Mass Isotopologue Distribution Analysis (MIDA)

Van Andel Research Institute Metabolomics Core
Brejnev Muhire, Christine Isaguirre and Ryan Sheldon

MIDA is a computer program used for isotope natural abundance correction using either labeled samples or theoretical mass isotopologues distribution values (MIDs). It generates natural abundance corrected values in comma-separated format and high-quality figures. It is written in R and is equipped with a user-friendly ShinyApp graphical interface. (pub.link)

1. Program structure

- app.R, server.R, ui.R : source code graphical user-interface
- mida.R : R script used to run MIDA in command line mode
- utils.R : functions used throughout the program
- MIDA.bat : batch file used to launch the MIDA ShinyApp on Windows OS

2. Input format

Input data consist of table of compound concentration values and a metadata describing samples and file listing names of all compounds in .csv, xls or xlsx format.

					Pyruvate Results	Pyruvate M+1 Results	Pyruvate M+2 Results	Pyruvate M+3 Results	Lactate Results	Lactate M+1 Results	Lactate M+2 Results	Lactate M+3 Results
Name	Data File	Type	Level	Acq. Date-Time	Resp.	Resp.	Resp.	Resp.	Resp.	Resp.	Resp.	Resp.
Blank_1	Blank_1.D	Sample		7/15/2019 16:31	15380	1980	614	1216	163959	32336	377893	40380
0min_a	0min_a.D	Sample		7/15/2019 17:14	6442646	868006	324565	24364	5217190	1250544	830266	103262
0min_b	0min_b.D	Sample		7/15/2019 17:57	5020177	662908	240234	16755	5064552	1234812	821446	119660
0min_c	0min_c.D	Sample		7/15/2019 18:40	3717213	478084	174101	12140	3601465	849116	698383	75078
Blank_2	Blank_2.D	Sample		7/15/2019 19:23	29424	2561	928	1353	333694	66642	416185	18587
20min_a	20min_a.D	Sample		7/15/2019 20:06	6701832	936370	344214	26422	4922736	1189268	821469	128954
20min_b	20min_b.D	Sample		7/15/2019 20:49	8707993	1223262	457281	37342	5477819	1295587	858876	99288
20min_c	20min_c.D	Sample		7/15/2019 21:32	7724890	1075413	388054	32863	4630630	1075568	797195	84972
Blank_3	Blank_3.D	Sample		7/15/2019 22:15	35560	4736	1959	76	394830	76273	420337	28706
1h_a	1h_a.D	Sample		7/15/2019 22:58	5181952	685477	258728	22756	4536691	1088041	797703	100957
1h_b	1h_b.D	Sample		7/15/2019 23:41	7096782	953857	363127	31163	5197828	1226366	842972	
1h_c	1h_c.D	Sample		7/16/2019 0:24	7841484	1044085	398633	33100	5483322	1347829	879188	201018
Blank_4	Blank_4.D	Sample		7/16/2019 1:07	32889	3144	1115	1482	487766	109553	414711	15110
2h_a	2h_a.D	Sample		7/16/2019 1:50	9105700	1303694	483999	43042	5864504	1409711	906430	116367
2h_b	2h_b.D	Sample		7/16/2019 2:33	6556746	873036	323530	30224	5589634	1333775	877744	105710
2h_c	2h_c.D	Sample		7/16/2019 3:16	7121191	951436	365710	30754	5431881	1359129	857430	137872

Fig. 1: Uncorrected data

Name	Labeling
Blank_1	blank1
0min_a	unlabeled
0min_b	unlabeled
0min_c	unlabeled
Blank_2	blank2
20min_a	labeled_sample1
20min_b	labeled_sample1
20min_c	labeled_sample1
Blank_3	blank3
1h_a	labeled_sample2
1h_b	labeled_sample2
1h_c	labeled_sample2
Blank_4	blank4
2h_a	labeled_sample3
2h_b	labeled_sample3
2h_c	labeled_sample3

Fig. 1: Meta Metadata

Compound	Included
Pyruvate	1
Lactate	1
L.Alanine	1
L.Glycine	1
X3.Hydroxybutyrate	0
L.2.Aminobutanoic.Acid	0
Beta.Alanine	0
N.Acetyl.glutamate	0
L.Valine	0
L.Leucine	0
L.Isoleucine	0
Succinate	1
L.Proline	1
Fumarate	0
Pyroglutamate	0
L.Methionine	1
Serine	1
alpha.Ketoglutarate	0
L.Threonine	0
L.Phenylalanine	0
Malate	1

Fig. 3: Compound list

Note:

All files need to in the same input directory and have the exact matching prefix but different suffix ("_uncorr.csv", mtd.csv, and 'cmpd.csv')
Eg: "BHB_uncorr.csv", "BHB_mtd.csv", "BHB_cmpd.csv"

Multiple experiment file can be put in one directory given that the have different prefix name.

