Supporting information for

Optimization of mass spectrometric parameters in data dependent acquisition for untargeted metabolomics on the basis of putative assignments

Hailemariam Abrha Assress[1, 2], Mario G. Ferruzzi[1, 2], Renny S. Lan*[1, 2]

[1]Arkansas Children's Nutrition Center, Little Rock, Arkansas, 72202

[2]Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, 72205

*Slan@uams.edu

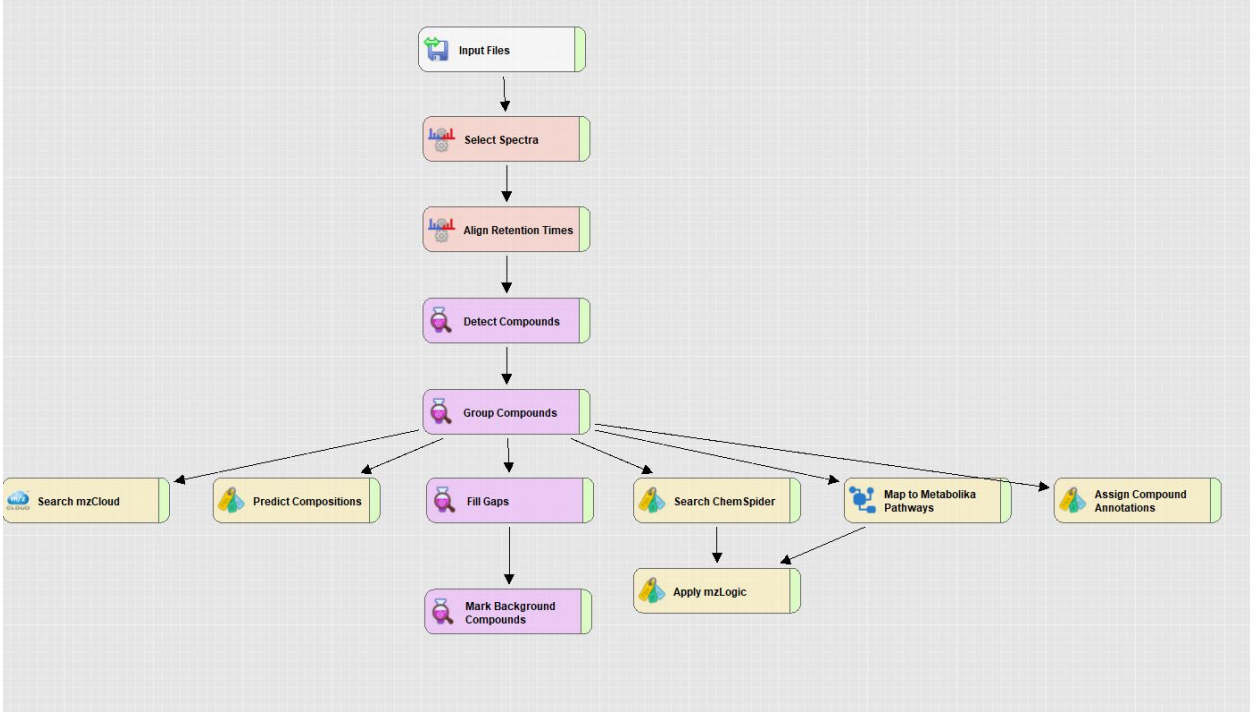


Figure S1: Workflow tree from Compound Discoverer 3.2.0 software displaying data processing nodes and the associated workflow connections. Nodes for initial data processing such as Input Files, Select Spectra nodes, and Align Retention times are included. The Detect Compounds node is used to implement peak detection for unknowns. The Group Compounds node integrates information from multiple raw data files. The Predict compositions, Search mzCloud, Search ChemSpider, and Apply mzLogic nodes were utilized as compound annotation nodes Based on both MS1 and MS2 spectrum data. The Assign Compound Annotations node was used to prioritize the annotation source. Lastly, the Mark Background Compounds node was used to indicate compounds arising from the experimental blank.

