SIS NMF Final: Diagnosis to DSD

March 21, 2015

1 Preparation

```
options(java.parameters = "-Xmx4G")

library(survival)
library(energy)
library(NMF)
library(nnls)

library(bnlearn)

library(glmulti)
library(glmnet)

library(RColorBrewer)
library(gplots)

library(xtable)
library(stargazer)

load("image.rda")
```

2 Cohort characteristics

```
cpvs.diag_dsd$Path.TumourLocation[cpvs.diag_dsd$Path.TumourLocation == ""] = NA
cpvs.diag_dsd$Path.Nodes.Regional.Involved.Fraction = cpvs.diag_dsd$Path.Nodes.Regional.Involved / cpvs.
cpvs.diag_dsd$Treat.Surgery.ExcisionStatus.Coarse = ordered(ifelse(cpvs.diag_dsd$Treat.Surgery.ExcisionS
cpvs.diag_dsd$Path.Grade.Coarse = ordered(ifelse(cpvs.diag_dsd$Path.Grade %in% c("1", "2"), "1or2", "3or
cpvs.diag_dsd$Path.TumourLocation.Coarse = factor(ifelse(cpvs.diag_dsd$Path.TumourLocation %in% c("Head"

summary(cpvs.diag_dsd)

##   Patient.ID        Patient.Gender                      Patient.Ethnicity
##  Length:110         Female:50      Asian                         :  5
##  Class :character   Male  :60      Asian, White/Caucasian        :  0
##  Mode  :character                  Black/African                 :  0
##                                    Black/African, White/Caucasian:  0
##                                    White/Caucasian               :104
##                                    NA's                          :  1
```

```
##
##                   Patient.Country History.LastFollowup.Date
##  Australia               :110    Min.   :2007-06-29
##  Italy                   :  0    1st Qu.:2011-08-19
##  New Zealand             :  0    Median :2013-03-12
##  Puerto Rico             :  0    Mean   :2012-10-16
##  United Kingdom          :  0    3rd Qu.:2014-04-24
##  United States of America:  0    Max.   :2014-09-23
##                                  NA's   :1
##  History.Smoking.PackYears History.Diagnosis.Date
##  Min.   : 0.75             Min.   :2007-06-04
##  1st Qu.: 9.00             1st Qu.:2010-01-28
##  Median :22.50             Median :2011-01-04
##  Mean   :26.89             Mean   :2011-01-14
##  3rd Qu.:43.75             3rd Qu.:2012-02-15
##  Max.   :70.00             Max.   :2012-10-17
##  NA's   :68
##  History.Diagnosis.AgeAtYears History.Surgery.Date
##  Min.   :36.0                 Min.   :2007-05-29
##  1st Qu.:61.0                 1st Qu.:2010-01-22
##  Median :67.0                 Median :2011-01-01
##  Mean   :66.4                 Mean   :2011-01-13
##  3rd Qu.:73.0                 3rd Qu.:2012-02-13
##  Max.   :87.0                 Max.   :2012-10-17
##
##                                            Treat.Surgery.Procedure
##  Classic Whipple                                        :79
##  Splenectomy, Subtotal Panc/L sided Panc or distal Panc: 6
##  Cholecystectomy, Classic Whipple                       : 5
##  Subtotal Panc/L sided Panc or distal Panc              : 4
##  Classic Whipple, Exploratory laparotomy                : 3
##  PPPD                                                   : 3
##  (Other)                                                :10
##  Treat.Surgery.ExcisionStatus Treat.Surgery.Margin.Pancreatic
##  R0:69                        <2 mm   : 4
##  R1:35                        Clear   :88
##  R2: 6                        Involved: 9
##                               NA's    : 9
##
##
##
##  Treat.Surgery.MarginSizeMm.Pancreatic Treat.Surgery.Margin.Periunc
##  Min.   : 0.0                          <2 mm   :20
##  1st Qu.: 5.0                          Clear   :52
##  Median :10.0                          Involved:15
##  Mean   :10.6                          NA's    :23
##  3rd Qu.:10.2
##  Max.   :40.0
##  NA's   :30
##  Treat.Surgery.MarginSizeMm.Periunc Treat.Surgery.Margin.PVGroove
##  Min.   : 0.00                      <2 mm   :23
##  1st Qu.: 1.00                      Clear   :55
##  Median : 3.00                      Involved:12
##  Mean   : 6.21                      NA's    :20
```

```
##  3rd Qu.:10.00
##  Max.   :40.00
##  NA's   :43
##  Treat.Surgery.MarginSizeMm.PVGroove Treat.Surgery.Margin.Retrop
##  Min.   : 0.00                       <2 mm   :21
##  1st Qu.: 1.00                       Clear   :68
##  Median : 3.00                       Involved: 9
##  Mean   : 4.08                       NA's    :12
##  3rd Qu.: 5.00
##  Max.   :30.00
##  NA's   :45
##  Treat.Surgery.MarginSizeMm.Retrop Treat.Surgery.Margin.CBD
##  Min.   : 0.10                     <2 mm   : 1
##  1st Qu.: 1.75                     Clear   :83
##  Median : 3.00                     Involved: 0
##  Mean   : 5.62                     NA's    :26
##  3rd Qu.:10.00
##  Max.   :25.00
##  NA's   :31
##  Treat.Surgery.MarginSizeMm.CBD Treat.Surgery.Margin.Duodenal
##  Min.   : 1.0                   Clear   :60
##  1st Qu.:11.8                   Involved: 1
##  Median :20.0                   NA's    :49
##  Mean   :23.6
##  3rd Qu.:32.5
##  Max.   :55.0
##  NA's   :47
##  Treat.Surgery.MarginSizeMm.Duodenal Treat.Surgery.Margin.Gastric
##  Min.   : 10.0                       Clear:59
##  1st Qu.: 40.0                       NA's :51
##  Median : 80.0
##  Mean   : 86.2
##  3rd Qu.:132.5
##  Max.   :190.0
##  NA's   :102
##  Treat.Surgery.MarginSizeMm.Gastric Treat.Surgery.Margin.Comments
##  Min.   : 10.0                      Length:110
##  1st Qu.: 50.0                      Class :character
##  Median : 70.0                      Mode  :character
##  Mean   : 67.9
##  3rd Qu.: 97.5
##  Max.   :100.0
##  NA's   :103
##                            Path.HistoType
##  Pancreatic Ductal Adenocarcinoma:110
##  Acinar Cell Carcinoma           :  0
##  Ampullary Adenocarcinoma        :  0
##  Carcinoid Tumour                :  0
##  Cholangiocarcinoma              :  0
##  Clear Cell Carcinoma            :  0
##  (Other)                         :  0
##                     Path.HistoType.Subtype Path.Grade
##  Gastric                       : 0         1: 8
##  Intestinal                    : 0         2:71
```

```
##  Mixed                        : 0         3:30
##  Not otherwise Specified (NOS):31         4: 1
##  Pancreatobiliary             :13
##  Squamous                     : 0
##  NA's                         :66
##       Path.TumourLocation Path.TumourSizeMm Path.Invasion.PN
##  Head           :83       Min.   :10.0      Absent :13
##  Head (Uncinate):10       1st Qu.:28.0      Present:96
##  Tail           : 9       Median :35.0      NA's   : 1
##  Body           : 7       Mean   :37.6
##                 : 0       3rd Qu.:45.0
##  (Other)        : 0       Max.   :90.0
##  NA's           : 1       NA's   :1
##  Path.Invasion.VS Path.Nodes.Regional.Total Path.Nodes.Regional.Involved
##  Absent :34       Min.   : 0.0              Min.   : 0.00
##  Present:72       1st Qu.:11.0              1st Qu.: 1.00
##  NA's   : 4       Median :16.0              Median : 2.00
##                   Mean   :18.1              Mean   : 3.18
##                   3rd Qu.:24.0              3rd Qu.: 4.00
##                   Max.   :46.0              Max.   :18.00
##
##  Path.Nodes.SepRec.Total Path.Nodes.SepRec.Involved
##  Min.   : 0.0            Min.   : 0.00
##  1st Qu.:11.0            1st Qu.: 1.00
##  Median :16.0            Median : 2.00
##  Mean   :18.1            Mean   : 3.18
##  3rd Qu.:24.0            3rd Qu.: 4.00
##  Max.   :46.0            Max.   :18.00
##
##                                    Staging.Version Staging.pM Staging.pN
##  pTNM AJCC 6th Ed 2002                     :14     M0  :  2   N0  :25
##  pTNM AJCC 7th Ed 2010                     :96     M1  :  6   N1  :84
##  pTNM AJCC 7th Ed 2010 (Ampulla)           : 0     NA's:102   NA's: 1
##  pTNM AJCC 7th Ed 2010 (Cholangiocarcinoma): 0
##  pTNM AJCC 7th Ed 2010 (Neuroendocrine)    : 0
##
##
##  Staging.pT Staging.Stage    History.Recurrence History.Recurrence.Date
##  Tis :  0   IA : 0       Not observed:24       Min.   :2007-10-14
##  T1  :  0   IB : 3       Suspected   : 4       1st Qu.:2010-12-11
##  T2  :  6   IIA:20       Confirmed   :78       Median :2012-02-22
##  T3  :102   IIB:80       NA's        : 4       Mean   :2012-01-21
##  T4  :  1   III: 1                             3rd Qu.:2012-12-29
##  NA's:  1   IV : 6                             Max.   :2014-08-27
##                                                NA's   :29
##  History.Recurrence.Site.Stomach History.Recurrence.Site.Peritoneum
##  Mode :logical                   Mode :logical
##  FALSE:110                       FALSE:94
##  NA's :0                         TRUE :16
##                                  NA's :0
##
##
##
##  History.Recurrence.Site.PancRemnant History.Recurrence.Site.PancBed
```

```
##  Mode :logical                       Mode :logical
##  FALSE:106                           FALSE:91
##  TRUE :4                             TRUE :19
##  NA's :0                             NA's :0
##
##
##
##  History.Recurrence.Site.Other History.Recurrence.Site.Omentum
##  Mode :logical                 Mode :logical
##  FALSE:102                     FALSE:109
##  TRUE :8                       TRUE :1
##  NA's :0                       NA's :0
##
##
##
##  History.Recurrence.Site.Mesentery History.Recurrence.Site.LymphNodes
##  Mode :logical                     Mode :logical
##  FALSE:108                         FALSE:88
##  TRUE :2                           TRUE :22
##  NA's :0                           NA's :0
##
##
##
##  History.Recurrence.Site.Lung History.Recurrence.Site.Liver
##  Mode :logical                Mode :logical
##  FALSE:88                     FALSE:72
##  TRUE :22                     TRUE :38
##  NA's :0                      NA's :0
##
##
##
##  History.Recurrence.Site.Brain History.Recurrence.Site.Bone
##  Mode :logical                 Mode :logical
##  FALSE:109                     FALSE:104
##  TRUE :1                       TRUE :6
##  NA's :0                       NA's :0
##
##
##
##                      History.Status History.Death.Date
##  Alive - With Disease       :15     Min.   :2007-11-21
##  Alive - Without Disease    :22     1st Qu.:2011-01-14
##  Deceased - Of Disease      :70     Median :2012-03-07
##  Deceased - Of Other Cause  : 3     Mean   :2012-02-21
##  Deceased - Of Unknown Cause: 0     3rd Qu.:2013-03-17
##                                     Max.   :2014-06-17
##                                     NA's   :37
##                        History.Death.Cause Surv.Event.Death
##  Cancer Death (Pancreatic)       :69       Min.   :0.000
##  Cancer Death (Other) - Lung ca  : 1       1st Qu.:0.000
##  Died of Treatment Complication  : 1       Median :1.000
##  Other (please specify)          : 1       Mean   :0.664
##  Other (please specify) - Suicide: 1       3rd Qu.:1.000
##  (Other)                         : 0       Max.   :1.000
```

```
##  NA's                             :37
##  Surv.EventTimeFromDiag.Death Surv.EventTimeFromSurg.Death
##  Min.   :  36                 Min.   :  36
##  1st Qu.: 402                 1st Qu.: 406
##  Median : 632                 Median : 634
##  Mean   : 674                 Mean   : 676
##  3rd Qu.: 912                 3rd Qu.: 917
##  Max.   :1778                 Max.   :1779
##
##  Surv.EventTimeFromRec.Death Surv.Event.DSDeath
##  Min.   :   7               Min.   :0.000
##  1st Qu.:  68               1st Qu.:0.000
##  Median : 183               Median :1.000
##  Mean   : 250               Mean   :0.636
##  3rd Qu.: 338               3rd Qu.:1.000
##  Max.   :1333               Max.   :1.000
##  NA's   :29
##  Surv.EventTimeFromDiag.DSDeath Surv.EventTimeFromSurg.DSDeath
##  Min.   :  36                   Min.   :  36
##  1st Qu.: 402                   1st Qu.: 406
##  Median : 632                   Median : 634
##  Mean   : 673                   Mean   : 675
##  3rd Qu.: 912                   3rd Qu.: 917
##  Max.   :1778                   Max.   :1779
##
##  Surv.EventTimeFromRec.DSDeath Surv.Event.Recurrence
##  Min.   :   7                  Min.   :0.000
##  1st Qu.:  68                  1st Qu.:0.000
##  Median : 183                  Median :1.000
##  Mean   : 250                  Mean   :0.736
##  3rd Qu.: 338                  3rd Qu.:1.000
##  Max.   :1333                  Max.   :1.000
##  NA's   :29                    NA's   :4
##  Surv.EventTimeFromDiag.Recurrence Surv.EventTimeFromSurg.Recurrence
##  Min.   :  34                      Min.   :  34
##  1st Qu.: 240                      1st Qu.: 240
##  Median : 392                      Median : 398
##  Mean   : 511                      Mean   : 512
##  3rd Qu.: 697                      3rd Qu.: 699
##  Max.   :1778                      Max.   :1779
##  NA's   :6                         NA's   :6
##  Path.Nodes.Regional.Involved.Fraction Treat.Surgery.ExcisionStatus.Coarse
##  Min.   :0.0000                        Clear   :69
##  1st Qu.:0.0435                        Involved:41
##  Median :0.1667
##  Mean   :0.2026
##  3rd Qu.:0.2727
##  Max.   :1.0000
##  NA's   :1
##  Path.Grade.Coarse Path.TumourLocation.Coarse
##  1or2:79           Head :93
##  3or4:31           Other:17
##
##
```

```
## 
## 
## 
```

```r
sort(apply(is.na(cpvs.diag_dsd), 2, sum))
```

```
##                            Patient.ID
##                                     0
##                        Patient.Gender
##                                     0
##                       Patient.Country
##                                     0
##                History.Diagnosis.Date
##                                     0
##          History.Diagnosis.AgeAtYears
##                                     0
##                  History.Surgery.Date
##                                     0
##               Treat.Surgery.Procedure
##                                     0
##          Treat.Surgery.ExcisionStatus
##                                     0
##         Treat.Surgery.Margin.Comments
##                                     0
##                       Path.HistoType
##                                     0
##                           Path.Grade
##                                     0
##             Path.Nodes.Regional.Total
##                                     0
##          Path.Nodes.Regional.Involved
##                                     0
##               Path.Nodes.SepRec.Total
##                                     0
##            Path.Nodes.SepRec.Involved
##                                     0
##                       Staging.Version
##                                     0
##                         Staging.Stage
##                                     0
##       History.Recurrence.Site.Stomach
##                                     0
##    History.Recurrence.Site.Peritoneum
##                                     0
##   History.Recurrence.Site.PancRemnant
##                                     0
##       History.Recurrence.Site.PancBed
##                                     0
##         History.Recurrence.Site.Other
##                                     0
##       History.Recurrence.Site.Omentum
##                                     0
##     History.Recurrence.Site.Mesentery
##                                     0
```

```
##    History.Recurrence.Site.LymphNodes
##                                     0
##          History.Recurrence.Site.Lung
##                                     0
##         History.Recurrence.Site.Liver
##                                     0
##         History.Recurrence.Site.Brain
##                                     0
##          History.Recurrence.Site.Bone
##                                     0
##                        History.Status
##                                     0
##                      Surv.Event.Death
##                                     0
##          Surv.EventTimeFromDiag.Death
##                                     0
##          Surv.EventTimeFromSurg.Death
##                                     0
##                    Surv.Event.DSDeath
##                                     0
##        Surv.EventTimeFromDiag.DSDeath
##                                     0
##        Surv.EventTimeFromSurg.DSDeath
##                                     0
##   Treat.Surgery.ExcisionStatus.Coarse
##                                     0
##                     Path.Grade.Coarse
##                                     0
##            Path.TumourLocation.Coarse
##                                     0
##                     Patient.Ethnicity
##                                     1
##              History.LastFollowup.Date
##                                     1
##                   Path.TumourLocation
##                                     1
##                     Path.TumourSizeMm
##                                     1
##                      Path.Invasion.PN
##                                     1
##                            Staging.pN
##                                     1
##                            Staging.pT
##                                     1
## Path.Nodes.Regional.Involved.Fraction
##                                     1
##                      Path.Invasion.VS
##                                     4
##                    History.Recurrence
##                                     4
##                 Surv.Event.Recurrence
##                                     4
##     Surv.EventTimeFromDiag.Recurrence
##                                     6
```

```
##     Surv.EventTimeFromSurg.Recurrence
##                                     6
##       Treat.Surgery.Margin.Pancreatic
##                                     9
##           Treat.Surgery.Margin.Retrop
##                                    12
##         Treat.Surgery.Margin.PVGroove
##                                    20
##          Treat.Surgery.Margin.Periunc
##                                    23
##              Treat.Surgery.Margin.CBD
##                                    26
##               History.Recurrence.Date
##                                    29
##           Surv.EventTimeFromRec.Death
##                                    29
##         Surv.EventTimeFromRec.DSDeath
##                                    29
## Treat.Surgery.MarginSizeMm.Pancreatic
##                                    30
##     Treat.Surgery.MarginSizeMm.Retrop
##                                    31
##                    History.Death.Date
##                                    37
##                   History.Death.Cause
##                                    37
##    Treat.Surgery.MarginSizeMm.Periunc
##                                    43
##   Treat.Surgery.MarginSizeMm.PVGroove
##                                    45
##        Treat.Surgery.MarginSizeMm.CBD
##                                    47
##         Treat.Surgery.Margin.Duodenal
##                                    49
##          Treat.Surgery.Margin.Gastric
##                                    51
##               Path.HistoType.Subtype
##                                    66
##             History.Smoking.PackYears
##                                    68
##   Treat.Surgery.MarginSizeMm.Duodenal
##                                   102
##                            Staging.pM
##                                   102
##    Treat.Surgery.MarginSizeMm.Gastric
##                                   103
```

