



MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

MICHIGAN REGIONAL COMPREHENSIVE METABOLOMICS RESOURCE CORE



MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

COMPUTATIONAL MEDICINE AND
BIOINFORMATICS

BINNER USER MANUAL

University of
Michigan
February 2019

2019 University of Michigan

This work is supported by the Michigan Regional Comprehensive Metabolomics Research Core through NIH Grant# U24DK097153 and the Advanced Proteome Informatics of Cancer Training Program (T32 CA140044).

Table of Contents

Terms of Use.....	3
General User Policy.....	3
For academic and non-profit institutions.....	3
For commercial and for-profit institutions.....	3
Citation Policy.....	3
Overview.....	4
Input File.....	5
Using Binner.....	5
Input Tab.....	6
Output Options Tab.....	7
Data Cleaning Tab.....	9
Feature Grouping Tab.....	10
Annotation.....	11
Annotation File.....	11
Annotation Tab.....	14
Output File.....	15
Summary Tab.....	15
Mass Difference Distribution.....	15
Bin Statistics.....	16
Correlations.....	16
Correlations by Cluster.....	17
Correlations by Bin.....	17
Mass Differences.....	18
Mass Differences by Cluster.....	18
Mass Differences by Bin.....	19
Intensities.....	19
Molecular Ions.....	19
Additional Columns.....	19
Supplemental Information.....	20
Cutoff Score.....	20
Clustering.....	20

Sub-clustering..... 21

Annotation Preferences..... 21

Isotope Detection..... 22

Terms of Use

General User Policy

For academic and non-profit institutions

- Permission is granted to access, use, and/or download the Tool for internal use only.
- If the user desires to create derivative works of the Tool, source code or access to data may be made available through request by sending an email to binner-help@umich.edu.
- Use of the Tool must be acknowledged in resulting publications (see citation policy below).

For commercial and for-profit institutions

- Permission is granted to access, use, and/or download the Tool for internal use only.
- To create derivative works of the Tool for commercial purposes, source code or access to data may be permitted through negotiation for a commercial license. Please send request to binner-help@umich.edu.

Citation Policy

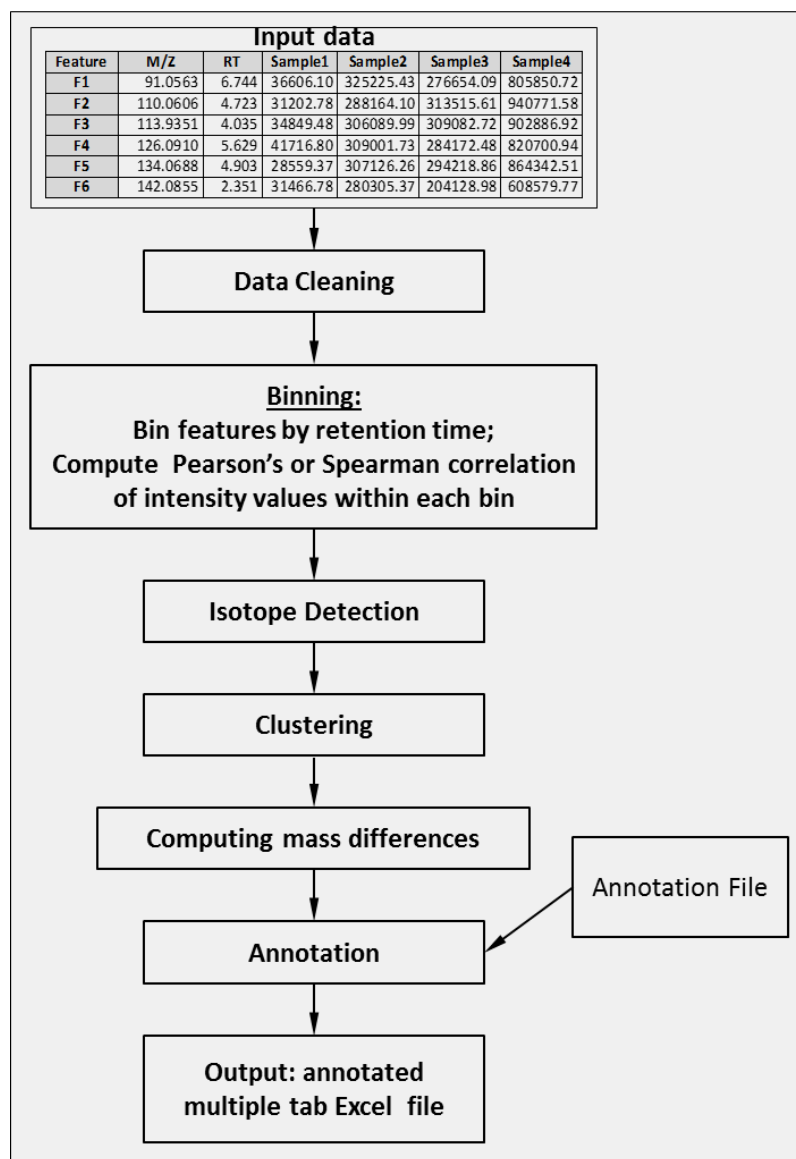
Please cite:

Please note that due to continuous software upgrades the images in this handout may differ from what you see on the screen.

Overview

When metabolites are analyzed by electrospray ionization-mass spectrometry (ESI-MS), they are often detected as multiple ion species due to the presence of isotopologues, adducts, and in-source fragments that cannot be readily identified. Binner is a Java application that takes a numerical feature table obtained from any preprocessing software (e.g. XCMS, MZmine) as input and outputs a file with clusters of closely eluting, highly correlated metabolite features and their putative annotations, along with their pairwise correlations and mass differences.

