Package 'MPAC'

September 7, 2024

Title Multi-omic Pathway Analysis of Cancer

Version 0.99.14

Description Multi-omic Pathway Analysis of Cancer (MPAC), integrates multi-omic data for understanding cancer mechanisms. It predicts novel patient groups with distinct pathway profiles as well as identifying key pathway proteins with potential clinical associations. From CNA and RNA-seq data, it determines genes' DNA and RNA states (i.e., repressed, normal, or activated), which serve as the input for PARADIGM to calculate Inferred Pathway Levels (IPLs). It also permutes DNA and RNA states to create a background distribution to filter IPLs as a way to remove events observed by chance. It provides multiple methods for downstream analysis and visualization.

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Depends R (>= 4.4.0)

URL https://github.com/pliu55/MPAC

BugReports https://github.com/pliu55/MPAC/issues

Imports data.table (>= 1.14.2), SummarizedExperiment (>= 1.30.2), BiocParallel (>= 1.28.3), fitdistrplus (>= 1.1), igraph (>= 1.4.3), BiocSingular (>= 1.10.0), S4Vectors (>= 0.32.3), SingleCellExperiment (>= 1.16.0), bluster (>= 1.4.0), fgsea (>= 1.20.0), scran (>= 1.22.1), ComplexHeatmap (>= 2.16.0), grid, stats

Suggests rmarkdown, knitr, svglite, bookdown(>= 0.34), testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

biocViews Software, Technology, Sequencing, RNASeq, Survival, Clustering, ImmunoOncology

SystemsRequirements The `runPrd()` function requires an external software named PARADIGM. For details, please see the 'Required external software' section in vignette's 'Run PARADIGM: runPrd()'.

git_url https://git.bioconductor.org/packages/MPAC

git_branch devel

git_last_commit b767726

git_last_commit_date 2024-06-18

Repository Bioconductor 3.20

Date/Publication 2024-09-06

Author Peng Liu [aut, cre] (<https://orcid.org/0000-0001-5655-2259>),
Paul Ahlquist [aut],
Irene Ong [aut],
Anthony Gitter [aut]

Maintainer Peng Liu <pliu55.wisc+bioconductor@gmail.com>

Contents

clSamp 2
colPermIPL 3
colRealIPL 4
conMtf 4
fltByPerm 5
ovrGMT 6
pltNeiStt 7
ppCnInp 7
ppPermInp 8
ppRealInp 9
ppRnaInp 9
runPermPrd 10
runPrd 11
subNtw 12

Index **13**

`clSamp` *Cluster samples by pathway over-representation*

Description

Cluster samples by pathway over-representation

Usage

```
clSamp(ovrmat, n_neighbors = 10, n_random_runs = 100, threads = 1)
```

Arguments

`ovrmat`	A matrix of gene set over-representation adjusted p-values with rows as gene sets and columns as samples. It is the output from `ovrGMT()`.
`n_neighbors`	Number of neighbors for clustering. A larger number is recommended if the size of samples is large. Default: 10.

n_random_runs	Number of random runs. Due to randomness introduced to the Louvain algorithm in R igraph 1.3.0 (https://github.com/igraph/rigraph/issues/539), a large number of runs are recommended to evaluate randomness in the clustering results. Default: 100.
threads	Number of threads to run in parallel. Default: 1

Value

A data table with each row representing one clustering result, and the first column denotes the number of occurrences of a clustering result and the rest of columns indicating each sample's cluster index. Rows are ordered by the number of occurrences from high to low.

Examples

```
fovr = system.file('extdata/clSamp/ovrmat.rds', package='MPAC')
ovrmat = readRDS(fovr)

clSamp(ovrmat)
```

---

colPermIPL *Collect Inferred Pathway Levels (IPLs) from PARADIGM runs on permuted data*

Description

Collect Inferred Pathway Levels (IPLs) from PARADIGM runs on permuted data

Usage

```
colPermIPL(indir, n_perms, sampleids = NULL)
```

Arguments

indir	Input folder that saves PARADIGM results. It should be set as the same as outdir as in runPrd().
n_perms	Number of permutations to collect.
sampleids	Sample IDs for which IPLs to be collected. If not provided, all files with suffix '_ipl.txt' in indir will be collected. Default: NULL.

Value

A data.table object with columns of permutation index, pathway entities and their IPLs.

