Supplemental Data File. Sweave documentation for microarray data analysis, Related to Figure 2

Introduction

This supplemental file includes R documentation for microarray analysis. We selected the significant gene list (FDR 0.1) from cell lines by fitting linear regression models on resistant cell line series (H1299 and H1355) using log transformed IC50 values. For xenograft microarray data, we performed student's t test for differential gene analysis with FDR 0.1. 35 genes (14 up regulated and 21 down regulated) were selected from the intersected genes for both cell lines and xenografts. We further tested the 35 gene preclinical signature on 65 patients who had received neoadjuvant chemotherapy. Unsupervised clustering using 35 genes separated the 65 patients into two groups. K-M curve for recurrence-free survival analysis showed that group 2 has significantly worse prognosis. Multivariate cox regression model for the 35 genes showed that the up-regulated gene **KDM3B** has the largest hazard ratio for poor cancer recurrence-free survival.

Microarray data were pre-processed by R package mbcb for background correction, then log-transformed and quantile-normalized with the R package preprocessCore. We summarized the gene-level expression by averaging the normalized probes intensity value if multiple probes mapped to the same gene.

Before running the code, put the data in the same folder with the code.

Statistics Analysis

H1299 and H1355 linear regression model

1. set the working enviroment and call the library package
   setwd("~/Documents/YY_Project/maithili/SWEAVE/")

```
library(ClassComparison)

## Loading required package: splines

## Loading required package: Biobase

## Loading required package: BiocGenerics

## Loading required package: parallel

##
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':
##
```

```
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB

## The following objects are masked from 'package:stats':
##
##     IQR, mad, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, as.vector, cbind,
##     colnames, do.call, duplicated, eval, evalq, Filter, Find, get,
##     grep, grepl, intersect, is.unsorted, lapply, lengths, Map,
##     mapply, match, mget, order, paste, pmax, pmax.int, pmin,
##     pmin.int, Position, rank, rbind, Reduce, rownames, sapply,
##     setdiff, sort, table, tapply, union, unique, unlist, unsplit

## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.

## Loading required package: oompaBase

library(survival)
library(VennDiagram)

## Loading required package: grid

## Loading required package: futile.logger

library(siggenes)

## Loading required package: multtest

library(gplots)

##
## Attaching package: 'gplots'

## The following object is masked from 'package:multtest':
##
##     wapply

## The following object is masked from 'package:oompaBase':
##
##     redgreen

## The following object is masked from 'package:stats':
##
##     lowess
```

2. Load normalized gene level expression data (preprocess procedure described in the method)

```
load("cell_line.RData")
```

3. run core function/analysis for H1299 Linear regression model for microarray at different time series with log transformed drug responses data IC 50.

```
# IC 50
ic50=c(rep(9.2, 5), rep(53, 2), rep(190, 2), rep(490, 2), rep(943, 3))
ic50=log(ic50) # log transform, otherwise est is too big.

# linear function
lm.ic=function(x){
  x=as.numeric(x)
  lm.x=lm(x~ic50)
  lm.co=summary(lm.x)$coefficients
  return(t(c(lm.co[2, 1], lm.co[2, 4])))
}

dim(df.1299)
```

```
## [1] 20549    19
```

```
head(df.1299[,1:5])
```

```
##   gene.id n gene.symbol H1299.Parental.1 H1299.Parental.2
## 1       1 2        A1BG         3.538114         3.530899
## 2       2 1         A2M         2.623364         3.244314
## 3       9 3        NAT1         3.791201         3.896611
## 4      10 1        NAT2         2.965261         3.086720
## 5      12 1    SERPINA3         2.933633         3.332563
## 6      13 1       AADAC         2.845070         3.086720
```

```
colnames(df.1299)
```

```
##  [1] "gene.id"          "n"                "gene.symbol"
##  [4] "H1299.Parental.1" "H1299.Parental.2" "H1299.Parental.3"
##  [7] "H1299.Untr.1"     "H1299.Untr.2"     "H1299.T5.1"
## [10] "H1299.T5.2"       "H1299.T10.1"      "H1299.T10.2"
## [13] "H1299.T15.1"      "H1299.T15.2"      "H1299.T18.1"
## [16] "H1299.T18.2"      "H1299.T18.3"      "est"
## [19] "p.value"
```

```
apply(df.1299[1:5, cell.1299], 1, lm.ic)
```

```
##                  1           2            3           4          5
## [1,] -0.005878157 -0.04970737 -0.166829025 0.008405362 0.79070808
## [2,]  0.884844954  0.20710959  0.003930657 0.826502466 0.00106438
```

```
est=apply(df.1299[, cell.1299], 1, lm.ic)
est=t(est)
```

```r
est=data.frame(est)
head(est)
```

```
##             X1          X2
## 1 -0.005878157 0.884844954
## 2 -0.049707375 0.207109593
## 3 -0.166829025 0.003930657
## 4  0.008405362 0.826502466
## 5  0.790708076 0.001064380
## 6  0.057528835 0.283359856
```

```r
names(est)=c("est", "p.value")
head(est)
```

```
##            est     p.value
## 1 -0.005878157 0.884844954
## 2 -0.049707375 0.207109593
## 3 -0.166829025 0.003930657
## 4  0.008405362 0.826502466
## 5  0.790708076 0.001064380
## 6  0.057528835 0.283359856
```

```r
df.1299=cbind(df.1299, est)

fdr=0.1
p.1299=cutoffSignificant(Bum(df.1299$p.value), fdr)
p.1299
```

```
## [1] 0.02927903
```

```r
table(df.1299$p.value < p.1299)
```

```
##
## FALSE  TRUE
## 16797  3752
```

```r
table(df.1299$p.value < p.1299 & df.1299$est < 0)
```

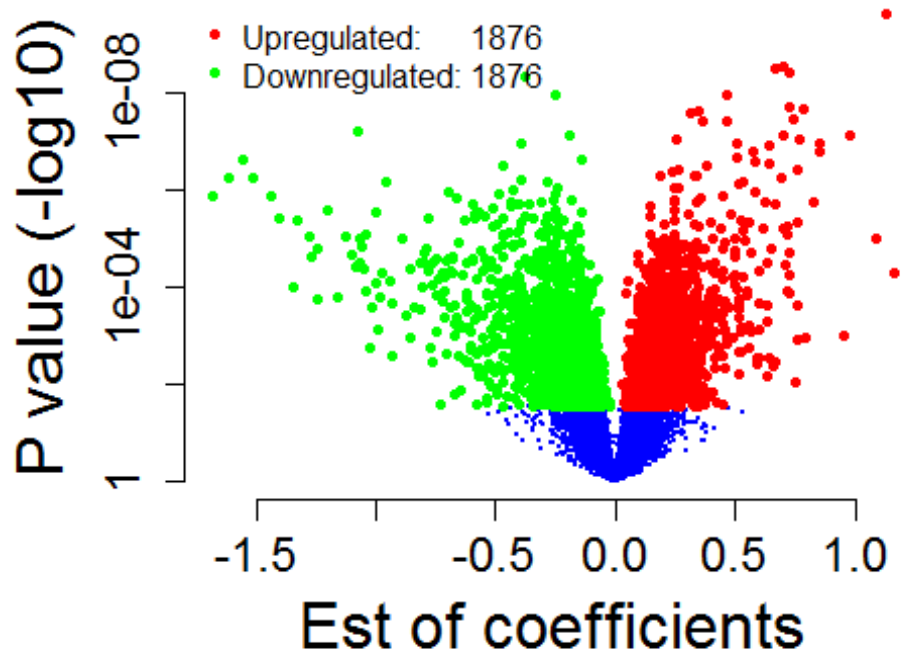
```
##
## FALSE  TRUE
## 18673  1876
```

```r
table(df.1299$p.value < p.1299 & df.1299$est > 0)
```

```
##
## FALSE  TRUE
## 18673  1876
```

```r
id.up.1299=df.1299$gene.id[df.1299$p.value < p.1299 & df.1299$est > 0]
id.down.1299=df.1299$gene.id[df.1299$p.value < p.1299 & df.1299$est < 0]
# volcano plot: p valu only
par(mar=c(5, 5, 2, 1))
plot(df.1299$est, -log10(df.1299$p.value), xlab="Est of coefficients",
```

```
cex=0.5,
     ylab="P value (-log10)", cex.lab=2, cex.axis=1.5, bty="n", col="blue",
pch=20,
     yaxt="n")
axis(2, at=c(0, 2, 4, 6, 8), labels=c(1, 0.01, 0.0001, 0.000001, 0.00000001),
cex.axis=1.5)
points(df.1299$est[df.1299$p.value < p.1299 & df.1299$est > 0],
       -log10(df.1299$p.value[df.1299$p.value < p.1299 & df.1299$est > 0]),
col="red",
       pch=20)
points(df.1299$est[df.1299$p.value < p.1299 & df.1299$est < 0],
       -log10(df.1299$p.value[df.1299$p.value < p.1299 & df.1299$est < 0]),
col="green",
       pch=20)
table(df.1299$p.value < p.1299 & df.1299$est > 0)

##
## FALSE  TRUE
## 18673  1876

table(df.1299$p.value < p.1299 & df.1299$est < 0)

##
## FALSE  TRUE
## 18673  1876

legend("topleft", c("Upregulated:      1876", "Downregulated: 1876"),
       col=c("red", "green"), pch=20, bty="n")
```



