Package 'MetaNetwork'

January 10, 2007

Version 1.0-0

Date 2007-01-10

Title A computational protocol for the genetic study of metabolic networks

Author Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Joost Keurentjes <Joost.Keurentjes@wur.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Maintainer Groningen Bioinformatics Center GBIC Jingyuan Fu <j.fu@rug.nl>

Depends qvalue

Description MetaNetwork maps metabolite quantitative trait loci (mQTLs) underlying variation in metabolite abundance in individuals of a segregating population with a two-part model to account for the often observed spike in the distribution of metabolite data. MetaNetwork predicts and visualizes potential associations between metabolites using correlations of mQTL profiles, rather than of abundance profiles. Simulation and permutation procedures are used to assess statistical significance.

URL http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

R topics documented:

`Overview` *MetaNetwork package*

Description

MetaNetwork maps metabolite quantitative trait loci (mQTLs) underlying variation in metabolite abundance in individuals of a segregating population with a two-part model to account for the often observed spike in the distribution of metabolite data. MetaNetwork predicts and visualizes potential associations between metabolites using correlations of mQTL profiles, rather than of abundance profiles. Simulation and permutation procedures are used to assess statistical significance.

Note

To install MetaNetwork package:
Download the MetaNetwork package from: `http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork`
(Windows/RGui) click 'Packages'; 'install package(s) from local zip files'; browse to the MetaNetwork zip file.

MetaNetwork requires the `qvalue` package for false discovery rate control:
(Windows/RGui) click 'Packages'; 'Install package(s)...'; select CRAN mirror; select qvalue.

The developers of MetaNetwork thank Gonzalo Vera, Bruno Tesson and Richard Scheltema for helping to develop this R-package, and Martijn Dijkstra and Rainer Breitling for helping to predict multiple peaks belonging to the same metabolite

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Joost Keurentjes <Joost.Keurentjes@wur.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Please cite publication below when reporting results produced using this package:
Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

`MetaNetwork` *genetic study of metabolites and network reconstruction*

Description

An integrated computational protocol to run a complete genetic analysis workflow on metabolites using diverse MetaNetwork methods for quantitative trait analysis, network reconstruction and Cytoscape network visualization.

Usage

```
MetaNetwork( markers, genotypes, traits, spike, qtlProfiles = NULL,
             qtlThres = NULL, qtlSumm = NULL, corrZeroOrder = NULL,
             corrSecondOrder = NULL, corrMethod = "qtl", corrThres = 0,
             cytoFiles = T, peaks = NULL, outputdir = "./MetaNetwork")
```

Arguments

`markers`	matrix of markers (rownames) and their chromosome numbers (column 1) and centi-Morgan positions (cM, column 2), ordered by position. See `markers` example data.
`genotypes`	matrix of genotypes for each marker (rownames) and individual (columnnames), as numeric values 1, 2 or NA when missing. See `genotypes` example data.
`traits`	matrix of phenotypes for each trait (rownames) and individual (columnnames), as numeric or NA when missing. See `traits` or `traits2` example data.
`spike`	numeric cut-off value to separate absent (qualitative) from available (quantitative) trait abundance.
`qtlProfiles`	(optional) matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values. If `qtlProfiles` is NULL (default), MetaNetwork will call function `qtlMapTwoPart` to generate the `qtlProfiles`. Otherwise, MetaNetwork will use the provided QTL results. See `qtlProfiles` example data.
`qtlThres`	(optional) numeric $-log_{10}(p)$ threshold value for significant QTLs. If `qtlThres` is NULL (default), the QTL significance threshold will be estimated by simulation using `qtlThreshold` at `alpha = 0.05` and `n.simulations = 1000`. Furthermore, the QTL significance threshold is also estimated by controlling the false discovery rate `fdrThres = 0.05` using `qtlFDR`. By default, the most stringent outcome of either `qtlThreshold` or `qtlFDR` is used. Otherwise, MetaNetwork will use provided threshold.
`qtlSumm`	(optional) data.frame with the summary of each QTL. If `qtlSumm` is NULL (default), MetaNetwork will call function `qtlSummary` to summarize QTL effects. See `qtlSumm` example data.
`corrZeroOrder`	(optional) the matrix of zero-order correlation coefficient between metabolites. If `corrZeroOrder` is NULL (default), zero-order correlation coefficient will

be calculated for QTL profiles using function qtlCorrZeroOrder.
See corrZeroOrder example data.

corrSecondOrder
: (optional) matrix of second-order partial correlation between metabolites.
If corrSecondOrder is NULL (default), second-order partial correlation will be calculated for QTL profiles using function qtlCorrSecondOrder.
See corrSecondOrder example data.

corrMethod
: (optional) character string indicating which correlation method, either "qtl" or "abundance".
If corrMethod is "qtl" (default), MetaNetwork will call function qtlCorrZeroOrder to calculate the correlation between QTL profiles. Otherwise, when corrMethod is "abundance", MetaNetwork will use Spearman corrrelation via function cor to calculate the correlation between metabolite abundance profiles.

corrThres
: (optional) numeric threshold for significant partial correlation coefficients.
If corrThres is NULL, the empirical threshold is estimated by permutation using function qtlCorrThreshold with n.permutations = 10000. Otherwise, the provided threshold is used. Default is 0.

peaks
: (optional) matrix of mass/charge peaks (column1) for each trait (rownames).
If peaks is set, MetaNetwork will call findMultiplePeaks to relate multiple mass peaks for correlated traits.
See peaks2 example peaks data for unidentified metabolite example traits data traits2.

cytoFiles
: (optional) boolean value that indicates if files for network visualization in Cytoscape should be created.
If TRUE (default) MetaNetwork will call function createCytoFiles to create two network files in outputdir for the significant correlations amongst metabolites: 'network.sif' and 'network.eda'.

outputdir
: (optional) output directory where generated data files will be stored. Default is "./MetaNetwork"

Details

First, MetaNetwork maps metabolite quantitative trait loci (mQTLs) underlying variation in metabolite abundance in individuals of a segregating population using a two-part model to account for the nature of metabolite data (step A, qtlMapTwoPart). This model combines the analysis of the binary traits (positive/not-available) with conditional analysis of the quantitative trait (numeric) among individuals with a positive binary phenotype. Simulation procedures are used to assess statistical significance (step B, qtlThreshold, qtlFDR). MetaNetwork will summarize the information about significant mQTLs (step C, qtlSummary).

