Package 'MIMP'

September 17, 2014

Type Package

Title Predicting the impact of single nucleotide variants on kinase-substrate phosphorylation

Version 1.0

Date 2013-10-29

Author Omar Wagih

Maintainer Omar Wagih <wagih@ebi.ac.uk>

Description No description

License LGPL

R topics documented:

`mimp` *Predict the impact of single variants on phosphorylation.*

Description

This function takes in mutation, sequence and phosphorylation data to predict the impact the mutation has on phosphorylation.

Usage

```
mimp(muts, seqs, psites, perc.bg = 90, perc.fg = 10, thresh.log2 = 0,
  display.results = T, include.cent = F, family.models = F)
```

Arguments

`muts` Mutation data file: a space delimited text file OR data frame containing two columns (1) gene and (1) mutation. Example:

TP53	R282W
CTNNB1	S33C
CTNNB1	S37F

`seqs` Sequence data file containing protein sequences in FASTA format OR named list of sequences where each list element is the uppercase sequence and the name of each element is that of the protein. Example: list(TP53="ABCXYZ", CDK2="HJKEWR")

`psites` Phosphorylation data file (optional): a space delimited text file OR data frame containing two columns (1) gene and (1) positions of phosphorylation sites. Example:

TP53	280
CTNNB1	29
CTNNB1	44

`perc.bg` Percentile value between 0 - 100. This value is used to compute a threshold, beta from the negative (background) distribution of scores. By default this is the 90th percentile of the background distribution of scores. Anything below the threshold is considered a negative hit.

`perc.fg` Percentile value between 0 - 100. This value is used to compute a threshold, alpha from the positive (foreground) distribution of scores. By default this is the 10th percentile of the foreground distribution of scores. Anything above the threshold is considered a positive hit.

`thresh.log2` Threshold for the absolute value of log ratio. Anything less than this value is discarded (default: 0).

`include.cent` If TRUE, gains and losses caused by mutation in the central STY residue are kept. Scores of peptides with a non-STY central residue is given a score of -1 (default: FALSE).

`family.models` By default, individual kinase specificity models used to scan for rewiring events. Set to TRUE If you would like to use specificity models of the kinase families (default: FALSE).

Value

The data is returned in a `data.frame` with the following columns:

`gene` gene with the rewiring event

`mut` mutation causing the rewiring event

`psite_pos` position of the central residue of the phosphosite

`mut_dist` distance of the mutation from the central phosphosite

`wt` sequence of the wildtype phosphosite (before the mutation)

`mt` sequence of the mutated phosphosite (after the mutation)

`score_wt` matrix similarity score of the wildtype phosphosite

`score_mt` matrix similarity score of the mutated phosphosite

`log_ratio` Log2 ratio between mutant and wildtype scores. A high positive log ratio represents a high confidence gain-of-signaling event. A high negative log ratio represents a high confidence loss-of-signaling event.

`pwm` name of the kinase being rewiried

`perc_wt` Percentile rank of the wt score

`perc_mt` Percentile rank of the mutant score

Examples

```
# Get the path to example mutation data
mut.file = system.file("extdata", "mutation_data.txt", package = "MIMP")

# Get the path to example FASTA sequence data
seq.file = system.file("extdata", "sequence_data.txt", package = "MIMP")

# View the files in a text editor
browseURL(mut.file)
browseURL(seq.file)

# Run rewiring analysis
results = mimp(mut.file, seq.file, display.results=TRUE)

# Show head of results
head(results)
```

`results2html` *Display MIMP results interactively in browser*

Description

Display MIMP results interactively in browser

Usage

```
results2html(x, max.rows = 5000)
```

Arguments

`x`	Data frame resulting from mimp call.
`max.rows`	If data contains more rows than this value, results won't be displayed.