

Package ‘intansv’

October 17, 2020

Title Integrative analysis of structural variations

Description This package provides efficient tools to read and integrate structural variations predicted by popular softwares. Annotation and visulation of structural variations are also implemented in the package.

Version 1.28.0

Author Wen Yao <ywhzau@gmail.com>

Maintainer Wen Yao <ywhzau@gmail.com>

biocViews Genetics, Annotation, Sequencing, Software

Depends R (>= 2.14.0), plyr, ggbio, GenomicRanges

Imports BiocGenerics, IRanges

License MIT + file LICENSE

git_url <https://git.bioconductor.org/packages/intansv>

git_branch RELEASE_3_11

git_last_commit 134d62a

git_last_commit_date 2020-04-27

Date/Publication 2020-10-16

R topics documented:

methodsMerge	2
plotChromosome	3
plotRegion	4
readBreakDancer	5
readCnvnator	7
readDelly	8
readLumpy	9
readPindel	10
readSoftSearch	11
readSvseq	12
svAnnotation	13
Index	15

 methodsMerge

Integrate structural variations predicted by different methods

Description

Integrate predictions of different tools to provide more reliable structural variations.

Usage

```
methodsMerge(..., others=NULL,
  overLapPerDel = 0.8, overLapPerDup = 0.8, overLapPerInv = 0.8,
  numMethodsSupDel = 2, numMethodsSupDup = 2, numMethodsSupInv = 2)
```

Arguments

...	results of different SVs predictions read in to R by intansv.
others	a data frame of structural variations predicted by other tools.
overLapPerDel	Deletions predicted by different methods that have reciprocal coordinate overlap larger than this threshold would be clustered together
overLapPerDup	Duplications predicted by different methods that have reciprocal coordinate overlap larger than this threshold would be clustered together
overLapPerInv	Inversions predicted by different methods that have reciprocal coordinate overlap larger than this threshold would be clustered together
numMethodsSupDel	Deletion clusters supported by no more than this threshold of read support would be discarded
numMethodsSupDup	Duplication clusters supported by no more than this threshold of read support would be discarded
numMethodsSupInv	Inversion clusters supported by no more than this threshold of read support would be discarded

Details

A structural variation (deletion, duplication, inversion et al.) may be reported by different tools. However, the boundaries of this structural variation predicted by different tools don't always agree with each other. Predictions of different methods with reciprocal overlap more than 80 percent were merged. Structural variations supported by only one method were discarded.

Value

A list with the following components:

del	the integrated deletions of different methods.
dup	the integrated duplications of different methods.
inv	the integrated inversions of different methods.

Author(s)

Wen Yao

Examples

```
breakdancer <- readBreakDancer(system.file("extdata/ZS97.breakdancer.sv",
                                           package="intansv"))
str(breakdancer)

cnvnator <- readCnvnator(system.file("extdata/cnvnator", package="intansv"))
str(cnvnator)

svseq <- readSvseq(system.file("extdata/svseq2", package="intansv"))
str(svseq)

delly <- readDelly(system.file("extdata/ZS97.DELLY.vcf", package="intansv"))
str(delly)

pindel <- readPindel(system.file("extdata/pindel", package="intansv"))
str(pindel)

sv_all_methods <- methodsMerge(breakdancer, pindel, cnvnator, delly, svseq)
str(sv_all_methods)

sv_all_methods.1 <- methodsMerge(breakdancer, pindel, cnvnator, delly, svseq,
                                overLapPerDel=0.7)
str(sv_all_methods.1)

sv_all_methods.2 <- methodsMerge(breakdancer, pindel, cnvnator, delly, svseq,
                                overLapPerDel=0.8, numMethodsSupDel=3)
str(sv_all_methods.2)
```

plotChromosome

Display the chromosome distribution of structural variations

Description

Display the chromosome distribution of structural variations by splitting the chromosomes into windows of specific size and counting the number of structural variations in each window.

Usage

```
plotChromosome(genome, structuralVariation, windowSize=1000000)
```

Arguments

genome A data frame with ID and length of all Chromosomes.
structuralVariation A list of structural variations.
windowSize A specific size (in base pair) to split chromosomes into windows.

