Package 'RAPIDR'

February 19, 2015

Title Reliable Accurate Prenatal non-Invasive Diagnosis R package

Description Package to perform non-invasive fetal testing for aneuploidies using sequencing count data from cell-free DNA

Version 0.1.1

Author Kitty Lo <kitty.lo@ucl.ac.uk>

Maintainer Kitty Lo <kitty.lo@ucl.ac.uk>

Depends R (>= 3.1.1)

Imports data.table, Biostrings, Rsamtools, GenomicRanges, GenomicAlignments, PropCIs

Suggests BSgenome.Hsapiens.UCSC.hg19

License GPL-3

LazyData false

NeedsCompilation no

Repository CRAN

Date/Publication 2014-11-20 18:16:37

R topics documented:

`BinBam` *BinBam*

Description

Given a list of bam files, this function writes the output to a text file after binning and doing gc Correction

Usage

```
BinBam(bam.file, index = bam.file, mask = NULL, k = 20000)
```

Arguments

`bam.file`	name of bamfile
`index`	index of bam file
`mask`	mask file in the bed format
`k`	bin size in kilobases

`createReferenceSetFromCounts` *createReferenceSetFromCounts*

Description

This function creates a reference set from a binned counts file

Usage

```
createReferenceSetFromCounts(binned.counts.file, outcomes,
  combined.counts.fname = NULL, method = "zscore", gcCorrect = FALSE,
  gcContentFile = NULL, filterBin = TRUE, removeOutlierSamples = FALSE,
  PCA = FALSE, numPC = 10, masked.counts.file = NULL,
  cleaned.binned.counts.fname = NULL)
```

Arguments

`binned.counts.file`	file name of the binned counts. The binned counts file should be comma delimited, and the first line need to be the chromosome names of each bin
`outcomes`	data.frame with column names: Dx, Gender, SampleID
`combined.counts.fname`	file name to write to for the combined counts per chromosomes, default is not to write result to file

`method`	either "zscore", "NCV" or "MAD", default is zscore
`gcCorrect`	whether to do gc correction or not (True = do the correction)
`gcContentFile`	file name of a Rdata object with the gcContent data
`filterBin`	whether to filter bins based on unusually high counts and high variance, default is to filter
`removeOutlierSamples`	whether to remove samples which has a low correlation value to the rest of the reference set, default is FALSE
`PCA`	whether to do PCA correction or not (True = do the correction)
`numPC`	number of principal components to discard (default = 10)
`masked.counts.file`	file name of the masked counts file
`cleaned.binned.counts.fname`	file name to write to for the corrected binned counts, default is not to write to file

Value

class of rapidr.ref which can then be used to test unknown samples

---

`evalPerformance` *evalPerformance*

Description

This function takes in the summed counts, the baseline and an outcomes table and calculates the overall performance on the dataset

Usage

```
evalPerformance(refset.calls, sample.outcomes)
```

Arguments

`refset.calls`	is the data.frame returned by the testUnknowns function. It contains column with sampleIDs, z-scores and calls for T21, T18, T13 and the fetal sex
`sample.outcomes`	data.frame with column names: Dx, Gender, SampleID

See Also

`testUnknowns`

gcContent *gcContent in 20kb bins*

Description

This is an example output from running makeGCContentData()

makeBinnedCountsFile *makeBinnedCountsFile*

Description

This function takes in a list of bam files and creates a binned counts file. If a mask file is provided, it will also create a masked binned counts file. The output file is comma separated and the first column is the sampleID, the header is the chromosome name of each bin.

