

# Package ‘enrichvs’

July 2, 2014

**Version** 0.0.5

**Date** 2011-06-29

**Title** Enrichment assessment of virtual screening approaches

**Author** Hiroaki YABUUCHI

**Maintainer** Hiroaki YABUUCHI <yabu.pharm@gmail.com>

**Depends** R (>= 0.99)

**Description** These programs are used for calculating enrichment factors, drawing enrichment curves to evaluate virtual screening approaches.

**License** BSD

**LazyLoad** yes

**Repository** CRAN

**Date/Publication** 2011-06-29 10:03:53

**NeedsCompilation** no

## R topics documented:

enrichvs-package . . . . .	2
auac . . . . .	2
auc . . . . .	3
bedroc . . . . .	4
dud_egfr . . . . .	5
enrichment_factor . . . . .	6
matplot_enrichment_curve . . . . .	7
plot_enrichment_curve . . . . .	7
plot_enrichment_curve_xlog . . . . .	8
rie . . . . .	9

<b>Index</b>	<b>11</b>
--------------	-----------

---

enrichvs-package

*Enrichment assessment of virtual screening approaches*

---

### Description

These programs are used for calculating enrichment factors, drawing enrichment curves to evaluate virtual screening approaches.

### Details

Package: enrichvs  
Type: Package  
Version: 0.0.4  
Date: 2011-05-03  
License: BSD  
LazyLoad: yes

In drug discovery processes, efficient virtual screening (VS) is essential to select candidate compounds. This package provides an easy-to-use interface to evaluate VS approaches. Especially, plotting an enrichment curve helps us to find the best approach.

### Author(s)

Hiroaki YABUUCHI

### References

Yabuuchi H. et al. Analysis of multiple compound-protein interactions reveals novel bioactive molecules. *Mol. Syst. Biol.* (2011) 7, 472.

---

auac

*Function to calculate the Area Under the Accumulation Curve (AUAC)*

---

### Description

Function to calculate the Area Under the Accumulation Curve (AUAC).

### Usage

```
auac(x, y, decreasing=TRUE, top=1.0)
```

**Arguments**

x	a vector for scores
y	a vector for labels (0:inactive, 1:active)
decreasing	TRUE if the compounds are ranked by decreasing score
top	threshold ratio of the false positives (when ROC analysis is performed on a top list)

**Value**

AUAC, in the range from 0 to 1.

**Author(s)**

Hiroaki YABUUCHI

**References**

Truchon et al. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* (2007) 47, 488-508.

**Examples**

```
x <- rnorm(1000) # random scores for 1000 compounds
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"
auc(x, y)

data(dud_egfr)
auc(dud_egfr$energy, dud_egfr$label, decreasing=FALSE)
```

---

auc

*Function to calculate the Area Under the ROC Curve (AUC)*

---

**Description**

Function to calculate the Area Under the ROC Curve (AUC)

**Usage**

```
auc(x, y, decreasing=TRUE, top=1.0)
```

**Arguments**

x	a vector for scores
y	a vector for labels
decreasing	TRUE if the compounds are ranked by decreasing score
top	threshold ratio of the false positives (when ROC analysis is performed on a top list)

**Value**

AUC, in the range from 0 to 1.

**Author(s)**

Hiroaki YABUUCHI

**References**

Truchon et al. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* (2007) 47, 488-508.

**Examples**

```
x <- rnorm(1000) # random scores for 1000 compounds
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"
auc(x, y)

data(dud_egfr)
auc(dud_egfr$energy, dud_egfr$label, decreasing=FALSE)

# AUC up to the first 10% of the false positives (ROC10%)
auc(dud_egfr$energy, dud_egfr$label, decreasing=FALSE, top=0.1)
```

---

bedroc

*Function to calculate the Boltzmann-Enhanced Discrimination of ROC (BEDROC)*

---

**Description**

Function to calculate the Boltzmann-Enhanced Discrimination of ROC (BEDROC)

**Usage**

```
bedroc(x, y, decreasing=TRUE, alpha=20.0)
```

**Arguments**

x	a vector for scores
y	a vector for labels
alpha	coefficient alpha
decreasing	TRUE if the compounds are ranked by decreasing score

**Value**

BEDROC, in the range from 0 to 1.

**Author(s)**

Hiroaki YABUUCHI

**References**

Truchon et al. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* (2007) 47, 488-508.

**Examples**

```
x <- rnorm(1000) # random scores for 1000 compounds
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"
bedroc(x, y)

data(dud_egfr)
bedroc(dud_egfr$energy, dud_egfr$label, decreasing=FALSE)
```

---

dud\_egfr

*Scores and labels for EGFR ligands/docoys from DUD dataset*

---

**Description**

Ligands and decoys for Epidermal Growth Factor Receptor (EGFR) were scored by DOCK programs (Huang et al., 2006).

**Usage**

```
data(dud_egfr)
```

**Format**

A list with 88888 observations on the following 2 variables.

```
energy1 -99.7 -98.3 -97.4 ...
label 0 0 0 ...
```

**Details**

energy : a numeric vector for scores (binding free energies) label : a numeric vector for labels (1:ligand, 0:decoy)

**Source**

A Directory of Useful Decoys (DUD): <http://dud.docking.org/r2/>

**References**

Huang N. et al. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* 49, 6789-6801, (2006).

**Examples**

```
data(dud_egfr)
enrichment_factor(dud_egfr$energy, dud_egfr$label)
```

---

enrichment\_factor      *Function to calculate the enrichment factor (EF)*

---

**Description**

Function to calculate the enrichment factor (EF)

**Usage**

```
enrichment_factor(x, y, top=0.05, decreasing=TRUE)
```

**Arguments**

x	a vector for scores
y	a vector for labels
top	thresholded ratio of top-ranked compound
decreasing	TRUE if the compounds are ranked by decreasing score

**Value**

EF, in the range from 0 to +Inf.

