Package 'pRolocdata'

April 20, 2016

Type Package

Title Data accompanying the pRoloc package

Version 1.8.0

Author Laurent Gatto and Lisa M. Breckels

Maintainer Laurent Gatto <lg390@cam.ac.uk>

Description Mass-spectrometry based spatial proteomics data sets from
Dunkley et al. (2006), Foster et al. (2006), Tan et
al. (2009), Hall et al. (2009), Trotter et al. (2010),
Ferro et al. (2010), Nikolovski et al. (2012, 2014),
Breckels et al. (2013), Groen et al. (2014) and
Christoforou 2015, and protein complex separation data
from Kristensen et al. (2012), Havugimana et al. (2012),
Kirkwood et al. (2013) and Fabre et al. (2015).

Depends R (>= 2.15), MSnbase

Imports Biobase

Suggests pRoloc

License GPL-2

BugReports https://github.com/lgatto/pRolocdata/issues

URL https://github.com/lgatto/pRolocdata

biocViews ExperimentData, Homo_sapiens_Data, MassSpectrometryData,
Arabidopsis_thaliana_Data, Drosophila_melanogaster_Data,
Mus_musculus_Data, StemCell

NeedsCompilation no

R topics documented:

1

---

andy2011   *Data from a LOPIT experiment on Human Embryonic Kidney fibroblast cells*

Description

This is a LOPIT dataset from a standard LOPIT experimental design on Human Embryonic Kidney (HEK293T) fibroblast cells. See below for more details.

Usage

```
data(andy2011)
data(andy2011goCC)
data(andy2011hpa)
```

Format

The data is an instance of class `MSnSet` from package `MSnbase`.

Details

This is a LOPIT experiment. Normalised intensities for 1371 proteins for eight iTRAQ 8-plex labelled fractions. This dataset was used in testing the phenotype discovery algorithm from Breckels et al., *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation*, J Proteomics, 2013, 88:129-40, see phenoDisco. New phenotype clusters identified from algorithm application are available as pd.2013 feature meta-data and the markers used as input for the analysis are available as markers feature meta-data.

The andy2011goCC instance contains binary assay data. Its columns represent GO CC terms that have been observed for the object's features. A 1 indicates that a GO term has been associated to a given feature (protein); a 0 means not such association was found in the GO ontology.

The andy2011hpa instance contains binary assay data. Its columns represent subcellular locations that have been observed from microscopy images from the Human Protein Atlas, for each protein.

A 1 indicates that a subcellular term has been associated to a given feature (protein); a 0 means not such association was found. This matrix of terms was generated from version 13, released on 11/06/2014 of the Human Protein Atlas.

Source

The data was generated by A. Christoforou at the Cambridge Centre for Proteomics.

<http://www.bio.cam.ac.uk/proteomics/>.

References

Breckels LM, Gatto L, Christoforou A, Groen AJ, Lilley KS and Trotter MWB. *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation.* J Proteomics. 2013 Aug 2;88:129-40. doi: 10.1016/j.jprot.2013.02.019. Epub 2013 Mar 21. PubMed PMID: 23523639

Examples

```
data(andy2011)
andy2011
pData(andy2011)
head(exprs(andy2011))
## Organelle marker proteins
table(fData(andy2011)$markers)
## PhenoDisco assignment results
table(fData(andy2011)$pd.2013)

data(andy2011goCC)
dim(andy2011goCC)
head(featureNames(andy2011goCC))
exprs(andy2011goCC)[1:10, 1:5]
```

---

at_chloro   *The AT_CHLORO data base*

Description

AT_CHLORO is a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins.

The `assayData` contains the raw spectral counts for 3 chloroplastic fractions (the envelope, the stroma and the thylakoids) and for a complete chloroplast sample. The percentage of occurrence in each of the sub-chloroplast fraction as calculated in Ferro et al. (2010) are available as feature meta data (`Percent_ENV`, `Percent_STR` and `Percent_THY`).

Usage

```
data(at_chloro)
```

Format

The data is an instance of class MSnSet from package MSnbase.

