Package 'regionReport'

April 15, 2017

Type Package

Title Generate HTML or PDF reports for a set of genomic regions or DESeq2/edgeR results

Version 1.8.2

Date 2017-03-06

Description Generate HTML or PDF reports to explore a set of regions such as the results from annotation-agnostic expression analysis of RNA-seq data at base-pair resolution performed by derfinder. You can also create reports for DESeq2 or edgeR results.

License Artistic-2.0

LazyData true

URL https://github.com/leekgroup/regionReport

 ${\bf BugReports}\ {\tt https://support.bioconductor.org/t/regionReport/}$

VignetteBuilder knitr

biocViews DifferentialExpression, Sequencing, RNASeq, Software, Visualization, Transcription, Coverage, ReportWriting, DifferentialMethylation, DifferentialPeakCalling

Depends R(>=3.2)

Imports derfinder (>= 1.1.0), DEFormats, DESeq2, GenomeInfoDb, GenomicRanges, knitcitations (>= 1.0.1), knitr (>= 1.6), knitrBootstrap (>= 0.9.0), methods, RefManageR, rmarkdown (>= 0.9.5), S4Vectors, SummarizedExperiment

Suggests BiocStyle, biovizBase, bumphunter (>= 1.7.6), Cairo, derfinderPlot (>= 1.3.2), devtools (>= 1.6), DT, DESeq, edgeR, ggbio (>= 1.13.13), ggplot2, grid, gridExtra, IRanges, mgcv, pasilla, pheatmap, RColorBrewer,

TxDb.Hsapiens.UCSC.hg19.knownGene, whisker

RoxygenNote 5.0.1

NeedsCompilation no

Author Leonardo Collado-Torres [aut, cre], Andrew E. Jaffe [aut], Jeffrey T. Leek [aut, ths]

Maintainer Leonardo Collado-Torres <lcollado@jhu.edu>

2 derfinderReport

R topics documented:

regionReport-pa	ckage	€.	 •	 					•							•	•	•	2
derfinderReport				 															2
DESeq2Report				 															5
edgeReport				 															8
load_install				 															10
renderReport .				 															11
with_wd				 															13

15

regionReport-package Generate HTML or PDF reports for a set of regions or DESeq2/edgeR

Description

Index

Generate an HTML reports to explore a set of regions such as the results from annotation-agnostic expression analysis of RNA-seq data at base-pair resolution performed by derfinder. The package can also be used to generate reports for DESeq2 or edgeR results. The HTML and PDF reports are generated using rmarkdown (http://rmarkdown.rstudio.com/).

Author(s)

Leonardo Collado-Torres <lcollado@jhsph.edu>

results.

derfinderReport Generate a HTML/PDF report exploring the basic results from derfinder

Description

This function generates a HTML report exploring the basic results from single base-level approach derfinder analysis results (www.bioconductor.org/packages/derfinder). The HTML report itself is generated using rmarkdown (http://rmarkdown.rstudio.com/). It works best after using mergeResults.

Usage

```
derfinderReport(prefix, outdir = "basicExploration",
  output = "basicExploration", project = prefix, browse = interactive(),
  nBestRegions = 100, makeBestClusters = TRUE, nBestClusters = 2,
  fullCov = NULL, hg19 = TRUE, p.ideos = NULL, txdb = NULL,
  device = "png", significantVar = "qvalue", customCode = NULL,
  template = NULL, theme = NULL, digits = 2, ...)
```

derfinderReport 3

Arguments

prefix The main data directory path where mergeResults was run. It should be the same

as mergeResults(prefix).

outdir The name of output directory relative to prefix.

output The name of output HTML file (without the html extension).

project The title of the project.

browse If TRUE the HTML report is opened in your browser once it's completed.

nBestRegions The number of region plots to make, ordered by area.

makeBestClusters

If TRUE, plotCluster is used on the nBestClusters regions by area. Note that

these plots take some time to make.

nBestClusters The number of region cluster plots to make by taking the nBestClusters re-

gions ranked by area of the cluster.

fullCov A list where each element is the result from loadCoverage used with cutoff=NULL.

