

README – MS-EmpiReS

The MS-EmpiReS package detects differential alternative splicing (DAS) in quantitative MS proteomics data. The package currently supports human and mouse data and the default is human.

The **necessary input files** are

1) A **tab-separated** file with quantified peptides in the following format:

peptide	charge	WT_Rep1	WT_Rep2	WT_Rep3	KO_Rep1	KO_Rep2	...
TLYDFPGNDAEDLPFK	4	215282	203230	238270	177317	155632	...
FASWALESDNNTALLL SK	2	165234	147356	185873	269853	292658	...
ATVLLSMSK	2	3170347	3233056	2923335	1276556	1695760	...
LQTLMSVDDSVVER	2	335778	301160	306921	675119	631553	...
HNEETGDNVGPLIK	2	367192	382423	408906	250343	291029	...

A header with the name peptide must be in the table. Adding a column with charge information (name charge) is optional. Alternatively, the charge can also be attached to the sequence with the separator `'_'` (for example `TLYDFPGNDAEDLPFK_2`). Charge information can increase the sensitivity. The names for the columns with the peptide intensities (WT_Rep1, WT_Rep2, ...) can be arbitrarily chosen. In the following, we will name this file **peptides.tsv**

2) A **tab-separated** file with a label mapping in the following format:

WT_Rep1	WT
WT_Rep2	WT
WT_Rep3	WT
KO_Rep1	KO
KO_Rep2	KO
KO_Rep3	KO

On the left side, the sample labels of the quantified peptides (i.e. headers in peptides.tsv) are listed. On the right side, the conditions to which the labels map are given. In the following, we will name this file **labelmap.tsv**

There are also two **optional input files**:

3) In the MS-EmpiReS folder, there is a subfolder **conf**, which contains a file termed **r.config**. In case you want to get plots and visualizations when using MS-EmpiReS, you need an R installation. You also need the R packages 'gplots' and 'cairo'. You then have to write the path to your R installation in your r.config. The config was written for and tested on Unix Systems (i.e. Linux/Mac)

4) A **tab-separated** file, where all conditions that should be compared are listed. For example, there could be a time-series measurement T1,T2,T3,...,Tn and you only want to compare relative to T1. In this case, you can provide the following file:

T1	T2
T1	T3
T1	T4
...	...

In the following, we will name this file **specified_comparisons.tsv**

Necessary Parameters

-peptides_tsv input table with quantified peptide, specified above

-labelmap_tsv sample to condition mappings, specified above

Optional Parameters

-minrep Minimum number of replicates necessary for a peptide to be included. Default is 2

-outdir you can specify an output directory, where tables and potentially plots are written

-plots Set this switch to produce overview plots and on-demand plots from an interactive table. Only works if you have R configured as described above.

-plots_interactive Set this switch if you only want to have on-demand plots and no overview plots

-specified_comparisons /path/to/specified_comparisons.tsv

-mouse MS-EmpireS can currently process data from Mus Musculus and Homo Sapiens. Homo Sapiens is set per default, use this switch if you have Mus Musculus data. In case that you want other organisms, please contact the developers.

-test_all_isoform_pairs As described in the paper, we use equivalence classes to map peptides to isoforms. A more speculative but also more sensitive mapping is to test all available pairs of isoforms. This switch enables pairwise testing.

-show_all_DAS_tests Often it is possible to test multiple equivalence classes for the same gene. Per default, we show a table where the best hit per gene has been automatically selected. In case you want to all browse all tested equivalence classes, select this option.

-analyze_series_of_conditions Set this switch if you want to compare multiple conditions to a single reference (e.g. timecourse). You need to specify the reference using the 'specified_comparisons' switch, where you specify that all conditions are compared to one reference

Example Method Calls

```
java -jar /path/to/MS-EmpiReS.jar -peptides_tsv /path/to/peptides.tsv -labelmap_tsv  
/path/to/labelmap.tsv
```

Minimum call, this produces an output table (see below) with information on gene-level DAS events. The OUTPUT directory is created in the directory, where peptides.tsv is located

```
java -jar /path/to/MS-EmpiReS.jar -peptides_tsv /path/to/peptides.tsv -labelmap_tsv  
/path/to/labelmap.tsv -plots -outdir /path/to/OUTPUT
```

Testing as above with additional detailed plots. r.config must be set.

Visualizations of significant proteins are automatically written and are stored in OUTPUT/EXAMPLES. An overview table OUTPUT/HTML_gene_level/index.html can be opened via a web browser (tested with Firefox)