

Package ‘GenCAT’

July 1, 2015

Type Package

Title Genetic Class Association Testing (GenCAT)

Version 1.0.1

Date 2015-06-12

Author Eric Reed, Sara Nu<c3><b1>ez, Jing Qian, Andrea Foulkes

Maintainer Eric Reed <reeder@bu.edu>

Description Implementation of the genetic class level association testing (GenCAT) method from SNP level association data. Refer to: ``Qian, J., Reed, E., Nunez, S., Qu, L., Reilly, M., Foulkes, A. (2015) Testing class-level genetic associations using single-element summary statistics. (Under review).".

License GPL-2

Suggests snpStats, knitr

Depends R (>= 2.10), stats, dplyr, doParallel, ggplot2, foreach, parallel, methods

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2015-07-01 16:06:25

R topics documented:

GenCAT-package	2
CardioData	3
CardioMapped	4
coords	5
GenCAT	6
GenCAT_manhattan	7
geno	9
map2class	10

Index	12
--------------	-----------

GenCAT-package

*Genetic Class Association Testing (GenCAT)***Description**

Implementation of the genetic class level association testing (GenCAT) method from SNP level association data. Refer to: "Qian, J., Reed, E., Nunez, S., Qu, L., Reilly, M., Foulkes, A. (2015) Testing class-level genetic associations using single-element summary statistics. (Under review).".

Details

Package: GenCAT
 Type: Package
 Title: Genetic Class Association Testing (GenCAT)
 Version: 1.0.1
 Date: 2015-06-12
 Author: Eric Reed, Sara Nuñez, Jing Qian, Andrea Foulkes
 Maintainer: Eric Reed <reeder@bu.edu>
 Description: Implementation of the genetic class level association testing (GenCAT) method from SNP level association
 License: GPL-2
 Suggests: snpStats, knitr
 Depends: stats, dplyr, doParallel, ggplot2, foreach, parallel, methods
 VignetteBuilder: knitr

Index of help topics:

CardioData	GWAS data from CARDIoGRAM on CAD
CardioMapped	CARDIoGRAM data mapped to classes
GenCAT	Running GenCAT
GenCAT-package	Genetic Class Association Testing (GenCAT)
GenCAT_manhattan	Create Manhattan Plot of GenCAT Results
coords	Protein coding gene coordinates (GRCh37/hg19)
geno	Genotype Data
map2class	Mapping SNPs to Classes

Author(s)

Eric Reed, Sara Nuñez, Jing Qian, Andrea Foulkes

Maintainer: Eric Reed <reeder@bu.edu>

References

Qian, J., Reed, E., Nunez, S., Qu, L., Reilly, M., Foulkes, A. (2015) Testing class-level genetic associations using single-element summary statistics. (Under review).

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hickey, G., Hinrichs, A. S., Hubley, R., Karolchik, D., Learned, K., Lee, B. T., Li, C. H., Miga, K.H. , Nguyen, N., Paten, B., Raney, B.J., Smit, A. F., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M., Kent, W.J. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D670-81. PMID: 25428374; PMC: PMC4383971

Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P. and et. al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 43 333-338.

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.

CardioData

GWAS data from CARDIoGRAM on CAD

Description

This is a subset GWAS results for risk of CAD.

Usage

```
data("CardioData")
```

Format

A data frame with 95157 observations on the following 6 variables.

SNP a character vector

effect_allele a character vector

other_allele a character vector

testStat a numeric vector

chr a numeric vector

position a numeric vector

Details

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

Source

www.CARDIOGRAMPLUSC4D.ORG.

References

Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P. and et. al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 43 333-338.

Examples

```
data("CardioData")
```

CardioMapped

CARDIoGRAM data mapped to classes

Description

CARDIoGRAM test statistics mapped to protein coding genes on chromosomes 13, 14, and 15.

Usage

```
data("CardioMapped")
```

Format

A data frame with 98582 observations on the following 7 variables.

SNP a character vector
effect_allele a character vector
other_allele a character vector
testStat a numeric vector
chr a numeric vector
position a numeric vector
class a character vector

Details

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

Source

www.CARDIOGRAMPLUSC4D.ORG.

References

Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P. and et. al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 43 333-338.

Examples

```
data("CardioMapped")
```

`coords`*Protein coding gene coordinates (GRCh37/hg19)*

Description

Protein coding gene coordinates adapted from UCSC Genomes Browser KnownCanonical track.

Usage

```
data("coords")
```

Format

A data frame with 18054 observations on the following 4 variables.

