

# Package ‘SPOTlight’

December 19, 2025

**Version** 1.15.0

**Type** Package

**Title** `SPOTlight`: Spatial Transcriptomics Deconvolution

**Description** `SPOTlight` provides a method to deconvolute spatial transcriptomics spots using a seeded NMF approach along with visualization tools to assess the results. Spatially resolved gene expression profiles are key to understand tissue organization and function. However, novel spatial transcriptomics (ST) profiling techniques lack single-cell resolution and require a combination with single-cell RNA sequencing (scRNA-seq) information to deconvolute the spatially indexed datasets. Leveraging the strengths of both data types, we developed SPOTlight, a computational tool that enables the integration of ST with scRNA-seq data to infer the location of cell types and states within a complex tissue. SPOTlight is centered around a seeded non-negative matrix factorization (NMF) regression, initialized using cell-type marker genes and non-negative least squares (NNLS) to subsequently deconvolute ST capture locations (spots).

**Depends** R (>= 4.5.0)

**Imports** ggplot2, Matrix, SingleCellExperiment, sparseMatrixStats, stats

**Suggests** BiocStyle, colorBlindness, DelayedArray, DropletUtils, ExperimentHub, ggcorrplot, grDevices, grid, igraph, jpeg, knitr, methods, png, rmarkdown, scater, scatterpie, scan, SpatialExperiment, SummarizedExperiment, S4Vectors, TabulaMurisSenisData, TENxVisiumData, testthat

**LinkingTo** Rcpp, RcppEigen

**biocViews** SingleCell, Spatial, StatisticalMethod

**License** GPL-3

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**VignetteBuilder** knitr

**URL** <https://github.com/MarcElosua/SPOTlight>

**BugReports** <https://github.com/MarcElosua/SPOTlight/issues>

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

**git\_branch** devel

**git\_last\_commit** b112ee3

**git\_last\_commit\_date** 2025-10-29

**Repository** Bioconductor 3.23

**Date/Publication** 2025-12-18

**Author** Marc Elosua-Bayes [aut, cre],  
Zachary DeBruine [aut],  
Helena L. Crowell [aut]

**Maintainer** Marc Elosua-Bayes <elosua.marc@gmail.com>

**Contents**

|                                 |           |
|---------------------------------|-----------|
| .filter . . . . .               | 2         |
| data . . . . .                  | 4         |
| plotCorrelationMatrix . . . . . | 5         |
| plotImage . . . . .             | 6         |
| plotInteractions . . . . .      | 7         |
| plotSpatialScatterpie . . . . . | 8         |
| plotTopicProfiles . . . . .     | 10        |
| runDeconvolution . . . . .      | 11        |
| SPOTlight . . . . .             | 13        |
| trainNMF . . . . .              | 15        |
| <b>Index</b>                    | <b>18</b> |

```
.filter      .init_nmf <- function(x, groups, mgs, n_top = NULL, gene_id
= "gene", group_id = "cluster", weight_id = "weight") #
check validity of input arguments if (is.null(n_top)) n_top <-
max(table(mgs[[group_id]])) stopifnot( is.character(gene_id),
length(gene_id) == 1, is.character(group_id), length(group_id) == 1,
is.character(weight_id), length(weight_id) == 1, c(gene_id, group_id,
weight_id) is.numeric(n_top), length(n_top) == 1, round(n_top) ==
n_top) ng <- nrow(x) nc <- ncol(x) names(ks) <- ks <- unique(groups)
# subset 'n_top' features mgs <- split(mgs, mgs[[group_id]]) mgs
<- lapply(mgs, function(df) o <- order(df[[weight_id]], decreas-
ing = TRUE) n <- ifelse(nrow(df) < n_top, nrow(df), n_top) df[o,
][seq_len(n), ] ) # subset unique features mgs <- lapply(ks, function(k)
g1 <- mgs[[k]][[gene_id]] g2 <- unlist(lapply(mgs[ks != k], '[' ,
gene_id)) mgs[[k]][!g1 ] # W is of dimension (#groups)x(#features)
with W(i,j) # equal to weight if j is marker for i, and ~0 otherwise
W <- vapply(ks, function(k) w <- numeric(ng) + 1e-12 names(w) <-
rownames(x) ws <- mgs[[k]][[weight_id]] w[mgs[[k]][[gene_id]]]
<- ws return(w) , numeric(ng)) # there is no need to initialize
H tp <- paste0("topic_", seq_len(length(ks))) dimnames(W) <-
list(rownames(x), tp) return(W) Filter features from expression matrix
```

---

## Description

Remove undetected features and optionally keep only shared features between the expression matrix and a reference set of features.

