

# Package ‘abyssEdge’

January 29, 2026

**Type** Package

**Title** Tsallis Entropy & Differential Diversity (abyssEdge fork)

**bioViews** Transcriptomics, RNASeq, DifferentialSplicing,  
AlternativeSplicing, TranscriptomeVariant

**Version** 1.3.1

**Author** Peter A. Szikora [aut], Tamas Pór [aut], Endre Sebestyén [aut, cre]

**Maintainer** Endre Sebestyén <endre.sebestyen@gmail.com>

**Description** This trimmed package focuses on calculating Tsallis entropy for transcript-level expression data and performing basic differential analyses (mean/median differences and log2 fold changes) between two sample conditions. Tsallis entropy generalizes Shannon entropy with a parameter `q` that controls sensitivity to transcript abundance; the package returns normalized Tsallis values and supports multiple `q` values per sample.

**RoxygenNote** 7.3.3

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

**Suggests** testthat, knitr, rmarkdown, tidyverse, dplyr

**URL** <https://github.com/your-org/abyssEdge>

**BugReports** <https://github.com/your-org/abyssEdge/issues>

**Depends** R (>= 4.1)

**License** GPL-3 + file LICENSE

**Encoding** UTF-8

**LazyData** true

**VignetteBuilder** knitr

## Contents

calculate_difference . . . . .	2
calculate_diversity . . . . .	4
calculate_fc . . . . .	5
calculate_lm_interaction . . . . .	5
calculate_method . . . . .	6
calculate_tsallis_entropy . . . . .	7
infer_sample_group . . . . .	7
label_shuffling . . . . .	8

plot_diversity_density . . . . .	8
plot_ma . . . . .	9
plot_mean_violin . . . . .	10
plot_tsallis_q_curve . . . . .	10
tcga_brca_luma_dataset . . . . .	11
wilcoxon . . . . .	11

**Index****13**

`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,
  coldata = NULL,
  group_col = NULL,
  control_opt = NULL,
  ...
)
```

**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 <code>p.adjust</code> function.
assayno	An integer value. In case of multiple assays in a <code>SummarizedExperiment</code> input, the argument specifies the assay number to use for difference calculations.
verbose	If TRUE, the function will print additional diagnostic messages.
coldata	Optional path to a tab-separated sample metadata file. The first column must contain sample identifiers which will be matched to the sample column names; if provided, the grouping column is taken from <code>group_col</code> or inferred from common names (e.g. Condition, Group).
group_col	Optional column name within coldata to use as the grouping factor. If NULL, a suitable column is auto-selected.
control_opt	Optional control group value used when coldata is provided. If set, this overrides control.
...	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
)
```

## Arguments

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

## Value

A [SummarizedExperiment](#) with assay ‘diversity’ containing per-gene diversity values.

---

calculate_fc	<i>Calculate splicing diversity changes between two conditions.</i>
--------------	---

---

## 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	<i>Linear-model interaction test for Tsallis entropy</i>
--------------------------	--

---

## 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)
```

**Arguments**

<code>se</code>	A ‘SummarizedExperiment’ containing a ‘diversity’ assay produced by ‘calculate_diversity(..., method = "tsallis", q = <vector>)’.
<code>sample_type_col</code>	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”).
<code>min_obs</code>	Minimum number of non-NA observations required to fit a model for a gene (default: 10).

**Value**

A data.frame with columns ‘gene’, ‘p\_interaction’, and ‘adj\_p\_interaction’, ordered by ascending ‘p\_interaction’.

<code>calculate_method</code>	<i>Calculate Tsallis diversity values for transcripts grouped by gene</i>
-------------------------------	---

**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
)
```

**Arguments**

<code>x</code>	Numeric matrix or data.frame of transcript-level expression values (rows = transcripts, columns = samples).
<code>genes</code>	Character vector with length equal to nrow(x) assigning each transcript to a gene.
<code>method</code>	Only “tsallis” is supported (default).
<code>norm</code>	Logical; if TRUE normalize Tsallis entropy values per gene.
<code>verbose</code>	Logical; show diagnostic messages when TRUE.
<code>q</code>	Numeric scalar or vector of q values to evaluate.

**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)
```

**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.

**Details**

The function calculates the Tsallis entropy, a generalization of Shannon entropy. For  $q \rightarrow 1$ , 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.

### Value

Raw and corrected p-values.

**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**

<code>se</code>	A ‘SummarizedExperiment‘ returned by ‘calculate_diversity‘.
<code>assay_name</code>	Name of the assay to use (default: "diversity").
<code>sample_type_col</code>	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**

<code>diff_df</code>	Data.frame returned by ‘calculate_difference‘ (or similar) containing mean columns and a ‘log2_fold_change‘ column, and ‘adjusted_p_values‘.
<code>mean_cols</code>	Optional character vector of length 2 with the names of the mean columns (defaults to first two columns that end with ‘_mean’).
<code>fold_col</code>	Name of the fold-change column (default: ‘log2_fold_change‘).
<code>padj_col</code>	Name of the adjusted p-value column (default: ‘adjusted_p_values‘).
<code>sig_alpha</code>	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\_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.

**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

**References**

The Cancer Genome Atlas Network (2012) Nature 490, 61–70 ([doi:10.1038/nature11412](https://doi.org/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

\* **datasets**  
tcga\_brca\_luma\_dataset, 11

calculate\_difference, 2  
calculate\_diversity, 4  
calculate\_fc, 5  
calculate\_lm\_interaction, 5  
calculate\_method, 6  
calculate\_tsallis\_entropy, 7

infer\_sample\_group, 7

label\_shuffling, 8

plot\_diversity\_density, 8  
plot\_ma, 9  
plot\_mean\_violin, 10  
plot\_tsallis\_q\_curve, 10

SummarizedExperiment, 4

tcga\_brca\_luma\_dataset, 11

wilcoxon, 11