

Package ‘CONCUR’

October 12, 2022

Type Package

Title Copy Number Profile Curve-Based Association Test

Version 1.4

Date 2022-06-05

Author Amanda Brucker, Shannon T. Holloway, Jung-Ying Tzeng

Maintainer Shannon T. Holloway <shannon.t.holloway@gmail.com>

Description Implements a kernel-based association test for copy number variation (CNV) aggregate analysis in a certain genomic region (e.g., gene set, chromosome, or genome) that is robust to the within-locus and across-locus etiological heterogeneity, and bypass the need to define a “locus” unit for CNVs.
Brucker, A., et al. (2020) <[doi:10.1101/666875](https://doi.org/10.1101/666875)>.

License GPL-2

NeedsCompilation no

Repository CRAN

Encoding UTF-8

Depends CompQuadForm, dplyr, mgcv

RoxygenNote 7.1.0

Collate 'commonAUC.R' 'segLength.R' 'popFunc.R' 'uniqueCombinations.R'
'createDF.R' 'deldup.R' 'CAUCkernel.R' 'cnvData.R' 'compFunc.R'
'dataChecks.R' 'test.qt.R' 'test.bt.R' 'pv_func.R' 'concur.R'

Date/Publication 2022-06-07 12:00:28 UTC

R topics documented:

cnvData	2
concur	2
covData	5
phenoData	6

Index	7
--------------	----------

`cnvData`*Pseudo Copy Number Variants Data*

Description

This data set includes simulated CNV data in PLINK CNV data format. The data are also available from the authors through the url provided below. These data were generated following the simulation study used to illustrate the method in the original manuscript also referenced below; it has been reduced to include only 600 individuals. These data are not meaningful and are intended for demonstration purposes only.

Usage`data(cnvData)`**Format**

`cnvData` is a data.frame containing 522 observations with 5 columns:

ID character patient identifier.

CHR CNV chromosome.

BP1 starting location in base pairs.

BP2 ending location in base pairs.

TYPE copy number (0,1,3,or 4).

References

Brucker, A., Lu, W., Marceau West, R., Yu, Q-Y., Hsiao, C. K., Hsiao, T-H., Lin, C-H., Magnusson, P. K. E., Holloway, S. T., Sullivan, P. F., Szatkiewicz, J. P., Lu, T-P., and Tzeng, J-Y. Association testing using Copy Number Profile Curves (CONCUR) enhances power in copy number variant analysis. <doi:10.1101/666875>.

<https://www4.stat.ncsu.edu/~jytzeng/Software/CONCUR/>

`concur`*Copy Number Profile Curve-Based Association Test*

Description

Implements a kernel-based association test for CNV aggregate analysis in a certain genomic region (e.g., gene set, chromosome, or genome) that is robust to the within-locus and across-locus etiological heterogeneity, and bypass the need to define a "locus" unit for CNVs.

Usage

```

concur(
  cnv,
  X,
  pheno,
  phenoY,
  phenoType,
  ...,
  nCore = 1L,
  outFileKernel = NULL,
  verbose = TRUE
)

```

Arguments

cnv	A character or data.frame object. If character, the name of the data file containing the CNV data (with a header). If data.frame, the CNV data. The data must contain the following columns: "ID", "CHR", "BP1", "BP2", "TYPE", where "ID" is a unique patient id, "CHR" is the CNV chromosome, "BP1" is the start location in base pairs or kilo-base pairs, "BP2" is the end location in base pairs or kilo-base pairs, and "TYPE" is the CNV copy number.
X	A character or data.frame object. If character, the name of the data file containing the covariate data (with a header). If data.frame, the covariate data. The data must contain a column titled "ID" containing a unique patient id. This column must contain the patient identifiers of the CNV data specified in input cnv; however, it can contain patient identifiers not contained in cnv. Further, inputs X and pheno must contain the same patient identifiers. Categorical variables must be translated into design matrix format.
pheno	A character or data.frame object. If character, the name of the data file containing the phenotype data (with a header). If data.frame, the phenotype data. The data must contain a column titled "ID" containing a unique patient id. This column must contain the patient identifiers of the CNV data specified in input cnv; however, it can contain patient identifiers not contained in cnv. Further, inputs X and pheno must contain the same patient identifiers.
phenoY	A character object. The column name in input pheno containing the phenotype of interest.
phenoType	A character object. Must be one of {"bin", "cont"} indicating if input phenoY (i.e., the phenotype of interest) is binary or continuous.
...	Ignored. Included to require named inputs.
nCore	An integer object. If nCore > 1, package parallel is used to calculate the kernel. Though the methods of package CompQuadForm dominate the time profile, setting nCore > 1L can improve computation times.
outFileKernel	A character object or NULL. If a character, the file in which the kernel is to be saved. If NULL, the kernel is returned by the function.
verbose	A logical object. If TRUE, progress information is printed to the screen.

