Package 'IMGTStatClonotype'

March 3, 2017

Type Package

Title Pairwise evaluation and visualization of IMGT clonotype (AA) diversity and expression from IMGT/HighV-QUEST output

Version 1.0.2

Author Safa Aouinti

Maintainer Safa Aouinti <safa.aouinti@igh.cnrs.fr>

Depends R(>= 3.3.0)

Imports reshape2 (>= 1.4.1), data.table (>= 1.9.6), multtest (>= 2.24.0), ggplot2 (>= 2.1.0), gridExtra (>= 2.0.0), DT (>= 0.1), shiny (>= 0.13.2), shinyjs(>= 0.4.0), colourpicker (>= 0.3), plotly(>= 3.4.13), d3heatmap(>= 0.6.1)

Description 'IMGTStatClonotype' developed by LIGM (Montpellier University, CNRS) and part of IMGT®, the international ImMunoGeneTics information system® (http://www.imgt.org) is an R package for statistical analysis of sets from IMGT/HighV-QUEST output. IMGT/HighV-QUEST is the IMGT web portal for next generation sequencing (NGS) analysis of immunoglobulins (IG) or antibodies and T cell receptor (TR) sequences. It provides a standardized and high quality output including the characterization of the IMGT clonotype (AA) diversity and expression and their comparison in up to one million sequences. 'IMGTStatClonotype' includes a generic and standardized procedure for evaluating the statistical significance of pairwise comparison between differences in proportions of the IMGT clonotypes (AA) diversity and expression per gene of a given IG or TR variable (V), diversity (D) or joining (J) group. The package 'IMGTStatClonotype' incorporates a user-friendly web interface, allowing use of the IMGT/StatClonotype tool, in users' own browser.

URL http://www.imgt.org/StatClonotype/

License LGPL

LazyData TRUE

Encoding UTF-8

RoxygenNote 5.0.1

NeedsCompilation no

R topics documented:

`clonNumDiv` *Numbers of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output for clonotype diversity*

Description

This function allows the creation of a matrix containing the number of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output with an IMGT gene of a given IMGT V, D or J group (IMGT clonotypes (AA) diversity).

Usage

```
clonNumDiv(data1, data2)
```

Arguments

`data1`	the first set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function `clonRem`
`data2`	the second set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function `clonRem`

Value

Matrix with IMGT gene names in rows and numbers of IMGT clonotypes (AA) in columns.

Examples

```
Ndiv<-clonNumDiv(MID1,MID2)
```

clonNumExp *Numbers of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output for clonotype expression*

Description

This function allows the creation of a matrix containing the number of sequences assigned to IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output with an IMGT gene of a given IMGT V, D or J group (IMGT clonotypes (AA) expression).

Usage

```
clonNumExp(data1, data2)
```

Arguments

data1	the first set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem
data2	the second set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem

Value

Matrix with IMGT gene names in rows and numbers of IMGT clonotypes (AA) in columns.

Examples

```
Nexp<-clonNumExp(MID1,MID2)
```

clonRem *CDR3-IMGT outlier length*

Description

This function removes IMGT clonoypes (AA) with CDR3-IMGT outlier lengths depending on the studied species (for *Homo sapiens* by default >=45 and <=4).

Usage

```
clonRem(set, min = 4, max = 45)
```

Arguments

set	the set from the IMGT/HighV-QUEST output to be compared
min	the lower level of CDR3-IMGT length
max	the upper level of CDR3-IMGT length

Value

This function returns sets from the IMGT/HighV-QUEST output without IMGT clonotypes (AA) having CDR3-IMGT outlier lengths. A matrix of 25 columns:

`cdr3aa`	CDR3-IMGT sequence (AA)
`expid`	Experimental ID
`clonoIndex`	Representative sequence index
`onecopy`	Nb of '1 copy'
`morethanone`	Nb of 'More than one'
`total`	Total nb of '1 copy' and 'More than one'
`indexes`	'1 copy' Indexes
`vgene`	V-gene
`vallele`	V-allele
`dgene`	D-gene
`dallele`	D-allele
`jgene`	J-gene
`jallele`	J-allele
`cdr1`	CDR1-IMGT
`cdr2`	CDR2-IMGT
`gcdr1`	CDR1-IMGT gapped sequence (AA)
`gcdr2`	CDR2-IMGT gapped sequence (AA)
`pid`	% identity with the closest germline IMGT V gene and allele
`length`	Sequence length
`c104`	C104 (1st-CYS)
`f118`	F118 or W118 (J-PHE or J-TRP)
`anchors`	Anchors (C104, F118 or W118)
`seqid`	Sequence ID
`functionality`	Functionality
`sequenceFileNumber`	Sequence file number
`sequenceClonoFileNumber`	Sequence clonotype file number

