**README file for LC-MS data:**

**Design\_file (metadata) – ‘study\_design\_lcms.txt’ – tab-delimited.** The design file can be found on Github (<https://github.com/artedison/metaanalysis/tree/main/LC-MS>)

The design file contains information as to the study design and analysis.

It is generated from the analyst who prepared the data and analyzed the samples

(Contact Brianna Garcia for further details: [brianna.garcia@uga.edu](mailto:amanda.shaver@uga.edu)).

Each column corresponds to the following:

|  |  |
| --- | --- |
| Sample Name | Unique sample name matching the data containing file |
| Raw file | Raw file name used to identify .raw Thermo LC-MS files |
| strain | C*. elegans* genotype for each sample |
| Sample type | Type of sample (sample, blank, or study pool (i.e., pool\_mutant, pool\_natural, and pool\_pd1074) |
| Batch | Batch identifier (1-6). Sample preparation was conducted on each batch individually and collected on the LC-MS. |
| Set | Set identifier (1-3) each genotype is contained within two adjacent sets. |

Each row corresponds to a sample as defined by the different column identifiers.

**Data Matrix –**

1. HILIC Pos: ‘ST002092\_AN003418\_Results.txt – tab delimited
2. RP Neg: ‘ST002092\_AN003417\_Results.txt’ – tab delimited
3. RP Pos: ‘ST002092\_AN003416\_Results.txt’ – tab delimited

The data matrices can be found on Metabolomics Workbench ([https://www.metabolomicsworkbench.org](https://www.metabolomicsworkbench.org/)) under the following Study ID: ST002092.

This file contains data from all the samples in the study, including quality controls.

Solvent Blanks and Process Blanks are **not** included.

The data matrix is composed of the following:

Columns:

1. **b1\_aos\_122** – Column named by ‘Sample Name’ containing the ranked position for each feature (row)

Rows:

1. **Features\_mz\_rt** – concatenated identifier consisting of the mass-to-charge ratio and the retention time of each feature (row)

Example:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| features\_mz\_rt | b1\_aos\_122 | b1\_aos\_127 | b1\_aos\_129 | b1\_aos\_142 | b1\_aos\_145 | b1\_aos\_149 | b1\_aos\_25 |
| 229.1295\_8.6 | 42 | 74 | 80 | 45 | 73 | 96 | NA |
| 234.1125\_1.03 | 145 | 160 | 87 | 177 | 144 | 94 | NA |

**Data processing:**

The data matrix enclosed has been obtained after the following transformations.

**Software used:**

**SLAW:** Slaw is an open-source untargeted metabolomics LC-MS data processing workflow and documentation by the developers can be found: <https://github.com/zamboni-lab/SLAW>

* Input: Centroided .mzML converted LC-MS files. Data can be found on Metabolomics Workbench under the following Study ID: ST002092.

**SECIMTools:** Data were then evaluated for QC and meta-analysis was conducted. Documentation can be found here: <https://github.com/McIntyre-Lab/papers/blob/master/shaver_metaanalysis_2022/documentation/bff_and_qc_slaw_output.xlsx>. Scripts can be found here: <https://github.com/McIntyre-Lab/papers/tree/master/shaver_metaanalysis_2022/scripts/massSpec>

**MATLAB:** Data were filtered to only retain features found in 100% of the PD1074 (individual samples and pool PD1074) and the solvent front was removed. MATLAB script can be found on Github: <https://github.com/artedison/metaanalysis/tree/main/LC-MS/scripts>

The data output files can be found on Metabolomics Workbench ([https://www.metabolomicsworkbench.org](https://www.metabolomicsworkbench.org/)) under the following Study ID: ST002092.