Package: aSPU (via r-universe)

September 4, 2024

Version 1.50

Date 2021-6-28

Title Adaptive Sum of Powered Score Test

Description R codes for the (adaptive) Sum of Powered Score ('SPU' and 'aSPU') tests, inverse variance weighted Sum of Powered score ('SPUw' and 'aSPUw') tests and gene-based and some pathway based association tests (Pathway based Sum of Powered Score tests ('SPUpath'), adaptive 'SPUpath' ('aSPUpath') test, 'GEEaSPU' test for multiple traits - single 'SNP' (single nucleotide polymorphism) association in generalized estimation equations, 'MTaSPUs' test for multiple traits - single 'SNP' association with Genome Wide Association Studies ('GWAS') summary statistics, Gene-based Association Test that uses an extended 'Simes' procedure ('GATES'), Hybrid Set-based Test ('HYST') and extended version of 'GATES' test for pathway-based association testing ('GATES-Simes').). The tests can be used with genetic and other data sets with covariates. The response variable is binary or quantitative. Summary; (1) Single trait-'SNP' set association with individual-level data ('aSPU', 'aSPUw', 'aSPUr'), (2) Single trait-'SNP' set association with summary statistics ('aSPUs'), (3) Single trait-pathway association with individual-level data ('aSPUpath'), (4) Single trait-pathway association with summary statistics ('aSPUsPath'), (5) Multiple traits-single 'SNP' association with individual-level data ('GEEaSPU'), (6) Multiple traits-single 'SNP' association with summary statistics ('MTaSPUs'), (7) Multiple traits-'SNP' set association with summary statistics('MTaSPUsSet'), (8) Multiple traits-pathway association with summary statistics('MTaSPUsSetPath').

Author Il-Youp Kwak and others (See Author(s) in each function manual)

Maintainer Il-Youp Kwak <ikwak2@cau.ac.kr>

VignetteBuilder knitr

Suggests rmarkdown, markdown, knitr

Depends R(>= 3.6.0), gee, MASS, mvtnorm, fields, matrixStats

License GPL-3

URL https://github.com/ikwak2/aSPU

RoxygenNote 7.1.0

Repository https://ikwak2.r-universe.dev

RemoteUrl https://github.com/ikwak2/aspu

RemoteRef HEAD

RemoteSha 6d6fab02967d1b9f409426e6e9fa66e667a060de

Contents

---

aSPU   *Sum of Powered Score (SPU) tests and adaptive SPU (aSPU) test for single trait - SNP set association.*

Description

It gives p-values of the SPU tests and aSPU test.

Usage

```
aSPU(
  Y,
  X,
  cov = NULL,
  resample = c("perm", "boot", "sim"),
  model = c("gaussian", "binomial"),
  pow = c(1:8, Inf),
  n.perm = 1000,
  threshold = 10^5
)
```

Arguments

Y	Response or phenotype data. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by k (n : number of observation, k : number of SNPs (or predictors)).
cov	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
resample	Use "perm" for residual permutations, "sim" for simulations from the null distribution, and "boot" for parametric bootstrap.
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
pow	power used in SPU test. A vector of the powers.
n.perm	number of permutations or bootstraps.
threshold	When n.perm is less than the threshold, we use faster version of aSPU but use more memory. (default is 10^5)

Value

A list object, Ts : test statistics for the SPU tests (in the order of the specified pow) and finally for the aSPU test. pvs : p-values for the SPU and aSPU tests.

Author(s)

Il-Youp Kwak, Junghi Kim, Yiwei Zhang and Wei Pan

References

Wei Pan, Junghi Kim, Yiwei Zhang, Xiaotong Shen and Peng Wei (2014) A powerful and adaptive association test for rare variants, Genetics, 197(4), 1081-95

Junghi Kim, Jeffrey R Wozniak, Bryon A Mueller, Xiaotong Shen and Wei Pan (2014) Comparison of statistical tests for group differences in brain functional networks, NeuroImage, 1;101:681-694

See Also

aSPUw

Examples

```
data(exdat)

## example analysis using aSPU test on exdat data.
# increase n.perm for better p.val
## Not run: out <- aSPU(exdat$Y, exdat$X, cov = NULL, resample = "boot",
            model = "binomial", pow = c(1:8, Inf), n.perm = 1000)
## End(Not run)


out$Ts
# This is a vector of Test Statistics for SPU and aSPU tests.
# SPU1 to SPUInf corresponds with the option pow=c(1:8, Inf)
# They are SPU test statistics.
# The last element aSPU is minimum of them, aSPU statistic.

out$pvs
# This is a vector of p-values for SPU and aSPU tests.
# SPU1 to SPUInf corresponds with the option pow=c(1:8, Inf)
# They are p-values for corresponding SPU tests.
# The last element is p-value of aSPU test.
```

---

aSPUd *Adaptive Sum of powered score (SPU) tests (SPU and aSPU) based on asymptotic distribution.*

Description

It gives the p-values of the SPU(1), SPU(2) and minP tests and aSPUd test based on the asymptotic distribution.

Usage

```
aSPUd(Y, X, cov = NULL, model = c("gaussian", "binomial"))
```

Arguments

Y	phenotype data. It can be disease lables; =0 for controls, =1 for cases. or It can be any quantitative traits. Vector with length n (number of observations)
X	genotype data; each row for a subject, and each column for an SNP. The value of each element is the # of the copies for an allele. Matrix with dimension n by k (n : number of observation, k : number of genotype data)
cov	covariates. Matrix with dimension n by p (n :number of observation, p : number of covariates)
model	Use "gaussian" for quantitative trait (Default) , and Use "binomial" for binary trait.

Value

p-values for SPU(1), SPU(2), minP tests and aSPU test.

References

Gongjun Xu, Lifeng Lin, Peng Wei and Wei Pan (2016) An adaptive two-sample test for high-dimensional means, Biometrika (2016) 103 (3): 609-624.

See Also

aSPU

Examples

```
data(exdat)
out <- aSPUd(exdat$Y, exdat$X, cov = NULL, model = "binomial")
out
```

---

aSPUpath	*Pathway based Sum of Powered Score tests (SPUpath) and adaptive SPUpath (aSPUpath) test for single trait - pathway association.*

Description

It gives p-values of the SPUpath tests and aSPUpath test.

