Introduction

- Graphical interface click-button tools within windows
  - very ergonomic
  - too ergonomic → lack of flexibility
  - don’t count on it! Have you ever seen a PhD student having the time to make beautiful green buttons?
  - paying for it!

- Tools available on the internet
  - very ergonomic
  - too ergonomic → lack of flexibility
  - A small part of the available tools
  - distributed on different universities locations
  - the submission size is often limited
  - must not be paranoid
Introduction

- Command line tools represent almost the majority of scientific tools and can be executed on high performance computers.
- Good parameters completeness.
- g33ks love it, since automatable, workflowsable, ...
- Minimum Linux knowledge is required.
- Cruel lack of ergonomics.
INTRODUCTION / GALAXY
## Galaxy / 4 / Metabolomics

### Tools

**Upload File** from your computer

**Export Data**

**Format Conversion**

**Preprocessing**

**Normalization**

**Quality Control**

**Statistical Analysis**

**Annotation**

**NMR**

**Preprocessing**

**Normalization**

**Quality Control**

**Statistical Analysis**

**Annotation**

**COMMON TOOLS**

**Data Handling**

**Text Manipulation**

**Filter and Sort**

**Join, Subtract and Group**

### Analyze Data

**Workflow**

**Shared Data**

**Visualization**

**Admin**

**Help**

**User**

---

### xcms.xcmsSet: version 2.0.1

#### Choose your inputs method:

- Zip file from your history containing your chromatograms

#### Zip file:

1. sacuri.zip

#### Extraction method for peaks detection:

- matchedFilter

- [method] See the help section below

#### Step size to use for profile generation:

0.01

- [step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size

#### Full width at half maximum of matched filtration gaussian model peak:

30

- [fwhm] Only used to calculate the actual sigma

### Advanced options:

- hide

---

### Authors

- Colin A. Smith: csmith@scripps.edu, Ralf Tautenhahn: rtautenh@gmail.com, Steffen Neumann: sneumann@ipb-halle.de, Paul Benton: paul.benton08@imperial.ac.uk and Christopher Conley: ccconley@ucdavis.edu


For details about this tool, please go to: [http://www.biolconductor.org/packages/release/bioc/html/xcms.html](http://www.biolconductor.org/packages/release/bioc/html/xcms.html)

### Galaxy integration

- ABIMS TEAM, Station biologique de Roscoff

Contact: support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.

---

### History

- **19:**
  - xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Diffreport)

- **18:**
  - xset.group.retcor.group.fillPeaks.annotate.negative.Rdata

- **17:**
  - xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

- **16:**
  - xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

- **15:**
  - xset.group.retcor.group.fillPeaks.RData

- **14:**
  - xset.group.retcor.group.Rplots.pdf

- **13:**
  - xset.group.retcor.group.RData

- **12:**
  - xset.group.retcor.BPCs_corrected.pdf

- **11:**

---

**Sacuri Zip**

- 19 shown

- 289.7 MB

---

**Galaxy**
Introduction / Galaxy

• Galaxy it’s …
  – No need to execute a command line through a terminal
  – Programming or scripting skills are not required
  – Submission of jobs is transparent through a high performance computer cluster
  – Secure histories and data manager
  – A data and protocols sharing system
  – Tool-boxes of several bioinformatics fields
    – NGS
    – Metabolomics
    – Statistics
    – Chemistry
    – Image analysis
    – Etc …
  – A web-based interface
RNA-Seq Analysis Tools

- CLC Bio
- Geneious
- Galaxy
- iPlant DE
- Command Line
- R

COST vs DIFFICULTY OF USE/LEARNING CURVE

Size of dot indicates flexibility/power
Why Galaxy?

