Package 'dinamic'

February 19, 2015

Title DiNAMIC A Method To Analyze Recurrent DNA Copy Number Aberrations in Tumors

Version 1.0

Date 2011-10-24

Author Vonn Walter, Andrew B. Nobel, and Fred A. Wright

Maintainer Vonn Walter <vwalter@email.unc.edu>

Description This function implements the DiNAMIC procedure for assessing the statistical significance of recurrent DNA copy number aberrations (Bioinformatics (2011) 27(5) 678 - 685).

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-29 08:58:34

NeedsCompilation no

R topics documented:

dinamic-package *Assessing the Statistical Significance of Recurrent DNA Copy Number Aberrations*

Description

The DiNAMIC method for assessing the statistical significance of recurrent DNA copy number aberrations was presented in Bioinformatics (2011) 27(5) 678 - 685. This package contains the functions required to perform both DiNAMIC's *Quick Look* and *Detailed Look* procedures.

Details

Package:	dinamic
Type:	Package
Version:	1.0
Date:	2011-10-25
License:	GPL (>= 2)
LazyLoad:	yes

DNA copy number gains and losses are commonly found in tumor tissue. Collectively, we refer to these changes as DNA copy number aberrations (CNAs). Because of underlying genomic instability, many CNAs occur at random locations throughout the genome. These CNAs are termed *sporadic*, and they are not associated with the tumor phenotype. Some CNAs provide a selective growth advantage, so one would expect to find these CNAs in multiple independent samples. CNAs of the latter type are termed *recurrent*, and distinguishing between sporadic and recurrent CNAs is largely a statistical issue.

Gains and losses are analyzed separately, and both of DiNAMIC's main functions `quickLook` and `detailedLook` assess the statistical significance of recurrent gains (losses) using permutation-based null distributions. The null distribution is produced by applying a novel *cyclic shift* permutation scheme, and this is performed by the `findNull` function. DiNAMIC's `peeling` function allows users to assess the significance of multiple gains (losses). The significance of a new gain (loss) is assessed conditionally on having detected previous gains (losses). The package includes DNA copy number data and associated marker information from the publicly available Wilms' tumor dataset of Natrajan et al. (J. Pathology (2006) 210: 49 - 58), as well as a cytoband annotation file.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Natrajan, R., Williams, R.D., Hing, S.N., et al., Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse, J. Pathology (2006) 210: 49 - 58.

Fujita P.A., Rhead B., Zweig A.S., et al., The UCSC Genome Browser database: update 2011, Nucleic Acids Res. (2010) 1 - 7 doi:10.1093/nar/gkq963.

Examples

```
data(wilms.data)
data(wilms.markers)
data(annot.file)
detailedLook(wilms.data, wilms.markers, annot.file, 100, 3)
#Produces the following output:
#" 1"       "155656176" "R:A-MEXP-192:RP11-393K10" "196"  "0.01"
#"12"       " 38270107" "R:A-MEXP-192:RP11-519E12" "2294" "0.01"
#" 8"       "  4554176" "R:A-MEXP-192:RP11-337D8"  "1659" "0.01"

quickLook(wilms.data, wilms.markers, annot.file, 100, 3)
#Produces the following output:
#" 1"       "155656176" "R:A-MEXP-192:RP11-393K10" "196"  "0.01"
#"12"       " 38270107" "R:A-MEXP-192:RP11-519E12" "2294" "0.01"
#" 8"       "  4554176" "R:A-MEXP-192:RP11-337D8"  "1659" "0.01"
```

---

annot.file *A Cytoband Annotation Data Frame*

Description

This four-column data frame contains cytoband annotation data that is used by the `makeCytoband` function. Each row corresponds to a distinct cytoband, and column 1 contains the chromosome number, column 2 contains the start position (in base pairs), column 3 contains the end position (in base pairs), and column 4 contains the cytoband name (e.g. p21.3). Additional columns may be present, but they are not used.

Usage

```
data(annot.file)
```

Format

A data frame with 811 observations on the following 4 variables.

`Chr` The chromosome for the cytoband

`Start` The start position (in base pairs) for the cytoband

`End` The end position (in base pairs) for the cytoband

`Band` The cytoband name (e.g. p13.1)

Source

The file cytoBand.txt.gz for the hg19 build can be downloaded from the UCSC Genome Browser at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/. The format of cytoBand.txt differs from that of annot.file, but it can be used by the function makeCytoband if reformat.cytoband = TRUE.

