Package 'POPPATHR'

September 3, 2019

Title Population-based pathway analysis of SNP-SNP coevolution

Version 1.0

Date 2019-08-08

Depends R (>= 3.5.0)

Imports plyr,
dplyr,
ggplot2,
data.table,
stringr,
reshape2,
gdata,
RColorBrewer,
gridExtra,
cowplot,
GenomicRanges,
snpStats,
RCy3

Description A bioinformatics package for determining pathway-level SNP coevolution driven by population-based positive selection.

License GPL (>= 2)

Encoding UTF-8

URL https://github.com/rosscm/POPPATHR

RoxygenNote 6.1.1

R topics documented:

calcFST — *Calculates SNP-level FST values for use in setupGSEArun.R*

Description

Calculates SNP-level FST values for use in setupGSEArun.R

Usage

```
calcFST(genoF, realFam, outDir, outF)
```

Arguments

genoF	(char) path to file with SNP genotype data (PLINK format).
realFam	(char) path to PLINK population coded fam file.
outDir	(char) directory to store output files.
outF	(char) path to write SNP-FST file.

Value

none

getPathStats — *Generates SNP lists per selection-enriched and unenriched pathway as determined by GSEA*

Description

Generates SNP lists per selection-enriched and unenriched pathway as determined by GSEA

Usage

```
getPathStats(genoF, resF, gseaStatF, snp2geneF, realFam, enrichNES = 0.3,
  unenrichNES = 0.1, enrichDir, unenrichDir)
```

Arguments

genoF	(char) path to file with SNP genotype data (PLINK format).
resF	(char) path to files with GSEA results. Strutured to compare results of two population analyses i.e., CEU vs. YRI and CEU vs. LWK.
gseaStatF	(char) path to GSEA statistics file.
snp2geneF	(char) path to SNP-gene mapping file.
realFam	(char) path to PLINK population coded fam file.
enrichNES	(integer) NES cutoff to select validated selection-enriched pathways (default=0.3)
unenrichNES	(integer) NES cutoff to select unenriched pathways (default=0.1)
enrichDir	(char) path to directory to store output files (PLINK files per validated selection-enriched pathway)
unenrichDir	(char) path to directory to store output files (PLINK files per unenriched pathway)

Value

none

`LDstatsBPM` *Calculate selection statistics (LD) and perform exploratory analyses for two sets of variants via R snpStats package*

Description

Calculate selection statistics (LD) and perform exploratory analyses for two sets of variants via R snpStats package

Usage

```
LDstatsBPM(enrichDir, unenrichDir, pop1, pop2, snp2geneF, outDir)
```

Arguments

`enrichDir`	(char) path to selection-enriched pathway SNP lists
`unenrichDir`	(char) path to unenriched pathway SNP lists
`pop1`	(char) character code for the first population (controls).
`pop2`	(char) character code for the second population (cases).
`snp2geneF`	(char) path to file with snp2gene mappings. Output of mapSNP2gene() (found in GWAS2Pathway)
`outDir`	(char) path to output directory

Value

none

`LDstatsWPM` *Calculate selection statistics (LD) and perform exploratory analyses for two sets of variants via R snpStats package*

Description

Calculate selection statistics (LD) and perform exploratory analyses for two sets of variants via R snpStats package

Usage

```
LDstatsWPM(enrichDir, unenrichDir, pop1, pop2, outDir)
```

Arguments

enrichDir	(char) path to selection-enriched pathway SNP lists
unenrichDir	(char) path to unenriched pathway SNP lists
pop1	(char) character code for the first population (controls).
pop2	(char) character code for the second population (cases).
outDir	(char) path to output directory

Value

none

---

plotEmap *Create EnrichmentMap in Cytoscape to visualize predictive pathways*

Description

Create a network where nodes are predictive pathways passing certain cutoff and edges indicate similarity in gene-sets. Pathways are then clustered to identify themes of predictive pathways. Generates one such network for each patient label.

Usage

```
plotEmap(gmtFile, nodeAttrFile, netName = "generic", outDir,
  minScore = 1, maxScore = 10, colorScheme = "cont_heatmap",
  imageFormat = "png", verbose = FALSE, createStyle = TRUE)
```

Arguments

gmtFile	(char) file path to GMT file (generated by writeEMap_file()). NOTE: This needs to be the absolute path name
nodeAttrFile	(list) file path to nodeAttr.txt file (generated by writeEMap_file())
netName	(char) name for network in Cytoscape. Using the patient class name is a good idea. (e.g. SURVIVE_YES and SURVIVE_NO).
outDir	(char) path to directory where file should be stored.
minScore	(int) minimum score of node to show
maxScore	(int) maximum score of node to show
colorScheme	(char) colour scheme for nodes. "cont_heatmap" sets a discrete map ranging from yellow to red for increasing scores. "netDx_ms" is the colour scheme used in the netDx methods paper. This map is (<=6: white; 7-9: orange; 10: red)
imageFormat	(char) one of PNG, PDF, SVG, or JPEG
verbose	(logical) print messages
createStyle	(logical) if generating more than one EMap, set to TRUE for first one and to FALSE for subsequent. Due to limitation in current version of RCy3

Value

Filename of image to which EnrichmentMap is exported. Also side effect of plotting the EnrichmentMap in an open session of Cytoscape.