Source

Myriam Ferro Exploring the Dynamics of Proteomes (EDyP) Laboratoire Biologie a Grande Echelle (BGE) U1038 INSERM/CEA/UJF Institut de Recherches en Technologies et Sciences pour le Vivant (iRTSV) CEA/Grenoble

References

Ferro M, Brugiere S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, Kieffer-Jaquinod S, Bruley C, Garin J, Joyard J, Masselon C, Rolland N. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol Cell Proteomics. 2010 Jun;9(6):1063-84. Epub 2010 Jan 10. PubMed PMID: 20061580; PubMed Central PMCID: PMC2877971

Examples

```
data(at_chloro)
dim(at_chloro)
pData(at_chloro)
head(exprs(at_chloro))
fvarLabels(at_chloro)
table(fData(at_chloro)$markers)
## check exprs data and 'TotalSpectralCount' feature meta data
all(fData(at_chloro)$TotalSpectralCount == rowSums(exprs(at_chloro)))
## create a set with the percentage of occurrence, as in Ferro et al. 2010
## rows that have no 'TOT' in the feature vars of interest
sel <- apply(fData(at_chloro)[, c("Percent_ENV", "Percent_STR", "Percent_THY")],
            1, function(.x) length(grep("TOT", .x)) == 0)
## new MSnSet
at_chloro2 <- at_chloro[sel, 1:3]
## columns of interest
perc <- c("Percent_ENV", "Percent_STR", "Percent_THY")
## create a new intensity matrix
exprs2 <-  matrix(as.numeric(as.matrix(fData(at_chloro2)[, perc])), ncol
= 3)
colnames(exprs2) <- sampleNames(at_chloro2)
rownames(exprs2) <- featureNames(at_chloro2)
summary(rowSums(exprs2))
exprs(at_chloro2) <- exprs2
validObject(at_chloro2)
```

---

dunkley2006 *LOPIT data from Dunkley et al. 2006*

Description

This is the data from Dunkley et al., *Mapping the Arabidopsis organelle proteome*, PNAS 2006 (PMID 16618929). See below for more details.

Usage

```
data(dunkley2006)
data(dunkley2006goCC)
```

Format

The data is an instance of class `MSnSet` from package `MSnbase`.

Details

This is a LOPIT experiment. Normalised intensities for 689 proteins for four iTRAQ 4-plex labelled fraction and 2 membrane preparation in duplicate (16 samples, see `phenoData(dunkley2006)` for more details) are provided.

Partial least square discriminant analysis (PLSDA) has originally been applied to the test data `fData(dunkley)$markers`); assignment results are available with `fData(dunkley)$assigned` for 5 organelles.

This dataset was also used in testing the phenotype discovery algorithm from Breckels et al., *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation*, J Proteomics, *In Press.*, see `phenoDisco`. New phenotype clusters identified from algorithm application are available as `pd.2013` feature meta-data.

The `dunkley2006goCC` instance contains binary assay data. Its columns represent GO CC terms that have been observed for the object's features. A 1 indicates that a GO term has been associated to a given feature (protein); a 0 means not such association was found in the GO ontology.

Source

Supporting Information on <http://www.pnas.org/content/103/17/6518.abstract>.

References

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. *Mapping the Arabidopsis organelle proteome.* Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518-23. Epub 2006 Apr 17. PubMed PMID: 16618929; PubMed Central PMCID: PMC1458916.

Breckels LM, Gatto L, Christoforou A, Groen AJ, Lilley KS and Trotter MWB. *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation* J Proteomics. *In Press.*

Examples

```
data(dunkley2006)
dunkley2006
phenoData(dunkley2006)
## Input training data (organelle markers)
```

```
table(fData(dunkley2006)$markers)
## PLSDA assignment results
table(fData(dunkley2006)$assigned)
## PhenoDisco results
table(fData(dunkley2006)$pd.2013)
```

E14TG2a   *Data from a LOPIT experiment on Mouse E14TG2a Embryonic Stem Cells*

Description

This is data from a standard LOPIT experimental design on Mouse E14TG2a embryonic stem cells. See below for more details.

Usage

```
data(E14TG2aS1)
data(E14TG2aS2)
data(E14TG2aR)
data(E14TG2aS1yLoc)
data(E14TG2aS1goCC)
```

Format

The data is an instance of class `MSnSet` from package `MSnbase`.