`chr` a numeric vector

`start` a numeric vector

`stop` a numeric vector

`class` a character vector

Details

These protein coding gene coordinates were adapted from the KnownCanonical track from the UCSC Genome Browser, GRCH37

Source

<https://genome.ucsc.edu/cgi-bin/hgTables>

References

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hickey, G., Hinrichs, A. S., Hubley, R., Karolchik, D., Learned, K., Lee, B. T., Li, C. H., Miga, K.H., Nguyen, N., Paten, B., Raney, B.J., Smit, A. F., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M., Kent, W.J. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D670-81. PMID: 25428374; PMC: PMC4383971

Examples

```
data("coords")
```

GenCAT

*Running GenCAT***Description**

This function runs the GenCAT approach on a data frame of SNP level test statistics.

Usage

```
GenCAT(SNPdata, genoData, snpInfo, pcCutoff = 0.95,
workers = getOption("mc.cores", 2L))
```

Arguments

SNPdata	A data table containing SNP level test statistics, chromosome, class, and allele assignment. This must have the column names 'SNP', 'testStat', 'chr', 'class', 'effect_allele', and 'other_allele'.
genoData	This is an object of class 'SnpMatrix' with genotype data to run GenCAT
snpInfo	If 'genoData' is provided this is a table which includes SNP data with columns 'chr', 'SNP', 'position', 'A1', and 'A2'."
pcCutoff	Threshold for the cumulative poportion of eigenvalues given by eigen() for the SNP-wise correlation matrix. By default this is set to 0.95.
workers	Specifies the number of parallel processes to run.

Details

Reference population should be representative of population used to generate SNP level test statistics. Genotype data is input using the arguments `genoData` and `snpInfo`. The former should be of class `SnpMatrix`. The latter should be a data frame similar to that of the `map` element of the list created by the `read.plink` or `read.pedfile` functions from the `snpStats` package.

The `pcCutoff` argument specifies the proportion of variability in the SNP wise correlation matrix used in the eigen decomposition and estimation of independent test statistics.

Value

An object of class "GenCATtest" with five elements.

GenCAT	A data frame containing GenCAT test results.
Used	A data frame containing SNP information for SNPs used in GenCAT analysis
notFound	A data frame containing SNP information for SNPs for which there was no reference genotype data
unMatched	A data frame containing SNP information for SNPs in which an allele assignment wasn't in reference genotype
TransStats	A data frame containing transformed test statistics from eigen decomposition of each class

Author(s)

Eric Reed, Sara Nunez, Jing Qian, Andrea Foulkes

References

Qian, J., Reed, E., Nunez, S., Qu, L., Reilly, M., Foulkes, A. (2015) Testing class-level genetic associations using single-element summary statistics. (Under review).

See Also

[snpStats](#)

Examples

```
#####  
#Running GenCAT  
#####  
data("CardioMapped")  
  
#Subset CardioMapped to decrease CPU time  
CardioMappedSub<-CardioMapped[CardioMapped$chr < 15,]  
set.seed(1)  
CardioMappedSub<-CardioMappedSub[sample(1:nrow(CardioMappedSub), 100),]  
  
print(head(CardioMappedSub))  
  
library(snpStats)  
data('geno')  
  
genoData<-geno$genotypes  
snpInfo<-geno$map  
  
print(str(genoData))  
  
colnames(snpInfo)<-c('chr', 'SNP', 'gen.dist', 'position', 'A1', 'A2')  
print(head(snpInfo))  
  
GenCATtest <- GenCAT(CardioMappedSub, genoData=genoData, snpInfo = snpInfo)
```

GenCAT_manhattan

Create Manhattan Plot of GenCAT Results

Description

This function will create a Manhattan Plot from output of GenCAT function

Usage

```
GenCAT_manhattan(GenCATout, sigThresh = NULL, highlightPosi = FALSE,
  labelPosi = FALSE, sepChr = 8e+05,
  plotTitle = "Manhattan Plot of GenCAT Results")
```

Arguments

GenCATout	An object of class, GenCATtest.
sigThresh	P-value threshold to highlight classes with strong signal from GenCAT test
highlightPosi	logical. If TRUE, classes with GenCAT p-value less than sigThresh will be shown in blue
labelPosi	logical. If TRUE, classes with GenCAT p-value less than sigThresh will be labelled.
sepChr	Specifies the space to put between chromosomes on the plot.
plotTitle	Character expression for plot title.

Details

GenCATtest is the class of the output of the GenCAT function.

