simExTargId (simultaneous experiment - MS/MS target identification)

WMB Edmands

August 19, 2015

The following illustrates the simExTargId works with example (.RAW) data files obtained on a Thermo FT-ICR-MS:

Create a dummy raw data directory on your computer's hard drive, move half the example data, to initiate the process then move the second half to simulate the actual real-time collection of LC-MS data.

```
# set your working directory to where example directories can be created and results
# output saved
# For example
setwd("C:/")
# create a dummy raw MS1 profiling data directory and analysis directory
# for example
todaysDate <- gsub(" ", "", format(Sys.time(), "%Y %m %d"))
dummyRawDir <- paste0(getwd(), "/", todaysDate, "_simExTargId_Example")
# create directory
dir.create(dummyRawDir)
# analysis directory where all simExTargId output will be stored
dummyAnalysisDir <- paste0(dummyRawDir, "_analysis")
# create directory
dir.create(dummyAnalysisDir)

# simExTargId external directory
extdataDir <- system.file("extdata", package="simExTargId")
# list blank files
blanksRaw <- list.files(extdataDir, pattern="blank", full.names=TRUE)
# list plasma IPA extract files
plasmaRaw <- list.files(extdataDir, pattern="IPA", full.names=TRUE)
# covariates file
coVariates <- paste(extdataDir, "coVariates.csv", sep="/")
# copy coVariates to analysisDir
coVariatesCopy <- paste(dummyAnalysisDir, basename(coVariates), sep="/")
file.copy(from=coVariates, to=coVariatesCopy)
## [1] TRUE
# identify number of slave for parallel processing using parallel package
nSlaves <- parallel::detectCores() - 1

# move the blank files into the temporary directory to start the process
blankRawCopies <- paste(dummyRawDir, basename(blanksRaw), sep="/")
file.copy(from=blanksRaw, to=blankRawCopies)
## [1] TRUE TRUE TRUE
# wind back the clock to simulate the files having been acquired at least 5 mins
# since last modification
windBackTheClock <- function(fileCopy, time){
  Sys.setFileTime(fileCopy, Sys.time() - time)
}
```

```
# apply to newly copied files
sapply(blankRawCopies, windBackTheClock, 300)
## C://20150819_simExTargId_Example/05blank.RAW
##                                        TRUE
## C://20150819_simExTargId_Example/09blank.RAW
##                                        TRUE
## C://20150819_simExTargId_Example/13blank.RAW
##                                        TRUE

# move the plasma samples and wind back the time less than 5 mins
plasmaRawCopies <- paste(dummyRawDir, basename(plasmaRaw), sep="/")
file.copy(from=plasmaRaw, to=plasmaRawCopies)
## [1] TRUE TRUE TRUE
# apply to newly copied files
sapply(plasmaRawCopies, windBackTheClock, 240)
## C://20150819_simExTargId_Example/18-IPAHG1.RAW
##                                         TRUE
## C://20150819_simExTargId_Example/19-IPAHG2.RAW
##                                         TRUE
## C://20150819_simExTargId_Example/20-IPAHG3.RAW
##                                         TRUE

# Start simExTargId function
simExTargId(rawDir=dummyRawDir, analysisDir=dummyAnalysisDir,
            coVar=coVariatesCopy, nSlaves=nSlaves)
## 6 raw data file(s) detected...
##
## Waiting for next raw data file...
##
## Converting to mzXML and initial peak picking:
## 05blank.RAW
## 09blank.RAW
## 13blank.RAW
##
## Starting SNOW cluster with 7 local sockets...
##
## Converting raw files and saving in mzXmlFiles directory...
##
## ...Done
##
## Attaching package: 'snow'
##
## The following objects are masked from 'package:BiocGenerics':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, clusterSplit, parApply, parCapply,
##     parLapply, parRapply, parSapply
##
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, clusterSplit, makeCluster,
##     parApply, parCapply, parLapply, parRapply, parSapply,
```

```
