Package 'BiSEp'

July 2, 2014

Type Package

Title Toolkit to identify synthetic lethality and cell line resistance relationships based on bimodality in gene expression data.

Version 1.0

Date 2013-10-21

Author Mark Wappett

Maintainer Mark Wappett <mark.wappett@astrazeneca.com>

Depends R (>= 3.0.1), mclust (>= 4.2.0), GOSemSim (>= 1.20.0), GO.db (>= 2.10.1), org.Hs.eg.db (>= 2.10.1)

Description Enables the user to infer potential synthetic lethal relationships by analysing relationships between bimodally distributed gene pairs in big gene expression datasets. Enables the user to look for enrichment of resistance to a compound in individual bimodal genes.

License Artistic-2.0

Repository CRAN

NeedsCompilation no

Date/Publication 2014-02-25 10:54:53

R topics documented:

`BiSEp-package` *BiSEp: Bimodality in gene expression to dissect tumours and reveal synthetic lethal drug targets and biomarkers*

Description

A set of tools that enable the user to accurately identify bimodality and non-normality in gene expression data and stratify samples as high or low expression for bimodal genes. Enables identification of candidate synthetic lethal gene pairs. Enables the user to assess functional redundancy between candidate synthetic lethal gene pairs. Enables identification of biomarkers of drug response.

Details

Package:	BiSEp
Type:	Package
Version:	1.0
Date:	2013-10-21
License:	GPL-2

This package has a mixture of CRAN and bioconductor packages listed as dependancies. Please ensure that you have Bioconductor installed.

Author(s)

Author: Mark Wappett

Maintainer: Mark Wappett <mark.wappett@astrazeneca.com>

`BEEM` *BEEM: Bimodal Expression Exclusive with Mutation*

Description

Takes a gene expression matrix and mutation matrix as input. The mutation matrix samples (columns) must mirror or significantly overlap with the gene expression matrix. The data in the mutation matrix must be a discreet 'WT' or 'MUT' call based on the status of each gene with each sample. Detects bimodality and non-normality in all genes across the dataset. Detects mutations of genes enriched in either the high or low gene expression modes.

Usage

```
BEEM(
data = data,
mutData = mutData,
confC = confC,
minMut = minMut
)
```

Arguments

data	This should be a log2 gene expression matrix with genes as rownames and samples as column names. Suitable for gene expression data from any platform - NGS datasets should be RPKM or RSEM values.
mutData	This should be a matrix with genes rownames and samples as column names. All cells should be made up of a discreet 'WT' or 'MUT' call. There should be good overlap (by sample) with the gene expression matrix
confC	The confidence with which you would like to call bimodality in the gene expression dataset. Either 'high' 'med' or 'low' can be input.
minMut	The minimum number of mutations you for a gene would consider for analysis.

Details

The lower confidence calls will dramatically affect the number of gene pairs that the tool produces and increase the false positive rate. The tool will take approximately 10 minutes to run a 5,000 row and 200 column input gene expression matrix using a 'medium,' confidence interval.

Value

A matrix containing 10 columns. Column 1 contains the bimodal genes from the expression data (gene 1) and column 2 contains the mutated candidate synthetic lethal gene pair (gene 2). Columns 3 and 4 contain the number of mutations of gene 2 in the low and high expression modes of gene 1. Column 5 contains the fishers p value that evaluates enrichment of mutation in either the high or low mode (indicated by column 10). Columns 6 and 7 contain the percentage of samples in the low and high expression modes of gene 1 that are mutated for gene 2. Columns 8 and 9 contain information on the overall size (in terms of sample) of the low and high expression modes of gene 1.

Author(s)

Mark Wappett

Examples

```
data(BEEM_EXPRS)
data(BEEM_MUT)
outBEEM <- BEEM(data=BEEM_EXPRS, mutData=BEEM_MUT, confC='high', minMut=10)
```

BEEM_EXPRS *A Log2 Gene Expression matrix*

Description

A Log2 Gene Expression matrix where rownames are genes and colnames are samples

Usage

```
data(BEEM_EXPRS)
```

Format

5,000 observations across 200 variables. All data is log2 transformed RSEM values.

BEEM_MUT *A matrix containing discreet mutation calls*

Description

A matrix containing discreet mutation calls of either 'WT' or 'MUT' where rownames are genes and column names are samples

Usage

```
data(BEEM_MUT)
```

Format

5000 observations across 229 variables.

BEEP *BEEP: Bimodal Expression Exclusive with Pharmacology*

Description

Detects bimodality and non-normality in all genes across the dataset. Integrates second matrix with discreet resistance/ sensitivity calls for any given compound for samples present in the gene expression matrix. Indicates if there is an enrichment of resistance in either the high or low expression modes of any of the genes

Usage

```
BEEP(
data = data,
sensData = sensData,
confC = confC
)
```

Arguments

data	This should be a log2 gene expression matrix with genes as rownames and samples as column names. Suitable for gene expression data from any platform - NGS datasets should be RPKM or RSEM values.
sensData	This should be a 2 column matrix where column 1 contains the sample names from the gene expression matrix and column 2 gives a set of discreet values describing the sensitivity of the sample to compound. A number of these values need to be called 'Resistant' to run the tool
confC	The confidence with which you would like to call bimodality in the gene expression dataset. Either 'high' 'med' or 'low' can be input.