– Accessibility
– Reproductibility
– Transparency
Introduction / Galaxy

```
[login@n0 -]$: cd projet
[login@n0 login]$: cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$: cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@ab-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err
tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
--GTF ../input/panda_v121029.gtf --b2-sensitive --r 300
--num-threads 8

[login@n0 mapping]$: qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```
Introduction / Galaxy

[lecorguille@n0 ~]$ e-PCR --help

e-PCR: invalid option -- 
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
- m ## Margin (default 50)
- w ## Wordsize (default 7)
- n ## Max mismatches allowed (default 0)
- g ## Max indels allowed (default 0)
- f ## Use ## discontiguos words, slow if ##>1
- o ## Set output file
- t ## Set output format:
  1 - classic, range (pos1..pos2)
  2 - classic, midpoint
  3 - tabular
  4 - tabular with alignment in comments (slow)
- d##-## Set default size range (default 100-350)
- p + Turn hits postprocess on/off
- v ## Verbosity flags
- a a|f Use presize alignmens (only if gaps>0), slow
  a - Allways or f - as Fallback
- x + Use 5'-end lowercase masking of primers (default -)
- u + Uppercase all primers (default -)

[...]
Introduction / Galaxy

xcmsSet.matchedFilter(object, fwhm = 30, sigma = f)

Arguments:

- object: xcmsRaw object
- fwhm: full width at half maximum of matched filtration gaussian model peak
- sigma: standard deviation (width) of matched filtration model peak
- max: maximum number of peaks per extracted ion chromatogram
- snthresh: signal to noise ratio cutoff
- step: step size to use for profile generation
- steps: number of steps to merge prior to filtration
- mzdiff: minimum difference in m/z for peaks with overlapping retention times
- index: return indices instead of values for m/z and retention times
- sleep: number of seconds to pause between plotting peak finding cycles
- scanrange: scan range to process

Choose your inputs method:

Zip file from your history containing your chromatograms

No unzip.zip dataset available.

Extraction method for peaks detection:

matchedFilter

Step size to use for profile generation:

0.01

Full width at half maximum of matched filtration gaussian model peak:

30

Maximum number of peaks per extracted ion chromatogram:

5

Signal to noise ratio cutoff:

10

Number of steps to merge prior to filtration:

2

Maximum number of peaks per extracted ion chromatogram

Signal to noise ratio cutoff

Number of steps to merge prior to filtration

Execute
Galaxy interface

Batch_correction (version 2.0.0)

Data Matrix file:
17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv

Sample metadata file:
3: sampleMetadata.tsv

must contain at least the three following columns: 'batch' + 'InjectionOrder' + 'sampleType'

Variable metadata file:
16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv

Type of regression model:
linear

To select between linear or non-linear (lowess or loess) methods to be used in Van der Kloet algorithm; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest:
batch

column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots:
basic

Amount of plots in the pdf file output. See Help section for more details.

Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

Contributors
Mélaine Potea - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)
GET HELP
Workflows for Metabolomics

W4M HowTo

<table>
<thead>
<tr>
<th>Description</th>
<th>PDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import datasets &lt; 2Go</td>
<td>Download</td>
</tr>
<tr>
<td>Import datasets &gt; 2Go</td>
<td>Download</td>
</tr>
<tr>
<td>Build And Configure A Workflow</td>
<td>Download</td>
</tr>
<tr>
<td>Share Histories And Workflows</td>
<td>Download</td>
</tr>
<tr>
<td>Format Data For Postprocessing</td>
<td>Download</td>
</tr>
<tr>
<td>Perform Xcms Preprocessing</td>
<td>Download</td>
</tr>
<tr>
<td>Perform Drift And Batch Correction</td>
<td>Download</td>
</tr>
<tr>
<td>Perform Univariate Analyzes</td>
<td>Download</td>
</tr>
<tr>
<td>Perform Multivariate Analyzes</td>
<td>Download</td>
</tr>
<tr>
<td>Perform LCMS Annotations</td>
<td>Download</td>
</tr>
<tr>
<td>Use NIST</td>
<td>Download</td>
</tr>
</tbody>
</table>

For requests, please fill in the webform here

