February 4, 2018

Data Processing, Pathway Mapping Multi-Omic Systems and Activity Analysis using XCMS Online
Supplementary Methods

Erica M. Forsberg[1,2], Tao Huan[1], Duane Rinehart[1], H. Paul Benton[1], Benedikt Warth[1,3], Brian Hilmers[1], Gary Siuzdak[1]*

[1]Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, CA, 92037, United States
[2] Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA, 92182, United States
[3] Department of Food Chemistry and Toxicology, University of Vienna, 1090 Vienna, Austria

*Author to whom correspondence should be addressed:

Gary Siuzdak
Tel: (858) 784-9415
E-mail: siuzdak@scripps.edu
Internet: https://masspec.scripps.edu/

Demonstration Dataset Experimental Design:

To test the efficacy of the multi-omic platform, a standard method for probing changes in metabolomics was developed using *E. coli* K12 MG1655 cultures grown in different carbon sources (glucose and adenosine carbon sources). Concurrently with metabolomic data acquisition, RNA transcriptomic data was generated on the same sample set and a set of literature-generated proteomic data[1-3] was acquired. This data was run using the base parameter set “HPLC / UHD QTOF (HILIC, neg)” with none of the settings altered, with the exception of the “identification tab”, where it was necessary to change the biosource selection to *E. coli* K12 MG1655, with pathway prediction settings of 20 ppm mass tolerance and a 5,000 signal intensity threshold to account for

Nature Protocols: doi:10.1038/nprot.2017.151

the expected error in the raw data. This was saved as a new parameter set called “Demo HPLC / UHD QTOF (HILIC, neg)”.

E. Coli Growth Methods.

Glucose and adenosine were prepared with equimolar concentrations of carbon in 10 mL aliquots of M9 minimal media in triplicate. Carbon concentration was based on a final concentration of 20 mM glucose or 0.12 M carbon. All carbon sources were prepared in sterile water, then filtered through a 0.22 µm syringe filter prior to addition to M9 salts. A 10 mL culture of *E. coli* was grown in LB media overnight at 37°C. To inoculate each condition, which was prepared in triplicate, a 1 mL aliquot of cells was centrifuged at 13,000 rpm for 1 minute, the supernatant media was removed and the pellet was washed 3X with sterile water. The cells were made up to a final volume of 1 mL in sterile water and a 1 µL aliquot was added to 8 mL of each carbon source. The cultures were grown until an OD_{600} ~1, or stopped after 72 hours if the growth rate was plateaued. Triplicate 1 mL aliquots were taken for each replicate for metabolomics analysis and duplicate 1 mL samples were taken for RNA sequencing.

E. coli RNA extraction and mRNA-seq.

RNA in E. coli samples were extracted using RNeasy Mini Kit (50, Cat. No. 74104) and the extraction procedures followed the protocols inside the extraction kit. In brief, cells were lysed with the working solution and then centrifuged. The supernatants were loaded onto spin column and spun down multiple times to purify RNAs. RNA-seq experiment was performed with 75 bp reads generated on the NextSeq Analyzer located at the Scripps DNA Sequencing Facility. The Genome Analyzer Pipeline Software (currently bcl2fastq/2.16.0.10) is used to perform the early data analysis of a sequencing run, which does the image analysis, base calling, and demultiplexing. Cutadapt software19 was used to trim the adapter and low base-pair called scores. For mRNA-Seq, STAR 2.3.0 was used to align to genome using the E. coli K12 genome reference. EdgeR as used with the method finds number of Differentially Expressed transcripts (DE) significantly changed for the comparisons of different carbon sources. The results are first filtered with False Discovery Rate (FDR) >0.15 and then by log

Counts Per Million (log2(CPM)) > 1.0. The log2 counts-per-million (log2 CPM) cut-off used to avoid undefined values and the poorly defined log-fold-changes for low counts shrunk towards zero. Further the deviation of the normalized counts within groups can be used to filter out the transcripts with higher variance. We noted that almost all the transcripts were significantly up or down changed.

