# Package 'rolypoly'

June 6, 2021

**Type** Package

**Title** Identifying Trait-Relevant Functional Annotations

**Version** 0.1.0

**Date** 2017-3-15

**Author** Diego Calderon

**Maintainer** Diego Calderon <dcal@stanford.edu>

**Description** Using enrichment of genome-wide association summary statistics to identify trait-relevant cellular functional annotations.

**Depends** R (>= 3.1.3),

**Imports** data.table (>= 1.9.6), dplyr (>= 0.4.3), foreach (>= 1.4.3), ggplot2 (>= 1.0.1), glmnet (>= 2.0-5), MASS (>= 7.3-45), Matrix (>= 1.2-6), matrixcalc (>= 1.0-3),

**LinkingTo**

**Suggests** CompQuadForm (>= 1.4.1), knitr, testthat, rmarkdown

**License** GPL-3

**VignetteBuilder** knitr

**LazyData** TRUE

**URL** https://github.com/dcalderon/rolypoly

**BugReports** https://github.com/dcalderon/rolypoly/issues

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-03-16 08:44:35

## R topics documented:

---

```
bootstrap_estimator
```
*Bootstrap parameter estimates for confidence intervals.*

---

### Description

Bootstrap parameter estimates for confidence intervals.

### Usage

```
bootstrap_estimator(rolypoly, bootstrap_iters, run_light, run_parallel)
```

### Arguments

| | |
|---|---|
| rolypoly | rolypoly object |
| bootstrap_iters | |
| | number of bootstrap iterations to run |
| run_light | if we throw away bootstrap data |
| run_parallel | if we want to collect bootstrap data in parallel |

### Examples

```
## Not run: bootstrap_estimator(rolypoly)
```

---

`calculate_annotation_block_heritability`

*Caclulate the contribution of block annotations to the heritability of a trait.*

---

### Description

A vector of indenpendent heritability contributions of block annotations is returned. Sum the vector to get total explained heritability and divide by sum to get proportion.

### Usage

```
calculate_annotation_block_heritability(block_data, params)
```

### Arguments

| | |
|---|---|
| `block_data` | functional information of blocks |
| `params` | parameter fit |

### Examples

```
## Not run: calculate_annotation_block_heritability(block_data, params)
```

---

`calculate_block_values`

*Caclulate predicted block values based on block information and model fit.*

---

### Description

From a model fit we can predict expected variance of a block based on information we have about the block. In the example of gene expression this would equate to predicting the importance of a gene based on its signature of expression.

### Usage

```
calculate_block_values(block_data, params)
```

### Arguments

| | |
|---|---|
| `block_data` | block_data |
| `params` | parameter fit |

### Value

returns block values

### Examples

```
## Not run: calculate_gene_values(block_data, params)
```

---

```
calculate_expected_block_values_given_ld
```
*Caclulate predicted block values based on block information and model fit.*

---

## Description

From a model fit we can predict expected variance of a block based on information we have about the block. In the example of gene expression this would equate to predicting the importance of a gene based on its signature of expression.

## Usage

```
calculate_expected_block_values_given_ld(rolypoly, block_values)
```

## Arguments

```
rolypoly        rolypoly object
block_values    estimated block values.
```

## Examples

```
## Not run: calculate_expected_block_values_given_ld(rolypoly, block_values)
```

---

```
cv_regularized_parameter_estimator
```
*Perform regularization inference.*

---

## Description

Use CV to find appropriate values of lambda for either feature selection or for prediction.

## Usage

```
cv_regularized_parameter_estimator(vectorized_rolypoly_data, n_folds = 10,
    ...)
```

## Arguments

```
vectorized_rolypoly_data
                rolypoly data used for inference
n_folds         number of folds for cross validation
...             other arguments to pass to cv.glmnet
```

## Value

results from cross validation

## Examples

```
## Not run: cv_regularized_parameter_estimator(vectorized_rolypoly_data)
```

| data_io | *Functions for opening and organizing data.* |

**Description**

We must open block annotation data, gwas data, snp annotations, gene annotations. Here, you'll find functions that to this and organize these data into a rolypoly object.

