Package 'facopy'

March 4, 2018

Type Package

Title Feature-based association and gene-set enrichment for copy number alteration analysis in cancer

Version 1.13.0

Date 2015-23-07

Author David Mosen-Ansorena

Maintainer David Mosen-Ansorena <dmosen.gn@cicbiogune.es>

Imports annotate, data.table, DOSE, FactoMineR, GO.db, GOstats, graphite, igraph, S4Vectors, IRanges, MASS, nnet, reshape2, Rgraphviz, scales

Depends R (>= 3.0), methods, cgdsr (>= 1.1.30), coin (>= 1.0), ggplot2, gridExtra, facopy.annot, grid

Description facopy is an R package for fine-tuned cancer CNA association modeling. Association is measured directly at the genomic features of interest and, in the case of genes, downstream gene-set enrichment analysis can be performed thanks to novel internal processing of the data. The software opens a way to systematically scrutinize the differences in CNA distribution across tumoral phenotypes, such as those that relate to tumor type, location and progression. Currently, the output format from 11 different methods that analyze data from whole-genome/exome sequencing and SNP microarrays, is supported. Multiple genomes, alteration types and variable types are also supported.

biocViews Software, CopyNumberVariation, GeneSetEnrichment, GenomicVariation, Genetics, Microarray, Sequencing, Visualization

R topics documented:

`facopy-package` *facopy: feature-based association and gene-set enrichment for copy number alteration analysis in cancer*

Description

facopy is an R package for fine-tuned cancer CNA association modeling. Association is measured directly at the genomic features of interest and, in the case of genes, downstream gene-set enrichment analysis can be performed thanks to novel internal processing of the data. The software opens a way to systematically scrutinize the differences in CNA distribution across tumoral phenotypes, such as those that relate to tumor type, location and progression. Currently, the output format from 11 different methods that analyze data from whole-genome/exome sequencing and SNP microarrays, is supported. Multiple genomes, alteration types and variable types are also supported.

Details

Package:	facopy
Type:	Package
License:	CC BY-NC 4.0

To analyze your data with facopy, first read copy number data through `readCNData`. Then, `addVariables` and `addFeatures`, and you are ready to start scrutinizing the data in multiple ways thanks to the many provided functions.

Author(s)

David Mosen-Ansorena

addFeatures *Add Feature Annotation*

Description

Adds feature annotation to a facopyInfo object and calculates the overlapping frequency in the sample set for every combination of copy number and feature.

Usage

```
addFeatures(fad, what = c("ensembl", "cancergene", "oncogene",
                          "tumorsupressor", "lincRNA", "mirnas")[1],
            genome = c("hg18", "hg19", "mm8")[1],
            lMargin = 0, rMargin = 0, minoverlap = 1,
            data = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data. Variable information should have been added beforehand on this object (see addVariables).
what	A character indicating the name of the feature set to use in subsequent analyses. If the name belongs to a set of features integrated in facopy, annotation is automatically loaded and the parameter data is not used.
genome	Reference genome build used in the copy number calling step prior to running facopy. Some feature sets may not be available depending on the selected genome.
lMargin	Number of base pairs with which to expand the left flanking side of each feature, increasing its chance to overlap alterations.
rMargin	Number of base pairs with which to expand the right flanking side of each feature, increasing its chance to overlap alterations.
minoverlap	Minimum overlap, in base pairs, between each feature and alteration.
data	If the genomic features of interest are not integrated within facopy, their information can be manually loaded using this parameter. Either a character or a data.frame. If a character, it indicates the file with the information, with headers and tab-delimited. In either case, the information shoul follow a structure with the following columns: - chr Chromosome in which the genomic feature lies. - bp_st Starting genomic position of the feature within the arm. - bp_en Ending genomic position of the feature within the arm. - feature Name of the feature. - chr_q_arm Chromosome arm in which the feature lies.

Value

facopyInfo object with information on both variable and feature annotations.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

myStudy = addFeatures(myStudy, "oncogene", "hg18")
```

---

addVariables *Add Variable Annotation*

Description

Attaches phenotypic information to a facopyInfo object that had been created from copy number data.

