4 Wm

Workflow4metabolomics

MS processing

Annotation

Y. Guitton, Alexis Delabrière, Franck Giacomoni, Jean-François Martin, Marion Landi

04/10/2018

v 3.0.0
Before

Will mass spectrometry bring me health, wealth and happiness?

A definite maybe, but first learn these....
Q, SIM, TOF, LC, ETD, ICR, EI, CI, ESI, GC, LIT, MS, APCI, DESI, CID, APPI, QIT, SRM, ECD, DART, FT, CE, MALDI, TIC and $m/z$
The annotation challenge

- Pre-processing
- Statistical analysis
- Raw data conversion
- Metabolite identifications
The annotation challenge

Induction of the novices

Abandon all hope, ye who enter here

9-circles stages of hell analysis
The annotation challenge

From mass spectra to molecule identification…an old problem

**Dendral** was an influential pioneer project in *artificial intelligence* (AI) of the 1960s, and the *computer software expert system* that it produced. Its primary aim was to study hypothesis formation and discovery in science. For that, a specific task in science was chosen: help *organic chemists* in identifying *unknown organic molecules*, by analyzing their *mass spectra* and using knowledge of chemistry.
Metabolomics is still faced with several significant challenges which currently limit its full scientific potential. The identification of metabolites is essential to convert analytical data into meaningful biological knowledge. However, identification confidence can vary widely because the process of identification is complex and dependent on the analytical platform and robustness of the methods applied, as well as the databases and resources used. Confident and unequivocal structure identification requires significant effort, which is multiplied dramatically in non-targeted metabolomics studies where 10–100s of metabolites can be deemed as biologically important and require identification.

Automated Annotation, the CASMI contest

CASMI contest
www.casmi-contest.org/ ▶ Traduire cette page

The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small...

Casmi 2016
CASMI 2016 is organised by Dr. Grégory Genta-Jouve ...

Important Dates
Important Dates. The schedule of CASMI 2017 is. Public release ...

Dates
Details about the CASMI 2013
Special Issue and dates are ...

Results of the Inaugural ...
... Small Molecule Identification:
Results of the Inaugural CASMI ...
Automated Annotation, other initiative

Metabolite Identification

Metabolomics aims to provide novel insights into the biochemical reactions of presence and concentrations of low molecular weight compounds from biologic for such high-throughput data collection are mass spectrometry (MS), often pre electrophoretic separation technologies, and nuclear magnetic resonance spect
The annotation challenge

Annotation vs Identification

- A world of known unknown or unknown unknown


The annotation challenge

It’s a long long way

Letzel, T. et al. In a class of its own. Lab & More International 4 (24-28)
The annotation challenge
It’s a long long way

• From m/z to compound name

Sample preparation → GC or LC-MS analyzer → Chromatogram

Separation of the compounds

Mass spectrum of peak at RT 8.5min

Compound X
The annotation challenge
It’s a long long way

- From m/z to compound name

Zoom on Isotopic profile

Annotated mass spectrum of peak at RT 8.5 min (LC-ESI-MS Positive Mode)
The annotation challenge
It’s a long long way

• Several thousands of metabolites are present in a biofluid
• Even with chromatography (LC or GC) it is not always possible to achieve a full separation of metabolites before the MS analysis (co-elution)
• An m/z (even with 1 ppm deviation) can correspond to many compounds (Kind et al 2006)

Within a 1 ppm range over the m/z 353.0781 ([M+H]⁺)
Match with 2702 compounds in Chemspider database
of which only 484 compounds contains exactly one Cl.

C\textsubscript{17}H\textsubscript{17}ClO\textsubscript{6} fit with 84 compounds in that database
The annotation challenge
It’s a long long way

Take home,
Remember MxTy Features can be... a lot of things

Adducts, in-source fragments, isotopes, noise...
The annotation challenge
It’s a long long way

Remember MxTy Features can be... a lot of things

Adducts, in-source fragments, isotopes, noise...