## Usage

```
.filter(x, y)
```

## Arguments

|   |                                            |
|---|--------------------------------------------|
| x | Expression matrix to filter                |
| y | Vector of feature names to keep (optional) |

## Details

This function:

- Removes features with zero expression across all samples
- Optionally filters to keep only features present in both datasets
- Ensures a minimum of 10 features remain after filtering

## Value

Filtered expression matrix

data

*Synthetic single-cell, mixture and marker data***Description**

`mockSC/mockSP()` are designed to generate synthetic single-cell and spatial mixture data. These data are not meant to represent biologically meaningful use-cases, but are solely intended for use in examples, for unit-testing, and to demonstrate SPOTlight's general functionality. Finally, `.get_mgs()` implements a statistically naive way to select markers from single-cell data; again, please don't use it in real life.

**Usage**

```
mockSC(ng = 200, nc = 50, nt = 3)
```

```
mockSP(x, ns = 100)
```

```
getMGS(x, n_top = 10)
```

**Arguments**

|                             |                                                                                             |
|-----------------------------|---------------------------------------------------------------------------------------------|
| <code>ng, nc, nt, ns</code> | integer scalar specifying the number of genes, cells, types (groups) and spots to simulate. |
| <code>x</code>              | Single cell experiment object                                                               |
| <code>n_top</code>          | integer specifying the number of marker genes to extract for each cluster.                  |

**Value**

- `mockSC` returns a `SingleCellExperiment` with rows = genes, columns = single cells, and cell metadata (`colData`) column type containing group identifiers.
- `mockSP` returns a `SingleCellExperiment` with rows = genes, columns = single cells, and cell metadata (`colData`) column type containing group identifiers.
- `getMGS` returns a `data.frame` with `nt*n_top` rows and 3 columns: gene and type (group) identifier, as well as the gene's weight = the proportion of counts accounted for by that type.

**Examples**

```
sce <- mockSC()
spe <- mockSP(sce)
mgs <- getMGS(sce)
```

---

plotCorrelationMatrix *Plot Correlation Matrix*


---

**Description**

This function takes in a matrix with the predicted proportions for each spot and returns a correlation matrix between cell types.

**Usage**

```
plotCorrelationMatrix(
  x,
  cor.method = c("pearson", "kendall", "spearman"),
  insig = c("blank", "pch"),
  colors = c("#6D9EC1", "white", "#E46726"),
  hc.order = TRUE,
  p.mat = TRUE,
  ...
)
```

**Arguments**

|                         |                                                                                                                                                                                                                                                                                      |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>x</code>          | numeric matrix with rows = samples and columns = cell types Must have at least two rows and two columns.                                                                                                                                                                             |
| <code>cor.method</code> | Method to use for correlation: <code>c("pearson", "kendall", "spearman")</code> . By default <code>pearson</code> .                                                                                                                                                                  |
| <code>insig</code>      | character, specialized insignificant correlation coefficients, <code>"pch"</code> , <code>"blank"</code> (default). If <code>"blank"</code> , wipe away the corresponding glyphs; if <code>"pch"</code> , add characters (see <code>pch</code> for details) on corresponding glyphs. |
| <code>colors</code>     | character vector with three colors indicating the lower, mid, and high color. By default <code>c("#6D9EC1", "white", "#E46726")</code> .                                                                                                                                             |
| <code>hc.order</code>   | logical value. If <code>TRUE</code> , correlation matrix will be <code>hc.ordered</code> using <code>hclust</code> function.                                                                                                                                                         |
| <code>p.mat</code>      | logical value. If <code>TRUE</code> (default), correlation significance will be used. If <code>FALSE</code> arguments <code>sig.level</code> , <code>insig</code> , <code>pch</code> , <code>pch.col</code> , <code>pch.cex</code> are invalid.                                      |
| <code>...</code>        | additional graphical parameters passed to <code>ggcorrplot</code> .                                                                                                                                                                                                                  |