workflow4metabolimcs.org
DATA IMPORT

< 2 GO
Data import < 2 Go

Welcome to workflow4metabolomics.org v2.0


Help and support: support@workflow4metabolomics.org

Latest news
01/06/2015 - Workflow4Metabolomics v2.0 starts today - Check the changelog section below

01/06/2015 - The W4M 2.0 release is presented in the June 2015 MetaboNews Spotlight [Link]
19/12/2014 - W4M publication in Bioinformatics is now available : Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics

Changelog
Tutorials
Past events

LC/MS
MS
Common
Data import < 2 Go

Copy / Paste data

1. Choose Paste/Fetch data
2. Select files (sampleMetadata class polarity batch, Blanc1S blank negative 1)
3. Enable Convert spaces to tabs
   - Option to use POSIX standard
Data import < 2 Go

From local files
Data import < 2 Go

From local files
Data import < 2 Go

From local files
Data import < 2 Go

From local files
Step 1: Choose a FTP Client

DATA IMPORT

< 2 GO
STEP 1: CHOOSE A FTP CLIENT

Avoid: Malwares inside

FileZilla
Cyberduck
WinSCP

Data import > 2 Go
Step 2: Easy!

DATA IMPORT

< 2 GO
Data import > 2 Go
Data import > 2 Go

Download data directly from web or upload files from your disk

FTP files
This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at ftp.workflow4metabolomics.org using your Galaxy credentials (email address and password).

⚠️ Your FTP directory does not contain any files.

Choose local file  Choose FTP file  Paste/Fetch data  Start  Pause  Reset  Close
Data import > 2 Go
Data import > 2 Go

Download data directly from web or upload files from your disk

Your FTP directory does not contain any files.
Data import > 2 Go

Download data directly from web or upload files from your disk

Connection opened
ftp.workflow4metabolomics.org
Data import > 2 Go
Data import > 2 Go

Download data directly from web or upload files from your disk

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Type</th>
<th>Genome</th>
<th>Settings</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>securi.zip</td>
<td>0.2 GB</td>
<td>Auto-det.</td>
<td>unspecified (?)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FTP files

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at ftp.workflow4metabolomics.org using your Galaxy credentials (email address and password).

Available files:

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Created</th>
</tr>
</thead>
<tbody>
<tr>
<td>securi.zip</td>
<td>0.2 GB</td>
<td>06/04/2015 06:13:33 PM</td>
</tr>
</tbody>
</table>
Data import > 2 Go

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Type</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>secun.zip</td>
<td>0.2 GB</td>
<td>Auto-det.</td>
<td>unspecified</td>
</tr>
</tbody>
</table>
Data import > 2 Go
Welcome to workflow4metabolomics.org v2.0


Help and support: support@workflow4metabolomics.org

Latest news

01/06/2015 - Workflow4Metabolomics v2.0 starts today - Check the changelog section below

01/06/2015 - The W4M 2.0 release is presented in the June 2015 MetaboNews Spotlight [link]


19/12/2015 - W4M publication in Bioinformatics is now available - *Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics*

- Changelog
- Tutorials
- Past events

---

Data import > 2 Go

---
DATA IMPORT
Data import

For HUGE public resources: genome, databank ...

--> Make a request to the support team
Data import

- Exercice
  - Fetch this file

http://tinyurl.com/w4mddata1

- For this exercice, consider that it is >2 Go
TOOLS
Tools - panel
Choose your inputs method:
Zip file from your history containing your chromatograms:

Zip file:
1: sacuri.zip

Extraction method for peaks detection:
matchedFilter
[method] See the help section below

Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size

Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma

Advanced options:
hide

Authors: Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton8@imperial.ac.uk and Christopher Conley cijconley@ucdavis.edu

For details about this tool, please go to http://www.biocductor.org/packages/release/bioc/html/xcms.html

Galaxy integration ABIMS TEAM, Station biologique de Roscoff.

Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
Tools can have some advanced options

A job has been successfully added to the queue - resulting in the following datasets:

2: xset.RData
3: sampleMetadata.tsv
4: xset.TICs_raw.pdf
