Package: MethScope (via r-universe)

October 26, 2026

Title Ultra-Fast Analysis of Sparse DNA Methylome via Recurrent Pattern Encoding

Version 1.0.4

Description Methods for analyzing DNA methylation data via Most Recurrent Methylation Patterns (MRMPs). Supports cell-type annotation, spatial deconvolution, unsupervised clustering, and cancer cell-of-origin inference. Includes C-backed summaries for YAME ``.cg/.cm" files (overlap counts, log2 odds ratios, beta/depth aggregation), an XGBoost classifier, NNLS deconvolution, and plotting utilities. Scales to large spatial and single-cell methylomes and is robust to extreme sparsity.

License AGPL-3

Encoding UTF-8

RoxygenNote 7.3.2

Imports xgboost, dplyr, utils, tidyr, stringr, caret, doParallel, parallel, ggplot2, uwot, magrittr, FNN, data.table, nnls

Suggests knitr, rmarkdown, spelling, testthat

VignetteBuilder knitr

Depends R (>= 4.0)

SystemRequirements zlib

NeedsCompilation yes

Language en-US

Author Hongxiang Fu [aut, cre] (ORCID: <https://orcid.org/0000-0002-9873-8606>), Wanding Zhou [cph, fnd], The SAMtools/HTSlib authors [ctb, cph] (BGZF components; see inst/COPYRIGHTS), Attractive Chaos [ctb, cph] (Author and copyright holder of khash.h (klib, MIT license))

Maintainer Hongxiang Fu <fhx@seas.upenn.edu>

Config/pak/sysreqs zlib1g-dev

Repository https://cran.r-universe.dev

Date/Publication 2026-08-26 23:46:18 UTC

RemoteUrl https://github.com/cran/MethScope

RemoteRef HEAD

RemoteSha e927782e5dbbbf4a3a895ba6ba114d88bae648d8

Contents

confidence_score . . . 2
confidence_score_top95 . . . 3
filter_cell . . . 3
GenerateInput . . . 4
GenerateReference . . . 4
imputeRowMean . . . 5
Input_training . . . 5
Liu2021_MouseBrain_P1000 . . . 6
nnls_deconv . . . 6
PlotConfusion . . . 7
PlotF1 . . . 7
PlotUMAP . . . 8
PlotUMAP_fixedwindow . . . 8
PredictCellType . . . 9
smooth_matrix . . . 10
Zhou2025_HumanAtlas_P1000 . . . 10

Index **11**

`confidence_score` *Produce confidence score for XGBoost prediction*

Description

Produce confidence score for XGBoost prediction

Usage

```
confidence_score(vec)
```

Arguments

`vec`	A vector of predicted probability for each cell type

Value

A numeric confidence score from 0 to 1.

confidence_score_top95 *Produce confidence score based on top 95 percent for XGBoost prediction*

Description

Produce confidence score based on top 95 percent for XGBoost prediction

Usage

```
confidence_score_top95(vec)
```

Arguments

vec	A vector of predicted probability for each cell type

Value

A numeric confidence score from 0 to 1.

filter_cell *Filter final prediction to reduce noise*

Description

Filter final prediction to reduce noise

Usage

```
filter_cell(pred_result, knn_res, KNeighbor = 5)
```

Arguments

pred_result	The prediction result from XGBoost
knn_res	knn graph from smooth_matrix
KNeighbor	Number of knn neighbors to use for smoothing (Default: 5)

Value

The final prediction result after dropping few cell types

GenerateInput *Generate pattern level data for cell type annotation*

Description

Generate pattern level data for cell type annotation

Usage

```
GenerateInput(query_fn, knowledge_fn)
```

Arguments

query_fn	File path to query .cg
knowledge_fn	File path to pattern file .cm

Value

A cell by pattern matrix.

GenerateReference *Generate reference pattern file*

Description

Generate reference pattern file

Usage

```
GenerateReference(binary_file, min_CG = 50, output_path = "./patterns.txt")
```

Arguments

binary_file	File path from the output of the bash script
min_CG	The minimum CpG # a pattern must have (Default: 50)
output_path	File path to store the output txt file (Default: current path)

Value

NULL, write out a patterns.txt file

`imputeRowMean` *Impute missing value for 100K window matrix*

Description

Impute missing value for 100K window matrix

Usage

```
imputeRowMean(mtx, na_percent = 30)
```

Arguments

`mtx`	A cell by 100K window data frame with missing values
`na_percent`	A na percent threshold to be filterd (Default: 30)

Value

A cell by 100K window data frame with imputed values

`Input_training` *Train XGBoost model to predict cell type*

Description

Train XGBoost model to predict cell type

Usage

```
Input_training(
  summary_results,
  cell_type_label,
  number_patterns = 1000,
  cross_validation = FALSE,
  xgb_parameters = list()
)
```

Arguments

`summary_results`	a wide cell by pattern matrix generated from GenerateInput function
`cell_type_label`	a vector of the corresponding cell type label for each row of the summary results
`number_patterns`	a numeric value to indicate number of patterns to be used (Default: 1000)

`cross_validation`
: a boolean varaible whether to perform cross_validation to obtain the best hyper parameters for the model

`xgb_parameters`
: an optional list for xgb model parameters provided by the user

Value

A MethScopeModel object containing the trained XGBoost booster and metadata.