In compound list 1 and 0 respectively indicate compounds which should or should not be corrected.

…

3. Workflow

- **Running MIDA**

Place the input *_uncorr.csv files and the associated *_mtd.csv and *_cmpd.csv files in one folder. Within the MIDA App GUI use the "Choose folder" button to select the input folder.

The settings tab allows to set the following:

- Correction method: whether to use labeled samples or theoretical MIDs
- Baseline detection level: The threshold detection level at which compounds are filtered
- List of compounds: correct for all compounds or a user-selected list (in "_cmpd.csv")
- Running mode: skip or overwrite the existing output directories of processed input files

To run the command using the command prompt

Rscript mida.R InputPath List 2000 overwrite or
Rscript mida.R InputPath All 1000 skip

- **Data processing**

1. Reads / checks input table for samples, metadata and compound list
   - Check the ordering of isotopes (M+0, M+1, M+2…M+n)
   - Select compounds as indicated in the compound list "_cmpd.csv" file
   - Read metadata and check whether samples names match those in the input table.
2. Substitute missing values with zero
3. Imputation of missing values (per group, per isotopologue, min val > 0)
4. Remove all compounds below baseline detection (eg. 2/3 of unlabeled samples M+0 are below 1000 or all labeled sample entries are < 1000)
5. Natural abundance correction: for each compound, create the correction matrix CM, from unlabeled data (for multiple sample, the mean of each isotopologue is used). Each labeled sample is corrected using the following formula:

$$\mathbf{A_{corr}} = \mathbf{[CM]^{-1}} \times \mathbf{A_{obs}}$$

   $[CM]^{-1}$: inverse of correction matrix, A_{obs}: labeled sample MIDs, Acorr : corrected MIDs
   Note: CM can also be obtained from theoretical MIDs for a given compound
6. Removing compounds for which 2/3 or more of unlabeled samples (M+0) have total fraction enrichment <95 or >105.
7. Generating figures and tables.

4. Output

The output is generated in same folder as input file with name prefix matching corresponding input plus a suffix "_out". Eg. "BHB_out". The output consisSt of the following:

Figures:
Total Enrichment Bar Plots.
Percent Enrichment Bar Plot
TCA Cycle show a bar plot of each compound
Heatmap compound x samples
Coefficient of variation across samples

Tables:
Corrected MIDs in .csv format
Excel formatted data for input for PRISM

The figures and tables can be visualized using the program interface. Example of figures are show bellow:

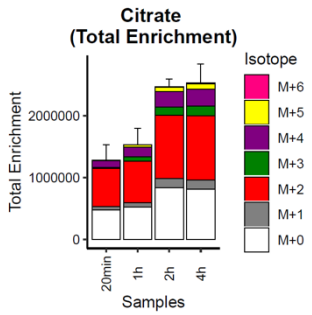

Fig 4 Total enrichment barplot

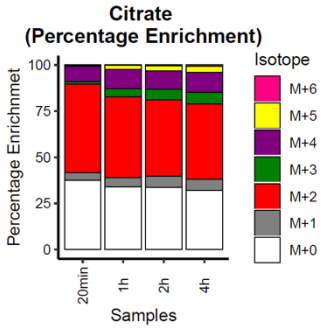

Fig 5 Percentage enrichment barplot.