**Value**

ggplot object

**Author(s)**

Marc Elosua Bayes & Helena L Crowell

## Examples

```
set.seed(321)
x <- replicate(m <- 25, runif(10, 0, 1))
rownames(x) <- paste0("spot", seq_len(nrow(x)))
colnames(x) <- paste0("type", seq_len(ncol(x)))

# The most basic example
plotCorrelationMatrix(x = x)

# Showing the non-significant correlatinos
plotCorrelationMatrix(x = x, insig = "pch")

# A more elaborated
plotCorrelationMatrix(
  x = x,
  hc.order = FALSE,
  type = "lower",
  outline.col = "lightgrey",
  method = "circle",
  colors = c("#64ccc9", "#b860bd", "#e3345d"))
```

---

plotImage

*Plot JP(E)G/PNG/Raster/RGB images*

---

## Description

This function takes in an image-related object - path to JP(E)G/PNG file, raster object, RGBarray. It returns a ggplot object with the selected image.

## Arguments

|       |                                                                                                                                                                                                                                                                                  |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| x     | A variety of objects can be passed: character string corresponding to an image file path, valid file types are JPG, JPEG and PNG. It can also take as input objects of class raster and RGB arrays. It can also take a SpatialExperiment from which the image will be extracted. |
| slice | Character string indicating which image slice to use when SpatialExperiment objects are passed. By default uses the first slice available.                                                                                                                                       |

## Value

ggplot object

## Author(s)

Marc Elosua Bayes & Helena L Crowell

**Examples**

```
# Filename
path <- file.path(
  system.file(package = "SPOTlight"),
  "extdata/SPOTlight.png")
plotImage(x = path)
# array
png_img <- png::readPNG(path)
plotImage(png_img)
# SpatialExperiment
```

---

|                  |                                |
|------------------|--------------------------------|
| plotInteractions | <i>Plot group interactions</i> |
|------------------|--------------------------------|

---

**Description**

This function takes in a matrix with the predicted proportions for each spot and returns a heatmap which = plotHeatmap or a network graph which = plotNetwork to show which cells are interacting spatially.

**Usage**

```
plotInteractions(
  x,
  which = c("heatmap", "network"),
  metric = c("prop", "jaccard"),
  min_prop = 0,
  ...
)
```

**Arguments**

|                       |                                                                                                                                                                                                                                                                                                                |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>x</code>        | numeric matrix with rows = samples and columns = groups. Must have at least one row and column, and at least two columns.                                                                                                                                                                                      |
| <code>which</code>    | character string specifying the type of visualization: one of "heatmap" or "network".                                                                                                                                                                                                                          |
| <code>metric</code>   | character string specifying which metric to show: one of "prop" or "jaccard".                                                                                                                                                                                                                                  |
| <code>min_prop</code> | scalar specifying the value above which a group is considered to be contributing to a given sample. An interaction between groups <i>i</i> and <i>j</i> is counted for sample <i>s</i> only when both <i>x</i> [ <i>s</i> , <i>i</i> ] and <i>x</i> [ <i>s</i> , <i>j</i> ] fall above <code>min_prop</code> . |
| <code>...</code>      | additional graphical parameters passed to <code>plot.igraph</code> when <code>which = "network"</code> (see <code>?igraph.plotting</code> ).                                                                                                                                                                   |

**Value**

base R plot

**Author(s)**

Marc Elosua Bayes & Helena L Crowell

**Examples**

```
library(ggplot2)
mat <- replicate(8, rnorm(100, runif(1, -1, 1)))
# Basic example
plotInteractions(mat)

### heatmap ###
# This returns a ggplot object that can be modified as such
plotInteractions(mat, which = "heatmap") +
  scale_fill_gradient(low = "#f2e552", high = "#850000") +
  labs(title = "Interaction heatmap", fill = "proportion")

### Network ###
# specify node names
nms <- letters[seq_len(ncol(mat))]
plotInteractions(mat, which = "network", vertex.label = nms)

# or set column names instead
colnames(mat) <- nms
plotInteractions(mat, which = "network")

# pass additional graphical parameters for aesthetics
plotInteractions(mat,
  which = "network",
  edge.color = "cyan",
  vertex.color = "pink",
  vertex.label.font = 2,
  vertex.label.color = "maroon")
```

---

plotSpatialScatterpie *Spatial scatterpie*

---

**Description**

This function takes in the coordinates of the spots and the proportions of the cell types within each spot. It returns a plot where each spot is a piechart showing proportions of the cell type composition.

