

Package ‘REGENT’

August 19, 2015

Title Risk Estimation for Genetic and Environmental Traits

Version 1.0.6

Date 2015-08-18

Author Daniel J.M. Crouch, Graham H.M. Goddard & Cathryn M. Lewis

Maintainer Daniel Crouch <djmcrouch@gmail.com>

Description Produces population distribution of disease risk and statistical risk categories, and predicts risks for individuals with genotype information.

Depends R (>= 2.14.0)

License GPL

NeedsCompilation no

Repository CRAN

Date/Publication 2015-08-19 13:49:39

R topics documented:

REGENT-package	2
EnvironmentalA	2
EnvironmentalB	3
GeneticA	3
GeneticB	4
Inds	4
REGENT.model	5
REGENT.predict	7
Index	10

REGENT-package

Risk Estimation for Genetic and Environmental Traits

Description

Provides risk estimation and categorisation for populations and individuals

Details

Package: REGENT
Type: Package
Version: 1.0.6
Date: 2015-08-18
License: GPL
LazyLoad: yes

Author(s)

Daniel Crouch, Graham Goddard & Cathryn Lewis.

Maintainer: Daniel Crouch - djmrcouch@gmail.com

References

Crouch, Goddard & Lewis (2011)

Goddard & Lewis, Risk categorization for complex disorders according to genotype relative risk and precision in parameter estimates (2010)

See Also

[REGENT.model,REGENT.predict,GeneticA,GeneticB,EnvironmentalA,EnvironmentalB,Inds](#)

EnvironmentalA*Example file for single level environmental factors*

Description

Example data for Crohns disease in correct input format. Also as text file "EnvironmentalA.txt" in the data folder for this package. data from Calkins, BM (1989).

Usage

```
data("REGENT")
```

Format

data frame

References

from Calkins, BM (1989) A meta-analysis of the role of smoking in inflammatory bowel disease

EnvironmentalB

Example file for single multiple level environmental factors

Description

Example data for Crohns disease in correct input format. Also as text file "EnvironmentalB.txt" in the data folder for this package. *PLEASE NOTE that the multilevel smoking data in EnvironmentalB is entirely artificial.*

Usage

```
data("REGENT")
```

Format

data frame

GeneticA

Example file for SNPs conferring multiplicative risks

Description

Example data for Crohns disease in correct input format. Also as text file "GeneticA.txt" in the data folder for this package. Data is from Franke et al (2010),

Usage

```
data("REGENT")
```

Format

data frame

References

Franke et al. (2010), Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci

GeneticB

Example file for SNPs conferring additive risks

Description

Example data for Crohns disease in correct input format. Also as text file "GeneticB.txt" in the data folder for this package. Data is from Franke et al (2010),

Usage

```
data("REGENT")
```

Format

data frame

References

Franke et al. (2010), Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci

Inds

Example of individual file for REGENT.predict

Description

Columns for risk factors, rows for individuals. Entries refer to genotypes (number of risk alleles as defined in the locus file, ie can also be protective alleles) or exposure levels

Usage

```
data("REGENT")
```

Format

data frame

REGENT.model

*REGENT.model***Description**

REGENT.model provides the population distribution of risk and proportion of the population in each risk category based on genetic(SNP) and environmental exposures.

Usage

```
REGENT.model(AnalysisName,LocusFile=NULL,EnvFile=NULL
,prev=0.001,cv=0.05,alpha=0.05,sims=100000
,indsims=100000,SmallSampAdjust=0.5,BaseRange=0.01
,PlotMax=5,Block=100)
```

Arguments

AnalysisName	String, must be provided. Output files will be named according to this argument. Running multiple analyses with the same name will cause previous files to be overwritten.
LocusFile	File path string. Location of file containing table of SNP input data. Required columns should have headers SNP, MAF, Ncase, Ncontrol. Risks should either be provided in one column with header RR, or two columns with headers RR_het and RR_hom. Other columns may be present but will not be used in the analysis. Each SNP is a row. Additional columns may be provided but will be ignored.
EnvFile	File path string. Location of file containing table of environmental risk data. Required columns should have headers Factor, Exposure, RR, SE. If multiple exposure levels exist, then the columns should be named Factor, RR1, Exposure1, SE1, RR2, Exposure2, SE2, etc. Each factor is a row. Additional columns may be provided but will be ignored
prev	Prevalance of the disease or trait. Default 0.001.
cv	Coefficient of variation. Default 0.05.
alpha	One minus the desired width of confidence intervals around multilocus risk estimates. Default 0.05 giving 95 percent confidence intervals.
sims	Number of simulations to perform for each single factor risk estimate, for obtaining confidence intervals. Default 100000.
indsims	Number of individuals in the simulated population, for obtaining multilocus genotype frequencies. Default 100000
SmallSampAdjust	Adjustment for small sample sizes, when calculating the standard error of homozygous risk genotypes. Default 0.5

BaseRange	Proportion of population used to calculate the baseline risk (the risk closest to the average in the population). This is to avoid choosing rare, uncertain risk estimates by chance. Default 0.01.
PlotMax	Value at which to truncate the Y-axis of risk distribution plots. High risks are typically rare and of less interest when assessing the distribution in the population. Default 5.
Block	Number of multilocus genotypes held in memory during confidence interval calculation. Higher values should decrease computation time. We advise increasing this substantially (10000+) on high performance systems. Default 100.

