

# Package ‘iBATCGH’

October 13, 2022

**Version** 1.3.1

**Date** 2015-07-06

**Title** Integrative Bayesian Analysis of Transcriptomic and CGH Data

**Author** Alberto Cassese

**Maintainer** Alberto Cassese <alberto.cassese@maastrichtuniversity.nl>

**Depends** R (>= 2.10)

**Imports** Rcpp (>= 0.11.2), msm

**LinkingTo** Rcpp, RcppArmadillo

**Description** Bayesian integrative models of gene expression and comparative genomic hybridization data. The package provides inference on copy number variations and their association with gene expression.

**License** GPL-2

**URL** <http://www.r-project.org>

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2020-10-14 16:34:40 UTC

## R topics documented:

iBATCGH-package . . . . .	2
Center . . . . .	3
iBAT . . . . .	4
iBATProbit . . . . .	6
Inference . . . . .	8
InferenceA . . . . .	10
InferenceMu . . . . .	11
InferenceR . . . . .	12
InferenceSd . . . . .	13
InferenceXi . . . . .	14
InitMu . . . . .	15
InitXi . . . . .	16

NCI_60 . . . . .	17
RListToVector . . . . .	17
Scenario1 . . . . .	18
Scenario2 . . . . .	19
TCGA_lung . . . . .	20
Tran . . . . .	21

<b>Index</b>	<b>22</b>
--------------	-----------

---

iBATCGH-package	<i>Integrative Bayesian Analysis of Transcriptomic and CGH data</i>
-----------------	---

---

## Description

Bayesian integrative models of gene expression and comparative genomic hybridization data. The package provides inference on copy number variations and their association with gene expression

## Details

Package: iBATCGH  
 Type: Package  
 Version: 1.3  
 Date: 2015-07-06  
 License: GNU

The package takes as inputs gene expression, Comparative Genomic Hybridization (CGH) data, and the physical distance between CGH probes. It returns posterior probabilities of inclusion for each potential association gene expression - CGH, and inference on the CGH latent states. Two different models have been considered, see references for details. Available functions are classified in four classes: Preprocessing, Main, Postprocessing, Simulated Data.

## Author(s)

Alberto Cassese, Marina Vannucci, Michele Guindani, Mahlet G. Tadesse.

Maintainer: Alberto Cassese <alberto.cassese@maastrichtuniversity.nl>

## References

- Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.
- Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.
- Guha S, Li Y, Neuberg D. Bayesian hidden Markov modelling of array cgh data. *JASA*. 2008;103(482):485-497.

**Examples**

```
## Not run:
data(NCI_60)

Y <- NCI_60$Affy
X <- NCI_60$aCGH
distance <- NCI_60$distance
disfix <- 146274826
xi <- InitXi(X)
tran <- Tran(xi)
mu <- InitMu()
d=0.2587288

Y <- Center(Y)

res <- iBAT(Y=Y,X=X,distance=distance,disfix=disfix,xi=xi,tran=tran,mu=mu,d=d)

summRes <- Inference(res,G=dim(Y)[[2]],M=dim(X)[[2]],niter=niter,burnin=bi,threshold=0.5)

## End(Not run)
```

---

Center

*Preprocessing - Center gene expression data*

---

**Description**

This function takes as argument a matrix of gene expression measurements  $Y$ , and returns the matrix obtained after centering each column with respect to its mean.

**Usage**

```
Center(Y)
```

**Arguments**

$Y$                       The matrix of data to be centered.

**Details**

This function center each column of the gene expression matrix.

**Value**

The matrix of data obtained after centering each column with respect to its mean.

**Author(s)**

Alberto Cassese

**Examples**

```
data(NCI_60)
Y <- NCI_60$Affy
Y<-Center(Y)
```

---

iBAT

---

*Main - Mixture selection prior*


---

**Description**

Perform MCMC iterations of the model, as described in the reference.