Can be generated using fullCoverage.

hg19 If TRUE then the reference is assumed to be hg19 and chromosome lengths as

well as the default transcription database (TxDb.Hsapiens.UCSC.hg19.knownGene)

will be used.

p. ideos A list where each element is the result of plotIdeogram. If it's NULL and hg19=TRUE

then they are created for the hg19 human reference.

txdb Specify the transcription database to use for making the plots for the top regions

by area. If NULL and hg19=TRUE then TxDb.Hsapiens.UCSC.hg19.knownGene

is used.

device The graphical device used when knitting. See more at http://yihui.name/knitr/options

(dev argument).

significantVar A character variable specifying whether to use the p-values, the FDR adjusted

p-values or the FWER adjusted p-values to determine significance. Has to be

either 'pvalue', 'qvalue' or 'fwer'.

customCode An absolute path to a child R Markdown file with code to be evaluated before

the reproducibility section. Its useful for users who want to customize the report by adding conclusions derived from the data and/or further quality checks and

plots.

template Template file to use for the report. If not provided, will use the default file found

in basicExploration/basicExploration.Rmd within the package source.

theme A ggplot2 theme to use for the plots made with ggplot2.

digits The number of digits to round to in the interactive table of the top nBestRegions.

Note that p-values and adjusted p-values won't be rounded.

... Arguments passed to other methods and/or advanced arguments. Advanced ar-

guments:

chrsStyle The naming style of the chromosomes. By default, UCSC. See se-

qlevelsStyle.

species Species name. See extendedMapSeqlevels for more information.

currentStyle Current naming style used. See extendedMapSeqlevels for more information.

fullRegions Part of the output of mergeResults. Specify it only if you have already loaded it in memory.

4 derfinderReport

fullNullSummary Part of the output of mergeResults. Specify it only if you have already loaded it in memory.

fullAnnotatedRegions Part of the output of mergeResults. Specify it only if you have already loaded it in memory.

optionsStats Part of the output of analyzeChr. Specify it only if you have already loaded it in memory.

optionsMerge Part of the output of mergeResults. Specify it only if you have already loaded it in memory.

overviewParams A two element list with base_size and areaRel that control the text size for the genomic overview plots.

the text size for the genomic overview plots.

output_format Either html_document, pdf_document or knitrBootstrap::bootstrap_document

clean Logical, whether to clean the results or not. Passed to render.

Passed to extendedMapSeqlevels.

unless you modify the YAML template.

Details

Set output_format to 'knitrBootstrap::bootstrap_document' or 'pdf_document' if you want a HTML report styled by knitrBootstrap or a PDF report respectively. If using knitrBootstrap, we recommend the version available only via GitHub at https://github.com/jimhester/knitrBootstrap which has nicer features than the current version available via CRAN.

If you modify the YAML front matter of template, you can use other values for output_format.

The HTML report styled with knitrBootstrap can be smaller in size than the default HTML report.

Value

An HTML report with a basic exploration of the derfinder results.

Author(s)

Leonardo Collado-Torres

See Also

mergeResults, analyzeChr, fullCoverage

Examples

DESeq2Report 5

```
## Collapse the coverage information
collapsedFull <- collapseFullCoverage(list(genomeData$coverage),</pre>
    verbose=TRUE)
## Calculate library size adjustments
sample Depths <- \ sample Depth (collapsed Full, \ probs=c(0.5), \ nonzero=TRUE,
    verbose=TRUE)
## Build the models
group <- genomeInfo$pop</pre>
adjustvars <- data.frame(genomeInfo$gender)</pre>
models <- makeModels(sampleDepths, testvars=group, adjustvars=adjustvars)</pre>
## Analyze chromosome 21
analyzeChr(chr='21', coverageInfo=genomeData, models=models,
    cutoffFstat=1, cutoffType='manual', seeds=20140330, groupInfo=group,
    mc.cores=1, writeOutput=TRUE, returnOutput=FALSE)
## Change the directory back to the original one
setwd(initialPath)
## End(Not run)
## Merge the results from the different chromosomes. In this case, there's
## only one: chr21
mergeResults(chrs = '21', prefix = 'derfinderReport-example',
    genomicState = genomicState$fullGenome)
## Load the options used for calculating the statistics
load(file.path('derfinderReport-example', 'chr21', 'optionsStats.Rdata'))
## Generate the HTML report
report <- derfinderReport(prefix='derfinderReport-example', browse=FALSE,
    nBestRegions=1, makeBestClusters=FALSE,
    fullCov=list('21'=genomeDataRaw$coverage), optionsStats=optionsStats)
if(interactive()) {
    ## Browse the report
    browseURL(report)
}
## Note that you can run the example using:
example('derfinderReport', 'regionReport', ask=FALSE)
## End(Not run)
```