3 Probe selection

```
table(cpss.sis$sel)

##
```

```
## FALSE  TRUE
## 12639   361
```

```r
mean(cpss.sis$sel)
```

```
## [1] 0.02777
```

```r
apply(cpss.sis.permuted, 2, sum)
```

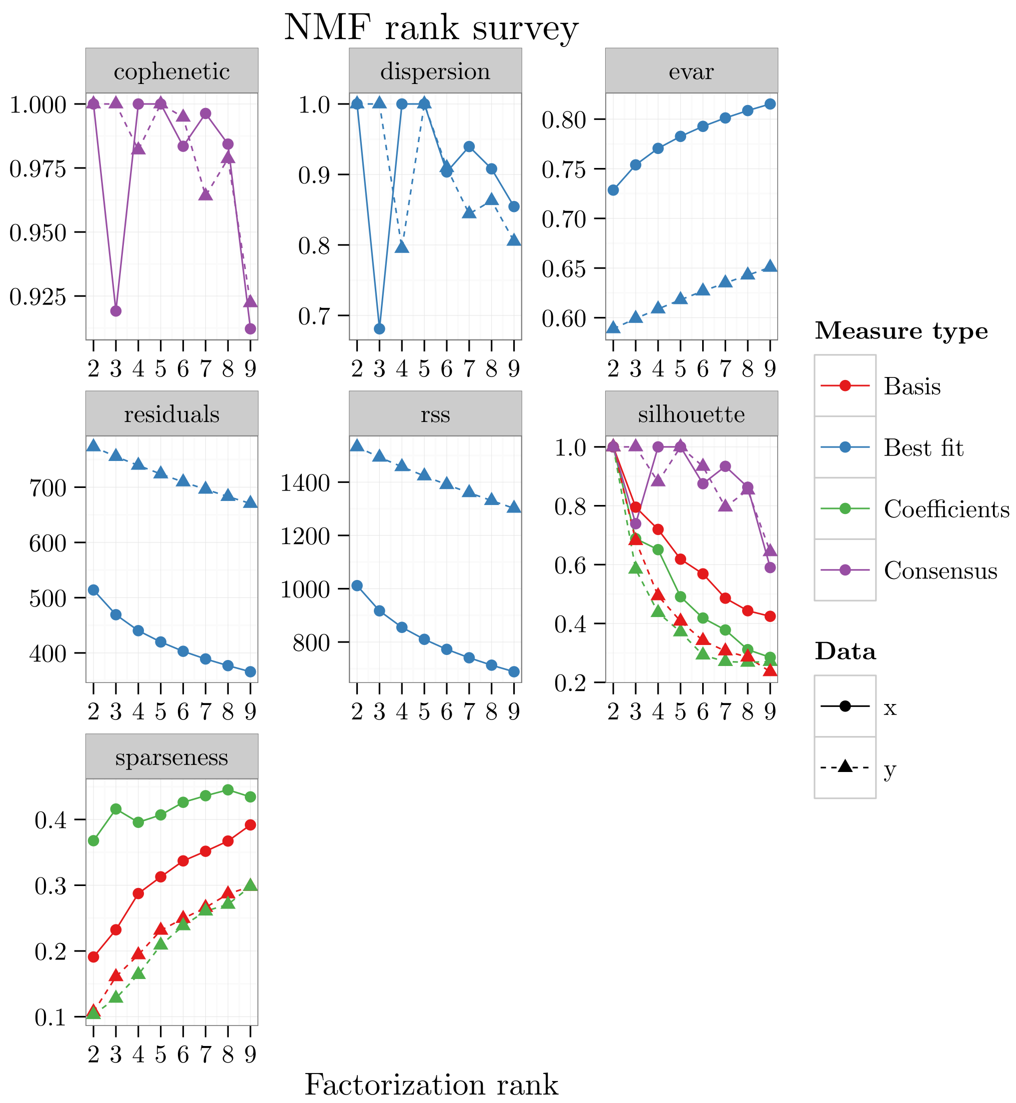
```
##  [1]  37 175  92  32 298  49  47 138  43 173  98  86 207 102 147  41  28
## [18] 160  75 273 154 124 415 109  41 141  50  63 107  63  64 237  84  52
## [35]  40 203  88  55  98  87  57 231  54  48  81 186 114  43  58 347
```

```r
median(apply(cpss.sis.permuted, 2, sum))
```

```
## [1] 87.5
```

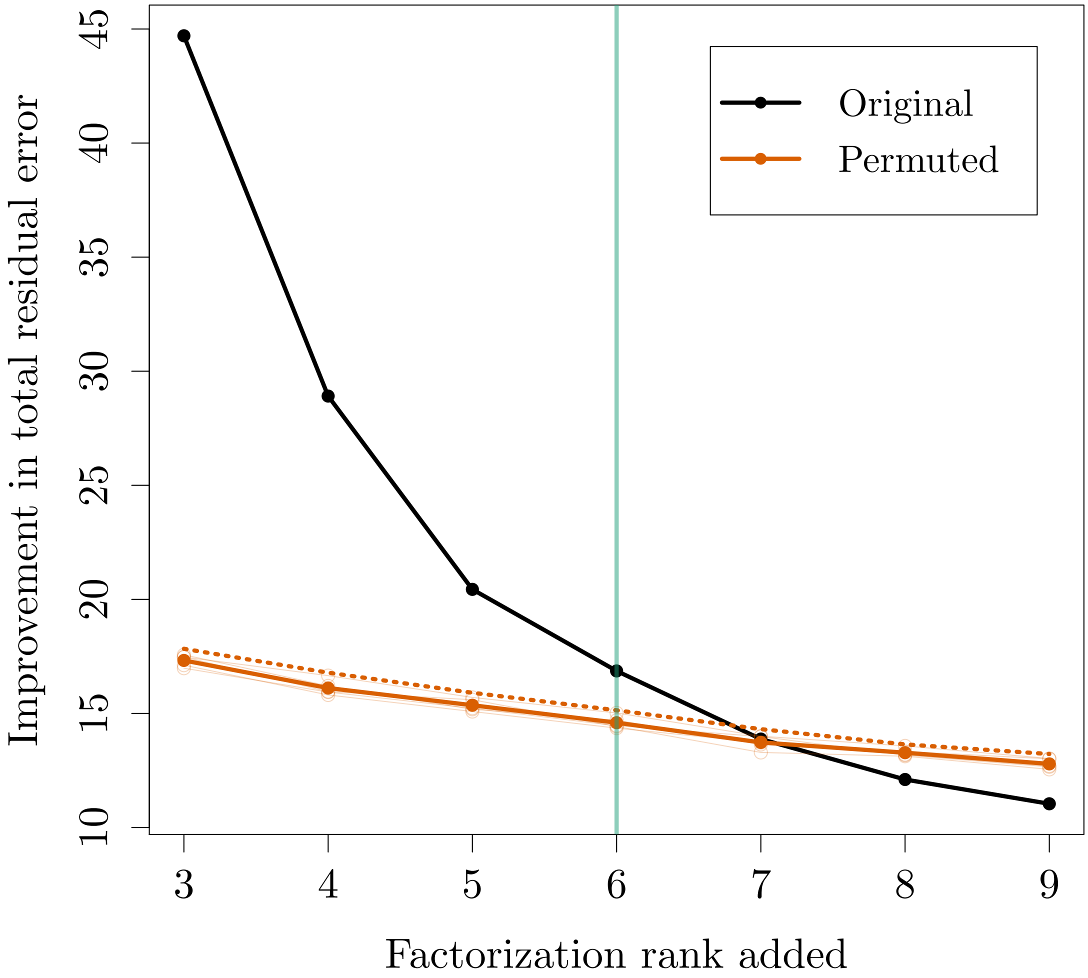
4 Factorization

```r
temp.pal = brewer.pal(3, "Dark2")
plot(nmf.runs.rank, nmf.runs.rank.random[[1]])
```



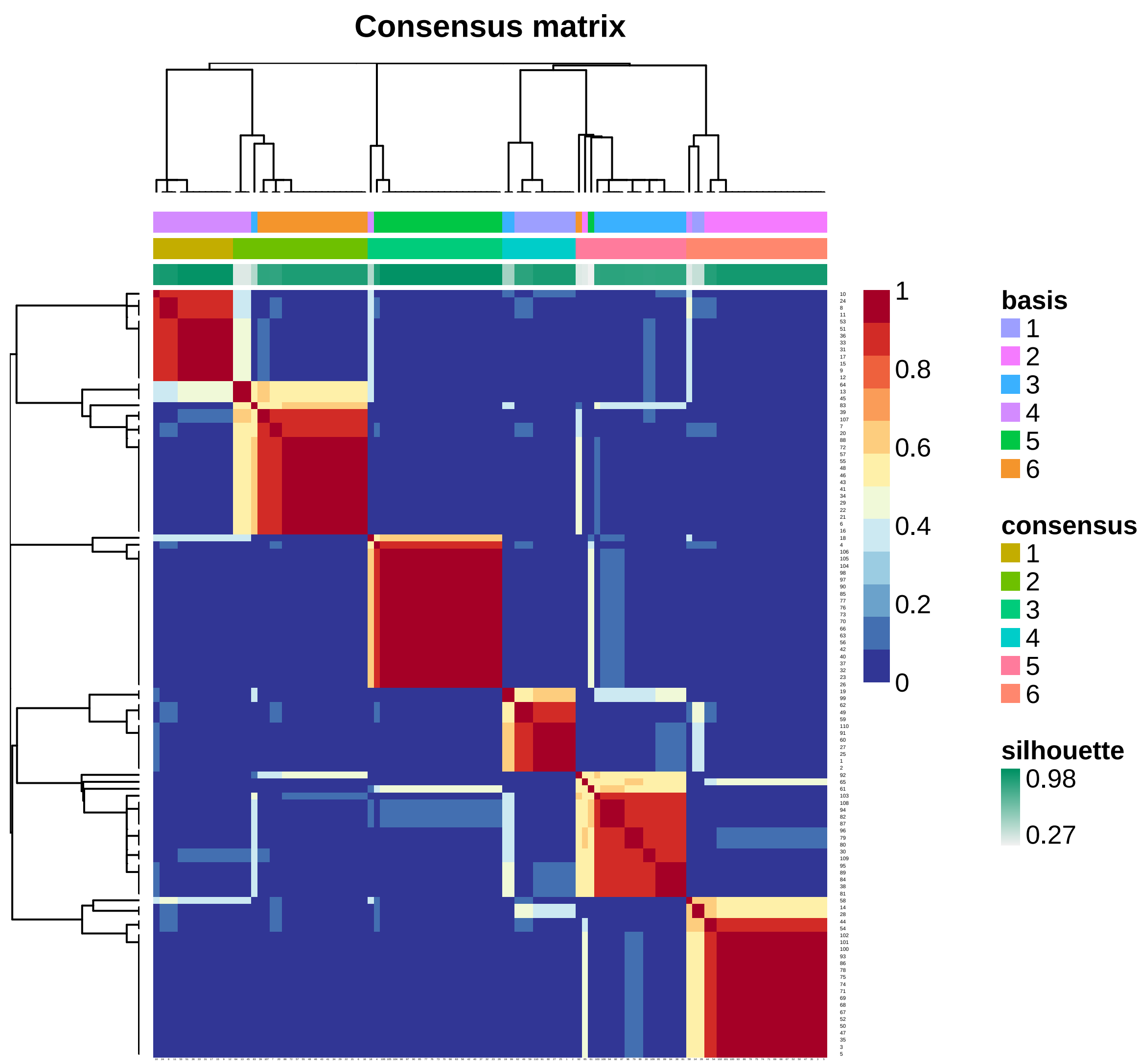

```
plot(nmf.rankrange[-1], -temp.orig_resids.delta,
        type = "o", col = "black", pch = 16, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.de
        xlab = "Factorization rank added", ylab = "Improvement in total residual error", lwd = 4)
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean, col = temp.pal[2], type = "o", pch = 16, lwd = 4)
for (i in 1:ncol(temp.perm_resids))
{
        lines(nmf.rankrange[-1], -temp.perm_resids.delta[,i], type = "o", col = do.call(rgb, as.list(c(c
}
lines(nmf.rankrange[-1], -temp.perm_resids.delta.threshold, col = temp.pal[2], lty = "dotted", lwd = 4)
if (nmf.rank.wasauto == TRUE)
{
        temp.col = temp.pal[1]
} else {
        temp.col = temp.pal[3]
}
abline(v = nmf.rank, col = do.call(rgb, as.list(c(col2rgb(temp.col)[,1]/255, alpha = 0.5))), lwd = 4)
#legend("topright", legend = c("Original data", "Permuted data", sprintf("Selected rank (%s)", ifelse(nm
```

```
legend("topright", legend = c("Original", "Permuted"), col = c("black", temp.pal[2]),  lty = "solid", po
```



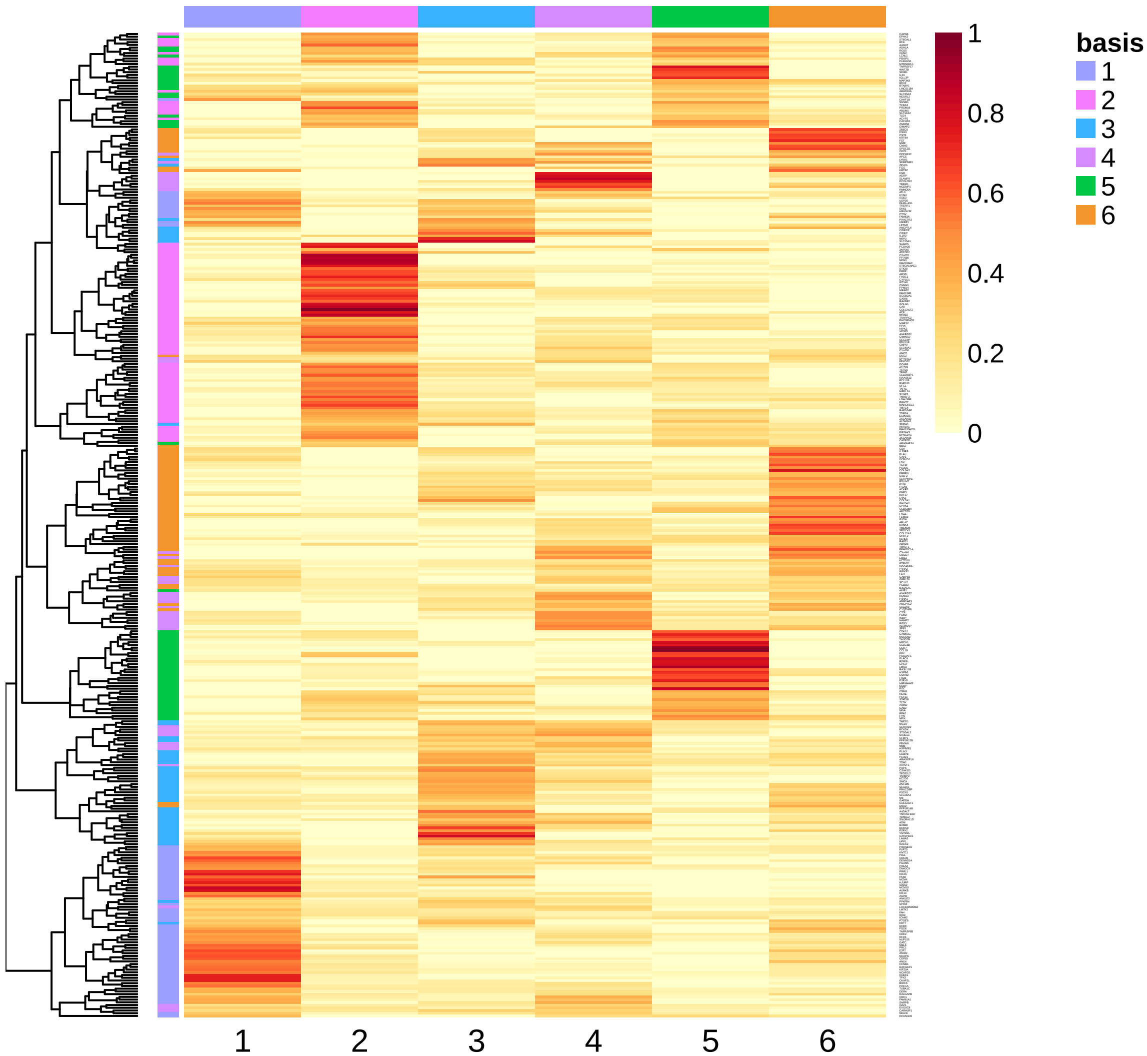
4.1 Fit

```
consensusmap(nmf.final)
```

