Introduction of InterFind

In this study, we designed InterFind to explore which significant co-occurrence or mutual exclusivity pathway pair can significantly influence the response of immunotherapy.

Input data

To run InterFind, genomic alteration matrices, category of genomic alteration, pathways list and immunotherapy response scores are needed. The sample order of the input data does not need to match. The DriverFind will match the order automatically. We suggest that do not use all genes to run InterFind. 72GB RAM is still not enough for all genes. We suggest transforming the genomic alteration matrices into pathways alteration matrices to save RAM. We have designed the transform function called alt_trans_path which can transform the genomic alteration matrices into pathways alteration matrices. If you have had pathway alteration matrices, you do not need genomic alteration matrices and pathways list. You can directly run InterFind using pathway alteration matrices, category of genomic alteration and immunotherapy response scores.

Principle

The InterFind will compared the response scores among different alteration status (both non-alteration, pathway 1 alteration, pathway 2 alteration and both pathways alteration) in pathways pair using Kruskal-Wallis test. Meanwhile, InterFind will judge whether pathways pair is significant co-occurrence or mutual exclusivity utilizing pair-wise fisher's exact test. If OR>1 the pathways pair is co-occurrence, else the pathways pair is mutual exclusivity. P values of Kruskal-Wallis test and pair-wise Fisher's Exact test are adjusted using FDR method respectively.

Imports

data.table; parallel; splitstackshape; dplyr; PMCMRplus; ggplot2; ggsci; ggpubr

Usage

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

Load omics data and source code.

```
> load("Example_data_for_InterFind.Rdata")
> source("InterFind.R")
```

Display genomic alteration matrices and their category.

```
> omics[[1]][1:4,1:4]
        TCGA-2Y-A9GS-01A TCGA-2Y-A9GT-01A TCGA-2Y-A9GU-01A TCGA-2Y-A9GV-01A
DDAH1                0                0                0                0
MET                  1                0                1                0
BPHL                 0                0                0                1
PTP4A1               0                0                0                0
> omics[[2]][1:4,1:4]
        TCGA-2Y-A9GS-01A TCGA-2Y-A9GT-01A TCGA-2Y-A9GU-01A TCGA-2Y-A9GV-01A
```

RB1 1 0 0 1
CDKN2A 1 0 1 1
CASP3 1 0 0 0
TCTE3 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
AXIN1 0 0 0 0
TP53 1 1 1 1
BAP1 0 0 0 0
> category
[1] "cna" "cnd" "met" "mut"

Display pathway 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"

Display the immunotherapy response scores.

```
> head(tide)
                                V1                 V2
TCGA-DD-AACC-01A TCGA-DD-AACC-01A 1.84
TCGA-DD-AAVQ-01A TCGA-DD-AAVQ-01A 1.75
TCGA-ED-A5KG-01A TCGA-ED-A5KG-01A 1.70
TCGA-K7-A5RG-01A TCGA-K7-A5RG-01A 1.68
TCGA-ED-A627-01A TCGA-ED-A627-01A 1.65
TCGA-CC-A7II-01A TCGA-CC-A7II-01A 1.62
```

Transform the genomic alteration matrices into pathways alteration matrices.

```
> pam=alist()
> for (i in 1:length(omics)){
+     pam [[i]]= alt_trans_path(pathway_list= pathway_list, data=omics[[i]])
+ }
```

Run InterFind

```
> res=InterFind(dat= pam, category=category, response=tide, returnAll=FALSE, fdr_KW=0.05, fdr_pair=0.01, cores=NULL)
```

The outcome of InterFind composed by three part. The first part is the pathways interaction information. The second part is the comparison of the response scores among different alteration status (both non-alteration, pathway 1 alteration, pathway 2 alteration and both pathways alteration) in pathways pair using Nemenyi test. The third part is the mean values of the response scores among different alteration status (both non- alteration, pathway 1 alteration, pathway 2 alteration and both pathways alteration) in pathways pair.

```
> head(res[[1]])
      gene1
3     "hsa00040: Pentose and glucuronate interconversions"
270 "hsa00040: Pentose and glucuronate interconversions"
271 "hsa00061: Fatty acid biosynthesis"
292 "hsa00510: N-Glycan biosynthesis"
294 "hsa00513: Various types of N-glycan biosynthesis"
295 "hsa00520: Amino sugar and nucleotide sugar metabolism"
      gene2                                                 type1 type2 00       11       01
3     "hsa00010: Glycolysis / Gluconeogenesis" "cnd" "cnd" "110" "109" " 78"
270 "hsa00020: Citrate cycle (TCA cycle)"      "cnd" "cnd" "138" " 93" " 50"
271 "hsa00020: Citrate cycle (TCA cycle)"      "cnd" "cnd" " 98" "108" " 35"
292 "hsa00020: Citrate cycle (TCA cycle)"      "cnd" "cnd" "111" "140" "  3"
294 "hsa00020: Citrate cycle (TCA cycle)"      "cnd" "cnd" "144" "134" "  9"
295 "hsa00020: Citrate cycle (TCA cycle)"      "cnd" "cnd" "123" " 82" " 61"
      10       fdr_KW             fdr_fisher        OR                  Event
3     " 37" "0.03356925" "2.425022e-09" "    4.135659" "Co_Occurance"
270 " 53" "0.02327398" "2.763101e-11" "    4.817587" "Co_Occurance"
```

```
271 " 93" "0.03939986" "8.588293e-07" "     3.239861" "Co_Occurance"
292 " 80" "0.01982360" "3.940167e-30" "    64.024223" "Co_Occurance"
294 " 47" "0.03538492" "8.085378e-39" "    44.899800" "Co_Occurance"
295 " 68" "0.03245598" "1.119796e-04" "     2.424832" "Co_Occurance"

> res[[2]][1:3]
[[1]]
                  00          01          10
01 0.30102355             NA          NA
10 0.01226440 0.3725264           NA
11 0.00533443 0.5838126 0.8963514

[[2]]
                  00          01          10
01 0.140544908            NA          NA
10 0.001638629 0.6391929          NA
11 0.015627303 0.9938382 0.6973415

[[3]]
                  00          01          10
01 0.999999960            NA          NA
10 0.242921501 0.52637652         NA
11 0.004705176 0.07927462 0.5110597

> head(res[[3]])
                  00          11                  01          10
[1,] 0.11881818     0.4148624     0.29653846    0.4762162
[2,] 0.12615942     0.3941935     0.37580000    0.4939623
[3,] 0.15357143     0.4578704     0.17142857    0.3068817
[4,] 0.09009009     0.3967857     -0.03333333   0.4198750
[5,] 0.16395833     0.4113433     0.03666667    0.4251064
[6,] 0.12211382     0.4300000     0.33098361    0.4201471
```

