Package 'orthos'

April 28, 2025

Type Package

Title `orthos` is an R package for variance decomposition using conditional variational auto-encoders

Version 1.6.0

Description `orthos` decomposes RNA-seq contrasts, for example obtained from a gene knock-out or compound treatment experiment, into unspecific and experiment-specific components. Original and decomposed contrasts can be efficiently queried against a large database of contrasts (derived from ARCHS4, https://maayanlab.cloud/archs4/) to identify similar experiments. `orthos` furthermore provides plotting functions to visualize the results of such a search for similar contrasts.

License MIT + file LICENSE

Depends R (>= 4.3), SummarizedExperiment

Imports AnnotationHub, basilisk, BiocParallel, colorspace, cowplot, DelayedArray, dplyr, ExperimentHub, ggplot2 (>= 3.4.0), ggpubr, ggrepel, ggsci, grDevices, grid, HDF5Array, keras, methods, orthosData, parallel, plyr, reticulate, rlang, S4Vectors, stats, tensorflow, tidyr

Suggests BiocManager, BiocStyle, htmltools, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

biocViews RNASeq, DifferentialExpression, GeneExpression

Config/testthat/edition 3

Encoding UTF-8

LazyData false

RoxygenNote 7.3.2

StagedInstall no

git_url https://git.bioconductor.org/packages/orthos

git_branch RELEASE_3_21

git_last_commit 9656fce

git_last_commit_date 2025-04-15

Repository Bioconductor 3.21

Date/Publication 2025-04-27

Author Panagiotis Papasaikas [aut, cre] (ORCID:
<https://orcid.org/0000-0002-1640-7636>),
Charlotte Soneson [aut] (ORCID:
<https://orcid.org/0000-0003-3833-2169>),
Michael Stadler [aut] (ORCID: <https://orcid.org/0000-0002-2269-4934>),
Friedrich Miescher Institute for Biomedical Research [cph]

Maintainer Panagiotis Papasaikas <panagiotis.papasaikas@fmi.ch>

Contents

orthos-package 2
decomposeVar 3
loadContrastDatabase 4
plotQueryResultsManh 5
plotQueryResultsViolin 6
queryWithContrasts 7
testOrthosEnv 9

Index **10**

`orthos-package` *orthos - variance decomposition using conditional variational auto-encoders*

Description

orthos decomposes RNA-seq contrasts, for example obtained from a genetic or compound treatment experiments, into non-specific and experiment-specific components. Original and decomposed contrasts can be efficiently queried against a large database of public contrasts (derived from ARCHS4, https://maayanlab.cloud/archs4/) to identify similar experiments. **orthos** furthermore provides plotting functions to visualize the results of such queries for similar contrasts.

Author(s)

Maintainer: Panagiotis Papasaikas <panagiotis.papasaikas@fmi.ch> (ORCID)

Authors:

- Charlotte Soneson <charlotte.soneson@fmi.ch> (ORCID)
- Michael Stadler <michael.stadler@fmi.ch> (ORCID)

Other contributors:

- Friedrich Miescher Institute for Biomedical Research [copyright holder]

decomposeVar *Decompose input contrasts to decoded and residual fractions*

Description

Decompose input contrasts (gene expression deltas) to decoded (generic) and residual (unique) components according to a contrast encoder-decoder pre-trained on a large corpus of public RNAseq experiments.

Usage

```
decomposeVar(
  M,
  MD = NULL,
  treatm = NULL,
  cntr = NULL,
  processInput = TRUE,
  organism = c("Human", "Mouse"),
  featureType = c("AUTO", "ENSEMBL_GENE_ID", "GENE_SYMBOL", "ENTREZ_GENE_ID",
    "ARCHS4_ID"),
  pseudocount = 4,
  verbose = TRUE
)
```

Arguments

M	Matrix of raw gene counts.
MD	Matrix of gene deltas (optional). If MD is specified, M is assumed to be a raw gene count matrix specifying context for contrasts specified in MD. MD is then a matrix of gene deltas with the same dimensions as M. If MD is specified, treatm and cntr have to be NULL.
treatm, cntr	Vectors indicating column indices in M corresponding to treatments and controls. If treatm and cntr are specified, MD has to be NULL.
processInput	If set to TRUE (default) the count matrix will be preprocessed (library normalized, log2-transformed after addition of a pseudocount, NA values will be set to 0).
organism	Selects the autoencoder model trained on data from this species. One of "Human" or "Mouse".
featureType	Set to "AUTO" for automatic feature id-type detection. Alternatively specify the type of supplied id features. Current supported types are "ENSEMBL_GENE_ID", "GENE_SYMBOL", "ENTREZ_GENE_ID" and "ARCHS4_ID".
pseudocount	Numerical scalar, added to raw counts in M when preprocessInput = TRUE.
verbose	Logical scalar indicating whether to print messages along the way.

