HOW TO PERFORM DRIFT AND BATCH CORRECTION?

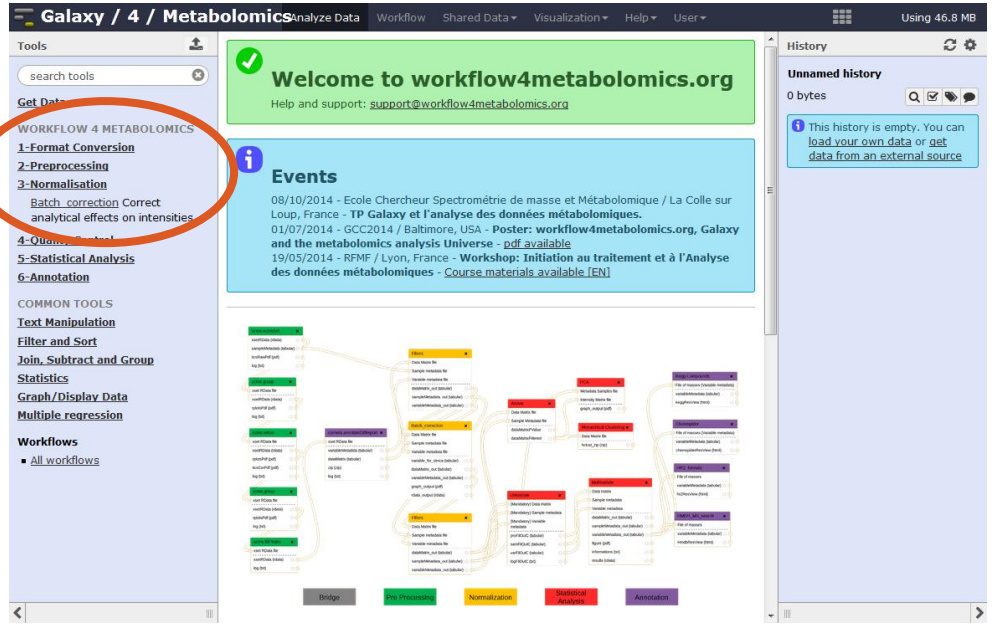
W4M Core Team

The "Batch correction" module

The Batch correction module provides correction of analytical effects inter and intra batch on intensities, using quality-control pooled samples (QC-pools), according to the algorithm mentioned by Van der Kloet (J Prot Res 2009).

This module is accessible via the left panel:

> WORKFLOW 4 METABOLOMICS
>> 3-Normalisation
>>> Batch correction



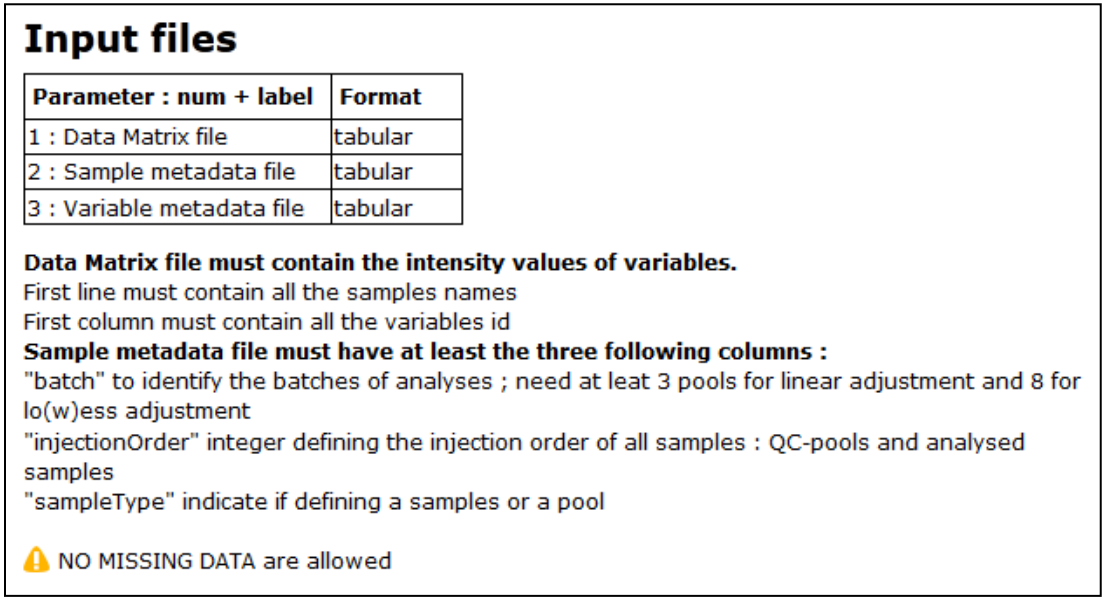
Mandatory information in your data

To run this module, your data must fit particular specifications:

