Package: acnr (via r-universe)

September 3, 2024

Type Package

Title Annotated Copy-Number Regions

Version 0.3.2

Date 2017-01-09

Author Morgane Pierre-Jean and Pierre Neuvial

Maintainer Morgane Pierre-Jean <`mpierrejean.pro@gmail.com`>

Description This data package provides SNP array data from different types of copy-number regions. These regions were identified manually by the authors of the package and may be used to generate realistic data sets with known truth.

License LGPL (>= 2.1)

Depends R (>= 2.10),

Suggests R.utils, xtable, knitr, rmarkdown, testthat

RoxygenNote 5.0.1

VignetteBuilder knitr

URL `https://github.com/mpierrejean/acnr`

BugReports `https://github.com/mpierrejean/acnr/issues`

Repository https://mpierrejean.r-universe.dev

RemoteUrl https://github.com/mpierrejean/acnr

RemoteRef HEAD

RemoteSha d65dbfa70315c6f8a2caa6f7077c4c19ce24190a

Contents

`acnr-package` *Annotated Copy-Number Regions*

Description

This data package contains SNP array data from different types of copy-number regions. These regions were identified manually by the authors of the package and may be used to generate realistic data sets with known truth.

Details

Package:	acnr
Type:	Package
Title:	Annotated Copy-Number Regions
Version:	0.2.2
Date:	2014-09-08
Author:	Morgane Pierre-Jean and Pierre Neuvial
Maintainer:	Morgane Pierre-Jean <morgane.pierrejean@genopole.cnrs.fr>
License:	LGPL (>= 2.1)
Depends:	R (>= 2.10), R.utils
Suggests:	RUnit, BiocGenerics
biocViews:	ExperimentData

Author(s)

Morgane Pierre-Jean and Pierre Neuvial

`getMinorMajorCopyNumbers` *Get minor and major copy number labels from region annotation labels*

Description

Get minor and major copy number labels from region annotation labels

Usage

```
getMinorMajorCopyNumbers(region)
```

Arguments

`region`	A character value, the annotation label for a copy number region. Should be encoded as `"(C1,C2)"`, where `C1` denotes the minor copy number, that is, the smallest of the two parent-specific copy numbers `C2` denotes the minor copy number, that is, the smallest of the two parent-specific copy numbers

Value

A `matrix` with `length(region)` rows and two columns: `C1` and `C2`, as described above.

References

Bengtsson H., Neuvial, P. and Speed, T. P. (2010) TumorBoost: normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays. BMC bioinformatics 11 (2010), p. 245.

Neuvial, P., Bengtsson H., and Speed, T. P. (2011) Statistical analysis of Single Nucleotide Polymorphism microarrays in cancer studies. Chapter 11 in *Handbook of Statistical Bioinformatics*, Springer.

Examples

```
dat <- loadCnRegionData(dataSet="GSE29172_H1395", tumorFraction=1)
regions <- unique(dat$region)
getMinorMajorCopyNumbers(regions)
```

`GSE11976_CRL2324` *Annotated copy-number regions from the GEO GSE11976 data set.*

Description

The GEO GSE11976 data set is a dilution series from the Illumina HumanCNV370v1 chip type (Staaf et al, 2008).

Format

A data frame with 770668 observations of 7 variables:

c total copy number (not log-scaled)

b allelic ratios in the diluted tumor sample (after TumorBoost)

genotype germline genotypes

region a character value, annotation label for the region. Should be encoded as `"(C1,C2)"`, where `C1` denotes the minor copy number and `C2` denotes the major copy number. For example,

- **(1,1)** Normal
- **(0,1)** Hemizygous deletion
- **(0,0)** Homozygous deletion
- **(1,2)** Single copy gain
- **(0,2)** Copy-neutral LOH
- **(2,2)** Balanced two-copy gain
- **(1,3)** Unbalanced two-copy gain
- **(0,3)** Single-copy gain with LOH

cellularity A numeric value between 0 and 1, the percentage of tumor cells in the sample.