Details

This is a LOPIT experiment. Normalised intensities of proteins from eight iTRAQ 8-plex labelled fractions are available for 2 replicates (indexed 1 and 2) using stringent and relaxed setting (S and R, respectively).

The `E14TG2aS1goCC` instance contains binary assay data. Its columns represent GO CC terms that have been observed for the object's features. A 1 indicates that a GO term has been associated to a given feature (protein); a 0 means not such association was found in the GO ontology.

The `E14TG2aS1yLoc` instance contains 34 sequence and annotation features obtained from a feature selection of the sequence and annotation features from the computational classifier YLoc. These features include: variants of psuedo amino acid counts, autocorrelation, sum of charge, prosite patterns, Gene Ontoloy terms and the number of signal peptides. These features are described in detail in Breckels et al (2015).

Source

The data was generated by A. Christoforou at the Cambridge Centre for Proteomics, Cambridge. http://www.bio.cam.ac.uk/proteomics/.

Examples

```
data(E14TG2aS1)
E14TG2aS1
pData(E14TG2aS1)
head(exprs(E14TG2aS1))
```

---

fabre2015r1 *Data from Fabre et al. 2015*

Description

Duplicated experimental data from Fabre et al. 2015, *Deciphering preferential interactions within supramolecular protein complexes: the proteasome case*. Protein complexes were separated by glycerol density gradient centrifugation. Proteins have been quantified by label-free (iBAQ) mass spectrometry.

Usage

```
data("fabre2015r1")
data("fabre2015r2")
```

References

Fabre B, Lambour T, Garrigues L, Amalric F, Vigneron N, Menneteau T, Stella A, Monsarrat B, Van den Eynde B, Burlet-Schiltz O, Bousquet-Dubouch MP. Deciphering preferential interactions within supramolecular protein complexes: the proteasome case. Mol Syst Biol. 2015 Jan 5;11(1):771. doi: 10.15252/msb.20145497. PubMed PMID: 25561571.

Examples

```
data(fabre2015r1)
experimentData(fabre2015r1)
library("pRoloc")
plot2D(fabre2015r1)
addLegend(fabre2015r1, where = "topright")
```

---

foster2006 *PCP data from Foster et al, 2006*

Description

This is the data from Foster et al., *A Mammalian Organelle Map by Protein Correlation Profiling*, Cell 2006 (PMID 16615899). See below for more details.

Usage

```
data(foster2006)
```

Format

The data is an instance of class `MSnSet` from package `MSnbase`.

Details

This is a PCP experiment. Label-free quantification has been done on a totla of 26 high and low density-separated fractions (see `pData(foster2006)`). A total of 1555 proteins have been quantified in a subset of the fractions. The proteins are described in the `featureData` slot. Chi^2 calculations, as defined in the PCP experiment, has been performed using marker proteins for a total of 8 organelles, as well as the authors' original assignment and notes are available in the `featureData` slot.

Source

Supplemental data on http://www.cell.com/abstract/S0092-8674(06)00369-2.

References

Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M. *A mammalian organelle map by protein correlation profiling.* Cell. 2006 Apr 7;125(1):187-99. PubMed PMID: 16615899.

Examples

```
data(foster2006)
foster2006
phenoData(foster2006)
featureData(foster2006)
## organelle marker proteins
table(fData(foster2006)$train)
```

---

groen2014	*Data from LOPIT experiments on Arabidopsis thaliana roots, from Groen et al (2014)*

Description

This is the data from Groen et al. *Identification of Trans Golgi Network proteins in Arabidopsis thaliana root tissue* J. Proteome Res, 2014, Feb 7; 13(2):763-776. See below for more details.

Usage

```
data(groen2014r1)
data(groen2014r2)
data(groen2014r3)
data(groen2014cmb)
data(groen2014r1goCC)
```

Format

An instance of class `MSnSet` from package `MSnbase`.

Details

This is a LOPIT experiment. Normalised intensities for proteins for four iTRAQ 4-plex labelled fractions are available for 3 replicates (`r1`, `r2` and `r3` respectively). The 3 replicates have also been combined as described in Groen et al. and Trotter et al. (2010) to generate a fourth dataset (`cmb`), also shown in the example code below.