- The 3 input .tsv data files (dataMatrix, sampleMetadata and variableMetadata) must be in the required W4M post-processing format (see *How to format your data for post-processing?*)
- The data matrix must not contain missing values
- The sample meta-data file must contain 3 specific columns with the following names:
  - batch
  - injectionOrder
  - sampleType

See the Help section of the module for more information.

Note: if you have only one batch, you still need to provide the « batch » column, that will contain a constant value.



Input files

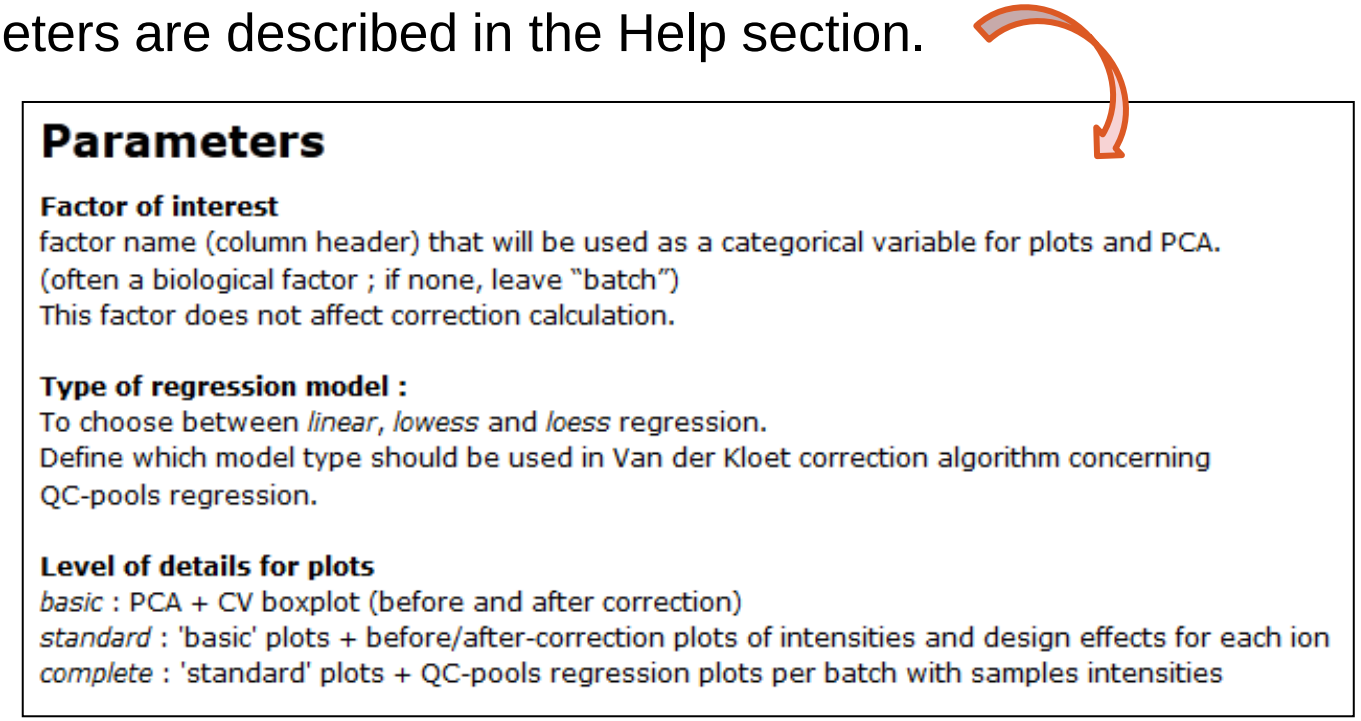
Parameter : num + label	Format
1 : Data Matrix file	tabular
2 : Sample metadata file	tabular
3 : Variable metadata file	tabular

Data Matrix file must contain the intensity values of variables.
First line must contain all the samples names
First column must contain all the variables id
Sample metadata file must have at least the three following columns :
"batch" to identify the batches of analyses ; need at leat 3 pools for linear adjustment and 8 for lo(w)ess adjustment
"injectionOrder" integer defining the injection order of all samples : QC-pools and analysed samples
"sampleType" indicate if defining a samples or a pool

⚠ NO MISSING DATA are allowed

Parameters

Parameters are described in the Help section.



