mQuest/mProphet Software Manual

STATISTICAL ANALYSIS OF MRM DATA

General

mQuest/*mProphet* is a software for the analysis of multiple/selected reaction monitoring MRM/SRM data. It was developed at the Institute for Molecular Systems Biology (IMSB) at the ETH Zurich.

The software provides fully automated processing, quality control and quantification of MRM data. *mQuest*/*mProphet* does not provide functionality to design MRM assays (transitions). In order for *mQuest*/*mProphet* to work, negative controls have to be included into the measurement. These so called decoy transitions are generated using the list of target transitions of the experiment.

mQuest/*mProphet* uses the full information content of MRM data to separate high from low quality signals and quantifies the quality in terms of false discovery rate (FDR). Further the software reports a number of different quantification types including the sum of peak heights/areas and an additional interference detection. *mQuest*/*mProphet* is broadly applicable because it includes the negative controls in the analysis for maximal recovery of information in each experiment. Therefore, *mQuest*/*mProphet* functions independent of the experiment design. For instance, it can handle data derived from metabolic labeled samples, non labeled samples or samples with synthetic reference peptides spiked in. Also, it can include external data derived from spectrum libraries in terms of relative fragment ion intensities.

mQuest/*mProphet* visualizes the recorded MRM transitions, the scoring of the peak groups and the statistical analysis.

To perform an experiment follow these steps:

1. convert your transition list into the format described below
2. add decoy transitions to this list using *mGen*
3. perform your measurement
4. convert your raw data into *mzXML*
5. use *mInteract* to analyze your data
6. browse through the raw data and analysis with the web browser

mQuest/mProphet package

The *mQuest/mProphet* software package mainly consists of *mGen*, *mMap*, *mQuest*, *mProphet* and *mInteract* and score_MRM. The functionality of each part is described in more detail below.

See www.mprophet.org for downloads, news and additional documentation.

Installation

SOFTWARE INSTALLATION

VIRTUAL MACHINE

We provide a virtual machine with the software already installed. The software is configured to run locally (e.g. on a windows laptop).

1. install VMware on your computer
2. enable virtualization in your BIOS (see the VMware help for more information)
3. download the virtual machine
4. unzip the virtual machine on your computer
5. choose "Open a Virtual Machine" in the VMware Player Menu and browse to the unzipped folder
6. choose "copied it" while importing the virtual machine
7. run the virtual machine (the user name is 'mprophet', the password is 'mprophet')

choose "example data analysis" on the desktop and click an .shtml file to browse through an example data set.

Visualization is enabled in Apache for the folder `/usr/local/apps/biognosys/mquest/www`. You can link any folder in there to enable visualization. This was already done for
`/home/mprophet/mProphet/mProphet_data_analysis`.

To connect the virtual machine to the host file system and analyze experimental data on the virtual machine choose:
Virtual Machine -> Virtual Machine Settings -> Options -> Shared Folders
The folder will then be mounted in `/mnt`

UBUNTU LINUX PACKAGE

Installing the package with apt-get from a package repository will automatically install all the dependencies needed for the software. As a prerequisite you need a freely available 64bit Ubuntu Linux 10.04 installed. If no such operating system is present, Ubuntu can be installed into a virtual machine which makes a tedious dual boot system unnecessary. As a virtualization software one can for instance use the freely available VMware or VirtualBox.

There are two package types to choose for installation. Type one installs only the command line tools, type two includes a configured Apache2 server to be able to browse through the results with a web browser.

Please follow these steps for the installation:

1. open a terminal
2. append the following line to `/etc/apt/sources.list`:
   `sudo gedit /etc/apt/sources.list`
   line：
   `deb http://systemsx03.ethz.ch/SybitDebRepo/binary/ubuntu10.04/DEBS/ /`
   sudo will give you super user rights which is needed to edit the system file sources.list
3. Generate the text file packager.key with the content from the link
   `gedit` <u>`packager.key`</u>
4. for security reasons it's obligatory to install a public key:
   `sudo apt-key add packager.key`

5. update the repositories with the command:
   `sudo apt-get update`
6. install either of the two package versions using the command:
   - `sudo apt-get install mquestweb`: recommended, Apache2 allows browsing through results
   - `sudo apt-get install mquest`: no Apache2 installation and hence no browsing through results

   for the web installation make sure to enter the correct domain name during installation (e.g. mquest.uni.edu or 131.63.230.61). If no host is entered when prompted, the visualization in the browser will be accessible locally via http://localhost.

Please be aware that the installation of mquestweb will modify your Apache configuration (httpd.conf,...)!