***E. coli* metabolome extraction and LC-MS based metabolomic profiling.**

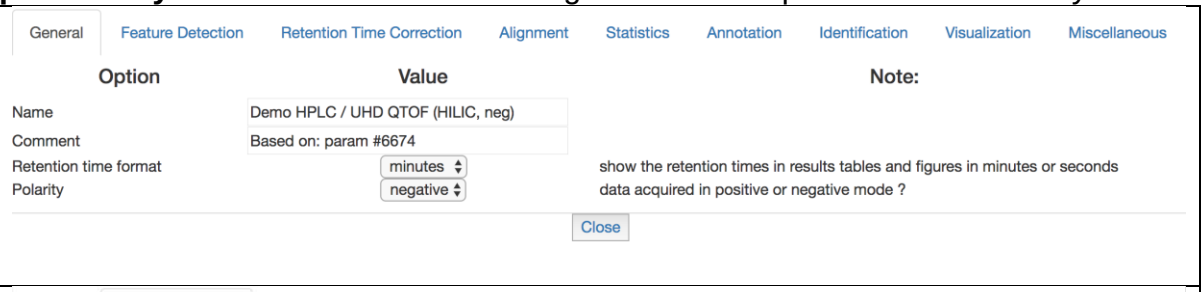
For metabolomics, each sample was spun down at 14,000xg for 5 min at 4ºC. The supernatant was removed and 1.0 mL of ice cold 2:2:1 v/v acetonitrile/methanol/water was added to each tube and vortexed. Cell lysis was performed by three freeze/thaw cycles in liquid nitrogen room temperature. Samples were sonicated and then placed in –20ºC overnight to allow protein and cell debris to precipitate. The samples were centrifuged again, supernatant was transferred to glass vials and evaporated at 8ºC to dryness. The dry extracts were then reconstituted in 100 µL of acetonitrile:water (1:1, v/v), sonicated for 10 min and centrifuged for 15 min at 13,000 rpm and 4°C to remove insoluble debris. The supernatants were transferred to HPLC vials and stored at −80°C prior to LC-MS analysis.

The LC-MS analysis was performed on each sample in triplicate using a Bruker Impact QTOF mass spectrometer (Billerica, MA, USA) linked to an Agilent 1200 series capillary HPLC system (Palo Alto, CA, USA). 2 µL sample was injected after injection volume optimization. A Phenomenex Luna NH2 column (1 mm × 15 cm, 3 µm particle size, 100 Å pore size) was used for LC separation. Mobile phase A was 20 mM ammonium acetate in H_2O with 5% acetonitrile (ACN), and mobile phase B was ACN with 5% H_2O. The gradient elution profile was as follows: t = 0 min, 95%B; t = 5 min, 95%B; t = 50 min, 5%B. t = 63min, 5%B. The flow rate was 50 µL/min. 20 min post acquisition time was also added with 95%B to re-equilibrium the LC column. Injection of 1 µL sodium formate (250 µM) occurred at t = 57 min and eluted on t = 61min. This calibration peak was used to internally calibrate the LC-MS data and also served as the MS peak intensity quality control. The m/z scan range was 25–1500, ionization was in negative mode and acquisition speed was 2 Hz. All data files were converted to mzXML format and uploaded to XCMS Online to perform a pairwise analysis on glucose control

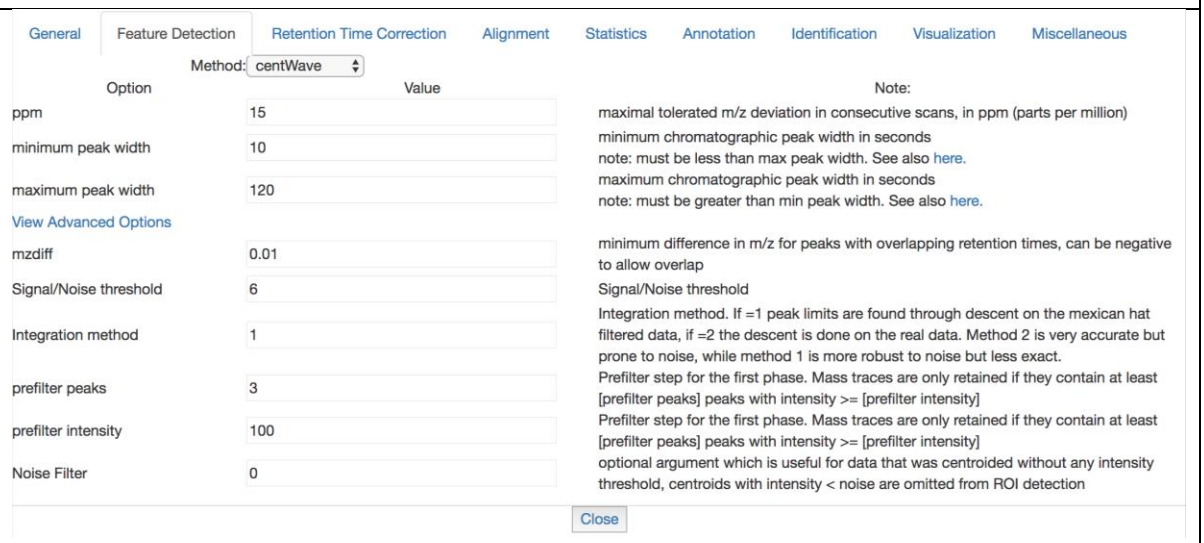
(n = 3) versus adenosine stressed (n = 3) using Welch's t-test. For parameter settings to run XCMS job and perform automated pathway analysis, see **Supplemental Table S1.**