Usage

```
addVariables(fad, varInfo, varTypes,
             varColumns = NULL, varValues = NULL,
             varColumnsNames = NULL, varValuesNames = NULL,
             ...)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
varInfo	A character or a data.frame. If it is a character, it should indicate the name of the file with information on the variables. If it is a codedata.frame, a column named code should contain the samples' unique identifiers.
varTypes	A vector of characters with the same length as the number of variables whose information will be loaded. Variables can be either categorical or quantitative. Planning ahead for subsequent analyses, it might be best to regard ordered variables as categorical. Vector elements should equal either "categorical" or "quantitative".
varColumns	An optional character vector. If the table with variable information contains extra columns, this parameter allows the specification of those columns with information on the variables to use.
varValues	An optional list of vectors. Each vector is either null or contains the subset of values to be used in subsequent analyses for the corresponding variable.
varColumnsNames	An optional character vector with the names of the variables, if different from the corresponding column names in the table.
varValuesNames	An optional list of vectors with the names of the variable values, if different from the corresponding values in the columns. If a vector is null, the corresponding value names are not changed, which is useful for quantitative variables with many possible values.
...	Extra parameters passed to the function are in turn passed to read.table, such as the presence of header and the column separation character. Only used if varInfo is a character that indicates the name of the file with variable information.

Value

facopyInfo object with incorporated phenotypic information.

Author(s)

David Mosen-Ansorena

Examples

```
data(myCalls)
data(myVariables)

myStudy = addVariables(myCalls, myVariables, c("continuous","categorical"))
```

---

alterationSummary *Alteration Summary*

Description

Summary of copy number calls in a facopyInfo object, by alteration type and chromosome arm.

Usage

```
alterationSummary(fad, filename = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
filename	Optional. A character specifying the name of the file to which to output the alteration summary table.

Details

A data.frame is always returned, and optionally written to a file.

Value

A data.frame with alteration frequencies by arm and alteration type.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

alterationSummary(myStudy)
```

calculateCor *Calculate Correlation Between Copy Number and Expression*

Description

Calculates the gene-wise correlation between copy number and expression data, which might come from: (1) the facopyInfo object and user-provided expression data or (2) from an external database.

Usage

```
calculateCor(fad, exprProfile, db = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
exprProfile	Either a character or a data.frame. If db is NULL, a data.frame with the expression data or a character that indicates the file with the expression data, with headers and tab-delimited. Otherwise, a character indicating the the name of the expression profile in the Cancer Genomics Data Server. See http://www.cbioportal.org/public-po
db	A character indicating the the name of the dataset, in the Cancer Genomics Data Server, from which to get the expression profile. See http://www.cbioportal.org/public-portal

Details

Only use this function and facopyEnrichment if you selected some kind of gene collection as genomic features.

Value

The returned object is used to select genes in the enrichment process encapsulated in the facopyEnrichment function.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

eCor = calculateCor(myStudy, "mrna_merged_median_Zscores", "coadread_tcga_pub")
head(eCor$cor)
```

facopy *Feature-Based Copy Number Association Analysis*

Description

Main function in the facopy package. It performs statistical association between copy number data and further variables at each genomic feature of interest.

Usage

