Package 'tcrl'

February 26, 2025

Type Package

Title Association Analysis of T-Cell Receptor Repertoire and Clinical Phenotypes

Version 0.1.0

Author Meiling Liu

Maintainer Meiling Liu <mliu@fredhutch.org>

Description The tcrl package offers a suite of functions for analyzing the association between the T-cell receptor (TCR) repertoire and clinical phenotypes. It includes tools for evaluating the diversity of the TCR repertoire and its relationship with clinical outcomes, incorporating methods for both continuous and binary phenotypes. The package is especially useful for researchers studying immune responses and disease associations in the context of TCR diversity.

Depends R (>= 3.5.0), Biostrings, matrixcalc, stats

License LGPL(>=2.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

R topics documented:

`AAfreq` *Compute Amino Acid Frequency in TCR Repertoire*

Description

The `AAfreq` function calculates the frequency of amino acids in the T-cell receptor (TCR) repertoire across multiple subjects. This function is specifically designed for understanding the distribution and abundance of amino acids in the TCR repertoire is essential for evaluating immune diversity and response.

Usage

```
AAfreq(Subject.ID, AAseq, Abundance)
```

Arguments

`Subject.ID`	A vector of subject identifiers. Each element corresponds to an individual subject in the study, ensuring that the data can be organized by subject.
`AAseq`	A vector of amino acid sequences representing the TCR repertoire. The sequences should be aligned with the `Subject.ID`, such that each sequence corresponds to a specific subject.
`Abundance`	A numeric vector of the abundance of each amino acid sequence. The order of `Abundance` must match the order of `Subject.ID` and `AAseq`, ensuring consistency in the data provided.

Details

The `AAfreq` function provides a simple yet efficient way to compute amino acid frequencies within a TCR repertoire. By inputting subject-specific TCR sequence data, users can quickly generate a frequency matrix for further analysis. This function is particularly useful in immunology and bioinformatics studies where researchers aim to characterize the TCR diversity and its potential associations with clinical phenotypes.

T-cell receptors (TCRs) play a critical role in recognizing antigens and mediating immune responses. The diversity of the TCR repertoire, characterized by variations in amino acid sequences, can provide insights into immune function and disease mechanisms. The `AAfreq` function helps quantify this diversity by computing the distribution of amino acids across multiple subjects' repertoires.

Value

The function returns a matrix where each row corresponds to a subject and each column represents an amino acid. The values in the matrix indicate the frequency of each amino acid for the corresponding subject.

`out` A numeric matrix representing the frequency distribution of amino acids per subject. This matrix can be used for downstream analysis, such as identifying patterns of amino acid diversity across the study cohort or correlating amino acid frequencies with clinical outcomes.

References

Liu, M., Goo, J., Liu, Y., Sun, W., Wu, M.C., Hsu, L. and He, Q., 2022. *TCR-L: an analysis tool for evaluating the association between the T-cell receptor repertoire and clinical phenotypes*. BMC Bioinformatics, 23(1), p.152.

Examples

```
# Load example dataset
data("example.data")
attach(example.data)

# Compute amino acid frequency
freq_matrix <- AAfreq(TCRdat[,1], TCRdat[,2], as.numeric(TCRdat[,3]))

# View the result
```

```
print(freq_matrix)

detach(example.data)
```

`example.data` *Simulated TCR Data with Continuous Response and Covariates*

Description

A simulated dataset consisting of T-cell receptor (TCR) information, continuous response variables, and covariates for 20 patients. This dataset is designed to demonstrate the usage of functions in analyzing the association between TCR repertoires and clinical or biological outcomes.

Usage

```
example.data
```

Format

A list containing the following components:

`TCRdat` A data frame with three columns representing TCR information:

- `Subject.ID` A vector of patient identifiers.
- `AAseq` A vector of amino acid sequences for the TCR repertoire.
- `Abundance` A numeric vector indicating the abundance of each amino acid sequence for the corresponding patient.

`Y` A numeric vector of continuous response variables (e.g., clinical or biological measurements) associated with each patient.

`X` A matrix of covariates, where each row corresponds to a patient, and each column represents a different covariate (e.g., age, gender, or other relevant clinical information).

`W` A numeric vector providing Kyte and Doolittle hydrophobicity information for each amino acid sequence.

Details

This simulated dataset is intended to demonstrate how TCR repertoire data, along with patient-specific covariates and response variables, can be used to evaluate associations between T-cell diversity and clinical outcomes. It includes TCR sequences and their corresponding abundance, as well as additional covariates and response data that are typically required in immunological and bioinformatics studies.

seqhom *Compute TCR Repertoire Homology Between Subjects*

Description

The seqhom function calculates the homology between T-cell receptor (TCR) repertoires of different subjects based on amino acid sequence similarities. This function uses specified substitution matrices to evaluate the degree of similarity between sequences, providing a quantitative measure of homology across subjects.

