Package 'wgsea'

July 2, 2014

Type Package

Title Wilcoxon based gene set enrichment analysis

Version 1.8

Date 2013-09-25

Maintainer Chris Wallace <chris.wallace@cimr.cam.ac.uk>

Author Chris Wallace [aut, cre], Olly Burren [ctb]

Description Non parametric alternative to Kolmogorov-Smirnov based standard GSEA testing.

License GPL

Depends snpStats (>= 1.8.1)

Suggests testthat

Repository CRAN

Date/Publication 2013-09-25 14:29:43

Collate 'Z.value.R' 'genperms.R' 'pairtest.R' 'plotting.R' 'scoretest.R' 'wilcoxon.R' 'wgsea-package.R'

NeedsCompilation no

R topics documented:

wgsea-package *Gene set enrichment analysis using Wilcoxon rank tests*

Description

Gene set enrichment analysis (GSEA) is typically based on tests derived from the Kolmogorov-Smirnov, which is underpowered and a need for simpler methods has been identified. The wgsea package contains functions for conducting GSEA using a Wilcoxon test to test for differences in the distribution of p values between SNPs within the gene set under test and a control set of SNPs.

Details

Package:	wgsea
Type:	Package
Version:	1.0
Date:	2012-04-18
License:	GPL

See the vignette for further details.

Author(s)

Chris Wallace <chris.wallace@cimr.cam.ac.uk>

References

Irizarry, R. A.; Wang, C.; Zhou, Y. & Speed, T. P. Gene set enrichment analysis made simple. Stat Methods Med Res 2009, 18, 565-575

Heinig, M.; Petretto, E.; Wallace, C.; et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature 2010, 467, 460-464

Examples

```
vignette(package="wgsea")
```

genperms *Generate permutations of a phenotype vector*

Description

Given a vector, generate n.perm samples and return a matrix with each permutation in each column.

Usage

```
genperms(pheno, n.perm = 0)
```

Arguments

pheno	a vector to be permuted
n.perm	the number of times to permute

Value

a matrix with dimensions length(pheno) x n.perm.

Author(s)

Chris Wallace <chris.wallace at cimr.cam.ac.uk>

Examples

```
y <- rbinom(50,2,0.3)
genperms(y,4)
```

---

pairtest *Generate p values for each SNP for case-control comparisons.*

Description

A wrapper for the snpStats function single.snp.tests. Generates p values for the association of each SNP with case or control status.

Usage

```
pairtest(case, control, n.perm = 0, pheno.perm = NULL,
  quiet = FALSE)
```

Arguments

case	SnpMatrix object holding genotypes of case subjects
control	SnpMatrix object holding genotypes of control subjects
n.perm	number of permutations of case control status required to generate permuted p value vectors. The default, given by n.perm=0, is not to permute.
pheno.perm	An alternative to specifying n.perm is to supply a matrix of alternative phenotypes, with each column relating to a different permutation.
quiet	set TRUE to suppress the printing of progress dots

Value

If `n.perm=0`, a vector of p values, one for each SNP (each column in the `case` and `control` objects. If `n.perm>0`, a matrix of p values, each column representing the results of a different permutation. a LIST, use

`comp2`	Description of 'comp2'

Author(s)

Chris Wallace

Examples

```
data(for.exercise,package="snpStats")
case <- snps.10[subject.support$cc==1,]
control <- snps.10[subject.support$cc==0,]
summary(pairtest(case,control))
```

`varplot` *Plot theoretical and estimated variances of Wstar*

Description

Given a vector of Wilcoxon statistics generated through permutation, plot theoretical and esitimated variance by cumulative number of permutations

Usage

```
varplot(Wstar, n1, n2)
```

Arguments

`Wstar`	the vector of Wilcoxon values generated by permutation
`n1`	number of items (SNPs) in regions to be tested.
`n2`	number of items (SNPs) in regions the control regions.

Value

None

Author(s)

Chris Wallace <chris.wallace at cimr.cam.ac.uk>

See Also

`wilcoxon`

Examples

```
x<- matrix(exp(-rexp(200000)),nrow=2000)
Wstar<-wilcoxon(p=x,snps.in=1:1000)
varplot(Wstar=Wstar,1000,1000)
```

`wilcoxon` *Wilcoxon test statistic, with optional weights.*

Description

Calculate a Wilcoxon two group rank test statistic, with optional propensity score weighting.

Usage

```
wilcoxon(p, snps.in, weights = NULL, binsize = 0.05)
```

Arguments

`p`	a numeric vector of observed p values from a list of SNPs or a matrix, with each column representing a vector under a different permutation of the dataset.
`snps.in`	a numeric vector indicating which members of p form the test group (their complement form the control group).
`weights`	optional propensity score weights. These are binned according to binsize, and a weight calculated which is inversely proportional to the probability of sampling a member of the test group in that bin.
`binsize`	see weights, above.

Value

A numeric value or, if p is a matrix, a numeric vector.

Author(s)

Chris Wallace

References

Propensity weights are described

Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. Biometrika, 1983, 70, 41-55

Rosenbaum, P. R. Model-based direct adjustment. Journal of the American Statistical Association, 1987, 82, 387-394

See Also

`Z.value`

Examples

```
x <- exp(-rexp(1000)) # uniform
y <- exp(-rexp(1000,0.8)) # skewed towards 0
wilcoxon(p=c(x,y),snps.in=1:1000)

## note, should be equal to
wilcox.test(x,y)
```

`Z.value` *Calculate a Z score from a Wilcoxon statistic and a set of random Wilcoxon statistics*

Description

The mean of a Wilcoxon statistic is unaffected by correlation within the variable under test, but its variance is. This function uses a set of Wilcoxon statistics generated from permuted data to estimate the variance empirically, and thus calculate a Z score.

Usage

```
Z.value(W, Wstar, n.in, n.out)
```

Arguments

`W`	Wilcoxon statistic for observed data.
`Wstar`	A vector of Wilcoxon statistics for a set of permuted data.
`n.in`	The number of items (SNPs) in the regions to be tested.
`n.out`	The number of items (SNPs) in the control regions.

Value

A list with two elements:

Z.theoretical which uses the theoretical mean of the Wilcoxon distribution under the null generated from n.in, n.out above

Z.empirical which uses Wstar to calculate an empirical estimate of the mean of the Wilcoxon distribution under the null

Note

The function can also deal with combining W statistics from multiple strata, as is typical in a meta analysis of GWAS data, using van Elteren's method. Strata may be defined by different geography or different SNP chips.

Author(s)

Chris Wallace

See Also

wilcoxon

Examples

```
x <- exp(-rexp(1000)) # uniform
y <- exp(-rexp(1000,0.8)) # skewed towards 0
W <- wilcoxon(p=c(x,y),snps.in=1:1000)

p.perm <- matrix(sample(c(x,y),replace=TRUE,size=10000),ncol=5)
Wstar <- wilcoxon(p=p.perm,snps.in=1:1000)

Z.value(W=W, Wstar=Wstar, n.in=1000, n.out=1000)
```

Index