Table S1: Description of the general parameter and their settings used for each data processing node during data processing using the compound Discoverer 3.2.0

Node	General parameter and settings
Input files	Raw data were imported
Select spectra	Precursor Selection: Use MS(n - 1) Precursor Use Isotope Pattern in Precursor Reevaluation: True Provide Profile Spectra: Automatic Store Chromatograms: False Lower RT limit: 0 Upper RT limit: 15 Total intensity threshold: Minimum peak count: Min. precursor mass: 50 Da Max. precursor mass: 1000 Da Scan type: any Min. Collision Energy: 0
Align Retention Times	Alignment Model: Adaptive curve Alignment Fallback: Use Linear Model Maximum Shift [min]: 2 Shift Reference File: True Mass Tolerance: 5 ppm Remove Outlier: True
Detect Compounds	Mass Tolerance [ppm]: 5 ppm Intensity Tolerance [%]: 30 S/N Threshold: 3 Min. Peak Intensity: 50,000 Ions: [2M+H]+1; [2M-H]-1; [M+2H]+2; [M+H]+1; [M+H+Na]+2; [M+H-H2O]+1; [M+Na]+1; [M-2H]-2;; [M-H]-1; [M-H-H2O]-1 Base Ions: [M+H]+1; [M-H]-1 Min. Element Counts: C H Max. Element Counts: C90 H190 Br3 Cl4 K2 N10 Na2 O18 P5 S5 Filter Peaks: True Max. Peak Width [min]: 1 Remove Singlets: True Min. Scans per Peak: 6 Min.Isotopes:2
Group Compounds	Mass Tolerance: 5 ppm RT Tolerance [min]: 0.2 Preferred Ions: [M+H]+1; [M-H]-1
Fill gaps	Mass Tolerance: 5 ppm S/N Threshold: 1.5

	Use Real Peak Detection: True
Mark Background Compounds	Max. Sample/Blank: 5 Max. Blank/Sample: 0 Hide Background: True
Search mzCloud	Compound Class: all Precursor Mass Tolerance: 10 ppm FT fragment Mass Tolerance: 10 ppm Library: Autoprocessed; Reference Max. # Results:10 Identity search: Cosine Match activation type: True Activation energy tolerance:20
Predict Compositions	Mass Tolerance: 5 ppm Min. Element Counts: C H Max. Element Counts: C90 H190 Br3 Cl4 N10 O18 P5 S5 Min. RDBE: 0 Max. RDBE: 40 Min. H/C: 0.1 Max. H/C: 4 Max. Candidates: 10 Max. Internal Candidates: 200 Intensity Tolerance [%]: 30 Intensity Threshold [%]: 0.1 S/N Threshold: 3 Min. Spectral Fit [%]: 30 Min. Pattern Cov. [%]: 90 Use Dynamic Recalibration: True Use Fragments Matching: True Mass Tolerance: 5 ppm
Search ChemSpider	Databases(s): BioCyc; Human Metabolome Database; KEGG Search Mode: By Formula or Mass Max. # of results per compound: 10 Result Order (for Max. # of results per compound): Order By Reference Count (DESC) Check all predicted Composition: False
Map to Metabolika Pathways	Metabolika Pathways: All Search Mode: By Formula or Mass Mass Tolerance: 5 ppm Max. # of results per compound: 3
Apply mzLogic	FT Fragment Mass Tolerance: 10 ppm IT Fragment Mass Tolerance: 0.4 Da Max.# Compounds: 0

	Max. # mzCloud Similarity Results to consider: 10 Match Factor Threshold: 30
Assign Compound Annotations	Mass Tolerance: 5 ppm Data Source 1: mzCloud Data Source 2: ChemSpider Data Source 3: Predicted Compositions Data Source 4: Metabolika Search Use mzLogic: True Use Spectral Distance: True SFit Threshold: 20 SFit Range: 20

