Package 'panoply'

April 27, 2018

Title Personalized Cancer Report

Version 0.98

Date 2018-04-24

URL http://kalarikrlab.org/Software/Panoply.html

Description
Personalized Cancer Report with prioritizing cancer drugs by the genomic networks they target.

Suggests xtable,
knitr

Depends R (>= 3.2.0),
stats,
MASS,
gage,
circlize,
Rgraphviz,
randomForest,
RColorBrewer

Imports methods,
utils,
graphics,
grDevices

VignetteBuilder knitr

License GPL (>= 2)

R topics documented:

---

clinPanBC *Clinical Data for Breast Cancer Subjects*

Description

Clinical Data of Breast Cancer patients to be used in Panoply

Usage

```
data("clinPanBC")
```

Format

A data frame with 6 subjects. One non-responder and 5 matched responders

PatientID character, patient identifier

Clinical.Molecular.Subtype character, Clinical Molecular Subtype Triple Negative

Breast.Nodal.CR.all character, Complete Response yes no

T.stage character, T stage T2

N.stage character, N stage N0

Percentage.tumor.cells.baseline.internal integer, Percentage Tumor Cells

Age.Group ordered factor, Age in 10 year categories 60-69 < 70+

Examples

```
data(clinPanBC)
str(clinPanBC)
```

---

clinPanCO *Clinical Data for TCGA Colon Cancer Patients*

Description

Clinical Data of TCGA Colon Cancer patients to be used in Panoply

Usage

```
data("clinPanCO")
```

Format

A data frame with 9 subjects. 5 cases (deceased) and 4 matched controls (alive)

bcr_patient_barcode character, patient identifier, TCGA barcode

bcr_patient_uuid character, TCGA uuid

patient_id character, simplified patient identifier

days_to_death character, days to death

days_to_last_followup character, days to last followup

race character, race, WHITE

vital_status character, vital status Alive Dead

gender character, gender FEMALE MALE

kras_mutation_found character, kras mutation presentYES

surv character, alive recur

CN.tumorpct numeric, Tumor Percentage

Examples

```
data(clinPanCO)
str(clinPanCO)
```

---

cnaPanCO *DNA CNV (germline) and CNA (tumor) data for the TCGA Colon Cancer Subjects*

Description

Per-Gene DNA Copy Number Variation (germline) and Copy Number Alteration (tumor) data for the TCGA Colon Cancer Subjects

Usage

```
data("cnaPanCO")
```

Format

A data frame with 6 subjects. One non-responder and 5 matched responders

CHROM chromosome of gene

START gene start position

STOP gene stop position

Gene.Symbol gene symbol, NCBI

TCGA-DM-A0XD patient identifier

Examples

```
data(cnaPanCO)
str(cnaPanCO)
```

cnaPanTNBC *DNA CNV (germline) and CNA (tumor) data for TCGA samples*

Description

Per-Gene DNA Copy Number Variation (germline) and Copy Number Alteration (tumor) data for TCGA samples

Usage

```
data("cnaPanTNBC")
```

Format

A data frame with 77 subjects.

CHROM chromosome of gene

START gene start position

STOP gene stop position

geneid Gene ID

Gene.Symbol gene symbol, NCBI

TCGA-A2-A04T patient identifier

Examples

```
data(cnaPanBC)
str(cnaPanBC)
```

dgidbPan *Druggable Genome Interaction Database for Cancer*

Description

Druggable Genome Interaction Database annotation for Cancer Drugs and their targets

Usage

```
data("dgidbPan")
```

Format

Gene Gene Symbol

Drugs All cancer drugs that target the gene

dgiSets *Druggable Genome Interaction Database for Cancer*

Description

Druggable Genome Interaction Database annotation for Cancer Drugs and their targets

Usage

```
data("dgiSets")
```

Format

The adjacency matrix dgi.adj has drugs (rows) and a 1 marked where the drug targets a gene (columns). The same information is stored as a list of gene names for each drug in dgi.gs. The data.frame dgiDrug has each drug in a row, with genes they target, and sources for which drug database contained it in these columns:

Drug Cancer drugs

GeneID Gene Symbols of genes targeted by the drug

Synonym Drug synonyms

DGIdb Logical, curated by DGI-db?