4. run core function/analysis for H1355

```
# IC 50
ic50=c(rep(2.2, 5), rep(15, 2), rep(25, 2), rep(245, 2), rep(315, 3))
ic50=log(ic50) # log transform, otherwise est is too big.
# significant p values will enrich.
ic50
```

```
##  [1] 0.7884574 0.7884574 0.7884574 0.7884574 0.7884574 2.7080502 2.7080502
##  [8] 3.2188758 3.2188758 5.5012582 5.5012582 5.7525726 5.7525726 5.7525726
```

```
# linear function
dim(df.1355)
```

```
## [1] 20549    19
```

```
head(df.1355[,1:5])
```

```
##   gene.id n gene.symbol H1355.Parental.1 H1355.Parental.2
## 1       1 2        A1BG         3.980461         3.570940
## 2       2 1         A2M         2.705274         2.888426
## 3       9 3        NAT1         4.303963         3.966284
## 4      10 1        NAT2         3.796795         3.593245
## 5      12 1    SERPINA3         3.229223         3.493975
## 6      13 1       AADAC         4.967962         5.713092
```

```
colnames(df.1355)
```

```
##  [1] "gene.id"          "n"                "gene.symbol"
##  [4] "H1355.Parental.1" "H1355.Parental.2" "H1355.Parental.3"
##  [7] "H1355.Untr.1"     "H1355.Untr.2"     "H1355.T4.1"
## [10] "H1355.T4.2"       "H1355.T8.1"       "H1355.T8.2"
## [13] "H1355.T13.1"      "H1355.T13.2"      "H1355.T16.1"
## [16] "H1355.T16.2"      "H1355.T16.3"      "est"
## [19] "p.value"
```

```
est=apply(df.1355[, cell.1355], 1, lm.ic)
est=t(est)
est=data.frame(est)
head(est)
```

```
##            X1           X2
## 1  0.02633836 0.4573952245
## 2 -0.03237285 0.1441376047
## 3  0.28051951 0.0004062189
## 4 -0.13380885 0.0753998578
## 5 -0.05697809 0.2763634625
## 6  0.14990967 0.2601679463
```

```
names(est)=c("est", "p.value")
head(est)
```

```
##           est      p.value
## 1  0.02633836 0.4573952245
## 2 -0.03237285 0.1441376047
## 3  0.28051951 0.0004062189
## 4 -0.13380885 0.0753998578
## 5 -0.05697809 0.2763634625
## 6  0.14990967 0.2601679463
```

```
df.1355=cbind(df.1355, est)
```

```
p.1355=cutoffSignificant(Bum(df.1355$p.value), fdr)
p.1355
```

```
## [1] 0.003536239
```

```
table(df.1355$p.value < p.1355)
```

```
##
## FALSE  TRUE
## 19954   595
```

```
table(df.1355$p.value < p.1355 & df.1355$est <0)
```

```
##
## FALSE  TRUE
## 20259   290
```

```
table(df.1355$p.value < p.1355 & df.1355$est > 0)
```

```
##
## FALSE  TRUE
## 20244   305
```

```
id.up.1355=df.1355$gene.id[df.1355$p.value < p.1355 & df.1355$est > 0]
id.down.1355=df.1355$gene.id[df.1355$p.value < p.1355 & df.1355$est < 0]
sum(id.up.1355 %in% id.up.1299)
```

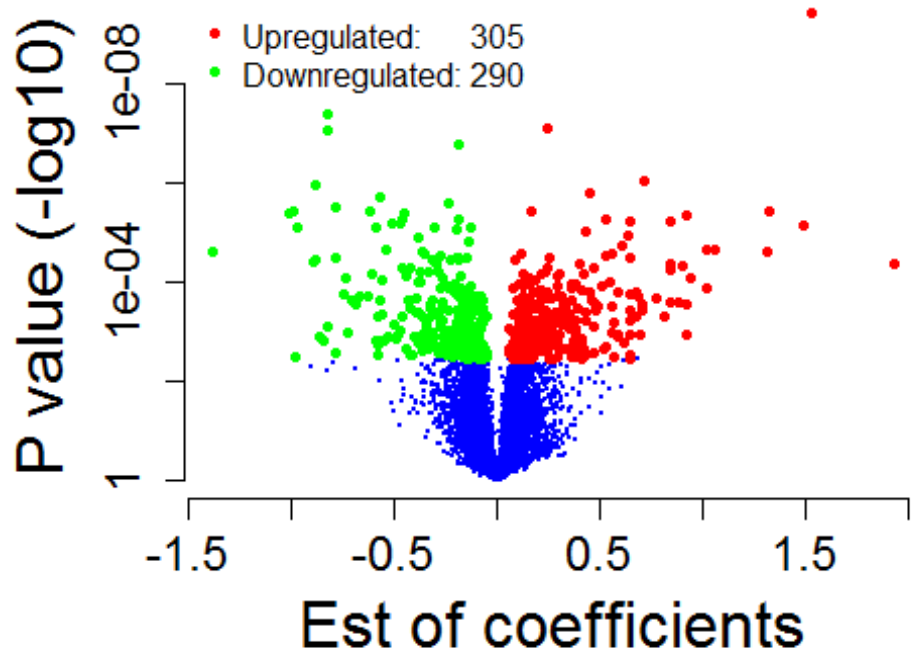
```
## [1] 51
```

```
sum(id.down.1355 %in% id.down.1299)
```

```
## [1] 59
```

```
par(mar=c(5, 5, 2, 1))
plot(df.1355$est, -log10(df.1355$p.value), xlab="Est of coefficients",
cex=0.5,
     ylab="P value (-log10)", cex.lab=2, cex.axis=1.5, bty="n", col="blue",
pch=20,
     yaxt="n")
axis(2, at=c(0, 2, 4, 6, 8), labels=c(1, 0.01, 0.0001, 0.000001, 0.00000001),
cex.axis=1.5)
points(df.1355$est[df.1355$p.value < p.1355 & df.1355$est > 0],
       -log10(df.1355$p.value[df.1355$p.value < p.1355 & df.1355$est > 0]),
```

```
col="red",
       pch=20)
points(df.1355$est[df.1355$p.value < p.1355 & df.1355$est < 0],
       -log10(df.1355$p.value[df.1355$p.value < p.1355 & df.1355$est < 0]),
col="green",
       pch=20)
table(df.1355$p.value < p.1355 & df.1355$est > 0)

##
## FALSE  TRUE
## 20244   305

table(df.1355$p.value < p.1355 & df.1355$est < 0)

##
## FALSE  TRUE
## 20259   290

legend("topleft", c("Upregulated:      305", "Downregulated: 290"),
       col=c("red", "green"), pch=20, bty="n")
```



5. venn diagramm for H1299 and H1355 intersected up and down regulated genes.

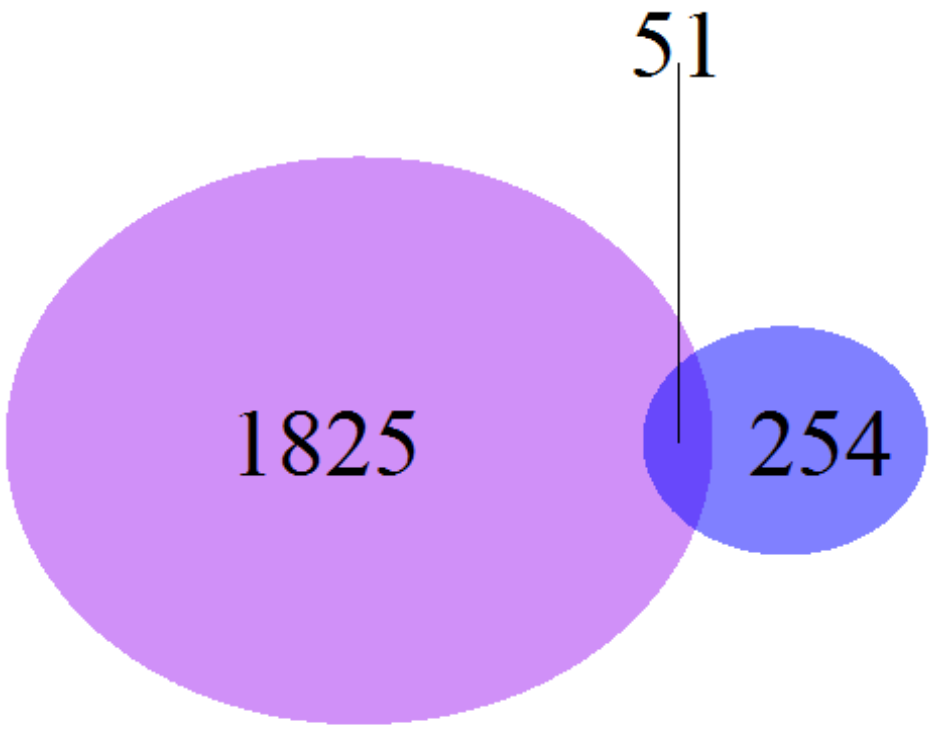
```
# upregulated
sum(df.1299$p.value < p.1299 & df.1299$est > 0)

## [1] 1876

sum(df.1355$p.value < p.1355 & df.1355$est > 0)
```

```
## [1] 305
```

```
sum((df.1299$p.value < p.1299 & df.1299$est > 0) &
      (df.1355$p.value < p.1355 & df.1355$est > 0))
```

```
## [1] 51
```

```
plot.new()
draw.pairwise.venn(1876, 305, 51, cex=3,
                   fill=c("purple", "blue"), lty="blank")
```




```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2],
polygon[GRID.polygon.3], polygon[GRID.polygon.4], text[GRID.text.5],
text[GRID.text.6], text[GRID.text.7], lines[GRID.lines.8], text[GRID.text.9],
text[GRID.text.10])
```

```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2],
polygon[GRID.polygon.3], polygon[GRID.polygon.4], text[GRID.text.5],
text[GRID.text.6], text[GRID.text.7], lines[GRID.lines.8], text[GRID.text.9],
text[GRID.text.10])
# down regulated
sum(df.1299$p.value < p.1299 & df.1299$est < 0)
```

```
## [1] 1876
```

