

Package ‘MetaCycle’

December 5, 2015

Type Package

Title Evaluate Periodicity in Large Scale Data

Version 1.1.0

Maintainer Gang Wu <wggucas@gmail.com>

Description Provides two functions-meta2d and meta3d for detecting rhythmic signals from time-series datasets. For analyzing time-series datasets without individual information, 'meta2d' is suggested, which could incorporate multiple methods from ARSER, JTK_CYCLE and Lomb-Scargle in the detection of interested rhythms. For analyzing time-series datasets with individual information, 'meta3d' is suggested, which takes use of any one of these three methods to analyze time-series data individual by individual and gives out integrated values based on analysis result of each individual.

Depends R (>= 3.0.2)

Imports gnm

Suggests knitr, rmarkdown, diagram, shape

VignetteBuilder knitr

License GPL (>= 2)

LazyData true

Author Gang Wu [aut, cre],
Ron Anafi [aut, ctb],
John Hogenesch [aut, ctb],
Michael Hughes [aut, ctb],
Karl Kornacker [aut, ctb]

Repository CRAN

NeedsCompilation no

Date/Publication 2015-12-05 00:08:22

R topics documented:

cycHumanBloodData 2

cycHumanBloodDesign	2
cycMouseLiverProtein	3
cycMouseLiverRNA	3
cycSimu4h2d	4
cycVignettesAMP	4
cycYeastCycle	5
meta2d	6
meta3d	9
MetaCycle	13

Index 14

cycHumanBloodData *cycHumanBloodData*

Description

This data set lists time-series profiles of 10 transcripts sampled from multiple individuals under different sleep conditions.

Usage

cycHumanBloodData

Format

A dataframe containing 439 columns (column 1 = transcript name, column 2 to 439 = samples from individuals at different time points and sleep conditions).

Source

Moller-Levet C. S., et al. (2013). Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci U S A*, **110**(12), E1132–1141.

cycHumanBloodDesign *cycHumanBloodDesign*

Description

This data set describes individual information, sleep condition and sampling time corresponding to each sample in 'cycHumanBloodData'.

Usage

cycHumanBloodDesign

Format

A dataframe containing 4 columns described as below:

[,1]	sample_library	character	sample ID
[,2]	subject	character	individual ID
[,3]	group	character	sleep condition
[,4]	time_hoursawake	numeric	hours after awake

Source

Moller-Levet C. S., et al. (2013). Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci U S A*, **110(12)**, E1132–1141.

cycMouseLiverProtein *cycMouseLiverProtein*

Description

This data set lists expression profiles of 5 circadian proteins with 3h-resolution covering two days.

Usage

cycMouseLiverProtein

Format

A dataframe containing 49 columns(column 1 = protein name, column 2 to 49 = time points from CT0 to CT45 with three replicates at each time point).

Source

Robles M. S., Cox J., Mann M. (2014). In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet*, **10(1)**, e1004047.

cycMouseLiverRNA *cycMouseLiverRNA*

Description

This data set lists expression profiles of 20 circadian transcripts with 1h-resolution covering two days.

Usage

cycMouseLiverRNA

Format

A dataframe containing 49 columns(column 1 = transcript name, column 2 to 49 = time points from CT18 to CT65).

Source

Hughes M. E., et al. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genet*, **5(4)**, e1000442.

cycSimu4h2d

cycSimu4h2d

Description

This data set lists 20 simulated profiles(periodic and non-periodic) with 4h-resolution covering two periods.

Usage

cycSimu4h2d

Format

A dataframe containing 13 columns(column 1 = curve ID, column 2 to 13 = time points from 0 to 44).

Source

Wu G., Zhu J., Yu J., Zhou L., Huang J. Z. and Zhang Z. (2014). Evaluation of five methods for genome-wide circadian gene identification. *Journal of Biological Rhythms*, **29(4)**, 231–242.

cycVignettesAMP

cycVignettesAMP

Description

This data set lists meta2d's analysis results of three circadian transcripts selected from the same source dataset used by cycMouseLiverRNA.