DrugBankCurate Logical, curated by Drug Bank list?

gcinfoPan *Gene information data for 7574 genes*

Description

Gene information Data for 7574 genes

Usage

```
data("gcinfoPan")
```

Format

A data frame with gene information

CHROM Chromosome

START Gene start position

STOP Gene stop position

Gene.Sybmol Gene

CodingLength Gene coding length

Examples

```
data(gcinfoPan)
str(gcinfoPan)
```

gcPanCO *Dataset with cqn-normalized RNA gene count data for TCGA colon cancer*

Description

Dataset with cqn-normalized RNA gene count data for TCGA colon cancer

Usage

```
data("gcPanCO")
```

Format

A data frame with genes (rows) and patients (columns).

TCGA-DM-A0XD normalized gene count data for patient TCGA-DM-A0XD

Examples

```
data(gcPanCO)
str(gcPanCO)
```

gcPanTNBC *Dataset with cqn-normalized RNA gene count data for TCGA samples*

Description

Dataset with cqn-normalized RNA gene count data for TCGA samples

Usage

```
data("gcPanTNBC")
```

Format

A data frame with 5490 genes (rows) and 79 patients (columns).

Examples

```
data(gcPanTNBC)
str(gcPanTNBC)
```

genelistPan *Curated Cancer Gene list for panoply*

Description

Curated Cancer Gene list for panoply

Usage

```
data("genelistPan")
```

Format

A data frame with gene information

Examples

```
data(genelistPan)
str(genelistPan)
```

panCircos *plot method for panPath/panDruggable objects*

Description

Create Circos Plot of copy number, RNA expression events, germline and somatic variants. By default, plots DNA mutation (variant/cna) and RNA expression events and shows connections between events in top pathways with a driver gene event. Allows either germline/somatic variants, cna, or all three.

Usage

```
panCircos(panGene, panDrug, caseids, variant = NULL,
  cna = NULL, gcount, gcinfo, tumorpct = 0.5,
  tailPct = 0.05, tailEnd = "upper", minTargets = 1, minPathPct = 0.05,
  minPathSize = 8, minPathways = 1, ...)
```

Arguments

panGene	A data.frame of drug test results from panGeneSets
panDrug	A data.frame of results from panDrugSets
caseids	identifiers of subjects that will match to variant column PatientID, and column name of cna and gcount.
variant	data.frame for somatic and/or germline variants. Must contain columns CHROM, POS, Gene.Symbol, SampleType (Germline, Tumor) and PatientID that must match the caseid and controlid
cna	data.frame with log2 of copy number aberrations. Required columns CHROM, START, STOP, Gene.Symbol, and columns named to match caseid and controlid

gcount	Normalized gene expression counts for patients (columns) at gene symbols (rows). Gene symbols are the row names and should match dataset gcinfoPan gene symbols.
tumorpct	Vector of approximate tumor percentage for each patient samples given in the order they appear in cases and controls. If a single value is passed, it is used for all patients. Copy number mutation events are called based on their approximate tumor percentage, transformed to the log2-ratio copy number call.
tailPct	percentile used to define the most extreme RNA expression events as cancer events per patient; only used when eventOnly=TRUE. Values 0 < tailPct < .5 allowed.
tailEnd	perform differential expression tests that are one-sided ("upper" or "lower"), or two-sided ("both"). Also used with tailPct to include both, upper, or lower expression events to be used as cancer events.
minTargets	Retain drug results for drugs that target at least minTargets genes in the cancer and network genes.
minPathPct	Of reactome pathways affected by gene networks targeted by the drug, only count those with total druggable by the drug in the pathway divided by the total pathway size >= minPathPct.
minPathSize	Of reactome pathways affected by gene networks targeted by the drug, only count the pathways that have total genes >= minPathSize.
minPathways	Filter drug results to those that target genes in at least minPathways, after accounting for minPathPct and minPathSize.

Details

circos plot of the driver gene events in outer bands, and gene expression events in the inner bands, colored blue if they are over- or under-expressed, and if one of the key drugs target them. Across the center circle, we show genes that are connected as the top cancer drivers connected to their druggable in-network genes.

Value

Nothing is returned

---

panDrugGraph *Drug Network Graph*

Description

Graph that shows the connection between the drugs, genes and pathways for the drugs that target the most genes

Usage

```
panDrugGraph(panDrug,usePathIDs = TRUE, ndrugs=8)
```

Arguments

panDrug	A data.frame of drug test results from panDrugSets
usePathIDs	If TRUE the shorter reactome pathway IDs will be used in the graph instead fo the complete name
ndrugs	Minimum number of Drugs to include in the graph