Supplementary Table S1. Parameter settings for *E. coli* experimental data analysis

1

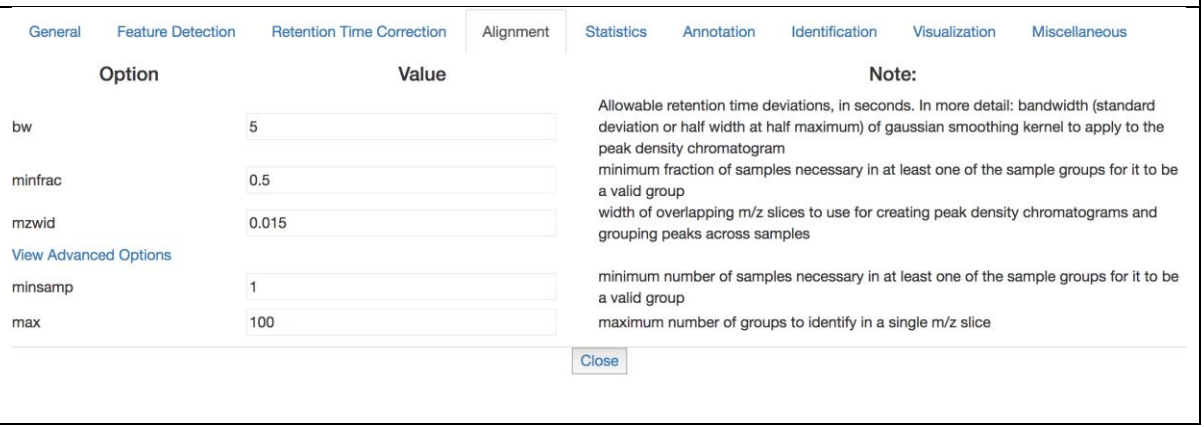


2

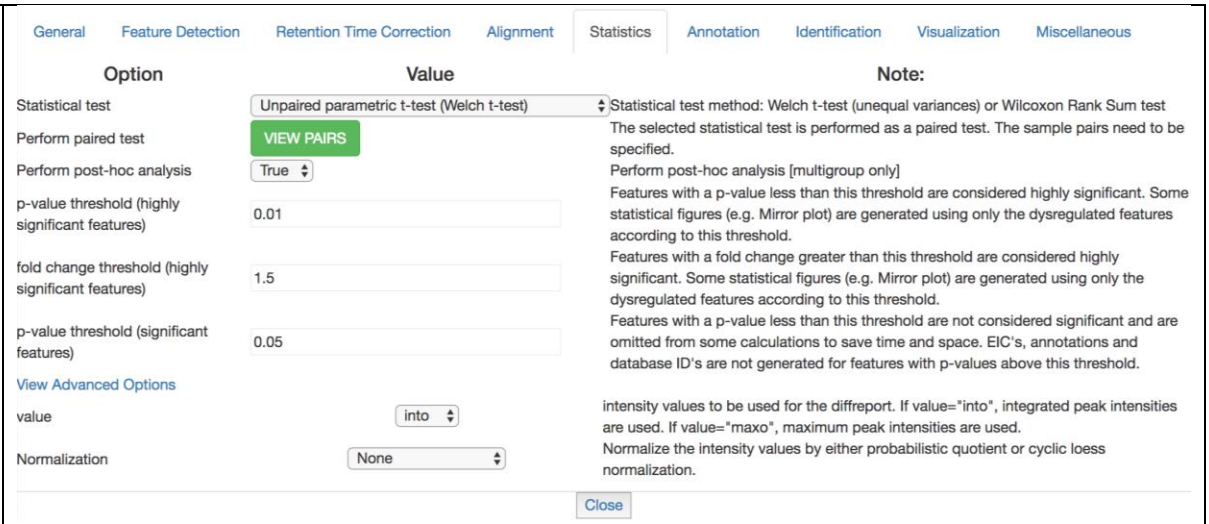


3

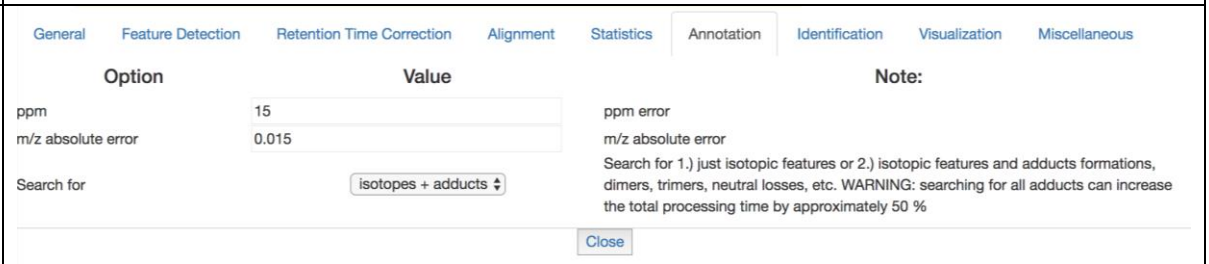
4



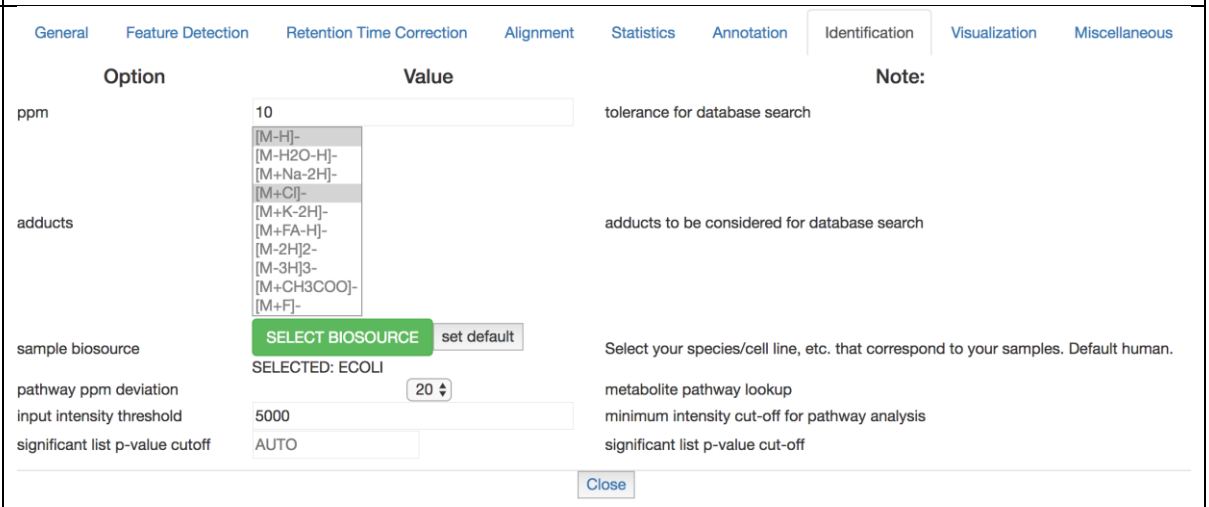
5


General
Feature Detection
Retention Time Correction
Alignment
Statistics
Annotation
Identification
Visualization
Miscellaneous
Option
Value
Note:
Statistical test
Unpaired parametric t-test (Welch t-test)
Statistical test method: Welch t-test (unequal variances) or Wilcoxon Rank Sum test
Perform paired test
VIEW PAIRS
The selected statistical test is performed as a paired test. The sample pairs need to be specified.
Perform post-hoc analysis
True
Perform post-hoc analysis [multigroup only]
p-value threshold (highly significant features)
0.01
Features with a p-value less than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
fold change threshold (highly significant features)
1.5
Features with a fold change greater than this threshold are considered highly significant. Some statistical figures (e.g. Mirror plot) are generated using only the dysregulated features according to this threshold.
p-value threshold (significant features)
0.05
Features with a p-value less than this threshold are not considered significant and are omitted from some calculations to save time and space. EIC's, annotations and database ID's are not generated for features with p-values above this threshold.
View Advanced Options
value
into
intensity values to be used for the diffreport. If value="into", integrated peak intensities are used. If value="maxo", maximum peak intensities are used.
Normalization
None
Normalize the intensity values by either probabilistic quotient or cyclic loess normalization.
Close

6


General
Feature Detection
Retention Time Correction
Alignment
Statistics
Annotation
Identification
Visualization
Miscellaneous
Option
Value
Note:
ppm
15
ppm error
m/z absolute error
0.015
m/z absolute error
Search for
isotopes + adducts
Search for 1.) just isotopic features or 2.) isotopic features and adducts formations, dimers, trimers, neutral losses, etc. WARNING: searching for all adducts can increase the total processing time by approximately 50 %
Close