Parameters

Factor of interest
factor name (column header) that will be used as a categorical variable for plots and PCA.
(often a biological factor ; if none, leave "batch")
This factor does not affect correction calculation.

Type of regression model :
To choose between *linear*, *lowess* and *loess* regression.
Define which model type should be used in Van der Kloet correction algorithm concerning QC-pools regression.

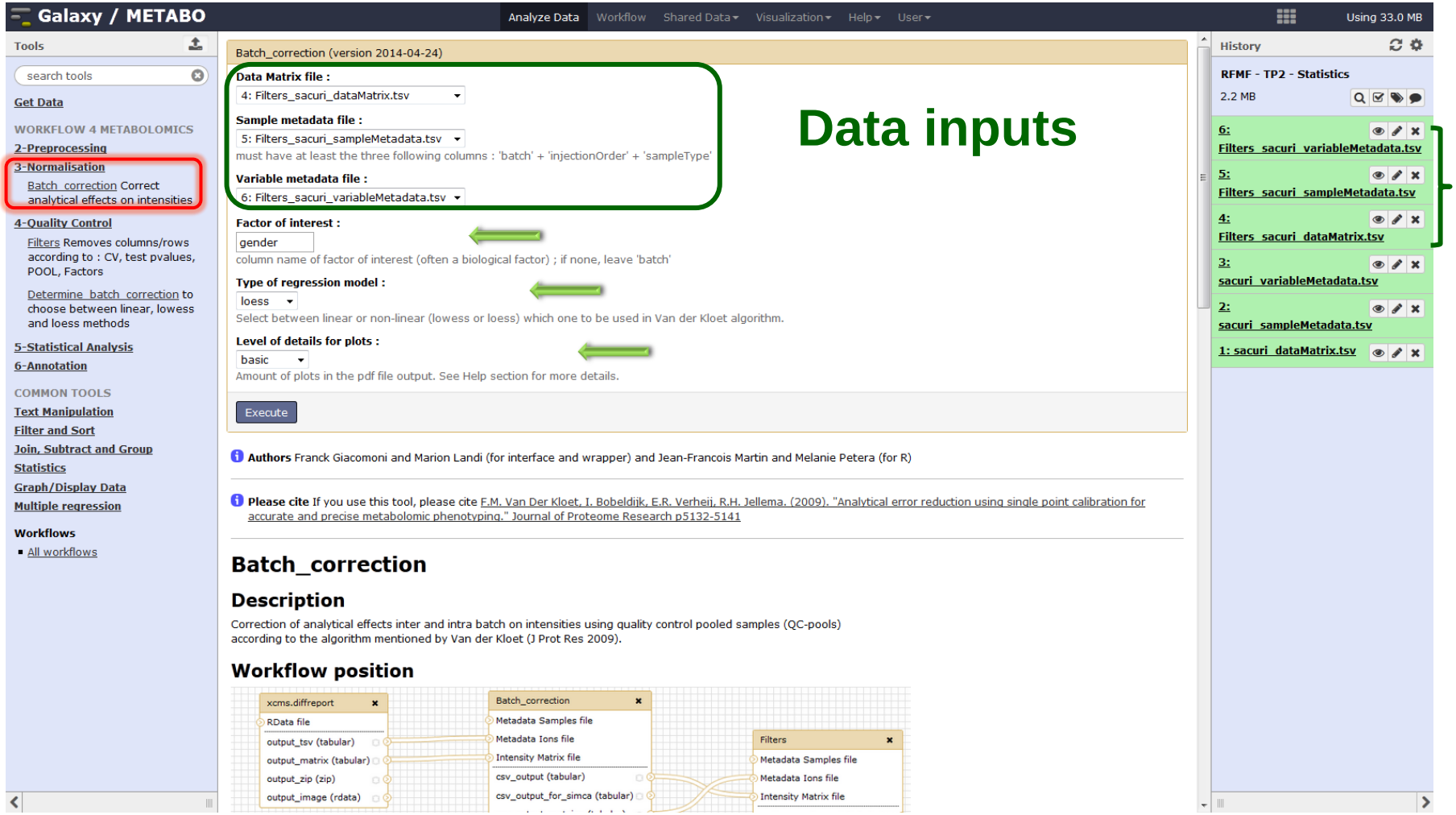
Level of details for plots
basic : PCA + CV boxplot (before and after correction)
standard : 'basic' plots + before/after-correction plots of intensities and design effects for each ion
complete : 'standard' plots + QC-pools regression plots per batch with samples intensities