Then, MetaNetwork predicts the network of potential associations between metabolites using correlations of mQTL profiles or abundance profiles (step D, qtlCorrZeroOrder; step E, qtlCorrSecondOrder). Optionally, permutation procedures can be used to assess statistical significance (step F, qtlCorrThreshold).

Finally, MetaNetwork generates files of predicted networks, which can be visualized using Cytoscape (step G, createCytoFiles), and optionally relates multiple mass peaks per metabolite that may be consequence of isotopes or charge difference (step H, findMultiplePeaks).

Analysis of about 24 metabolites takes a few minutes on a desktop computer (Pentium 4). Analysis of a metabolome of about 2000 metabolites will take around four days. In addition, MetaNetwork is able to integrate high-throughput data from future metabolomics, transcriptomics and proteomics experiments in conjunction with phenotypic data.

After running MetaNetwork with defaults, the R console will show:

```
>MetaNetwork (markers=markers, genotypes=genotypes, traits=traits, spike=4
              qtlThres=3.79)
Step A: QTL mapping....
         result in R object 'qtlProfiles'
         result in ./MetaNetwork/qtlProfiles.csv
         process time 29.25 sec

Step B: Simulation test for QTL significance threshold....skipped
         using user-provided QTL threshold: 3.79

Step C: QTL summary....
         result in R object: 'qtlSumm'
         result in ./MetaNetwork/qtlSumm.csv
         process time 1.66 sec

Step D: Zero-order correlation ....
         result in R object: 'corrZeroOrder'
         result in ./MetaNetwork/corrZeroOrder.csv
         process time 2.97 sec

Step E: 2nd-order correlation ....
         result in R object: 'corrSecondOrder'
         result in ./MetaNetwork/corrSecondOrder.csv
         process time 9.58 sec

Step F: Permutation test for 2nd-order correlation significance threshold...skip
         using user-provided correlation threshold: 0

Step G: Create Cytoscape network files...
         SIF file is: ./MetaNetwork/network.sif
         EDA file is: ./MetaNetwork/network.eda

Step H: Find Multiple Peaks....skipped
```

Value

`qtlProfiles` matrix of QTL mapping of traits (rownames) to markers (columnnames) as log-transformed "p values" [$-log_{10}(p)$], see `qtlMapTwoPart`. A +/- sign is added to indicate the direction of the additive effect: values are positive if the QTL has higher metabolite abundance for individuals carrying the genotype 2 than those carrying the genotype 1; values are negative otherwise.
See `qtlProfiles` example data.

`qtlThres` estimated QTL significance threshold.
See function `qtlThreshold`.

`qtlSumm` data frame with QTL summary.
See `qtlSumm` example data.

`corrZeroOrder`
matrix of zero order correlation of QTL profiles.
See `corrZeroOrder` example data.

`corrSecondOrder`
matrix of 2nd order correlation of QTL profiles.
See `corrSecondOrder` example data.

corrPermutations
: vector of the permutations of maximum, absolute correlation values.
See function qtlCorrThreshold.

corrThres
: numeric correlation threshold.
See function qtlCorrThreshold.

cytoFiles
: network files "network.sif" and "network.eda" for cytoscape are produced in outputdir.
See function createCytoFiles.

multiplePeaks
: If peaks is not NULL, data frame with Multiple Peak summary.
See multiplePeaks example data.

resultFiles
: If outputdir is not NULL, the above outputs will be also saved in files "qtlProfiles.csv", "qtlSumm.csv", "corrZeroOrder.csv","corrSecondOrder.csv", "corrPermutations.csv", "multiplePeaks.csv", respectively. A summary of analysis processing, results objects and output files can be seen in the R console and is saved in file "output.txt".

Note

The names of individuals (columnnames) must be consistent over genotypes and traits. The names of peaks (rownames) must be consistent over peaks and traits.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use markers, genotypes and traits as example data sets or use loadData to load your own data.
Use qtlMapTwoPart for the calculation of qtlProfiles.
Use qtlThreshold and qtlFDR for the estimation of qtlThres QTL significance threshold.
Use qtlCorrZeroOrder and qtlCorrSecondOrder for the calculation of zero order and second order correlation for corrZeroOrder and corrSecondOrder respectively.
Use qtlCorrThreshold for the estimation of corrThres correlation significance threshold.
Use qtlSummary for the generation of qtlSumm QTL summary.
Use createCytoFiles for the generation of Cytoscape network files.
Use findMultiplePeaks for the relation of isotopic or differentially charged metabolites.

Examples

```
## load the example data provided with this package
data(genotypes)
data(traits)
data(markers)

#set qtlThres
```

```
qtlThres    <- 3.79

#run metanetwork with predefined thresholds
MetaNetwork (markers=markers, genotypes=genotypes, traits=traits, spike=4,
             qtlThres=qtlThres)

##OR: load data from csv
#genotypes <- loadData("genotypes.csv")
#traits    <- loadData("traits.csv")
#markers   <- loadData("markers.csv")
#MetaNetwork (markers=markers, genotypes=genotypes, traits=traits2,
#             qtlThres=qtlThres, spike=4)

##OR: let MetaNetwork estimate qtlThres and identify multiple peaks
#data(genotypes)
#data(traits2)
#data(markers)
#data(peaks2)
#MetaNetwork (markers=markers, genotypes=genotypes, traits=traits2,
#             peaks=peaks2, spike=4)

##show part of the qtlProfiles
qtlProfiles[1:5,1:5]

##show part of the qtl summary
qtlSumm[1:5,]

##show part of the zero order correlation
corrZeroOrder[1:5,1:5]

##show part of the second order correlation
corrSecondOrder[1:5,1:5]

##plot the qtlProfiles
qtlPlot(markers, qtlProfiles, qtlThres)

##load network.sif and network.eda into Cytoscape
```

