Package ‘denovolyzeR’

October 13, 2022

Title  Statistical Analyses of De Novo Genetic Variants
Version  0.2.0
Date  2016-08-01

Description  An integrated toolset for the analysis of de novo (sporadic) genetic sequence variants. denovolyzeR implements a mutational model that estimates the probability of a de novo genetic variant arising in each human gene, from which one can infer the expected number of de novo variants in a given population size. Observed variant frequencies can then be compared against expectation in a Poisson framework. denovolyzeR provides a suite of functions to implement these analyses for the interpretation of de novo variation in human disease.

Depends  R (>= 3.1.0)
Imports  dplyr (>= 0.3), reshape2 (>= 1.4)
License  GPL-3
LazyData  true
Suggests  knitr, rmarkdown
VignetteBuilder  knitr

URL  http://denovolyzeR.org

BugReports  http://github.com/jamesware/denovolyzeR/issues
RoxygenNote  5.0.1
NeedsCompilation  no

Author  James Ware [aut, cre],
        Jason Homsy [ctb],
        Kaitlin Samocha [ctb]

Maintainer  James Ware <j.ware@imperial.ac.uk>
Repository  CRAN
Date/Publication  2016-08-01 14:55:41
**R topics documented:**

- autismDeNovos .................................................. 2
- denovolyze ...................................................... 2
- denovolyzeMultiHits ........................................... 5
- denovolyzeR ..................................................... 6
- fmrpGenes ......................................................... 7
- parseInput ....................................................... 7
- PermuteMultiHits ............................................... 8
- viewProbabilityTable ........................................... 9

**autismDeNovos**

*de novo variants found in 1,078 autism trios*

**Description**

*de novo variants found in 1,078 autism trios, published in Nature Genetics [http://www.nature.com/doifinder/10.1038/ng.3050]*

**Format**

A data frame with 1096 obs of 2 variables:

- **gene** Gene symbol of gene containing de novo variant
- **class** Functional class of variant: "syn" = synonymous, "mis" = missense, "non" = nonsense, "splice" = canonical splice site, "frameshift" = frameshift indel

**References**

[http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222185/]

---

**denovolyze**

*Evaluates burden of de novo variation against expectation*

**Description**

Determines whether the test population carry more *de novo* variants than expected. Variants may be grouped by variant class (e.g. are there more LOF variants than expected, across the whole dataset?), or by gene (are there more variants of a given class in SCN2A?).
Usage

denovolyze(genes, classes, nsamples, groupBy = "class",
includeGenes = "all", includeClasses = c("syn", "mis", "misD", "non",
"stoploss", "startgain", "splice", "frameshift", "lof", "prot", "protD",
"all"), geneId = "geneName", signifP = 3, roundExpected = 1,
probTable = NULL, misD = NULL)

denovolyzeByClass(genes, classes, nsamples, groupBy = "class",
includeGenes = "all", includeClasses = c("syn", "mis", "lof", "prot",
"all"), geneId = "geneName", signifP = 3, roundExpected = 1,
probTable = NULL)

denovolyzeByGene(genes, classes, nsamples, groupBy = "gene",
includeGenes = "all", includeClasses = c("lof", "prot"),
geneId = "geneName", signifP = 3, roundExpected = 1, probTable = NULL)

Arguments

genes A vector of genes containing de novo variants.
classes A vector of classes of de novo variants. Standard supported classes are "syn" (synonymous), "mis" (missense), "non" (nonsense), "splice" (splice), "frameshift" (frameshift) and "lof" (loss of function = non + splice + frameshift). Additional classes that are supported by the code, but are not included in the built-in probability tables, are "stoploss", "startloss", "misD" (damaging missense). These labels may be used for user-supplied probability tables. If "misD" is present, then "mis" (in the input) implies non-damaging missense.
nsamples Number of individuals considered in de novo analysis.
groupBy Results can be tabulated by "gene", or by variant "class"
includeGenes Genes to include in analysis. "all" or a vector of gene names.
includeClasses Determines which variant classes are tabulated in output. In addition to the input classes, summaries can be produced for "prot" (protein-altering = mis + lof), "all", and "protD" (protein damaging = misD + lof, only available if misD included in user-specified probability table). If "misD" is present, then "mis" will return statistics for all missense. Non-damaging missense are not analysed separately.
geneId Gene identifier used. One of "hgncID", "hgncSymbol", "enstID", "ensgID" or "geneName" (default, equals ensembl "external_gene_name")
signifP Number of significant figures used to round p-values in output.
roundExpected Number of decimal places used to round expected burdens in output.
probTable Probability table. A user-defined table of probabilities can be provided here, to replace the probability table included in the package.
misD If the user-specified probability table contains probabilities for a sub-category of missense variants (e.g. predicted to be damaging by an in silico algorithm), this column should be called misD, or the alternative name should be specified here.
Details