DESeq2Report

Generate a HTML/PDF report exploring DESeq2 results

Description

This function generates a HTML report with exploratory data analysis plots for DESeq2 results created with DESeq. Other output formats are possible such as PDF but lose the interactivity. Users

6 DESeq2Report

can easily append to the report by providing a R Markdown file to customCode, or can customize the entire template by providing an R Markdown file to template.

Usage

```
DESeq2Report(dds, project = "", intgroup, colors = NULL, res = NULL,
   nBest = 500, nBestFeatures = 20, customCode = NULL,
   outdir = "DESeq2Exploration", output = "DESeq2Exploration",
   browse = interactive(), device = "png", template = NULL,
   searchURL = "http://www.ncbi.nlm.nih.gov/gene/?term=", theme = NULL,
   digits = 2, ...)
```

Arguments

dds A DESeqDataSet object with the results from running DESeq.

project The title of the project.

interesting groups: a character vector of names in colData(x) to use for group-

ing. This parameter is passed to functions such as plotPCA.

colors vector of colors used in heatmap. If NULL, then a a default set of colors will be

used. This argument is passed to pheatmap.

res A DESeqResults object. If NULL, then results will be used on dds with default

parameters.

nBest The number of features to include in the interactive table. Features are ordered

by their adjusted p-values.

nBestFeatures The number of best features to make plots of their counts. We recommend a

small number, say 20.

customCode An absolute path to a child R Markdown file with code to be evaluated before

the reproducibility section. Its useful for users who want to customize the report by adding conclusions derived from the data and/or further quality checks and

plots.

outdir The name of output directory.

output The name of output HTML file (without the html extension).

browse If TRUE the HTML report is opened in your browser once it's completed.

device The graphical device used when knitting. See more at http://yihui.name/knitr/options

(dev argument).

template Template file to use for the report. If not provided, will use the default file found

in DESeq2Exploration/DESeq2Exploration.Rmd within the package source.

searchURL A url used for searching the name of the features in the web. By default http://www.ncbi.nlm.nih.

is used which is the recommended option when features are genes. It's only used

when the output is a HTML file.

theme A ggplot2 theme to use for the plots made with ggplot2.

digits The number of digits to round to in the interactive table of the top nBestFeatures.

Note that p-values and adjusted p-values won't be rounded.

... Arguments passed to other methods and/or advanced arguments. Advanced ar-

guments:

software The name of the package used for performing the differential expression analysis. Either DESeq2 or edgeR.

DESeq2Report 7

```
dge A DGEList object. NULL by default and only used by edgeReport.
```

theCall The function call. NULL by default and only used by edgeReport.

output_format Either html_document, pdf_document or knitrBootstrap::bootstrap_documer
unless you modify the YAML template.

clean Logical, whether to clean the results or not. Passed to render.

Details

Set output_format to 'knitrBootstrap::bootstrap_document' or 'pdf_document' if you want a HTML report styled by knitrBootstrap or a PDF report respectively. If using knitrBootstrap, we recommend the version available only via GitHub at https://github.com/jimhester/knitrBootstrap which has nicer features than the current version available via CRAN.

If you modify the YAML front matter of template, you can use other values for output_format.

Value

An HTML report with a basic exploration for the given set of DESeq2 results.

Author(s)

Leonardo Collado-Torres

Examples

```
## Load example data from the pasilla package
library('pasilla')
library('DESeq')
library('DESeq2')
## Create DESeqDataSet object from the pasilla package
data('pasillaGenes')
countData <- counts(pasillaGenes)</pre>
colData <- pData(pasillaGenes)[, c('condition', 'type')]</pre>
dds <- DESeqDataSetFromMatrix(countData = countData,</pre>
  colData = colData,
   design = \sim condition)
dds <- DESeq(dds)
## The output will be saved in the 'DESeq2Report-example' directory
dir.create('DESeq2Report-example', showWarnings = FALSE, recursive = TRUE)
## Generate the HTML report
report <- DESeq2Report(dds, 'DESeq2-example', c('condition', 'type'),</pre>
    outdir = 'DESeq2Report-example')
if(interactive()) {
    ## Browse the report
    browseURL(report)
}
## Not run:
## Note that you can run the example using:
example('DESeq2Report', 'regionReport', ask=FALSE)
```

8 edgeReport

```
## End(Not run)
```

edgeReport

Generate a HTML/PDF report exploring edgeR results

Description

This function generates a HTML report with exploratory data analysis plots for edgeR results created. Other output formats are possible such as PDF reports but they lose the interactivity. Users can easily append to the report by providing a R Markdown file to customCode, or can customize the entire template by providing an R Markdown file to template.

