Package 'eGST'

Title Leveraging eQTLs to Identify Individual-Level Tissue of Interest
for a Complex Trait

Version 1.0.0

Date 2019-06-30

Description Genetic predisposition for complex traits is often manifested through multiple tissues of interest at different time points in the development. As an example, the genetic predisposition for obesity could be manifested through inherited variants that control metabolism through regulation of genes expressed in the brain and/or through the control of fat storage in the adipose tissue by dysregulation of genes expressed in adipose tissue. We present a method eGST (eQTL-based genetic subtyper) that integrates tissue-specific eQTLs with GWAS data for a complex trait to probabilistically assign a tissue of interest to the phenotype of each individual in the study. eGST estimates the posterior probability that an individual's phenotype can be assigned to a tissue based on individual-level genotype data of tissue-specific eQTLs and marginal phenotype data in a genome-wide association study (GWAS) cohort. Under a Bayesian framework of mixture model, eGST employs a maximum a posteriori (MAP) expectation-maximization (EM) algorithm to estimate the tissue-specific posterior probability across individuals. Methodology is available from: A Majumdar, C Giambartolomei, N Cai, MK Freund, T Haldar, T Schwarz, J Flint, B Pasaniuc (2019) <doi:10.1101/674226>.

Depends R (>= 3.2.0)

License GPL-3

Encoding UTF-8

LazyData true

URL https://github.com/ArunabhaCodes/eGST

BugReports https://github.com/ArunabhaCodes/eGST/issues

RoxygenNote 6.1.1

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

Imports purrr, mvtnorm, MASS, utils, stats, matrixStats

NeedsCompilation no

Contents

---

eGST-package   *eGST (eQTL-based Genetic Sub-Typer): Leveraging eQTLs to identify*
                                   individual-level tissue of interest for a complex trait.

Description

Genetic predisposition for complex traits is often manifested through multiple tissues of interest
at different time points in the development. As an example, the genetic predisposition for obesity
could be manifested through inherited variants that control metabolism through regulation of genes
expressed in the brain and/or through the control of fat storage in the adipose tissue by dysregula-
tion of genes expressed in adipose tissue. We present a method eGST that integrates tissue-specific
eQTLs with GWAS data for a complex trait to probabilistically assign a tissue of interest to the phe-
notype of each individual in the study. eGST estimates the posterior probability that an individual's
phenotype can be assigned to a tissue based on individual-level genotype data of tissue-specific
eQTLs and marginal phenotype data in a GWAS cohort. Under a Bayesian framework of mixture
model, eGST employs a maximum a posteriori (MAP) expectation-maximization (EM) algorithm
to estimate the tissue-specific posterior probability across individuals.

Functions

eGST   It estimates the posterior probability that the genetic susceptibility of the phenotype of an indi-
       vidual in the study is mediated through eQTLs specific to a tissue of interest. The phenotype
       across individuals can be classified into tissues under consideration based on the estimated
       tissue-specific posterior probability across individuals.

Author(s)

Maintainer: Arunabha Majumdar <statgen.arunabha@gmail.com>

Authors:

- Tanushree Haldar <tanushree.haldar@gmail.com>
- Bogdan Pasaniuc <pasaniuc@ucla.edu>

References

Majumdar A, Giambartolomei C, Cai N, Freund MK, Haldar T, J Flint, Pasaniuc B (2019) Leveraging eQTLs to identify tissue-specific genetic subtype of complex trait. bioRxiv.

See Also

Useful links:

- <https://github.com/ArunabhaCodes/eGST>
- Report bugs at <https://github.com/ArunabhaCodes/eGST/issues>

---

eGST	*Run eGST.*

Description

Run eGST to estimate the posterior probability that the genetic susceptibility of the phenotype of an individual in the study is mediated through eQTLs specific to a tissue of interest. To create sets of tissue-specific eQTLs in your context, please see our manuscript: Majumdar A, Giambartolomei C, Cai N, Freund MK, Haldar T, J Flint, Pasaniuc B (2019) Leveraging eQTLs to identify tissue-specific genetic subtype of complex trait, bioRxiv.