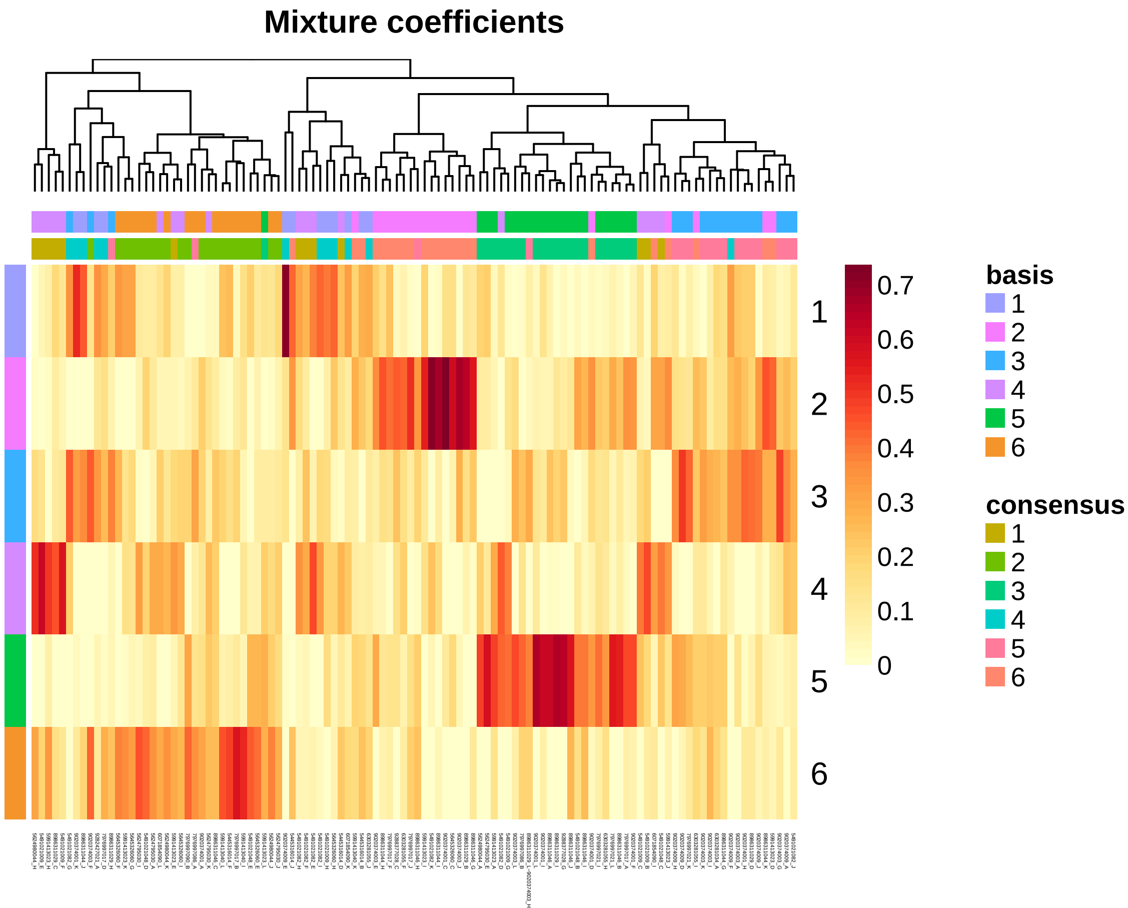



```r
basismap(nmf.final)
```

Basis components



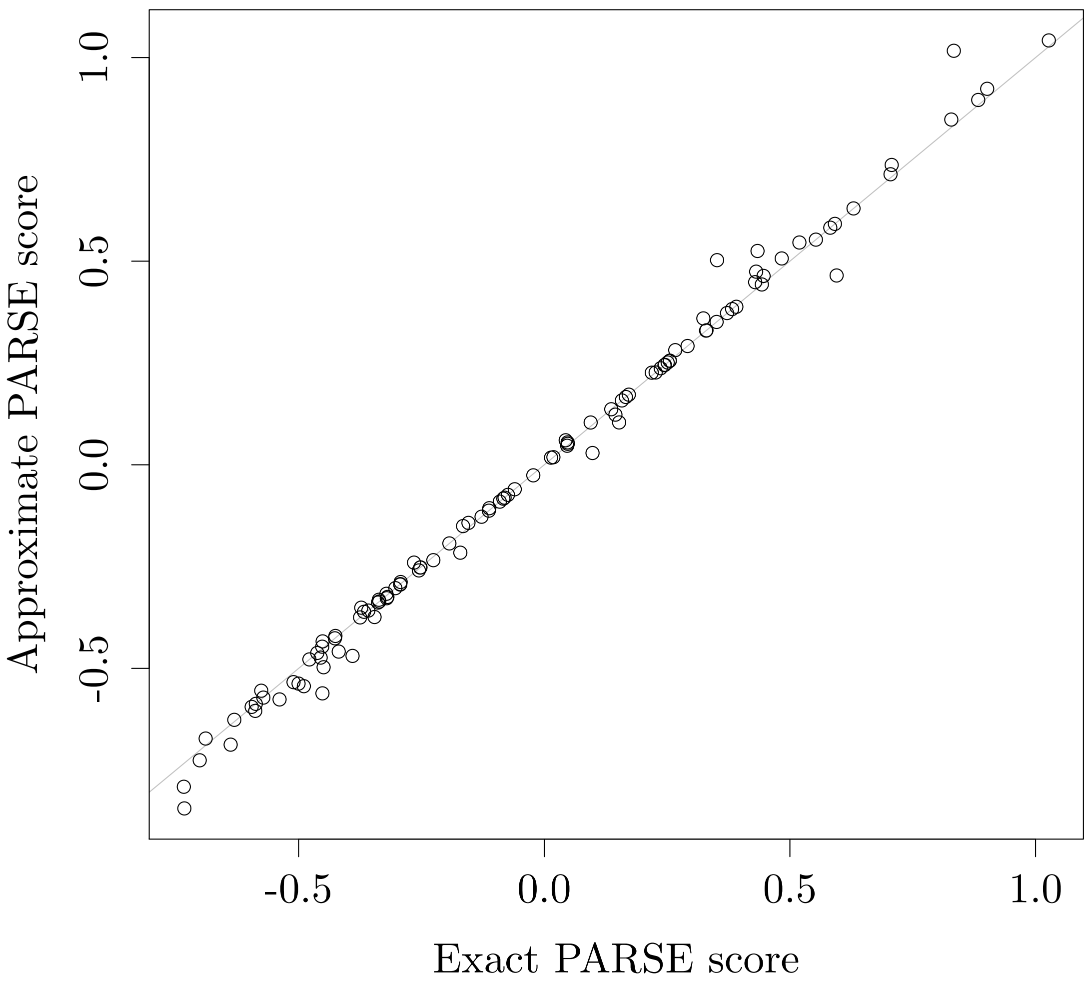

```r
coefmap(nmf.final)
```



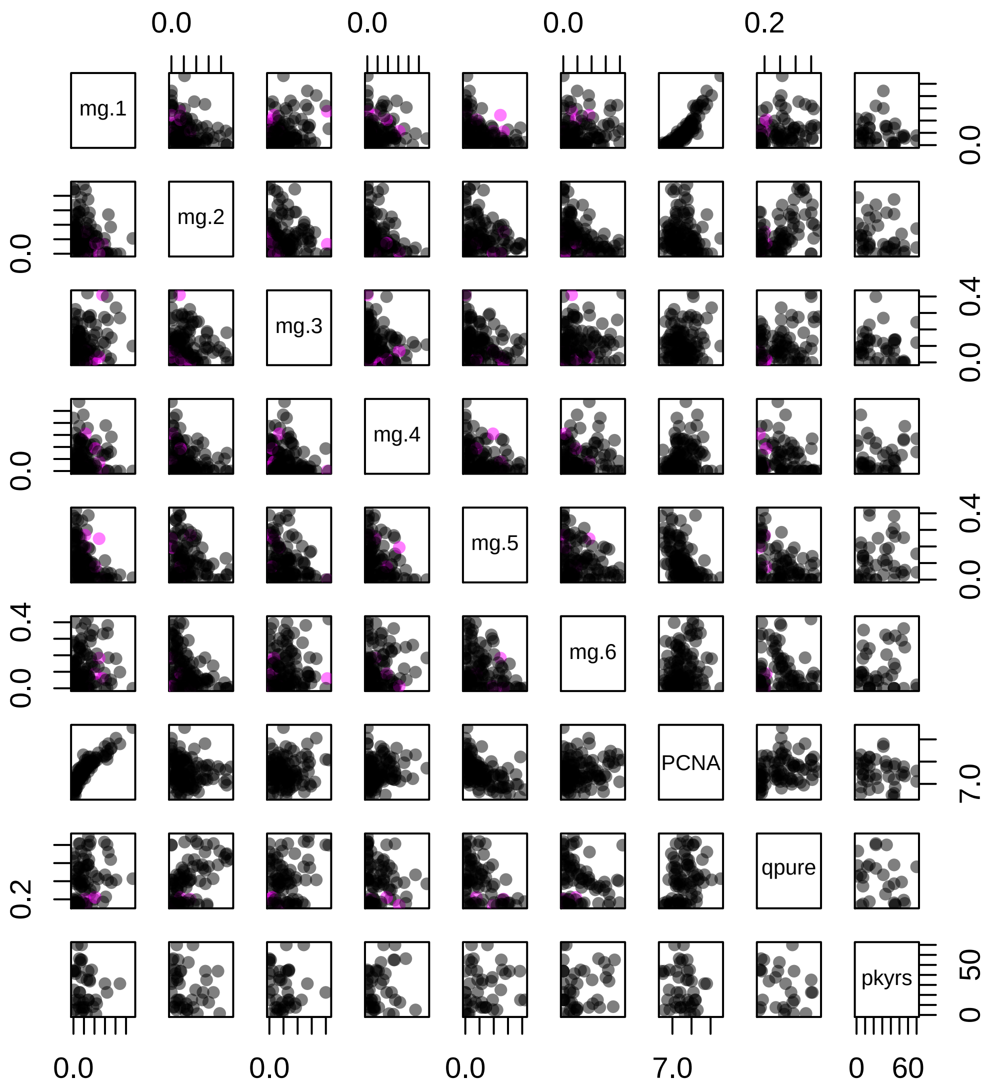

```
coefs.diag_dsd = apply(xlin.diag_dsd.sel, 2, function(xcol) nnls(basis(nmf.final), xcol)$x)
coefs.diag_rec = apply(xlin.diag_rec.sel, 2, function(xcol) nnls(basis(nmf.final), xcol)$x)
coefs.recr_dsd = apply(xlin.recr_dsd.sel, 2, function(xcol) nnls(basis(nmf.final), xcol)$x)
coefs.pdac_au = apply(xlin.pdac_au.sel, 2, function(xcol) nnls(basis(nmf.final), xcol)$x)
axis_coefs.diag_dsd = as.matrix(cbind(axis1 = coefs.diag_dsd[1,] - coefs.diag_dsd[5,], axis2 = coefs.dia
```

```
library(MASS)
W_plus = ginv(basis(nmf.final))

A1 = W_plus[1,] - W_plus[5,]
A2 = W_plus[6,] - W_plus[2,]
PARSE_approx = matrix(1.354*A1 + 1.548*A2, ncol = 1)

rownames(PARSE_approx) = rownames(basis(nmf.final))

PARSE_approx_scores = t(xlin.diag_dsd.sel) %*% PARSE_approx
```

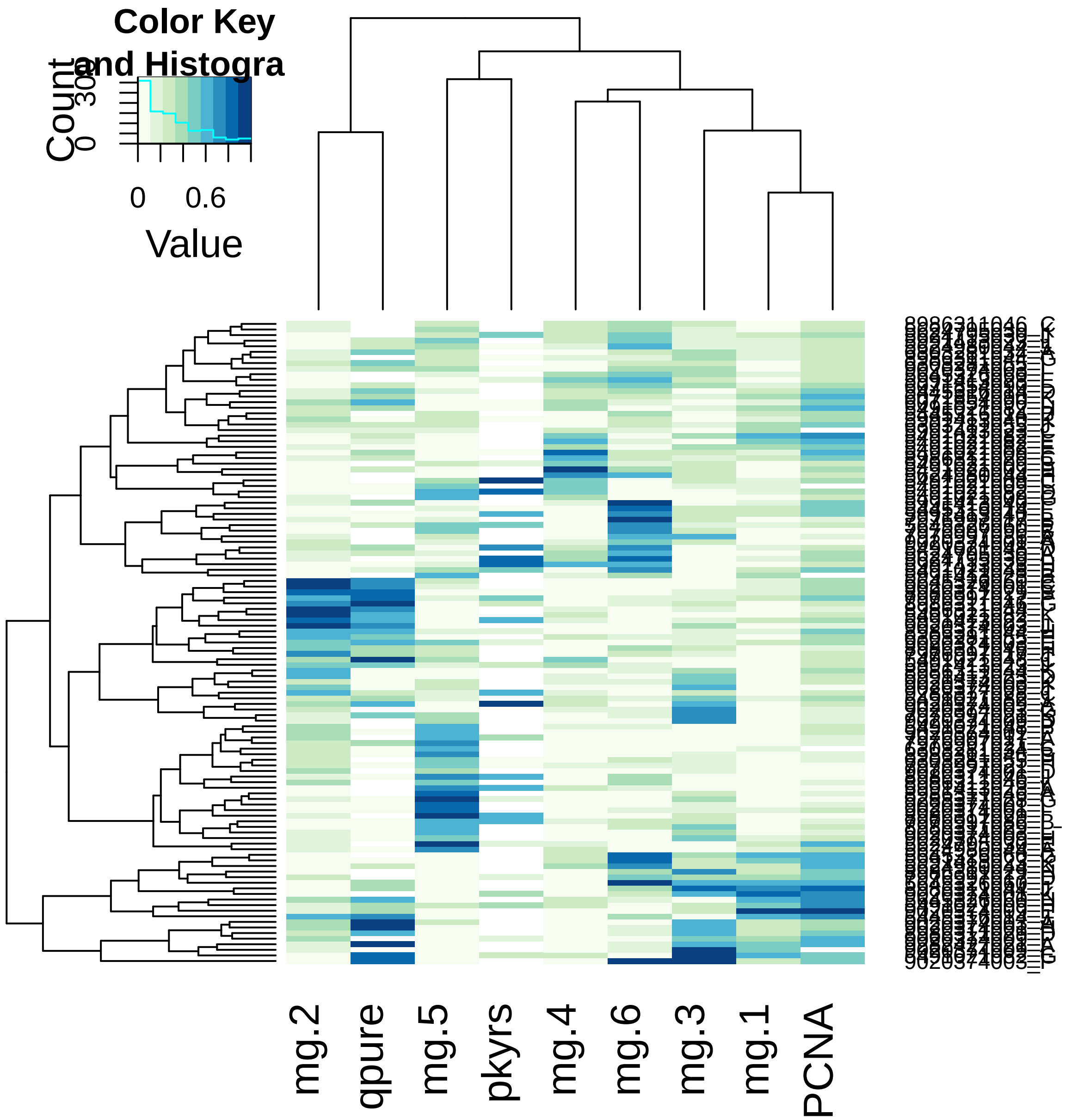
```
PARSE_exact_scores = 1.354*(coefs.diag_dsd[1,] - coefs.diag_dsd[5,]) + 1.548*(coefs.diag_dsd[6,] - coefs

plot(PARSE_exact_scores, PARSE_approx_scores, xlab = "Exact PARSE score", ylab = "Approximate PARSE scor
abline(0, 1, col = rgb(0, 0, 0, 0.25))
```



