Help in the identification of compounds

Daniel Jacob

INRA UMR 1332 BFP, Metabolomics Facility – 10 JS RFMS– May 2016
http://nmrprocflow.org/
A clustering-based workflow applied on the NMR spectra buckets (NMRProcFlow)

http://biostatflow.org/?session=S27974
Cluster Analysis: Highlighting of latent variables

Optimized in order to

Cuttree = 0.200, Nb Clust = 56, Nb Vars = 226, Clust Max Size = 21

Maximize

Minimize

Metabolomics Facility – 10 JS RFMS– May 2016
Highlighting biomarkers

- By choosing a good “cut threshold”, clusters link mainly the buckets that have a *between-groups* variance,
- Hoping that these "groups" corresponds to factor levels.
Clusters mainly located at the periphery of a circle => biomarkers are highlighted

- The "between-groups" variance within clusters seems mostly correspond to this factor

$Q^2 = 0.949$
The "between-groups" variance within clusters seems NOT (or very slightly) correspond to this factor.
Clusters cover the whole range of intensities (here we have 4 levels of magnitudes)
Export the Cluster file

BioStatFlow

Dataset
Workflow
Results
New Session
Save Session

previous step
Exports:
CSV
Association
0_dataset
0_factors
Scaling
Clustering of Variables
HCA_CLUSTERS
HCA_Samples
HCA_Variables
MDS_Clusters
MDS_Samples
MDS_Variables
Stats_Clustering
boclusters
clusters

Download

p6_MDS_Clusters.txt

<table>
<thead>
<tr>
<th>VAR</th>
<th>CLID</th>
<th>PPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>B9_1277</td>
<td>C1</td>
<td>9.1272</td>
</tr>
<tr>
<td>B4_4425</td>
<td>C1</td>
<td>4.4425</td>
</tr>
<tr>
<td>B3_5608</td>
<td>C2</td>
<td>5.6080</td>
</tr>
<tr>
<td>B9_8215</td>
<td>C2</td>
<td>8.2150</td>
</tr>
<tr>
<td>B6_3419</td>
<td>C2</td>
<td>6.3419</td>
</tr>
<tr>
<td>B7_0035</td>
<td>C3</td>
<td>7.0035</td>
</tr>
<tr>
<td>B7_0615</td>
<td>C3</td>
<td>7.0615</td>
</tr>
<tr>
<td>B7_2285</td>
<td>C3</td>
<td>7.2285</td>
</tr>
<tr>
<td>B7_1505</td>
<td>C3</td>
<td>7.1505</td>
</tr>
<tr>
<td>B7_3615</td>
<td>C3</td>
<td>7.3615</td>
</tr>
<tr>
<td>B8_5040</td>
<td>C3</td>
<td>5.0400</td>
</tr>
<tr>
<td>B8_4400</td>
<td>C3</td>
<td>4.4000</td>
</tr>
<tr>
<td>B7_4071</td>
<td>C3</td>
<td>7.4071</td>
</tr>
<tr>
<td>B7_4582</td>
<td>C4</td>
<td>7.4582</td>
</tr>
<tr>
<td>B7_4582</td>
<td>C4</td>
<td>7.4582</td>
</tr>
<tr>
<td>B7_4582</td>
<td>C4</td>
<td>7.4582</td>
</tr>
</tbody>
</table>

INRA UMR 1332 BFP, Metabolomics Facility – 10 JS RFMS– May 2016
ANOVA-PCA

C3
From the cluster file, copy-paste the corresponding PPM values to C3 cluster.
chlorogenic acid; 6.0; proton-500MHz;

https://peakforest.org/PFs000478
C35

21 buckets: { 2.04 – 2.84 }

Glutamic acid; 6.0; proton-500MHz;

https://peakforest.org/PFs000578

INRA UMR 1332 BFP, Metabolomics Facility
To summarize ...

Manual Peak Matching

https://peakforest.org/
Towards Automatic Peak Matching
Automatic Peak Matching

Hierarchical Cluster Analysis (HCA)

ANOVA-PCA

Clustering

Apply a “cut-tree”

Optimal Cut Height

NB Clusters

NB Variables within a cluster

Size of the biggest Clusters

CutHeight

1 cluster ↔ 1 peak list

https://peakforest.org/