5: xset.BPCs_raw.pdf
6: xset.log.txt

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.
Tools - Job status

- Status

Job is waiting to run

= the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs or on the requested number of processors
Tools - Job status

- Status

Job is currently running

= the job is being executed on the computing cluster

Duration time of this status depends completely on the job’s attributes and the computing resources allocated.

Some programs are executed with several processors (using 4, 8 or 16 Gb of RAM).

And others are mono-threaded 😞
Tools - Job status

- Status

Job is finished

It’s status is OK

but warnings or errors can be hidden behind. Ah hum!
Tools - Job status

- Status

Job is finished but with an error status

= the program sends an error

The error is often explained by the program and sometimes … not.
Tools - Job status

- Status

Job is finished but with an error status

Error causes:
- The user :P
  - Bad usage: input file, format or option
  - Wrong porting of the program through Galaxy … sorry :/
  - Non anticipated crash of the program
Exercice

TOOLS
• Aim of this exercise
  – Import data into Galaxy into a new history
  – Execute and chain example of little Galaxy friendly tools together.
Tools - Exercice

- Fetch these two tabular files (< 2Go)
  - Link1:  
    http://tinyurl.com/w4mdata2  
  - Link2:  
    http://tinyurl.com/w4mdata3  

- Tabular files (data separated by tab delimiters)
  - VariableMetadata.tsv
  - DataMatrix.tsv

- Check their contents and datatypes through Galaxy.
Tools - Exercice

• First tool:
  – Search for the tool « Compute an expression on every row » in the toolbar)
  – Calculate the average for each metabolite by sample:

• Set the parameters
  – Add expression:
    \[(c2+c3+c4+c5+c6+c7)/6\]
  – as a new column to: Choose the DataMatrix.tsv
• Second tool:
  – Search for the tool « Cut columns from a table » in the toolbar)
  – Keep only columns 1 and 8:

  • Set the parameters
    – Cut columns: $c_1, c_8$
    – Delimited by: Tab
    – From?: compute on Data 1
Tools - Exercice

• Third tool:
  – Search for the tool « Join two Datasets side by side on a specified field » in the toolbar)
  – Join the two tab files by the metabolite name:

  • Set the parameters
    – Join: **Cut on Data 3**
    – Using column: **column 1**
    – with: **variableMetadata.tsv**
    – And column: **column 1**
Part II

TOOLS
Tools – Handle errors

- 07-06-13: Metabolomic: Workflow 4 Metabolomics, updated to version 2.1.0 (2013_06_07)
- 30-04-13: RNASeq: DESeq is now available for RNASeq expression data with reference (with gtf input).
- 26-04-13: RNASeq: DESeq is now available for de novo RNASeq expression data (without gtf input).
- 25-04-13: RNASeq: sam2counts is now available to count the reads coverage by transcript. It's also a requirement for DESeq denovo.
- 26-04-13: Metabolomic: Workflow Metabolomic by ABiMS, updated to version 2.0.0 (2013_04_18)

An error occurred with this dataset: Fatal error: Exit code 10! ERROR: There is a problem with the group of condition (presence of NA). You may need to use change the mode (column/row) Current groups: NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
Tools – Handle errors

Dataset generation errors

Dataset 7: m2XML_copper_stress.group_retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv_anova_pvalue.tabular

Tool execution generated the following error message:

Fatal error: Exit code 10 ()
ERROR: There is a problem with the group of condition (presence of NA). You may need to use change the mode.
Current groups:
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA
NA NA NA NA NA

An error occurred with this dataset: Fatal error: Exit code 10 () ERROR: There is a problem with the group of condition (presence of NA). You may need to use change the mode (column/row). Current groups:
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA

View or report this error
Tools – Handle errors

Sent to the support team
HISTORY
Both inputs and outputs

Galaxy / 4 / Metabolomics

Tools
- Upload File from your computer
- Export Data
- LC-MS
- Format Conversion
- Preprocessing
- Normalisation
- Batch_correction: Corrects intensities for signal drift and batch-effects
- Determine_batch_correction: to choose between linear, loess and loess methods