**Usage**

```
plotSpatialScatterpie(
  x,
  y,
  cell_types = colnames(y),
  img = FALSE,
  slice = NULL,
```



```

    scatterpie_alpha = 1,
    pie_scale = 0.4,
    degrees = NULL,
    axis = NULL,
    ...
)

```

## Arguments

|                               |                                                                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>x</code>                | Object containing the spots coordinates, it can be an object of class <code>SpatialExperiment</code> , <code>dataframe</code> or <code>matrix</code> . For the latter two rownames should have the spot barcodes to match <code>x</code> . If a <code>matrix</code> it has to of dimensions <code>nrow(y) x 2</code> where the columns are the <code>x</code> and <code>y</code> coordinates in that order. |
| <code>y</code>                | <code>Matrix</code> or <code>dataframe</code> containing the deconvoluted spots. rownames need to be the spot barcodes to match to <code>x</code> .                                                                                                                                                                                                                                                         |
| <code>cell_types</code>       | Vector of cell type names to plot. By default uses the column names of <code>y</code> .                                                                                                                                                                                                                                                                                                                     |
| <code>img</code>              | Logical <code>TRUE</code> or <code>FALSE</code> indicating whether to plot the image or not. Objects of classes accepted by <code>plotImage</code> can also be passed and that image will be used. By default <code>FALSE</code> .                                                                                                                                                                          |
| <code>slice</code>            | Character string indicating which slice to plot if <code>img</code> is <code>TRUE</code> . By default uses the first image.                                                                                                                                                                                                                                                                                 |
| <code>scatterpie_alpha</code> | Numeric scalar to set the alpha of the pie charts. By default 1.                                                                                                                                                                                                                                                                                                                                            |
| <code>pie_scale</code>        | Numeric scalar to set the size of the pie charts. By default 0.4.                                                                                                                                                                                                                                                                                                                                           |
| <code>degrees</code>          | From <code>SpatialExperiment rotateImg</code> . For clockwise ( <code>degrees &gt; 0</code> ) and counter-clockwise ( <code>degrees &lt; 0</code> ) rotation. By default <code>NULL</code> .                                                                                                                                                                                                                |
| <code>axis</code>             | From <code>SpatialExperiment mirrorImg</code> . When a <code>SpatialExperiment</code> object is passed as the image return the mirror image. For horizontal ( <code>axis = "h"</code> ) and vertical ( <code>axis = "v"</code> ) mirroring. By default <code>NULL</code> .                                                                                                                                  |
| <code>...</code>              | additional parameters to <code>geom_scatterpie</code>                                                                                                                                                                                                                                                                                                                                                       |

## Value

ggplot object

## Author(s)

Marc Elosua Bayes & Helena L Crowell

## Examples

```

set.seed(321)

# Coordinates
x <- replicate(2, rnorm(100))
rownames(x) <- paste0("spot", seq_len(nrow(x)))
colnames(x) <- c("imagecol", "imagerow")

```

```
# Proportions
y <- replicate(m <- 5, runif(nrow(x), 0, 1))
y <- prop.table(y, 1)

rownames(y) <- paste0("spot", seq_len(nrow(y)))
colnames(y) <- paste0("type", seq_len(ncol(y)))

(plt <- plotSpatialScatterpie(x = x, y = y))
```

---

|                   |                                |
|-------------------|--------------------------------|
| plotTopicProfiles | <i>Plot NMF topic profiles</i> |
|-------------------|--------------------------------|

---

## Description

This function takes in the fitted NMF model and returns the topic profiles learned for each cell `facet = FALSE` or cell type `facet = TRUE`. Ideal training will return all the cell from the same cell type to share a unique topic profile.

## Usage

```
plotTopicProfiles(x, y, facet = FALSE, min_prop = 0.01, ncol = NULL)
```

## Arguments

|                       |                                                                                                                                                                                                                                                                                           |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>x</code>        | list object obtained from SPOTlight.                                                                                                                                                                                                                                                      |
| <code>y</code>        | vector of group labels. Should be of length <code>ncol(res_lv11\$NMF\$h)</code> .                                                                                                                                                                                                         |
| <code>facet</code>    | logical indicating whether to stratify by group. If <code>FALSE</code> (default), weights will be the median across cells for each group (point = topic weight for a given cell type). If <code>TRUE</code> , cell-specific weights will be shown (point = topic weight of a given cell). |
| <code>min_prop</code> | scalar in <code>[0,1]</code> . When <code>facet = TRUE</code> , only cells with a weight <code>&gt; min_prop</code> will be included.                                                                                                                                                     |
| <code>ncol</code>     | integer scalar specifying the number of facet columns.                                                                                                                                                                                                                                    |