| inference | *Inference functions.* |

**Description**

Functions for inferring relevant annotations using the polyTest model.

| main_wrapper | *Main wrapper functions.* |

**Description**

The main endpoint user functions.

| make_ld_matrix | *Helper function to pull LD data from NCBI.* |

**Description**

Given the path of a gwas file open it into a data.table object

**Usage**

```
make_ld_matrix(all_snps, ld_data)
```

**Arguments**

| all_snps | The snps that were queried |
| ld_data | A returned LD matrix with SNP, Proxy, and RSquared columns |

**Value**

an LD matrix where query snps will be the first columns in the correct order

**Examples**

```
## Not run: make_ld_matrix(all_snps, ld_data)
```

---

| | |
|---|---|
| make_results_df | *Helper function to make a summary table of results from bootstrap data.* |

---

### Description

Helper function to make a summary table of results from bootstrap data.

### Usage

```
make_results_df(value_collection, annotations, model_estimates)
```

### Arguments

```
value_collection
```
          collection of bootstrapped value estimates

```
annotations
```
   vector of annotation names

```
model_estimates
```
          estimates for bias parameter estimates

### Value

data frame with results summary

### Examples

```
## Not run: make_results_df(value_collection)
```

---

| | |
|---|---|
| parameter_estimator | |
| | *Find parameter estimates for the data.* |

---

### Description

Find parameter estimates for the data.

### Usage

```
parameter_estimator(vectorized_rolypoly_data)
```

### Arguments

```
vectorized_rolypoly_data
```
          rolypoly data that has been vectorized

### Value

results of inference

### Examples

```
## Not run: parameter_estimator(rolypoly)
```

```
plot_rolypoly_annotation_estimates
```
*Visualize parameter estimates after running inference*

### Description

Visualize parameter estimates after running inference

### Usage

```
plot_rolypoly_annotation_estimates(rolypoly)
```

### Arguments

```
rolypoly
```
a rolypoly object

### Value

ggplot2 object

### Examples

```
## Not run: plot_rolypoly_annotation_estimates(rolypoly)
```

```
plot_rolypoly_annotation_ranking
```
*Rank annotations by p-value after running inference*

### Description

Rank annotations by p-value after running inference

### Usage

```
plot_rolypoly_annotation_ranking(rolypoly)
```

### Arguments

```
rolypoly
```
a rolypoly object

### Value

ggplot2 object

### Examples

```
## Not run: plot_rolypoly_annotation_ranking(rolypoly)
```

---

regularized_inference
                                  *Inference functions that include regularization*

---

### Description

Functions for inferring relevant annotations using the polyTest model.

---

robust_parameter_estimator
                                  *Find robust parameter estimates for the data.*

---

### Description

Find robust parameter estimates for the data.

### Usage

```
robust_parameter_estimator(vectorized_rolypoly_data, outlier_threshold = 10)
```

### Arguments

```
vectorized_rolypoly_data
```
                     vectorized rolypoly data
```
outlier_threshold
```
                     outlier threshold for robust inference

### Examples

```
## Not run: parameter_estimator(rolypoly)
```

---

rolypoly                        *The rolypoly package for identifying annotations associated with com-*
                                *plex traits.*

---

### Description

The rolypoly package for identifying annotations associated with complex traits.

---

```
rolypoly_add_ld_corrected_gwas_block_scores
```
*Add LD corrected block scores to rolypoly.*

---

### Description

Add LD corrected block scores to rolypoly.

### Usage

```
rolypoly_add_ld_corrected_gwas_block_scores(rolypoly, fast_calculation = T)
```

### Arguments

| | |
|---|---|
| `rolypoly` | rolypoly data |
| `fast_calculation` | |
| | if F then LD deconvolution else quadratic form. |

### Value

rolypoly object with LD corrected gwas scores attached

### Examples

```
## Not run: rolypoly_add_ld_corrected_gwas_block_scores(rolypoly)
```

---

```
rolypoly_link_blocks_and_gwas
```
*Link blocks and gwas*

---

### Description

Takes block information, potentially independent LD blocks or gene blocks, and gwas data and organizes the data for interenal processing