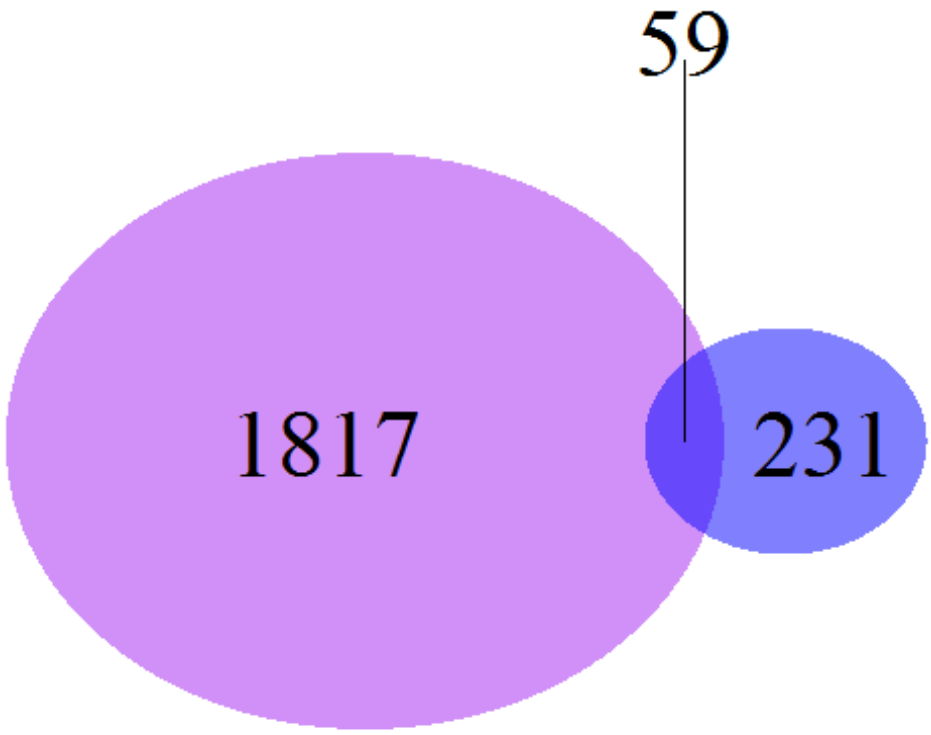
```
sum(df.1355$p.value < p.1355 & df.1355$est < 0)
```

```
## [1] 290
```

```
sum((df.1299$p.value < p.1299 & df.1299$est < 0) &
      (df.1355$p.value < p.1355 & df.1355$est < 0))
```

```
## [1] 59
```

```
plot.new()
draw.pairwise.venn(1876, 290, 59, cex=3,
                   fill=c("purple", "blue"), lty="blank")
```




```
## (polygon[GRID.polygon.11], polygon[GRID.polygon.12],
polygon[GRID.polygon.13], polygon[GRID.polygon.14], text[GRID.text.15],
text[GRID.text.16], text[GRID.text.17], lines[GRID.lines.18],
text[GRID.text.19], text[GRID.text.20])
```

```
## (polygon[GRID.polygon.11], polygon[GRID.polygon.12],
polygon[GRID.polygon.13], polygon[GRID.polygon.14], text[GRID.text.15],
text[GRID.text.16], text[GRID.text.17], lines[GRID.lines.18],
text[GRID.text.19], text[GRID.text.20])
```

Xenografts

1. load the xenograft data

```
load("xeno.RData")
head(df.xeno[,1:5])
```

```
##   gene.id n gene.symbol H1299.Parental.Cis+Doc.871
## 1       1 2        A1BG                   3.633839
## 2       2 1         A2M                   2.229475
```

```
## 3       9 3        NAT1                    3.012080
## 4      10 1        NAT2                    2.419627
## 5      12 1    SERPINA3                    5.481228
## 6      13 1       AADAC                    2.488485
##   H1299.Parental.Cis+Doc.873
## 1                   3.136163
## 2                   1.674401
## 3                   3.092054
## 4                   2.585305
## 5                   7.112472
## 6                   2.548242
```

```
dim(df.xeno)
```

```
## [1] 20549    15
```

```
colnames(df.xeno)
```

```
##  [1] "gene.id"                    "n"
##  [3] "gene.symbol"                "H1299.Parental.Cis+Doc.871"
##  [5] "H1299.Parental.Cis+Doc.873" "H1299.Parental.Cis+Doc.878"
##  [7] "H1299.Parental.Saline.872"  "H1299.Parental.Saline.877"
##  [9] "H1299.Parental.Saline.891"  "H1299.T18.Cis+Doc.868"
## [11] "H1299.T18.Cis+Doc.882"      "H1299.T18.Cis+Doc.889"
## [13] "H1299.T18.Saline.862"       "H1299.T18.Saline.870"
## [15] "H1299.T18.Saline.886"
```

```
xeno.line
```

```
##  [1] "H1299.Parental.Cis+Doc.871" "H1299.Parental.Cis+Doc.873"
##  [3] "H1299.Parental.Cis+Doc.878" "H1299.Parental.Saline.872"
##  [5] "H1299.Parental.Saline.877"  "H1299.Parental.Saline.891"
##  [7] "H1299.T18.Cis+Doc.868"      "H1299.T18.Cis+Doc.882"
##  [9] "H1299.T18.Cis+Doc.889"      "H1299.T18.Saline.862"
## [11] "H1299.T18.Saline.870"       "H1299.T18.Saline.886"
```

```
id.up=df.1299$gene.id[df.1299$p.value < p.1299 & df.1299$est > 0 &
                        df.1355$p.value < p.1355 & df.1355$est > 0]
id.down=df.1299$gene.id[df.1299$p.value < p.1299 & df.1299$est < 0 &
                          df.1355$p.value < p.1355 & df.1355$est < 0]
```

2. significant differental expression gene analysis (FDR 0.1)

```
xeno.untr=xeno.line[c(4:6, 10:12)]
xeno.untr
```

```
## [1] "H1299.Parental.Saline.872" "H1299.Parental.Saline.877"
## [3] "H1299.Parental.Saline.891" "H1299.T18.Saline.862"
## [5] "H1299.T18.Saline.870"      "H1299.T18.Saline.886"
```

```
head(df.xeno)
```

```
##   gene.id n gene.symbol H1299.Parental.Cis+Doc.871
## 1       1 2        A1BG                  3.633839
## 2       2 1         A2M                  2.229475
## 3       9 3        NAT1                  3.012080
## 4      10 1        NAT2                  2.419627
## 5      12 1    SERPINA3                  5.481228
## 6      13 1       AADAC                  2.488485
##   H1299.Parental.Cis+Doc.873 H1299.Parental.Cis+Doc.878
## 1                   3.136163                   2.745621
## 2                   1.674401                   1.610231
## 3                   3.092054                   2.863332
## 4                   2.585305                   2.481384
## 5                   7.112472                   5.059355
## 6                   2.548242                   3.361560
##   H1299.Parental.Saline.872 H1299.Parental.Saline.877
## 1                  3.536566                  2.910598
## 2                  2.307369                  1.766422
## 3                  3.045042                  2.973530
## 4                  2.596068                  3.350789
## 5                  2.822787                  6.649452
## 6                  3.526960                  2.935540
##   H1299.Parental.Saline.891 H1299.T18.Cis+Doc.868 H1299.T18.Cis+Doc.882
## 1                  3.209863              3.693003              3.064912
## 2                  1.827652              2.748023              2.080255
## 3                  3.317923              2.552347              2.659500
## 4                  2.897194              2.441063              2.553360
## 5                  3.574117             10.439048             11.141053
## 6                  2.858179              2.535265              2.235517
##   H1299.T18.Cis+Doc.889 H1299.T18.Saline.862 H1299.T18.Saline.870
## 1              3.453592             3.537467             3.438834
## 2              2.221931             1.940983             2.200645
## 3              2.819919             2.557727             2.305327
## 4              2.247488             2.308160             1.967581
## 5             11.685980            12.776207            12.126267
## 6              3.207945             2.885458             3.018145
##   H1299.T18.Saline.886
## 1             3.450917
## 2             1.727244
## 3             2.749388
## 4             2.017007
## 5            11.371214
## 6             2.583509
```

```
dat.validate=df.xeno[, c("gene.id", "n", "gene.symbol", xeno.untr)]
head(dat.validate)
```

```
##   gene.id n gene.symbol H1299.Parental.Saline.872
## 1       1 2        A1BG                  3.536566
## 2       2 1         A2M                  2.307369
## 3       9 3        NAT1                  3.045042
```

```
## 4      10 1        NAT2                  2.596068
## 5      12 1    SERPINA3                  2.822787
## 6      13 1       AADAC                  3.526960
##   H1299.Parental.Saline.877 H1299.Parental.Saline.891 H1299.T18.Saline.862
## 1                  2.910598                  3.209863             3.537467
## 2                  1.766422                  1.827652             1.940983
## 3                  2.973530                  3.317923             2.557727
## 4                  3.350789                  2.897194             2.308160
## 5                  6.649452                  3.574117            12.776207
## 6                  2.935540                  2.858179             2.885458
##   H1299.T18.Saline.870 H1299.T18.Saline.886
## 1             3.438834             3.450917
## 2             2.200645             1.727244
## 3             2.305327             2.749388
## 4             1.967581             2.017007
## 5            12.126267            11.371214
## 6             3.018145             2.583509
```

```
dat.validate$fold.change=apply(dat.validate[,7:9], 1, mean)-
apply(dat.validate[, 4:6], 1, mean)

xeno.p=function(x){
  x=as.numeric(x)
  return(t.test(x[1:3], x[4:6], alternative="two.sided")$p.value)
}
xeno.p(dat.validate[1, 4:9])
```

```
## [1] 0.2900682
```

```
dat.validate$p.value=apply(dat.validate[, 4:9], 1, xeno.p)
```

Gene signatures

1. Venn diagram up and down regulated genes for both cell lines and xenografts

```
#venn diagram
p.xeno=cutoffSignificant(Bum(dat.validate$p.value), 0.1)
p.xeno
```

```
## [1] 0.01136181
```

```
table(dat.validate$p.value < p.xeno)
```

```
##
## FALSE  TRUE
## 19178  1371
```

```
id.up.xeno=dat.validate$gene.id[dat.validate$p.value < p.xeno &
                                  dat.validate$fold.change > 0]
length(id.up)
```

```
## [1] 51
```

