

February 4, 2018

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

Erica M. Forsberg^{1,2}, Tao Huan¹, Duane Rinehart¹, H. Paul Benton¹, Benedikt Warth^{1,3}, Brian Hilmers¹, Gary Siuzdak^{1*}

¹Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, CA, 92037, United States

²Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA, 92182, United States

³Department 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 data¹⁻³ 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

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 $\text{OD}_{600} \sim 1$, 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 ($\log_2(\text{CPM})$) > 1.0. The \log_2 counts-per-million (\log_2 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 4°C. 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 –20°C overnight to allow protein and cell debris to precipitate. The samples were centrifuged again, supernatant was transferred to glass vials and evaporated at 8°C 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 \times 15 cm, 3 μm particle size, 100 Å pore size) was used for LC separation. Mobile phase A was 20 mM ammonium acetate in H₂O with 5% acetonitrile (ACN), and mobile phase B was ACN with 5% H₂O. 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 $\mu\text{L}/\text{min}$. 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	<p>General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous</p> <p>Option Value Note:</p> <p>Name Demo HPLC / UHD QTOF (HILIC, neg)</p> <p>Comment Based on: param #6674</p> <p>Retention time format <input type="text" value="minutes"/> show the retention times in results tables and figures in minutes or seconds</p> <p>Polarity <input type="text" value="negative"/> data acquired in positive or negative mode ?</p> <p><input type="button" value="Close"/></p>
2	<p>General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous</p> <p>Method: <input type="text" value="centWave"/></p> <p>Option Value Note:</p> <p>ppm <input type="text" value="15"/> maximal tolerated m/z deviation in consecutive scans, in ppm (parts per million)</p> <p>minimum peak width <input type="text" value="10"/> minimum chromatographic peak width in seconds note: must be less than max peak width. See also here.</p> <p>maximum peak width <input type="text" value="120"/> maximum chromatographic peak width in seconds note: must be greater than min peak width. See also here.</p> <p>View Advanced Options</p> <p>mzdiff <input type="text" value="0.01"/> minimum difference in m/z for peaks with overlapping retention times, can be negative to allow overlap</p> <p>Signal/Noise threshold <input type="text" value="6"/> Signal/Noise threshold</p> <p>Integration method <input type="text" value="1"/> Integration method. If =1 peak limits are found through descent on the mexican hat filtered data, if =2 the descent is done on the real data. Method 2 is very accurate but prone to noise, while method 1 is more robust to noise but less exact.</p> <p>prefilter peaks <input type="text" value="3"/> Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]</p> <p>prefilter intensity <input type="text" value="100"/> Prefilter step for the first phase. Mass traces are only retained if they contain at least [prefilter peaks] peaks with intensity >= [prefilter intensity]</p> <p>Noise Filter <input type="text" value="0"/> optional argument which is useful for data that was centroided without any intensity threshold, centroids with intensity < noise are omitted from ROI detection</p> <p><input type="button" value="Close"/></p>
3	<p>General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous</p> <p>Method: <input type="text" value="obiwarp"/></p> <p>Option Value Note:</p> <p>profStep <input type="text" value="1"/> step size (in m/z) to use for profile generation from the raw data files</p> <p><input type="button" value="Close"/></p>
4	<p>General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous</p> <p>Option Value Note:</p> <p>bw <input type="text" value="5"/> Allowable retention time deviations, in seconds. In more detail: bandwidth (standard deviation or half width at half maximum) of gaussian smoothing kernel to apply to the peak density chromatogram</p> <p>minfrac <input type="text" value="0.5"/> minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group</p> <p>mzwid <input type="text" value="0.015"/> width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples</p> <p>View Advanced Options</p> <p>minsamp <input type="text" value="1"/> minimum number of samples necessary in at least one of the sample groups for it to be a valid group</p> <p>max <input type="text" value="100"/> maximum number of groups to identify in a single m/z slice</p> <p><input type="button" value="Close"/></p>