**Usage**

```
iBAT(Y, X, distance, disfix, intercept=1, xi,
R=-1, tran, mu, sigma=((rgamma(4,1,1))^(-0.5)),
cmu=1/1000000, c=10, delta=3, d, e=0.001, f=0.999,
alpha=20, deltak=c(-1,0,0.58,1), tauk=c(1,1,1,2),
upp_bounds=c(-0.1, 0.1, 0.73, Inf),
low_bounds=c(-Inf, -0.1, 0.1, 0.73),
alpha_IG=c(1,1,1,1), beta_IG=c(1,1,1,1),
low_IG=c(0.41,0.41,0.41,1), a=c(1,1,1,1),
niter=500000, burnin=200000, Cout=1000,
phi=0.5, pR=0.4, selectioncgh=-1, pXI=0.6, indep=0)
```

**Arguments**

Y	Matrix of gene expression data
X	Matrix of CGH data
distance	Vector of distance between CGH probes
disfix	Length of the chromosome under investigation
intercept	If set to one an intercept is included in the regression model
xi	Initialized matrix of latent states
R	Initialized association matrix in a vector form. Default set to -1, that automatically creates a vector with all the positions set to zero
tran	Initialized transition matrix
mu	Initialized state specific mean vector
sigma	Initialized state specific standard deviation vector
cmu	Parameter that controls the variance of the prior on the intercept
c	Parameter that determines the shrinkage in the model
delta	Parameter of the Inverse-Gamma prior on the error variance
d	Parameter of the Inverse-Gamma prior on the error variance
e	Parameter of the Beta prior on the inclusion probability

f	Parameter of the Beta prior on the inclusion probability
alpha	Parameter that regulates the strength of the independent part of the mixture
deltak	Vector of mean of the prior on the state specific mean
tauk	Vector of sd of the prior on the state specific mean
upp_bounds	Vector of upper bounds of the prior on the state specific mean
low_bounds	Vector of lower bounds of the prior on the state specific mean
alpha_IG	Vector of parameters of the prior on the state specific standard deviation
beta_IG	Vector of parameters of the prior on the state specific standard deviation
low_IG	Truncation of the prior on the state specific standard deviation
a	Vector of parameters of the prior on the transition matrix
niter	Number of Monte Carlo Markov Chain iterations
burnin	Burn-in
Cout	Print the number of iterations ran every Cout iterations
phi	Probability of an A/D step
pR	Parameter of the distribution used to select the rows to be updated at every MCMC iteration
selectioncgh	Number of samples not in neutral state in order to consider a CGH as a potential candidate for association with gene expression. Default set to -1 that automatically set it to 10% of the samples
pXI	Parameter of the distribution used to select the rows to be updated at every MCMC iteration
indep	If set to an integer different from zero, run the analysis with an independent prior, see reference.

### Value

The output consists of an R list composed by  $4 \cdot \text{niter} + 3$  elements objects, where niter is the number of MCMC iterations. The first niter objects of the list are vectors, each containing the positions of the association matrix set to one, at the corresponding MCMC iteration. Each of the following niter objects of the list are the transition matrices at the corresponding MCMC iteration, while the third and the fourth set of niter objects are the vectors of state specific mean and state specific variance, respectively. The last three objects of the list consist of three matrices counting the number of times the corresponding latent state has been set to 1,3 and 4, respectively.

### Author(s)

Alberto Cassese

### References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

**Examples**

```
## Not run:
data(NCI_60)

Y <- NCI_60$Affy
X <- NCI_60$aCGH
distance <- NCI_60$distance
disfix <- 146274826
xi <- InitXi(X)
tran <- Tran(xi)
mu <- InitMu()
d=0.2587288

Y <- Center(Y)

res <- iBAT(Y=Y,X=X,distance=distance,disfix=disfix,xi=xi,tran=tran,mu=mu,d=d)

summRes <- Inference(res,G=dim(Y)[[2]],M=dim(X)[[2]],niter=niter,burnin=bi,threshold=0.5)

## End(Not run)
```

---

iBATProbit

*Main - Probit selection prior*


---

**Description**

Perform MCMC iterations of the model described in the reference.