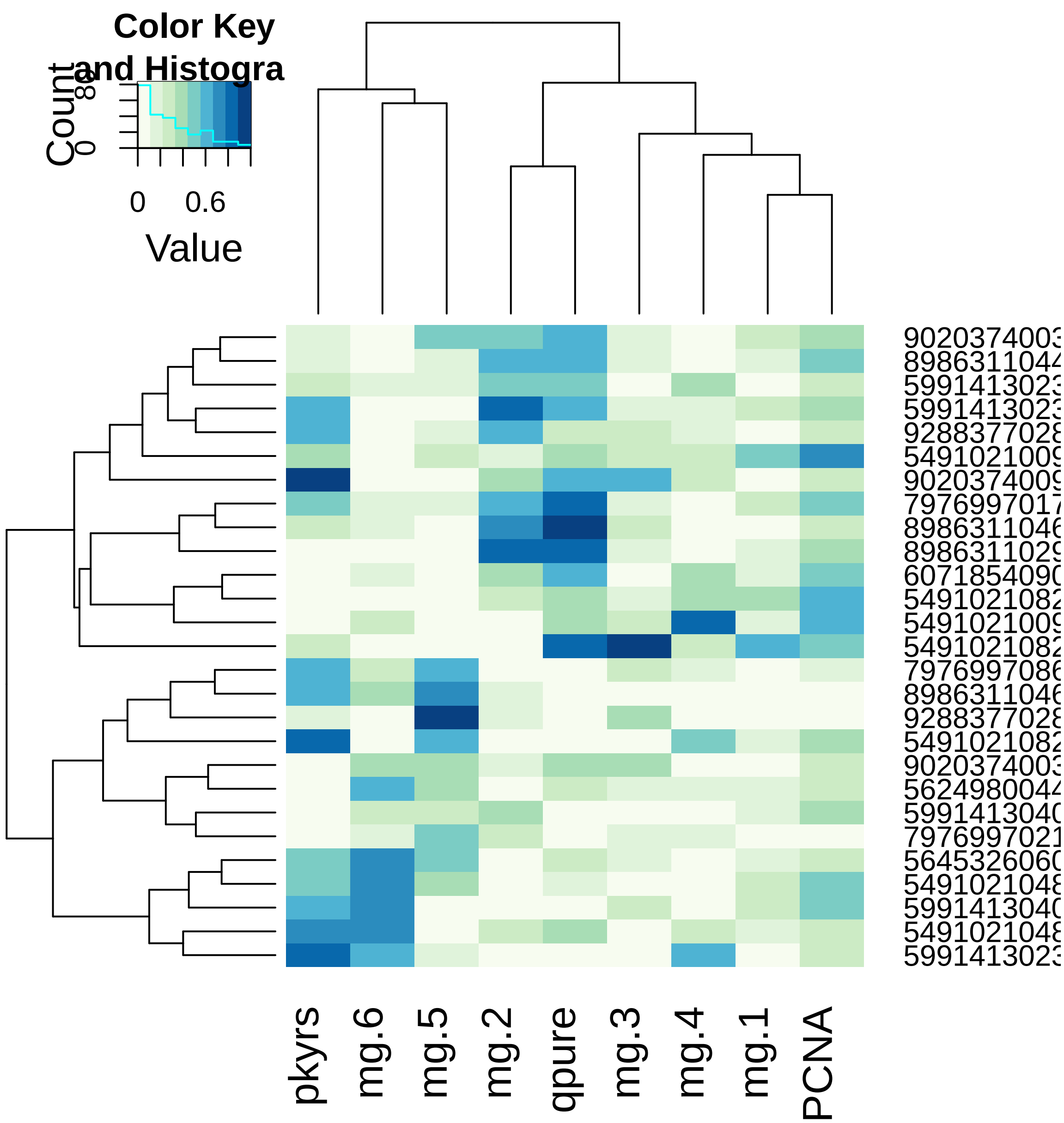

```
temp.pred.pairs = t(rbind(coefs.pdac_au, metapcna.scores[colnames(coefs.pdac_au)]))
colnames(temp.pred.pairs) = paste("mg", 1:ncol(temp.pred.pairs), sep = ".")
colnames(temp.pred.pairs)[ncol(temp.pred.pairs)] = "PCNA"
temp.pred.pairs = cbind(temp.pred.pairs, qpure = samps.pdac_au$purity_qpure, pkyrs = cpvs.pdac_au$Histor
pairs(temp.pred.pairs, pch = 16, cex = 1, col = ifelse(rownames(temp.pred.pairs) %in% colnames(xlin.diag
```



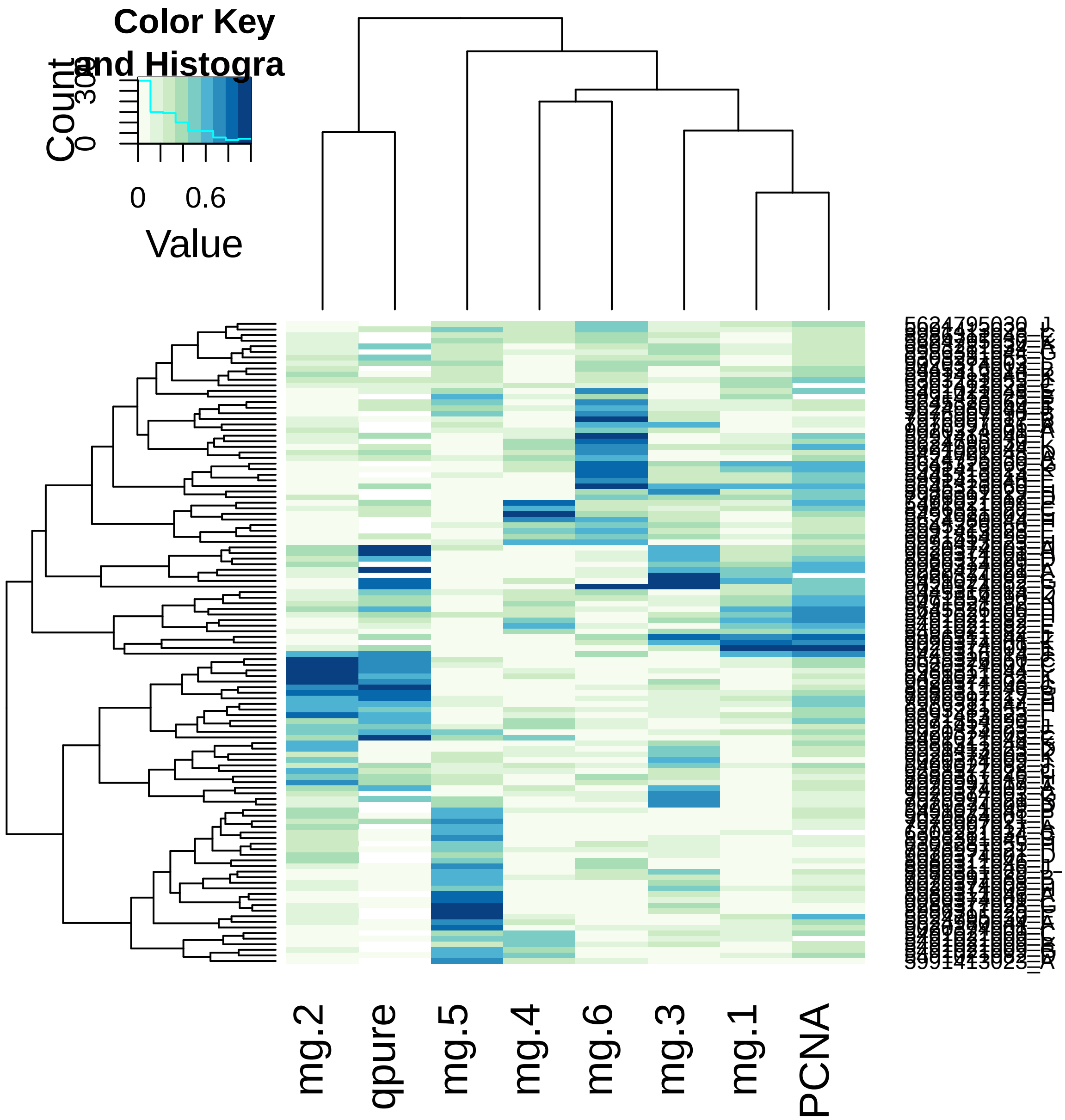

```
temp.pred.pairs.rescaled = t((t(temp.pred.pairs) - apply(temp.pred.pairs, 2, min, na.rm = TRUE)) / (appl
heatmap.2(temp.pred.pairs.rescaled, trace = "none", scale = "none", col = brewer.pal(9, "GnBu"))
```



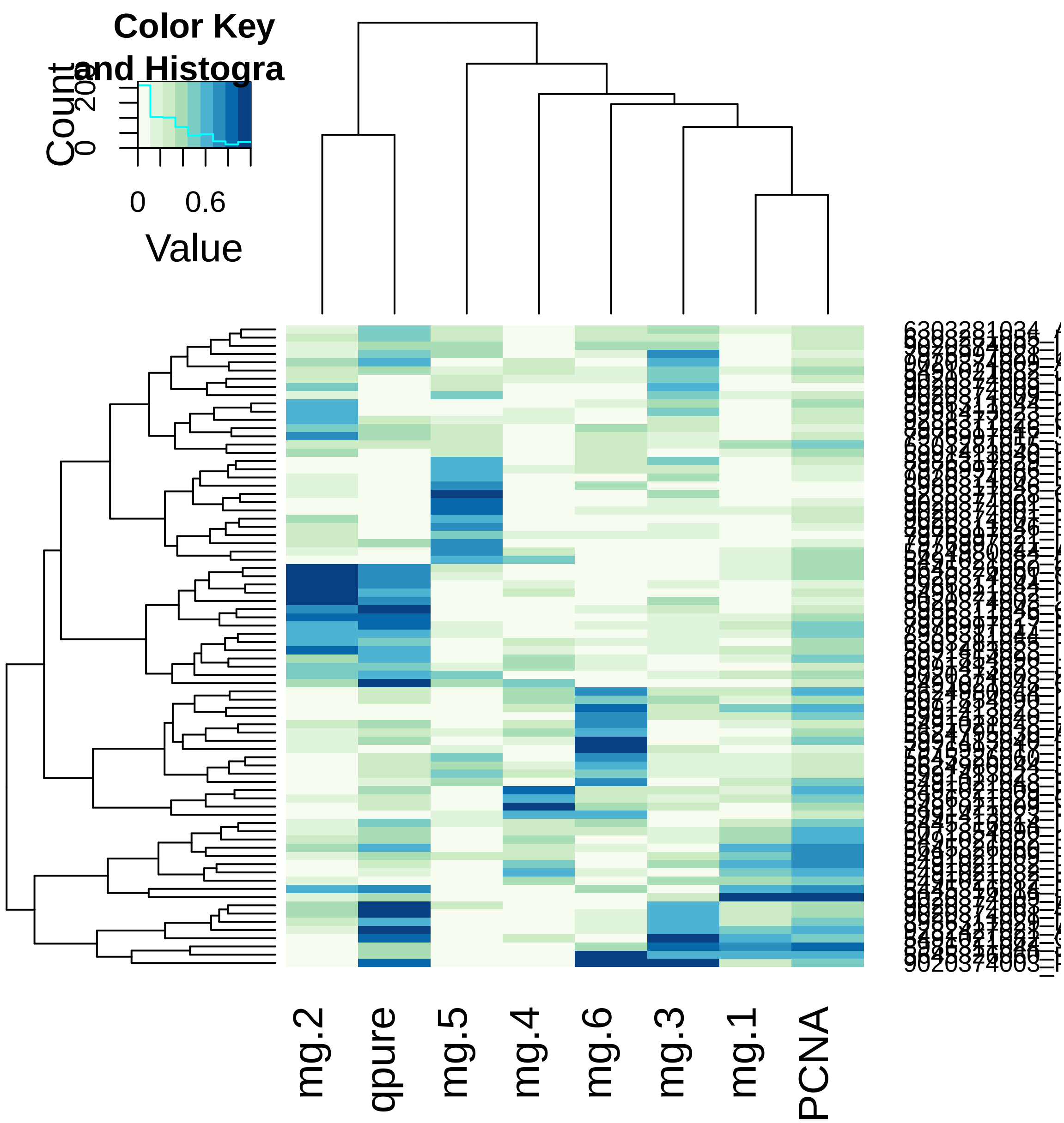

```
heatmap.2(temp.pred.pairs.rescaled[apply(!is.na(temp.pred.pairs.rescaled), 1, all),], trace = "none", sc
```



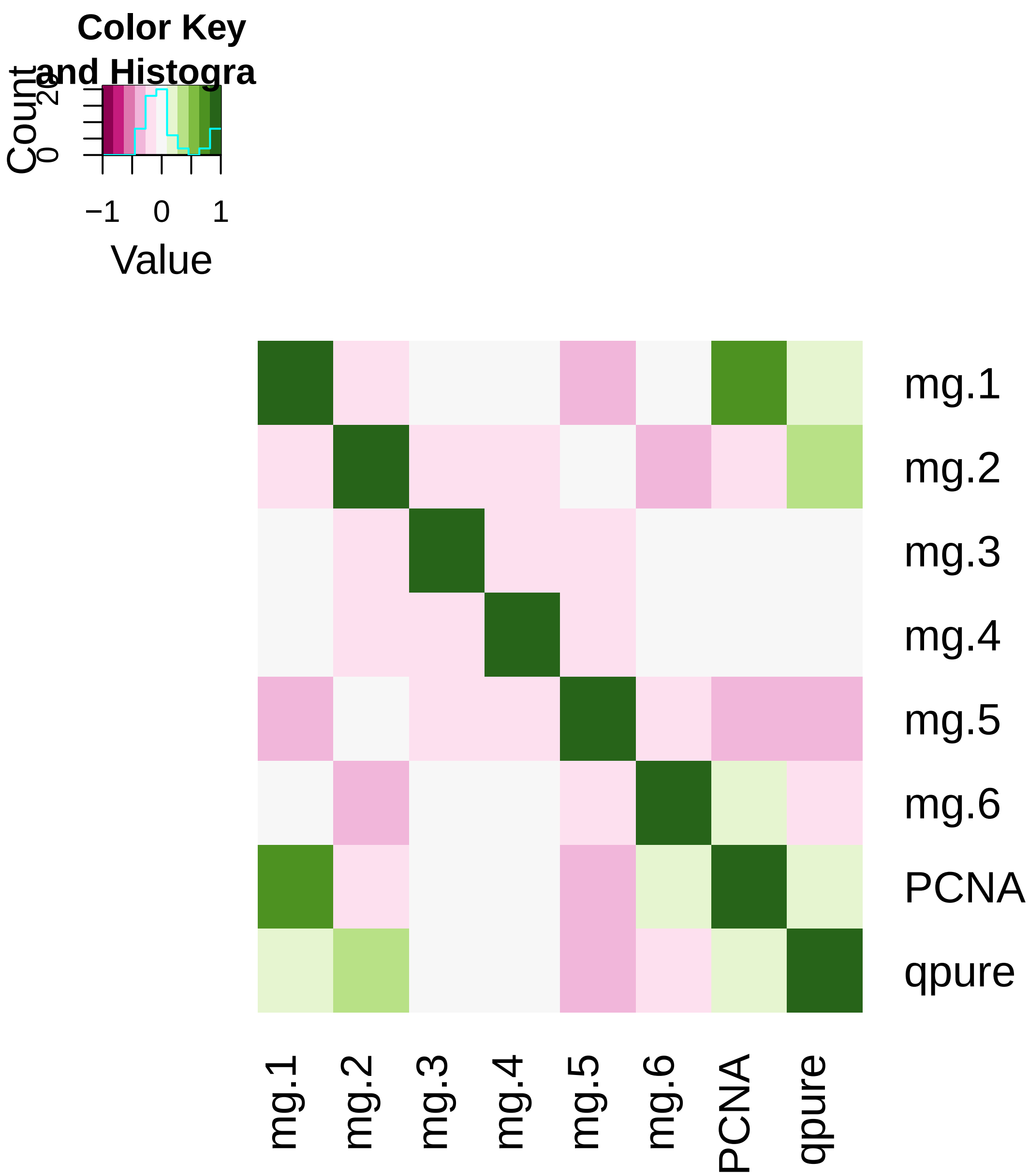

```
temp.pred.pairs.rescaled2 = temp.pred.pairs.rescaled[,colnames(temp.pred.pairs.rescaled) != "pkyrs"]
heatmap.2(temp.pred.pairs.rescaled2, trace = "none", scale = "none", col = brewer.pal(9, "GnBu"))
```



