MSPLIT-DIA package user manual:

MSPLIT-DIA is a spectral library search tool that identifies peptides from multiplexed MS/MS spectra in DIA data (e.g. SWATH).

1) Creating library from DDA runs

1.1 Creating library from msgfdb search results

MSPLIT creates spectral library from DDA runs based on MSGFDB search results. It is very likely one would build spectral library from multiple DDA runs. Therefore we need to first process the raw database search results from multiple DDA data. Remember to have the "-tda " option on when running MSGFDB search, this will allow the estimation of false discovery rate (FDR) using the Target/Decoy Approach (TDA). First put all the search result files in a folder and run the following command to generate a combined result list. This script picks a representative PSM (currently the one with lowest msgf probability) for each unique peptide precursor (i.e. peptide sequence and charge state combination) and ensure the overall peptide-level FDR for the combined search results is low. Note that we cannot simply accept all search results since the number of false ID in each individual search maybe low, in the combined result the number of false ID will scale as the number of MS/MS data and you combine .

java -cp MSPLIT-DIAvXXX.jar org.Spectrums.TDAStat MSGFDB-Dir <search result file folder> <merged result file> <fdr>

Note in the command, the java jar package usually has some version label (e.g "MSPLIT-DIAv08122014.jar"), this will changes with time, so just substitute "XXX" with whatever the version number the newest software happen to be. Text in "< >" brackets are parameters that need to be passed to the program. For fdr value please use fractional value (i.e. 0.01 stands for 1% FDR).

Next, to create spectral library, put all the relevant DDA mzXML file in the same folder. Then run the following command:

java -cp MSPLIT-DIAvXXX.jar UI.CreateSpectralLibrary <spectrum file directory> <msgfdb combined result file> <output library file>

This creates a library in annotated mgf format, which can be used to search SWATH or DIA data. An example spectrum file is also include in the package, one can creates library base on other search tools and create a spectral library based on the mgf format (http://www.matrixscience.com/help/data\_file\_help.html). It is relatively straightforward, an example is shown below:

BEGIN IONS

TITLE=Scan Number: 12826 Retention Time: PT2401.3S PROTEIN:UPS\_sp|P04040|CATA\_HUMAN\_Catalase\_OS=

CHARGE=2+

PEPMASS=907.3960571289062

SEQ=+42.011ADSRDPASDQMQHWK

SCAN=1

159.088 50.36

227.116 80.00

316.166 64.06

333.192 123.69

... (note: more listing of peaks go here)

END IONS

The important field to have is SEQ=+42.011ADSRDPASDQMQHWK, which tells MSPLIT the peptide for the spectral library entry.

1.2 Creating decoy spectral library

In order to estimate FDR, before using the library first create a decoy version of the library using the following command:

java -cp MSPLIT-DIAvXXX.jar org.Spectrums.DecoySpectrumGenerator <library file> <output file name> <fragment mass tolerance in Da>

This command will appends a decoy version of the spectral library to the original library and also perform some noise filtering of the target spectrum based on the fragment ion annotation from the identified peptides.

2) Searching DIA data using MSPLIT-DIA

2.1 Performing spectral library search

To search a swath file using the library created above, run the following command:

This will return a result file that contains all the spectrum-spectrum-matches (SSM) to each multiplexed spectra in SWATH with a minimum cosine similarity (e.g 0.7).

java -Xmx2500M -cp MSPLIT-DIAvXXX.jar org.Spectrums.SWATHMSPLITSearch <parent mass tolerance in Da> <fragment mass tolerance in ppm> <number of scans in one cycle> <SWATH file> <library file> <outfile>

Note parent mass tolerance should be specified in Da and fragment mass tolerance should be specified in PPM. The third parameter is the number of MS scans per cycle of DIA run (e.g. in SWATH, 1 MS1 is followed by 34 MS/MS to cover a mass range of 400 -- 1250, thus the number of scan in one cycle is 35). If you are unsure about the cycle, you can also input “0” and MSPLIT-DIA will try to deduce this parameter from the given spectrum file.