Details

Graph where Drugs=Red Ellipse, Genes=Blue Ellipse, Pathways=Green Rectangles

Value

Nothing is returned

Examples

```
data(clinPanTNBC)
data(genelistPan)
data(cnaPanTNBC)
data(gcPanTNBC)
data(gcinfoPan)
data(variantPanTNBC)
data(dgidbPan)
data(dgiSets)
data(reactome)
patient <- "TCGA-B6-A0IK"
match.index <- which(clinPanTNBC$Vital.Status=="Alive" & clinPanTNBC$PatientID!=patient & clinPanTNBC$ag
ptmatch <- clinPanTNBC[match.index,"PatientID"]
drivGenes <- panGeneSets(caseid=patient, controlid=ptmatch, eventOnly=TRUE, variant=variantPanTNBC,
  cna=cnaPanTNBC, gcount=gcPanTNBC, tumorpct=0.3, tailEnd="upper", tailPct=0.1)
set.seed(1000)
drugResults <- panDrugSets(drivGenes, caseids=patient,
  controlids=ptmatch, gcount=gcPanTNBC,nsim=200, tailEnd="upper")
drugResults<- drugResults[!is.na(drugResults$Pathways) & !is.na(drugResults$N.Cancer.Genes),]
panDrugGraph(drugResults)
```

---

panDrugSets *Tests on drugs for their effectiveness in targeting cancer genes in cases*

Description

Tests on drugs for their effectiveness in targeting activated cancer genes in cases against their matched controls.

Usage

```
panDrugSets(panGene, caseids, controlids, gcount, minTargets=1, minPathPct=.05, minPathSize=8,
        minPathways=1, nsim=1000, tailEnd="both", gene.gs=NULL,
        gene.adj=NULL, drug.gs=NULL, drug.adj=NULL,
        gageCompare = ifelse(length(caseids) > 1, "as.group", "unpaired"))
```

Arguments

panGene	A data.frame of drug test results from panGeneSets
caseids	identifiers of subjects case subjects that will match to variant column PatientID, and column name of cna and gcount.
controlids	identifiers of subjects in row names of variant and column names of gcount who are controls
gcount	Normalized gene expression counts for patients (columns) at gene symbols (rows). Gene symbols are the row names and should match dataset gcinfoPan gene symbols.
minTargets	Retain drug results for drugs that target at least minTargets genes in the cancer and network genes.
minPathPct	Of reactome pathways affected by gene networks targeted by the drug, only count those with total druggable by the drug in the pathway divided by the total pathway size >= minPathPct.
minPathSize	Of reactome pathways affected by gene networks targeted by the drug, only count the pathways that have total genes >= minPathSize.
minPathways	Filter drug results to those that target genes in at least minPathways, after accounting for minPathPct and minPathSize. gene.gs=NULL, gene.adj=NULL, drug.gs=NULL, drug.adj=NULL,
drug.gs	A list of genes, where the genes in each set are targeted by each drug in drug.adj.
drug.adj	Adjacency matrix (values of 1 or 0) for drug-gene interactions with the drug in the row directly impacting the gene in the column. The framework is equipped for coding for other interaction types of the drug and genes.
gene.gs	List of genes per pathway/network to annotate the pathways targeted by the drug. Can be user-defined, but panoply contains reactome sets.
gene.adj	Adjacency matrix of genes connected via a graph of nodes and edges. Panoply contains reactome.adj, which is directed acyclic graph of Reactome nodes (genes) and edges (relationships).
tailEnd	For Drug Network tests, perform differential expression tests that are one-sided ("upper" or "lower"), or two-sided ("both"). Also used with tailPct to include both, upper, or lower expression events to be used as cancer events.
nsim	Number of simulations to perform for evalulating significance of meta drug tests
gageCompare	Character string for the R gage package to specify how to do within-network comparisons of the case(s) versus the controls across genes

Details

If tailEnd is upper, then testing for drugs that target over-expressed genes and gene networks in case(s) versus controls.

Value

A data.frame with the following columns

- Drug: Drug Name
- N.Cancer.Genes: number of cancer genes that are targeted by the drug
- Cancer.Genes: cancer genes that are targeted by the drug

- N.Network.Genes: number of network genes targeted by the drug
- Network.Genes: network genes targeted by the drug
- N.Pathways: number of pathways with genes targeted by the drug (meeting minPathPct and minPathSize settings)
- Pathways: pathway names of pathways with a gene targeted by the drug
- Network: Number of genes directly targeted by the drug
- DNT.pval: p-value comparing gene expression between caseids and controlids patients in the genes directly targeted by the drug
- DMT.Stat: Meta Z-statistic of combined gene networks for any gene targeted by the drug
- DMT.pval: p-value for DMT.Stat Z-statistic, based on multiplying s_i~N(0,1) by gene network test Z-statistics for i=1..nsim
- Z.stat: Meta Z-statistic of combined gene networks for any gene targeted by the drug, weighted by number of genes targeted by the drug divided by gene network size
- PScore: A score used to sort the drug test results, which is a sum of the -log10(pval) of Network.pval and ZSim.pval