Details

When calling `decomposeVar()`, you may see an `ImportError` on the console. This most likely does not have any negative consequences, rather it means that R and python may not be library compatible and that an automated fallback approach is being used (for more details see `testload` argument of `basiliskStart`).

Value

A `SummarizedExperiment` object with the decomposed contrasts in the assays and the decomposed variance as the `colData`.

Author(s)

Panagiotis Papasaikas

Examples

```
MKL1_human <- readRDS(system.file("extdata", "GSE215150_MKL1_Human.rds",
package = "orthos"))

# Specifying M, treatm and cntr:
dec_MKL1_human <- decomposeVar(M = MKL1_human, treatm = c(2, 3), cntr = c(1, 1),
                               organism = "Human", verbose = FALSE)


# Alternatively by specifying M and MD:
pseudocount <- 4
M  <- sweep(MKL1_human, 2,
            colSums(MKL1_human), FUN = "/") * 1e+06
M  <- log2(M + pseudocount)
DeltaM <- M[,c("MKL1","caMKL1")]-M[,"Ctrl"] # Matrix of contrasts
ContextM <- M[,c("Ctrl","Ctrl")] # Matrix with context for the specified contrasts
colnames(ContextM) <- colnames(DeltaM) # M and MD need identical dimnames
RES <- decomposeVar(M = ContextM, MD = DeltaM, processInput = FALSE)
```

---

loadContrastDatabase *Load contrast database*

Description

Load a pre-calculated, organism-specific contrast database and return it as a `SummarizedExperiment`.

Usage

```
loadContrastDatabase(
  organism = c("Human", "Mouse"),
  mode = c("ANALYSIS", "DEMO"),
  mustWork = TRUE
)
```

Arguments

organism	Character scalar selecting the organism for which to load the contrast database. One of "Human" or "Mouse".
mode	When in "ANALYSIS" mode (default) the complete contrast DB is queried. "DEMO" mode employs a small "toy" database for the queries. "DEMO" should only be used for testing/demonstration purposes and never for actual analysis purposes.
mustWork	Logical scalar. If FALSE and the contrast database is not available, return an empty SummarizedExperiment object. If TRUE (the default) and the contrast database is not available, loadContrastDatabase throws an error.

Details

Organism-specific databases are compiled in HDF5SummarizedExperiment objects. The first time 'loadContrastDatabase()' is called for a database either directly or via 'queryWithContrasts()' the required objects will be automatically downloaded from 'ExperimentHub' and cached in the user ExperimentHub directory (see 'ExperimentHub::getExperimentHubOption("CACHE")') using the 'orthosData' companion data-package.

Value

A SummarizedExperiment with pre-calculated contrasts as assays.

Author(s)

Panagiotis Papasaikas, Michael Stadler

Examples

```
# !!!Note!!! mode="DEMO" for demonstration purposes only. Default is mode="ANALYSIS"
SE_mouse_demoDB <- loadContrastDatabase (organism="Mouse", mode="DEMO")
SE_mouse_demoDB

SE_human_demoDB <- loadContrastDatabase (organism="Human", mode="DEMO")
SE_human_demoDB
```

---

plotQueryResultsManh *Visualize query results as a composite manhattan/density plot.*

Description

Visualize query results as a composite manhattan/density plot.

Usage

```
plotQueryResultsManh(queryResults, doPlot = TRUE)
```

Arguments

queryResults	A list containing the results of a query performed with queryWithContrasts
doPlot	Logical scalar specifying if a plot should be generated.

Value

A composite manhattan/density plot for the scores of queries using different contrast components against the respective contrast DBs.

Author(s)

Panagiotis Papasaikas, Michael Stadler

Examples

```
MKL1_human <- readRDS(system.file("extdata", "GSE215150_MKL1_Human.rds",
package = "orthos"))

# Decompose contrasts:
dec_MKL1_human <- decomposeVar(M = MKL1_human, treatm = c(2, 3), cntr = c(1, 1),
                               organism = "Human", verbose = FALSE)

# Perform query against contrast DB with the decomposed fractions.
# !!!Note!!! mode="DEMO" for demonstration purposes only.
params <- BiocParallel::MulticoreParam(workers = 2)
query.res.human <- queryWithContrasts(dec_MKL1_human, organism = "Human",
                                      BPPARAM = params, verbose = FALSE,
                                      mode = "DEMO")

# plot results for individual contrasts using composite Manhattan/Density plots:
ManhDensPlots <- plotQueryResultsManh(query.res.human, doPlot = FALSE)
ManhDensPlots[["caMKL1"]]
```

---

plotQueryResultsViolin

Visualize query results as violin plots

Description

Visualize query results as violin plots

Usage

```
plotQueryResultsViolin(queryResults, doPlot = TRUE)
```

Arguments

queryResults	A list containing the results of a query performed with queryWithContrasts
doPlot	Logical scalar specifying if a plot should be generated.

