Package 'oompaData'

July 22, 2025

Title Data to Illustrate OOMPA Algorithms

Version 3.1.5

Date 2025-04-06

Description This is a data-only package to provide example data for other packages that are part of the ``Object-Oriented Microrray and Proteomics Analysis'' suite of packages. These are described in more detail at the package URL.

Depends R (>= 4.4)

License Apache License (== 2.0)

LazyData yes

URL http://oompa.r-forge.r-project.org/

NeedsCompilation no

Author Kevin R. Coombes [aut, cre]

Maintainer Kevin R. Coombes <krc@silicovore.com>

Repository CRAN

Date/Publication 2025-04-06 20:40:02 UTC

Contents

---

clinical.info   *Experimental info for the prostate cancer data set*

Description

This data set provides experimental and clinical information about the (partial) prostate cancer data set included for demonstration purposes as part of the tail.rank.test package. The experiments were two-color glass microarrays printed at Stanford.

Usage

```
data(clinical.info)
```

Format

A data frame with 112 observations on the following 6 variables.

Arrays A factor containing the barcode of the microarray on which the experiment was performed. Each of the 112 entries should be distinct.

Reference A factor describing the reference sample used in each experiment. This was a common reference, so the identifiers here are not meaningful.

Sample A factor identifying the test sample in each experiment. These match the codes published in the original paper.

Status A factor with three levels identifying normal prostate (N), prostate cancer (T), or lymph node metastasis (L).

Subgroups A factor with five levels: I II III N O. These correspond to the groups found in the original paper using clustering.

ChipType a factor with levels new or old. At least two different print designs of microarrays were used in this experiment; this factor identifies the design.

Source

The data was originally described in the paper by Lapointe et al., and downloaded from the Stanford Microarray Database https://bio.tools/stanfordmicroarraydb.

References

Lapointe J et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A*, 101, 811–816.

See Also

expression.data, gene.info

Examples

```
data(clinical.info)
summary(clinical.info)
```

| expression.data | *Microarray expression data on prostate cancer* |

Description

A subset of the microarray data from a study of prostate cancer at Stanford is supplied as demo data with the tail.rank.test package.

Usage

```
data(expression.data)
```

Format

A data frame with 2000 observations on the 112 variables. Each column represent a different patient sample, as described in the accompanying data.frame called clinical.info.

Details

This data set contains normalized microarray expression data on 2000 randomly selected genes from a prostate cancer data set. The study was originially described in a publication by Lapointe et al. The experiments were performed on two-color glass microarrays printed at Stanford and available from the Stanford Microarray Database. We downloaded the raw data and preprocessed it. In particular,after background correction and loess normalization, we computed log ratios between the channels. We then randomly selected 2000 of the 42129 spots to include as demonstration data here.

Source

https://bio.tools/stanfordmicroarraydb

References

Lapointe J et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A*, 101, 811–816.

See Also

clinical.info, gene.info

Examples

```
data(expression.data)
summary(expression.data)
```

gene.info

Gene information for the prostate cancer data set

Description

This data set provides information about the genes included with the (partial) prostate cancer data set as part of the tail.rank.test package.

Usage

```
data(gene.info)
```

Format

A data frame with 2000 observations on the following 6 variables.

ArrayI.Spot a numeric vector; where is this clone spotted on the old arrays

ArrayII.Spot a numeric vector; where is this clone spotted on the new arrays

Clone.ID a factor; the IMAGE clone identifier

Gene.Symbol a factor; the official gene symbol

Cluster.ID a factor; the UniGene cluster number

Accession a factor; the GenBanlk accession number

Source

The data was originally described in the paper by Lapointe et al., and downloaded from the Stanford Microarray Database <https://bio.tools/stanfordmicroarraydb>. We randomly selected 2000 of the 42129 spots to include as demonstration data here.

References

Lapointe J et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc Natl Acad Sci U S A*, 101, 811–816.

See Also

[clinical.info](), [expression.data]()

Examples

```
data(gene.info)
summary(gene.info)
```

lungData-data	*Lung Cancer Gene Expression Dataset*

Description

This data set contains clinical annotations and the log expression of 150 genes for a set of 444 lung cancer patients. The 150 genes were selected randomly from a larger Affymetrix U133A dataset.

Usage

```
data(lungData)
```

Format

A data matrix (`lung.dataset`) containing the log expression of 150 genes (rows) in 444 lung tumor samples (columns), along with a data frame (lung.clinical) containing clinical annotations of the patients.

Source

Supporting data for the Nature Medicine paper by Shedden et al. was downloaded from the (now defunct) caArray web site. The original data used to be available by FTP from the NIH, but can now only be found in the Gene Expression Omnibus at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68571`. The data were log transformed by mapping the expression value x to log2(1+x). A subset of genes and of clinical annotation columns were selected to form this data set.

References

Abrams ZB, Zucker M, Wang M, Asiaee Taheri A, Abruzzo LV, Coombes KR.
Thirty biologically interpretable clusters of transcription factors distinguish cancer type.
BMC Genomics. 2018 Oct 11;19(1):738. doi: 10.1186/s12864-018-5093-z.

Asiaee A, Abrams ZB, Nakayiza S, Sampath D, Coombes KR.
Explaining Gene Expression Using Twenty-One MicroRNAs.
J Comput Biol. 2020 Jul;27(7):1157-1170. doi: 10.1089/cmb.2019.0321.

Index