Nov 24, 2025

LC-MS/MS Untargeted Metabolomics Data Processing

DOI
dx.doi.org/10.17504/protocols.io.bf4wjqxe

Kevin Contrepois[1]

[1]Stanford University

Human BioMolecular Atlas Program (HuBMAP) Method Development Community
Tech. support email: **Jeff.spraggins@vanderbilt.edu**

Kevin Contrepois

DOI: https://dx.doi.org/10.17504/protocols.io.bf4wjqxe

Protocol Citation: Kevin Contrepois 2025. LC-MS/MS Untargeted Metabolomics Data Processing. **protocols.io https://dx.doi.org/10.17504/protocols.io.bf4wjqxe**

Protocol status: Working

Created: May 06, 2020

Last Modified: May 13, 2020

Protocol Integer ID: 36726

Keywords: HuBMAP, Metabolomics, Stanford University, ms untargeted metabolomics data processing scope, untargeted metabolomics data, metabolomics data, metabolomics dataset, metabolite annotation, m, spectral database,

Abstract

Scope:
To describe the procedure to process untargeted metabolomics data.

Expected outcome/data:
Metabolomics dataset is processed using commercial software and custom R scripts. Metabolite annotation was performed using an in-house library and spectral databases.

Troubleshooting

1 Data from each mode were independently analyzed using Progenesis QI software (v2.3) (Nonlinear Dynamics, Durham, NC).

2 Metabolic features from blanks and those that didn't show sufficient linearity upon dilution in QC samples ($r < 0.6$) were discarded.

3 Only metabolic features present in >2/3 of the samples were kept for further analysis.

4 Median normalization was applied to correct for differential starting material quantity.

5 Missing values were imputed by drawing from a random distribution of low values in the corresponding sample.

6 Data from each mode were merged and metabolites were formally identified by matching fragmentation spectra and retention time to analytical-grade standards when possible or matching experimental MS/MS to fragmentation spectra in publicly available databases.

7 Metabolite abundances were reported as spectral counts.

8 Peak annotation was first performed by matching experimental m/z, retention time and MS/MS spectra to an in-house library of analytical-grade standards. Remaining peaks were identified by matching experimental m/z and fragmentation spectra to publicly available databases including HMDB (**http://www.hmdb.ca/**), MoNA (**http://mona.fiehnlab.ucdavis.edu/**) and MassBank (**http://www.massbank.jp/**) using the R package 'MetID' (v0.2.0) (PMID: 30944337). Briefly, metabolic feature tables from Progenesis QI were matched to fragmentation spectra with a m/z and a retention time window of ±15 ppm and ±30 s (HILIC) and ± 20 s (RPLC), respectively. When multiple MS/MS spectra match a single metabolic feature, all matched MS/MS spectra were used for the identification. Next, MS1 and MS2 pairs were searched against public databases and a similarity score was calculated using the forward dot–product algorithm which takes into account both fragments and intensities (PMID: 24222034). Metabolites were reported if the similarity score was above 0.4. Spectra from metabolites of interest were further manually investigated to confirm identification.