References

Fujita P.A., Rhead B., Zweig A.S., et al., The UCSC Genome Browser database: update 2011, Nucleic Acids Res. (2010) 1 - 7 doi:10.1093/nar/gkq963.

Examples

```
data(annot.file)
annot.file[1:10,]
#Produces the following output
#Chr    Start      End   Band
#1    1        0  2300000 p36.33
#2    1  2300000  5300000 p36.32
#3    1  5300000  7100000 p36.31
#4    1  7100000  9200000 p36.23
#5    1  9200000 12600000 p36.22
#6    1 12600000 16100000 p36.21
#7    1 16100000 20300000 p36.13
#8    1 20300000 23800000 p36.12
#9    1 23800000 27800000 p36.11
#10   1 27800000 30000000  p35.3
```

---

detailedLook *Assessing the Significance of Recurrent DNA Copy Number Aberrations*

Description

This function applies the "Detailed Look" version of DiNAMIC's cyclic shift procedure to assess the statistical significance of recurrent DNA copy number aberrations. The statistical significance of recurrent gains (gain.loss = "gain") or recurrent losses (gain.loss = "loss") are assessed using an empirical null distribution produced by num.perms cyclic shifts of the DNA copy number matrix x. The null distribution is produced by findNull, which is called internally.

Usage

```
detailedLook(x, marker.data, annot.file, num.perms, num.iters,
gain.loss = "gain", reformat.annot = FALSE, random.seed = NULL)
```

Arguments

x	An n by m numeric matrix containing DNA copy number data from n subjects at m markers.
marker.data	A dataframe containing marker position data for markers in the autosomes. Column 1 contains the chromosome number for each marker, and column 2 contains the position (in base pairs) for each markers. Additional columns, if present, represent information about the markers (e.g. probe names).
annot.file	A cytoband annotation dataframe. Each row corresponds to a distinct cytoband, and column 1 contains the chromosome number, column 2 contains the start position (in base pairs), column 3 contains the end position (in base pairs), and column 4 contains the cytoband name (e.g. p21.3). Additional columns may be present, but they are not used.
num.perms	A positive integer that represents the number of cyclic shifts used to create the empirical null distribution.
num.iters	A positive integer that represents the number of distinct gain (loss) loci that will be assessed. See Details for more information.
gain.loss	A character string that indicates whether recurrent gains (gain.loss = "gain") or recurrent losses (gain.loss = "loss") are assessed.
reformat.annot	A logical value that indicates whether annot.file needs to be reformatted (default = FALSE). See the "note" section of makeCytoband for additional information.
random.seed	An optional random seed (default = NULL).

Details

This function applies the *Detailed Look* version of DiNAMIC's cyclic shift procedure to assess the statistical significance of recurrent DNA copy number aberrations. Either recurrent gains (gain.loss = "gain") or recurrent losses (gain.loss = "loss") are assessed using a null distribution based on num.perms cyclic shifts of x. Iterative calls to DiNAMIC's *peeling* procedure (implemented here in the peeling function) allow users to assess the statistical significance of num.iters distinct gains (losses). As noted in Bioinformatics (2011) 27(5) 678 - 685, the Detailed Look procedure recalculates the null distribution after each iteration of the peeling procedure. While this approach is more computationally intensive, simulations suggest that it provides more power to detect recurrent gains (losses).

Value

A matrix with num.iters rows. The entries of each row correspond to the marker that is being assessed. More specifically, the entries are (1) the chromosome number, (2) the marker position (in base pairs), (3) additional marker information present in marker.data, (4) the marker number, and (5) the p-value obtained from the null distribution, (6) the endpoints of the peak interval (in base pairs), as described in Bioinformatics (2011) 27(5) 678 - 685.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Examples

```
data(wilms.data)
data(wilms.markers)
data(annot.file)
detailedLook(wilms.data, wilms.markers, annot.file, 100, 3)
#Produces the following output:
#" 1"       "155656176" "R:A-MEXP-192:RP11-393K10" "196"  "0.01"
#"12"       " 38270107" "R:A-MEXP-192:RP11-519E12" "2294" "0.01"
#" 8"       "  4554176" "R:A-MEXP-192:RP11-337D8"  "1659" "0.01"
```

`findNull` *Find DiNAMIC's Null Distribution*

Description

This function is used internally by DiNAMIC's `detailedLook` and `quickLook` functions. It uses the cyclic shift procedure to create an empirical distribution that provides an approximation to the distribution of `max(colSums(x))` or `min(colSums(x))` under the null hypothesis that no underlying CNAs are present. The empirical distribution is based on `num.perms` cyclic shifts of `x`.

