Package 'geneXtendeR'

October 17, 2017

Type Package

Version 1.2.0

Title Optimal Gene Extensions From Histone Modification ChIP-seq Data

Description geneXtendeR is designed to optimally annotate a histone
modification ChIP-seq peak input file with functionally
important genomic features (e.g., genes associated with peaks)
based on optimization calculations. geneXtendeR optimally
extends the boundaries of every gene in a genome by some
genomic distance (in DNA base pairs) for the purpose of
flexibly incorporating cis-regulatory elements (CREs), such as
enhancers and promoters, as well as downstream elements that
are important to the function of the gene relative to an
epigenetic histone modification ChIP-seq dataset. geneXtender
computes optimal gene extensions tailored to the broadness of
the specific epigenetic mark (e.g., H3K9me1, H3K27me3), as
determined by a user-supplied ChIP-seq peak input file. As
such, geneXtender maximizes the signal-to-noise ratio of
locating genes closest to and directly under peaks. By
performing a computational expansion of this nature, ChIP-seq
reads that would initially not map strictly to a specific gene
can now be optimally mapped to the regulatory regions of the
gene, thereby implicating the gene as a potential candidate,
and thereby making the ChIP-seq experiment more successful.
Such an approach becomes particularly important when working
with epigenetic histone modifications that have inherently
broad peaks.

Maintainer Bohdan Khomtchouk <khomtchoukmed@gmail.com>

URL https://github.com/Bohdan-Khomtchouk/geneXtendeR

BugReports https://github.com/Bohdan-Khomtchouk/geneXtendeR/issues

Depends rtracklayer, R (>= 3.3.1)

Imports data.table, dplyr, graphics, utils

Suggests BiocStyle, knitr, rmarkdown

VignetteBuilder knitr

License GPL (>= 3)

LazyData TRUE

biocViews ChIPSeq, Genetics, Annotation, GenomeAnnotation,
        DifferentialPeakCalling, Coverage, PeakDetection, ChipOnChip,
        HistoneModification, DataImport

RoxygenNote 5.0.1

NeedsCompilation yes

Author Bohdan Khomtchouk [aut, cre]

R topics documented:

---

annotate   *Annotate peaks file.*

Description

Annotate a user's peaks file (which has been preprocessed with the peaksInput() command) with gene information based on optimally chosen geneXtendeR upstream extension file. This command requires a preprocessed "peaks.txt" file (generated using peaksInput()) to be present in the user's working directory, otherwise the user is prompted to rerun the peaksInput() command in order to regenerate it.

Usage

```
annotate(organism, extension)
```

Arguments

organism   Object name assigned from readGFF() command.

extension   Desired upstream extension.

Value

The gene coordinates are extended by 'extension' at the 5-prime end, and by 500 bp at the 3-prime end. The peaks file is then overlayed on these new gene coordinates, producing a file of peaks annotated with gene ID, gene name, and gene-to-peak genomic distance (in bp). Distance is calculated between 5-prime end of gene and 3-prime end of peak.

Examples

```
rat <- readGFF("ftp://ftp.ensembl.org/pub/release-84/gtf/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.84.gt
fpath <- system.file("extdata", "somepeaksfile.txt", package="geneXtendeR")
peaksInput(fpath)
annotate(rat, 2500)
```

---

barChart   *Produces bar charts.*

Description

Makes bar graphs showing the number of genes under peaks at various upstream extension levels.

Usage

```
barChart(organism, start, end, by)
```

Arguments

organism	Object name assigned from readGFF() command.
start	Lower bound of upstream extension.
end	Upper bound of upstream extension.
by	Interval between consecutive extensions.

Value

Creates bar charts.

Examples

```
rat <- readGFF("ftp://ftp.ensembl.org/pub/release-84/gtf/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.84.gt
fpath <- system.file("extdata", "somepeaksfile.txt", package="geneXtendeR")
peaksInput(fpath)
barChart(rat, 1000, 3000, 100)
```

---

distinct   *Finds unique genes under peaks.*

Description

Determines what genes directly under peaks are actually unique between two different upstream extension levels.