Examples

```
indir = system.file('/extdata/runPrd/', package='MPAC')
n_perms = 3

colPermIPL(indir, n_perms)
```

colRealIPL *Collect Inferred Pathway Levels (IPLs) from PARADIGM runs on real data*

Description

Collect Inferred Pathway Levels (IPLs) from PARADIGM runs on real data

Usage

```
colRealIPL(indir, sampleids = NULL)
```

Arguments

indir	Input folder that saves PARADIGM results. It should be set as the same as outdir as in runPrd().
sampleids	Sample IDs for which IPLs to be collected. If not provided, all files with suffix '_ipl.txt' in indir will be collected. Default: NULL.

Value

A data.table object with columns of pathway entities and their IPLs.

Examples

```
indir = system.file('/extdata/runPrd/', package='MPAC')

colRealIPL(indir)
```

conMtf *Find consensus pathway motifs from a list of pathways*

Description

Find consensus pathway motifs from a list of pathways

Usage

```
conMtf(subntwl, omic_genes = NULL, min_mtf_n_nodes = 5)
```

Arguments

subntwl	A list of igraph objects representing input pathways from different samples. It is the output from subNtw()
omic_genes	A vector of gene symbols to narrow down over-representation calculation to only those with input genomic data. If not provided, all genes in the GMT file will be considered. Default: NULL.
min_mtf_n_nodes	Number of minimum nodes in a motif. Default: 5

Value

A list of igraph objects representing consensus pathway motifs

Examples

```
fsubntwl = system.file('extdata/conMtf/subntwl.rds', package='MPAC')
subntwl = readRDS(fsubntwl)

fomic_gns = system.file('extdata/TcgaInp/inp_focal.rds', package='MPAC')
omic_gns = rownames(readRDS(fomic_gns))

conMtf(subntwl, omic_gns, min_mtf_n_nodes=50)
```

---

fltByPerm *Filter IPLs from real data by distribution from permuted data*

Description

Filter IPLs from real data by distribution from permuted data

Usage

```
fltByPerm(realdt, permdt)
```

Arguments

realdt	A data.table object containing entities and their IPLs from real data. It is the output from `colRealIPL()`.
permdt	A data.table object containing permutation index, entities and their IPLs from permuted data. It is the output from `colPermIPL()`.

Value

A matrix of filtered IPLs with rows as entities and columns as samples. Entities with IPLs observed by chance are set to NA.

Examples

```
freal = system.file('extdata/fltByPerm/real.rds', package='MPAC')
fperm = system.file('extdata/fltByPerm/perm.rds', package='MPAC')
realdt = readRDS(freal)
permdt = readRDS(fperm)

fltByPerm(realdt, permdt)
```

ovrGMT *Calculate over-representation of gene sets in each sample by genes from sample's largest sub-pathway*

Description

Calculate over-representation of gene sets in each sample by genes from sample's largest sub-pathway

Usage

```
ovrGMT(subntwlist, fgmt, omic_genes = NULL, threads = 1)
```

Arguments

subntwlist	A list of igraph objects represented the largest sub-pathway for each sample. It is the output of subNtw().
fgmt	A gene set GMT file. This will be the same file used for the gene set over-representation calculation in the next step. It is used here to ensure output sub-pathway contains a minimum number of genes from to-be-used gene sets.
omic_genes	A vector of gene symbols to narrow down over-representation calculation to only those with input genomic data. If not provided, all genes in the GMT file will be considered. Default: NULL.
threads	Number of threads to run in parallel. Default: 1

Value

A matrix containing over-representation adjusted P with rows as gene set names and columns as sample IDs.

Examples

```
fsubntwl  = system.file('extdata/subNtw/subntwl.rds',    package='MPAC')
fgmt      = system.file('extdata/ovrGMT/fake.gmt',       package='MPAC')
fomic_gns = system.file('extdata/TcgaInp/inp_focal.rds', package='MPAC')
subntwl  = readRDS(fsubntwl)
omic_gns = rownames(readRDS(fomic_gns))

ovrGMT(subntwl, fgmt, omic_gns)
```

pltNeiStt — *Plot a heatmap of pathway and omic states of a protein and its pathway neighbors*

Description

Plot a heatmap of pathway and omic states of a protein and its pathway neighbors

Usage

```
pltNeiStt(real_se, fltmat, fpth, protein)
```

Arguments

real_se	A SummarizedExperiment object of PARADIGM CNA and RNA states. It is the same matrix as the output from ppRealInp().
fltmat	A matrix contains filterd IPL with rows as 'entity' and column as samples. This is the output from fltByPerm().
fpth	Name of a pathway file for PARADIGM.
protein	Name of the protein to plot. It requires to have CN and RNA state data, as well as pathway data from the input.

