

Metandem Online Tool User Manual

July 2017

1 Overview

Metandem is a free online platform for the data analysis of mass spectrometry-based isobaric labeling metabolomics. It allows users to upload and process MS or hyphenated MS data files for isobaric-labeling-based quantitative and qualitative metabolomics analyses. Generally, isobaric labeling reagents are comprised of an isotopically-coded reporter, a balance group, and an amine reactive group, such as TMT, ^{1,2} iTRAQ, ^{3,4}, aminoxy TMT ⁵, DiLeu ⁶⁻⁹, DiAla ¹⁰, DiVal ¹⁰, and DiART ¹¹. The balanced structure of each label shifts precursor by the same nominal mass. Upon MS/MS fragmentation, the reporter is cleaved from the balance group, and each labeled sample produces a unique reporter ion whose intensity represents the relative abundance of the labeled molecule from the original sample.

Metandem provides a comprehensive data analysis pipeline integrating feature extraction, metabolite quantification, metabolite identification, batch processing of multiple data files, parameter optimization, median normalization, and statistical analysis. Quantification was achieved on the MS/MS level by extracting reporter ion intensities from the isobaric labeling experiment. Metabolite identification was achieved on the MS1 level by accurate mass matching against the Human Metabolome Database (HMDB) ¹². All results including quantification, identification, parameter optimization, and statistical analysis can be downloaded for offline examination. All graphs are interactive, which can be visualized and edited through Plotly web platform. No user account needs to be set up to use the platform, and the entire data analysis typically takes less than 30 min to be finished.

Isobaric Labeling for Metabolite Analysis

Reporter Ions

# of Ions:			Purity
1	<input type="text" value="115.1252"/>	Da	<input type="text" value="1.0"/>
2	<input type="text" value="115.1315"/>	Da	<input type="text" value="1.0"/>
3	<input type="text" value="116.1286"/>	Da	<input type="text" value="1.0"/>
4	<input type="text" value="116.1349"/>	Da	<input type="text" value="1.0"/>
5	<input type="text" value="116.1407"/>	Da	<input type="text" value="1.0"/>
6	<input type="text" value="117.1318"/>	Da	<input type="text" value="1.0"/>
7	<input type="text" value="117.1377"/>	Da	<input type="text" value="1.0"/>
8	<input type="text" value="117.144"/>	Da	<input type="text" value="1.0"/>
9	<input type="text" value="118.1352"/>	Da	<input type="text" value="1.0"/>
10	<input type="text" value="118.1411"/>	Da	<input type="text" value="1.0"/>

Mass tolerance:

mDa

Mass shift caused by labeling:

Da

Only output data containing all reporter ions

Data Processing

Input files: Uploaded Files One Demo Data File Three Demo Data Files

You can select and upload multiple files. If no data file is uploaded, demo data file will be used.

Batch Processing

Mass tolerance:

ppm

Retention Time tolerance:

min

Median Normalization

Output Parameter Optimization Graphs

Output Statistical Graphs

Metabolite Identification

Perform Metabolite Identification

We have excluded metabolite matches which do not contain N in the chemical formula. (e.g., C₁₁H₈O)

Accurate mass tolerance:

ppm

Test Run (process partial data with Precursor (M+H) between 180 - 220)

Note: this program may run over 10 minutes. Please be patient.

2 Data Analysis

Metandem online tool can analyze both individual data file and multiple data files as batch processing following the same procedure, including data upload, define reporter ion information and parameters, job submission, feature extraction, parameter optimization, quantification, identification, median normalization, and statistical analysis. Users can conduct the data analysis all at once or step by step by unchecking some function boxes.

2.1 Data Upload

Metandem currently supports .txt and .mgf format. Users need to convert raw data files into .txt or .mgf before submitting a job. For thermo .raw data file, we recommend using

COMPASS¹³ to convert the data format into .txt. For other vendor data files, ProteoWizard¹⁴ can be used to convert into .mgf file. Demo data files (both individual demo file and three demo files) are also available for testing.

