Package 'intansv'

April 5, 2014

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 1.2.0

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 Artistic-2.0

R topics documented:

`geneAnnotation` *Annotation of genes affected by structural variations*

Description

Report the details of genes affected by structural variations.

Usage

```
geneAnnotation(structuralVariation,genomeAnnotation)
```

Arguments

`structuralVariation`	A data frame of structural variations.
`genomeAnnotation`	A genomic ranges of the genome annotation.

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.. And a specific gene might be affected by multiple SVs. This function gives the detailed effects caused by structural variations to genes and its elements from the point of genes.

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

- chr the chromosome of a structural variation.
- start the start coordinate of a structural variation.
- end the end coordinate of a structural variation.

Value

A data frame with the following columns:

`locus`	the gene affected by structural variations.
`exon`	the effect of structural variations to exons of a specific gene.
`intron`	the effect of structural variations to introns of a specific gene.
`cds`	the effect of structural variations to cdss of a specific gene.
`utr`	the effect of structural variations to utrs of a specific gene.

Author(s)

Wen Yao

Examples

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

load(system.file("extdata/genome.anno.RData",package="intansv"))
str(msu_gff_v7)
gene.breakdancer.anno <- llply(breakdancer,geneAnnotation,
                                genomeAnnotation=msu_gff_v7)
str(gene.breakdancer.anno)
```

`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 supportted by no more than this threshold of read support would be discarded
`numMethodsSupDup`	Duplication clusters supportted by no more than this threshold of read support would be discarded
`numMethodsSupInv`	Inversion clusters supportted 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 varions 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/delly",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(genomeAnnotation, structuralVariation, windowSize=1000000)
```

Arguments

genomeAnnotation	GenomicRanges of the chromosome length.
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 splitted 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/delly",package="intansv"))
str(delly)

load(system.file("extdata/genome.anno.RData",package="intansv"))
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 genomic ranges of the genome annotation.
`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/delly",package="intansv"))
str(delly)

load(system.file("extdata/genome.anno.RData",package="intansv"))
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=1000000,
                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 occurence 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)
```

`readCnvnator` *Read in the structural variations predicted by CNVnator*

Description

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

Usage

```
readCnvnator(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

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

---

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

Arguments

dataDir	a directory containing the prediction results of DELLY.
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 read pairs supporting the occurence 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 DELLY. The paired-end deletions output files should be named using the suffix ".del" and the corresponding split-read output files should be named using the suffix ".del.br". The paired-end duplications output files should be named using the suffix ".dup" and the corresponding split-read output files should be named using the suffix ".dup.br". The paired-end inversions output files should be named using the suffix ".inv" and the corresponding split-read output files should be named using the suffix ".inv.br". See the example dataset for more details.

Value

A list with the following components:

`del`	the deletions predicted by DELLY.
`dup`	the duplications predicted by DELLY.
`inv`	the inversions predicted by DELLY.

Author(s)

Wen Yao

Examples

```
delly <- readDelly(system.file("extdata/delly",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, readsSupport=3,
          method="Lumpy", regSizeUpperCutoff=1000000,
          breakpointThres=200, scoreCut=0.1, ...)
```

Arguments

file	the file containing the prediction results of Lumpy.
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.
breakpointThres	a threshold to remove SVs with breakpoint with too large interval.
scoreCut	predictions with score larger than this threshold will be discarded.
...	parameters passed to read.table.

Details

The predicted SVs could be further filtered by the number of reads supporting the occurence 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 Lumpy. 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 Lumpy.
dup	the duplications predicted by Lumpy.
inv	the inversions predicted by Lumpy.

Author(s)

Wen Yao

Examples

```
lumpy <- readLumpy(system.file("extdata/ZS97.lumpy.pesr.bedpe",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 occurence 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` *Read in the structural variations predicted by SoftSearch*

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 occurence 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 SoftSearch. 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 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

`dataDir`	a directory containing the predictions of SVseq2.
`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 occurence 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:

`del` the deletions predicted by SVseq2.

Author(s)

Wen Yao

Examples

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

`svAnnotation` *Annotation of structural variations*

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 genomic ranges of the genome annotation.

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:

chr	the chromosome of a structural variation.
start	the start coordinate of a structural variation.
end	the end coordinate of a structural variation.
overlap	the overlap length between a structural variation and a specific gene or its element.
annotation	the annotation of a specific gene that overlap with the structural variation.
parent	the ID of a specific gene that overlap with the structural variation.

Author(s)

Wen Yao

Examples

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

load(system.file("extdata/genome.anno.RData",package="intansv"))
str(msu_gff_v7)
breakdancer.anno <- llply(breakdancer,svAnnotation,
                          genomeAnnotation=msu_gff_v7)
str(breakdancer.anno)
```

Index