Usage

```
makeBinnedCountsFile(bam.file.list, sampleIDs, binned.counts.fname,
  mask = NULL, k = 20000)
```

Arguments

bam.file.list	list of bam file names
sampleIDs	list of sampleIDs, assumed to be in the same order as the bam files in bam.file.list
binned.counts.fname	file name of the output binned counts file
mask	file name of a bed file with the regions to be masked out. Default is no mask file
k	bin size in number of bases. Default is 20,000 bp

See Also

createReferenceSetFromCounts

makeGCContentData *makeGCContentData*

Description

This function calculates the GC content using the hg19 reference genome in bins of user-defined size

Usage

```
makeGCContentData(gc.fname, k = 20000)
```

Arguments

gc.fname	file name of the output file. The file will be written as a .RData object
k	bin size, default is 20,000 bp

makeGCContentPerPos *makeGCContentPerPos*

Description

This function calculates the GC content using the hg19 reference genome for each position in the chromosome

Usage

```
makeGCContentPerPos(gc.fname, k = 200)
```

Arguments

gc.fname	file name of the output file. The file will be written as a .RData object
k	bin size, default is 200 bp

outcomes *set of simulated outcomes data*

Description

This is just a simulated data set for testing purpose

plotTestSample *plotTestSample*

Description

This function plots the QC information. Note that it can only plot one sample at a time

Usage

```
plotTestSample(rapidr.test, input.sampleID, ordering = "normal")
```

Arguments

rapidr.test	rapidr.test object which contains the results from testing a sample
input.sampleID	the sampleID that you like to plot
ordering	normal or gc, default is normal, when set to GC, the chromosomes are ordered according to GC content

print.rapidr.ref *print.rapidr.ref*

Description

Pretty print the rapidr.ref object

Usage

```
## S3 method for class 'rapidr.ref'
print(x, ...)
```

Arguments

x	rapidr.ref object which contains the baselines and the type of correction used to create the baselines
...	other options

print.rapidr.test *print.rapidr.test*

Description

Pretty print the rapidr.test object

Usage

```
## S3 method for class 'rapidr.test'
print(x, ...)
```

Arguments

x	rapidr.test object which contains the results of the test samples
...	other options

ReadBed *ReadBed*

Description

This function loads a bed file

Usage

```
ReadBed(bedfile)
```

Arguments

bedfile	name of the bed file

`testUnknowns` *testUnknowns*

Description

This function uses a reference set to test the unknown samples The unknown samples needs to be a binned counts file which can be created using the makeBinnedCountsFile function.

Usage

```
testUnknowns(ref.data.set, unknowns.counts.file, gcContentFile = NULL,
  masked.counts.file = NULL, combined.counts.fname = NULL)
```

Arguments

`ref.data.set`	rapidr.ref object which contains the baselines and the corrections used to create the baselines
`unknowns.counts.file`	file name of the file with the binned counts of the unknowns, first column needs to be the sampleID, first row needs to be the chromosome names of each bin
`gcContentFile`	file name of a .Rdata object which contains information on GC content in the genome
`masked.counts.file`	optional file of the binned counts after masking
`combined.counts.fname`	file name to write to for the combined counts per chromosome

Value

data.frame with z-scores for chr21, chr18, chr13, and the fetal sex which can be male, female or monosomy X. For males, there is also an estimate of the fetal fraction using the deficit of chrX

`writeCleanedCountsFile` *writeCleanedCountsFiles*

Description

This function takes a binned.counts.file and applies GC correction or PCA correction and writes out the resulting binned counts as a a new file

Usage

```
writeCleanedCountsFile(binned.counts.file, cleaned.binned.counts.fname,
  gcContentFile, gcCorrect = FALSE, PCA = FALSE)
```

Arguments

binned.counts.file	file name of the binned counts. The binned counts file should be comma delimited, and the first line need to be the chromosome names of each bin
cleaned.binned.counts.fname	file name to write to for the corrected binned counts, default is not to write to file
gcContentFile	file name of a Rdata object with the gcContent data
gcCorrect	whether to do gc correction or not (True = do the correction)
PCA	whether to do PCA correction or not (True = do the correction)

---

writeResultsToFile *writeResultsToFile*

Description

This function writes the rapidr.test object to a file

Usage

```
writeResultsToFile(rapidr.test, output.fname)
```

Arguments

rapidr.test	rapidr.test object which contains the results from testing a sample
output.fname	file name of the output file (it will be tab delimited)

Index