`markers` *Example marker data*

Description

Example matrix of `markers` (rownames) for Arabidopsis thaliana and their chromosome numbers (column 1) and centi-Morgan (cM, column 2) positions, ordered by position:

```
> data(markers)
> markers[1:5,]
```

	chr	cM
PVV4	1	0.0
AXR-1	1	6.4
HH.335C-Col	1	10.8
DF.162L/164C-Col	1	12.9
EC.480C	1	15.059

Usage

```
data(markers)
```

Format

117 rows of markers and two columns for chromosome number and centi-Morgan position

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Joost Keurentjes <Joost.Keurentjes@wur.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use MetaNetwork for automated analysis of this data set as part a genetic analysis protocol on metabolites.

---

genotypes *Example genotype data*

Description

Example matrix of genotypes for each marker (rownames) and 162 Arabidopsis thaliana individuals (columnnames), as numeric values 1, 2 or NA when missing:

```
> data(genotypes)
> genotypes[1:5,1:5]
```

	X1	X3	X4	X5	X6
PVV4	1	1	2	1	2
AXR-1	1	1	2	1	2
HH.335C-Col	1	1	1	1	2
DF.162L/164C-Col	1	1	1	1	2
EC.480C	1	1	1	1	2

Usage

```
data(genotypes)
```

Format

117 markers by 162 samples (RILS).

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Joost Keurentjes <Joost.Keurentjes@wur.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use MetaNetwork for automated analysis of this data set as part a genetic analysis protocol on metabolites. Use qtlMapTwoPart for the calculation of qtlProfiles form genotypes.

---

traits *Example traits data*

Description

The traits dataset contains measured abundance of 18 glucosinates and 6 flavonols for 162 Arabidopsis thaliana individuals. The traits2 dataset contains abundance of 20 unidentified mass peaks (mass peak numbers) for 162 Arabidopsis thaliana individuals. Example matrice of phenotypes for each trait (rownames) and sample/individual (columnnames), as numeric or NA when missing:

```
> data(traits)
> traits[1:5,1:5]
```

	X1	X3	X4	X5	X6
3-Hydroxypropyl	NA	942	2402	602	213
4-Hydroxybutyl	NA	4	10	183	198
4-Methylsulfinylbutyl	NA	55	62	13386	1671
3-Butenyl	NA	84	32	18	4339
3-Methylthiopropyl	NA	3108	569	4	7

Usage

```
data(traits)
data(traits2)
```

Format

the number of traits (rows) by the number of samples (columns, RILS).

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Joost Keurentjes <Joost.Keurentjes@wur.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `MetaNetwork` for automated analysis of this data set as part a genetic analysis protocol on metabolites. Use `qtlMapTwoPart` for the calculation of `qtlProfiles` form `traits`.

`peaks` *Example peak data*

Description

Example matrix of mass/charge peaks (column1) for each trait (rownames) that accompanies the untargeted `traits2` metabolite abundance dataset from Arabidopsis thaliana:

```
> data(peaks2)
> peaks2[1:5,]
```

	MassOverCharge
LCavg.1539	378
LCavg.1549	379
LCavg.1555	753
LCavg.1570	424
LCavg.1610	436

Usage

```
data(peaks2)
```

Format

20 rows of peaks with mass over charge values (column 1) for example data set `traits2`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `MetaNetwork` for automated analysis of this data set as part a genetic analysis protocol on metabolites.
Use `findMultiplePeaks` to relate multiple peaks per metabolite.

`qtlProfiles` *Example qtl profiles*

Description

Example matrix of QTL mapping profiles of traits (rownames) on to markers (columnnames), calculated from Arabidopsis thaliana `genotypes` and `traits` example datasets, as $-log_{10}(p)$ values. The +/- sign indicates the direction of the additive effect: values are positive if the QTL has higher metabolite abundance for individuals carrying the genotype 2 than those carrying the genotype 1; values are negative otherwise.

```
>data(qtlProfiles)
>qtlProfiles[1:5,1:5]
```

	PVV4	AXR-1	HH.335C-Col	DF.162L/164C-Col	EC.480C
3-Hydroxypropyl	1.1003510	1.52622550	0.7823755	0.69234969	1.0440796
4-Hydroxybutyl	0.3007378	1.86596872	2.3859329	2.24693677	1.7367161
4-Methylsulfinylbutyl	-0.1738533	0.08042058	0.1171418	0.02829792	-0.1050181
3-Butenyl	-0.1875411	-1.16887454	-0.8904340	-0.80973239	-0.9517352
3-Methylthiopropyl	0.5516505	0.33930888	0.3652804	0.36563859	0.2445772

This matrix can be produced by functions `qtlMapTwoPart` or `MetaNetwork`.

Usage

```
data(qtlProfiles)
```

Format

A matrix of QTL $-log_{10}(p)$ values for 24 traits (rows) and 117 markers(columns).

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `qtlMapTwoPart` to generate `qtlProfiles`.
Use `MetaNetwork` for automated analysis of this data set as part a genetic analysis protocol on metabolites.

`qtlSumm` *Example QTL summary report*

Description

Example of the QTL summary for each trait as data frame that can be produced by functions `qtlSummary` or `MetaNetwork`.

```
> data(qtlSumm)
> qtlSumm[1:5,]
```

	traitName	QTLchr	QTLmk	QTLleftcm	QTLpeakcm	QTLrightcm	logP
1	3-Hydroxypropyl	4	GA1	3.6640	9.027	11.1280	11.1
2	3-Hydroxypropyl	5	GH.117C	32.4675	35.356	37.6390	13.5
3	4-Hydroxybutyl	3	DF.77C	0.0000	0.000	9.4365	3.9
4	4-Hydroxybutyl	4	GA1	3.6640	9.027	16.5810	4.5
5	4-Methylsulfinylbutyl	5	GH.117C	32.356	35.356	37/6390	16.3

(cont'd) VarP1 VarP2 additive

(1 cont'd)	0.0	26.0	-2784.0
(2 cont'd)	0.0	31.2	-3088.8
(3 cont'd)	10.4	1.3	-16.9
(4 cont'd)	5.5	6.6	28.0
(5 cont'd)	12.0	34.2	61.3

Usage

```
data(qtlSumm)
```

Format

A data frame.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `qtlSummary` to generate `qtlSumm`.
Use `MetaNetwork` to produce this dataset as part a genetic analysis protocol on metabolites.