Details

The lower confidence calls will dramatically affect the number of gene pairs that the tool produces and increase the false positive rate. The tool will take approximately 10 minutes to run a 5,000 row and 200 column input gene expression matrix using a 'medium,' confidence interval.

Value

A list of 2 matrices.

Author(s)

Mark Wappett

Examples

```
data(BEEP_input_data)
data(BEEP_sens_data)
outputBEEP <- BEEP(data=BEEP_input_data, sensData=BEEP_sens_data, confC='high')
```

BEEP_input_data *Gene by sample input data matrix.*

Description

Gene by sample input data matrix where rownames are genes and colnames are samples. Scaled log2 gene expression data.

Usage

```
data(BEEP_input_data)
```

Format

1,000 observations across 257 variables.

BEEP_sens_data *Matrix containing sample names and discreet pharmacology calls.*

Description

A matrix containing 2 columns. Column 1 contains sample names and must at least partially match column names of associated gene expression matrix. Column 2 contains discreet pharmacology calls of which a subset should be labelled 'Resistant'.

Usage

```
data(BEEP_sens_data)
```

Format

257 observations across 2 variables.

BIGEE *BIGEE: Bimodal Gene Expression Exclusivity.*

Description

Part of the Synthetic Lethality detection in Genomics toolkit. Detects bimodality and non-normality in all genes across the dataset. Compares all pairwise combinations of bimodal genes and searches for mutually exclusive low expression as evidence of potential synthetic lethality. Scores gene-pairs based on the presence of mutual exclusive bimodality and the distribution of signal intensity across the rest of the dataset.

Usage

```
BIGEE(
data = data,
confC = confC
)
```

Arguments

data	This should be a log2 gene expression matrix with genes as rownames and samples as column names. Suitable for gene expression data from any platform - NGS datasets should be RPKM or RSEM values.
confC	The confidence with which you would like to call bimodality in the gene expression dataset. Either 'high' 'med' or 'low' can be input.

Details

The lower confidence calls will dramatically affect the number of gene pairs that the tool produces and increase the false positive rate. The tool will take approximately 10 minutes to run a 5,000 row and 200 column input matrix using a 'medium,' confidence interval.

Value

A matrix containing three columns. Columns 1 and 2 are the gene symbols that make up the candidate synthetic lethal gene pairs. Column 3 is the score calculated the tool to rank the statistical significance of the gene pairs.

Author(s)

Mark Wappett

Examples

```
data(BIGEE_example_input)
outBIGEE <- BIGEE(data=BIGEE_example_input, confC='high')
```

BIGEE_example_input *A data matrix containing log2 gene expression data*

Description

A data matrix containing log2 gene expression data where genes are rownames and samples are column names

Usage

```
data(BIGEE_example_input)
```

Format

1000 variables across 200 samples.

BISEP *BISEP: Bimodality in Gene Expression data.*

Description

Detects bimodality and non-normality in all genes across the dataset.

Usage

```
BISEP(
data = data,
confC = confC
)
```

Arguments

data	This should be a log2 gene expression matrix with genes as rownames and samples as column names. Suitable for gene expression data from any platform - NGS datasets should be RPKM or RSEM values.
confC	The confidence with which you would like to call bimodality in the gene expression dataset. Either 'high' 'med' or 'low' can be input.

Details

The lower confidence calls will dramatically affect the number of gene pairs that the tool produces and increase the false positive rate. The tool will take approximately 10 minutes to run a 5,000 row and 200 column input matrix using a 'medium,' confidence interval.

Value

A list containing two matrices. Matrix 1 contains the output of the BISEP algorithm - including the midpoint of the bimodal distribution and the associated p value. Matrix 2 contains the output from the BI algorithm - including the delta, pi and BI values.

Author(s)

Mark Wappett

Examples

```
data(BIGEE_example_input)
outBISEP <- BISEP(data=BIGEE_example_input, confC='high')
```

FURE *FURE: Functional redundancy between synthetic lethal gene pairs*

Description

Utilises gene ontology information from the GO database bioconductor package. Assesses gene pairs output from the SLinG and BEEM tools for gene ontology functional redundancy. Performs semantic similarity scoring utilising the GOSemSim bioconductor package

Usage

```
FURE(
x7 = data,
inputType = inputType)
```

Arguments

x7	This should be the output matrix (or similar) from the SLinG and BEEM tools. Columns 1 and 2 should be gene symbols.
inputType	Either 'S' for SLING or 'B' for BEEM based on origin of the input matrix.

Value

A list of matrices containing gene pairs with associated synthetic lethal statistical significance values + gene ontology annotation/ scores.

Author(s)

Mark Wappett

Examples

```
data(FURE_example_input)
outputFURE <- FURE(x7=FURE_example_input, inputType='S')
```

FURE_example_input *Output matrix from the SLinG tool*

Description

Output matrix from the SLinG tool.

Usage

```
data(FURE_example_input)
```

Format

A data frame with 5000 observations across 3 variables.

Index