- Transformation: Transforms the dataMatrix intensity values
- Quality Control
- Statistical Analysis
- Annotation
- GC-MS
- Preprocessing
- Normalisation
- Quality Control
- Statistical Analysis
- Annotation
- NMR
- Preprocessing
- Normalisation

Analysis Data
- Workflow
- Shared Data
- Visualization
- Admin
- Help
- User

History panel

Batch_correction (version 2.0.0)

Data Matrix file:
17: xset.group.recor.group.fillPeaks.annotate.dataMatrix.tsv

Sample metadata file:
3: sampleMetadata.tsv

must contain at least the following columns: 'batch' + 'InjectionOrder' + 'sampleType'

Variable metadata file:
16: xset.group.recor.group.fillPeaks.annotate.variableMetadata.tsv

Type of regression model:
linear

To select between linear or non-linear (loess or loess) methods to be used in Van Der Kloe algorithm; when using loess, you can choose to use pools or samples to model batch effect.

Factor of interest:
batch

column name of factor of interest (often a biological factor); if none, leave 'batch'

Level of details for plots:
basic

Amount of plots in the pdf file output. See Help section for more details.

Execute

Authors
Jean-Francois Martin - PF MetaToul-AXIOM ; INRA ; MetaboHUB (for original version of this tool and overall development of the R script)

Contributors
Melanie Potera - PFEM ; INRA ; MetaboHUB (for R wrapper and R script improvement)
Etienne Thevenot - LIST/LADIS ; CEA ; MetaboHUB (for R script and wrapper concerning "all loess pool" and "all loess sample" methods)

History

19: xset.group.recor.group.fillPeaks.annotate.variableMetadata.tsv (DiffReport)
18: xset.group.recor.group.fillPeaks.annotate.negative.Rdata
17: xset.group.recor.group.fillPeaks.annotate.dataMatrix.tsv
16: xset.group.recor.group.fillPeaks.annotate.variableMetadata.tsv
15: xset.group.recor.group.fillPeaks.RData
14: xset.group.recor.group.RPlots.pdf
13: xset.group.recor.group.RData
12: xset.group.recor.BPCs_corrected.pdf
11: 

Sacurizip
19 shown
289.7 MB

File History
History panel renaming and annotation
## Saved Histories

### History panel

**Saved histories: Rename, Delete, Delete Permanently**

<table>
<thead>
<tr>
<th>Name</th>
<th>Datasets</th>
<th>Tags</th>
<th>Size on Disk</th>
<th>Created</th>
<th>Last Updated</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacuri</td>
<td>19</td>
<td>2 Tags</td>
<td>289.7 MB</td>
<td>Sep 02, 2015</td>
<td>~3 days ago</td>
<td>current history</td>
</tr>
<tr>
<td>Sacuri Lib</td>
<td>30</td>
<td>0 Tags</td>
<td>17.3 MB</td>
<td>May 14, 2014</td>
<td>Sep 02, 2015</td>
<td></td>
</tr>
<tr>
<td>Cooper Stress Lib</td>
<td>19</td>
<td>0 Tags</td>
<td>7.8 MB</td>
<td>May 13, 2014</td>
<td>Sep 02, 2015</td>
<td></td>
</tr>
</tbody>
</table>

For 0 selected histories: Rename, Delete, Delete Permanently, Undelete

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.
### Saved Histories

<table>
<thead>
<tr>
<th>Name</th>
<th>Datasets</th>
<th>Tags</th>
<th>Sharing</th>
<th>Size on Disk</th>
<th>Created</th>
<th>Last Updated</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sacuri</td>
<td>19</td>
<td>2 Tags</td>
<td></td>
<td>289.7 MB</td>
<td>Sep 02, 2015</td>
<td>~3 days ago</td>
<td>current history</td>
</tr>
<tr>
<td>Sacuri Lib</td>
<td>30</td>
<td>0 Tags</td>
<td></td>
<td>17.3 MB</td>
<td>May 14, 2014</td>
<td>Sep 02, 2015</td>
<td></td>
</tr>
<tr>
<td>Cooper Stress Lib</td>
<td>19</td>
<td>0 Tags</td>
<td></td>
<td>7.8 MB</td>
<td>May 13, 2014</td>
<td>Sep 02, 2015</td>
<td></td>
</tr>
</tbody>
</table>

For 0 selected histories: Rename, Delete, Delete Permanently, Undelete

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.
Dataset

Both inputs and outputs
Dataset

Informations

Galaxy / 4 / Metabolomics

**Tool: xcms.group**

- Name: xset.group.retocor.group.RData
- Created: Wed Sep 2 09:11:46 2015 (UTC)
- Filesize: 5.2 MB
- Object: xset.group.retocor.group.RData
- Format: rdata.xcms.group
- Galaxy Tool ID: toolshed.france-bioinformatique.fr/repos/lecorguille/xcms_group/alims_xcms_group2.0.1
- Galaxy Tool Version: 2.0.1
- Tool Version: toolset
- Output: toolset
- Tool Standard Error: toolset
- Tool Exit Code: 0
- API ID: 9265a1b3d61fdeb
- History ID: f7c05917f701f7
- UUID: 72a13a4b-6e2e-47a6-b152-2768b7e876

**Input Parameter**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Note for rerun</th>
</tr>
</thead>
<tbody>
<tr>
