Package ‘bmass’

May 17, 2019

Encoding UTF-8
Type Package
Version 1.0.3
Date 2019-05-15
Title Bayesian Multivariate Analysis of Summary Statistics
Description Multivariate tool for analyzing genome-wide association study results in the form of univariate summary statistics. The goal of 'bmass' is to comprehensively test all possible multivariate models given the phenotypes and datasets provided. Multivariate models are determined by assigning each phenotype to being either Unassociated (U), Directly associated (D) or Indirectly associated (I) with the genetic variant of interest. Test results for each model are presented in the form of Bayes factors, thereby allowing direct comparisons between models. The underlying framework implemented here is based on the modeling developed in "A Unified Framework for Association Analysis with Multiple Related Phenotypes", M. Stephens (2013) <doi:10.1371/journal.pone.0065245>.
License GPL (>= 3)
URL https://github.com/mturchin20/bmass
BugReports https://github.com/mturchin20/bmass/issues
Depends R (>= 3.3.0)
Imports utils, stats
Suggests testthat, knitr, rmarkdown
LazyData true
NeedsCompilation no
RoxygenNote 6.1.1
VignetteBuilder knitr
Author Michael Turchin [aut, cre],
Matthew Stephens [aut],
Peter Carbonetto [ctb]
Bayesian multivariate analysis of summary statistics (bmass)

Description

Run bmass on a set of phenotypes that each have univariate GWAS statistics on the same set of SNPs

Usage

```r
bmass(DataSources, GWASsnps = NULL,
      SNPMarginalUnivariateThreshold = 1e-06,
      SNPMarginalMultivariateThreshold = 1e-06, GWASThreshFlag = TRUE,
      GWASThreshValue = 5e-08, NminThreshold = 0,
      PrintMergedData = FALSE, PrintProgress = FALSE, ...)
```

Arguments

- **DataSources**
  A string indicating the variable names of the input datafiles and phenotypes. No default value.

- **GWASsnps**
  A data.table containing rows of SNPs that were univariate genome-wide significant in the phenotypes being used for analysis; GWASsnps input file should have two columns, one for chromosome and another for basepair position (with headers of Chr and BP). No default value.

- **SNPMarginalUnivariateThreshold**
  A numerical value indicating the univariate p-value threshold to use when collecting marginally significant SNPs for final bmass analysis. Default is 1e-6.
SNPMarginalMultivariateThreshold
A numerical value indicating the basic multivariate p-value threshold to use when collecting marginally significant SNPs for final bmass analysis. Default is 1e-6.

GWASThreshFlag
A logical TRUE/FALSE flag that indicates whether to threshold input GWASSnps list by a univariate GWAS p-value or not (e.g., the input GWASSnps list contains variants that are significant from discovery + replication data, but the input summary statistics are just from the discovery cohort). Default is TRUE.

GWASThreshValue
A numerical value indicating the univariate p-value threshold to use in conjunction with the GWASThreshFlag. Default is 5e-8.

NminThreshold
A numerical value that indicates a sample size threshold to use where SNPs below which are removed. Default is 0.

PrintMergedData
A logical TRUE/FALSE flag that indicates whether the intermediary ‘merged datafile’ should be included in the final bmass output; this file combines all the phenotypes for every SNP provided just prior to thresholding for marginally significant SNPs. Default is FALSE.

PrintProgress
A logical TRUE/FALSE flag that indicates whether progress statements should be printed to stderr during the course of running bmass or not. Default is FALSE.

Value
A list containing model, SNP, and posterior information for both the previously significant univariate SNPs (PreviousSNPs) and the newly significant multivariate SNPs (NewSNPs). For a full breakdown of the bmass output list structure, please see the associated vignettes.

Other Examples
bmass(c("HDL", "LDL", "TG", "TC"), GWASSnps, NminThreshold = 50000) bmass(c("HDL", "LDL", "TG", "TC"), GWASSnps, GWASThreshFlag = FALSE, SNPMarginalUnivariateThreshold = 1e-4)

bmassOutput <- bmass(c("HDL", "LDL", "TG", "TC"), GWASSnps, NminThreshold = 50000)

Examples
Phenotypes <- c("bmass_SimulatedData1", "bmass_SimulatedData2")
bmassOutput <- bmass(Phenotypes, bmass_SimulatedSigSNPs)
summary(bmassOutput)
bmassOutput$NewSNPs$SNPs
**bmass_SimulatedData1  bmass Simulated Dataset 1**