Details

4 files are created by REGENT.model.A)All model details, inputs and log information are written to the main output file which is named after the argument provided to AnalysisName.B)Colour and C)greyscale plots of the risk distribution are also provided, and D)the raw data used to create these in a text file.

See the example folder included in this package for the correct input file format.

Value

A list including elements

categories	Table giving upper and lower boundaries for each risk category: Reduced, Average, Elevated and High.
baseline	Single value specifying the baseline risk before rebasing; required when passing the object to REGENT.predict
LocusFile	Table of genetic data used for analysis. NULL if argument LocusFile was set to NULL.
EnvFile	Table of environmental data used for analysis. NULL if argument EnvFile was set to NULL.

Author(s)

Graham Goddard, Daniel Crouch and Cathryn Lewis. Email: djmcrouch@gmail.com

See Also

[REGENT.predict,GeneticA,GeneticB,EnvironmentalA,EnvironmentalB,Inds](#)

Examples

```
library(REGENT)

#Load example data from package

data("REGENT")

write.table(GeneticA,file="GeneticA.txt")
```

```

write.table(GeneticB,file="GeneticB.txt")
write.table(EnvironmentalA,file="EnvironmentalA.txt")
write.table(EnvironmentalB,file="EnvironmentalB.txt")

x=REGENT.model(AnalysisName="Example",LocusFile="GeneticA.txt",EnvFile="EnvironmentalA.txt")

x

```

REGENT.predict

REGENT.predict

Description

REGENT.predict takes genotype and exposure information for individuals and calculates their absolute and relative risk of disease, and categorises them as reduced, average, elevated or high risk based on the risk categorisation model determined by REGENT.model.

Usage

```

REGENT.predict(AnalysisName,model,ind,prev=0.001,cv=0.05,sims=100000,Block=100,alpha=0.05,
SmallSampAdjust=0.5)

```

Arguments

AnalysisName	String, must be provided. The output file will be named according to this argument, with the suffix "_Predictions.txt". Running multiple analyses with the same name will cause previous files to be overwritten.
model	Must be provided. Either a file path string giving the location of a file created by REGENT.model (the main file containing model information), or a variable containing the object returned by REGENT.model.
ind	Must be provided. File path giving the location of individual file, which should have columns for each risk factor (with header of SNP names or Factor names as provided to REGENT.model) and a row for each individual. Genotypes are encoded 0, 1 or 2 describing the number of copies of the risk allele (as defined in the model). Environmental factors are encoded 0, 1, 2, 3 etc. depending on how many exposure levels were modelled. The row header contains individual names.
prev	Prevalance of the disease or trait. Default 0.001.
cv	Coefficient of variation. Default 0.05.
sims	Number of simulations to perform for each single factor risk estimate, for obtaining confidence intervals. Default 100000.
Block	Number of multilocus genotypes held in memory during confidence interval calculation. Higher values should decrease computation time. We advise increasing this substantially (10000+) on high performance systems. Default 100.

alpha	One minus the desired width of confidence intervals around multilocus risk estimates. Default 0.05 giving 95 percent confidence intervals.
SmallSampAdjust	Adjustment for small sample sizes, when calculating the standard error of homozygous risk genotypes. Default 0.5.

Details

Email: djmcrouch@gmail.com

One file is created by REGENT.predict, with the contents of the returned object and the input parameters/data, plus analysis log.

See the example folder included in this package for the correct input file format.

Value

Table with columns: Absolute risk, genotype relative risk, lower confidence interval, upper confidence interval, risk category, and borderline category status.

Author(s)

Graham Goddard, Daniel Crouch and Cathryn Lewis

See Also

[REGENT.model](#), [GeneticA](#), [GeneticB](#), [EnvironmentalA](#), [EnvironmentalB](#), [Inds](#)

Examples

```
#Load example data from package

library(REGENT)

data("REGENT")

write.table(GeneticA, file="GeneticA.txt")
write.table(GeneticB, file="GeneticB.txt")
write.table(EnvironmentalA, file="EnvironmentalA.txt")
write.table(EnvironmentalB, file="EnvironmentalB.txt")
write.table(Inds, file="Inds.txt")

#Create model

x=REGENT.model(AnalysisName="Example", LocusFile="GeneticB.txt", EnvFile="EnvironmentalA.txt")

#Option 1, read model from object

y=REGENT.predict(AnalysisName="Example", model=x, ind="Inds.txt")

#Option 2, read model from file
```



```
y=REGENT.predict(AnalysisName="Example",model="Example.txt",ind="Inds.txt")
```

Index

*Topic **datasets**

EnvironmentalA, [2](#)

EnvironmentalB, [3](#)

GeneticA, [3](#)

GeneticB, [4](#)

Inds, [4](#)

EnvironmentalA, [2](#), [2](#), [6](#), [8](#)

EnvironmentalB, [2](#), [3](#), [6](#), [8](#)

GeneticA, [2](#), [3](#), [6](#), [8](#)

GeneticB, [2](#), [4](#), [6](#), [8](#)

Inds, [2](#), [4](#), [6](#), [8](#)

REGENT (REGENT-package), [2](#)

REGENT-package, [2](#)

REGENT.model, [2](#), [5](#), [8](#)

REGENT.predict, [2](#), [6](#), [7](#)