Data-Dependent-Assisted Data-Independent Acquisition (DaDIA.R) User Manual

(Version 4.0; Sep 21th, 2020)

Jian Guo, Sam Shen, Tao Huan

- DaDIA.R is a metabolic feature extraction and annotation workflow for DaDIA metabolomics workflow.
- The program is written in the language 'R' and is publicly available at GitHub (https://github.com/HuanLab/DaDIA.git)
- Please see below for instructions on using the DaDIA.R code:

1) Download the R-scrip "DaDIA_SWATHprocessing.R" from Github
2) Install libraries "XCMS", "MSnbase", "dplyr", "doParallel", "foreach", "metaMS", and "CAMERA" if previously not installed (R Version 4.0 or above and XCMS Version 3.11.4 is required; all other packages should be updated to the newest available version)
3) In line 17 – 63, set required parameters to desired value according to onscreen prompts (See below).

```
############################################################
#Part 1: Parameters for feature extraction
DDA.directory <- "C:/Users/User/Desktop/SAM DONT TOUCH DONT DELETE/DaDIAtestSciexData/DDA"
DIA.directory <- "C:/Users/User/Desktop/SAM DONT TOUCH DONT DELETE/DaDIAtestSciexData/DIA"
cwpDDA <- CentWaveParam(ppm=10,
                        peakwidth=c(5,60),
                        mzdiff = 0.01,
                        snthresh = 6,
                        integrate = 1,
                        prefilter = c(3,100),
                        noise = 100) #XCMS parameters for DDA feature extraction
cwpDIA <- CentWaveParam(ppm=10,
                        peakwidth=c(5,60),
                        mzdiff = 0.01,
                        snthresh = 6,
                        integrate = 1,
                        prefilter = c(3,100),
                        noise = 100) #XCMS parameters for DIA feature extraction
mass.tol <- 10 #mz tolerance in ppm: used in feature dereplication and MS2 matching
mass.const.tol <- 0.05 #mz tolerance in constant value: used in feature rescue
rt.tol <- 60 #rt tolerance in seconds
num.samples <- 11 #enter how many DIA samples here
plot.DaDIA <- TRUE #plot DaDIA features
plot.DaDIA.mztol <- 0.5 #DaDIA feature plotting mz window width
plot.DaDIA.rttol <- 30 #DaDIA feature plotting rt window width
#Parameters for alignment
bw <- 5 #bandwidth of gaussian smoothing kernel to apply to the peak density chromatogram
minfrac <- 0.5 #minimum fraction of samples necessary in sample groups for it to be a valid group
mzwid <- 0.025 #width of overlapping m/z slices to use for grouping peaks across samples
max <- 100 #maximum number of groups to identify in a single m/z slice
quantitative.method <- "maxo"
# "maxo" = peak height
# "into" = peak area
############################################################
#Part 2: Parameters for database search (dot product)
feature.annotation <- TRUE #annotate DaDIA features
db.name <- "convertedLibrarPos.Rds" #annotation library name
RDS <- TRUE
# "TRUE" = database is in RDS format
# "FALSE" = database is in MSP format
ms1.tol <- 0.01 #dot product calculation ms1 tolerance
ms2.tol <- 0.02 #dot product calculation ms2 tolerance
dot.product.threshold <- 0.1 #dot product annotation threshold
match.number.threshold <- 1 #annotation match number threshold
adduct_isotope.annotation <- TRUE #perform CAMERA annotation
export.mgf <- TRUE #export individual MS2 spectra as .mgf
combine.mgf <- TRUE #combine all exported .mgf files
annotation.plot <- TRUE #plot all annotated MS2 spectrum against library
############################################################
```

Table 1. DaDIA parameters needed to be changed according to the user's preference.

