Predicts elemental compositions for all compounds, fills gaps across all samples, and hides chemical background (using Blank samples). Identifies compounds using mzCloud (ddMS2) and ChemSpider (formula or exact mass). Also performs similarity search for all compounds with ddMS2 data using mzCloud. Applies mzLogic algorithm to rank order ChemSpider results. Maps compounds to biological pathways using Metabolika. Applies QC-based batch normalization if QC samples are available. Calculates differential analysis (t-test or ANOVA), determines p-values, adjusted p-values, ratios, fold change, CV, etc.).

Created with Compound Discoverer version: 3.2.0.421

[Input Files (6)]
-->Select Spectra (33)

[Select Spectra (33)]
-->Align Retention Times (26)

[Align Retention Times (26)]
-->Detect Compounds (9)

[Detect Compounds (9)]
-->Group Compounds (31)

[Group Compounds (31)]
-->Fill Gaps (32)
-->Search ChemSpider (23)
-->Map to Metabolika Pathways (34)
-->Predict Compositions (29)
-->Assign Compound Annotations (25)
-->Search mzCloud (38)

[Fill Gaps (32)]
-->Apply QC Correction (43)

[Apply QC Correction (43)]
-->Apply Missing Value Imputation (44)

[Apply Missing Value Imputation (44)]
-->Mark Background Compounds (42)

[Search ChemSpider (23)]
-->Apply mzLogic (35)

[Map to Metabolika Pathways (34)]
-->Apply mzLogic (35)

[Mark Background Compounds (42)]

[Apply mzLogic (35)]

[Predict Compositions (29)]

[Assign Compound Annotations (25)]

[Search mzCloud (38)]

[Differential Analysis (17)]

Processing node 6: Input Files

Processing node 33: Select Spectra
---
1. Spectrum Properties Filter:
- Lower RT Limit: 0
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Ignore Specified Scans: (not specified)
- Lowest Charge State: 0
- Highest Charge State: 0
- Min. Precursor Mass: 0 Da
- Max. Precursor Mass: 5000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

2. Scan Event Filters:
- Mass Analyzer: (not specified)
- MS Order: Any
- Activation Type: (not specified)
- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Any
- Polarity Mode: (not specified)

3. Peak Filters:
- S/N Threshold (FT-only): 1.5

4. Replacements for Unrecognized Properties:

- Unrecognized Charge Replacements: 1
- Unrecognized Mass Analyzer Replacements: ITMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: CID
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

5. General Settings:
- Precursor Selection: Use MS(n - 1) Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic
- Store Chromatograms: False

Processing node 26: Align Retention Times
---
1. General Settings:
- Alignment Model: Adaptive curve
- Alignment Fallback: Use Linear Model
- Maximum Shift [min]: 2
- Shift Reference File: True
- Mass Tolerance: 10 ppm
- Remove Outlier: True

Processing node 9: Detect Compounds
---
1. General Settings:
- Mass Tolerance [ppm]: 10 ppm
- Intensity Tolerance [%]: 30
- S/N Threshold: 2
- Min. Peak Intensity: 500000
- Ions:
    [2M+ACN+H]+1
    [2M+ACN+Na]+1
    [2M+FA-H]-1
    [2M+H]+1
    [2M+K]+1
    [2M+Na]+1
    [2M+NH4]+1
    [2M-H]-1
    [2M-H+HAc]-1
    [M+2H]+2
    [M+ACN+2H]+2
    [M+ACN+H]+1
    [M+ACN+Na]+1
    [M+Cl]-1

[M+DMSO+H]+1
[M+FA-H]-1
[M+H]+1
[M+H+K]+2
[M+H+MeOH]+1
[M+H+Na]+2
[M+H+NH4]+2
[M+H-H2O]+1
[M+H-NH3]+1
[M+K]+1
[M+Na]+1
[M+NH4]+1
[M-2H]-2
[M-2H+K]-1
[M-H]-1
[M-H+HAc]-1
[M-H+TFA]-1
[M-H-H2O]-1
- Base Ions: [M+H]+1; [M-H]-1
- Min. Element Counts: C H
- Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O15 P3 S5

2. Peak Detection:
- Filter Peaks: True
- Max. Peak Width [min]: 0.5
- Remove Singlets: True
- Min. # Scans per Peak: 5
- Min. # Isotopes: 1

3. Isotope Grouping:
- Min. Spectral Distance Score: 0
- Remove Potentially False Positive Isotopes: True

Processing node 31: Group Compounds
---
1. Compound Consolidation:
- Mass Tolerance: 10 ppm
- RT Tolerance [min]: 1