After installation the program will be located per default in /usr/local/apps/biognosys/mquest/bin. To be able to use shorter commands you can edit .profile with `gedit ~/.profile` and add the following line to the end of the file:

`PATH=/usr/local/apps/biognosys/mquest/bin:$PATH`
Activate the new profile：

`source .profile`

The standard path for data analysis is /usr/local/apps/biognosys/mquest/www. You can link any folder to this directory to be able to visualize the data e.g. with:

`ln -s /home/<user>/mProhpet /usr/local/apps/biognosys/mquest/www`

The path for this analysis in the web browser is `http://localhost/mQuest-web/mProphet` for a local installation.

We'd like to specially acknowledge **Adam Srebniak** for general help and making the package! adam.srebniak (at) systemsx (pt) ch

MANUAL INSTALLATION (WINDOWS)

We recommend the installation using the package repository or to download the virtual machine (see above). However, the software can also be installed manually on Windows, Linux or Mac OS X. This, implies manual installation of dependencies. For data visualization of the raw data, a webserver (e.g. Apache) supporting server side includes must also be installed. In the following we will not describe how to setup a webserver on windows. Instead, a quick description on how to setup the command line tools on a windows machine:

1. download the mQuest/mProphet software package from the download section
2. install a perl interpreter (e.g. from http://www.activestate.com/activeperl)
3. this is a list of non standard perl modules that are used by the software. they can be installed using the perl package manager `ppm`. Depending on the system, some modules might already be installed. Here is a screenshot with the typical modules that need to be installed on windows.
4. install the R interpreter (http://cran.r-project.org/)
5. start the R console and type `install.packages("MASS")` on the R command line Select a mirror and the R package will be installed

After this step all programs will properly run from the command line. To run the programs, change to the mQuest/bin directory and type `perl mInteract.pl -help` for a quick help. You can also add the path, where the scripts are, to the environment variable and run the programs from the command line without specifying the full path. For further information on how to use the software see the manual in the download section.

For data visualization a server that supports server side includes (SSI) must be installed. Although, in principle an Apache server can be installed and configured on windows we will not explain this here. Unless you are experienced with setting up a server we suggest to either install the package on an Ubuntu Linux or to install the virtual machine with the installed software.

mGen

mGen generates decoy transitions based on a list of target transitions. The decoy transitions are used as negative controls in the downstream statistical analysis for optimal recovery of information and for quality assessment.

INSTALLATION

mGen can be either installed with the whole package (recommended, see above) or it can be installed and run separately. To run *mGen* a Perl interpreter must be installed on the system. For instance the ActivePerl (www.activestate.com) distribution can be used for windows. Unpack *mGen* to a new folder and run from within this folder.

INPUT

As input the final and full list of transitions of the experiment should be used. As minimal required information the algorithm needs Q1, Q3 and a unique transition group id. The transition group id has to be a unique identifier for all transitions of a peptide including its modifications, labeling and the charge state of the peptide. Do not use the same transition group id for a peptide with two different modifications or charge states.

Besides the minimal information needed to generate the decoy transitions other important information should also be in the transition table if it is available: For instance the retention time if the MRM measurement is scheduled. Further, the relative intensities of transitions derived from e.g. spectrum libraries should be in the input table if they are used later on for scoring.

INPUT FORMAT

Two input formats are recognized by *mGen*.

1. A tab separated table format (*mMap* format).
2. Transitions exported from Skyline using the *mProphet*.skyr. The necessary skyline export report format (required by Skyline) can be downloaded from the *mQuest/mProphet* website. *mGen* converts the Skyline format into the *mMap* format

column name	required?	type	description
Q1	yes	number	**precursor m/z setting for the Quadrupole one**
Q3	yes	number	**fragment ion m/z setting for the Quadrupole three (fragment m/z)**
transition_group_id	yes	text	**an unique id for the transitions of one Precursor**
decoy	yes	[0,1]	*1* for decoy transitions, *0* for target transitions
sequence	yes	text	(deprecated: stripped_sequence,modification) the peptide sequence or a unique sequence identifier e.g. *sequence1*
prec_z	yes	integer	the precursor charge state
frg_type	yes	[a,b,c,x,y,z]	the fragment ion type
frg_nr	yes	integer	the fragment number
frg_z	yes	integer	the fragment charge state
CE	no	number	the collision energy used for this precursor
protein_name	no	text	the name of the corresponding protein
relative_intensity	no	number	the relative intensity of this fragment ion compared to the basepeak, needed for the relative intensity based scores – is ignored if not available.
transition_name	no	text	a name for the transition (e.g. AZGP1.FGCEIENDR.2.y8.1.light). Not used by *mQuest*, only for annotation,
isotype	only in label exp.	[light,heavy]	a string for the labeling (*light* or *heavy*), only needed in samples with labeling
Tr_recalibrated	no	number in minutes	retention time for scheduled MRM measurement

DECOY GENERATION FUNCTIONALITY

The decoy transitions are created such that they resemble the target transitions in most characteristics but do not target expected peptides. Also the decoy transitions should not record any signal related to the peptide targeted by the target transitions. To accomplish this, a random integer between 3 and 10 is subtracted from Q1 for one transition group and a random integer between -5 and 5 is added to alter Q3.