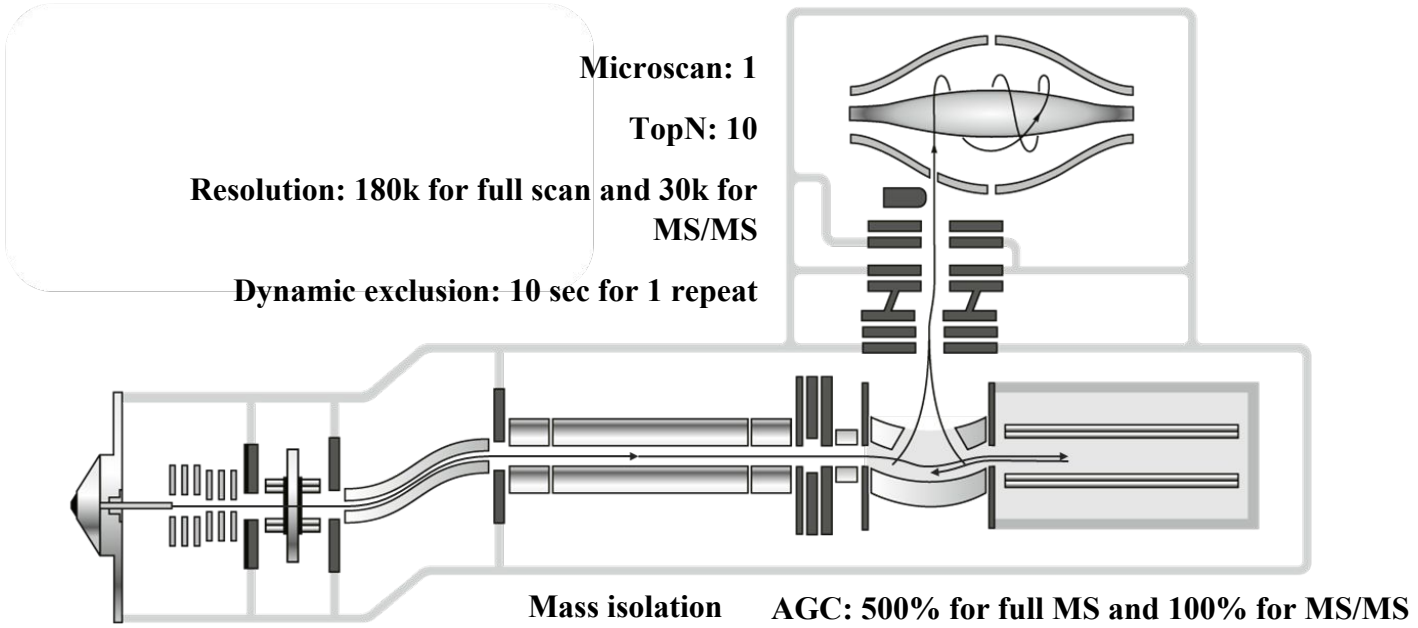


Figure S2: Summary of the Optimum values for the investigated mass spectrometric parameters

