Lipidomics

MassHunter Pre-processing script

This script takes .csv files from MassHunter and prepares multiple .csv files as output.

Input requirements

Columns

As a minimum, retention time (RT) and Area columns both must be present for each compound. These and the Data File (filename) columns are the only ones that are used; other columns may be present but will be discarded.

File names

- Single entry for 'Standard', labelled with 'std' in the file name
  e.g. 'Mix-std-12.d' and 'Std.98.d' are acceptable
- Group names must be the first part of the file name, followed by an underscore
  e.g. 'Flower_1.d' and 'leaf.upper_24.d' will become part of groups 'Flower' and 'leaf.upper' respectively.

Compound names

- Abbreviated compound class followed by '('
  e.g. CE (), PG(34:1), LPC(std) are acceptable, as whitespace is removed automatically
- Internal standards must have 'IS' in the compound name. Case is important:
  If a compound's name has 'is' in it, for example 'h**is**tidinyl choline', it will not be categorised as an internal standard. However, HIStidinyl choline will – the this will need to be edited, otherwise it will be used *only* as an internal standard, and won't show up in the results
- The compound that will be used for normalisation must also have IS in the compound name, with the above stipulations.
- Compound names must be unique.

Function

The output file from MassHunter is a remarkably ugly .csv file. Provided the requirements above are satisfied, this script will take the file name column and use that as the basis for generating the sample and group names, and then split out the retention time and area columns for each compound to separate .csv files.

The internal standards are then used to calculate concentrations for each compound by dividing it by the internal standard for that class of compound. This data is then saved as a .csv file. A separate .csv file contains the average and standard error of the mean for the groups, and another file has the sum of the means for each class of compound for each group.

For example, if you had the following matrix with 8 samples in 2 groups (flower and leaf), with two compound classes (PG and PC):

Sample	Group	CE(17:0)/IS	CE(21:0)	CE(24:6)	CE(25:4)
flower_1	flower	14931.34	12790.79	281425.69	0.00
flower_2	flower	13614.70	11873.99	331619.55	2810.99
flower_3	flower	15464.46	16288.37	43352.18	2544.43
flower_4	flower	7891.76	17209.21	40726.98	0.00
leaf_1	leaf	5904.68	9531.50	490888.53	1635.70
leaf_2	leaf	4983.84	23206.79	428420.84	3352.18
leaf_3	leaf	6712.44	13873.18	332677.71	0.00
leaf_4	leaf	5492.73	29297.25	258263.33	0.00

Table 1: Concentrations of lipids in samples.

This would then become:

Group	DataType	CE(17:0)/IS	CE(21:0)	CE(24:6)	CE(25:4)
flower	Mean	5773.42	18977.18	377562.60	1246.97
flower	SEM	365.29	4469.01	51377.93	800.67
leaf	Mean	86870.96	82441.44	130418.01	9399.23
leaf	SEM	3672.17	15470.52	15107.89	1646.03

Table 2: Average of concentrations (with standard error) for each group.

with the means and standard errors for the two groups, which is then further processed to:

Group	DataType	CE
flower	Mean	403560.18
flower	SEM	51579.43
leaf	Mean	309129.64
leaf	SEM	21994.99

Table 3: Sum of average concentrations in each class of compound..

showing the sum for each of the groups. This is also the case with the normalised values, where Table 1 above is divided by the normalisation factor (the response factor for the specified reference compound), and then the same method is applied to generate the averages and sum of averages.

Usage

The script asks for the value for the standard concentration (usually 5000nM), whether the internal standards are to be kept in the output matrices and whether to perform normalisation. The script is called by typing:

```
> source(“mhproc.r”)
```

and following the prompts in the console window.