Binner aims to accurately annotate suspected chemical adducts and decipher the underlying metabolite neutral mass. Below is an overview of the Binner workflow.



Input File

Binner accepts .txt and .csv files where each row represents one feature. The file must include the following columns:

- A column containing the feature name. Any name or identifier, including chemical formula, compound name, or other identifier are accepted.
Note: The names need to be machine-readable, as certain special symbols and characters may cause a problem.
- A column containing mass to charge ratios (m/z).
- A column containing retention times (RT). The recommendation is to provide retention time in minutes.
Note: If seconds are used, the user must update Binner's default settings accordingly.
- A column for each sample containing the intensity for each associated feature. A minimum of 3 samples is necessary, although 20 or more is recommended.
- Additional columns can be included in the input data file and then carried over to the output file.

The order of the above listed columns is flexible, as Binner requires column mapping on the [Input Tab](#).

Feature Name	Mass-to-Charge Ratio	Retention Time	Samples				
Compound	Mass	RT	Samp1	Samp2	Samp3	Samp4	Samp5
C3 H7 N	57.0581	1.0384			427096	477359	
C3 H6 O	58.0424	6.1384		2424		21882	
TRIMETHYLAMINE (M+H)+	59.0733	0.6873	1271	1909	8877	3546	6118
ETHANOLAMINE (M+Na)+	61.0531	0.6636					
64.0392@5.353634	64.0392	5.3536	15405	12421	14390	15748	14430
65.0122@0.5439432	65.0122	0.5439	180375	176556	182122	189830	191664
65.0855@0.5426082	65.0855	0.5426	2744		6017	1461	1670
67.9755@0.5450444	67.9755	0.5450	135195	134462	135359	138674	141783
C5 H8	68.0623	1.6897	48589	53364	58211	8989	51034
C4 H7 N	69.0582	0.7086	194307	209721	191421	194343	194460
C4 H9 N	71.0741	0.8910	490019	419006	1642372	1820177	2198530
C3 H4 S	72.0038	1.8654	49671	34737	51251	46175	46610
C4 H8 O	72.0582	10.0960		17742		37948	
C4 H11 N	73.0897	3.7548	24835		19658	19085	22057
C6 H6	78.0475	4.4775	13941	12728	13869	13274	12392
81.0068@0.5461972	81.0068	0.5462	4668		4697	4184	5329
C5 H6 O	82.0425	10.6468	111581	107889	111566	125945	119531

Using Binner

After launching Binner, the user will see a series of tabs. The user can step through Binner using the tabs at the top of the window or the Next>> button at the bottom right of the window.

Input Tab

Binner v0.9.13.1

Input Output Options Data Cleaning Feature Grouping Annotation

Import Experimental Data

Select File (must be .txt)

stellaXCMS_pos.txt Import...

Specify Columns

Compound Compound

Mass mz

Retention Time rt

First Sample X20150811.EX00468.A003.IN0016.S00018999.P

Last Sample X20150811.EX00468.A003.IN0016.S00019078.P

80 samples will be analyzed

Exclude Samples...