Data Processing

Input files: Uploaded Files One Demo Data File Three Demo Data Files

You can select and upload multiple files. If no data file is uploaded, demo data file will be used.

2.2 Reporter Ion information

After uploading the data files, users must define the reporter ion information based on their isobaric labeling experiment. N-plex of isobaric labeling can be used to analyze N samples at the same time, generating N reporter ions upon MS/MS fragmentation with unique m/z values. Metandem currently supports up to 20 reporter ion channels for quantification. Because of instrument variations, the recommended input reporter ion masses are the experimental m/z values (with four decimal points) which can be obtained by averaging across several MS/MS spectra. If the theoretical reporter ion masses are used instead, a larger reporter ion mass tolerance might be needed for best results.

When using reporter ion intensities for quantification, purity correction is often necessary to account for the isotopic interference and impurities from synthetic reagents. Metandem provides the function of purity correction for isobaric labeling experiment. The purity of each label can be input along with the accurate masses of reporter ions. The default setting for purity is 1. If commercial isobaric tags are used, such as TMT or iTRAQ, the purity information can be found in the assay package or contacting technical support. For custom isobaric tags like DiLeu^{6,8}, purity values can be calculated with similar method described in our previous publications^{8,15}.

Mass shift caused by labeling is used for the calculation of molecular weight for metabolite identification. Although multiplex isobaric labels should induce the same nominal mass difference, high resolution mass spectrometers might be able to distinguish these subtle differences. Therefore, mass shift caused by labeling is the average accurate masses calculated among all used labels.

In most cases, we recommend select the option of only output data containing all reporter ions to ensure the best data quality. However, if some target molecules are known to be not present in some samples or with an extremely high fold change compared to other samples, this box can be unselected.

Reporter Ions

# of Ions: 4	Purity	
1	115.1252 Da	1.0
2	115.1315 Da	1.0
3	116.1286 Da	1.0
4	116.1349 Da	1.0

Mass tolerance:
0.4 mDa

Mass shift caused by labeling:
145.12801 Da

Only output data containing all reporter ions

2.3 Single Job Creation

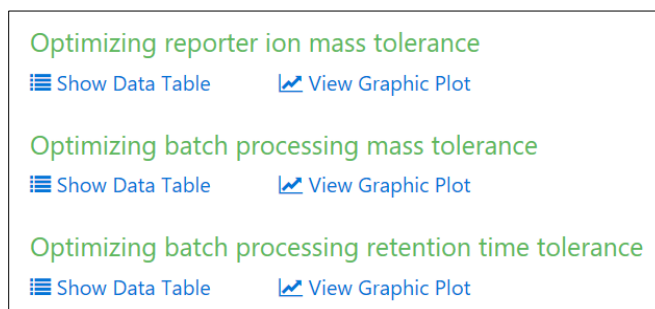
Single job is for the analysis of only one data file. Adjustable data analysis parameters are the reporter ion mass tolerance and accurate mass tolerance for metabolite identification.

2.4 Batch Processing Job Creation

For the analysis of batch data files, users can simply upload multiple data files at once. Those files are recommended to be technical or biological replicates for the best merging results. Merging multiple data files on the feature level is performed based on certain tolerance of precursor masses and retention time of the detected feature.

