Package 'RareVariantVis'

October 12, 2016

Type Package

Title Visualization of rare variants in whole genome sequencing data

Version 1.6.2

Date 2016-04-23

Author Tomasz Stokowy

Maintainer Tomasz Stokowy <tomasz.stokowy@k2.uib.no>

Description Genomic variants can be analyzed and visualized using many tools. Unfortunately, number of tools for global interrogation of variants is limited. Package RareVariantVis aims to present genomic variants (especially rare ones) in a global, per chromosome way. Visualization is performed in two ways - standard that outputs png figures and interactive that uses JavaScript d3 package. Interactive visualization allows to analyze trio/family data, for example in search for causative variants in rare Mendelian diseases.

License Artistic-2.0

LazyData TRUE

Depends BiocGenerics, VariantAnnotation, googleVis

Imports S4Vectors, IRanges, GenomeInfoDb, GenomicRanges

Suggests knitr, AshkenazimSonChr21

VignetteBuilder knitr

biocViews GenomicVariation, Sequencing, WholeGenome

NeedsCompilation no

R topics documented:

`CentromeresHg19` *Positions of centromeres on all chromosomes, human genome 19*

Description

CentromeresHg19 is a data frame providing positions of centromeres on all chromosomes of human genome 19 and its sizes. Size of centromere is also approximated in WinSize column to allow proper visualization of particular chromosomes.

Usage

```
data(CentromeresHg19)
```

Format

data frame

Value

test data

Source

UCSC table browser

`chromosomeVis` *Visualization of variants on the chromosome*

Description

Reads file containing table of variants (or alternatively, vcf file) from one chromosome and provides adequate visualization. Function outputs visualization png figure in current working directory. Figure illustrates variants (dots) in their genomic coordinates (x axis). Ratio of alternative reads and depth (y axis) gives information about type of variant: homozygous alternative (expected ratio 1) and heterozygous (expected ratio 0.5). Green dots represent rare variants that pass filters: coding/UTR, nonsynonymous variant with dbSNP frequency < 0.01. Orange vertical lines depict position of centromere. Red curve illustrates moving average of alternative reads/depth ratio. High values of this curve (exceedint 0.75) can suggest potential homozygous/deleterious regions. In addition, file containing table with rare variants only is generated. Table includes the same columns as input file.

Usage

```
chromosomeVis(file, sampleName, chromosome, centromeres, pngWidth,
pngHeight, plot, frFilter, dpFilter, vcf, posFilter)
```

Arguments

file	a connection or a character string giving the name of the text file/vcf to load. File should include following information: Start.position - for location on chromosome, SNP.Frequency - for dbSNP frequency, DP - sequencing depth, AD - allelic depths for reference and alternative alleles, Gene.name - gene symbol, Gene.component and Variant.type. Example of the table is provided in attached SonVariantsChr21.txt file. Example file can be loaded using command data(SonVariantsChr21) from library AshkenazimSonChr21.
sampleName	character string giving the sample name
chromosome	a numeric value giving the analyzed chromosome number
centromeres	a data frame providing location of centromeres on chromosomes
pngWidth	a numeric value setting width of output figure
pngHeight	a numeric value setting height of output figure
plot	logic value indicating if plot should be displayed or saved on the current working directory
frFilter	numeric value indicating dbSNP frequency filtering threshold (default 0.01). It is recommended not to change this value in the first phase of experiments.
dpFilter	numeric value indicating DP filtering threshold (default 10). It is recommended not to change this value in the first phase of experiments.
vcf	logic value indicating if input file is in vcf format. Vcf is supported, however data frame inputs are recommended.
posFilter	vector of values indicating positions of variants to be filtered out in addition to other filtering settings (default 0). Could be used for external database or in-house database filtering.