### Usage

```
rolypoly_link_blocks_and_gwas(rolypoly, ld_folder, r2_threshold = 0.2,
  run_parallel = F)
```

### Arguments

| | |
|---|---|
| `rolypoly` | a rolypoly object |
| `ld_folder` | path to a folder with ld data |
| `r2_threshold` | LD threshold to look at data |
| `run_parallel` | check if user wants to run in parallel |

### Value

rolypoly object with data attached

## Examples

```
## Not run: rolypoly_link_blocks_and_gwas(rolypoly, ld_folder, r2_threshold)
```

---

```
rolypoly_load_block_annotation
                 Load annotations for blocks of LD, in some cases this is a gene anno-
                 tation with a window around a gene.
```

---

## Description

Load annotations for blocks of LD, in some cases this is a gene annotation with a window around a gene.

## Usage

```
rolypoly_load_block_annotation(rolypoly, block_annotation, genes = T)
```

## Arguments

rolypoly          rolypoly data object

block_annotation
                  annotation information for block

genes             if these are genes

## Value

rolypoly data with block annotations attached

## Examples

```
## Not run: rolypoly_load_block_annotation(rolypoly, block_annotation)
```

---

```
rolypoly_load_block_data
                 Block annotations, usually gene model.
```

---

## Description

Block annotations, usually gene model.

## Usage

```
rolypoly_load_block_data(rolypoly, block_data)
```

## Arguments

rolypoly          a rolypoly object

block_data        a data frame of block information, usually gene expression. Requires rownames
                  that are identitcal to block labels loaded previously.

**Value**

a rolypoly object with block information loaded

**Examples**

```
## Not run: rolypoly_load_block_data(rolypoly, block_data)
```

---

rolypoly_load_gwas *Load gwas data*

---

**Description**

Load gwas data

**Usage**

```
rolypoly_load_gwas(rolypoly, gwas_data, snp_annotations = NULL,
  gwas_z_filter = -1, add_spline = F, n_knots = 1, add_poly = F,
  n_degree = 2)
```

**Arguments**

| | |
|---|---|
| rolypoly | rolypoly data |
| gwas_data | gwas data |
| snp_annotations | |
| | if there are additional snp annotations included |
| gwas_z_filter | |
| | if we want to remove large effect SNPs |
| add_spline | for fitting a spline to maf |
| n_knots | number of knots for spline |
| add_poly | for fitting a polynomial to maf |
| n_degree | degree of polynomial to fit |

**Value**

rolypoly object with gwas data loaded

**Examples**

```
## Not run: rolypoly_load_gwas(rolypoly, gwas_data)
```

---

```
rolypoly_perform_inference
```
                              *Run inference.*

---

### Description

Coordinates running inference.

### Usage

```
rolypoly_perform_inference(rolypoly, bootstrap_iters = 50,
  outlier_threshold = -1, run_light = F, run_parallel = F)
```

### Arguments

| | |
|---|---|
| `rolypoly` | rolypoly object |
| `bootstrap_iters` | |
| | number of bootstrap iterations to perform |
| `outlier_threshold` | |
| | threshold for performing robust regression, still experimental. |
| `run_light` | if we throw out bootstrap data |
| `run_parallel` | if we collect bootstraps in parallel |

### Value

rolypoly object with inference information attached

### Examples

```
## Not run: run_inference(rolypoly)
```

---

```
rolypoly_perform_regularized_inference
```
                     *Run inference with added regularization.*

---

### Description

If p-values are desired use the other inference function. This for prediction purposes.

### Usage

```
rolypoly_perform_regularized_inference(rolypoly, ...)
```

### Arguments

| | |
|---|---|
| `rolypoly` | a rolypoly object |
| `...` | other arguments to pass to cv.glmnet |

**Value**

rolypoly object with regularization results

**Examples**

```
## Not run: rolypoly_perform_regularized_inference(rolypoly)
```

---

rolypoly_plots          *Plot functions.*

---

**Description**

Functions for rolypoly results.