##     splitIndices, stopCluster
## Starting snow cluster with 7 local sockets.
## Detecting features in file # 1 : 05blank.mzXML
## Detecting features in file # 2 : 09blank.mzXML
## Detecting features in file # 3 : 13blank.mzXML
## Saving 01.xcmsSet.RData file...
##
## There are not yet sufficient samples to conduct statistical analyses...
##
## subsequent steps of XCMS processing and data analysis will not yet occur...
##
## Converting to mzXML and initial peak picking:
## 20-IPAHG3.RAW
##
## ...Done
## Detecting features in file : 20-IPAHG3.mzXML
##
##  Detecting mass traces at 10 ppm ...
##  % finished: 0 10 20 30 40 50 60 70 80 90 100 Warning: There were 68 peak data insertion problems.
##  Please try lowering the "ppm" parameter.
##
##  4541 m/z ROI's.
##
##  Detecting chromatographic peaks ...
##  % finished: 0 10 20 30 40 50 60 70 80 90 100
##  987  Peaks.
## Saving 01.xcmsSet.RData file...
##
## There are not yet sufficient samples to conduct statistical analyses...
##
## subsequent steps of XCMS processing and data analysis will not yet occur...
##
## Converting to mzXML and initial peak picking:
## 18-IPAHG1.RAW
## 19-IPAHG2.RAW
##
## Starting SNOW cluster with 7 local sockets...
##
## Converting raw files and saving in mzXmlFiles directory...
##
## ...Done
## Starting snow cluster with 7 local sockets.
## Detecting features in file # 1 : 18-IPAHG1.mzXML
## Detecting features in file # 2 : 19-IPAHG2.mzXML
## Saving 01.xcmsSet.RData file...
##
## obiwarp retention time correction...
## center sample:  18-IPAHG1
## Processing: 05blank  09blank  13blank  20-IPAHG3  19-IPAHG2
## Saving 02.retcor.RData file...
##
## Peak grouping...
## 355 359 362 366 370 374 377 381 385 389 392 396 400 404 407 411 415 419 422 426 430 434 437 441 445 4
```

```
## Saving 03.group.RData file...
##
## zero filling...
## Starting snow cluster with 7 local sockets.
## 1 : 05blank.mzXML
## 2 : 09blank.mzXML
## 3 : 13blank.mzXML
## 4 : 20-IPAHG3.mzXML
## 5 : 18-IPAHG1.mzXML
## 6 : 19-IPAHG2.mzXML
## Saving 04.fillPeaks.RData file...
##
## 01. Saving xcms peak table to 01.peakTables directory...
##
## Preprocessing, PCA outlier removal, statistical analysis and MS2 target identification...
##
##
## 02. Preprocessing peak table...
##
## zero filling...
## log transforming to the base 2.718...
## saving pre-processed peak table to 02.preProc directory...
##
## 03. PCA projection and score cluster/ batch effect identification...
##
## Calculating PCA model 1...
## No outliers identified PCA model 1
## saving PCA results to 03.PCA/006_samples directory...
##
## No outliers were removed therefore no outlier removed peak table will be saved in the 03.PCA/006_samp
##
## Identifying PCA scores clusters with co-variates table...
##
## No PCA score plot clusters were identified...
## 04. statistical analysis, 3 co-variates...
##
## co-variate table columns:
## class
## replicateInjection
## runOrder
##
## 1964 multiple comparisons
##
## Starting SNOW cluster with 7 local sockets...
##
## deconvoluting xcms peak data by retention time and correlation clustering...
##
## hierarchical clustering peak group retention times...
##
## intra RT group correlation clustering 780 rt groups...
##
## Calculating weighted mean for 993 pseudospectra accounting for 1565 of 1964 total features
## Writing statistical results output, 006 samples...
```

```
##
## ...done
##
## Generating final diffreport and EICs...
##
## Loading required package: multtest
## 05blank 09blank 13blank 20-IPAHG3 18-IPAHG1 19-IPAHG2
## finished...
```

simExTargId will wait until at least five minutes after the raw data file was last modified, to ensure that the file acquisition has completed.

After at least 3 samples of each class found in the second column of the co-variates table, retention time alignment, grouping, zero-filling, then pre-processing, PCA analysis, stats analysis and data-deconvolution will occur.

During a run the output of the statistical analyses can be viewed using the shiny application a zip file containing, a seperate .csv file for each sample to reinject can be downloaded and used to guide further MS/MS experimentation.

```
# this command will open the application in your web-browser
targIdShiny(analysisDir=paste(dummyAnalysisDir, "output/04.stats", sep="/"))
```