Details

The CNV data must adhere to the following conditions:

- CNVs must be at least 1 unit long.
- CNVs cannot end at the exact location another begins

Violations of these conditions typically occur when data are rounded to a desired resolution. For example

ID	CHR	BP1	BP2	TYPE
1	13	10112087	10112414	3

becomes upon rounding to kilo

ID	CHR	BP1	BP2	TYPE
1	13	10112	10112	3

These cases should either be discarded or modified to be of length 1, e.g.,

ID	CHR	BP1	BP2	TYPE
1	13	10112	10113	3

As an example of condition 2

ID	CHR	BP1	BP2	TYPE
1	13	100768	101100	3
1	13	101100	101299	1

should be modified to one of

ID	CHR	BP1	BP2	TYPE
1	13	100768	101100	3
1	13	101101	101299	1

or

ID	CHR	BP1	BP2	TYPE
1	13	100768	101099	3
1	13	101100	101299	1

Additionally,

ID	CHR	BP1	BP2	TYPE
1	13	100768	101100	3
1	13	101100	101299	3

should be combined as

ID	CHR	BP1	BP2	TYPE
1	13	100768	101299	3

Value

A list containing the kernel (or its file name) and the p-value.

References

Brucker, A., Lu, W., Marceau West, R., Yu, Q-Y., Hsiao, C. K., Hsiao, T-H., Lin, C-H., Magnusson, P. K. E., Holloway, S. T., Sullivan, P. F., Szatkiewicz, J. P., Lu, T-P., and Tzeng, J-Y. Association testing using Copy Number Profile Curves (CONCUR) enhances power in copy number variant analysis. <doi:10.1101/666875>.

Examples

```
data(cnvData)

# limit data for examples
exCNV <- cnvData$ID %in% paste0("P", 1:150)
exCOV <- covData$ID %in% paste0("P", 1:150)
exPHE <- phenoData$ID %in% paste0("P", 1:150)

# binary phenoType
results <- concur(cnv = cnvData[exCNV,],
                 X = covData[exCOV,],
                 pheno = phenoData[exPHE,],
                 phenoY = 'PHEB',
                 phenoType = 'bin',
                 nCore = 1L,
                 outFileKernel = NULL,
                 verbose = TRUE)

# continuous phenoType
results <- concur(cnv = cnvData[exCNV,],
                 X = covData[exCOV,],
                 pheno = phenoData[exPHE,],
                 phenoY = 'PHEC',
                 phenoType = 'cont',
                 nCore = 1L,
                 outFileKernel = NULL,
                 verbose = TRUE)
```

covData

Pseudo Covariate Data

Description

This data set includes simulated covariate data. These data were generated as draws from a Binom(1,0.5) distribution for the 800 individuals in the example data provided with the package. These data are not meaningful and are intended for demonstration purposes only.

Usage

```
data(cnvData)
```

Format

covData is a data.frame containing 400 observations with 2 columns

ID character patient identifier.

SEX binary indicator of M/F.

References

Brucker, A., Lu, W., Marceau West, R., Yu, Q-Y., Hsiao, C. K., Hsiao, T-H., Lin, C-H., Magnusson, P. K. E., Holloway, S. T., Sullivan, P. F., Szatkiewicz, J. P., Lu, T-P., and Tzeng, J-Y. Association testing using Copy Number Profile Curves (CONCUR) enhances power in copy number variant analysis. <doi:10.1101/666875>.

phenoData

Pseudo Phenotype Data

Description

This data set includes simulated phenotype data. These data include a binary phenotype and a normally distributed continuous phenotype that are randomly generated independent of the CNV data. These data are not meaningful and are intended for demonstration purposes only.

Usage

```
data(cnvData)
```

Format

phenoData is a data.frame containing 400 observations with 3 columns

ID character patient identifier.

PHEB binary phenotype.

PHEC continuous phenotype.

References

Brucker, A., Lu, W., Marceau West, R., Yu, Q-Y., Hsiao, C. K., Hsiao, T-H., Lin, C-H., Magnusson, P. K. E., Holloway, S. T., Sullivan, P. F., Szatkiewicz, J. P., Lu, T-P., and Tzeng, J-Y. Association testing using Copy Number Profile Curves (CONCUR) enhances power in copy number variant analysis. <doi:10.1101/666875>.

Index

* datasets

cnvData, [2](#)

covData, [5](#)

phenoData, [6](#)

cnvData, [2](#)

concur, [2](#)

covData, [5](#)

phenoData, [6](#)