# Credentialing Features: A Benchmarking Platform to Optimize Untargeted Metabolomic Methods

Nathaniel Guy Mahieu,[1,2] Xiaojing Huang,[2] Ying-Jr Chen,[1,2] and Gary J. Patti[1,2,*]

[1]Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri 63130, United States

[2]Departments of Genetics and Medicine, Washington University School of Medicine, St. Louis, Missouri 63110, United States

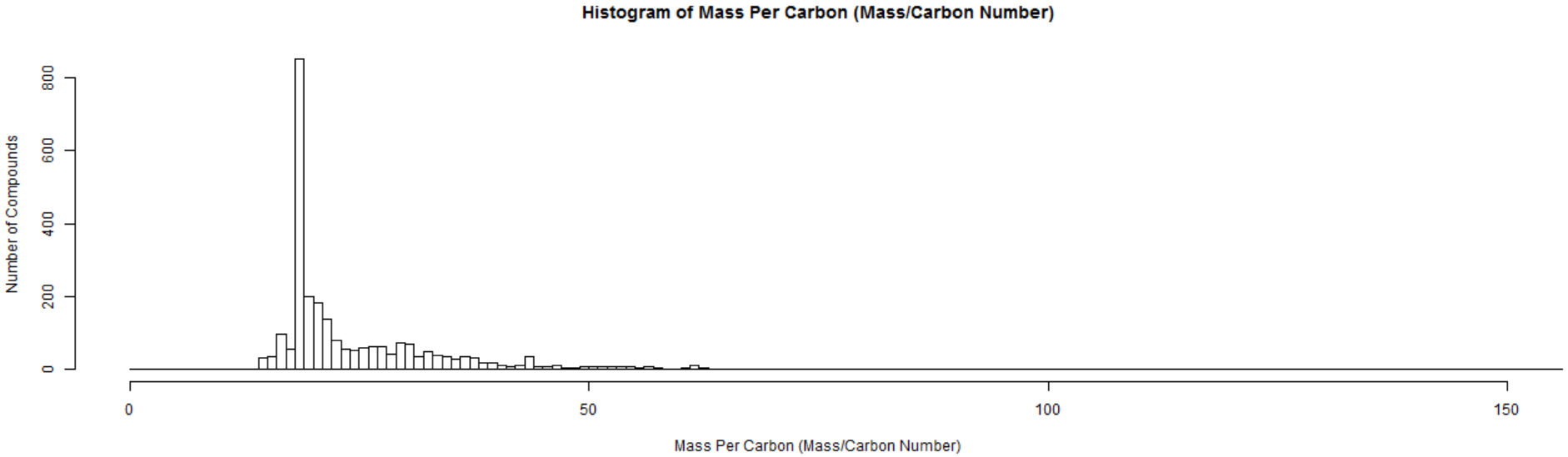
*To whom correspondence should be addressed: gjpattij@wustl.edu