Usage

```
aSPUpath(
  Y,
  X,
  cov = NULL,
  model = c("binomial", "gaussian"),
  snp.info,
  gene.info,
  pow = c(1:8, Inf),
  pow2 = c(1, 2, 4, 8),
  n.perm = 200,
  usePCs = F,
  varprop = 0.95
)
```

Arguments

Y	Response or phenotype data. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by k (n : number of observation, k : number of SNPs (or predictors)).
cov	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.
pow	SNP specific power(gamma values) used in SPUpath test.
pow2	GENE specific power(gamma values) used in SPUpath test.
n.perm	number of permutations.
usePCs	indicating whether to extract PCs and then use PCs of X.
varprop	the proportion of the variations explained (cutoff) that determines how many top PCs to use.

Value

P-values for SPUpath tests and aSPUpath test.

Author(s)

Il-Youp Kwak and Wei Pan

References

Wei Pan, Il-Youp Kwak and Peng Wei (2015) A Powerful and Pathway-Based Adaptive Test for Genetic Association With Common or Rare Variants, The American Journal of Human Genetics, 97, 86-98

See Also

[simPathAR1Snp](#)

Examples

```
## Not run: dat1<-simPathAR1Snp(nGenes=20, nGenes1=5, nSNPlim=c(1, 20),
          nSNP0=1, LOR=.2, n=100, MAFlim=c(0.05, 0.4), p0=0.05 )
## End(Not run)


# p-values of SPUpath and aSPUpath tests.
## Not run: p.pathaspu<- aSPUpath(dat1$Y, dat1$X, snp.info = dat1$snp.info,
          gene.info = dat1$gene.info,
          model = "binomial", pow=1:8, pow2=c(1, 2, 4, 8), n.perm=1000)
## End(Not run)

p.pathaspu
## pow = 1:8 and pow2 = 1,2,4,8
## So, there are 8*4 = 32 SPUpath p-values.
## SPUpathi,j corresponds pow = i , pow2 = j
## The last element, aSPUpath gives aSPUpath p-value.
```

---

aSPUpath2   *Pathway based Sum of Powered Score tests (SPUpath) and adaptive*
                              SPUpath (aSPUpath) test for single trait - pathway association. (vec-
                              tor version, fast when n is large)

Description

It gives p-values of the SPUpath tests and aSPUpath test. Faster than aSPUsPath function when n is large (N > 10^4).

Usage

```
aSPUpath2(
  Y,
  X,
  cov = NULL,
  model = c("binomial", "gaussian"),
  snp.info,
  gene.info,
```

```
    pow = c(1:8, Inf),
    pow2 = c(1, 2, 4, 8),
    n.perm = 200,
    usePCs = F,
    varprop = 0.95
)
```

Arguments

Y	Response or phenotype data. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by k (n : number of observation, k : number of SNPs (or predictors)).
cov	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.
pow	SNP specific power(gamma values) used in SPUpath test.
pow2	GENE specific power(gamma values) used in SPUpath test.
n.perm	number of permutations.
usePCs	indicating whether to extract PCs and then use PCs of X.
varprop	the proportion of the variations explained (cutoff) that determines how many top PCs to use.

Value

P-values for SPUpath tests and aSPUpath test.

Author(s)

Il-Youp Kwak and Wei Pan

References

Wei Pan, Il-Youp Kwak and Peng Wei (2015) A Powerful and Pathway-Based Adaptive Test for Genetic Association With Common or Rare Variants, The American Journal of Human Genetics, 97, 86-98

See Also

simPathAR1Snp

Examples

```
## Not run: dat1<-simPathAR1Snp(nGenes=20, nGenes1=5, nSNPlim=c(1, 20),
        nSNP0=1, LOR=.2, n=100, MAFlim=c(0.05, 0.4), p0=0.05 )
## End(Not run)

# p-values of SPUpath and aSPUpath tests.
## Not run: p.pathaspu<- aSPUpath2(dat1$Y, dat1$X, snp.info = dat1$snp.info,
        gene.info = dat1$gene.info,
        model = "binomial", pow=1:8, pow2=c(1, 2, 4, 8), n.perm=1000)
## End(Not run)
#p.pathaspu
## pow = 1:8 and pow2 = 1,2,4,8
## So, there are 8*4 = 32 SPUpath p-values.
## SPUpathi,j corresponds pow = i , pow2 = j
## The last element, aSPUpath gives aSPUpath p-value.
```

---

aSPUr	*Robust Sum of powered score (SPU) tests and aSPU test for a quantitative trait*

Description

The test is based on the Huber loss function and using the parametric bootstrap for inference (i.e. bootstrapping residuals).

Usage

```
aSPUr(Y, X, cov = NULL, pow = c(1:8, Inf), B = 1000, C = 1.345)
```

Arguments

Y	a vector of quantitative traits (QTs).
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by k (n : number of observation, k : number of SNPs (or predictors)).
cov	Covariates. A matrix with dimension n by k2 (n :number of observation, k2 : number of covariates).
pow	power used in SPUr test. A vector of the powers.
B	number of bootstraps.
C	Constant in huber loss function. C = 1.345 is chosen to maintain a high efficiency for a Normal error.

Value

p-values of the SPUr tests in the order of supplied pow values; finally, the p-value of the aSPUr test (that combines the SPUs tests with pow by taking their min P-value and adjust for multiple testing).

Author(s)

Yiwei Zhang and Wei Pan

References

Peng Wei, Ying Cao, Yiwei Zhang, Zhiyuan Xu, Il-Youp Kwak, Eric Boerwinkle, Wei Pan (2016) On Robust Association Testing for Quantitative Traits and Rare Variants, G3, 6(12) 3941-3950.

See Also

aSPU

Examples

```
data(exdat)

## example analysis using aSPU test on exdat data.

QT <- jitter(exdat$Y)

out <- aSPUr(Y = QT, X = exdat$X, cov = NULL, B = 100)
out
## This is a vector of p-values for SPUr and aSPUr tests.
## SPU1 to SPUInf corresponds with the option pow=c(1:8, Inf)
## They are p-values for corresponding SPUr tests.
## The last element is p-value of aSPUr test.
```

---

aSPUs   *Sum of Powered Score (SPUs) tests and adaptive SPU (aSPUs) test for*
                             single trait - SNP set association with GWAS summary statistics.

Description

It gives p-values of the SPUs tests and aSPUs test with GWAS summary statistics.

