Package 'ABC.RAP'

October 12, 2022

Title Array Based CpG Region Analysis Pipeline

Version 0.9.0

Author Abdulmonem Alsaleh [cre, aut], Robert Weeks [aut], Ian Morison [aut], RStudio [ctb]

Maintainer Abdulmonem Alsaleh <a.alsaleh@hotmail.co.nz>

Contact Ian Morison <ian.morison@otago.ac.nz>

Description It aims to identify candidate genes that are "differentially methylated" between cases and controls. It applies Student's t-test and delta beta analysis to identify candidate genes containing multiple "CpG sites".

License GPL-3

Depends R (>= 3.1.0)

Imports graphics, stats, utils

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

SystemRequirements GNU make

RoxygenNote 5.0.1.9000

NeedsCompilation no

Repository CRAN

Date/Publication 2016-10-20 10:52:16

R topics documented:

`annotate_data` *Annotating the filtered probes*

Description

This function annotates each filtered probe with gene name, chromosome number, probe location, distance from transcription start site (TSS), and relation to CpG islands. The annotation file is based on "UCSC platform" annotation format and was obtained from Illumina GPL13534_HumanMethylation450_15017482_v1.1 file (BS0010894-AQP_content.bpm).

Usage

```
annotate_data(x)
```

Arguments

x	the filtered probes from filter_data

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
test_data_filtered <- filter_data(test_data)
test_data_annotated <- annotate_data(test_data_filtered)
```

`annotation_file` *annotation file for the 450k probes*

Description

UCSC annotation for the 450k DNA methylation probes. The annotation was obtained from "Illumina GPL13534_HumanMethylation450_15017482_v1.1" file with few amendments on the gene names

Usage

```
data("annotation_file")
```

Format

A data frame

`CpG_hits` *Identifying genes for which multiple CpG sites show significant methylation difference*

Description

This function calculates the number of significantly different CpG sites between cases and controls for each gene and produces a frequency table with genes that have more than one CpG site.

Usage

```
CpG_hits(x)
```

Arguments

`x`	Results from the overlap_data function

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
test_data_filtered <- filter_data(test_data)
test_data_ttest <- ttest_data(test_data_filtered, 1, 2, 3, 4, 1e-3)
test_data_delta_beta <- delta_beta_data(test_data_filtered, 1, 2, 3, 4, 0.5, -0.5, 0.94, 0.06)
test_overlapped_data <- overlap_data(test_data_ttest, test_data_delta_beta)
test_CpG_hits <- CpG_hits(test_overlapped_data)
```

`delta_beta_data` *Applying delta beta analysis to calculate the difference between cases and controls*

Description

This function calculates the delta beta value for the filtered probes. It calculates the difference in mean DNA methylation between cases and controls for each probe. Also, it selects probes with DNA methylation differences that are higher in cases than controls by a user specified meth_cutoff value and differences that are lower in cases than controls by the unmeth_cutoff value. In addition, the function provides the option to specify probes where the average beta value of the cases or controls is greater than a high_meth cutoff value or less than a low_meth cutoff value.

Usage

```
delta_beta_data(x, cases_column_1, cases_column_n, controls_column_1,
  controls_column_n, meth_cutoff, unmeth_cutoff, high_meth, low_meth)
```

Arguments

x	the filtered 450k probes from filter_data function
cases_column_1	The first column (column number) for cases in the filtered dataset
cases_column_n	The last column (column number) for cases in the filtered dataset
controls_column_1	The first column (column number) for controls in the filtered dataset
controls_column_n	The last column (column number) for controls in the filtered dataset
meth_cutoff	The cutoff level for the methylation difference between cases and controls (cases minus controls)
unmeth_cutoff	The cutoff level for the methylation difference between controls and cases (cases minus controls). Consequently, it requires a negative value.
high_meth	The upper margin for the highly methylated probes
low_meth	The lower margin for the low methylation

Examples

```
data(test_data)
data(nonspecific_probes)
test_data_filtered <- filter_data(test_data)
test_data_delta_beta <- delta_beta_data(test_data_filtered, 1, 2, 3, 4, 0.5, -0.5, 0.94, 0.06)
```

---

filter_data *Filtering DNA methylation 450k non_specific probes*

Description

This function filters the reported nonspecific probes, and also filters probes that interrogate SNPs of minor allele frequency (MAF) > 0.1. A list of nonspecific probes was obtained from Chen et al (2013) supplementary files.