**Usage**

```
iBATProbit(Y, X, distance, disfix, intercept=1, xi,
R=-1, tran, mu, sigma=((rgamma(4,1,1))^(-0.5)),
cmu=1/1000000, c=10, delta=3, d, alpha0=2.32,
alpha1=1, deltak=c(-1,0,0.58,1), tauk=c(1,1,1,2),
upp_bounds=c(-0.1, 0.1, 0.73, Inf),
low_bounds=c(-Inf, -0.1, 0.1, 0.73),
alpha_IG=c(1,1,1,1), beta_IG=c(1,1,1,1),
low_IG=c(0.41,0.41,0.41,1), a=c(1,1,1,1),
niter=500000, burnin=200000, Cout=1000,
phi=0.5, pR=0.4, selectioncgh=-1, pXI=0.6)
```

**Arguments**

Y	Matrix of gene expression data
X	Matrix of CGH data
distance	Vector of distance between CGH probes
disfix	Length of the chromosome under investigation

intercept	If set to one an intercept is included in the regression model
xi	Initialized matrix of latent states
R	Initialized association matrix in a vector form. Default set to -1, that automatically creates a vector with all the positions set to zero
tran	Initialized transition matrix
mu	Initialized state specific mean vector
sigma	Initialized state specific standard deviation vector
cmu	Parameter that controls the variance of the prior on the intercept
c	Parameter that determines the shrinkage in the model
delta	Parameter of the Inverse-Gamma prior on the error variance
d	Parameter of the Inverse-Gamma prior on the error variance
alpha0	Baseline intercept of the selection prior
alpha1	Parameter that regulates the strength of the spatially informed dependence
deltak	Vector of mean of the prior on the state specific mean
tauk	Vector of sd of the prior on the state specific mean
upp_bounds	Vector of upper bounds of the prior on the state specific mean
low_bounds	Vector of lower bounds of the prior on the state specific mean
alpha_IG	Parameter of the prior on the state specific standard deviation
beta_IG	Parameter of the on the state specific standard deviation
low_IG	Truncation of the prior on the state specific standard deviation
a	Vector of parameters of the prior on the transition matrix
niter	Number of Monte Carlo Markov Chain iteration
burnin	Burn-in
Cout	Print the number of iterations ran every Cout iterations
phi	Probability of an A/D step
pR	Parameter of the distribution used to select the rows to be updated at every MCMC iteration
selectioncgh	Number of samples not in neutral state in order to consider a CGH as a potential candidate for association with gene expression. Default set to -1 that automatically set it to 10% of the samples
pXI	Parameter of the distribution used to select the rows to be updated at every MCMC iteration

### Value

The output consists of an R list composed by  $4 \cdot \text{niter} + 3$  objects, where niter is the number of MCMC iterations. The first niter objects of the list are vectors, each containing the positions of the association matrix set to one, at the corresponding MCMC iteration. Each of the following niter objects of the list are the transition matrices at the corresponding MCMC iteration, while the third and the fourth set of niter objects are the vectors of state specific mean and state specific variance, respectively. The last three objects of the list consist of three matrices counting the number of times the corresponding latent state has been set to 1, 3 and 4, respectively.

**Author(s)**

Alberto Cassese

**References**

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**Examples**

```
## Not run:
data(TCGA_lung)

Y <- TCGA_lung$Affy
X <- TCGA_lung$aCGH
distance <- TCGA_lung$distance
disfix <- 199446827
xi <- InitXi(X)
tran <- Tran(xi)
mu <- InitMu()
d=0.2587288

Y <- Center(Y)

res <- iBATProbit(Y=Y,X=X,distance=distance,disfix=disfix,xi=xi,tran=tran,mu=mu,d=d)

summRes <- Inference(res,G=dim(Y)[[2]],M=dim(X)[[2]],niter=niter,burnin=bi,threshold=0.5)

## End(Not run)
```

---

Inference

*Postprocessing - Posterior Inference*

---

**Description**

Performs posterior inference on the output of the main function

**Usage**

```
Inference(listComplete, G, M, niter, burnin, threshold = 0.5)
```

**Arguments**

listComplete	Output of the main function
G	Number of gene expression probes
M	Number of CGH probes
niter	Number of Monte Carlo Markov Chain iterations
burnin	Burn-in
threshold	Threshold on the posterior probabilities of inclusion of the association matrix



**Details**

Wrapper function which calls `InferenceR`, `InferenceXi`, `InferenceA`, `InferenceMu`, `InferenceSd` on the output of the main function.

