Package 'orthosData'

October 27, 2026

Type Package

Title Data for the orthos package

Version 1.10.0

Description `orthosData` is the companion ExperimentData package to the `orthos` R package for mechanistic studies
using differential gene expression experiments. It provides functions for retrieval from ExperimentHub
and local caching of the models and datasets used internally in orthos.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Imports AnnotationHub, BiocFileCache, ExperimentHub, HDF5Array, stringr, SummarizedExperiment

URL https://github.com/fmicompbio/orthosData

BugReports https://github.com/fmicompbio/orthosData/issues

VignetteBuilder knitr

biocViews ExperimentData, RNASeqData, ExperimentHub

Suggests BiocStyle, ggplot2, knitr, rmarkdown, testthat

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

git_branch RELEASE_3_23

git_last_commit 1bf590e

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-10-27

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

orthosData-package 2
GetorthosContrastDB 3
GetorthosModels 4

Index **7**

`orthosData-package` *orthosData - Data for the orthos package*

Description

orthosData is the companion ExperimentData package to the 'orthos' R package for mechanistic studies using differential gene expression experiments. It provides functions for retrieval from ExperimentHub and local caching of the models and datasets used internally in 'orthos'.

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]

See Also

Useful links:

- https://github.com/fmicompbio/orthosData
- Report bugs at https://github.com/fmicompbio/orthosData/issues

GetorthosContrastDB *Cache an orthosData contrast DB*

Description

Download in cache HDF5 and RDS component files for an othosData contrast database from ExperimentHub. As these are components of a single HDF5SummarizedExperiment object they HAVE to be cached with the exact prefix used at creation time (see HDF5Array::saveHDF5SummarizedExperiment)

Usage

```
GetorthosContrastDB(
  organism = c("Human", "Mouse"),
  mode = c("ANALYSIS", "DEMO"),
  ARCHS4v = "v212",
  verbose = 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 cached. "DEMO" mode caches a small "toy" database for the queries. "DEMO" should only be used for testing/demonstration purposes and never for actual analysis purposes.
ARCHS4v	Version of ARCHS4 used to build the contrastDB.
verbose	Logical scalar indicating whether to print messages along the way.

Details

The orthosData contrast database contains over 100,000 differential gene expression experiments compiled from the ARCHS4 database* of publicly available expression data. Each entry in the database corresponds to a pair of RNAseq samples contrasting a treatment vs a control condition.

A combination of metadata-semantic and quantitative analyses was used to determine the proper assignment of samples to such pairs in ‘orthosData‘.

The ~20,000 gene features/organism used in the database are "sanctioned" according to several criteria (located on canonical chromosomes, no pseudogenes, no ribosomal protein genes, detected in at least a small fraction of the ARCHS4 database).

The orthosData contrast database contains assays with the original contrasts in the form of gene expression log2 CPM ratios (i.e log2 fold changes, log2FCs), precalculated, decoded and residual components of those contrasts using the orthosData models as well as the gene expression context of those contrasts in the form of log2-transformed library normalized counts (i.e log2 counts per million, log2CPMs). It also contains extensive annotation on both the ‘orthos‘ feature genes and the contrasted conditions.

For each organism the DB is stored as an HDF5SummarizedExperiment with an HDF5 component that contains the gene assays and an rds component that contains gene annotation in the rowData and the contrast annotation in the colData.

Note that because of the way that HDF5 datasets and serialized SummarizedExperiments are linked in an HDF5SummarizedExperiment, the two components -although relocatable- need to have the exact same filenames as those used at creation time. In other words the files can be moved (or copied) to a different directory or to a different machine and they will retain functionality as long as both live in the same directory and are never renamed.

All orthosData datasets are publicly available through zenodo.org according to open data standards and licenses under the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Value

the `dirname` of the cached objects.

Author(s)

Panagiotis Papasaikas

References

*Lachmann, Alexander, et al. "Massive mining of publicly available RNA-seq data from human and mouse." Nature communications 9.1 (2018): 1366

See Also

`HDF5Array::saveHDF5SummarizedExperiment`

https://doi.org/10.5281/zenodo.7554914

Examples

```
GetorthosContrastDB(organism = "Mouse", mode="DEMO")

se <- HDF5Array::loadHDF5SummarizedExperiment(dir = ExperimentHub::getExperimentHubOption("CACHE"),
prefix = "mouse_v212_NDF_c100_DEMO")
```

`GetorthosModels` *Cache orthosData models*

Description

Download in cache a set of orthosData keras models from ExperimentHub.

Usage

```
GetorthosModels(
  organism = c("Human", "Mouse"),
  ARCHS4v = "v212",
  verbose = TRUE
)
```

Arguments

organism	Character scalar selecting the organism for which to load the contrast database. One of "Human" or "Mouse".
ARCHS4v	Version of ARCHS4 used to build the contrastDB.
verbose	Logical scalar indicating whether to print messages along the way.

Details

The function pre-caches a set of pre-trained orthosData keras models for a given organism. These are the models required to perform inference in 'orthos'. For each organism they are of three types:

ContextEncoder: The encoder component of a Variational Autoencoder (VAE). Used to produce a latent encoding of a given gene expression profile (i.e context).

- Input is a gene expression vector (shape=N, where N is the number or 'orthos' gene features) in the form of log2-transformed library normalized counts (log2 counts per million, log2CPMs).

- Output is a 64-d latent representation of the context.

DeltaEncoder: The encoder component of a conditional Variational Autoencoder (cVAE).

Used to produce a latent encoding of a contrast between two conditions (i.e delta).

- Input is a vector of gene expression contrasts (shape=N) in the form of gene log2 CPM ratios (log2 fold changes, log2FCs), concatenated with the corresponding context encoding.

- Output is a 512-d latent representation of the contrast, conditioned on the context.

DeltaDecoder: The decoder component of the same cVAE as above. Used to produce the decoded version of the contrast between two conditions.

- Input is the concatenated vector of the delta and context latent encodings.

- Output is the decoded contrast vector (shape=N), conditioned on the context.

For more details on model architecture and use of these models in 'orthos' please refer to the 'orthos' package vignette: vignette("orthosIntro", package = "orthos").

Value

A character vector with the full paths to the cached models

Author(s)

Panagiotis Papasaikas

Examples

```
GetorthosModels(organism = "Mouse")
```

Index

GetorthosContrastDB, 3
GetorthosModels, 4

HDF5Array::saveHDF5SummarizedExperiment, *3, 4*

orthosData *(*orthosData-package*)*, 2
orthosData-package, 2