```
facopy(fad, alteration, model, nullModel = NULL,
       modelPart = c("response","predictor","unknown","whole")[1],
       strata = NULL, toOrdered = NULL, toIntervals = NULL,
       sel = NULL, plot = FALSE, pvalThr = 0.05, db = NULL,
       link = c("logit", "probit")[1], parametric = FALSE,
       design = c("binary", "versus", "lvog")[1],
       FUN, ...)
```

Arguments

fad	facopyInfo object with a certain study's facopy data. Feature information should have been added beforehand on this object (see addFeatures).
alteration	A character, the name of the combination of alterations to be considered for the analysis. It should be one of the following: - amplifications All amplifications (CN>2). - deletions All deletions (CN<2). - loh All loss of heterozygosity (LOH), regardless of copy number. - cnas All copy number alterations (CN<>2). - any Any kind of alteration. - all Any kind of alteration, same as any. - onlygain Only non-LOH amplifications. - someloss All deletions plus LOH alterations.
model	A character. Model, or part of it (response or predictor), whose association will be measured. If modelPart="response", the name of a single variable, representing the response of the association model. If modelPart="predictor", a linear predictor of the copy number, using a combination of the variables. If modelPart="unknown", the association is measured as the strenght of the relationship between copy number and the specified combination of variables. The model is limited to x1 + ... + xn | strata, see independence_test. If modelPart="whole", a character representation of a formula in compact symbolic form. Use the @ (at) symbol to refer to the copy number variable.
nullModel	A character. The null model against which to evaluate the fitness of the association model. Only used if modelPart="predictor" or modelPart="whole".
modelPart	A character. Indicates what part of the association model is defined in the model parameter.

strata	A character. The name of a categorical variable that defines stratification blocks in the association model. Only used if modelPart="response" or modelPart="unknown".
toOrdered	Certain categorical variables can also be understood as ordered. This parameter takes a list of named vectors, where the name of each vector is a variable name and its contents reflect the quantification of the variable values, in the same order as defined in addVariables.
toIntervals	Quantitative variables can be broken down into intervals. This parameter takes a list of named vectors, where the name of each vector is a variable name and its contents reflect the breaks of the variable values (excluding bottom and top limits).
sel	A vector of feature names on which to perform the association. Leave to NULL for genome-wide association over all the features specified in addFeatures. Plotting will only be done if sel=NULL.
plot	A logicil indicating whether to output a composite plot with an arm-wise display of genome-wide alteration frequencies. If the model consists of a single variable, frequencies will be broken down by variable value. Features with significant associations and additional information pulled from external databases can be displayed as overlaid layers.
pvalThr	Significant associations under this threshold will be shaded in the output plot. Only used if plot=TRUE.
db	An optional string representing the name of a database whose data will be overlaid in the output plot. Typically, the format is "[database]_[dataset]". The total amplification plus deletion frequencies will be displayed unless alteration indicates either amplification or deletion. In such cases, only the matching alterations are displayed. Only used if plot=TRUE. See getFacopyInfo for a list of available data sets.
link	A character, link function to be used with the multinomial error distribution in logistic regression models. See glm.
parametric	A logical that indicates whether to perform one-way ANOVA instead of Kruskal-Wallis in the association of copy number with quantitative variables.
design	Depending on the chosen alteration, different designs are available. The simplest design is binary: an alteration exists or it does not. The versus design, for CNAs, assigns a value of -1, 0 or 1 depending on whether a deletion, no copy number change or an amplification exists for a given feature. The vlog design, for all (any) alterations, assigns a value of -1, 0 or 1 depending on whether a deletion or LOH, no copy number change or an amplification without LOH exists.
FUN	A function that tests a model. Only used if modelPart="whole". Functions from the coin package and those that inherit from either the lm (such as glm) are supported, as well as those that directly return a pvalue. Thus, the function of interest can be wrapped in a wrapper function that provides the pvalue.
...	Further arguments for the FUN function. Only used if modelPart="whole".

Details

Only the facopyInfo object, the alteration type and a simple model (e.g. the name of a variable) are required. The rest of the parameters tune up the association model and control the graphical output.
Alterations in the selected external database, if selected, are depicted as grey overlaid bars. Significant regions are depicted in turn as overlaid rectangles that go from top to bottom.

Value

A data.frame with the following columns:

feature	Name of the genomic feature.
p_value	Pvalue from the association test under the given model at the genomic feature.
chr_q_arm	Chromosome and arm in which the genomic feature lies.
bp_st	Starting genomic position of the feature within the chromosome.
bp_en	Ending genomic position of the feature within the chromosome.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

genes = facopy(myStudy, "amp", "stage")
head(genes)
```

---

facopyEnrichment *Gene-set Enrichment Analysis*

Description

Performs gene-set enrichment analysis of cancer copy number data, based on a gene table, a facopyInfo object and the correlations between gene copy number and expression data.