Details

To visualize the distribution of structural variations in the whole genome, chromosomes were split into windows of specific size (default 1 Mb) and the number of structural variations in each window were counted. The number of structural variations were shown using circular barplot.

Value

A circular plot with five layers:

- the circular view of genome ideogram.
- the chromosome coordinates labels.
- the circular barplot of number of deletions in each chromosome window.
- the circular barplot of number of duplications in each chromosome window.
- the circular barplot of number of inversions in each chromosome window.

Author(s)

Wen Yao

Examples

```
delly <- readDelly(system.file("extdata/ZS97.DELLY.vcf", package="intansv"))
str(delly)

genome.file.path <- system.file("extdata/chr05_chr10.genome.txt", package="intansv")
genome <- read.table(genome.file.path, head=TRUE, as.is=TRUE)
str(genome)

plotChromosome(genome, delly, 1000000)
```

plotRegion

Display structural variations in a specific genomic region

Description

Display the structural variations in a specific genomic region in circular view.

Usage

```
plotRegion(structuralVariation, genomeAnnotation,
            regionChromosome, regionStart, regionEnd)
```

Arguments

structuralVariation	A list of structural variations.
genomeAnnotation	A data frame of genome annotations.
regionChromosome	The chromosome identifier of a specific region to view.
regionStart	The start coordinate of a specific region to view.
regionEnd	The end coordinate of a specific region to view.

Details

Different SVs were shown as rectangles in different layers. See the package vignette and the example dataset for more details.

Value

A circular plot of all the structural variations and genes in a specific region with four layers:

- The composition of genes of a specific genomic region.
- The composition of deletions of a specific genomic region.
- The composition of duplications of a specific genomic region.
- The composition of inversions of a specific genomic region.

Author(s)

Wen Yao

Examples

```
delly <- readDelly(system.file("extdata/ZS97.DELLY.vcf", package="intansv"))
str(delly)

anno.file.path <- system.file("extdata/chr05_chr10.anno.txt", package="intansv")
msu_gff_v7 <- read.table(anno.file.path, head=TRUE, as.is=TRUE)
str(msu_gff_v7)

plotRegion(delly,msu_gff_v7,"chr05",1,200000)
```

readBreakDancer

Read in the structural variations predicted by breakDancer

Description

Reading in the structural variations predicted by breakDancer, filtering low quality predictions and merging overlapping predictions.

Usage

```
readBreakDancer(file="", scoreCutoff=60, readsSupport=3,  
                regSizeLowerCutoff=100, regSizeUpperCutoff=10000000,  
                method="BreakDancer", ...)
```

Arguments

file	the output file of breakDancer.
scoreCutoff	the minimum score for a structural variation to be read in.
readsSupport	the minimum read pair support for a structural variation to be read in.
regSizeLowerCutoff	the minimum size for a structural variation to be read in.
regSizeUpperCutoff	the maximum size for a structural variation to be read in.
method	a tag to assign to the result of this function.
...	parameters passed to read.table.

Details

The predicted SVs could be further filtered by score, number of read pairs supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details.

Value

A list with the following components:

del	the deletions predicted by breakDancer.
inv	the inversions predicted by breakDancer.

Author(s)

Wen Yao

Examples

```
breakdancer <- readBreakDancer(system.file("extdata/ZS97.breakdancer.sv",  
                                           package="intansv"))  
str(breakdancer)
```

readCnvator	<i>Read in the structural variations predicted by CNVnator</i>
-------------	----------------------------------------------------------------

Description

Reading the structural variations predicted by CNVnator, filtering low quality predictions and merging overlapping predictions.

Usage

```
readCnvator(dataDir=".", regSizeLowerCutoff=100, regSizeUpperCutoff=1000000,
            method="CNVnator")
```

Arguments

dataDir	the directory that contain the output files of CNVnator.
regSizeLowerCutoff	the minimum size for a structural variation to be read.
regSizeUpperCutoff	the maximum size for a structural variation to be read.
method	a tag to assign to the result of this function.

Details

The predicted SVs could be further filtered by the predicted size of SVs to get more reliable SVs. See our paper for more details. The directory that specified by the parameter "dataDir" should only contain the predictions of CNVnator. See the example dataset for more details.

Value

A list with the following components:

del	the deletions predicted by CNVnator.
dup	the duplications predicted by CNVnator.

