

Package: RBPSpecificity (via r-universe)

July 22, 2026

Type Package

Title RBP Inherent Specificity and Variation Sensitivity Analysis Tool

Version 0.99.6

Description Provides tools to analyze RNA Binding Protein (RBP) binding specificities from high-throughput sequencing data. Functions include calculating K-mer enrichment, Inherent Specificity (IS), and Mutational Sensitivity (VS), along with visualization of IS and VS. For detailed methodology and applications, please refer to our manuscript (see URL field or vignette).

License GPL-3

Encoding UTF-8

biocViews Software, MotifAnnotation, Sequencing, Coverage, GenomeAnnotation, Visualization, GeneRegulation, KEGG

BugReports <https://github.com/S00NYI/RBPSpecificity/issues>

URL <https://www.biorxiv.org/content/10.1101/2025.03.28.646018v2>,
<https://github.com/S00NYI/RBPSpecificity>,
https://github.com/S00NYI/BITS_Specificity

Depends R (>= 4.5.0)

Imports Biostrings, BSgenome, GenomeInfoDb, GenomicRanges, ggplot2, methods, reshape2, S4Vectors, stats, utils

Suggests BiocStyle, BSgenome.Hsapiens.UCSC.hg38, IRanges, knitr, rmarkdown, testthat

VignetteBuilder knitr

Config/roxygen2/version 8.0.0

Config/pak/sysreqs

make libbz2-dev libicu-dev liblzma-dev libxml2-dev libssl-dev xz-utils zlib1g-dev

Repository <https://bioc.r-universe.dev>

Date/Publication 2026-07-19 01:16:56 UTC

RemoteUrl <https://github.com/bioc/RBPSpecificity>

RemoteRef HEAD
RemoteSha e1b8712b78890e3dc3299ab7f9a074b7adc35431

Contents

RBPSpecificity-package	2
motifEnrichment	3
plotSensitivity	5
plotSpecificity	6
returnSensitivity	7
returnSpecificity	8
Index	9

RBPSpecificity-package	<i>RBPSpecificity: Analyze RBP Binding Specificity and Variation Sensitivity</i>
------------------------	--

Description

The ‘RBPSpecificity’ package provides a suite of tools for researchers working with RNA Binding Proteins (RBPs) to analyze binding specificity from high-throughput sequencing data such as CLIP (CrossLinking and ImmunoPrecipitation) and RBNS (RNA Bind-n-Seq).

It allows users to quantify K-mer enrichment within RBP binding sites, assess the inherent specificity (IS) and the variation sensitivity (VS) of RBP towards the target motifs. The package aims to streamline bioinformatic analyses and provide clear visualizations for interpreting RBP specificity.

Note: This package is under active development.

Key Functions

- The main workflow of the package revolves around the following exported functions:
- **motifEnrichment**: Calculates K-mer enrichment from peak data...
 - **returnSpecificity**: Quantifies the specificity of RBP...
 - **plotSpecificity**: Visualizes the distribution of motif enrichment scores with annotation...
 - **returnSensitivity**: Quantifies the sensitivity of RBP...
 - **plotSensitivity**: Generates plots to visualize sensitivity

Author(s)

Soon Yi <cu.soonyi@gmail.com>

See Also

Useful links:

- <https://github.com/S00NYI/RBPSpecificity> (Development repository)
- Report bugs at <https://github.com/S00NYI/RBPSpecificity/issues>

motifEnrichment

Calculate K-mer Motif Enrichment from Genomic Coordinates

Description

Main workflow function for RBPSpecificity. Processes RBP binding peak data to calculate background-corrected K-mer enrichment scores using either ANR (total occurrence) or ZOOPS (per-sequence presence) statistical models.

Usage

```
motifEnrichment(
  coordinates,
  species_or_build,
  K,
  extension = c(0, 0),
  method = "anr",
  bkg_iter = 100,
  bkg_min_dist = 500,
  bkg_max_dist = 1000,
  scramble_bkg = FALSE,
  nucleic_acid_type = "DNA"
)
```

Arguments

coordinates	A data frame or GRanges object with genomic coordinates. Data frames need chr, start, end columns.
species_or_build	Character, genome build (e.g. "hg38").
K	Integer, K-mer length (e.g. 5).
extension	Numeric vector of length 2: c(five_prime, three_prime). Positive extends, negative trims. Default: c(0, 0).
method	Character, enrichment model: "anr" (default) for total occurrences or "zoops" for per-sequence presence.
bkg_iter	Integer, background iterations. Default: 100.
bkg_min_dist	Integer, min shift distance. Default: 500.
bkg_max_dist	Integer, max shift distance. Default: 1000.

scramble_bkg Logical, scramble background sequences to control for nucleotide composition. Default: FALSE.

nucleic_acid_type Character, "DNA" or "RNA". Default: "DNA".

