MS data processing

Report creation and Annotations

Yann GUITTON

04/10/2018

v 2.0.0
CAMERA an other R package integrated in Galaxy

The R-package **CAMERA** is a **C**ollection of **A**lgorithms for **ME**tabolite **pRofile** **A**nnotation.

Its primary purpose is the annotation and evaluation of LC-MS data. It includes algorithms for annotation of isotope peaks, *adducts and fragments* in peak lists.

Additional methods cluster mass signals that originate from a single metabolite, based on rules for mass differences and peak shape comparison.
xcms diffreport & CAMERA

CAMERA.annotate = CAMERA::annotateDiffreport

In details:

1- xcms::diffreport : Generates features list, EICs, BoxPlot and statistics
**xcms diffreport & CAMERA**

\[ \text{CAMERA.annotate} = \text{CAMERA::annotateDiffreport} \]

**In details:**

1- `xcms::diffreport`: Generates features list, EICs, BoxPlot and statistics

---

**EICs**

Extracted Ion Chromatogram: 462.3047 - 462.3057 m/z

**Boxplots**

Feature 462.3/715

\[ m/z = M_x \]
xCMS diffreport & CAMERA

CAMERA.annotate = CAMERA::annotateDiffreport

In LC-MS ESI Features are usually not alone

Number of features is not equal to number of detected molecules

Search in raw data

m/z = M_x

m/z

C13 isotopes
CAMERA.annotate = CAMERA::annotateDiffreport

In details:
2- CAMERA::xsAnnotate: read xcms object
3- CAMERA::groupFWHM: search co-eluting features RT based
4- CAMERA::findIsotopes: search for isotopic relation between features (C_{12}/C_{13})
5- CAMERA::groupCorr : try to improve co-elution separation
6- CAMERA::findAdduct: search for known adducts and fragments [M+Na]^+, [M+H-H_2O]^+, ....

Non annotated, but low intensity
CAMERA adduct annotation: defining rules

Green = Annotation OK
Red = No annotation or wrong or adduct not in CAMERA

[M+H-NH$_3$]$^+$

[M+H-NH$_3$-H$_2$O]$^+$?
CAMERA.annotate = CAMERA::annotateDiffreport

In details:
2- CAMERA::xsAnnotate: read xcms object
3- CAMERA::groupFWHM: search co-eluting features RT based
4- CAMERA::findIsotopes: search for isotopic relation between features (C\textsubscript{12}/C\textsubscript{13})
5- CAMERA::groupCorr: try to improve co-elution separation
6- CAMERA::findAdduct: search for known adducts and fragments [M+Na]^+, [M+H-H\textsubscript{2}O]^+; ....

Some times .... Co-elution are not fully resolved, have a look to your data
Additional information added to the diffreport by CAMERA
**xCMS diffreport & CAMERA**

CAMERA.annotate = CAMERA::annotateDiffreport

**In details:**

1- xcms::diffreport : Generates features list, EICs, BoxPlot and statistics

2- CAMERA::xsAnnotate : read xcms object

3- CAMERA::groupFWHM : search co-eluting features RT based

4- CAMERA::findIsotopes : search for isotopic relation between features ($C_{12}/C_{13}$)

5- CAMERA::groupCorr : try to improve co-elution separation

6- CAMERA::findAdduct : search for known adducts and fragments $[M+Na]^+$, $[M+H-H_2O]^+$, ....

Many steps = quite a lot of parameters
CAMERA annotation function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition. (Galaxy Version 2.2.0)

**RData file**
- No rdata.xcms.fillpeaks or rdata dataset available.
- Output file from another function xcms (fillPeaks)

**Group co-eluted peaks based on RT [groupFWHM]**
- Multiplier of the standard deviation
  - Value: 6
  - [sigma]
- Percentage of FWHM width
  - Value: 0.6
  - [perfwhm]

**Annotation general options**
- **General ppm error**
  - Value: 5
  - [ppm]
- **General absolut error in m/z**
  - Value: 0.015
  - [mzabs]
xcmms diffreport & CAMERA more params

**Format Conversion**

**Preprocessing**

- CAMERA.annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.