Examples

```
data(MID1)
data(MID2)
MID1<-clonRem(MID1)
MID2<-clonRem(MID2)
```

diffpropGph *Differences in proportions graph*

Description

This function draws the graph of differences in proportions of IMGT clonotypes (AA) (or sequences assigned to IMGT clonotypes (AA)) with significance and confidence interval (CI) bars.

Usage

```
diffpropGph(data, ...)
```

Arguments

data	the data isssued from the fuction sigrepDiv for clonotype diversity and from the function sigrepExp for clonotype expression
...	optional parameters

Value

Graph of differences in proportions of IMGT clonotypes (AA) (or sequences assigned to IMGT clonotypes (AA)) in the two compared sets from the IMGT/HighV-QUEST output with significance and confidence interval (CI) bars.

Examples

```
diffpropGph(div)$Vgenes
diffpropGph(div)$Jgenes
diffpropGph(div)$Dgenes
diffpropGph(exp)$Vgenes
diffpropGph(exp)$Jgenes
diffpropGph(exp)$Dgenes
```

GeneList *List of IMGT genes (data)*

Description

The data contains the list of IMGT gene names ordered by their positions in the locus. This order is considered in the results visualization.

Usage

```
data("GeneList")
```

Format

A data frame with 4031 observations (raws) and 4 variables (columns). Rows represent the list of IMGT gene names (first column) ordered depending on species (second column), IMGT gene names with functionality (3rd column) and the locus (4th column).

Source

http://www.imgt.org/IMGTrepertoire/LocusGenes/

References

Giudicelli V., Chaume D., Lefranc M-P. (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 33(Database issue):D256-61. doi: 10.1093/nar/gki010. PMID: 15608191.

Lefranc M-P. and Lefranc G. (2001) The T cell receptor FactsBook. Academic Press, London, UK (398 pages).

Lefranc M-P. and Lefranc G. (2001) The Immunoglobulin FactsBook. Academic Press, London, UK (458 pages).

Lefranc, M-P. (2014) Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. Front. Immunol. 5:22. doi: 10.3389/fimmu.2014.00022. PMID: 24600447.

---

launch *IMGT/StatClonotype launch*

Description

This function launches the application IMGT/StatClonotype pairwise comparisons through a user-friendly interface in the web browser.

Usage

```
launch()
```

Value

An interaction tool to visualize pairwise comparison between sets from IMGT/HighV-QUEST output. If no errors occurred this function returns (NULL) else it returns error(s) message(s) shown in the R console.

Examples

```
launch()
```

`meltgene` *List of IMGT V and J genes*

Description

This function allows the creation of a new matrix grouping the columns vgene, dgene and jgene in one column with their identifier (ID).

Usage

```
meltgene(data, ..., variable.name = "Gene_Type", value.name = "Gene_Name")
```

Arguments

`data`	a set from the IMGT/HighV-QUEST output
`...`	further arguments passed to or from other methods
`variable.name`	name of variable used to store measured variable names, by default "Gene_Type"
`value.name`	name of variable used to store values, by default "Gene_Name"

Details

This function must be applied for the two sets from the IMGT/HighV-QUEST output to be compared. It is based on the function melt of the package reshape2.

Value

New matrix with 4 columns: "expid" (Experimental ID), "total" (Total nb of ‘1 copy‘ and ‘More than one‘), "Gene_Type" (V D or J genes), "Gene_Name" (Gene names).

References

Wickham H. (2007) Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. http://www.jstatsoft.org/v21/i12/

Examples

```
set1<-meltgene(MID1)
set2<-meltgene(MID2)
```

MID1 *CD4- population at Pre-vaccination IMGT/HighV-QUEST output*

Description

The data of one set from IMGT/HighV-QUEST output used here as an example.