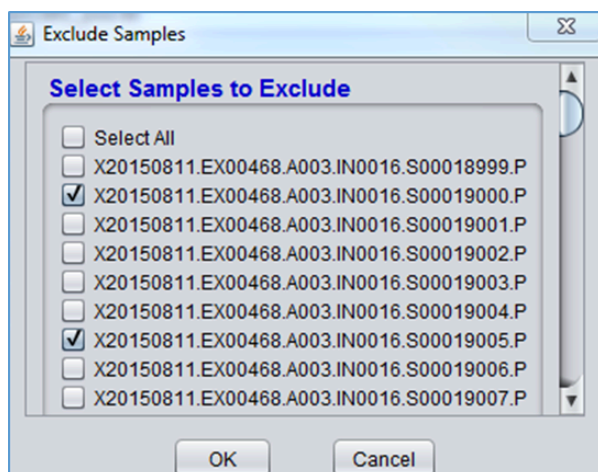
No samples excluded from analysis

Specify Ionization Mode

☒ Positive ☐ Negative

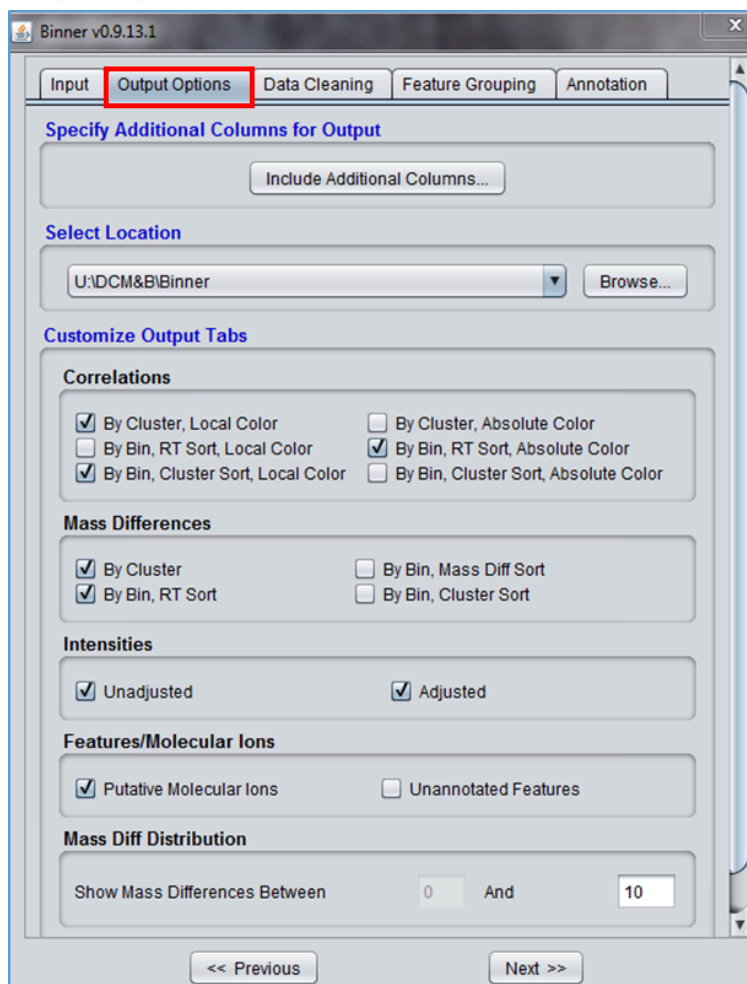
<< Previous Next >>

1. Under “Import Experimental Data,” use the “Import...” button to find and load the input data file.
2. Under “Specify Columns,” use the dropdown arrows next to each column type to select the specific column from the input file that maps to the column type.
3. To exclude one or more samples from the analysis, click the “Exclude Samples...” button.
 - a. A popup window will appear listing the sample names from the input file (see image below).
 - b. Check the boxes next to those samples that are to be excluded.
 - c. Click the “OK” button.

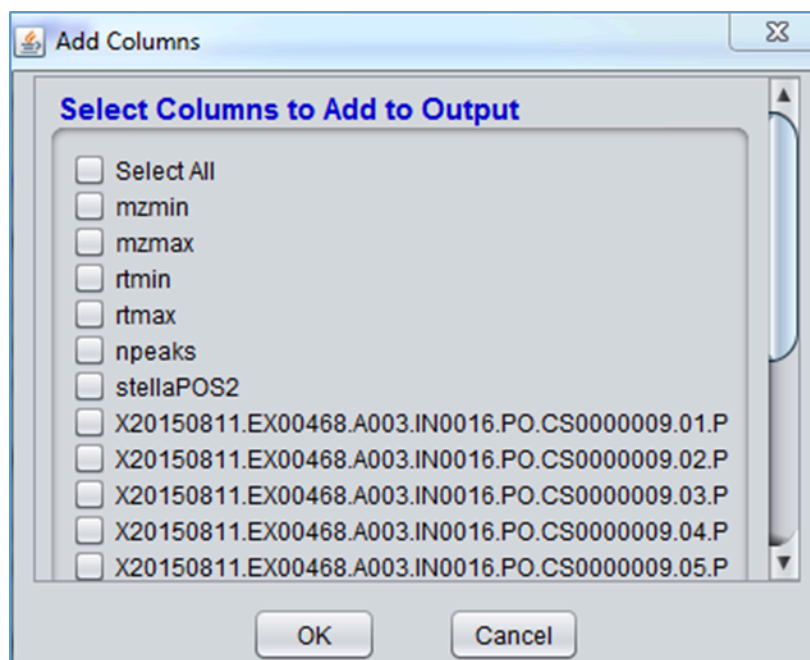


4. Under “Specify Ionization Mode” on the Input tab, select whether the mass spectrometry ionization mode was positive or negative.
5. To move to the next section, click the “Next>>” button or click on a different tab at the top of the window.

Output Options Tab



1. If the input file contained additional columns to be included in the output file, click the “Include Additional Columns...” button under “Specify Additional Columns for Output.”
 - a. A popup window will appear listing the column headings from the input file (see image below).
 - b. Check the box next to a column to add it to the output file.
 - c. Click the “OK” button.



2. Under “Select Location” on the Output Options tab, use the “Browse...” button to choose where to save the output file.
3. Under “Customize Output Tabs,” check the boxes next to the information to be included in the output file. Each option will appear as a separate tab in the output file. To learn more about each tab option, see the [Output File](#) section of this document.

Note: The number and types of tabs selected will impact processing time.

- a. [Correlations](#) – contains correlations between features.
 - b. [Mass Differences](#) – contains the mass differences between features.
 - c. [Intensities](#) – contains the signal intensity (abundance) for each feature.
 - d. [Features/Molecular Ions](#) – contains (a) only the putative molecular ions annotated by Binner; (b) unannotated features.
 - e. [Mass Diff Distribution](#) – contains the distribution of the mass differences across the entire data set.
4. To move to the next section, click the “Next>>” button or click on a different tab at the top of the window.

Data Cleaning Tab

The screenshot shows the 'Data Cleaning' tab in the Binner v0.9.14 software. The interface is divided into several sections: 'Set Filtering Options', 'Set Normalization Options', and 'Set Deisotoping Options'. The 'Data Cleaning' tab is highlighted with a red box. The 'Set Filtering Options' section includes 'Filter Outliers' (Remove Intensities More Than 4.0 Std. Dev. From Feature Mean), 'Filter Features by Missingness' (Remove Features Missing More Than 30.0 % Of Intensity Measurements), and 'Missing Symbols' (Treat 0 Character as Missing Intensity Value). The 'Set Normalization Options' section includes a checkbox for 'Log-Transform Data (using ln(1+x))'. The 'Set Deisotoping Options' section includes checkboxes for 'Deisotope Data' and 'Deisotope Mass Difference Distribution', and input fields for 'Mass Tolerance' (0.002 Da), 'Correlation Cutoff' (0.6), and 'Maximum Retention Time Difference' (0.1). Navigation buttons '<< Previous' and 'Next >>' are at the bottom.

