Package 'CNVseq'

December 6, 2024

Version 1.0.0

Date 2024-12-05

Type Package

Title CNVseq: A tool for CNV-seq data calling variants

Author Zhan-Ni Chen

Maintainer Zhan-Ni Chen <sanadamakomi@gmail.com>

Depends R (>= 3.4.0)

Imports utils,
S4Vectors,
IRanges,
GenomicRanges,
Biostrings,
bitops,
GenomeInfoDb,
Rsamtools,
GenomicAlignments,
DNAcopy,
data.table

Description Detect CNVs in CNV-seq data. By default, it identifies DUP and DEL larger than 100 kb with a mosaicism level above 0.5, and CNVs larger than 5 Mb with a mosaicism level above 0.1. The mosaicism level is adjustable.

License Artistic-2.0

Encoding UTF-8

URL https://github.com/sanadamakomi/CNVseq

BugReports https://github.com/sanadamakomi/CNVseq/issues

LazyData true

RoxygenNote 7.3.2

Contents

`batch_call_cnv` *Call CNVs from read count files*

Description

Call CNVs from read count files

Usage

```
batch_call_cnv(
  cnm_path,
  cnn_paths,
  cnn_dirs,
  out_dir,
  test_id,
  ref_ids,
  by_gender = FALSE,
  fraction = c(0.5, 0.5)
)
```

Arguments

`cnm_path`	Path of merged CNN file.
`cnn_paths`	Path of CNN files, separated by comma(,).
`cnn_dirs`	Path of CNN directory, separated by comma(,).

out_dir	Output directory path.
test_id	Sample id to call CNV.
ref_ids	Reference ids to create a baseline, separated by comma(,).
by_gender	A bool value to call by gender.
fraction	Fraction of CNV >=100Kb and >=5Mb, default:c(0.5, 0.5).

---

batch_cov *Calculate read count in bams*

Description

Calculate read count in bams

Usage

```
batch_cov(bam_dirs, bam_paths, bed_path, out_dir, thread = 4, mapq.filter = 0)
```

Arguments

bam_dirs	Path of BAM directory, separated by comma(,).
bam_paths	Path of BAM file, separated by comma(,).
bed_path	Path of BED file.
out_dir	A character string of directory to output coverage files.
thread	An integer providing the number of thread, default: 4.
mapq.filter	A non-negative integer specifying the minimum mapping quality to include. BAM reads with mapping qualities less than mapqFilter are discarded, default: 0.

---

batch_gender *Calculate gender and total read from a coverage file.*

Description

Calculate gender and total read from a coverage file.

Usage

```
batch_gender(cnn_paths, cnn_dirs, out_dir)
```

Arguments

cnn_paths	Path of CNN files, separated by comma(,).
cnn_dirs	Path of CNN directory, separated by comma(,).
out_dir	Output directory path.

batch_merge *Merge CNN, VCF, CNR, CNS file.*

Description

Input cnn_paths or cnn_dirs, it will merge multiple samples' read count result files(.cnn) and create a merged file. Input auto_file and sex_file, it will merge autosome and sex chromosome result, the VCF, CNR, CNS file can be input.

Usage

```
batch_merge(cnn_paths, cnn_dirs, auto_file, sex_file, out_path)
```

Arguments

cnn_paths	Path of CNN files, separated by comma(,).
cnn_dirs	Path of CNN directory, separated by comma(,).
auto_file	path of VCF, CNR, CNS file.
sex_file	path of VCF, CNR, CNS file.
out_path	Output file path.

batch_split *Split wgs region into bins*

Description

Split wgs region into bins

Usage

```
batch_split(path, bin = 10000, access_bed = "")
```

Arguments

path	Path of output, a BED file.
bin	An integer of bin size, default: 1E4.
access_bed	BED file include sequence-accessible region. If NULL it will use whole genome region.

bed_to_gr *Read BED file and output grange*

Description

Read BED file and output grange

Usage

```
bed_to_gr(file)
```

Arguments

file	Path of BED file.

call_cnv *Calling CNV*

Description

Calling CNV

Usage

```
call_cnv(data, s_id, is_male = TRUE, fraction = c(0.5, 0.5))
```

Arguments

data	A data frame of log2ratio to do CBS. It must has three column: chromosome, end, log2.
s_id	A charactor string of sample id.
is_male	A bool value, TRUE when the gender is male. change-points, default: 0.05.
fraction	Fraction of CNV >=100Kb and >=5Mb, default:c(0.5, 0.5).

check_bam *Check BAM file.*

Description

It will stop if BAM file is illegal.