@source http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11976 @references Staaf, J., Lindgren, D., Vallon-Christersson, J., Isaksson, A., Goransson, H., Juliusson, G., ... & Ringn\'er, M. (2008). Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. Genome Biol, 9(9), R136.

Details

These data have been processed from the files available at http://cbbp.thep.lu.se/~markus/software/BAFsegmentation/ using scripts that are included in the 'inst/preprocessing/GSE11976' directory of this package.

Examples

```
dat <- loadCnRegionData("GSE11976_CRL2324")
unique(dat$region)
```

`GSE13372_HCC1143` *Annotated copy-number regions from the GEO GSE13372 data set.*

Description

The GEO GSE13372 data set is from the Affymetrix GenomeWideSNP_6 chip type. We have extracted one tumor/normal pair corresponding to the breast cancer cell line HCC1143. For consistency with the other data sets in the package the tumor and normal samples are labeled according to their tumor cellularity, that is, 100

Format

A data frame with 205842 observations of 7 variables:

c total copy number (not log-scaled)

b allelic ratios in the diluted tumor sample (after TumorBoost)

genotype germline genotypes

bT allelic ratios in the diluted tumor sample (before TumorBoost)

bN allelic ratios in the matched normal sample

region a character value, annotation label for the region. Should be encoded as `"(C1,C2)"`, where `C1` denotes the minor copy number and `C2` denotes the major copy number. For example,

- **(1,1)** Normal
- **(0,1)** Hemizygous deletion
- **(0,0)** Homozygous deletion
- **(1,2)** Single copy gain
- **(0,2)** Copy-neutral LOH
- **(2,2)** Balanced two-copy gain
- **(1,3)** Unbalanced two-copy gain
- **(0,3)** Single-copy gain with LOH

genotype the (germline) genotype of SNPs. By definition, rows with missing genotypes are interpreted as non-polymorphic loci (a.k.a. copy number probes).

cellularity A numeric value between 0 and 1, the percentage of tumor cells in the sample.

Details

These data have been processed from the files available from GEO using scripts that are included in the 'inst/preprocessing/GSE13372' directory of this package. This processing includes normalization of the raw CEL files using the CRMAv2 method implemented in the aroma.affymetrix package.

Source

`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13372` `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13372`

References

Chiang DY, Getz G, Jaffe DB, O'Kelly MJ et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods 2009 Jan;6(1):99-103. PMID: 19043412

Bengtsson, H., Wirapati , P. & Speed, T.P. (2009). A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6, Bioinformatics 25(17), pp. 2149-56.

Bengtsson H., Neuvial, P. and Speed, T. P. (2010) TumorBoost: normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays. BMC bioinformatics 11 (2010), p. 245.

Examples

```
dat <- loadCnRegionData("GSE13372_HCC1143")
unique(dat$region)
```

`GSE29172_H1395` *Annotated copy-number regions from the GEO GSE29172 (and GSE26302) data sets.*

Description

The GEO GSE29172 data set is a dilution series from the Affymetrix GenomeWideSNP_6 chip type. The GEO GSE26302 data set contains the experiment corresponding to the matched normal (i.e. 0% dilution).

Format

A data frame with 770668 observations of 7 variables:

c total copy number (not log-scaled)

b allelic ratios in the diluted tumor sample (after TumorBoost)

genotype germline genotypes

bT allelic ratios in the diluted tumor sample (before TumorBoost)

bN allelic ratios in the matched normal sample

region a character value, annotation label for the region. Should be encoded as `"(C1,C2)"`, where `C1` denotes the minor copy number and `C2` denotes the major copy number. For example,

- **(1,1)** Normal
- **(0,1)** Hemizygous deletion
- **(0,0)** Homozygous deletion
- **(1,2)** Single copy gain
- **(0,2)** Copy-neutral LOH
- **(2,2)** Balanced two-copy gain
- **(1,3)** Unbalanced two-copy gain
- **(0,3)** Single-copy gain with LOH

genotype the (germline) genotype of SNPs. By definition, rows with missing genotypes are interpreted as non-polymorphic loci (a.k.a. copy number probes).

cellularity A numeric value between 0 and 1, the percentage of tumor cells in the sample.

