Package 'VegaMC'

December 5, 2025

Type Package

Title VegaMC: A Package Implementing a Variational Piecewise Smooth Model for Identification of Driver Chromosomal Imbalances in Cancer

Version 3.49.0

Date 2015-03-03

Author S. Morganella and M. Ceccarelli

Maintainer Sandro Morganella <morganellaalx@gmail.com>

Description This package enables the detection of driver chromosomal imbalances including loss of heterozygosity (LOH) from array comparative genomic hybridization (aCGH) data. VegaMC performs a joint segmentation of a dataset and uses a statistical framework to distinguish between driver and passenger mutation. VegaMC has been implemented so that it can be immediately integrated with the output produced by PennCNV tool. In addition, VegaMC produces in output two web pages that allows a rapid navigation between both the detected regions and the altered genes. In the web page that summarizes the altered genes, the link to the respective Ensembl gene web page is reported.

License GPL-2

Depends R (>= 2.10.0), biomaRt, Biobase

Imports methods

LazyLoad yes

biocViews aCGH, CopyNumberVariation

git_url https://git.bioconductor.org/packages/VegaMC

git_branch devel

git_last_commit ad29267

git_last_commit_date 2025-10-29

Repository Bioconductor 3.23

Date/Publication 2025-12-04

Contents

`VegaMC-package` *VegaMC: A Package Implementing a Variational Piecewise Smooth Model for Identification of Driver Chromosomal Imbalances in Cancer.*

Description

VegaMC enables the detection of driver chromosomal imbalances (deletions, amplifications and loss of heterozygosities (LOHs)) from array comparative genomic hybridization (aCGH) data. VegaMC performs a joint segmentation of aCGH data. Segmented regions are then used into a statistical framework to distinguish between driver and passenger mutations. In this way, significant imbalances can be detected by the associated p-value. VegaMC has been implemented to be easily integrated with the output produced by PennCNV. VegaMC produces in output two web pages allowing a rapid navigation between both detected regions and altered genes. In the web page summarizing the altered genes, the user finds the link to the respective Ensembl gene web page.

Details

Package:	VegaMC
Type:	Package
Version:	3.9.3
License:	GPL-2
LazyLoad:	yes

Examples

```
## Copy the example dataset in current folder
file.copy(system.file("example/breast_Affy500K.txt", package="VegaMC"),
                      ".")

## Analyse data and save results in sorted.txt file
results <- vegaMC("breast_Affy500K.txt", "results", html=FALSE,
                  getGenes=FALSE)
```

sortData *This function sorts a dataset file by the genomic position of the probes.*

Description

This function sorts a dataset file by the genomic position of the probes. This function makes very easy the integration of VegaMC with the output of PennCNV tool.

Usage

```
sortData(dataset, output_file_name = "")
```

Arguments

dataset	Dataset file.
output_file_name	Name of the file in which sorted data are stored.

Value

This function returns the input matrix ordered by the genomic position of the probes.

Note

This function allows to sort a dataset by the genomic position. The input file must have the chromosome and the position in column two and three respectively. This format follows the standard output of PennCNV. An example of file can be found in inst/example folder.

Author(s)

Sandro Morganella

References

Morganella S., and Ceccarelli M. VegaMC: a R/bioconductor package for fast downstream analysis of large array comparative genomic hybridization datasets. Bioinformatics, 28(19):2512-4 (2012).

Examples

```
## Copy the example dataset in current folder
file.copy(system.file("example/breast_Affy500K.txt", package="VegaMC"),
                      ".")


## Sort data and save results in sorted.txt file
sortData("breast_Affy500K.txt", "sorted.txt");
```

vegaMC-methods *Class aggregator of VegaMC.*

Description

VegaMC enables the detection of driver chromosomal imbalances (deletions, amplifications and loss of heterozygosities (LOHs)) from array comparative genomic hybridization (aCGH) data. VegaMC performs a joint segmentation of aCGH data. Segmented regions are then used into a statistical framework to distinguish between driver and passenger mutations. In this way, significant imbalances can be detected by the associated p-value. VegaMC has been implemented to be easily integrated with the output produced by PennCNV and with the Genoset eSet Objects. VegaMC produces in output two web pages allowing a rapid navigation between both detected regions and altered genes. In the web page summarizing the altered genes, the user finds the link to the respective Ensembl gene web page.