Note:
The choice of the type of regression model is left to the expert assessment of the user:

- the module "Determine batch correction" flags variables for which correction is not appropriate (see slide #8)
- on an indicative basis, one may consider using « linear » when pools intensities vary linearly according to the injection order, and « loess » otherwise;

Still the user is the only judge regarding which model is the more appropriate for his data.

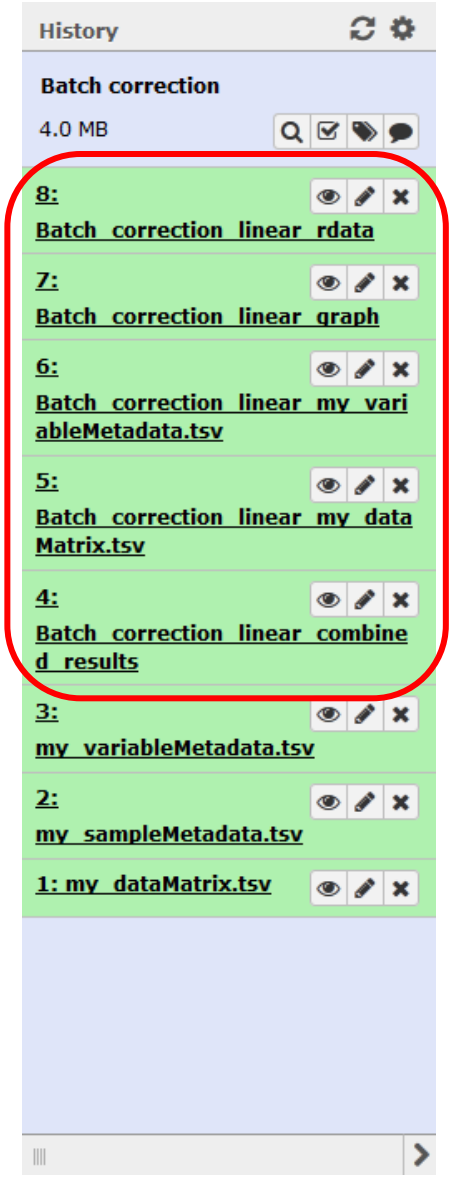
Example


Galaxy / METABO
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 33.0 MB
Tools
search tools
Get Data
WORKFLOW 4 METABOLOMICS
2-Preprocessing
3-Normalisation
Batch_correction Correct analytical effects on intensities
4-Quality Control
Filters Removes columns/rows according to : CV, test pvalues, POOL, Factors
Determine_batch_correction to choose between linear, lowess and loess methods
5-Statistical Analysis
6-Annotation
COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Statistics
Graph/Display Data
Multiple regression
Workflows
All workflows
Batch_correction (version 2014-04-24)
Data Matrix file :
4: Filters_sacuri_dataMatrix.tsv
Sample metadata file :
5: Filters_sacuri_sampleMetadata.tsv
must have at least the three following columns : 'batch' + 'injectionOrder' + 'sampleType'
Variable metadata file :
6: Filters_sacuri_variableMetadata.tsv
Data inputs
Factor of interest :
gender
column name of factor of interest (often a biological factor) ; if none, leave 'batch'
Type of regression model :
loess
Select between linear or non-linear (lowess or loess) which one to be used in Van der Kloet algorithm.
Level of details for plots :
basic
Amount of plots in the pdf file output. See Help section for more details.
Execute
Authors Franck Giacomoni and Marion Landi (for interface and wrapper) and Jean-Francois Martin and Melanie Petera (for R)
Please cite If you use this tool, please cite F.M. Van Der Kloet, I. Bobeldijk, E.R. Verheij, R.H. Jellema. (2009). "Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping." Journal of Proteome Research p5132-5141
Batch_correction
Description
Correction of analytical effects inter and intra batch on intensities using quality control pooled samples (QC-pools) according to the algorithm mentioned by Van der Kloet (J Prot Res 2009).
Workflow position
xcms.diffreport
RData file
output_tsv (tabular)
output_matrix (tabular)
output_zip (zip)
output_image (rdata)
Batch_correction
Metadata Samples file
Metadata Ions file
Intensity Matrix file
csv_output (tabular)
csv_output_for_simca (tabular)
Filters
Metadata Samples file
Metadata Ions file
Intensity Matrix file
History
RFMF - TP2 - Statistics
2.2 MB
6: Filters_sacuri_variableMetadata.tsv
5: Filters_sacuri_sampleMetadata.tsv
4: Filters_sacuri_dataMatrix.tsv
3: sacuri_variableMetadata.tsv
2: sacuri_sampleMetadata.tsv
1: sacuri_dataMatrix.tsv
Data inputs