Value

comp1	function returns html file with visualization and txt file with rare variants

Author(s)

Tomasz Stokowy

Examples

```
## example data
library(AshkenazimSonChr21)

## example of data frame input
chromosomeVis(SonVariantsChr21, "1", 21, CentromeresHg19, 1600, 1200)
```

```
## example of vcf file input
fl <- system.file("extdata", "SonVariantsChr21.vcf.gz", package="AshkenazimSonChr21")
chromosomeVis(fl, "1", 21, CentromeresHg19, 1600, 1200, vcf = TRUE)
```

FatherRareVariantsChr21

Annotated rare variants on the chromosome 21, human genome 19

Description

FatherRareVariantsChr21 is a data frame consisting of annotated genomic variants coming from CompleteGenomics whole genome sequencing. Variants were filtered from GIAB Ashkenazim Trio father file using following filters: dbSNP frequency < 0.01, coding, nonsynonymous or nonsense variant, sequencing depth > 10

Usage

```
data(FatherRareVariantsChr21)
```

Format

data frame

Value

test data

Source

https://sites.stanford.edu/abms/content/giab-reference-materials-and-data

---

MotherRareVariantsChr21

Annotated rare variants on the chromosome 21, human genome 19

Description

MotherRareVariantsChr21 is a data frame consisting of annotated genomic variants coming from CompleteGenomics whole genome sequencing. Variants were filtered from GIAB Ashkenazim Trio mother file using following filters: dbSNP frequency < 0.01, coding, nonsynonymous or nonsense variant, sequencing depth > 10

Usage

```
data(MotherRareVariantsChr21)
```

Format

data frame

Value

test data

Source

https://sites.stanford.edu/abms/content/giab-reference-materials-and-data

---

movingAverage *Computation of moving average*

Description

Function calculates moving average from a vector of numeric values.

Usage

```
movingAverage(x, n, centered)
```

Arguments

x	a vector of numeric values for which moving average is computed
n	numeric value giving the frame length for moving average
centered	logic variable indicating if moving average should be centered (default = FALSE)

Value

comp1	function returns vector of moving average values

Author(s)

Winston Chang

Examples

```
movingAverage(1:20, n=3, centered=FALSE)
```

`multipleVis` *Interactive visualization of rare variants on the chromosome, applicable for multiple files*

Description

Reads files containing table of rare variants from one chromosome and provides adequate multiple sample visualization. Input files can be obtained from function chromosomeVis. Function outputs visualization html figure in current working directory. Figure depicts samples in subfigures. Subfigures illustrate variants (dots) in their genomic coordinates (x axis). Ratio of alternative reads and depth (y axis) gives information about type of variant: homozygous alternative (expected ratio 1) and heterozygous (expected ratio 0.5). Zoom to the figures is possible, by marking the region of interest with mouse left click. Right click induces zoom out and return to the original plot. Pointing on variants provides basic information about the variant - gene name and position on chromosome.

Usage

```
multipleVis(filesList,filesOrder,sampleNames,chromosome,centromeres,inHouse,inHouseSampNr)
```

Arguments

`filesList`	a string indicating files names that includes input files
`filesOrder`	vector including integers specifying order of displayed files. Default is 1:n, where n is number of files
`sampleNames`	a vector of strings including sample names. Default is names of the files
`chromosome`	a numeric value giving the analyzed chromosome number
`centromeres`	a data frame providing location of centromeres on chromosomes
`inHouse`	a logical value, if TRUE, creating inhouse table in .txt file, if FALSE, case is ignored
`inHouseSampNr`	a numeric value giving the threshold of variant repetition. Samples above this value are stored in inhouse table

Value

`comp1`	function returns html visualization file for trio and txt file with inheritance annotated rare variants

Author(s)

Mateusz Garbulowski

Examples

```
dummyMother = MotherRareVariantsChr21
dummyFather = FatherRareVariantsChr21
filesList = c(MotherRareVariantsChr21, SonRareVariantsChr21,
FatherRareVariantsChr21, dummyMother, dummyFather)
filesOrder = c(1, 2, 3, 4, 5)
sampleNames = c('Mother', 'Son', 'Father', 'dummyMother', 'dummyFather')
multipleVis(filesList, filesOrder, sampleNames, 21, centromeres = CentromeresHg19,
inHouse = TRUE, inHouseSampNr = 5)
```

---

rareVariantVis *Interactive visualization of rare variants on the chromosome*

Description

Reads file containing table of rare variants from one chromosome and provides adequate visualization. Input file can be obtained from function chromosomeVis. Function outputs visualization html figure in current working directory. Figure illustrates variants (dots) in their genomic coordinates (x axis). Ratio of alternative reads and depth (y axis) gives information about type of variant: homozygous alternative (expected ratio 1) and heterozygous (expected ratio 0.5). Zoom to the figures is possible, by marking the region of interest with mouse left click. Right click induces zoom out and return to the original plot. Pointing on variants provides basic information about the variant - gene name and position on chromosome.