## Value

ggplot object

## Author(s)

Marc Elosua Bayes & Helena L Crowell

**Examples**

```

library(ggplot2)
x <- mockSC()
y <- mockSP(x)
z <- getMGS(x)

res <- SPOTlight(x, y,
  groups = x$type,
  mgs = z,
  group_id = "type",
  verbose = FALSE)

plotTopicProfiles(res[[3]], x$type, facet = TRUE)
plotTopicProfiles(res[[3]], x$type, facet = FALSE)

```

runDeconvolution

*Run Deconvolution using NNLS model***Description**

This function takes in the mixture data, the trained model & the topic profiles and returns the proportion of each cell type within each mixture

**Usage**

```

runDeconvolution(
  x,
  mod,
  ref,
  scale = TRUE,
  min_prop = 0.01,
  verbose = TRUE,
  slot = "counts",
  L1_nnls_topics = 0,
  L2_nnls_topics = 0,
  L1_nnls_prop = 0,
  L2_nnls_prop = 0,
  threads = 0,
  ...
)

```

**Arguments**

|     |                                                                                                    |
|-----|----------------------------------------------------------------------------------------------------|
| x   | mixture dataset. Can be a numeric matrix, SingleCellExperiment or SpatialExperiment                |
| mod | object as obtained from trainNMF.                                                                  |
| ref | object of class matrix containing the topic profiles for each cell type as obtained from trainNMF. |

|                              |                                                                                                                                                                                                                                                             |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| scale                        | logical specifying whether to scale single-cell counts to unit variance. This gives the user the option to normalize the data beforehand as you see fit (CPM, FPKM, ...) when passing a matrix or specifying the slot from where to extract the count data. |
| min_prop                     | scalar in [0,1] setting the minimum contribution expected from a cell type in x to observations in y. By default 0.                                                                                                                                         |
| verbose                      | logical. Should information on progress be reported?                                                                                                                                                                                                        |
| slot                         | If the object is of class SpatialExperiment indicates matrix to use. By default "counts".                                                                                                                                                                   |
| L1_nnls_topics, L1_nnls_prop | LASSO penalty in the range (0, 1] for NNLS when computing cell type topic profiles and cell type proportions respectively. Larger values remove "noisy" contributions more aggressively.                                                                    |
| L2_nnls_topics, L2_nnls_prop | RIDGE penalty >0 for NNLS when computing cell type topic profiles and cell type proportions respectively. Larger values remove "noisy" contributions more aggressively.                                                                                     |
| threads                      | number of threads to use, default 0 (all threads)                                                                                                                                                                                                           |
| ...                          | additional parameters.                                                                                                                                                                                                                                      |

**Value**

base a list where the first element is a list giving the NMF model and the second is a matrix containing the topic profiles learnt.

**Author(s)**

Marc Elosua Bayes, Zach DeBruine, and Helena L Crowell

**Examples**

```
set.seed(321)
# mock up some single-cell, mixture & marker data
sce <- mockSC(ng = 200, nc = 10, nt = 3)
spe <- mockSP(sce)
mgs <- getMGS(sce)

res <- trainNMF(
  x = sce,
  y = rownames(spe),
  groups = sce$type,
  mgs = mgs,
  weight_id = "weight",
  group_id = "type",
  gene_id = "gene")
# Run deconvolution
decon <- runDeconvolution(
  x = spe,
  mod = res[["mod"]],
  ref = res[["topic"]])
```

---

SPOTlight

*Deconvolution of mixture using single-cell data*


---

## Description

This is the backbone function which takes in single cell expression data to deconvolute spatial transcriptomics spots.

## Usage

```
SPOTlight(
  x,
  y,
  groups = NULL,
  mgs,
  n_top = NULL,
  gene_id = "gene",
  group_id = "cluster",
  weight_id = "weight",
  hvg = NULL,
  scale = TRUE,
  min_prop = 0.01,
  verbose = TRUE,
  slot_sc = "counts",
  slot_sp = "counts",
  L1_nmf = 0,
  L2_nmf = 0,
  maxit = 100,
  threads = 0,
  tol = 1e-05,
  L1_nnlst_topics = 0,
  L2_nnlst_topics = 0,
  L1_nnlst_prop = 0,
  L2_nnlst_prop = 0,
  ...
)
```