Usage

```
filter_data(x)
```

Arguments

x	The normalised beta values in a data matrix format, where conditions are arranged in columns and cg probes are arranged in rows.

References

Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8:203-9.

Examples

```
data(test_data)
data(nonspecific_probes)
test_data_filtered <- filter_data(test_data)
```

---

nonspecific_probes *450k DNA methylation non specific probes*

Description

data frame of the non specific probes that need to be filtered out from 450k array datasets

Usage

```
data("nonspecific_probes")
```

Format

A data frame

Details

These non specific probes interrogates SNPs with mean allelic frequency (MAF) > 0.1, and also those that don't align uniquely on the genome. The list of nonspecific probes was obtained from Chen et al (2013) supplementary files

References

Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8;203-9

overlap_data *Overlapping Student's t-test and delta beta results*

Description

This function overlaps the results from both Student's t-test and delta beta analyses to identify probes (CpG sites) that are highly and significantly different between cases and controls.

Usage

```
overlap_data(x, y)
```

Arguments

x	Results from t-test or delta beta analyses
y	Results from t-test or delta beta analyses

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
test_data_filtered <- filter_data(test_data)
test_data_ttest <- ttest_data(test_data_filtered, 1, 2, 3, 4, 1e-3)
test_data_delta_beta <- delta_beta_data(test_data_filtered, 1, 2, 3, 4, 0.5, -0.5, 0.94, 0.06)
test_overlapped_data <- overlap_data(test_data_ttest, test_data_delta_beta)
```

plot_candidate_genes *Plotting highly different and significant probes annotated by their corresponding gene names*

Description

This function plots the potential candidate genes for which multiple CpG sites show significant difference.

Usage

```
plot_candidate_genes(x)
```

Arguments

x	Results from the overlap_data function

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
test_data_filtered <- filter_data(test_data)
test_data_ttest <- ttest_data(test_data_filtered, 1, 2, 3, 4, 1e-3)
test_data_delta_beta <- delta_beta_data(test_data_filtered, 1, 2, 3, 4, 0.5, -0.5, 0.94, 0.06)
test_overlapped_data <- overlap_data(test_data_ttest, test_data_delta_beta)
plot_candidate_genes(test_overlapped_data)
```

---

plot_data *Overview description of the DNA methylation pattern for cases and controls*

Description

This function produces four distribution plots that summarise the DNA methylation patterns for cases (top left) and controls (top right). The top two histograms show the pattern of mean DNA methylation levels for cases and controls. The bottom two plots show the difference in DNA methylation between cases and controls (a boxplot comparing methylation profile for cases and controls, and a delta beta plot describing the methylation difference between cases and controls). The function also provides summary statistics for the delta beta analysis that can be used to select cutoff values for the delta_beta_data function.

Usage

```
plot_data(x, cases_column_1, cases_column_n, controls_column_1,
  controls_column_n)
```

Arguments

x	The filtered 450k probes from filter_data() function
cases_column_1	The first column (column number) for cases in the filtered dataset
cases_column_n	The last column (column number) for cases in the filtered dataset
controls_column_1	The first column (column number) for controls in the filtered dataset
controls_column_n	The last column (column number) for controls in the filtered dataset

Examples

```
data(test_data)
data(nonspecific_probes)
test_data_filtered <- filter_data(test_data)
plot_data(test_data_filtered, 1, 2, 3, 4)
```

plot_gene *Plotting and exporting methylation profile for candidate genes*

Description

This function explores the DNA methylation profile for any gene. The function generates four plots: the top plots show the difference in DNA methylation between cases and controls (a bar chart of the delta beta values for all probes arranged from 5' to 3' positions and a plot showing the difference in mean DNA methylation between cases and controls). The bottom plots show the distribution of DNA methylation for each probe that interrogates a CpG site in the investigated gene, for cases (left) and controls (right), respectively. Also, an annotation table for the arranged probes is generated with the following columns: probe names, gene name, distance from TSS, mean methylation for cases, mean methylation for controls, delta beta values (cases minus controls), and t-test p.values.

