

# Package ‘StarBioTrek’

September 28, 2023

**Type** Package

**Title** StarBioTrek

**Version** 1.27.0

**Date** 04-16-2019

**Author** Claudia Cava,  
Isabella Castiglioni

**Maintainer** Claudia Cava <claudia.cava@ibfm.cnr.it>

**Depends** R (>= 3.3)

**Imports** SpidermiR, graphite, AnnotationDbi, e1071, ROCR, MLmetrics,  
grDevices, igraph, reshape2, ggplot2

## Description

This tool StarBioTrek presents some methodologies to measure pathway activity and cross-talk among pathways integrating also the information of network data.

**License** GPL (>= 3)

**biocViews** GeneRegulation, Network, Pathways, KEGG

**Suggests** BiocStyle, knitr, rmarkdown, testthat, devtools, roxygen2,  
qgraph, png, grid

**VignetteBuilder** knitr

**LazyData** true

**URL** <https://github.com/claudiacava/StarBioTrek>

**BugReports** <https://github.com/claudiacava/StarBioTrek/issues>

**RoxygenNote** 6.1.1

**git\_url** <https://git.bioconductor.org/packages/StarBioTrek>

**git\_branch** devel

**git\_last\_commit** 3404841

**git\_last\_commit\_date** 2023-04-25

**Date/Publication** 2023-09-27

**R topics documented:**

average	2
circleplot	3
ConvertedIDgenes	4
Data_CANCER_normUQ_fil	4
dsscorectrlk	5
euclidstctrlk	5
GetData	6
getNETdata	6
GetPathData	7
GetPathNet	8
GE_matrix	8
GE_matrix_mean	9
GOChord	9
IPPI	10
listpathnet	11
netw	11
netw_IPPI	11
norm	12
path	12
pathnet	12
pathway	13
pathway_matrix	13
path_KEGG	13
plotcrosstalk	14
score_euc_dista	14
SelectedSample	15
select_class	15
StarBioTrek	16
stdv	16
svm_classification	17
tumo	17
<b>Index</b>	<b>18</b>

---

average	<i>For TCGA data get human pathway data and creates a matrix with the average of genes for each pathway.</i>
---------	--------------------------------------------------------------------------------------------------------------

---

**Description**

average creates a matrix with a summarized value for each pathway

**Usage**

average(pathwayexpsubset)

**Arguments**

pathwayexpsubset  
list of pathway data

**Value**

a matrix value for each pathway

**Examples**

```
list_path_gene<-GE_matrix(DataMatrix=Data_CANCER_normUQ_fil,genes.by.pathway=pathway[1:50])  
score_mean<-average(pathwayexpsubset=list_path_gene)
```

---

circleplot                      *Preparation for circle plot*

---

**Description**

circleplot function takes as input data derived by the function plotcrosstalk and pLOt a circle plot.

**Usage**

```
circleplot(preplot, scoregene)
```

**Arguments**

preplot                      a list as obtained from the function plotcrosstalk  
scoregene                    a score for each gene with values included between -10 e +10

**Value**

a list with correlation matrix and gene set for each gene

**Examples**

```
formatplot<-plotcrosstalk(pathway_plot=pathway[1:6],gs_expre=tumo)  
score<-runif(length(formatplot[[2]]), min=-10, max=+10)  
circleplot(preplot=formatplot,scoregene=score)
```

---

ConvertedIDgenes      *Get interacting genes inside pathways.*

---

**Description**

GetPathNet creates a list of genes inside the pathways.

**Usage**

```
ConvertedIDgenes(path_ALL)
```

**Arguments**

path\_ALL      variable. The user can select the variable as obtained by GetData function

**Value**

a list of pathways

**Examples**

```
pathway<-ConvertedIDgenes(path_ALL=path[1:3])
```

---

Data\_CANCER\_normUQ\_fil  
*pathway data list*

---

**Description**

pathway data list

**Format**

A dataframe with gene expression profiles

---

dsscorecrtlk	<i>For TCGA data get human pathway data and creates a measure of discriminating score among pathways</i>
--------------	----------------------------------------------------------------------------------------------------------

---

**Description**

dsscorecrtlk creates a matrix with discriminating score for pathways

**Usage**

```
dsscorecrtlk(dataFilt, pathway_exp)
```

**Arguments**

dataFilt	TCGA matrix
pathway_exp	a list of pathway data

**Value**

a matrix value for each pathway

**Examples**

```
cross_talk_st_dv<-dsscorecrtlk(dataFilt=tumo[,1:2],pathway_exp=pathway[1:5])
```

---

euclidcrtlk	<i>For TCGA data get human pathway data and creates a measure of cross-talk among pathways</i>
-------------	------------------------------------------------------------------------------------------------