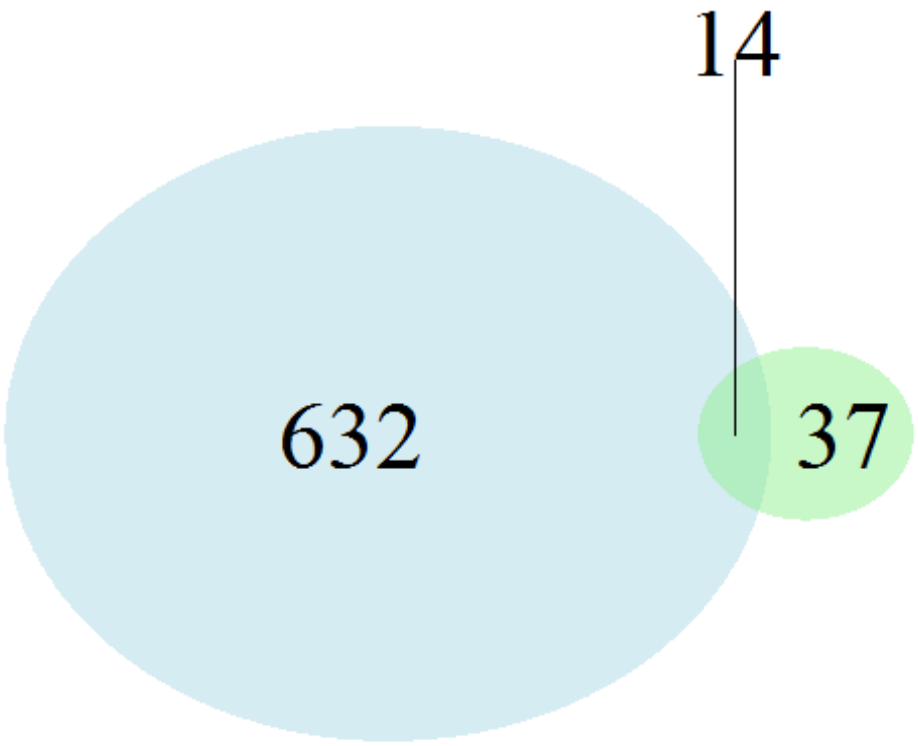
```
length(id.up.xeno)
```

```
## [1] 646
```

```
sum(id.up %in% id.up.xeno)
```

```
## [1] 14
```

```
plot.new()
draw.pairwise.venn(646, 51, 14, cex=3,
                   fill=c("light blue", "light green"), lty="blank")
```




```
## (polygon[GRID.polygon.21], polygon[GRID.polygon.22],
polygon[GRID.polygon.23], polygon[GRID.polygon.24], text[GRID.text.25],
text[GRID.text.26], text[GRID.text.27], lines[GRID.lines.28],
text[GRID.text.29], text[GRID.text.30])
```

```
## (polygon[GRID.polygon.21], polygon[GRID.polygon.22],
polygon[GRID.polygon.23], polygon[GRID.polygon.24], text[GRID.text.25],
text[GRID.text.26], text[GRID.text.27], lines[GRID.lines.28],
text[GRID.text.29], text[GRID.text.30])
```

```
id.down.xeno=dat.validate$gene.id[dat.validate$p.value < p.xeno &
                                  dat.validate$fold.change < 0]
length(id.down)
```

```
## [1] 59
```

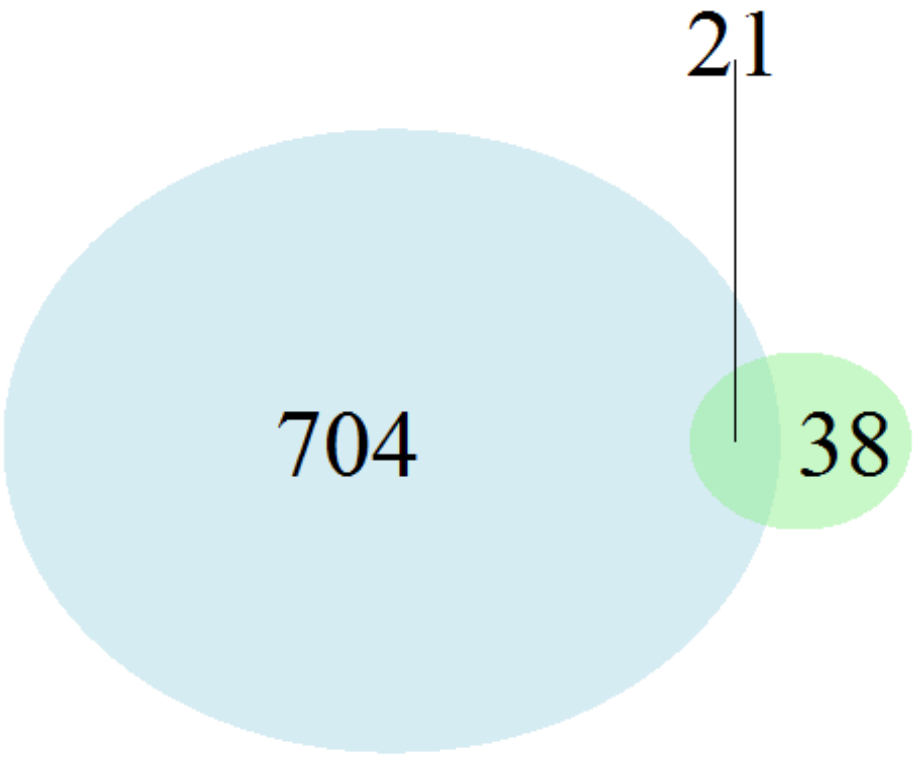
```
length(id.down.xeno)
```

```
## [1] 725
```

```
sum(id.down %in% id.down.xeno)
```

```
## [1] 21
```

```
plot.new()
draw.pairwise.venn(725, 59, 21, cex=3,
                   fill=c("light blue", "light green"), lty="blank")
```




```
## (polygon[GRID.polygon.31], polygon[GRID.polygon.32],
polygon[GRID.polygon.33], polygon[GRID.polygon.34], text[GRID.text.35],
text[GRID.text.36], text[GRID.text.37], lines[GRID.lines.38],
text[GRID.text.39], text[GRID.text.40])
```

```
## (polygon[GRID.polygon.31], polygon[GRID.polygon.32],
polygon[GRID.polygon.33], polygon[GRID.polygon.34], text[GRID.text.35],
text[GRID.text.36], text[GRID.text.37], lines[GRID.lines.38],
text[GRID.text.39], text[GRID.text.40])
id.up.regulate=id.up[id.up %in% id.up.xeno]
id.down.regulate=id.down[id.down %in% id.down.xeno]
sig.gene=dat.validate[dat.validate$gene.id %in%
c(id.up.regulate,id.down.regulate),c("gene.id", "gene.symbol")]
gene35=as.character(sig.gene$gene.symbol)
up=dat.validate[dat.validate$gene.id %in% id.up.regulate,c("gene.id",
"gene.symbol")]
```

2.xenografts vocano plot

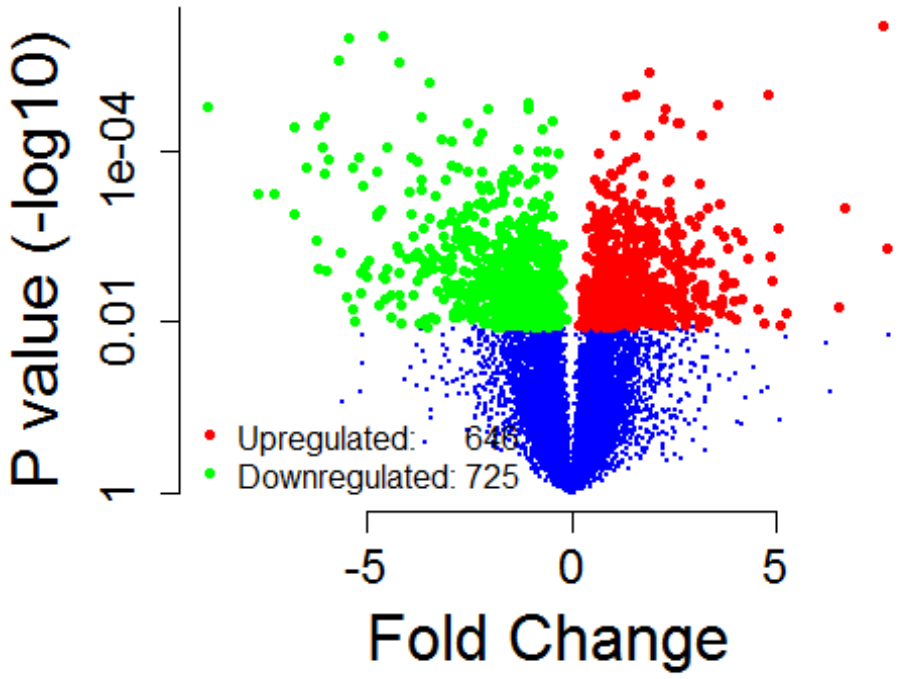
```r
par(mar=c(5, 5, 2, 1))
plot(dat.validate$fold.change, -log10(dat.validate$p.value), xlab="Fold
Change", cex=0.5,
     ylab="P value (-log10)", cex.lab=2, cex.axis=1.5, bty="n", col="blue",
pch=20,
     yaxt="n")
axis(2, at=c(0, 2, 4, 6, 8), labels=c(1, 0.01, 0.0001, 0.000001, 0.00000001),
cex.axis=1.5)
points(dat.validate$fold.change[dat.validate$p.value < p.xeno &
dat.validate$fold.change > 0],
       -log10(dat.validate$p.value[dat.validate$p.value < p.xeno &
dat.validate$fold.change > 0]), col="red",
       pch=20)
points(dat.validate$fold.change[dat.validate$p.value < p.xeno &
dat.validate$fold.change < 0],
       -log10(dat.validate$p.value[dat.validate$p.value < p.xeno &
dat.validate$fold.change < 0]), col="green",
       pch=20)
table(dat.validate$p.value < p.xeno & dat.validate$fold.change > 0)
```

```
##
## FALSE  TRUE
## 19903   646
```

```r
table(dat.validate$p.value < p.xeno & dat.validate$fold.change < 0)
```

```
##
## FALSE  TRUE
## 19824   725
```

```r
legend("bottomleft", c("Upregulated:      646", "Downregulated: 725"),
       col=c("red", "green"), pch=20, bty="n")
```