Usage

```
plot_gene(x, b, cases_column_1, cases_column_n, controls_column_1,
  controls_column_n)
```

Arguments

x	The filtered and annotated 450k probes
b	Gene name between quotation marks
cases_column_1	The first column (column number) for cases in the filtered dataset
cases_column_n	The last column (column number) for cases in the filtered dataset
controls_column_1	The first column (column number) for controls in the filtered dataset
controls_column_n	The last column (column number) for controls in the filtered dataset

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
test_data_filtered <- filter_data(test_data)
test_data_annotated <- annotate_data(test_data_filtered)
KLHL34 <- plot_gene(test_data_annotated, 'KLHL34', 1, 2, 3, 4)
```

process.ABC.RAP *An automated analysis applying all ABC.RAP functions in one script*

Description

This function processes the ABC.RAP workflow automatically

Usage

```
process.ABC.RAP(x, cases_column_1, cases_column_n, controls_column_1,
  controls_column_n, ttest_cutoff, meth_cutoff, unmeth_cutoff, high_meth,
  low_meth)
```

Arguments

x	The normalised beta values in a data matrix format, where conditions are arranged in columns and cg probes are arranged in rows.
cases_column_1	The first column (column number) for cases in the filtered dataset
cases_column_n	The last column (column number) for cases in the filtered dataset
controls_column_1	The first column (column number) for controls in the filtered dataset
controls_column_n	The last column (column number) for controls in the filtered dataset
ttest_cutoff	The cutoff level to filter insignificant p-values
meth_cutoff	The cutoff level for the methylation difference between cases and controls (cases minus controls)
unmeth_cutoff	The cutoff level for the methylation difference between controls and cases (controls minus cases). Consequently, it requires a negative value.
high_meth	The upper margin for the highly methylated probes
low_meth	The lower margin for the low methylation

Examples

```
data(test_data)
data(nonspecific_probes)
data(annotation_file)
process.ABC.RAP(test_data, 1, 2, 3, 4, 1e-3,  0.5, -0.5, 0.94, 0.06)
```

test_data *test dataset of 450k DNA methylation*

Description

This is a small dataset of 450k DNA methylation array with 10000 probes. The dataset has four columns; columns 1 and 2 contain normalised beta values for paediatric B ALL cases, and columns 3 and 4 contain beta values for controls (remission cases)

Usage

```
data("test_data")
```

Format

A data frame

Details

a small test dataset

References

Busche S, Ge B, Vidal R, etc. Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukaemia. Cancer Research 2013; 73(14); 4323-4336

ttest_data *applying t-test analysis*

Description

This function applies "two.sided", unequal variance Student's t-test analysis for each probe comparing cases and controls. A cutoff for p-values can be entered to minimise multiple testing bias to filter insignificant p-values.

Usage

```
ttest_data(x, cases_column_1, cases_column_n, controls_column_1,
  controls_column_n, ttest_cutoff)
```

Arguments

x	The filtered 450k probes from filter_data() function
cases_column_1	The first column (column number) for cases in the filtered dataset
cases_column_n	The last column (column number) for cases in the filtered dataset
controls_column_1	The first column (column number) for controls in the filtered dataset
controls_column_n	The last column (column number) for controls in the filtered dataset
ttest_cutoff	The cutoff level to filter insignificant p-values

Examples

```
data(test_data)
data(nonspecific_probes)
test_data_filtered <- filter_data(test_data)
test_data_ttest <- ttest_data(test_data_filtered, 1, 2, 3, 4, 1e-3)
```

Index