---

**Description**

euclidcrtlk creates a matrix with euclidean distance for pairwise pathways

**Usage**

```
euclidcrtlk(dataFilt, pathway_exp)
```

**Arguments**

dataFilt	TCGA matrix
pathway_exp	list of pathway data

**Value**

a matrix value for each pathway

**Examples**

```
score_euc_dista_t<-eucdistcrtlk(dataFilt=tumo[,1:2],pathway_exp=pathway[1:5])
```

---

 GetData

*Get general information inside pathways.*


---

**Description**

GetData creates a list with genes inside the pathways.

**Usage**

```
GetData(species, pathwaydb)
```

**Arguments**

species            variable. The user can select the species of interest from SELECT\_path\_species(path\_spec)  
 pathwaydb        variable. The user can select the pathway database of interest from SELECT\_path\_graphite(path\_spec)

**Value**

a list of pathways

**Examples**

```
## Not run:
species="hsapiens"
pathwaydb="pharmgkb"
path<-GetData(species,pathwaydb)
## End(Not run)
```

---

 getNETdata

*Get network data from GeneMania.*


---

**Description**

getNETdata creates a data frame with network data. Network category can be filtered among: physical interactions, co-localization, genetic interactions and shared protein domain.

**Usage**

```
getNETdata(network, organismID = NULL)
```

**Arguments**

network            variable. The user can use the following parameters based on the network types to be used. PHint for Physical\_interactions, COloc for Co-localization, GENint for Genetic\_interactions and SHpd for Shared\_protein\_domains

organismID        organism==NULL default value is homo sapiens.

**Value**

list with gene-gene (or protein-protein interactions)

**Examples**

```
## Not run:  
organismID="Saccharomyces_cerevisiae"  
netw<-getNETdata(network="SHpd",organismID)  
## End(Not run)
```

---

GetPathData

*Get genes inside pathways.*

---

**Description**

GetPathData creates a list of genes inside the pathways.

**Usage**

```
GetPathData(path_ALL)
```

**Arguments**

path\_ALL            variable. The user can select the variable as obtained by GetData function

**Value**

a list of pathways

**Examples**

```
pathway_ALL_GENE<-GetPathData(path_ALL=path[1:3])
```

---

GetPathNet	<i>Get interacting genes inside pathways.</i>
------------	-----------------------------------------------

---

**Description**

GetPathNet creates a list of genes inside the pathways.

**Usage**

```
GetPathNet(path_ALL)
```

**Arguments**

path\_ALL            variable. The user can select the variable as obtained by GetData function

**Value**

a list of pathways

**Examples**

```
pathway_net<-GetPathNet(path_ALL=path[1:3])
```

---

GE_matrix	<i>Get human KEGG pathway data and a gene expression matrix in order to obtain a list with the gene expression for only pathways given in input .</i>
-----------	-------------------------------------------------------------------------------------------------------------------------------------------------------

---

**Description**

GE\_matrix creates a list of gene expression for pathways given by the user.

**Usage**

```
GE_matrix(DataMatrix, genes.by.pathway)
```

**Arguments**

DataMatrix        gene expression matrix (eg.TCGA data)  
genes.by.pathway        a list of pathway data as provided by GetData and ConvertedID\_genes

**Value**

a list for each pathway ( gene expression level belong to that pathway)



**Examples**

```
list_path_gene<-GE_matrix(DataMatrix=tumo[,1:2],genes.by.pathway=pathway[1:5])
```

---

GE_matrix_mean	<i>Get human KEGG pathway data and a gene expression matrix in order to obtain a matrix with the mean gene expression for only pathways given in input .</i>
----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------

---

**Description**

GE\_matrix creates a matrix of mean gene expression levels for pathways given by the user.

**Usage**

```
GE_matrix_mean(DataMatrix, genes.by.pathway)
```

**Arguments**

DataMatrix      gene expression matrix (eg.TCGA data)  
genes.by.pathway      list of pathway data as provided by getKEGGdata

**Value**

a matrix for each pathway (mean gene expression level belong to that pathway)

**Examples**

```
list_path_plot<-GE_matrix_mean(DataMatrix=tumo[,1:2],genes.by.pathway=pathway[1:5])
```

---

GOChord	<i>Displays the relationship between genes and terms.</i>
---------	-----------------------------------------------------------

---

**Description**

The GOChord function generates a circularly composited overview of selected/specific genes and their assigned processes or terms. More generally, it joins genes and processes via ribbons in an intersection-like graph.

