NAME

ionFinder - perform high a throughput search for fragment ions in ms2 spectra.

SYNOPSIS

ionFinder [options] [input_dir ...]

ionFinder [options] --inputMode tsv <input_file_path> [...]

DESCRIPTION

ionFinder Reads peptides from one or more DTASelect-filter files, calculates theoretical B and Y peptide fragments, and searches parent MS-2 scans for theoretical fragments. If no argument is specified for *input_dir*, the current working directory is used.

The envisioned use case of **ionFinder** is to search for neutral loss ions which are not considered by database searching software. The program can also be used to automatically generate annotated MS-2 spectra for an entire mass spec run. Options are available to add neutral loss fragments to the search, and to print annotated MS-2 spectra for each peptide.

OPTIONS

Command line options are processed from left to right. Options can be specified more than once. If conflicting options are specified, later specifications override earlier ones.

ION SEAECH SETTINGS

-mt, --matchTolerance *<tolerance>*

Tolerance in Thomson (m/z units) or ppm for predicted fragment ions in .ms2 files. By default *<tolerance>* is interpreted as Th. To change the units of *<tolerance>*, use the **--matchTyep** argument. 10 ppm is the default.

--matchType

Specify how the **--matchTolerance** argument is interpreted. **ppm** is the default.

mz, th

Interpret **--matchTolerance** as Th or m/z units.

ppm

Interpret **--matchTolerance** as ppm.

--citStats

Search ms2s with default settings for citrulline. Specifically **--citStats** sets the following flags:

--modMass 0.984 **--lossMass** 43.0058 **--calcNL** 1 **--isoAA** NQ **--cTermMod** 0 **-o** peptide_cit_stats.tsv

Later specifications of the previous flags will override those set by **--citStats**.

--modMass *<mass>*

Specify mass of '*' on modified peptides.

--isoAA *<amino_acids>*

Specify amino acids whose mass is ambiguous with the amino acids with the dynamic modification. *<amino_acids>* should be single letter amino acid codes with no space between them.

--lossMass *<mass>*

Specify mass of neutral loss to search for.

--cTermMod *<0/1>*

Specify whether to allow c terminally modified peptides. Default is **1**.

0
Do not allow c terminally modified peptides.

1
Allow c terminally modified peptides.

--calcNL *<0/1>*
Specify whether neutral loss fragment ions should be search for. If this option is set to **1** without any argument for **--lossMass** given, it is ignored. Default is **0**.

0
Do not search for neutral loss ions.

1
Search for neutral loss fragment ions.

-minC *<charge>*
Minimum charge state to be considered. 1 is the default.

-maxC *<charge>*
Maximum charge state to be considered. 1 is the default.

-mmComp *<compare_method>*
When multiple candidate fragment ions are found within **--matchTolerance**, how should a tie be broken? Default is **intensity**.

intensity
The most intense ion will be used.

ms
The ion with the closest mz to the predicted mz for the fragment will be used.

PEPTIDE FILTERING OPTIONS

-rev *<0/1>*
Specify whether to include reverse matches in spectrum output file. **0** is the default.

0
Only include forward matches in spectrum output file.

1
Include all peptides in spectrum output file.

-m, --modFilter *<filter>*
Specify how to include modified peptides in output. **1** is the default.

0
Require peptides to be modified.

1
Include peptides regardless of modification.

2
Exclude modified peptides.

MS2 FILTERING OPTIONS

-minMZ *<mz>*
Minimum ion MZ to be considered from *.ms2* files. By default all identified ions are considered.

-minMZ *<mz>*
Maximum ion MZ to be considered from *.ms2* files. By default all identified ions are considered.

-minInt *<relative_intensity>*
Minimum relative intensity to include from *.ms2* files. Ion intensities are normalized to 100 before labeling, so **-minInt** should be supplied as a relative intensity. If this option is set, an intensity

filter will be applied to spectra before annotation. By default, all ion intensities are included.

-n, --artifactNLIntPerc *<percentage>*
Percentage of ion intensity allowed for artifact NL ions. The intensity cutoff for neutral loss ions ions is dynamically set for each spectrum such that the total ion intensities for all neutral loss ions is no more than *n* percent from artifact neutral loss ions. Argument should be supplied as a percentage. Default is 1.0 percent.

-minLabInt *<relative_intensity>*
Minimum relative intensity to include from *.ms2* files. If this option is set, only ion with relative intensities above *<relative_intensity>* will be labeled. By default, all ions are considered in labeling regardless of intensity.

-minNlLabInt *<relative_intensity>*
Same as **-minLabInt**, but only applies to neutral loss ions.