Usage

cycVignettesAMP

Format

A dataframe containing 71 columns described as below:

[,1]	CycID	character	transcript name
[,2]	ARS_pvalue	numeric	pvalue from ARS
[,3]	ARS_BH.Q	numeric	FDR from ARS
[,4]	ARS_period	numeric	period from ARS
[,5]	ARS_adjphase	numeric	adjusted phase from ARS
[,6]	ARS_amplitude	numeric	amplitude from ARS
[,7]	JTK_pvalue	numeric	pvalue from JTK
[,8]	JTK_BH.Q	numeric	FDR from JTK
[,9]	JTK_period	numeric	period from JTK
[,10]	JTK_adjphase	numeric	adjusted phase from JTK
[,11]	JTK_amplitude	numeric	amplitude from JTK
[,12]	LS_pvalue	numeric	pvalue from LS
[,13]	LS_BH.Q	numeric	FDR from JTK
[,14]	LS_period	numeric	period from LS
[,15]	LS_adjphase	numeric	adjusted phase from LS
[,16]	LS_amplitude	numeric	amplitude from LS
[,17]	meta2d_pvalue	numeric	integrated pvalue
[,18]	meta2d_BH.Q	numeric	FDR based on integrated pvalue
[,19]	meta2d_period	numeric	averaged period of three methods
[,20]	meta2d_phase	numeric	integrated phase
[,21]	meta2d_Base	numeric	baseline value given by meta2d
[,22]	meta2d_AMP	numeric	amplitude given by meta2d
[,23]	meta2d_rAMP	numeric	relative amplitude
[,24:71]	CT18 to CT65	numeric	sampling time point

Source

Hughes M. E., et al. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genet*, **5(4)**, e1000442.

cycYeastCycle

cycYeastCycle

Description

This data set lists expression profiles of 10 cycling transcripts with 16-minutes resolution covering about two yeast cell cycles.

Usage

cycYeastCycle

Format

A dataframe containing 12 columns(column 1 = transcript name, column 2 to 12 = time points from 2 minutes to 162 minutes after recovery phase).

Source

Orlando D. A., et al. (2008). Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature*, **453**(7197), 944–947.

meta2d	<i>Detect rhythmic signals from time-series datasets with multiple methods</i>
--------	--

Description

This is a function that incorporates ARSER, JTK_CYCLE and Lomb-Scargle to detect rhythmic signals from time-series datasets.

Usage

```
meta2d(infile, outdir = "metaout", filestyle, timepoints, minper = 20,
maxper = 28, cycMethod = c("ARS", "JTK", "LS"),
analysisStrategy = "auto", outputFile = TRUE, outIntegration = "both",
adjustPhase = "predictedPer", combinePvalue = "fisher",
weightedPerPha = FALSE, ARSmle = "auto", ARSdefaultPer = 24,
outRawData = FALSE, releaseNote = TRUE, outSymbol = "")
```

Arguments

infile	a character string. The name of input file containing time-series data.
outdir	a character string. The name of directory used to store output files.
filestyle	a character vector(length 1 or 3). The data format of input file, must be "txt", or "csv", or a character vector containing field separator character(sep), quoting character (quote), and the character used for decimal points(dec, for details see read.table).
timepoints	a numeric vector corresponding to sampling time points of input time-series data; if sampling time points are in the first line of input file, it could be set as a character sting-"Line1" or "line1".
minper	a numeric value. The minimum period length of interested rhythms. The default is 20 for circadian rhythms.
maxper	a numeric value. The maximum period length of interested rhythms. The default is 28 for circadian rhythms.
cycMethod	a character vector(length 1 or 2 or 3). User-defined methods for detecting rhythmic signals, must be selected as any one, any two or all three methods(default from "ARS"(ARSER), "JTK"(JTK_CYCLE) and "LS"(Lomb-Scargle).
analysisStrategy	a character string. The strategy used to select proper methods from cycMethod for analyzing input time-series data, must be "auto"(default), or "selfUSE". See Details part for more information.

