Package 'tLOH'

April 3, 2022

Version 1.2.0

Type Package

Date 2021-05-5

Title Assessment of evidence for LOH in spatial transcriptomics pre-processed data using Bayes factor calculations

Description tLOH, or transcriptomicsLOH, assesses evidence for loss of heterozygosity (LOH) in pre-processed spatial transcriptomics data. This tool requires spatial transcriptomics cluster and allele count information at likely heterozygous single-nucleotide polymorphism (SNP) positions in VCF format. Bayes factors are calculated at each SNP to determine likelihood of potential loss of heterozygosity event. Two plotting functions are included to visualize allele fraction and aggregated Bayes factor per chromosome. Data generated with the 10X Genomics Visium Spatial Gene Expression platform must be pre-processed to obtain an individual sample VCF with columns for each cluster. Required fields are allele depth (AD) with counts for reference/alternative alleles and read depth (DP).

License MIT + file LICENSE

URL https://github.com/USCDTG/tLOH

Encoding UTF-8

Suggests knitr, rmarkdown

Depends R (>= 4.0)

Imports scales, stats, utils, ggplot2, data.table, purrr, dplyr, VariantAnnotation, GenomicRanges, MatrixGenerics

VignetteBuilder knitr

BugReports https://github.com/USCDTG/tLOH/issues

biocViews CopyNumberVariation, Transcription, SNP, GeneExpression, Transcriptomics

RoxygenNote 7.1.1

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

git_branch RELEASE_3_14

git_last_commit 628ddd6

git_last_commit_date 2021-10-26

Date/Publication 2022-04-03

Author Michelle Webb [cre, aut],
David Craig [aut]

Maintainer Michelle Webb <michelgw@usc.edu>

R topics documented:

`aggregateCHRPlot` *Visualization of data output from the tLOHCalc function, aggregated per chromosome*

Description

Output is a plot of the sum of Log10(1/K) values (K is a Bayes factor) per chromosome for each cluster. The dotted line at y=3 represents threshold for substantial evidence toward Model 2

Arguments

`df`	An input dataframe with merged cluster data output by tLOHCalc
`sample`	Name of sample for plot title

Value

Output is a plot where the y axis is sum of Log10(1/K) values (K is a Bayes factor) per chromosome and the x axis is chromosome

Author(s)

Michelle Webb

Examples

```
data('humanGBMsampleAC')
df <- tLOHCalc(humanGBMsampleAC)
aggregateCHRPlot(df,"Example")
```

`alleleFrequencyPlot` *Visualization of data output from the tLOHCalc function*

Description

Creates a plot with panels for each cluster. The x-axis is chromosome, y-axis is allele frequency. Point color is Log10(1/K) where K is a Bayes factor

Arguments

`df`	An input dataframe with merged cluster data
`sample`	Name of sample for plot title

Value

Output is a plot of allele frequency for each cluster. Can be assigned to object and visualized individually. For each panel, the y axis has a min of 0 and max of 1

Author(s)

Michelle Webb

Examples

```
data('humanGBMsampleAC')
df <- tLOHCalc(humanGBMsampleAC)
alleleFrequencyPlot(df,"Example")
```

`humanGBMsampleAC` *Imported dataset of a human glioblastoma spatial transcriptomics sample processed with tLOHImportData.*

Description

A dataset of a human glioblastoma sample containing the allele count (AC) information for 9 spatial transcriptomics clusters

Usage

```
data("humanGBMsampleAC")
```

Format

A data frame with 34601 rows and 7 variables:

rsID dbSNP rs identifier

CLUSTER cluster number

TOTAL total number of counts

REF counts for the reference allele

ALT counts for the alternative allele

CHR chromosome number

POS genomic position

Source

Craig Lab data repository

Examples

```
data("humanGBMsampleAC")
```

`marginalM1Calc` *Calculate marginal of Model 1*

Description

This function takes the number of counts for a reference allele as x, and the number of total allele counts as y.

Arguments

`x`	Number of counts for the reference allele.
`y`	Number of counts total at this SNP position.

Details

The reference and total counts should come from a .csv output by the spatial LOH pre-processing pipeline.

Value

The value returned from marginalM1Calc is numeric

Author(s)

Michelle Webb

Examples

```
marginalM1Calc(10, 0.5)
```

marginalM2CalcBHET *Calculation of marginal M2 het*

Description

Calculation of marginal M2 het

Usage

```
marginalM2CalcBHET(x, a, b)
```

Arguments

x	Number of counts for the reference allele
a	Alpha value
b	Beta value

Value

The value returned from marginalM2CalcBHET is numeric

Author(s)

Michelle Webb

Examples

```
save <- data.frame(REF=c(10,2,3,4,5,10),TOTAL=c(20,20,20,20,20,20))
apply(save, MARGIN = 1, FUN = marginalM2CalcBHET, a = 10,b = 10)
```

marginalM2CalcBLOH *Marginal M2 Calculation*

Description

Calculation of the marginal for Model 2

Usage

```
marginalM2CalcBLOH(x, a, b)
```

Arguments

x	Counts for the reference allele
a	Alpha value
b	Beta value

Value

The value returned from marginalM2CalcBLOH is numeric

Author(s)

Michelle Webb

Examples

```
test <- data.frame(REF=c(10,2,3,4,5,10),TOTAL=c(20,20,20,20,20,20))
apply(test, MARGIN = 1, FUN = marginalM2CalcBLOH, a = 10,b = 10)
```

---

removeOutlierFromCalc *Removes outliers*

Description

Take rows with a total count greater than 2000 and sets to NA

Arguments

dataframe	input dataframe
cols	which column
rows	which row
newValue	what to replace

Value

Dataframe returned

Author(s)

Michelle Webb

Examples

```
test <- data.frame(TOTAL=c(2000,20,20,20,20,20))
removeOutlierFromCalc(test,"TOTAL",test[test$TOTAL > 2000,],NA)
```

`tLOHCalc` *Assement of evidence for LOH in clusters from spatial transcriptomics data*

Description

Calculates Bayes factors for allele fractions at each SNP position. Uses dataframe output by tLOHDataImport

Usage

```
tLOHCalc(forCalcDF)
```

Arguments

`forCalcDF`	Input dataframe generated from the tLOHDataImport function

Value

Output is a dataframe with values that can be visualized with alleleFrequencyPlot() or aggregateCHRPlot()

Author(s)

Michelle Webb

Examples

```
data('humanGBMsampleAC')
df <- tLOHCalc(humanGBMsampleAC)
head(df)
```

`tLOHDataImport` *Import VCF for tLOHCalc*

Description

Import a VCF with per-cluster allele count information at heterozygous SNP positions for the tLOHCalc calculation function.

Arguments

`vcf`	An input VCF file. Spatial transcriptomics clusters make up the sample columns. AD and DP fields are required. Each SNP should be annotated with dbSNP rsIDs.

Value

Output is a dataframe with required fields for tLOHCalc

Author(s)

Michelle Webb

Examples

```
    ## Not run:
    R.utils::gunzip("inst/extdata/Example.vcf.gz","inst/extdata/Example.vcf")
    exampleDF <- tLOHDataImport("inst/extdata/Example.vcf")
## End(Not run)
```

Index