

# Package ‘STADyUM’

April 7, 2026

**Type** Package

**Title** Statistical Transcriptome Analysis under a Dynamic Unified Model

**Version** 1.1.3

**Description** STADyUM is a package with functionality for analyzing nascent RNA read counts to infer transcription rates. This includes utilities for processing experimental nascent RNA read counts as well as for simulating PRO-seq data. Rates such as initiation, pause release and landing pad occupancy are estimated from either synthetic or experimental data. There are also options for varying pause sites and including steric hindrance of initiation in the model.

**URL** <https://github.com/rhassett-cshl/STADyUM>

**BugReports** <https://github.com/rhassett-cshl/STADyUM>

**biocViews** StatisticalMethod, Transcriptomics, Transcription,  
Sequencing

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** false

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.3

**Imports** GenomicRanges, IRanges, S4Vectors, methods, tibble, dplyr,  
ggplot2, progress, GenomeInfoDb, Rcpp, data.table, purrr,  
rtracklayer, tidyr, rlang, MASS

**LinkingTo** Rcpp

**Suggests** testthat (>= 3.0.0), knitr, rmarkdown, devtools

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**Depends** R (>= 4.5.0)

**git\_url** <https://git.bioconductor.org/packages/STADyUM>

**git\_branch** devel

**git\_last\_commit** c0a84f5

**git\_last\_commit\_date** 2025-12-04

**Repository** Bioconductor 3.23

**Date/Publication** 2026-04-06

**Author** Yixin Zhao [aut] (ORCID: <<https://orcid.org/0000-0002-7856-2818>>),  
 Lingjie Liu [aut] (ORCID: <<https://orcid.org/0000-0002-5765-1213>>),  
 Rebecca Hassett [aut, cre] (ORCID:  
 <<https://orcid.org/0000-0001-5133-9387>>)

**Maintainer** Rebecca Hassett <hassett@cshl.edu>

## Contents

estimateTranscriptionRates . . . . .	2
ExperimentTranscriptionRates-class . . . . .	3
rates . . . . .	7
show,ExperimentTranscriptionRates-method . . . . .	7
SimulatePolymerase-class . . . . .	8
SimulationTranscriptionRates-class . . . . .	15
stericHindrance . . . . .	18
TranscriptionRates-class . . . . .	18
TranscriptionRatesLRT-class . . . . .	21
<b>Index</b>	<b>25</b>

---

estimateTranscriptionRates

*Generic function for estimating transcription rates*

---

## Description

Generic function that estimates transcription rates from either simulation data (SimulatePolymerase object) or experimental data (bigwig files and genomic regions).

## Usage

```
estimateTranscriptionRates(x, ...)
```

## Arguments

x                    The input data (either a SimulatePolymerase object or bigwig files)  
 ...                  Additional arguments passed to the specific methods

## Value

An object containing estimated transcription rates

## Examples

```
load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
expRates <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control",
  stericHindrance = TRUE,
  omegaScale = 1000
)
```

---

ExperimentTranscriptionRates-class

*Constructor for ExperimentTranscriptionRates object*

---

## Description

Class ExperimentTranscriptionRates has read counts, pause and gene body genomic region coordinates, steric hindrance and omega scale factor parameters used to estimate the transcription rates as well as the estimated rates

Estimates the transcription rates using Expectation Maximization likelihood formula. Estimates initiation, pause-release rates, average read depth along gene body and pause regions. If steric hindrance is enabled, also estimates the landing pad occupancy. Estimated from experimental data, such as nascent RNA sequencing read counts and genomic coordinates, and constructs an object that holds these rates. Considers models where pause sites are fixed or varied. Considers models where steric hindrance is enabled or disabled.

Accessor for the estimated rates

Accessor for the read counts

Accessor for the pause region coords [GRanges-class](#)

Accessor for the gene body regions coords [GRanges-class](#)

Accessor for the omega scale used to scale the omega value based on prior studies.

Accessor for the steric hindrance flag. If TRUE, the landing-pad occupancy is inferred in the rates held in this object.

## Usage

```
## S4 method for signature 'character'
estimateTranscriptionRates(
  x,
  bigwigMinus,
  pauseRegions,
  geneBodyRegions,
```

```

    name,
    stericHindrance = FALSE,
    omegaScale = NULL
)

## S4 method for signature 'ExperimentTranscriptionRates'
rates(object)

counts(object)

## S4 method for signature 'ExperimentTranscriptionRates'
counts(object)

pauseRegions(object)

## S4 method for signature 'ExperimentTranscriptionRates'
pauseRegions(object)

geneBodyRegions(object)

## S4 method for signature 'ExperimentTranscriptionRates'
geneBodyRegions(object)

omegaScale(object)

## S4 method for signature 'ExperimentTranscriptionRates'
omegaScale(object)

## S4 method for signature 'ExperimentTranscriptionRates'
stericHindrance(object)

```

### Arguments

x	The path to a bigwig file from the plus strand recording PRO-seq read counts
bigwigMinus	the path to a bigwig file from the minus strand recording PRO-seq read counts
pauseRegions	a <a href="#">GRanges-class</a> object that must contain a gene_id
geneBodyRegions	a <a href="#">GRanges-class</a> object that must contain a gene_id
name	a character value for the name of the experiment
stericHindrance	a logical value to determine whether to infer landing-pad occupancy or not. Defaults to FALSE.
omegaScale	a numeric value for scaling omega. Defaults to NULL.
object	an <a href="#">ExperimentTranscriptionRates</a> object