outputFile	logical. If TRUE, analysis results will be wrote in the output files. If FALSE, analysis results will be returned as an R list.
outIntegration	a character string. This parameter controls what kinds of analysis results will be outputted, must be one of "both" (default), "onlyIntegration"(only output integration file), or "noIntegration"(not output integration file).
adjustPhase	a character string. The method used to adjust original phase calculated by each method in integration file, must be one of "predictedPer"(adjust phase with predicted period length) or "notAdjusted"(not adjust phase).
combinePvalue	a character string. The method used to integrate multiple p-values, must be one of "bonferroni"(Bonferroni correction), or "fisher"(Fisher's method).
weightedPerPha	logical. If TRUE, weighted scores based on p-value given by each method will be used to calculate the integrated period length and phase.
ARSmle	a character string. The strategy of using MLE method in ar fit of "ARS", must be one of "auto"(use MLE depending the number of time points), "mle" (always use MLE), or "nomle"(never use MLE).
ARSdefaultPer	a numeric value. The expected period length of interested rhythm, which is a necessary parameter for ARS. The default is 24(for circadian rhythms). Set it to another proper numeric value for other rhythms.
outRawData	logical. If TRUE, raw time-series data will be added in the output files.
releaseNote	logical. If TRUE, reminding or warning notes during the analysis will be released on the screen.
outSymbol	a character string. A common prefix exists in the names of output files.

Details

ARSER(Yang, 2010), **JTK_CYCLE**(Hughes, 2010), and **Lomb-Scargle**(Glynn, 2006) are three popular methods of detecting rhythmic signals. ARS can not analyze unevenly sampled datasets, or evenly sampled datasets but with missing values, or with replicate samples, or with non-integer sampling interval. JTK is not suitable to analyze unevenly sampled datasets or evenly sampled datasets but with non-integer sampling interval. If set analysisStrategy as "auto"(default), meta2d will automatically select proper method from cycMethod for each input dataset. If the user clearly know that the dataset could be analyzed by each method defined by cycMethod and do not hope to output integrated values, analysisStrategy can be set as "selfUSE".

ARS used here is translated from its python version which always uses "yule-walker", "burg", and "mle" methods(see [ar](#)) to fit autoregressive models to time-series data. Fitting by "mle" will be very slow for datasets with many time points. If ARSmle = "auto" is used, meta2d will only include "mle" when number of time points is smaller than 24. In addition, one evaluation work(Wu, 2014) indicates that ARS shows relative high false positive rate in analyzing high-resolution datasets (1h/2days and 2h/2days). JTK(version 3) used here is the latest version, which improves its p-value calculation in analyzing datasets with missing values.

The power of detecting rhythmic signals for an algorithm is associated with the nature of data and interested periodic pattern(Deckard, 2013), which indicates that integrating analysis results from multiple methods may be helpful to rhythmic detection. For integrating p-values, Bonferroni correction("bonferroni") and Fisher's method("fisher") (Fisher, 1925; implementation code from **MADAM**) could be selected, and "bonferroni" is usually more conservative than "fisher".

The integrated period is arithmetic mean of multiple periods. For integrating phase, meta2d takes use of **mean of circular quantities**. Integrated period and phase is further used to calculate the baseline value and amplitude through fitting a constructed periodic model.

Phases given by JTK and LS need to be adjusted with their predicted period (adjustedPhase = "predictedPer") before integration. If adjustedPhase = "notAdjusted" is selected, no integrated phase will be calculated. If set weightedPerPha as TRUE, weighted scores will be used in averaging periods and phases. Weighted scores for one method are based on all its reported p-values, which means a weighted score assigned to any one profile will be affected by all other profiles. It is always a problem of averaging phases with quite different period lengths(eg. averaging two phases with 16-hours' and 30-hours' period length). Currently, setting minper, maxper and ARSdefaultPer to a same value may be the only way of completely eliminating such problem.

This function is originally aimed to analyze large scale periodic data(eg. circadian transcriptome data) without individual information. Please pay attention to data format of input file(see Examples part). Except the first column and first row, others are time-series experimental values(setting missing values as NA).

