Package 'SASPECT'

February 19, 2015

Version 0.1-1

Date 2008-06-02

Title Significant AnalysiS of PEptide CounTs.

Author Pei Wang <pwang@fhcrc.org>, Yan Liu <yanliu@fhcrc.org>

Maintainer Pei Wang <pwang@fhcrc.org>

Description A statistical method for significant analysis of comparative proteomics based on LC-MS/MS Experiments

License GPL (>= 2)

Repository CRAN

Date/Publication 2008-06-23 07:54:00

NeedsCompilation no

R topics documented:

`mouseTissue` *LC MS/MS Data from Mouse Model*

Description

An Example data for package `SASPECT`

Usage

```
data(mouseTissue)
```

Details

This data example is from the mouse study described in Whiteaker et. al. 2007. It contains the information of 333 peptides from 20 LC-MS/MS experiments (10 from the normal group and 10 from the control group).

Value

`mouseTissue` is a list of four components:

`peptideData`	a list of two numeric matrices, `PeptideCount` and PeptideConfidence. The columns correspond to 20 LC-MS/MS experiments, and rows correspond to 333 peptides. `PeptideCount` records the total spectral counts of each peptide in each experiment. `PeptideConfidence` tracks the highest PeptideProphet score of each peptide identification in each experiment in the database search procedure.
`pep.set`	a character vector of length 333, recording the peptide IDs.
`pep.pro.name`	a character matrix consisting of 15579 rows and 2 columns. The first column is a vector of mouse protein IDs (IPI numbers), while the second column gives the names of the peptides matching to the mouse proteins in the first column.
`run.group.info`	a data frame consiting of 2 rows and 2 columns, which indicates the case status and the sample size of each group.

References

Whiteaker, J. R., Zhang, H., Zhao, L., Wang, P., Kelly-Spratt, K. S., Ivey, R. G., Piening, B. D., Feng, L., Kasarda, E., Gurley, K. E., Eng, J. K., Chodosh, L. A., Kemp, C. J., McIntosh, M. W., Paulovich, A. G (2007) Integrated Pipeline for Mass Spectrometry-Based Discovery and Confirmation of Biomarkers Demonstrated in a Mouse Model of Breast Cancer. J. Proteome Res., 6(10); 3962-3975.

Wang, P., Liu, Y., McIntosh, M. W., Paulovich, A. G (2008) Significant analysis for comparative proteomics studies using label free LC-MS/MS experiments.

SASPECT *Significant AnalysiS of PEptide CounTs*

Description

A function for identifying differentially expressed proteins between two sample groups using spectral counts from LC-MS/MS Experiments

Usage

```
SASPECT(peptideData, pep.set, pep.pro.name, run.group.info,
        permu.iter=50, filter.run=2, filter.score=0.95)
```

Arguments

peptideData	a list of two components: PeptideCount and PeptideConfidence. Both are numeric matrices with p rows each representing one peptide and n1+n2 columns each representing one sample (n1=sample size of the first group, and n2=sample size of the second group). PeptideCount records the peptide spectral counts of all p peptides in all n1+n2 samples. PeptideConfidence tracks the confidence score of each peptide identification in the database search procedure (e.g. the PeptideProphet score). Both matrics need to be arranged in the way that the first n1 columns represents samples from the first group and the rest columns are for the second group.
pep.set	a character vector of length p. The ith element is the peptide ID corresponding to the ith row of peptideData$PeptideCount and peptideData$PeptideConfidence.
pep.pro.name	a character matrix with 2 columns. The first column gives the protein IDs, and the second column gives the names of the peptides matching to the proteins in the first column.
run.group.info	a data frame with two columns. The first column (run.group.info$label) is a character vector of length 2, giving the group names of the two groups. The second column (run.group.info$count) is a numeric vector of length 2, giving the number of samples in the first group (n1) and the second group (n2).
permu.iter	an integer. It is the number of permutation iterations for estimating FDR. The default value is 50.
filter.run	an integer. It is the filter criteria for removing peptides observed in too few samples. The default value is 2.
filter.score	a scale. PeptideConfidence scores above this value are counted in the filtering process. The default value is 0.95

Details

This function implements the SASPECT-hybrid method (Wang et. al. 2008, in preparation), which is a modified version of the original SASPECT mothod proposed in Whiteaker et. al. 2007. The Score1 column in the returned matrix gives test statistics using the original SASPECT method.

Value

SASPECT generates a data frame with 7 columns:

Protein	Protein groups' ID.
ProteinsInGroup	Names of proteins in each protein group (separated by .).
Score1	test score based on Appear-Absent (AA) measurements. A positive value suggests the abundence level in the second group is higher than the first group. A negative value suggests the opposite.
Score2	test score based on non zero total Spectral count (SpecC) measurements. A positive value suggests the abundence level in the second group is higher than the first group. A negative value suggests the opposite.
Score	final SASPECT score (sum square of Score1 and Score2).

Qvalue	FDR resulted from permutation test based on Score.
PeptideNumber	number of peptides observed for each protein(protein group).

Author(s)

Wang, P. and Liu, Y.

References

Whiteaker, J. R., Zhang, H., Zhao, L., Wang, P., Kelly-Spratt, K. S., Ivey, R. G., Piening, B. D., Feng, L., Kasarda, E., Gurley, K. E., Eng, J. K., Chodosh, L. A., Kemp, C. J., McIntosh, M. W., Paulovich, A. G (2007) Integrated Pipeline for Mass Spectrometry-Based Discovery and Confirmation of Biomarkers Demonstrated in a Mouse Model of Breast Cancer. J. Proteome Res., 6(10); 3962-3975.

Wang, P., Liu, Y., McIntosh, M. W., Paulovich, A. G (2008) Significant analysis for comparative proteomics studies using label free LC-MS/MS experiments (in preparation).

Examples

```
library(SASPECT)
data(mouseTissue)

SASPECT.result<-SASPECT(peptideData=mouseTissue$peptideData,
        pep.set=mouseTissue$pep.set,
        pep.pro.name=mouseTissue$pep.pro.name,
        run.group.info=mouseTissue$run.group.info,
        permu.iter=50,
        filter.run=2,
        filter.score=0.95)
### it takes about 1 minute to run this example.

### check the qvalue distribution
qvalue=as.numeric(SASPECT.result[,"Qvalue"])
plot(sort(qvalue))

### output the result into a table file
write.table(SASPECT.result, file="SASPECT.result.txt", row.names=FALSE, sep="\t")
```

Index