Package 'BCellMA'

March 8, 2017

Type Package

Title B Cell Receptor Somatic Hyper Mutation Analysis

Version 0.3.4

Date 2016-10-27

Author Olena Ohlei <olena.ohlei@student.uni-luebeck.de>

Maintainer Olena Ohlei <olena.ohlei@student.uni-luebeck.de>

Description Includes a set of functions to analyze for instance nucleotide frequencies as well as transition and transversion. Can reconstruct germline sequences based on the international ImMunoGeneTics information system (IMGT/HighV-QUEST) outputs, calculate and plot the difference (%) of nucleotides at 6 positions around a mutation to identify and characterize hotspot motifs as well as calculate and plot average mutation frequencies of nucleotide mutations resulting in amino acid substitution.

License GPL (>= 2)

Depends R (>= 3.2.5), ggplot2, reshape2, grid

Suggests graphics, utils, knitr, rmarkdown

LazyData true

URL https://www.r-project.org

RoxygenNote 6.0.1

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2017-03-08 10:58:44

R topics documented:

`aa_cdr3_dist` *Average frequency of nucleotide mutations resulting in amino acid substitution in CDR3*

Description

Calculate average frequency of nucleotide mutations resulting in amino acid substitution in CDR3.

Usage

```
aa_cdr3_dist(data)
```

Arguments

`data`	A column CDR3_IMGT from IMGT/HighV-Quest table 7 called "7_V-REGION-mutation-and-AA-change-table.txt"

Value

Output is the data matrix of average frequency of nucleotide mutations resulting in amino acid substitution.

References

Zuckerman NS., Hazanov H., Barak M., Edelman H., Hess S., Shcolnik H., Dunn-Walters D.,and Mehr R. Somatic hypermutation and antigen-driven selection of B cells are altered inautoimmune diseases.J Autoimmun, 35(4):325 - 335, 2010. doi: 10.1016/j.jaut.2010.07.004.

Examples

```
data(IMGTtab7)
cdr3_matrix<-aa_cdr3_dist(data=IMGTtab7$CDR3_IMGT)
```

aa_dist — *Average frequency of nucleotide mutations resulting in amino acid substitution in FRs and CDRs*

Description

Calculate average frequency of nucleotide mutations resulting in amino acid substitution in CDRs and/or in FRs.

Usage

```
aa_dist(data)
```

Arguments

data	CDRs and FRs colums from IMGT/HighV-Quest table 7 coled "7_V-REGION-mutation-and-AA-change-table.txt"

Value

Output is the data matrix of average mutation frequency of nucleotide mutations resulting in amino acid substitution.

References

Zuckerman NS., Hazanov H., Barak M., Edelman H., Hess S., Shcolnik H., Dunn-Walters D.,and Mehr R. Somatic hypermutation and antigen-driven selection of B cells are altered inautoimmune diseases.J Autoimmun, 35(4):325 - 335, 2010. doi: 10.1016/j.jaut.2010.07.004.

Examples

```
data(IMGTtab7)
Regions<-cbind(IMGTtab7$FR1_IMGT,IMGTtab7$CDR1_IMGT, IMGTtab7$FR2_IMGT)
Regions_matrix<-aa_dist(data=Regions)
```

aa_plot — *Plot of average frequency of nucleotide mutations resulting in amino acid substitution*

Description

The average frequency of nucleotide mutations resulting in amino acid substitution can be plotted as an aa_plot().

Usage

```
aa_plot(data, text, legend.position, characteristics)
```

Arguments

data	Results from the function aa_dist or aa_cdr3_dist.
text	The legend of the plot.
legend.position	The position of the legend. It can be "none","left", "right", "bottom", "top".
characteristics	is a data set "data(Klassen)" in this package

Value

Output is the plot of average frequency of nucleotide mutations resulting in amino acid substitution. The classes are divided according to IMGT.

References

Pommie C., Levadoux S., Sabatier R., Lefranc G., and Lefranc MP. IMGT standardizedcriteria for statistical analysis of immunoglobulin V-REGION amino acid properties.J MolRecognit, 17(1):17-32, 2004.

Examples

```
data(IMGTtab7)
Regions<-cbind(IMGTtab7$FR1_IMGT,IMGTtab7$CDR1_IMGT)
allRegions_matrix<-aa_dist(Regions)
data(Klassen)
aa_plot(allRegions_matrix, "Amino acid Distribution", "right", Klassen)
```

---

gene_comb_funk *gene/gene ratios of two gene families*

Description

The calculation was obtained according bcRep package in R language. Every combination of two families was summarized and divided by the sum of all combinations across each group.