Author(s)

Wen Yao

Examples

```
cnvator <- readCnvator(system.file("extdata/cnvator", package="intansv"))
str(cnvator)
```

`readDelly`*Read in the structural variations predicted by DELLY*

Description

Reading the structural variations predicted by DELLY, filtering low quality predictions and merging overlapping predictions.

Usage

```
readDelly(file="", regSizeLowerCutoff=100, regSizeUpperCutoff=1000000,
readsSupport=3, method="Delly", ...)
```

Arguments

<code>file</code>	the file containing the prediction results of DELLY.
<code>regSizeLowerCutoff</code>	the minimum size for a structural variation to be read.
<code>regSizeUpperCutoff</code>	the maximum size for a structural variation to be read.
<code>readsSupport</code>	the minimum read pair support for a structural variation to be read.
<code>method</code>	a tag to assign to the result of this function.
<code>...</code>	parameters passed to <code>read.table</code> .

Details

The predicted SVs could be further filtered by the number of read pairs supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details.

Value

A list with the following components:

<code>del</code>	the deletions predicted by DELLY.
<code>dup</code>	the duplications predicted by DELLY.
<code>inv</code>	the inversions predicted by DELLY.

Author(s)

Wen Yao

Examples

```
delly <- readDelly(system.file("extdata/ZS97.DELLY.vcf", package="intansv"))
str(delly)
```

`readLumpy`*Read in the structural variations predicted by Lumpy*

Description

Reading the structural variations predicted by Lumpy, filtering low quality predictions and merging overlapping predictions.

Usage

```
readLumpy(file="", regSizeLowerCutoff=100, regSizeUpperCutoff=1000000,
readsSupport=3, method="Lumpy", ...)
```

Arguments

<code>file</code>	the file containing the prediction results of Lumpy.
<code>regSizeLowerCutoff</code>	the minimum size for a structural variation to be read.
<code>regSizeUpperCutoff</code>	the maximum size for a structural variation to be read.
<code>readsSupport</code>	the minimum read pair support for a structural variation to be read.
<code>method</code>	a tag to assign to the result of this function.
<code>...</code>	parameters passed to <code>read.table</code> .

Details

The predicted SVs could be further filtered by the number of reads supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details.

Value

A list with the following components:

<code>del</code>	the deletions predicted by Lumpy.
<code>dup</code>	the duplications predicted by Lumpy.
<code>inv</code>	the inversions predicted by Lumpy.

Author(s)

Wen Yao

Examples

```
lumpy <- readLumpy(system.file("extdata/ZS97.LUMPY.vcf", package="intansv"))
str(lumpy)
```

readPindel *Read in the structural variations predicted by Pindel*

Description

Reading the structural variations predicted by Pindel, filtering low quality predictions and merging overlapping predictions.

Usage

```
readPindel(dataDir=".", regSizeLowerCutoff=100,
           regSizeUpperCutoff=1000000, readsSupport=3,
           method="Pindel")
```

Arguments

dataDir the directory containing the prediction results of Pindel.
 regSizeLowerCutoff the minimum size for a structural variation to be read.
 regSizeUpperCutoff the maximum size for a structural variation to be read.
 readsSupport the minimum read pair support for a structural variation to be read.
 method a tag to assign to the result of this function.

Details

The predicted SVs could be further filtered by the number of reads supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details. The directory that specified by the parameter "dataDir" should only contain the predictions of Pindel. The deletions output files should be named using the suffix "_D", the duplications output files should be named using the suffix "_TD", and the inversions output files should be named using the suffix "_INV". See the example dataset for more details.

Value

A list with the following components:

del the deletions predicted by Pindel.
 dup the duplications predicted by Pindel.
 inv the inversions predicted by Pindel.

Author(s)

Wen Yao

Examples

```
pindel <- readPindel(system.file("extdata/pindel", package="intansv"))
str(pindel)
```

readSoftSearch	<i>Read in the structural variations predicted by SoftSearch</i>
----------------	------------------------------------------------------------------

Description

Reading the structural variations predicted by SoftSearch, filtering low quality predictions and merging overlapping predictions.

