survivalHelp Demo

Hope Mumme

12/19/2021

```
knitr::opts_chunk$set(echo = TRUE)
#devtools::install("hmumme/survivalHelp") # uncomment if you have not installed survivalHelp
library(survivalHelp)
load("~/GitHub/survivalHelp/demo/data/mpalData.RData") # load demo data
```

The file mpalData.RData contains the necessary objects to run survival analysis on the MPAL samples from the TARGET-ALL-P3 (ALAL) datasets. It includes the bulkRNA-seq expression matrices for both the B/Myeloid and T/Myeloid pediatric MPAL samples, the ensmbl id matrices, the ensmbl to gene symbol reference data frame, and the survival data for both subsets.

```
ls()
```

```
## [1] "bENS" "bMat" "osB"  "osT"  "ref"  "tENS" "tMat"
```

Single Gene First, we will show how to obtain survival plots for a single gene's expression in B/Myeloid MPAL samples. You can use either median or cutP as the cut method here.

```
gene = "TP53"
out = mCut(gene, bMat, osB, cut = "cutP")
```

```
## [1] "Cutoff = 4.18"
```

```
out
```

```
## $cutoff
## [1] 4.18
##
## $geneExp
##  [1] 2.764898 2.174534 4.274291 4.183876 4.867383 3.501388 3.650339 2.814366
##  [9] 3.378220 3.432946 3.808004 5.423713 4.202089 3.995177 4.235658 4.217549
## [17] 5.004767 4.677813 4.430098 3.476395 2.450538 4.615487 3.443859 2.820383
## [25] 3.122325
```

The mCut function returns the gene expression cutoff to use to split samples into High and Low groups based on cutP method and the gene expression values for each sample.

```
osB["group"] = cutGroups(osB, out)
osB
```

```
## # A tibble: 25 x 4
##    sample                           OS Vital.Status group
##    <chr>                         <dbl> <chr>        <fct>
##  1 TARGET-15-SJMPAL012419_diagnosis 40.4  Alive        Low
##  2 TARGET-15-SJMPAL012422_diagnosis 68.1  Alive        Low
##  3 TARGET-15-SJMPAL012425_diagnosis 19.4  Dead         High
##  4 TARGET-15-SJMPAL016448_diagnosis 13.2  Dead         High
##  5 TARGET-15-SJMPAL017975_diagnosis 96.2  Alive        High
##  6 TARGET-15-SJMPAL040024_diagnosis 34.0  Dead         Low
##  7 TARGET-15-SJMPAL040025_diagnosis 70.0  Alive        Low
##  8 TARGET-15-SJMPAL040030_diagnosis  2.17 Dead         Low
##  9 TARGET-15-SJMPAL040031_diagnosis 81.9  Alive        Low
## 10 TARGET-15-SJMPAL040033_diagnosis 54.5  Dead         Low
## # ... with 15 more rows
```

The samples have been split into High and Low categories in the osB group column.

```
plotSurv(osB, name = "TP53")
```

Survival Curves by TP53 Expression
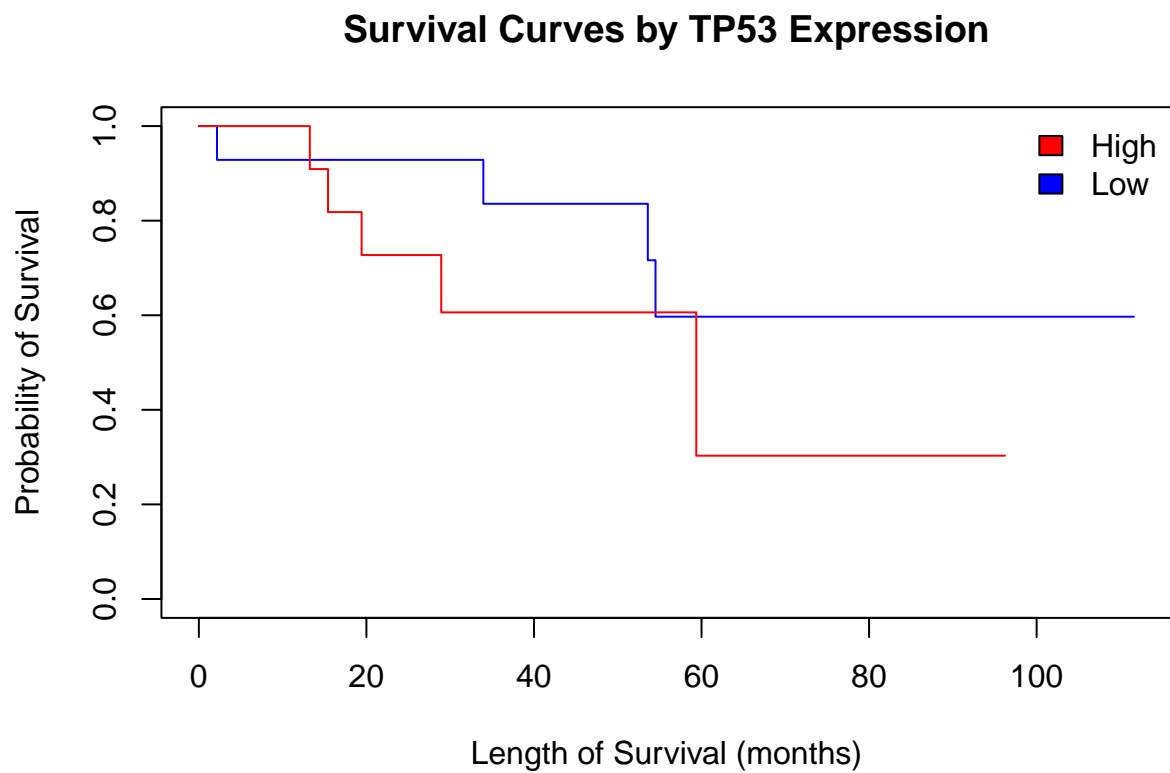
We can also calculate the survival statistics.

```
survStats(osB)
```

```
##   Comparison survdiffP    coxHR      coxP
## 1            0.2367862 2.202731 0.247774
```