**Description**
A manually created sample dataset for use in Roxygen2 documents and vignettes.

**Format**
A data frame with 11 rows and 9 variables:

- **Chr** chromosome
- **BP** basepair position
- **Marker** rsID# or other identifier
- **MAF** Minor Allele Frequency
- **A1** reference allele
- **A2** alternative allele
- **Direction** direction of association effect size, + or -
- **pValue** p-Value of GWAS association
- **N** sample size

**Source**
Manually created

---

**bmass_SimulatedData2  bmass Simulated Dataset 2**

**Description**
A manually created sample dataset for use in Roxygen2 documents and vignettes.

**Format**
A data frame with 11 rows and 9 variables:

- **Chr** chromosome
- **BP** basepair position
- **Marker** rsID# or other identifier
- **MAF** Minor Allele Frequency
- **A1** reference allele
- **A2** alternative allele
- **Direction** direction of association effect size, + or -
- **pValue** p-Value of GWAS association
- **N** sample size
**Source**

Manually created

---

**bmass_SimulatedSigSNPs**

*bmass Simulated GWAS SNPs*

**Description**

A manually created list of GWAS significant SNPs to be used in conjunction with 'bmass_SimulatedData1' and 'bmass_SimulatedData2'.

**Format**

A data frame with 2 rows and 2 variables:

- **Chr** chromosome
- **BP** basepair position

**Source**

Manually created

---

**bmass_TestData1**

*bmass Test Dataset 1*

**Description**

A manually created sample dataset for use in unit tests.

**Format**

A data frame with 11 rows and 9 variables:

- **Chr** chromosome
- **BP** basepair position
- **Marker** rsID# or other identifier
- **MAF** Minor Allele Frequency
- **A1** reference allele
- **A2** alternative allele
- **Direction** direction of association effect size, + or -
- **pValue** p-Value of GWAS association
- **N** sample size

**Source**

Manually created
bmass_TestData2  bmass Test Dataset 2

Description
A manually created sample dataset for use in unit tests.

Format
A data frame with 11 rows and 9 variables:

- **Chr**  chromosome
- **BP**  basepair position
- **Marker**  rsID# or other identifier
- **MAF**  Minor Allele Frequency
- **A1**  reference allele
- **A2**  alternative allele
- **Direction**  direction of association effect size, + or -
- **pValue**  p-Value of GWAS association
- **N**  sample size

Source
Manually created

bmass_TestSigSNPs  bmass Test GWAS SNPs

Description
A manually created list of GWAS significant SNPs to be used in conjunction with 'bmass_TestData1' and 'bmass_TestData2'.

Format
A data frame with 2 rows and 2 variables:

- **Chr**  chromosome
- **BP**  basepair position

Source
Manually created
GetMarginalPosteriors  

Get Marginal \{U,D,I\} Posteriors

Description

Get marginal posteriors for how much every individual phenotype belongs to categories \{U,D,I\} across each SNP.

Usage

`GetMarginalPosteriors(DataSources, ListSNPs, Models, LogFile)`

Arguments

- **DataSources**: A string indicating the variable names of the input datafiles and phenotypes.
- **ListSNPs**: A list produced from running `bmass` containing the SNPs of interest to get marginal posteriors for.
- **Models**: A matrix describing the models being explored (default output from running `bmass`).
- **LogFile**: A matrix of string outputs for function logging purposes (default output from running `bmass`).

Value

A list containing three matrices of SNPs x Phenotypes marginal posteriors for each category \{U,D,I\}; this list is appended to the input ListSNPs as a new object, Marginals (the full returned object is a list containing the input ListSNPs and the input LogFile).

Examples

```r
Phenotypes <- c("bmass_SimulatedData1", "bmass_SimulatedData2")
bmassOutput <- bmass(Phenotypes, bmass_SimulatedSigSNPs)
bmassOutput[c("PreviousSNPs", "LogFile")]
GetMarginalPosteriors(Phenotypes, bmassOutput$PreviousSNPs, bmassOutput$Models, bmassOutput$LogFile)
bmassOutput$PreviousSNPs$Marginals
```
GetModelPriorMatrix

Get Model Prior Matrix

Description

Creates a matrix containing the model descriptions and their associated priors.

