

# Package ‘BasicSTARRseq’

February 3, 2025

**Type** Package

**Title** Basic peak calling on STARR-seq data

**Version** 1.34.0

**Date** 2015-01-11

**Author** Annika Buerger

**Maintainer** Annika Buerger <annika.buerger@ukmuenster.de>

**Imports** S4Vectors,methods,IRanges,GenomeInfoDb,stats

**Depends** GenomicRanges,GenomicAlignments

**Description** Basic peak calling on STARR-seq data based on a method introduced in “Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq” Arnold et al. Science. 2013 Mar 1;339(6123):1074-7. doi: 10.1126/science. 1232542. Epub 2013 Jan 17.

**License** LGPL-3

**LazyData** TRUE

**Suggests** knitr

**VignetteBuilder** knitr

**biocViews** PeakDetection, GeneRegulation, FunctionalPrediction, FunctionalGenomics, Coverage

**NeedsCompilation** no

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

**git\_branch** RELEASE\_3\_20

**git\_last\_commit** 21c0bef

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.20

**Date/Publication** 2025-02-03

## Contents

|                              |          |
|------------------------------|----------|
| getPeaks . . . . .           | 2        |
| STARRseqData-class . . . . . | 3        |
| <b>Index</b>                 | <b>5</b> |

---

getPeaks *Peak calling on STARR-seq data*

---

### Description

Performs basic peak calling on STARR-seq data based on a method introduced in "Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq" Arnold et al. [1]

### Usage

```
getPeaks(object, minQuantile = 0.9, peakWidth = 500, maxPval = 0.001,
         deduplicate = TRUE, model = 1)
```

### Arguments

|             |   |
|-------------|---|
| object      | A <a href="#">STARRseqData</a> object for which the peaks should be calculated. |
| minQuantile | Which quantile of coverage height should be considered as peaks.                |
| peakWidth   | The width (in base pairs) that the peaks should have.                           |
| maxPval     | The maximal p-value of peaks that is desired.                                   |
| deduplicate | Whether the sequences should be deduplicated before calling peaks or not.       |
| model       | Which binomial model should be applied to calculate the p-values.               |

### Details

The peak calling works the following way: All genomic positions having a STARR-seq coverage over the quantile `minQuantile` are considered to be the center of a peak with width `peakWidth`. If then two or more peaks overlap, the lower one is discarded. If then the binomial p-Value of the peak is higher than `maxPval` the peak is discarded as well.

The binomial model 1 for calculating the p-Value is: number of trials = total number of STARR-seq sequences, number of successes = STARR-seq coverage, estimated success probability in each trial = input coverage/total number of input sequences.

The binomial model 2 for calculating the p-Value is: number of trials = STARR-seq coverage plus input coverage, number of successes = STARR-seq coverage, estimated success probability in each trial = total number of STARR-seq sequences/(total number of STARR-seq sequences plus total number of input sequences). This model is used in [1].

The enrichment of STARR-seq over input coverage is then calculated as follows: (STARR-seq coverage of peak/total number of STARR-seq sequences)/(input coverage of peak/total number of input sequences), the numerator and denominator corrected conservatively to the bounds of the 0.95 binomial confidence interval corresponding to model 1.

### Value

The method `getPeaks` return a [GRanges](#) object. The contained ranges are the found peaks with desired width `peakWidth`. The metadata columns of the ranges contain four elements:

|            |   |
|------------|---|
| sampleCov  | The maximal and central STARR-seq coverage of the peak.   |
| controlCov | The maximum of the central and the median input coverage of the peak.   |
| pVal       | The binomial p-Value of the coverage height of the peak normalised to total number of sequences in STARR-seq and input. |

enrichment      The enrichment of STARR-seq over input coverage height normalised to total number of sequences in STARR-seq and input corrected conservatively to the bounds of a confidence interval.

### Author(s)

Annika Buerger

### References

[1] *Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq*. Arnold et al. Science. 2013 Mar 1;339(6123):1074-7. doi: 10.1126/science.1232542. Epub 2013 Jan 17.

### See Also

[GRanges STARRseqData-class](#)

### Examples

```
# create a small sample STARRseqData object
starrseqFileName <- system.file("extdata", "smallSTARR.bam",
                                package="BasicSTARRseq")
inputFileName <- system.file("extdata", "smallInput.bam",
                              package="BasicSTARRseq")
data <- STARRseqData(sample=starrseqFileName, control=inputFileName,
                    pairedEnd=TRUE)

# call peaks with default parameters
peaks = getPeaks(data)

# call peaks with no deduplication and no restriction concerning p-value
peaks = getPeaks(data, maxPval = 1, deduplicate = FALSE)

# call peaks with other binomial model and width 700
peaks = getPeaks(data, peakWidth = 700, model = 2)

# call peaks assuming less regions as potential peaks
peaks = getPeaks(data, minQuantile = 0.99)
```

---

STARRseqData-class      *Class "STARRseqData"*

---

### Description

The STARR-seq data class is a container for STARR-sequencing data.

### Details

STARRseqData contains two GRanges objects that store the STARR-seq sequences and the input sequences respectively of an STARR-seq experiment.

**Slots**

sample: Object of class "GRanges" which contains STARR-seq sequences.

control: Object of class "GRanges" which contains input sequences.

**Constructor**

STARRseqData(sample, control): Create a STARRseqData object.

sample: An GRanges object.

control: An GRanges object.

**Accessors**

In the following code snippets, x is an STARRseqData object.

sample(x), sample(x) <- value: Get or set the STARR-seq sequences.

control(x), control(x) <- value: Get or set the input sequences.

**Methods**

**getPeaks** signature(object = "STARRseqData"): Performs basic peak calling on data.

**Author(s)**

A. Buerger

**References**

*Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq.* Arnold et al. Science. 2013 Mar 1;339(6123):1074-7. doi: 10.1126/science.1232542. Epub 2013 Jan 17.

**See Also**

[GRanges getPeaks](#)

**Examples**

```
# create small sample dataset
starrseqFileName <- system.file("extdata", "smallSTARR.bam", package="BasicSTARRseq")
inputFileName <- system.file("extdata", "smallInput.bam", package="BasicSTARRseq")
STARRseqData(sample=starrseqFileName, control=inputFileName, pairedEnd=TRUE)
```

# Index

## \* classes

- STARRseqData-class, 3
- control (STARRseqData-class), 3
- control, STARRseqData-method
  - (STARRseqData-class), 3
- control<- (STARRseqData-class), 3
- control<- , STARRseqData, GRanges-method
  - (STARRseqData-class), 3
- getPeaks, 2, 4
- getPeaks, STARRseqData-method
  - (getPeaks), 2
- GRanges, 2–4
- sample (STARRseqData-class), 3
- sample, STARRseqData-method
  - (STARRseqData-class), 3
- sample<- (STARRseqData-class), 3
- sample<- , STARRseqData, GRanges-method
  - (STARRseqData-class), 3
- STARRseqData, 2
- STARRseqData (STARRseqData-class), 3
- STARRseqData, character, character, logical-method
  - (STARRseqData-class), 3
- STARRseqData, GRanges, GRanges, ANY-method
  - (STARRseqData-class), 3
- STARRseqData, GRanges, GRanges-method
  - (STARRseqData-class), 3
- STARRseqData-class, 3