Usage

```
check_bam(x)
```

Arguments

x	A character string or vector of BAM File path.

count_reads_of_region *Calculating read count in region*

Description

Calculating read count in region

Usage

```
count_reads_of_region(region, bamPath, mapq.filter = 0, minoverlap = 75L)
```

Arguments

region	A grange object of region to extract reads in BAM file.
bamPath	A character string of the BAM path.
mapq.filter	A non-negative integer specifying the minimum mapping quality to include. BAM reads with mapping qualities less than mapqFilter are discarded.
minoverlap	Minimum overlap size for region and reads, default: 75L.

count_reads_of_region_multicore *Calculating read count in multiple thread*

Description

Calculating read count in multiple thread

Usage

```
count_reads_of_region_multicore(
  region,
  bamPath,
  thread,
  batch,
  tmpDir = NULL,
  mapq.filter = 0,
  minoverlap = 75L
)
```

Arguments

region	A grange object of region to extract reads in BAM file.
bamPath	A character string of the BAM path.
thread	An integer providing the number of thread.
batch	An integer giving how many GRanges are performed in a batch.
tmpDir	A character string of directory to output coverage files (<sampleid>.cnn). Default is the current folder.

mapq.filter	A non-negative integer specifying the minimum mapping quality to include. BAM reads with mapping qualities less than mapqFilter are discarded.
minoverlap	Minimum overlap size for region and reads, default: 75L.

---

creat_chunk *Create grange into chunk*

Description

Create grange into chunk

Usage

```
creat_chunk(gr, bin = 10000)
```

Arguments

gr	A grange object.
bin	Bin size, default: 1E4.

---

do_cbs *Circular Binary Segmentation*

Description

Circular Binary Segmentation

Usage

```
do_cbs(data, s_id, alpha = 0.05, min.width = 2)
```

Arguments

data	A data frame of log2ratio to do CBS. It must has three column: chromosome, end, log2.
s_id	A charactor string of sample id.
alpha	A numeric value of significance levels for the test to accept change-points, default: 0.05.
min.width	An integer value of the minimum number of markers for a changed segment, default: 2.

get_hg19_seqinfo *Hg19 seqinfo*

Description

Hg19 seqinfo

Usage

```
get_hg19_seqinfo()
```

get_id *Extract sample id from file path*

Description

Extract sample id from file path

Usage

```
get_id(file, ptn = NULL)
```

Arguments

file	Path of file.
ptn	A charactor for spliting string. By default NULL, it will split string by '.'.

get_wgs_bin *Split wgs region into bins*

Description

Split wgs region into bins

Usage

```
get_wgs_bin(bin = 10000, access_bed = "")
```

Arguments

bin	An integer of bin size, default: 1E4.
access_bed	BED file include sequence-accessible region. If NULL it will use whole genome region.

make_vcf_format *Create VCF FORMAT column*

Description

Create VCF FORMAT column

Usage

```
make_vcf_format(x)
```

Arguments

x	A data frame of CNV result.

make_vcf_header *Create VCF header*

Description

Create VCF header

Usage

```
make_vcf_header()
```

make_vcf_info *Create VCF INFO column*

Description

Create VCF INFO column

Usage

```
make_vcf_info(x)
```

Arguments

x	A data frame of CNV result.

make_vcf_matrix *Create VCF matrix*

Description

Create VCF matrix

Usage

```
make_vcf_matrix(x)
```

Arguments

x	A data frame of CNV result.

merger_cov_files *Merge coverage files.*

Description

Align and merge coverage files (<filename>.cnn) with chromosome, start and end position. Four required fields in a coverage file are chromosome name, start and end position, depth of coverage.

Usage

```
merger_cov_files(files, path = NULL)
```

Arguments

files	A character vector contains several coverage files path.
path	Path to write to.

Value

A data frame, of which columns are chromosome, start position, end position, and depths in input coverage files.

merge_bin_df *Merge log2ratio with a specific interval of bins*

Description

Merge log2ratio with a specific interval of bins

Usage

```
merge_bin_df(data, group_size = 50)
```

Arguments

data	A data frame of cnr file.
group_size	A interval to merge, default: 50.

merge_result_files *Merge autosome and sex chromosome result*

Description

Merge autosome and sex chromosome result

Usage

```
merge_result_files(auto_file, sex_file, out_path)
```

Arguments

auto_file	path of VCF, CNR, CNS file.
sex_file	path of VCF, CNR, CNS file.
out_path	path of output file.

merge_segment *Filtering transcripts*

Description

Filtering transcripts

Usage

```
merge_segment(x)
```

Arguments

x	A data.frame of CNV calling result.

output_vcf *Output CNV calling result to VCF file*

Description

Output CNV calling result to VCF file

Usage

```
output_vcf(dat, path)
```

Arguments

dat	A data.frame of CNV calling result.
path	Path of VCF file.

ploidy_to_cn *Count copy numer with different mosaicism level*

Description

Count copy numer with different mosaicism level

Usage

```
ploidy_to_cn(f, ploidy, cn)
```

Arguments

f	Fraction of variants.
ploidy	the ploidy of chromosome.
cn	Copy number.

read_db_file *Read annovation database file and return granges*

Description

Read annovation database file and return granges

Usage

```
read_db_file(prefix, db_path, p = 1, header = FALSE, file_encode = "UTF-8")
```

Arguments

prefix	A prefix of database file, e.g. centromere_telomere for hg19_centromere_telomere.txt.
db_path	Path of database directory.
p	A integer for the column index of chromosome, default: 1.
header	A bool for the header of database file, default: FALSE.
file_encode	The file endode of database file, default: UTF-8.

---

write_cov_file *Write read count result file*

Description

Write read count result file

Usage

```
write_cov_file(gr, path)
```

Arguments

gr	A GRange with a column named rc.
path	Path to write to.

Index