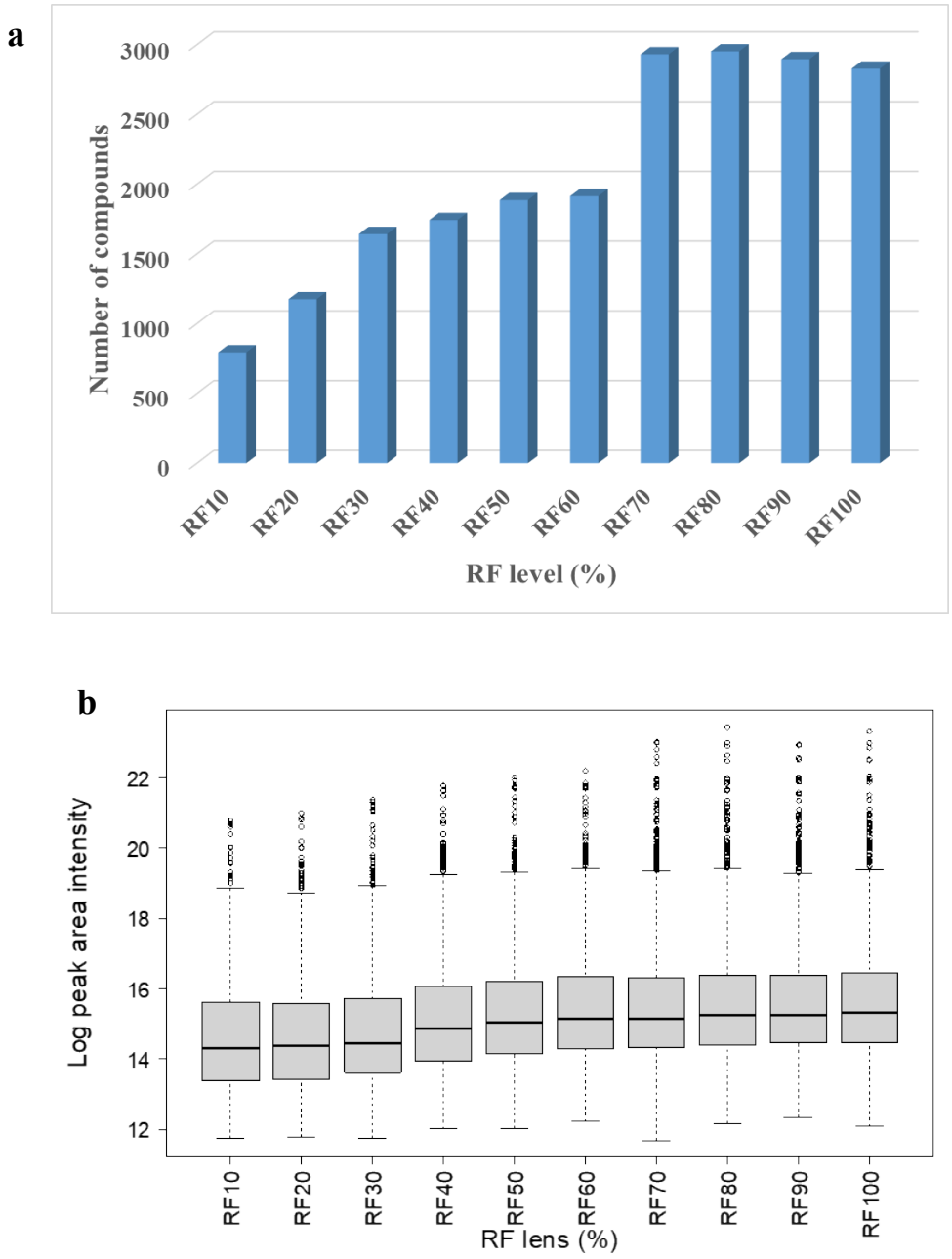


Figure S3: a) Bar graph representation of the effect of % RF lens on the number of compound annotation, (b) boxplot for peak area intensity. The bars in the box plot show the minimum and maximum values, while the line within the box shows the median value.

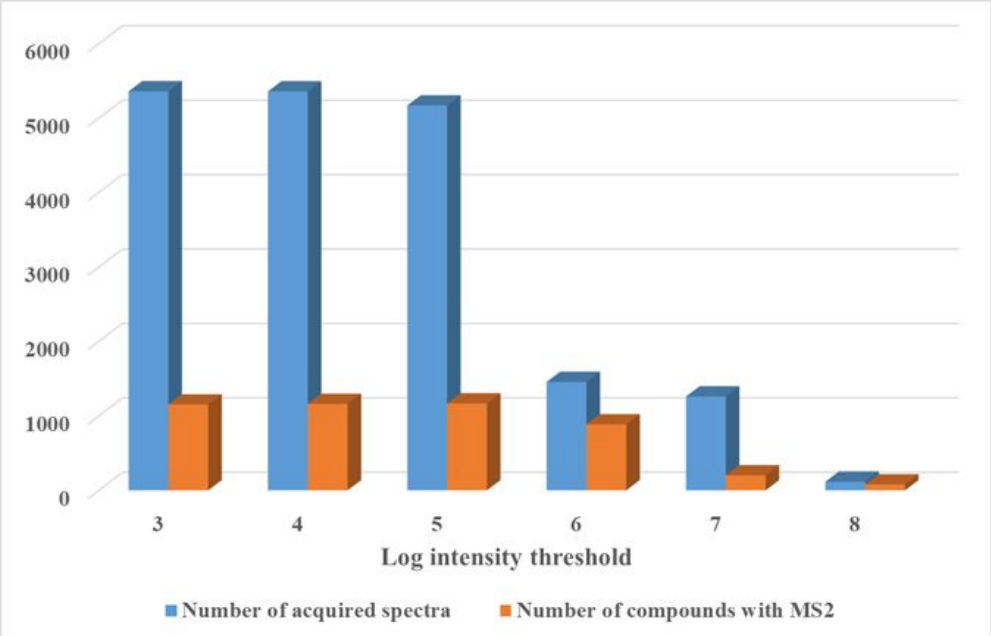
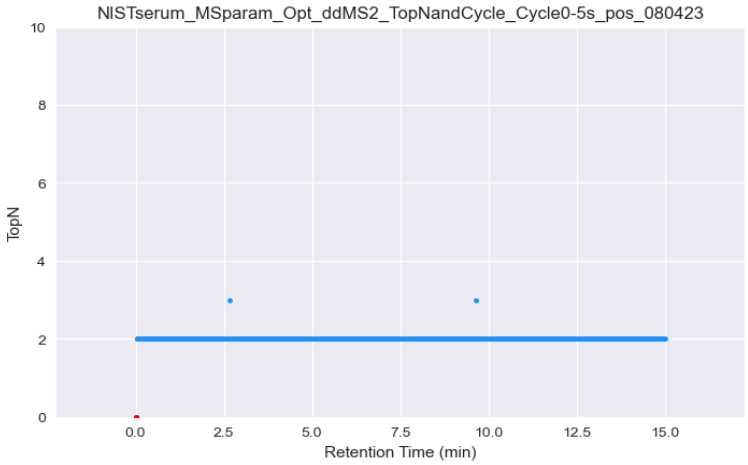
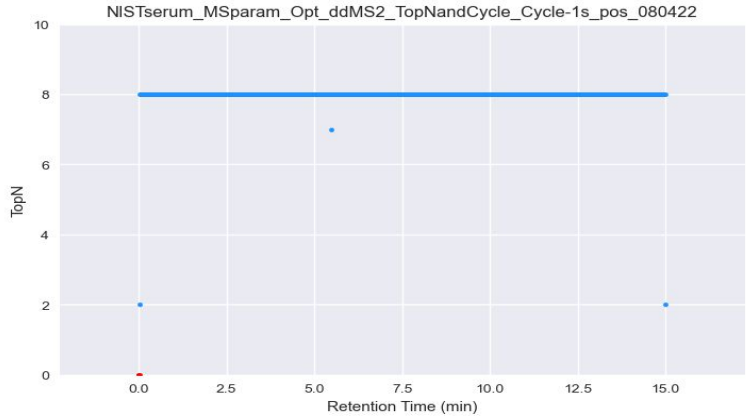


