Lipidomics

MassHunter Pre-processing script

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

Input requirements

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 (whitespace is removed automagically)
- Internal standards must have 'IS' in the compound name
  If a compound's name has 'is' in it, for example '**his**tidinyl choline', 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 be labelled with “d” and a number (i.e. the number of deuterated components) before the compound class (e.g. d4LPC, d9.PC, d6_PE are all acceptable).
  - If it isn't a deuterated compound, labelling it with a pretend “d4” or similar at the start (“d” and any number) will cause it to be treated as the compound to use for normalisation.

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):

Group	CE(17:0)/IS	CE(21:0)	CE(24:6)	CE(25:4)	PG(15:0)	PG(16:0)	PG(17:0)	PG(34:1)/IS
Flower	14931.34	12790.79	281425.69	0	1115.49	2455.79	554.92	16.87
Flower	13614.7	11873.99	331619.55	2810.99	958.83	2733.26	476.71	0
Flower	15464.46	16288.37	43352.18	2544.43	1187.39	3901.16	589.52	0
Flower	7891.76	17209.21	40726.98	0	1103.31	4334.2	460.6	0
Leaf	5904.68	9531.5	490888.53	1635.7	526.51	3041.75	455.6	18.09
Leaf	4983.84	23206.79	428420.84	3352.18	628.41	2017.29	369.62	0
Leaf	6712.44	13873.18	332677.71	0	762.29	1962.85	532.46	0
Leaf	5492.73	29297.25	258263.33	0	600.39	3131.61	212.05	0

Table 1: Concentrations of lipids in samples.

This would then become:

Group	StatType	CE(17:0)/IS	CE(21:0)	CE(24:6)	CE(25:4)	PG(15:0)	PG(16:0)	PG(17:0)	PG(34:1)/IS
Flower	Mean	5773.425	18977.181	377562.601	1246.971	629.398	2538.376	392.432	4.522
Flower	SEM	365.286	4469.015	51377.927	800.671	49.235	317.289	68.711	4.522
Leaf	Mean	86870.961	82441.438	130418.013	9399.233	903.409	22385.504	635.803	199.763
Leaf	SEM	3672.166	15470.517	15107.889	1646.032	58.945	1290.825	29.936	48.789

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

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

Group	StatType	CE	PG
Flower	Mean	403560.178	3564.728
Flower	SEM	51579.434	328.387
Leaf	Mean	309129.645	24124.479
Leaf	SEM	21994.987	1293.437

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 assumes that the value for the standard concentration is 5000nmol, the internal standards are to be kept in the output matrices and that there will be no normalisation. If this is the case, the script can be called by typing:

```
> source("mhproc.r")
> mhproc("filename.csv")
```

This is equivalent to:

```
> mhproc("filename.csv",std_conc=5000,keep_IS=TRUE,norm=FALSE,norm_cpd="")
```

If you would like to change the defaults, you can replace mhproc() with any of the following:

Don't keep the internal standards in the output matrices:

```
> mhproc("filename.csv",keep_IS=FALSE)
```

Change the standard concentration:

```
> mhproc("filename.csv",std_conc=2500)
```

Do perform normalisation, using d1LPC(15:0)/IS as the internal standard:

```
> mhproc("filename.csv",norm=TRUE,norm_cpd="d1LPC")
```

Change all defaults:

```
> mhproc("filename.csv",std_conc=2500,keep_IS=FALSE,norm=TRUE,norm_cpd="d7PE")
```

Note: The name for the norm_cpd is the compound *class*, not the whole name.

Further

To incorporate all the output .csv files into a single .xls file, use the import_multi_csv macro. This is available at http://code.google.com/p/ma-bioinformatics/ under Excel-macros

Note: Macro only available for Microsoft Excel.