Usage

```
findNull(x, num.perms, random.seed = NULL)
```

Arguments

`x`	An n by m numeric matrix containing DNA copy number data from n subjects at m markers.
`num.perms`	A positive integer that represents the number of cyclic shifts used to create the empirical distribution.
`random.seed`	An optional random seed (default = NULL).

Details

The cyclic shift procedure is detailed in Bioinformatics (2011) 27(5) 678 - 685. Briefly, cyclic shift is a permutation procedure for DNA copy number data that largely preserves the underlying correlation of the markers. This function uses `num.perms` cyclic shifts of the copy number matrix `x` to create an approximate null distribution for `max(colSums(x))` or `min(colSums(x))`. The statistical significance of the observed value of `max(colSums(x))` or `min(colSums(x))` is assessed by the functions `quickLook` and `detailedLook`.

Value

A numerical vector of length num.perms.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Examples

```
random.seed = 12345
set.seed(random.seed)
x = matrix(rnorm(50), 5, 10)
num.perms = 10
example.null = findNull(x, num.perms, random.seed)
#round(example.null, 2)
#Returns 5.50 4.93 5.84 5.01 4.11 4.54 3.72 4.13 4.12 6.59
```

---

makeCytoband *Find the Chromosome Arm for Each Marker*

Description

This function is used internally by DiNAMIC's `peeling` function. It finds the chromosome arm (p or q) for each marker in the matrix `marker.data`.

Usage

```
makeCytoband(marker.data, annot.file, reformat.annot = FALSE)
```

Arguments

`marker.data`	A two-column numeric matrix of marker position data for markers in the autosomes. Column 1 contains the chromosome number for each marker, and column 2 contains the position (in base pairs) for each markers. This is a submatrix of the marker position matrix used by `quickLook` and `detailedLook`.
`annot.file`	A dataframe containing cytoband annotation for the autosomes. Each row corresponds to a distinct cytoband, and column 1 contains the chromosome number, column 2 contains the start position (in base pairs), column 3 contains the end position (in base pairs), and column 4 contains the cytoband name (e.g. p21.3). Additional columns may be present, but they are not used.
`reformat.annot`	A logical value that indicates whether `annot.file` needs to be reformatted. See "Note" for additional information.

Details

DiNAMIC's peeling procedure is detailed in Bioinformatics (2011) 27(5) 678 - 685, and it is performed by the `peeling` function. By construction, the peeling procedure only affects markers in a given chromosome arm. This function is used internally by the `peeling` function to restrict the peeling procedure to the chromosome arm containing the marker that corresponds to `max(colSums(x))`.

Value

A character vector of length m, where m is the number of markers.

Note

A four-column cytoband annotation file called `annot.file` is included in the package. However, users who wish to use other cytoband annotation files can download five-column annotation files from the UCSC Genome Browser. For example, the file cytoBand.txt.gz for the hg19 build can be found at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/. The entries in the first column of cytoBand.txt do not have the correct form, and this file also contains cytoband annotation data for the X and Y chromosomes. Thus users should change `reformat.annot` to TRUE when using these files.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Examples

```
data(wilms.markers)
data(annot.file)
wilms.pq = makeCytoband(wilms.markers, annot.file)
#A character vector of length 3288, and each entry is either
#"p" or "q", depending on the chromosome arm of the given marker.
table(wilms.pq)
#Produces the following output:
#wilms.pq

#1147 2141
```

peeling *Apply the Peeling Procedure at a Given Marker*

Description

This function is used internally by DiNAMIC's `detailedLook` and `quickLook` functions. Briefly, `detailedLook` and `quickLook` assess the statistial significance of the most aberrant gain (loss). Once this is done, the `peeling` function produces a new matrix of copy number data in which the original aberrant gain (loss) has been nullified. This allows users to assess the statistical significance of subsequent gains (losses) conditional on having found and removed previous gains (losses).