3 Parameter Optimization

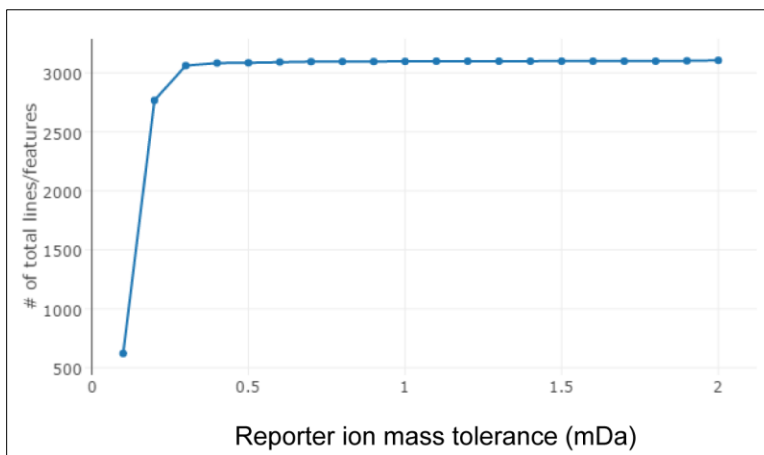
Metandem provides straightforward online parameter optimization function. Different datasets generated from diverse mass spectrometers may require unique parameter settings for the best performance, we therefore recommend perform parameter optimization prior to the actual data analysis. However, users can also skip this step by simply unchecking the box of Output Parameter Optimization Graphs. The default settings are optimized settings for orbitrap mass spectrometry data sets. There are three parameter optimization graphs (POGs) which can be online visualized to select the best parameter settings for custom dataset.



3.1 Reporter Ion Mass Tolerance

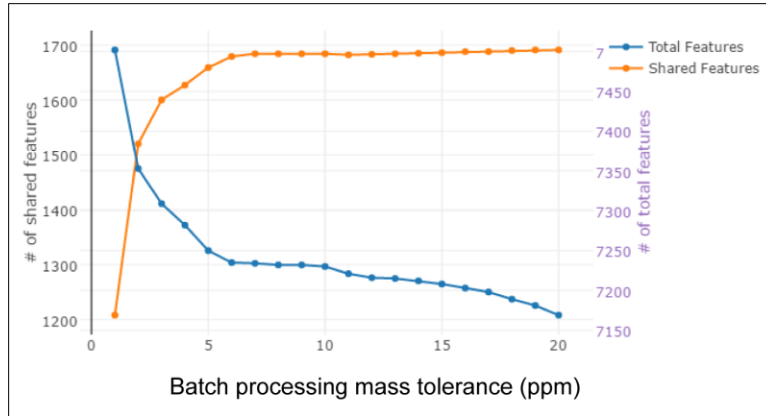
Reporter ion mass tolerance is used to find specific reporter ions based on the input reporter ion masses and extract their intensities from the dataset for quantification. Metandem runs the program automatically with different reporter ion mass tolerance (from 0.1 to 2 mDa) and calculates the corresponding number of total lines/features to generate the POG curve. With

the increase of mass tolerance, the total number of features rises and then reaches a plateau. As shown in the current graph for the demo data files, the optimal reporter ion mass tolerance would be 0.4 mDa.



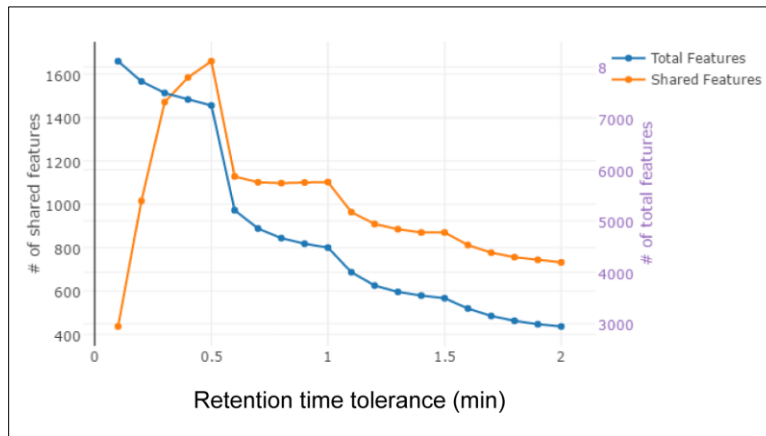
3.2 Batch Processing Mass Tolerance

When multiple data files are uploaded for analysis, quantitative information is generated for each file and merged among multiple files on the feature level. With the increase of batch processing mass tolerance (from 0.1 to 20 ppm), the total number of features dropped as more features can be merged together and the number of shared features rises and then reaches a plateau. As with the demo data files, the optimal batch processing mass tolerance would be 6 ppm.