**Value**

A list made by the following items

R	Binary matrix of estimated association
Xi	Matrix of estimated copy number states
A	Estimated transition matrix
Mu	Estimated vector of state specific means
Sd	Estimated vector of state specific standard deviations

**Author(s)**

Alberto Cassese

**References**

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**See Also**

See Also [InferenceR](#), [InferenceXi](#)

**Examples**

```
## Not run:
data(NCI_60)

Y <- NCI_60$Affy
X <- NCI_60$aCGH
distance <- NCI_60$distance
disfix <- 146274826
xi <- InitXi(X)
tran <- Tran(xi)
mu <- InitMu()
d=0.2587288

Y <- Center(Y)

res <- iBAT(Y=Y,X=X,distance=distance,disfix=disfix,xi=xi,tran=tran,mu=mu,d=d)

summRes <- Inference(res,G=dim(Y)[[2]],M=dim(X)[[2]],niter=niter,burnin=bi,threshold=0.5)
```

```
## End(Not run)
```

---

InferenceA

*Postprocessing - Inference on the transition matrix*

---

## Description

This function returns a matrix obtained as the elementwise mean of the association matrices, output of the Monte Carlo Markov Chain iterations, after removing burn-in.

## Usage

```
InferenceA(listA, niter, burnin)
```

## Arguments

listA	Second niter objects of the output of the main function
niter	Number of Monte Carlo Markov Chain iterations
burnin	Burn-in

## Value

Estimated transition matrix.

## Author(s)

Alberto Cassese

## References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

## See Also

See Also as [Inference](#)

## Examples

```
##See Inference
```

---

InferenceMu

*Postprocessing - Inference on the vector of state specific mean*

---

### Description

This function returns a vector obtained as the elementwise mean of the vectors of state specific mean, output of the Monte Carlo Markov Chain iterations, after removing burn-in.

### Usage

```
InferenceMu(listMu, niter, burnin)
```

### Arguments

listMu	Third niter objects of the output of the main function
niter	Number of Monte Carlo Markov Chain iteration
burnin	Burn-in

### Value

Estimated vector of state specific mean.

### Author(s)

Alberto Cassese

### References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

### See Also

See Also as [Inference](#)

### Examples

```
##See Inference
```

---

InferenceR

*Postprocessing - Inference on the association matrix*

---

### Description

This function performs posterior inference on the association matrix, returning the matrix of posterior probabilities of inclusion for each association gene expression - CGH. It also returns the binary matrix of significant links that exceed a threshold given as argument.

### Usage

```
InferenceR(listR, G, M, niter, burnin, threshold = 0.5)
```

### Arguments

listR	First niter objects of the output of the main function
G	Number of gene expression probes
M	Number of aCGH probes
niter	Number of Monte Carlo Markov Chains iterations
burnin	Burn-in
threshold	Threshold on the posterior probability of inclusion

### Value

A list made of the following items

FreqMat	Matrix of posterior probabilities of inclusion
final	Binary association matrix

### Author(s)

Alberto Cassese

### References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

### See Also

See Also as [Inference](#)

### Examples

```
##See Inference
```

---

InferenceSd

*Postprocessing - Inference on the vector of state specific sd*

---

### Description

This function returns a vector obtained as the elementwise mean of the vectors of state specific standard deviation, output of the Monte Carlo Markov Chain iterations, after removing burn-in.

### Usage

```
InferenceSd(listSd, niter, burnin)
```

### Arguments

listSd	Fourth niter objects of the output of the main function
niter	Number of Monte Carlo Markov Chain iterations
burnin	Burn-in

### Value

Estimated vector of state specific sd.