`corrZeroOrder and corrSecondOrder`

Example correlation data

Description

Example of a QTL correlation matrix of `traits` that can be produced using functions `qtlCorrZeroOrder` or `qtlCorrSecondOrder`.

```
> data(corrZeroOrder)
> corrZeroOrder[1:3,1:3]
```

	3-Hydroxypropyl	4-Hydroxybutyl	4-Methylsulfinylbutyl
3-Hydroxypropyl	1.0000000	0.8205883	0.8471439
4-Hydroxybutyl	0.8205883	1.0000000	0.9182831
4-Methylsulfinylbutyl	0.8471439	0.9182831	1.0000000

Usage

```
data(corrZeroOrder)
data(corrSecondOrder)
```

Format

24x24 matrix of correlation coefficients between `traits`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `qtlCorrZeroOrder`, `qtlCorrSecondOrder` or `cor` to reproduce this data set.
Use `MetaNetwork` to produce this dataset as part a genetic analysis protocol on metabolites.

`multiplePeaks` *Example Multiple Peaks report*

Description

Example of the Multiple Peaks report that relates metabolites represented by multiple mass peaks (because of isotopic or differential charges) for `peaks2` example data set using information from `traits2`. This report can be produced by functions `findMultiplePeaks` or `MetaNetwork`.

```
> data(multiplePeaks)
> multiplePeaks[1:5,]
```

cluster	peak1	mz1	peak2	mz2	corrCoef	massDiff	massRatio	relationship
1	LCavg.1539	378	LCavg.1538	377	0.9998261	1	1.0	isotope
2	LCavg.1570	424	LCavg.1566	423	0.9989359	1	1.0	isotope
3	LCavg.1610	436	LCavg.1596	873	0.9912291	-437	0.5	diffCharged
4	LCavg.1611	437	LCavg.1610	436	0.9612162	1	1.0	isotope
5	LCavg.1612	438	LCavg.1596	873	0.9513867	-435	0.5	diffCharged

Usage

```
data(multiplePeaks)
```

Format

A data frame.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `findMultiplePeaks` to generate `multiplePeaks`.
Use `MetaNetwork` to produce this dataset as part a genetic analysis protocol on metabolites.

`loadData` *Load data from csv files*

Description

Load a comma separated values (csv) file and convert the contents of the file into a matrix object, taking the 1st column as the rownames and the 1st row as columnnames. See also `read.csv`.

Usage

```
mydata <- loadData(filename)
```

Arguments

`filename`	path of the file where the data will be loaded from.

Value

Returns a matrix containing the data from the indicated file.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use genotypes, traits and markers as example data sets.
Use MetaNetwork for automated analysis of loaded data sets as part of a genetic analysis protocol on metabolites.

Examples

```
#not run
##load genotypes
genotypes <- loadData("genotypes.csv")

##load traits
traits    <- loadData("traits.csv")

##load markers
markers   <- loadData("markers.csv")

## end not run
```

---

qtlMapTwoPart *QTL test statistic using two-part model*

Description

QTL model for a mixed phenotype with qualitative and quantitative variance.

Part one tests whether the presence/absence of metabolites has a genetic basis, i.e. whether different genotype classes at a given marker differ in their number of non-zero observations.

Part two tests whether quantitative variation in non-zero abundance has a genetic basis: whether the non-zero observations for each of these genotype classes at a given marker differ in mean abundance.

The 'p-value' of the QTL is computed as the product of the two 'p-values' in the two parts.

Usage

```
qtlMapTwoPart(genotypes, traits, spike, filename=NULL)
```

Arguments

genotypes	matrix of genotypes for each marker (rownames) and individual (columnnames), as numeric values 1, 2 or NA when missing. See genotypes example data.
traits	matrix of phenotypes for each trait (rownames) and individual (columnnames), as numeric or NA when missing. See traits or traits2 example data.

`spike`	numeric cut-off value to separate absent (qualitative) from available (quantitative) trait abundance.
`filename`	(optional) path of the file where the QTLs are to be stored. Default is NULL.

Details

The analysis model for `traits` data with both qualitative and quantitative variance. If the distribution of `traits` is mixed, a two part model is used.

The first part describes a binominal model which tests for association of markers with observed noise value. For each trait: Let y_i denote the mass intensity for i^{th} RIL and let $z_i = 0$ if $y_i \leq$ `spike` and $z_i = 1$ if $y_i >$`spike`. We then test each marker for significant differences between the two genotypes for the probability of presence of the trait (mass) peak:
$H_0 : p\{z = 1|g = 1\} = p\{z = 1|g = 2\}$
versus the alternative hypothesis
$H_1 : p\{z = 1|g = 1\} \neq p\{z = 1|g = 2\}$,
where g is the genotype (1 or 2) of a marker under analysis.

The second part describes a parametric model which tests for association of markers with quantitative variances for those individuals where $y_i >$`spike`. Under the assumption of normal distribution, we test each marker for significant differences in the mean values between two genotypes:
$H_0 : u\{g = 1\} = u\{g = 2\}$
versus the alternative hypothesis
$H_1 : u\{g = 1 \neq u\{g = 2\}$.

When there is binary data only (no quantitative data) or quantitative data only (no spike), the 'p-value' of the missing part is set to one. The p value of the two-part model is then determined by the multiple of the p values from the two parts.

