Package ‘SlideCNA’

January 23, 2025

Title Calls Copy Number Alterations from Slide-Seq Data

Version 0.1.0

Description This takes spatial single-cell-type RNA-seq data (specifically designed for Slide-seq v2) that calls copy number alterations (CNAs) using pseudo-spatial binning, clusters cellular units (e.g. beads) based on CNA profile, and visualizes spatial CNA patterns. Documentation about 'SlideCNA' is included in the the preprint by Zhang et al. (2022, <doi:10.1101/2022.11.25.517982>). The package 'enrichR' (>= 3.0), conditionally used to annotate SlideCNA-determined clusters with gene ontology terms, can be installed at <`https://github.com/wjawaid/enrichR`> or with install_github(``wjawaid/enrichR'').

Imports data.table, reshape2, dplyr, ggplot2, scales, pheatmap, cluster, factoextra, dendextend, Seurat, tidyselect, stringr, magrittr, tibble, futile.logger, mltools, utils

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0), enrichR (>= 3.0)

Config/testthat/edition 3

NeedsCompilation no

Author Diane Zhang [aut, cre] (<`https://orcid.org/0000-0002-0268-4413`>),
Johanna Klughammer [aut] (<`https://orcid.org/0000-0002-3628-9278`>),
Jan Watter [aut] (<`https://orcid.org/0000-0001-7519-731X`>),
Broad Institute of MIT and Harvard [cph, fnd]

Maintainer Diane Zhang <`dkzhang711@gmail.com`>

Repository CRAN

Date/Publication 2025-01-23 17:40:05 UTC

Contents

`bin` *Subfunction of bin_metadata() for expression/positional binning*

Description

This function computes a pseudospatial distance between beads that combines spatial distance and distance from the expression space, then using the silhouette score and hierarchical clustering, segregates beads into bins

Usage

```
bin(
  dat,
  md,
  k,
  pos = TRUE,
  pos_k = 55,
  ex_k = 1,
  hc_function = "ward.D2",
  plot_directory
)
```

Arguments

`dat`	data.table of smoothed relative expression intensities
`md`	data.table of metadata of each bead
`k`	number of malignant bins to set
`pos`	TRUE if doing spatial and expressional binning, FALSE if just expressional binning
`pos_k`	positional weight
`ex_k`	expressional weight
`hc_function`	hierarchical clustering function
`plot_directory`	output plot directory path

Value

A data.table of bead metadata combined with bin designations

`bin_metadata` *Spatio-molecular binning of relative expression intensities*

Description

This function combines metadata with binned relative expression intensities

Usage

```
bin_metadata(
  md,
  dat,
  avg_bead_per_bin = 12,
  pos = TRUE,
  pos_k = 55,
  ex_k = 1,
  hc_function = "ward.D2",
  plot_directory
)
```

Arguments

`md`	data.table of metadata of each bead
`dat`	data.table of smoothed relative expression intensities
`avg_bead_per_bin`	integer of average number of beads there should be per bin
`pos`	TRUE if doing spatial and expressional binning, FALSE if just expressional binning
`pos_k`	positional weight

`ex_k`	expressional weight
`hc_function`	hierarchical clustering function
`plot_directory`	output plot directory path

Value

A data.table of bead metadata combined with binned expression intensities for all genes for all beads

---

`center_rm` *Center expression intensities*

Description

Take in a data.table of genomic positions and smoothed expression intensities counts and center by subtracting average intensity across all beads for each gene

Usage

```
center_rm(rm)
```

Arguments

`rm`	data.table of smoothed expression intensities counts

Value

centered_rm data.table of smoothed, centered expression intensities

---

`clone_so` *Add clone information to meta data of seurat object and bin the beads*

Description

This function adds another column for cluster designation to a seurat object's meta data and bins beads

Usage

```
clone_so(so, hcl_sub, md, mal = FALSE)
```

Arguments

`so`	Seurat object of beads and their meta data
`hcl_sub`	hierarchical clustering object of cluster assignemnt as outputted from SlideCNA::plot_clones()
`md`	data.table of metadata of each bead
`mal`	TRUE if only using malignant beads