2.2 Filtering search results

Finally, the results are filtered using a Target/Decoy Approach by enforcing a peptide level false discovery rate (FDR). Use the following command:

java -cp MSPLIT-DIAv12102014.jar UI.SWATHFilter –r <search result file> -o <output file> -fdr <fdr>

The full command option list are as followed:

usage: SWATHFilter

 -fdr <fdr> fdr for filtering result

 -help help

 -maxRT <maxRT> max RT used in build RT corrlation

 -minRT <minRT> min RT used to build RT correlation

 -o <filtered output file> output for filtered result

 -r <resultFile> search result file

 -rt < 0 | 1 > use Reteintion time to filter result

This will compute an peptide-level FDR for each SSM in the search result and any results with fdr greater than the specify the fdr value is filter out. Two example outputs are include in the package for user to see what the output looks like. FDR is a fractional value (i.e. 0.01 stands for 1% FDR).

2.2.1 Using retention-time to filter results

When spectral library has accurate retention-time (RT) available, MSPLIT-DIA can additionally use them filter out false positive IDs and thus improve sensitivity of peptide identification at the same fdr level. First of all, the RT should be encoded in the mgf file of the spectral library, an example is shown below:

BEGIN IONS

TITLE=Scan Number: 12826 Retention Time: PT2401.3S PROTEIN:UPS\_sp|P04040|CATA\_HUMAN\_Catalase\_OS=

CHARGE=2+

PEPMASS=907.3960571289062

SEQ=+42.011ADSRDPASDQMQHWK

RTINSECONDS=1230.233333333333334

SCAN=1

159.088 50.36

227.116 80.00

316.166 64.06

333.192 123.69

... (note: more listing of peaks go here)

END IONS

 Notice the highlighted field “RTINSECONDS=” which specify the RT for the peptide.

To utilize the RT information run the “SWATHFilter” command with “-rt” option on:

java -cp MSPLIT-DIAv12102014.jar UI.SWATHFilter –r <search result file> -o <output file> -fdr <fdr> -rt 1

MSPLIT-DIA will generate three results files, the file the most care about is the one that end in suffix “rtfiltered\_fdrfiltered.txt”, which contains the list of identified peptides that are both below the specified fdr threshold and have RT that are at expected location.

2.3 Performing spectral library search with PTMs

MSPLTI-DIA allow for the detection of PTMs by predicting the library spectra of modified peptides from those of unmodified peptides. The list of PTMs can be specified in a file. An example of searching for methionine oxidation is shown below:

#comments here, do not consider by MSPLIT-DIA

#

#

maxPTM,1

15.995,M,any,opt,Oxidation on Methionine

The format to specify a PTM is similar to InsPecT or MSGFDB, each line describes a type of PTM as follow:

mass,residues,positions,type,name/labels/comments

mass - mass of the modification

residues - residues the modification can go on; \* means any residues

position - positions of the modification: N-term, C-term, any, or a position

type - optional

name/labels/comments - description of the mods

The first line set the maximum numbers of mods per peptide, currently MSPLIT-DIA only support single mod, so please do not change the first line. An example of this modification file name “Mod\_MSPLIT-DIA.txt” are included in the package.

The modification file can be input to MSPLIT-DIA as an additional argument:

java -Xmx2500M -cp MSPLIT-DIAvXXX.jar org.Spectrums.SWATHMSPLITSearch <parent mass tolerance in Da> <fragment mass tolerance in ppm> <number of scans in one cycle> <SWATH file> <library file> <outfile> <PTMs list file>

3) Generating input library for quantification analysis

3.1 Intensity-based quantification

After obtaining a search result from SWATH file, the identified peptides can be used to generate a sample-specific assay library which can be used as input to targeted tools (e.g. PeakView/Skyline/OpenSWATH) for quantification. Run the command as follow:

For peakview:

java -Xmx2000M -cp MSPLIT-DIAvXXX.jar UI.SWATHQuant PeakViewInput

<library file> <mpslit search result> <fasta file> <output file>

For openswath:

java -Xmx2000M -cp MSPLIT-DIAvXXX.jar UI.SWATHQuant OpenSwath

<library file> <mpslit search result> <fasta file> <output file>

For skyline:

java -Xmx2000M -cp MSPLIT-DIAvXXX.jar UI.SWATHQuant Skyline

<library file> <mpslit search result> <fasta file> <output file>

3.2 Spectral counting

MSPLIT-DIA can generate quantification information for spectral-count based methods (e.g. SAINT analysis for detection of protein-protein interactions). To do this run the following command:

java -Xmx1000M -cp MSPLIT-DIAvXXX.jar UI.SWATHQuant SAINTInput

<mpslit search result> <fasta file> <experiment name> <Bait name> <output file>

Bait name is the name of the bait protein and experiment name is a text label for the particular experiments, they can be any strings.