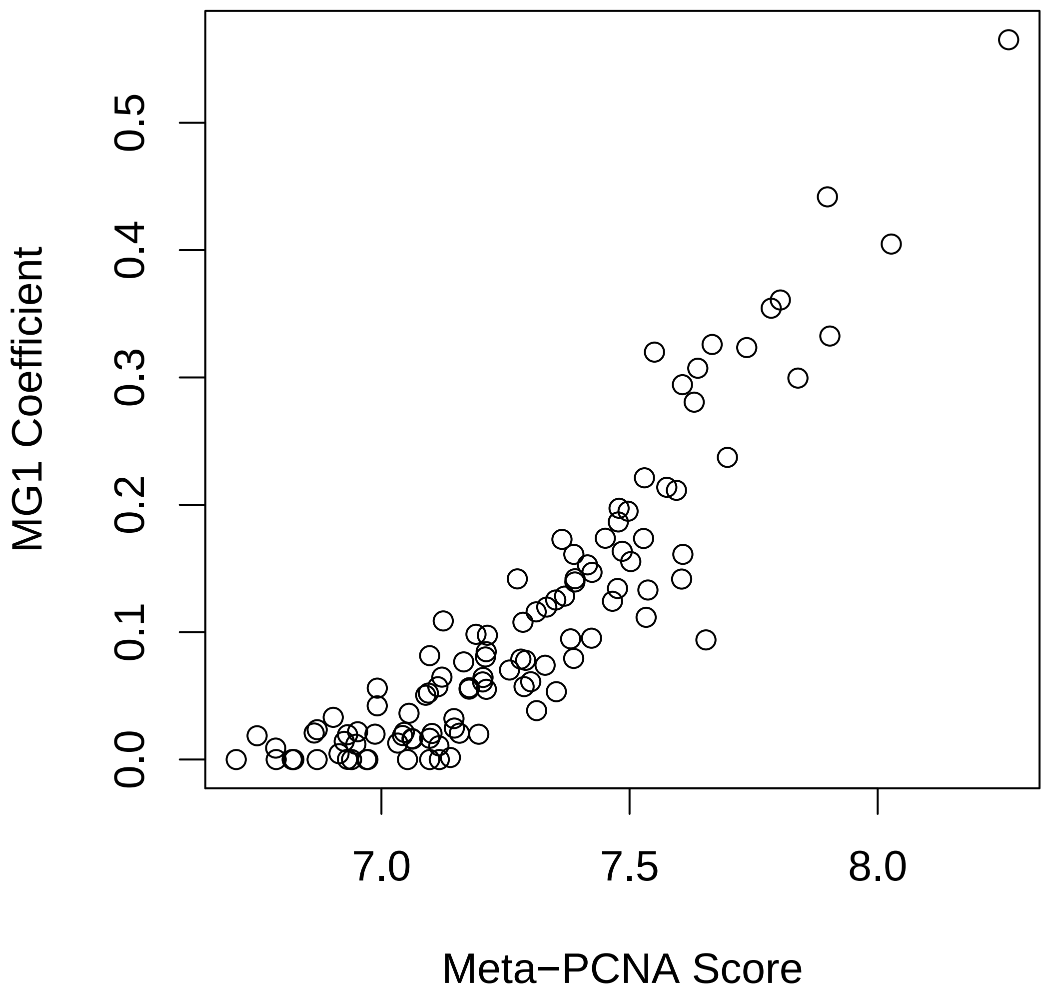

```
heatmap.2(temp.pred.pairs.rescaled2[apply(!is.na(temp.pred.pairs.rescaled2), 1, all),], trace = "none",
```



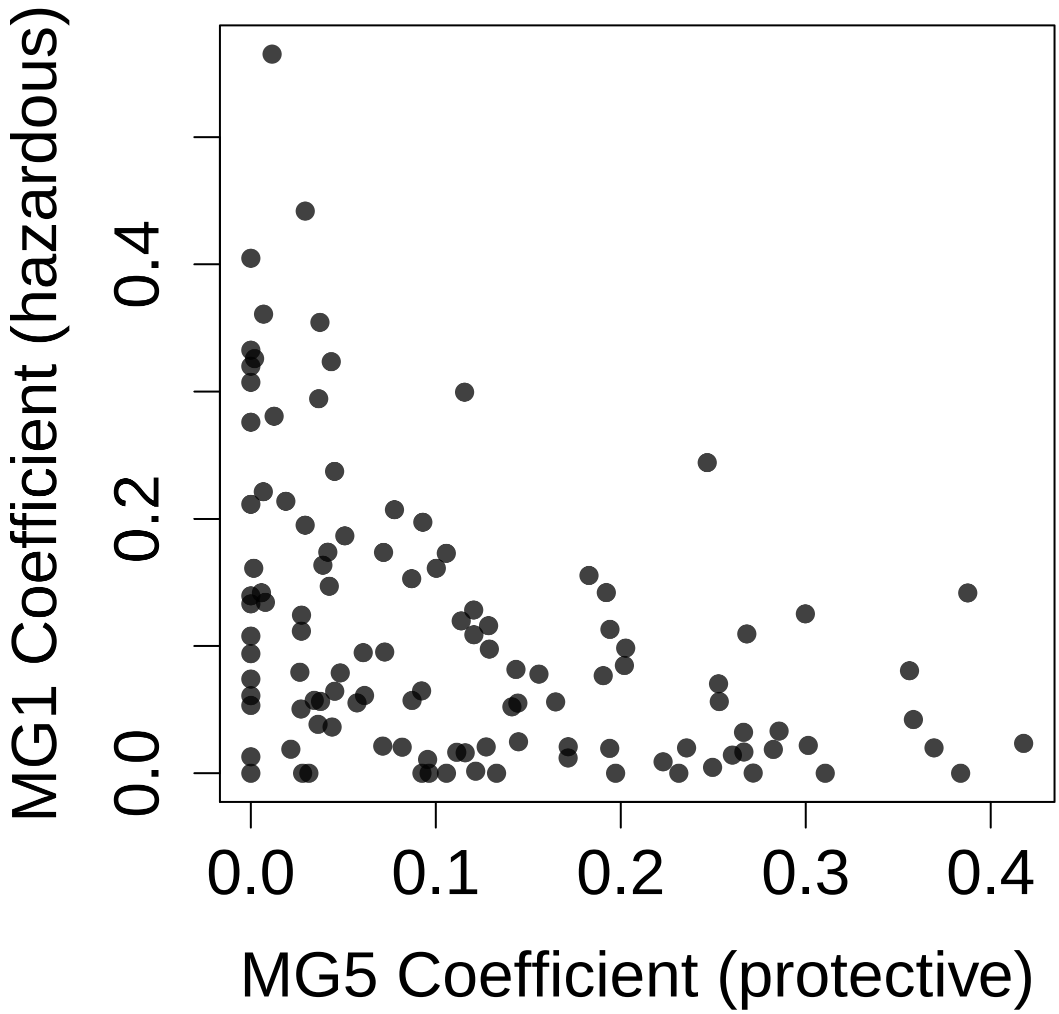

```
temp.cors = apply(temp.pred.pairs[,colnames(temp.pred.pairs) != "pkyrs"], 2, function(x) apply(temp.pred
#diag(temp.cors) = NA
heatmap.2(temp.cors, trace = "none", Rowv = FALSE, Colv = FALSE, col = brewer.pal(11, "PiYG"), dendrogra
```



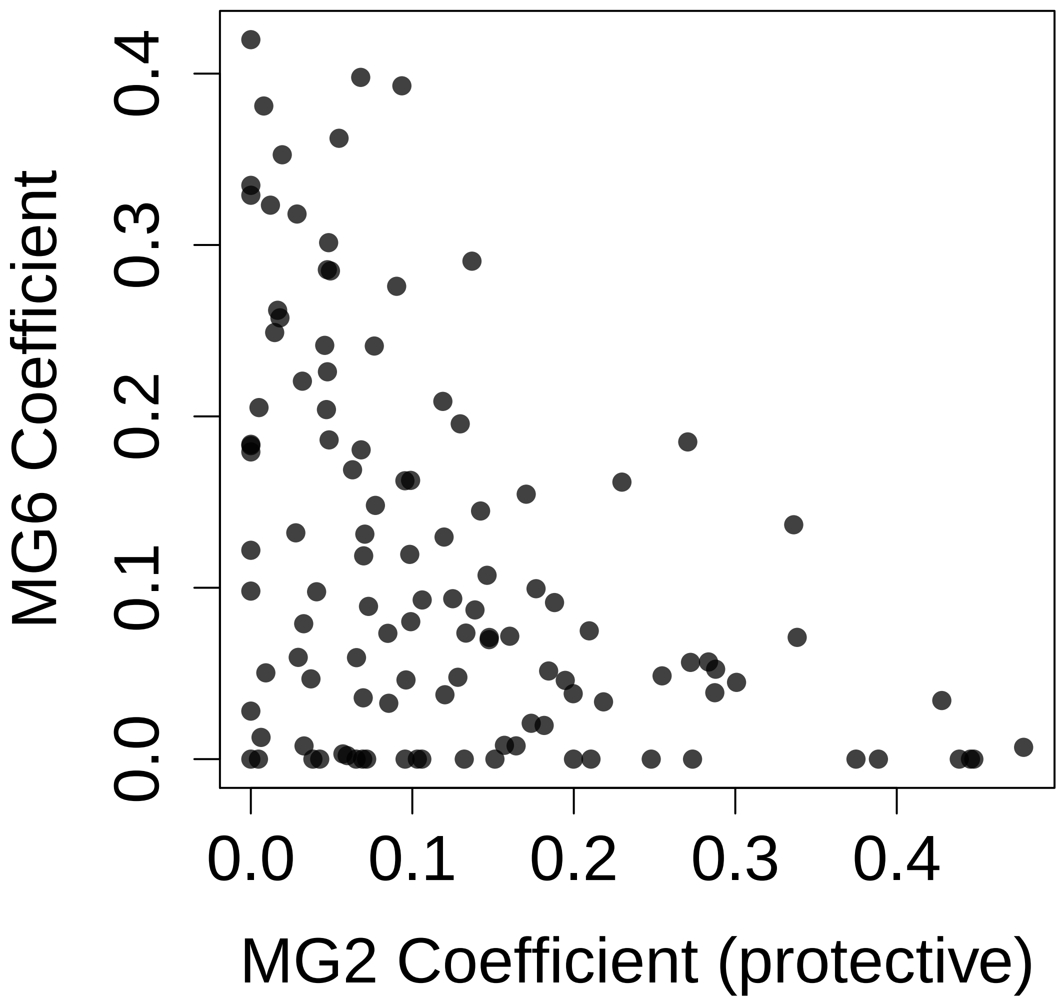

```
plot(temp.pred.pairs[,"mg.1"] ~ temp.pred.pairs[,"PCNA"], col = ifelse(rownames(temp.pred.pairs) %in% co
```




```
plot(temp.pred.pairs[,"mg.5"], temp.pred.pairs[,"mg.1"], xlab = "MG5 Coefficient (protective)", ylab = '
```




```
plot(temp.pred.pairs[,"mg.2"], temp.pred.pairs[,"mg.6"], xlab = "MG2 Coefficient (protective)", ylab = '
```




```
#scatter.smooth(temp.pred.pairs[,"mg.5"], temp.pred.pairs[,"mg.1"], xlab = "MG5 Coefficient (protective
#scatter.smooth(temp.pred.pairs[,"mg.2"], temp.pred.pairs[,"mg.6"], xlab = "MG2 Coefficient (protective
#smoothScatter(temp.pred.pairs[,"mg.5"], temp.pred.pairs[,"mg.1"], xlab = "MG5 Coefficient (protective)
#smoothScatter(temp.pred.pairs[,"mg.2"], temp.pred.pairs[,"mg.6"], xlab = "MG2 Coefficient (protective)

temp.coefs.pdcor = apply(coefs.diag_dsd, 1, function(x1) apply(coefs.diag_dsd, 1, function(x2) dcov.test
temp.coefs.pfisher = apply(coefs.diag_dsd, 1, function(x1) apply(coefs.diag_dsd, 1, function(x2) fisher.
diag(temp.coefs.pdcor) = NA
temp.coefs.pdcor[lower.tri(temp.coefs.pdcor)] = NA
diag(temp.coefs.pfisher) = NA
temp.coefs.pfisher[lower.tri(temp.coefs.pfisher)] = NA
temp.coefs.pdcor.holm = matrix(p.adjust(temp.coefs.pdcor, "holm"), nrow = nrow(temp.coefs.pdcor))
temp.coefs.pfisher.holm = matrix(p.adjust(temp.coefs.pfisher, "holm"), nrow = nrow(temp.coefs.pfisher))
temp.coefs.pdcor.holm
```

```
##      [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
## [1,]   NA 0.2016 0.4500 1.0000 0.0015 1.0000
## [2,]   NA     NA 0.3066 0.0130 0.1800 0.0015
## [3,]   NA     NA     NA 0.0336 0.0451 1.0000
## [4,]   NA     NA     NA     NA 0.0480 1.0000
## [5,]   NA     NA     NA     NA     NA 0.0480
## [6,]   NA     NA     NA     NA     NA     NA
```

```
temp.coefs.pfisher.holm
```

```
##      [,1] [,2]   [,3] [,4]    [,5]    [,6]
## [1,]   NA    1 1.0000    1 0.03203 1.00000
## [2,]   NA   NA 0.7286    1 1.00000 0.03203
## [3,]   NA   NA     NA    1 1.00000 1.00000
## [4,]   NA   NA     NA   NA 0.72858 1.00000
## [5,]   NA   NA     NA   NA      NA 1.00000
## [6,]   NA   NA     NA   NA      NA      NA
```

```
dcov.test(coefs.diag_dsd[5,], coefs.diag_dsd[1,], R = 19999)
```

```
##
##  dCov test of independence
##
## data:  index 1, replicates 19999
## nV^2 = 0.1291, p-value = 5e-05
## sample estimates:
##    dCov
## 0.03426
```

```
dcov.test(coefs.diag_dsd[2,], coefs.diag_dsd[6,], R = 19999)
```

```
##
##  dCov test of independence
##
## data:  index 1, replicates 19999
## nV^2 = 0.1396, p-value = 5e-05
## sample estimates:
##    dCov
## 0.03562
```

```
cor.test(coefs.diag_dsd[5,], coefs.diag_dsd[1,], method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  coefs.diag_dsd[5, ] and coefs.diag_dsd[1, ]
## z = -4.97, p-value = 6.694e-07
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##     tau
## -0.3243
```

```
cor.test(coefs.diag_dsd[2,], coefs.diag_dsd[6,], method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  coefs.diag_dsd[2, ] and coefs.diag_dsd[6, ]
## z = -4.931, p-value = 8.195e-07
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##     tau
## -0.3236
```

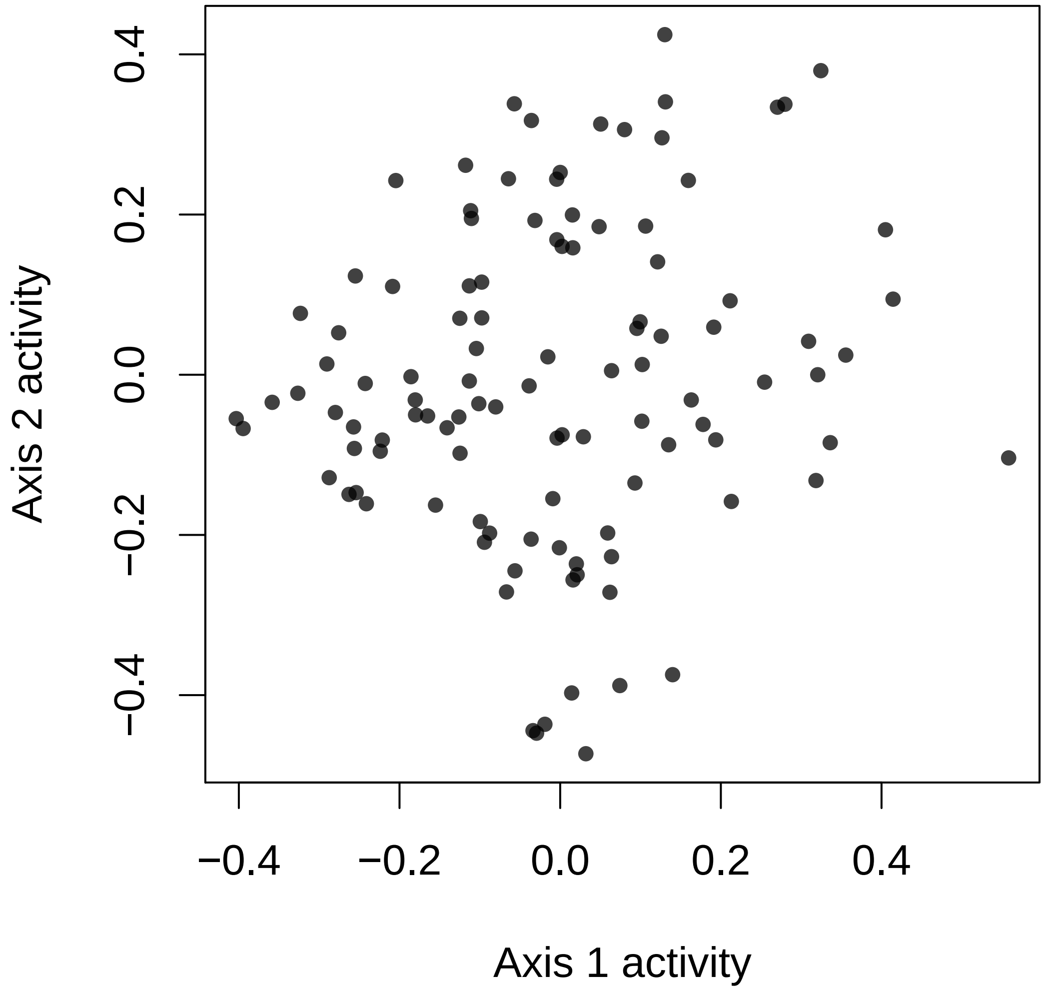
```
temp.axis1 = coefs.diag_dsd[1,] - coefs.diag_dsd[5,]
temp.axis2 = coefs.diag_dsd[6,] - coefs.diag_dsd[2,]
dcov.test(temp.axis1, temp.axis2, R = 19999)
```

```
##
##  dCov test of independence
##
## data:  index 1, replicates 19999
## nV^2 = 0.1074, p-value = 0.0197
## sample estimates:
##    dCov
## 0.03124
```