Usage

```
gene_comb_funk(family1, family2)
```

Arguments

family1	"V_GENE_and_allele" colums from IMGT/HighV-Quest table 1 coled "1_Summary.txt"
family2	"J_GENE_and_allele" or "D_GENE_and_allele" colums from IMGT/HighV-Quest table 1 coled "1_Summary.txt"

Value

Output is the data matrix of gene/gene ratios of two gene families.

References

Bischof J. and Ibrahim SM. bcRep: R Package for Comprehensive Analysis of B Cell Receptor Repertoire Data. PLoS One. 11(8):e0161569, 2016. doi: 10.1371/journal.pone.0161569.

Examples

```
data(IMGTtab1)
gane_comb<-gene_comb_funk(family1 = IMGTtab1$V_GENE_and_allele,
                          family2 = IMGTtab1$J_GENE_and_allele)
gane_comb
```

---

gene_comb_plot *Plot of gene/gene ratios of two gene families*

Description

Gene/gene ratios of two gene families can be plotted as an gene_comb_plot().

Usage

```
gene_comb_plot(data, text, legend_position, a, b)
```

Arguments

data	Results from the function gene_comb_funk().
text	The legend of the plot.
legend_position	The position of the legend. It can be "none","left", "right", "bottom", "top".
a	The angle of X-axis legend.
b	The distance from X-axis to legend.

Value

Output is the plot of gene/gene ratios of two gene families.

References

Bischof J. and Ibrahim SM. bcRep: R Package for Comprehensive Analysis of B Cell Receptor Repertoire Data. PLoS One. 11(8):e0161569, 2016. doi: 10.1371/journal.pone.0161569.

Examples

```
data(IMGTtab1)
gane_comb<-gene_comb_funk(family1 = IMGTtab1$V_GENE_and_allele,
                          family2 = IMGTtab1$J_GENE_and_allele)
gene_comb_plot(gane_comb, "Plot of IGHV and IGHJ ratio", legend_position = "right", a = 35, b = 0.5)
```

---

germlineReconstr *Function to reconstruction of germline sequencies based on IMGT outputs.*

Description

Function to reconstruction of germline sequencies based on IMGT outputs.

Usage

```
germlineReconstr(data_tab2, data_tab7)
```

Arguments

data_tab2	Column V_Region in the IMGT table 2
data_tab7	Column V_Region in the IMGT table 7

Value

output is a vector of germline sequencies.

References

Brochet X., Lefranc MP., and Giudicelli V. IMGT/V-QUEST: the highly customized andintegrated system for IG and TR standardized V-Jand V-D-J sequence analysis.NucleicAcids Res., 36(Web Server issue):W503 - W508, 2008. doi: 10.1093/nar/gkn316

Examples

```
data(IMGTtab2)
data(IMGTtab7)
germline<-germlineReconstr(IMGTtab2$V_REGION, IMGTtab7$V_REGION)
germline
```

IMGTtab1 *IMGT Table 1*

Description

Data from the Excel file 1 are named Summary and are included in IMGTtab1

Usage

```
data(IMGTtab1)
```

Format

A data frame with 9 rows and 29 variables

Details

- Sequence_number. number of Sequence
- Sequence_ID. ID frome gene bank
- Functionality. productive/unproductive
- V_GENE_and_allele. germline genes of V Region
- J_GENE_and_allele. germline genes of J Region
- D_GENE_and_allele. germline genes of D Region

IMGTtab2 *IMGT Table 2*

Description

Data from the Excel file 2 are named IMGT-gapped-nt-sequences and are included in IMGTtab2

Usage

```
data(IMGTtab2)
```

Format

A data frame with 9 rows and 18 variables

Details

- Sequence_number. number of Sequence
- Sequence_ID. ID frome gene bank
- Functionality. productive/unproductive
- V_GENE_and_allele. germline genes of V Region
- J_GENE_and_allele. germline genes of J Region
- D_GENE_and_allele. germline genes of D Region
- V_D_J_REGION. nucleotides in VDJ region in mutated sequence
- V_J_REGION. NA
- V_REGION. nucleotides in V region in mutated sequence
- FR1_IMGT. nucleotides in FR1 in mutated sequence
- CDR1_IMGT. nucleotides in CDR1 in mutated sequence
- FR2_IMGT. nucleotides in FR2 in mutated sequence
- CDR2_IMGT. nucleotides in CDR2 in mutated sequence
- FR3_IMGT. nucleotides in FR3 in mutated sequence
- CDR3_IMGT. nucleotides in CDR3 in mutated sequence
- JUNCTION. nucleotides in JUNCTION region in mutated sequence
- J_REGION. nucleotides in J region in mutated sequence
- FR4_IMGT. nucleotides in FR4 in mutated sequence

`IMGTtab7` *IMGT Table 7*

Description

Data from the Excel file 7 are called the V-REGION-mutation-table and are included in IMGTtab7.