Usage

```
seqhom(Subject.ID, AAseq, Abundance, substitutionMatrix)
```

Arguments

Subject.ID	A vector of subject identifiers. Each entry corresponds to a specific individual in the dataset, aligning with the provided TCR sequences and their abundances.
AAseq	A vector of amino acid sequences representing the TCR repertoire for each subject. The order of sequences must match the order of Subject.ID.
Abundance	A numeric vector indicating the abundance of each TCR sequence. The abundance values should correspond to the same order as Subject.ID and AAseq.
substitutionMatrix	A character string specifying the substitution matrix to use for sequence alignment. Options include: "BLOSUM45", "BLOSUM50", "BLOSUM62", "BLOSUM80", "BLOSUM100", "PAM30", "PAM40", "PAM70", "PAM120", and "PAM250". These matrices provide different scoring systems for determining the similarity between amino acids in sequences.

Details

The seqhom function provides an efficient method for comparing TCR repertoires across subjects by leveraging amino acid substitution matrices. By aligning sequences using substitution scores such as BLOSUM or PAM matrices, researchers can quantify the similarity between different individuals' immune repertoires. This homology measure is essential for studying the diversity and shared features of TCR repertoires in immunological studies. The selection of the substitution matrix can impact the resulting homology scores, as different matrices are optimized for various evolutionary distances and similarity metrics.

Value

The function returns a homology matrix:

S A symmetric matrix where each entry S[i,j] represents the computed homology score between the TCR repertoires of subjects i and j. Higher values indicate greater similarity (homology) between the subjects' TCR repertoires.

Examples

```
# Load the example dataset
data("example.data")
attach(example.data)

# Compute the homology matrix using BLOSUM62
S <- seqhom(TCRdat[,1], TCRdat[,2], as.numeric(TCRdat[,3]), "BLOSUM62")

# View the homology matrix
print(S)

detach(example.data)
```

TCRL *Score Test for TCR Repertoire and Phenotypes*

Description

The TCRL_bin and TCRL_cont functions implement score-based association tests to evaluate the relationship between T-cell receptor (TCR) repertoire features and clinical phenotypes. These tests can be applied to binary and continuous phenotypes, incorporating both fixed and random variant effects.

Usage

```
TCRL_bin(Y, X, fR, W, S)

TCRL_cont(Y, X, fR, W, S)
```

Arguments

Y	A vector representing the outcome variable. Y can be either a continuous variable (for TCRL_cont) or a binary variable (for TCRL_bin).
X	A covariate matrix, where each row corresponds to a subject and each column represents a covariate (e.g., clinical or demographic variables).
fR	A feature matrix where each row represents a subject and each column corresponds to an extracted feature of the TCR repertoire (e.g., amino acid frequencies).
W	A feature annotation matrix where each row represents a TCR feature, and each column corresponds to a variant. The number of rows in W should match the number of columns in fR.
S	A homology matrix representing the correlation between subjects, which can be calculated using the seqhom function. This matrix accounts for subject-specific similarities in the TCR repertoire.

Details

The TCRL_bin function is used for binary outcomes, while TCRL_cont is used for continuous outcomes. Both functions perform score-based tests that incorporate both fixed and random effects, allowing researchers to assess the association between TCR features and clinical phenotypes, such as disease status or quantitative biomarkers.

The functions make use of the homology matrix `S` to account for correlations between subjects' TCR repertoires, enhancing the robustness of the statistical tests.

Value

The function returns a list with the following components:

`fix.effect` P-value for testing the fixed variant effect, which assesses whether the TCR features have a consistent association with the phenotype across all subjects.

`random.effect` P-value for testing the random variant effect, which allows the variant effects to vary across subjects.

`Overall.pval` The overall p-value for testing the association between the TCR repertoire and the phenotype. This p-value is calculated by combining `fix.effect` and `random.effect` using Fisher's procedure.

Examples

```
# Load the example dataset
data("example.data")
attach(example.data)

# Extract TCR repertoire features
fR <- AAfreq(TCRdat[,1], TCRdat[,2], as.numeric(TCRdat[,3]))

# Compute the homology matrix
S <- seqhom(TCRdat[,1], TCRdat[,2], as.numeric(TCRdat[,3]), 'BLOSUM62')

# Perform score test for continuous outcome
TCRL_cont(Y, X, fR, W, S)

detach(example.data)
```

`tcrl` *tcrl: Association Analysis of T-Cell Receptor Repertoire and Clinical Phenotypes*

Description

The `tcrl` package offers a suite of functions for analyzing the associations between T-cell receptor (TCR) repertoire features and clinical phenotypes. This package is specifically designed for immunological and bioinformatics studies, providing tools to explore how the diversity and composition of the TCR repertoire are related to disease states, treatment responses, or other clinical outcomes.

Index