Usage

```
aSPUs(Zs, corSNP, pow = c(1:8, Inf), n.perm = 1000, Ps = FALSE, prune = TRUE)
```

Arguments

Zs	Z-scores for each SNPs. It could be P-values if the Ps option is TRUE.
corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
pow	power used in SPU test. A vector of the powers.

n.perm	number of permutations or bootstraps.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.
prune	if it is TRUE, do pruing before the test using pruneSNP function.

Value

A list object, Ts : test statistics for the SPU tests (in the order of the specified pow) and finally for the aSPU test. pvs : p-values for the SPUs and aSPUs tests.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2015) Adaptive Gene- and Pathway-Trait Association Testing with GWAS Summary Statistics, Bioinformatics, 32(8), 1178-1184.

See Also

aSPUw aSPU aSPUsPath

Examples

```
data(kegg9)
## example analysis using aSPUM test.
g <- kegg9$gene.info[1,1]  #  SOAT1
## Take snps mapped on gene "SOAT1" from the information of gene.info and snp.info.
snps <- which( ( kegg9$snp.info[,2] == kegg9$gene.info[kegg9$gene.info[,1] == g, 2] ) &
               (kegg9$snp.info[,3] > kegg9$gene.info[kegg9$gene.info[,1] == g, 3] ) &
               (kegg9$snp.info[,3] < kegg9$gene.info[kegg9$gene.info[,1] == g, 4] )  )
## Take subsets
newP <- kegg9$nP[snps] ;
ldsub <- kegg9$ldmatrix[snps, snps];

## Get p-value for gene SOAT1. Read vignette for details.
out <- aSPUs(newP, corSNP=ldsub , pow=c(1:8, Inf), n.perm=100, Ps=TRUE)

out$Ts
# This is a vector of Test Statistics for SPUM and aSPUM tests.
# SPUs1 to SPUsInf corresponds with the option pow=c(1:8, Inf)
# They are SPUs test statistics.
# The last element aSPUs is minimum of them, aSPUs statistic.

out$pvs
# This is a vector of p-values for SPUs and aSPUs tests.
# SPUs1 to SPUsInf corresponds with the option pow=c(1:8, Inf)
# They are p-values for corresponding SPUs tests.
# The last element is p-value of aSPUs test.
```

---

aSPUsD *Sum of Powered Score (SPUs) tests and adaptive SPU (aSPUs) test for single trait - SNP set association with GWAS summary statistics (distribution based).*

Description

It gives p-values of the SPUs tests and aSPUs test with GWAS summary statistics.

Usage

```
aSPUsD(Zs, corrSNP, Ps = FALSE)
```

Arguments

Zs   Z-scores for each SNPs. It could be P-values if the Ps option is TRUE.

corrSNP   Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).

Ps   TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.

Value

pvs : p-values for the SPUsD and aSPUsD tests.

Author(s)

Il-Youp Kwak and Wei Pan

References

Gongjun Xu, Lifeng Lin, Peng Wei and Wei Pan (2016) An adaptive two-sample test for high-dimensional means, Biometrika (2016) 103 (3): 609-624. Il-Youp Kwak, Wei Pan (2015) Adaptive Gene- and Pathway-Trait Association Testing with GWAS Summary Statistics, Bioinformatics, 32(8), 1178-1184

See Also

[aSPUs](#) [aSPU](#)

Examples

```
data(kegg9)
## example analysis using aSPUM test.
g <- kegg9$gene.info[1,1]  #  SOAT1
## Take snps mapped on gene "SOAT1" from the information of gene.info and snp.info.
snps <- which( ( kegg9$snp.info[,2] == kegg9$gene.info[kegg9$gene.info[,1] == g, 2] ) &
                (kegg9$snp.info[,3] > kegg9$gene.info[kegg9$gene.info[,1] == g, 3] ) &
                (kegg9$snp.info[,3] < kegg9$gene.info[kegg9$gene.info[,1] == g, 4] ) )
```

```
## Take subsets
newP <- kegg9$nP[snps] ;
ldsub <- kegg9$ldmatrix[snps, snps];
## Get p-value for gene SOAT1. Read vignette for details.
out <- aSPUsD(newP, corrSNP=ldsub, Ps=TRUE)

out
```

---

aSPUsPath	*Pathway based Sum of Powered Score tests (SPUsPath) and adaptive SPUpath (aSPUsPath) test for single trait - pathway association with GWAS summary statistics.*

Description

It gives p-values of the SPUsPath tests and aSPUsPath test with GWAS summary statistics.

Usage

```
aSPUsPath(
  Zs,
  corSNP,
  pow = c(1, 2, 4, 8, Inf),
  pow2 = c(1, 2, 4, 8),
  snp.info,
  gene.info,
  n.perm = 1000,
  Ps = FALSE,
  prune = TRUE
)
```

Arguments

Zs	Z-scores for each SNPs. It could be P-values if the Ps option is TRUE.
corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
pow	SNP specific power(gamma values) used in SPUpath test.
pow2	GENE specific power(gamma values) used in SPUpath test.
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.
n.perm	number of permutations.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.
prune	if it is TRUE, do pruing before the test using pruneSNP function.

Value

P-values for SPUMpath tests and aSPUMpath test.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2015) Adaptive Gene- and Pathway-Trait Association Testing with GWAS Summary Statistics, Bioinformatics, 32(8):1178-1184

See Also

[aSPUs](#)

Examples

```
data(kegg9)

# p-values of SPUpath and aSPUpath tests.
out.a <- aSPUsPath(kegg9$nP, corSNP = kegg9$ldmatrix, pow=c(1:8, Inf),
                   pow2 = c(1,2,4,8),
                   snp.info=kegg9$snp.info, gene.info = kegg9$gene.info,
                   n.perm=5, Ps = TRUE)

out.a
```

---

aSPUsPath2   *Pathway based Sum of Powered Score tests (SPUsPath) and adaptive*
                                    SPUpath (aSPUsPath) test for single trait - pathway association with
                                    GWAS summary statistics. (vector version, fast when n is large)

Description

It gives p-values of the SPUsPath tests and aSPUsPath test with GWAS summary statistics. Faster than aSPUsPath function when n is large (N > 10^4).