Usage

```
peeling(x, marker.data, cytoband, k)
```

Arguments

`x`	An n by m numeric matrix containing DNA copy number data from n subjects at m markers.
`marker.data`	A two-column numeric matrix of marker position data for markers in the autosomes. Column 1 contains the chromosome number for each marker, and column 2 contains the position (in base pairs) for each markers. This is a submatrix of the marker position matrix used by `quickLook` and `detailedLook`.
`cytoband`	A character vector of length m that contains the chromosome arm (p or q) for each marker. This is produced by the `makeCytoband` function.
`k`	A positive integer between 1 and m that represents the most aberrant marker.

Details

The peeling procedure is detailed in Algorithm 2 of Bioinformatics (2011) 27(5) 678 - 685, but here we provide a brief overview. By construction, marker `k` represents the most aberrant gain (loss). The peeling procedure rescales all copy number values in `x` that contribute to making marker `k` aberrant, so that after applying the peeling procedure marker `k` is "null." By construction, the rescaling procedure is restricted to entries in `x` that correspond to markers in the same chromosome arm as `k`. This allows users to assess the statistical significance of multiple gains (losses) throughout the genome.

Value

A list containing two components: (1) the n by m matrix produced by applying the peeling algorithm to the matrix `x` at marker `k`, and (2) the peak interval around marker `k`, as described in Bioinformatics (2011) 27(5) 678 - 685.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

`quickLook` *Assessing the Significance of Recurrent DNA Copy Number Aberrations*

Description

This function applies the "Quick Look" version of DiNAMIC's cyclic shift procedure to assess the statistical significance of recurrent DNA copy number aberrations. The statistical significance of recurrent gains (`gain.loss` = "gain") or recurrent losses (`gain.loss` = "loss") is assessed using an empirical null distribution produced by `num.perms` cyclic shifts of `x`.

Usage

```
quickLook(x, marker.data, annot.file, num.perms, num.iters, gain.loss = "gain",
  reformat.annot = FALSE, random.seed = NULL)
```

Arguments

`x`	An n by m numeric matrix containing DNA copy number data from n subjects at m markers.
`marker.data`	A dataframe containing marker position data for markers in the autosomes. Column 1 contains the chromosome number for each marker, and column 2 contains the position (in base pairs) for each markers. Additional columns, if present, represent information about the markers (e.g. probe names).
`annot.file`	A cytoband annotation dataframe. Each row corresponds to a distinct cytoband, and column 1 contains the chromosome number, column 2 contains the start position (in base pairs), column 3 contains the end position (in base pairs), and column 4 contains the cytoband name (e.g. p21.3). Additional columns may be present, but they are not used.
`num.perms`	A positive integer that represents the number of cyclic shifts used to create the empirical distribution.
`num.iters`	A positive integer that represents the number of distinct gain (loss) loci that will be assessed. See "Details" for more information.
`gain.loss`	A character string that indicates whether recurrent gains (`gain.loss = "gain"`) or recurrent losses (`gain.loss = "loss"`) are assessed.
`reformat.annot`	A logical value that indicates whether annot.file needs to be reformatted (default = FALSE). See the "Note" section of `makeCytoband` for additional information.
`random.seed`	An optional random seed (default = NULL).

Details

This function applies the "Quick Look" version of DiNAMIC's cyclic shift procedure to assess the statistical significance of recurrent DNA copy number aberrations. Either recurrent gains (`gain.loss = "gain"`) or recurrent losses (`gain.loss = "loss"`) are assessed using a null distribution based on num.perms cyclic shifts of `x`. Iterative calls to DiNAMIC's peeling procedure (implemented here in the `peeling` function) allow users to assess the statistical significance of num.iters distinct gains (losses). As noted in Bioinformatics (2011) 27(5) 678 - 685, the "Quick Look" procedure calculates the null distribution once, and the same distribution is used to assess the statistical significance of the most aberrant gain or loss after each iteration of the peeling procedure. This approach is less computationally intensive than "Detailed Look" because the null distribution is only computed once, but simulations suggest that it provides less power to detect recurrent gains (losses). The resulting p-values are corrected for multiple comparisons because the null distribution is based on computing `max(colSums(x))` or `min(colSums(x))`.