Value

Returns a matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values. The $-log_{10}(p)$ values are assigned negative if traits for individuals carrying genotype "1" are larger on average than traits for individuals carrying genotype "2". Otherwise, the $-log_{10}(p)$ value is positive.

Note

The individual columns of `genotypes` and `traits` must have the same order. The names of individuals must be consistent over `genotypes` and `traits`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use genotypes and traits as example data sets or use loadData to load your own data.
Use MetaNetwork for automated application of this function as part a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(genotypes)
data(traits)

##OR: load your own data
#genotypes  <- loadData("genotypes.csv")
#traits     <- loadData("traits.csv")

##calculate the two part qtl using a cutoff spike of 4
qtlProfiles <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

#show part of the results
qtlProfiles[1:5,1:5]
```

---

qtlThreshold *Estimate QTL significance threshold.*

Description

Simulation test to estimate empirical threshold for QTL significance. The trait values are simulated using the median number of noise, median mean and standard deviation from the observed trait data under the null hypothesis of no QTL. At each simulation test, the highest absolute $-log_{10}(p)$ value is recorded. The threshold is set at desired alpha level (i.e. take the $-log_{10}(p)$ value at the alpha proportion position of the permutation set).

Usage

```
qtlThreshold(genotypes, traits, spike,
             n.simulations = 1000, alpha = 0.05)
```

Arguments

genotypes	matrix of genotypes for each marker (rownames) and individual (columnnames), as numeric values 1, 2 or NA when missing. See genotypes example data.
traits	matrix of phenotypes for each trait (rownames) and individual (columnnames), as numeric or NA when missing. See traits example data.
spike	numeric cut-off value to separate absent (qualitative) from available (quantitative) trait abundance.
n.simulations	(optional) number of simulations. Default is 1000 times.
alpha	(optional) numeric alpha level for the threshold. Default is 0.05.

Value

Returns the $-log_{10}(p)$ significance threshold value for QTLs.

Note

The individual columns of `genotypes` and `traits` must have the same order. The names of individuals should be consistent over `genotypes` and `traits`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

Examples

```
## load the example data provided with this package
data(genotypes)
data(traits)

##OR: load your own data
#genotypes  <- loadData("genotypes.csv")
#traits     <- loadData("traits.csv")

##estimate qtl threshold for significance with low count of simulations (advised: 1000)
qtlThres    <- qtlThreshold(genotypes, traits, 4, n.simulations = 10)

##show the threshold
qtlThres
```

---

qtlFDR *Estimate QTL thresholds for false discovery rate (FDR)*

Description

Calculate thresholds to control the false discovery rate in QTL analysis. At given fdrThres(desired proportion of false positives), estimate the corresponding qtlThreshold (qtlThres). And/or at given qtlThres (desired treshold for significance), estimate the corresponding proportion of false positives (fdrThres).

Usage

```
qtlFDR <- function( qtlProfiles, fdrThres=0.05, qtlThres=NULL )
```

Arguments

qtlProfiles	matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values. See qtlProfiles example data.
fdrThres	the desired qvalue proportion of false positives incured (called the false discovery rate). Default is 0.05. See qvalue package.
qtlThres	numeric $-log_{10}(p)$ desired threshold value for significant QTLs. See qtlThreshold function.

Value

A matrix with three columns and two rows:

c:	qValue	proportion of false positives (fdr).
c:	pValue	$10^{-1*qtlThres}$ transformation of qtlThres/qtlProfiles to match qvalue package.
c:	-log10P	$-log_{10}(p)$ estimation of qtlThres.
r:	fdrThres	a pValue and $-log_{10}(p)$ qtlThres for the given fdrThres.
r:	qtlThres	a qValue (fdr) and pValue for the given qtlThres.

Example:

	qValue	pValue	-log10P
[1,]	0.050000000	0.0809722708	1.091664
[2,]	0.001344755	0.0006845554	3.164591

If fdrThres or qtlThres is NULL then the respective rows are ommitted.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl. Acd. Sci. USA 100, 9440-9445 (2003).

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use qtlThreshold to estimate an QTL threshold based on simulation.
Use qtlMapTwoPart to calculate qtlProfiles.
Use MetaNetwork for automated application of this function as part a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers        <- loadData("markers.csv")
#genotypes      <- loadData("genotypes.csv")
#traits         <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles     <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres        <- 3.79

##OR: estimate the threshold yourself
#qtlThres       <- qtlThreshold(genotypes, traits, spike=4)

##estimate FDR
qtlFDR          <- qtlFDR(qtlProfiles, fdrThres=0.05, qtlThres=qtlThres)

##show FDR for both fdrThres of 0.05 and qtlThres of qtlThres
qtlFDR
```

---

qtlSummary *Summarize QTL effects*

Description

Summarize QTL profiles including peak position (chromosome, centi-Morgan), support interval (calculated via the function `qtlSupportInterval`), proportion of QTL variation explained by part 1 and 2 of the `qtlMapTwoPart` model, and allele substitution effect.

Usage

```
qtlSummary(markers, genotypes, traits, qtlProfiles, spike, qtlThres,
           interval.dropoff = 1.5, filename = NULL)
```

Arguments

`markers`	matrix of markers (rownames) and their chromosome numbers (column 1) and centi-Morgan positions (cM, column 2), ordered by position. See `markers` example data.
`genotypes`	matrix of genotypes for each marker (rownames) and individual (columnnames), as numeric values 1, 2 or NA when missing. See `genotypes` example data.
`traits`	matrix of phenotypes for each trait (rownames) and individual (columnnames), as numeric or NA when missing. See `traits` example data.

`qtlProfiles` matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values.
See `qtlProfiles` example data.

`spike` numeric cut-off value to separate absent (qualitative) from available (quantitative) trait abundance.

`qtlThres` numeric $-log_{10}(p)$ threshold value for significant QTLs.

`interval.dropoff` (optional) drop-off value for QTL support intervals. Default is 1.5.

`filename` (optional) path of the file where the qtlSummary is to be stored. Default is NULL.

Value

Returns a data frame with a QTL summary which contains the following headers:

`traitName` name of trait.