OUTPUT

The output table contains the target transitions and the decoy transitions are appended at the end of the file. The format is the *mMap* format described above. Decoy transitions are indicated with a „1" in the new column „decoy".

SYNOPSIS

```
mGen.pl –TRS transitions.txt
mGen.pl –SKY skyline_exported_transitions.txt
mGen.pl –TRS transitions.txt –workflow SPIKE_IN
mGen.pl –TRS transitions.txt –num_samples 3 –num_decoys 33
```

OPTIONS

If reference peptides are used in the measurements only half of the decoy transitions corresponding to the endogenous peptide need to be measured. This is because the data from the reference peptide is also used for the decoys in order to have a fair scoring. *mGen* provides this option with the –workflow SPIKE_IN flag.

For one experiment only roughly 100 decoy data points are needed, i.e. 100 measurements of a decoy transition group. This means that in large experiments only a few decoys need to be measured in each run. *mGen* provides the option to generate only a user specified number of decoy transitions with the –*num_decoys* option.

If the same transitions need to be measured on a number of samples *mGen* can be used to generate –*num_samples* transition lists each with –*num_decoys* randomly selected decoy transition groups.

mInteract

FUNCTION

This tool combines the functionality of the other data analysis programs. It allows batch processing of a whole experiment consisting of a large number of LC-MS runs including conversion of the raw *mzXML* (*mMap*), scoring (*mQuest*) and statistical analysis (*mProphet*). See below for a description for the individual programs.

INSTALLATION

mInteract can either be installed with the whole package (recommended, see above) or it can be installed and run separately. To run *mInteract* a Perl interpreter must be installed on the system. For instance the ActivePerl (www.activestate.com) distribution can be used for windows.

INPUT

mInteract takes as input the *mzXML* file and the transition list(s) with the format as described above in the *mGen* paragraph.

OUTPUT

mInteract generates *mrml*x files, which contain the scored transition group records, as well as .shtml files for browsing through the data. Further, a pdf file containing graphical display of the statistical analysis is generated. Finally, a table containing all the peak groups including all the information for quality filtering as well as quantification is written out.

SYNOPSIS

mInteract.pl –mMap –mQuest –mProphet –input ./ -workflow SPIKE_IN
For more options enter:
mInteract.pl –help or –manual for even more information

OPTIONS

The most important options of *mInteract* are the workflow and the machine type as well as the cycle time for QTRAP data. The *–mMap*_explicit together with the *–mMap*_match_regex option can be used to explicitly pair one *mzXML* file with one transition list file which reduces the chances of ambiguities in mapping the meta data back to the raw data.

mMap

FUNCTION

mMap combines the raw data from the mass spectrometer with the meta data of the transition (*mMap* format). This is necessary because the metadata of the transitions is lost during the conversion process to *mzXML*.

INSTALLATION

mMap can be either installed with the whole package (recommended, see above) or it can be installed and run separately. To run *mQuest* a Perl interpreter must be installed on the system. For instance the ActivePerl (www.activestate.com) distribution can be used for windows.

INPUT

mMap requires as input the *mzXML* file and the transition list in the *mMap* format as described above in the *mGen* paragraph. The transition-list must be tab separated (.txt in Excel).

The raw data can be converted into the open source *mzXML* format with the following converters:

Machine	converter	Download link
Thermo TSQ	readw.exe	http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW#Current_Version
AB Sciex QTRAP	mzwiff	http://tools.proteomecenter.org/wiki/index.php?title=Software:mzWiff

Note that the vendor software (Xcalibur or Analyst respectively) needs to be installed on the computer where the converters are installed. It is therefore most straightforward to run the conversion on the computer attached to the mass spectrometer.

THERMO, QUANTUM OR VANTAGE

The Xcalibur raw machine file format is *.raw*

Conversion, single file: readw.exe --mzXML file.raw
*Conversion, multiple files: readw.exe --mzXML *.raw*

AB-SCIEX, QTRAP 5500 OR QTRAP 4000

The native file format is .wiff and wiff.scan. Both files are needed for the conversion, the wiff.scan file contains the XIC information.

Conversion, single file: mzWiff.exe --mzXML file.wiff
*Conversion, multiple files: mzWiff.exe --mzXML *.wiff*

OUTPUT

mMap generates *mrml* files, which combine the raw data with the meta data of the transitions. The data is grouped into transition group records. The *mrml* files serve as input for the search engine *mQuest*.