---

rolypoly_roll          *Main rolypoly wrapper function.*

---

**Description**

The entry point for rolypoly analysis. If no expression data, we assume that we are running just the vegas score processing.

**Usage**

```
rolypoly_roll(rolypoly = NULL, gwas_data = NULL, block_annotation = NULL,
  block_data = NULL, ld_folder = NULL, bootstrap_iters = 50,
  outlier_threshold = -1, perform_cv = F, n_folds = 10,
  gwas_z_filter = -1, add_spline = F, n_knots = 1, add_poly = F,
  n_degree = 2, run_light = T, gwas_link_parallel = F,
  bootstrap_parallel = F, keep_model = F, keep_gwas = F, ...)
```

**Arguments**

| | |
|---|---|
| rolypoly | Previous rolypoly run to parts of pipeline. |
| gwas_data | Gwas data for a trait, including snp annotations. |
| block_annotation | |
| | Start and end points for blocks |
| block_data | Information about blocks. |
| ld_folder | Folder with LD information. |
| bootstrap_iters | |
| | Number bootstrap iterations to perform for inference. |
| outlier_threshold | |
| | Set to positive if we want to run robusted regression. |
| perform_cv | If we want to interpret annotation effects do not set this to T. However, if our goal is prediction accuracy then set this to T. |
| n_folds | number of folds for cross validation |

gwas_z_filter

> Z-score filter for SNPs, helps prevent large effects biasing inference.

add_spline       If we want to fit a spline to maf.

n_knots          number of knots to add to the spline.

add_poly         If we want to fit a polynomial to maf.

n_degree         the degree of the polynomial.

run_light        if we want to throw away bootstrap data, and save memory

gwas_link_parallel

> if user wants to run in gwas linking in parallel, registerDoParallel must have been run in advance.

bootstrap_parallel

> if user wants to run in bootstraps in parallel, registerDoParallel must have been run in advance.

keep_model       if we should keep the regression model, can be large.

keep_gwas        set to T if we want to include gwas in returned rolypoly object.

...              other arguments to pass to cv.glmnet

## Value

rolypoly object

## Examples

```
## Not run: rolypoly_roll(rolypoly)
```

---

sim_block_annotation

*Simulated block data annotation.*

---

## Description

A dataset containing simulated block data annotation for use in the rolypoly vignette.

## Usage

```
sim_block_annotation
```

## Format

A data frame with 1000 rows and 4 variables:

**chrom** chromosome, we only use autosomes

**start** base pair position of variant

**end** rsid identifier of variant

**label** effect size, univariate regression coefficient

## Source

I generated these fields to link with SNP positions

---

```
sim_expression_data_normalized
```
*Simulated expression data.*

---

## Description

A dataset containing simulated expression data for use in the rolypoly vignette.

## Usage

```
sim_expression_data_normalized
```

## Format

A data frame with 1000 rows and 5 variables:

**Liver** simulated expression data for tissue. This tissue has a significant effect on the simulated gwas.

**Adrenal.Gland** simulated expression data for tissue

**Blood** simulated expression data for tissue. This tissue has a significant effect on the simulated gwas.

**Heart** simulated expression data for tissue

**Lung** simulated expression data for tissue

## Source

I generated this dataset

---

```
sim_gwas_data
```
*Simulated GWAS summary statistics*

---

## Description

A dataset containing simulated genome-wide association summary statistics for use in the rolypoly vignette.

## Usage

```
sim_gwas_data
```

## Format

A data frame with 14934 rows and 6 variables:

**chrom** chromosome, we only use autosomes

**pos** base pair position of variant

**rsid** rsid identifier of variant

**beta** effect size, univariate regression coefficient

**se** standard error of effect size

**maf** minor allele frequency

**Source**

rsids were from 1000g and I generated the other fields

---

vectorize_rolypoly *Take a list of rolypoly data and vectorize it for inference.*

---

**Description**

Take a list of rolypoly data and vectorize it for inference.

**Usage**

```
vectorize_rolypoly(data)
```

**Arguments**

data            the list of block information from rolypoly object

**Value**

list of necessary information for inference

**Examples**

```
## Not run: vectorize_rolypoly(data)
```

# Index