<td>xset.RData</td>
<td>9: xset.group.retocor.RData</td>
<td></td>
</tr>
<tr>
<td>Method to use for grouping</td>
<td>density</td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Minimum fraction of samples necessary</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Width of overlapping mz slices</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Advanced options</td>
<td>show</td>
<td></td>
</tr>
<tr>
<td>Maximum number of groups to identify in a single mz slice</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

**Inheritance Chain**

- xset.group.retocor.group.RData

---

**History**

- xset.group.retocor.group.RData
- 5.2 MB
- format: rdata.xcms.group
- 9265a1b3d61fdeb
- 72a13a4b-6e2e-47a6-b152-2768b7e876

ARGUMENTS INFO
- xset.group.retocor
- function group
- m/z resolution
- Rcpp 0.10.9
- Bioconductor 2.28.0
- mR 2.0.0
- igraph 0.7.1
- xcms 1.42.0
- snow 0.3.13
- batch 1.1.4

---

**View Details**

- xset.group.retocor.RPCs_corrected.pdf
- xset.group.retocor.TICs_corrected.pdf
- xset.group.retocor.RData
- xset.group.retocor.Rplots.pdf
Dataset

Renaming and annotation

**Edit Attributes**

- **Name:** xset.group.recor.group.fillPeaks.annotate

- **Info:**
  - adding: bio_vs_blank.annot.tsv (deflated 52%)
  - adding: bio_vs_blank_box (stored 0%)

- **Annotation / Notes:**

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

**Database/Build:**

- unspecified (?)

**Save**

**Auto-detect**

This will inspect the dataset and attempt to correct the above column values if they are not accurate.
Change the Datatype of the Dataset

- New Type: tabular

Dataset but not modify its contents. Use this if Galaxy has incorrectly guessed the type of your dataset.
Re-run a job
Cleanup

DATASET
Delete a dataset
The dataset isn’t really deleted
It’s in the Trash
“Empty Trash”: to free up disk space
WORKFLOW
A workflow is a sequence of tool operations and parameters.

Can match the experiment protocol.

A workflow is built to be replayed (more or less strict).
Our workflow
Our workflow with Galaxy
From a history
Workflow

From a history
## Workflow manager

### Your workflows

<table>
<thead>
<tr>
<th>Name</th>
<th># of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS-MS</td>
<td>7</td>
</tr>
<tr>
<td>Copy of 'gigaXml' shared by '<a href="mailto:ethevenot@sb-roscoff.fr">ethevenot@sb-roscoff.fr</a>'</td>
<td>13</td>
</tr>
<tr>
<td>Workflow LC/MS</td>
<td>6</td>
</tr>
<tr>
<td>Community</td>
<td>10</td>
</tr>
<tr>
<td>Full workflow</td>
<td>19</td>
</tr>
<tr>
<td>Workflow XCMS</td>
<td>8</td>
</tr>
</tbody>
</table>

### Workflows shared with you by others

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
<th># of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>demo_workflow_06_annotation</td>
<td><a href="mailto:mlandi@sb-roscoff.fr">mlandi@sb-roscoff.fr</a></td>
<td>6</td>
</tr>
<tr>
<td>cohort</td>
<td><a href="mailto:ethevenot@sb-roscoff.fr">ethevenot@sb-roscoff.fr</a></td>
<td>15</td>
</tr>
<tr>
<td>gigaRaw-convert</td>
<td><a href="mailto:ethevenot@sb-roscoff.fr">ethevenot@sb-roscoff.fr</a></td>
<td>1</td>
</tr>
</tbody>
</table>

### Other options

- Configure your workflow menu
Edit a workflow

Your workflows

<table>
<thead>
<tr>
<th>Name</th>
<th># of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISM</td>
<td>7</td>
</tr>
<tr>
<td>Run</td>
<td>13</td>
</tr>
<tr>
<td>Share or Publish</td>
<td>6</td>
</tr>
<tr>
<td>Download or Export</td>
<td>10</td>
</tr>
<tr>
<td>Copy</td>
<td>19</td>
</tr>
<tr>
<td>Rename</td>
<td>8</td>
</tr>
<tr>
<td>View</td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td></td>
</tr>
</tbody>
</table>

Workflows shared with you by others

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
<th># of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>demo_workflow_06_annotation</td>
<td><a href="mailto:mlandi@sb-roscoff.fr">mlandi@sb-roscoff.fr</a></td>
<td>6</td>
</tr>
<tr>
<td>cohort</td>
<td><a href="mailto:ethevenot@sb-roscoff.fr">ethevenot@sb-roscoff.fr</a></td>
<td>15</td>
</tr>
<tr>
<td>gigaRaw-convert</td>
<td><a href="mailto:ethevenot@sb-roscoff.fr">ethevenot@sb-roscoff.fr</a></td>
<td>1</td>
</tr>
</tbody>
</table>

Other options

Configure your workflow menu
Workflow

Edit a workflow: drag and drop
Edit a workflow: drag and drop
Edit a workflow: delete a noodle
Edit a workflow: add a tool