## Arguments

|                     |                                                                                                                                                                                                                                                                            |
|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>x, y</code>   | single-cell and mixture dataset, respectively. Can be a numeric matrix or <code>SingleCellExperiment</code> ..                                                                                                                                                             |
| <code>groups</code> | character vector of group labels for cells in <code>x</code> . When <code>x</code> is a <code>SingleCellExperiment</code> ., defaults to <code>colLabels(x)</code> and <code>Idents(x)</code> , respectively. Make sure <code>groups</code> is not a <code>Factor</code> . |
| <code>mgs</code>    | <code>data.frame</code> or <code>DataFrame</code> of marker genes. Must contain columns holding gene identifiers, group labels and the weight (e.g., <code>logFC</code> , <code>-log(p-value)</code> ) a feature has in a given group.                                     |

|                                           |                                                                                                                                                                                                                                                                          |
|-------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>n_top</code>                        | integer scalar specifying the number of markers to select per group. By default NULL uses all the marker genes to initialize the model.                                                                                                                                  |
| <code>gene_id, group_id, weight_id</code> | character specifying the column in <code>mgs</code> containing gene identifiers, group labels and weights, respectively.                                                                                                                                                 |
| <code>hvg</code>                          | character vector containing <code>hvg</code> to include in the model. By default NULL.                                                                                                                                                                                   |
| <code>scale</code>                        | logical specifying whether to scale single-cell counts to unit variance. This gives the user the option to normalize the data beforehand as you see <code>fit</code> (CPM, FPKM, ...) when passing a matrix or specifying the slot from where to extract the count data. |
| <code>min_prop</code>                     | scalar in $[0,1]$ setting the minimum contribution expected from a cell type in <code>x</code> to observations in <code>y</code> . By default 0.                                                                                                                         |
| <code>verbose</code>                      | logical. Should information on progress be reported?                                                                                                                                                                                                                     |
| <code>slot_sc, slot_sp</code>             | If the object is of class <code>SingleCellExperiment</code> indicates matrix to use. By default "counts".                                                                                                                                                                |
| <code>L1_nmf</code>                       | LASSO penalty in the range $(0, 1]$ for NMF, larger values increase sparsity of each factor                                                                                                                                                                              |
| <code>L2_nmf</code>                       | RUDGE penalty $>0$ for NMF, larger values increase angle between factors and thus sparsity.                                                                                                                                                                              |
| <code>maxit</code>                        | maximum number of NMF iterations for fitting                                                                                                                                                                                                                             |
| <code>threads</code>                      | number of threads to use, default 0 (all threads)                                                                                                                                                                                                                        |
| <code>tol</code>                          | tolerance of the NMF model at convergence, the Pearson correlation distance between models across consecutive iterations ( $1e-5$ is publication quality)                                                                                                                |
| <code>L1_nnls_topics, L1_nnls_prop</code> | LASSO penalty in the range $(0, 1]$ for NNLS when computing cell type topic profiles and cell type proportions respectively. Larger values remove "noisy" contributions more aggressively.                                                                               |
| <code>L2_nnls_topics, L2_nnls_prop</code> | RIDGE penalty $>0$ for NNLS when computing cell type topic profiles and cell type proportions respectively. Larger values remove "noisy" contributions more aggressively.                                                                                                |
| <code>...</code>                          | additional parameters.                                                                                                                                                                                                                                                   |

## Details

SPOTlight uses a Non-Negative Matrix Factorization approach to learn which genes are important for each cell type. In order to drive the factorization and give more importance to cell type marker genes we previously compute them and use them to initialize the basis matrix. This initialized matrices will then be used to carry out the factorization with the single cell expression data. Once the model has learn the topic profiles for each cell type we use non-negative least squares (NNLS) to obtain the topic contributions to each spot. Lastly, NNLS is again used to obtain the proportion of each cell type for each spot by finding the fitting the single-cell topic profiles to the spots topic contributions.