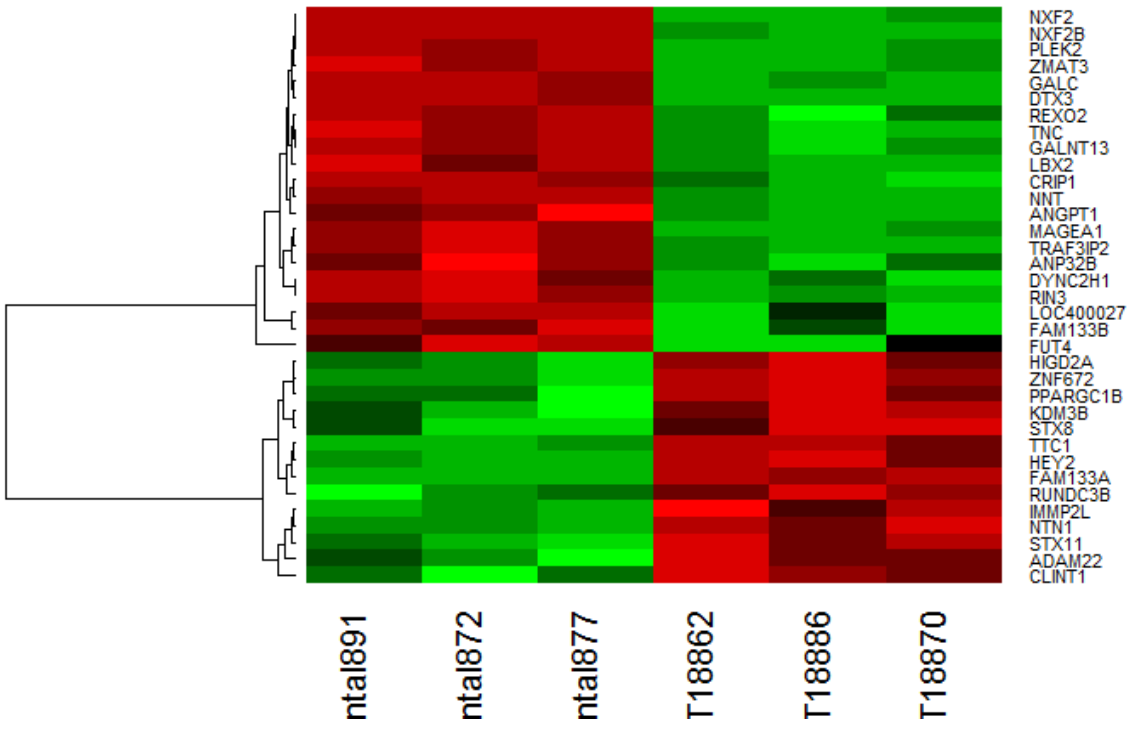


3. heatmaps for 35 significance gene signature in Cell lines H1355,H1299 and Xenografts

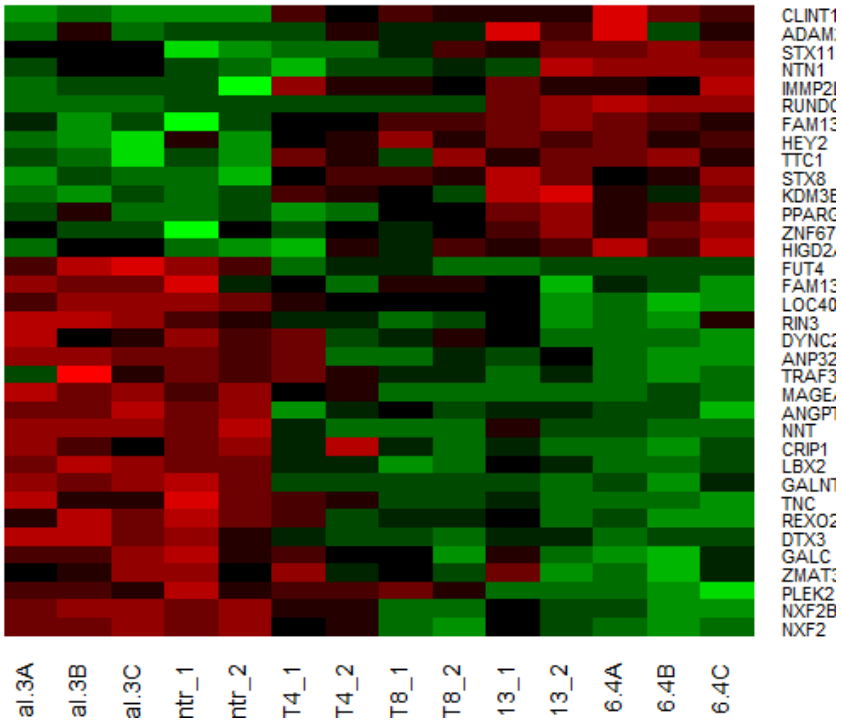
```
# load 35 genes signature

xeno <- read.table("xenografts.txt",sep="\t",head=TRUE)
CL <- read.table("cellLines.txt",sep="\t",head=TRUE)


# In xenografts, select probes with largest absolute Fold change to represent
the genes expression
xeno35 <- xeno[xeno$Symbol %in% gene35,]
dat <- NULL

for (id in unique(xeno35$Symbol)){
  tmp <- xeno35[xeno35$Symbol %in% id,]
  if (nrow(tmp)>1){

    mm =
tmp[tmp$H1299.T18.vs.H1299.P==max(abs(tmp$H1299.T18.vs.H1299.P))|tmp$H1299.T1
8.vs.H1299.P==-max(abs(tmp$H1299.T18.vs.H1299.P)) ,]
    dat <- rbind(dat,mm)

  } else {
    dat=rbind(dat,tmp)

  }
```

```
}
xe <- dat[,10:21]
rownames(xe) <- dat$Symbol
xe=as.matrix(xe)
xe=xe[,c(1,2,3,7,8,9)]


#################################################################
# In cell lines, select probes with largest absolute Fold change to represent
the genes expression

# cell lines
CL35 <- CL[CL$Symbol %in% gene35,]
CL35.H1355 <-
data.frame(Symbol=CL35$Symbol,ProbeID=CL35$Probe.ID,CL35[,grep("H1355",colnam
es(CL35))])
CL35.H1299 <-
data.frame(Symbol=CL35$Symbol,ProbeID=CL35$Probe.ID,CL35[,grep("H1299",colnam
es(CL35))])
H1355.dat <- NULL

for (id in unique(CL35.H1355$Symbol)){
  tmp <- CL35.H1355[CL35.H1355$Symbol %in% id,]
  if (nrow(tmp)>1){

    mm =
tmp[tmp$H1355.T16.vs.H1355.P==max(abs(tmp$H1355.T16.vs.H1355.P))|tmp$H1355.T1
6.vs.H1355.P==-max(abs(tmp$H1355.T16.vs.H1355.P)) ,]
    H1355.dat <- rbind(H1355.dat,mm)

  } else {
    H1355.dat=rbind(H1355.dat,tmp)

  }
}

H1355.cb <- as.matrix(H1355.dat[,4:17])
rownames(H1355.cb) <- as.character(H1355.dat$Symbol)


H1299.dat <- NULL

for (id in unique(CL35.H1299$Symbol)){
  tmp <- CL35.H1299[CL35.H1299$Symbol %in% id,]
  if (nrow(tmp)>1){

    mm =
tmp[tmp$H1299.T18.vs.H1299.P==max(abs(tmp$H1299.T18.vs.H1299.P))|tmp$H1299.T1
8.vs.H1299.P==-max(abs(tmp$H1299.T18.vs.H1299.P)) ,]
```

```
    H1299.dat <- rbind(H1299.dat,mm)

  } else {
    H1299.dat=rbind(H1299.dat,tmp)

  }
}
H1299.cb <- as.matrix(H1299.dat[,4:17])
rownames(H1299.cb) <- H1299.dat$Symbol


# heatmaps

mm=heatmap.2(xe, col=greenred, scale="row",distfun=function(x) {as.dist(1-
cor(t(x)))},tracecol=NULL,,Colv=FALSE,key=FALSE)

## Warning in heatmap.2(xe, col = greenred, scale = "row", distfun =
## function(x) {: Discrepancy: Colv is FALSE, while dendrogram is `row'.
## Omitting column dendogram.
```