`QTLchr` the chromosome number where a QTL locates.

`QTLmk` the marker where the trait maps to.

`QTLleftcm` the cM position of left border of the QTL support interval.

`QTLpeakcm` the cM position of the QTL peak.

`QTLrightcm` the cM of right border of the QTL support interval.

`logp` the $-log_{10}(p)$ value of a QTL.

`VarP1` the percentage of qualitative variance explained by a QTL.

`VarP2` the percentage of quantitative variance explained by a QTL.

`additive` the allele substitution effect (=half the difference of metabolite abundance between genotype 1 and 2).

See `qtlSumm` example data.

Note

The individual columns of genotypes and traits must have the same order. The markers should be ordered sequentially. The names of markers, traits and individuals should be consistent over `markers`, `genotypes`, `traits` and `qtlProfiles`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use markers, genotypes and traits as example data sets or use loadData to load your own data.
Use qtlMapTwoPart to calculate qtlProfiles.
Use qtlThreshold to estimate qtlThres QTL threshold for significance.
Use MetaNetwork for automated application of this function as part of a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers    <- loadData("markers.csv")
#genotypes  <- loadData("genotypes.csv")
#traits     <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres    <- 3.79

##OR: estimate the threshold yourself
#qtlThres   <- qtlThreshold(genotypes, traits, spike=4)

##summarize the qtlProfiles
qtlSumm     <- qtlSummary(markers, genotypes, traits, qtlProfiles, spike=4,
                          qtlThres=qtlThres)

##show the summary
qtlSumm[1:5,]
```

---

qtlSupportInterval *Estimate QTL support interval*

Description

For one QTL profile, significant QTLs are selected based on qtlThres. The regions within the interval.dropoff of these QTL peaks are defined as a support interval. The most left and the most right marker within each support interval are returned as matrix.

Usage

```
qtlSupportInterval(markers, oneQtlProfile, qtlThres, interval.dropoff = 1.5)
```

Arguments

markers	matrix of markers (rownames) and their chromosome numbers (column 1) and centi-Morgan positions (cM, column 2), ordered by position. See markers example data.
oneQtlProfile	one row from the qtlProfiles matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values. See qtlProfiles example data.
qtlThres	numeric $-log_{10}(p)$ threshold value for significant QTLs.
interval.dropoff	numeric drop-off $-log_{10}(p)$ value from the QTL peak that defines left and right border of support interval (optional). Default is 1.5.

Value

Returns a matrix of the markers number that define the left and right borders of each support interval for oneQtlProfile.

Note

The markers should be ordered sequentially. The names of markers should be consistent over markers and qtlProfiles.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use markers as example data set or use loadData to load your own data.
Use qtlMapTwoPart for the calculation of qtlProfiles.
Use qtlThreshold for the estimation of qtlThres QTL significance threshold.
Use qtlSummary for automated application of this function to produce a support interval summary for a set qtlProfiles.
Use MetaNetwork for automated appliction of this function as part a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers    <- loadData("markers.csv")
#genotypes  <- loadData("genotypes.csv")
```

```
#traits     <- loadData("traits.csv")

##calculate the two part qtl for only the first trait
qtlProfiles <- qtlMapTwoPart(genotypes=genotypes, traits=traits[1,], spike=4)

##set the qtl threshold
qtlThres    <- 3.79

##OR: estimate the threshold yourself
#qtlThres   <- qtlThreshold(genotypes, traits, spike=4)

##calculate qtl confidence interval for the first qtl profile
qtlSuppInt  <- qtlSupportInterval(markers, qtlProfiles[1,], qtlThres=qtlThres,
                                  interval.dropoff = 1.5)

##show the interval
qtlSuppInt
```

`qtlPlot` *Plot QTL profiles*

Description

Plot QTL profiles.

Usage

```
qtlPlot(markers, qtlProfiles, qtlThres, addTitle=NULL, addMarkerLabels=F,
        addLegend=T, color=NULL)
```

Value

Draws a plots that visualizes the $-log_{10}(p)$ QTL values (y-axis) along the markers on the genome (x-axis), parameterized based on user-defined settings.

Note

The `markers` should be ordered sequentially. The names of makers should be consistent over `markers` and `qtlProfiles`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers    <- loadData("markers.csv")
#genotypes  <- loadData("genotypes.csv")
#traits     <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres    <- 3.79

##OR: estimate the threshold yourself
#qtlThres   <- qtlThreshold(genotypes, traits, spike=4)

##plot the qtlProfiles
qtlPlot(markers, qtlProfiles, qtlThres)
```

---

qtlCorrZeroOrder *Calculate the zero-order correlation*

Description

Calculate the zero-order correlation on QTL profiles.

Usage

```
qtlCorrZeroOrder(markers, qtlProfiles, qtlThres, filename=NULL)
```

Arguments

markers	matrix of markers (rownames) and their chromosome numbers (column 1) and centi-Morgan positions (cM, column 2), ordered by position. See markers example data.
qtlProfiles	matrix of QTL mapping of traits (rownames) to markers (columnnames), as $-log_{10}(p)$ values. See qtlProfiles example data.
qtlThres	numeric $-log_{10}(p)$ threshold value for significant QTLs.
filename	(optional) path of the file where the correlations are to be stored. Default NULL.

Details

QTL support intervals are determined (via qtlSupportInterval with interval.dropoff = 1.5) and the $-log_{10}(p)$ values outside of the borders of these intervals are set to zero. Pairwise correlation coefficients between any two traits are then calculated as

$$r_{xy} = \frac{2\sum_{i=1}^{n} x_i * y_i}{\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2}$$

where r_{xy} is the correlation coefficient between qtlProfiles x and y and $i(i = 1...n)$ is a marker. x_i and y_i represent $-log_{10}(p)$ QTL profile values for marker i.

Value

Returns a matrix of correlation coefficients.

Note

The markers should be ordered sequentially. The names of markers and traits should be consistent over `qtlProfiles` and `markers`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `markers` as example data set or use `loadData` to load your own data.
Use `qtlSupportInterval` to calculate support intervals.
Use `qtlMapTwoPart` to calculate `qtlProfiles`.
Use `qtlThreshold` to estimate `qtlThres` QTL significance threshold .
Use `MetaNetwork` for automated application of this function as part a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers     <- loadData("markers.csv")
#genotypes   <- loadData("genotypes.csv")
```

```
#traits      <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles  <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres     <- 3.79

##OR: estimate the threshold yourself
#qtlThres    <- qtlThreshold(genotypes, traits, spike=4)

##calculate zero order correlation
qtlZeroOrder <- qtlCorrZeroOrder(markers, qtlProfiles, qtlThres)

##show the correlations
qtlZeroOrder[1:5,1:5]
```