2. Fragment Data Selection:
- Preferred Ions: [M+H]+1; [M-H]-1

Processing node 32: Fill Gaps
---
1. General Settings:
- Mass Tolerance: 10 ppm

- S/N Threshold: 1.5
- Use Real Peak Detection: True

Processing node 43: Apply QC Correction
---
1. General Settings:
- Regression Model: Linear
- Min. QC Coverage [%]: 25
- Max. QC Area RSD [%]: 50
- Max. Corrected QC Area RSD [%]: 50
- Max. # Files Between QC Files: 20

Processing node 44: Apply Missing Value Imputation
---
1. General Settings:
- Imputation Method: Automatic Selection
- Fill Blanks with min value: False

2. Random Forest Settings:
- Number of trees: 100
- Max Number of Iterations: 10

Processing node 42: Mark Background Compounds
---
1. General Settings:
- Max. Sample/Blank: 2
- Max. Blank/Sample: 0
- Hide Background: True

Processing node 23: Search ChemSpider
---
1. Search Settings:
- Database(s): BioCyc; Human Metabolome Database; KEGG
- Search Mode: By Formula or Mass
- Mass Tolerance: 10 ppm
- Max. # of results per compound: 100
- Max. # of Predicted Compositions to be searched per Compound: 3
- Result Order (for Max. # of results per compound): Order By Reference Count (DESC)

2. Predicted Composition Annotation:
- Check All Predicted Compositions: False

Processing node 35: Apply mzLogic

---
1. Search Settings:
- FT Fragment Mass Tolerance: 10 ppm
- IT Fragment Mass Tolerance: 0.4 Da
- Max. # Compounds: 0
- Max. # mzCloud Similarity Results to consider per Compound: 10
- Match Factor Threshold: 30

Processing node 34: Map to Metabolika Pathways
---
1. Search Settings:
-Metabolika Pathways
- Search Mode: By Formula or Mass

2. By Mass Search Settings:
- Mass Tolerance: 10 ppm

3. By Formula Search Settings:
- Max. # of Predicted Compositions to be searched per Compound: 3

4. Display Settings:
- Max. # Pathways in 'Pathways' column: 20

Processing node 29: Predict Compositions
---
1. Prediction Settings:
- Mass Tolerance: 10 ppm
- Min. Element Counts: C H
- Max. Element Counts: C90 H190 Br3 Cl4 N10 O18 P3 S5
- Min. RDBE: 0
- Max. RDBE: 40
- Min. H/C: 0.1
- Max. H/C: 4
- Max. # Candidates: 10
- Max. # Internal Candidates: 200

2. Pattern Matching:
- Intensity Tolerance [%]: 30
- Intensity Threshold [%]: 0.1
- S/N Threshold: 3
- Min. Spectral Fit [%]: 30
- Min. Pattern Cov. [%]: 90
- Use Dynamic Recalibration: True

3. Fragments Matching:
- Use Fragments Matching: True

- Mass Tolerance: 15 ppm
- S/N Threshold: 3

Processing node 25: Assign Compound Annotations
---
1. General Settings:
- Mass Tolerance: 10 ppm

2. Data Sources:
- Data Source #1: mzCloud Search
- Data Source #2: Predicted Compositions
- Data Source #3: MassList Search
- Data Source #4: ChemSpider Search
- Data Source #5: Metabolika Search
- Data Source #6: (not specified)
- Data Source #7: (not specified)

3. Scoring Rules:
- Use mzLogic: True
- Use Spectral Distance: True
- SFit Threshold: 20
- SFit Range: 20

Processing node 38: Search mzCloud
---
1. General Settings:
- Compound Classes: All
- Precursor Mass Tolerance: 10 ppm
- FT Fragment Mass Tolerance: 10 ppm
- IT Fragment Mass Tolerance: 0.4 Da
- Library: Autoprocessed; Reference
- Post Processing: Recalibrated
- Max. # Results: 10
- Annotate Matching Fragments: True

2. DDA Search:
- Identity Search: Cosine
- Match Activation Type: True
- Match Activation Energy: Match with Tolerance
- Activation Energy Tolerance: 20
- Apply Intensity Threshold: True
- Similarity Search: Confidence Forward
- Match Factor Threshold: 60

3. DIA Search:
- Use DIA Scans for Search: False

- Max. Isolation Width [Da]: 500
- Match Activation Type: False
- Match Activation Energy: Any
- Activation Energy Tolerance: 100
- Apply Intensity Threshold: False
- Match Factor Threshold: 20

Processing node 17: Differential Analysis
---
1. General Settings:
- Log10 Transform Values: True