Package ‘PANACEA’

August 19, 2023

Title  Personalized Network-Based Anti-Cancer Therapy Evaluation
Version  1.0.1
Maintainer  Ege Ulgen <egeulgen@gmail.com>
Description  Identification of the most appropriate pharmacotherapy for each patient based on genomic alterations is a major challenge in personalized oncology. ‘PANACEA’ is a collection of personalized anti-cancer drug prioritization approaches utilizing network methods. The methods utilize personalized “driverness” scores from ‘driveR’ to rank drugs, mapping these onto a protein-protein interaction network. The “distance-based” method scores each drug based on these scores and distances between drugs and genes to rank given drugs. The “RWR” method propagates these scores via a random-walk with restart framework to rank the drugs. The methods are described in detail in Ulgen E, Ozisik O, Sezerman OU. 2023. PANACEA: network-based methods for pharmacotherapy prioritization in personalized oncology. Bioinformatics <doi:10.1093/bioinformatics/btad022>.
License  MIT + file LICENSE
Encoding  UTF-8
RoxygenNote  7.2.3
URL  https://github.com/egeulgen/PANACEA,
     https://egeulgen.github.io/PANACEA/
BugReports  https://github.com/egeulgen/PANACEA/issues
biocViews
Imports  org.Hs.eg.db, DBI, igraph, reshape2
Suggests rmarkdown, knitr, testthat (>= 3.0.0), covr
Config/testthat/edition  3
Depends  R (>= 4.0)
LazyData  true
LazyDataCompression  xz
VignetteBuilder  knitr
NeedsCompilation  yes
Description

Add Drugs as Nodes

Usage

```r
add_drugs_as_nodes(W_mat, drug_target_interactions, edge_weight = 1000)
```

Arguments

- `W_mat` : adjacency matrix for the chosen PIN
- `drug_target_interactions` : data frame containing (processed) drugs and target genes
- `edge_weight` : edge weight for drug-target gene interaction (default = 1000)

Value

adjacency matrix with the drugs added as nodes
**adj_list2mat**

**Turn Adjacency List into Adjacency Matrix**

**Description**

Turn Adjacency List into Adjacency Matrix

**Usage**

adj_list2mat(adj_list)

**Arguments**

adj_list  Adjacency list

**Value**

Adjacency matrix

---

**convert2alias**

Convert Input Gene Symbols to Alias

**Description**

Convert Input Gene Symbols to Alias

**Usage**

convert2alias(input_genes, target_genes)

**Arguments**

input_genes  vector of input genes
target_genes  vector of target genes

**Value**

vector of converted gene symbols (if any alias in target genes)
DGIdb_interactions_df  DGIdb Interactions Expert-curated Sources

Description

Data frame containing drug-gene interactions from expert-curated sources (CancerCommons, CGI, ChemblInteractions, CIViC, ClearityFoundationBiomarkers, ClearityFoundationClinicalTrial, COSMIC, DoCM, MyCancerGenome, MyCancerGenomeClinicalTrial, TALC, TdgClinicalTrial, TEND) from DGIdb.

Usage

DGIdb_interactions_df

Format

a data frame containing 11323 rows and 2 variables:

- **drug_name**  Drug name
- **gene_name**  HGNC gene symbol for the interacting gene

example_driveR_res  Example driveR Result

Description

Data frame containing 'driveR' results for a lung adenocarcinoma case.

Usage

example_driveR_res

Format

a data frame containing 106 rows and 3 variables:

- **gene_symbol**  HGNC gene symbol
- **driverness_prob**  'driverness' probability
- **prediction**  driveR’s prediction for whether the gene is a 'driver' or 'non-driver'
example_scores_dist  

Example PANACEA "distance-based" Method Result

Description
Vector containing 'PANACEA' "distance-based" results for a lung adenocarcinoma case. Names are drug names, values are scores

Usage
example_scores_dist

Format
named vector of 1423 values

example_scores_RWR  

Example PANACEA "RWR" Method Result

Description
Vector containing 'PANACEA' "RWR" results for a lung adenocarcinoma case. Names are drug names, values are scores

Usage
example_scores_RWR

Format
named vector of 1423 values

Laplacian.norm  

Graph Laplacian Normalization

Description
Graph Laplacian Normalization

Usage
Laplacian.norm(W)
network_propagation

Arguments

\( W \)
- square symmetric adjacency matrix

Value

normalized adjacency matrix

network_propagation

Network Propagation (Random-walk with Restart)