`qtlCorrSecondOrder` *Calculate the second-order correlation*

Description

Calculate the second-order partial correlation between any pair of QTL profiles to decompose the confounded effect of multiple QTLs.

Usage

```
qtlCorrSecondOrder(corrZeroOrder, topCorNo=20, filename = NULL)
```

Arguments

`corrZeroOrder`	a zero order correlation matrix that has been calculated before. See `corrZeroOrder` example data.
`topCorNo`	(optional) to reduce computing time, limit the pairwise conditioning to only the `topCorNo` top correlated compounds. Default is 20. To condition on all pairs, set topCorNo to nrow(corrZeroOrder).
`filename`	(optional) path of the file where the correlations are to be stored. Default NULL.

Details

First-order correlation between x and y conditional on a single variable z are calculated by

$$r_{xy|z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

where r_{xy}, r_{xz} and r_{yz} are zero-order correlation coefficients between traits x and y, x and z, and y and z, respectively (provided as `corrZeroOrder`.

Then second-order partial correlation between x and y, conditional on a pair of variables z and k is a function of the first-order coefficients.

$$r_{xy|zk} = \frac{r_{xy|z} - r_{xk|z}r_{yk|z}}{\sqrt{(1 - r^2_{xk|z})(1 - r^2_{yk|z})}}$$

For each pair x and y, the second-order partial correlations are calculated conditional on each pair z and k and the minimal value is stored.

In order to save computing time, conditional variables z and k can be chosen from `topCorNo` top correlated variable x and y (e.g. `topCorNo` = 20).

Value

Returns a matrix of second-order partial correlations.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

Source

Keurentjes JJB, FU J, de vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, and Koornneef M. The genetics of plant metabolism. Nature Genetics (2006) 7: 842-849.

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `qtlCorrZeroOrder` to calculate `corrZeroOrder` zero order correlation.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers        <- loadData("markers.csv")
#genotypes      <- loadData("genotypes.csv")
#traits         <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles     <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres        <- 3.79

##OR: estimate the threshold yourself
#qtlThres       <- qtlThreshold(genotypes, traits, spike=4)
```

```
##calculate zero order correlation
corrZeroOrder   <- qtlCorrZeroOrder(markers, qtlProfiles, qtlThres)

##calculate the second order correlation
corrSecondOrder <- qtlCorrSecondOrder(corrZeroOrder=corrZeroOrder)

##view the correlations
corrSecondOrder[1:5,1:5]
```

`qtlCorrThreshold` *Estimate correlation threshold*

Description

Permutation test to estimate threshold for significant QTL correlation. In each permutation, the relation between genotype and trait are disturbed and then correlation is calculated. Thus, any detected genetic correlation is assumed to be false. Compute the partial correlation coefficient in each permutation and record highest value. Then after `n.permutations`, the threshold is set at desired `alpha` level, Bonferroni corrected by the number of edges per trait (the number of traits-1).

Usage

```
qtlCorrThreshold(markers, genotypes, traits, spike, qtlThres,
                 n.permutations = 10000, alpha = 0.05, method = "qtl")
```

Arguments

`markers`	matrix of markers (rownames) and their chromosome numbers (column 1) and centi-Morgan positions (cM, column 2), ordered by position. See `markers` example data.
`genotypes`	matrix of genotypes for each marker (rownames) and individual (columnnames), as numeric values 1, 2 or NA when missing. See `genotypes` example data.
`traits`	matrix of phenotypes for each trait (rownames) and individual (columnnames), as numeric or NA when missing. See `traits` example data.
`spike`	numeric cut-off value to separate absent (qualitative) from available (quantitative) trait abundance.
`qtlThres`	numeric $-log_{10}(p)$ threshold value for significant QTLs.
`n.permutations`	(optional) number of iterations of permutation. Default is 10000 times.
`alpha`	(optional) Bonferonni corrected significant level. Default is 0.05.
`method`	(optional) correlation method string, "qtl" or "abundance" for correlation on QTL profiles or metabolite abundance level, respectively. If `method` is "qtl" (default), qtlCorrThreshold will call `qtlCorrZeroOrder` to calculate the correlation between QTL profiles. Otherwise, when `method` is "abundance", qtlCorrThreshold will use Spearman correlation with the `cor` function to calculate the correlation between metabolite abundance profiles.

Value

Returns: threshold value for correlation coefficient.

`corrPermutations`
: a vector of the permutations of maximum, absolute correlation values.

Note

The individual columns of `genotypes` and `traits` must have the same order. The `markers` should be ordered sequentially. The names of markers, traits and individuals should be consistent over `markers`, `genotypes`, and `traits`.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use `markers`, `genotypes` and `traits` as example data sets or use `loadData` to load your own data.
Use `qtlThreshold` to estimate `qtlThres` QTL significance threshold.
Use `qtlCorrZeroOrder` and `qtlCorrSecondOrder` for QTL correlations that can be test against the resulting correlation threshold.
Use `MetaNetwork` for automated appliction of this function as part a genetic analysis protocol on metabolites.

Examples

```
## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers    <- loadData("markers.csv")
#genotypes  <- loadData("genotypes.csv")
#traits     <- loadData("traits.csv")