Usage

```
distinct(organism, start, end)
```

Arguments

`organism`	Object name assigned from readGFF() command.
`start`	Lower bound of upstream extension.
`end`	Upper bound of upstream extension.

Details

V1-V3 denote the chromosome/start/end positions of the peaks, V4-V6 denote the respective values of the genes, V7 is the gene ID (e.g., Ensembl ID), V8 is the gene name, and V9 is the distance of peak to nearest gene.

Value

A table of unique genes located under peaks between two upstream extension levels.

Examples

```
rat <- readGFF("ftp://ftp.ensembl.org/pub/release-84/gtf/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.84.gt
fpath <- system.file("extdata", "somepeaksfile.txt", package="geneXtendeR")
peaksInput(fpath)
distinct(rat, 2000, 3000)
```

---

linePlot   *Produces line plots.*

Description

Makes differential line plots showing the differences in the number of genes under peaks at consecutive upstream extension levels.

Usage

```
linePlot(organism, start, end, by)
```

Arguments

`organism`	Object name assigned from readGFF() command.
`start`	Lower bound of upstream extension.
`end`	Upper bound of upstream extension.
`by`	Interval between consecutive extensions.

Value

Creates differential line plots.

Examples

```
rat <- readGFF("ftp://ftp.ensembl.org/pub/release-84/gtf/rattus_norvegicus/Rattus_norvegicus.Rnor_6.0.84.gt
fpath <- system.file("extdata", "somepeaksfile.txt", package="geneXtendeR")
peaksInput(fpath)
linePlot(rat, 1000, 3000, 100)
```

---

peaksInput   *Preprocesses a peaks input file.*

Description

Takes your tab-delimited 3-column (chromosome number, peak start, and peak end) input file (see ?samplepeaksinput) consisting of peaks called from a peak caller (e.g., MACS2 or SICER) and sorts the file by chromosome and start position, thereby creating a preprocessed file for downstream geneXtendeR analysis. This file (called "peaks.txt") is a preprocessed file of the original input and is deposited in the user's working directory and used for the remainder of the analysis. In this "peaks.txt" file, peaks are sorted by chromosome number and start position, where the X chromosome is designated by the integer 100, the Y chromosome by the integer 200, and the mitochondrial chromosome by the integer 300.

Usage

```
peaksInput(filename)
```

Arguments

filename   Name of file containing peaks that have been generated from a peak caller (e.g., MACS2, SICER). See ?samplepeaksinput for an example of such an input file.

Value

Returns a formatted file (called "peaks.txt") that has been preprocessed in preparation for usage with barChart(), linePlot(), distinct(), and other downstream commands and deposited in the user's working directory.

Examples

```
?samplepeaksinput  #Documentation of some exemplary sample input
data(samplepeaksinput)
head(samplepeaksinput)
tail(samplepeaksinput)
fpath <- system.file("extdata", "somepeaksfile.txt", package="geneXtendeR")
peaksInput(fpath)
```

---

rat	*Gene transfer format (GTF) file for rat (Rattus_norvegicus.Rnor_6.0.84)*

Description

A dataset downloaded from Ensembl that contains the entries of a GTF file for Rattus norvegicus.

Usage

```
data(rat)
```

Format

A data frame with 748514 rows and 28 variables corresponding to the entries of a GTF file. Column names are standardized and can be found here: http://www.ensembl.org/info/website/upload/gff.html.

Value

Demonstrates a rat GTF file downloaded from: ftp://ftp.ensembl.org/pub/release-84/gtf/rattus_norvegicus/Rattus_norveg

Examples

```
head(rat)
tail(rat)
```

---

samplepeaksinput	*Sample peaks list to be used as input to geneXtendeR*

Description

A dataset containing the chromosome number, start and stop positions of ChIP-seq peaks along the Rattus norvegicus genome (rn6 assembly). A dataset like this may be used as input to the peaksInput() command, which will sort the dataset by chromosome number and start position.

Usage

```
data(samplepeaksinput)
```

Format

A data frame with 25089 rows and 3 variables:

chr Chromosome number

start Peak start position [in units of base pairs]

end Peak end position [in units of base pairs]

Value

Demonstrates a sample peaks file used as input.

Examples

```
head(samplepeaksinput)
tail(samplepeaksinput)
```

Index