Usage

```
facopyEnrichment(fad, geneTable, cor, outFolder,
                 pvalThr = 0.05, corThr = 0.1, plotThr=0.001)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
geneTable	A table with the format of facopy function's output, containing at least the columns regarding feature name, p-value and chromosome arm.
cor	A list generated by a call to calculateCor, containing information on correlation between gene copy number and expression data.
outFolder	The folder to which to write the enrichment results.
pvalThr	Maximum p-value required to consider a gene to be significantly associated.
corThr	Minimum R2 required to consider a gene's copy number and expression to be sufficiently correlated.
plotThr	Pathways from Biocarta, KEGG and Reactome with lower p-values than this integer will have their graphs plotted in the output folder.

Details

Only use this function and `calculateCor` if you selected some kind of gene collection as genomic features.

Value

The input gene table is returned with correlation values attached, ordered by p-value and then by R2.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

# then perform an association analysis
# genes = facopy(myStudy, "amp", "stage")

# calculate expression-CN correlations
# eCor = calculateCor(myStudy, "mrna_merged_median_Zscores", "coadread_tcga_pub")

# and run facopy enrichment
# facopyEnrichment(myStudy, genes, eCor, "~/myFolder/stageAmpEnrichment")
```

---

facopyInfo-class *Class* "facopyInfo"

Description

`facopyInfo` is the main unit of information used through the analysis of your data with `facopy`.

Content

These objects are generated by methods in the package and their basic information can be recovered by the `summary` method.

Author(s)

David Mosen-Ansorena

Examples

```
showClass("facopyInfo")
```

facopyPlot *Plot a Genome-Wide Overview of Association Results*

Description

The function plots alteration frequencies by genomic feature, broken down by variable value (if the variable is discrete). It also allows to plot relevant alteration frequencies within a specified external dataset.

Usage

```
facopyPlot(fad, alteration, varName, db = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
alteration	A character describing the kinds of alteration to include. It should be one of the following: - amplifications All amplifications (CN>2). - deletions All deletions (CN<2). - loh All loss of heterozygosity (LOH), regardless of copy number. - cnas All copy number alterations (CN<>2). - any Any kind of alteration. - all Any kind of alteration, same as any. - onlygain Only non-LOH amplifications. - someloss All deletions plus LOH alterations.
varName	A character indicating the variable of interest within the facopyInfo object. Call summary on your facopyInfo object to see the names of defined variables.
db	An optional string representing the name of a database whose data will be overlaid in the output plot. Typically, the format is "[database]_[dataset]". The total amplification plus deletion frequencies will be displayed unless alteration indicates either amplification or deletion. In such cases, only the matching alterations are displayed.

Details

Alterations in the selected external database, if selected, are depicted as grey overlaid bars.

Value

A plot is generated in the graphics device. Nothing is returned.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

facopyPlot(myStudy, "amp", "stage", db="gsk_colon")
```

getFacopyInfo *List Available facopy Input and Characteristics*

Description

Lists of supported items are printed for reference.

Usage

```
getFacopyInfo()
```

Details

The following lists of supported items are printed: variable types, alteration combinations, input formats for CN calls, genome builds, genomic features.

Author(s)

David Mosen-Ansorena

Examples

```
getFacopyInfo()
```

myCalls *Example facopy Object with Just Copy Number Calls*

Description

Example `facopyInfo` object with just the copy number calls, generated using the `readCNData` function.

Usage

```
data(myCalls)
```

Format

The format is: formal class 'facopyInfo'

Examples

```
data(myCalls)

summary(myCalls)
```

myStudy *Example Complete facopy Object*

Description

Example facopyInfo object, generated using the readCNData, addVariables and addFeatures functions.

