Introduction of DriverFind

In this study, we are interested in the effect of genomics alterations within the pathways on the pathways. Thus, we development the algorithm called DriverFind. DriverFind can identify the genomic alterations which significantly alter the pathway level using Kruskal-Wallis test and Nemenyi test.

Input data

To run DriverFind, mRNA expression data, genomic alteration matrices and pathway list are needed. DriverFind can select the gene alterations in the pathways automatically. The sample order of the input data does not need to match. The DriverFind will match the order automatically.

Principle

First, GSVA R package was used to assess the level of the pathways in the pathways list. The default GSVA method was gsva. Users can change to ssgsea, plage or zscore.

For CNA (copy number alteration) which contained copy number amplification and copy number deletion, first we used Kruskal-Wallis test to screen genes which alterations can significantly influence the level of pathways. For selected CNA, we used Nemenyi test to further compare whether copy number amplification or copy number deletion can significantly altered the level of pathways. For mutation or methylation, we used Wilcoxon sum rank test to screen the genes which can significantly altered the level of pathways. P values will be adjusted using FDR method.

Imports

data.table; parallel; splitstackshape; GSVA; PMCMRplus; ggplot2

Usage

```
> rm(list=ls())
```

Load the mRNA expression data and genomic alterations data.

```
> load("Example_data_for_DriverFind.Rdata")
```

Display the category of the data

```
> category
[1] "mrna" "cnv"    "met"    "mut"
```

Display the mRNA expression data

```
> omics[[1]][1:4,1:4]
```

	TCGA-2Y-A9GS-01A	TCGA-2Y-A9GT-01A	TCGA-2Y-A9GU-01A	TCGA-2Y-A9GV-01A
A1BG	14.466771	16.245087	13.18666	16.36228
A1CF	10.619945	10.323539	10.29122	10.68075
A2BP1	1.792564	0.000000	0.00000	0.00000
A2LD1	6.638434	7.749313	11.50892	8.66103

Display the genomic alterations data: CNA, methylation and mutation

```
> omics[[2]][1:4,1:4]
       TCGA-2V-A95S-01A TCGA-2Y-A9GS-01A TCGA-2Y-A9GT-01A TCGA-2Y-A9GU-01A
ACAP3                0                0                0                0
ACTRT2               0                0                0                0
AGRN                 0                0                0                0
ANKRD65              0                0                0                0
> omics[[3]][1:4,1:4]
       TCGA-EP-A26S-01A TCGA-DD-AAVX-01A TCGA-ED-A7PZ-01A TCGA-LG-A6GG-01A
TBX15                0                1                0                0
DAB2IP               0                1                1                1
CD58                 0                1                0                0
PDE2A                0                0                0                0
> omics[[4]][1:4,1:4]
       TCGA-BC-A5W4-01A TCGA-RG-A7D4-01A TCGA-DD-A1EE-01A TCGA-DD-AACB-01A
CTNNB1                1                1                1                1
TP53                  1                1                1                1
TTN                   1                1                0                0
MUC16                 0                0                1                1
```

Display the pathways list.

```
> pathway_list[1:2]
$`hsa00010: Glycolysis / Gluconeogenesis`
 [1] "HK3"       "HK1"       "HK2"       "HKDC1"     "GCK"       "GPI"       "PFKM"
 [8] "PFKP"      "PFKL"      "FBP1"      "FBP2"      "ALDOC"     "ALDOA"     "ALDOB"
[15] "TPI1"      "GAPDH"     "GAPDHS"    "PGK2"      "PGK1"      "PGAM1"     "PGAM2"
[22] "PGAM4"     "ENO3"      "ENO2"      "ENO1"      "ENO4"      "PKM"       "PKLR"
[29] "PDHA2"     "PDHA1"     "PDHB"      "DLAT"      "DLD"       "LDHAL6A"   "LDHAL6B"
[36] "LDHA"      "LDHB"      "LDHC"      "ADH1A"     "ADH1B"     "ADH1C"     "ADH7"
[43] "ADH4"      "ADH5"      "ADH6"      "AKR1A1"    "ALDH2"     "ALDH3A2"   "ALDH1B1"
[50] "ALDH7A1"   "ALDH9A1"  "ALDH3B1"   "ALDH3B2"   "ALDH1A3"   "ALDH3A1"   "ACSS1"
[57] "ACSS2"     "GALM"      "PGM1"      "PGM2"      "G6PC"      "G6PC2"     "G6PC3"
[64] "ADPGK"     "BPGM"      "MINPP1"    "PCK1"      "PCK2"

$`hsa00020: Citrate cycle (TCA cycle)`
 [1] "CS"        "ACLY"      "ACO2"      "ACO1"      "IDH1"      "IDH2"      "IDH3B"     "IDH3G"
 [9] "IDH3A"     "OGDHL"     "OGDH"      "DLST"      "DLD"       "SUCLG1"    "SUCLG2"    "SUCLA2"
[17] "SDHA"      "SDHB"      "SDHC"      "SDHD"      "FH"        "MDH1"      "MDH2"      "PC"
```

[25] "PCK1" "PCK2" "PDHA2" "PDHA1" "PDHB" "DLAT"

Run DriverFind

```
> source("DriverFind.R")
> out=DriverFind(omics=omics, category=category, cores=NULL, gsva_method="gsva", pathway_list=pathway_list)
```

Finally, we can analyze the result. We use pathway hsa00030: Pentose phosphate pathway to show how to analyze the result. The outcome of the analysis is a list which contained two parts. The first part is the plot of the outcome. The second part is the difference of the mean value between alteration and normal of the potential driver alterations.

```
> analysis(out [which(names(out)=="hsa00030: Pentose phosphate pathway")])
[[1]]

[[2]]
      V1  V2          V3
5 ALDOA cnd -0.2337661
3  PGM1 cnd -0.2074395
2   PGD cnd -0.1994832
1  H6PD cnd -0.1894414
4 GLYCTK cnd -0.1261604
```