Usage

```
aSPUsPath2(
  Zs,
  corSNP,
  pow = c(1, 2, 4, 8, Inf),
  pow2 = c(1, 2, 4, 8),
  snp.info,
  gene.info,
```

```
    n.perm = 1000,
    Ps = FALSE,
    prune = TRUE
)
```

Arguments

Zs	Z-scores for each SNPs. It could be P-values if the Ps option is TRUE.
corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
pow	SNP specific power(gamma values) used in SPUpath test.
pow2	GENE specific power(gamma values) used in SPUpath test.
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.
n.perm	number of permutations.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.
prune	if it is TRUE, do pruing before the test using pruneSNP function.

Value

P-values for SPUMpath tests and aSPUMpath test.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2015) Adaptive Gene- and Pathway-Trait Association Testing with GWAS Summary Statistics, Bioinformatics, 32(8):1178-1184

See Also

aSPUs

Examples

```
data(kegg9)

# p-values of SPUpath and aSPUpath tests.
## Not run: out.a <- aSPUsPath2(kegg9$nP, corSNP = kegg9$ldmatrix, pow=c(1:8, Inf),
                pow2 = c(1,2,4,8),
                snp.info=kegg9$snp.info, gene.info = kegg9$gene.info,
                n.perm=1000, Ps = TRUE)
## End(Not run)
```

```
#out.a
```

---

aSPUw   *Inverse variance weighted Sum of Powered Score tests (SPUw) and adaptive SPUw (aSPUw) test for single trait - SNP set association.*

Description

It gives the p-values of the SPUw tests and aSPUw test based on the permutations of the residuals or simulations from the null distripution.

Usage

```
aSPUw(
  Y,
  X,
  cov = NULL,
  resample = c("perm", "boot", "sim"),
  model = c("gaussian", "binomial"),
  pow = c(1:8, Inf),
  n.perm = 1000
)
```

Arguments

Y	Response or phenotype data. It can be a disease indicator; =0 for controls, =1 for cases. Or it can be a quantitative trait. A vector with length n (number of observations).
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by k (n : number of observation, k : number of SNPs (or predictors)).
cov	Covariates. A matrix with dimension n by p (n :number of observation, p : number of covariates).
resample	Use "perm" for residual permutations, "sim" for simulations from the null distribution, and "boot" for parametric bootstrap.
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
pow	power used in SPU test. A vector of the powers.
n.perm	number of permutations or bootstraps.

Value

A list object, Ts : Test Statistics for the SPUw and aSPUw test. pvs : p-values for the SPUw and aSPUw test.

Author(s)

Il-Youp Kwak, Junghi Kim and Wei Pan

References

Junghi Kim, Jeffrey R Wozniak, Bryon A Mueller, Xiaotong Shen and Wei Pan (2014) Comparison of statistical tests for group differences in brain functional networks, Neuroimage, 1;101:681-694

See Also

[aSPU](#)

Examples

```
data(exdat)

# increase n.perm for accurate p-value
## Not run: out <- aSPUw(exdat$Y, exdat$X, pow = c(1:8, Inf), n.perm = 1000)


out$Ts
# This is a vector of Test Statistics for SPU and aSPU tests.
# SPU1 to SPUInf corresponds with the option pow=c(1:8, Inf)
# They are SPU test statistics.
# The last element aSPU is minimum of them, aSPU statistic.

out$pvs
# This is a vector of p-values for SPU and aSPU tests.
# SPU1 to SPUInf corresponds with the option pow=c(1:8, Inf)
# They are p-values for corresponding SPU tests.
# The last element is p-value of aSPU test.
```

---

estcov	*estcov*

Description

Estimate the covariance matrix of multiple traits based on their (null) summary Z-scores.

Usage

```
estcov(allZ, Ps = FALSE)
```

Arguments

allZ	matrix of summary Z-scores for all SNP. each row for SNP; each column for single trait.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.

Value

estimated correlation matrix.

Author(s)

Junghi Kim, Yun Bai and Wei Pan

References

Junghi Kim, Yun Bai and Wei Pan (2015) An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics, Genetic Epidemiology, 8:651-663

See Also

MTaSPUs minP

Examples

```
# -- n.snp: number of SNPs
# -- n.trait: number of traits
# -- n.subject: number of subjects

n.snp <- 100
n.traits <- 10
n.subjects <- 1000
traits <- matrix(rnorm(n.subjects*n.traits), n.subjects, n.traits)
v <- cov(traits)
allZ <- rmvnorm(n.snp, Sigma=v)
colnames(allZ) <- paste("trait", 1:n.traits, sep="")
rownames(allZ) <- paste("snp", 1:n.snp, sep="")


r <- estcov(allZ)
MTaSPUs(Z = allZ, v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
MTaSPUs(Z = allZ[1,], v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
minP(Zi= allZ[1,], r = r)
```

---

exdat   *An example data set*

Description

The exdat data set is list of three objects. exdat$Y is a vector of length 1000, 500 0s and 500 1s. exdat$X is a matrix of 1000 by 10. This simulated X matrix is assumed to be rare variants. All eliments are 0, 1 or 2. exdat$SNP0indx is a vector of length 10. The values are 1 or 0, 0 indicate corresponding column of X matrix have no association with Y vector.

Usage

```
data(exdat)
```

Examples

```
data(exdat)
exdat$X[c(1:10, 501:510),]
exdat$Y[c(1:10, 501:510)]

# 3rd and 10th column of X have no association with Y
exdat$SNP0indx
```

GATES2	*Gene-based Association Test that uses an extended Simes procedure (GATES) for single trait - SNP set association*

Description

Get the p-value of GATES. Usually it is used to get genomewise p-values. This function is taken from postgwas package. There is a little modification of the code GATES in postgwas package. 1) The approximated matrix may have negative eigen value, we modified it not to have negative values; 2) we added one more return (the key gene location) for Hyst method.

Usage

```
GATES2(ldmatrix, p)
```

Arguments

ldmatrix   numeric. A correlation matrix of SNPs, dimensions matching the p and snps arguments.

p   p-value for each SNPs.

Value

A p-value of GATES and the key gene location (to be used by Hyst).