`Liu2021_MouseBrain_P1000` *Load the Liu et al. mouse brain pretrained model*

Description

Load the Liu et al. mouse brain pretrained model

Usage

```
Liu2021_MouseBrain_P1000()
```

Value

A MethScopeModel object for use with `PredictCellType`.

`nnls_deconv` *Estimate cell type relative proportion*

Description

Estimate cell type relative proportion

Usage

```
nnls_deconv(ref, mixture_matrix, number_patterns = 1000, var_threshold = 0.01)
```

Arguments

`ref`
: An imputed wide cell by pattern matrix generated from GenerateInput function using reference Pseudobulk

`mixture_matrix`
: An imputed wide cell by pattern matrix generated from GenerateInput function

`number_patterns`
: a numeric value to indicate number of patterns to be used (Default: 1000)

`var_threshold`
: a numeric value to indicate variance that should filter the patterns (Default: 0.1)

Value

A cell type by cell matrix showing the relative cell type proportion estimate for each cells

PlotConfusion *Generate confusion table for the final prediction*

Description

Generate confusion table for the final prediction

Usage

```
PlotConfusion(prediction_result, actual_label, log2 = F)
```

Arguments

prediction_result	Prediction result from PredictCellType
actual_label	Ground truth cell label
log2	Log scale count (Default: False)

Value

A ggplot2 confusion table object.

PlotF1 *Generate F1 score barplot for each class*

Description

Generate F1 score barplot for each class

Usage

```
PlotF1(prediction_result, actual_label)
```

Arguments

prediction_result	Prediction result from PredictCellType
actual_label	Ground truth cell label

Value

A ggplot2 object.

PlotUMAP *Generate UMAP for the final prediction based on cell patterns*

Description

Generate UMAP for the final prediction based on cell patterns

Usage

```
PlotUMAP(predictMatrix, prediction_result, n_component = 30, seed = 123, ...)
```

Arguments

predictMatrix	a wide cell by pattern matrix generated from GenerateInput function
prediction_result	Prediction result from PredictCellType
n_component	Number of PCA components to use (Default: 30)
seed	A number for random seed (Default: 123)
...	Additional arguments passed to uwot::umap (e.g., n_neighbors, metric).

Value

A list of two ggplot2 UMAP object.

PlotUMAP_fixedwindow *Generate UMAP for the final prediction based on fixed window eg.100kb bin widows*

Description

Generate UMAP for the final prediction based on fixed window eg.100kb bin widows

Usage

```
PlotUMAP_fixedwindow(
  query_fn,
  knowledge_fn,
  prediction_result,
  n_component = 30,
  seed = 123,
  ...
)
```

Arguments

query_fn	File path to query .cg
knowledge_fn	File path to 100bk bins window or reference pattern
prediction_result	Prediction result from PredictCellType
n_component	Number of PCA components to use (Default: 30)
seed	A number for random seed (Default: 123)
...	Additional arguments passed to uwot::umap (e.g., n_neighbors, metric).

Value

A list of two ggplot2 UMAP object.

PredictCellType *Predict cell type annotation from the trained model*

Description

Predict cell type annotation from the trained model

Usage

```
PredictCellType(bst_model, predictMatrix, smooth = FALSE, KNeighbor = 5)
```

Arguments

bst_model	The boosting model trained from ModelTrain
predictMatrix	A wide cell by pattern matrix generated from GenerateInput function
smooth	A Boolean variable to indicate whether smooth the matrix (Default: FALSE)
KNeighbor	number of knn neighbors to use for smoothing (Default: 5)

Value

A cell by cell type matrix with confidence score and labeled cell type.

Examples

```
## Not run:
# Run from the root of a cloned zhou-lab/MethScope repository.
input_pattern <- GenerateInput("inst/extdata/example.cg",
                               "inst/extdata/mm10_Liu2021.cm")
model <- Liu2021_MouseBrain_P1000()
prediction <- PredictCellType(model, input_pattern)

## End(Not run)
```

smooth_matrix *Smooth cell by pattern matrix to reduce noise*

Description

Smooth cell by pattern matrix to reduce noise

Usage

```
smooth_matrix(predictMatrix, KNeighbor = 5)
```

Arguments

predictMatrix	A wide cell by pattern matrix generated from GenerateInput function
KNeighbor	Number of knn neighbors to use for smoothing (Default: 5)

Value

A wide cell by pattern matrix after smoothing and knn graph

Zhou2025_HumanAtlas_P1000 *Load the Zhou et al. human atlas pretrained model*

Description

Load the Zhou et al. human atlas pretrained model

Usage

```
Zhou2025_HumanAtlas_P1000()
```

Value

A MethScopeModel object for use with PredictCellType.

Index

confidence_score, 2
confidence_score_top95, 3

filter_cell, 3

GenerateInput, 4
GenerateReference, 4

imputeRowMean, 5
Input_training, 5

Liu2021_MouseBrain_P1000, 6

nnls_deconv, 6

PlotConfusion, 7
PlotF1, 7
PlotUMAP, 8
PlotUMAP_fixedwindow, 8
PredictCellType, *6*, 9, *10*

smooth_matrix, 10

Zhou2025_HumanAtlas_P1000, 10