

Package ‘geneAttribution’

January 3, 2025

Type Package

Title Identification of candidate genes associated with genetic variation

Version 1.33.0

Date 2016-10-16

Author Arthur Wuster

Maintainer Arthur Wuster <wustera@gene.com>

Description Identification of the most likely gene or genes through which variation at a given genomic locus in the human genome acts. The most basic functionality assumes that the closer gene is to the input locus, the more likely the gene is to be causative. Additionally, any empirical data that links genomic regions to genes (e.g. eQTL or genome conformation data) can be used if it is supplied in the UCSC .BED file format.

License Artistic-2.0

biocViews SNP, GenePrediction, GenomeWideAssociation, VariantAnnotation, GenomicVariation

Imports utils, GenomicRanges, org.Hs.eg.db, BiocGenerics, GenomeInfoDb, GenomicFeatures, IRanges, rtracklayer

Suggests TxDb.Hsapiens.UCSC.hg38.knownGene, TxDb.Hsapiens.UCSC.hg19.knownGene, knitr, rmarkdown, testthat

RoxygenNote 5.0.1

VignetteBuilder knitr

git_url <https://git.bioconductor.org/packages/geneAttribution>

git_branch devel

git_last_commit fec55bb

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2025-01-03

Contents

| | |
|---------------------------|----------|
| geneAttribution | 2 |
| geneModels | 3 |
| loadBed | 4 |
| normP | 4 |
| Index | 6 |

| | |
|-----------------|---|
| geneAttribution | <i>geneAttribution: Identification of candidate genes associated with noncoding genetic variation</i> |
|-----------------|---|

Description

Identification of the most likely gene or genes through which variation at a given genomic locus in the human genome acts. The most basic functionality assumes that the closer gene is to the input locus, the more likely the gene is to be causative. Additionally, any empirical data that links genomic regions to genes (e.g. eQTL or genome conformation data) can be used if it is supplied in UCSC .bed file format. A typical workflow requires loading gene models and empirical data, then running geneAttribution() on the locus of interest

Given genomic coordinate, return normalized probability for each gene

Usage

```
geneAttribution(chr, pos, geneCoordinates, empiricalData, lambda = 7.61e-06,
  maxDist = 1e+06, minPP = 0.01)
```

Arguments

| | |
|-----------------|--|
| chr | A character string representing a chromosome (e.g. "chr2") |
| pos | An integer representing a genomic position in the same genome build that gene models |
| geneCoordinates | A GenomicRanges object, as generated by geneModels() |
| empiricalData | A list of GenomicRanges objects, as generated by loadBed(). Optional |
| lambda | Float. Variable for exponential distribution. Default based on empirical eQTL data from multiple tissues. Optional |
| maxDist | Integer. Only genes within this distance of query locus are considered. Optional |
| minPP | Float. Genes with a posterior probability < minPP are lumped as "Other". Can be set to 0 when all genes should be reported. Optional |

Value

A sorted, numeric vector of normalized gene probabilities

Examples

```
geneLocs <- geneModels()
fileName <- system.file("extdata", "eqtlHaplotypeBlocks.b38.bed", package="geneAttribution")
empirical <- loadBed(fileName)
geneAttribution("chr2", 127156000, geneLocs, empirical)
```

| | |
|------------|-------------------------|
| geneModels | <i>Load gene models</i> |
|------------|-------------------------|

Description

Get gene models as a GenomicRanges object, with gene names in the symbol column For hg19, you may want to use TxDb.Hsapiens.UCSC.hg19.knownGene and for GRCh38, TxDb.Hsapiens.UCSC.hg38.knownGene (set as default)

Usage

```
geneModels(txdb = TxDb.Hsapiens.UCSC.hg38.knownGene::TxDb.Hsapiens.UCSC.hg38.knownGene,
maxGeneLength = 1e+06, genesToInclude, genesToExclude)
```

Arguments

| | |
|----------------|---|
| txdb | GenomicFeatures TxDb object containing genomic coordinates of genes |
| maxGeneLength | An integer. Gene models that are longer than this are excluded. Optional |
| genesToInclude | A character vector of gene symbols of genes to include (e.g. only protein coding genes). Optional |
| genesToExclude | A character vector of gene symbols of genes to exclude. Optional |

Value

A GenomicRanges object containing human gene models

Examples

```
geneModels()
geneModels(genesToInclude = c("CYBR1", "ADAMTS1", "ADAMTS5", "N6AMT1", "LTN1"))
```

| | |
|---------|--|
| loadBed | <i>Load UCSC *.BED files containing empirical data</i> |
|---------|--|

Description

Required *.BED file format (tab-separated): chr start end name (optional column: score). Sample files supplied with package are limited to chromosome 2.

Usage

```
loadBed(files, weights)
```

Arguments

| | |
|---------|--|
| files | A character vector containing *.BED file names |
| weights | An integer vector containing weighting for each bed file. Optional |

Value

A list of GenomicRanges objects containing the data from the *.BED files, with weightings in the score column

Examples

```
fileName1 <- system.file("extdata", "hiCRegions.b38.bed", package="geneAttribution")
fileName2 <- system.file("extdata", "eqlHaplotypeBlocks.b38.bed", package="geneAttribution")
loadBed(c(fileName1, fileName2), c(2, 5))
```

| | |
|-------|---|
| normP | <i>Normalize likelihoods and return probabilities</i> |
|-------|---|

Description

Normalize likelihoods and return probabilities

Usage

```
normP(pVector, minPP = 0)
```

Arguments

| | |
|---------|--|
| pVector | A numeric vector of pre-normalization likelihoods |
| minPP | A float. Genes with a posterior probability < minPP are lumped as "Other". Optional |

normP

5

Value

A sorted, numeric vector of normalized probabilities

Examples

```
normP(c(5, 1, 1, 1, 1, 1, 0.1))  
normP(c(5, 1, 1, 1, 1, 1, 0.1), minPP=0.1)
```

Index

geneAttribution, [2](#)
geneAttribution-package
 (geneAttribution), [2](#)
geneModels, [3](#)

loadBed, [4](#)

normP, [4](#)