Usage

```
edgeReport(dge, object, project = "", intgroup, colors = NULL,
  pAdjustMethod = "BH", alpha = 0.1, independentFiltering = FALSE, filter,
  theta, filterFun, nBest = 500, nBestFeatures = 20, customCode = NULL,
  outdir = "edgeRexploration", output = "edgeRexploration",
  browse = interactive(), device = "png", template = NULL,
  searchURL = "http://www.ncbi.nlm.nih.gov/gene/?term=", theme = NULL,
  digits = 2, ...)
```

Arguments

dge A DGEList object.

object A DGEExact or DGELRT object that contains p-values stored in object\$table\$PValue.

project The title of the project.

integroup interesting groups: a character vector of names in colData(x) to use for group-

ing. This parameter is passed to functions such as plotPCA.

colors vector of colors used in heatmap. If NULL, then a a default set of colors will be

used. This argument is passed to pheatmap.

pAdjustMethod the method to use for adjusting p-values, see p.adjust. This argument will be

passed to results.

alpha the significance cutoff used for optimizing the independent filtering (by default

0.1). If the adjusted p-value cutoff (FDR) will be a value other than 0.1, alpha

should be set to that value. This argument will be passed to results.

independentFiltering

logical, whether independent filtering should be applied automatically. By default it's set to FALSE in contrast with the default used in results to match edgeR's

behavior.

filter the vector of filter statistics over which the independent filtering will be opti-

mized. By default the logCPM will be used if independentFiltering is set to TRUE. It can also be a length 1 character vector specifying one of the column

names of object\$table.

theta the quantiles at which to assess the number of rejections from independent fil-

tering. This argument is passed results.

edgeReport 9

filterFun an optional custom function as described in results.

nBest The number of features to include in the interactive table. Features are ordered

by their adjusted p-values.

nBestFeatures The number of best features to make plots of their counts. We recommend a

small number, say 20.

customCode An absolute path to a child R Markdown file with code to be evaluated before

the reproducibility section. Its useful for users who want to customize the report by adding conclusions derived from the data and/or further quality checks and

plots.

outdir The name of output directory.

output The name of output HTML file (without the html extension).

browse If TRUE the HTML report is opened in your browser once it's completed.

device The graphical device used when knitting. See more at http://yihui.name/knitr/options

(dev argument).

template Template file to use for the report. If not provided, will use the default file found

in DESeq2Exploration/DESeq2Exploration.Rmd within the package source.

in B250422Aptotation/B250422Aptotation/time within the purkage source.

is used which is the recommended option when features are genes. It's only used

A url used for searching the name of the features in the web. By default http://www.ncbi.nlm.nih.

when the output is a HTML file.

theme A ggplot2 theme to use for the plots made with ggplot2.

digits The number of digits to round to in the interactive table of the top nBestFeatures.

Note that p-values and adjusted p-values won't be rounded.

... Arguments passed to other methods and/or advanced arguments. Advanced ar-

guments:

software The name of the package used for performing the differential expres-

sion analysis. Either DESeq2 or edgeR.

dge A DGEList object. NULL by default and only used by edgeReport.

theCall The function call. NULL by default and only used by edgeReport. **output_format** Either html_document, pdf_document or knitrBootstrap::bootstrap_document

unless you modify the YAML template.

clean Logical, whether to clean the results or not. Passed to render.

Details

searchURL

Set output_format to 'knitrBootstrap::bootstrap_document' or 'pdf_document' if you want a HTML report styled by knitrBootstrap or a PDF report respectively. If using knitrBootstrap, we recommend the version available only via GitHub at https://github.com/jimhester/knitrBootstrap which has nicer features than the current version available via CRAN.