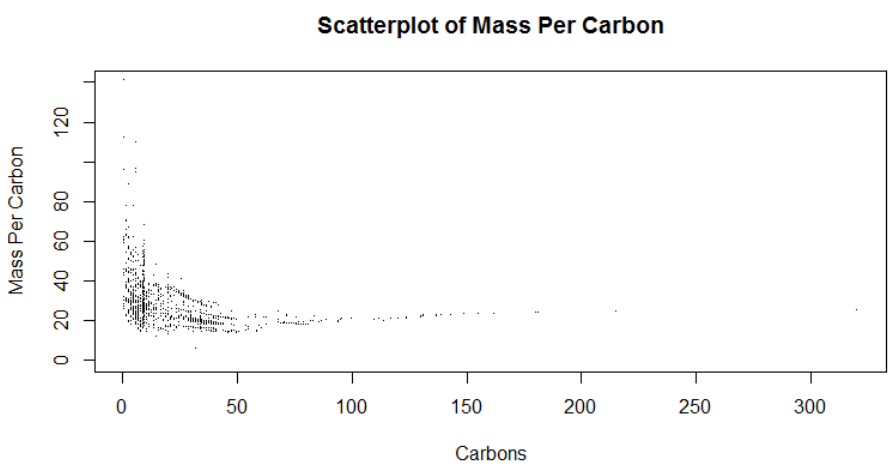
## Abstract

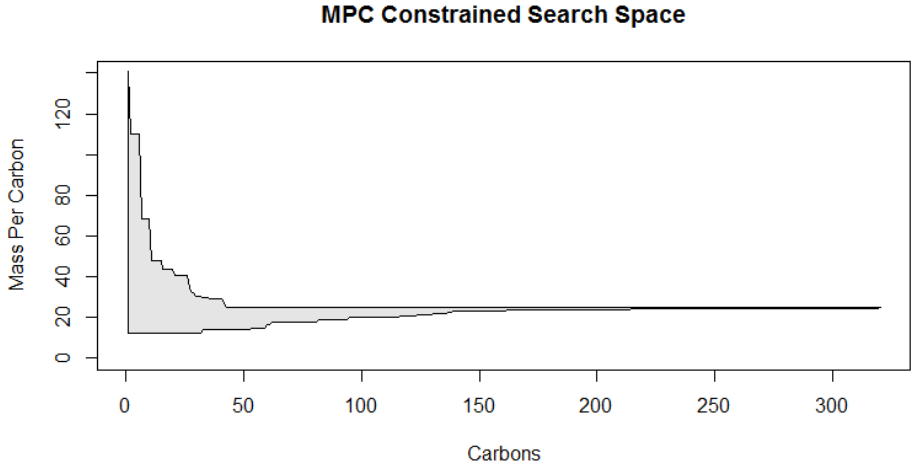
The aim of untargeted metabolomics is to profile as many metabolites as possible, yet a major challenge is comparing experimental method performance on the basis of metabolome coverage. To date, most published approaches have compared experimental methods by counting the total number of features detected. Due to artifactual interference, however, this number is highly variable and therefore is a poor metric for comparing metabolomic methods. Here we introduce an alternative approach to benchmarking metabolome coverage which relies on mixed Escherichia coli extracts from cells cultured in regular and 13C-enriched media. After mass spectrometry-based metabolomic analysis of these extracts, we "credential" features arising from E. coli metabolites on the basis of isotope spacing and intensity. This credentialing platform enables us to accurately compare the number of nonartifactual features yielded by different experimental approaches. We highlight the value of our platform by reoptimizing a published untargeted metabolomic method for XCMS data processing. Compared to the published parameters, the new XCMS parameters decrease the total number of features by 15% (a reduction in noise features) while increasing the number of true metabolites detected and grouped by 20%. Our credentialing platform relies on easily generated E. coli samples and a simple software algorithm that is freely available on our laboratory Web site (http://pattilab.wustl.edu/software/credential/). We have validated the credentialing platform with reversed-phase and hydrophilic interaction liquid chromatography as well as Agilent, Thermo Scientific, AB SCIEX, and LECO mass spectrometers. Thus, the credentialing platform can readily be applied by any laboratory to optimize their untargeted metabolomic pipeline for metabolite extraction, chromatographic separation, mass spectrometric detection, and bioinformatic processing.

## Contents

**Supporting Figure S-2. Calculation of Mass Pe Carbon (mpc) From ECMD.** A histogram of mass in Daltons divided by carbon number (mpc) are shown below. The mass of a methylene ($CH_2$, 14 Da) unit is a logical lower bound for mass per carbon. ECMDB contains four compounds which have an mpc lower than 14, all of which are more reduced (contain more rings and double bonds). An mpc of 141 is the largest in ECMDB and corresponds to carbamoyl phosphate. The most common mass per carbon is 18-19 Da/C with 850 compounds falling in this range. Based on the data, a carbon number dependent limit is placed on the mass range in which to search for isotopes. This is depicted in the lower plots.



**Supporting Table S-3. Suggested Parameters for Various Instrumentation Platforms.** The credentialing technique is flexible and can be applied to many types of instrumentation and chromatography. Below are suggested values for different instrumentation that have been shown to be effective experimentally.