Output files
The annotation challenge
It’s a long long way

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>isotopes</td>
<td>adduct</td>
<td>pcgroup</td>
</tr>
<tr>
<td>M91.0533T294</td>
<td>91.053313583267</td>
<td>293.66607766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M103.0532T294</td>
<td>103.053210073</td>
<td>293.7915000766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M104.0565T294</td>
<td>104.05656973859</td>
<td>293.961005249023</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M111.0563T293</td>
<td>111.056392499</td>
<td>292.906007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M118.064T293</td>
<td>118.064027181733</td>
<td>293.486997947693</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M120.079T294</td>
<td>120.07977503358</td>
<td>293.68507766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M121.0828T294</td>
<td>121.082801452764</td>
<td>293.61007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M131.0479T294</td>
<td>131.047905554428</td>
<td>293.89701252808</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M149.0584T294</td>
<td>149.058389292059</td>
<td>293.626007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M166.0851T294</td>
<td>166.085116054389</td>
<td>293.626007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M167.0881T294</td>
<td>167.088130017022</td>
<td>293.61007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M355.2599T293</td>
<td>355.259991953212</td>
<td>293.486997947693</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M569.2058T294</td>
<td>569.205847544972</td>
<td>293.576007766724</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>M981.3261T293</td>
<td>981.326133270392</td>
<td>292.906007766724</td>
<td>NA</td>
<td>14</td>
</tr>
</tbody>
</table>
### Phenylalanine MS/MS HCD 25 NCE

<table>
<thead>
<tr>
<th>Name</th>
<th>m/z</th>
<th>mzmin</th>
<th>mzmax</th>
<th>rt</th>
<th>rtmnin</th>
<th>rtmmax</th>
<th>npeaks</th>
<th>adduct</th>
<th>isotopes</th>
<th>pgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>M91.0533T294</td>
<td>91.0533113853267</td>
<td>91.052889047859</td>
<td>91.0542767530706</td>
<td>203.686007766724</td>
<td>202.907090124756</td>
<td>205.102066070856</td>
<td>0 0 0 0</td>
<td>[4][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M103.0532T294</td>
<td>103.053210100747</td>
<td>103.053267422947</td>
<td>103.0545455854753</td>
<td>293.701510009766</td>
<td>202.247012252808</td>
<td>322.947010166777</td>
<td>18 0 12 4</td>
<td>[4][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M104.0532T294</td>
<td>104.053267422947</td>
<td>104.05361164341</td>
<td>104.0537014081481</td>
<td>293.961005249023</td>
<td>293.37981288662</td>
<td>294.421505249023</td>
<td>4 0 4 0</td>
<td>[4][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M117.0532T293</td>
<td>117.05325232409</td>
<td>117.05593050134</td>
<td>117.057724985541</td>
<td>292.906007766724</td>
<td>292.97912252808</td>
<td>293.37981288662</td>
<td>5 0 5 0</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M118.064T293</td>
<td>118.064027181733</td>
<td>118.063655955061</td>
<td>118.065335413064</td>
<td>293.489997947693</td>
<td>291.976007766724</td>
<td>294.449007766724</td>
<td>14 0 11 3</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M120.079T294</td>
<td>120.079775203358</td>
<td>120.077746621476</td>
<td>120.081432711575</td>
<td>293.866007766724</td>
<td>292.247012252808</td>
<td>296.675060406961</td>
<td>23 0 12 11</td>
<td>[13][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M131.082T294</td>
<td>131.082804152794</td>
<td>131.082841205154</td>
<td>131.0832187401154</td>
<td>293.636007766724</td>
<td>292.247012252808</td>
<td>295.151005249023</td>
<td>16 0 12 4</td>
<td>[13][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M131.084T294</td>
<td>131.084790554428</td>
<td>131.084790554428</td>
<td>131.084790554428</td>
<td>293.636007766724</td>
<td>295.151005249023</td>
<td>295.151005249023</td>
<td>7 0 6 1</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M149.058T294</td>
<td>149.0583802092059</td>
<td>149.0583802092059</td>
<td>149.059115224874</td>
<td>293.626007766724</td>
<td>291.976007766724</td>
<td>295.102066070856</td>
<td>13 0 11 2</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M156.085T294</td>
<td>156.085116054389</td>
<td>166.082511200787</td>
<td>166.085962757471</td>
<td>293.626007766724</td>
<td>292.247012252808</td>
<td>318.7800592948</td>
<td>23 0 10 9</td>
<td>[33][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M157.081T294</td>
<td>157.08181301702</td>
<td>167.087286250763</td>
<td>167.08817122005</td>
<td>293.611005249023</td>
<td>292.257012252808</td>
<td>295.102066070856</td>
<td>12 0 10 2</td>
<td>[33][M]+</td>
<td>NA 14</td>
<td></td>
</tr>
<tr>
<td>M356.259T293</td>
<td>356.25989193312</td>
<td>356.258869682692</td>
<td>356.258869682692</td>
<td>293.489997947693</td>
<td>291.958996147446</td>
<td>294.421005249023</td>
<td>4 0 4 0</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M569.208T294</td>
<td>569.2084745494972</td>
<td>569.2094252524099</td>
<td>569.2094252524099</td>
<td>293.578007766724</td>
<td>291.976007766724</td>
<td>294.421005249023</td>
<td>5 0 4 1</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M981.3261T293</td>
<td>981.326133270392</td>
<td>981.325578700556</td>
<td>981.33009300144</td>
<td>292.906007766724</td>
<td>292.357012252808</td>
<td>393.626007766724</td>
<td>6 0 4 2</td>
<td>NA 14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phenylalanine MS/MS HCD 25 NCE.
The annotation challenge
It's a long long way