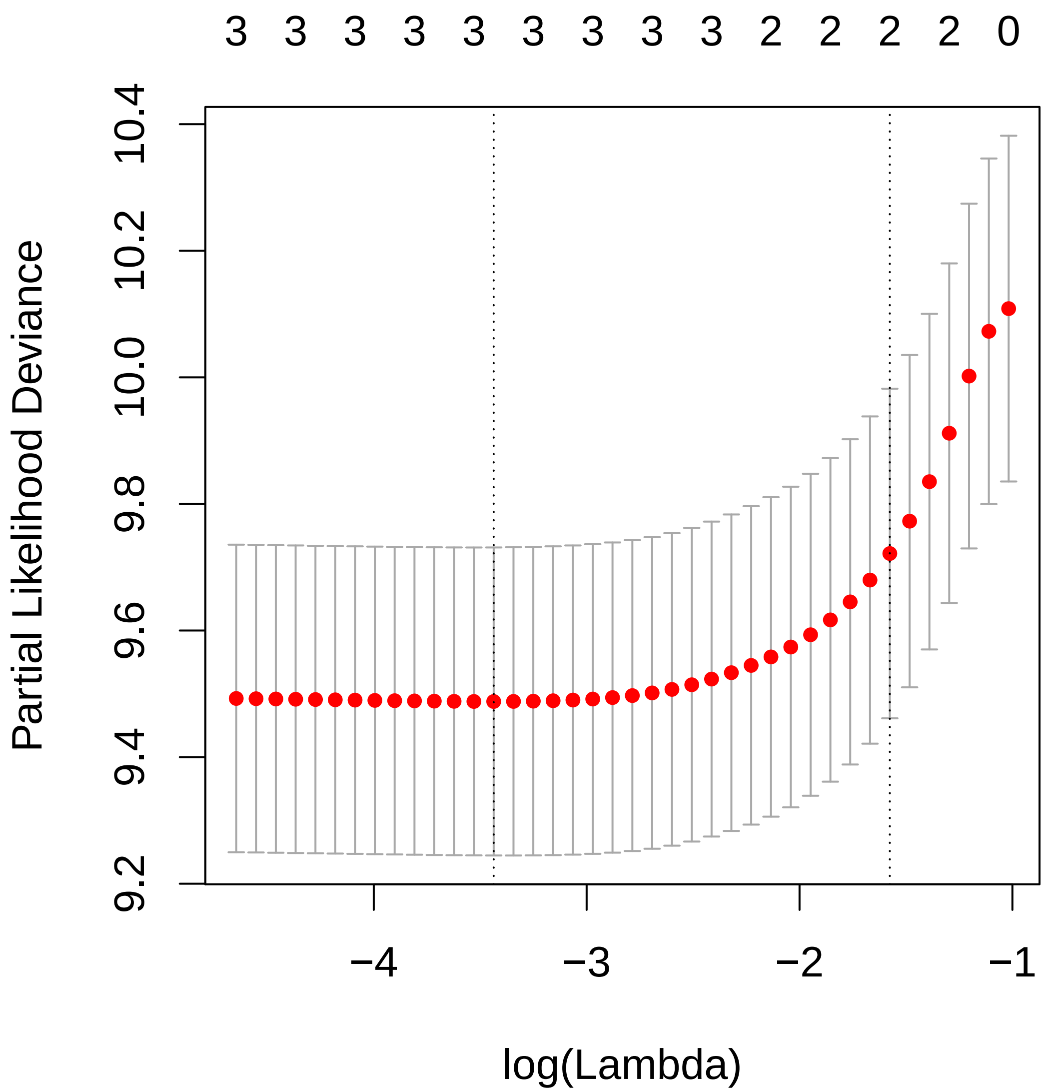
```
cor.test(temp.axis1, temp.axis2, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  temp.axis1 and temp.axis2
## z = 1.253, p-value = 0.2103
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##    tau
## 0.0809
```

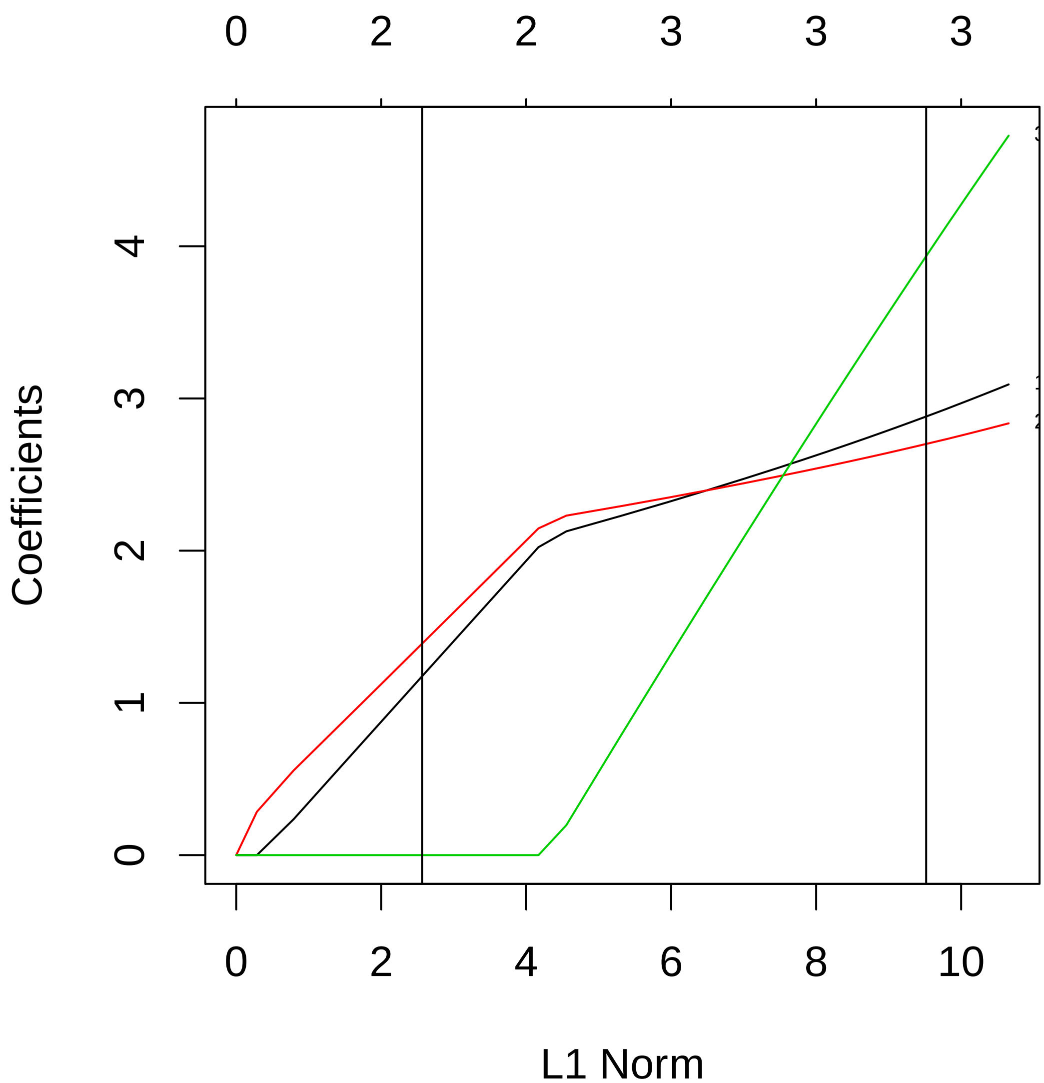
```
plot(temp.axis2 ~ temp.axis1, xlab = "Axis 1 activity", ylab = "Axis 2 activity", pch = 16, col = rgb(0,
```




```
coxph(y.diag_dsd ~ temp.axis1 * temp.axis2)

## Call:
## coxph(formula = y.diag_dsd ~ temp.axis1 * temp.axis2)
##
##
##                         coef exp(coef) se(coef)    z       p
## temp.axis1              3.19      24.2    0.676 4.72 2.4e-06
## temp.axis2              2.89      18.0    0.657 4.40 1.1e-05
## temp.axis1:temp.axis2 5.03     153.1    4.189 1.20 2.3e-01
##
## Likelihood ratio test=48  on 3 df, p=2.12e-10  n= 110, number of events= 70

temp = cv.glmnet(cbind(temp.axis1, temp.axis2, temp.axis1*temp.axis2), y.diag_dsd, family = "cox", nfolc
plot(temp)
```




```r
plot(temp$glmnet.fit, label = TRUE)
abline(v = sum(abs(coef(temp$glmnet.fit, s = temp$lambda.1se))))
abline(v = sum(abs(coef(temp$glmnet.fit, s = temp$lambda.min))))
```

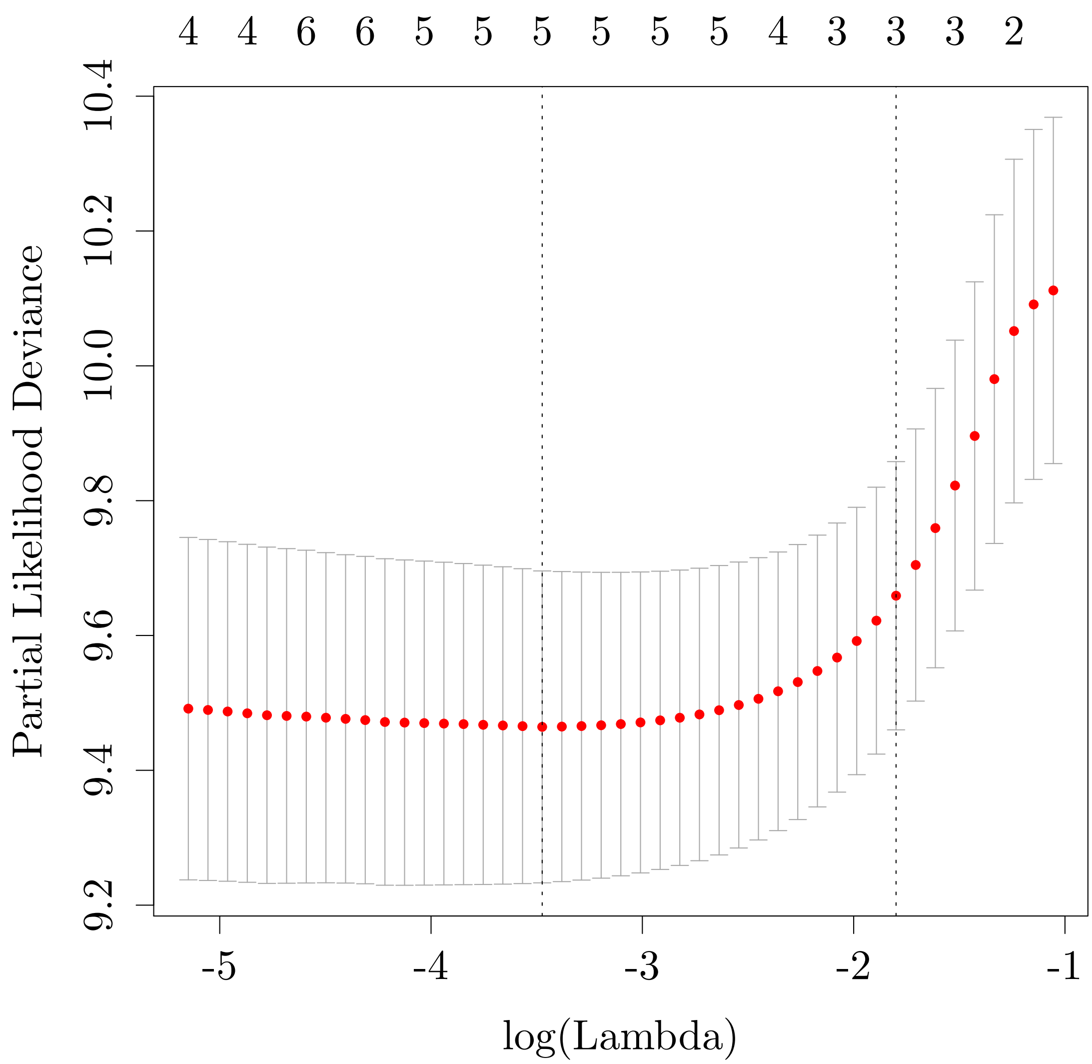



```
coef(temp$glmnet.fit, s = temp$lambda.1se)

## 3 x 1 sparse Matrix of class "dgCMatrix"
##                 1
## temp.axis1 1.176
## temp.axis2 1.390
##             .
```

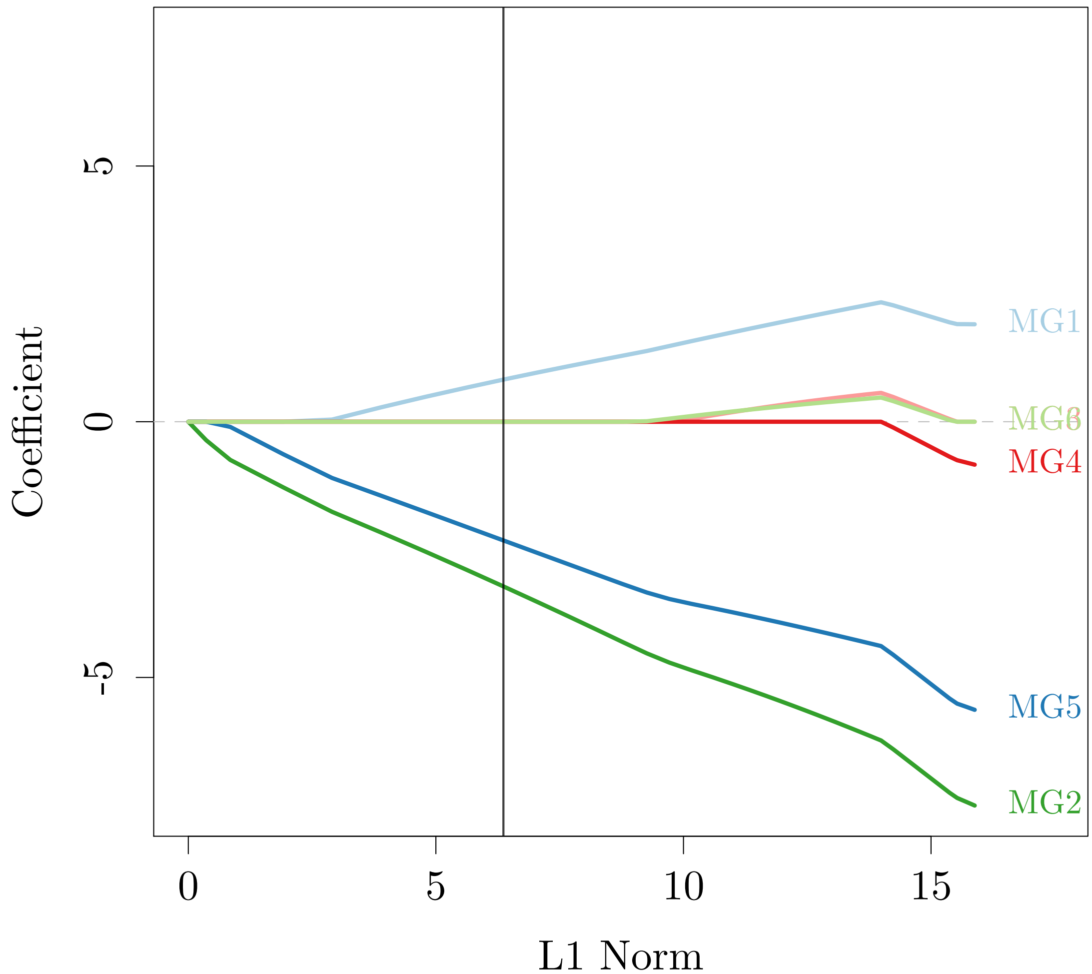
4.2 LASSO on training set

```
glmnet.fit.cv.diag_dsd = cv.glmnet(t(coefs.diag_dsd), y.diag_dsd, family = "cox", nfolds = 10)
glmnet.fit.cv.diag_rec = cv.glmnet(t(coefs.diag_rec), y.diag_rec, family = "cox", nfolds = 10)
glmnet.fit.cv.recr_dsd = cv.glmnet(t(coefs.recr_dsd), y.recr_dsd, family = "cox", nfolds = 10)
```