1. Under “Set Filtering Options,” adjust the filtering options to handle outliers and missing data.
 - a. By default, Binner identifies intensities more than 4.0 standard deviations from the feature mean and marks these as missing. The user can change this filter by entering a different number in the text box.
 - b. By default, Binner identifies missing intensity values and removes features that have more than 30% of data missing (including outliers identified above). The user can change this filter by entering a different number in the text box.
 - c. Any intensities marked as missing that still remain in the data set are imputed with the feature median.
2. If data are not yet normalized, under “Set Normalization Options,” check the box next to “Log-Transform Data (using $\ln(1+x)$).”

Note: Data normalization is recommended. If data are already normalized, uncheck this box.

3. Under “Set Deisotoping Options,” click the check box next to “Deisotope Data” to perform deisotoping as part of the data cleaning process. The user can also choose to eliminate isotopes from the mass difference distribution using the “Deisotope Mass Difference Distribution” check box. Select the appropriate parameters to use in the detection and removal of isotopes within a bin.
 - a. By default, Binner removes isotopes with a mass tolerance of 0.002 Daltons or lower. The user can change this parameter by entering a different number in the “Mass Tolerance” text box.
 - b. By default, Binner removes isotopes with a correlation of 0.6 or higher. The user can change this parameter by entering a different number in the “Correlation Cutoff” text box.
 - c. By default, Binner removes isotopes that elute within a retention time gap of 0.1 minutes or less. The user can change this parameter by entering a different number in the “Maximum Retention Time Difference” text box.
4. To move to the next section, click the “Next>>” button or click on a different tab at the top of the window.

Feature Grouping Tab

Binner v0.9.14

Input Output Options Data Cleaning **Feature Grouping** Annotation

Set Binning Gap

Retention Time Gap Size 0.03

Override Bin Size Limits (Not Recommended)

☐ Override Bin Size Limit for Analysis 4000

☐ Override Bin Size Limit for Binwise Output 3000

Select Correlation Measure

☒ Pearson's Correlation ☐ Spearman's Rank Correlation

Choose Bins to Cluster

☒ Bins Below Cutoff Score ☐ Bins Above Cutoff Size ☐ All Bins

Cutoff Value

0 1 2 3 4 5 6 7 8 9 10

Sub-Divide Clusters

Select Grouping Method

☐ Cluster on RT ☒ Rebin on RT

Set Gap Policy

Always break on gaps larger than Retention Time Gap Size

<< Previous Next >>

1. Under “Set Binning Gap,” specify the retention time gap size to be used in determining the bins. A gap size of 0.03 means that a new bin will be formed whenever there is a gap of 0.03 minutes or more between two consecutive retention times.
Note: The gap size is in minutes (default = 0.03 minutes).
2. Under “Select Correlation Measure,” select either Pearson’s Correlation or Spearman’s Rank Correlation.
3. Under “Choose Bins to Cluster,” select the appropriate options for the data. This parameter helps to avoid unnecessary clustering of very highly correlated bins with closely eluting features.
 - a. Bins Below Cutoff Score – will not cluster bins above a specified cutoff score.
 - i. The default score is 2, but the user can change this using the Cutoff Value slider.
 - ii. See the Supplemental [Cutoff Score](#) section for more details about the cutoff score.
 - b. Bins Above Cutoff Size – will not cluster bins below a specified cutoff size.
 - i. The default size is 5, but the user can change this using the Cutoff Value slider.

- c. All Bins – all bins will be clustered.

Note: Clusters are determined using Weighted Silhouette. See the Supplemental [Weighted Silhouette](#) section for more details.

4. Under “Sub-Divide Clusters,” select whether the clusters will be sub-divided by hierarchical clustering on retention time (“Cluster on RT”) or re-binned based on retention time gaps (“Rebin on RT”). This parameter helps sub-divide clusters that encompass a wide retention time range and could include multiple molecular ions.
 - a. If “Cluster on RT” is selected, set the desired parameters for gap size (See the [Sub-clustering](#) section for details).
 - b. If “Rebin on RT” is selected, Binner will break on gaps larger than the RT gap.
5. To move to the next section, click the “Next>>” button or click on a different tab at the top of the window.

Annotation

Annotation File

The annotation file represents existing knowledge about common adducts, neutral losses, etc. For some users, this file will be minimal, while for other users it will be extensive. The annotation file is a .txt file that includes:

- A column containing the annotations.
- Individual columns for mass, mode, and charge for each annotation.
 - If the Mass column contains a negative number, Binner will interpret it as a loss group.
- A column containing tier information (“1” or “2”).
 - Tier 1 charge carriers:
 - Are allowed as a base of an annotation group, e.g., [M+H] for the most intense feature of the annotation group is allowed only if H is a tier 1 charge carrier.
 - Allow complex annotations to appear even in the absence of simpler ones. In particular, an annotation of the form [M+CC+NM] is allowed even when [M+CC] does not appear in that annotation group. Here, M is the putative molecular ion, CC is the tier 1 charge carrier (the same one in both instances), and NM is a putative neutral mass (gain or loss).
 - Tier 2 charge carriers:
 - Are not allowed as a base of an annotation group.
 - Do not allow complex annotations ([M+CC+NM]) to appear in the absence of simpler ones ([M+CC]).
 - Note that in neither case is a charge carrier allowed as a multimer base. In other words, base annotations such as [2M+H] or [3M+K] are not considered for any annotation group.

