Package 'BubbleTree'

October 8, 2015

Type Package

Title A method to elucidate purity and clonality in tumors using copy number ratio and allele frequency

Version 1.0.0

Date 2014-12-04

Author Wei Zhu <zhuw@medimmune.com> Michael Kuziora <kuzioram@medimmune.com> Brandon Higgs <higgsb@medimmune.com>

Maintainer Wei Zhu <zhuw@medimmune.com>

Description BubbleTree utilizes homogenous pertinent somatic copy number alterations (SCNAs) as markers of tumor clones to extract estimates of tumor ploidy, purity and clonality.

Depends R (>= 3.0)

Imports GenomicRanges, IRanges, plyr, geosphere, mixdist, dplyr, shape

biocViews CopyNumberVariation

License GPL (>=3.1)

Suggests BiocStyle, knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

NeedsCompilation no

R topics documented:

bubbletree 2
calc.prev 3
cnv.gr 4
compareBubbles 5
drawBranches 6
drawBubble 6
getRBD 7
hcc.rbd.lst 8
hetero.gr 9
plotBubbles 9

Index **11**

---

bubbletree *BubbleTree: a method to elucidate purity and clonality in tumors using copy number ratio and allele frequency.*

Description

BubbleTree utilizes homogenous pertinent somatic copy number alterations (SCNAs) as markers of tumor clones to extract tumor ploidy, purity and clonality estimates.

A list of provided functions:

- `drawBranches` the function to draw branches of BubbleTree
- `plotBubbles` the function to draw BubbleTree
- `calc.prev` the function to calculate the prevalence of the tumor clones

Author(s)

Wei Zhu <ZhuW@MedImmune.com>

Michael Kuziora <KuzioraM@MedImmune.com>

Brandon Higgs <HiggsB@MedImmune.com>

References

BubbleTree: an intuitive visualization to elucidate tumoral aneuploidy and clonality in somatic mosaicism using copy number ratio and allele frequency. Wei Zhu, Michael Kuziora, Christopher Morehouse, Tianwei Zhang, Yinong Sebastian, Zheng Liu, Dong Shen, Jiaqi Huang, Zhengwei Dong, Yi Gu, Feng Xue, Liyan Jiang, Yihong Yao, Brandon W. Higgs. *Genome Biology*, Submitted (2015)

Examples

```
data(hetero.gr) #loads sequence variants
data(cnv.gr) #loads copy number variation data
rbd<-getRBD(hetero.gr, cnv.gr)
plotBubbles(rbd) #plot BubbleTree
pur <- calc.prev(rbdx=rbd,heurx=FALSE,modex=3,plotx="prev_model.pdf")

# extract the genotype (branch) and frequency for each segment
 pur[[1]]$ploidy_prev
 pur[[2]][nrow(pur[[2]]),2]
```

`calc.prev` *Calculate tumor cell prevalence in a sample, an indication of sample purity*

Description

A model fitting of component distributions calculated from somatic copy number segments within the BubbleTree diagram.

Usage

```
calc.prev(rbdx,heurx=FALSE,modex=3,plotx="~/prev_model.pdf")
```

Arguments

`rbdx`	a matrix or data frame generated by the function plotBubble().
`heurx`	a logical value indicating if only using A/B and AA/BB branches for purity calculation or all branches.
`modex`	an integer value providing the expected number of modes in mixture distribution.
`plotx`	a character string specifying the mixture model plot file name.

Details

The top n (user defined) most frequent prevalence estimates are seeded as means in this finite mixture model, to predict the tumor (sub)clonal prevalences. This allows a user-defined expected number of modes within the component distribution, though overlapping modes converge to the same estimate. Then these SCNA segment frequencies are used to estimate the means (+/- standard deviations) of the component distributions using the R package mixdist [Macdonald et al, 1988]. The standard deviations are constrained by the Poisson relation given by $\alpha = \sqrt{u_i}, i = 1, ..., k.$.

Value

List object of two elements: 1) the rbdx data frame with two addition columns (2^seg.mean and genotype with frequency for each segment), and 2) a data matrix of modex rows and two columns indicating the seeding modes (column 1) and estimated modes (column 2), each with the number of segments supporting each mode, separated by an underscore. The largest mode in column two is the estimated tumor purity

References

Macdonald PDM. and Green PEJ: User's Guide to Program MIX: An Interactive Program for Fitting Mixtures of Distributions. ICHTHUS DATA SYSTEMS 1988. Macdonald, PDM (1988). *Demonstration Examples for MIX 2.3*. Ichthus Data Systems, Hamilton, Ontario. 13 pp. ISBN 0-9692305-1-4.