Usage

```
data(IMGTtab7)
```

Format

A data frame with 9 rows and 11 variables

Details

- Sequence_number. number of Sequence
- Sequence_ID. ID frome gene bank
- Functionality. productive/unproductive
- V_GENE_and_allele. germline genes of V Region
- V_REGION. nucleotide changes and aminoacid changes in V region
- FR1_IMGT. nucleotide changes and aminoacid changes in FR1
- CDR1_IMGT. nucleotide changes and aminoacid changes in CDR1
- FR2_IMGT. nucleotide changes and aminoacid changes in FR2
- CDR2_IMGT. nucleotide changes and aminoacid changes in CDR2
- FR3_IMGT. nucleotide changes and aminoacid changes in FR3
- CDR3_IMGT. nucleotide changes and aminoacid changes in CDR3

`IMGTtab8` *IMGT Table 8*

Description

Data from the Excel file 8 are called the V-REGION-nt-mutation-statistics and are included in IMGTtab8.

Usage

```
data(IMGTtab8)
```

Format

A data frame with 9 rows and 130 variables

Details

- Sequence_number. number of Sequence
- Sequence_ID. ID frome gene bank
- Functionality. productive/unproductive
- V_GENE_and_allele. germline genes of V Region

Klassen *Klassen*

Description

A dataset for usage of Funktion aa_plot()

Usage

```
data(Klassen)
```

Format

A data frame with 400 rows and 2 variables

Details

- Egenschaften. Physio-chemical properties of aminoacids
- Klassen. Classification in classes according to IMGT/HighV-QUEST classes

lengthCDR3 *Frequency distribution of the CDR3 length*

Description

Calculate length of CDR3.

Usage

```
lengthCDR3(data)
```

Arguments

data	"CDR3_IMGT_length" colums from IMGT/HighV-Quest table 1 coled "1_Summary.txt"

Value

Output Output is the number of amino acids in CDR3.

References

references Bischof J. and Ibrahim SM. bcRep: R Package for Comprehensive Analysis of B Cell Receptor Repertoire Data. PLoS One. 11(8):e0161569, 2016. doi: 10.1371/journal.pone.0161569.

Examples

```
data(IMGTtab1)
lenght_tab<-lengthCDR3(as.numeric(IMGTtab1$CDR3_IMGT_length))
```

`na_funktion` *The function to check for reconsrtruction errors*

Description

The function to check for reconsrtruction errors

Usage

```
na_funktion(data)
```

Arguments

`data`	from function germlineReconstr().

Value

total number of NAs. If the answer is null then there are no errors in the germline reconstruction.

Examples

```
germline<-germlineReconstr(IMGTtab2$V_REGION, IMGTtab7$V_REGION)
na_funktion(germline)
```

`nucleotide_mutation` *The frequencies of nucleotide mutations and the frequencies of transition and transversion*

Description

Calculate nucleotide (A, T, C and G) mutations, transition, transversion and their frequency.

Usage

```
nucleotide_mutation(data)
```

Arguments

`data`	A colums from 11 to 22 in the IMGT Table 8 caled "8_V-REGION-nt-mutation-statistics.txt".

Value

Output is four different values:

vregion_nt_mut number of nucleotide mutations in V region.

transition_rel the frequencies of transition.

transverion_rel the frequencies of transversion.

trans_transv_anzahl number of transition and transversion in V region.

Examples

```
data(IMGTtab8)
bm_proband<- nucleotide_mutation(IMGTtab8[,11:22])
percentlabels<- round(bm_proband$vregion_nt_mut/sum(bm_proband$vregion_nt_mut)*100)
pielabels<- paste(percentlabels, "%", sep="")
pie(bm_proband$vregion_nt_mut, col=c("grey50","black","grey90","white"),
    labels=pielabels, cex=1.5, radius = 0.3)
legend("right", c(" prod. IGHV"), cex=1.5)
legend("topleft", c("A","G","T","C"), cex=2, fill=c("grey50","black",
                                                    "grey90","white"))
```

---

targetingMatrix *Function to the calculation of the number of Nucleotids at three positions before and after a mutation*

Description

Function to the calculation of the number of Nucleotids at three positions before and after a mutation

Usage

```
targetingMatrix(data_tab2, data_tab_germline, data_tab7)
```

Arguments

data_tab2	Column V_Region in th IMGT table 2
data_tab_germline	Output from function germlineReconstr()
data_tab7	Column V_Region in th IMGT table 7

Value

Result is a list of 8 matrices with the numbers of nucleotides at three positions before and after a mutation from A, T, C, and G, as well as in a mutated sequence and in the germline sequence

References

Spencer J. and Dunn-Walters DK. Hypermutation at A-T base pairs: the A nucleotidereplacement spectrum is affected by adjacent nucleotides and there is no reverse comple-mentary of sequences flanking muated A and T nucleotides.J Immunol, 175(8):5170 - 5177,2005.