**Author(s)**

Hiroaki YABUUCHI

**Examples**

```
x <- rnorm(1000) # random scores for 1000 compounds
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"
enrichment_factor(x, y, top=0.1)

data(dud_egfr)
enrichment_factor(dud_egfr$energy, dud_egfr$label, decreasing=FALSE)
```

---

`matplot_enrichment_curve`*Function to plot multiple enrichment curves for virtual screening*

---

**Description**

Function to plot multiple enrichment curves for virtual screening

**Usage**

```
matplot_enrichment_curve(x, y)
```

**Arguments**

`x` score matrix for compounds (row) by several approaches (column)  
`y` activity label for compounds (0:inactive, 1:active)

**Value**

call plots

**Author(s)**

Hiroaki YABUUCHI

**Examples**

```
x <- matrix(rnorm(5000), ncol=5) # random 5 scores for 1000 compounds  
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"  
matplot_enrichment_curve(x, y)
```

---

`plot_enrichment_curve` *Function to plot a enrichment curve for virtual screening*

---

**Description**

Function to plot a enrichment curve for virtual screening

**Usage**

```
plot_enrichment_curve(x, y, decreasing=TRUE, npoint=100,  
colbarwidth=1, colorize=FALSE, add=FALSE, col="black", lwd=2)
```

**Arguments**

x	a vector for scores
y	a vector for labels (0:inactive, 1:active)
decreasing	TRUE if the compounds are ranked by decreasing score
npoint	the number of points to plot the curve
colbarwidth	the width of color bar (only if colorize==TRUE)
colorize	TRUE if the plot line is colored by the corresponding scores
add	TRUE if the plot is add to the previous plot
col	color of the plot
lwd	line width of the plot

**Value**

call a plot

**Author(s)**

Hiroaki YABUUCHI

**Examples**

```
data(dud_egfr)
x <- rnorm( 88888 ) # virtual scores for 1000 compounds
y <- c(rep(1,88), rep(0,88800)) # activity labels for "x"
plot_enrichment_curve(x, y, col="blue")

plot_enrichment_curve(dud_egfr$energy, dud_egfr$label,
add=TRUE, decreasing=FALSE, col="red") # Add a plot
legend("bottomright", legend = c("ideal", "random", "score1", "score2"),
lty=c(2,3,1,1), col=c("black", "grey", "blue", "red"), bty="n")
```

---

plot\_enrichment\_curve\_xlog

*Function to plot a enrichment curve with a logarithmic scale on the x axis*

---

**Description**

Function to plot a enrichment curve with a logarithmic scale on the x axis

**Usage**

```
plot_enrichment_curve_xlog(x, y, decreasing=TRUE, npoint=100,
colbarwidth=1, colorize=FALSE, add=FALSE, col="black", lwd=2)
```



**Arguments**

x	a vector for scores
y	a vector for labels (0:inactive, 1:active)
decreasing	TRUE if the compounds are ranked by decreasing score
npoint	the number of points to plot the curve
colbarwidth	the width of color bar (only if colorize==TRUE)
colorize	TRUE if the plot line is colored by the corresponding scores
add	TRUE if the plot is add to the previous plot
col	color of the plot
lwd	line width of the plot

**Value**

call a plot

**Author(s)**

Hiroaki YABUUCHI

**Examples**

```
data(dud_egfr)
x <- rnorm( 88888 ) # virtual scores for 1000 compounds
y <- c(rep(1,88), rep(0,88800)) # activity labels for "x"
plot_enrichment_curve_xlog(x, y, col="blue")

plot_enrichment_curve_xlog(dud_egfr$energy, dud_egfr$label,
add=TRUE, decreasing=FALSE, col="red") # Add a plot
legend("topleft", legend = c("ideal", "random", "score1", "score2"),
lty=c(2,3,1,1), col=c("black", "grey", "blue", "red"), bty="n")
```

---

rie

*Function to calculate the Robust Initial Enhancement (RIE)*

---

**Description**

Function to calculate the Robust Initial Enhancement (RIE)

**Usage**

```
rie(x, y, decreasing=TRUE, alpha=20.0)
```

**Arguments**

x	a vector for scores
y	a vector for labels (0:inactive, 1:active)
alpha	coefficient alpha
decreasing	TRUE if the compounds are ranked by decreasing score

**Value**

RIE, in the range from 0 to +Inf.

**Note**

Though RIE was originally calculated through a Monte Carlo simulation, this function adopts a simplified formulation proposed by Truchon et al. to reduce computational cost.

**Author(s)**

Hiroaki YABUUCHI

**References**

Sheridan RP et al. Protocols for bridging the peptide to nonpeptide gap in topological similarity searches. *J. Chem. Inf. Comput. Sci.* (2001) 41, 1395-1406.

Truchon et al. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* (2007) 47, 488-508.

**Examples**

```
x <- rnorm(1000) # random scores for 1000 compounds
y <- c(rep(1,50), rep(0,950)) # activity labels for "x"
rie(x, y)

data(dud_egfr)
rie(dud_egfr$energy, dud_egfr$label, decreasing=FALSE)
```

# Index

\*Topic **datasets**

dud\_egfr, [5](#)

\*Topic **package**

enrichvs-package, [2](#)

auac, [2](#)

auc, [3](#)

bedroc, [4](#)

dud\_egfr, [5](#)

enrichment\_factor, [6](#)

enrichvs (enrichvs-package), [2](#)

enrichvs-package, [2](#)

matplotlib\_enrichment\_curve, [7](#)

plot\_enrichment\_curve, [7](#)

plot\_enrichment\_curve\_xlog, [8](#)

rie, [9](#)