Usage

```
readSoftSearch(file="", regSizeLowerCutoff=100, readsSupport=3,
               method="softSearch", regSizeUpperCutoff=1000000,
               softClipsSupport=3, ...)
```

Arguments

file	the file containing the prediction results of SoftSearch.
regSizeLowerCutoff	the minimum size for a structural variation to be read.
regSizeUpperCutoff	the maximum size for a structural variation to be read.
readsSupport	the minimum read pair support for a structural variation to be read.
method	a tag to assign to the result of this function.
softClipsSupport	the minimum soft clip support for a structural variation to be read.
...	parameters passed to read.table

Details

The predicted SVs could be further filtered by the number of reads supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details.

Value

A list with the following components:

del	the deletions predicted by SoftSearch.
dup	the duplications predicted by SoftSearch.
inv	the inversions predicted by SoftSearch.

Author(s)

Wen Yao

Examples

```
softSearch <- readSoftSearch(system.file("extdata/ZS97.softsearch", package="intansv"))
str(softSearch)
```

`readSvseq`*Read in the structural variations predicted by SVseq2*

Description

Reading the structural variations predicted by SVseq2, filtering low quality predictions and merging overlapping predictions.

Usage

```
readSvseq(dataDir=".", regSizeLowerCutoff=100, method="SVseq2",
          regSizeUpperCutoff=1000000, readsSupport=3)
```

Arguments

<code>dataDir</code>	a directory containing the predictions of SVseq2.
<code>regSizeLowerCutoff</code>	the minimum size for a structural variation to be read.
<code>regSizeUpperCutoff</code>	the maximum size for a structural variation to be read.
<code>readsSupport</code>	the minimum read pair support for a structural variation to be read.
<code>method</code>	a tag to assign to the result of this function.

Details

The predicted SVs could be further filtered by the number of reads supporting the occurrence of a specific SV, and the predicted size of SVs to get more reliable SVs. See our paper for more details. The directory that specified by the parameter "dataDir" should only contain the predictions of SVseq2. The deletions output files should be named using the suffix ".del". See the example dataset for more details.

Value

A list with the following components:

<code>del</code>	the deletions predicted by SVseq2.
------------------	------------------------------------

Author(s)

Wen Yao

Examples

```
svseq <- readSvseq(system.file("extdata/svseq2", package="intansv"))
str(svseq)
```

svAnnotation	<i>Annotation of structural variations</i>
--------------	--------------------------------------------

Description

Annotate the effect caused by structural variations to genes and elements of genes.

Usage

```
svAnnotation(structuralVariation, genomeAnnotation)
```

Arguments

structuralVariation	A data frame of structural variations.
genomeAnnotation	A data frame of genome annotations.

Details

A structural variation (deletion, duplication, inversion et al.) could affect the structure of a specific gene, including deletion of introns/exons, deletion of whole gene, et al.. This function gives the detailed effects caused by structural variations to genes and elements of genes.

The parameter "structuralVariation" should be a data frame with three columns:

- chromosome the chromosome of a structural variation.
- pos1 the start coordinate of a structural variation.
- pos2 the end coordinate of a structural variation.

Value

A data frame with the following columns:

chromosome	the chromosome of a structural variation.
pos1	the start coordinate of a structural variation.
pos2	the end coordinate of a structural variation.
size	the size of a structural variation.
info	information on a structural variation.
tag	the tag of a genomic element.
start	the start coordinate of a genomic element.
end	the end coordinate of a genomic element.
strand	the strand of a genomic element.
ID	the ID of a genomic element.

Author(s)

Wen Yao

Examples

```
breakdancer <- readBreakDancer(system.file("extdata/ZS97.breakdancer.sv",
                                           package="intansv"))
str(breakdancer)

msu_gff_v7 <- read.table(system.file("extdata/chr05_chr10.anno.txt", package="intansv"),
                        head=TRUE, as.is=TRUE, sep="\t")
breakdancer.anno <- llply(breakdancer, svAnnotation,
                          genomeAnnotation=msu_gff_v7)
str(breakdancer.anno)
```

Index

`methodsMerge`, 2

`plotChromosome`, 3

`plotRegion`, 4

`readBreakDancer`, 5

`readCnvator`, 7

`readDelly`, 8

`readLumpy`, 9

`readPindel`, 10

`readSoftSearch`, 11

`readSvseq`, 12

`svAnnotation`, 13