Figure S4: The effect of signal intensity threshold on the number of acquired MS/MS spectra and annotation of metabolites.

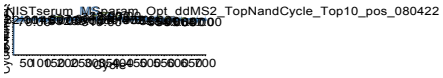
a


NISTserum_MSparam_Opt_ddMS2_TopNandCycle_Cycle0-5s_pos_080423
TopN
Retention Time (min)
0
2
4
6
8
10
0.0
2.5
5.0
7.5
10.0
12.5
15.0

b


NISTserum_MSparam_Opt_ddMS2_TopNandCycle_Cycle-1s_pos_080422
TopN
Retention Time (min)
0
2
4
6
8
10
0.0
2.5
5.0
7.5
10.0
12.5
15.0

c


NISTserum_MSparam_Opt_ddMS2_TopNandCycle_Top10_pos_080422

d

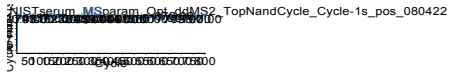

NISTserum_MSparam_Opt_ddMS2_TopNandCycle_Cycle-1s_pos_080422

Figure S5: Number of data dependent scans (TopN) performed at 0.5 s cycle time (a) and 1 s cycle time (b). Experimental cycle time when Top10 is used (c) and when a cycle time of 1 s (d) is used.

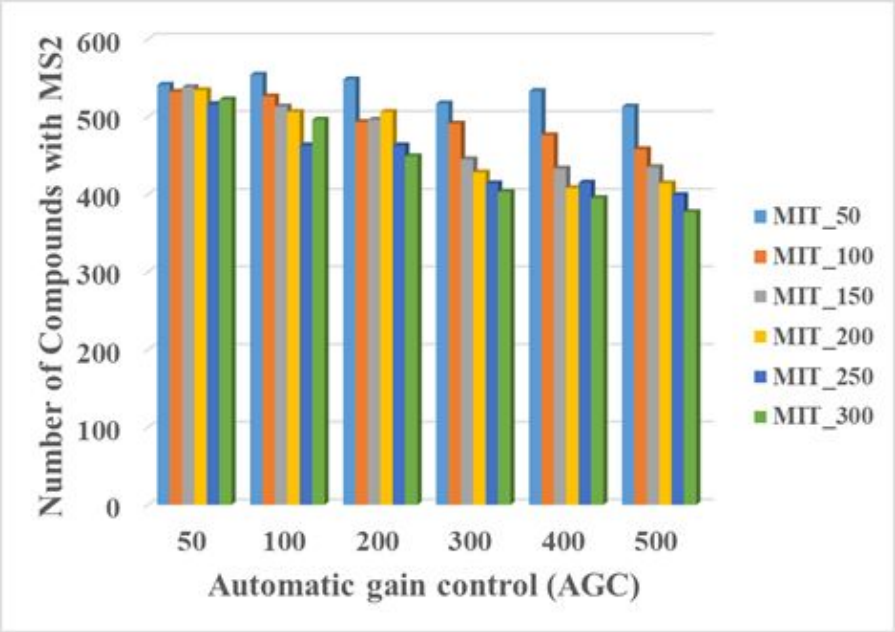


Figure S6: The effect of MIT on the number of acquired MS/MS spectra and compound annotation.