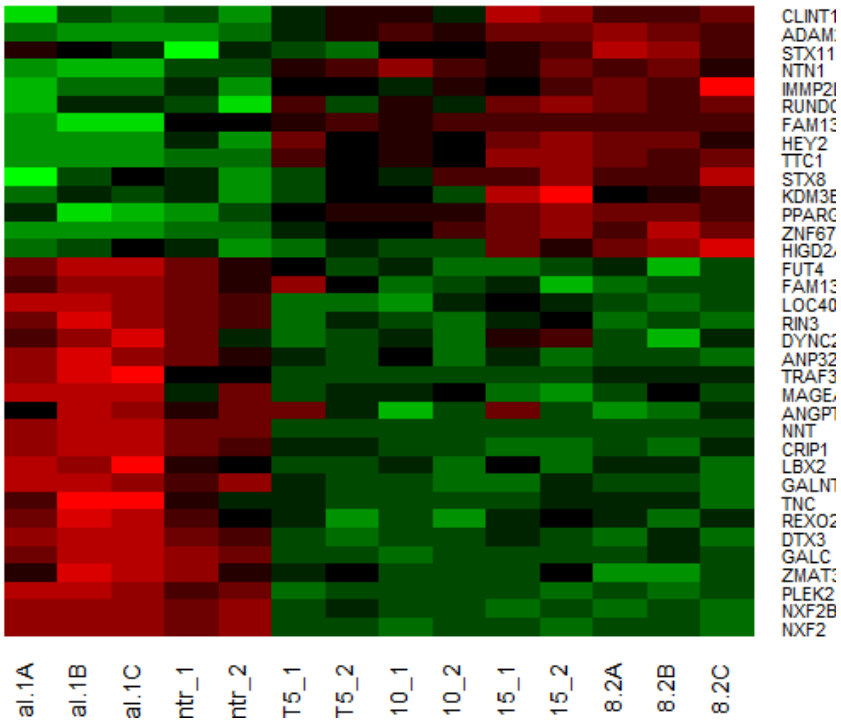


```
## Warning in heatmap.2(xe, col = greenred, scale = "row", distfun =
## function(x) {: Discrepancy: Colv is FALSE, while dendrogram is `row'.
## Omitting column dendogram.

go=rownames(xe)[mm$rowInd] # keep the gene rows in the same order as
xenografts

heatmap.2(H1355.cb[go,], col=greenred,
margin=c(3,3),scale="row",Rowv=FALSE,Colv=FALSE,tracecol=NULL,key=FALSE)

## Warning in heatmap.2(H1355.cb[go, ], col = greenred, margin = c(3, 3),
## scale = "row", : Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
```




```
## Warning in heatmap.2(H1355.cb[go, ], col = greenred, margin = c(3, 3),
## scale = "row", : Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
heatmap.2(H1299.cb[go,], col=greenred,
margin=c(3,3),scale="row",Rowv=FALSE,Colv=FALSE,tracecol=NULL,key=FALSE)
```

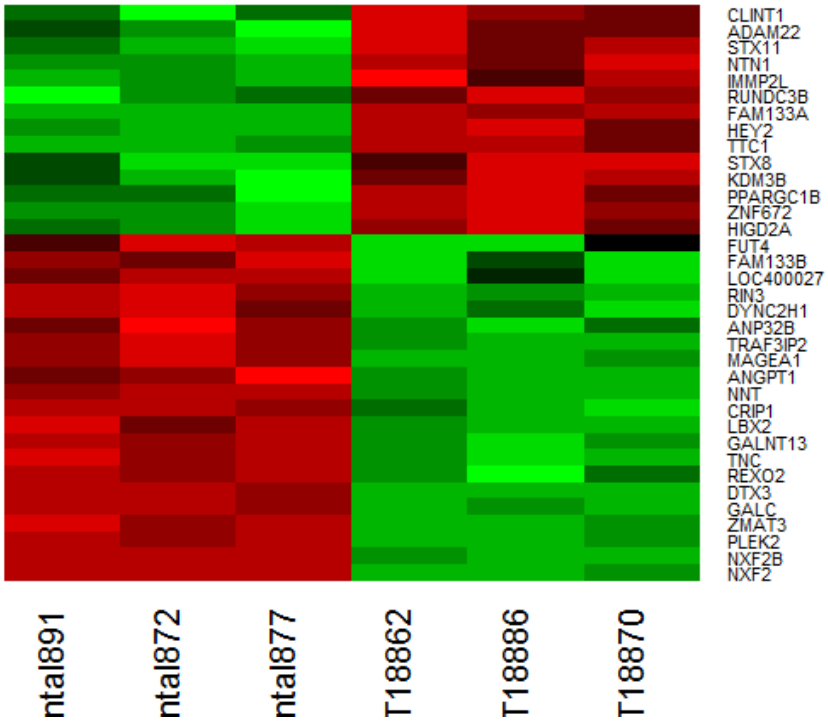
```
## Warning in heatmap.2(H1299.cb[go, ], col = greenred, margin = c(3, 3),
## scale = "row", : Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
```




```
## Warning in heatmap.2(H1299.cb[go, ], col = greenred, margin = c(3, 3),
## scale = "row", : Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
```


```
heatmap.2(xe[go,], col=greenred, scale="row",distfun=function(x) {as.dist(1-
cor(t(x)))},tracecol=NULL,Colv=FALSE,Rowv=FALSE,key=FALSE)
```

```
## Warning in heatmap.2(xe[go, ], col = greenred, scale = "row", distfun =
## function(x) {: Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
```




```
## Warning in heatmap.2(xe[go, ], col = greenred, scale = "row", distfun =
## function(x) {: Discrepancy: Rowv is FALSE, while dendrogram is `none'.
## Omitting row dendogram.
```

Gene signatures in patient data

Unsupervise learning for 35 gene signautures on 65 patients data who got the neoadjuvant therapy. 1. load the data

```
load("patient expr.RData")

names(dat.pati)[4:278] <- sub("\\.", "-", names(dat.pati[4:278]))

# load clinical data
dat.clin <- read.csv("03-28-2014_Dalvi M_65 Neoadjuvant patients with UPDATED
annotation.csv", as.is=T)
dat.clin$id %in% names(dat.pati)
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [15] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [29] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
## [43] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## [57] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

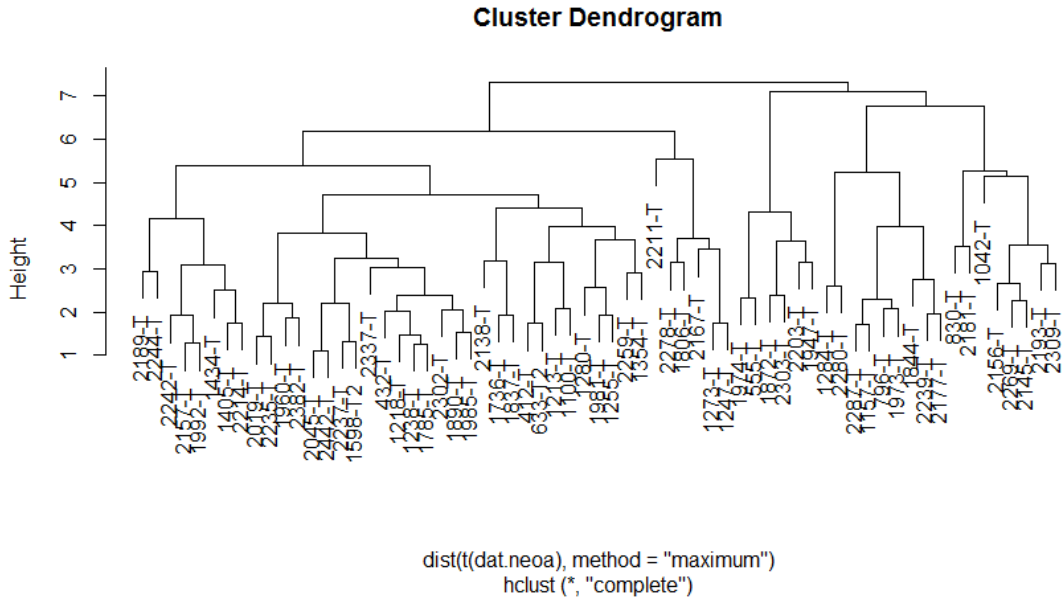
```
dat.neoa <- dat.pati[, c(names(dat.pati[1:3]), dat.clin$id)]
id <- c(id.up.regulate, id.down.regulate)

dat.neoa <- dat.neoa[dat.neoa$gene.id %in% id, ]
row.names(dat.neoa) <- dat.neoa$gene.id
row.names(dat.neoa)
```

```
##  [1] "284"    "1396"   "2526"   "2581"   "3371"   "4100"   "7265"
##  [8] "8676"   "9423"   "9482"   "9685"   "10541"  "10758"  "23493"
## [15] "23530"  "25996"  "26499"  "51780"  "53616"  "56001"  "64393"
## [22] "79659"  "79890"  "79894"  "83943"  "85474"  "114805" "133522"
## [29] "154661" "192286" "196403" "257415" "286499" "400027" "728343"
```

```
dat.neoa <- dat.neoa[, 4:68]
```

2. cluster plot and K-M curve recurrence free survival analysis

```
# cluster plot
hc.cell=hclust(dist(t(dat.neoa), method="maximum"))
plot(hc.cell)
```




```
hc.cell$order
```

```
##  [1] 30 52 32  3 50 12 15 29 37 38 10 22  4 19  5  8 53 18 21  2 45  1  7
## [24] 44 14 11 17 16 20 43 46 39 34 36  6  9 49 13 33 47 48 51 57 65 28 54
## [47] 42 55 27 61 25 58 40 63 56 59 64 31 41 62 60 26 35 23 24
```

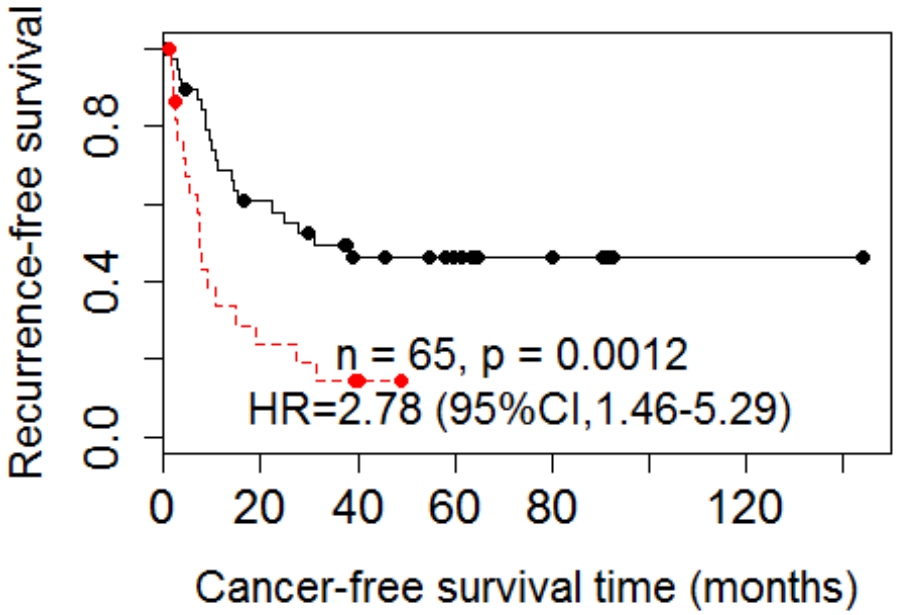
```
dat.surv <- data.frame(id=hc.cell$labels[hc.cell$order], group=c(rep(1, 42),
rep(2, 23)))
dat.surv <- merge(dat.surv, dat.clin, by="id")
```