2

We calculated the logRank p-value (survdiffP) for the survival plot, the Cox Hazard Ratio (coxHR), and the Cox HR p-value (coxP). From these statistics, we can see that the expression of TP53 does not have a significant effect on B/Myeloid MPAL pediatric survival.

Gene Set

Second, we will show how to obtain survival plots for a combined gene set's expression in T/Myeloid MPAL samples. You can use either median or cutP as the cut method here. Instead of individual expression values, we take the GSEA scores for a gene set and use this when splitting samples into High/Low groups.

```
gs = c("TP53","ICAM1","HPGD")
out = mSet(gs = gs, ENS = tENS, os = osT, ref = ref, cut = "cutP")
```

```
## Warning in MulticoreParam(progressbar = verbose, workers = parallel.sz, :
## MulticoreParam() not supported on Windows, use SnowParam()
```

```
## [1] "Cutoff = 0.615"
```

```
out
```

```
## $cutoff
## [1] 0.615
##
## $scores
##                                  GSVA score
## TARGET-15-SJMPAL011911_diagnosis -1.5143186
## TARGET-15-SJMPAL011912_diagnosis -0.6591309
## TARGET-15-SJMPAL011913_diagnosis -0.6922743
## TARGET-15-SJMPAL011914_diagnosis  0.2660055
## TARGET-15-SJMPAL011915_diagnosis -1.6432046
## TARGET-15-SJMPAL012420_diagnosis -1.3422825
## TARGET-15-SJMPAL012421_diagnosis -1.9336741
## TARGET-15-SJMPAL016340_diagnosis -1.1024330
## TARGET-15-SJMPAL016342_diagnosis  1.1053439
## TARGET-15-SJMPAL016343_diagnosis  0.9934946
## TARGET-15-SJMPAL016344_diagnosis -1.1994352
## TARGET-15-SJMPAL016447_diagnosis -0.5609134
## TARGET-15-SJMPAL017978_diagnosis  1.3965067
## TARGET-15-SJMPAL022667_diagnosis -1.0397229
## TARGET-15-SJMPAL040027_diagnosis  1.4847092
## TARGET-15-SJMPAL040037_diagnosis  0.4313799
## TARGET-15-SJMPAL040459_diagnosis  0.4460435
## TARGET-15-SJMPAL042791_diagnosis -0.1535142
## TARGET-15-SJMPAL042792_diagnosis -0.9037000
## TARGET-15-SJMPAL042793_diagnosis  0.5357397
## TARGET-15-SJMPAL042794_diagnosis -1.0056206
## TARGET-15-SJMPAL042796_diagnosis  0.9312591
## TARGET-15-SJMPAL042797_diagnosis  1.8474669
## TARGET-15-SJMPAL042801_diagnosis -0.5053526
## TARGET-15-SJMPAL043505_diagnosis  1.8338670
## TARGET-15-SJMPAL043510_diagnosis  1.9461764
## TARGET-15-SJMPAL043514_diagnosis  0.6146132
```

```
## TARGET-15-SJMPAL043773_diagnosis  0.9792731
## TARGET-15-SJMPAL044948_diagnosis -0.5563021
```