We designed a function called analysis to analyze the outcome. The analysis function is designed to screened the significant outcome from all pathways pairs. The outcome of the analysis is a list composed by two parts. The first part is the detailed information of pathways pairs. The second part is the comparison of the response scores among different alteration status (both non-alteration, pathway 1 alteration, pathway 2 alteration and both ##pathways alteration) in pathways pair using Nemenyi test.

```
> w=analysis(outcome=res, cutoff=0.05)

  72 pairs can significantly increase the scores
  5 pairs can significantly decrease the scores
> head(w[[1]])
```

```
      gene1
845   "hsa00760: Nicotinate and nicotinamide metabolism"
1202 "hsa04614: Renin-angiotensin system"
4050 "hsa00770: Pantothenate and CoA biosynthesis"
4059 "hsa01040: Biosynthesis of unsaturated fatty acids"
5651 "hsa00760: Nicotinate and nicotinamide metabolism"
5672 "hsa03060: Protein export"
      gene2                                                          type1 type2
845   "hsa00061: Fatty acid biosynthesis"                          "cnd" "cnd"
1202 "hsa00062: Fatty acid elongation"                             "cnd" "cnd"
4050 "hsa00280: Valine, leucine and isoleucine degradation" "cnd" "cnd"
4059 "hsa00280: Valine, leucine and isoleucine degradation" "cnd" "cnd"
5651 "hsa00410: beta-Alanine metabolism"                           "cnd" "cnd"
5672 "hsa00410: beta-Alanine metabolism"                           "cnd" "cnd"
      00       11       01       10       fdr_KW             fdr_fisher        OR
845   " 65" "161" " 40" " 68" "0.01853599" "6.196297e-08" "     3.830074"
1202 " 80" "110" "120" " 24" "0.03484261" "2.577916e-05" "     3.045607"
4050 " 58" "147" "116" " 13" "0.01022712" "1.522233e-08" "     5.625742"
4059 " 60" "151" "112" " 11" "0.01022712" "2.119669e-10" "     7.312148"
5651 " 85" " 98" " 20" "131" "0.02512712" "2.870444e-05" "     3.168963"
5672 "150" " 73" " 45" " 66" "0.03084991" "5.749145e-08" "     3.670980"
      Event              Effect
845   "Co_Occurance" "increase"
1202 "Co_Occurance" "increase"
4050 "Co_Occurance" "increase"
4059 "Co_Occurance" "increase"
5651 "Co_Occurance" "increase"
5672 "Co_Occurance" "increase"
> head(w[[2]])
              00         11         01          10
[1,] 0.1589231  0.4535404  0.1242500  0.15764706
[2,] 0.1773750  0.4690000  0.2693333  0.03916667
[3,] 0.1362069  0.4782993  0.1949138 -0.13692308
[4,] 0.1316667  0.4700662  0.1958929 -0.16181818
[5,] 0.1611765  0.5030612  0.0800000  0.26290076
[6,] 0.2413333  0.5500000  0.2388889  0.18090909
```

Next, we can plot the violin plot to display the comparison of the response scores among different alteration status (both non-alteration, pathway 1 alteration, pathway 2 alteration and both pathways alteration) in pathways pair. i represents the row number of the outcome of analysis. The outcome of plot_analysis composed by two parts. The first part is the violin plot and the second part is the outcome of Nemenyi test.

```
> plot_analysis(outcome=w[[1]], i=1)
[[1]]
```

[[2]]

Pairwise comparisons using Tukey-Kramer-Nemenyi all-pairs test with Tukey-Dist approximation

data: as.numeric(as.matrix(tag$V2)) by as.factor(tag$V1)

	00	01	10
01	0.9985	-	-
10	0.9990	1.0000	-
11	0.0094	0.0270	0.0045

P value adjustment method: single-step
alternative hypothesis: two.sided

Warning message:
In kwAllPairsNemenyiTest.default(c(-0.01, 1.36, -0.43, -1.51, -1.03, :
Ties are present, p-values are not corrected.