**Usage**

```
GOChord(data, title, space, gene.order, gene.size, gene.space, nlf = 1,  
lfc.col, lfc.min, lfc.max, ribbon.col, border.size, process.label, limit)
```

**Arguments**

<code>data</code>	The matrix represents the binary relation (1= is related to, 0= is not related to) between a set of genes (rows) and processes (columns); a column for the logFC of the genes is optional
<code>title</code>	The title (on top) of the plot
<code>space</code>	The space between the chord segments of the plot
<code>gene.order</code>	A character vector defining the order of the displayed gene labels
<code>gene.size</code>	The size of the gene labels
<code>gene.space</code>	The space between the gene labels and the segment of the logFC
<code>nlogfc</code>	Defines the number of logFC columns (default=1)
<code>logfc.col</code>	The fill color for the logFC specified in the following form: <code>c(color for low values, color for the mid point, color for the high values)</code>
<code>logfc.min</code>	Specifies the minimum value of the logFC scale (default = -3)
<code>logfc.max</code>	Specifies the maximum value of the logFC scale (default = 3)
<code>ribbon.col</code>	The background color of the ribbons
<code>border.size</code>	Defines the size of the ribbon borders
<code>process.label</code>	The size of the legend entries
<code>limit</code>	A vector with two cutoff values (default= <code>c(0,0)</code> ).

---

 IPPI

*Multilayer analysis Cava et al. BMC Genomics 2017*


---

**Description**

IPPI function takes as input pathway and network data in order to select genes with central role in that pathway. Please see Cava et al. 2017 BMC Genomics

**Usage**

```
IPPI(pathax, netwa)
```

**Arguments**

<code>pathax</code>	pathway matrix Please see example path for format
<code>netwa</code>	a dataframe Please see example path for format netw

**Value**

a list with driver genes for each pathway

**Examples**

```
## Not run:
DRIVER_SP<-IPPI(pathax=pathway_matrix[,1:3],netwa=netw_IPPI[1:50000,])
## End(Not run)
```

---

listpathnet	<i>Get human KEGG pathway data and the output of list_path_net define the common genes.</i>
-------------	---------------------------------------------------------------------------------------------

---

**Description**

listpathnet creates a list of interacting genes for each human pathway.

**Usage**

```
listpathnet(lista_net, pathway_exp)
```

**Arguments**

lista_net	output of path_net
pathway_exp	pathway data as provided by getKEGGdata

**Value**

a list of genes for each pathway (interacting genes belong to that pathway)

**Examples**

```
lista_network<-pathnet(genes.by.pathway=pathway[1:5],data=netw)
list_path<-listpathnet(lista_net=lista_network,pathway=pathway[1:5])
```

---

netw	<i>network data</i>
------	---------------------

---

**Description**

network data

**Format**

A data frame with rows and variables

---

netw_IPPI	<i>network data for IPPI fuction</i>
-----------	--------------------------------------

---

**Description**

network data for IPPI fuction

**Format**

A list

---

norm	<i>TCGA data with normal samples</i>
------	--------------------------------------

---

**Description**

TCGA data with normal samples

**Format**

A data frame with rows and variables

---

path	<i>pathway data list</i>
------	--------------------------

---

**Description**

pathway data list

**Format**

A list of dataframe

---

pathnet	<i>Get human KEGG pathway data and creates a network data.</i>
---------	----------------------------------------------------------------

---

**Description**

pathnet creates a list of network data for each human pathway. The network data will be generated when interacting genes belong to that pathway.

**Usage**

```
pathnet(genes.by.pathway, data)
```

**Arguments**

genes.by.pathway	a list of pathway data as provided by <code>ConvertedIDgenes</code>
data	a list of network data as provided by <code>getNETdata</code>

**Value**

a list of network data for each pathway (interacting genes belong to that pathway)

**Examples**

```
lista_net<-pathnet(genes.by.pathway=pathway[1:5],data=netw)
```

---

pathway	<i>pathway data</i>
---------	---------------------

---

**Description**

pathway data

**Format**

A data frame with rows and variables

---

pathway_matrix	<i>network data</i>
----------------	---------------------

---

**Description**

network data

**Format**

A data frame with rows and variables

---

path_KEGG	<i>All pathways data from KEGG</i>
-----------	------------------------------------

---

**Description**

All pathways data from KEGG

**Format**

A list of pathways with the involved genes

---

plotcrosstalk	<i>Preparation for plotting cross-talk</i>
---------------	--------------------------------------------

