Untargeted metabolomics of full scan MS hyphenated data – mining for minor peaks with Finnee, risks and benefit

Guillaume Erny, Arminda Alves, Faculdade de Engenharia da Universidade do Porto, LEPABE
THE PROJECT – FINNEE FROM SPECTRA TO FORMULAE

Chronic Diseases Research Center
- Prof. Dr. Nuno Neupart
- Dr. Pedro Martins

Instituto de Biologia Experimental e Tecnológica
- Dr. Patricia Gomes-Alves
- Ricardo Gomes

Laboratory for Process Engineering, Environment, Biotechnology and Energy
- Dr. Guillaume Erny
- Dr. Monica Santos
Exhaled breath condensate (EBC) is the exhalate from breath, that has been condensed, typically via cooling using a collection device (commonly to 4 °C or subzero temperatures using a refrigerating device). EBC reflects changes in the respiratory fluid that lines the airways and is an inexpensive, non-invasive tool that has potential for scientific research. Despite its promises, it has not been proven for screening or diagnosing diseases of the lung and other conditions.
We can have a look!
NEED OF AN UNTARGETED ANALYSIS

CLASSICAL WORKFLOW FOR SPECTRAL PROCESSING (xcMS & mzMINE)

1. .mzXML data import
2. Chromatogram building
3. Profile to centroid MS scans
4. Chromatogram deconvolution
5. Feature alignment
6. MS row filter (min. n features)
7. Isotope filter (isotope peaks)
8. Normalization (TIC or Standard)
9. Manual validation
10. Feature .csv export

Profile to centroid MS scans

I can help!

https://xcmsonline.scripps.edu
GARbage IN, GARBAGE OUT (GIGO)
FINNEE WORKFLOW

Original file
- CompassXport
- msConvert (ProteoWizard)
- Proprietary software

mzML format

Matlab Raw dataset

Scan

Savitzky-Golay and data

Selection of profiles and baseline correction

Background noise at each m/z

S/N normalisation and noise removal

Centroid scans
Export as mzML

Pure Ions

(>300,000 m/z interval)

Intersection axis (1,000 points)

1ST IBERIAN MEETING IN SEPARATION SCIENCES & MASS SPECTROMETRY,
October 8, 2019
SELECTION OF INDIVIDUAL mz INCREMENT TO CORRECT
Baseline Drift Correction, Noise Estimation and Noise Removal

Peak to peak noise

Mass / m/z

Intensity / counts

1ST IBERIAN MEETING IN SEPARATION SCIENCES & MASS SPECTROMETRY, October 8, 2019
RESULTING TOTAL ION PROFILES

Data file size: ~ 300,000 KB

Data file size: ~ 50,000 KB
PEAK MINING
ALIGMENT AND VALIDATION WITH QC SAMPLES

**QC Samples**: The quality control (QC) sample should qualitatively and quantitatively representative the entire collection of samples included in the study, providing an average of all of the metabolomes analysed in the study. The QC samples are analysed intermittently for the duration of the analytical study to assess the variance observed in the data throughout the sample preparation, data acquisition and data processing step.

1. Peaks present in all QC samples
2. Peak whose RSD (peak Area) is below 20 %
## RESULTS

**xcms:**  [https://xcmsonline.scripps.edu/](https://xcmsonline.scripps.edu/)

**msDial:**  [http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/](http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/)

<table>
<thead>
<tr>
<th></th>
<th>Peaks per dts</th>
<th>Aligned peaks(^1)</th>
<th>Aligned peaks(^2)</th>
<th>Aligned peaks(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finnee</strong></td>
<td>5000</td>
<td>4610</td>
<td>3627</td>
<td>1773</td>
</tr>
<tr>
<td><strong>Xcms</strong></td>
<td>8000</td>
<td>4132</td>
<td>4123</td>
<td>1587</td>
</tr>
<tr>
<td><strong>msDial</strong></td>
<td>&gt;20,000</td>
<td>19859</td>
<td>16972</td>
<td>1454</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Finnee</th>
<th>xcms</th>
<th>msDial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finnee</strong></td>
<td>438</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>xcms</strong></td>
<td>0</td>
<td>273</td>
<td>0</td>
</tr>
<tr>
<td><strong>msDial</strong></td>
<td>0</td>
<td>0</td>
<td>564</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Finnee^</th>
<th>xcms^</th>
<th>msDial^</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finnee</strong></td>
<td>574</td>
<td>574</td>
<td>0</td>
</tr>
<tr>
<td><strong>xcms</strong></td>
<td>150</td>
<td>0</td>
<td>150</td>
</tr>
<tr>
<td><strong>msDial</strong></td>
<td>0</td>
<td>129</td>
<td>129</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finnee</strong></td>
<td>611</td>
</tr>
<tr>
<td><strong>xcms</strong></td>
<td>611</td>
</tr>
<tr>
<td><strong>msDial</strong></td>
<td>611</td>
</tr>
</tbody>
</table>

1. Peaks with at least 9 match out of 11
2. Peaks with a least 11 match out of 11
3. RSD of the area below 20%
FINNEE: AN OPEN-SOURCE MATLAB® TOOLBOX

RELEASE V2.0 ALPHA

This new release introduces the mz master axis to improve the correction of the profile MS scans. The wiki has not been updated yet, more information is available on the quick start_2.pdf. If you are interested in this work and would like to participate, I am looking for beta testers and help for the wiki.

Posted in Finnee2016. Programming on July 18, 2017. Leave a comment - Edit

RELEASE V1.0 ALPHA

Pure ion profiles can now be obtained from a centroid dataset. The pure ion profiles are stored in a PeakList object with figures of merit. PeakList can also be used for the centergram and Clusters plot representation. More information can be found on the wiki or in the two screenshots below.
ACKNOWLEDGEMENTS

This work was financially supported by the projects:

(i) project UID/EQU/00511/2019 - Laboratory for Process Engineering, Environment, Biotechnology and Energy – LEPABE funded by national funds through FCT/MCTES (PIDDAC);

(ii) Project POCI-01-0145-FEDER-029702, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES;

(iii) Project “LEPABE-2-ECO-INNOVATION” – NORTE-01-0145-FEDER-000005, funded by Norte Portugal Regional Operational Programme (NORTE 2020), under PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).