Value

A heatmap of pathway and omic states of a protein and its pathway neighbors

Examples

```
fpth = system.file('extdata/Pth/tiny_pth.txt', package='MPAC')

freal = system.file('extdata/pltNeiStt/inp_real.rds', package='MPAC')
fflt  = system.file('extdata/pltNeiStt/fltmat.rds',   package='MPAC')

real_se = readRDS(freal)
fltmat = readRDS(fflt)
protein = 'CD86'

pltNeiStt(real_se, fltmat, fpth, protein)
```

ppCnInp — *Prepare input copy-number (CN) alteration data to run PARADIGM*

Description

Prepare input copy-number (CN) alteration data to run PARADIGM

Usage

```
ppCnInp(cn_tumor_mat)
```

Arguments

cn_tumor_mat	A matrix of tumor CN focal data with rows as genes and columns as samples

Value

A SummarizedExperiment object of CN state for PARADIGM

Examples

```
fcn = system.file('extdata/TcgaInp/focal_tumor.rds', package='MPAC')
cn_tumor_mat = readRDS(fcn)

ppCnInp(cn_tumor_mat)
```

---

ppPermInp *Permute input genomic state data between genes in the same sample*

Description

Permute input genomic state data between genes in the same sample

Usage

```
ppPermInp(real_se, n_perms=3, threads=1)
```

Arguments

real_se	A SummarizedExperiment object of CN and RNA states from real samples with rows as genes and columns as samples. It is the output from ppRealInp().
n_perms	Number of permutations. Default: 3
threads	Number of threads to run in parallel. Default: 1

Value

A list of SummarizedExperiment objects of permuted CN and RNA states. The metadata i in each obbect denotes its permutation index.

Examples

```
freal = system.file('extdata/TcgaInp/inp_real.rds', package='MPAC')
real_se = readRDS(freal)

ppPermInp(real_se, n_perms=3)
```

ppRealInp — *Prepare input copy-number (CN) alteration and RNA data to run PARADIGM*

Description

Prepare input copy-number (CN) alteration and RNA data to run PARADIGM

Usage

```
ppRealInp(cn_tumor_mat, rna_tumor_mat, rna_normal_mat, threads = 1)
```

Arguments

cn_tumor_mat	A matrix of tumor CN focal data with rows as genes and columns as samples
rna_tumor_mat	A matrix of RNA data from tumor samples with rows as genes and columns as samples
rna_normal_mat	A matrix of RNA data from normal samples with rows as genes and columns as samples
threads	Number of threads to run in parallel. Default: 1

Value

A SummarizedExperiment object of CN and RNA state for PARADIGM

Examples

```
fcn = system.file('extdata/TcgaInp/focal_tumor.rds', package='MPAC')
ftumor = system.file('extdata/TcgaInp/log10fpkmP1_tumor.rds', package='MPAC')
fnorm = system.file('extdata/TcgaInp/log10fpkmP1_normal.rds', package='MPAC')

cn_tumor_mat = readRDS(fcn)
rna_tumor_mat = readRDS(ftumor)
rna_norm_mat  = readRDS(fnorm)

ppRealInp(cn_tumor_mat, rna_tumor_mat, rna_norm_mat)
```

ppRnaInp — *Prepare input RNA data to run PARADIGM*

Description

Prepare input RNA data to run PARADIGM

Usage

```
ppRnaInp(rna_tumor_mat, rna_normal_mat, threads = 1)
```

Arguments

rna_tumor_mat	A matrix of RNA data from tumor samples with rows as genes and columns as samples
rna_normal_mat	A matrix of RNA data from normal samples with rows as genes and columns as samples
threads	Number of threads to run in parallel. Default: 1

Value

A SummarizedExperiment of RNA state for PARADIGM

Examples

```
ftumor = system.file('extdata/TcgaInp/log10fpkmP1_tumor.rds', package='MPAC')
fnorm = system.file('extdata/TcgaInp/log10fpkmP1_normal.rds', package='MPAC')
rna_tumor_mat = readRDS(ftumor)
rna_norm_mat  = readRDS(fnorm)

ppRnaInp(rna_tumor_mat, rna_norm_mat, threads=2)
```

