|  |  |
| --- | --- |
| runMultiTraitGwas {genStatPipeline} | R Documentation |

**Perform multi-trait GWAS**

**Description**

runMultiTraitGwas performs multi-trait or multi-environments Genome Wide Association mapping on phenotypic and genotypic data contained in a gData object.

**Usage**

runMultiTraitGwas(gData, environments = NULL, covar = NULL,

snpCov = NULL, kin = NULL, kinshipMethod = c("astle", "GRM", "IBS",

"vanRaden"), GLSMethod = c("single", "multi"), subsetMarkers = FALSE,

markerSubset = "", MAF = 0.01, fitVarComp = TRUE,

covModel = c("unst", "pw", "fa"), VeDiag = TRUE, tolerance = 1e-06,

maxIter = 2e+05, maxDiag = 10000, mG = 1, mE = 1, CmHet = TRUE,

DmHet = TRUE, stopIfDecreasing = TRUE, computeLogLik = TRUE,

Vg = NULL, Ve = NULL, reduceK = FALSE, nPca = NULL,

estCom = FALSE, parallel = FALSE, nCores = NULL)

**Arguments**

|  |  |
| --- | --- |
| gData | An object of class gData containing at least map, markers and pheno. |
| environments | A vector of environments on which to run multi-trait GWAS. These can be either numeric indices or character names of list items in pheno. If NULL GWAS is run for all environments. For multi-environment GWAS, all environments should be columns in a single list item (i.e. a single ‘environment’). ==> confusing.... change ? see details ? |
| covar | An optional vector of covariates taken into account when running GWAS. These can be either numeric indices or character names of columns in covar in gData. If NULL no covariates are used. |
| snpCov | An optional character vector of snps to be included as covariates. |
| kin | An optional kinship matrix or list of kinship matrices. These matrices can be from the matrix class as defined in the base package or from the dsyMatrix class, the class of symmetric matrices in the Matrix package. If GLSMethod = "single" then one matrix should be provided, if GLSMethod = "multi" a list of chromosome specific matrices of lenght equal to the number of chromosomes in map in gData. If NULL then matrix kinship in gData is used.  If both kin is provided and gData contains a matrix kinship then kin is used. |
| kinshipMethod | An optional character indicating the method used for calculating the kinship matrix(ces). Currently "astle" (Astle and Balding, 2009), "GRM", "IBS" and "vanRaden" (VanRaden, 2008) are supported. If a kinship matrix is supplied either in gData or in parameter kin kinshipMethod is ignored. |
| GLSMethod | A character string indicating the method used to estimate the marker effects. Either single for using a single kinship matrix. or multi for using chromosome specific kinship matrices. |
| subsetMarkers | Should the marker data be subsetted? |
| markerSubset | A numeric or character vector used for subsetting the markers. Ignored if subsetMarkers = FALSE. |
| MAF | A numerical value between 0 and 1. SNPs with minor allele frequency below this value are not taken into account for the analysis, i.e. p-values and effect sizes are put to missing (NA). Ignored if useMAF is FALSE. |
| fitVarComp | Should the variance components be fitted? If FALSE they should be supplied in Vg and Ve |
| covModel | A character string indictating the model used when fitting the variance components. Either unst for unstructured for both Vg and Ve (as in Zhou and Stephens (2014)), pw for unstructered for both Vg and Ve (pairwise, as in Furlotte and Eskin (2013)) or fa for factor-analytic for both Vg and Ve. Ignored if fitVarComp = FALSE |
| VeDiag | Should there be environmental correlations if covModel = "unst" or "pw"? If traits are measured on the same individuals putTRUE. |
| tolerance | A numerical value. The iterating process stops if the difference in conditional log-likelihood between two consecutive iterations drops below tolerance. Only used when covModel = "fa". |
| maxIter | An integer for the maximum number of iterations. Only used when covModel = "fa". |
| maxDiag | A numical value. The maximal value of the diagonal elements in the precision matrices Cm and Dm (ignoring the low-rank part W W^t). Only used when covModel = "fa". |
| mG | An integer. The order of the genetic part of the factor analytic model. Only used when covModel = "fa". |
| mE | An integer. The order of the environmental part of the factor analytic model. Only used when covModel = "fa". |
| CmHet | Should an extra diagonal part be added in the model for the precision matrix Cm? Only used when covModel = "fa". |
| DmHet | Should an extra diagonal part be added in the model for the precision matrix Dm? Only used when covModel = "fa". |
| stopIfDecreasing | Should the iterating process in the factor analytic model stop if after 50 iterations the log-likelihood decreases between two consecutive iterations? Only used when covModel = "fa". |
| Vg | An optional matrix with genotypic variance components. Vg should have row names column names corresponding to the column names of Y. It may contain additional rows and colums which will be ignored. Ignored if fitVarComp = TRUE. |
| Ve | An optional matrix with environmental variance components. Ve should have row names column names corresponding to the column names of Y. It may contain additional rows and colums which will be ignored. Ignored if fitVarComp = TRUE. |
| reduceK | Should the kinship matrix be reduced? See [reduceKinship](file:///M:\willem\research\STATISTICAL_GENETICS\statgenpipeline\statGenPipeline\documentation_update\reduceKinship.html) |
| nPca | An integer giving the number of Pcas used whe reducing the kinship matrix. Ignored if reduceK = FALSE. |
| estCom | Should the common SNP-effect model be fitted? If TRUE not only the SNP-effects but also the common SNP-effect and QTL x E effect are estimated. |
| parallel | Should the computation of variance components be done in parallel? Only used if covModel = "pw". A parallel computing environment has to be setup by the user. |
| nCores | A numerical value indicating the number of cores to be used by the parallel part of the algorithm. If NULL the number of cores used will be equal to the number of cores available on the machine - 1. |

**Value**

An object of class [GWAS](file:///M:\willem\research\STATISTICAL_GENETICS\statgenpipeline\statGenPipeline\documentation_update\GWAS.html).

**Details**

**References**

Dahl et al. (2013). Network inference in matrix-variate Gaussian models with non-independent noise. arXiv preprint arXiv:1312.1622.

Furlotte, N.A. and Eskin, E. (2015). Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. Genetics, May 2015, Vol.200-1, p. 59-68.

Kruijer et al. (2015) Marker-based estimation of heritability in immortal populations. Genetics. February 2015, Vol. 199-2, p. 379-398.

Zhou, X. and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods, February 2014, Vol. 11, p. 407–409.

Millet et al. (2016)

Thoen et al. (2017)

Korte et al. (2012)

[Package *genStatPipeline* version 0.0.0.9000 [Index](file:///M:\willem\research\STATISTICAL_GENETICS\statgenpipeline\statGenPipeline\documentation_update\00Index.html)]

Issues:

* MAF: based on all genotypic data, or only those genotypes with phenotypic data ?
* environments : multi-trait vs multi-env ...
* Too many options for factor analytic. Just fix to the default: Cmhet, DmHet, StopIf, ComputeLog...
* Tolerance + max.diag should become a function of the data