Value

A seurat object updated with clone information

`cnv_heatmap` *Plot CNV scores on a heat map*

Description

This function prepares data for plotting and makes a heat map of CNV scores per bead across all genes

Usage

```
cnv_heatmap(
  cnv_data,
  md,
  chrom_colors,
  hc_function = "ward.D2",
  plot_directory
)
```

Arguments

`cnv_data`	list object of cnv data from SlideCNA::prep_cnv_dat()
`md`	data.table of metadata of each bead
`chrom_colors`	vector of colors labeled by which chromosome they correspond to
`hc_function`	character for which hierarchical clustering function to use
`plot_directory`	output plot directory path

Value

None

`dat_to_long` *Convert data to long format and add in metadata*

Description

This function will create rows for each bead and gene combination, adding in new metadata with bin designations

Usage

```
dat_to_long(dat, md)
```

Arguments

`dat`	data.table of smoothed relative expression intensities
`md`	data.table of metadata per bead

Value

A data.table of bead expression intensities per gene with metadata in long format

---

`find_cluster_markers` *Find and plot top n DEGs per cluster*

Description

This function uses Seurat's marker finding capability to find DEGs of each cluster

Usage

```
find_cluster_markers(
  so_clone,
  type,
  logfc.threshold = 0.2,
  min.pct = 0,
  only.pos = TRUE,
  n_markers = 5,
  value = "log2_expr",
  text_size = 16,
  title_size = 18,
  legend_size_pt = 4,
  p_val_thresh = 0.05,
  bin = TRUE,
  plot_directory = None
)
```

Arguments

`so_clone`	seurat object with 'clone' (SlideCNA-designated cluster) and bin annotations
`type`	character string that is 'all' if using malignant and normal clusters and 'malig' if just using malignant clusters
`logfc.threshold`	numeric float that is seurat parameter, representing the minimum log2 fold change for DEGs to be significant
`min.pct`	numeric Seurat function parameter
`only.pos`	TRUE if only using DEGs with positive log2 fold change
`n_markers`	integer of number of top DEGs to plot/use

`value`	expression value of DEGs; one of ("log2_expr", "avg_expr", and "avg_log2FC") for log2-normalized aerage epxression, average expression, or log2 fold change
`text_size`	Ggplot2 text size
`title_size`	Ggplot2 title size
`legend_size_pt`	Ggplot2 legend_size_pt
`p_val_thresh`	value for p value cutoff for DEGs
`bin`	TRUE if using binned beads
`plot_directory`	output plot directory path

Value

A list object with cluster marker information markers_clone = data.table of all cluster markers top_markers_clone = data.table of just top cluster markers top_clone_vis = data.frame formatted for plot visualization of top cluster markers

---

`find_go_terms` *Find and plot top n GO-enriched terms per cluster*

Description

This function utilizes cluster-specific DEGs to identify cluster-specifc GO biological processes and plots these if they occur

Usage

```
find_go_terms(
  cluster_markers_obj,
  type,
  n_terms = 5,
  text_size,
  title_size,
  plot_directory
)
```

Arguments

`cluster_markers_obj`	list object with cluster marker information
`type`	character string that is 'all' if using malignant and normal clusters and 'malig' if just using malignant clusters
`n_terms`	integer of number of top DEGs to plot/use
`text_size`	integer of text size for ggplot
`title_size`	integer of title size for ggplot
`plot_directory`	output plot directory path

Value

A list object with cluster GO term information en_clone = data.table of cluster GO terms top_en_clone = data.table of just top cluster GO terms

---

`get_num_clust` *Find optimal number of clusters*

Description

This function uses the Silhouette Method applied to CNV scores to determine the best number of clusters to divide the binned beads into

Usage

```
get_num_clust(
  data,
  hc_func = "ward.D2",
  max_k = 10,
  plot = TRUE,
  malig = FALSE,
  k = NA,
  plot_directory
)
```