Analyses can be restricted to a subset of genes, and/or a subset of variant classes
See vignette("denovolyzeR_intro") for more information.

Value

Returns a data frame

Functions

- **denovolyzeByClass**: denovolyzeByClass
- **denovolyzeByGene**: denovolyzeByGene

Examples

```r
### denovolyze
denovolyze(genes=autismDeNovos$gene,  
classes=autismDeNovos$class,  
nsamples=1078)

### denovolyzeByClass
denovolyzeByClass(genes=autismDeNovos$gene,  
classes=autismDeNovos$class,  
nsamples=1078)

# this convenience function is identical to:
denovolyze(genes=autismDeNovos$gene,  
classes=autismDeNovos$class,  
nsamples=1078,  
groupBy="class",  
includeClasses=c("syn","mis","lof","prot","all"),  
includeGenes="all"  
)

### denovolyzeByGene
denovolyzeByGene(genes=autismDeNovos$gene,  
classes=autismDeNovos$class,  
nsamples=1078)

# this is identical to:
denovolyze(genes=autismDeNovos$gene,  
classes=autismDeNovos$class,  
nsamples=1078,  
groupBy="gene",  
includeClasses=c("lof","prot"),
```
denovolyzeMultiHits

Description

Are there more genes containing >1 de novos than expected?

Usage

denovolyzeMultiHits(genes, classes, nsamples, nperms = 100,
includeGenes = "all", includeClasses = c("syn", "mis", "lof", "prot", "all"), nVars = "actual", geneId = "geneName", probTable = NULL,
misD = NULL, signifP = 3, roundExpected = 1)

Arguments

genes A vector of genes containing de novo variants.

classes A vector of classes of de novo variants. Standard supported classes are "syn" (synonymous), "mis" (missense), "non" (nonsense), "splice" (splice), "frameshift" (frameshift) and "lof" (loss of function = non + splice + frameshift). Additional classes that are supported by the code, but are not included in the built-in probability tables, are "stoploss", "startloss", "misD" (damaging missense). These labels may be used for user-supplied probability tables. If "misD" is present, then "mis" (in the input) implies non-damaging missense.

nsamples Number of individuals considered in de novo analysis.

nperms Number of permutations

includeGenes Genes to include in analysis. "all" or a vector of gene names.

includeClasses Determines which variant classes are tabulated in output. In addition to the input classes, summaries can be produced for "prot" (protein-altering = mis + lof), "all", and "protD" (protein damaging = misD + lof, only available if misD included in user-specified probability table). If "misD" is present, then "mis" will return statistics for all missense. Non-damaging missense are not analysed separately.

nVars Select whether expected number of multihits is determined by "expected" total number of variants, or "actual" total. Actual (default) is more conservative.

geneId Gene identifier used. One of "hgncID", "hgncSymbol", "enstID", "ensgID" or "geneName" (default, equals ensembl "external_gene_name")

probTable Probability table. A user-defined table of probabilities can be provided here, to replace the probability table included in the package.
misD  If the user-specified probability table contains probabilities for a sub-category of missense variants (e.g., predicted to be damaging by an in silico algorithm), this column should be called misD, or the alternative name should be specified here.

signifP  Number of significant figures used to round p-values in output.

roundExpected  Number of decimal places used to round expected burdens in output.

Details

See vignette (denovostats_intro) for more information.

