Convert a FASTQ file to an indexed BAM

Description

This function takes in FASTQ files and performs quality control before aligning to a reference genome. It assumes paired-end samples are of the same file name with an underscore (_) and a number to signify different reads of the same sample. Each sample's outputs will be stored in a separate directory.

Usage

```
fastqProcession(
  file = list_files_with_exts(dir = dir, exts = "fastq"),
  dir = getwd(),
  genome,
  genomedir,
  paired = FALSE,
  threads = 4L,
  molecule = "RNA"
)
```

Arguments

`file`	A list of paths to FASTQ files. If no paths are entered, defaults to all fastq files in dir.
`dir`	Output directory.
`genome`	A BSgenome object, GmapGenome object, or a character string indicating the genome name eg. "hg19".
`genomedir`	A directory containing the reference genome. Otherwise, it is the parent directory of the reference genome is genome is a character string or BSgenome object.
`paired`	Indicates whether the samples are from paired-end or single-end reads.
`threads`	The number of cores to use in the process.
`molecule`	A character string indicating either DNA or RNA samples.

Value

A list of file paths to created BAM files

Analyse differential gene expression

Description

This function analyses differentially expressed genes between sample groups. It assumes that sample BAM files are ordered in a directory structure such as "group/sample/" as fastqProcession() would order it. There should be more than one sample per group or else differential expression analysis won't work It outputs a CSV file showing differential expression (ordered by p-value). It works similarly to gene_Counting(), but then further analyses those counts to obtain differential expression data.

Usage

```
geneExpression(
  dir = getwd(),
  groups,
  TxDb,
  orgDb,
  outputdir = getwd(),
  threads = 4L,
  paired = FALSE
)
```

Arguments

`dir`	The parent directory of the sample groups.
`groups`	Folder names of the sample groups. The default is all folders in dir.
`TxDb`	A TxDb object upon which regions of the genome are counted.
`orgDb`	An orgDb object for annotating the CSV with gene symbols and Ensembl IDs.
`outputdir`	Output directory of CSV file.
`threads`	Number of cores to use.
`paired`	Indicates whether the samples are from paired-end reads.

Value

A data frame list containing all of the differential expression comparison.

Count genes

Description

This function counts reads of gene regions between sample groups. It assumes that sample BAM files are ordered in a directory structure such as "group/sample/" as fastqProcession() would order it. It outputs a CSV file showing gene counts. Works similarly to geneExpression(), but outputs count data instead of differential expression data.

Usage

```
gene_Counting(
  dir = getwd(),
  groups,
  TxDb,
  orgDb,
  outputdir = getwd(),
  threads = 4L,
  paired = FALSE
)
```

Arguments

`dir`	The parent directory of the sample groups.
`groups`	Folder names of the sample groups. The default is all folders in dir.
`TxDb`	A TxDb object upon which regions of the genome are counted.
`orgDb`	An orgDb object for annotating the CSV with gene symbols and Ensembl IDs.
`outputdir`	Output directory of CSV file.
`threads`	Number of cores to use.
`paired`	Indicates whether the samples are from paired-end reads.

Value

A data frame containing gene counts.

Call single nucleotide polymorphism variants

Description

This function calls SNP variants from BAM files. The results are formatted into a VCF file and the ID column is populated with dbSNP IDs.

Usage

```
callSNP(bam, genome, genomedir, SNPlocs, threads = 4L, outputdir = getwd())
```

Arguments

`bam`	A list of paths to BAM files
`genome`	A BSgenome object, a GmapGenome object, or a character string indicating the reference genome eg. "hg19"
`genomedir`	The directory containing the reference genome or, if genome is a character string, the parent directory of the reference genome directory.
`SNPlocs`	An SNPlocs object containing dbSNP IDs.
`threads`	The number of cores to use.
`outputdir`	The output directory for the VCF file.

Value

A list of file paths to the VCF files.

Call insertion/deletion variants

Description

This function calls indel variants from BAM files. The results are formatted into a VCF file and the ID column is populated with dbSNP IDs.

Usage

```
callIndel(bam, genome, genomedir, SNPlocs, threads, outputdir = getwd())
```

Arguments

`bam`	A list of paths to BAM files
`genome`	A BSgenome object, a GmapGenome object, or a character string indicating the reference genome eg. "hg19"
`genomedir`	The directory containing the reference genome or, if genome is a character string, the parent directory of the reference genome directory.
`SNPlocs`	An SNPlocs object containing dbSNP IDs. For indels, this may be an XtraSNPlocs object.
`threads`	The number of cores to use. Cores should equal a factor of reference genome sequence levels ie. chromosome contigs should be equally divisible between cores.
`outputdir`	The output directory for the VCF file.

Value

A list of file paths to the VCF files.

Call copy number variants

Description

This function calls copy number variants from sample BAM files compared to control BAM files. It assumes that BAM files are stored in separate folders as is created by fastqProcession(). This function requires that control BAM files are provided. Once complete, it creates a CSV file containing copy number information.

Usage

```
callCNV(
  controldir,
  control = NULL,
  experimentdir,
  experiment = NULL,
  bed,
  outputdir = getwd()
)
```

Arguments

`controldir`	The parent directory of the sample directories.
`control`	The names of the folders in which control BAM files are. If NULL, all folders in controldir will be checked for BAM files.
`experimentdir`	The parent directory of sample on which to investigate copy numbers.
`experiment`	The names of the folders in which sample BAM files are. If NULL, all folders in experimentdir will be checked for BAM files.
`bed`	A character string indicating BED file path or a TxDb object from which to extract a BED file.
`outputdir`	The directory in which to place copy number call.

Value

A data frame containing copy number calls.