-minSNR *<snr>*
Minimum signal to noise ratio (SNR) to include from *.ms2* files. If this option is set, only ion with SNRs above *<snr>* will be labeled. By default, all ions are considered in labeling regardless of SNR.

--snrConf *<conf>*
Confidence interval to use when estimating signal and noise threshold as a fraction of 1. Default is 0.9.

--incAllIons *<0/1>*
Specify whether to include unlabeled ions in spectrum output file. **1** is the default.

0
Only include labeled fragment ions in spectrum output file.

1
Include all ions in spectrum output file.

MS2 PLOTTING OPTIONS

-p, --printSpectra
Print *.spectrum* files for each peptide analyzed?

-y, --plotHeight *<height>*
Specify ms2 plot height in inches to calculate label positions for in *.spectrum* output files. Default is **4** inches.

-w, --plotWidth *<height>*
Specify ms2 plot width in inches to calculate label positions for in *.spectrum* output files. Default is **12** inches.

--labelArtifactNL *<filter>*
Specify whether to include artifact neutral loss ions in ms2 plots. **0** is the default.

0
Do not include artifact NL ions.

1
Include all theoretical NL ions.

INPUT / OUTPUT OPTIONS

-i, --inputMode *<input_mode>*
Specify how peptides to search for in .ms2 file will be supplied. Default is **dtafilter**.

dtafilter
Use DTAFilter-file(s) as input.

tsv

Supply peptide list as .tsv formatted file. Required columns are: "sampleName", "sequence", "precursorFile", "scanNum".

Optional columns are: "formula", "parentId", "parentProtein", "parentDescription", "matchDirection", "fullSequence", "unique", "charge", "score", "precursorMZ", "precursorScan".

-d, --dir *<path>*

Set working directory from which to run program. By default working directory at runtime is used.

-o, --ofname *<ofname>*

Set name of summary output file. By default, *<ofname>* is *peptide_cit_stats.tsv*.

-dta *<name>*

Set default name of DTASelect-filter files. By default, *<name>* is *DTASelect-filter.txt*.

OTHER

--fastaFile *<path>*

Specify .fasta formatted file to lookup numbers of modified residues in *peptide_cit_stats.tsv*.

-I, --printInt*<0/1>*

Should peptide fragment ion intensities be included in tsv output? **0** is the default.

0

Do not include fragment intensity columns.

1

Include fragment intensity columns.

-u, --peptideUID *<0/1>*

Specify whether to include the "peptide_unique_ID" column in the output file. **0** is the default.

0

Do not include the "peptide_unique_ID" column.

1

Include the "peptide_unique_ID" column.'

-g, --groupMod *<group_method>*

Choose how to group peptides with multiple modifications. **1** is the default.

0

Do not group modifications. The output file will have a separate line for each modification site in a peptide.

1

Group modifications onto a single line for each peptide.

--parallel

The part of **ionFinder** which searches .ms2 files for fragment ions is written to run concurrently on multiple threads. By default only a single thread is used. If this option is set, the number of threads returned by std::thread::hardware_concurrency() are used.

--nThread *<n_thread>*

Manually set the number of threads to use.

-v, --version

Print binary version number and exit program.

-h, --help

Display this help file.

PROGRAM OUTLINE

ionFinder has 3 phase. See *ionFinder_userGuide.pdf* for a detailed description of each phase.

1) INPUT

- All *input_dir* folders are searched for DTASelect-filter files each peptide is read into a data structure in memory.
- Get peptide modification parameters from existing *sequest.params* files.

2) ANALYSIS

- Calculate b, y and neutral loss ions for each peptide.
- Load the corresponding *.ms2* file into a buffer in memory
- Find the location of the scan in the *.ms2* file buffer and parse it into a *ms2::Spectrum* object
- Search the *ms2::Spectrum* object for fragment ions within a user defined tolerance.
- Classify the fragment ions which were found.
- Determine whether fragments found unambiguously show that the peptide contains the modification.

3) OUTPUT

- Write out *peptide_cit_stats.tsv* with a summary of the results.

EXAMPLES

ionFinder

Run ionFinder from current working directory using default parameters.

ionFinder --citStats

Run ionFinder from current working directory using default parameters for citrulline.

ionFinder --modMass 79.9799 --lossMass 97.9770 --calcNL 1

Run ionFinder from current working directory, searching for a neutral loss of 97.9770 on residues with a residues with a modification (indicated by a '*') with a mass of 79.9799.

ionFinder --citStats --printSpectra

Run ionFinder from current working directory using default parameters for citrulline and printing intermediate *.spectra* files.

AUTHOR

ionFinder was written by Aaron Maurais. Email questions or bugs to: aaron.maurais@.bc.edu