The annotation challenge
It’s a long long way

Remember MxTy Features can be...a lot of things

Adducts, in-source fragments, isotopes, noise...

Output files

<table>
<thead>
<tr>
<th>Category 5</th>
<th>Category 4</th>
<th>Category 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>mz</td>
<td>mzmin</td>
</tr>
<tr>
<td>M134.0676</td>
<td>134.061005</td>
<td>134.060756</td>
</tr>
<tr>
<td>M178.0576</td>
<td>178.050333</td>
<td>178.050386</td>
</tr>
<tr>
<td>M179.0576</td>
<td>179.053611</td>
<td>179.053429</td>
</tr>
<tr>
<td>M215.0358</td>
<td>215.000584</td>
<td>214.999805</td>
</tr>
<tr>
<td>M357.1176</td>
<td>357.108799</td>
<td>357.108545</td>
</tr>
<tr>
<td>M358.1176</td>
<td>358.112119</td>
<td>358.112347</td>
</tr>
<tr>
<td>M359.1176</td>
<td>359.114597</td>
<td>359.114206</td>
</tr>
</tbody>
</table>
The annotation challenge

It’s a long long way

Remember MxTy Features can be…a lot of things

Adducts, in-source fragments, isotopes, noise…

Output files
The annotation challenge
It’s a long long way

Letzel, T. et al. In a class of its own. Lab & More International 4 (24-28)
Annotation with W4M

HOW TO
• Strategies with accurate masses
  – Classic search with only one library or a mix of chemical compounds libraries
    • « Bank Of Interest » : species linked to studies
    • « InDesperateBank » : find me something, please!
  – Complexe : chain complementary tools
    • Combine types of banks, formula elucidator.
    • Great to build workflow...
• The « LCMS Annotation » modules allow you to perform:
  • « Adducts » annotation: CAMERA
  • « Chemical » annotation:
    ✓ Meta-engine (Chemspider)
    ✓ General banks (Kegg, PubChem...)
    ✓ Spectral banks (MassBank, Golm)
    ✓ Specialized banks (HMDB, Lipidmaps, ...)
    ✓ In-House Database
    ✓ « De novo » tools (HR2 for chemical composition)
« LC/GC-MS Annotation » modules in W4M

Current version: 3.0


Help and support: support@workflow4metabolomics.org

Latest news:
- 10/05/2017 - LC-MS: A new tutorial video explain how to run xcmsSet in parallel on single files [link]
- 20/04/2017 - Workflow4Metabolomics v3.0 starts today - Check the changelog section below