Usage

```
vegaMC(dataset, output_file_name="output", beta=0.5,
min_region_bp_size=1000, correction=FALSE,
loss_threshold=-0.2, gain_threshold=0.2,
baf=TRUE, loh_threshold=0.75, loh_frequency=0.8, bs=1000,
pval_threshold=0.05, html=TRUE, getGenes=TRUE,
mart_database="ensembl", ensembl_dataset="hsapiens_gene_ensembl")
```

Arguments

dataset	Dataset file following the PennCNV format: The first three columns describe the name, the chromosome and the position respectively. The other columns of the matrix report the LRR and the BAF (if available) of each sample. Note that observations must be ordered by the respective genomic position.
output_file_name	(Default codeoutput) File name used to save the results.
beta	(Default 0.5) This parameter is used to compute the stop condition. It is used to calculate the maximum jump allowed in scale parameter updating. If beta=0 then the resulting segmentation will be composed of a region for each probe (all regions will contain just a probe). In contrast, if beta is very large, then the segmentation will contain just a region for each chromosome.
min_region_bp_size	(Default 1000) VegaMC deletes from the list the regions shorter then this size (in bp).
correction	(Default FALSE) If this parameter is TRUE multiple testing corrections is performed.
loss_threshold	(Default -0.2) Values used to mark a region as a deletion (loss). If the wighted mean of a region is lower than this threshold, then the region is marked as a deletion (loss).

`gain_threshold` (Default 0.2) Values used to mark a region as an amplification (gain). If the wighted mean of a region is greater than this threshold, then the region is marked as an amplification (gain).

`baf` (Default `TRUE`) This parameter specifies if the dataset contains BAF measurements (default `TRUE`. If BAF is available, then VegaMC is able to compute LOH imbalances.

`loh_threshold` (Default 0.75) Threshold used to distinguish between homozygous and heterozygous genotypes. If the BAF is greater than `loh_threshold` or lower then (1-`loh_threshold`) then the respective probe is considered to be homozygous.

`loh_frequency` (Default 0.8) Minimum fraction of homozygous probes needed for marking a region as LOHs. Regions with a fraction of homozygous probes greater than this threshold are marked as LOH.

`bs` (Default 1000) Number of permutation bootstraps performed to compute the null distribution.

`pval_threshold` (Default 0.05) Significance level used to reject the null hypothesis. If the p-value of an aberration (loss, gain, LOH) is not greater than this threshold, then the region is considered to be significant and, consequently, it is considered a driver mutation.

`html` (Default `TRUE`) If this value is `TRUE`, then in output will be produced an html file called `output_file_name`.html in which a summary of all detected regions is reported.

`getGenes` (Default `TRUE`) If this value is `TRUE`, then in output will be produced an html file called `output_file_name`Genes.html in which the list of all genes overlapping the significant regions is reported.

`mart_database` (Default `ensembl`) BioMart database name you want to connect to. Possible database names can be retrieved with the function listMarts of biomaRt package.

`ensembl_dataset` (Default `hsapiens_gene_ensembl`) BioMart dataset used to get information from Ensembl BioMart database. In order to obtain the list of all available dataset use the command `listDatasets` of biomaRt package.

Value

After the execution of this function, a matrix containing all information on the detected regions is returned. This object is a matrix having a row for each detected regions described by the following columns:

`Chromosome` The chromosome in which the region is located.

`bp Start` The position in which the region starts (in bp).

`bp End` The position in which the region ends (in bp).

`Region Size` The size of the regions (in bp).

`Mean` The weighted mean of the region computed on all samples.

`Loss p-value` The p-value associated to the probability to have a driver deletion.

`Gain p-value` The p-value associated to the probability to have a driver amplification.

`LOH p-value` The p-value associated to the probability to have a driver LOH.

`% Loss` The percentage of samples showing a deletion for this region.

`% Gain` The percentage of samples showing an amplification for this region.

`% LOH` The percentage of samples showing a LOH for this region.

`Probe Size` The number of probes composing the region.

`Loss Mean` Mean of LRR computed only on the samples that show a loss.

`Gain Mean` Mean of LRR computed only on the samples that show a gain.

`LOH Mean` Mean of LRR computed only on the samples that show a LOH.

`Focal-score Loss`
Focal Score associated to deletion.

`Focal-score Gain`
Focal Score associated to amplification.

`Focal-score LOH`
Focal Score associated to LOH.

This matrix is automatically saved in the current work directory as a tab delimited file. For default the name used to asave the file is 'output'.

Methods

`signature(dataset = "character")` This method allows to run VegaMC on a data file in PennCNV format.

`signature(dataset = "GenoSet")` This method allows to run VegaMC on a GenoSet object of `genoset` package.

Author(s)

Sandro Morganella

References

Morganella S., and Ceccarelli M. VegaMC: a R/bioconductor package for fast downstream analysis of large array comparative genomic hybridization datasets. Bioinformatics, 28(19):2512-4 (2012).

Examples

```
## Run VegaMC
file.copy(system.file("example/breast_Affy500K.txt", package="VegaMC"), ".")
results <- vegaMC("breast_Affy500K.txt", "results", html=FALSE,
                  getGenes=FALSE)
```

Index