

# Package ‘NETI2’

October 8, 2019

**Type** Package

**Title** Joint reconstruction of multiple gene networks by simultaneously capturing inter-tumor and intra-tumor heterogeneity

**Version** 1.0.0

**Author** Jia-Juan Tu

**Maintainer** Jia-Juan Tu <adaline\_juan@mails.ccnu.edu.cn>

**Description** NETI2 is a method designed for jointly estimating multiple gene networks across multiple cancer subtypes. NETI2 can simultaneously characterize inter-tumor and intra-tumor heterogeneity at the network level.

**Depends** R(>= 3.4.2)

**Imports** MASS, Matrix, QUIC, igraph, mvtnorm, foreach

**Suggests** knitr,  
rmarkdown

**VignetteBuilder** knitr

**RoxygenNote** 6.1.1

**License** GPL(>= 2)

**Encoding** UTF-8

**LazyData** true

## R topics documented:

generate.data . . . . .	1
NETI2 . . . . .	3
TCGA.BRCA . . . . .	4
<b>Index</b>	<b>6</b>

---

generate.data	<i>Generate simulated data</i>
---------------	--------------------------------

---

## Description

The complete procedure for generating simulated data. For details, refer to simulation study (Section 3.1 in the main text).

**Usage**

```
generate.data(p, n, K, network.type, umin, umax)
```

**Arguments**

p	The number of genes.
n	The sample size. A positive integer or a vector with length equal to the number of cancer subtypes.
K	The number of cancer subtypes.
network.type	A character string indicating which network type is generated. "ER" (Erdos-Renyi) and "SF" (scale-free) can be used.
umin	The lower limits of the edge values.
umax	The upper limits of the edge values.

**Details**

The function is used to generate the gene expression datasets.

**Value**

x	A list (length = $K$ ) of data matrices ( $n_k \times p$ ), where $K$ is the number of cancer subtypes and $n_k$ is the sample size of k-th cancer subtype.
theta.y	A list (length = $K$ ) of precision matrices ( $p \times p$ ) from cancerous cells of different cancer subtypes, where $K$ is the number of cancer subtypes and $n_k$ is the sample size of k-th cancer subtype.
theta.z	A matrix of precision matrix ( $p \times p$ ) from non-cancerous cells shared by all cancer subtypes.
purity	A list (length = $K$ ) of the tumor purity information vectors ( $n_k \times 1$ ), where $K$ is the number of cancer subtypes and $n_k$ is the sample size of k-th cancer subtype.

**Author(s)**

Jia-Juan Tu

**References**

Jia-Juan Tu, Le Ou-Yang, Hong Yan, Xiao-Fei Zhang and Hong Qin (2019), Joint reconstruction of multiple gene networks by simultaneously capturing inter-tumor and intra-tumor heterogeneity.

**See Also**

[NETI2](#), [TCGA.BRCA](#)

**Examples**

```
# Simulation data
data.x = generate.data(p = 100, n = 100, K = 4, network.type = "ER", umin = 0.5, umax = 1)
```

NETI2

*Joint reconstruction of multiple gene NETWORKs by simultaneously capturing Inter-tumor and Intra-tumor heterogeneity***Description**

The complete procedure for reconstructing multiple gene networks using NETI2. For details, refer to Supplementary Section S3.2.

**Usage**

```
NETI2(X,purity,lambda, tau, delta)
```

**Arguments**

X	A list (length = $K$ ) of data matrices ( $n_k \times p$ ), where $K$ is the number of cancer subtypes and $n_k$ is the sample size of k-th cancer subtype.
purity	A list (length = $K$ ) of the tumor purity information vectors ( $n_k \times 1$ ), where $K$ is the number of cancer subtypes and $n_k$ is the sample size of k-th cancer subtype.
lambda	The common tuning parameter for controlling the overall degree sparsity of the estimated gene networks. For details, refer to Supplementary TableS7.
tau	The tuning parameter balances the network size between non-cancerous and cancerous networks. For details, refer to Supplementary TableS7.
delta	The tuning parameter balances the network size between subtype specific networks. For details, refer to Supplementary TableS7.