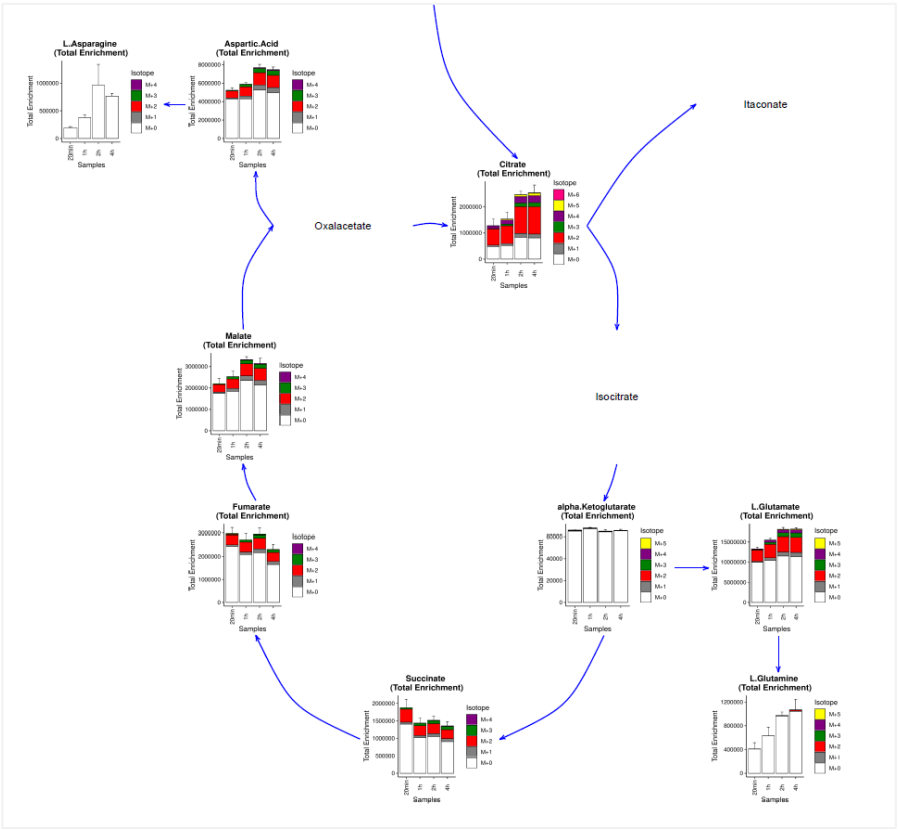

Fig 6 TCA cycle plot.

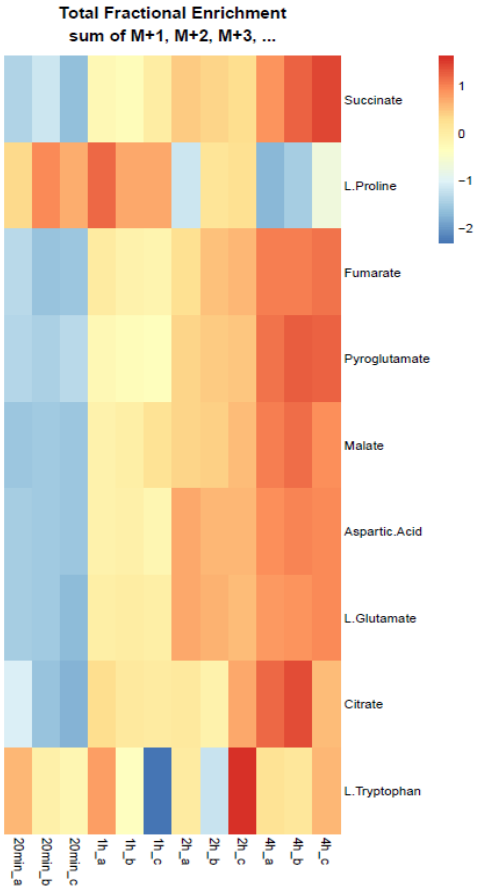

Fig. 7: heatmap

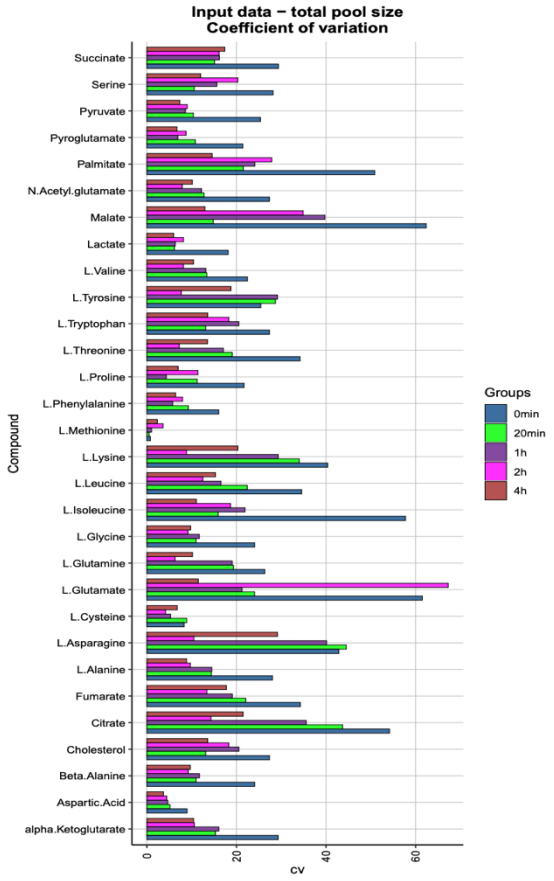

Fig. 8: Input total pool size coefficient of variation