The `groen2014r1goCC` instance contains binary assay data. Its columns represent GO CC terms that have been observed for the object's features. A 1 indicates that a GO term has been associated to a given feature (protein); a 0 means not such association was found in the GO ontology.

Source

http://pubs.acs.org/doi/abs/10.1021/pr4008464

References

Groen AJ, Sancho-Andres G, Breckels LM, Gatto L, Aniento F, and Lilley KS. *Identification of Trans Golgi Network proteins in Arabidopsis thaliana root tissue*. J. Proteome Res, 2014, Feb 7; 13(2):763-776. DOI:10.1021/pr4008464, PMID: 24344820.

Trotter MWB, Sadowski PG, Dunkley TPJ, Groen AJ and Lilley KS. *Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions.* Proteomics 2010 10(23):4213-4219. PMID 21058340.

Sadowski PG, Groen AJ, Dupree P and Lilley KS. *Sub-cellular localization of membrane proteins.* Proteomics 2008 8(19):3991-4011. PMID 18780351.

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. *Mapping the Arabidopsis organelle proteome.* Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518-23. Epub 2006 Apr 17. PubMed PMID: 16618929; PubMed Central PMCID: PMC1458916.

Examples

```
data(groen2014r1)
data(groen2014r2)
data(groen2014r3)
data(groen2014cmb)

## The combine dataset can generated manually using
cmb <- combine(groen2014r1, updateFvarLabels(groen2014r2))
cmb <- filterNA(cmb)
cmb <- combine(cmb, updateFvarLabels(groen2014r3))
cmb <- filterNA(cmb)
fData(cmb) <- fData(cmb)[, c(1,2,5)]
cmb

## or can simply be loaded directly
data(groen2014cmb)
```

```
## check datsets are the same
all.equal(cmb, groen2014cmb, check.attributes=FALSE)
```

---

hall2009   *LOPIT data from Hall et al. 2009*

Description

This is the data from Hall et al. *The Organelle Proteome of the DT40 Lymphocyte Cell Line* Mol Cell Proteomics. 2009 Jun;8(6):1295-305. (PMID: PMC2690488).

Usage

```
data(hall2009)
```

Format

An instance of class MSnSet from package MSnbase.

Details

See reference.

Source

http://www.mcponline.org/content/8/6/1295.abstract

References

Hall SL, Hester S, Griffin JL, Lilley KS, Jackson AP. *The organelle proteome of the DT40 lymphocyte cell line* Mol Cell Proteomics. 2009 Jun;8(6):1295-305. doi: 10.1074/mcp.M800394-MCP200. Epub 2009 Jan 30. PubMed PMID: 19181659; PubMed Central PMCID: PMC2690488.

Examples

```
data(hall2009)
pData(hall2009)
library("pRoloc")
plot2D(hall2009)
```

---

havugimana2012 *Data from Havugimana et al. 2012*

Description

Data from Havugimana et al. 2012, *A census of human soluble protein complexes*. The protein complexes were fractionated by ion exchange chromatography, Isoelectric focusing and sucrose density gradient centrifugation. Proteins were quantified by spectral counting.

Usage

```
data("havugimana2012")
```

References

Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, Wan C, Vlasblom J, Dar VU, Bezginov A, Clark GW, Wu GC, Wodak SJ, Tillier ER, Paccanaro A, Marcotte EM, Emili A. A census of human soluble protein complexes. Cell. 2012 Aug 31;150(5):1068-81. doi: 10.1016/j.cell.2012.08.011. PubMed PMID: 22939629; PubMed Central PMCID: PMC3477804.

Examples

```
data(havugimana2012)
experimentData(havugimana2012)
```

---

hyperLOPIT2015 *Data from the hyperLOPIT technology on Mouse E14TG2a embryonic stem cells*

Description

This is a spatial proteomics dataset from a hyperLOPIT experimental design on Mouse E14TG2a embryonic stem cells

Usage

```
data(hyperLOPIT2015)
data(hyperLOPIT2015ms3r1)
data(hyperLOPIT2015ms3r2)
data(hyperLOPIT2015ms3r3)
data(hyperLOPIT2015ms2)
```

Format

The data is an instance of class MSnSet from package MSnbase.