The file is split into 2 sections:

- Charge-carrying adducts that help form the hypothesis for what the underlying adducts are in the data.
- Neutral additions and losses that complement previously found adducts.

Annotation	Mass	Mode	Charge	Tier
H	1.007276	Positive	1	1
Na	22.9892	Positive	1	1
K	38.9631	Positive	1	2
NH4	18.0345	Positive	1	2
2H	2.014552	Positive	2	1
H+Na	23.99648	Positive	2	1
H+K	39.97059	Positive	2	1
2Na	45.9784	Positive	2	1
3H	3.021828	Positive	3	1
H	-1.007276	Negative	-1	1
Cl-	34.9694	Negative	-1	1
COOH-	44.9982	Negative	-1	1
2H	-2.014552	Negative	-2	1
3H	-3.021828	Negative	-3	1
H2O	-18.0106	Both	0	1
NH3	-17.0254	Both	0	1
Na-H	21.9821	Both	0	1
K-H	37.956	Both	0	1
H+Cl	35.97668	Both	0	1
Acetonitrile	41.02655	Both	0	1
HCOOH	-46.0037	Both	0	1
NaCOOH	67.98742	Both	0	1
KCOOH	83.9513	Both	0	1
NH3+H2O	-35.0359	Both	0	1

Charge-Carrying Adducts

Neutral Differences,
Additions/Losses

Annotation Tab

Binner v0.9.14

Input Output Options Data Cleaning Feature Grouping **Annotation**

Set Parameter Values

Annotation Mass Tolerance 0.002 Da

Annotation RT Tolerance 0.1

Select Annotation File (must be .txt)

annotations_with_tier.txt Import...

Specify Columns

Annotation Annotation

Mass Mass

Mode Mode

Charge Charge

Tier Tier

Set Annotation Preferences

☒ Use Neutral Masses To Help Determine Best Charge Carrier

☐ Allow Variable Charge Without Isotope Information

Generate Annotated Report

Run Analysis

<< Previous Next >>

1. Under “Set Parameter Values,” change the parameters as appropriate for the data.
 - a. By default, Binner looks for annotations that have a mass error of 0.002 Daltons or less. The user can change this parameter by entering a different number in the “Annotation Mass Tolerance” text box.
 - b. By default, Binner looks for annotations that have a retention time tolerance of 0.1 minutes. The user can change this parameter by entering a different number in the “Annotation RT Tolerance” text box.
2. Under “Select Adduct/Fragment File (must be .txt),” use the “Import...” button to find and load the annotation file.
3. Under “Specify Columns,” use the dropdown arrows next to each column type to select the specific column from the annotation file that maps to the column type.
4. Under “Set Annotation Preferences,” select options as appropriate.
 - a. If “Use Neutral Masses To Help Determine Best Charge Carrier” is selected, Binner will include neutral masses to determine the carrier with the maximum number of

annotation “hits.” See the Supplemental [Annotation Preferences](#) section for more information.

- b. If “Allow Variable Charge Without Isotope Information” is selected, all charge carriers in the annotation file will be included in the candidate set regardless of the presence or absence of evidence for isotopes. If it is not selected, Binner excludes multiply charged carriers from the candidate set unless evidence for isotopes suggests their inclusion.

See the Supplemental [Annotation Preferences](#) section for more information.

5. Under “Generate Annotated Report,” click the “Run Analysis” button.
6. A progress bar will appear, keeping the user informed of the progress of the analysis.

Output File

Binner generates an .xlsx file with multiple tabs. This file includes the Summary tab, the Mass Diff Distribution tab, and the Bin Statistics tab, plus tabs specified by the user on the Binner [Output Options Tab](#). The file will be saved at the location specified by the user on the Binner [Output Options Tab](#).

Summary Tab

The first tab is the Summary tab that lists parameters used during the analysis. This includes:

- The version of Binner used during the analysis.
- Information about the input file.
- Data cleaning information, such as the number of removed features, whether or not normalization was performed, and parameters used for deisotoping.
- Information about the Annotation file, including the ionization mode.
- Annotation preferences.
- The location of the saved output file.
- Data analysis information, such as the gap size between bins, the clustering method, and the number of groups and annotations.

Mass Difference Distribution

The Mass Difference Distribution tab (“Mass Diff Distribution”) shows a distribution of the bin-wise mass differences for the entire data set (or the range specified in the Binner [Output Options Tab](#)), highlighting the common mass differences. This allows the user to see which mass differences appear more frequently than would be expected by chance. If not already included in the annotation file, the user can add them for future use. Additionally, this allows for assessment of mass accuracy for the input feature table.

- The Mass Differences column shows the distribution of mass differences.
 - Each dash represents a number of times a specific mass difference appears in the data set.
 - Each color represents a continuous range of mass.
- The Count column shows the actual number of times the mass difference appears in the data set.

- The Annotation column provides the annotation, if it exists in the user-provided annotation file, otherwise “no annotation” is used.