---

**Description**

plot\_cross\_talk function takes as input pathway data and prepares the data to visualize (e.g. ggplot2, qqgraph, igraph)

**Usage**

```
plotcrosstalk(pathway_plot, gs_expre)
```

**Arguments**

pathway_plot	pathway
gs_expre	a gene expression matrix

**Value**

a list with correlation matrix and gene set for each gene

**Examples**

```
formatplot<-plotcrosstalk(pathway_plot=pathway[1:6],gs_expre=tumo)
```

---

score_euc_dista	<i>Score Matrix of pairwise pathway using euclidean distance</i>
-----------------	------------------------------------------------------------------

---

**Description**

Score Matrix of pairwise pathway using euclidean distance

**Format**

A data frame with rows and variables

---

SelectedSample	<i>Select the class of TCGA data</i>
----------------	--------------------------------------

---

**Description**

select two labels from ID barcode

**Usage**

```
SelectedSample(Dataset, typesample)
```

**Arguments**

Dataset	gene expression matrix
typesample	the labels of the samples (e.g. tumor,normal)

**Value**

a gene expression matrix of the samples with specified label

**Examples**

```
tumo<-SelectedSample(Dataset=Data_CANCER_normUQ_fil,typesample="tumour")[,2]
```

---

select_class	<i>Select the class of TCGA data</i>
--------------	--------------------------------------

---

**Description**

select best performance

**Usage**

```
select_class(performance_matrix, cutoff)
```

**Arguments**

performance_matrix	list of AUC value
cutoff	cut-off for AUC value

**Value**

a gene expression matrix with only pairwise pathway with a particular cut-off

---

StarBioTrek	<i>Download data</i>
-------------	----------------------

---

### Description

StarBioTrek allows you to Download data of samples from StarBioTrek

### Details

The functions you're likely to need from **StarBioTrek** is `path_star` Otherwise refer to the vignettes to see how to format the documentation.

---

stdv	<i>For TCGA data get human pathway data and creates a measure of standard deviations among pathways</i>
------	---------------------------------------------------------------------------------------------------------

---

### Description

stdv creates a matrix with standard deviation for pathways

### Usage

```
stdv(gslist)
```

### Arguments

gslist            pathway data

### Value

a matrix value for each pathway

### Examples

```
list_path_gene<-GE_matrix(DataMatrix=tumo[,1:2],genes.by.pathway=pathway[1:5])
score_stdev<-stdv(gslist=list_path_gene)
```



---

svm\_classification      *SVM classification for each feature*

---

**Description**

svm class creates a list with AUC, Accuracy, Sensitivity, Specificity values

**Usage**

```
svm_classification(TCGA_matrix, tumour, normal, nfs)
```

**Arguments**

TCGA_matrix	gene expression matrix where the first two columns represent the interacting pathways.
tumour	barcode samples for a class
normal	barcode samples for another class
nfs	nfs split data into a training and test set
Target	label for the classes

**Value**

a list with AUC value for pairwise pathway

**Examples**

```
## Not run:
nf <- 60
res_class<-svm_classification(TCGA_matrix=score_euc_dista[1:30,],nfs=nf,
normal=colnames(norm[,1:10]),tumour=colnames(tumo[,1:10]))
## End(Not run)
```

---

tumo                      *TCGA data with tumour samples*

---

**Description**

TCGA data with tumour samples

**Format**

A data frame with rows and variables

# Index

## \* internal

- Data\_CANCER\_normUQ\_fil, 4
  - netw, 11
  - netw\_IPPI, 11
  - norm, 12
  - path, 12
  - path\_KEGG, 13
  - pathway, 13
  - pathway\_matrix, 13
  - score\_euc\_dista, 14
  - tumo, 17
- average, 2
- circleplot, 3
- ConvertedIDgenes, 4
- Data\_CANCER\_normUQ\_fil, 4
- dsscorectrlk, 5
- euclidctrlk, 5
- GE\_matrix, 8
- GE\_matrix\_mean, 9
- GetData, 6
- getNETdata, 6
- GetPathData, 7
- GetPathNet, 8
- GOChord, 9
- IPPI, 10
- listpathnet, 11
- netw, 11
- netw\_IPPI, 11
- norm, 12
- path, 12
- path\_KEGG, 13
- pathnet, 12
- pathway, 13
- pathway\_matrix, 13
- plotcrosstalk, 14
- score\_euc\_dista, 14
- select\_class, 15
- SelectedSample, 15
- StarBioTrek, 16
- StarBioTrek-package (StarBioTrek), 16
- stdv, 16
- svm\_classification, 17
- tumo, 17