If you modify the YAML front matter of template, you can use other values for output_format.

This report is similar to the one created by DESeq2Report with two additional plots exclusive for edgeR results. We designed the reports to be very similar intentionally and use the Bioconductor package DEFormats to achieve this goal.

Value

An HTML report with a basic exploration for the given set of edgeR results.

10 load_install

Author(s)

Leonardo Collado-Torres

Examples

```
## Create example data using DEFormats
library('DEFormats')
set.seed(20160407)
counts <- simulateRnaSeqData()</pre>
group \leftarrow rep(c("A", "B"), each = 3)
## Create DGEList object
library('edgeR')
dge <- DGEList(counts, group = group)</pre>
## Perform DE analysis with edgeR
design <- model.matrix( ~ group)</pre>
dge <- estimateDisp(dge, design)</pre>
fit <- glmFit(dge, design)</pre>
lrt <- glmLRT(fit, coef = 2)</pre>
## The output will be saved in the 'edgeReport-example' directory
dir.create('edgeReport-example', showWarnings = FALSE, recursive = TRUE)
## Generate the HTML report
report <- edgeReport(dge, lrt, project = 'edgeR-example', intgroup = 'group',</pre>
    outdir = 'edgeReport-example')
if(interactive()) {
    ## Browse the report
    browseURL(report)
}
## Not run:
## Note that you can run the example using:
example('edgeReport', 'regionReport', ask=FALSE)
## End(Not run)
```

load_install

Attempt to load the namespace of a package and install it if it's missing

Description

This function uses requireNamespace to try to load a package. But if it's misssing it will then install it via Bioconductor.

Usage

```
load_install(pkg, quietly = TRUE)
```

renderReport 11

Arguments

pkg A single character vector with the name of the package.

quietly Whether to run requireNamespace and biocLite quietly or not.

renderReport

Generate a HTML/PDF report exploring a set of genomic regions

Description

This function generates a HTML report with quality checks, genome location exploration, and an interactive table with the results. Other output formats are possible such as PDF but lose the interactivity. Users can easily append to the report by providing a R Markdown file to customCode, or can customize the entire template by providing an R Markdown file to template.

Usage

```
renderReport(regions, project = "", pvalueVars = c(`P-values` = "pval"),
  densityVars = NULL, significantVar = mcols(regions)$pval <= 0.05,
  annotation = NULL, nBestRegions = 500, customCode = NULL,
  outdir = "regionExploration", output = "regionExploration",
  browse = interactive(), txdb = NULL, device = "png",
  densityTemplates = list(Pvalue = templatePvalueDensity, Common =
  templateDensity, Manhattan = templateManhattan), template = NULL,
  theme = NULL, digits = 2, ...)

templatePvalueDensity

templatePvalueDensity

templatePvalueHistogram

templateHistogram</pre>
```

Arguments

regions The set of genomic regions of interest as a GRanges object. All sequence lengths

must be provided.

project The title of the project.

pvalueVars The names of the variables with values between 0 and 1 to plot density values by

chromosome and a table for commonly used cutoffs. Most commonly used to explore p-value distributions. If a named character vector is provided, the names

are used in the plot titles.

densityVars The names of variables to use for making density plots by chromosome. Com-

monly used to explore scores and other variables given by region. If a named

character vector is provided, the names are used in the plot titles.

significantVar A logical variable differentiating statistically significant regions from the rest.

When provided, both types of regions are compared against each other to see

differences in width, location, etc.

12 renderReport

annotation The output from matchGenes used on regions. Note that this can take time for

a large set of regions so it's better to pre-compute this information and save it.

nBestRegions The number of regions to include in the interactive table.

customCode An absolute path to a child R Markdown file with code to be evaluated before

the reproducibility section. Its useful for users who want to customize the report by adding conclusions derived from the data and/or further quality checks and

plots.

outdir The name of output directory.

output The name of output HTML file (without the html extension).

browse If TRUE the HTML report is opened in your browser once it's completed.

txdb Specify the transcription database to use for identifying the closest genes via

matchGenes. If NULL it will use TxDb.Hsapiens.UCSC.hg19.knownGene by

default.

device The graphical device used when knitting. See more at http://yihui.name/knitr/options

(dev argument).

densityTemplates

A list of length 3 with templates for the p-value density plots (variables from pvalueVars), the continuous variables density plots (variables from densityVars), and Manhattan plots for the p-value variables (pvalueVars). These templates are processed by whisker.render. Check the default templates for more information. The densityTemplates argument is available for those users interested in customizing these plots. For example, to show histograms instead of density plots use templatePvalueHistogram and templateHistogram instead of

templatePvalueDensity and templateDensity respectively.

template Template file to use for the report. If not provided, will use the default file found

in regionExploration/regionExploration.Rmd within the package source.

theme A ggplot2 theme to use for the plots made with ggplot2.

digits The number of digits to round to in the interactive table of the top nBestRegions.