Zuckerman NS., Hazanov H., Barak M., Edelman H., Hess S., Shcolnik H., Dunn-Walters D.,and Mehr R. Somatic hypermutation and antigen-driven selection of B cells are altered inautoimmune diseases.J Autoimmun, 35(4):325 - 335, 2010. doi: 10.1016/j.jaut.2010.07.004.

Examples

```
data(IMGTtab2)
data(IMGTtab7)
germline<-germlineReconstr(IMGTtab2$V_REGION, IMGTtab7$V_REGION)
data<-targetingMatrix(data_tab2=IMGTtab2, data_tab_germline=germline, data_tab7=IMGTtab7)
data
```

---

targeting_motiv *Function to calculate the difference at three positions before and after a mutation to identify the hotspot motifs*

Description

The change around the mutated nucleotide flanking three bases each is expressed as the difference (

Usage

```
targeting_motiv(data)
```

Arguments

data	Output from the function hotspotmutMat().

Value

Output is a list of four mutation environments A, C, T and G with the difference (

References

Spencer J. and Dunn-Walters DK. Hypermutation at A-T base pairs: the A nucleotidereplacement spectrum is affected by adjacent nucleotides and there is no reverse comple-mentary of sequences flanking muated A and T nucleotides.J Immunol, 175(8):5170 - 5177,2005.

Zuckerman NS., Hazanov H., Barak M., Edelman H., Hess S., Shcolnik H., Dunn-Walters D.,and Mehr R. Somatic hypermutation and antigen-driven selection of B cells are altered inautoimmune diseases.J Autoimmun, 35(4):325 - 335, 2010. doi: 10.1016/j.jaut.2010.07.004.

Examples

```
data(IMGTtab2)
data(IMGTtab7)
germline<-germlineReconstr(IMGTtab2$V_REGION, IMGTtab7$V_REGION)
data<-targetingMatrix(data_tab2=IMGTtab2, data_tab_germline=germline, data_tab7=IMGTtab7)
targeting_motiv_data<-targeting_motiv(data)
targeting_motiv_data
```

---

targeting_motive_plot *Polt of the difference at three positions before and after a mutation to identify the hotspot motifs.*

Description

The the difference (

Usage

```
targeting_motive_plot(pwm, xaxis = TRUE, yaxis = TRUE, xfontsize = 15,
  yfontsize = 15, xlim)
```

Arguments

pwm	Result of the function hotspot().
xaxis	Accept TRUE/FALSE parameter. TRUE draw X-axis.
yaxis	Accept TRUE/FALSE parameter. TRUE draw Y-axis.
xfontsize	The plot size of the X-axis.
yfontsize	The plot size of the Y-axis.
xlim	Limit of the drowed Y-axis.

Value

Output is plot the sequence logos around the mutation.

Note

The Function are based on functions from package segLogo.

References

Spencer J. and Dunn-Walters DK. Hypermutation at A-T base pairs: the A nucleotidereplacement spectrum is affected by adjacent nucleotides and there is no reverse comple-mentary of sequences flanking muated A and T nucleotides.J Immunol, 175(8):5170 - 5177,2005.

Zuckerman NS., Hazanov H., Barak M., Edelman H., Hess S., Shcolnik H., Dunn-Walters D.,and Mehr R. Somatic hypermutation and antigen-driven selection of B cells are altered inautoimmune diseases.J Autoimmun, 35(4):325 - 335, 2010. doi: 10.1016/j.jaut.2010.07.004.

Bembom O. seqLogo: Sequence logos for DNA sequence alignments, Status 10.08.2016. URL-http://www.bioconductor.org/packages/release/bioc/html/seqLogo.html.

Schneider TD. and Stephens RM. Sequence logos: a new way to display consensus sequences.Nucleic Acids Res, 18(20):6097 - 6100, 1990.

Examples

```
data(IMGTtab2)
data(IMGTtab7)
germline<-germlineReconstr(IMGTtab2$V_REGION, IMGTtab7$V_REGION)
data<-targetingMatrix(data_tab2=IMGTtab2, data_tab_germline=germline, data_tab7=IMGTtab7)
targeting_motiv_data<-targeting_motiv(data)
targeting_motive_plot(targeting_motiv_data$A, xfontsize = 15, yfontsize = 15, xlim=60 )
```

Index