Arguments

`data`	cnv_data list object of cnv data from SlideCNA::prep_cnv_dat()
`hc_func`	character string for which hierarchical clustering function to use
`max_k`	integer of number max number of clusters to evaluate (2:max_k)
`plot`	TRUE if plotting silhoutte scores per cluster
`malig`	TRUE if only using malignant bins and FALSE if using all bins
`k`	integer of optimal number of clusters, if known, and NA if not known
`plot_directory`	output plot directory path

Value

An integer representing the number of clusters that optimizes the silhouette score

---

`long_to_bin` *Convert to wide bin x genes + metadata format*

Description

This function will combine beads into bins, taking the average expression intensities, average positions, most common cluster seurat cluster, and most common cluster/tissue type of constituent beads

Usage

```
long_to_bin(dat_long, plot_directory, spatial = TRUE)
```

Arguments

`dat_long`	data.table of bead expression intensities per gene with metadata in long format
`plot_directory`	output plot directory path
`spatial`	True if using spatial information

Value

data.table of expression intensities at aggregated bin level

`make_seurat_annot` *Creation of Seurat object*

Description

This function takes in raw counts (and potentially meta data) to make a Seurat object and process it

Usage

```
make_seurat_annot(
  cb,
  md = NULL,
  seed_FindClusters = 0,
  seed_RunTSNE = 1,
  seed_RunUMAP = 42
)
```

Arguments

`cb`	sparse counts matrix (genes x cells/beads)
`md`	data.frame of meta data for cells/beads if specific annotations known
`seed_FindClusters`	seed number for FindCLusters
`seed_RunTSNE`	seed number for RunTSNE
`seed_RunUMAP`	seed number for RunUMAP

Value

A Seurat object with specific Seurat features run

`make_so_bin` *Make a binned version of a Seurat object*

Description

Aggregate Seurat object counts by bin to create a new Seurat object with binned beads as units instead of beads

Usage

```
make_so_bin(so, md, hcl_sub, mal = FALSE)
```

Arguments

`so`	Seurat object of beads and their meta data
`md`	data.frame of metadata for Seurat object
`hcl_sub`	hierarchical clustering object of cluster assignemnt as outputted from SlideCNA::plot_clones()
`mal`	TRUE if using malignant beads only

Value

A Seurat object with binned beads as units and corresponding binned metadata

---

`mean_cnv_plot` *Plot mean CNV scores per bin and per chromosome*

Description

This function colors and plots each bin by its mean CNV score on spatial coordinates for each chromosome

Usage

```
mean_cnv_plot(
  cnv_data,
  text_size,
  title_size,
  legend_height_bar,
  plot_directory
)
```

Arguments

`cnv_data`	list object of cnv data from SlideCNA::prep_cnv_dat()
`text_size`	integer of text size for ggplot
`title_size`	integer of title size for ggplot
`legend_height_bar`	integer of bar height of legend for ggplot
`plot_directory`	output plot directory path

Value

None

`mode` *Subfunction of long_to_bin() that finds mode of vector/column*

Description

This function finds the mode of a vector

Usage

```
mode(x)
```

Arguments

`x`	vector (column in data.table) to calculate the mode from

Value

mode of the vector

`plot_clones` *Plot cluster/clone information*

Description

This function plots cluster dendrograms, spatial assignment, and the CNV heat map

Usage

```
plot_clones(
  cnv_data,
  md,
  k,
  type,
  chrom_colors,
  text_size,
  title_size,
  legend_size_pt,
  legend_height_bar,
  hc_function = "ward.D2",
  plot_directory,
  spatial = TRUE
)
```