Frequent Mass Differences (without isotopes)	Count	"Peak"(Empirical)	"Peak"(Annotation)	Annotation	Group Count
0.000	8	0.000	0.000	KCOOH + C2H3N+ 2X Acetonitrile	27
0.001	5	-	-	-	-
0.002	1	-	-	-	-
0.003	2	-	-	-	-
0.004	1	-	-	-	-
0.005	1	-	-	-	-
0.006	2	-	-	-	-
0.007	3	-	-	-	-
0.008	3	-	-	-	-
0.009	1	-	-	-	-
0.500	58	-	-	-	-
0.501	94	0.501	NA	No annotation	152
0.994	14	-	-	-	-
0.995	21	-	-	-	-
0.996	28	-	-	-	-
0.997	38	-	-	-	-
0.998	32	-	-	-	-
0.999	41	-	-	-	-
1.000	67	-	-	-	-
1.001	139	-	-	-	-
1.002	301	-	-	-	-
1.003	399	1.003	1.003	isotope	1476
1.004	87	-	-	-	-

Bin Statistics

The Bin Statistics tab provides valuable statistics on the Binner results. This includes:

- Total number of bins and clusters, plus summary statistics on the features per bin and per cluster.
- Cluster size distribution.
- Bin size distribution.
- Clusters per bin distribution.

Correlations

Correlation analysis results are visualized as a heatmap. There are several options for the correlations output, each appearing as a separate tab. Data can be presented on a by cluster or by bin basis. There are 2 color options: local or absolute. Absolute color uses a range of correlation coefficients found in the entire dataset, whereas local color uses the range of the actual data values found within a cluster or bin. Within each Correlations by Bin tab, data can be sorted by retention time or by cluster.

Select the desired output options on the [Output Options Tab](#) in Binner:

Correlations

☐ By Cluster, Local Color
☐ By Cluster, Absolute Color
☐ By Bin, RT Sort, Local Color
☐ By Bin, RT Sort, Absolute Color
☐ By Bin, Cluster Sort, Local Color
☐ By Bin, Cluster Sort, Absolute Color

Correlations by Cluster

The Correlations by Cluster tab (“Corrs by clust”) should be used for initial review of the annotation results.

- See the Supplemental [Isotope Detection](#) section for information on how isotopes are detected within a bin.
- In the Isotopes column:
 - Parent ions are highlighted green.
 - Isotopes are highlighted blue.
- Isotopes are labeled with a lower case “i.”
- Adducts are labeled with an upper case “M.”
- In the Annotations and Derivations columns:
 - Putative molecular ions are in cells highlighted green.
 - Likely redundant features are in cells highlighted yellow.
- In the Correlations columns:
 - The redder the cell, the stronger the correlation.
 - The greener the cell, the weaker the correlation.

Feature	Mass	RT	KMD	Isotopes	Annotations	Derivations	Mass Error	Bin	Corr Cluster	Rebin Subcluster	RT Subcluster	Correlations			
												C2206	C2207	C2208	C2209
C2206	586.274206	15.3961	0.380433	[i283 + 1]				221	2	3	3	1	0.99	0.97	0.95
C2207	585.271052	15.3962	0.382467	[i283]	[M265 + H]	[584.2638 + 1.007276]	0	221	2	3	3	0.99	1	0.98	0.96
C2208	587.277657	15.3962	0.378103	[i283 + 2]				221	2	3	3	0.97	0.98	1	0.95
C2209	607.253078	15.3987	0.424986		[M265 + Na]	[584.2638 + 22.9892]	0.000102	221	2	3	3	0.86	0.82	0.87	0.84

Correlations by Bin

The Correlations by Bin tab (“Corrs by bin”) has similar information to the [Correlations by Cluster](#) tab, except the information is organized by bin instead of by cluster. This tab allows users to evaluate feature relationships in a broader context. Binner provides an option to sort the information in this tab by retention time or cluster.

- See the Supplemental [Isotope Detection](#) section for information on how isotopes are detected within a bin.
- In the Isotopes column:
 - Parent ions are highlighted green.
 - Isotopes are highlighted blue.

- Isotopes are labeled with a lower case “i.”
- Adducts are labeled with an upper case “M.”
- In the Adducts/NLs and Derivations columns:
 - Putative molecular ions are in cells highlighted green.
 - Likely redundant features are in cells highlighted yellow.
- In the Correlations columns:
 - The redder the cell, the stronger the correlation.
 - The greener the cell, the weaker the correlation.
- When sorted by cluster, the clusters are color-coded to easily determine which features belong to the same cluster (see image below).

Feature	Mass	RT	KMD	Isotopes	Annotations	Derivations	Mass Error	Bin	Corr Cluster	Rebin Subcluster	RT Subcluster	Correlations			
												C1341	C1342	C1343	C1344
C1341	245.046741	4.4472	0.226880					51	1	1	1	1	0.19	0.16	0.16
C1342	203.053894	4.4568	0.172837	[i136]	[M169 + Na]	[180.0647 + 22.9892]		0 51	2	2	2	0.19	1	0.93	0.96
C1343	204.057108	4.4585	0.170744	[i136 + 1]				51	2	2	2	0.16	0.93	1	0.87
C1344	383.118808	4.4618	0.308986		[2M169 + Na]	[2(180.0647) + 22.9892]	0.000219	51	2	2	2	0.16	0.96	0.87	1
C1345	219.028945	4.4634	0.215625		[M169 + K]	[180.0647 + 38.9631]	0.001150	51	2	2	2	0.16	0.92	0.86	0.88
C1346	136.072252	4.4805	0.079687					51	3	3	3	-0.17	-0.46	-0.24	-0.47

Mass Differences

The mass differences output is provided by cluster or by bin. When provided by bin, the user can sort the information by retention time, mass difference, or cluster.

Select the desired output options on the [Output Options Tab](#) in Binner:

Mass Differences

☒ By Cluster
 ☐ By Bin, Mass Diff Sort

☒ By Bin, RT Sort
 ☐ By Bin, Cluster Sort

Mass Differences by Cluster

The Mass Differences by Cluster tab (“Mass diffs by clust”) shows the differences in masses between features within the same cluster.

