Package 'archipelago'

May 9, 2025

Title Archipelago plot method for illustration of variant set association test statistics

Version 0.1.9000

Author Dylan Lawless [aut, cre] (<<https://orcid.org/0000-0001-8496-3725>>)

Maintainer Dylan Lawless <dylan.lawless@protonmail.com>

Description Archipelago plot, a graphical method for interpreting both VSAT p-values and individual variant contributions. The Archipelago method assigns a meaningful genomic coordinate to the VSAT p- value, enabling its simultaneous visualization alongside individual variant p-values. This results in an intuitive and rich illustration akin to an archipelago of clustered islands, enhancing the understanding of both collective and individual impacts of variants. The Archipelago plot is applicable in any genetic association study that uses variant collapse to evaluate both individual variants and variant sets, and its customizability facilitates clear communication of complex genetic data. By integrating two dimensions of genetic data into a single visualization, VSAT results can be easily read and aid in identification of potential causal variants in variant sets such as protein pathways.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Imports ggplot2, dplyr

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Depends R (>= 2.10)

LazyData true

Contents

---

archipelago_plot   *Archipelago Plot*

Description

Plot p-values from variant and variant-set tests.

Arguments

df1	A dataframe of variant-set data.
df2	A dataframe of variant data.
plot_title	Title for the plot.
add_title	Logical; add title if TRUE.
plot_subtitle	Subtitle for the plot.
add_subtitle	Logical; add subtitle if TRUE.
chr_ticks	Logical; show chromosome ticks if TRUE.
show_legend	Logical; display the legend if TRUE.
color_theme	Name of the colour theme.
custom_colors	Vector of custom colours.
color_labels	Labels for the colour groups.
crit_val	Critical p-value threshold.
point_size	Size of the points.
point_size_large	
	Size of the large points of interest such as VSAT.
fig_width	Width of the archipelago plot.
fig_height	Height of the archipelago plot.
raw_fig_width	Width of the raw plot.
raw_fig_height	Height of the raw plot.
output_path	File path for the plot.
output_raw	File path for the raw plot.
file_type	Use .png, .jpg, or .pdf. Defaults to png. pdf is slow and large for many SNPs.
alpha_point	Use the alpha_point value to set point transparency.
alpha_seg	Use the alpha_segvalue to set line segment transparency.
better_space	Use better_space = TRUE to make sure that x-axis chr do not squash.
legend_position	
	Default right, allows bottom top left right.

Value

A ggplot object.

Examples

```
## Not run:
  # Load example data for df1 (VSAT) and df2 (individual variant)
  data("vsat_pval")        # same structure as df1
  data("variant_pval")     # same structure as df2

  # Basic usage with defaults
  archipelago_plot(
    df1 = vsat_pval,
    df2 = variant_pval
  )

  # Specify a built-in colour theme
  archipelago_plot(
    df1 = vsat_pval,
    df2 = variant_pval,
    color_theme = "alice"
  )

  # More customised usage
  output_path <- "./archipelago_plot_custom_color.pdf"
  output_raw <- "./vsat_raw_plot.pdf"
  color_labels <- c("Label_1", "Label_2", "Label_3", "Label_4")
  custom_colors <- c("#9abfd8", "#cac1f3", "#371c4b", "#2a5b7f")

  archipelago_plot(
    df1 = vsat_pval,
    df2 = variant_pval,
    add_title = TRUE,
    plot_title = "My Archipelago Plot",
    add_subtitle = TRUE,
    plot_subtitle = "VSAT vs Single Variant",
    show_legend = TRUE,
    chr_ticks = FALSE,
    point_size = 0.5,
    color_theme = NULL,          # ignore built-in theme
    custom_colors = custom_colors,
    color_labels = color_labels,
    crit_val_VSAT = 0.05 / 300, # for highlighting VSAT p-values
    crit_val_single_variant = 5e-8, # typical single-variant threshold
    output_path = output_path,
    output_raw = output_raw,
    file_type = "pdf"            # save as PDF instead of PNG
  )

## End(Not run)
```

---

df2   *Example variant data*

Description

Example variant-level p-values from a genome-wide association study.

Usage

```
df2
```

Format

A data frame with columns:

set_ID Set ID
BP Base pair position
P p-value
CHR Chromosome
SNP SNP ID

---

vsat_pval   *Example variant set data*

Description

Example variant set-level p-values.

Usage

```
vsat_pval
```

Format

A data frame with columns:

set_ID Set ID
P p-value

Index