# Package 'TSENAT'

January 31, 2026

**Type** Package

**Title** Tsallis Entropy Analysis Toolbox

**Version** 0.1

**Maintainer** Gallardo Cristóbal <gallardoalba@pm.me>

**Description** Tsallis entropy generalizes Shannon entropy with a parameter q that controls sensitivity to transcript abundance. This package allows the calculation of Tsallis entropy for transcript-level expression data and differential analyses between two sample conditions.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.3.3

**Depends** R (>= 4.1)

**Imports** SummarizedExperiment, methods, stats, magrittr, ggplot2

**Suggests** testthat, knitr, rmarkdown, tidyr, dplyr, mgcv, S4Vectors, cowplot, ggrepel, patchwork

**VignetteBuilder** knitr

**URL** https://github.com/gallardoalba/TSENAT

**BugReports** https://github.com/gallardoalba/TSENAT/issues

**biocViews** Transcriptomics, RNASeq, DifferentialSplicing, AlternativeSplicing, TranscriptomeVariant

**NeedsCompilation** no

**Author** Gallardo Cristobal [aut, cre]

## Contents

---

`calculate_difference` *Calculate splicing diversity changes between two conditions.*

---

## Description

Calculate splicing diversity changes between two conditions.

## Usage

```
calculate_difference(
  x,
  samples,
  control,
  method = "mean",
  test = "wilcoxon",
  randomizations = 100,
  pcorr = "BH",
  assayno = 1,
  verbose = FALSE,
  ...
)
```

## Arguments

| | |
|---|---|
| `x` | A `SummarizedExperiment` with splicing diversity values for each gene in each sample or a `data.frame` with gene names in the first column and splicing diversity values for each sample in additional columns. |
| `samples` | A vector of length one, specifying the column name of the `colData` annotation column from the `SummarizedExperiment` object, that should be used as the category column or a character vector with an equal length to the number of columns in the input dataset, specifying the category of each sample in the case of a `data.frame` input. |
| `control` | Name of the control sample category, defined in the `samples` vector, e.g. `control = 'Normal'` or `control = 'WT'`. |
| `method` | Method to use for calculating the average splicing diversity value in a condition. Can be `'mean'` or `'median'`. |

| | |
|---|---|
| test | Method to use for p-value calculation: use 'wilcoxon' for Wilcoxon rank sum test or 'shuffle' for a label shuffling test. |
| randomizations | Number of random shuffles, used for the label shuffling test (default = 100). |
| pcorr | P-value correction method applied to the Wilcoxon rank sum test or label shuffling test results, as defined in the p.adjust function. |
| assayno | An integer value. In case of multiple assays in a SummarizedExperiment input, the argument specifies the assay number to use for difference calculations. |
| verbose | If TRUE, the function will print additional diagnostic messages. |
| ... | Further arguments to be passed on for other methods. |

## Details

The function calculates diversity changes between two sample conditions. It uses the output of the diversity calculation function, which is a SummarizedExperiment object of splicing diversity values. Additionally, it can use a data.frame as input, where the first column contains gene names, and all additional columns contain splicing diversity values for each sample. A vector of sample conditions also serves as input, used for aggregating the samples by condition.

It calculates the mean or median of the splicing diversity data per sample condition, the difference of these values and the log2 fold change of the two conditions. Furthermore, the user can select a statistical method to calculate the significance of the changes. The p-values and adjusted p-values are calculated using a Wilcoxon sum rank test or label shuffling test.

The function will exclude genes of low sample size from the significance calculation, depending on which statistical test is applied.

## Value

A data.frame with the mean or median values of splicing diversity across sample categories and all samples, log2(fold change) of the two different conditions, raw and corrected p-values.