Value

meta2d will write analysis results in different files under outdir if set outputFile = TRUE. Files named with "ARSresult", "JTKresult" and "LSresult" store analysis results from ARS, JTK and LS respectively. The file named with "meta2d" is the integration file, and it stores integrated values in columns with a common name tag-"meta2d". The integration file also contains p-value, FDR value, period, phase(adjusted phase if adjustedPhase = "predictedPer") and amplitude values calculated by each method. If outputFile = FALSE is selected, meta2d will return a list containing the following components:

ARS	analysis results from ARS method
JTK	analysis results from JTK method
LS	analysis results from LS method
meta	the integrated analysis results as mentioned above

References

- Yang R. and Su Z. (2010). Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*, **26(12)**, i168–i174.
- Hughes M. E., Hogenesch J. B. and Kornacker K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological Rhythms*, **25(5)**, 372–380.
- Glynn E. F., Chen J. and Mushegian A. R. (2006). Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics*, **22(3)**, 310–316.
- Wu G., Zhu J., Yu J., Zhou L., Huang J. Z. and Zhang Z. (2014). Evaluation of five methods for genome-wide circadian gene identification. *Journal of Biological Rhythms*, **29(4)**, 231–242.
- Deckard A., Anafi R. C., Hogenesch J. B., Haase S.B. and Harer J. (2013). Design and analysis of large-scale biological rhythm studies: a comparison of algorithms for detecting periodic signals in biological data. *Bioinformatics*, **29(24)**, 3174–3180.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Oliver and Boyd (Edinburgh).

Kugler K. G., Mueller L.A. and Graber A. (2010). MADAM - an open source toolbox for meta-analysis. *Source Code for Biology and Medicine*, **5**, 3.

Examples

```
# write 'cycMouseLiverProtein' into a 'txt' file
write.table(cycMouseLiverProtein, file="cycMouseLiverProtein.txt",
  sep="\t", quote=FALSE, row.names=FALSE)
# write 'cycSimu4h2d' and 'cycYeastCycle' into two 'csv' files
write.csv(cycSimu4h2d, file="cycSimu4h2d.csv", row.names=FALSE)
write.csv(cycYeastCycle, file="cycYeastCycle.csv", row.names=FALSE)

# analyze 'cycMouseLiverProtein.txt' with JTK_CYCLE and Lomb-Scargle
meta2d(infile="cycMouseLiverProtein.txt", filestyle="txt",
  outdir="example", timepoints=rep(seq(0, 45, by=3), each=3),
  cycMethod=c("JTK","LS"), outIntegration="noIntegration")

# analyze 'cycSimu4h2d.csv' with ARSER, JTK_CYCLE and Lomb-Scargle and
# output integration file with analysis results from each method
meta2d(infile="cycSimu4h2d.csv", filestyle="csv", outdir="example",
  timepoints="Line1")

# analyze 'cycYeastCycle.csv' with ARSER, JTK_CYCLE and Lomb-Scargle to
# detect transcripts associated with cell cycle, and return analysis
# results instead of output them into files
cyc <- meta2d(infile="cycYeastCycle.csv", filestyle="csv",
  minper=80, maxper=96, timepoints=seq(2, 162, by=16),
  outputFile=FALSE, ARSdefaultPer=85, outRawData=TRUE)
head(cyc$ARS)
head(cyc$JTK)
head(cyc$LS)
head(cyc$meta)
```

meta3d

Detect rhythmic signals from time-series datasets with individual information

Description

This is a function that takes use of any one method from ARSER, JTK_CYCLE and Lomb-Scargle to detect rhythmic signals from time-series datasets containing individual information.