Description

Network Propagation (Random-walk with Restart)

Usage

\[
\text{network\_propagation}(\text{prior\_vec}, W_{\text{prime}}, \alpha, \text{max.\_iter} = 1000, \epsilon = 1e^{-04})
\]

Arguments

- \text{prior\_vec}
  - vector of prior knowledge on selected genes (names are gene symbols)
- \text{W\_prime}
  - (Laplacian-normalized, symmetric) adjacency matrix
- \text{alpha}
  - restart parameter, controlling trade-off between prior information and network smoothing
- \text{max.\_iter}
  - maximum allowed number of iterations (default = 1000)
- \text{eps}
  - epsilon value to assess the L2 norm of the difference between iterations (default = 1e-4)

Details


Value

vector of propagation values
PANACEA: Personalized Network-based Anti-Cancer Therapy Evaluation

Description

Identification of the most appropriate pharmacotherapy for each patient based on genomic alterations is a major challenge in personalized oncology. PANACEA is a collection of personalized anti-cancer drug prioritization approaches utilizing network methods. The methods utilize personalized "driveness" scores from 'driveR' to rank drugs, mapping these onto a protein-protein interaction network (PIN). The "distance-based" method scores each drug based on these scores and distances between drugs and genes to rank given drugs. The "RWR" method propagates these scores via a random-walk with restart framework to rank the drugs.

Author(s)

Maintainer: Ege Ulgen <egeulgen@gmail.com> (ORCID) [copyright holder]

See Also

score_drugs for the wrapper function for scoring of drugs via network-based methods

process_drug_target_interactions

Process Drug-Target Interactions

Description

Process Drug-Target Interactions

Usage

process_drug_target_interactions(
  drug_target_interactions,
  PIN_genes,
  drug_name_col = "drug_name",
  target_col = "gene_name"
)

Arguments

drug_target_interactions
  data frame containing drugs and target genes

PIN_genes
  gene symbols for the chosen PIN

drug_name_col
  name of the column containing drug names (default = "drug_name")

target_col
  name of the column containing drug targets (default = "converted_target_gene")
processed drug-target interactions. Processing involves converting symbols missing in the PIN, merging drugs that have the same target gene(s).

**score_drugs**  
**Scoring of Drugs via Network-based Methods**

**Description**  
Scoring of Drugs via Network-based Methods

**Usage**  
```
score_drugs(driveR_res, drug_interactions_df, W_mat, method, ...)
```

**Arguments**
- `driveR_res` data frame of driveR results
- `drug_interactions_df` data frame of drug-gene interactions
- `W_mat` adjacency matrix for the PIN
- `method` scoring method (one of ‘distance-based’ or ‘RWR’)
- `...` additional arguments for `score_drugs_distance_based` or `score_drugs_RWR_based`

**Details**
This is the wrapper function for the two proposed methods for personalized scoring of drugs for individual cancer samples via network-based methods. The available methods are ‘distance-based’ and ‘RWR’. For the ‘distance-based’ method, the score between a gene (g) and drug (d) is formulated as:

\[
score(g, d) = \frac{\text{driver}(g)}{(d(g, d) + 1)^2}
\]

where driver(g) is the driverness probability of gene g, as predicted by ‘driveR’ and d(g, d) is the distance within the PIN between gene g and drug d. The final score of the drug d is then the average of the scores between each altered gene and d:

\[
score(d) = \Sigma score(g, d)/|\text{genes}|
\]

For the ‘RWR’ method, a random-walk with restart framework is used to propagate the driverness probabilities.

By default `DGIdb_interactions_df` is used as the `drug_interactions_df`.

If the `W_mat` argument is not supplied, the built-in STRNG data `STRING_adj_df` is used to generate `W_mat`.