- Commonly recognized mass differences are highlighted with different colors, representing different mass relationships. For example, the mass difference between a sodium adduct and a protonated adduct is 21.98. In the below image, all mass differences close to 21.98 are the same color.

Feature	Mass	RT	Isotopes	Annotations	Derivations	Mass Diffs			
						C1379	C1380	C1381	C1382
C1379	259.039945	4.8005		[M176 + H + Acetonitrile]	[217.0066 + 1.007276 + 41.026549]	0	56.992409	145.055397	35.011471
C1380	202.047537	4.8009	[i140]	[M174 + Na]	[179.0583 + 22.9892]	-56.992409	0	88.062989	-21.980938
C1381	113.984548	4.8011		[M175 + H]	[112.9773 + 1.007276]	-145.055397	-88.062989	0	-110.043927
C1382	224.028475	4.8012		[M174 + Na + Na-H]	[179.0583 + 22.9892 + 21.9821]	-35.011471	21.980938	110.043927	0
C1383	105.033874	4.8019	[i141]	[M174 + H - (H2O+CO+HCN+H2)]	[179.0583 + 1.007276 - 75.032028]	-154.006072	-97.013663	-8.950674	-118.994601
C1384	180.065937	4.8020		[M174 + H]	[179.0583 + 1.007276]	-78.974009	-21.981600	66.081389	-43.962538
C1385	162.055498	4.8026		[M174 + H - H2O]	[179.0583 + 1.007276 - 18.0106]	-96.984448	-39.992039	48.070950	-61.972977
C1386	218.013873	4.8036		[M176 + H]	[217.0066 + 1.007276]	-41.026073	15.966336	104.029324	-6.014602
C1387	106.037112	4.8038	[i141 + 1]			-153.002834	-96.010425	-7.947436	-117.991363
C1388	155.013042	4.8039		[M175 + H + Acetonitrile]	[112.9773 + 1.007276 + 41.026549]	-104.026903	-47.034494	41.028494	-69.015432
C1389	203.051565	4.8039	[i140 + 1]			-55.988380	1.004029	89.067017	-20.976909

Mass Differences by Bin

The Mass Differences by Bin tab (“Mass diffs by bin”) has similar information to the [Mass Differences by Cluster](#) tab, except it shows the differences in masses between features within the same bin. In this tab, the user can sort the information by retention time, mass difference, or cluster.

- Commonly recognized mass differences are highlighted with different colors, representing different mass relationships (see below image).

Feature	Mass	RT	Annotations	Derivations	Mass Diffs				
						C1	C2	C3	C4
C1	137.964096	0.1228			0	10.991903	9.012562	-35.012415	37.994820
C2	126.972193	0.1233	[M1 + H - (H2O+CO)]	[171.9692 + 1.007276 - 46.0037]	-10.991903	0	-1.979341	-46.004318	27.002917
C3	128.951534	0.1337			-9.012562	1.979341	0	-44.024977	28.982258
C4	172.976511	0.1338	[M1 + H]	[171.9692 + 1.007276]	35.012415	46.004318	44.024977	0	73.007235
C5	99.969276	0.1341			-37.994820	-27.002917	-28.982258	-73.007235	0
C6	112.956543	0.1342	[M2 + H - H2O]	[129.9591 + 1.007276 - 18.0106]	-25.007553	-14.015650	-15.994991	-60.019968	12.987267

Intensities

The Intensities tabs (“Unadj intensities” and “Adj intensities”) provide the signal intensity for each feature. The intensity values in the adjusted tab contain the imputed and log-transformed data. The intensity values in the unadjusted tab are the original input data.

Molecular Ions

Depending on the user’s Binner [Output Options Tab](#) selection, this will include putative molecular ions, unannotated features, or both (each as a separate tab).

Additional Columns

Each tab includes the below additional columns. To view these columns, it may be necessary to manually adjust the column widths as some are mostly “hidden” by default. The user will find these collapsed columns between the “Mass Error” and “Bin” columns.

- Molecular Ion Number – reports the “index” of the corresponding molecular ion.
- Charge Carrier – shows the charge carrier for the corresponding annotation.

- Adduct/NL – reports the neutral addition or loss portion of the annotation.
- Mass Error – provides the difference between the observed m/z value and the theoretical m/z value.

Supplemental Information

Cutoff Score

If a bin already has highly correlated and closely eluting features, there is no advantage to clustering. As a result, the [Feature Grouping](#) tab has the “Bins Below Cutoff Score” option which tells Binner not to cluster bins above a specific cutoff score. Binner’s default cutoff score is 2. The cutoff score is equal to:

$$\frac{(Corr_{avg})^2}{\log_2(n)\sqrt{rt_{max} - rt_{min}}}$$

Where:

- $Corr_{avg}$ is the average of the off-diagonal correlations of the bin correlation matrix.
- n is the number of features within the bin.
- rt_{max} is the maximum retention time (in minutes) of the bin features.
- rt_{min} is the minimum retention time (in minutes) of the bin features.

Clustering

After binning and deisotoping, Binner clusters features within each bin using a hierarchical clustering algorithm with average linkage. Feature-to-feature distances are represented by the Euclidian distance between correlation vectors. If ρ_{kl} is the Pearson correlation between feature k and feature l , then the distance between feature i and feature j is calculated as:

$$\sqrt{\sum_k (p_{ik} - p_{jk})^2}$$

Stopping condition: To determine the optimal number of clusters, Binner utilizes a weighted variant of the classic silhouette diagnostic. In the traditional approach, at each clustering step a silhouette value (s_k) is calculated for each clustered feature

$$s_k = \frac{(a_k - b_k)}{\max\{a_k, b_k\}}$$

where:

a_k is the average distance between feature k and all other features in the same cluster

b_k is the minimum average distance between feature k and all features *not* in the same cluster

Typically the cluster configuration with the highest average silhouette is chosen as optimal.