Author(s)

Milan Hiersche(taken from pastgwas package), Il-Youp Kwak(modified a little)

References

Miao-Xin Li, Hong-Sheng Gui, Johnny S.H. Kwan and Pak C. Sham (2011) GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure The American Journal of Human Genetics 88, 283-293

See Also

[Hyst](#) [GatesSimes](#)

Examples

```
#simula <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 20), nSNP0=1:3,
#                          LOR=.2, rholim=c(0,0),
#                          n=30, MAFlim=c(0.05, 0.4), p0=0.05)
#Ps <- getlogitp(simula$Y, simula$X)

## get correlation of SNPs using controls
#ldmat <- cor(simula$X[ simula$Y == 0, ])

#o.pvec = order(Ps)
# ldmat <- ldmat[o.pvec, o.pvec]
#(gatesp <- GATES2(ldmat, sort(Ps))[1])
```

GatesSimes   *GATES-Simes test for single trait - pathway association.*

Description

Get the p-value of GATES-Simes. It uses an extended Simes procedure to combine GATES p-values
across multiple genes in a pathway.

Usage

```
GatesSimes(pvec, ldmatrix, snp.info, gene.info)
```

Arguments

pvec	p-values for each SNP.
ldmatrix	numeric. A correlation matrix of SNPs with dimensions matching the length of pvec (the number of SNPs).
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.

Value

A p-value.

Author(s)

Il-Youp Kwak and Wei Pan

References

Hongsheng Gui, Miaoxin Li, Pak C Sham and Stacey S Cherny (2011) Comparisons of seven algorithms for pathway analysis using the WTCCC Crohn's Disease BMC Research Notes, 4:386

See Also

Hyst GATES2

Examples

```
#simula <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 20), nSNP0=1:3,
#                         LOR=.2, rholim=c(0,0),
#                         n=30, MAFlim=c(0.05, 0.4), p0=0.05)
#logitp <- getlogitp(simula$Y, simula$X)

## get correlation of SNPs using controls
#ldmat <- cor(simula$X[ simula$Y == 0, ])
#out <- GatesSimes(pvec = logitp, ldmatrix = ldmat, snp.info = simula$snp.info,
#                  gene.info = simula$gene.info)
#out
```

GEEaSPU   *The SPU and aSPU tests for multiple traits - single SNP association in generalized estimating equations.*

Description

It gives p-values of the GEESPU tests and GEEaSPU test.

Usage

```
GEEaSPU(
  traits,
  geno,
  Z = NULL,
  model = c("binomial", "gaussian"),
  gamma = c(1:8, Inf),
  n.sim = 1000,
  corstr = "independence"
)
```

Arguments

traits	trait matrix. The row for individuals and the column for traits.
geno	A matrix of genetic information.

Z	covariates.
model	Use "gaussian" for a quantitative trait, and use "binomial" for a binary trait.
gamma	power used in GEEaSPU test. A vector of the powers.
n.sim	number of simulations.
corstr	a character string specifying the correlation structure. The following are permitted: "independence", "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M" and "unstructured"

Value

p-values for the GEE-SPU and GEE-aSPU test.

Author(s)

Junghi Kim, Wei Pan and Il-Youp Kwak

References

Yiwei Zhang, Zhiyuan Xu, Xiaotong Shen, Wei Pan (2014) Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. Neuroimage. 96:309-25

Examples

```
traits <- matrix(rnorm(100*5, 0,1), ncol=5)
Z <- rnorm(100, 2, 0.5)
geno <- rbinom(100, 2, 0.5)
out <- GEEaSPU(traits, geno, Z = NULL, model = "gaussian",
   gamma = c(1:8,Inf), n.sim = 100)
```

---

getlogitp	*Get the p-values of the multiple SNPs based on the Wald test in a logistic regresion model.*

Description

Get p-value using logistic regresion for each of the multiple SNPs

Usage

```
getlogitp(Y, X)
```

Arguments

Y	Response or phenotype data. It can be a disease indicator; =0 for controls, =1 for cases.
X	Genotype or other data; each row for a subject, and each column for an SNP (or a predictor). The value of each SNP is the # of the copies for an allele. A matrix with dimension n by p (n : number of observation, p : number of SNPs (or predictors)).

Value

p-values for each SNPs.

See Also

[GATES2](#) [GatesSimes](#) [Hyst](#)

Examples

```
#simula <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 20), nSNP0=1:3,
#                           LOR=.2, rholim=c(0,0),
#                           n=10, MAFlim=c(0.05, 0.4), p0=0.05)
#logitp <- getlogitp(simula$Y, simula$X)
```

Hyst	*HYST (Hybrid Set-based Test) for single trait - pathway association*

Description

Get a p-value using HYST.

Usage

```
Hyst(pvec, ldmatrix, snp.info, gene.info)
```

Arguments

pvec	p-values for each SNP.
ldmatrix	numeric. A correlation matrix of SNPs with dimensions matching the length of pvec (the number of SNPs).
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.

Value

A p-value.

Author(s)

Il-Youp Kwak and Wei Pan

References

Miao-Xin Li, Johnny S.H. Kwan and Pak C. Sham (2012) HYST: A Hybrid Set-Based Test for Genome-wide Association Studies, with Application to Protein-Protein Interaction-Based Association Analysis The American Journal of Human Genetics, 91, 478-488.

See Also

GatesSimes GATES2

Examples

```
#simula <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 20), nSNP0=1:3,
#                            LOR=.2, rholim=c(0,0),
#                            n=30, MAFlim=c(0.05, 0.4), p0=0.05)
#logitp <- getlogitp(simula$Y, simula$X)

## get correlation of SNPs using controls
#ldmat <- cor(simula$X[ simula$Y == 0, ])
#out <- Hyst(pvec = logitp, ldmatrix = ldmat, snp.info = simula$snp.info,
#            gene.info = simula$gene.info)
```

---

kegg9   *A subset of coronary artery disease (CAD) data*

Description

Data on coronary artery disease, myocardial infarction have been contributed by CARDIoGRAM-plusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG

The data set contains P value data for coronary artery disease (CAD). We mapped these SNPs to the 9th KEGG pathway.

gene.info is a contains 16 gene information of 9th Kegg pathway. The 1st column is the name of the gene, 2nd column is the chromosome number, 3rd column is where the gene starts and 4th column is where the gene ends.

snp.info contains 330 snp informations mapped on this Kegg pathway. The 1st column is the rsID of the SNP, 2nd column is the chromosome number and the 3rd column is the location of the SNP.

Ps is a vector containing p-value information for mapped SNPs.