```r
# K-M plot
kmplot <- function(survival,groups, title.lab="",xlab="",ylab="",
               survivallimit=c(60, 120), display=TRUE, cex.axis=1.5,
cex.lab=1.4, cex.main=1.5,
               mar=c(5.1 , 5.3, 4.1, 1.1), sig=NULL, ...)
{
  require(survival)
  survival<-survival[!is.na(groups),]
  groups<-groups[!is.na(groups)]
  if(length(levels(factor(groups)))<2)
  { cat("error in kmplot\n"); return() }
  logrank<-survdiff(survival ~ groups,  ...)
  pv <- pchisq(logrank$chisq,1, lower.tail=F)

  summary_coxph <- summary(coxph(survival ~ groups, ...))
  ci <-summary_coxph$conf.int


  col=c("black", "red")
  if (display) {
      sfit= survfit(survival ~ groups, ...)

      plot(sfit, col=col, lty=1:2, main=title.lab, xlab=xlab,ylab=ylab,
mark.time=TRUE,mark=19,
         cex.axis=cex.axis, cex.lab=cex.lab, cex.main=cex.main, mar=mar, ...)

      ### add two vertical line represent 5 year and 10 year
#      sapply(survivallimit, function(x) abline(v=x, col="grey"))

       ### add results on plot
      stat=paste("n = ",length(groups),", ",pv.expr(pv) , "\n
HR=",format(ci[1],digits=3),
                 " (95%CI,",format(ci[3],digits=3),"-",
                 format(ci[4],digits=3),")",sep="")
      x=min(survival[,1])+0.5*(max(survival[,1])-min(survival[,1]))
      text(x, 0.15 ,stat , cex=cex.lab)
  }
  return(list(group_table=table(groups),logrank.p=pv,
              hr=ci[1],hr.5=ci[3],hr.95=ci[4], n=length(groups),
              ebeta=summary_coxph$coef[1],
z=summary_coxph$coef[4],pr.z=summary_coxph$coef[5],
              groups=groups))
}

         ##Function to format pvalues in K-M plot
pv.expr <- function(x, digits = 2) {
    if (!x) return(0)
    exponent <- floor(log10(x))
    base <- round(x / 10^exponent, digits)
```

```
    ifelse(x > 0.0001,
        paste("p = ", base*(10^exponent), sep=""),
        paste("p = ", base, "E", exponent, sep=""))
}


survival <-
Surv(time=as.numeric(as.character(dat.surv$cancer.free.survival.month)),
              event=dat.surv$recurrence == 'Y')
survdiff(survival ~ dat.surv$group)
```

```
## Call:
## survdiff(formula = survival ~ dat.surv$group)
##
##                     N Observed Expected (O-E)^2/E (O-E)^2/V
## dat.surv$group=1 42        20    28.56      2.56      10.5
## dat.surv$group=2 23        18     9.44      7.75      10.5
##
##  Chisq= 10.5  on 1 degrees of freedom, p= 0.0012
```

```
kmplot(survival, dat.surv$group,xlab="Cancer-free survival time
(months)",ylab="Recurrence-free survival")
```




```
## $group_table
## groups
##  1  2
## 42 23
```

```
##
## $logrank.p
## [1] 0.001200184
##
## $hr
## [1] 2.776687
##
## $hr.5
## [1] 1.458712
##
## $hr.95
## [1] 5.28548
##
## $n
## [1] 65
##
## $ebeta
## [1] 1.021258
##
## $z
## [1] 3.109546
##
## $pr.z
## [1] 0.001873753
##
## $groups
##  [1] 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 2 1 2 2 1 1 1 1 1 2 2 1
## [36] 1 2 2 1 2 2 1 1 1 1 2 1 1 1 2 1 2 2 1 2 2 1 1 1 1 2 1 2 2
```

3. Multivariate cox regression model

```
# data for 65 neoajuvant patients
load("dat65.RData")
head(dat65[,1:5])
```

```
##   gene.id n gene.symbol    2302-T    1238-T
## 1       1 2        A1BG  3.638638  2.605372
## 2       2 1         A2M 10.282485 11.270556
## 3       9 3        NAT1  3.201879  3.562783
## 4      10 1        NAT2  2.183924  2.484070
## 5      12 1    SERPINA3 10.435610  9.637462
## 6      13 1       AADAC  8.201044  3.566616
```

```
dim(dat65)
```

```
## [1] 19579    68
```

```
cli65=dat.clin
sig.id <- c(id.up.regulate, id.down.regulate)
sur65 <- dat65[dat65$gene.id %in% sig.id,]
row.names(sur65) <- sur65$gene.symbol
sur65 <- sur65[, -c(1:3)]
```

```r
sur65 <- data.frame(t(sur65))
head(sur65)
```

```
##          ANGPT1    CRIP1     FUT4     GALC      TNC   MAGEA1     TTC1
## 2302-T 8.235965 10.73370 5.422461 7.486111 8.190325 2.634824 8.216957
## 1238-T 7.010597 11.35512 5.125421 7.166048 8.693179 2.708898 7.949370
## 2157-T 5.381803 10.67481 4.150018 6.272858 7.075845 2.896832 8.054993
## 2045-T 7.963517 11.84460 4.616963 6.794380 7.304523 2.662581 7.866941
## 2237-T 7.489364 12.19631 5.336181 7.265428 7.057829 2.768827 8.055649
## 2259-T 6.587808 12.08331 4.885976 6.528196 8.851127 3.828562 8.342618
##           STX11     NTN1     STX8   CLINT1   ANP32B TRAF3IP2     HEY2
## 2302-T 8.806033 3.866315 7.807953 9.964402 11.97116 5.748529 4.538813
## 1238-T 8.990390 3.934511 8.164496 9.348762 11.80851 5.224289 6.463278
## 2157-T 7.266282 4.691052 7.932429 8.789924 12.60085 5.632375 4.672881
## 2045-T 9.024002 3.782765 7.853039 9.588817 11.81897 4.804227 5.344224
## 2237-T 8.642615 3.650787 7.420597 9.801010 11.90997 5.340369 6.014473
## 2259-T 6.644821 3.998293 7.765663 9.865428 11.74342 5.591911 8.132118
##             NNT     REXO2    PLEK2    KDM3B   ADAM22     NXF2     ZMAT3
## 2302-T 4.551442 10.626628 6.297819 7.387870 2.473813 2.654559  9.680279
## 1238-T 4.751584 10.565241 6.575559 7.741217 2.452605 2.820190  9.842212
## 2157-T 5.521030 10.975039 4.895468 7.082318 2.813311 2.837278  9.086257
## 2045-T 5.020791  9.837021 3.924777 7.125601 2.421299 2.879944 10.220259
## 2237-T 4.159578 10.301812 4.765241 7.175600 2.352307 2.790665  9.797409
## 2259-T 4.410768 10.430360 5.056644 8.019820 2.826604 2.525510  9.548210
##         DYNC2H1     RIN3   ZNF672   IMMP2L     LBX2  GALNT13 PPARGC1B
## 2302-T 4.191597 4.099950 8.520529 6.935128 3.066943 2.871684 3.334735
## 1238-T 5.409916 5.833676 8.051448 6.965085 3.123558 2.592927 4.696052
## 2157-T 4.979832 4.487548 8.851127 7.713463 3.093184 3.093184 4.472508
## 2045-T 5.705767 5.067450 7.956891 6.827816 2.898155 2.682759 5.601410
## 2237-T 5.783142 5.163301 8.259766 6.386431 2.781739 2.855904 5.743148
## 2259-T 7.879810 3.347565 8.959878 6.676347 3.190928 3.695353 4.047072
##         RUNDC3B    HIGD2A     DTX3  FAM133B  FAM133A LOC400027    NXF2B
## 2302-T 4.391510  9.798051 5.078655 6.594635 2.419880  8.763619 2.653797
## 1238-T 3.793786 10.840024 6.443461 7.581591 3.507960  8.564467 2.406528
## 2157-T 2.499195 10.096499 5.335023 7.793973 2.466545  9.510188 2.593304
## 2045-T 5.482069 11.104236 5.138590 7.859911 2.350893  8.217731 2.534494
## 2237-T 3.759741 10.754040 6.324054 6.933605 3.132086  8.316771 2.501597
## 2259-T 6.139675 10.715144 6.606734 7.584588 2.957714  8.965241 2.644243
```

```r
sur65$id <- row.names(sur65)
sur65 <- merge(sur65, cli65, by="id")
sur65$event <- ifelse(sur65$recurrence == "Y", 1, 0)
paste(names(sur65)[2:36], collapse = " + ")
```

```
## [1] "ANGPT1 + CRIP1 + FUT4 + GALC + TNC + MAGEA1 + TTC1 + STX11 + NTN1 +
STX8 + CLINT1 + ANP32B + TRAF3IP2 + HEY2 + NNT + REXO2 + PLEK2 + KDM3B +
ADAM22 + NXF2 + ZMAT3 + DYNC2H1 + RIN3 + ZNF672 + IMMP2L + LBX2 + GALNT13 +
PPARGC1B + RUNDC3B + HIGD2A + DTX3 + FAM133B + FAM133A + LOC400027 + NXF2B"
```

```r
# multivariate cox regression model
fit65 <- coxph(Surv(cancer.free.survival.month, event)~ ANGPT1 + CRIP1 + FUT4
```