### Author(s)

Alberto Cassese

### References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

### See Also

See Also as [Inference](#)

### Examples

```
##See Inference
```

---

InferenceXi

*Postprocessing - Inference on the latent states*

---

### Description

This function returns the modal latent states.

### Usage

```
InferenceXi(listXi, niter, burnin)
```

### Arguments

listXi	Last three objects of the output of the main function
niter	Number of Monte Carlo Markov Chain iterations
burnin	Burn-in

### Details

Must use the same burn-in as in the main function.

### Value

Matrix of modal latent states, i.e. estimated Copy Number Variants. A four class classification is considered:

1. Loss
2. Neutral
3. Gain
4. Amplification

### Author(s)

Alberto Cassese

### References

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

### See Also

See Also as [Inference](#)

**Examples**

```
##See Inference
```

---

InitMu	<i>Preprocessing - Initialize state specific mean vector</i>
--------	--

---

**Description**

Initializes the state specific mean vector, by sampling each element independently from its prior, i.e. truncated normal distribution..

**Usage**

```
InitMu(deltak = c(-1, 0, 0.58, 1), tauk = c(1, 1, 1, 2),  
low_bounds = c(-Inf, -0.1, 0.1, 0.73), upp_bounds = c(-0.1, 0.1, 0.73, Inf))
```

**Arguments**

deltak	Vector of means of the truncated Normal distributions
tauk	Vector of standard deviations of the truncated Normal distributions
low_bounds	Vector of lower bounds of the truncated Normal distributions
upp_bounds	Vector of upper bounds of the truncated Normal distributions

**Value**

A vector of state specific mean, that could be used as input of the main function.

**Author(s)**

Alberto Cassese

**References**

Guha S, Li Y, Neuberg D. Bayesian hidden Markov modelling of array cgh data. JASA. 2008;103(482):485-497.

**Examples**

```
mu <- InitMu()
```

---

`InitXi`*Preprocessing - Initialize matrix of latent states*

---

**Description**

This function takes a matrix of CGH data as the only argument and returns a crude estimate of the corresponding latent copy number states.

**Usage**

```
InitXi(X, bounds = c(-0.5, 0.29, 0.79))
```

**Arguments**

<code>X</code>	Matrix of aCGH data
<code>bounds</code>	Vector of threshold used to estimate the latent states

**Details**

Given as argument a vector of threshold bounds the function simply applies the thresholding to the data and groups them into four subsets. Each subset is associated to a specific latent state.

**Value**

Return a matrix of estimated latent states, that could be used as input of the main function.

**Author(s)**

Alberto Cassese

**References**

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**See Also**

See Also [Tran](#)

**Examples**

```
data(NCI_60)
X <- NCI_60$aCGH
xi <- InitXi(X)
```



---

`NCI_60`*NCI-60 cancer cell lines data*

---

**Description**

Processed and filtered NCI-60 cancer cell lines data, as described in the reference.

**Usage**

```
data(NCI_60)
```

**Format**

The format is a list of 3 objects

\$ aCGH

\$ Affy

\$ distance

**Source**

Full matrix of data downloaded from [discover.nci.nih.gov/cellminer](http://discover.nci.nih.gov/cellminer).

**References**

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

**Examples**

```
data <- data(NCI_60)
```

---

`RListToVector`*Internal function*

---

**Description**

Internal function

**Usage**

```
RListToVector(xList, G, T)
```

**Arguments**

xList	List of included associations (C format)
G	Number of gene expression probes
T	Number of CGH probes

**Details**

This is an internal function.

**Value**

Returns a vector whose elements are the number of times, across the MCMC iterations, the corresponding position of the transition matrix has been set to one.

**Author(s)**

Alberto Cassese

---

Scenario1	<i>Simulated data - Scenario 1</i>
-----------	------------------------------------

---

**Description**

Simulates the data as described in the reference provided below (Scenario 1).