Details

These data have been processed from the files available from GEO using scripts that are included in the 'inst/preprocessing/GSE29172' directory of this package. This processing includes normalization of the raw CEL files using the CRMAv2 method implemented in the aroma.affymetrix package.

Source

`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29172` `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26302`

References

Rasmussen, M., Sundstr\"om, M., Kultima, H. G., Botling, J., Micke, P., Birgisson, H., Glimelius, B. & Isaksson, A. (2011). Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. Genome Biology, 12(10), R108.#'

Bengtsson, H., Wirapati , P. & Speed, T.P. (2009). A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6, Bioinformatics 25(17), pp. 2149-56.

Bengtsson H., Neuvial, P. and Speed, T. P. (2010) TumorBoost: normalization of allele-specific tumor copy numbers from a single pair of tumor-normal genotyping microarrays. BMC bioinformatics 11 (2010), p. 245.

Examples

```
dat <- loadCnRegionData("GSE29172_H1395")
unique(dat$region)
```

`listDataSets` *List available data sets*

Description

List available data sets

Usage

```
listDataSets()
```

Value

name of one of the data sets of the package, see listDataSets

Examples

```
listDataSets()
```

`listTumorFractions` *List of available tumor fractions for a data set*

Description

List of available tumor fractions for a data set

Usage

```
listTumorFractions(dataSet)
```

Arguments

`dataSet`	The name of a data set from the package, see listDataSets

Value

A numeric vector, the available tumor fractions for a data set

Examples

```
dataSets <- listDataSets()
fracs <- listTumorFractions(dataSets[1])
```

`loadCnRegionData` *loadCnRegionData*

Description

Load real, annotated copy number data

Usage

```
loadCnRegionData(dataSet, tumorFraction = 1)
```

Arguments

`dataSet`	name of one of the data sets of the package, see `listDataSets`
`tumorFraction`	proportion of tumor cells in the "tumor" sample (a.k.a. tumor cellularity). See `listTumorFractions`.

Details

This function is a wrapper to load real genotyping array data taken from

* a dilution series from the Affymetrix GenomeWideSNP_6 chip type (Rasmussen et al, 2011), see `GSE29172_H1395` * a dilution series from the Illumina HumanCNV370v1 chip type (Staaf et al, 2008), see `GSE11976_CRL2324` * a tumor/normal pair from the Affymetrix GenomeWideSNP_6 chip type (Chiang et al, 2008), see `GSE13372_HCC1143`

Value

a data.frame containing copy number data for different types of copy number regions. Columns:

c Total copy number

b Allele B fraction (a.k.a. BAF)

region a character value, annotation label for the region. Should be encoded as `"(C1,C2)"`, where `C1` denotes the minor copy number and `C2` denotes the major copy number. For example,

- **(1,1)** Normal
- **(0,1)** Hemizygous deletion
- **(0,0)** Homozygous deletion
- **(1,2)** Single copy gain
- **(0,2)** Copy-neutral LOH
- **(2,2)** Balanced two-copy gain
- **(1,3)** Unbalanced two-copy gain
- **(0,3)** Single-copy gain with LOH

muN the (germline) genotype of SNPs. By definition, rows with missing genotypes are interpreted as non-polymorphic loci (a.k.a. copy number probes).

Author(s)

Morgane Pierre-Jean and Pierre Neuvial

Examples

```
affyDat <- loadCnRegionData(dataSet="GSE29172_H1395", tumorFraction=1)
str(affyDat)

illuDat <- loadCnRegionData(dataSet="GSE11976_CRL2324", tumorFraction=.79)
str(illuDat)

affyDat2 <- loadCnRegionData(dataSet="GSE13372_HCC1143", tumorFraction=1)
str(affyDat2)
```

Index