Package 'fabiaData'

November 10, 2026

Title Data sets for FABIA (Factor Analysis for Bicluster Acquisition)

Version 1.50.0

Date 2012-08-15

Author Sepp Hochreiter <hochreit@bioinf.jku.at>

Maintainer Sepp Hochreiter <hochreit@bioinf.jku.at>

Depends R (>= 2.10.0), Biobase

Imports utils

Suggests fabia

Description Supplying gene expression data sets for the demos of the biclustering method ``Factor Analysis for Bicluster Acquisition'' (FABIA). The following three data sets are provided: A) breast cancer (van't Veer, Nature, 2002), B) multiple tissues (Su, PNAS, 2002), and C) diffuse large-B-cell lymphoma (Rosenwald, N Engl J Med, 2002).

License LGPL (>= 2.1)

URL http://www.bioinf.jku.at/software/fabia/fabia.html

biocViews CancerData, BreastCancerData, MicroarrayData

LazyLoad yes

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

git_branch RELEASE_3_23

git_last_commit d6f2b0a

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-10

Contents

Breast_A 2
CBreast 3
CDLBCL 4

CMulti 4
DLBCL_B 5
fabiaData 6
fabiaDataVersion 6
Multi_A 7
XBreast 8
XDLBCL 9
XMulti 9

Index **11**

`Breast_A` *Subclasses of van't Veer breast cancer microarray data set*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by van't Veer et al. 2002. Array S54 was removed because it is an outlier.

Goal was to find a gene signature to predict the outcome of a cancer therapy that is to predict whether metastasis will occur. A 70 gene signature has been discovered.

Here we want to find subclasses in the data set.

Hoshida et al. 2007 found 3 subclasses and verified that 50/61 cases from class 1 and 2 were ER positive and only in 3/36 from class 3.

`bA` is the data set with 97 samples and 1213 genes, `bAc1` are the three subclasses from Hoshida et al. 2007.

Usage

```
Breast_A
```

Format

Matrix `XBreast`: 97 samples and 1213 probe sets, Vector `CBreast`: three subclasses from Hoshida

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. 'Gene expression profiling predicts clinical outcome of breast cancer', Nature 415:530-536, 2002.

CBreast *Subclasses of van't Veer breast cancer microarray data set*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by van't Veer et al. 2002. Array S54 was removed because it is an outlier.

Goal was to find a gene signature to predict the outcome of a cancer therapy that is to predict whether metastasis will occur. A 70 gene signature has been discovered.

Here we want to find subclasses in the data set.

Hoshida et al. 2007 found 3 subclasses and verified that 50/61 cases from class 1 and 2 were ER positive and only in 3/36 from class 3.

`XBreast` is the data set with 97 samples and 1213 genes, `CBreast` are the three subclasses from Hoshida et al. 2007.

Usage

```
CBreast
```

Format

Vector `CBreast` of 97 samples giving the three subclasses from Hoshida.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. 'Gene expression profiling predicts clinical outcome of breast cancer', Nature 415:530-536, 2002.

CDLBCL *Microarray data set of Rosenwald diffuse large-B-cell lymphoma*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Rosenwald et al. 2002.

Goal was to predict the survival after chemotherapy

Hoshida divided the data set into three classes: "OxPhos" (oxidative phosphorylation), "BCR" (B-cell response), and "HR" (host response).

We want to identify these subclasses.

Usage

```
CDLBCL
```

Format

Vector `CDLBCL` of 180 samples giving the three subclasses according to Hoshida et al. 2007.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', New Engl. J. Med. 346: 1937-1947, 2002.

CMulti *Microarray data set of Su on different mammalian tissue types*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Su et al. 2002.

Gene expression from human and mouse samples across a diverse array of tissues, organs, and cell lines have been profiled. The goal was to have a reference for the normal mammalian transcriptome.

Here we want to identify the subclasses which correspond to the tissue types.

Usage

```
CMulti
```

Format

Vector `CMulti` of 102 samples giving the four classes of tissue types.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, et al. 'Large-scale analysis of the human and mouse transcriptomes', Proc. Natl. Acad. Sci. USA 99:4465-4470, 2002.

DLBCL_B *Microarray data set of Rosenwald diffuse large-B-cell lymphoma*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Rosenwald et al. 2002.

Goal was to predict the survival after chemotherapy.

Hoshida divided the data set into three classes: "OxPhos" (oxidative phosphorylation), "BCR" (B-cell response), and "HR" (host response).

We want to identify these subclasses.