### Value

an [ExperimentTranscriptionRates-class](#) object

a DataFrame with the following columns:

geneId	a character vector of gene IDs
chi	a numeric vector of RNAP density along gene body
betaOrg	a numeric vector
betaAdp	a numeric vector
fkMean	a numeric vector of the mean position of pause sites
fkVar	a numeric vector of the variance of pause site positions
phi	a numeric vector of the landing-pad occupancy
betaZeta	a numeric vector of the pause-escape rate
alphaZeta	a numeric vector of the potential initiation rate
omega	a numeric vector of the effective initiation rate
omegaZeta	a numeric vector of the effective initiation rate
expectedPauseSiteCounts	a numeric vector of the expected pause site counts
actualPauseSiteCounts	a numeric vector of the actual pause site counts that are observed
likelihood	a numeric vector of the likelihood of the model

a DataFrame with the following columns:

geneId	a character vector of gene IDs
summarizedGbCounts	a numeric vector of the sum of gene body read counts
gbLength	a numeric vector of the gene body length
summarizedPauseCounts	a numeric vector of the sum of pause region read counts

a [GRanges-class](#) object with the pause regions

a [GRanges-class](#) object with the gene body regions

a numeric value for the scaling factor for omega

a logical value to determine whether to infer landing-pad occupancy or not

## Slots

counts a data.frame with five columns gene\_id, summarizedPauseCounts, pauseLength, summarizedGbCounts, gbLength

bigwigPlus a path to bigwig for plus strand recording PRO-seq read counts. This can be generated with the proseq2.0 pipeline.

bigwigMinus a path to bigwig for minus strand recording PRO-seq read counts. This can be generated with the proseq2.0 pipeline.

pauseRegions a [GRanges-class](#) object that holds the pause regions coordinates

geneBodyRegions a [GRanges-class](#) object that holds the gene body regions coordinates

`name` a character value for the name of the experiment

`stericHindrance` a logical value indicating whether to infer landing-pad occupancy

`omegaScale` a numeric value for scaling omega, or NULL if steric hindrance is disabled

`rates` a `tbl_df` containing the estimated rates with columns:

**geneId** Character. Gene identifier

**chi** Numeric. RNAP density along gene body given as estimate for the gene body elongation rate [RNAPs/bp]

**betaOrg** Numeric. Ratio of gene body RNAP density to pause region RNAP density with fixed pause sites given as an estimate for the pause-escape rate

**betaAdp** Numeric. Ratio of gene body RNAP density to pause region RNAP density from adapted model which allows pause sites to vary across cells given as an estimate for the pause-escape rate

**fkMean** Numeric. Mean position of pause sites

**fkVar** Numeric. Variance of pause site positions

**phi** Numeric. Landing-pad occupancy (only if steric hindrance is enabled)

**betaZeta** Numeric. Pause-escape rate (only if steric hindrance is enabled)

**alphaZeta** Numeric. Potential initiation rate (only if steric hindrance is enabled)

**omegaZeta** Numeric. Effective initiation rate (only if steric hindrance is enabled)

**totalGbRc** Numeric. Total gene body read counts

**gbLength** Numeric. Gene body length

**expectedPauseSiteCounts** Numeric. Expected pause site counts

**actualPauseSiteCounts** Numeric. Observed pause site counts

## Examples

```
# Create an ExperimentTranscriptionRates object
load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
expRates <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control",
  stericHindrance = TRUE,
  omegaScale = 1000
)
```

---

rates	<i>Accessor for estimated rates</i>
-------	-------------------------------------

---

### Description

Generic accessor for the estimated rates from any `TranscriptionRates` object

### Usage

```
rates(object)
```

### Arguments

`object`            A `TranscriptionRates` object

### Value

A tibble containing the estimated rates

### Examples

```
# Create an ExperimentTranscriptionRates object
load(system.file("extdata",
"granges_for_read_counting_DLD1_chr21.RData", package = "STADyUM"))
expRates <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control",
  stericHindrance = TRUE,
  omegaScale = 1000
)
rates(expRates)
```

---

`show,ExperimentTranscriptionRates-method`

*Show method for ExperimentTranscriptionRates objects*

---

### Description

Enhanced `show` method for `ExperimentTranscriptionRates` objects that displays summary statistics and provides guidance on data access

**Usage**

```
## S4 method for signature 'ExperimentTranscriptionRates'
show(object)
```

**Arguments**

object                    An ExperimentTranscriptionRates object

**Value**

NULL (invisibly)

---

SimulatePolymerase-class

*Constructor for SimulatePolymerase object*

---

**Description**

Class `SimulatePolymerase` tracks the movement of RNAPs along the DNA templates. It contains the parameters for the simulator as well as the simulator results including the position of RNAPs for the last step, the pause sites, a probability vector containing probability RNAPs move forward or not, and a combined cells data vector containing the total number of RNAPs at each site across all cells

Runs the simulator that tracks the movement of RNAPs along the DNA templates of a large number of cells. It accepts several key user-specified parameters, including the initiation rate, pause-escape rate, a constant or variable elongation rate, the mean and variance of pause sites across cells, as well as the center-to-center spacing constraint between RNAPs, the number of cells being simulated, the gene length, and the total time of transcription. The simulator simply allows each RNAP to move forward or not, in time slices of  $1e-4$  minutes, according to the specified position-specific rate parameters. It assumes that at most one movement of each RNAP can occur per time slice. The simulator monitors for collisions between adjacent RNAPs, prohibiting one RNAP to advance if it is at the boundary of the allowable distance from the next. After running for specified time, it outputs the position of RNAPs for the last specified number of steps.

