

QuaSAR Documentation

Description: A Comprehensive Pipeline for Quantitative and Statistical Analysis of Reaction Monitoring Results

Author: Rushdy Ahmad, rahmad@broadinstitute.org
D. R. Mani, manidr@broadinstitute.org

Summary

QuaSAR implements a comprehensive, easy to use pipeline for quantitative MRM-MS data analysis. For every monitored peptide the module provides statistical measures such as, coefficient of variation (CV), regression slope and intercept (with confidence intervals), interference detection, and limits of detection (LOD) and quantification (LOQ). It also provides succinct visual summaries of various results including reproducibility, interferences and detection limits. Statistics and visualization are integrated to enable effective data interpretation, understanding and insight.

Parameters

Name	Description
skyline file	This is a Skyline export in csv file format. The file has to include the following 10 columns with the exact column names given here (in no particular order): SampleName, PeptideSequence, ProteinName, ReplicateName, PrecursorCharge, ProductCharge, FragmentIon, AverageMeasuredRetentionTime, light Area, heavy Area. The file may contain many more columns, but it has to include at least these 10 columns.
concentration file	Concentration values given in csv file format. The file has to include the following 5 columns with the exact column names given here (in no particular order): <i>SampleName</i> , <i>SampleID</i> , <i>Analyte Concentration</i> , <i>IS Conc</i> , <i>Concentration Ratio</i> . Notice the single space between "Analyte Concentration", "IS Conc" and "Concentration Ratio". The space has to be maintained.
title	The title to be displayed on each calibration plot.
analyte	User must specify the exact analyte column header name as it is appears in the input CSV file. The default is "light Area".

GenePattern

standard	User must specify the exact internal standard column header name as it appears in the input CSV file. The default is "heavy Area".
generate cv table	If "yes" then generate CV (coefficient of variation) table.
generate calcurves	If "yes" then generate calibration curves.
generate lodloq table	If "yes" then generate LOD/LOQ table.
generate peak area plots	If "yes" then generate peak area plots with Peak Area units on the y-axis and analyte concentration on the x-axis.
use par	If "yes" then use peak area ratio (PAR) for analysis (instead of concentration) with PAR on the y-axis and analyte concentration on the x-axis.
max calcurve linear scale	The maximum value for linear scale in fmols/ul.
max calcurve log scale	The maximum value for log scale in fmols/ul.
perform audit	If "yes" then perform AuDIT for interference detection.
audit cv threshold	For AuDIT the threshold for coefficient of variation below which transition is quantification-worthy. Default is 0.2.
output prefix	Prefix used for all output file names.

Input Files

- A) The input file specified by "**skyline 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 AB Sciex) or universal (e.g., Skyline from the MacCoss Lab, University of Washington). QuaSAR is agnostic to the pre-processing software used, as long as the following columns are present in the input file, in no particular ordering of the columns.
1. *SampleName*: The actual sample ID (excluding replicate notation). This is usually derived from the SampleName column output by MRM processing software. *SampleName* must be unique for different

GenePattern

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 *SampleName* column must be identical for all the replicates.

2. *ProteinName*: Protein name for the peptide that is being monitored.
3. *ReplicateName*: Replicate name for the Sample. Maybe an extension of the *SampleName*, and is unique for each replicate (ie, A1, A2, A3 are replicates of sample A).
4. *PeptideSequence*: Peptide name and/or sequence for the peptide that is being monitored.
5. *PrecursorCharge*: The charge of the precursor.
6. *ProductCharge*: The charge of the product.
7. *FragmentIon*: The transition that is being monitored (i.e. y8).
8. *AverageMeasuredRetentionTime*: The measured average retention time
9. *analyte*: Integrated peak area for the analyte for the specified peptide and transition. The default is “light Area”. Please notice the space between “light” and “Area”. The space needs to be maintained.
10. *standard*: Integrated peak area for the SIS for the specified peptide and transition. The default is “heavy Area”. Please notice the space between “heavy” and “Area”. The space needs to be maintained.

Notes:

- QuaSAR requires both an analyte and an internal standard (IS) to be measured – i.e., both *light Area* and *heavy Area* must be present, and have legitimate values. Heavy Area can be replaced with a different column header, as long as this is specified on the input form in GenePattern. For example, “15N Area” might be an additional form of the peptides in the document. To use these values for the internal standard, “15N Area” should be specified and match the exact column header containing that data.
 - 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 *SampleName* field should not contain the *FragmentIon*. This will result in errors since a given “sample” will then have only one transition.
 - If using AuDIT within QuaSAR then AuDIT requires that peptides have 3 or more transitions.
 - For a given sample and peptide replicate, *FragmentIon* 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 QuaSAR, using the *FragmentIon*, *PrecursorCharge* and *ProductCharge* columns.
- B) The input file specified by “**concentration file**” is a comma delimited (csv) file that needs to be constructed and provided by the user. The file must contain the following five columns with these exact column names in no particular ordering of the columns.

GenePattern

1. *SampleName*: This is identical to what is provided in the “skyline file”. This is usually derived from the SampleName column output by MRM processing software.
2. *SampleID*: This identifier groups samples together in particular groups of interest. For example, process replicates of a given sample can be defined by a root sample name (i.e. “patient0102”), and when defined by this root name will be grouped together.
3. *Analyte Concentration*: The concentration of the analyte. Notice the space in “Analyte Concentration”. If the concentration is unknown then the field needs to be left empty or blank. The unit is in fmols/ul.
4. *IS Conc*: The concentration of the internal standard. Notice the space in “IS Conc”. The internal standard concentration need not be uniform for a given experiment, but must be the same for all peptides in the input file. The unit is in fmols/ul.
5. *Concentration Ratio*: The ratio of analyte and internal standard concentration. The values in this column may be left empty. Notice the space in “Concentration Ratio”.

