Package 'consensus'

December 10, 2024

Title Cross-platform consensus analysis of genomic measurements via interlaboratory testing method

Version 1.25.0

Date 2020-05-22

Author Tim Peters

Maintainer Tim Peters <t.peters@garvan.org.au>

Description An implementation of the American Society for Testing and Materials (ASTM) Standard E691 for interlaboratory testing procedures, designed for cross-platform genomic measurements. Given three (3) or more genomic platforms or laboratory protocols, this package provides interlaboratory testing procedures giving per-locus comparisons for sensitivity and precision between platforms.

Depends R (>= 3.5), RColorBrewer

Imports matrixStats, gplots, grDevices, methods, graphics, stats, utils

biocViews QualityControl, Regression, DataRepresentation, GeneExpression, Microarray, RNASeq

Suggests knitr, RUnit, rmarkdown, BiocGenerics

License BSD_3_clause + file LICENSE

VignetteBuilder knitr

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

git_branch devel

git_last_commit c612c54

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2024-12-10

Contents

`consensus-package` *Cross-platform consensus analysis of genomic measurements via interlaboratory testing method*

Description

An implementation of the American Society for Testing and Materials (ASTM) Standard E691 for interlaboratory testing procedures, designed for cross-platform genomic measurements. Given three (3) or more genomic platforms or laboratory protocols, this package provides interlaboratory testing procedures giving per-locus comparisons for sensitivity and precision between platforms.

Author(s)

Tim J. Peters <t.peters@garvan.org.au>

Examples

```
data("TCGA")
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNA-Seq"),
   data=list(U133A, Huex, Agilent, RNASeq))
fit <- fitConsensus(tcga_mm)
```

`Agilent` *Agilent microarray gene expression data*

Description

Gene expression data from 27 Glioblastoma Multiforme (GBM) patients measured on a custom Agilent Gene Expression Microarray.

Usage

```
data("TCGA")
```

Format

Numeric matrix.

Source

https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/UNC202.txt

References

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., ..., Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 2010, 17(1), 98-110.

Examples

```
data("TCGA")
```

---

consensus-internal *Internal consensus objects and functions*

Description

Internal consensus objects and functions

ConsensusFit-class *Row-linear fit from multiple platforms/conditions - class*

Description

An S4 class that stores parameter value output from `fitConsensus`.

Slots

This class has eight slots, each containing parameters from the row-linear fit:

- `a_i`: Platform-wise average (intercepts).
- `b_i`: Platform-wise sensitivity (slopes).
- `d_i`: Platform-wise precision (residual mean squares). Note that higher values correspond to lower precision.
- `V_a`: Variance of `a_i`. High values indicate high discordance in dynamic range.
- `V_b`: Variance of `b_i`. High values indicate high discordance in sensitivity.
- `V_d`: Averaged precision across platforms.

`z0`: Point of approximate concurrence for all regression lines. Only applicable when `a_i` and `b_i` are highly correlated. See Equations 13.16 and 13.39 of Mandel (2012).

`Vdelta`: Residual variance about the line when `b_i` is regressed against `a_i`. Lower values indicate a higher degree of concurrence, assuming that `a_i` and `b_i` are highly correlated. See Equation 13.36 of Mandel (2012).

Methods

`ConsensusFit` objects have a `show` method that describes the dimensions of the data, in the form: `"ConsensusFit object with i platforms/conditions and k measured loci"`.

Author(s)

Tim Peters <t.peters@garvan.org.au>

References

Mandel, J. (2012). The statistical analysis of experimental data. Courier Corporation. Chapter 13: *The Systematic Evaluation of Measuring Processes*.

See Also

`fitConsensus`: outputs `ConsensusFit` objects.

`fitConsensus` *Fit row-linear models to all loci*

Description

The main function of this package. Fits a number of row-linear models from a `MultiMeasure` object, one for each matching row of the data matrices contained within it. Outputs a `ConsensusFit` object containing per-platform, per locus consensus values for average, sensitivity and precision.

Usage

```
fitConsensus(multimeas)
```

Arguments

`multimeas`	An object of class `MultiMeasure`.