Usage

```
rareVariantVis(file, sample, chromosome, centromeres)
```

Arguments

file	a connection or a character string giving the name of the file to load. File should include tab delimited text file with columns: Start.position - for location on chromosome, SNP.Frequency - for dbSNP frequency, DP - sequencing depth, AD - allelic depths for reference and alternative alleles, Gene.name - gene symbol.
sample	character string giving the sample name
chromosome	a numeric value giving the analyzed chromosome number
centromeres	a data frame providing location of centromeres on chromosomes

Value

comp1	function returns html file with visualization of rare variants

Author(s)

Tomasz Stokowy

Examples

```
rareVariantVis(SonRareVariantsChr21, "1", 21, CentromeresHg19)
```

`SonRareVariantsChr21` *Annotated rare variants on the chromosome 21, human genome 19, subset of SonVariantsChr21*

Description

SonRareVariantsChr21 is a data frame consisting of annotated genomic variants coming from CompleteGenomics whole genome sequencing. Variants were filtered from SonVariantsChr21 file using following filters: dbSNP frequency < 0.01, coding, nonsynonymous or nonsense variant, sequencing depth > 10

Usage

```
data(SonRareVariantsChr21)
```

Format

data frame

Value

test data

Source

https://sites.stanford.edu/abms/content/giab-reference-materials-and-data

---

`trioVis` *Interactive visualization of rare variants on the chromosome, applicable for trio data*

Description

Reads files containing table of rare variants from one chromosome and provides adequate trio visualization. Input files can be obtained from function chromosomeVis. Function outputs visualization html figure in current working directory. Figure depicts three samples in subfigures: mother, index and father. Subfigures illustrate variants (dots) in their genomic coordinates (x axis). Ratio of alternative reads and depth (y axis) gives information about type of variant: homozygous alternative (expected ratio 1) and heterozygous (expected ratio 0.5). Zoom to the figures is possible, by marking the region of interest with mouse left click. Right click induces zoom out and return to the original plot. Pointing on variants provides basic information about the variant - gene name and position on chromosome. Some variants are clustering, especially in polymorphic genomic regions

like HLA or MUC. Some of this regions are also technical sequencing artefacts. If variant belongs to such region (has more than 3 rare non-synonymous coding variants in region 100 000 bases to the left and to the right from variant) it is denoted as problematic region variant. It is also denoted as such in output text file.

Usage

```
trioVis(fileMother, fileIndex, fileFather, sampleMother,
    sampleIndex, sampleFather, chromosome, centromeres)
```

Arguments

`fileMother`	a connection or a character string giving the name of the file with mother variants to load. File should include tab delimited text file with columns: Start.position - for location on chromosome, SNP.Frequency - for dbSNP frequency, DP - sequencing depth, AD - allelic depths for reference and alternative alleles, Gene.name - gene symbol.
`fileIndex`	a connection or a character string giving the name of the file with mother variants to load. File should include tab delimited text file with columns: Start.position - for location on chromosome, SNP.Frequency - for dbSNP frequency, DP - sequencing depth, AD - allelic depths for reference and alternative alleles, Gene.name - gene symbol.
`fileFather`	a connection or a character string giving the name of the file with mother variants to load. File should include tab delimited text file with columns: Start.position - for location on chromosome, SNP.Frequency - for dbSNP frequency, DP - sequencing depth, AD - allelic depths for reference and alternative alleles, Gene.name - gene symbol.
`sampleMother`	character string giving the mother sample name
`sampleIndex`	character string giving the index sample name
`sampleFather`	character string giving the father sample name
`chromosome`	a numeric value giving the analyzed chromosome number
`centromeres`	a data frame providing location of centromeres on chromosomes

Value

`comp1`	function returns html visualization file for trio and txt file with inheritance annotated rare variants

Author(s)

Tomasz Stokowy

Examples

```
trioVis(MotherRareVariantsChr21, SonRareVariantsChr21, FatherRareVariantsChr21, "Mother", "Son",
"Father", 21, CentromeresHg19)
```

Index