Arguments

`cnv_data`	list object of cnv data from SlideCNA::prep_cnv_dat()
`md`	data.table of metadata of each bead
`k`	integer of number of clusters/clones
`type`	character string, being "all" if using all binned beads, or "malig" if just malignant binned beads
`chrom_colors`	vector of colors labeled by which chromosome they correspond to
`text_size`	Ggplot2 text size
`title_size`	Ggplot2 title size
`legend_size_pt`	Ggplot2 legend_size_pt
`legend_height_bar`	Ggplot2 legend_height_bar
`hc_function`	character string for which hierarchical clustering function to use
`plot_directory`	output plot directory path
`spatial`	TRUE if using spatial information

Value

A hierarchical clustering object of the clusters

`prep` *Infercnv-based preparation of relative gene expression intensities*

Description

This function takes in a data table of raw counts and a vector of reference/normal beads to normalize counts and adjust for reference expression.

Usage

```
prep(so, normal_beads, gene_pos, chrom_ord, logTPM = FALSE)
```

Arguments

`so`	Seurat object of Slide-seq data with raw counts
`normal_beads`	vector of names of normal beads
`gene_pos`	data.table with columns for GENE, chr, start, end, rel_gene_pos (1 : # of genes on chromosome)
`chrom_ord`	vector of the names of chromosomes in order
`logTPM`	TRUE if performing adjustment with logTPM

Value

A data.table of normalized, capped, and ref-adjusted counts with genomic psoition info

---

`prep_cnv_dat` *Prepare data for CNV heat map*

Description

This function caps CNV scores, adds annotation columns for plotting, performs hierarchical clustering of bins based on similar CNV score, and plots nUMI per bin

Usage

```
prep_cnv_dat(
  dat_bin,
  lower = 0.6,
  upper = 1.4,
  hc_function = "ward.D2",
  plot_directory
)
```

Arguments

`dat_bin`	data.table of CNV scores per bin
`lower`	numeric float to represent the lower cap for CNV scores
`upper`	numeric float to represent the upper cap for CNV scores
`hc_function`	character for which hierarchical clustering function to use
`plot_directory`	output plot directory path

Value

A list object for downstream cnv plotting and analysis all = data.table of CNV scores of all bins x (metadata + genes) malig = data.table of CNV scores of just malignant bins x (metadata + genes) all_wide = data.frame in wide format of CNV scores of all bins x (metadata + genes) malig_wide = data.frame in wide format of CNV scores of just malignant bins x (metadata + genes) hcl = hclust object that describes the hierarchical clustering for malignant bins hcl_all = hclust object that describes the hierarchical clustering for all bins

---

`quantile_plot` *Plot CNV score quantiles per bin and per chromosome*

Description

This function colors and plots each bin by its CNV score quantiles (min, 1st quartile, median, 3rd quartile, max) on spatial coordinates for each chromosome

Usage

```
quantile_plot(
  cnv_data,
  cluster_label = "seurat_clusters",
  text_size,
  title_size,
  legend_height_bar,
  plot_directory
)
```

Arguments

`cnv_data`	list object of cnv data from SlideCNA::prep_cnv_dat()
`cluster_label`	character string of which column name to keep
`text_size`	integer of text size for ggplot
`title_size`	integer of title size for ggplot
`legend_height_bar`	integer of bar height of legend for ggplot
`plot_directory`	output plot directory path

Value

None

`ref_adj` *Adjust for Reference (Normal) Beads*

Description

Take in a data.table of genomic positions and smoothed, centered expression intensities counts and adjust for reference beads by subtracting average intensities of reference beads for each gene. This is the second reference adjustment.

Usage

```
ref_adj(centered_rm, normal_beads)
```

Arguments

`centered_rm`	data.table of smoothed, centered expression intensities counts
`normal_beads`	vector of names of normal beads

Value

rm_adj data.table of smoothed relative expression intensities

---

`run_enrichr` *Subfunction to get significantly enriched GO terms given a set of signfiicant beads and genes*

Description

This function finds the GO biological processes associated with the top n genes using enrichR

Usage

```
run_enrichr(genes, n_genes)
```

Arguments

`genes`	vector of differentially expressed genes
`n_genes`	number of the most significantly enriched DEGs to base gene enrichment from

