

AuDIT Documentation

Description: Automated detection of inaccurate and imprecise transitions in

Multiple Reaction Monitoring Mass Spectrometry (MRM-MS)

Author: D. R. Mani, manidr@broadinstitute.org

Summary

Multiple reaction monitoring-mass spectrometry (MRM-MS) of peptides with stable isotope-labeled internal standards (SIS) is a quantitative assay for measuring proteins in complex biological matrices. These assays can be highly precise and quantitative, but the frequent occurrence of interferences require that MRM-MS data be manually reviewed by an expert, a time intensive process that is subject to human error. The AuDIT module implements an algorithm that, in an automated manner, identifies inaccurate transition data based on the presence of interfering signal or inconsistent recovery between replicate samples.

The algorithm objectively evaluates MRM-MS data using two orthogonal approaches. First, it compares the relative product ion intensities of the analyte peptide to those of the SIS peptide and uses a t-test (in conjunction with a p-value threshold) to determine if they are significantly different. Then, coefficient of variation is calculated from the ratio of analyte and SIS peak areas from sample replicates, with those transitions having excessive variation flagged as being unsuitable.

The algorithm for Automated Detection of Inaccurate and imprecise Transitions (AuDIT) in SID-MRM-MS data greatly reduces the time required for manual, subjective inspection of data, improves the overall accuracy of data analysis, and is easily implemented into the standard data analysis workflow. AuDIT currently works with exported results from MRM-MS data processing software packages.

References

Susan E Abbatiello, D. R. Mani, Hasmik Keshishian, Steven A. Carr. Automated Detection of Inaccurate and Imprecise Transitions in Quantitative Assays of Peptides by Multiple Reaction Monitoring Mass Spectrometry. *Clinical Chemistry* 56:2 (2010).



Parameters

Name	Description
data file	Input file with pre-processed MRM data in csv format. Contents of this file are described below.
skyline export	If set to yes, input data file is assumed to be a Skyline export, and is preprocessed to automatically create a data file with appropriate input format for AuDIT.
	If set to no, input data file must contain the specified (see below) columns in the data set in the same order, with the indicated column names.
pvalue threshold	A threshold for the multiple hypothesis corrected t- test p-value. When the corrected p-value is greater than this threshold, the MRM-MS transition is deemed to be free of interference and is therefore a valid transition to use for quantitation. If the p-value is less than this threshold (i.e., the t-test result is significant), then the transition has some form of interference or imprecision.
cv threshold	A threshold for the coefficient of variation for the replicates of a transition for a specific peptide precursor. If the actual CV of the replicates is larger than this CVThreshold, the transition is too variable, it is considered unsuitable for quantitation.
all pairs	If this option is set to yes, all possible pairs of relative ratios for the transitions of a given peptide are calculated and used in determining problematic transitions. This is useful only when the MRM-MS runs have more than 3 transitions per peptide. See reference for details.
output intermediate results	When this option is TRUE, several intermediate results files are generated before the final results file. This can aid in understanding the steps executed by the algorithm or to help with debugging.
output prefix	A single .csv results file is created using the outputFilePrefix as the filename prefix. When intermediateResults = TRUE, then several



intermediate files are also created using this as the filename prefix. The contents of the output results file is described below.

Input File

The input file specified by *data file* is a comma delimited (csv) file that is derived from software used to pre-process raw MRM-MS data. The pre-processing software could be vendor provided (e.g., MultiQuant from Applied Biosystems) or universal (e.g., Skyline from the MacCoss Lab, University of Washington).

When the "skyline export" parameter is set to "no", AuDIT is agnostic to the pre-processing software used, as long as the following columns are present in the input file, in the specified order:

- Sample: The actual sample ID (excluding replicate notation). This is
 usually derived from the SampleName column output by MRM processing
 software. Sample must be unique for different concentrations (if any), and
 must be the same for all replicates of that sample. In other words, for a
 given peptide and transition, the value in the Sample column must be
 identical for all the replicates.
- 2. Replicate: Replicate number for the Sample.
- 3. *Peptide:* Peptide name and/or sequence for the peptide that is being monitored.
- 4. *Transition.ID:* An indication of the transition being monitored. This may be a number or some other notation (e.g., b- or y-fragment number with charge state). While different peptides may have the same transition.id, these must be unique for a given peptide.
- 5. Area: Integrated peak area for the analyte for the specified peptide and transition.
- 6. *IS.Area:* Integrated peak area for the SIS for the specified peptide and transition.

The input file must contain a column header and the names used in the column header must appear as listed above.

When "skyline export" is set to "yes", AuDIT assumes that the column naming used is Skyline specific. The following columns must be present:

PeptideSequence, ReplicateName, Fragmentlon, PrecursorCharge,

ProductCharge, light Area, heavy Area. In addition, the Sample column must be derived from the Skyline SampleName column, to satisfy the conditions stated above (for skyline export = "no"). When the input data set is treated as a skyline export, the columns can be in any order (and interspersed with other



columns), as long as they are present in the data set (with those exact names). When using N15, or other variants of reference standards, the appropriate light or heavy area column may need to be renamed.

Notes:

- AuDIT requires both an analyte and an internal standard (IS) to be measured – i.e., both Area and IS.Area must be present, and have legitimate values.
- Rows in the input table that represent only an IS transition must be removed. Data from both analyte and IS transitions must be integrated into a single row in the input file.
- The Sample field should not contain the transition id. This will result in errors since a given "sample" will then have only one transition.
- AuDIT requires that peptides have 3 or more transitions.
- For a given sample and peptide replicate, transition ids must not be replicated. Thus, for a (sample, peptide, replicate) combination, if n transitions are recorded, the resulting n Transition.IDs must be unique. For skyline exports, this will be automatically handled by AuDIT, using the Fragmentlon, PrecursorCharge and ProductCharge columns.

Output File

The output file generated on successful completion of AuDIT has the following columns:

- 1. peptide The Peptide column from the input data.
- 2. sample The Sample column from the input data.
- 3. *transition.id* The Transition.ID column from input data. For skyline exports, this will be a concatenation of the Fragmentlon and PrecursorCharge columns.
- 4. *pvalue.final* The multiple testing corrected t-test p-value for the transition under consideration.
- 5. *status* This column is the result of applying the p-value threshold to pvalue.final, and is 'good' if pvalue.final > p-value threshold; the transition is marked 'bad' otherwise
- 6. *cv* The calculated coefficient of variation for the replicates of this peptide/transition.
- 7. *cv.status* Whether the CV is less than the CV threshold; 'good' if CV is less than threshold, 'bad' otherwise.
- 8. *final.call* The final decision on whether the transition under consideration is imprecise or has interferences. The final.call is 'good' if both status and cv.status are 'good'. If either status or cv.status is 'bad', final.call is 'bad'.



Example Data

Example input data from MultiQuant processed data is available in site56-7.1-multiquant-data.csv. Data from Skyline pre-processing is in site56-7.1-skyline-data.csv. Results from running AuDIT on these datasets are available in site56-7.1-skyline-data-result.csv. The Skyline example data has been preprocessed to generic format, and should be invoked with skyline export set to "no".

Platform Dependencies

Module type: Proteomics

CPU type: Any OS: Any Language: R