Details

The workflow proceeds as follows: 1. Parses input coordinates into a GRanges object. 2. Loads the specified BSgenome for sequence retrieval. 3. Extracts peak sequences with optional extension/trimming. 4. Counts K-mers (both total and per-sequence presence). 5. Generates background K-mer profiles from shifted regions. 6. Computes enrichment with statistical testing.

ANR vs. ZOOPS Motif Occurrence Models: Two statistical models are available via the method parameter:

- **ANR (Any Number of Repetitions):** Counts every occurrence of a motif within each sequence (retains multiplicity). Enrichment is evaluated using a conditional binomial test on aggregate occurrence counts, under the assumption that occurrences are Poisson-distributed.
- **ZOOPS (Zero-or-One Occurrence Per Sequence):** Treats each sequence as a binary outcome (motif present or absent). Enrichment is evaluated using the cumulative hypergeometric distribution.

ANR is sensitive to motif multiplicity and cooperative binding hotspots, whereas ZOOPS is robust to repetitive elements and compositional noise.

Value

A data frame with 10 columns:

MOTIF K-mer sequence.

Score Min-max normalized log2FC, range [0, 1].

log2FC Log2 fold-change (method-specific).

pvalue Statistical significance (method-specific).

padj Benjamini-Hochberg adjusted p-value.

target_fraction Fraction of peak sequences containing the motif (ZOOPS).

bkg_fraction Fraction of background sequences containing the motif (ZOOPS).

target_occurrence Total motif occurrences in peaks.

bkg_occurrence Background occurrence rate (per sequence).

multiplicity Average occurrences per containing sequence.

Examples

```
if (requireNamespace("BSgenome.Hsapiens.UCSC.hg38", quietly = TRUE)) {
  peaks_file <- system.file("extdata",
    "ENCODE_eCLIP_HNRNPC_K562_narrowPeak.bed",
    package = "RBPSpecificity"
  )
  peaks <- read.table(peaks_file,
```

```

      header = FALSE, sep = "\t"
    )
    colnames(peaks) <- c(
      "chr", "start", "end",
      "name", "score", "strand"
    )
    result <- motifEnrichment(peaks, "hg38",
      K = 5,
      method = "anr", bkg_iter = 5
    )
    head(result)
  }

```

plotSensitivity	<i>Plot Sensitivity Matrix</i>
-----------------	--------------------------------

Description

Visualizes the sensitivity scores for a motif, typically showing sensitivity to SNVs at each position.

Usage

```
plotSensitivity(motif_enrichment, motif = NULL, ...)
```

Arguments

motif_enrichment	A data frame containing motif enrichment scores (output from 'motifEnrichment()'). Requires 'MOTIF' and 'Score' columns.
motif	Character string, the reference motif for which sensitivity was calculated. If NULL or "top", the top-scoring motif is used (default: NULL).
...	Additional arguments passed to ggplot theme layers or geoms.

Value

A ggplot object visualizing the sensitivity matrix (e.g., using points sized or colored by sensitivity).

Examples

```

# Mock data with variants
motifs <- c("AAAA", "CAAA", "GAAA", "TAAA")
scores <- c(10, 5, 4, 3)
df <- data.frame(MOTIF = motifs, Score = scores)

# Plot sensitivity for "AAAA"
plotSensitivity(df, motif = "AAAA")

```

plotSpecificity	<i>Plot Specificity Distribution</i>
-----------------	--------------------------------------

Description

Generates a histogram of motif enrichment scores, annotated with the median score, the score of a specific motif, and its specificity value.