Value

A data.table of the most significant GO terms and their meta data

`run_slide_cna` *Run SlideCNA workflow*

Description

Take a raw expression counts, cell type annotations, and positional cooridnates to identify CNA patterns across space and CNA-based clustering patterns

Usage

```
run_slide_cna(
  counts,
  beads_df,
  gene_pos,
  output_directory,
  plot_directory,
  spatial = TRUE,
  roll_mean_window = 101,
  avg_bead_per_bin = 12,
  pos = TRUE,
  pos_k = 55,
  ex_k = 1,
  hc_function_bin = "ward.D2",
  spatial_vars_to_plot = c("seurat_clusters", "bin_all", "N_bin", "umi_bin",
    "cluster_type"),
  scale_bin_thresh_hard = TRUE,
  lower_bound_cnv = 0.6,
  upper_bound_cnv = 1.4,
  hc_function_cnv = "ward.D2",
  hc_function_cnv_heatmap = "ward.D2",
  quantile_plot_cluster_label = "seurat_clusters",
  hc_function_silhouette = "ward.D2",
  max_k_silhouette = 10,
  plot_silhouette = TRUE,
  hc_function_plot_clones = "ward.D2",
  use_GO_terms = TRUE,
  chrom_ord = c("chr1", "chr2", "chr3", "chr4", "chr5", "chr6", "chr7", "chr8", "chr9",
   "chr10", "chr11", "chr12", "chr13", "chr14", "chr15", "chr16", "chr17", "chr18",
    "chr19", "chr20", "chr21", "chr22", "chr23", "chrX", "chrY", "chrM"),
  chrom_colors = c(chr1 = "#8DD3C7", chr2 = "#FFFFB3", chr3 = "#BEBADA", chr4 =
   "#FB8072", chr5 = "#80B1D3", chr6 = "#FDB462", chr7 = "#B3DE69", chr8 = "#FCCDE5",
   chr9 = "#D9D9D9", chr10 = "#BC80BD", chr11 = "#CCEBC5", chr12 = "#FFED6F", chr13 =
    "#1B9E77", chr14 = "#D95F02", chr15 = "#7570B3", chr16 = "#E7298A", chr17 =
    "#66A61E", chr18 = "#E6AB02", chr19 = "#A6761D", chr20 = "#666666", chr21 =
    "#A6CEE3", chr22 = "#1F78B4", chrX = "#B2DF8A"),
  text_size = 16,
  title_size = 18,
```

```
  legend_size_pt = 4,
  legend_height_bar = 1.5
)
```

Arguments

`counts`	data.frame of raw counts (genes x beads)
`beads_df`	data.frame of annotation of each bead (beads x annotations); contains columns 'bc' for bead names, 'cluster_type' for annotations of 'Normal' or 'Malignant', 'pos_x' for x-coordinate bead positions, and 'pos_y' for y-coordinate bead positions
`gene_pos`	data.frame with columns for GENE, chr, start, end, rel_gene_pos (1 : # of genes on chromosome)
`output_directory`	output directory path
`plot_directory`	output plot directory path
`spatial`	TRUE if using spatial information FALSE if not
`roll_mean_window`	integer number of adjacent genes for which to average over in pyramidal weighting scheme
`avg_bead_per_bin`	integer of average number of beads there should be per bin
`pos`	TRUE if doing spatial and expressional binning, FALSE if just expressional binning
`pos_k`	positional weight
`ex_k`	expressional weight
`hc_function_bin`	hierarchical clustering function for binning; to feed hclust's method argument, one of "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid"
`spatial_vars_to_plot`	character vector of features to plot/columns of metadata
`scale_bin_thresh_hard`	TRUE if using strict thresholds for expression thresholds and FALSE if adjusting thresholds based on 1 + or - the mean of absolute min and max vlaues
`lower_bound_cnv`	numeric float to represent the lower cap for CNV scores
`upper_bound_cnv`	numeric float to represent the upper cap for CNV scores
`hc_function_cnv`	character for which hierarchical clustering function to use for CNV-calling; to feed hclust's method argument, one of "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid"