Value

A matrix with `num.iters` rows. The entries of each row correspond to the marker that is being assessed. More specifically, the entries are (1) the chromosome number, (2) the marker position (in base pairs), (3) additional marker information present in marker.data, (4) the marker number, and (5) the p-value obtained from the null distribution, (6) the endpoints of the peak interval (in base pairs), as described in Bioinformatics (2011) 27(5) 678 - 685.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Examples

```
data(wilms.data)
data(wilms.markers)
data(annot.file)
quickLook(wilms.data, wilms.markers, annot.file, 100, 3)
#Produces the following output:
#" 1"       "155656176" "R:A-MEXP-192:RP11-393K10" "196"  "0.01"
#"12"       " 38270107" "R:A-MEXP-192:RP11-519E12" "2294" "0.01"
#" 8"       "  4554176" "R:A-MEXP-192:RP11-337D8"  "1659" "0.01"
```

recodeBinary *Recode Binary Vectors*

Description

This function is called internally by DiNAMIC's peeling function, and by construction the kth entry of binary.vec is 1, where k is described below. If length(binary.vec) = m, then the function produces a binary vector of length m that contains a single contiguous string of 1's, namely the string that contains the 1 in the kth position of binary.vec.

Usage

```
recodeBinary(binary.vec, k)
```

Arguments

binary.vec	A binary vector of length m (>= 1) whose kth entry is 1.
k	A positive integer.

Value

A binary vector of length m that contains a single contiguous string of 1's, namely the string that contains the 1 in the kth position of binary.vec.

Author(s)

Vonn Walter, Andrew B. Nobel, Fred A. Wright

Maintainer: <vwalter@email.unc.edu> Vonn Walter

References

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

Examples

```
test = c(1, 0, 0, 1, 1, 0, 0, 1, 0)
recodeBinary(test, 5)
#Returns (0, 0, 0, 1, 1, 0, 0, 0, 0)
```

wilms.data *DNA Copy Number Data from Natrajan et al. (2006)*

Description

Natrajan et al. (J. Pathology (2006) 210: 49 - 58) used array comparative genomic hybridization to obtain genome-wide DNA copy number data from 97 Wilms' tumor samples at 3288 markers. This matrix contains the DNA copy number data after applying the bias-correction procedure outlined in Bioinformatics (2011) 27(5) 678 - 685. Each row corresponds to DNA copy number from one subject at 3288 markers, while each column contains DNA copy number data for 97 subjects at one marker.

Usage

```
data(wilms.data)
```

Format

A 97 by 3288 numeric matrix containing DNA copy number data, as described above.

Source

http://www.ebi.ac.uk/arrayexpress/ accession number E-TABM-10.

References

Natrajan, R., Williams, R.D., Hing, S.N., et al., Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse, J. Pathology (2006) 210: 49 - 58.

Walter, V., Nobel, A.B., and Wright, F.A., DiNAMIC: a method to identify recurrent DNA copy number aberrations in tumors, Bioinformatics (2011) 27(5) 678 - 685.

wilms.markers *aCGH Marker Data from Natrajan et al. (2006)*

Description

Natrajan et al. (J. Pathology (2006) 210: 49 - 58) used array comparative genomic hybridization to obtain genome-wide DNA copy number data from 97 Wilms' tumor samples at 3288 markers. This data frame contains the marker information for the arrays. Each row corresponds to a marker, and column 1 lists the chromosome number, column 2 is the marker position (in base pairs), and column 3 is the marker name.

Usage

```
data(wilms.markers)
```

Format

A data frame with 3288 observations on the following 3 variables.

`Chromosome` The chromosome for the given marker

`Position` The position (in bp) for the given marker

`Name` The name of the marker (e.g. R:A-MEXP-192:RP11-465B22)

Source

http://www.ebi.ac.uk/arrayexpress/ accession number E-TABM-10.

References

Natrajan, R., Williams, R.D., Hing, S.N., et al., Array CGH profiling of favourable histology Wilms tumours reveals novel gains and losses associated with relapse, J. Pathology (2006) 210: 49 - 58.

Examples

```
data(wilms.markers)
wilms.markers[1:10,]
#Produces the following output:
#Chromosome Position                      Name
#1           1  1036185  R:A-MEXP-192:RP11-465B22
#2           1  2078912 R:A-MEXP-192:RP11-82D16_1
#3           1  3588274 R:A-MEXP-192:RP11-62M23_2
#4           1  4366573 R:A-MEXP-192:RP11-111O5_1
#5           1  5877817    R:A-MEXP-192:RP11-49J3
#6           1  6062011  R:A-MEXP-192:RP11-426J21
#7           1  6293700   R:A-MEXP-192:RP11-51B04
#8           1  6896255  R:A-MEXP-192:RP11-402E10
#9           1  7041726 R:A-MEXP-192:RP11-60J11_1
#10          1  7653234  R:A-MEXP-192:RP11-338N10
```

Index