Usage

```
eGST(pheno, geno, tissues, logLimprovement = 5 * 10^(-8),
  seed_choice = sample(1:1000, size = 1), nIter = 100)
```

Arguments

pheno	A numeric vector of length N where N is the number of individuals. It contains the GWAS phenotype values of individuals. No default.
geno	A list with K elements where K is the number of tissues. Each element of geno is the genotype matrix of the eQTLs specific to a tissue in the GWAS cohort. So j-th element of geno is N by Mj matrix containing the genotype data of N individuals (rows) at Mj eQTLs (columns) specific to j-th tissue. Each eQTL is a bi-allelic SNP with minor allele frequency > 0.01. Genotypes at each eQTL must be normalized across N individuals. If 0/1/2 valued genotype matrix is provided, it is internally normalized. No default.
tissues	A character vector of length K. It contains the names of tissues of interest in the analysis. The order of tissues in this vector must match the order of tissues in the previous argument 'geno'. No default.
logLimprovement	
	A positive real number specifying the minimum possible improvement of data log-likelihood in MAP-EM stopping criterion. Default $5 * 10^{(-8)}$.
seed_choice	An integer providing the choice of random seed for initialization in MAP-EM algorithm. Default is an integer randomly selected in (1,...,1000).
nIter	An integer providing the maximum number of iterations allowed in the MAP-EM algorithm. Default is 100.

Value

The output produced by eGST is a list which consists of various components.

gamma   A N by K matrix providing the tissue-specific posterior probability of N indi-
               viduals across K tissues.

alfa   Baseline tissue-specific intercepts/means of the trait.

beta   Tissue-specific eQTLs' genetic effect on the trait.

sigma_g   Square root of the variance of tissue-specific per-eQTL genetic effect on the
               trait.

sigma_e   Square root of the error variance of tissue-specific subtype of the trait which
               remains unexplained by the tissue-specific eQTLs.

m   Number of tissue-specific eQTLs.

logL   log-likelihood of the data.

References

A Majumdar, C Giambartolomei, N Cai, MK Freund, T Haldar, T Schwarz, J Flint, B Pasaniuc
(2019) Leveraging eQTLs to identify tissue-specific genetic subtype of complex trait. bioRxiv.

Examples

```
data(ExamplePhenoData)
pheno <- ExamplePhenoData
head(pheno)
data(ExampleEQTLgenoData)
geno <- ExampleEQTLgenoData
geno[[1]][1:5,1:5]
geno[[2]][1:5,1:5]
tissues <- paste("tissue", 1:2, sep = "")
result <- eGST(pheno, geno, tissues)
str(result)
```

ExampleEQTLgenoData   *An example of tissue-specific eQTLs genotype data.*

Description

ExampleEQTLgenoData is a list with two elements each containing a 1000 by 100 ordered genotype
matrix. Each matrix provides the genotype data of the 1000 individuals at 100 tissue-specific eQTLs
for each tissue.

Usage

```
data(ExampleEQTLgenoData)
```

Format

A list of two numeric matrix each having 1000 rows (individuals) and 100 columns (eQTLs):

Examples

```
data(ExampleEQTLgenoData)
geno <- ExampleEQTLgenoData
```

ExamplePhenoData   *An example of phenotype data.*

Description

ExamplePhenoData is a simulated vector of phenotype for 1000 individuals. In this simulated example dataset, we have considered two tissues and corresponding sets of 100 tissue-specific eQTLs each. First half of 1000 individuals' phenotypes were simulated to have genetic effect from the first tissue specific eQTLs, but no effect from the second tissue-specific eQTLs. Hence first 500 individuals were assigned to the first tissue. Similarly, second half of the 1000 individuals were simulated to have genetic effect from the second-tissue specific eQTLs and hence they were assigned to the second tissue.

Usage

```
data(ExamplePhenoData)
```

Format

A numeric vector of length 1000 containing the phenotype data of 1000 individuals.

Examples

```
data(ExamplePhenoData)
pheno <- ExamplePhenoData
```

Index