## Examples

```
# data.frame with splicing diversity values
x <- data.frame(Genes = letters[seq_len(10)], matrix(runif(80), ncol = 8))

# sample categories
samples <- c(rep('Healthy', 4), rep('Pathogenic', 4))

# To calculate the difference of splicing diversity changes between the
# 'Healthy' and 'Pathogenic' condition together with the significance values,
# using mean and Wilcoxon rank sum test, use:
calculate_difference(x, samples, control = 'Healthy', method = 'mean', test = 'wilcoxon')
```

---

# calculate_diversity *Calculate Tsallis diversity per gene across samples*

---

## Description

Calculate Tsallis diversity per gene across samples

## Usage

```
calculate_diversity(
  x,
  genes = NULL,
  method = "tsallis",
  norm = TRUE,
  tpm = FALSE,
  assayno = 1,
  verbose = FALSE,
  q = 2,
  what = c("S", "D")
)
```

## Arguments

| | |
|---|---|
| x | A numeric matrix or data.frame of transcript-level expression values (rows = transcripts, columns = samples), or a SummarizedExperiment-like object. |
| genes | Character vector assigning each transcript (row) to a gene. Must have length equal to nrow(x) or the number of transcripts in ‘x‘. |
| method | Character; method to use for diversity calculation. Only ’tsallis’ is currently supported. |
| norm | Logical; if TRUE, normalize Tsallis entropy to [0,1] per gene. |
| tpm | Logical; if TRUE and ‘x‘ is a tximport-style list, use the ‘$abundance‘ matrix instead of ‘$counts‘. |
| assayno | Integer assay index to use when ‘x‘ is a SummarizedExperiment. |
| verbose | Logical; print diagnostic messages when TRUE. |
| q | Numeric scalar or vector of Tsallis q values to evaluate (q > 0). If length(q) > 1, the result will contain separate columns per sample and q. |
| what | Character; which quantity to return: "S" for Tsallis entropy, "D" for Hill numbers. |

## Value

A SummarizedExperiment with assay ‘diversity‘ containing per-gene diversity values.

---

# calculate_fc *Calculate splicing diversity changes between two conditions.*

---

## Description

Calculate splicing diversity changes between two conditions.

## Usage

```
calculate_fc(x, samples, control, method = "mean")
```

### Arguments

| | |
|---|---|
| x | A matrix with the splicing diversity values. |
| samples | Character vector with an equal length to the number of columns in the input dataset, specifying the category of each sample. |
| control | Name of the control sample category, defined in the samples vector, e.g. control = 'Normal' or control = 'WT'. |
| method | Method to use for calculating the average splicing diversity value in a condition. Can be 'mean' or 'median'. |

### Details

The function uses a matrix of splicing diversity values in order to calculate mean or median differences and log2 fold changes between two conditions.

### Value

A data.frame with mean or median value of splicing diversity across sample categories, the difference between these values and the log2 fold change values.

---

## calculate_lm_interaction

*Linear-model interaction test for Tsallis entropy*

---

### Description

For each gene, fit a linear model of the form ‘entropy ~ q * group‘ and extract the p-value for the interaction term (whether the effect of ‘q‘ differs between groups). The function expects a ‘SummarizedExperiment‘ produced by ‘calculate_diversity()‘ when ‘method = "tsallis"‘ and multiple ‘q‘ values have been computed (column names contain ‘_q=‘).

### Usage

```
calculate_lm_interaction(se, sample_type_col = NULL, min_obs = 10,
  method = c("linear", "gam", "fpca"), nthreads = 1, assay_name = "diversity")
```

### Arguments

| | |
|---|---|
| se | A ‘SummarizedExperiment‘ containing a ‘diversity‘ assay produced by ‘calculate_diversity(..., method = "tsallis", q = <vector>)‘. |
| sample_type_col | Optional column name in ‘colData(se)‘ that contains a grouping factor for samples (character). If ‘NULL‘, the function will attempt to infer group from column names (suffix ‘_N‘ interpreted as "Normal"). |
| min_obs | Minimum number of non-NA observations required to fit a model for a gene (default: 10). |
| method | Character; fitting method: "linear", "gam", or "fpca". |
| nthreads | Integer; number of threads to use for parallel processing (currently unused). |
| assay_name | Character; name of the assay in ‘se‘ to use (default: "diversity"). |

## Value

A data.frame with columns ‘gene‘, ‘p_interaction‘, and ‘adj_p_interaction‘, ordered by ascending ‘p_interaction‘.

