A picture containing clipart

Description automatically generated**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

1. **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.

A screenshot of a cell phone

Description automatically generated

Fig. 1: Uncorrected data

**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.

…

**A screenshot of a cell phone

Description automatically generated**A screenshot of a cell phone

Description automatically generated

**Fig. 1: Meta Metadata Fig. 3: Compound list**

1. **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.

1. Substitute missing values with zero
2. Imputation of missing values (per group, per isotopologue, min val > 0)
3. Remove all compounds below baseline detection (eg. 2/3 of unlabeled samples M+0 are below 1000 or all labeled sample entries are < 1000)
4. 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:

**Acorr** = **[CM] -1 x** **Aobs**

[CM]-1: inverse of correction matrix, Aobs: labeled sample MIDs, Acorr : corrected MIDs

Note: CM can also be obtained from theoretical MIDs for a given compound

1. Removing compounds for which 2/3 or more of unlabeled samples (M+0) have total fraction enrichment <95 or >105.
2. Generating figures and tables.
3. **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:**

**Tables:**

Corrected MIDs in .csv format

Excel formatted data for input for PRISM

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

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

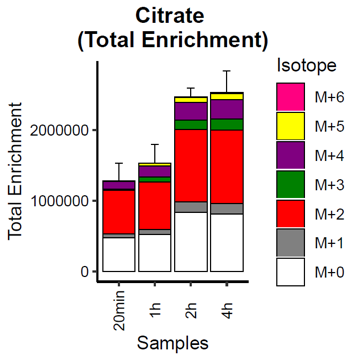
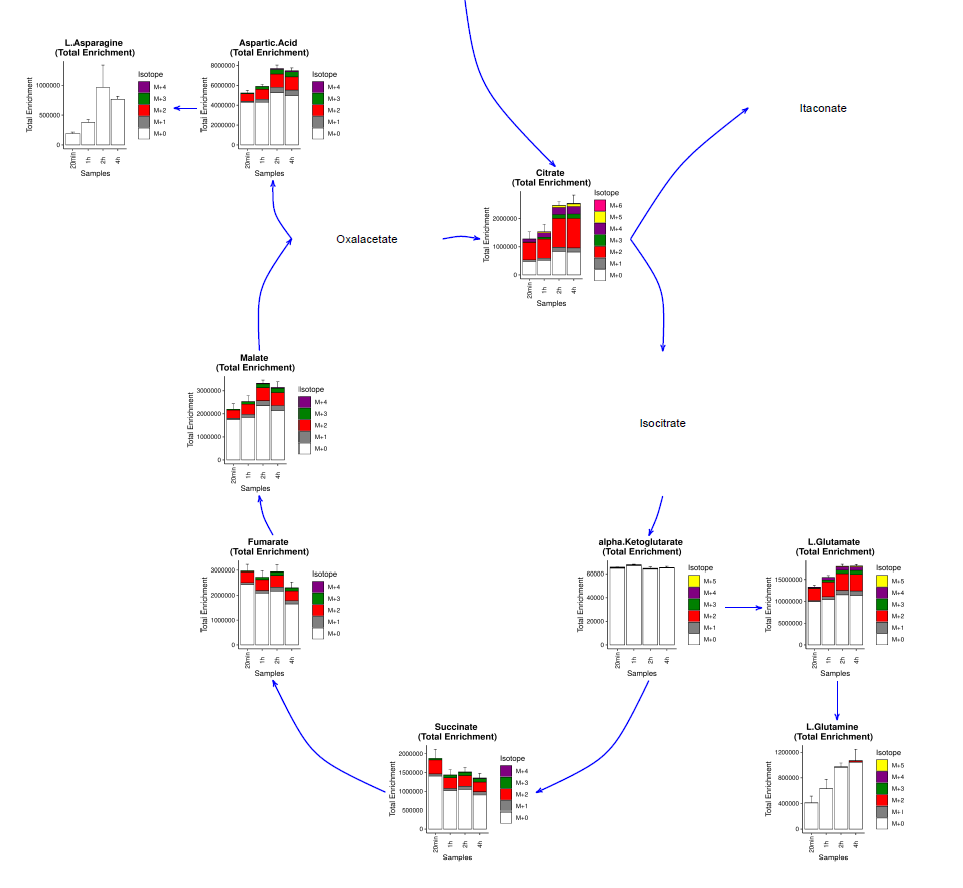


Fig 4 Total enrichment barplot

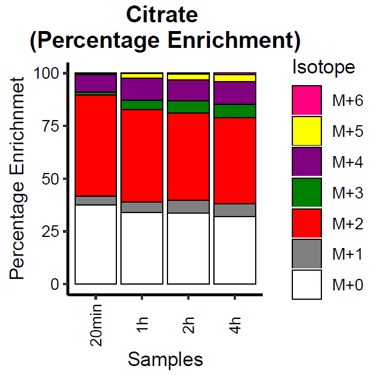


Fig 5 Percentage enrichment barplot. Fig 6 TCA cycle plot.

A screenshot of a computer

Description automatically generatedA screenshot of a cell phone

Description automatically generated

Fig. 7: heatmap Fig. 8: Input total pool size coefficient of variation