Sample read counts from a `SimulatePolymerase` object. To match our simulated read counts to reality, we need to compute a scaling factor  $\lambda$ . One way of doing it is computing the read density based on real experiments. For example, we have computed the read density within gene body in *Dukler et al (2017)* for genes with median expression (i.e., 0.0489). If we assume the read counts following a Poisson distribution, we can then sample the read counts with mean equals to the RNAP frequency multiplied by  $\lambda$ .

Show method for `SimulatePolymerase` object in human readable format including summary statistics

Accessor for the pause sites numeric vector from a `SimulatePolymerase` object.

Accessor for the probability numeric matrix from a `SimulatePolymerase` object. The matrix has dimensions (sites x cells) where each element represents the transition probability for a specific site and cell.

Accessor for the combined cells data numeric vector from a SimulatePolymerase object.

Accessor for the position matrices from a SimulatePolymerase object.

Accessor for the final position matrix from a SimulatePolymerase object. This represents the last state of the simulation.

Accessor for the simulation parameters from a SimulatePolymerase object.

Accessor for the read counts numeric vector sampled from a SimulatePolymerase object.

Get the position matrix for a specific time point from a SimulatePolymerase object.

Get all available time points from a SimulatePolymerase object.

Get the combined cells data vector for a specific time point from a SimulatePolymerase object. This function calculates the total number of polymerases at each site across all cells for the specified time point, similar to the final combinedCellsData but for intermediate time points.

Plot the distribution of pause sites across the gene as a histogram.

Plot position matrices as a ggplot2 heatmap showing polymerase positions across all cells at specific time points. Each cell in the heatmap represents whether a polymerase is present (1) or absent (0) at a specific site in a specific cell. By default, shows the final position matrix. Plots for pause region which is from site 1 to user-defined kMax.

Create a ggplot2 PCA plot of the polymerase position matrix, showing the first two principal components for each cell. Useful for exploring heterogeneity and clustering among cells. Useful for exploring the site occupancy patterns for each cell.

Plot combined cells data as a ggplot2 plot. For SimulatePolymerase objects, can plot data from specific time points by calculating combined cells data from position matrices.

## Usage

```
simulatePolymerase(
  k = 50,
  ksd = 25,
  kMin = 17,
  kMax = 200,
  geneLen = 1950,
  alpha = 1,
  beta = 1,
  zeta = 2000,
  zetaSd = 1000,
  zetaMin = 1500,
  zetaMax = 2500,
  zetaVec = NULL,
  cellNum = 1000,
  polSize = 33,
  addSpace = 17,
  time = 1,
  timesToRecord = NULL
)

sampleReadCounts(object, readDensity = 0.0489)
```

```
## S4 method for signature 'SimulatePolymerase'  
sampleReadCounts(object, readDensity = 0.0489)  
  
## S4 method for signature 'SimulatePolymerase'  
show(object)  
  
pauseSites(object)  
  
## S4 method for signature 'SimulatePolymerase'  
pauseSites(object)  
  
siteProbabilities(object)  
  
## S4 method for signature 'SimulatePolymerase'  
siteProbabilities(object)  
  
combinedCellsData(object)  
  
## S4 method for signature 'SimulatePolymerase'  
combinedCellsData(object)  
  
positionMatrices(object)  
  
## S4 method for signature 'SimulatePolymerase'  
positionMatrices(object)  
  
finalPositionMatrix(object)  
  
## S4 method for signature 'SimulatePolymerase'  
finalPositionMatrix(object)  
  
parameters(object)  
  
## S4 method for signature 'SimulatePolymerase'  
parameters(object)  
  
readCounts(object)  
  
## S4 method for signature 'SimulatePolymerase'  
readCounts(object)  
  
getPositionMatrixAtTime(object, timePoint)  
  
## S4 method for signature 'SimulatePolymerase'  
getPositionMatrixAtTime(object, timePoint)  
  
getAvailableTimePoints(object)
```

```
## S4 method for signature 'SimulatePolymerase'
getAvailableTimePoints(object)

getCombinedCellsDataAtTime(object, timePoint = NULL)

## S4 method for signature 'SimulatePolymerase'
getCombinedCellsDataAtTime(object, timePoint = NULL)

plotPauseSites(object, file = NULL, width = 8, height = 6)

## S4 method for signature 'SimulatePolymerase'
plotPauseSites(object, file = NULL, width = 8, height = 6)

plotPositionHeatmap(
  object,
  timePoint = NULL,
  maxCells = NULL,
  file = NULL,
  width = 10,
  height = 8,
  dpi = 300
)

## S4 method for signature 'SimulatePolymerase'
plotPositionHeatmap(
  object,
  timePoint = NULL,
  maxCells = NULL,
  file = NULL,
  width = 10,
  height = 8,
  dpi = 300
)

plotPolymerasePCA(
  object,
  timePoint = NULL,
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)

## S4 method for signature 'SimulatePolymerase'
plotPolymerasePCA(
  object,
  timePoint = NULL,
```

```

    file = NULL,
    width = 8,
    height = 6,
    dpi = 300
)

plotCombinedCells(
  object,
  start = NULL,
  end = NULL,
  timePoint = NULL,
  file = NULL
)

## S4 method for signature 'SimulatePolymerase'
plotCombinedCells(
  object,
  start = NULL,
  end = NULL,
  timePoint = NULL,
  file = NULL
)