Usage

```
data("MID1")
```

Format

A data frame with 2348 observations on the following 26 variables:

cdr3aa: CDR3-IMGT sequence (AA)

expid: Experimental ID

clonoIndex: Representative sequence index

onecopy: Nb of '1 copy'

morethanone: Nb of 'More than one'

total: Total nb of '1 copy' and 'More than one'

indexes: '1 copy' Indexes

vgene: V-gene

vallele: V-allele

dgene: D-gene

dallele: D-allele

jgene: J-gene

jallele: J-allele

cdr1: CDR1-IMGT

cdr2: CDR2-IMGT

gcdr1: CDR1-IMGT gapped sequence (AA)

gcdr2: CDR2-IMGT gapped sequence (AA)

pid: % identity with the closest germline IMGT V gene and allele

length: Sequence length

c104: C104 (1st-CYS)

f118: F118 or W118 (J-PHE or J-TRP)

anchors: Anchors (C104, F118 or W118)

seqid: Sequence ID

functionality: Functionality

sequenceFileNumber: Sequence file number

sequenceClonoFileNumber: Sequence clonotype file number

Source

Sequencing data used for this example is available in the NCBI Sequence Read Archive under the accession code SRX326382. The description of this data is available in Li S. et al. (2013).

References

Alamyar E., Giudicelli V., Li S., Duroux P., Lefranc M-P. (2012) IMGT/High V-QUEST: The IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Research. 8:1:2. doi: 10.4172/1745-7580.1000056. PMID: 22647994.

Alamyar E., Duroux P., Lefranc M.-P. and Giudicelli V. (2012) IMGT tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT HighV-QUEST for NGS. Methods Mol. Biol. 882:569-604. doi: 10.1007/978-1-61779-842-9_32. PMID: 22665256.

Li S., Lefranc M.-P., Miles J.J., Alamyar E., Giudicelli V., Duroux P., et al. (2013) IMGT/HighV-QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. Nature Communications. 4:2333. doi: 10.1038/ncomms3333. PMID: 23995877.

Examples

```
data(MID1)
str(MID1)
```

`MID2` *CD4+ population at Pre-vaccination IMGT/HighV-QUEST output*

Description

The data of one set from IMGT/HighV-QUEST output used here as an example.

Usage

```
data("MID2")
```

Format

A data frame with 1882 observations on the following 26 variables:

cdr3aa: CDR3-IMGT sequence (AA)

expid: Experimental ID

clonoIndex: Representative sequence index

onecopy: Nb of '1 copy'

morethanone: Nb of 'More than one'

total: Total nb of '1 copy' and 'More than one'

indexes: '1 copy' Indexes

vgene: V-gene

vallele: V-allele

dgene: D-gene

dallele: D-allele

jgene: J-gene

jallele: J-allele

cdr1: CDR1-IMGT

cdr2: CDR2-IMGT

gcdr1: CDR1-IMGT gapped sequence (AA)

gcdr2: CDR2-IMGT gapped sequence (AA)

pid: % identity with the closest germline IMGT V gene and allele

length: Sequence length

c104: C104 (1st-CYS)

f118: F118 or W118 (J-PHE or J-TRP)

anchors: Anchors (C104, F118 or W118)

seqid: Sequence ID

functionality: Functionality

sequenceFileNumber: Sequence file number

sequenceClonoFileNumber: Sequence clonotype file number

Source

Sequencing data used for this example is available in the NCBI Sequence Read Archive under the accession code SRX326382. The description of this data is available in Li S. et al. (2013)

References

Alamyar E., Giudicelli V., Li S., Duroux P., Lefranc M-P. (2012) IMGT/High V-QUEST: The IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Research. 8:1:2. doi: 10.4172/1745-7580.1000056. PMID: 22647994.

Alamyar E., Duroux P., Lefranc M.-P. and Giudicelli V. (2012) IMGT tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT HighV-QUEST for NGS. Methods Mol. Biol. 882:569-604. doi: 10.1007/978-1-61779-842-9_32. PMID: 22665256.

Li S., Lefranc M.-P., Miles J.J., Alamyar E., Giudicelli V., Duroux P., et al. (2013) IMGT/HighV-QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. Nature Communications. 4:2333. doi: 10.1038/ncomms3333. PMID: 23995877.

Examples

```
data(MID2)
str(MID2)
```

`multprocPlot` *Multiples testing procedures displays*

Description

This function draws graphs from multiple testing results. It displays the number of rejected hypotheses plotted against the Type I error rate for each of the procedures and the ordered adjusted *p*-values plotted for each of the procedures obtained by using the functions mt.plot of the package multtest (plottype: "rvsa" and "pvst" respectively).

