# Package 'iNOTE'

June 8, 2017

**Type** Package

**Title** Integrative Network Omnibus Total Effect Test

**Version** 1.0

**Date** 2017-06-05

**Author** Su H. Chu <su_chu@brown.edu>
Yen-Tsung Huang <ythuang@stat.sinica.edu.tw>

**Maintainer** Su H. Chu <su_chu@brown.edu>

**Description** Integrated joint analysis of multiple platform genomic data across biological gene sets or pathways using powerful variance-component based testing procedures.

**License** GPL (>= 2)

**Depends** CompQuadForm, plyr, mixtools

**NeedsCompilation** no

**Encoding** UTF-8

**Repository** CRAN

**Date/Publication** 2017-06-07 22:14:26 UTC

## R topics documented:

## iNOTE-package *Integrative Network Omnibus Total Effect Test*

### Description

Integrated joint analysis of multiple platform genomic data across biological gene sets or pathways using powerful variance-component based testing procedures.

### Author(s)

Su H. Chu <su_chu@brown.edu> Yen-Tsung Huang <ythuang@stat.sinica.edu.tw>

Maintainer: Su H. Chu <su_chu@brown.edu>

### References

Chu S.H. and Huang Y-T. (2017) Integrative genomic analysis of biological gene sets with applications in lung cancer. (Revise and Resubmit)

Huang Y-T, Lin X. (2013) Gene set analysis using variance component tests. *BMC Bioinformatics*. **14**(1):210. PMCID: PMC3776447

Huang Y-T, Vanderweele TJ, Lin X. (2014) Joint analysis of SNP and gene expression data in genetic association studies of complex diseases. *The Annals of Applied Statistics*. **8**(1):352–76.

## CPG *SRC Signaling Pathway CpG Methylation Data in TCGA*

### Description

TCGA lung cancer patient Illumina 450K CpG methylation data for genes in the SRC signaling pathway geneset identified Gautschi et al (2008).

### Usage

```
data(CPG)
```

### Format

A list object with eight indices, containing the matrix of CpG methylation levels for sites associated with each individual gene in the SRC Signaling pathway for 249 TCGA subjects.

`ID1` A named matrix with 249 subject rows and 14 CpG columns.

`ID2` A named matrix with 249 subject rows and 15 CpG columns.

`ID3` A named matrix with 249 subject rows and 14 CpG columns.

`ID4` A named matrix with 249 subject rows and 8 CpG columns.

`SERPINE1` A named matrix with 249 subject rows and 14 CpG columns.

SMAD6 A named matrix with 249 subject rows and 36 CpG columns.

SMAD7 A named matrix with 249 subject rows and 24 CpG columns.

TGFB1 A named matrix with 249 subject rows and 21 CpG columns.

## Details

Genomic data were extracted from lung cancer patients (adenocarcinoma or squamous cell carcinoma) in The Cancer Genome Atlas and adjusted for batch effects. The CPG dataset represents a subset of the CpG methylation data for patients with non-missing survival status at one year after initial diagnosis in genes identified by Gautschi et al (2008) in the Molecular Signatures Database (MsigDB; Subramanian et al (2005)) as lung-cancer associated members of the SRC signaling pathway.

## Source

**TCGA** http://cancergenome.nih.gov/

**MsigDB** http://software.broadinstitute.org/gsea/msigdb/

## References

Gautschi O, Tepper CG, Purnell PR, Izumiya Y, Evans CP, Green TP, et al. (2008) Regulation of Id1 expression by SRC: implications for targeting of the bone morphogenetic protein pathway in cancer. *Cancer Research*. **68**(7):2250–8.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. **102**(43):15545–50.

## Examples

```
data(CPG)
str(CPG)
```

---

# GE *SRC Signaling Pathway mRNA Expression Data in TCGA*

---

## Description

TCGA lung cancer patient mRNA expression data for genes in the SRC signaling pathway geneset identified Gautschi et al (2008).

## Usage

```
data(GE)
```

## Format

A named matrix with 249 subject rows and 8 mRNA expression columns (one column per gene in the SRC signaling pathway).

## Details

Genomic data were extracted from lung cancer patients (adenocarcinoma or squamous cell carcinoma) in The Cancer Genome Atlas and adjusted for batch effects. The GE dataset represents a subset of the genomic mRNA expression data for patients with non-missing survival status at one year after initial diagnosis in genes identified by Gautschi et al (2008) in the Molecular Signatures Database (MsigDB; Subramanian et al (2005)) as lung-cancer associated members of the SRC signaling pathway.

## Source

**TCGA** http://cancergenome.nih.gov/

**MsigDB** http://software.broadinstitute.org/gsea/msigdb/

## References

Gautschi O, Tepper CG, Purnell PR, Izumiya Y, Evans CP, Green TP, et al. (2008) Regulation of Id1 expression by SRC: implications for targeting of the bone morphogenetic protein pathway in cancer. *Cancer Research*. **68**(7):2250–8.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. **102**(43):15545–50.

## Examples

```
data(GE)
str(GE)
```

---

# inote *Integrative Network Omnibus Total Effect Test*

---

## Description

Omnibus variance-component based testing procedure to test for the total effect of methylation loci and mRNA expression across a network without requiring the specification of an underlying disease risk model.