```

### Arguments

<code>k</code>	an integer value for the mean of pause sites across cells.
<code>ksd</code>	a numeric value for the standard deviation of pause sites across cells.
<code>kMin</code>	an integer value for the upper bound of pause sites allowed.
<code>kMax</code>	an integer value for the lower bound of pause sites allowed.
<code>geneLen</code>	an integer value for the length of the gene.
<code>alpha</code>	a numeric value for the initiation rate.
<code>beta</code>	a numeric value for the pause release rate.
<code>zeta</code>	a numeric value for the mean elongation rate across sites.
<code>zetaSd</code>	a numeric value for the standard deviation of pause sites across sites.
<code>zetaMin</code>	a numeric value for the minimum elongation rate.
<code>zetaMax</code>	a numeric value for the maximum elongation rate.
<code>zetaVec</code>	a character value for the path to the <code>zetaVec</code> file.
<code>cellNum</code>	an integer value for the number of cells to simulate.
<code>polSize</code>	an integer value for the polymerase II size.
<code>addSpace</code>	an integer value for the additional space in addition to RNAP size.
<code>time</code>	a numeric value for the time to simulate.
<code>timesToRecord</code>	a numeric vector of specific time points to record position matrices for, or NULL to record no extra position matrices. Final position matrix is always recorded. Default is NULL.

object	A SimulatePolymerase-class object
readDensity	A numeric value for the read density within gene body in <i>Dukler et al. (2017)</i> for genes with median expression (i.e., 0.0489).
timePoint	Optional numeric value specifying the time point being plotted (used for automatic title generation and documentation)
file	Optional file path to save the plot to
width	Plot width in inches
height	Plot height in inches
maxCells	Maximum number of cells to display (for performance with large datasets). If NULL, shows all cells.
dpi	Plot resolution in DPI
start	Integer, starting position for plotting (default: NULL, uses full range)
end	Integer, ending position for plotting (default: NULL, uses full range)

**Value**

a SimulatePolymerase object

The SimulatePolymerase object with read counts sampled given the readDensity parameter used

A list containing all simulation parameters including k, ksd, kMin, kMax, geneLen, alpha, beta, zeta, zetaSd, zetaMin, zetaMax, zetaVec, cellNum, polSize, addSpace, time, and stepsToRecord

A 2D matrix of polymerase positions for the specified time point

A numeric vector of available time points

An integer vector representing the total number of polymerases at each site across all cells for the specified time point

A ggplot object showing the distribution of pause sites

A ggplot object showing heatmap of polymerase positions

A ggplot object showing the PCA plot of cells

A ggplot object

**Slots**

k an integer value for the mean of pause sites across cells.

ksd a numeric value for the standard deviation of pause sites across cells.

kMin an integer value for the upper bound of pause sites allowed.

kMax an integer value for the lower bound of pause sites allowed.

geneLen an integer value for the length of the gene.

alpha a numeric value for the initiation rate.

beta a numeric value for the pause release rate.

zeta a numeric value for the mean elongation rate across sites.

zetaSd a numeric value for the standard deviation of pause sites across sites.

zetaMin a numeric value for the minimum elongation rate.  
 zetaMax a numeric value for the maximum elongation rate.  
 zetaVec a character value for the path to the zetaVec file.  
 cellNum an integer value for the number of cells to simulate.  
 polSize an integer value for the polymerase II size.  
 addSpace an integer value for the additional space in addition to RNAP size.  
 time a numeric value for the time to simulate.  
 timesToRecord a numeric vector of specific time points to record position matrices for, or NULL to record no extra position matrices. Final position matrix is always recorded. Default is NULL.  
 pauseSites a numeric vector of pause sites  
 siteProbabilities a numeric matrix representing the probability that the polymerase move forward or not at each site for each cell (sites x cells)  
 combinedCellsData an integer vector representing the total number of RNAPs at each site across all cells  
 positionMatrices a list of position matrices  
 finalPositionMatrix a matrix representing the final position matrix  
 readCounts a numeric vector for read counts per nucleotide

### Note

The simulatePolymerase function can take a few minutes to run for large numbers of cells (> 1000) and long simulation times (> 200).

### Examples

```

# Create a SimulatePolymerase object
sim <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500,
  zetaMax = 2500, zetaVec = NULL, cellNum = 100, polSize = 33,
  addSpace = 17, time = 1, timesToRecord = NULL
)
# Create a SimulatePolymerase object with specific time points recorded
sim2 <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500, zetaMax
= 2500,
  zetaVec = NULL, cellNum = 1000, polSize = 33, addSpace = 17, time = 1,
  timesToRecord = c(0.5, 1.0)
)
# Create a SimulatePolymerase object
sim <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500,
  zetaMax = 2500, zetaVec = NULL, cellNum = 100, polSize = 33,
  addSpace = 17, time = 1, timesToRecord = NULL