SYNOPSIS

mMap.pl –mz "*.*mzXML*" –trans transitions.txt

For more options enter:

mMap.pl -h

OPTIONS

mMap needs the information about the mass spectrometer because this influences the structure of the raw data. Use the option –*mach* with either QTRAP (ABI Sciex), TSQ (Thermo Scientific). For the QTRAP the cycle time has to be specified in seconds with the –*cyc* option.

mQuest

mQuest processes *mrml* files and performs the scoring of the signals that were recorded. Several scoring schemas can be selected to accommodate different experimental set-ups (e.g. with/without reference peptide).

Shortly, the algorithm performs the following steps: First the peak detection finds the peaks in the recorded transitions. Then the peaks in a transition group record are grouped to peak groups. Each peak group is scored for a number of characteristics like shape correlations, co-elution correlations and correlation to predicted relative intensities of fragment ions if available.

INSTALLATION

mQuest can be either installed with the whole package (recommended, see above) or it can be installed and run separately. To run *mQuest* a Perl interpreter must be installed on the system. For instance the ActivePerl (www.activestate.com) distribution can be used for windows.

Following open source Perl modules have to be installed (e.g. from CPAN). More Perl modules might be required depending on the local Perl installation.

File::Basename, XML::TreeBuilder, File::Spec, File::Spec::Functions, XML::Writer, Data::Dumper, DBI, SVG, GD

INPUT

The input of *mQuest* are the *mrml* files which combine the raw data with the transition meta information.

OUTPUT

mQuest generates *mrml*x and _scores.xls files that contain a list of peak groups with associated scores. The data also contains the information that can be used for quantitation afterwards.

SYNOPSIS

mQuest.pl -mrm "*_*mrml.xml*"

OPTIONS

mQuest takes as input a parameter file which specifies the workflow and other parameters. A couple of parameter files can be found in the conf directory.

mProphet

mProphet combines the peak group scores for optimal recovery of signals also making use the data derived from the decoy transitions. The decoy data is further used for quality control of the data (false discovery rate estimation). After running *mProphet*, the data can be filtered according to a user defined quality and the resulting data can be further analyzed or used for quantitation.

INSTALLATION

mProphet can be either installed with the whole package (recommended, see above) or it can be installed and run separately. If manually installed, R must be on the system. Download the R package from www.r-project.org and install it on your system. Install the package MASS (at least Built 2.8.1).

INPUT

mProphet uses a tab separated table format as input. Each row in the table represents a peak group. Following columns are necessary in the input table:

1	transition_group_record	a unique identifier for a transition group recorded once
2	decoy	a logical indicator for whether the transition group is decoy or not
3	main_var…	the main peak group score used for discrimination
4	var…	1 to n score columns starting with var…

Rows with identical transition group record ids are interpreted as peak groups from the same transition group record. In the results only the top scoring peak groups are considered.

FUNCTION

mProphet uses a semi-supervised learning approach to combine peak group scores for optimal separation. In the process of the semi-supervised learning data derived from the decoy transitions can be used as false by definition making this flexible learning approach possible. Prior to the learning step the data is split into training and testing data set to avoid bias in false discovery rate estimation.

OUTPUT

mProphet generates two major output tables. The first table contains all the top scoring peak groups as rated by their *mProphet* score. The *mProphet* score is a weighted linear combination of all peak group scores for optimal separation of true and false peak groups. The second table contains statistical characteristics, such as false

discovery rate and sensitivity, of the data set for an array of *mProphet* score cutoffs. Further *mProphet* generates a pdf containing graphically displayed information about the single scores and the *mProphet* score.

SYNOPSIS

R can be invoked directly:
```
R --slave –args bin_dir=C:\mProphet data_file=peak_groups.tsf < mProphet.R
```
or the wrapper ca be used:
```
mProphet.pl –workflow SPIKE_IN –mQuest ./
```

Software Authors

Lukas Reiter
current address
  reiter (at) biognosys (pt) ch
  Biognosys AG
  c/o IMSB ETH Zurich
  Wolfgang-Pauli Strasse 16
  HPT C119
  CH-8093 Zurich
  Switzerland

former address
  Hengartner Laboratory
  Institute of Molecular Life Sciences
  Winterthurerstrasse 190
  University of Zurich - Irchel
  CH-8057 Zurich
former located
  Aebersold Laboratory
  Institute for Molecular Systems Biology
  Wolfgang-Pauli-Str. 16
  ETH Hönggerberg, HPT C 75
  CH-8093 Zurich

Oliver Rinner
current address
  rinner (at) biognosys (pt) ch
  Biognosys AG
  c/o IMSB ETH Zurich
  Wolfgang-Pauli Strasse 16
  HPT C119
  CH-8093 Zurich
  Switzerland

former address
  Aebersold Laboratory
  Institute for Molecular Systems Biology
  Wolfgang-Pauli-Str. 16
  ETH Hönggerberg
  CH-8093 Zurich

Disclaimer

License

This software is licensed under the Apache License Version 2.0.

Warranty

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.