---

`calculate_method` *Calculate Tsallis diversity values for transcripts grouped by gene*

---

## Description

This helper computes per-gene Tsallis entropy across samples. The trimmed package only supports the Tsallis method; other diversity metrics were removed.

## Usage

```
calculate_method(
  x,
  genes,
  method = "tsallis",
  norm = TRUE,
  verbose = FALSE,
  q = 2,
  what = c("S", "D", "both")
)
```

## Arguments

| | |
|---|---|
| `x` | Numeric matrix or data.frame of transcript-level expression values (rows = transcripts, columns = samples). |
| `genes` | Character vector with length equal to nrow(x) assigning each transcript to a gene. |
| `method` | Only "tsallis" is supported (default). |
| `norm` | Logical; if TRUE normalize Tsallis entropy values per gene. |
| `verbose` | Logical; show diagnostic messages when TRUE. |
| `q` | Numeric scalar or vector of q values to evaluate. |
| `what` | Character; which quantity to compute: "S" (entropy), "D" (Hill numbers), or "both". |

## Value

A data.frame with genes in the first column and per-sample (and per-q) Tsallis entropy values in subsequent columns.

## calculate_tsallis_entropy

*Calculate Tsallis entropy for a vector of transcript-level expression values of one gene.*

### Description

Calculate Tsallis entropy for a vector of transcript-level expression values of one gene.

### Usage

```
calculate_tsallis_entropy(x, q = 2, norm = TRUE, what = c("S", "D", "both"), log_base = exp(1))
```

### Arguments

| | |
|---|---|
| x | Vector of expression values. |
| q | Tsallis entropy parameter (q > 0). Can be a single value or a numeric vector. Default is 2. If q = 1, the function returns the Shannon entropy. |
| norm | If TRUE, the entropy values are normalized to the number of transcripts for each gene. The normalized entropy values are always between 0 and 1. If FALSE, genes cannot be compared to each other, due to possibly different maximum entropy values. |
| what | Which quantity to return: "S" for Tsallis entropy (S_q), "D" for Hill numbers (D_q), or "both". |
| log_base | Numeric; logarithm base used for the Shannon limit and normalization (default: e). |

### Details

The function calculates the Tsallis entropy, a generalization of Shannon entropy. For q -> 1, Tsallis entropy converges to Shannon entropy.

### Value

A single gene-level Tsallis entropy value (or a numeric vector when multiple q values provided).

## infer_sample_group *Infer sample group from sample names*

### Description

Infer sample group from sample names

### Usage

```
infer_sample_group(sample_names)
```

### Arguments

| | |
|---|---|
| sample_names | Character vector of sample names. |

### Value

Character vector of group labels (e.g. 'Normal'/'Tumor') or NA.

---

## label_shuffling *Calculate p-values using label shuffling.*

---

### Description

Calculate p-values using label shuffling.

### Usage

```
label_shuffling(
  x,
  samples,
  control,
  method,
  randomizations = 100,
  pcorr = "BH"
)
```

### Arguments

| | |
|---|---|
| x | A matrix with the splicing diversity values. |
| samples | Character vector with an equal length to the number of columns in the input dataset, specifying the category of each sample. |
| control | Name of the control sample category, defined in the samples vector, e.g. control = 'Normal' or control = 'WT'. |
| method | Method to use for calculating the average splicing diversity value in a condition. Can be 'mean' or 'median'. |
| randomizations | The number of random shuffles. |
| pcorr | P-value correction method applied to the results, as defined in the p.adjust function. |

### Details

The permutation p-values are computed two-sided as the proportion of permuted log2 fold-changes at least as extreme as the observed value, with a pseudocount added: (count + 1) / (n_perm + 1).

### Value

Raw and corrected p-values.

### Note

The permutation test returns two-sided empirical p-values using a pseudocount to avoid zero p-values for small numbers of permutations. See the function documentation for details.