```

```

)
# Sample read counts per nucleotide
sim <- sampleReadCounts(sim)
# Print the read counts per nucleotide
print(readCounts(sim))
# Create a SimulatePolymerase object
sim <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500,
  zetaMax = 2500, zetaVec = NULL, cellNum = 100, polSize = 33,
  addSpace = 17, time = 1, timesToRecord = c(0.5, 1.0))
# Plot final position matrix
plotPositionHeatmap(sim, file="position_heatmap.png")
# Plot position matrix at time 0.5
plotPositionHeatmap(sim, timePoint = 0.5, file="position_heatmap.png")
# Plot position matrix with maxCells = 100
plotPositionHeatmap(sim, timePoint = 0.5, maxCells = 100,
file="position_heatmap.png")
# Create a SimulatePolymerase object
sim <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500,
  zetaMax = 2500, zetaVec = NULL, cellNum = 100, polSize = 33,
  addSpace = 17, time = 1, timesToRecord = c(0.5, 1.0))
# Plot final combined cells data
plotCombinedCells(sim, file="combined_cells.png")
# Plot specific time point data
plotCombinedCells(sim, timePoint = 0.5, file="combined_cells.png")
# Plot with custom range and title
plotCombinedCells(sim, timePoint = 0.5, start = 100, end = 500,
file="combined_cells.png")

```

---

### SimulationTranscriptionRates-class

*Constructor for SimulationTranscriptionRates object*

---

### Description

Class containing the simulated data, such as the nascent RNA sequencing reads sampled from the simulated polymerase movement in the `SimulatePolymerase` object. It also contains the estimated average read depths along the gene body and pause regions, given fixed or varied pause sites, as well as the landing pad occupancy estimate. These can be under a model with or without steric hindrance.

Estimates transcription rates from a `SimulatePolymerase` object using Expectation Maximization likelihood formula.

Show method for `SimulationTranscriptionRates` object in human readable format including summary statistics

Accessor for the `SimulatePolymerase` object from a `SimulationTranscriptionRates` object.

Accessor for the steric hindrance flag from a SimulationTranscriptionRates object.

Accessor for the tibble containing the estimated rates from a SimulationTranscriptionRates object.

### Usage

```
## S4 method for signature 'SimulatePolymerase'
estimateTranscriptionRates(x, name, stericHindrance = FALSE)

## S4 method for signature 'SimulationTranscriptionRates'
show(object)

simpol(object)

## S4 method for signature 'SimulationTranscriptionRates'
simpol(object)

## S4 method for signature 'SimulationTranscriptionRates'
stericHindrance(object)

## S4 method for signature 'SimulationTranscriptionRates'
rates(object)
```

### Arguments

`x` a [SimulatePolymerase](#) object

`name` a string providing information about the simulation

`stericHindrance` a logical value to determine whether to infer landing-pad occupancy or not. Defaults to FALSE.

`object` a SimulationTranscriptionRates object

### Value

a [SimulationTranscriptionRates](#) object

### Slots

`simpol` a [SimulatePolymerase](#) object

`name` a character value for the name of the experiment

`stericHindrance` a logical value to determine whether to infer landing-pad occupancy or not. Defaults to FALSE.

`rates` a `tbl_df` containing the estimated rates with columns:

- trial** Numeric. Trial number, each trial samples 5000 cells
- chi** Numeric. RNAP density along gene body given as estimate for the gene body elongation rate [RNAPs/bp]
- betaOrg** Numeric. Ratio of gene body RNAP density to pause region RNAP density with fixed pause sites given as an estimate for the pause-escape rate

- betaAdp** Numeric. Ratio of gene body RNAP density to pause region RNAP density from adapted model which allows pause sites to vary across cells given as an estimate for the pause-escape rate
- phi** Numeric. Landing-pad occupancy representing the fraction of time in the simulation that the landing pad is occupied by RNA polymerase. The landing pad, also known as the initiation site, is the region where the RNA polymerase binds to the DNA and begins transcription. It is the `polSize + addSpace` length (only applicable if steric hindrance is enabled)
- fk** list. Likelihood of pausing at each pause region position
- fkMean** Numeric. Mean position of pause sites
- fkVar** Numeric. Variance of pause site positions
- totalTssRc** Numeric. Total RNAP read counts in the TSS region (across all cells)
- totalGbRc** Numeric. Total RNAP read counts in the gene body region (across all cells)
- totalLandingRc** Numeric. Total RNAP read counts in the landing pad region (across all cells)
- avgRcPerCell** Numeric. Average number of RNAPs per cell for gene body and TSS regions
- avgTssRcPerCell** Numeric. Average number of RNAPs per cell for pause regions (across all cells)
- avgLandingRcPerCell** Numeric. Average number of RNAPs per cell for landing pad regions (across all cells)
- actualPauseSiteCounts** list. Simulated pause site counts for each cell
- expectedPauseSiteCounts** list. Expected pause site counts for each cell estimated from the EM algorithm
- expectationMaximizationStatus** character. Status of the expectation maximization algorithm. "normal": converged normally, "single\_site": converged to single pause site, "max\_iterations": reached maximum iterations without convergence. Note max # of iterations is 500.
- likelihood** a numeric vector of the likelihood of the model

## Examples

```
# Create a SimulatePolymerase object
sim <- simulatePolymerase(
  k = 50, ksd = 25, kMin = 17, kMax = 200, geneLen = 1950,
  alpha = 1, beta = 1, zeta = 2000, zetaSd = 1000, zetaMin = 1500,
  zetaMax = 2500, zetaVec = NULL, cellNum = 100, polSize = 33,
  addSpace = 17, time = 1)
# Estimate transcription rates
estRates <- estimateTranscriptionRates(sim, name="sim_beta1_k50")
# Print the estimated rates
print(estRates)
```

stericHindrance      *Accessor for steric hindrance flag*

---

### Description

Generic accessor for the steric hindrance flag from any TranscriptionRates object

### Usage

```
stericHindrance(object)
```

### Arguments

object      A TranscriptionRates object

### Value

A logical value indicating whether steric hindrance was modeled

### Examples

```
# Create an ExperimentTranscriptionRates object
load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
expRates <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control",
  stericHindrance = TRUE,
  omegaScale = 1000
)
stericHindrance(expRates)
```

---

TranscriptionRates-class

*Base class for TranscriptionRates objects*

---

**Description**

Virtual base class that defines the common interface for transcription rate objects. Both `ExperimentTranscriptionRates` and `SimulationTranscriptionRates` inherit from this class.