3.3 Batch Processing Retention Time Tolerance

When merging multiple data files for batch processing, the number of total features and shared features are highly dependent on the precursor retention time tolerance. If the retention time tolerance is over high, different features are incorrectly merged together, causing the decreased number of both total features and shared features. As shown in the current graph for the demo dataset, the optimal tolerance would be 0.5 min.



4 Results

After the job submission, the program will start running. The entire data analysis pipeline typically takes from several seconds to 30 min to be finished. Users can download all the tables and graphs for offline analysis or visualize the interactive graphs online through Plotly web platform. The web address (URL) for output results can also be saved for future view/edit of the graphs.

4.1 Test Run

Test run function is available to provide a quick preliminary evaluation of the dataset by processing partial data with precursor (M+H) between m/z 180-220.

Test Run (process partial data with Precursor (M+H) between 180 - 220)

[Submit](#)

4.2 Results Tables

The output result tables (.csv) include each individual file reporter ion results, merged results, normalized merged results (if median normalization function is selected), and metabolite identification. The input data files can also be downloaded.

Demo input files:

1. [1.txt](#)
2. [2.txt](#)
3. [3.txt](#)

Individual file reporter ion results:

1. [Process_1.csv](#)
2. [Process_2.csv](#)
3. [Process_3.csv](#)

Merged results:

- [merge.csv](#)

Merged results after Median Normalization:

- [merge_median_normalized.csv](#)

Metabolite identification:

- [metabolite_ID.csv](#)

In the table of reporter ion results of each individual file, the first four columns are precursor M+H (m/z), charge, molecular weight (calculated based on the charge and mass shift caused by labeling), and retention time. Intensity of each reporter ion (RP) is extracted and listed in the table, together with reporter ion ratios calculated with RP1 as the denominator.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Precursor (M+H)	Charge	MW	Time	RP1	RP2	RP3	RP4	RP5	RP6	RP7	RP8	RP9	RP10	Ratio RP1	Ratio RP2	Ratio RP3	Ratio RP4	Ratio RP5	Ratio RP6	Ratio RP7	Ratio RP8	Ratio RP9	Ratio RP10
2	221.1734	1	75.0376	5.40	1.E+07	1.E+07	1.E+07	1.E+07	8.E+06	3.E+07	3.E+07	3.E+07	8.E+07	8.E+07	1	0.93	0.78	0.91	0.54	2.19	1.97	2.12	5.80	5.55
3	251.1782	1	105.0423	5.41	2.E+06	2.E+06	2.E+06	2.E+06	2.E+06	4.E+06	3.E+06	4.E+06	1.E+07	1.E+07	1	0.96	0.84	0.79	0.75	1.80	1.59	1.84	4.65	4.67
4	251.1845	1	105.0487	5.41	2.E+06	2.E+06	2.E+06	2.E+06	2.E+06	4.E+06	3.E+06	4.E+06	1.E+07	1.E+07	1	0.82	0.79	0.70	0.72	1.87	1.56	1.89	4.61	4.76
5	613.1592	2	321.8954	5.50	3.E+04	2.E+04	2.E+04	3.E+04	4.E+04	6.E+04	5.E+04	6.E+04	2.E+05	1.E+05	1	0.73	0.79	0.81	1.15	1.77	1.68	1.76	4.73	3.72
6	219.1555	1	73.0196	5.50	2.E+06	2.E+06	1.E+06	2.E+06	1.E+06	9.E+04	5.E+04	7.E+04	1.E+05	9.E+04	1	0.85	0.71	0.96	0.60	0.05	0.03	0.04	0.08	0.05
7	280.1973	1	134.0614	5.51	1.E+05	1.E+05	1.E+05	1.E+05	1.E+05	3.E+05	2.E+05	3.E+05	7.E+05	6.E+05	1	0.95	1.03	1.04	1.07	2.33	1.96	2.65	6.50	5.66
8	809.2445	2	517.9806	5.51	3.E+04	4.E+04	3.E+04	3.E+04	4.E+04	6.E+04	7.E+04	1.E+05	2.E+05	2.E+05	1	1.19	0.97	1.06	1.36	2.04	2.39	3.27	7.16	7.34