The 'PPs' is a list object contains the SNP information for each genes. For example, PPs$CEL contains snp.info matrix mapped on gene 'CEL'. The 1st column is the rsID of the SNP, 2nd

column is the chromosome number, the 3rd column is the location of the SNP and the 4th column is p-value of the SNP.

nP is a subvector of Ps. These are SNPs on reference population(Hapmap CEU phase 2 data, downloadable from plink, SNPs with MAF less than 5 percent omitted)

ldmatrix is a correaltion matrix of SNPs from the reference population.

Usage

```
data(kegg9)
```

References

Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H et al. (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 43: 333-338

Examples

```
data(kegg9)

## gene informations
kegg9$gene.info

## SNPs mapped on 3rd and 4th gene in 9th Kegg pathway
kegg9$PPs[3:4]

## snp information
kegg9$snp.info

## The 1st 10 P-values of SNPs mapped on 9th Kegg pathway.
kegg9$nP[1:10]

## correlation matrix among those SNPs
kegg9$ldmatrix[1:10,1:10]
```

---

minP   *minP test.*

Description

Return exact minP test p-value for multiple traits - single SNP association.

Usage

```
minP(Zi, r)
```

Arguments

Zi	a vector of summary Z-scores for single SNP
r	estimated correlation matrix based on the summary Z-scores (output of estcov)

Value

return exact minP test

Author(s)

Junghi Kim, Yun Bai and Wei Pan

References

Junghi Kim, Yun Bai and Wei Pan (2015) An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics, Genetic Epidemiology, 8:651-663

See Also

estcov MTaSPUs

Examples

```
# -- n.snp: number of SNPs
# -- n.trait: number of traits
# -- n.subject: number of subjects

n.snp <- 100
n.traits <- 10
n.subjects <- 1000
traits <- matrix(rnorm(n.subjects*n.traits), n.subjects, n.traits)
v <- cov(traits)
allZ <- rmvnorm(n.snp, Sigma=v)
colnames(allZ) <- paste("trait", 1:n.traits, sep="")
rownames(allZ) <- paste("snp", 1:n.snp, sep="")


r <- estcov(allZ)
MTaSPUs(Z = allZ, v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
MTaSPUs(Z = allZ[1,], v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
minP(Zi= allZ[1,], r = r)
```

MTaSPUs	*The SPU and aSPU tests for multiple traits - single SNP association with GWAS summary statistics.*

Description

SNP based adaptive association test for multiple phenotypes with GWAS summary statistics.

Usage

```
MTaSPUs(Z, v, B, pow, transform = FALSE, Ps = FALSE)
```

Arguments

Z	matrix of summary Z-scores, SNPs in rows and traits in columns. Or a vector of summary Z-scores for a single snp
v	estimated correlation matrix based on the summary Z-scores (output of estcov)
B	number of Monte Carlo samples simulated to compute p-values, the maximum number of MC simulations is 1e8
pow	power used in SPU test. A vector of the powers.
transform	if TRUE, the inference is made on transformed Z
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.

Value

compute p-values for SPU(gamma) i.e. pow=1:8, and infinity aSPU, based on the minimum p-values over SPU(power) each row for single SNP

Author(s)

Junghi Kim, Yun Bai and Wei Pan

References

Junghi Kim, Yun Bai and Wei Pan (2015) An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics, Genetic Epidemiology, 8:651-663

See Also

minP estcov

Examples

```
# -- n.snp: number of SNPs
# -- n.trait: number of traits
# -- n.subject: number of subjects

n.snp <- 100
n.traits <- 10
n.subjects <- 1000
traits <- matrix(rnorm(n.subjects*n.traits), n.subjects, n.traits)
v <- cov(traits)
allZ <- rmvnorm(n.snp, Sigma=v)
colnames(allZ) <- paste("trait", 1:n.traits, sep="")
rownames(allZ) <- paste("snp", 1:n.snp, sep="")


r <- estcov(allZ)
MTaSPUs(Z = allZ, v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
MTaSPUs(Z = allZ[1,], v = r, B = 100, pow = c(1:4, Inf), transform = FALSE)
minP(Zi= allZ[1,], r = r)
```

MTaSPUsSet	*gene-Multitrait Sum of Powered Score (MTSPUsSet) tests and adaptive MTSPUsSet (MTaSPUsSet) test for multi trait - SNP set association with GWAS summary statistics.*

Description

It gives p-values of the MTSPUsSet tests and MTaSPUsSet test with GWAS summary statistics.

Usage

```
MTaSPUsSet(
  Zs,
  corSNP,
  corPhe,
  pow = c(1, 2, 4, 8),
  pow2 = c(1, 2, 4, 8),
  n.perm = 5000,
  Ps = FALSE,
  prune = TRUE
)
```

Arguments

Zs   Z-score matrix. row represent SNPs and column represent traits. It could be
              P-values if the Ps option is TRUE.

corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
corPhe	Correlation matirx of phenotypes to be tested; Estimated from Z-scores.
pow	SNP specific power(gamma values) used in MTSPUsSet test.
pow2	GENE specific power(gamma values) used in MTSPUsSet test.
n.perm	number of permutations or bootstraps.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.
prune	if it is TRUE, do pruing before the test using pruneSNP function.

Value

A vector object, MTSPUsSet test P values and MTaSPUsSet P value.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2017) Gene- and pathway-based association tests for multiple traits with GWAS summary statistics, Bioinformatics, 33(1), 64-71.

Examples

```
data(SAMD11)
attach(SAMD11)
## example analysis using MTaSPUsSet test.
(outFZ <- MTaSPUsSet(ZsF, corSNP=corSNPF, corPhe = corPheF,
     pow=c(1,2,4,8),  pow2 = c(1,2,4,8), n.perm=10, Ps=FALSE))
```

MTaSPUsSetPath	*Multitraits - Pathway based Sum of Powered Score tests (MTSPUsSet-Path) and adaptive MTSPUsSetpath (MTaSPUsSetPath) test for multi trait - pathway association with GWAS summary statistics.*

Description

It gives p-values of the MTSPUsSetPath tests and MTaSPUsSetPath test with GWAS summary statistics.