Usage

```
multprocPlot(data, ...)
```

Arguments

`data`	the data isssued from the function `sigrepDiv` or `sigrepExp`
`...`	optional parameters

Value

Graphs from multiple testing results.

Source

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics R. Genome Biology, Vol. 5, R80, https://www.bioconductor.org/

References

Pollard K.S., Dudoit S., van der Laan M.J. (2005). Multiple testing procedures: R multtest package and applications to genomics. *In*: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman R., Carey V.J., Huber W., Irizarry R.A., Dudoit S. (Eds) Springer (Statistics for Biology and Health Series), pp. 251-272.

Examples

```
dev.new(width=6.7, height=3.14)
multprocPlot(div)
multprocPlot(exp)
```

normjuxBars *Normalized bar graph of the proportions*

Description

This function draws the normalized bar graph of the proportions of IMGT clonotypes (AA) (or sequences assigned to IMGT clonotypes (AA)).

Usage

```
normjuxBars(data, ...)
```

Arguments

data	the data isssued from the function sigrepDiv or sigrepExp
...	optional parameters

Value

Normalized juxtaposed bar graphs of the proportions of IMGT clonotypes (AA) (or sequences assigned to IMGT clonotypes (AA)) in two compared sets from IMGT/HighV-QUEST output.

Examples

```
normjuxBars(div)$BarGphV
normjuxBars(div)$BarGphD
normjuxBars(div)$BarGphJ
normjuxBars(exp)$BarGphV
normjuxBars(exp)$BarGphD
normjuxBars(exp)$BarGphJ
```

preabsDiv *Gene presence/absence in the IMGT clonotypes (AA) for clonotype diversity*

Description

This function allows the creation of a new boolean matrix indicating the presence (coded by 1) or the absence (coded by 0) of genes in the IMGT clonotypes (AA) for clonotype diversity.

Usage

```
preabsDiv(datag, data)
```

Arguments

datag	IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem
data	issued from the function meltgene

Value

Boolean matrix with IMGT clonotypes (AA) in rows and gene names in columns.

Examples

```
b1<-preabsDiv(MID1,set1)
b2<-preabsDiv(MID2,set2)
```

`preabsExp` *Gene presence/absence in the IMGT clonotypes (AA) for clonotype expression*

Description

This function allows the creation of a new matrix indicating the presence or the absence of genes in the sequences assigned to IMGT clonotypes (AA) for clonotype expression.

Usage

```
preabsExp(datag, data)
```

Arguments

`datag`	IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function `clonRem`
`data`	issued from the function `meltgene`

Value

Matrix with sequences assigned to IMGT clonotypes (AA) in rows and gene names in columns.

Examples

```
e1<-preabsExp(MID1,set1)
e2<-preabsExp(MID2,set2)
```

S1 *Memory IgD+ B cell IMGT/HighV-QUEST output*

Description

The data of one set from IMGT/HighV-QUEST output used here as an example.

Usage

```
data("S1")
```

Format

A data frame with 27731 observations on the following 26 variables:

cdr3aa: CDR3-IMGT sequence (AA)

expid: Experimental ID

clonoIndex: Representative sequence index

onecopy: Nb of '1 copy'

morethanone: Nb of 'More than one'

total: Total nb of '1 copy' and 'More than one'

indexes: '1 copy' Indexes

vgene: V-gene

vallele: V-allele

dgene: D-gene

dallele: D-allele

jgene: J-gene

jallele: J-allele

cdr1: CDR1-IMGT

cdr2: CDR2-IMGT

gcdr1: CDR1-IMGT gapped sequence (AA)

gcdr2: CDR2-IMGT gapped sequence (AA)

pid: % identity with the closest germline IMGT V gene and allele

length: Sequence length

c104: C104 (1st-CYS)

f118: F118 or W118 (J-PHE or J-TRP)

anchors: Anchors (C104, F118 or W118)

seqid: Sequence ID

functionality: Functionality

sequenceFileNumber: Sequence file number

sequenceClonoFileNumber: Sequence clonotype file number

Source

Sequencing data used for this example is available in the NCBI Sequence Read Archive under the accession code SRX470417. The description of this data is available in Mroczek E.S. et al. (2014).