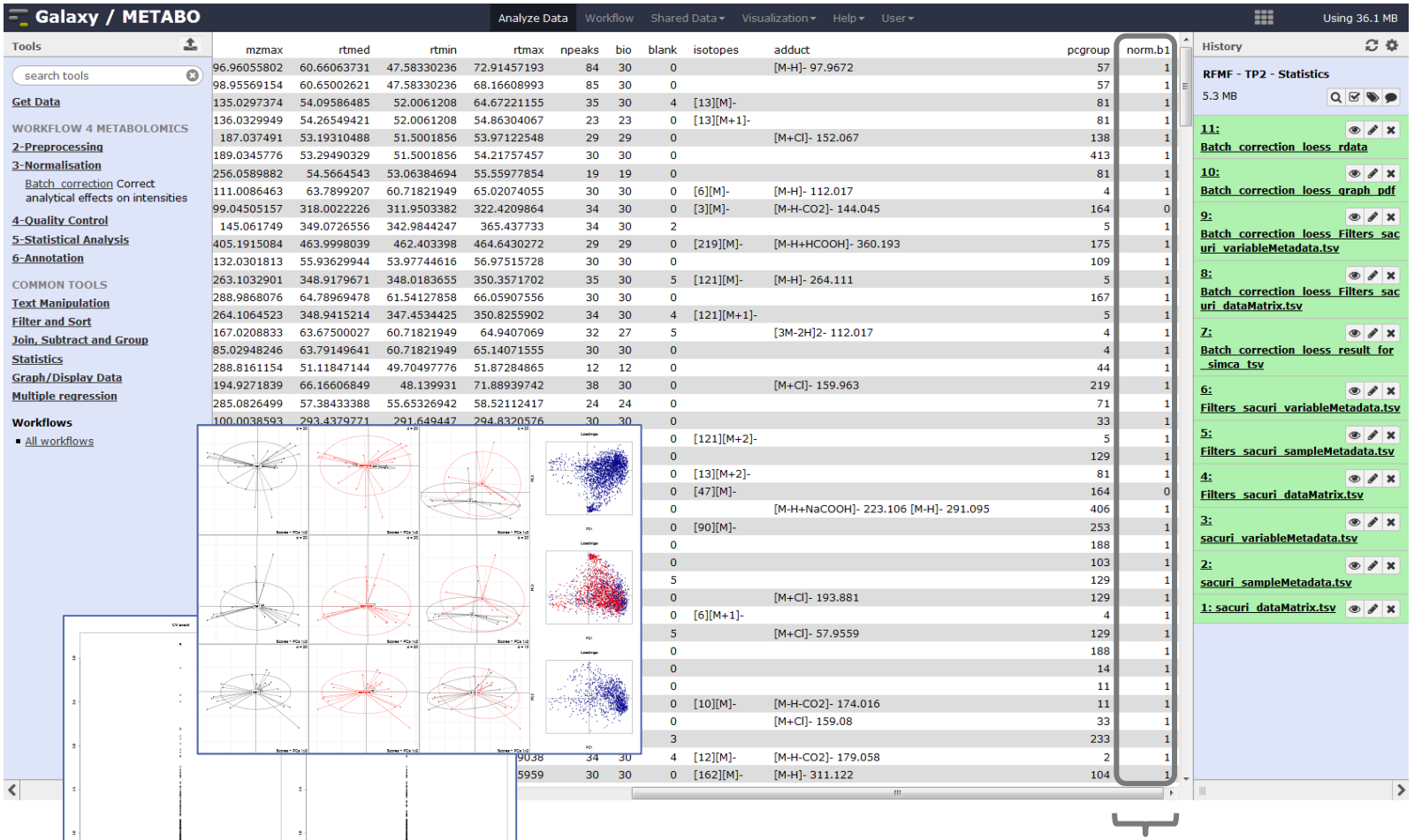
Outputs

The module generates 5 files:

- An Rdata file containing the results for further export and processing with R
- A pdf file containing diagnosis plots of the correction
  - The level of detail can be selected as one of the arguments of the module
- Two tables corresponding to the data matrix and the variable meta-data file
  - The data matrix contains the new intensities
  - The variable meta-data table is the original one plus columns indicating if intra-batch correction as been applied for each ion and each batch (0=no, 1=yes)
- A combined table that corresponds to the junction of the sample meta-data file and the data matrix of corrected intensities



Example



Graphical output

Variable meta-data

Normalised intensities (data matrix)