Creates a histogram plot showing the distribution of observed mean pause site positions across all genes. This visualization helps identify the range and shape of pause site positions

Creates a scatter plot comparing actual pause site counts against expected pause site counts from the EM algorithm. It is comparing the number of polymerase at the pause site in the simulated/experimental data against the number of polymerase at the pause site expected by the model. This visualization assesses the goodness-of-fit of the pause site model by showing how well the model predictions align with the actual data. A perfect fit would show all points on the diagonal line ( $y=x$ ). The  $R^2$  value is calculated and displayed on the plot to quantify the model fit quality. This plot is useful for validating the accuracy of the pause site estimation and identifying any systematic biases in the model predictions.

Creates a density plot showing the distribution of gene body RNAP density ( $\chi$ ) across all genes. This visualization helps identify the range and shape of RNA polymerase density in gene bodies, which can reveal patterns in transcriptional activity.

Plot a scatter plot with gene body RNAP density on the x-axis and beta (ratio of gene body RNAP density to pause region RNAP density) on the y-axis. Fits a linear model to the data and plots the line. Can plot beta for either the adapted model or the single pause site model.

Plot a contour map with mean pause site position on the x-axis and pause site variance on the y-axis.

**Usage**

```
plotMeanPauseDistrib(object, file = NULL, width = 8, height = 6, dpi = 300)
```

```
## S4 method for signature 'TranscriptionRates'
```

```
plotMeanPauseDistrib(object, file = NULL, width = 8, height = 6, dpi = 300)
```

```
plotExpectedVsActualPauseSiteCounts(
  object,
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)
```

```
## S4 method for signature 'TranscriptionRates'
```

```
plotExpectedVsActualPauseSiteCounts(
  object,
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)
```

```
plotChiDistrib(object, file = NULL, width = 8, height = 6, dpi = 300)
```

```

## S4 method for signature 'TranscriptionRates'
plotChiDistrib(object, file = NULL, width = 8, height = 6, dpi = 300)

plotBetaVsChi(
  object,
  betaType = "betaAdp",
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)

## S4 method for signature 'TranscriptionRates'
plotBetaVsChi(
  object,
  betaType = "betaAdp",
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)

plotPauseSiteContourMap(object, file = NULL, width = 8, height = 6, dpi = 300)

## S4 method for signature 'TranscriptionRates'
plotPauseSiteContourMap(object, file = NULL, width = 8, height = 6, dpi = 300)

```

### Arguments

object	an <a href="#">TranscriptionRates</a> object
file	the path to a file to save the plot to
width	the width of the plot in inches
height	the height of the plot in inches
dpi	the resolution of the plot in dpi
betaType	the type of beta to plot. Can be "betaAdp" for the adapted model or "betaOrg" for the single pause site model. Defaults to "betaAdp".

### Value

an [ggplot2](#) object  
 an [ggplot2](#) object  
 an [ggplot2](#) object  
 an [ggplot2](#) object  
 an [ggplot2](#) object

**Slots**

rates a `tbl_df` containing the estimated rates  
 name a character value for the name of the experiment  
 stericHindrance a logical value indicating whether steric hindrance was modeled

**Examples**

```
# Create an ExperimentTranscriptionRates object
load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
expRates <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control"
)
plotPauseSiteContourMap(expRates, file="pause_sites_contour_map.png")
```

---

 TranscriptionRatesLRT-class

*Constructor for TranscriptionRatesLRT object*

---

**Description**

Constructs results of likelihood ratio test comparing transcription data estimated from two different sets of experimental read counts.

Likelihood ratio test comparing aspects of transcriptional dynamics of the same TU under different conditions using Likelihood Ratio Test Statistics. Uses read counts and rate estimates estimated from `estimateTranscriptionRates`. The method also requires scaling factors to determine changes in  $\chi$  estimates. They can be the numbers of total mapped reads or spike-in reads from the samples. Likelihood ratio test computes the log 2 fold change in  $\chi$  estimates between conditions, the log 2 fold change in beta estimates between conditions, the t-statistics for the likelihood ratio tests, and the adjusted p-values based on the "BH" method.

Accessor for the first `ExperimentTranscriptionRates` object from a `TranscriptionRatesLRT` object.

Accessor for the second `ExperimentTranscriptionRates` object from a `TranscriptionRatesLRT` object.

Accessor for the spike-in scaling factor from a `TranscriptionRatesLRT` object.

Accessor for the chi table from a `TranscriptionRatesLRT` object.

Accessor for the beta table from a `TranscriptionRatesLRT` object.

Plot a contour map with mean pause site position on the x-axis and pause site standard deviation on the y-axis. The plot is a comparison between two conditions.