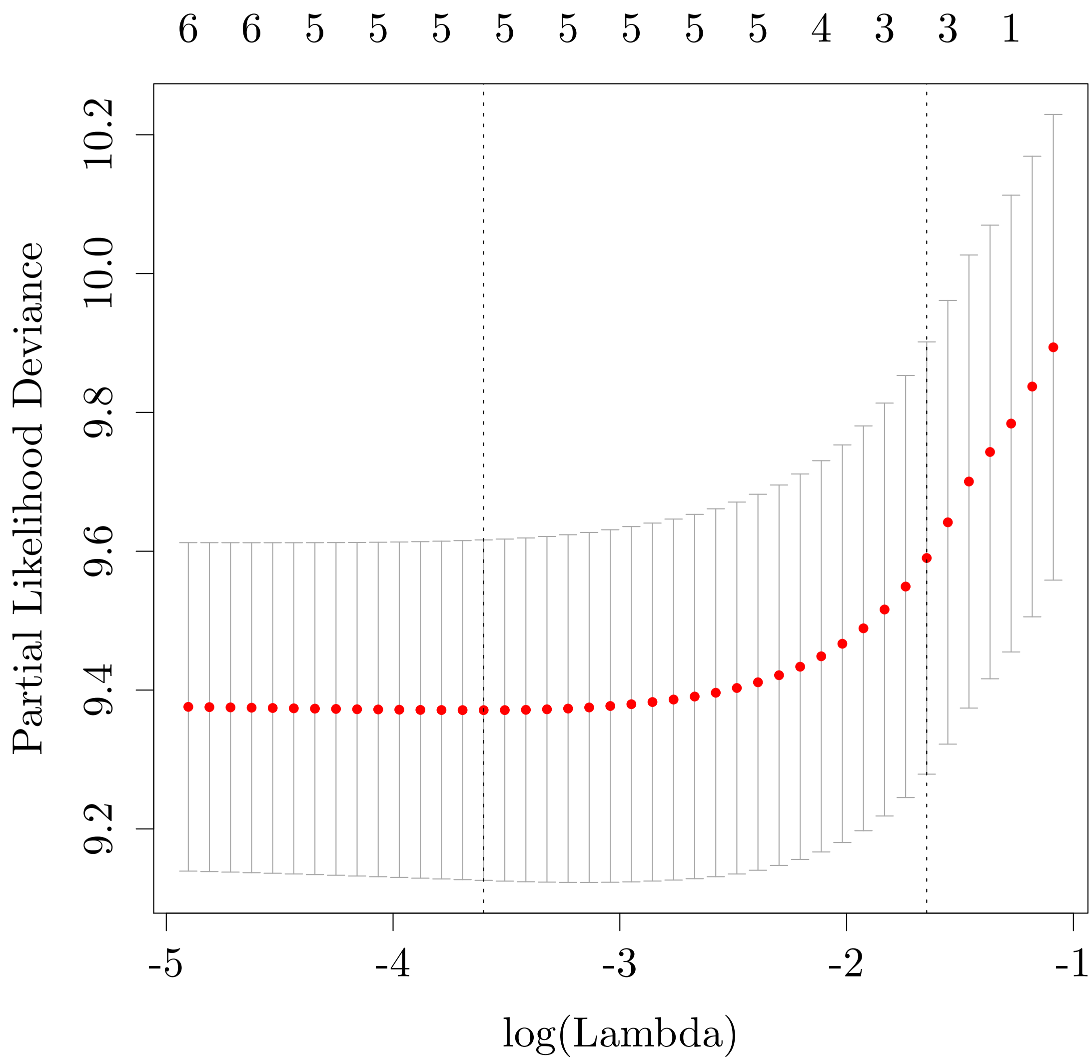
```r
plot(glmnet.fit.cv.diag_dsd)
```



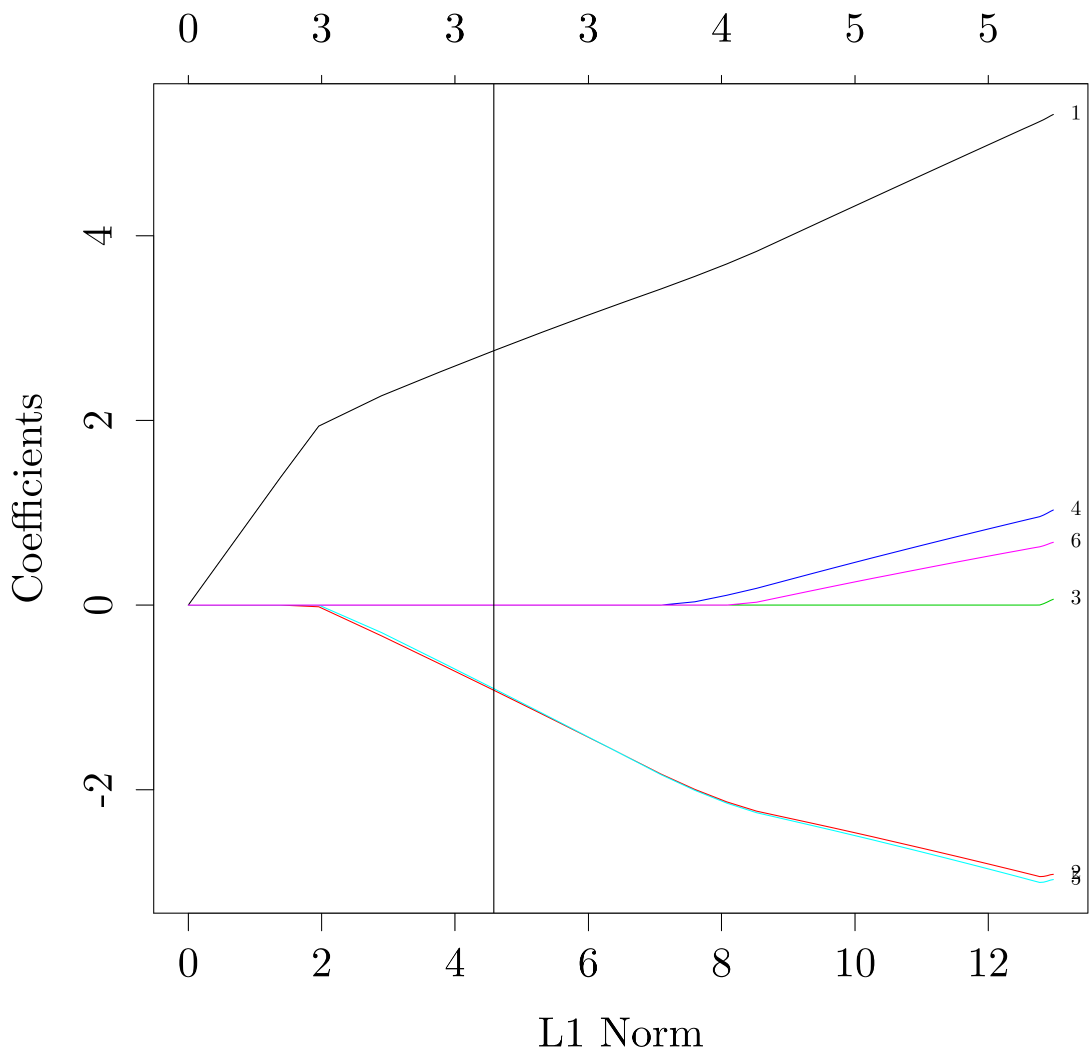

```r
#plot(glmnet.fit.cv.diag_dsd£glmnet.fit, label = TRUE)
temp = glmnet.fit.cv.diag_dsd$glmnet.fit
temp.betamax = max(abs(temp$beta))
temp.l1 = colSums(abs(temp$beta))
plot(0 ~ 0, xlim = c(0, max(temp.l1) * 1.1), ylim = c(-temp.betamax, temp.betamax), type = "n", xlab = '
abline(h = 0, lty = "dashed", col = "grey")
temp.pal = brewer.pal(nrow(temp$beta), "Paired")[c(1,4,5,6,2,3)]
for (i in 1:nrow(temp$beta))
{
        lines(temp$beta[i,] ~ temp.l1, col = temp.pal[i], lwd = 4)
        text(max(temp.l1) * 1.02, temp$beta[i,length(temp.l1)], paste("MG", i, sep = ""), col = temp.pal
}
abline(v = sum(abs(coef(glmnet.fit.cv.diag_dsd$glmnet.fit, s = glmnet.fit.cv.diag_dsd$lambda.1se))), col
```



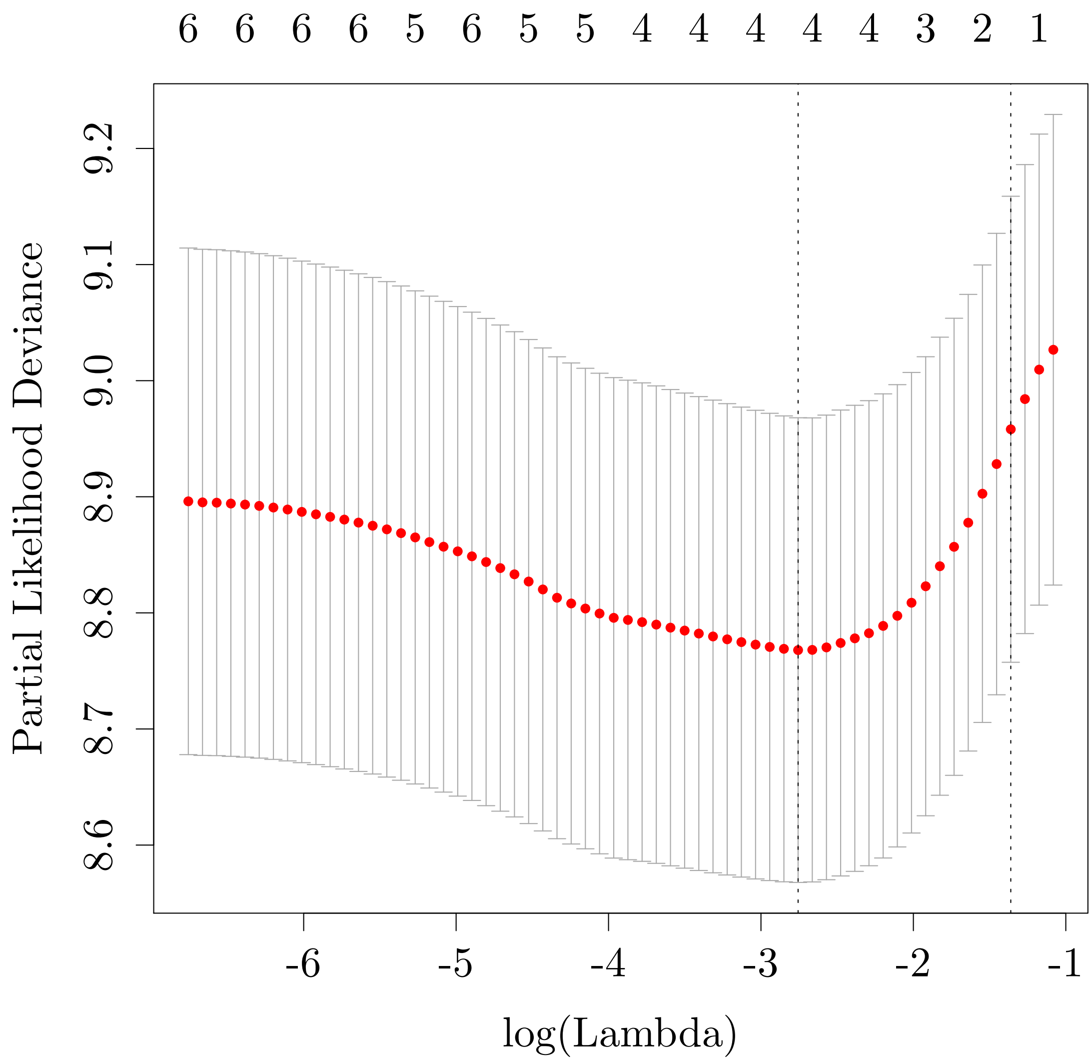

```r
#abline(v = sum(abs(coef(glmnet.fit.cv.diag_dsd£glmnet.fit, s = glmnet.fit.cv.diag_dsd£lambda.min))))
coef(glmnet.fit.cv.diag_dsd$glmnet.fit, s = glmnet.fit.cv.diag_dsd$lambda.1se)

## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## V1  0.8238
## V2 -3.2195
## V3  .
## V4  .
## V5 -2.3208
## V6  .

plot(glmnet.fit.cv.diag_rec)
```



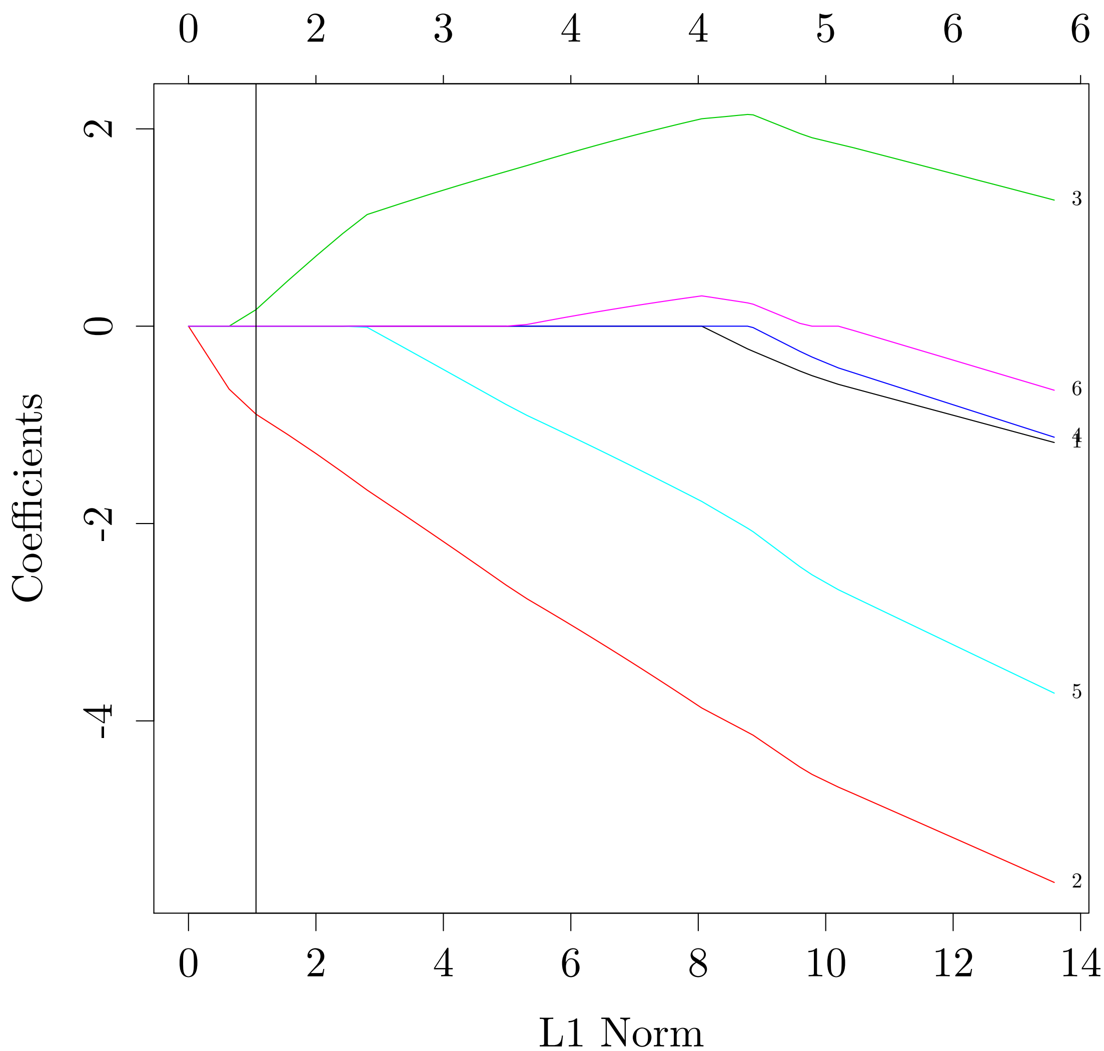

```r
plot(glmnet.fit.cv.diag_rec$glmnet.fit, label = TRUE)
abline(v = sum(abs(coef(glmnet.fit.cv.diag_rec$glmnet.fit, s = glmnet.fit.cv.diag_rec$lambda.1se))))
```




```
#abline(v = sum(abs(coef(glmnet.fit.cv.diag_rec£glmnet.fit, s = glmnet.fit.cv.diag_rec£lambda.min))))
coef(glmnet.fit.cv.diag_rec$glmnet.fit, s = glmnet.fit.cv.diag_rec$lambda.1se)

## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## V1  2.7555
## V2 -0.9230
## V3  .
## V4  .
## V5 -0.9055
## V6  .

plot(glmnet.fit.cv.recr_dsd)
```




```
plot(glmnet.fit.cv.recr_dsd$glmnet.fit, label = TRUE)
abline(v = sum(abs(coef(glmnet.fit.cv.recr_dsd$glmnet.fit, s = glmnet.fit.cv.recr_dsd$lambda.1se))))
```




```
#abline(v = sum(abs(coef(glmnet.fit.cv.recr_dsd£glmnet.fit, s = glmnet.fit.cv.recr_dsd£lambda.min))))
coef(glmnet.fit.cv.recr_dsd$glmnet.fit, s = glmnet.fit.cv.recr_dsd$lambda.1se)

## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## V1  .
## V2 -0.8920
## V3  0.1676
## V4  .
## V5  .
## V6  .
```