Usage

```
plotSpecificity(motif_enrichment, motif = NULL, bins = 50, ...)
```

Arguments

motif_enrichment	A data frame containing motif enrichment scores (output from 'motifEnrichment()'). Requires 'MOTIF' and 'Score' columns.
motif	Character string, the specific motif to highlight and calculate specificity for. If NULL or "top", the top-scoring motif is used (default: NULL).
bins	Integer, number of bins for the histogram (default: 50).
...	Additional arguments passed to ggplot theme layers or geoms.

Value

A ggplot object representing the annotated histogram.

Examples

```
# Dummy data
df <- data.frame(MOTIF = c("AAAA", "CCCC", "GGGG", "UUUU"), Score = c(10, 2, 5, 1))

# Plot specific motif
plotSpecificity(df, motif = "AAAA")

# Plot top motif (AAAA)
plotSpecificity(df)

# Change number of bins
plotSpecificity(df, bins = 20)
```

returnSensitivity	<i>Calculate Sensitivity</i>
-------------------	------------------------------

Description

Calculates the sensitivity for a target motif or for all motifs. Sensitivity reflects how much the enrichment score changes upon single nucleotide variations (SNVs).

Usage

```
returnSensitivity(
  motif_enrichment,
  motif = NULL,
  return_type = "specific",
  output_type = "matrix",
  sensitivity_method = "1_minus_norm_score"
)
```

Arguments

motif_enrichment	A data frame with 'MOTIF' and 'Score' columns.
motif	Character string, the reference motif. Only used if 'return_type = "specific"'. If NULL or "top", the top-scoring motif is used.
return_type	Character string, either "specific" (default) or "all". If "specific", calculates sensitivity for one motif. If "all", calculates an average sensitivity score for every motif.
output_type	Character string, specifying the output format if 'return_type = "specific"': "matrix" (default) or "number". Ignored if 'return_type = "all"'.
sensitivity_method	Character string defining how sensitivity is calculated. Default: "1_minus_norm_score".

Value

Depends on inputs: - If 'return_type = "specific"' and 'output_type = "matrix"': A sensitivity matrix. - If 'return_type = "specific"' and 'output_type = "number"': A single numeric sensitivity score. - If 'return_type = "all"': A data frame with 'MOTIF' and 'Sensitivity' columns.

Examples

```
# Mock data with a motif (AAAAA) and some variants
motifs <- c("AAAAA", "CAAAA", "GAAAA", "TAAAA", "ACAAA")
scores <- c(100, 50, 40, 30, 60)
df <- data.frame(MOTIF = motifs, Score = scores)

# Calculate sensitivity matrix
sens_mat <- returnSensitivity(df, motif = "AAAAA", output_type = "matrix")
```

```
# Calculate average sensitivity score
sens_score <- returnSensitivity(df, motif = "AAAAA", output_type = "number")
```

returnSpecificity	<i>Calculate Specificity</i>
-------------------	------------------------------

Description

Calculates the specificity for a given motif based on its enrichment score relative to the median enrichment score of all motifs.

Usage

```
returnSpecificity(motif_enrichment, motif = NULL, return_type = "specific")
```

Arguments

motif_enrichment	A data frame containing motif enrichment scores, typically output from 'motifEnrichment()'. Must have columns 'MOTIF' and 'Score'.
motif	Character string, the specific motif for which to calculate specificity. Only used if 'return_type = "specific"'. If NULL or "top", the top-scoring motif is used. Default: NULL.
return_type	Character string, either "specific" (default) or "all". If "specific", returns a single specificity value for the target motif. If "all", returns a data frame with specificity values for every motif.

Value

A single numeric specificity value if 'return_type = "specific"', or a data frame with 'MOTIF' and 'Specificity' columns if 'return_type = "all"'.

Examples

```
# Create a dummy enrichment dataframe
df <- data.frame(MOTIF = c("AAAAA", "CCCCC", "GGGGG"), Score = c(10, 2, 5))

# Calculate specificity for a specific motif
returnSpecificity(df, motif = "AAAAA")

# Calculate specificity for all motifs
returnSpecificity(df, return_type = "all")
```

Index

`motifEnrichment`, [2](#), [3](#)

`plotSensitivity`, [2](#), [5](#)

`plotSpecificity`, [2](#), [6](#)

`RBPSpecificity`

(`RBPSpecificity-package`), [2](#)

`RBPSpecificity-package`, [2](#)

`returnSensitivity`, [2](#), [7](#)

`returnSpecificity`, [2](#), [8](#)