Details

For each locus, a row-linear model (Mandel 1994) is fit of the form

$$Z_{ij} = a_i + b_i(x_j - \bar{x}) + d_{ij}$$

where Z_{ij} is a matrix of measurements at the same genomic locus k, the row index $i = 1, \ldots, p$ labels the platform or condition (microarray, library prep method for sequencing assay etc.) used and the column index $j = 1, \ldots, n$ labels the biological samples that are interrogated at that locus on each of the p platforms. Hence a_i is the intercept (row averages of Z_{ij}), b_i the slope of the regression line (sensitivity) and $d_i = (n-2)^{-1}\Sigma_j d_{ij}^2$ the residual mean square (precision) about the *i*th fitted line, noting that higher d_i corresponds to lower precision. Values of a_i, b_i and d_i can be found in the slots of the `ConsensusFit` object.

For `MultiMeasure` objects with 10,000 loci or more, a progress message is printed for every 10,000 loci fitted.

Value

A `ConsensusFit` object with slots containing various parameter values from the row-linear fits. More information can be found in the linked class description. Output from this function can then be passed to various plotting functions for data exploration.

Author(s)

Tim Peters <t.peters@garvan.org.au>

References

Mandel, J. (1994). Analyzing Interlaboratory Data According to ASTM Standard E691. In *Quality and Statistics: Total Quality Management* (pp. 59-59-12). 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959: ASTM International.

Mandel, J. (2012). The statistical analysis of experimental data. Courier Corporation. Chapter 13: "The Systematic Evaluation of Measuring Processes".

Ku, H.H. (1969). Precision Measurement and Calibration. Volume 1. Statistical Concepts and Procedures (No. NBS-SP-300-VOL-1). Issued February 1969. US Department of Commerce. Chapter 3.7: "The Interlaboratory Evaluation of Testing Methods". Mandel, J. and Lashof, T.W. p. 170.

Examples

```
data("TCGA")
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNA-Seq"),
   data=list(U133A, Huex, Agilent, RNASeq))
fit <- fitConsensus(tcga_mm)
```

Huex — *Affymetrix Huex gene expression data*

Description

Gene expression data from 27 Glioblastoma Multiforme (GBM) patients measured on the Affymetrix HuEx GeneChip.

Usage

```
data("TCGA")
```

Format

Numeric matrix.

Source

https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/LBL202.txt

References

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., ..., Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 2010, 17(1), 98-110.

Examples

```
data("TCGA")
```

MultiMeasure — *MultiMeasure Constructor*

Description

Creates a `MultiMeasure` object from a set of 3 or more numeric matrices, in preparation to pass to `fitConsensus`.

Usage

```
MultiMeasure(names=NA_character_, data=list())
```

Arguments

names	character vector contains the names of each data type (e.g. RNA-Seq, Agilent etc.). Must be the same length as data.
data	list of numeric matrices of identical dim, rownames and colnames where each matrix contains the measurements from the platform/condition described in names. Rows of each matrix correspond to genomic features and columns to samples. Must be the same length as, and have order correspond to, names.

Details

A MultiMeasure contains a list of numeric matrices with identical dimensions and matching row names and column names, to which multiple row-linear models can be fit using fitConsensus. Users should pass a vector of names describing the platform/conditions the genomic measurements are made under, and a corresponding list of matrices to the data argument. A series of validity checks will be made on data correctness and a helpful error message will be returned if the structure does not conform to the above description.

Value

a MultiMeasure object

Author(s)

Tim Peters <t.peters@garvan.org.au>

See Also

MultiMeasure-class

Examples

```
data(TCGA)
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNA-Seq"),
   data=list(U133A, Huex, Agilent, RNASeq))
```

MultiMeasure-class *Multi-platform genomic measurements across the same samples - class*

Description

An S4 class that stores normalised matched genomic data from multiple platforms and/or laboratory conditions (e.g. from microarrays, RNA-Seq and other sequencing assays).

List Components

This class has two slots, names and data.

names: character vector contains the names of each data type (e.g. RNA-Seq, Agilent etc.). Must be the same length as data.

data: list of numeric matrices of identical dim, rownames and colnames where each matrix contains the measurements from the platform/condition described in names. Rows of each matrix correspond to genomic features and columns to samples. Must be the same length as names.

Methods

MultiMeasure objects have a show method that describes the dimensions of the data, in the form: MultiMeasure object with i platforms/conditions, j samples and k measured loci.