Indicator for intra-batch normalisation (added in variable meta-data table)

Notes

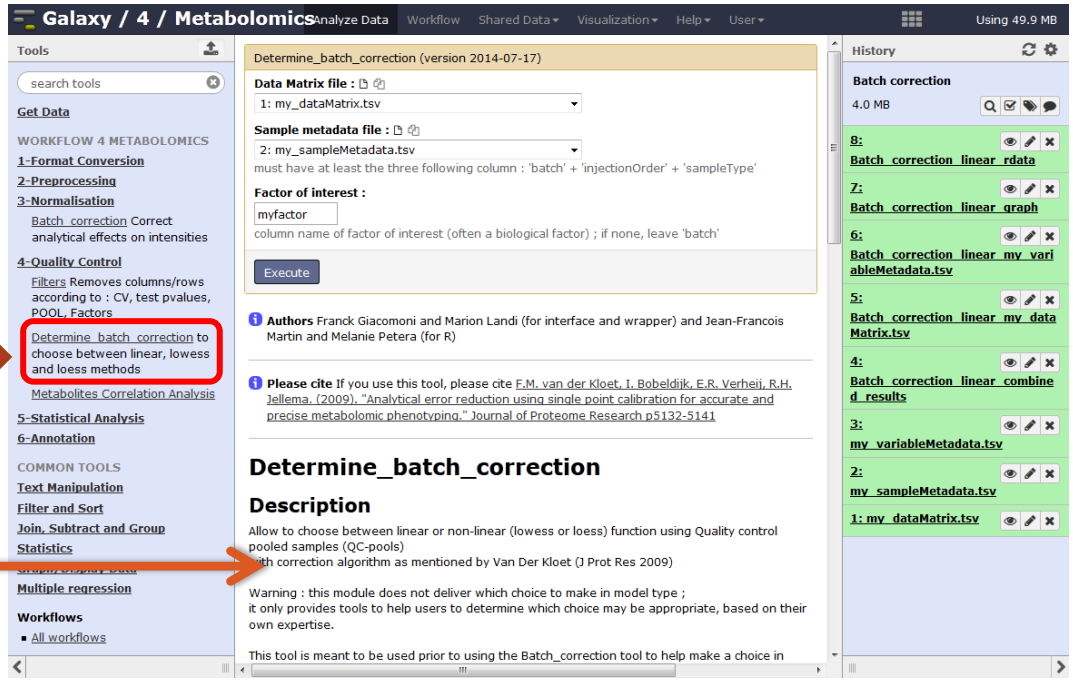
To help you determine which type of regression model you should use with your data, you can use the “Determine batch correction” module.

- It generates plots of intensities according to injection order and design effects for each ion.
- Each type of regression is displayed on intensity plots to enable comparison between the methods.
- A table indicating whether intra-batch correction is possible for each ion is generated.

4 – Quality Control

Determine batch correction

See the Help section of the module for more information



References

F.M. Van Der Kloet, I. Bobeldijk, E.R. Verheij, R.H. Jellema (2009). "Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping." Journal of Proteome Research p5132-5141