In the merge table, each feature is given a number to correspond with the results of metabolite identification, followed by average precursor, charge, average molecular weight, and average retention time across all input files. The average reporter ion ratios and the relative standard deviations across all files are provided in the table, followed by the original RP intensities and ratios from each data file. If normalization function is selected, median normalization is performed for the averaged ratios displayed in a new merge file.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	R
1	Number	Average Precursor (M+H)	Charge	Average MW	Average Time	Average _Ratio_ RP1	Average _Ratio_ RP2	Average _Ratio_ RP3	Average _Ratio_ RP4	Average _Ratio_ RP5	Average _Ratio_ RP6	Average _Ratio_ RP7	Average _Ratio_ RP8	Average _Ratio_ RP9	Average _Ratio_ RP10	RSD_ RP1	RSD_ RP2	RSD_ RP3	RSD_ RP4	RSD_ RP5	RSD_ RP6	RSD_ RP7	RSD_ RP8	RSD_ RP9	RSD_ RP10
2	1	221.1733	1	75.0374	5.48	1	0.95	0.72	0.90	0.52	2.13	1.95	2.06	5.59	5.48	0	0.04	0.09	0.06	0.02	0.02	0.03	0.07	0.03	
3	2	251.1781	1	105.0423	5.57	1	0.86	0.79	0.69	0.73	1.68	1.51	1.76	4.26	4.40	0	0.09	0.08	0.11	0.04	0.07	0.07	0.07	0.09	
4	3	251.1844	1	105.0486	5.41	1	0.86	0.75	0.69	0.71	1.75	1.56	1.77	4.40	4.60	0	0.03	0.04	0.04	0.02	0.06	0.01	0.05	0.03	
5	4	193.1983	1	47.0625	5.47	1	0.14	5.27	43.52	46.35	15.43	13.27	112.67	15.75	23.11	0	0.18	0.05	0.03	0.03	0.05	0.03	0.02	0.07	
6	5	231.1706	1	85.0347	5.60	1	1.33	0.81	0.84	1.18	1.47	1.64	1.72	4.70	4.25	0	0.15	0.14	0.08	0.27	0.05	0.13	0.20	0.09	
7	6	337.0866	1	190.9508	5.43	1	2.56	2.23	2.67	3.86	5.13	4.29	6.02	10.47	11.75	0	0.77	0.87	0.82	0.91	0.77	0.86	0.81	0.69	
8	7	295.1153	1	148.9795	5.63	1	1.09	1.30	1.41	0.79	3.51	3.53	4.00	7.33	10.30	0	0.07	0.07	0.05	0.07	0.01	0.01	0.05	0.04	
9	8	190.1782	1	44.0423	5.69	1	0.39	1.26	0.20	0.20	0.24	0.19	0.22	0.32	0.33	0	0.08	0.02	0.01	0.02	0.01	0.05	0.04	0.03	
10	9	216.1598	1	70.0239	5.49	1	3.78	2.02	1.79	2.86	2.76	2.49	2.82	6.23	5.62	0	0.36	0.41	0.28	0.32	0.19	0.27	0.22	0.29	

Metabolite identification is achieved by accurate mass matching against the Human Metabolome Database (HMDB). Accurate mass tolerance (ppm) can be defined by the user. Input number, average precursor M+H mass, charge, average MW, and average time from the identification table match with the merge file. Monoisotopic MW is the theoretical accurate MW. Metabolite ID number from KEGG, PubChem, and HMDB are provided in the result.

