

# Package ‘GOTHiC’

April 10, 2015

**Title** Binomial test for Hi-C data analysis

**Description** This is a Hi-C analysis package using a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance in Hi-C experiments. It takes mapped paired NGS reads as input and gives back the list of significant interactions for a given bin size in the genome.

**Version** 1.2.2

**Date** 2013-06-07

**Author** Borbala Mifsud and Robert Sugar

**Maintainer** Borbala Mifsud <Borbala.Mifsud@cancer.org.uk>

**Depends** R (>= 2.15.1), methods, utils, GenomicRanges, Biostrings, BSgenome, data.table

**Imports** BiocGenerics, S4Vectors, IRanges, ShortRead, rtracklayer, ggplot2

**Suggests** HiCDataLymphoblast

**Enhances** parallel

**License** GPL-3

**biocViews** Sequencing, Preprocessing, Epigenetics

## R topics documented:

filtered . . . . .	2
GOTHiC . . . . .	2
mapReadsToRestrictionSites . . . . .	5
pairReads . . . . .	6
<b>Index</b>	<b>8</b>

---

filtered	<i>A GenomicRangesList object used as an example in the GOTHiC package</i>
----------	--

---

### Description

filtered is a GenomicRangesList example object used as an example for the binomialHiC package. This GenomicRangesList contains reads from a human lymphoblastoid cell line HiC experiment (Lieberman-Aiden et al. 2009) for chr20, that were mapped to the genome, paired and PCR duplicate-filtered.

### Usage

```
data(lymphoid_chr20_paired_filtered)
```

### Format

The format is: GenomicRangesList with 2 slots: \$paired\_reads\_1 contains the coordinates for one end of the paired reads \$paired\_reads\_2 contains the coordinates for the other end of the paired reads

### Author(s)

Borbala Gerle and Robert Sugar

### See Also

mapReadsToRestrictionSites

### Examples

```
data(lymphoid_chr20_paired_filtered)
```

---

GOTHiC	<i>Genome Organisation Through HiC</i>
--------	--

---

### Description

GOTHiC performs a cumulative binomial test to detect interactions between distal genomic loci that have significantly more reads than expected by chance in Hi-C experiments. It takes mapped paired NGS reads as input and gives back the list of significant interactions for a given bin size in the genome.

**Usage**

```
GOTHIC(fileName1, fileName2, sampleName, res,
        BSgenomeName=BSgenome.Hsapiens.UCSC.hg19,
        genome=BSgenome.Hsapiens.UCSC.hg19, restrictionSite=A^AGCTT,
        enzyme=HindIII, cistrans=all, filterdist=10000,
        DUPLICATETHRESHOLD=1, fileType=BAM, parallel=FALSE, cores=NULL)
```

**Arguments**

fileName1	File containing the mapped reads of the first fragment ends (BAM or Bowtie format)
fileName2	File containing the mapped reads of the second fragment ends (BAM or Bowtie format)
sampleName	A character string that will be used to name the exported BedGraph file containing the coverage, R object files with paired and mapped reads, and the final data frame with the results from the binomial test. They will be saved in the current directory.
res	An integer that gives the required bin size or resolution of the contact map e.g. 1000000.
BSgenomeName	A character string of the BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build 'BSgenome.Hsapiens.UCSC.hg19'.
genome	The BSgenome package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build BSgenome.Hsapiens.UCSC.hg19.
restrictionSite	A character string that specifies the enzymes recognition site, ^ indicating where the enzyme actually cuts. The default is the HindIII restriction site: 'A^AGCTT'.
enzyme	A character string containing the name of the enzyme used during the Hi-C experiment (i.e. "HindIII", "NcoI"). The default is "HindIII".
cistrans	A character string with three possibilities. "all" runs the binomial test on all interactions, "cis" runs the binomial test only on intrachromosomal/cis interactions, "trans" runs the binomial test only on interchromosomal/trans interactions.
filterdist	An integer specifying the distance between the midpoint of fragments under which interactions are filtered out in order to filter for those read-pairs where the digestion was incomplete. The default is 10000.
DUPLICATETHRESHOLD	An integer specifying the maximum amount of duplicated paired-end reads allowed, over that value it is expected to be PCR bias. The default is 1.
fileType	A character string specifying the format of the aligned reads. The default is 'BAM'. Other accepted format is 'Bowtie'.
parallel	Logical argument. If TRUE the mapping and the binomial test will be performed faster using multiple cores. The default is FALSE.

**cores** An integer specifying the number of cores used in the parallel processing if `parallel=TRUE`. The default is `NULL`.

### Value

A `data.frame` containing elements

`chr1 / chr2` chromosome(s) containing interacting regions 1 and 2  
`locus1 / locus2` start positions of the interacting regions 1 and 2 in the corresponding chromosome(s)  
`relCoverage1 / relCoverage2` relative coverage corresponding to regions 1 and 2  
`probability` expected frequency  
`expected` expected number of reads  
`readCount` observed reads number  
`pvalue` binomial p-value  
`qvalue` binomial p-value corrected for multi-testing with Benjamini-Hochberg  
`logObservedOverExpected` observed/expected read numbers log ratio

### Author(s)

Borbala Mifsud and Robert Sugar

### See Also

`binom.test`, `pairReads`, `mapReadsToRestrictionSites`

### Examples

```
library(GOTHIC)
dirPath <- system.file("extdata", package="HiCDataLymphoblast")
fileName1 <- list.files(dirPath, full.names=TRUE)[1]
fileName2 <- list.files(dirPath, full.names=TRUE)[2]
binom=GOTHIC(fileName1, fileName2, sampleName=lymphoid_chr20, res=1000000,
BSgenomeName=BSgenome.Hsapiens.UCSC.hg18, genome=BSgenome.Hsapiens.UCSC.hg18,
restrictionSite=A^AGCTT, enzyme=HindIII,cistrans=all, filterdist=10000,
DUPLICATETHRESHOLD=1, fileType=Bowtie, parallel=FALSE, cores=NULL)
```

---

`mapReadsToRestrictionSites`*Function to map aligned and paired reads to the restriction fragments*

---

## Description

This function takes mapped paired NGS reads in the format of a `GenomicRangesList` object where the two end of the reads are in the `GenomicRanges` `paired_reads_1` and `paired_reads_2`. It prepares the digestion file from the genome supplied to it with the given restriction enzyme and specificity and maps the reads to the fragments.