Details

This is a hyperLOPIT experiment. Normalised intensities for proteins for TMT 10-plex labelled fractions are available for 3 replicates acquired in MS3 mode (hyperLOPIT2015ms3r1, hyperLOPIT2015ms3r2 and hyperLOPIT2015ms3r3) using an Orbitrap Fusion mass-spectrometer. The first two replicates have also been combined as described in Trotter et al (2010) to generate dataset hyperLOPIT2015 to increase organellar resolution. A dataset acquired in MS2 mode has also been acquired (hyperLOPIT2015ms2) which was also generated using the Orbitrap Fusion and using a TMT 10-plex experimental design.

Source

The data was generated by A. Christoforou and C. Mulvey in the Cambridge Centre for Proteomics. http://www.bio.cam.ac.uk/proteomics/.

References

Christoforou A, Mulvey CM, Breckels LM, Hayward PC, Geladaki E, Hurrell T, Naake T, Gatto L, Viner R, Martinez Arias A, Lilley KS. *A draft map of the mouse pluripotent stem cell spatial proteome* Nature Communications, Accepted, October 2015

Examples

```
data(hyperLOPIT2015)
hyperLOPIT2015
pData(hyperLOPIT2015)
head(exprs(hyperLOPIT2015))
```

---

kirkwood2013   *Data from Kirkwood et al. 2013.*

Description

Data from Kirkwood et al. 2013, *Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics*. Protein complexes were separated by size exclusion chromatography and proteins were quantified by spectral counting.

Usage

```
data("kirkwood2013")
```

References

Kirkwood KJ, Ahmad Y, Larance M, Lamond AI. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. Mol Cell Proteomics. 2013 Dec;12(12):3851-73. doi: 10.1074/mcp.M113.032367. Epub 2013 Sep 16. PubMed PMID: 24043423; PubMed Central PMCID: PMC3861729.

Examples

```
data(kirkwood2013)
experimentData(kirkwood2013)
```

---

kristensen2012r1   *Data from Kristensen et al. 2012*

Description

Triplicated experimental data from Kristensen et al. 2012, *A high-throughput approach for measuring temporal changes in the interactome*. Protein complexes were separated by size exclusion chromatography and protein were quantified using SILAC.

Usage

```
data("kristensen2012r1")
data("kristensen2012r2")
data("kristensen2012r3")
```

References

Kristensen AR, Gsponer J, Foster LJ. A high-throughput approach for measuring temporal changes in the interactome. Nat Methods. 2012 Sep;9(9):907-9. doi: 10.1038/nmeth.2131. Epub 2012 Aug 5. PubMed PMID: 22863883; PubMed Central PMCID: PMC3954081.

Examples

```
data(kristensen2012r1)
experimentData(kristensen2012r1)
```

---

nikolovski2012   *Data from Nikolovski et al. 2012*

Description

This is the data used in Nikolovksi et al. (2012). See below for details and references.

Usage

```
data(nikolovski2012)
data(nikolovski2012imp)
```

Format

The data is an instance of class MSnSet from package MSnbase.

Details

These data are a concatenation of 4 LOPIT experiments. Experiments 1 and 2 are from Dunkley et al. 2006 (see also dunkley2006). Exepriments 3 and 4 are new.

In the LOPIT experiments by Dunkley et al. (2006), peripheral membrane proteins were removed by carbonate washing of the isolated membranes, while for experiments 3 and 4, no carbonate wash was performed and are, as such, enriched in peripheral and luminal proteins. See figure 1 in Nikolovski 2012 for a description of the design.

In nikolovksi2012imp missing values have been imputed using partial least-squares regression.

The training set used for the Naive Bayesian classifier is available as the markers feature meta-data. Note that Nikolovksi included a group of markers labelled 'others', which has been retained in these data sets. The results produced in this work are available in the preds feature variable (note that some organelles are marked with a '*', which is undefined here).

Source

Supporting Information on http://www.plantphysiol.org/content/160/2/1037.long, also available in the package's extdata directory.