Changelog

3.0.0 - 20/04/2017

LC-MS
- Preprocessing
  - UPGRADE - xcms.r (2.1.0): upgrade the xcms version from 1.44.0 to 1.46.0
  - NEW - xcms.r (2.1.0): The W4M tools will be able to take as input a single *xcms.xcmsSet Merger* before *xcms.group*.
  - xcms.xcmsSet (2.1.0): the default value of "matchedFilter" -> "Step": default values to 0.1
  - xcms.xcmsSet (2.1.0): propose scanrange for all methods
  - BUGFIX - xcms.group and xcms.fillPeaks (2.1.0): Add an option to expose
  - BUGFIX - CAMERA.annotate (2.2.0): the diffreport ids didn't convert the rt in minutes
  - UPDATE - CAMERA.annotate (2.2.0): the settings (digits, conversion in minutes) are conservative and because it can be dangerous for the data integrity during a full namecustom within the variableMetadata.
LC-MS Annotation

WORKFLOW- 1
Annotation workflow - 1

« Chemical » annotation workflow:
- Chain a specific and well adapted bank and a formula elucidator

- Specific bank
  - HMDB MS search
    - File of masses
    - variableMetadata (tabular)
    - HmdbResView (html)

- « De novo » tool
  - HR2 formula
    - File of masses
    - variableMetadata (tabular)
    - hr2ResView (html)

- 6-Annotation
  - HMDB MS search: search on HMDB by file selection
  - Lipidmaps: search on LM online with masses
  - HR2 formula: find a formula for the masses
    - Find a mol file with a kegg id
  - Kegg Compounds: a Kyoto Encyclopedia of Genes and Genomes small molecules database.
  - Chemspider: Search and share chemistry.
  - CAMERA.annotateDiffreport: Wrapper function for the xcms diffreport and the annotate function. Returns a diffreport within the annotation results.
Annotation workflow - 1

« Chemical » annotation workflow:

- **Human Metabolome DataBase:**
  - Bank containing small molecule metabolites found in the human body. V 3.6 with 41,815 entries + 5,688 proteins.

- **HR2:**
  - Molecular formula generator: HiRes MS by Joerg Hau.
  - Generate random chemical formula and filter them with “the seven golden rules”
Annotation workflow - 1

Tasks:
- Import « published Workflow » called: DEMO_workflow_annotation_HMDB
Annotation workflow – 1: your input DATA

• « Chemical Annotation » modules take as input either:
  – a list of masses entered manually
  or
  – an input file
Annotation workflow – 1
Preparing your input file (1/3)

- The input file format is TSV (tabulation separated values) or TABULAR. It is simple text file that can be prepared by using Excel and saved using the tabulated type format.
Annotation workflow – 1
Preparing your files (2/3)

- You can then rename your file with the .tsv extension (instead of .txt) by right-clicking on the file (and ignoring the warning):
Annotation workflow – 1
Preparing your input file (3/3)

- Such « .tsv » files (i.e. tabular separated; e.g. variableMetadata.tsv) can be handled correctly both by Excel and Galaxy.

- The input file structure must contain:
  - column with ID of the ion - example : M(mz)T(rt)
  - column with masses of ions
  - Decimal separator must be "."

- Such « .tsv » files (i.e. tabular separated; e.g. variableMetadata.tsv) can be handled correctly both by Excel and Galaxy.
Annotation workflow – 1

Loading your files into Galaxy

- Upload your file (variableMetadata.tsv) either by using the icon and « drag & drop » the file:
Annotation workflow – 1
Check that your data have been uploaded correctly
Annotation workflow – 1

Use « LCMS Annotation » modules

Open the module of your choice and select your input file. Select position of your id and mass columns.

You are now ready to configure your annotation tool.
Annotation workflow – 1: HMDB TOOLS

- HMDB input and configuration:

Data input: a metadataVariable file with Mandatory ids and mz

Search parameters: delta and molecular species

Authors: Marion Landi and Franck Giacomoni

Please cite if you use this tool, please cite
Tasks:

- Import « published Histories » called: DEMO_Annotation_hmdb_res

For outputs viewing...
### Annotation workflow – 1: OUTPUT FORMATS

#### Annotation tool results – HTML view

<table>
<thead>
<tr>
<th>ID from input</th>
<th>Mass (m/z)</th>
<th>Compound_1_</th>
<th>Formula</th>
<th>Compound MW (Da)</th>
<th>Adduct</th>
<th>Adduct_Type</th>
<th>Adduct MW (Da)</th>
<th>Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>M99T318</td>
<td>99.04492968</td>
<td>C4H6</td>
<td>54.04695</td>
<td>M+FA-H</td>
<td>-</td>
<td></td>
<td>99.045151</td>
<td>0.00022132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C10H16O4</td>
<td>200.104859</td>
<td>M-2H</td>
<td>-</td>
<td></td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td>M145T349</td>
<td>145.0616508</td>
<td>C4H8N2O</td>
<td>100.063663</td>
<td>M+FA-H</td>
<td>-</td>
<td></td>
<td>145.061864</td>
<td>0.0002132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C4H8N2O</td>
<td>100.063663</td>
<td>M+FA-H</td>
<td>-</td>
<td></td>
<td>145.061864</td>
<td>0.0002132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td></td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2H3O06</td>
<td>438.204239</td>
<td>M-3H</td>
<td>-</td>
<td></td>
<td>145.060804</td>
<td>0.0008468</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2H3O06</td>
<td>438.204239</td>
<td>M-3H</td>
<td>-</td>
<td></td>
<td>145.060804</td>
<td>0.0008468</td>
</tr>
<tr>
<td>M405T464</td>
<td>405.1910066</td>
<td>C22H32O8</td>
<td>424.209718</td>
<td>M-H20-H</td>
<td>-</td>
<td></td>
<td>405.191328</td>
<td>0.0003214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C22H32O8</td>
<td>424.209718</td>
<td>M-H20-H</td>
<td>-</td>
<td></td>
<td>405.191328</td>
<td>0.0003214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td></td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td></td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td></td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td></td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
</tbody>
</table>

Results of HMDB queries - Search params: Molecular specie = negative / delta = 0.001
Annotation workflow – 1: OUTPUT FORMATS

- Annotation tool results – HTML view

Results of HMDB queries - Search params: Molecular specie = negative / delta = 0.001

<table>
<thead>
<tr>
<th>ID from input</th>
<th>Mass (m/z)</th>
<th>Compound_ID</th>
<th>Formula</th>
<th>Compound MW (Da)</th>
<th>Adduct</th>
<th>Adduct_Type</th>
<th>Adduct MW (Da)</th>
<th>Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>M99T318</td>
<td>99.04492968</td>
<td>HMDB41792</td>
<td>C4H6</td>
<td>54.04695</td>
<td>M+FA-H</td>
<td>-</td>
<td>99.045151</td>
<td>0.00022132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB91470</td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB32351</td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB01386</td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB32459</td>
<td>C5H8O2</td>
<td>100.05243</td>
<td>M-H</td>
<td>-</td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB30989</td>
<td>C10H16O4</td>
<td>200.104859</td>
<td>M-2H</td>
<td>-</td>
<td>99.045154</td>
<td>0.00022432</td>
</tr>
<tr>
<td>M145T349</td>
<td>145.0616508</td>
<td>HMDB33952</td>
<td>C4H8N2O</td>
<td>100.063663</td>
<td>M+FA-H</td>
<td>-</td>
<td>145.061864</td>
<td>0.0002132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB31642</td>
<td>C4H8N2O</td>
<td>100.063663</td>
<td>M+FA-H</td>
<td>-</td>
<td>145.061864</td>
<td>0.0002132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB02031</td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB00641</td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB06899</td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB28687</td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB03423</td>
<td>C5H10N2O3</td>
<td>146.069142</td>
<td>M-H</td>
<td>-</td>
<td>145.061866</td>
<td>0.0002152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB40613</td>
<td>C26H30O6</td>
<td>438.204239</td>
<td>M-3H</td>
<td>-</td>
<td>145.060804</td>
<td>0.0008468</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB36597</td>
<td>C26H30O6</td>
<td>438.204239</td>
<td>M-3H</td>
<td>-</td>
<td>145.060804</td>
<td>0.0008468</td>
</tr>
<tr>
<td>M405T464</td>
<td>405.1910066</td>
<td>HMDB39013</td>
<td>C22H32O8</td>
<td>424.209718</td>
<td>M-H2O-H</td>
<td>-</td>
<td>405.191328</td>
<td>0.0003214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB34492</td>
<td>C22H32O8</td>
<td>424.209718</td>
<td>M-H2O-H</td>
<td>-</td>
<td>405.191328</td>
<td>0.0003214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB40795</td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB12613</td>
<td>C21H28O5</td>
<td>360.193674</td>
<td>M+FA-H</td>
<td>-</td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB41795</td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMDB36901</td>
<td>C20H26O5</td>
<td>346.178024</td>
<td>M+Hac-H</td>
<td>-</td>
<td>405.191875</td>
<td>0.0008684</td>
</tr>
</tbody>
</table>
Annotation workflow – 1: OUTPUT