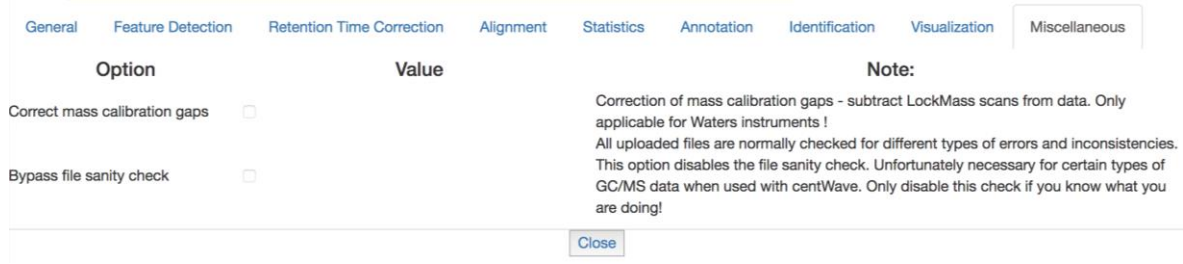
7


General
Feature Detection
Retention Time Correction
Alignment
Statistics
Annotation
Identification
Visualization
Miscellaneous
Option
Value
Note:
ppm
10
tolerance for database search
[M-H]-
[M-H2O-H]-
[M+Na-2H]-
[M+Cl]-
[M+K-2H]-
[M+FA-H]-
[M-2H]2-
[M-3H]3-
[M+CH3COO]-
[M+F]-
adducts
adducts to be considered for database search
SELECT BIOSOURCE
set default
sample biosource
Select your species/cell line, etc. that correspond to your samples. Default human.
SELECTED: ECOLI
pathway ppm deviation
20
metabolite pathway lookup
input intensity threshold
5000
minimum intensity cut-off for pathway analysis
significant list p-value cutoff
AUTO
significant list p-value cut-off
Close

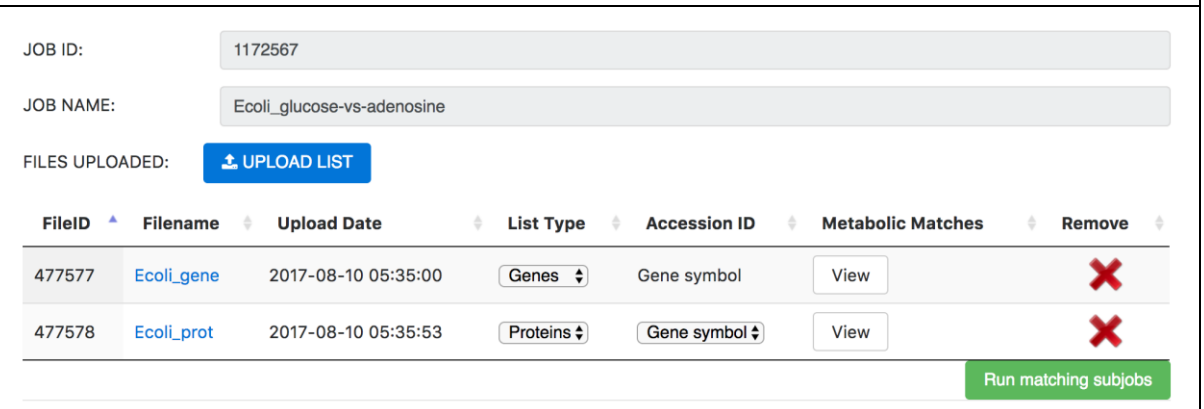
8


General
Feature Detection
Retention Time Correction
Alignment
Statistics
Annotation
Identification
Visualization
Miscellaneous
Option
Value
Note:
EIC width
200
Default width for extracted ion chromatograms in seconds
Close

9



10



1. Du, Z., Nandakumar, R., Nickerson, K.W. & Li, X. Proteomic adaptations to starvation prepare Escherichia coli for disinfection tolerance. *Water Res.* **69**, 110-119 (2015).
2. Vijayendran, C., Burgemeister, S., Friehs, K., Niehaus, K. & Flaschel, E. 2DBase: 2D-PAGE database of Escherichia coli. *Biochemical and Biophysical Research Communications* **363**, 822-827 (2007).
3. Vijayendran, C. et al. The plasticity of global proteome and genome expression analyzed in closely related W3110 and MG1655 strains of a well-studied model organism, Escherichia coli-K12. *J. Biotechnol.* **128**, 747-761 (2007).