Examples

```
#load sequence variants
data('hetero.gr', package='BubbleTree', envir = environment())
#load copy number variation data
data('cnv.gr', package='BubbleTree', envir = environment())
rbd<-getRBD(hetero.gr, cnv.gr) #plot BubbleTree
pur <- calc.prev(rbdx=rbd,heurx=FALSE,modex=3,plotx="prev_model.pdf")

# extract the genotype (branch) and frequency for each segment
 pur[[1]]$ploidy_prev
 pur[[2]][nrow(pur[[2]]),2]
```

---

cnv.gr *A sample dataset of tumor segmented copy number variations*

Description

A GRanges object containing segmented copy number log2 ratios and number of markers/segemnt in a tumor/normal sample pair from a patient with NSCLC

Usage

```
data(cnv.gr)
```

Format

The format is: Formal class 'GRanges' [package "GenomicRanges"] with 6 slots

Details

Metadata columns include

- **num.mark** number of markers within segment
- **seg.mean** mean log2 copy number ratio within a segment

Examples

```
data(cnv.gr)
```

compareBubbles *compare bubbles from two samples*

Description

TBA

Usage

```
compareBubbles(rbd1, rbd2, min.mark=500, min.dist=0.2, max.dist=100, main="")
```

Arguments

rbd1	RBD (R-score BAF Dataframe) from the sample 1
rbd2	RBD data.frame from the sample 2
min.mark	integer segments with mininum markers to be compared
min.dist	numeric minimum distance of the overlapped segments to be displayed
max.dist	numeric maximum distance of the overlapped segments to be displayed
main	character string for the plot title

Details

The segments (larger than min.mark) from the two samples are compared to each other.

Value

A list of the detailed information of the overlapped segments

Examples

```
data('hcc.rbd.lst', package='BubbleTree', envir = environment())

# show the SCNV changes between the recurrent tumor and the primary tumor
compareBubbles(hcc.rbd.lst$HCC11.Primary.Tumor,
    hcc.rbd.lst$HCC11.Recurrent.Tumor,min.dist=0.05, min.mark=2000)

# show the similarity in the recurrent tumors between two subjects
# Interestingly, 17p- and 17q+ are conserved.
compareBubbles(hcc.rbd.lst$HCC4.Recurrent.Tumor,
    hcc.rbd.lst$HCC11.Recurrent.Tumor,
    min.dist=0.0, max.dist=0.1, min.mark=500)
```

drawBranches *Plot branches of BubbleTree plot*

Description

Plot branches of BubbleTree plot

Usage

```
drawBranches(xmax=3.2, main="")
```

Arguments

xmax	define the upper limit of the x-axis
main	title of the plot

Details

The branches of BubbleTree plot stand for interger copy number change. For example, "B" and "BB" indicates LOH and copy-number neutral LOH, repectively.

Value

A plot showing branches of a BubbleTree

Examples

```
drawBranches(xmax=2.6)
drawBranches()
```

drawBubble *Draw a bubble*

Description

Draw single bubble to BubbleTree plot with customized label, size and color

Usage

```
drawBubble(seg.mean, hds.median, num.mark, col, min.cex=0.3,
  size=1, info=NULL ,adj=0.5)
```

Arguments

seg.mean	copy ratio score of the segment
hds.median	median HDS score of the segment
num.mark	number of the marks harbored by the segment
col	color of the bubble
min.cex	minimum font size
size	size of the bubble to scale
info	label of the bubble
adj	adjusted postion of the label

Value

Plots a single bubble on the BubbleTree Plot

Examples

```
drawBranches(main="Demo")
drawBubble(0.5, 0.3, 5000, col="blue", size=2, info="PTEN", adj=-0.5)
```

---

getRBD *Get RBD (R-score BAF Dataframe) of the homogeneous SCNV segments*

Description

Get RBD (R-score BAF Dataframe) of the homogeneous SCNV segments

Usage

```
getRBD(snp.gr, cnv.gr, max.sd = 0.1)
```

Arguments

snp.gr	a GRanges object containing BAF (B-allele frequency) of the germline heterogenous loci
cnv.gr	a GRanges object containing num.mark and seg.mean, generated from the CNV call
max.sd	Numeric value indicating the maximum standard deviation of Homozygous Deviation Scores (HDS) within a cnv segment. Segments with SD above this cutoff will be omitted.