References

Mroczek E.S., Ippolito G.C., Rogosch T., Hoi K.H., Hwangpo T.A., Brand M.G. et al. (2014) Differences in the composition of the human antibody repertoire by B cell subsets in the blood. Front Immunol. 5:96. doi: 10.3389/fimmu.2014.00096. PMID: 24678310.

Examples

```
data(S1)
str(S1)
```

S2 *Memory IgD- B cell IMGT/HighV-QUEST output*

Description

The data of one set from IMGT/HighV-QUEST output used here as an example.

Usage

```
data("S2")
```

Format

A data frame with 17308 observations on the following 26 variables:

cdr3aa: CDR3-IMGT sequence (AA)

expid: Experimental ID

clonoIndex: Representative sequence index

onecopy: Nb of '1 copy'

morethanone: Nb of 'More than one'

total: Total nb of '1 copy' and 'More than one'

indexes: '1 copy' Indexes

vgene: V-gene

vallele: V-allele

dgene: D-gene

dallele: D-allele

jgene: J-gene

jallele: J-allele

cdr1: CDR1-IMGT

cdr2: CDR2-IMGT

gcdr1: CDR1-IMGT gapped sequence (AA)

gcdr2: CDR2-IMGT gapped sequence (AA)

pid: % identity with the closest germline IMGT V gene and allele

length: Sequence length

c104: C104 (1st-CYS)

f118: F118 or W118 (J-PHE or J-TRP)

anchors: Anchors (C104, F118 or W118)

seqid: Sequence ID

functionality: Functionality

sequenceFileNumber: Sequence file number

sequenceClonoFileNumber: Sequence clonotype file number

Source

Sequencing data used for this example is available in the NCBI Sequence Read Archive under the accession code SRX470416. The description of this data is available in Mroczek E.S. et al. (2014).

References

Mroczek E.S., Ippolito G.C., Rogosch T., Hoi K.H., Hwangpo T.A., Brand M.G. et al. (2014) Differences in the composition of the human antibody repertoire by B cell subsets in the blood. Front Immunol. 5:96. doi: 10.3389/fimmu.2014.00096. PMID: 24678310.

Examples

```
data(S2)
str(S2)
```

`sigrepDiv` *Significance of the difference in proportions with 95% confidence interval (CI) for IMGT clonotype (AA) diversity between two sets from IMGT/HighV-QUEST output*

Description

This function tests the significance of the difference in proportions with 95% confidence interval (CI) for IMGT clonotype (AA) diversity.

Usage

```
sigrepDiv(Data, data1, data2)
```

Arguments

Data	the matrix issued from the function clonNumDiv containing the number of IMGT clonotypes (AA)
data1	the first set from IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem
data2	the second set from IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem

Value

A matrix of 21 columns:

Gene_Name	The list of IMGT gene names found in the two compared sets from the IMGT/HighV-QUEST output
Gene_Type	The type of genes (V, D or J)
Nb_IMGT_clonotype_AA.set1	The nb of IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Proportion.set1	The proportion of IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Normalized_proportion.set1	The normalized proportion for 10000 IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Nb_IMGT_clonotype_AA.set2	The nb of IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Proportion.set2	The proportion of IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Normalized_proportion.set2	The normalized proportion for 10000 IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Difference_proportion	The difference in proportions of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output with the corresponding gene indicated in the first column "Gene_Name"
z	The z-score values to determine the significance of the difference between two proportions
Lower_bound_IC_diff_prop	The lower bound of the 95% confidence interval (CI) for the difference in proportions of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output
Upper_bound_IC_diff_prop	The upper bound of the 95% confidence interval (CI) for the difference in proportions of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output

rawp	The *p*-values obtained from z-scores to evaluate the significance of difference in proportions of IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output
Bonferroni	The adjusted *p*-values issued from the Bonferroni multiple testing procedure
Holm	The adjusted *p*-values issued from the Holm multiple testing procedure
Hochberg	The adjusted *p*-values issued from the Hochberg multiple testing procedure
SidakSS	The adjusted *p*-values issued from the Sidak single-step (SS) multiple testing procedure
SidakSD	The adjusted *p*-values issued from the Sidak single-down (SD) multiple testing procedure
BH	The adjusted *p*-values issued from the Benjamini & Hochberg (BH) multiple testing procedure
BY	The adjusted *p*-values issued from the Benjamini & Yekutieli (BY) multiple testing procedure
Test_interpretation	The test interpretation: before adjustment of *p*-values (rawp) non-significant and after adjustment by the multiple testing procedure: significant differences in proportions validated by the seven procedures (All_p), by two or more procedures (Min_2p) and only by BH (Only_BH)

