
DO-MS-Processing

Release 1.0

Georg Wallmann

Apr 19, 2022

CONTENTS

- 1 Command Line Usage** **1**
 - 1.1 Positional Arguments 1
 - 1.2 options 1
- 2 Integration with Search Pipelines** **3**

COMMAND LINE USAGE

Command line tool for feature detection in shotgun MS experiments. Can be used together with DIA-NN to provide additional information on the peptide like features identified in the MS1 spectra.

```
usage: FeatureDetection [-h] --raw-parser-location RAW_PARSER_LOCATION
                        [--dinosaur-location DINOSAUR_LOCATION] [-m] [-d] [-v]
                        [-t TEMPORARY_FOLDER] [-r RAW_FILE_LOCATION]
                        [--no-feature-detection] [--no-fill-times] [--no-tic]
                        [--no-sn] [--no-mzml-generation]
                        [--mz-bin-size MZ_BIN_SIZE]
                        [--rt-bin-size RT_BIN_SIZE] [--resolution RESOLUTION]
                        [-p PROCESSES] [--isotopes-sn]
                        report
```

1.1 Positional Arguments

report Path pointing to report.tsv output from DIA-NN.
Default: /home/runner/work/DO-MS-DIA/DO-MS-DIA/docs

1.2 options

--raw-parser-location Path pointing to the ThermoRawFileParser executable.

--dinosaur-location Path pointing to the dinosaur jar executable.

-m, --mono Use mono for ThermoRawFileParser under Linux and OSX.
Default: False

-d, --delete Delete generated mzML and copied raw files after successfull feature generation.
Default: False

-v, --verbose Show verbose output.
Default: False

-t, --temporary-folder Input Raw files will be temporarilly copied to this folder. Required for use with Google drive.

- r, --raw-file-location** By default, raw files are loaded based on the File.Name column in the report.tsv. With this option, a different folder can be specified.
Default: ""
- no-feature-detection** All steps are performed as usual but Dinosaur feature detection is skipped. No features.tsv file will be generated.
Default: False
- no-fill-times** All steps are performed as usual but fill times are not extracted. No fill_times.tsv file will be generated.
Default: False
- no-tic** All steps are performed as usual but binned TIC is not extracted. No tic.tsv file will be generated.
Default: False
- no-sn** Signal to Noise ratio is not estimated for precursors
Default: False
- no-mzml-generation** Raw files are not converted to .mzML. Nevertheless, mzML files are expected in their theoretical output location and loaded. Should be only be carefully used for repeated calculations or debugging
Default: False
- mz-bin-size** Bin size over the mz dimension for TIC binning.
Default: 10.0
- rt-bin-size** Bin size over the RT dimension for TIC binning in minutes. If a bin size of 0 is provided, binning will not be applied and TIC is given per scan.
Default: 1
- resolution** Set the resolution used for estimating counts from S/N data
Default: 70000
- p, --processes** Number of Processes
Default: 10
- isotopes-sn** Use all isotopes from the same scan as the highest intensity datapoint for estimating the SN and copy number.
Default: False

INTEGRATION WITH SEARCH PIPELINES