Usage

```
data(myStudy)
```

Format

The format is: formal class 'facopyInfo'

Examples

```
data(myStudy)

summary(myStudy)
```

myVariables *Example Phenotypic Annotation*

Description

This data.frame contains three variables with annotation on phenotypes for 20 samples.

Usage

```
data(myVariables)
```

Format

A data frame with 20 observations (samples) on the following 3 variables.

code: a numeric vector
age: a numeric vector
stage: a numeric vector

Examples

```
data(myVariables)

head(myVariables)
```

plotBar *Overview of Alterations by Chromosome Arm*

Description

The function plots amplification and deletion frequencies by chromosome arm, as well as LOH frequencies

Usage

```
plotBar(fad, byFeature = TRUE, sel = NULL, selColors = NULL,
        ylim = c(-1, 1), baseColor = "black",
        varName = NULL, value = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
byFeature	A logical. If TRUE, arm frequencies are calculated as the mean alteration frequencies across genomic features in each arm. If TRUE, they are calculated as the mean proportion of altered base pairs in each arm.
sel	A character vector with the chromosome arms that will be highlighted using the colors in selColors. Example: c("1q","9p").
selColors	A vector of colors with the same length as the parameter sel. Each color in the vector will be used for the selected chromosome arm in the same position within sel.
ylim	A two integer vector indicating frequency limits in the plot. Deletion frequency is indicated with a negative number. Default is c(-1,1).
baseColor	Base color of the triangles that indicate frequencies in the plot. Default is "black".
varName	Either NULL or a character indicating the variable of interest within the facopyInfo object. Call summary on your facopyInfo object to see the names of defined variables. If NULL, the frequencies across all samples are plotted. Otherwise, only samples with the variable value in the parameter value are selected for the computation. In this latter case, only categorical variables are allowed.
value	Value that the samples should have for the selected variable. Only these samples are selected for the frequency calculation.

Details

Amplification frequencies are depicted by chromosome arm over the horizontal line, while deletion frequencies appear below it. The frequencies of alterations with LOH are depicted as smaller triangles.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

# select some chromosome arms to highlight
myArms = c("8q","13q","20q","8p","18q")
myColors = c(rainbow(15)[1:3], rainbow(15)[10:11])

plotBar(myStudy, TRUE, myArms, myColors, ylim=c(-0.5,1))
```

`plotHist` *Stacked Histograms*

Description

The function plots a stacked histogram of chromosome arm alterations by variable value

Usage

```
plotHist(fad, alteration, varName,
         sel = NULL, selColors = NULL, selOnly = FALSE,
         baseColor = "black", bin = 0.05, xmax = 1, ymax)
```

Arguments

`fad`	`facopyInfo` object with a certain study's `facopy` data.
`alteration`	A character describing the kinds of alteration to include. It should be one of the following: - `amplifications` All amplifications (CN>2). - `deletions` All deletions (CN<2). - `loh` All loss of heterozygosity (LOH), regardless of copy number. - `cnas` All copy number alterations (CN<>2). - `any` Any kind of alteration. - `all` Any kind of alteration, same as `any`. - `onlygain` Only non-LOH amplifications. - `someloss` All deletions plus LOH alterations.
`varName`	A character indicating the variable of interest within the `facopyInfo` object. A stacked histogram will be generated for every value of the variable. Only available for discrete variables. Call `summary` on your `facopyInfo` object to see the names of defined variables.
`sel`	A character vector with the chromosome arms that will be highlighted using the colors in `selColors`. Example: `c("1q","9p")`.
`selColors`	A vector of colors with the same length as the parameter `sel`. Each color in the vector will be used for the selected chromosome arm in the same position within `sel`.
`selOnly`	A logical indicating whether to only plot the frequencies of selected chromosome arms.
`baseColor`	Base color of the triangles that indicate frequencies in the plot. Default is `"black"`.

bin	The width of the histogram cells. Default is 0.05.
xmax	The maximum limit of the x axis in the plot. Default is 1.
ymax	The maximum limit of the y axis in the plot.