Source

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics R. Genome Biology, Vol. 5, R80, https://www.bioconductor.org/

References

Pollard K.S., Dudoit S., van der Laan M.J. (2005). Multiple testing procedures: R multtest package and applications to genomics. *In*: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman R., Carey V.J., Huber W., Irizarry R.A., Dudoit S. (Eds) Springer (Statistics for Biology and Health Series), pp. 251-272.

Examples

```
div<-sigrepDiv(Ndiv,MID1,MID2)
```

---

sigrepExp *Significance of the difference in proportions with 95% confidence interval (CI) for IMGT clonotype (AA) expression between two sets from IMGT/HighV-QUEST output*

Description

This function tests the significance of the difference in proportions with 95% confidence interval (CI) for IMGT clonotype (AA) expression.

Usage

```
sigrepExp(Data, data1, data2)
```

Arguments

Data	the matrix issued from the function clonNumExp containing the number of IMGT clonotypes (AA)
data1	the first set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem
data2	the second set from the IMGT/HighV-QUEST output without CDR3-IMGT outlier lengths issued from the function clonRem

Value

A matrix of 21 columns:

Gene_Name	The list of IMGT gene names found in the two compared sets from the IMGT/HighV-QUEST output
Gene_Type	The type of genes (V, D or J)
Nb_IMGT_clonotype_AA.set1	The nb of sequences assigned to IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Proportion.set1	The proportion of sequences assigned to IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Normalized_proportion.set1	The normalized proportion for 10000 sequences assigned to IMGT clonotypes (AA) in the first IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Nb_IMGT_clonotype_AA.set2	The nb of sequences assigned to IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Proportion.set2	The proportion of sequences assigned to IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Normalized_proportion.set2	The normalized proportion for 10000 sequences assigned to IMGT clonotypes (AA) in the second IMGT/HighV-QUEST output (set) with the corresponding gene indicated in the first column "Gene_Name"
Difference_proportion	The difference in proportions of sequences assigned to IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output with the corresponding gene indicated in the first column "Gene_Name"
z	The z-score values to determine the significance of the difference between two proportions

`Lower_bound_IC_diff_prop`
: The lower bound of the 95% confidence interval (CI) for the difference in proportions of sequences assigned to IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output

`Upper_bound_IC_diff_prop`
: The upper bound of the 95% confidence interval (CI) for the difference in proportions of sequences assigned to IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output

`rawp`
: The p-values obtained from z-scores to evaluate the significance of difference in proportions of sequences assigned to IMGT clonotypes (AA) in the two compared sets from the IMGT/HighV-QUEST output

`Bonferroni`
: The adjusted p-values issued from the Bonferroni multiple testing procedure

`Holm`
: The adjusted p-values issued from the Holm multiple testing procedure

`Hochberg`
: The adjusted p-values issued from the Hochberg multiple testing procedure

`SidakSS`
: The adjusted p-values issued from the Sidak single-step (SS) multiple testing procedure

`SidakSD`
: The adjusted p-values issued from the Sidak single-down (SD) multiple testing procedure

`BH`
: The adjusted p-values issued from the Benjamini & Hochberg (BH) multiple testing procedure

`BY`
: The adjusted p-values issued from the Benjamini & Yekutieli (BY) multiple testing procedure

`Test_interpretation`
: The test interpretation: before adjustment of p-values (rawp) non-significant and after adjustment by the multiple testing procedure: significant differences in proportions validated by the seven procedures (All_p), by two or more procedures (Min_2p) and only by BH (Only_BH)

Source

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., Dudoit S., et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics R. Genome Biology, Vol. 5, R80, https://www.bioconductor.org/

References

Pollard K.S., Dudoit S., van der Laan M.J. (2005). Multiple testing procedures: R multtest package and applications to genomics. *In*: Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman R., Carey V.J., Huber W., Irizarry R.A., Dudoit S. (Eds) Springer (Statistics for Biology and Health Series), pp. 251-272.

Examples

```
exp<-sigrepExp(Nexp,MID1,MID2)
```

Index