Value

Returns a data.frame

Examples

denovolyzeMultiHits(genes=autismDeNovos$gene,
                    classes=autismDeNovos$class,
                    nsamples=1078)

--

denovolyzeR  A package for the analysis of de novo sequencing variants

Description

A package for the analysis of de novo sequencing variants

Author(s)

James Ware <j.ware@imperial.ac.uk>

References

http://github.com/jamesware/denovolyzeR
Description

837 genes found to interact with the fragile X mental retardation protein (FMRP)

Format

A vector of gene symbols

References

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4222185/
http://dx.doi.org/10.1016/j.cell.2011.06.013

parseInput  Checks input for errors

Description

An internal function to check inputs

Usage

parseInput(genes = genes, classes = classes, nsamples = nsamples,
groupBy = groupBy, includeGenes = includeGenes,
includeClasses = includeClasses, geneId = geneId, signifP = signifP,
roundExpected = roundExpected, probTable = NULL)

Arguments

genes A vector of genes containing de novo variants.

classes A vector of classes of de novo variants. Standard supported classes are "syn" (synonymous), "mis" (missense), "non" (nonsense), "splice" (splice), "frameshift" (frameshift) and "lof" (loss of function = non + splice + frameshift). Additional classes that are supported by the code, but are not included in the built-in probability tables, are "stoploss", "startloss", "misD" (damaging missense). These labels may be used for user-supplied probability tables. If "misD" is present, then "mis" (in the input) implies non-damaging missense.

nsamples Number of individuals considered in de novo analysis.

groupBy Results can be tabulated by "gene", or by variant "class"

includeGenes Genes to include in analysis. "all" or a vector of gene names.
includeClasses
Determines which variant classes are tabulated in output. In addition to the input classes, summaries can be produced for "prot" (protein-altering = mis + lof), "all", and "protD" (protein damaging = misD + lof, only available if misD included in user-specified probability table). If "misD" is present, then "mis" will return statistics for all missense. Non-damaging missense are not analysed separately.

geneId
Gene identifier used. One of "hgncID", "hgncSymbol", "enstID", "ensgID" or "geneName" (default, equals ensembl "external_gene_name")

signifP
Number of significant figures used to round p-values in output.

roundExpected
Number of decimal places used to round expected burdens in output.

probTable
Probability table. A user-defined table of probabilities can be provided here, to replace the probability table included in the package.

Value
warning or error if any invalid input, else assigns variables back to parent function

Description
An internal function called by denovolyzeMultiHits

Usage
PermuteMultiHits(x, y, nperms = 100, class = "lof", geneId = "geneName", includeGenes = "all", probTable = pDNM)

Arguments

x
Total number of de novo variants observed in dataset

y
Number of genes with >1 de novo variant (of class "class") in the population

nperms
Number permutations

class
In c("lof","mis","syn","prot")

geneId
Gene identifier used. One of "hgncID", "hgncSymbol", "enstID", "ensgID" or "geneName" (default, equals ensembl "external_gene_name")

includeGenes
Genes to include in analysis. "all" or a vector of gene names.

probTable
Probability table. A user-defined table of probabilities can be provided here, to replace the probability table included in the package.

Value
Returns a named vector of 5 values
viewProbabilityTable

See Also
denovolyzeMultiHits

viewProbabilityTable Displays underlying de novo probability tables

Description
Tabulates probability of \textit{de novo} variant for each protein-coding variant class, for each gene. Values are probability of a \textit{de novo} variant per chromosome per generation. i.e. expected number of \textit{de novos} for a given gene/class = $p \times 2 \times nsamples$.

Usage
viewProbabilityTable(format = "wide")

Arguments
format option to display table in wide format (default; one line per gene), or long format
Index

* datasets
  autismDeNovos, 2
  fmrpGenes, 7
* keywords
  denovolyzeMultiHits, 5

autismDeNovos, 2
denovolyze, 2
denovolyzeFromClass (denovolyze), 2
denovolyzeByGene (denovolyze), 2
denovolyzeMultiHits, 5, 9
denovolyzeR, 6
denovolyzeR-package (denovolyzeR), 6

fmrpGenes, 7
parseInput, 7
PermuteMultiHits, 8

viewProbabilityTable, 9