## plot_diversity_density *Plot diversity distributions (density) by sample type*

### Description

Plot diversity distributions (density) by sample type

### Usage

```
plot_diversity_density(se, assay_name = "diversity", sample_type_col = NULL)
```

### Arguments

| | |
|---|---|
| se | A 'SummarizedExperiment' returned by 'calculate_diversity'. |
| assay_name | Name of the assay to use (default: "diversity"). |
| sample_type_col | Optional column name in 'colData(se)' that contains sample types. If missing, sample type will be inferred from column names (suffix after last underscore) or classified as 'Group'. |

### Value

A 'ggplot' object with layered density plots.

## plot_ma *Plot MA plot for difference results*

### Description

Plot MA plot for difference results

### Usage

```
plot_ma(
  diff_df,
  mean_cols = NULL,
  fold_col = "log2_fold_change",
  padj_col = "adjusted_p_values",
  sig_alpha = 0.05
)
```

### Arguments

| | |
|---|---|
| diff_df | Data.frame returned by 'calculate_difference' (or similar) containing mean columns and a 'log2_fold_change' column, and 'adjusted_p_values'. |
| mean_cols | Optional character vector of length 2 with the names of the mean columns (defaults to first two columns that end with '_mean'). |
| fold_col | Name of the fold-change column (default: 'log2_fold_change'). |
| padj_col | Name of the adjusted p-value column (default: 'adjusted_p_values'). |
| sig_alpha | Threshold for significance (default: 0.05). |

### Value

A ‘ggplot‘ MA-plot object.

---

`plot_mean_violin` *Plot violin of per-gene mean diversity by sample type*

---

### Description

Plot violin of per-gene mean diversity by sample type

### Usage

```
plot_mean_violin(se, assay_name = "diversity", sample_type_col = NULL)
```

### Arguments

| | |
|---|---|
| `se` | A ‘SummarizedExperiment‘ returned by ‘calculate_diversity‘. |
| `assay_name` | Name of the assay to use (default: "diversity"). |
| `sample_type_col` | Optional column name in ‘colData(se)‘ containing sample types. |

### Value

A ‘ggplot‘ violin plot object.

---

`plot_top_transcripts` *Plot top transcripts for a gene*

---

### Description

For a given gene, find transcripts using a tx->gene mapping, compute per-transcript statistics between two sample groups, select the top N transcripts by p-value and plot their expression across groups.

### Arguments

| | |
|---|---|
| `counts` | A numeric matrix of transcript-level expression (rows = transcripts, columns = samples). |
| `gene` | Character; gene symbol to inspect. |
| `samples` | Character vector of sample group labels (length = ncol(counts)). |
| `tx2gene` | Path to a two-column tab-delimited file with columns ‘Transcript‘ and ‘Gen‘, or a data.frame with those columns. Required. |
| `top_n` | Integer; number of transcripts to show (default = 3). If NULL, all transcripts for the gene are plotted. |
| `pseudocount` | Numeric value added when computing log2 fold-change to avoid division by zero (default = 1e-6). |
| `output_file` | Optional path to save the plot (ggsave will be used). If NULL the ggplot object is returned. |

### Value

A ‘ggplot‘ object (or invisibly saved file if ‘output_file‘ provided).

---

`plot_tsallis_density_multq` *Density plot of Tsallis entropy for multiple q values*

---

### Description

Density plot of Tsallis entropy for multiple q values

### Usage

```
plot_tsallis_density_multq(se, assay_name = "diversity")
```

### Arguments

| | |
|---|---|
| `se` | A ‘SummarizedExperiment‘ returned by ‘calculate_diversity‘ with multiple q values (colnames like ’Sample_q=0.01’). |
| `assay_name` | Name of the assay to use (default: "diversity"). |

### Value

A ‘ggplot‘ density plot object faceted by q and colored by group.

---

`plot_tsallis_q_curve` *Plot median +- IQR of Tsallis entropy across q values by group*

---

### Description

This reproduces the ‘tsallis-q-curve-mean-sd‘ plot from the vignette: for each q value, compute per-gene Tsallis entropy per sample, then summarize across genes by group (median and IQR) and plot median with a ribbon spanning median +- IQR/2.