**Usage**

```
Scenario1(sigmak = 0.1)
```

**Arguments**

sigmak	Standard deviation of the error term
--------	--------------------------------------

**Value**

Return a list made of the following items

Y	Matrix of simulated gene expression
X	Matrix of simulated CGH
$X_i$	True matrix of hidden states
A	Empirical transition matrix
mu	True vector of state specific mean
Sd	True vector of state specific sd
coeff	True matrix of association coefficients between gene expression and CGH probes
distance	Vector of distance between CGH probes
disfix	Length of the chromosome

**Author(s)**

Alberto Cassese

**References**

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**Examples**

```
data <- Scenario1(sigmak = 0.1)
```

---

Scenario2

*Simulated data - Scenario 2*

---

**Description**

Simulates the data as described in the reference provided below (Scenario 2).

**Usage**

```
Scenario2(sigmak = 0.1)
```

**Arguments**

sigmak                      Standard deviation of the error term

**Value**

Return a list made of the following items

Y	Matrix of simulated gene expression
X	Matrix of simulated CGH
Xi	True matrix of hidden states
A	Empirical transition matrix
mu	True vector of state specific mean
Sd	True vector of state specific sd
coeff	True matrix of association coefficients between gene expression and CGH probes
distance	Vector of distance between CGH probes
disfix	Length of the chromosome

**Author(s)**

Alberto Cassese

**References**

Cassese A, Guindani M, Tadesse M, Falciani F, Vannucci M. A hierarchical Bayesian model for inference of copy number variants and their association to gene expression. *Annals of Applied Statistics*, 8(1), 148-175.

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**Examples**

```
data <- Scenario2(sigmak = 0.1)
```

---

TCGA\_lung

*TCGA lung squamous cell carcinoma data*

---

**Description**

Processed and filtered TCGA lung squamous cell carcinoma data, as described in the reference.

**Usage**

```
data(TCGA_lung)
```

**Format**

The format is a list of 3 objects

\$ aCGH

\$ Affy

\$ distance

**Source**

Full matrix of data obtained from <https://tcga-data.nci.nih.gov/tcga/>.

**References**

Cassese A, Guindani M, Vannucci M. A Bayesian integrative model for genetical genomics with spatially informed variable selection. *Cancer Informatics*.

**Examples**

```
data <- data(TCGA_lung)
```

---

Tran

*Preprocessing - Tran*

---

### Description

This function computes the transition matrix corresponding to a specific matrix of latent states.

### Usage

```
Tran(xi)
```

### Arguments

`xi` Matrix of estimated latent states

### Details

The matrix of latent states must consider only 4 states:

1. Loss
2. Neutral
3. Gain
4. Amplification

### Value

A 4 by 4 matrix with transition probabilities derived from the input matrix of latent states.

### Author(s)

Alberto Cassese

### References

Guha S, Li Y, Neuberg D. Bayesian hidden Markov modelling of array cgh data. *JASA*. 2008;103(482):485-497.

### Examples

```
data(NCI_60)

X <- NCI_60$aCGH
xi <- InitXi(X)
tran <- Tran(xi)
```

# Index

## \* Main

iBAT, 4  
iBATProbit, 6

## \* Package

iBATCGH-package, 2

## \* Postprocessing

Inference, 8  
InferenceA, 10  
InferenceMu, 11  
InferenceR, 12  
InferenceSd, 13  
InferenceXi, 14

## \* Preprocessing

Center, 3  
Tran, 21

## \* Prerocessing

InitMu, 15

## \* Simulated data

Scenario1, 18  
Scenario2, 19

## \* datasets

NCI\_60, 17  
TCGA\_lung, 20

RListToVector, 17

Scenario1, 18  
Scenario2, 19

TCGA\_lung, 20  
Tran, 16, 21

Center, 3

iBAT, 4  
iBATCGH (iBATCGH-package), 2  
iBATCGH-package, 2  
iBATProbit, 6  
Inference, 8, 10–14  
InferenceA, 10  
InferenceMu, 11  
InferenceR, 9, 12  
InferenceSd, 13  
InferenceXi, 9, 14  
InitMu, 15  
InitXi, 16

NCI\_60, 17