Row #	Parameter Name	Parameter Function
18	*DDA.directory*	Set directory containing all DDA .mzxml files
19	*DIA.directory*	Set directory containing all DIA(SWATH) .mzxml files, SWATH isolation window labeling .txt file, and annotation library .msp file
20	*cwpDDA*	Set XCMS parameters for DDA feature extraction
27	*cwpDIA*	Set XCMS parameters for DIA feature extraction

34	*mass.tol*	Set m/z tolerance (± ppm) for feature dereplication and MS2 matching
35	*mass.const.tol*	Set m/z tolerance (± constant value) for feature rescue
36	*rt.tol*	Set retention time tolerance (± sec) for identifying same features
37	*num.samples*	Set number of DIA(SWATH) samples to run
38	*plot.DaDIA*	Set whether to plot extracted DaDIA MS1 features
39	*plot.DaDIA.mztol*	Set DaDIA feature plotting mz window width
40	*plot.DaDIA.rttol*	Set DaDIA feature plotting rt window width
42	*bw*	Set XCMS feature alignment bandwidth
43	*minfrac*	Set XCMS feature alignment minimum sample fraction
44	*mzwid*	Set XCMS feature alignment m/z slice width
45	*max*	Set XCMS feature alignment maximum # of groups / slice
46	*quantitative.method*	Set whether to use max or integrated intensity for calculations; pick between peak height or peak area
51	*feature.annotation*	Set whether to perform MS2 extraction and DaDIA feature annotation
52	*db.name*	Set the name of the library used for dot product annotation
53	*RDS*	Set whether the annotation library is in RDS of MSP format
56	*ms1.tol*	Set MS1 tolerance in dot product calculation
57	*ms2.tol*	Set MS2 tolerance in dot product calculation
58	*dot.product.threshold*	Set annotation dot product score threshold
59	*match.number.threshold*	Set annotation match number score threshold
60	*adduct_isotope.annotation*	Set whether to perform CAMERA adduct and isotope annotation
61	*export.mgf*	Set whether to export MS2 spectrum as individual .mgf files
62	*combine.mgf*	Set whether to concatenate all exported .mgf files into a single .mfg file
63	*annotation.plot*	Set whether to plot all annotated MS2 spectrum against library

4) In line 7, set the directory in the user's computer that contains all DDA samples (Note: there should be only .mzxml files in this folder). In line 8, set the directory in the user's computer that contains all SWATH samples, SWATH pocket definitions in .txt format, and the annotation library in .msp format (Note: there should only be .mzxml, .txt, and .msp files in this folder). See below for a sample format of the SWATH pocket definition file; note that the column headers should be kept the same as the example shown:

RP(+)SWATHwindowsetting.txt

Experiment	MSType	Minmz	Maxmz
0	SCAN	45	1170
1	SWATH	45	70.5
2	SWATH	70.5	95.5
3	SWATH	95.5	120.5
4	SWATH	120.5	145.5
5	SWATH	145.5	170.5
6	SWATH	170.5	195.5
7	SWATH	195.5	220.5
8	SWATH	220.5	245.5
9	SWATH	245.5	270.5
10	SWATH	270.5	295.5
11	SWATH	295.5	320.5
12	SWATH	320.5	345.5
13	SWATH	345.5	370.5
14	SWATH	370.5	395.5
15	SWATH	395.5	420.5
16	SWATH	420.5	445.5
17	SWATH	445.5	470.5
18	SWATH	470.5	495.5
19	SWATH	495.5	520.5
20	SWATH	520.5	545.5
21	SWATH	545.5	570.5
22	SWATH	570.5	595.5
23	SWATH	595.5	620.5
24	SWATH	620.5	645.5
25	SWATH	645.5	670.5
26	SWATH	670.5	695.5

5) Note: if users wish to use MS-Dial library for annotation, they must first convert the raw MS-Dial library to a readable .msp format using the included R script *"convertMSP.R"*
6) Click on "→Source" in the R-Studio interface to begin the DaDIA workflow
7) After running the scrip for single sample runs, a csv file "DaDIAtable.csv" containing all MS1 level features extracted and a csv file "annotated_output.csv" containing feature annotation results will be generated. After running the scrip for multi-sample runs, multiple csv files "n_DaDIAtable.csv" (n is number of DIA(SWATH) samples) containing all MS1 level features extracted for each DIA(SWATH) sample, a csv file "alignedDaDIAtable.csv" containing all MS1-features in different samples aligned together, and a csv file "annotated_output.csv" that adds annotation to the "alignedDaDIAtable" will be generated. If the user wishes to plot DaDIA features or output .mgf files, additional folders and files will be generated. All results generated by the program will be in the *DIA.directory* set by the user.