##set the qtl threshold
qtlThres    <- 3.79

##OR: estimate the threshold yourself
#qtlThres   <- qtlThreshold(genotypes, traits, spike=4)

##estimate qtlCorrThreshold with low number of permutations (advised: 10000)
cat("warning: simulating correlations takes a few minutes")
flush.console()
corrThres   <- qtlCorrThreshold(markers, genotypes, traits, spike=4,
                                qtlThres=qtlThres, n.permutations=10)
```

```
##show threshold
corrThres
```

`createCytoFiles` *Write Cytoscape visualization files*

Description

Create visualization files for Cytoscape, the network visualization software. Any correlation higher than `simThres` is written into Cytoscape network files and can be loaded into Cytoscape for visualization.

Usage

```
createCytoFiles(corrMatrix, filename, simThres = NULL, hideNodes = T)
```

Arguments

`corrMatrix`	matrix for pair-wise correlation. See `corrSecondOrder` example data.
`filename`	network file name, without extension. Extensions .sif and .eda will be added for network and edge attribute files, respectively.
`simThres`	(optional) numeric similarity threshold if the values in matrix are similarity values such as correlation coefficients. Default is NULL.
`hideNodes`	(optional) logical value to hide nodes without any significant links. Default is TRUE

Value

A network file (`filename`.sif) and edge attribute file (`filename`.eda) are generated.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

http://www.cytoscape.org/

See Also

Use `cor`, `qtlCorrZeroOrder` and `qtlCorrSecondOrder` to calculate `corrMatrix` correlation matrix.
Use `MetaNetwork` for automated appliction of this function as part a genetic analysis protocol on metabolites.

Examples

```
##NOTE: this method can be used on any correlation matrix.
#Here we use MetaNetwork methods.

## load the example data provided with this package
data(markers)
data(genotypes)
data(traits)

##OR: load your own data
#markers        <- loadData("markers.csv")
#genotypes      <- loadData("genotypes.csv")
#traits         <- loadData("traits.csv")

##calculate the two part qtl
qtlProfiles     <- qtlMapTwoPart(genotypes=genotypes, traits=traits, spike=4)

##set the qtl threshold
qtlThres        <- 3.79

##OR: estimate the threshold yourself
#qtlThres       <- qtlThreshold(genotypes, traits, spike=4)

##calculate zero order correlation
corrZeroOrder   <- qtlCorrZeroOrder(markers, qtlProfiles, qtlThres)

##calculate second order correlation
corrSecondOrder <- qtlCorrSecondOrder( corrZeroOrder )

##set the correlation threshold
corrThres       <- 0.14

##OR: estimate qtlCorrThreshold yourself
#corrThres      <- qtlCorThreshold(markers, genotypes, traits, spike=4, qtlThres=qtlThres

##create cytoscape files "mynetwork.sif" and "mynetwork.eda"
createCytoFiles(corrSecondOrder, "mynetwork", simThres = corrThres)

cat("cytofiles mynetwork.sif and mynetwork.eda created\n")
```

`findMultiplePeaks` *Cluster multiple mass peaks based on correlation and mass*

Description

Predict metabolites that may be represented by multiple mass peaks based on correlation and m/z value. If two mass peaks are highly correlated (r>0.95) and their mass difference is 1 or 2 (as isotopes) or mass ratio is 2, 3, or 1/2, 1/3 (as different charged), they are predicted as multiple peaks of a same metabolite.

Usage

```
findMultiplePeaks <- function( corrZeroOrder, peaks, corrThres=0.95,
                               filename=NULL) )
```

Arguments

corrZeroOrder	a zero order peak correlation matrix that has been calculated before. See corrZeroOrder example data.
peaks	matrix of mass/charge peaks (column1) for each trait (rownames). See peaks2 example peaks data for unidentified example traits data traits2.
corrThres	(optional) numeric threshold for significant correlated peaks that should be tested for multiple peaks relationships. Default is 0.95.
filename	(optional) path of the csv file where the multiple peak summary is to be stored. Default NULL.

Value

Returns a data frame with Multiple Peaks summary containing the following headers:

cluster	cluster id of related multiple peaks.
peak1	mass over charge peak.
mz1	mass over charge ratio of peak1.
peak2	mass over charge peak that is related to peak1.
mz2	mass over charge ratio of peak2.
corrCoef	correlation coefficient between peak1 and peak2.
massDiff	peak difference: mz1 - mz2.
massRatio	peak ratio: mz1/mz2.
relationship	putative isotope or diffCharged relationship.

See multiplePeaks example data set.

Note

The names of traits should be consistent over corrZeroOrder and peaks.

Author(s)

Jingyuan Fu <j.fu@rug.nl>, Morris Swertz <m.a.swertz@rug.nl>, Ritsert Jansen <r.c.jansen@rug.nl>

References

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. MetaNetwork: a computational tool for the genetic study of metabolism. Nature Protocols (2007).

http://gbic.biol.rug.nl/supplementary/2007/MetaNetwork

See Also

Use traits2 and peaks2 as example unidintified peak data set or use loadData to load your own data.
Use qtlCorrZeroOrder to calculate corrZeroOrder zero order correlation.

Examples

```
## load the example data provided with this package
data(traits2)
data(peaks2)
data(genotypes)
data(markers)

##OR: load your own data
#traits2      <- loadData("traits2.csv")
#peaks2       <- loadData("peaks2.csv")
#genotypes    <- loadData("genotypes.csv")
#markers      <- loadData("markers.csv")

##calculate the two part qtl
qtlProfiles   <- qtlMapTwoPart(genotypes=genotypes, traits=traits2, spike=4)

##set the qtl threshold
qtlThres      <- 3.79

##OR: estimate the threshold yourself
#qtlThres     <- qtlThreshold(genotypes, traits, spike=4)

##calculate zero order correlation
corrZeroOrder <- qtlCorrZeroOrder(markers, qtlProfiles, qtlThres)

##find multiple peaks
multiplePeaks <- findMultiplePeaks(corrZeroOrder, peaks2)

##show summary
multiplePeaks
```

Index