Usage

```
meta3d(datafile, designfile, outdir = "metaout", filestyle, design_libColm,
  design_subjectColm, minper = 20, maxper = 28, cycMethodOne = "JTK",
  timeUnit = "hour", design_hrColm, design_dayColm = NULL,
  design_minColm = NULL, design_secColm = NULL, design_groupColm = NULL,
  design_libIDrename = NULL, adjustPhase = "predictedPer",
```

```
combinePvalue = "fisher", weightedMethod = TRUE,
outIntegration = "both", ARSmle = "auto", ARSdefaultPer = 24,
dayZeroBased = FALSE, outSymbol = "")
```

Arguments

datafile	a character string. The name of data file containing time-series experimental values of all individuals.
designfile	a character string. The name of experimental design file, at least containing the library ID(column names of datafile), subject ID(the individual corresponding to each library ID), and sampling time information of each library ID.
outdir	a character string. The name of directory used to store output files.
filestyle	a character vector(length 1 or 3). The data format of input files, must be "txt", or "csv", or a character vector containing field separator character(sep), quoting character(quote), and the character used for decimal points(dec, for details see read.table).
design_libColm	a numeric value. The order index(from left to right) of the column storing library ID in designfile.
design_subjectColm	a numeric value. The order index(from left to right) of the column storing subject ID in designfile.
minper	a numeric value. The minimum period length of interested rhythms. The default is 20 for circadian rhythms.
maxper	a numeric value. The maximum period length of interested rhythms. The default is 28 for circadian rhythms.
cycMethodOne	a character string. The selected method for analyzing time-series data of each individual, must be one of "ARS"(ARSER), "JTK"(JTK_CYCLE), or "LS"(Lomb-Scargle).
timeUnit	a character string. The basic time-unit, must be one of "day", "hour"(default for circadian study), "minute", or "second" depending on specific experimental design.
design_hrColm	a numeric value. The order index(from left to right) of the column storing time point value-sampling hour information in designfile. If there is no such column in designfile, set it as NULL.
design_dayColm	a numeric value. The order index(from left to right) of the column storing time point value-sampling day information in designfile. If there is no such column in designfile, set it as NULL(default).
design_minColm	a numeric value. The order index(from left to right) of the column storing time point value-sampling minute information in designfile. If there is no such column in designfile, set it as NULL(default).
design_secColm	a numeric value. The order index(from left to right) of the column storing time point value-sampling second information in designfile. If there is no such column in designfile, set it as NULL(default).

design_groupColm	a numeric value. The order index(from left to right) of the column storing experimental group information of each individual in designfile. If there is no such column in designfile, set it as NULL(default) and take all individuals as one group.
design_libIDrename	a character vector(length 2) containing a matchable character string in each library ID of designfile, and a replacement character string. If it is not necessary to replace characters in library ID of designfile, set it as NULL(default).
adjustPhase	a character string. The method used to adjust each calculated phase before getting integrated phase, must be one of "predictedPer"(adjust phase with predicted period length) or "notAdjusted"(not adjust phase).
combinePvalue	a character string. The method used to integrate p-values of multiple individuals, currently only "fisher"(Fisher's method) could be selected.
weightedMethod	logical. If TRUE(default), weighted score based on p-value of each individual will be used to integrate period, phase and amplitude values of multiple individuals.
outIntegration	a character string. This parameter controls what kinds of analysis results will be outputted, must be one of "both", "onlyIntegration", or "noIntegration". See meta2d for more information.
ARSmle	a character string. The strategy of using MLE method in "ARS", must be one of "auto", "mle", or "nomle". See meta2d for more information.
ARSdefaultPer	a numeric value. The expected period length of interested rhythm, which is a necessary parameter for ARS. See meta2d for more information.
dayZeroBased	logical. If TRUE, the first sampling day is recorded as day zero in the designfile.
outSymbol	a character string. A common prefix exists in the names of output files.

Details

This function is originally aimed to analyze large scale periodic data with individual information. Please pay attention to the data format of datafile and designfile(see Examples part). Time-series experimental values(missing values as NA) from all individuals should be stored in datafile, with the first row containing all library ID(unique identification number for each sample) and the first column containing all detected molecular names(eg. transcript or gene name). The designfile should at least have three columns-library ID, subject ID and sampling time column. Experimental group information of each subject ID may be in another column. In addition, sampling time information may be stored in multiple columns instead of one column. For example, sampling time-"36 hours" may be recorded as "day 2"(sampling day column, design_dayColm) plus "12 hours"(sampling hour column, design_hrColm). The library ID in datafile and designfile should be same. If there are different characters between library ID in these two files, try design_libIDrename to keep them same.

