Package ‘echo.find’

June 20, 2019

Type Package
Title Finding Rhythms Using Extended Circadian Harmonic Oscillators (ECHO)
Version 3.0
Description Provides a function (echo_find()) designed to find rhythms from data using extended harmonic oscillators. For more information, see H. De los Santos et al. (2017) <doi:10.1145/3107411.3107420>.
License MIT + file LICENSE
Encoding UTF-8
LazyData true
Imports minpack.lm (>= 1.2.1), boot (>= 1.3-22)
URL https://github.com/delosh653/ECHO
RoxygenNote 6.1.1
Suggests knitr, rmarkdown, ggplot2
VignetteBuilder knitr
NeedsCompilation no
Author Hannah De los Santos [aut, cre], Emily Collins [aut], Kristin Bennett [aut], Jennifer Hurley [aut]
Maintainer Hannah De los Santos <delosh@rpi.edu>
Repository CRAN
Date/Publication 2019-06-20 19:50:03 UTC

R topics documented:
echo.find ................................................................. 2
echo_find ............................................................... 2
expressions .............................................................. 4

Index 5
### echo_find

**Function to calculate the results for all genes using the extended circadian harmonic oscillator (ECHO) method.**

#### Description

To read more about our initial work on this project and cite us, see *Circadian Rhythms in Neurospora Exhibit Biologically Relevant Driven and Damped Harmonic Oscillations* by H. De los Santos et al. (2017)

#### Usage

```r
echo_find(genes, begin, end, resol, num_reps, low = 1, high = 2,
          run_all_per, paired, rem_unexpr, rem_unexpr_amt = 70,
          rem_unexpr_amt_below = 0, is_normal, is_de_linear_trend, is_smooth,
          run_conf = F, which_conf = "Bootstrap", harm_cut = 0.03,
          over_cut = 0.15, seed = 30)
```

#### Arguments

- **genes**: data frame of genes with the following specifications: first row is column labels, first column has gene labels/names, and all other columns have expression data. This expression data must be ordered by time point then by replicate, and must have evenly spaced time points. Any missing data must have cells left blank.
- **begin**: first time point for dataset
- **end**: last time point for dataset
- **resol**: resolution of time points
- **num_reps**: number of replicates
- **low**: lower limit when looking for rhythms, in hours. May be unused if finding rhythms of any length within timecourse (run_all_per is TRUE).
- **high**: upper limit when looking for rhythms, in hours. May be unused if finding rhythms of any length within timecourse (run_all_per is TRUE).
- **run_all_per**: boolean which indicates whether or not rhythms of any length within timecourse should be searched for.
paired if replicate data, whether the replicates are related (paired) or not (unpaired)
rem_unexpr boolean indicating whether genes with less than rem_unexpr_amt percent expression should not be considered
rem_unexpr_amt percentage of expression for which genes should not be considered if rem_unexpr is TRUE
rem_unexpr_amt_below cutoff for expression
is_normal boolean that indicates whether data should be normalized or not
is_de_linear_trend boolean that indicates whether linear trends should be removed from data or not
is_smooth boolean that indicates whether data should be smoothed or not
run_conf boolean of whether or not to run confidence intervals
which_conf string of which type of confidence interval to compute ("Bootstrap" or "Jackknife")
harm_cut positive number indicating the cutoff for a gene to be considered harmonic
over_cut positive number indicating the cutoff for a gene to be considered repressed/overexpressed
seed number for random seed to fix for bootstrapping for confidence intervals

Value
results, a data frame which contains:
Gene Name gene name
Convergence did the fit converge, or descriptor of type of data (constant, unexpressed, etc.)
Iterations number of iterations
Amplitude.Change.Coefficient Amplitude change coefficient value for fit
Oscillation Type Type of oscillation (damped, driven, etc.)
Initial Amplitude Initial amplitude value for fit
Radian.Frequency Radian frequency for fit
Period Period for fit (in time units)
Phase Shift Phase shift for fit (radians)
Hours Shifted Phase shift for fit (hours)
Equilibrium Value Equilibrium shift for fit
Slope Slope value of original data, if linear baseline is removed
Tau Kendall’s tau between original and fitted values
P-value P-value calculated based on Kendall’s tau
BH Adj P-Value Benjamini-Hochberg adjusted p-values
expressions

BY Adj P-Value    Benjamini-Yekutieli adjusted p-values
CI.PARAM.Low      Lower confidence interval bound for all parameters, if calculated
CI.PARAM.High     Higher confidence interval bound for all parameters, if calculated
Original TPX.Y    Processed values for gene expression at time point X, replicate Y
Fitted TPX        Fitted values for gene expression at time point X

Examples

# for more elaboration, please see the vignette
# "expressions" is the example echo.find data frame
# long example - commented out
echo_find(genes = expressions, begin = 2, end = 48, resol = 2,
         num_reps = 3, low = 26, high = 26, run_all_per = FALSE,
         paired = FALSE, rem_unexpr = FALSE, rem_unexpr_amt = 70, rem_unexpr_amt_below=0,
         is_normal = FALSE, is_de_linear_trend = FALSE, is_smooth = FALSE)

expressions

Synthetic expression data for 12 genes.

Description

A dataset containing the names and expression values for 12 synthetically generated samples. This example data has time points from 2 to 48 hours with 2 hour resolution and 3 replicates. Random missing data is also included. Synthetic data was created by randomly selecting parameters for the extended harmonic oscillator equation (see journal paper link in vignette for the equation), then adding random uniform noise to each expression.

Usage

expressions

Format

A data frame with 12 rows and 73 variables (column 1: sample labels, columns to 2 to 73: numerical values for gene expression in the format CTX.Y (time point X, replicate Y)).

Details

Note the data format: its first column has gene labels/names, and all other columns have expression data. This expression data is ordered by time point then by replicate, and has evenly spaced time points. Any missing data has cells left blank.
Index

*Topic datasets
  expressions, 4

echo.find, 2
echo.find-package (echo.find), 2
echo_find, 2
expressions, 4