Silhouette weights: One drawback of the s_k silhouette statistic is that it may select a configuration with only 2 clusters when the data naturally falls into a grouping where one cluster is much larger than the rest. In this case, a variant silhouette method² that downweights large clusters can be corrective:

$$sila_k = \frac{s_k}{1+a_k}$$

However, for the problem at hand, a disadvantage of the fully weighted ($sila_k$) diagnostic is that it may overcorrect, splitting the data into many extremely small groups with only 2 or 3 features per cluster. To avoid both under- and over-splitting, Binner uses a hybrid statistic:

$$hsila_k = \frac{s_k}{1+\omega a_k}$$

By examining the cluster size distribution as ω varies from 0.0 (corresponding to s_k) to $\omega = 1.0$ (corresponding to $sila_k$), Binner selects both a weight and a corresponding cluster configuration that best captures the underlying correlation structure.

Sub-clustering

Some correlation-based clusters determined at the clustering step may still span a retention time range not typically observed in co-eluting features. To account for this, initial clusters are further divided by rebinning (splitting on any within cluster gaps greater than the specified RT gap) or by sub-clustering on RT. In the latter case, Binner again utilizes its hierarchical clustering algorithm with feature-to-feature distance measure $|RT_i - RT_j|$ where RT_j is the retention time for feature j .

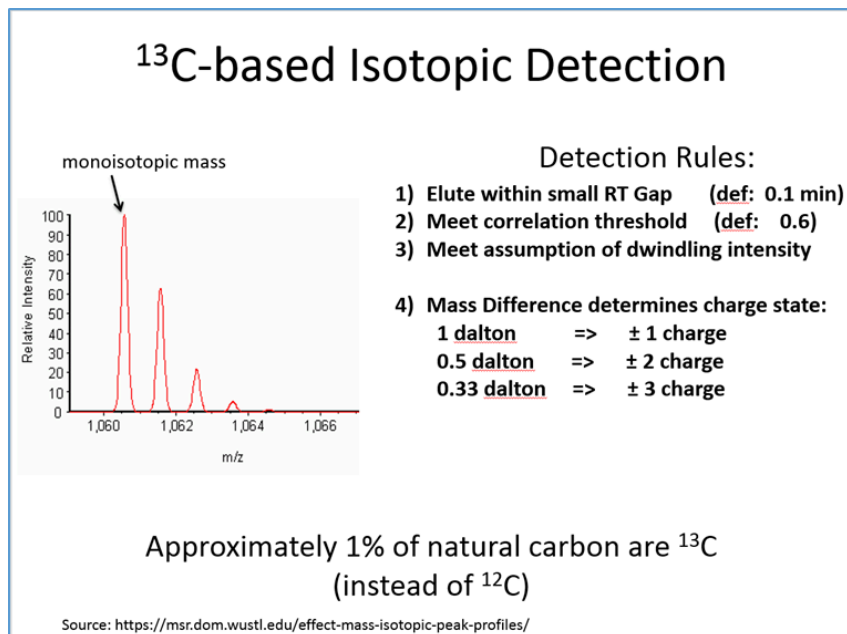
Gap rule: Binner's RT sub-clustering is tunable via a "gap rule" option. This rule specifies both a minimum and a maximum gap size; any natural gaps identified by RT clustering that are smaller than the minimum will not be split; conversely, the data will be split on any gap larger than the maximum even if the clustering algorithm does not identify this gap as a natural split point.

Annotation Preferences

Binner provides two options for users who wish to customize the global annotation behavior of the program. The first option controls the use of neutral masses in the annotation process. If this option is selected, the process of determining the best base adduct of an annotation group will include consideration of secondary annotations of the form [M+CC+NM] in addition to those of the form [M+CC]. Here, M is the putative molecular ion, CC is the putative charge carrier, and NM is a putative neutral mass (gain or loss). If the option is not selected, these types of secondary annotations will not be considered. The second option controls the handling of multiple charges in annotation bases. By default, this option is not selected. This means that multiple charges ($z=2$ or $z=3$) are only allowed for the bases of annotation groups when isotope information specifically indicates that they are present. See the [Isotope Detection](#) section for details. When a putative annotation base has isotope information indicating a multiple charge, this charge is specifically assigned to the feature. In the absence of such isotope information, the feature will be assigned a charge of $z=1$. However, if the user wishes to loosen

this restriction, selecting the second option will allow all charges ($z=1$, $z=2$ and $z=3$) to be considered regardless of any information provided by isotope detection.

Isotope Detection



As shown above, the rules used to detect isotopes within a bin are:

- Elute within a small retention time gap. By default, this is 0.1 minutes
- Meet a correlation threshold. By default, this is 0.6.
- Meet the assumption of dwindling intensity, so each subsequent isotope is smaller than the previous one.
- Determine the charged state on a compound using the difference in mass between two suspected isotopes. For example, if the mass difference is 0.5 Daltons, it is a double charged compound.

References

- (1) Rousseeuw, P. J. *Journal of Computational and Applied Mathematics* **1987**, 20, 53-65.
- (2) Starczewski, A.; Krzyżak, A.; Springer International Publishing: Cham, 2016, pp 114-124.