References

Nikolovski N, Rubtsov D, Segura MP, Miles GP, Stevens TJ, Dunkley TP, Munro S, Lilley KS, Dupree P. *Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics.* Plant Physiol. 2012 Oct;160(2):1037-51. doi: 10.1104/pp.112.204263. Epub 2012 Aug 24. PMID: 22923678; PMCID: PMC3461528.

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. *Mapping the Arabidopsis organelle proteome.* Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518-23. Epub 2006 Apr 17. PubMed PMID: 16618929; PubMed Central PMCID: PMC1458916.

Examples

```
data(nikolovski2012)
data(nikolovski2012imp)
table(is.na(nikolovski2012))
table(is.na(nikolovski2012imp))
phenoData(nikolovski2012)
table(fData(nikolovski2012)$markers)
all.equal(sort(featureNames(nikolovski2012)),
          sort(featureNames(nikolovski2012imp)))
library("pRoloc")
plot2D(nikolovski2012imp)
addLegend(nikolovski2012imp, where = "topright", bty = "n", cex =.7)
```

---

nikolovski2014 *Data from Nikolovski et al. 2014*

Description

This is the data used in Nikolovksi et al. (2014). See below for details and references.

Usage

```
data(nikolovski2014)
```

Format

The data is an instance of class `MSnSet` from package `MSnbase`.

Details

Abstract: The proteomic composition of the Arabidopsis Golgi apparatus is currently reasonably well documented; however little is known about the relative abundances between different proteins within this compartment. Accurate quantitative information of Golgi resident proteins is of great importance: it facilitates a better understanding of the biochemical processes which take place within this organelle, especially those of different polysaccharide synthesis pathways. Golgi resident proteins are challenging to quantify since the abundance of this organelle is relatively low within the cell. In this study an organelle fractionation approach, targeting the Golgi apparatus, was combined with a label free quantitative mass spectrometry (MS), data-independent acquisition (DIA) method employing ion mobility separation known as LC-IMS-MSE (or HDMSE), to simultaneously localize proteins to the Golgi apparatus and assess their relative quantity. In total 102 Golgi localised proteins were quantified. These data provide new insight into Golgi apparatus organization and demonstrate that organelle fractionation in conjunction with label free quantitative MS is a powerful and relatively simple tool to access protein organelle localisation and their relative abundances. The findings presented open a unique view on the organization of the plant Golgi apparatus, leading towards novel hypotheses centered on the biochemical processes of this organelle.

These data are a concatenation of 2 LOPIMS gradients, labelled gradient A and B, each with 10 fractions.

Source

Supplemental Data downloaded from http://www.plantphysiol.org/content/early/2014/08/13/pp.114.245589/suppl/DC1, also available in the package's `extdata` directory.

References

Nikolovski N, Shliaha PV, Gatto L, Dupree P, Lilley KS. Label free protein quantification for plant Golgi protein localisation and abundance. Plant Physiol. 2014 Aug 13. pii: pp.114.245589. [Epub ahead of print] PubMed PMID: 25122472.

Examples

```
data(nikolovski2014)
pData(nikolovski2014)
library("pRoloc")
plot2D(nikolovski2014)
addLegend(nikolovski2014, where = "topright", bty = "n", cex =.7)

A <- pData(nikolovski2014)$gradient == "A"
par(mfrow = c(1, 2))
plot2D(nikolovski2014[, A], main = "Gradient A")
plot2D(nikolovski2014[, !A], main = "Gradient B")
```

---

pRolocdata   *List of pRolocdata data sets*

Description

This function lists the data sets available in pRolocdata package by calling data(package = "pRolocdata").

Usage

```
pRolocdata()
```

Author(s)

Laurent Gatto <lg390@cam.ac.uk>

References

See in the respective data sets' manual pages for references to publications.

Examples

```
pRolocdata()
```

---

pRolocmetadata   *Extract pRoloc metadata*

Description

Extracts relevant metadata from an MSnSet instance. See README.md for a description and explanation of the metadata fields.

Usage

```
pRolocmetadata(x)
```

Arguments

x   A pRolocdata data.

Value

An instance of class pRolocmetadata.