**FORMATS**

- Annotation tool results – HTML links
Annotation workflow – 1: OUTPUT FORMATS

- TSV output format

<table>
<thead>
<tr>
<th>DELTA</th>
<th>COMPOUND_NAME</th>
<th>HMDB_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>M97T51</td>
<td>47</td>
<td>0.003604744502:</td>
</tr>
<tr>
<td>M99T51</td>
<td>52</td>
<td>0.003605255498:</td>
</tr>
<tr>
<td>M138T54</td>
<td>179</td>
<td>0.0002606:</td>
</tr>
<tr>
<td>M139T54</td>
<td>52</td>
<td>0.0001435:</td>
</tr>
<tr>
<td>M187T53</td>
<td>487</td>
<td>0.0007276:</td>
</tr>
<tr>
<td>M190T53</td>
<td>504</td>
<td>0.0002232:</td>
</tr>
<tr>
<td>M256T55</td>
<td>1076</td>
<td>0.0002297:</td>
</tr>
<tr>
<td>M111T64</td>
<td>83</td>
<td>0.001965:</td>
</tr>
<tr>
<td>M99T318</td>
<td>53</td>
<td>0.000672:</td>
</tr>
<tr>
<td>M145T318</td>
<td>223</td>
<td>0.0002213:</td>
</tr>
<tr>
<td>M405T464</td>
<td>2027</td>
<td>0.0003214:</td>
</tr>
<tr>
<td>M132T56</td>
<td>168</td>
<td>0.0001569:</td>
</tr>
<tr>
<td>M268T349</td>
<td>1138</td>
<td>0.0000378:</td>
</tr>
<tr>
<td>M283T65</td>
<td>1335</td>
<td>0.000072:</td>
</tr>
<tr>
<td>M264T349</td>
<td>1143</td>
<td>0.000072:</td>
</tr>
</tbody>
</table>

"No_result_found_on_HMDB"

<< METLIN FORMAT LIKE >>

0.0002606:|(C3H6O3)::HMDB01882 0.0002606:|(C3H6O3)

0.003604744502:|(Clitocine)::HMDB33718
0.003605255498:|(Citrusinine II)::HMDB30373
Annotation workflow – 1: HR2 tool

- **HR2 input and configuration:**

  - **Data input:** a metadataVariable file with mandatory ids and mz
  
  - **Search parameters:** delta, molecular species and molecule initial charge

---

**Authors:** Marion Landi, marion.landl@clermont.inra.fr and Franck Giacomoni, franck.giacomoni@clermont.inra.fr

HR2 original program and its documentation are Copyright (c) 1992-2005 by Joerg Heu under GNU General Public License ("GPL").

**Please cite:** If you use this tool, please cite Tobias Kind and Oliver Piehoh, (2007), "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry," BMC Bioinformatics 8:105.
Each module proposes the same input and output formats, and the same type of parameters:

- **The common input and output formats** allow you to run multiple annotation tools sequentially or in parallel.
- **Commons outputs:**
  - HTML view: includes hyperlinks; multiple annotation results for the same variable are split into consecutive rows; next releases will provide more interactivity with the results.
  - .tsv format:
Parallel annotation workflow:
- Combine chemical compounds libraries
- and a spectral repository.
● **KEGG Compounds:**
  ● Sub-bank of KEGG containing small molecules, biopolymers, and other chemical substances, relevant to biological systems. V2014-05 with 17,254 entries.

● **MassBank:**
  ● Public repository of mass spectra of small chemical compounds for life sciences. V2014-03 with 40,889 spectra for 15,775 compounds.