Examples

```
data(clinPanTNBC)
data(genelistPan)
data(cnaPanTNBC)
data(gcPanTNBC)
data(gcinfoPan)
data(variantPanTNBC)
data(dgiSets)
data(reactome)
patient <- "TCGA-B6-A0IK"
match.index <- which(clinPanTNBC$Vital.Status=="Alive" & clinPanTNBC$PatientID!=patient & clinPanTNBC$ag
ptmatch <- clinPanTNBC[match.index,"PatientID"]
drivGenes <- panGeneSets(caseid=patient, controlid=ptmatch, eventOnly=TRUE, variant=variantPanTNBC,
  cna=cnaPanTNBC, gcount=gcPanTNBC, tumorpct=0.3, tailEnd="upper", tailPct=0.1)
set.seed(1000)
drugResults <- panDrugSets(drivGenes, caseids=patient,
controlids=ptmatch, gcount=gcPanTNBC,nsim=200, tailEnd="upper",
drug.gs=dgi.gs, drug.adj=dgi.adj, gene.gs=reactome.gs, gene.adj=reactome.adj)

drugResults[1:10,!grepl("^Pathway", names(drugResults))]
```

---

panGeneGraph *Creates reactome network graph for Genes connected to top Drugs*

Description

Uses Reactome data to create graph that shows the connection between Driver Genes and Connected Genes for the top Drugs in panDrug

Usage

```
panGeneGraph(panGene, panDrug, minTargets=2, minPathways=2, ndrugs=4, ndrivers=20, ...)
```

Arguments

panGene	A data.frame of drug test results from panGeneSets
panDrug	A data.frame of drug test results from panDrugSets
minTargets	Retain drug results for drugs that target at least minTargets genes in the cancer and network genes.
minPathways	Filter drug results to those that target genes in at least minPathways, after accounting for minPathPct and minPathSize.
ndrugs	Number of Drugs to use to find Cancer Driver and Network genes in graph
ndrivers	Number of unique Cancer Drive genes in graph
...	Dynamic parameter for the values of additional parameters for the graph plot.

Details

A graph that shows the Cancer Driver Network for the top Drugs in pan Drug. Circles=Druggable, Ellipse=Expressed Druggable Drivers, Red=Cancer Driver Genes, Blue=Network Genes

Value

nothing is returned

Examples

```
data(clinPanTNBC)
data(genelistPan)
data(cnaPanTNBC)
data(gcPanTNBC)
data(gcinfoPan)
data(variantPanTNBC)
data(dgidbPan)
data(dgiSets)
data(reactome)
patient <- "TCGA-B6-A0IK"
match.index <- which(clinPanTNBC$Vital.Status=="Alive" & clinPanTNBC$PatientID!=patient & clinPanTNBC$ag
ptmatch <- clinPanTNBC[match.index,"PatientID"]
drivGenes <- panGeneSets(caseid=patient, controlid=ptmatch, eventOnly=TRUE, variant=variantPanTNBC,
  cna=cnaPanTNBC, gcount=gcPanTNBC, tumorpct=0.3, tailEnd="upper", tailPct=0.1)
set.seed(1000)
drugResults <- panDrugSets(drivGenes, caseids=patient, controlids=ptmatch, gcount=gcPanTNBC,nsim=200, tailEnd

panGeneGraph(drivGenes,drugResults)
```

---

panGeneSets *Test gene networks for a set of cases versus a set of controls*

Description

Test differential gene expression in gene networks for a set of cases versus a set of controls, with gene networks defined by reactome directed graph for cancer genes, and genomic events determined from variant, cna, or gcount.