---

runPermPrd *Run PARADIGM on permuted data*

Description

Run PARADIGM on permuted data

Usage

```
runPermPrd(perml, fpth, outdir,
    PARADIGM_bin=NULL, nohup_bin=NULL, sampleids=NULL, threads=1)
```

Arguments

perml	A list of SummarizedExperiment objects of permuted CNA and RNA states generated by ppPermInp().
fpth	Name of a pathway file for PARADIGM.
outdir	Output folder to save all results.
PARADIGM_bin	PARADIGM binary, which can be downloaded from https://github.com/sng87/paradigm-scripts/tree/master/public/exe. Note that the binary is only available for Linux or MacOS. Default: NULL
nohup_bin	nohup binary, which is used for long running PARADIGM jobs. Default: NULL
sampleids	A vector of sample IDs to run PARADIGM on. If not provided, all the samples that exist in both copy-number alteration and RNA files will be ran. Default: NULL
threads	Number of threads to run in parallel. Default: 1

Value

None

Examples

```
fperm = system.file('extdata/TcgaInp/inp_perm.rds', package='MPAC')
perml = readRDS(fperm)
fpth = system.file('extdata/Pth/tiny_pth.txt', package='MPAC')
outdir = tempdir()
paradigm_bin = '/path/to/PARADIGM'  ## change to binary location
pat = 'TCGA-CV-7100'

# depends on external PARADIGM binary, do not run
runPermPrd(perml, fpth, outdir, paradigm_bin, sampleids=c(pat))
```

---

runPrd *Run PARADIGM on multi-omic data*

Description

Run PARADIGM on multi-omic data

Usage

```
runPrd(real_se, fpth, outdir, PARADIGM_bin=NULL, nohup_bin=NULL,
    sampleids=NULL, threads=1)
```

Arguments

real_se	A SummarizedExperiment object of PARADIGM CNA and RNA states. It is the same matrix as the output from ppRealInp().
fpth	Name of a pathway file for PARADIGM.
outdir	Output folder to save all results.
PARADIGM_bin	PARADIGM binary, which can be downloaded from https://github.com/sng87/paradigm-scripts/tree/master/public/exe. Note that the binary is only available for Linux or MacOS. Default: NULL
nohup_bin	nohup binary, which is used for long running PARADIGM jobs. Default: NULL
sampleids	A vector of sample IDs to run PARADIGM on. If not provided, all the samples that exist in both copy-number alteration and RNA files will be ran. Default: NULL
threads	Number of threads to run in parallel. Default: 1

Value

None

Examples

```
freal = system.file('extdata/TcgaInp/inp_real.rds', package='MPAC')
real_se  = readRDS(freal)

fpth = system.file('extdata/Pth/tiny_pth.txt', package='MPAC')
outdir = tempdir()
paradigm_bin = '/path/to/PARADIGM'  ## change to binary location

# depends on external PARADIGM binary
runPrd(real_se, fpth, outdir, paradigm_bin, sampleids=c('TCGA-CV-7100'))
```

---

subNtw *Subset pathways by IPL results*

Description

Subset pathways by IPL results

Usage

```
subNtw(fltmat, fpth, fgmt, min_n_gmt_gns = 2, threads = 1)
```

Arguments

fltmat	A matrix contains filterd IPL with rows as 'entity' and column as samples. This is the output from `fltByPerm()`.
fpth	Name of a pathway file for PARADIGM.
fgmt	A gene set GMT file. This will be the same file used for the gene set over-representation calculation in the next step. It is used here to ensure output sub-pathway contains a minimum number of genes from to-be-used gene sets.
min_n_gmt_gns	Minimum number of genes from the GMT file in the output sub-pathway. Default: 2.
threads	Number of threads to run in parallel. Default: 1

Value

A list of igraph objects representing the largest sub-pathway for each sample.

Examples

```
fflt = system.file('extdata/fltByPerm/flt_real.rds', package='MPAC')
fltmat = readRDS(fflt)
fpth = system.file('extdata/Pth/tiny_pth.txt',       package='MPAC')
fgmt = system.file('extdata/ovrGMT/fake.gmt',        package='MPAC')

subNtw(fltmat, fpth, fgmt, min_n_gmt_gns=1)
```

Index

clSamp, 2
colPermIPL, 3
colRealIPL, 4
conMtf, 4

fltByPerm, 5

ovrGMT, 6

pltNeiStt, 7
ppCnInp, 7
ppPermInp, 8
ppRealInp, 9
ppRnaInp, 9

runPermPrd, 10
runPrd, 11

subNtw, 12