Usage

```
MTaSPUsSetPath(
  Zs,
  corPhe,
  corSNP,
  pow1 = c(1, 2, 4, 8),
  pow2 = c(1, 2, 4, 8),
  pow3 = c(1, 2, 4, 8),
  snp.info,
  gene.info,
  n.perm = 1000,
  Ps = FALSE,
  prune = TRUE
)
```

Arguments

Zs	Z-score matrix. row represent SNPs and column represent traits. It could be P-values if the Ps option is TRUE.
corPhe	Correlation matirx of phenotypes to be tested; Estimated from Z-scores.
corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
pow1	SNP specific power(gamma values) used in MTSPUsSetPath test.
pow2	GENE specific power(gamma values) used in MTSPUsSetPath test.
pow3	Trait specific power(gamma values) used in MTSPUsSetPath test.
snp.info	SNP information matrix, the 1st column is SNP id, 2nd column is chromosome #, 3rd column indicates SNP location.
gene.info	GENE information matrix, The 1st column is GENE id, 2nd column is chromosome #, 3rd and 4th column indicate start and end positions of the gene.
n.perm	number of permutations.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.
prune	if it is TRUE, do pruing before the test using pruneSNP function.

Value

P-values for MTSPUsSetpath tests and MTaSPUsSetpPath test.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2017) Gene- and pathway-based association tests for multiple traits with GWAS summary statistics, Bioinformatics, 33(1), 64-71

See Also

MTaSPUsSet

Examples

```
Zs <- cbind ( c( 0.3, 0.2, 0.1,0.5,1.2),
          c(-.1, .3,-.1,.1,1.2) )

varSNP = cbind( c( 1, .1,0, 0, .11),
               c(.1,1, 0, 0, 0),
               c(0,0, 1, 0, 0),
               c(0,0, 0, 1, 0),
               c(.11,0,0,0,1) )

varPhe = cbind( c( 1, -.1),
               c(-.1,1) )

gene.info = data.frame( Gnm = c( "G1", "G2"), chr = c(1,3),
                      loc1 = c(0, 0), loc2 = c(10,10) )
snp.info = data.frame( rsid = c("rs1", "rs2", "rs3", "rs4", "rs5"),
                chr = c(1,1,3,3,3), loc = c(1,2,1,2,3) )
 out <- MTaSPUsSetPath(Zs, corPhe = varPhe, corSNP=varSNP,
            n.perm = 100, snp.info = snp.info, gene.info = gene.info)
out
```

MTaSPUsSetScore	*gene-Multitrait Sum of Powered Score (MTSPUsSetScore) tests and adaptive MTSPUsSet (MTaSPUsSet) test for multi trait - SNP set association with GWAS summary statistics. (Score version added)*

Description

It gives p-values of the MTSPUsSet tests MTScore and MTaSPUsSet_Score test with GWAS summary statistics.

Usage

```
MTaSPUsSetScore(
  Zs,
  corSNP,
  corPhe,
  pow = c(1, 2, 4, 8),
  pow2 = c(1, 2, 4, 8),
  n.perm = 5000,
  Ps = FALSE
)
```

Arguments

Zs	Z-score matrix. row represent SNPs and column represent traits. It could be P-values if the Ps option is TRUE.
corSNP	Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).
corPhe	Correlation matirx of phenotypes to be tested; Estimated from Z-scores.
pow	SNP specific power(gamma values) used in MTSPUsSetScore test.
pow2	GENE specific power(gamma values) used in MTSPUsSetScore test.
n.perm	number of permutations or bootstraps.
Ps	TRUE if input is p-value, FALSE if input is Z-scores. The default is FALSE.

Value

A vector object, MTSPUsSet test, MTScore test P values and MTaSPUsSet_Score P value.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2017) Gene- and pathway-based association tests for multiple traits with GWAS summary statistics, Bioinformatics. 33(1), 64-71

Examples

```
data(SAMD11)
attach(SAMD11)
## example analysis using aSPUM test.
(outFZ <- MTaSPUsSetScore(ZsF, corSNP=corSNPF, corPhe = corPheF,
     pow=c(1,2,4,8),  pow2 = c(1,2,4,8), n.perm=10, Ps=FALSE))
```

---

plotPmat	*Image plot of P-value matrix.*

Description

It gives a P-value image for a given P-value matrix.

Usage

```
plotPmat(
  Ps,
  zlim = NULL,
  main = NULL,
  yt = NULL,
  xlab = "SNPs",
  thresh = -log(5e-08, 10),
  trait.names = NULL
)
```

Arguments

Ps	P-value matrix. Row represent traits, column represent SNPs
zlim	-log 10 transformed p value range.
main	main title.
yt	it controls the location of trait names.
xlab	lable for x axis. Default is "SNPs".
thresh	Default is set to widely used genome wise threshold -log(5e-8,10).
trait.names	A vector of trait names.

Value

Image of P-values matrix.

Author(s)

Il-Youp Kwak and Wei Pan

References

Il-Youp Kwak, Wei Pan (2017) Gene- and pathway-based association tests for multiple traits with GWAS summary statistics, Bioinformatics. 33(1), 64-71

See Also

[aSPUs](#)

Examples

```
## Say we have 3 traits and their p-values at 5 SNPs.
Ps <- rbind( c(0.001, 0.4, 0.5, 0.00000001, .1),
             c(0.03, 0.3, 0.3, 0.00001, .2),
             c(0.01, 0.2, 0.4, 0.001, .0001) )

## We can visualize it using plotPmat function.
plotPmat(Ps)
```

---

pruneSNP *Prune SNPs (related to aSPUs, aSPUsPath, MTaSPUsSet)*

Description

When correlation matirx have highly correlated SNPs, the performance of aSPUs, aSPUsPath and MTaSPUsSet are not very good. Do pruning using this function.

Usage

```
pruneSNP(corSNP, rup = 0.95, rdown = -0.95)
```

Arguments

corSNP   Correlation matirx of the SNPs to be tested; estimated from a reference panel (based on the same set of the reference alleles as used in calculating Z-scores).

rup   pruning criteria, erase one SNP when correlation between two SNPs are larger than this value.

rdown   pruning criteria, erase one SNP when correlation between two SNPs are smaller than this value.