Usage

```
DLBCL_B
```

Format

Matrix `XDLBCL`: 180 samples and 661 probe sets, Vector `CDLBCL`: three subclasses according to Hoshida et al. 2007.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', New Engl. J. Med. 346: 1937-1947, 2002.

`fabiaData` *Display available data sets*

Description

`fabiaData` available data sets.

Usage

```
fabiaData()
```

Author(s)

Sepp Hochreiter

See Also

`Breast_A`, `DLBCL_B`, `Multi_A`,

Examples

```
fabiaData()
```

`fabiaDataVersion` *Display version info for package fabiaData*

Description

`fabiaDataVersion` displays version information about the package.

Usage

```
fabiaDataVersion()
```

Author(s)

Sepp Hochreiter

See Also

Breast_A, DLBCL_B, Multi_A,

Examples

```
fabiaDataVersion()
```

Multi_A *Microarray data set of Su on different mammalian tissue types*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Su et al. 2002.

Gene expression from human and mouse samples across a diverse array of tissues, organs, and cell lines have been profiled. The goal was to have a reference for the normal mammalian transcriptome.

Here we want to identify the subclasses which correspond to the tissue types.

Usage

```
Multi_A
```

Format

Matrix `XMulti`: 102 samples and 5565 probe sets, Vector `CMulti`: four classes of tissue types.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, et al. 'Large-scale analysis of the human and mouse transcriptomes', Proc. Natl. Acad. Sci. USA 99:4465-4470, 2002.

XBreast *Subclasses of van't Veer breast cancer microarray data set*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by van't Veer et al. 2002. Array S54 was removed because it is an outlier.

Goal was to find a gene signature to predict the outcome of a cancer therapy that is to predict whether metastasis will occur. A 70 gene signature has been discovered.

Here we want to find subclasses in the data set.

Hoshida et al. 2007 found 3 subclasses and verified that 50/61 cases from class 1 and 2 were ER positive and only in 3/36 from class 3.

`XBreast` is the data set with 97 samples and 1213 genes, `CBreast` are the three subclasses from Hoshida et al. 2007.

Usage

```
XBreast
```

Format

Matrix `XBreast`: 97 samples and 1213 probe sets.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. 'Gene expression profiling predicts clinical outcome of breast cancer', Nature 415:530-536, 2002.

XDLBCL *Microarray data set of Rosenwald diffuse large-B-cell lymphoma*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Rosenwald et al. 2002.

Goal was to predict the survival after chemotherapy

Hoshida divided the data set into three classes: "OxPhos" (oxidative phosphorylation), "BCR" (B-cell response), and "HR" (host response).

We want to identify these subclasses.

Usage

```
XDLBCL
```

Format

Matrix `XDLBCL`: 180 samples and 661 probe sets.

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. 'The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma', New Engl. J. Med. 346: 1937-1947, 2002.

XMulti *Microarray data set of Su on different mammalian tissue types*

Description

Microarray data from Broad Institute "Cancer Program Data Sets" which was produced by Su et al. 2002.

Gene expression from human and mouse samples across a diverse array of tissues, organs, and cell lines have been profiled. The goal was to have a reference for the normal mammalian transcriptome.

Here we want to identify the subclasses which correspond to the tissue types.

Usage

```
XMulti
```

Format

Matrix `XMulti`: 102 samples and 5565 probe sets

Source

Broad Institute "Cancer Program Data Sets": http://www.broadinstitute.org/cgi-bin/cancer/datasets.cgi

References

Hoshida Y, Brunet J-P, Tamayo P, Golub TR, Mesirov JP, 'Subclass Mapping: Identifying Common Subtypes in Independent Disease Data Sets', PLoS ONE 2(11): e1195, 2007.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, et al. 'Large-scale analysis of the human and mouse transcriptomes', Proc. Natl. Acad. Sci. USA 99:4465-4470, 2002.

Index

∗ **datasets**
Breast_A, 2
CBreast, 3
CDLBCL, 4
CMulti, 4
DLBCL_B, 5
Multi_A, 7
XBreast, 8
XDLBCL, 9
XMulti, 9
∗ **models**
fabiaData, 6
fabiaDataVersion, 6

Breast_A, 2, *6, 7*

CBreast, 3
CDLBCL, 4
CMulti, 4

DLBCL_B, 5, *6, 7*

fabiaData, 6
fabiaDataVersion, 6

Multi_A, *6, 7*, 7

XBreast, 8
XDLBCL, 9
XMulti, 9