	A	B	C	D	E	F	G	H	I	J	K	L
	Input	Average Precursor		Average	Average			chemical_	monisotopic		pubchem_	
1	Number (M+H)	Charge	MW	Time	dppm	name		formula	_MW	kegg_id	_id	compound
2												hmdb_id
2	20	301.2075	1	155.0717	0.35	14.03	L-Histidine	C6H9N3O2	155.0695	C00135	6274	HMDB00177
3	20	301.2075	1	155.0717	0.35	11.91	4-Phenylpyridine	C11H9N	155.0735	C11310	13651	HMDB33123
4	32	235.1831	1	89.0473	0.56	4.29	L-Alanine	C3H7NO2	89.0477	C00041	5950	HMDB00161
5	32	235.1831	1	89.0473	0.56	4.29	Sarcosine	C3H7NO2	89.0477	C00213	1088	HMDB00271
6	32	235.1831	1	89.0473	0.56	4.29	Beta-Alanine	C3H7NO2	89.0477	C00099	239	HMDB00056
7	32	235.1831	1	89.0473	0.56	4.29	D-Alanine	C3H7NO2	89.0477	C00133	71080	HMDB01310

4.3 Interactive Statistical Graphs

Histogram for detected precursor distribution

[View Graphic Plot](#)

Histogram for retention time distribution

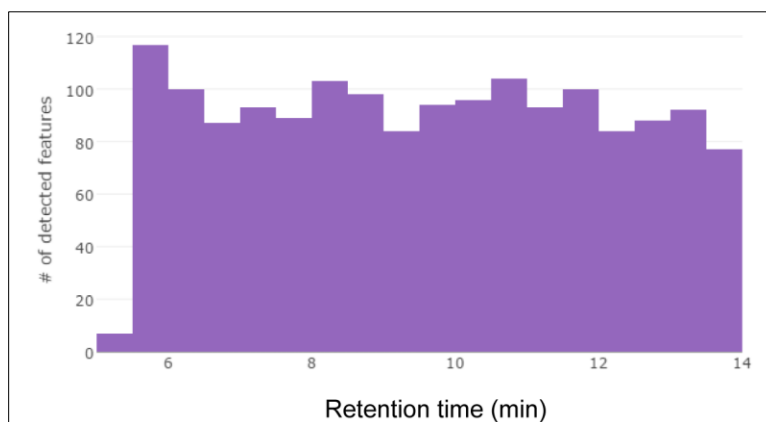
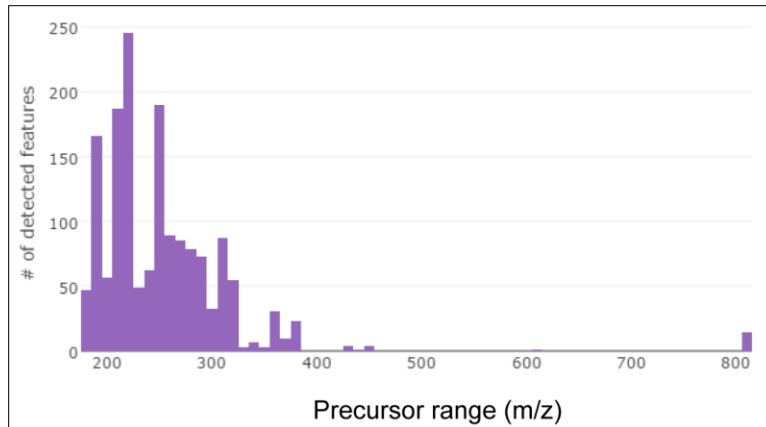
[View Graphic Plot](#)

Box plots for log₂(reporter ion ratio)

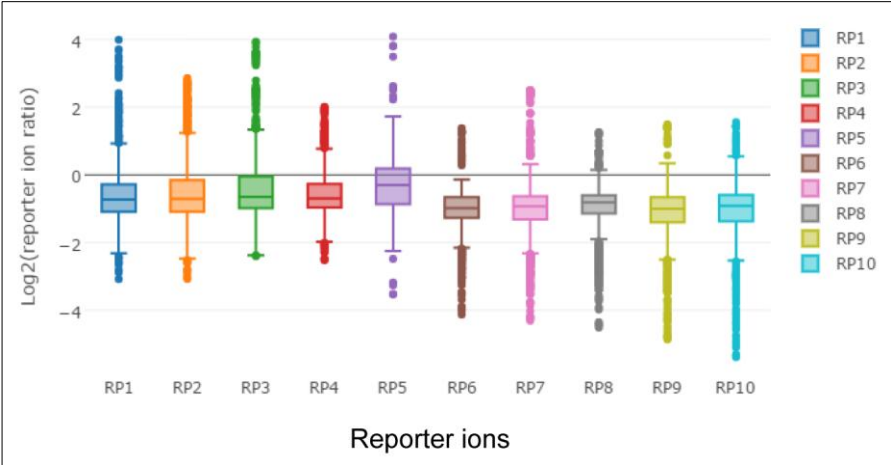
[View Graphic Plot](#)