Details

This function merge BAF and CNV call results into one data frame. The segments with high HDS variation are omitted. The RBD of the remaining "homogeneous" segments are returned.

Value

A data frame to be called by plotBubble

Examples

```
#load sequence variants
data('hetero.gr', package='BubbleTree', envir = environment())
#load copy number variation data
data('cnv.gr', package='BubbleTree', envir = environment())
rbd <- getRBD(snp.gr=hetero.gr, cnv.gr=cnv.gr)
plotBubbles(rbd, main="BubbleTree Plot")
```

---

hcc.rbd.lst *Rbd data from the HCC samples*

Description

A list of RBD (R-BAF Dataframe) of the four HCC samples:
HCC4.Primary.Tumor
HCC4.Recurrent.Tumor
HCC11.Primary.Tumor
HCC11.Recurrent.Tumor

Usage

```
data(hcc.rbd.lst)
```

Format

The format is: A list of 4 data frames

Details

These data frames are produced by the getRBD() function and include columns for:

- **seg.id** copy number segment identifier
- **hds.median** median homozygosity deviation score withing the segment
- **hds.sd** standard deviation of hds within the segment
- **num.mark** number of markers within the segment
- **seg.mean** mean copy number of markers within segment
- **chr** segment chromosome
- **start** segment start coordinate
- **end** segment end coordinate
- **cytoband** segment cytoband

Examples

```
data(hcc.rbd.lst)
```

`hetero.gr` *BAF of the germline heterozygous loci in GRanges format*

Description

A sample data of B-allele frquencies of the germline heterozygous loci

Usage

```
data(hetero.gr)
```

Format

The format is: Formal class 'GRanges' [package "GenomicRanges"] with 6 slots

Details

Metadata columns include

- **freq** B allele frequency

Examples

```
data(hetero.gr)
```

`plotBubbles` *Plot Bubbles*

Description

Plot Bubbles

Usage

```
plotBubbles(rbd,  min.cex=0.3, show.cyto=TRUE, no.bayes=FALSE,
 xmax=3.2, size=1, main="BubbleTree Plot")
```

Arguments

rbd	a data.frame containing tumor allele frequency and segmented CNV previously generated by the getRBD() function
min.cex	minimum size of bubble annotation on the plot
show.cyto	Logical; indicating if cytoband information should be displayed on plot. Default is TRUE.
no.bayes	Logical: contol labels
xmax	maximum value for R score plotted on x-axis
size	scaling factor that controls relative size of bubbles appearing on plot
main	character string for the plot title

Details

For each segment iteratively calculate the median and standard deviation (SD) of the homozygous deviation score (HDS) of the heterozygous-loci and filter those SCNAs with high HDS variation (empirically, SD>0.2). The median of HDS, the copy ratio and segment size for each homogenous SCNA are used to define the X-Y coordinates and sizes of the bubbles in the diagram.

Value

Creates a bubbletree plot using the current graphics device and returns a data.frame object containing summary information on genomic regions affected by chromosome loss or gain used in the plot

Examples

```
#load sequence variants
data('hetero.gr', package='BubbleTree', envir = environment())
#load copy number variation data
data('cnv.gr', package='BubbleTree', envir = environment())
rbd <- getRBD(snp.gr=hetero.gr, cnv.gr=cnv.gr)
plotBubbles(rbd)

data('hcc.rbd.lst', package='BubbleTree', envir = environment())
pdf(file="hcc.bubbletree.pdf", width=8, height=6)
lapply(names(hcc.rbd.lst), function(sample) plotBubbles(hcc.rbd.lst[[sample]],
 size=2, main=sample))
dev.off()
```

Index

∗Topic **datasets**
cnv.gr, 4
hcc.rbd.lst, 8
hetero.gr, 9

addBubble *(*drawBubble*)*, 6

BubbleTree *(*bubbletree*)*, 2
bubbletree, 2
bubbletree-package *(*bubbletree*)*, 2

calc.prev, *2*, 3
cnv.gr, 4
compareBubbles, 5

drawBranches, *2*, 6
drawBubble, 6

getRBD, 7

hcc.rbd.lst, 8
hetero.gr, 9

plotBubbles, *2*, 9