Plot a violin plot comparing the beta values between two conditions.

Plot a violin plot comparing the chi values between two conditions.

Plot a MA plot for log fold change treated/control vs log mean beta\*zeta.

### Usage

```
likelihoodRatioTest(
  transcriptionRates1,
  transcriptionRates2,
  scaleFactor = 1,
  spikeInFile = NULL
)

transcriptionRates1(object)

## S4 method for signature 'TranscriptionRatesLRT'
transcriptionRates1(object)

transcriptionRates2(object)

## S4 method for signature 'TranscriptionRatesLRT'
transcriptionRates2(object)

spikeInFile(object)

## S4 method for signature 'TranscriptionRatesLRT'
spikeInFile(object)

chiTbl(object)

## S4 method for signature 'TranscriptionRatesLRT'
chiTbl(object)

betaTbl(object)

## S4 method for signature 'TranscriptionRatesLRT'
betaTbl(object)

plotPauseSiteContourMapTwoConditions(
  object,
  file = NULL,
  width = 8,
  height = 6,
  dpi = 300
)

## S4 method for signature 'TranscriptionRatesLRT'
plotPauseSiteContourMapTwoConditions(
```

```

    object,
    file = NULL,
    width = 8,
    height = 6,
    dpi = 300
)

BetaViolinPlot(object, file = NULL, width = 8, height = 6, dpi = 300)

## S4 method for signature 'TranscriptionRatesLRT'
BetaViolinPlot(object, file = NULL, width = 8, height = 6, dpi = 300)

ChiViolinPlot(object, file = NULL, width = 8, height = 6, dpi = 300)

## S4 method for signature 'TranscriptionRatesLRT'
ChiViolinPlot(object, file = NULL, width = 8, height = 6, dpi = 300)

plotLfcMa(object, file = NULL, width = 8, height = 6, dpi = 300)

## S4 method for signature 'TranscriptionRatesLRT'
plotLfcMa(object, file = NULL, width = 8, height = 6, dpi = 300)

```

### Arguments

transcriptionRates1	an <a href="#">TranscriptionRates</a> object
transcriptionRates2	an <a href="#">TranscriptionRates</a> object
scaleFactor	a numeric value to scale the beta estimates. Defaults to 1
spikeInFile	path to a csv file containing scale factors based on total or spike-in reads or NULL if not provided. Defaults to NULL.
object	an <a href="#">TranscriptionRatesLRT</a> object
file	the path to a file to save the plot to
width	the width of the plot in inches
height	the height of the plot in inches
dpi	the resolution of the plot in dpi

### Value

a [TranscriptionRatesLRT](#) object  
 tbl\_df  
 tbl\_df  
 an [ggplot2](#) object  
 an [ggplot2](#) object  
 an [ggplot2](#) object  
 an [ggplot2](#) object

**Examples**

```

load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
transcriptionRates1 <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control"
)
transcriptionRates2 <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Treated"
)
spikeInFile <- system.file("extdata", "spikein_scaling_factor.csv",
package = "STADyUM")
lrts <- likelihoodRatioTest(transcriptionRates1, transcriptionRates2,
scaleFactor=1, spikeInFile)
# Print the likelihood ratio test object
print(lrts)
# Create an ExperimentTranscriptionRates object
load(system.file("extdata", "granges_for_read_counting_DLD1_chr21.RData",
package = "STADyUM"))
transcriptionRates1 <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Control"
)
transcriptionRates2 <- estimateTranscriptionRates(system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_plus_chr21.bw", package = "STADyUM"),
bigwigMinus = system.file("extdata",
"PROseq-DLD1-aoi-NELFC_Auxin_Ctrl-SE_minus_chr21.bw", package = "STADyUM"),
  pauseRegions = bw_pause_filtered,
  geneBodyRegions = bw_gb_filtered,
  name = "Treated"
)
spikeInFile <- system.file("extdata", "spikein_scaling_factor.csv",
package = "STADyUM")
lrts <- likelihoodRatioTest(transcriptionRates1, transcriptionRates2,
scaleFactor=1, spikeInFile)
plotPauseSiteContourMapTwoConditions(lrts,
file="pause_sites_contour_map.png")

```

# Index

betaTbl (TranscriptionRatesLRT-class),  
21

betaTbl, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class),  
21

BetaViolinPlot  
(TranscriptionRatesLRT-class),  
21

BetaViolinPlot, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class),  
21

chiTbl (TranscriptionRatesLRT-class), 21

chiTbl, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class),  
21

ChiViolinPlot  
(TranscriptionRatesLRT-class),  
21

ChiViolinPlot, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class),  
21

combinedCellsData  
(SimulatePolymerase-class), 8

combinedCellsData, SimulatePolymerase-method  
(SimulatePolymerase-class), 8

counts  
(ExperimentTranscriptionRates-class),  
3

counts, ExperimentTranscriptionRates-method  
(ExperimentTranscriptionRates-class),  
3

estimateTranscriptionRates, 2

estimateTranscriptionRates, character-method  
(ExperimentTranscriptionRates-class),  
3

estimateTranscriptionRates, SimulatePolymerase-method 3  
(SimulationTranscriptionRates-class),  
15