If single job is created, the individual file of reporter ion results is used for statistical analysis. For batch processing of multiple input data files, Metandem uses the shared features among all merged files to conduct statistical analysis and output as interactive statistical graphs including histogram distributions of precursor mass and retention time, and box plots of reporter ion ratios before and after median normalization. All graphs can be viewed/edited online or downloaded for offline analysis. Parameter optimization graphs are introduced in the previous section 3.

The first histogram is for the distribution of precursor with the X axis as precursor mass range (m/z) and Y axis as the number of detected features shared among merged files. The second histogram represents the distribution of features' retention time (min). The histograms below are generated by the batch processing of three demo data files.



Box plots represent the measured reporter ion ratios across all shared features in the merge .csv file (before and after median normalization). Reporter ion ratios are calculated with intensity of RP1 as the denominator. Box denotes 25th and 75th percentiles; line inside the box denotes the median; whisker denotes standard deviation. The box plot below is generated using batch processing demo data after median normalization.



References

- (1) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. *Anal Chem* **2003**, *75*, 4942-4942.
- (2) Murphy, J. P.; Everley, R. A.; Coloff, J. L.; Gygi, S. P. *Anal Chem* **2014**.
- (3) Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlett-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. *Molecular & Cellular Proteomics* **2004**, *3*, 1154-1169.
- (4) Kaspar, H.; Dettmer, K.; Chan, Q.; Daniels, S.; Nimkar, S.; Davignus, M. L.; Stamler, J.; Elliott, P.; Oefner, P. J. *J Chromatogr B* **2009**, *877*, 1838-1846.
- (5) Zhong, X. F.; Chen, Z. W.; Snovida, S.; Liu, Y.; Rogers, J. C.; Li, L. J. *Anal Chem* **2015**, *87*, 6527-6534.
- (6) Xiang, F.; Ye, H.; Chen, R. B.; Fu, Q.; Li, L. J. *Anal Chem* **2010**, *82*, 2817-2825.
- (7) Frost, D. C.; Greer, T.; Li, L. *Anal Chem* **2015**, *87*, 1646-1654.
- (8) Hao, L.; Zhong, X. F.; Greer, T.; Ye, H.; Li, L. J. *Analyst* **2015**, *140*, 467-475.
- (9) Hao, L.; Johnson, J.; Lietz, C. B.; Buchberger, A.; Frost, D.; Kao, W. J.; Li, L. J. *Anal Chem* **2017**, *89*, 1138-1146.
- (10) Yu, Q.; Shi, X. D.; Greer, T.; Lietz, C. B.; Kent, K. C.; Li, L. J. *Journal of Proteome Research* **2016**, *15*, 3420-3431.
- (11) Yuan, W.; Zhang, J. X.; Li, S. W.; Edwards, J. L. *Journal Of Proteome Research* **2011**, *10*, 5242-5250.
- (12) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res* **2013**, *41*, D801-D807.
- (13) Wenger, C. D.; Phanstiel, D. H.; Lee, M. V.; Bailey, D. J.; Coon, J. J. *Proteomics* **2011**, *11*, 1064-1074.
- (14) Holman, J. D.; Tabb, D. L.; Mallick, P. *Curr Protoc Bioinformatics* **2014**, *46*, 13 24 11-19.
- (15) Sturm, R. M.; Lietz, C. B.; Li, L. J. *Rapid Commun Mass Sp* **2014**, *28*, 1051-1060.