

# Package ‘PEACH’

October 12, 2022

**Title** Pareto Enrichment Analysis for Combining Heterogeneous Datasets

**Version** 0.1.1

**Author** Jinyan Chan [aut, cre], Jinghua Gu [aut, cre]

**Maintainer** Jinyan Chan <Jinyan\_Chan@baylor.edu>

**Description** A meta gene set analysis tool developed based on principles of Pareto dominance (William B T Mock (2011) <[doi:10.1007/978-1-4020-9160-5\\_341](https://doi.org/10.1007/978-1-4020-9160-5_341)>). It is designed to combine gene set analysis p-values from multiple transcriptome datasets (e.g., microarray and RNA-Seq). The novel Pareto method for p-value combination allows PEACH to properly model heterogeneity and correlation in Omics datasets.

**Depends** R (>= 3.5.0), MASS (>= 7.3-50), mnormt (>= 1.5-6), metap (>= 1.3)

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.0

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-01-07 17:00:05 UTC

## R topics documented:

KEGG . . . . .	2
peach . . . . .	2
TCGA.input . . . . .	3
<b>Index</b>	<b>5</b>

---

KEGG

*KEGG pathways*

---

### **Description**

This data set gives the geneset list of KEGG pathways

### **Usage**

KEGG

### **Format**

A large list containing 186 pathways.

### **Source**

KEGG genesets from GSEA MSigDB Collections

---

peach

*Pareto Enrichment Analysis for Combining Heterogeneous datasets*

---

### **Description**

This function is for pathway enrichment meta-analysis with Pareto dominance based method. The input of this function is the gene level test-statistics (e.g. t statistics) from multiple datasets on which meta analysis will be performed, and a pathway (or geneset) list. It outputs the pathway p-values for each individual dataset as well as the pareto combined pathway p-values in a data frame.

### **Usage**

```
peach(  
  input.data = NULL,  
  nsample = 1000,  
  input.pathway = NULL,  
  direction = "up",  
  is.Fisher.Stouffer = TRUE  
)
```

**Arguments**

<code>input.data</code>	The test statistics of each gene from multiple datasets (the test statistics is from the case versus control statistical test, e.g. t-test). The rows are genes where the rownames are gene names (official gene symbols). The columns are the individual datasets.
<code>nsample</code>	The number of random sampling times for Pareto meta-analysis p-value calculation. As Pareto based meta-analysis is a non-parametric method, this parameter decides the NULL distribution size of meta-pathway p-value computing.
<code>input.pathway</code>	the pathways or genesets in the format of lists. The pathways or genesets should be defined by official gene symbols. An example KEGG pathway can be obtained with <code>data('KEGG')</code> . (The pathway input format is the same with the output from the <code>'gmtPathways'</code> function from <code>fgsea</code> package.)
<code>direction</code>	"up" or "down" denoting if the pathway p-value is calculated by accounting for pathway up-regulation or down-regulation. The default is "up", which means the peach function calculated combined p-value indicates if a pathway is up-regulated across the datasets being combined.
<code>is.Fisher.Stouffer</code>	Logical indication. If TRUE, peach function will output the combined meta-pathway-p-value from non-parametric Fisher's and Stouffer's method. The combined p-value will not be the same with the original Fisher's or Stouffer's method, as this version has the Monte Carlo implementation of these two methods that accounts for the correlation from the input dataset.

**Examples**

```
## load example input data (TCGA cancer versus control test t statistics)
data('TCGA.input')
## load the KEGG pathways
data('KEGG')
## Run peach

res = peach(input.data=TCGA.input,input.pathway=KEGG,direction="up",is.Fisher.Stouffer = TRUE)
```

---

<code>TCGA.input</code>	<i>A sample of 16 cancers gene differential expression analysis t statistic score data</i>
-------------------------	--

---

**Description**

This data set gives the TCGA 16 cancer types tumor versus normal t-test scores of each gene.

**Usage**

```
TCGA.input
```

**Format**

A matrix containing the t statistics of 20501 genes/rows and 16 cancer types/columns

**Source**

TCGA

# Index

- \* **datasets**
    - TCGA.input, 3
  - \* **function**
    - peach, 2
  - \* **pathways**
    - KEGG, 2
  - \* **peach**
    - peach, 2
- KEGG, 2
- peach, 2
- TCGA.input, 3