**Details**

The function is used to jointly reconstruction of multiple gene networks by simultaneously capturing inter-tumor and intra-tumor heterogeneity. For each cancer subtype, the observed gene expression levels of tumor samples are assumed to be a mixture of expressions from non-cancerous and cancerous cells. The gene expression levels of cancerous cells are assumed to follow subtype specific multivariate normal distributions, and the gene expression levels of non-cancerous cells across all subtypes are assumed to follow the same multivariate normal distribution. The precision matrices of these multivariate normal distributions are used to build the non-cancerous and subtype specific cancerous networks. Given the observed gene expression data and tumor purity information data, we use a penalized likelihood approach to estimate the model parameters. We develop an efficient iterative procedure based on Expectation Maximization (EM) algorithm to solve the optimization problem with latent (unobserved) variables. Each iteration of the EM algorithm consists of two steps: E-step and M-step. For details, refer to Supplementary Section S3.2.

**Value**

theta.y	A list (length = $K$ ) of estimated precision matrices from cancerous cells of different cancer subtypes.
theta.z	A matrix of estimated precision matrix from non-cancerous cells shared by all cancer subtypes.
LL.temp	Log-likelihood of the data for different EM iterations.

**Author(s)**

Jia-Juan Tu

**References**

Jia-Juan Tu, Le Ou-Yang, Hong Yan, Xiao-Fei Zhang and Hong Qin (2019), Joint reconstruction of multiple gene networks by simultaneously capturing inter-tumor and intra-tumor heterogeneity

**See Also**

[generate.data](#), [TCGA.BRCA](#)

**Examples**

```
# Simulation data
data.x= generate.data(p = 100, n = 100, K = 4, network.type ="ER", umin = 0.5, umax = 1)
result = NETI2(data.x$X,data.x$purity, lambda = 0.9, tau = 0.5,delta = 0.5)

# TCGA breast cancer data
data("TCGA.BRCA")
result = NETI2(TCGA.BRCA$BRCA.data,TCGA.BRCA$BRCA.purity, lambda = 1.6, tau = 0.4,delta = 0.2)
```

---

TCGA.BRCA

---

*TCGA breast cancer data and tumor purity information data*


---

**Description**

The TCGA breast cancer gene expression data and tumor purity information data are used in our study. The gene expression data are obtained from the TCGA database and collected from mRNA expression (Agilent G450 microarray). The breast cancer is a heterogeneous disease that can be mainly divided into four molecular subtypes: Luminal A, Luminal B, Basal-like and HER2-enriched. The data only include expression measurements of genes that overlap with the breast cancer pathway collected from the Kyoto Encyclopedia of Genes and Genomes database. The tumor purity information of these samples are collected from Aran et al (2015). This data is an object of class list of length two. The first element (BRCA.data) is the gene expression data and the second element (BRCA.purity) is the tumor purity information data.

**Usage**

```
data("TCGA.BRCA")
```

**Format**

An object of class list of length 2.

**Author(s)**

Jia-Juan Tu

**Source**

- [1] The Cancer Genome Atlas Research Network (2012), Comprehensive molecular portraits of human breast tumors. *Nature*. 490 (7418), 61-70. (<http://cancergenome.nih.gov/>)
- [2] Aran, D. et al. (2015). Systematic pan-cancer analysis of tumour purity. *Nature Communications*, 6, 8971.

**References**

Jia-Juan Tu, Le Ou-Yang, Hong Yan, Xiao-Fei Zhang and Hong Qin (2019), Joint reconstruction of multiple gene networks by simultaneously capturing inter-tumor and intra-tumor heterogeneity.

**See Also**

[NETI2](#), [generate.data](#)

**Examples**

```
data("TCGA.BRCA")
```

# Index

\*Topic **datasets**

TCGA.BRCA, [4](#)

generate.data, [1](#), [4](#), [5](#)

NETI2, [2](#), [3](#), [5](#)

TCGA.BRCA, [2](#), [4](#), [4](#)