5	<div style="display: flex; justify-content: space-between; border-bottom: 1px solid black; padding-bottom: 5px;"> General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous </div> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%; text-align: left;">Option</th> <th style="width: 40%; text-align: left;">Value</th> <th style="width: 30%; text-align: left;">Note:</th> </tr> </thead> <tbody> <tr> <td>Statistical test</td> <td>Unpaired parametric t-test (Welch t-test)</td> <td>Statistical test method: Welch t-test (unequal variances) or Wilcoxon Rank Sum test</td> </tr> <tr> <td>Perform paired test</td> <td>VIEW PAIRS</td> <td>The selected statistical test is performed as a paired test. The sample pairs need to be specified.</td> </tr> <tr> <td>Perform post-hoc analysis</td> <td>True</td> <td>Perform post-hoc analysis [multigroup only]</td> </tr> <tr> <td>p-value threshold (highly significant features)</td> <td>0.01</td> <td>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.</td> </tr> <tr> <td>fold change threshold (highly significant features)</td> <td>1.5</td> <td>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.</td> </tr> <tr> <td>p-value threshold (significant features)</td> <td>0.05</td> <td>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.</td> </tr> <tr> <td colspan="3">View Advanced Options</td> </tr> <tr> <td>value</td> <td>into</td> <td>intensity values to be used for the diffreport. If value="into", integrated peak intensities are used. If value="maxo", maximum peak intensities are used.</td> </tr> <tr> <td>Normalization</td> <td>None</td> <td>Normalize the intensity values by either probabilistic quotient or cyclic loess normalization.</td> </tr> <tr> <td colspan="3" style="text-align: right;">Close</td> </tr> </tbody> </table>	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		
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	<div style="display: flex; justify-content: space-between; border-bottom: 1px solid black; padding-bottom: 5px;"> General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous </div> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%; text-align: left;">Option</th> <th style="width: 40%; text-align: left;">Value</th> <th style="width: 30%; text-align: left;">Note:</th> </tr> </thead> <tbody> <tr> <td>ppm</td> <td>15</td> <td>ppm error</td> </tr> <tr> <td>m/z absolute error</td> <td>0.015</td> <td>m/z absolute error</td> </tr> <tr> <td>Search for</td> <td>isotopes + adducts</td> <td>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 %</td> </tr> <tr> <td colspan="3" style="text-align: right;">Close</td> </tr> </tbody> </table>	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																				
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	<div style="display: flex; justify-content: space-between; border-bottom: 1px solid black; padding-bottom: 5px;"> General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous </div> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%; text-align: left;">Option</th> <th style="width: 40%; text-align: left;">Value</th> <th style="width: 30%; text-align: left;">Note:</th> </tr> </thead> <tbody> <tr> <td>ppm</td> <td>10</td> <td>tolerance for database search</td> </tr> <tr> <td>adducts</td> <td> <div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;"> [M-H]- [M-H2O-H]- [M+Na-2H]- [M+C]- [M+K-2H]- [M+FA-H]- [M-2H]2- [M-3H]3- [M+CH3COO]- [M+F]- </div> </td> <td>adducts to be considered for database search</td> </tr> <tr> <td>sample biosource</td> <td>SELECT BIOSOURCE set default SELECTED: ECOLI</td> <td>Select your species/cell line, etc. that correspond to your samples. Default human.</td> </tr> <tr> <td>pathway ppm deviation</td> <td>20</td> <td>metabolite pathway lookup</td> </tr> <tr> <td>input intensity threshold</td> <td>5000</td> <td>minimum intensity cut-off for pathway analysis</td> </tr> <tr> <td>significant list p-value cutoff</td> <td>AUTO</td> <td>significant list p-value cut-off</td> </tr> <tr> <td colspan="3" style="text-align: right;">Close</td> </tr> </tbody> </table>	Option	Value	Note:	ppm	10	tolerance for database search	adducts	<div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;"> [M-H]- [M-H2O-H]- [M+Na-2H]- [M+C]- [M+K-2H]- [M+FA-H]- [M-2H]2- [M-3H]3- [M+CH3COO]- [M+F]- </div>	adducts to be considered for database search	sample biosource	SELECT BIOSOURCE set default SELECTED: ECOLI	Select your species/cell line, etc. that correspond to your samples. Default human.	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											
Option	Value	Note:																																
ppm	10	tolerance for database search																																
adducts	<div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;"> [M-H]- [M-H2O-H]- [M+Na-2H]- [M+C]- [M+K-2H]- [M+FA-H]- [M-2H]2- [M-3H]3- [M+CH3COO]- [M+F]- </div>	adducts to be considered for database search																																
sample biosource	SELECT BIOSOURCE set default SELECTED: ECOLI	Select your species/cell line, etc. that correspond to your samples. Default human.																																
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	<div style="display: flex; justify-content: space-between; border-bottom: 1px solid black; padding-bottom: 5px;"> General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous </div> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%; text-align: left;">Option</th> <th style="width: 40%; text-align: left;">Value</th> <th style="width: 30%; text-align: left;">Note:</th> </tr> </thead> <tbody> <tr> <td>EIC width</td> <td>200</td> <td>Default width for extracted ion chromatograms in seconds</td> </tr> <tr> <td colspan="3" style="text-align: right;">Close</td> </tr> </tbody> </table>	Option	Value	Note:	EIC width	200	Default width for extracted ion chromatograms in seconds	Close																										
Option	Value	Note:																																
EIC width	200	Default width for extracted ion chromatograms in seconds																																
Close																																		

9

General Feature Detection Retention Time Correction Alignment Statistics Annotation Identification Visualization Miscellaneous

Option **Value** **Note:**

Correct mass calibration gaps Correction of mass calibration gaps - subtract LockMass scans from data. Only applicable for Waters instruments!

Bypass file sanity check All uploaded files are normally checked for different types of errors and inconsistencies. This option disables the file sanity check. Unfortunately necessary for certain types of GC/MS data when used with centWave. Only disable this check if you know what you are doing!

Close

10

JOB ID: 1172567

JOB NAME: Ecoli_glucose-vs-adenosine

FILES UPLOADED: [↑ UPLOAD LIST](#)

FileID	Filename	Upload Date	List Type	Accession ID	Metabolic Matches	Remove
477577	Ecoli_gene	2017-08-10 05:35:00	Genes	Gene symbol	View	✖
477578	Ecoli_prot	2017-08-10 05:35:53	Proteins	Gene symbol	View	✖

Run matching subjobs

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).