Usage

GetModelPriorMatrix(DataSources, Models, ModelPriors, LogFile,  
SigmaAlphas = c(0.005, 0.0075, 0.01, 0.015, 0.02, 0.03, 0.04, 0.05,  
0.06, 0.07, 0.08, 0.09, 0.1, 0.15))

Arguments

DataSources    A string indicating the variable names of the input datafiles and phenotypes.
Models         A matrix describing the models being explored (default output from running  
bmass).
ModelPriors    A vector containing the priors on each model across each tranche of sigma alpha  
(default output from running bmass; length is number of models times number  
of sigma alphas).
LogFile        A matrix of string outputs for function logging purposes (default output from  
running bmass).
SigmaAlphas    A vector containing the different values traversed for this 'effect size controlling'  
hyperparameter (see "Prior on Sigma_Alpha" in Stephens 2013 PLoS ONE,  
https://doi.org/10.1371/journal.pone.0065245).

Value

A matrix containing the original description of each model sort by prior, each model’s trained prior,  
the cumulative prior distribution, and the model’s original order position.

Examples

Phenotypes <- c("bmass_SimulatedData1", "bmass_SimulatedData2")  
bmassOutput <- bmass(Phenotypes, bmass_SimulatedSigSNPs)  
bmassOutput[c("ModelPriorMatrix", "LogFile")]<-  
  GetModelPriorMatrix(Phenotypes, bmassOutput$Models,  
  bmassOutput$ModelPriors, bmassOutput$LogFile)  
head(bmassOutput$ModelPriorMatrix)
GetTopModelsPerSNPViaPosterior

Get Top Multivariate Models

Description

Get a summary of the top models per SNP across all multivariate \{U,D,I\} combinations based on posterior probabilities.

Usage

GetTopModelsPerSNPViaPosterior(DataSources, ListSNPs, ModelPriorMatrix, LogFile)

Arguments

- **DataSources**: A string indicating the variable names of the input datafiles and phenotypes.
- **ListSNPs**: A list produced from running `bmass` containing the SNPs of interest to get marginal posteriors for.
- **ModelPriorMatrix**: A matrix detailing the models being explored and their associated priors (obtained by running `getModelPriorMatrix`).
- **LogFile**: A matrix of string outputs for function logging purposes (default output from running `bmass`).

Value

A matrix containing each model that was a SNP’s top model at least once, along with related information; this matrix is appended to the input ListSNPs as a new object, TopModels (the full returned object is a list containing the input ListSNPs and the input LogFile).

Examples

```r
Phenotypes <- c("bmass_SimulatedData1", "bmass_SimulatedData2")
bmassOutput <- bmass(Phenotypes, bmass_SimulatedSigSNPs)
bmassOutput[c("ModelPriorMatrix", "LogFile")]) <-
    GetModelPriorMatrix(Phenotypes, bmassOutput$Models,
bmassOutput$ModelPriors, bmassOutput$LogFile)
bmassOutput[c("PreviousSNPs", "LogFile")]) <-
    GetTopModelsPerSNPViaPosterior(Phenotypes,
bmassOutput$PreviousSNPs, bmassOutput$ModelPriorMatrix, bmassOutput$LogFile)
head(bmassOutput$PreviousSNPs$TopModels)
```
GlobalLipids2013.GWASsnps

Description
A list of the univariate GWAS significant SNPs from the GlobalLipids2013 dataset to be used in the second introductory bmass vignette.

Format
A data frame with 157 rows and 2 variables:

- **Chr**  chromosome
- **BP**  basepair position

Source
Supplementary Tables 2 and 3 from https://doi.org/10.1038/ng.2797.
Index

*Topic data
  bmass_SimulatedData1, 4
  bmass_SimulatedData2, 4
  bmass_SimulatedSigSNPs, 5
  bmass_TestData1, 5
  bmass_TestData2, 6
  bmass_TestSigSNPs, 6
  GlobalLipids2013.GWASnps, 10

bmass, 2, 7–9
bmass_SimulatedData1, 4
bmass_SimulatedData2, 4
bmass_SimulatedSigSNPs, 5
bmass_TestData1, 5
bmass_TestData2, 6
bmass_TestSigSNPs, 6

GetMarginalPosteriors, 7
GetModelPriorMatrix, 8, 9
GetTopModelsPerSNPViaPosteriors, 9
GlobalLipids2013.GWASnps, 10