```
+ GALC + TNC + MAGEA1 + TTC1 + STX11 + NTN1 + STX8 + CLINT1 + ANP32B +
TRAF3IP2 + HEY2 + NNT + REXO2 + PLEK2 + KDM3B + ADAM22 + NXF2 + ZMAT3 +
DYNC2H1 + RIN3 + ZNF672 + IMMP2L + LBX2 + GALNT13 + PPARGC1B + RUNDC3B +
HIGD2A + DTX3 + FAM133B + FAM133A + LOC400027 + NXF2B, data=sur65)
sum65 <- summary(fit65)
coe65=sum65$coefficients
coe65
```

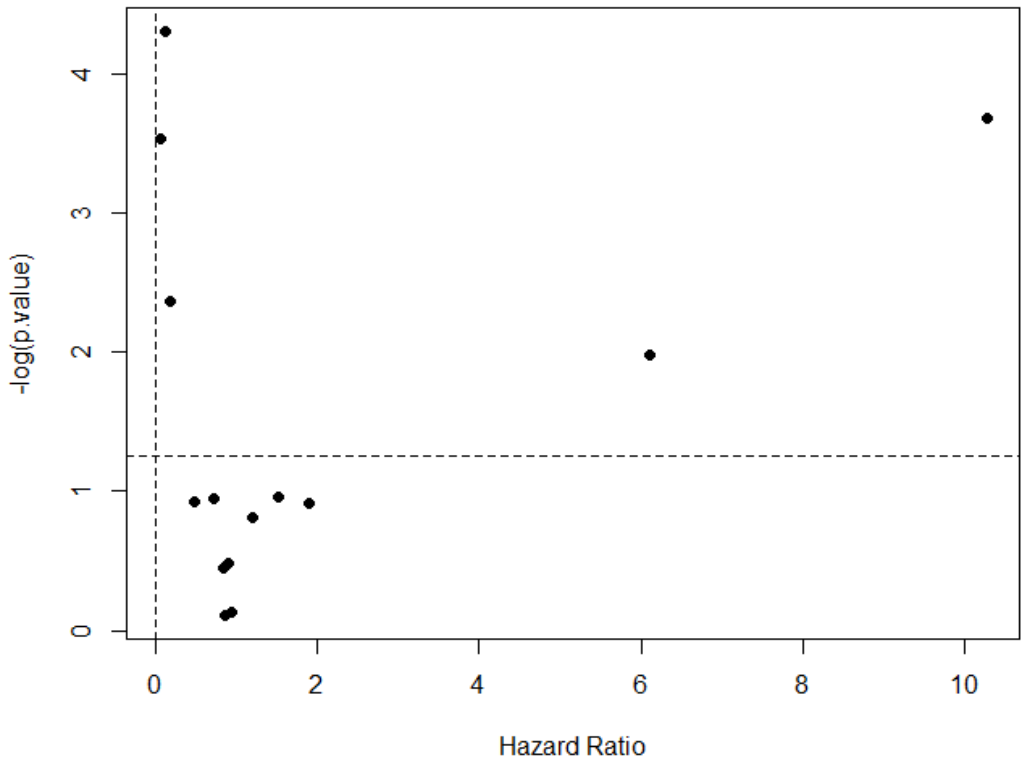
```
##                coef   exp(coef)  se(coef)          z     Pr(>|z|)
## ANGPT1     0.2091745  1.23266002 0.3522183  0.5938772 0.5525942229
## CRIP1      0.4827728  1.62056161 0.4083533  1.1822427 0.2371093988
## FUT4       0.6953699  2.00445029 0.8489477  0.8190962 0.4127315352
## GALC       0.1018298  1.10719499 0.6486271  0.1569928 0.8752505211
## TNC        0.4542830  1.57504372 0.2696067  1.6849841 0.0919916316
## MAGEA1     0.3877544  1.47366785 0.2178626  1.7798117 0.0751067903
## TTC1      -0.7462477  0.47414233 0.8774816 -0.8504426 0.3950790600
## STX11     -0.0624412  0.93946831 0.4083143 -0.1529243 0.8784579357
## NTN1       0.4174976  1.51815770 0.4800938  0.8696166 0.3845099402
## STX8      -2.1243707  0.11950815 0.8603591 -2.4691674 0.0135427840
## CLINT1    -2.9858164  0.05049826 1.3702253 -2.1790697 0.0293264880
## ANP32B    -0.8004983  0.44910512 0.7571715 -1.0572219 0.2904103310
## TRAF3IP2   1.3593645  3.89371789 0.8955198  1.5179613 0.1290241587
## HEY2      -0.1132994  0.89288327 0.2281820 -0.4965310 0.6195198115
## NNT        3.0160368 20.41024063 0.8867630  3.4011757 0.0006709668
## REXO2      0.8232530  2.27789773 1.0772530  0.7642150 0.4447391043
## PLEK2     -0.0854132  0.91813283 0.2689104 -0.3176270 0.7507679273
## KDM3B      2.3300661 10.27862073 1.0422492  2.2356132 0.0253771209
## ADAM22     1.8088556  6.10345840 1.2200918  1.4825569 0.1381921728
## NXF2      -1.9577895  0.14117013 1.3088808 -1.4957738 0.1347126236
## ZMAT3     -0.2423971  0.78474446 0.8924421 -0.2716111 0.7859210966
## DYNC2H1   -0.7245371  0.48454882 0.4513348 -1.6053205 0.1084232849
## RIN3       0.1752646  1.19156148 0.4063743  0.4312886 0.6662585233
## ZNF672    -1.7224168  0.17863391 1.0290374 -1.6738135 0.0941672569
## IMMP2L     0.6351963  1.88739254 0.7552601  0.8410299 0.4003311966
## LBX2       0.6945857  2.00287903 0.3475644  1.9984375 0.0456692455
## GALNT13    0.5186821  1.67981236 0.3777974  1.3729106 0.1697801307
## PPARGC1B  -0.3403402  0.71152822 0.3936462 -0.8645839 0.3872672093
## RUNDC3B   -0.1737039  0.84054572 0.3701324 -0.4693021 0.6388537108
## HIGD2A    -0.1515406  0.85938303 1.1451527 -0.1323322 0.8947215570
## DTX3       0.8800273  2.41096560 0.3218440  2.7343286 0.0062507638
## FAM133B   -1.6762593  0.18707245 1.3213149 -1.2686297 0.2045731611
## FAM133A    0.1851519  1.20340119 0.2420717  0.7648639 0.4443526181
## LOC400027 -0.3161092  0.72897986 0.5113653 -0.6181670 0.5364652793
## NXF2B      2.7053611 14.95971778 1.6524065  1.6372249 0.1015834781
```

```
dm=coe65[as.character(up$gene.symbol),]
dim(dm)
```

```
## [1] 14  5
```

14 up regulated genes

```
plot(dm[,2],-log(dm[,5]),type="p",pch=19,xlab="Hazard Ratio",ylab="-
log(p.value)")
abline(h=1.25,v=0,lty=2)
```




```
dat.surv$neoadj <- ifelse(dat.surv$Neoadjuvant.Drugs %in% c("Cisplatin
Docetaxel", "Carboplatin Paclitaxel", "Carboplatin Docetaxel", "Cisplatin"),
1, 0)

table(dat.surv$neoadj)

##
##  0  1
## 10 55

dat.surv$path[grep("IA", dat.surv$pathology)] <- "I"
dat.surv$path[grep("IB", dat.surv$pathology)] <- "I"
dat.surv$path[grep("IIA", dat.surv$pathology)] <- "II"
dat.surv$path[grep("IIB", dat.surv$pathology)] <- "II"
dat.surv$path[grep("IIIA", dat.surv$pathology)] <- "III"
dat.surv$path[grep("IIIB", dat.surv$pathology)] <- "III"
dat.surv$path[grep("IV", dat.surv$pathology)] <- "IV"
```

```r
coxph(Surv(cancer.free.survival.month, recurrence == 'Y') ~ group,
data=dat.surv)
```

```
## Call:
## coxph(formula = Surv(cancer.free.survival.month, recurrence ==
##     "Y") ~ group, data = dat.surv)
##
##
##         coef exp(coef) se(coef)    z      p
## group 1.021     2.777    0.328 3.11 0.0019
##
## Likelihood ratio test=9.15  on 1 df, p=0.00249
## n= 65, number of events= 38
```

```r
cox.fit <- coxph(Surv(cancer.free.survival.month, recurrence == 'Y') ~ group
+ histology + age + smoke + Gender + Race + Adjuvant.Therapy + neoadj + path,
data=dat.surv)

cox.fit
```

```
## Call:
## coxph(formula = Surv(cancer.free.survival.month, recurrence ==
##     "Y") ~ group + histology + age + smoke + Gender + Race +
##     Adjuvant.Therapy + neoadj + path, data = dat.surv)
##
##
##                                   coef exp(coef) se(coef)     z      p
## group                           1.6292    5.0997   0.4858  3.35 0.0008
## histologyOther                  0.3710    1.4492   0.4745  0.78 0.4343
## histologySquamous              -0.2292    0.7952   0.4969 -0.46 0.6446
## age                             0.0222    1.0224   0.0251  0.88 0.3765
## smokeY                         -0.9275    0.3955   0.6676 -1.39 0.1647
## GenderM                        -0.2108    0.8099   0.4312 -0.49 0.6249
## RaceAsian or Pacific Islander -0.2788    0.7567   1.5221 -0.18 0.8547
## RaceCaucasian                  -0.8713    0.4184   0.8005 -1.09 0.2764
## RaceHispanic                   -0.1321    0.8762   1.2877 -0.10 0.9183
## Adjuvant.TherapyY              -1.0292    0.3573   0.4962 -2.07 0.0380
## neoadj                          0.5252    1.6909   0.5182  1.01 0.3107
## pathII                         -0.2370    0.7890   0.5893 -0.40 0.6875
## pathIII                         1.0031    2.7268   0.4973  2.02 0.0437
## pathIV                          0.9530    2.5935   0.6674  1.43 0.1533
##
## Likelihood ratio test=23.2  on 14 df, p=0.0566
## n= 65, number of events= 38
```