Details

Alteration frequencies are calculated genomic feature-wise within each chromosome arm.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

# select some chromosome arms to highlight
myArms = c("8q","13q","20q","8p","18q")
myColors = c(rainbow(15)[1:3], rainbow(15)[10:11])

plotHist(myStudy, "amp", "stage", myArms, myColors, bin=0.1, ymax=80)
```

---

plotPCA *facopy PCA*

Description

Plot a PCA of the samples based on their similarity given a certain set of alterations.

Usage

```
plotPCA(fad, alteration, varName, sel = NULL,
        design = c("binary", "versus", "lvog")[1],
        do.plot = TRUE, by.size = TRUE, cex = 4)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
alteration	A character describing the kinds of alteration to include. It should be one of the following: - amplifications All amplifications (CN>2). - deletions All deletions (CN<2). - loh All loss of heterozygosity (LOH), regardless of copy number. - cnas All copy number alterations (CN<>2). - any Any kind of alteration. - all Any kind of alteration, same as any. - onlygain Only non-LOH amplifications. - someloss All deletions plus LOH alterations.

varName	A character indicating the variable of interest within the facopyInfo object. Points representing samples in the PCA will be colored according to the classification in such variable. Call summary on your facopyInfo object to see the names of defined variables.
sel	A character vector with the chromosome arms that will be taken into account in the analysis. Example: c("1q","9p").
design	Depending on the chosen alteration, different designs are available. The simplest design is binary: an alteration exists or it does not. The versus design, for CNAs, assigns a value of -1, 0 or 1 depending on whether a deletion, no copy number change or an amplification exists for a given feature. The vlog design, for all (any) alterations, assigns a value of -1, 0 or 1 depending on whether a deletion or LOH, no copy number change or an amplification without LOH exists.
do.plot	A logical indicating whether to produce graphical output.
by.size	A logical indicating whether the width of the points' border represents the frequency of the selected alterations.
cex	A numerical value giving the amount by which points in the plot should be scaled relative to the base size. Default is 4.

Value

The results of the PCA are returned and the graphical output is optional.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

pca = plotPCA(myStudy, "any", "stage")
head(pca$eig)
```

---

plotZoom *Zoom In to Plot a Chromosome Arm or a Genomic Feature*

Description

The function shows the selected alterations either in a chromosome arm or near a genomic feature, color-coded by variable value.

Usage

```
plotZoom(fad, what = c("feature", "arm"), name,
         alteration, varName, margin = 0)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
what	Charcter indicating whether to plot selected alterations either near a genomic feature or in a chromosome arm. Default is feature.
name	Name of the genomic feature or chromosome arm of interest. Examples: "MIRN181A2", "14q".
alteration	A character describing the kinds of alteration to include. It should be one of the following: - amplifications All amplifications (CN>2). - deletions All deletions (CN<2). - loh All loss of heterozygosity (LOH), regardless of copy number. - cnas All copy number alterations (CN<>2). - any Any kind of alteration. - all Any kind of alteration, same as any. - onlygain Only non-LOH amplifications. - someloss All deletions plus LOH alterations.
varName	A character indicating the variable of interest within the facopyInfo object. Bars representing alterations will be colored according to the classification of the corresponding samples in such variable. Call summary on your facopyInfo object to see the names of defined variables.
margin	If parameter what is "feature", a margin downstream and upstream of the genomic feature to consider when searching for overlapping alterations.

Details

One line per sample in the study is displayed in the case of chromosome arms, whereas only those samples with alterations are displayed in the case of genomic features.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

plotZoom(myStudy, "feat", "RAD51L1", "amp", "stage")

plotZoom(myStudy, "arm", "8p", "del", "stage")
```

---

preview *Alteration and Varible Preview*

Description

A wrapper function for variableSummary, alterationSummary and variableCor.

Usage

```
preview(fad, folder = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
folder	Optional. A character specifying the name of the file to which to output the table with variable correlations.