The mSet function returns the GSEA score cutoff to use to split samples into High and Low groups based on cutP method and the GSEA values for each sample.

```
osT["group"] = cutGroups(osT, out)
osT
```

```
## # A tibble: 29 x 4
##    sample                          OS Vital.Status group
##    <chr>                        <dbl> <chr>        <fct>
##  1 TARGET-15-SJMPAL011911_diagnosis  66.9  Alive        Low
##  2 TARGET-15-SJMPAL011912_diagnosis 104.   Alive        Low
##  3 TARGET-15-SJMPAL011913_diagnosis 112.   Alive        Low
##  4 TARGET-15-SJMPAL011914_diagnosis  92.5  Alive        Low
##  5 TARGET-15-SJMPAL011915_diagnosis   6.51 Dead         Low
##  6 TARGET-15-SJMPAL012420_diagnosis 138.   Alive        Low
##  7 TARGET-15-SJMPAL012421_diagnosis  56.5  Alive        Low
##  8 TARGET-15-SJMPAL016340_diagnosis   7.33 Alive        Low
##  9 TARGET-15-SJMPAL016342_diagnosis   9.70 Dead         High
## 10 TARGET-15-SJMPAL016343_diagnosis  16.8  Dead         High
## # ... with 19 more rows
```

The samples have been split into High and Low categories in the osT group column.

```
plotSurv(osT, name = "TP53, ICAM1, and HPGD")
```

Survival Curves by TP53, ICAM1, and HPGD Expression
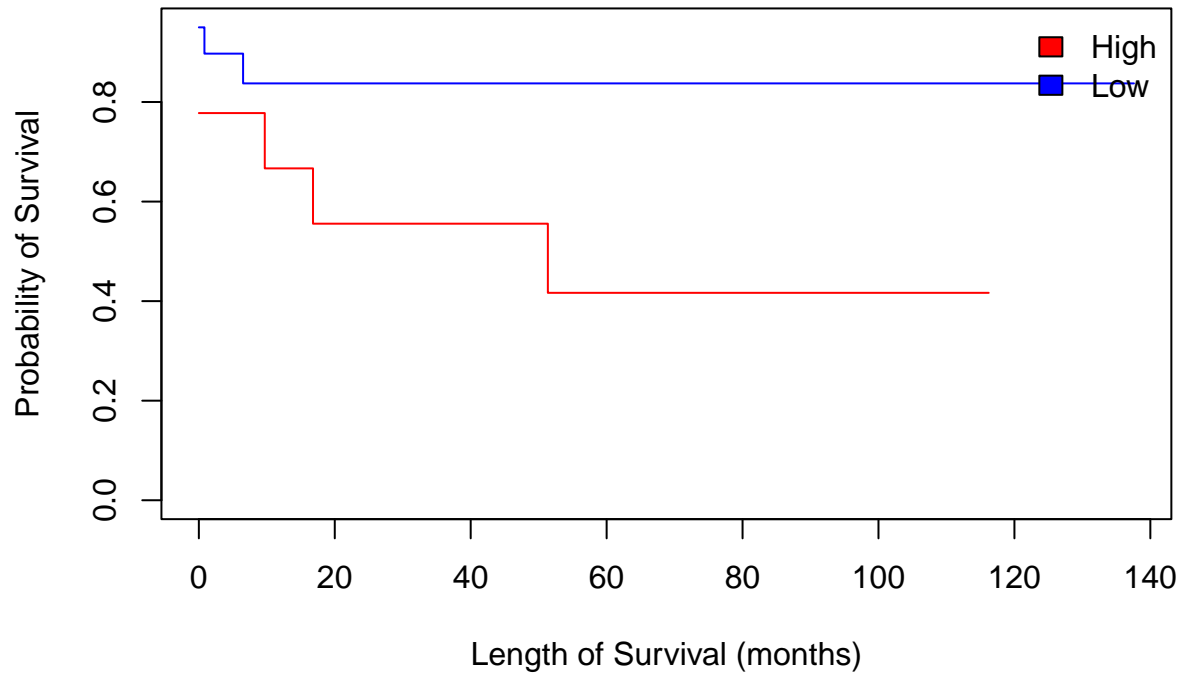
We can calculate the survival statistics for the gene set

```
survStats(osT)
```

```
##   Comparison  survdiffP    coxHR        coxP
## 1             0.06036603 3.643607 0.07719813
```

We calculated the logRank p-value (survdiffP) for the survival plot, the Cox Hazard Ratio (coxHR), and the Cox HR p-value (coxP). From these statistics, we can see that the gene set of TP53, ICAM1, and HPGD does not have a significant effect on T/Myeloid MPAL pediatric survival.