Author(s)

Tim Peters <t.peters@garvan.org.au>

See Also

MultiMeasure constructs MultiMeasure objects.

---

plotMarginals *Density plots of per-platform marginal distributions*

Description

Plots a series of marginal densities for each platform for either (a) average, (b) sensitivity or (c) precision.

Usage

```
plotMarginals(consfit,
       param=c("average", "sensitivity", "precision"),
       pal=palette(), xlim=NULL, ...)
```

Arguments

consfit	An object of class ConsensusFit.
param	Whether average (a_i), sensitivity (b_i) or precision (d_i) is plotted.
pal	Colour palette. Length must be at least the number of platforms/conditions.
xlim	Range of values to be plotted. If NULL then the entire density is plotted.
...	Extra arguments passed to legend().

Details

Precision is plotted on the log scale.

Value

A plot to the current device.

Author(s)

Tim Peters <t.peters@garvan.org.au>

Examples

```
data("TCGA")
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNASeq"),
   data=list(U133A, Huex, Agilent, RNASeq))
fit <- fitConsensus(tcga_mm)
plotMarginals(fit, "sensitivity", brewer.pal(n = 4, name = "Dark2"))
```

---

plotMostDiscordant *Plot a heatmap showing a selection of loci*

Description

Plots a heatmap of a specified number of loci showing per-platform, values for either (a) average (a_i), (b) sensitivity (b_i) or (c) precision (d_i) for the most discordant for each. Discordance is ranked by $V(a_i)$, $V(b_i)$ or $\frac{\Sigma(d_i)}{p-1}$ where $p =$ the number of platforms/conditions.

Usage

```
plotMostDiscordant(consfit, param=c("average", "sensitivity", "precision"),
                   numloci=20, pal=colorRampPalette(brewer.pal(9, "RdYlGn")))
```

Arguments

consfit	An object of class ConsensusFit.
param	Whether average (a_i), sensitivity (b_i) or precision (d_i) is plotted.
numloci	The number of loci to plot.
pal	Colour palette. Length must be at least the number of platforms/conditions.

Value

A plot to the current device.

Author(s)

Tim Peters <t.peters@garvan.org.au>

Examples

```
data("TCGA")
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNASeq"),
   data=list(U133A, Huex, Agilent, RNASeq))
fit <- fitConsensus(tcga_mm)
plotMostDiscordant(fit, "sensitivity", 25)
```

`plotOneFit` *Plot a single row-linear fit from a genomic locus*

Description

Plots a series of regressions of platform measurements against their consensus mean.

Usage

```
plotOneFit(multimeas, idx, pal=palette(), ...)
```

Arguments

`multimeas`	An object of class `MultiMeasure`.
`idx`	Row index of the set of matrices in `multimeas`.
`pal`	Color palette. Length must be at least the length of multimeas@data.
`...`	Extra arguments passed to `legend()`.

Details

Visualises a row-linear fit explicitly in the measurement space. Steeper (positive) slopes mean greater sensitivity, and greater scatter around the regression line indicates lower precision.

Value

A plot to the current device.

Author(s)

Tim Peters <t.peters@garvan.org.au>

Examples

```
data("TCGA")
tcga_mm <- MultiMeasure(names=c("U133A", "Huex", "Agilent", "RNASeq"),
   data=list(U133A, Huex, Agilent, RNASeq))
plotOneFit(tcga_mm, "TP53", brewer.pal(n = 4, name = "Dark2"))
```

RNASeq *RNA-Seq gene expression data*

Description

Limma-voom normalised gene expression data from 27 Glioblastoma Multiforme (GBM) patients measured via RNA-Seq.

Usage

```
data("TCGA")
```

Format

Numeric matrix.

Source

https://portal.gdc.cancer.gov/

References

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., ..., Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 2010, 17(1), 98-110.

Examples

```
data("TCGA")
```

U133A *Affymetrix U133A gene expression data*

Description

Log-transformed gene expression data from 27 Glioblastoma Multiforme (GBM) patients measured on the Affymetrix-HT-HG-U133A GeneChip.

Usage

```
data("TCGA")
```

Format

Numeric matrix.

Source

https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/Broad202.txt

References

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., ..., Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, 2010, 17(1), 98-110.

Examples

```
data("TCGA")
```

Index