`hc_function_cnv_heatmap`	character for which hierarchical clustering function to use for visualzing CNV heat map; to feed hclust's method argument, one of "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid"
`quantile_plot_cluster_label`	character string of which column name to keep in quantile plot
`hc_function_silhouette`	character string for which hierarchical clustering function to use for the Silhouette method; to feed hclust's method argument, one of "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median" or "centroid"
`max_k_silhouette`	integer of number max number of clusters to evaluate (2:max_k_silhouette) . in Silhouette method
`plot_silhouette`	TRUE if plotting silhouette scores for clustering
`hc_function_plot_clones`	character string for which hierarchical clustering function to use in plotting clones
`use_GO_terms`	TRUE if using enrichR to get Gene Ontology terms for SlideCNA-defined clusters
`chrom_ord`	character vector of order and names of chromosomes
`chrom_colors`	character vector of which colors each chromosome should be in heat map
`text_size`	integer of size of text in some ggplots
`title_size`	integer of size of title in some ggplots
`legend_size_pt`	integer of size of legend text size in some ggplots
`legend_height_bar`	integer of height of legend bar in some ggplots

Value

None

`scalefit` *Subfunction for scale_nUMI that normalizes a given bin for UMI count and centers the mean CNV score at 1*

Description

This function re-scales expression intensities to be in a smaller range, normalizes for nUMI per bin, and centers the CNV scores to have a mean of 1

Usage

```
scalefit(obj, nbin, start, end)
```

Arguments

`obj`	data.table of relative expression intensities per bin
`nbin`	nUMIs in that specific bin
`start`	lower bound of CNV scores
`end`	upper bound of CNV scores

Value

vector of adjusted CNV scores for that bins with nbin number of nUMIs within the range (inclusive) of start to end

---

`scale_nUMI` *Scale for nUMI (UMI Count) to generate CNV scores*

Description

This function re-scales expression intensities to be in a smaller range, normalizes for nUMI per bin, and subtracts reference bead signal

Usage

```
scale_nUMI(dat_bin, thresh_hard = FALSE)
```

Arguments

`dat_bin`	data.table of relative expression intensities per bin
`thresh_hard`	TRUE if using strict thresholds for expression thresholds and FALSE if adjusting thresholds based on 1 + or - the mean of absolute min and max values

Value

data.table of CNV scores per bin

`SpatialPlot` *Spatial plots of meta data*

Description

This function will plot information about beads and bins on x and y coordinates

Usage

```
SpatialPlot(
  dat_long,
  vars = NULL,
  text_size,
  title_size,
  legend_size_pt,
  legend_height_bar,
  plot_directory
)
```

Arguments

`dat_long`	data.table of bead expression intensities per gene with metadata in long format
`vars`	character vector of features to plot/columns of metadata
`text_size`	Ggplot2 text size
`title_size`	Ggplot2 title size
`legend_size_pt`	Ggplot2 legend_size_pt
`legend_height_bar`	Ggplot2 legend_height_bar
`plot_directory`	output plot directory path

Value

None

`weight_rollmean` *Expressional smoothing along a chromosome using a weighted pyramidal moving average*

Description

Take in a data.table of genomic positions and bead normalized/modified counts and apply pyramidal weighting with a window size k to create smoothed expression intensities

Usage

```
weight_rollmean(dat, k = 101)
```

Arguments

`dat`	data.table of normalized/adjusted counts
`k`	size of window for weighting

Value

A data.table of expression intensities

`weight_rollmean_sub` *Subfunction of weight_rollmean*

Description

Take in a counts matrix and apply pyramidal weighting with a window size k to create smoothed expression intensities

Usage

```
weight_rollmean_sub(mat, k)
```

Arguments

`mat`	matrix of normalized/adjusted counts
`k`	size of window for weighting

Value

A matrix of smoothed counts

Index