Edit a workflow: add a noodle
Workflow

Edit a workflow: set or release a parameter
Workflow

Run a workflow

Running workflow "LS-MS"

Step 1: xcms.xcmsset (version 2.0.1)

Choose your inputs method
Zip file from your history containing your chromatograms

**Zip file**
1: sacuri.zip

Extraction method for peaks detection
matchedFilter

Step size to use for profile generation
0.01 Lf

Full width at half maximum of matched filtration gaussian model peak
4 Lf

Advanced options
show

Maximum number of peaks per extracted ion chromatogram
50

Signal to noise ratio cutoff
3 Lf

Number of steps to merge prior to filtration
2 Lf

Step 2: xcms.group (version 2.0.1)

xset RData file
Output dataset 'xsetRData' from step 1

Method to use for grouping
density

**Bandwidth**
30

Minimum fraction of samples necessary
0.3 Lf
Workflow

- Possible

- Impossible (until now)
biologist ↔ biologist

• Sharing histories or datasets
  – With or without linked workflow
bioanalyst ↔ biologist

• Sharing workflows
  – Pre-configured parameters
  – With or without release parameters (set at runtime)

• According to the user-end knowledge
bioinformatician ↔ bioinformatician

• Sharing tools, scripts and wrappers
  – Toolshed
Datasets

Share

### Saved Histories

<table>
<thead>
<tr>
<th>Name</th>
<th>Datasets</th>
<th>Tags</th>
<th>Sharing</th>
<th>Size on Disk</th>
<th>Created</th>
<th>Last Updated</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preprocessing</td>
<td>8</td>
<td>1</td>
<td></td>
<td>45.6 MB</td>
<td>~19 hours ago</td>
<td>~less than ago</td>
<td>current history</td>
</tr>
<tr>
<td>Switch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share or Publish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rename</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete Permanently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After Preprocessing</td>
<td>3</td>
<td>0</td>
<td></td>
<td>1.4 MB</td>
<td>~37 minutes ago</td>
<td>~7 minutes ago</td>
<td></td>
</tr>
<tr>
<td>Rename</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete Permanently</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.
Workflow

Your workflows

<table>
<thead>
<tr>
<th>Name</th>
<th># of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>complete_workflow_RFME</td>
<td>17</td>
</tr>
</tbody>
</table>

Workflow shared with you by others

No workflows have been shared with you by others.

Other options

- Configure your workflows
- Edit
- Run
- Share or Publish
- Download or Export
- Copy
- Rename
- View
- Delete
Share or Publish Workflow 'complete_workflow_RFMF'

**Make Workflow Accessible via Link and Publish It**
This workflow is currently restricted so that only you and the users listed below can access it. You can:

- Make Workflow Accessible via Link
  Generates a web link that you can share with other people so that they can view and import the workflow.

- Make Workflow Accessible and Publish
  Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy’s Published Workflows section, where it is publicly listed and searchable.

**Share Workflow with Individual Users**
You have not shared this workflow with any users. 

- Share with a user

---

- **Restricted community**
- **All the Galaxy server users**
- **Designated community (login@sb-roscoff.fr)**
Running workflow "Workflow XCMS"

**Step 1: xcms.xcmsgSet** (version 20140507)

**Library directory name**

```
scurl
```

**Method**

matchedFilter

**step**

0.01

**fwhm**

30

**Advanced options**

hide

**Step 2: xcms.group** (version 20140507)

**RData file**

Output dataset 'output' from step 1

**Method**

density

**bw**

30

**minfrac**

0.5

**mzwid**

0.25

**Advanced options**

hide

**Step 3: xcms.recor** (version 20140507)
• Get shared histories
Share

• Get shared workflows
Share

- Import shared

**Histories**

**Workflows**

**Workflows shared with you by others**

- **Workflow.mmonsoor**
  - Owner: mmonsoor@sb-rostock.de
  - # of Steps: 7

Other:
- View
- Run
- Copy
- Remove
Share

Level 5
- Share of tools and descriptions in the ToolShed

Level 4
- Launch autonomously tools
- Use advanced parameters
- Use the Galaxy API
- Provide workflow for colleagues Level 1-3

Level 3
- Launch autonomously tools
- Use workflow more or less presetted

Level 2
- Use presetted workflow

Level 1
- Share his data to colleagues Level 2-5