Details

A list of data.frame is always returned, and optionally written to files in a specified folder.

Value

A list of data.frame with the results of calling the functions variableSummary, alterationSummary and variableCor.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

myCallsPreview = preview(myStudy)

myCallsPreview$byVar
head(myCallsPreview$byAlt)
myCallsPreview$varCor
```

---

readCNData *Read Copy Number Data*

Description

The function reads the copy number output generated by one of the supported copy number analysis tools

Usage

```
readCNData(folder, method = NULL, sex = c("X"), FUN = NULL, version = NULL,
           window = NULL, rankThr = NULL, pfbFilename, lengthThr = 10, ...)
```

Arguments

folder	A character indicating the path where the files with copy number information are located. In order to avoid possible conflicts, try not to place any additional files within the folder.
method	Either NULL or a character. If a character, it specifies the method used generate the copy number output. Currently supported: "seqcna", "cnanorm", "patchwork", "freec", "oncosnp", "oncosnp-seq", "gap" and "exomecnv". If NULL, a generic input will be assumed with the following columns: chromosome, segment start, segment end, segment length and segment type. The

segment types should be coded as 1 (deletion), 2 (normal), 3 (amplification), 11 (LOH deletion), 12 (LOH normal) or 13 (LOH amplification).

`sex` Either `NULL` or a character vector. If a character vector, it should contain the names of the sexual chromosomes that should be read.

`FUN` A pattern replacement function. Some copy number analysis tools add suffixes and file extension characters to the names of their output files. This function allows keeping just the sample names. Leave to `NULL` for default replacement based on the selected method.

`version` A character with the software version of the copy number analysis tool. The output may vary between versions of the same method. Leave to `NULL` if unknown and check your software version if reading the copy number data is not successful.

`window` An integer. Required for output files from CNAnorm, which do not contain information on the window length used during the copy number analysis process.

`rankThr` An integer. Threshold on the Rank parameter provided by OncoSNP-SEQ and the newer versions of OncoSNP (>1.3).

`pfbFilename` For GAP, name of a file with population B allele frequencies (PFB), the same one used during the copy number analysis process.

`lengthThr` An integer. For GAP, filter those alterations that span less SNPs than this threshold.

`...` If a method is not selected (`NULL`), further arguments passed to the `read.delim` function, in charge of reading the input files from the selected folder.

Value

A `facopyInfo` object with copy number data.

Author(s)

David Mosen-Ansorena

Examples

```
# myCalls = readCNData("~/myFolder/", "seqcna")
# myCalls = readCNData("~/myFolder/", "gap", pfbFilename="~/myPfb.pfb")
# myCalls = readCNData("~/myFolder/", "cnanorm", window=50000)
```

`variableCor` *Variable Correlations*

Description

Correlations between pairs of variables in a `facopyInfo` object.

Usage

```
variableCor(fad, filename = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
filename	Optional. A character specifying the name of the file to which to output the table with variable correlations.

Details

A data.frame is always returned, and optionally written to a file.

Value

A data.frame with alteration variable correlations, with a combination of the appropriate parametric and non-parametric tests.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

variableCor(myStudy)
```

---

variableSummary *Variable Summary*

Description

For each pair of values in a variable, it performs a statistical test to see whether samples with such variable values present significantly different spans of a certain alteration type.

Usage

```
variableSummary(fad, filename = NULL)
```

Arguments

fad	facopyInfo object with a certain study's facopy data.
filename	Optional. A character specifying the name of the file to which to output the table with variable correlations.

Details

A data.frame is always returned, and optionally written to a file.

Value

A data.frame with alteration spans and significance for each pair of values in a variable and alteration type.

Author(s)

David Mosen-Ansorena

Examples

```
data(myStudy) # load example study

variableSummary(myStudy)
```

Index