ExperimentTranscriptionRates-class, 3

finalPositionMatrix  
(SimulatePolymerase-class), 8

finalPositionMatrix, SimulatePolymerase-method  
(SimulatePolymerase-class), 8

geneBodyRegions  
(ExperimentTranscriptionRates-class),  
3

geneBodyRegions, ExperimentTranscriptionRates-method  
(ExperimentTranscriptionRates-class),  
3

getAvailableTimePoints  
(SimulatePolymerase-class), 8

getAvailableTimePoints, SimulatePolymerase-method  
(SimulatePolymerase-class), 8

getCombinedCellsDataAtTime  
(SimulatePolymerase-class), 8

getCombinedCellsDataAtTime, SimulatePolymerase-method  
(SimulatePolymerase-class), 8

getPositionMatrixAtTime  
(SimulatePolymerase-class), 8

getPositionMatrixAtTime, SimulatePolymerase-method  
(SimulatePolymerase-class), 8

ggplot2, 20, 23

GRanges-class, 3–5

likelihoodRatioTest  
(TranscriptionRatesLRT-class),  
21

omegaScale  
(ExperimentTranscriptionRates-class),  
3

omegaScale, ExperimentTranscriptionRates-method  
(ExperimentTranscriptionRates-class),  
3

parameters (SimulatePolymerase-class), 8

- parameters, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- pauseRegions  
(ExperimentTranscriptionRates-class), 3
- pauseRegions, ExperimentTranscriptionRates-method  
(ExperimentTranscriptionRates-class), 3
- pauseSites (SimulatePolymerase-class), 8
- pauseSites, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- plotBetaVsChi  
(TranscriptionRates-class), 18
- plotBetaVsChi, TranscriptionRates-method  
(TranscriptionRates-class), 18
- plotChiDistrib  
(TranscriptionRates-class), 18
- plotChiDistrib, TranscriptionRates-method  
(TranscriptionRates-class), 18
- plotCombinedCells  
(SimulatePolymerase-class), 8
- plotCombinedCells, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- plotExpectedVsActualPauseSiteCounts  
(TranscriptionRates-class), 18
- plotExpectedVsActualPauseSiteCounts, TranscriptionRates-method  
(TranscriptionRates-class), 18
- plotLfcMa  
(TranscriptionRatesLRT-class), 21
- plotLfcMa, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class), 21
- plotMeanPauseDistrib  
(TranscriptionRates-class), 18
- plotMeanPauseDistrib, TranscriptionRates-method  
(TranscriptionRates-class), 18
- plotPauseSiteContourMap  
(TranscriptionRates-class), 18
- plotPauseSiteContourMap, TranscriptionRates-method  
(TranscriptionRates-class), 18
- plotPauseSiteContourMapTwoConditions  
(TranscriptionRatesLRT-class), 21
- plotPauseSiteContourMapTwoConditions, TranscriptionRatesLRT-method  
(TranscriptionRatesLRT-class), 21
- plotPauseSites  
(SimulatePolymerase-class), 8
- plotPauseSites, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- plotPolymerasePCA  
(SimulatePolymerase-class), 8
- plotPolymerasePCA, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- plotPositionHeatmap  
(SimulatePolymerase-class), 8
- plotPositionHeatmap, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- positionMatrices  
(SimulatePolymerase-class), 8
- positionMatrices, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- rates, 7
- rates, ExperimentTranscriptionRates-method  
(ExperimentTranscriptionRates-class), 3
- rates, SimulationTranscriptionRates-method  
(SimulationTranscriptionRates-class), 15
- readCounts (SimulatePolymerase-class), 8
- readCounts, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- sampleReadCounts  
(SimulatePolymerase-class), 8
- sampleReadCounts, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- show, ExperimentTranscriptionRates-method, 7
- show, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- show, SimulationTranscriptionRates-method  
(SimulationTranscriptionRates-class), 15
- simpl  
(SimulationTranscriptionRates-class), 15
- simpl, SimulationTranscriptionRates-method  
(SimulationTranscriptionRates-class), 15
- SimulatePolymerase, 16
- SimulatePolymerase, SimulatePolymerase-method  
(SimulatePolymerase-class), 8
- SimulatePolymerase-class, 8
- SimulationTranscriptionRates, 16

SimulationTranscriptionRates-class, [15](#)  
siteProbabilities  
    (SimulatePolymerase-class), [8](#)  
siteProbabilities, SimulatePolymerase-method  
    (SimulatePolymerase-class), [8](#)  
spikeInFile  
    (TranscriptionRatesLRT-class),  
    [21](#)  
spikeInFile, TranscriptionRatesLRT-method  
    (TranscriptionRatesLRT-class),  
    [21](#)  
stericHindrance, [18](#)  
stericHindrance, ExperimentTranscriptionRates-method  
    (ExperimentTranscriptionRates-class),  
    [3](#)  
stericHindrance, SimulationTranscriptionRates-method  
    (SimulationTranscriptionRates-class),  
    [15](#)  
  
tbl\_df, [6](#), [16](#), [21](#)  
TranscriptionRates, [20](#), [23](#)  
TranscriptionRates-class, [18](#)  
transcriptionRates1  
    (TranscriptionRatesLRT-class),  
    [21](#)  
transcriptionRates1, TranscriptionRatesLRT-method  
    (TranscriptionRatesLRT-class),  
    [21](#)  
transcriptionRates2  
    (TranscriptionRatesLRT-class),  
    [21](#)  
transcriptionRates2, TranscriptionRatesLRT-method  
    (TranscriptionRatesLRT-class),  
    [21](#)  
TranscriptionRatesLRT, [23](#)  
TranscriptionRatesLRT-class, [21](#)