Output Files

If all options that generate an output are selected then upon successful completion QuaSAR will generate the following output files:

1. Filename with suffix: **-parameters-summary.csv**
A CSV file containing all parameters used by user to generate results
2. Filename with suffix: **-calcurves.pdf**
A PDF file with the calibration plots ordered by peptide sequence
3. Filename with suffix: **-calcurves.csv**
A CSV file with associated calibration plot information
4. Filename with suffix: **-calcurves-peakareas.pdf**
A PDF file with the calibration plots using peak areas instead of concentrations and ordered by peptide sequence
5. Filename with suffix: **-calcurves-peakareas.csv**
A CSV file with associated calibration plot information using peak areas instead of concentrations
6. Filename with suffix: **-by_protein_calcurves.pdf**
A PDF file with the calibration plots ordered by protein name
7. Filename with suffix: **-by_proteins_calcurves-peakareas.pdf**
A PDF file with the calibration plots using peak areas instead of concentrations and ordered by protein name
8. Filename with suffix: **-cvplot.pdf**
A PDF file showing overall assay reproducibility by plotting CV vs. theoretical concentration values.
9. Filename with suffix: **-lodboxplot.pdf**
A PDF file showing distribution of LOD values with a box-whiskers plot.
10. Filename with suffix: **-lod-loq-raw.csv**

GenePattern

A CSV file with lod-loq raw values.

11. Filename with suffix: **-lod-loq-final.csv**

A CSV file with final lod-loq values, which is a subset of the raw file. The lowest LOD and LOQ values are selected for a given peptide.

12. Filename with suffix: **-cvtable.csv**.

A CSV file with a CV table for all peptide/sample/transitions.

13. Filename with suffix: **-cvtable-final.csv**.

A CSV file with the minimum CV reported in table for all peptide/sample/transitions

14. Filename with suffix: **-audit.csv**

A CSV file with the AuDIT results for all transitions.

15. Filename : **Rplots.pdf**

This is a system generated empty file and needs to be ignored by the user. In a future release of QuaSAR this file will be removed.

Directions for Preparing Input Files for QuaSAR

Files to be generated:

- Skyline export file from integrated data (*.csv format)
- Concentration map and sample grouping map (*.csv format)

To Generate Skyline Export File:

1. Obtain Skyline report template, "QuaSAR_Report_Template" from Skyline website
 - a. Save to local drive
 - b. Open Skyline and go to File/Export/Report, then click "Import"
 - c. Point to location of QuaSAR_Report_Template.skyr and click OK
 - d. Skip to Step 3
2. To generate a new report template from scratch:
 - a. In Skyline, go to File/Export/Report
 - b. Click Edit List, then Add
 - c. Select the following 10 parameters for the report:
 - i. FileName
 - ii. SampleName
 - iii. PeptideSequence
 - iv. ProteinName
 - v. ReplicateName
 - vi. PrecursorCharge
 - vii. ProductCharge
 - viii. FragmentIon
 - ix. AverageMeasuredRetentionTime
 - x. Area
 - d. Click "Pivot isotope label"
 - e. Click OK, then OK again to get back to "Export Report" window.
3. To generate a Skyline export file with the QuaSAR_Report_Template:

GenePattern

- a. File/Export/Report
- b. Select "QuaSAR_Report_Template", then select OK
- c. Save file

To Generate Concentration and Sample Grouping Map file:

1. Create an Excel Spreadsheet
2. Populate the first 5 columns with the following headers (please use exact spelling. Please note single space in "Analyte Concentration", "IS Conc" and "Concentration Ratio")
 - a. SampleName (no space)
 - b. SampleID (no space)
 - c. Analyte Concentration
 - d. IS Conc
 - e. Concentration Ratio
3. Save file as "Concentration_Map_Template" in *.csv format

To Populate Concentration and Sample Grouping Map file (MAP file):

1. Open the Skyline export file from your experiment and select the first iteration of each sample name (SampleName column) so that only one instance of each sample is selected, and copy. Skyline directly exports the results in the proper order. If the Skyline file has been sorted after export from Skyline, please re-sort using the following three criteria: PeptideSequence (Values, A to Z), FragmentIon (Values, A to Z), and SampleName (Values, A to Z). Then select the first iteration of each sample name and copy.
2. Paste into the "SampleName" column in the MAP file
3. A sample may have multiple replicates. The "SampleID" column groups these replicates together. Define a "sample ID" or grouping mechanism for your sample to delineate what replicates belong to which sample and populate the column.
 - a. You can use something simple (like, A, B, C, etc), or other unique identifier that will link together replicates of the same sample
 - b. Replicates can refer to technical replicates, process replicates or biological replicates
4. If Analyte Concentration is known (for example, as in a calibration curve), populate this column with the analyte concentration for each sample in fmol/uL values.
 - a. For samples where analyte concentration is unknown, leave it blank
5. For IS Conc, list the concentration of your internal standard in fmol/uL
 - a. QuaSAR is unable to handle multiple concentrations for IS targets at this time
6. For Concentration Ratio column, if you prefer your calculations are done using only peak area ratio and concentration ratio (for the calibration curves), you can populate this column for your calibration curves by dividing column 3 by column 4. If you don't want to use this metric, just leave the column blank. If you're unsure, just leave it blank.

GenePattern

Platform Dependencies

Module type:	Proteomics
CPU type:	Any
OS:	Any
Language:	R