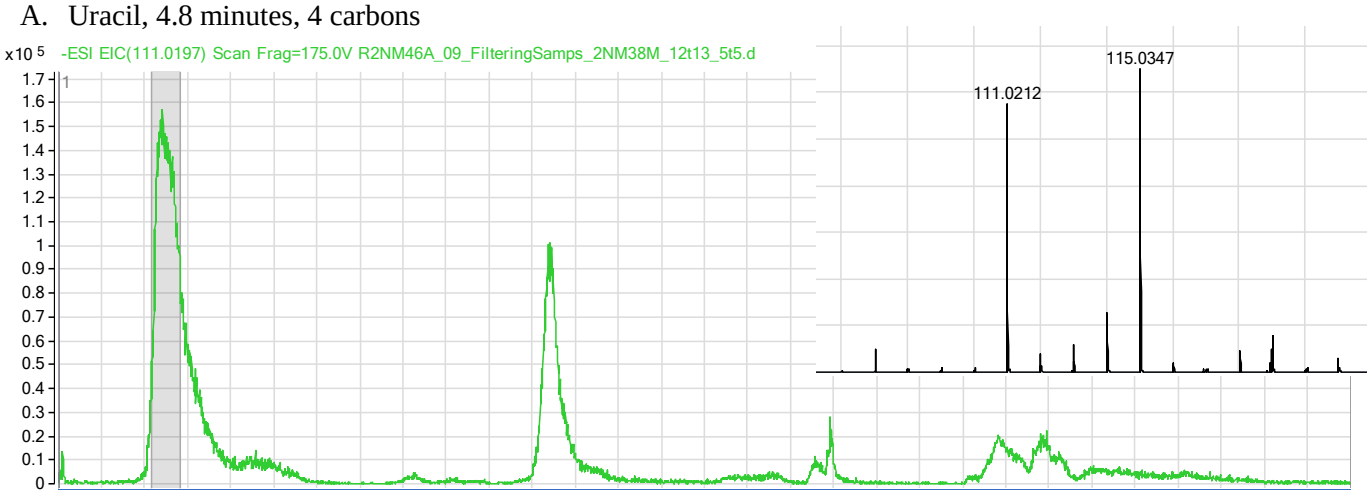
| Parameter | Suggested Defaults | | Explanation |
|---|---|---|---|
| ***iso_ppm*** | Time of Flight*:<br>Orbitrap**:<br>FT-ICR: | 4<br>1<br>0.1 | This is the mass error allowed when considering the difference between a $^{12}C$ and $^{13}C$ peak. This should be set according to the intra-scan mass error, rather than the absolute mass error of the instrument. |
| ***mix_tol*** | 4 | | This is a coarse filter that ensures the $^{12}C$ peak and $^{13}C$ peak are of comparable intensity to their mixed ratios. This should allow a large error as many effects cause the $U^{12}C$ and $U^{13}C$ peaks to vary in intensity. A stricter filter is applied in the second round. |
| **ratio_tol** | 1.8 | | This is a fine filter which ensures the intensity ratio between the two samples approaches the ratio of mixing (See Data Analysis). This is the most sensitive parameter and can be set according to the user's needs. Values approach 1 are more selective. 1.8 offers a false positive rate of approximately 0.6% |
| ***iso_rt*** | HILIC: 0.1 x (peak fwhm)<br>C18: 0.05 x (peak fwhm) | | This is the acceptable tolerance (in seconds) when matching a $U^{12}C$ peak to a $U^{13}C$ peak. Ideally the peaks have an identical retention times but in some cases poor peak shape causes the detected retention time to vary between isotopes. For chromatography which generates consistant peak shapes this can be lowered. |
| ***mpc_tol*** | 1 | | Mass per carbon (mpc) is calculated as described above in Supplement S-2. The *mpc_tol* parameter is useful if a user is attempting to credential peaks with extremely large masses per number of carbons such as highly phosphorylated or metal containing compounds. |

***Agilent QTOF, AB SCIEX TripleTOF, LECO Pegasus**
****Thermo QE**

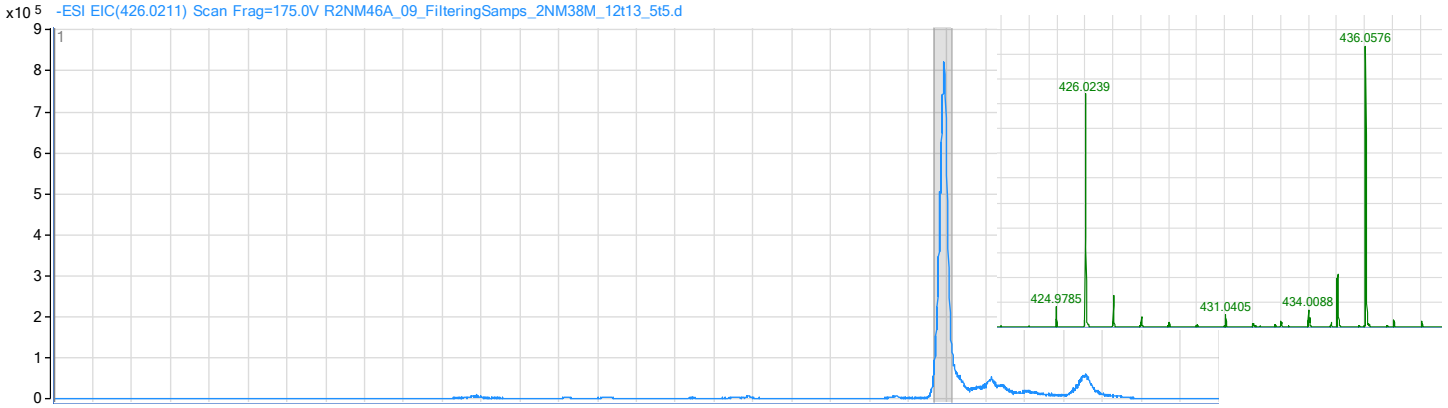
**Supporting Figure S-4. Raw Data Credentialed Features**

Mass spectra and extracted ion chromatograms are shown for the three knowns targeted for MS/MS. The labeling pattern exhibited by credentialed features can be seen in the inset. (A) Uracil, (B) ADP, (C) UDP-GlcA. Inset mass spectra are averaged over the highlighted region of each chromatogram.

A. Uracil, 4.8 minutes, 4 carbons




B. ADP, 48.5 minutes, 10 carbons




C. UDP-5'-GlcA, 43.8 minutes, 15 carbons