Examples

```
#refer to writeEMapInput_many.R for working writeEMapInput_many() example
data(featScores)
pathFile <- sprintf("%s/extdata/Human_160124_AllPathways.gmt",
          path.package("netDx.examples"))
pathwayList <- readPathways(pathFile)
pathwayList <- pathwayList[c(1:5)]
netInfoFile <- sprintf("%s/extdata/KIRC_output/inputNets.txt",
     path.package("netDx.examples"))
netTypes <- read.delim(netInfoFile,sep="\t",h=FALSE,as.is=TRUE)
outDir <- paste(tempdir(),"plots",sep="/")
if (!file.exists(outDir)) dir.create(outDir)
EMap_input <- writeEMapInput_many(featScores,pathwayList,
     netTypes,outDir=outDir)
outDir <- paste(getwd(),"plots",sep="/")
if (!file.exists(outDir)) dir.create(outDir)
gmtFile <- EMap_input[[1]][1]
nodeAttrFile <- EMap_input[[1]][2]

# not run because requires Cytoscape to be installed and open
# plotEmap(gmtFile = gmtFile, nodeAttrFile = nodeAttrFile, netName="HighRisk",
# outDir=outDir)
```

---

popPCA *Calculate population stratification between different ancestry groups (e.g., CEU vs YRI) via complete linkage clustering*

Description

Calculate population stratification between different ancestry groups (e.g., CEU vs YRI) via complete linkage clustering

Usage

```
popPCA(genoF, famF, pop1, pop2, dimensions = 3L, outF)
```

Arguments

genoF	(char) path to file with SNP genotype data (PLINK format).
famF	(char) path to PLINK fam file (case/control population coded).
pop1	(char) character code for the first population (controls).
pop2	(char) character code for the second population (cases).
dimensions	(integer, default=3).
outF	(char) path to write PNG image.

Value

none

`recodeFAM` *Recodes PLINK fam file to case/control format by population adapted from SP plink_baseSetup.R (part of GWAS2pathway package)*

Description

Recodes PLINK fam file to case/control format by population adapted from SP plink_baseSetup.R (part of GWAS2pathway package)

Usage

```
recodeFAM(genoF, pop1, pop2, popsF, setSeed = 42L, caseCode = 2,
  ctrlCode = 1, outF = "case-ctrl")
```

Arguments

genoF	(char) path to file with SNP genotype data (PLINK format)
pop1	(char) character code for the first population (controls).
pop2	(char) character code for the second population (cases).
popsF	(char) path to file with population information. Gives the number of samples per population in the dataset.
setSeed	(integer) value for set.seed() before shuffling (default=42).
caseCode	(integer) value for case samples in fam phenotype column.
ctrlCode	(integer) value for control samples in fam phenotype column.
outF	(char) optional - name for fam file (default=case-ctrl). file extension added.

Value

none

`setupGSEArun` *Sets up and runs GSEA using population-based FST table containing GSEA results (Geneset, Size, ES, NES, NominalP, FDR, FWER)*

Description

Sets up and runs GSEA using population-based FST table containing GSEA results (Geneset, Size, ES, NES, NominalP, FDR, FWER)

Usage

```
setupGSEArun(realF, pathF, snp2geneF, setPerm = 1000L,
  snp2genedist = 500000L, minGene = 10L, maxGene = 300L,
  setSeed = 42L, outDir)
```

Arguments

realF	(char) path to file with real SNP association statistics
pathF	(char) path to pathway definitions GMT file
snp2geneF	(char) path to file with snp2gene mappings. Output of mapSNP2gene() (found in GWAS2Pathway).
setPerm	(integer) set cycle of permutations to run default=1000
snp2genedist	(integer) value for GSEA –distance. Max. distance between SNP and gene for their association in snp2geneF
minGene	(integer) value for GSEA –setmin. Min. number of genes in a gene set to be considered
maxGene	(integer) value for GSEA –setmax. Max. number of genes in a gene set to be considered
setSeed	(integer) value for GSEA –seed

Value

none

SNP2gene *Maps SNPs to their nearest genes (for use in setupGSEArun.R) (adapted from map_SNP2gene from SP's GWAS2pathway package)*

Description

NOTE: updated GenomicRanges to version 1.30.1 on 01/30/2018 Requires ignore.strand=TRUE param to properly run distanceToNearest() given unknown strand assignment for SNP array-based genotyping info

Usage

```
SNP2gene(inF, geneF, marg = 0L, outF)
```

Arguments

inF	(char) path to PLINK .bim file (only CHR, SNP, and BP columns considered.
geneF	(char) path to refseq table with header.
marg	(integer) region upstream and downstream of [txStart,txEnd].
outF	(char) path to write snp2gene mapping.

Value

none

Index