**Mode**

- All functions
- [quick] If TRUE, use only groupFWHM and findIsotopes functions. Else if FALSE, use also groupCorr.

**Camera Annotation**

- **Annotate Isotopes [findIsotopes]**
  - Max. ion charge
    - 3
  - Max. number of expected isotopes
    - 4
  - The percentage number of samples, which must satisfy the C12/C13 rule for isotope annotation
    - 0.5

- **Verifying grouping co-eluted peaks [groupCorr]**
  - groupCorr: correlation threshold (0..1)
    - 0.75
  - groupCorr: Method selection for grouping peaks after correlation analysis into pseudospecies
    - hcs
    - [graphMethod]
  - groupCorr: significant correlation threshold
    - 0.05
xcms diffreport & CAMERA more params

**LC-MS**

**Format Conversion**

**Preprocessing**

`CAMERA.annotate` CAMER annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.

---

**Flowchart Diagram**

1. Raw LC/MS Data
   - NetCDF
   - mzXML
   - mzData

2. Filter and Identify Peaks `xcmsSet()`
3. Match Peaks Across Samples `group()`
   - Retention Time Correction `retcor()`
4. Fill in Missing Peak Data `fillPeaks()`
5. Statistically Analyze Results `diffreport()`
6. Visualize Important Peaks `getEIC()`

---

**Annotate Adducts [findAdducts]**

Which polarity mode was used for measuring of the ms sample

- negative polarity

How much peaks will be calculated in every thread using the parallel mode

- 100
- [max_peaks]

Use a personal ruleset file

- TRUE

[No csv dataset available.]

**Checklist**

- Manual
- Text
- Graphical
- Hyperlink
- Specialized

**Tools**

- xcms
- CAMERA
- diffreport

**Annotations**

- groupCorr: Use correlation inside samples for peak grouping
  - Yes
  - No
  - [calcCiS]

- groupCorr: Use isotopic relationship for peak grouping
  - Yes
  - No
  - [calcIso]

- groupCorr: Use correlation across samples for peak grouping
  - Yes
  - No
  - [calcCaS]
CAMERA

Annotation of Adduct Fragments and isotopes

More parameters

xcms diffreport & CAMERA more params

CAMERA.annotate: CAMERA annotate function. Returns annotation results (isotope peaks, adducts and fragments) and a diffreport if more than one condition.

Statistics and results export: [diffreport]

Number of condition

One condition

Export options

Convert retention time (seconds) into minutes

Yes [ ] No [ ]

Convert the columns rtmed, rtmin and rtmax into minutes

Number of decimal places for mass values reported in ions' identifiers.

4

A minimum of 4 decimal places is recommended. Useful to avoid duplicates within identifiers

Number of decimal places for retention time values reported in ions' identifiers.

0

Useful to avoid duplicates within identifiers

General used intensity value

[intval] See the help section below

Resubmit your raw dataset or your zip file

[ ] Execute
Output files

- **xset.annotate.variableMetadata.tsv**
  
  For each metabolite (row): the value of the intensity in each sample, fold, anova, mzmed, mzmin, mzmax, rtmed, rtmin, rtmax, npeaks, isotopes, adduct and pcgroup

- **xset.annotate.dataMatrix.tsv**
  
  A tabular file which represents for each metabolite (row), the value of the intensity in each sample (column).

- **xset.annotate.zip**
  
  It contains filebase_eic, filebase_box and filebase.tsv for one condition vs another (Anova analysis).

- **xset.annotate.Rdata rdata.camera.quick or rdata.camera.positive or rdata.camera.negative**
  
  Rdata file, that be used outside Galaxy in R.
Output files

- xset.annotate.variableMetadata.tsv

For each metabolite (row): the value of the intensity in each sample, fold, anova, mzmed, mzmin, mzmax, rtmed, rtmin, rtmax, npeaks, isotopes, adduct and pcgroup
And next...database search!