Note that p-values and adjusted p-values won't be rounded.

. . Arguments passed to other methods and/or advanced arguments. Advanced ar-

guments:

overviewParams A two element list with base_size and areaRel that control

the text size for the genomic overview plots.

output_format Either html_document, pdf_document or knitrBootstrap::bootstrap_documer

unless you modify the YAML template.

clean Logical, whether to clean the results or not. Passed to render.

Format

An object of class character of length 1.

Details

Set output_format to 'knitrBootstrap::bootstrap_document' or 'pdf_document' if you want a HTML report styled by knitrBootstrap or a PDF report respectively. If using knitrBootstrap, we recommend the version available only via GitHub at https://github.com/jimhester/knitrBootstrap which has nicer features than the current version available via CRAN.

If you modify the YAML front matter of template, you can use other values for output_format.

with_wd 13

Value

An HTML report with a basic exploration for the given set of genomic regions.

Author(s)

Leonardo Collado-Torres

Examples

```
## Load derfinder for an example set of regions
library('derfinder')
regions <- genomeRegions$regions</pre>
## Assign chr length
library('GenomicRanges')
seqlengths(regions) <- c('chr21' = 48129895)
## The output will be saved in the 'renderReport-example' directory
dir.create('renderReport-example', showWarnings = FALSE, recursive = TRUE)
## Generate the HTML report
report <- renderReport(regions, 'Example run', pvalueVars = c(</pre>
    'Q-values' = 'qvalues', 'P-values' = 'pvalues'), densityVars = c(
    'Area' = 'area', 'Mean coverage' = 'meanCoverage'),
    significantVar = regions$qvalues <= 0.05, nBestRegions = 20,</pre>
    outdir = 'renderReport-example')
if(interactive()) {
    ## Browse the report
    browseURL(report)
}
## Not run:
## Note that you can run the example using:
example('renderReport', 'regionReport', ask=FALSE)
## End(Not run)
## Check the default templates. For users interested in customizing these
## plots.
## For p-value variables:
cat(templatePvalueDensity)
## For continous variables:
cat(templateDensity)
## For Manhattan plots
cat(templateManhattan)
```

14 with_wd

Description

Temporarily evaluate an expression in a directory, then set the directory back to the original.

Usage

```
with_wd(dir, expr)
```

Arguments

dir a directory to perform an expression within

expr expression to evaluate

Details

See here: http://plantarum.ca/code/setwd-part2/

Author(s)

Tyler Smith, contributed to regionReport by David Robinson https://github.com/dgrtwo

Index

```
*Topic datasets
                                                 theme, 3, 6, 9, 12
    renderReport, 11
                                                 whisker.render, 12
*Topic package
                                                 with_wd, 13
    regionReport-package, 2
analyzeChr, 4
derfinderReport, 2
DESeq, 5, 6
DESeq2Report, 5, 9
DESeqDataSet, 6
DESeqResults, 6
DGEExact, 8
DGEList, 7-9
DGELRT, 8
edgeReport, 7, 8, 9
extendedMapSeqlevels, 3, 4
fullCoverage, 3, 4
load_install, 10
loadCoverage, 3
matchGenes, 12
mergeResults, 2-4
p.adjust, 8
pheatmap, 6, 8
plotCluster, 3
plotIdeogram, 3
plotPCA, 6, 8
regionReport-package, 2
render, 4, 7, 9, 12
renderReport, 11
results, 6, 8, 9
seqlevelsStyle, 3
templateDensity(renderReport), 11
templateHistogram (renderReport), 11
templateManhattan (renderReport), 11
templatePvalueDensity(renderReport), 11
templatePvalueHistogram(renderReport),
        11
```