Author(s)

Eric Reed, Sara Nunez, Jing Qian, Andrea Foulkes

Examples

```
#####
#Running GenCAT
#####
data("CardioMapped")

#Subset CardioMapped to decrease CPU time
CardioMappedSub<-CardioMapped[CardioMapped$chr < 15,]
set.seed(1)
CardioMappedSub<-CardioMappedSub[sample(1:nrow(CardioMappedSub), 100),]

library(snpStats)
data('geno')

genoData<-geno$genotypes
snpInfo<-geno$map

colnames(snpInfo)<-c('chr', 'SNP', 'gen.dist', 'position', 'A1', 'A2')
print(head(snpInfo))

GenCATtest <- GenCAT(CardioMappedSub, genoData=genoData, snpInfo = snpInfo)

#####
#Create Manhattan Plot
```



```
#####

print(str(GenCATtest))
GenCAT_manhattan(GenCATtest, sigThresh = (0.05/nrow(GenCATtest$GenCAT)),
highlightPosi = TRUE, labelPosi = TRUE)
```

geno *Genotype Data*

Description

This is genotype data as read in by read.plink (snpStats). It contains genotypes for 99 individuals from the CEU population (Utah residents with northern and western European ancestry). It only includes chromosomes 13-15.

Usage

```
data("geno")
```

Format

The format is: List of 3 \$ genotypes: Formal class 'SnpMatrix' [package "snpStats"] with 1 slot ..
 ..@ .Data: raw [1:99, 1:84195] 01 03 02 03 attr(*, "dimnames")=List of 2 \$: chr [1:99] "CEU_1" "CEU_2" "CEU_3" "CEU_4" \$: chr [1:84195] "rs624673" "rs9511877" "rs638773" "rs9511880" ... \$ fam : 'data.frame': 99 obs. of 6 variables: ..\$ pedigree: chr [1:99] "CEU_1" "CEU_2" "CEU_3" "CEU_4"\$ member : int [1:99] 1 1 1 1 1 1 1 1 1\$ father : logi [1:99] NA NA NA NA NA NA\$ mother : logi [1:99] NA NA NA NA NA NA\$ sex : logi [1:99] NA NA NA NA NA NA\$ affected: logi [1:99] NA NA NA NA NA NA ... \$ map : 'data.frame': 84195 obs. of 6 variables: ..\$ chromosome: int [1:84195] 13 13 13 13 13 13 13 13 13 13\$ snp.name : chr [1:84195] "rs624673" "rs9511877" "rs638773" "rs9511880"\$ cM : logi [1:84195] NA NA NA NA NA NA\$ position : num [1:84195] 19743996 19744070 19744848 19745096 19745251\$ allele.1 : chr [1:84195] "G" "A" "A" "G"\$ allele.2 : chr [1:84195] "A" "G" "G" "T" ...

Details

This data was filtered for: minor allele frequency < 0.01, call rate < 0.95, and Hardy-Weinberg Equilibrium $p < 10^{-6}$.

Source

<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20130502/>

References

The 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422), 56-65.

Examples

```
data("geno")
```

`map2class`*Mapping SNPs to Classes*

Description

The function maps SNPs to classes based on genome coordinates

Usage

```
map2class(coords, SNPs, extend.boundary = 0)
```

Arguments

<code>coords</code>	A data frame containing chromosome and upper and lower coordinates for a each class to be mapped. This must include column names 'chr', 'start', 'stop', and 'class'.
<code>SNPs</code>	A data frame containing genome coordinates of SNPs. This must include column names 'SNP', 'chr', 'position'.
<code>extend.boundary</code>	Number of base pairs to extend the class coordinates for mapping.

Details

Genomic coordinates defining the class boundaies in object, `coords`, and SNP location in object, `SNPs` should be annotated within the same build.

Note: That if using the `map2class` function to map SNPs to classes to use in the `GenCAT` function, then additional columns for `effect_allele` and `other_allele` can be used.

Value

A data frame with SNP information including class label.

Author(s)

Eric Reed, Sara Nunez, Jing Qian, Andrea Foulkes

Examples

```
data('CardioData')
data('coords')

#####
#Mapping SNPs to genes
#####
print(head(coords))

#Subset CardioData to decrease CPU time
CardioDataSub<-CardioData[CardioData$chr < 15,]
set.seed(1)
CardioDataSub<-CardioDataSub[sample(1:nrow(CardioDataSub), 100),]

print(head(CardioDataSub))

CardioMapped<-map2class(coords, CardioDataSub, extend.boundary = 5000)
```

Index

*Topic **\textasciitildekwd1**

GenCAT, [6](#)

map2class, [10](#)

*Topic **\textasciitildekwd2**

GenCAT, [6](#)

map2class, [10](#)

*Topic **datasets**

CardioData, [3](#)

CardioMapped, [4](#)

coords, [5](#)

geno, [9](#)

*Topic **package**

GenCAT-package, [2](#)

CardioData, [3](#)

CardioMapped, [4](#)

coords, [5](#)

GenCAT, [6](#)

GenCAT-package, [2](#)

GenCAT_manhattan, [7](#)

geno, [9](#)

map2class, [10](#)

snpStats, [7](#)