**Value**
vector of scores per drug.
Examples

```r
toy_data <- data.frame(
  gene_symbol = c("TP53", "EGFR", "KDR", "ATM"),
  driverness_prob = c(0.94, 0.92, 0.84, 0.72)
)
toy_interactions <- DGIdb_interactions_df[1:25, ]
res <- score_drugs(
  driveR_res = toy_data,
  drug_interactions_df = toy_interactions, # leave blank for default
  W_mat = toy_W_mat, # leave blank for default
  method = "distance-based",
  verbose = FALSE
)
```

score_drugs_distance_based

**Distance-based Scoring of Drugs**

Description

Distance-based Scoring of Drugs

Usage

```r
score_drugs_distance_based(
  driveR_res, 
  drug_interactions_df, 
  W_mat, 
  driver_prob_cutoff = 0.05, 
  drug_name_col = "drug_name", 
  target_col = "gene_name", 
  verbose = TRUE
)
```

Arguments

- `driveR_res` data frame of driveR results
- `drug_interactions_df` data frame of drug-gene interactions
- `W_mat` adjacency matrix for the PIN
- `driver_prob_cutoff` cut-off value for 'driverness_prob' (default = 0.05)
- `drug_name_col` for 'drug_interactions_df', the column name containing drug names/identifiers
- `target_col` for 'drug_interactions_df', the column name containing target gene symbols
- `verbose` boolean to control verbosity (default = TRUE)
Value
vector of scores per drug. Drugs with the same target gene(s) are merged (via process_drug_target_interactions).

Examples
```
toy_data <- data.frame(
  gene_symbol = c("TP53", "EGFR", "KDR", "ATM"),
  driverness_prob = c(0.94, 0.92, 0.84, 0.72)
)
toy_interactions <- DGIdb_interactions_df[1:100, ]
res <- score_drugs_distance_based(
  driveR_res = toy_data,
  drug_interactions_df = toy_interactions,
  W_mat = toy_W_mat, verbose = FALSE
)
```

score_drugs_RWR_based  RWR-based Scoring of Drugs

Description
RWR-based Scoring of Drugs

Usage
```
score_drugs_RWR_based(
  driveR_res, drug_interactions_df, W_mat, alpha = 0.05,
  max.iter = 1000, eps = 1e-04, drug_name_col = "drug_name",
  target_col = "gene_name", verbose = TRUE
)
```

Arguments
- driveR_res  data frame of driveR results
- drug_interactions_df  data frame of drug-gene interactions
- W_mat  adjacency matrix for the PIN
- alpha  restart parameter, controlling trade-off between prior information and network smoothing
- max.iter  maximum allowed number of iterations (default = 1000)
eps  epsilon value to assess the L2 norm of the difference between iterations (default = 1e-4)

drug_name_col  for 'drug_interactions_df', the column name containing drug names/identifiers

target_col  for 'drug_interactions_df', the column name containing target gene symbols

verbose  boolean to control verbosity (default = TRUE)

Value  vector of scores per drug. Drugs with the same target gene(s) are merged (via process_drug_target_interactions)

Examples

toy_data <- data.frame(
    gene_symbol = c("TP53", "EGFR", "KDR", "ATM"),
    driverness_prob = c(0.94, 0.92, 0.84, 0.72)
)

toy_interactions <- DGIdb_interactions_df[1:100, ]

res <- score_drugs_RWR_based(
    driver_res = toy_data,
    drug_interactions_df = toy_interactions,
    W_mat = toy_W_mat, verbose = FALSE
)

STRING_adj_df  Adjacency List for STRING v11.5 - High Confidence Interactions

Description  Data frame of adjacency list for STRING v11.5 interactions with combined score > 700 (high confidence)

Usage  STRING_adj_df

Format  a data frame with 887797 rows and 3 variables:

protein1  Interactor 1
protein2  Interactor 2
value  edge weight(combined score)
toy_W_mat  

*Toy Adjacency Matrix (for examples)*

---

**Description**

Symmetric matrix containing example adjacency data

**Usage**

`toy_W_mat`

**Format**

Matrix of 84 rows and 84 columns
Index

* datasets
  - DGIdb_interactions_df, 4
  - example_driveR_res, 4
  - example_scores_dist, 5
  - example_scores_RWR, 5
  - STRING_adj_df, 11
  - toy_W_mat, 12

  add_drugs_as_nodes, 2
  adj_list2mat, 3

  convert2alias, 3
  DGIdb_interactions_df, 4, 8
  example_driveR_res, 4
  example_scores_dist, 5
  example_scores_RWR, 5

  Laplacian.norm, 5

  network_propagation, 6

  PANACEA, 7
  PANACEA-package (PANACEA), 7
  process_drug_target_interactions, 7, 10, 11

  score_drugs, 7, 8
  score_drugs_distance_based, 8, 9
  score_drugs_RWR_based, 8, 10
  STRING_adj_df, 8, 11

  toy_W_mat, 12