Usage

```
panGeneSets(caseids, controlids, variant = NULL, cna = NULL,
        gcount, tumorpct = 0.5, tailPct = 0.1, tailEnd = "both",
        eventOnly = FALSE,  gene.adj = NULL, drug.adj = NULL,
        gageCompare = ifelse(length(caseids) > 1, "as.group", "unpaired"))
drivDNA(ids, variant = NULL, cna = NULL, tumorpct = 0.5, gene.adj=NULL)
outRNA(ids, gcount, tailPct = 0.1, tailEnd = "both")
```

Arguments

caseids	identifiers of subjects case subjects that will match to variant column PatientID, and column name of cna and gcount.
controlids	identifiers of subjects in row names of variant and column names of gcount who are controls
ids	identifiers of subjects in row names of variant and column names of gcount and cna matrices
variant	data.frame for somatic and/or germline variants. Must contain columns CHROM, POS, Gene.Symbol, SampleType (Germline, Tumor) and PatientID that must match the caseid and controlid
cna	data.frame with log2 of copy number aberrations. Required columns CHROM, START, STOP, Gene.Symbol, and columns named to match caseid and controlid
gcount	Normalized gene expression counts for patients (columns) at gene symbols (rows). Gene symbols are the row names and should match dataset gcinfoPan gene symbols.
tumorpct	Vector of approximate tumor percentage for each patient samples given in the order they appear in cases and controls. If a single value is passed, it is used for all patients. Copy number mutation events are called based on their approximate tumor percentage, transformed to the log2-ratio copy number call.
tailPct	percentile used to define the most extreme RNA expression events as cancer events per patient; only used when eventOnly=TRUE. Values 0 < tailPct < .5 allowed.
tailEnd	perform differential expression tests that are one-sided ("upper" or "lower"), or two-sided ("both"). Also used with tailPct to include both, upper, or lower expression events to be used as cancer events.
gene.adj	Adjacency matrix of genes connected via a graph of nodes and edges. Panoply contains reactome.adj, which is directed acyclic graph of Reactome nodes (genes) and edges (relationships).
drug.adj	Adjacency matrix (values of 1 or 0) for drug-gene interactions with the drug in the row directly impacting the gene in the column. The framework is equipped for coding for other interaction types of the drug and genes.
gageCompare	Character string telling the gage package, which performs the gene network tests, to perform tests of 1-vs-M ("unpaired") or N-vs-M ("as.group")
eventOnly	Logical, test only gene networks that have a cancer event (variant, cna, and rna-expression outside tailPct)).

Details

If eventOnly=FALSE, then test all gene networks for differential expression between cases and controls. Otherwise, cancer events are determined per gene from variant (any variant in the gene

for the case(s)), cna (copy gain or loss for case(s) given tumorpct), and gcount (outlying expression from tailPct and tailEnd settings) and only those gene networks with a cancer event are tested.

Value

A data.frame with the following columns

- Cancer.Gene: genes whose networks were tested
- Network: total connections of the cacer gene to other genes via reactome adjacency matrix
- Network.pval: p-value comparing gene expression between caseids and controlids patients
- Network.mean: Z-statistic of differential expression between caseids and controlids
- Druggable: Is the cancer gene druggable? 1=TRUE
- Total.Druggable: Number of the connected network genes that are druggable
- NetGenes: Network genes that are over- or under-expressed
- NetDrugs: Drug names that would target the genes in Total.Druggable
- Drugs: Drug names that would target the cancer gene

Examples

```
data(clinPanTNBC)
data(genelistPan)
data(cnaPanTNBC)
data(gcPanTNBC)
data(gcinfoPan)
data(variantPanTNBC)
data(dgidbPan)
data(dgiSets)
data(reactome)

patient <- "TCGA-B6-A0IK"
match.index <- which(clinPanTNBC$Vital.Status=="Alive" & clinPanTNBC$PatientID!=patient & clinPanTNBC$ag
ptmatch <- clinPanTNBC[match.index,"PatientID"]
drivGenes <- panGeneSets(caseid=patient, controlid=ptmatch, eventOnly=TRUE, variant=variantPanTNBC,
  cna=cnaPanTNBC, gcount=gcPanTNBC, tumorpct=0.3, tailEnd="upper", tailPct=0.1)
drivGenes[1:10,1:6]
```

---

variantPanTNBC *Dataset with somatic and germline DNA variants for TCGA patients*

Description

Dataset with somatic and germline DNA variants for TCGA patients patients

Usage

```
data("variantPanTNBC")
```

Format

A data frame with the following columns:

Gene.Symbol Gene Symbol

CHROM Chromosome

POS Start Position

Stop Stop Position

Ref Reference Allele

Alt Alternate Allele

Effect Effect

PatientID Patient ID

V1B.refDP Germline Reference allele depth

V1B.altDP Germline Alternate allele depth

V1T.refDP Tumor Reference allele depth

V1T.altDP Tumor Alternate allele depth

AD AD

GT Genotype

SampleType Sample Type Blood Tissue

Details

Non-synonymous somatic variants, and high-impact germline variants.

Examples

```
data(variantPanTNBC)
str(variantPanTNBC)
```

Index