### Usage

```
plot_tsallis_q_curve(
  readcounts,
  genes,
  q_values = seq(0.01, 2, by = 0.01),
  group_pattern = "_N$",
  group_names = c("Normal", "Tumor")
)
```

### Arguments

| | |
|---|---|
| `readcounts` | Numeric matrix or data.frame with transcripts as rows and samples as columns. |
| `genes` | Character vector assigning a gene id to each row of ‘readcounts‘. |
| `q_values` | Numeric vector of q values to evaluate (default ‘seq(0.01,2,by=0.01)‘). |
| `group_pattern` | Regular expression used to detect the first group in sample names (default ‘"_N$"‘). |
| `group_names` | Character vector of length 2 with names for groups (default ‘c("Normal","Tumor")‘). |

## Value

A ‘ggplot‘ object showing median +- IQR across q values by group.

---

`plot_tsallis_violin_multq` *Violin plot of Tsallis entropy for multiple q values*

---

## Description

Violin plot of Tsallis entropy for multiple q values

## Usage

```
plot_tsallis_violin_multq(se, assay_name = "diversity")
```

## Arguments

| | |
|---|---|
| `se` | A ‘SummarizedExperiment‘ returned by ‘calculate_diversity‘ with multiple q values (colnames like ’Sample_q=0.01’). |
| `assay_name` | Name of the assay to use (default: "diversity"). |

## Value

A ‘ggplot‘ violin plot object faceted/colored by group and q.

---

`plot_volcano` *Volcano plot for differential results*

---

## Description

Volcano plot for differential results

## Usage

```
plot_volcano(
  diff_df,
  x_col = "mean_difference",
  padj_col = "adjusted_p_values",
  label_thresh = 0.1,
  padj_thresh = 0.05,
  top_n = 5
)
```

### Arguments

| | |
|---|---|
| `diff_df` | Data.frame from ‘calculate_difference()‘ containing at least ‘mean_difference‘ and an adjusted p-value column (default ‘adjusted_p_values‘). |
| `x_col` | Column name for x-axis (default ‘mean_difference‘). |
| `padj_col` | Column name for adjusted p-values (default ‘adjusted_p_values‘). |
| `label_thresh` | Absolute x threshold to mark significance (default 0.1). |
| `padj_thresh` | Adjusted p-value cutoff (default 0.05). |
| `top_n` | Integer; number of top labeled genes to display (default 5). |

### Value

ggplot volcano plot.

---

`tcga_brca_luma_dataset`

*TCGA Luminal A breast cancer dataset*

---

### Description

Data from The Cancer Genome Atlas, downloaded on 08th September, 2020. It contains transcript level read counts of 20 patients with Luminal A type breast cancer (primary tumor and solid normal samples).

### Usage

```
data(tcga_brca_luma_dataset)
```

### Format

A data frame with 996 rows and 41 columns. The first column contains gene names, all additional columns contain RNA-sequencing read counts for samples.

### Source

TCGA Legacy (archived)

### References

The Cancer Genome Atlas Network (2012) Nature 490, 61–70 (doi:10.1038/nature11412)

wilcoxon *Calculate p-values using Wilcoxon rank sum test.*

## Description

Calculate p-values using Wilcoxon rank sum test.

## Usage

```
wilcoxon(x, samples, pcorr = "BH", paired = FALSE, exact = FALSE)
```

## Arguments

| | |
|---|---|
| x | A matrix with the splicing diversity values. |
| samples | Character vector with an equal length to the number of columns in the input dataset, specifying the category of each sample. |
| pcorr | P-value correction method applied to the results, as defined in the p.adjust function. |
| paired | If TRUE, the Wilcox-test will be paired, and therefore it will be a signed rank test instead of the rank sum test. |
| exact | If TRUE, an exact p-value will be computed. |

## Value

Raw and corrected p-values in a matrix.

# Index