Value

A list of ggplot violin plots (one for each dataset) for the scores of queries using different contrast components against the respective contrast DBs.

Author(s)

Panagiotis Papasaikas, Michael Stadler

Examples

```
MKL1_human <- readRDS(system.file("extdata", "GSE215150_MKL1_Human.rds",
package = "orthos"))

# Decompose contrasts:
dec_MKL1_human <- decomposeVar(M = MKL1_human, treatm = c(2, 3), cntr = c(1, 1),
                               organism = "Human", verbose = FALSE)

# Perform query against contrast DB with the decomposed fractions.
# !!!Note!!! mode="DEMO" for demonstration purposes only.
params <- BiocParallel::MulticoreParam(workers = 2)
query.res.human <- queryWithContrasts(dec_MKL1_human, organism = "Human",
                                      BPPARAM = params, verbose = FALSE,
                                      mode = "DEMO")

# plot results for individual contrasts using violin plots::
ViolinPlots <- plotQueryResultsViolin(query.res.human, doPlot = FALSE)
ViolinPlots[["caMKL1"]]
```

---

queryWithContrasts *Query the contrast database with a set of contrasts*

Description

Query the contrast database with a set of contrasts

Usage

```
queryWithContrasts(
  contrasts,
  use = c("expressed.in.both", "all.genes"),
  exprThr = 0.25,
  organism = c("Human", "Mouse"),
```

```
  plotType = c("violin", "manh", "none"),
  detailTopn = 10,
  verbose = TRUE,
  BPPARAM = BiocParallel::bpparam(),
  chunk_size = 500,
  mode = c("ANALYSIS", "DEMO")
)
```

Arguments

contrasts	A SummarizedExperiment object with assays containing contrasts named INPUT_CONTRASTS, DECODED_CONTRASTS and RESIDUAL_CONTRASTS (at least one should be present) and context information in an assay named CONTEXT. The latter is only required when use="expressed.in.both". This is typically generated using decomposeVar.
use	Determines if all.genes or genes expressed in both query and target context will be used. Note that "expressed.in.both", though more accurate, is slower.
exprThr	is the quantile in the provided context that determines the expression value above which a gene is considered to be expressed. This same value is then used for thresholding the contrast database. Only applies when use="expressed.in.both".
organism	Uses the 'orthosData' contrast database from this species. One of "Human" or "Mouse".
plotType	Select the type of visualization for the query results "violin", "manh" or "none" to suppress the plotting.
detailTopn	specifies the number of top hits for which metadata will be returned in the TopHits slot of the results.
verbose	Logical scalar indicating whether to print messages along the way.
BPPARAM	BiocParallelParam object specifying how parallelization is to be performed using e.g. MulticoreParam) or SnowParam).
chunk_size	Column dimension for the grid used to read blocks from the HDF5 Matrix. Sizes between 250 and 1000 are recommended. Smaller sizes reduce memory usage.
mode	When in "ANALYSIS" mode (default) the complete contrast DB is queried. "DEMO" mode employs a small "toy" database for the queries. "DEMO" should only be used for testing/demonstration purposes and never for actual analysis purposes.

Value

A list with three elements called "pearson.rhos", "zscores" and "TopHits", containing raw and z-scored Pearson's rho correlation coefficients between the query contrast(s) and the contrasts in the database, as well as detailed metadata for the detailTopn best hits.

Author(s)

Panagiotis Papasaikas

Examples

```
MKL1_human <- readRDS(system.file("extdata", "GSE215150_MKL1_Human.rds",
package = "orthos"))

# Decompose contrasts:
dec_MKL1_human <- decomposeVar(M = MKL1_human, treatm = c(2, 3), cntr = c(1, 1),
                               organism = "Human", verbose = FALSE)

# Perform query against contrast DB with the decomposed fractions.
# !!!Note!!! mode="DEMO" for demonstration purposes only.
params <- BiocParallel::MulticoreParam(workers = 2)
query.res.human <- queryWithContrasts(dec_MKL1_human, organism = "Human",
                                      BPPARAM = params, verbose = FALSE,
                                      mode = "DEMO")
```

`testOrthosEnv` *Test conda environment*

Description

Test conda environment

Usage

```
testOrthosEnv()
```

Value

A list indicating whether keras is available, and the version of TensorFlow.

Author(s)

Charlotte Soneson

Examples

```
testOrthosEnv()
```

Index

basiliskStart, *4*

colData, *4*

decomposeVar, 3

loadContrastDatabase, 4

MulticoreParam, *8*

orthos *(*orthos-package*)*, 2
orthos-package, 2

plotQueryResultsManh, 5
plotQueryResultsViolin, 6

queryWithContrasts, 7

SnowParam, *8*
SummarizedExperiment, *4*

testOrthosEnv, 9