Author(s)

Laurent Gatto

Examples

```
library("pRolocdata")
data(dunkley2006)
data(dunkley2006)
pRolocmetadata(dunkley2006)
```

---

tan2009   *LOPIT data from Tan et al. 2009*

Description

This is the data from Tan et al., *Mapping organelle proteins and protein complexes in Drosophila melanogaster*, J Proteome Res. 2009 Jun;8(6):2667-78. See below for more details.

Usage

```
data(tan2009r1)
data(tan2009r2)
data(tan2009r3)
data(tan2009r1goCC)
```

Format

The data is an instance of class MSnSet from package MSnbase.

Details

This is a LOPIT experiment. Normalised intensities for proteins for four iTRAQ 4-plex labelled fractions are available for 3 replicates (r1, r2 and r3 respectively). The partial least square discriminant analysis results from the paper are available as PLSDA feature meta-data and the markers used in analysis are available as markers feature meta-data (Note: the ER and Golgi organelle markers were combined in original PLSDA analysis).

Replicate 1 was also used in testing the phenotype discovery algorithm from Breckels et al., *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation*, J Proteomics, *In Press.*, see phenoDisco. New phenotype clusters identified from algorithm application are available as pd.2013 feature meta-data.

The `tan2009r1goCC` instance contains binary assay data. Its columns represent GO CC terms that have been observed for the object's features. A 1 indicates that a GO term has been associated to a given feature (protein); a 0 means not such association was found in the GO ontology.

Source

Supporting Information on <http://pubs.acs.org/doi/full/10.1021/pr800866n>

References

Mapping organelle proteins and protein complexes in Drosophila melanogaster. Tan DJ, Dvinge H, Christoforou A, Bertone P, Martinez Arias A, Lilley KS. J Proteome Res. 2009 Jun;8(6):2667-78. PMID: 19317464

Breckels LM, Gatto L, Christoforou A, Groen AJ, Lilley KS and Trotter MWB. *The Effect of Organelle Discovery upon Sub-Cellular Protein Localisation* J Proteomics. *In Press.*

Examples

```
data(tan2009r1)
tan2009r1
pData(tan2009r1)
head(exprs(tan2009r1))
# Organelle markers
table(fData(tan2009r1)$markers)
# PLSDA assignment results
table(fData(tan2009r1)$PLSDA)
```

---

trotter20010 *LOPIT data sets used in Trotter et al. 2010.*

Description

The two Arabidobsis LOPIT data sets `trotter2010shallow` and `trotter2010steep` have been used in Trotter et al. (2010) to illustrate improvement of sub-cellular resolution upon data fusion. The data have originally been published in Dunkley et al. (2006) and Sadowski et al. (2008), respectively.

Usage

```
data(trotter2010shallow)
data(trotter2010steep)
```

Format

The data are instances of class `MSnSet` from package `MSnbase`.

Source

Supporting information available on http://onlinelibrary.wiley.com/doi/10.1002/pmic.201000359/abstract

References

Trotter MWB, Sadowski PG, Dunkley TPJ, Groen AJ and Lilley KS. *Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions.* Proteomics 2010 10(23):4213-4219. PMID 21058340.

Sadowski PG, Groen AJ, Dupree P and Lilley KS. *Sub-cellular localization of membrane proteins.* Proteomics 2008 8(19):3991-4011. PMID 18780351.

Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS. *Mapping the Arabidopsis organelle proteome.* Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6518-23. Epub 2006 Apr 17. PubMed PMID: 16618929; PubMed Central PMCID: PMC1458916.

Examples

```
library(pRoloc)
## Replication of figure 4 from Trotter et al.
## individual data sets
data(trotter2010steep)
data(trotter2010shallow)
## combined data
combined <- combine(trotter2010shallow,
                    trotter2010steep)
par(mfrow = c(1,3))
plot2D(trotter2010shallow, fcol = "TAIR8", main = "Shallow")
plot2D(trotter2010steep, fcol = "TAIR8", main = "Steep")
plot2D(combined, fcol = "TAIR8", main = "Combined")
addLegend(combined, where = "bottomleft",
          fcol = "TAIR8", bty = "n", ncol = 2)
```

Index