**Value**

a numeric matrix with rows corresponding to samples and columns to groups

**Author(s)**

Marc Elosua Bayes, Zach DeBruine, and Helena L Crowell

**Examples**

```
library(scater)
library(scraper)

# Use Mock data
# Refer to the vignette for a full workflow
sce <- mockSC(ng = 200, nc = 10, nt = 3)
spe <- mockSP(sce)
mgs <- getMGS(sce)

res <- SPOTlight(
  x = counts(sce),
  y = counts(spe),
  groups = as.character(sce$type),
  mgs = mgs,
  hvg = NULL,
  weight_id = "weight",
  group_id = "type",
  gene_id = "gene")
```

---

trainNMF

*train NMF model*


---

**Description**

This is the training function used by SPOTLight. This function takes in single cell expression data, trains the model and learns topic profiles for each cell type

**Usage**

```
trainNMF(
  x,
  y = NULL,
  groups = NULL,
  mgs,
  n_top = NULL,
  gene_id = "gene",
  group_id = "cluster",
  weight_id = "weight",
  hvg = NULL,
```

```

    scale = TRUE,
    verbose = TRUE,
    L1_nmf = 0,
    L2_nmf = 0,
    tol = 1e-05,
    maxit = 100,
    threads = 0,
    slot_sc = "counts",
    ...
)

```

### Arguments

|                              |                                                                                                                                                                                                                                                             |
|------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| x                            | single-cell dataset. Can be a numeric matrix, Can be a numeric matrix or SingleCellExperiment.                                                                                                                                                              |
| y                            | Null if you want to train the model with all the genes in the SC data or a character vector with the rownames of the mixture dataset to subset the gene set used to the intersection between them.                                                          |
| groups                       | character vector of group labels for cells in x. When x is a SingleCellExperiment., defaults to colLabels(x) and Idents(x), respectively. Make sure groups is not a Factor.                                                                                 |
| mgs                          | data.frame or DataFrame of marker genes. Must contain columns holding gene identifiers, group labels and the weight (e.g., logFC, -log(p-value) a feature has in a given group.                                                                             |
| n_top                        | integer scalar specifying the number of markers to select per group. By default NULL uses all the marker genes to initialize the model.                                                                                                                     |
| gene_id, group_id, weight_id | character specifying the column in mgs containing gene identifiers, group labels and weights, respectively.                                                                                                                                                 |
| hvg                          | character vector containing hvg to include in the model. By default NULL.                                                                                                                                                                                   |
| scale                        | logical specifying whether to scale single-cell counts to unit variance. This gives the user the option to normalize the data beforehand as you see fit (CPM, FPKM, ...) when passing a matrix or specifying the slot from where to extract the count data. |
| verbose                      | logical. Should information on progress be reported?                                                                                                                                                                                                        |
| L1_nmf                       | LASSO penalty in the range (0, 1] for NMF, larger values increase sparsity of each factor                                                                                                                                                                   |
| L2_nmf                       | RUDGE penalty >0 for NMF, larger values increase angle between factors and thus sparsity.                                                                                                                                                                   |
| tol                          | tolerance of the NMF model at convergence, the Pearson correlation distance between models across consecutive iterations (1e-5 is publication quality)                                                                                                      |
| maxit                        | maximum number of NMF iterations for fitting                                                                                                                                                                                                                |
| threads                      | number of threads to use, default 0 (all threads)                                                                                                                                                                                                           |
| slot_sc                      | If the object is of class SingleCellExperiment indicates matrix to use. By default "counts".                                                                                                                                                                |
| ...                          | additional parameters.                                                                                                                                                                                                                                      |



**Value**

a list where the first element is a list with the NMF model information and the second is a matrix containing the topic profiles learnt per cell type.

**Author(s)**

Marc Elosua Bayes & Helena L Crowell

**Examples**

```
set.seed(321)
# mock up some single-cell, mixture & marker data
sce <- mockSC(ng = 200, nc = 10, nt = 3)
spe <- mockSP(sce)
mgs <- getMGS(sce)

res <- trainNMF(
  x = sce,
  y = rownames(spe),
  groups = sce$type,
  mgs = mgs,
  weight_id = "weight",
  group_id = "type",
  gene_id = "gene")
# Get NMF model
res[["mod"]]
# Get topic profiles
res[["topic"]]
```

# Index

`.filter`, [2](#)

`data`, [4](#)

`getMGS (data)`, [4](#)

`mockSC (data)`, [4](#)

`mockSP (data)`, [4](#)

`plotCorrelationMatrix`, [5](#)

`plotHeatmap (plotInteractions)`, [7](#)

`plotImage`, [6](#)

`plotInteractions`, [7](#)

`plotNetwork (plotInteractions)`, [7](#)

`plotSpatialScatterpie`, [8](#)

`plotTopicProfiles`, [10](#)

`runDeconvolution`, [11](#)

`SPOTlight`, [13](#)

`trainNMF`, [15](#)