## Usage

```
mapReadsToRestrictionSites(pairedReadsFile, sampleName,  
BSgenomeName, genome, restrictionSite, enzyme, parallel=F, cores=1)
```

## Arguments

<code>pairedReadsFile</code>	R object of <code>GenomicRangesList</code> containing <code>paired_reads_1</code> and <code>paired_reads_2</code> <code>GenomicRanges</code> with the paired mapped reads from a Hi-C experiment.
<code>sampleName</code>	A character string that will be used to name the exported R object file with the mapped reads containing a <code>GenomicRangesList</code> with slots <code>locus1</code> and <code>locus2</code> . It will be saved in the current directory.
<code>BSgenomeName</code>	A character string of the <code>BSgenome</code> package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build <code>'BSgenome.Hsapiens.UCSC.hg19'</code> .
<code>genome</code>	The <code>BSgenome</code> package required to make the restriction fragment file containing information for both the organism the experiment was made in, and the genome version the reads were mapped to. The default is the current human genome build <code>BSgenome.Hsapiens.UCSC.hg19</code> .
<code>restrictionSite</code>	A character string that specifies the enzymes recognition site, <code>^</code> indicating where the enzyme actually cuts. The default is the <code>HindIII</code> restriction site: <code>'A^AGCTT'</code> .
<code>enzyme</code>	A character string containing the name of the enzyme used during the Hi-C experiment (i.e. <code>"HindIII"</code> , <code>"NcoI"</code> ). The default is <code>"HindIII"</code> .
<code>parallel</code>	Logical argument. If <code>TRUE</code> the mapping will be performed faster using multiple cores. The default is <code>FALSE</code> .
<code>cores</code>	An integer specifying the number of cores used in the parallel processing if <code>parallel=TRUE</code> . The default is 1.

**Value**

A GenomicRangesList

locus1	GenomicRanges with the coordinates of the start of the fragment where one end of the read mapped
locus2	GenomicRanges with the coordinates of the start of the fragment where the other end of the read mapped

**Author(s)**

Borbala Mifsud and Robert Sugar

**See Also**

pairReads, GOTHic

**Examples**

```
library(GOTHic)
data(lymphoid_chr20_paired_filtered)
mapped=mapReadsToRestrictionSites(filtered, sampleName=lymphoid_chr20,
BSgenomeName=BSgenome.Hsapiens.UCSC.hg18, genome=BSgenome.Hsapiens.UCSC.hg18,
restrictionSite=A^AGCTT, enzyme=HindIII, parallel=FALSE, cores=1)
```

---

pairReads

*Function pairs aligned paired NGS reads*

---

**Description**

This function takes bowtie output files, pairs the reads, only keeps those where both ends mapped, filters for perfect duplicates to avoid PCR bias, and saves and returns a GenomicRangesList object that contains the paired\_reads\_1 and paired\_reads\_2 GenomicRanges with the paired reads

**Usage**

```
pairReads(fileName1, fileName2, sampleName, DUPLICATETHRESHOLD = 1,
fileType=BAM)
```

**Arguments**

fileName1	File containing the mapped reads of the first fragment ends (BAM or Bowtie format)
fileName2	File containing the mapped reads of the second fragment ends (BAM or Bowtie format)
sampleName	A character string that will be used to name the exported BedGraph file containing the coverage, and the R object file with paired reads. They will be saved in the current directory.

**DUPLICATETHRESHOLD**

An integer specifying the maximum amount of duplicated paired-end reads allowed, over that value it is expected to be PCR bias. The default is 1.

**fileType**

A character string specifying the format of the aligned reads. The default is 'BAM'. Other accepted format is 'Bowtie'.

**Value**

A GenomicRangesList called filtered

paired\_reads\_1 GenomicRanges with the coordinates of where one end of the read mapped

paired\_reads\_2 GenomicRanges with the coordinates of where the other end of the read mapped

**Author(s)**

Borbala Mifsud and Robert Sugar

**See Also**

mapReadsToRestrictionSites, GOTHIC

**Examples**

```
library(GOTHIC)
dirPath <- system.file("extdata", package="HiCDataLymphoblast")
fileName1 <- list.files(dirPath, full.names=TRUE)[1]
fileName2 <- list.files(dirPath, full.names=TRUE)[2]
paired <- pairReads(fileName1, fileName2, sampleName=lymphoid_chr20,
DUPLICATETHRESHOLD = 1, fileType=Bowtie)
```

# Index

\*Topic **datasets**

filtered, [2](#)

\*Topic **manip**

GOTHIC, [2](#)

mapReadsToRestrictionSites, [5](#)

pairReads, [6](#)

\*Topic **package**

GOTHIC, [2](#)

filtered, [2](#)

GOTHIC, [2](#)

lymphoid\_chr20\_paired\_filtered  
(filtered), [2](#)

mapReadsToRestrictionSites, [5](#)

pairReads, [6](#)