● **ChemSpider**
  ● Meta engine providing text and structure search to over 30 millions structures from ~500 data sources.
**Annotation workflow – 2: ADJUST data**

- **Manage positive / negative mode:**
  - Some tools are not mass spectrometry oriented (ex: KEGG)
  - Possible to adjust mass with COMMON TOOLS / Compute

- positive mode: C3 - 1.007825 (Proton) + 0.0005486 (electron)
- negative mode: C3 + 1.007825 (Proton) - 0.0005486 (electron)

<table>
<thead>
<tr>
<th>Mzmed [M+H]</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>132.0300721</td>
<td>131.0227957</td>
</tr>
<tr>
<td>263.1031444</td>
<td>262.095868</td>
</tr>
<tr>
<td>288.9865308</td>
<td>287.9792544</td>
</tr>
<tr>
<td>264.106299</td>
<td>263.0990226</td>
</tr>
</tbody>
</table>
Tasks:

- Import « published Histories » called: DEMO_Annotation_ kegg_res
Annotation

MORE OPTIONS
You can use the multi-view tool of W4M and visualize your annotation files simultaneously.
An annotation with in-house Database

### bank_inhouse search by masses on local bank (Galaxy Tool Version 1.0.0)

**Would you use a file**

- **YES**

  If 'NO' is selected then one or more mass(es) must be entered manually.

**FILE OF MASSES (format: tabular)**

- 12: variableMetadata_NEG_06RT.tsv

  Generally variable metadata file

**Do you have a header**

- **YES**

  If 'YES' is selected then enter your number of header lines

**Number of header lines**

- 1

  Number of lines not containing masses

**Column of Id**

- Column: 1

**Would you use retention time (RT)**

- **YES**

**Column of retention time (RT)**

- Column: 2

---

Use your personnal database (as .tsv files)
 ProbMetab : your input DATA

- "ProbMetab" module take as input either:
  - output file from annotatediffreport and
  - xcmsSet xcms object after missing data replacement, to retrieve SNR to isotopic peaks.
Chaining The « Annotation » modules

- The « LCMS Annotation » modules can be chained with « statistical » modules and the Filters module. ProbMetab can be directly connected with XCMS outputs modules.
- Query modules can be run sequentially or in parallel.
### Annotation of GC-MS data in Galaxy

#### 6: peakspectra.msp

A MSP export of pseudospectras readable by NIST or Golm Metabolome

#### 20: GolmResult.html

<table>
<thead>
<tr>
<th>Cname</th>
<th>AnalyteName</th>
<th>Ri</th>
<th>DotProductDistance</th>
<th>EuclideanDistance</th>
<th>HammingDistance</th>
<th>JaccardDistance</th>
<th>S12GowerLegend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown 1</td>
<td>Citric acid (4TMS)</td>
<td>1803.92065</td>
<td>0.01</td>
<td>0.01</td>
<td>15</td>
<td>0.28</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Isocitric acid (4TMS)</td>
<td>1803.39966</td>
<td>0.04</td>
<td>0.01</td>
<td>109</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Gluconic acid (6TMS)</td>
<td>1984.52283</td>
<td>0.11</td>
<td>0.02</td>
<td>85</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Malic acid, 3-oxalo- (1MEOX) (4TMS) BP</td>
<td>2200</td>
<td>0.12</td>
<td>0.02</td>
<td>868</td>
<td>0.96</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Galactaric acid (6TMS)</td>
<td>2030.46814</td>
<td>0.12</td>
<td>0.02</td>
<td>461</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Arabinoseptulosonic acid enol, 3-deoxy- (5TMS) MP</td>
<td>1926.63843</td>
<td>0.11</td>
<td>0.02</td>
<td>499</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Malic acid, 3-oxalo- (1MEOX) (4TMS) MP</td>
<td>2190.4812</td>
<td>0.12</td>
<td>0.02</td>
<td>478</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>myo-Inositol-1-phosphate (7TMS)</td>
<td>2414.291</td>
<td>0.12</td>
<td>0.02</td>
<td>403</td>
<td>0.92</td>
<td>0.85</td>
</tr>
</tbody>
</table>

http://web11.sb-roscoff.fr/download/w4m/howto/w4m_HowToUseNIST_V01.pdf
References