Value

a list object pruned.corSNP and to.erase. pruned.corSNP is pruned correlation matrix. to.erase is SNP index to erase to get purned object. (i.e. corSNP[-to.erase, -to.erase] = pruned.corSNP)

Author(s)

Il Youp Kwak

See Also

aSPUs aSPUsPath MTaSPUsSet

Examples

```
data(kegg9)
## example analysis using aSPUM test.
g <- kegg9$gene.info[1,1]  #  SOAT1
## Take snps mapped on gene "SOAT1" from the information of gene.info and snp.info.
snps <- which( ( kegg9$snp.info[,2] == kegg9$gene.info[kegg9$gene.info[,1] == g, 2] ) &
                 (kegg9$snp.info[,3] > kegg9$gene.info[kegg9$gene.info[,1] == g, 3] ) &
                 (kegg9$snp.info[,3] < kegg9$gene.info[kegg9$gene.info[,1] == g, 4] ) )
## Take subsets
newP <- kegg9$nP[snps] ;
ldsub <- kegg9$ldmatrix[snps, snps];
```

```
prSNP <- pruneSNP(ldsub)
newP <- newP[-prSNP$to.erase]
ldsub <- ldsub[-prSNP$to.erase, -prSNP$to.erase ]

## Get p-value for gene SOAT1. Read vignette for details.
out <- aSPUs(newP, corSNP=ldsub , pow=c(1:8, Inf), n.perm=100, Ps=TRUE)

out$Ts
# This is a vector of Test Statistics for SPUM and aSPUM tests.
# SPUs1 to SPUsInf corresponds with the option pow=c(1:8, Inf)
# They are SPUs test statistics.
# The last element aSPUs is minimum of them, aSPUs statistic.

out$pvs
# This is a vector of p-values for SPUs and aSPUs tests.
# SPUs1 to SPUsInf corresponds with the option pow=c(1:8, Inf)
# They are p-values for corresponding SPUs tests.
# The last element is p-value of aSPUs test.
```

SAMD11   *An example data for MTaSPUsSet test*

Description

Genetic Investigation of ANthropometric Traits (GIANT) consortium data contain P-values of 2.7 million SNPs with six anthropometric traits that are well established to represent body size and shape: height, weight, BMI, waist circumference (WC), hip circumference(HIP).

We mapped SNPs on gene SAMD11. This subdata contains P values, Z scores of SNPs mapped on gene SAMD11. It also contains correlation among SNPs and correlation among phenotypes for demostrated example of MTaSPUsSet test.

Usage

```
data(SAMD11)
```

References

Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, Strawbridge R, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JM, et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature 518: 187-196.

Examples

```
data(SAMD11)

## Z-scores for man
SAMD11$ZsM
```

```
## P-values for man
SAMD11$PsM

## correlation among SNPs for man
SAMD11$corSNPM

## correlation among phenotypes for man
SAMD11$corPheM
```

---

simPathAR1Snp   *Simulating a pathway with multiple SNPs.*

Description

It gives a simulated SNPs consisting of multiple genes in a pathway. Each SNPs from a latent
multivariate Gaussian variable with an AR1 correlation structure.

Usage

```
simPathAR1Snp(
  nGenes = 10,
  nGenes1 = 5,
  nSNPs = NULL,
  ncSNPs = NULL,
  nSNPlim = c(1, 20),
  nSNP0 = 1:3,
  LOR = 0.3,
  n = 100,
  MAFlim = c(0.05, 0.4),
  rholim = c(0, 0),
  p0 = 0.05,
  noncausal = FALSE
)
```

Arguments

nGenes	The number of total genes.
nGenes1	The number of causal genes.
nSNPs	A vector, length matched with total number of genes. Each elements of vector indicate the number of SNPs in the gene. Default is nSNPs = NULL, in this case the number of nSNPs randomly selected from nSNPlow to nSNPup.
ncSNPs	A vector, length matched with total number of genes. Each elements of vector indicate the number of causal SNPs in the gene. Default is ncSNPs = NULL, in this case the number of ncSNPs are randomly selected from nSNP0.
nSNPlim	If nSNPs = NULL, the number of SNPs in Gene randomly selected in nSNPlim.

nSNP0	If ncSNPs = NULL, the number of causal SNPs in Gene randomly selected from nSNP0. Default is 1:3.
LOR	Association in log OR between a causal SNP and outcome.
n	# of cases (= # of controls).
MAFlim	MAF's of the SNPs are drawn from Unif(MAFlim[1], MAFlim[2]).
rholim	the SNPs in eahc gene are from a latent Normal variable with a AR(rho) corr structure, rho's are drawn from Unif(rholim[1], rholim[2]); the SNPs in diff genes are independant.
p0	background disease prevalence;i.e. intercept=log(p0/(1-p0)).
noncausal	exclude causal SNPs if TRUE, it is the simulation set up d in the paper(Pan et al 2015).

Value

a list of the binary outcome Y (=0 or 1) and SNPs (=0, 1 or 2); Y is a vector of length 2n; X is a matrix of 2n by nSNP.

See Also

[aSPUpath](#)

Examples

```
# Simulation set up A a) in the paper (Pan et al 2015)
## Not run:  simula <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 20), nSNP0=1:3,
                         LOR=.2, rholim=c(0,0),
                         n=100, MAFlim=c(0.05, 0.4), p0=0.05)
## End(Not run)


# Simulation set up A b) in the paper
#simulb <- simPathAR1Snp(nGenes=20, nGenes1=1, nSNPlim=c(1, 100), nSNP0=1:3,
#                         LOR=.2, rholim=c(0,0),
#                         n=100, MAFlim=c(0.05, 0.4), p0=0.05)
```

---

someGs   *An example data for MTaSPUsSet test*

Description

Genetic Investigation of ANthropometric Traits (GIANT) consortium data contain P-values of 2.7 million SNPs with six anthropometric traits that are well established to represent body size and shape: height, weight, BMI, waist circumference (WC), hip circumference(HIP).

We mapped SNPs on gene LCORL, RASA2, STK33 and RPGRIP1L. This subdata contains P values, correlation matrices for demostrated example of MTaSPUsSet test.

Usage

```
data(someGs)
```

References

Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, Strawbridge R, Pers TH, Fischer K, Justice AE, Workalemahu T, Wu JM, et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature 518: 187-196.

Examples

```
data(someGs)

## P-values for LCORL
someGs$LCORL[[1]]

## correlation matrix for LCORL
someGs$LCORL[[2]]

## P-values for RASA2
someGs$RASA2[[1]]

## correlation matrix for LCORL
someGs$RASA2[[2]]

# See vignettes for more details.
```

Index