ARS, JTK or LS could be used to analyze time-series profiles individual by individual. meta3d requires that all individuals should be analyzed by the same method before integrating calculated p-value, period, phase, baseline value, amplitude and relative amplitude values group by group. However, the sampling pattern among individuals may be different and the requirement of sampling pattern for each method is not same(see more information about these methods and their limitations

in [meta2d](#)). Please carefully select a proper method for the specific dataset. meta3d also help users select the suitable method through warning notes.

P-values from different individuals are integrated with Fisher's method ("fisher")(Fisher,1925; implementation code from **MADAM**).For short time-series profiles(eg. 10 time points or less), p-values given by Lomb-Scargle may be over conservative, which will also lead to conservative integrated p-values. The integrated period, baseline, amplitude and relative amplitude values are arithmetic mean of multiple individuals, respectively. The phase is **mean of circular quantities**(adjustPhase = "predictedPer") or a arithmetic mean (adjustPhase = "notAdjusted") of multiple individual phases. For completely removing the potential problem of averaging phases with quite different period length(also mentioned in [meta2d](#)), setting minper, maxper and ARSdefaultPer to a same value may be the only known way. If weightedMethod = TRUE is selected, weighted scores($-\log_{10}(p\text{-values})$) will be taken into account in integrating period, phase, baseline, amplitude and relative amplitude.

Value

meta3d will write analysis results to outdir instead of returning them as objects. Output files with "meta3dSubjectID" in the file name are analysis results for each individual. Files named with "meta3dGroupID" store integrated p-values, period, phase, baseline, amplitude and relative amplitude values from multiple individuals of each group and calculated FDR values based on integrated p-values.

References

Glynn E. F., Chen J., and Mushegian A. R. (2006). Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. *Bioinformatics*, **22(3)**, 310–316

Fisher, R.A. (1925). *Statistical methods for research workers*. Oliver and Boyd (Edinburgh).

Kugler K. G., Mueller L.A., and Graber A. (2010). MADAM - an open source toolbox for meta-analysis. *Source Code for Biology and Medicine*, **5**, 3.

Examples

```
# write 'cycHumanBloodData' and 'cycHumanBloodDesign' into two 'csv' files
write.csv(cycHumanBloodData, file="cycHumanBloodData.csv",
  row.names=FALSE)
write.csv(cycHumanBloodDesign, file="cycHumanBloodDesign.csv",
  row.names=FALSE)

# detect circadian transcripts with JTK in studied individuals
meta3d(datafile="cycHumanBloodData.csv", cycMethodOne="JTK",
  designfile="cycHumanBloodDesign.csv", outdir="example",
  filestyle="csv", design_libColm=1, design_subjectColm=2,
  design_hrColm=4, design_groupColm=3)
```

MetaCycle

MetaCycle Evaluate periodicity in large scale data

Description

MetaCycle Evaluate periodicity in large scale data

Author(s)

Gang Wu <wggucas@gmail.com>, Ron Anafi <ranafi@gmail.com>, John Hogenesch <hogenesch@gmail.com>, Michael Hughes <michael.evan.hughes@gmail.com>, Karl Kornacker <kornacker@midohio.twcbc.com>
Maintainer: Gang Wu <wggucas@gmail.com>

Index

`ar`, [7](#)

`cycHumanBloodData`, [2](#)

`cycHumanBloodDesign`, [2](#)

`cycMouseLiverProtein`, [3](#)

`cycMouseLiverRNA`, [3](#)

`cycSimu4h2d`, [4](#)

`cycVignettesAMP`, [4](#)

`cycYeastCycle`, [5](#)

`meta2d`, [6](#), [11](#), [12](#)

`meta3d`, [9](#)

`MetaCycle`, [13](#)

`MetaCycle-package (MetaCycle)`, [13](#)

`read.table`, [6](#), [10](#)