## Usage

```
inote(iCPG, iGE, iY, iX, i.omniseed=NA, no.pert = 1000, imethod = "chi")
```

## Arguments

| | |
|---|---|
| iCPG | A list of J CpG matrices with dimensions nx$p_j$. |
| iGE | An nxJ matrix of gene expression values. |
| iY | An nx1 dichotomous outcome vector. |
| iX | An nxr covariate matrix. |

| | |
|---|---|
| i.omniseed | A Jx1 seed vector. |
| no.pert | No. perturbations per gene level test; defaults to 1000. |
| imethod | Omnibus testing method – 'chi' or 'uni' |

## Value

A list with the following components:

| | |
|---|---|
| p | The omnibus test p-value for the joint, integrative total effect test for the gene set. |
| p.00 | The null distribution of the omnibus test statistic p-value. |
| method | The omnibus method specified by the user. |
| gs.uni.mod | If applicable, the omnibus gene-set model selected by the iNOTE-uni method. |

## Author(s)

Su H. Chu & Yen-Tsung Huang

## References

Chu S.H. and Huang Y-T. (2017) Integrative genomic analysis of biological gene sets with applications in lung cancer. (Revise and Resubmit)

## Examples

```
data(X); data(Y); data(CPG); data(GE)
## Not run: inote(iCPG=CPG, iGE=GE, iY=Y, iX=X, no.pert=1000, imethod='chi')
```

---

# itegs *Integrative Total Effect of a Gene Set Test*

---

## Description

A variance-component based testing procedure to test for the total effect of methylation loci and mRNA expression across a network after specifying an underlying disease risk model which applies to all genes in the gene set of interest.

## Usage

```
itegs(iCPG, iGE, iY, iX, imodel = "mgc", iapprox = "pert",
  i.omniseed = NA, no.pert = 1000, gsp.emp = TRUE)
```

## Arguments

| | |
|---|---|
| iCPG | A list of J CpG matrices with dimensions nx$p_j$. |
| iGE | An nxJ matrix of gene expression values. |
| iY | An nx1 dichotomous outcome vector. |
| iX | An nxr covariate matrix. |
| imodel | The specified disease risk model for the whole gene set: 'mgc', 'mg', or 'm'. |
| iapprox | The preferred method of gene-set p-value calculation: 'pert' or 'davies' |
| i.omniseed | A Jx1 seed vector. |
| no.pert | If using 'pert', the no. perturbations per gene level test; defaults to 1000. |
| gsp.emp | If using 'pert': The method of calculating the perturbation based p-value: empirical (TRUE) or parametric (FALSE). |

## Value

A list with the following components:

| | |
|---|---|
| p | The p-value for the joint, integrative total effect test for the gene set under a pre-specified disease risk model for the whole gene set. |
| iapprox | The method of gene-set p-value calculation: 'pert' or 'davies'. |
| imodel | The user-specified disease risk model for the whole gene set. |
| p.emp | If using 'pert', and if gsp.emp is set to 'FALSE', returns the empirical p-value in addition to the approximated p-value. |
| gsp.emp | A logical value to indicate the method of calculation for the perturbation based p-value: TRUE for empirical, FALSE for parametric. |

## Author(s)

Su H. Chu & Yen-Tsung Huang

## References

Chu S.H. and Huang Y-T. (2017) Integrative genomic analysis of biological gene sets with applications in lung cancer. (Revise and Resubmit)

## Examples

```
data(X); data(Y); data(CPG); data(GE)
itegs(iCPG=CPG, iGE=GE, iY=Y, iX=X, imodel='mgc', iapprox='pert', gsp.emp=FALSE);
itegs(iCPG=CPG, iGE=GE, iY=Y, iX=X, imodel='mgc', iapprox='davies');
```

test.pw.info *Annotation Information for Gautschi SRC CpG Sites*

## Description

Includes the name of the pathway within MsigDB, the full list of lung-cancer associated genes originally reported by Gautschi et al (2008) in the SRC signaling pathway, and limited annotation information for the CpG sites which both mapped to the members of the gene set and were also available for testing in the integrative analyses.

## Usage

```
data(test.pw.info)
```

## Format

An R list object with three indices: `pw.name` (the name of the MsigDB pathway), `pw.genes` (the list of genes in the pathway), and `annot` (limited annotation information on the CpG sites that are associated with gene members of the pathway).

## Source

**TCGA** http://cancergenome.nih.gov/

**MsigDB** http://software.broadinstitute.org/gsea/msigdb/

## References

Gautschi O, Tepper CG, Purnell PR, Izumiya Y, Evans CP, Green TP, et al. (2008) Regulation of Id1 expression by SRC: implications for targeting of the bone morphogenetic protein pathway in cancer. *Cancer Research*. **68**(7):2250–8.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. **102**(43):15545–50.

## Examples

```
data(test.pw.info)
ls(test.pw.info)
```

## X — *Lung Cancer Patient Clinical Data*

### Description

TCGA lung cancer patient clinical data including sex, race, age at diagnosis, pathological stage of tumor at diagnosis, cancer type, and smoking history.

### Usage

```
data(X)
```

### Format

A named R matrix with 249 subject rows and 6 clinical phenotyp columns.

### Details

Clinical data were obtained from lung cancer patients (adenocarcinoma or squamous cell carcinoma) in The Cancer Genome Atlas among those who had one-year survival information, as well as both CpG methylation and mRNA expression data available. Recorded variables in this test dataset include sex, race (white or non-white), age at initial clinical diagnosis, pathological stage of tumor at biopsy, cell-type (adenocarcinoma or squamous cell), and smoking history in packyears.

### Source

**TCGA** http://cancergenome.nih.gov/

### Examples

```
data(X)
str(X)
```

## Y — *Lung Cancer Patient 1-Year Survival Status*

### Description

Lung cancer survival status at one year after initial diagnosis (0=non-survivor, 1=survivor).

### Usage

```
data("Y")
```

### Format

An R vector with length of 249.

## Source

**TCGA** http://cancergenome.nih.gov/

## Examples

```
data(Y)
str(Y)
```

# Index