4.3 Prediction on validation sets

```r
load("../../data/15_validation.rda")
```

```r
val.basis = basis(nmf.final)
rownames(GSE21501.lingex) = GSE21501.feat$Gene.symbol
rownames(GSE28735.lingex) = GSE28735.feat$Gene.symbol
GSE21501.lingex.for_basis = GSE21501.lingex[match(rownames(val.basis), rownames(GSE21501.lingex)),]
GSE28735.lingex.for_basis = GSE28735.lingex[match(rownames(val.basis), rownames(GSE28735.lingex)),]
GSE21501.lingex.for_basis[is.na(GSE21501.lingex.for_basis)] = 0
GSE28735.lingex.for_basis[is.na(GSE28735.lingex.for_basis)] = 0

GSE21501.coefs = apply(GSE21501.lingex.for_basis, 2, function(xcol) nnls(val.basis, xcol)$x)
GSE28735.coefs = apply(GSE28735.lingex.for_basis, 2, function(xcol) nnls(val.basis, xcol)$x)

GSE21501.axis1 = GSE21501.coefs[1,] - GSE21501.coefs[5,]
GSE21501.axis2 = GSE21501.coefs[6,] - GSE21501.coefs[2,]
GSE28735.axis1 = GSE28735.coefs[1,] - GSE28735.coefs[5,]
GSE28735.axis2 = GSE28735.coefs[6,] - GSE28735.coefs[2,]

GSE21501.score = 1.354*GSE21501.axis1 + 1.548*GSE21501.axis2
GSE28735.score = 1.354*GSE28735.axis1 + 1.548*GSE28735.axis2

GSE21501.pcna = apply(GSE21501.gex[match(metapcna.sig, GSE21501.feat$Gene.symbol),], 2, median, na.rm =
GSE28735.pcna = apply(GSE28735.gex[match(metapcna.sig, GSE28735.feat$Gene.symbol),], 2, median, na.rm =
```

```r
temp = coxph(Surv(GSE21501.samp$time, GSE21501.samp$event) ~ GSE21501.score)
summary(temp)

## Call:
## coxph(formula = Surv(GSE21501.samp$time, GSE21501.samp$event) ~
##     GSE21501.score)
##
##   n= 102, number of events= 66
##
##                 coef exp(coef) se(coef)    z Pr(>|z|)
## GSE21501.score 1.81      6.13     1.14 1.59     0.11
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## GSE21501.score      6.13      0.163     0.655      57.3
##
## Concordance= 0.577  (se = 0.042 )
## Rsquare= 0.024   (max possible= 0.993 )
## Likelihood ratio test= 2.49  on 1 df,   p=0.115
## Wald test            = 2.52  on 1 df,   p=0.112
## Score (logrank) test = 2.54  on 1 df,   p=0.111

temp = coxph(Surv(GSE28735.samp$time, GSE28735.samp$event) ~ GSE28735.score)
summary(temp)

## Call:
## coxph(formula = Surv(GSE28735.samp$time, GSE28735.samp$event) ~
##     GSE28735.score)
##
```

```
##   n= 42, number of events= 29
##
##                   coef exp(coef) se(coef)    z Pr(>|z|)
## GSE28735.score 1.867     6.471    0.752 2.48    0.013
##
##                exp(coef) exp(-coef) lower .95 upper .95
## GSE28735.score      6.47      0.155      1.48      28.2
##
## Concordance= 0.655  (se = 0.064 )
## Rsquare= 0.132   (max possible= 0.981 )
## Likelihood ratio test= 5.92  on 1 df,   p=0.0149
## Wald test            = 6.17  on 1 df,   p=0.013
## Score (logrank) test = 6.46  on 1 df,   p=0.011
```

```
anova(coxph(Surv(GSE21501.samp$time, GSE21501.samp$event) ~ GSE21501.axis1 + GSE21501.axis2))
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(GSE21501.samp$time, GSE21501.samp$event)
## Terms added sequentially (first to last)
##
##                 loglik Chisq Df Pr(>|Chi|)
## NULL              -255
## GSE21501.axis1   -254  1.44  1       0.23
## GSE21501.axis2   -254  1.09  1       0.30
```

```
anova(coxph(Surv(GSE28735.samp$time, GSE28735.samp$event) ~ GSE28735.axis1 + GSE28735.axis2))
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(GSE28735.samp$time, GSE28735.samp$event)
## Terms added sequentially (first to last)
##
##                 loglik Chisq Df Pr(>|Chi|)
## NULL             -83.1
## GSE28735.axis1  -81.4  3.43  1      0.064
## GSE28735.axis2  -80.2  2.51  1      0.113
```

```
load("../../data/validation/tcga-clin-gex.20141118.rda")
```

```
doValForSingleCancer = function(cancer_id)
{
        # nevents, ntotal, score_p, anova_pcna, anova_score, anova_axis1, anova_axis2
        message(cancer_id)
        cancer_data = data.merged[[cancer_id]]
        if (!"illuminahiseq_rnaseqv2" %in% names(cancer_data$gex))      { return(c(0, 0, NA, NA, NA, NA,

        gex = cancer_data$gex$illuminahiseq_rnaseqv2
        clin = cancer_data$clin

        days_to_death = clin$days_to_death
        days_to_death[days_to_death == "[Not Applicable]"] = NA
        days_to_death = as.numeric(as.character(days_to_death))

        days_to_initial_pathologic_diagnosis = clin$days_to_initial_pathologic_diagnosis
```

```
        days_to_initial_pathologic_diagnosis[days_to_initial_pathologic_diagnosis == "[Not Applicable]"]
        days_to_initial_pathologic_diagnosis = as.numeric(as.character(days_to_initial_pathologic_diagno

        days_to_last_followup = clin$days_to_last_followup
        days_to_last_followup[days_to_last_followup == "[Not Applicable]"] = NA
        days_to_last_followup = as.numeric(as.character(days_to_last_followup))

        time_event = days_to_death - days_to_initial_pathologic_diagnosis
        time_lfu = days_to_last_followup - days_to_initial_pathologic_diagnosis
        time_obs = time_event
        time_obs[is.na(time_obs)] = time_lfu[is.na(time_obs)]
        time_obs[!is.na(time_obs) & !is.na(time_lfu)] = pmin(time_obs[!is.na(time_obs) & !is.na(time_lfu
        event = (time_event <= time_lfu & !is.na(time_event) & !is.na(time_lfu)) | (!is.na(time_event) &

        y = Surv(time_obs, event)

        gex = gex[!grepl("^\\?\\|", rownames(gex)),]
        rownames(gex) = gsub("\\|.*", "", rownames(gex))

        pcna = apply(log2(gex[rownames(gex) %in% metapcna.sig,] + 1), 2, median)

        gex.axes = gex[match(rownames(val.basis), rownames(gex)),]
        gex.axes[apply(is.na(gex.axes), 1, all),] = 0
        gex.axes = gex.axes - apply(gex.axes, 1, min, na.rm = TRUE)
        gex.axes = gex.axes / apply(gex.axes, 1, max, na.rm = TRUE)
        gex.axes[is.na(gex.axes)] = 0

        coefs = apply(gex.axes, 2, function(xcol) nnls(val.basis, xcol)$x)

        axis1 = coefs[1,] - coefs[5,]
        axis2 = coefs[6,] - coefs[2,]
        score = 1.354*axis1 + 1.548*axis2

        valid = !is.na(score) & !is.na(pcna) & !is.na(y[,1]) & !is.na(y[,2])
        axis1 = axis1[valid]
        axis2 = axis2[valid]
        score = score[valid]
        pcna = pcna[valid]
        y = y[valid,]

        nevents = sum(y[,2])
        ntotal = nrow(y)

        score_p = pchisq(2*diff(coxph(y ~ score)$loglik), 1, lower.tail = FALSE)
        anova_pcna = anova(coxph(y ~ pcna + score))[,"Pr(>|Chi|)"][2]
        anova_score = anova(coxph(y ~ pcna + score))[,"Pr(>|Chi|)"][3]
        anova_axis1 = anova(coxph(y ~ axis1 + axis2))[,"Pr(>|Chi|)"][2]
        anova_axis2 = anova(coxph(y ~ axis1 + axis2))[,"Pr(>|Chi|)"][3]

        c(nevents, ntotal, score_p, anova_pcna, anova_score, anova_axis1, anova_axis2)
}
```

```r
val_pvals = sapply(names(data.merged), doValForSingleCancer)

## acc
## blca
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## brca
## cesc
## coad
## dlbc
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :  Loglik converged
before variable  1,2 ; beta may be infinite.
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :  Loglik converged
before variable  1,2 ; beta may be infinite.
## gbm
## hnsc
## kich
## kirc
## kirp
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## lgg
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## lihc
## luad
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## lusc
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## meso
## ov
## paad
## prad
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :  Ran out of
iterations and did not converge
## Warning in fitter(X, Y, strats, offset, init, control, weights = weights, :  Ran out of
iterations and did not converge
## read
## sarc
## skcm
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## thca
## ucec
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
```

```
## Warning in FUN(c("acc", "blca", "brca", "cesc", "coad", "dlbc", "gbm", "hnsc", :  NAs introduced
by coercion
## ucs
```

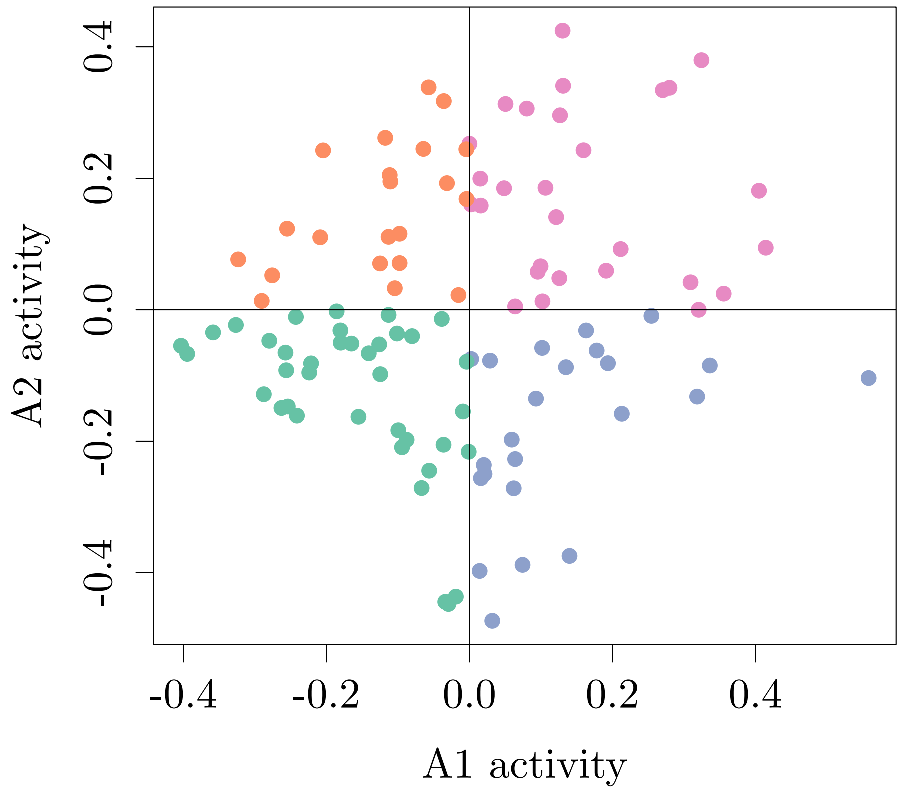
```
rownames(val_pvals) = c("nevents", "ntotal", "p.score", "p.anova.pcna", "p.anova.pcna_score", "p.anova.a
val_pvals = as.data.frame(t(val_pvals))

val_pvals[val_pvals$nevents >= 50 | rownames(val_pvals) == "paad",]
```

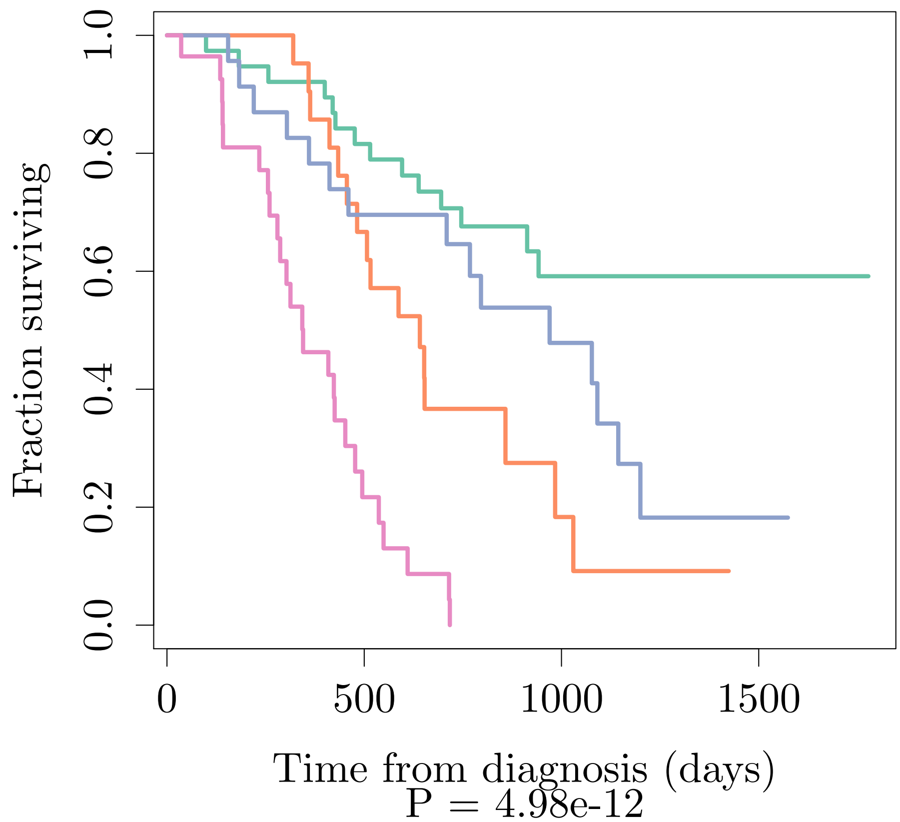
```
##      nevents ntotal   p.score p.anova.pcna p.anova.pcna_score
## gbm       54    143 2.287e-01    8.185e-01          0.1587102
## hnsc     124    367 8.075e-03    4.719e-01          0.0107907
## kirc     153    497 2.034e-12    9.569e-11          0.0028892
## lgg       53    272 1.493e-05    6.316e-04          0.0078542
## luad     106    431 8.336e-06    7.205e-03          0.0001042
## lusc     117    395 9.624e-01    7.035e-02          0.4109578
## ov       115    251 2.380e-02    5.903e-01          0.0178108
## paad      17     58 4.952e-03    8.549e-02          0.0239990
##      p.anova.axis1 p.anova.axis1_axis2
## gbm      9.252e-01           6.877e-02
## hnsc     4.367e-02           8.341e-02
## kirc     2.673e-08           1.639e-05
## lgg      1.593e-04           3.350e-02
## luad     1.238e-03           1.543e-03
## lusc     1.597e-01           2.559e-01
## ov       3.655e-01           3.298e-02
## paad     1.562e-02           1.249e-01
```

```
plot_km_axes = function(axis1, axis2, y, mc = FALSE, ...)
{
        t1 = t2 = 0
        if (mc == TRUE)
        {
                t1 = median(axis1, na.rm = TRUE)
                t2 = median(axis2, na.rm = TRUE)
        }

        class = paste(c("L", "H")[I(axis1 >= t1)+1], c("L", "H")[I(axis2 >= t2)+1], sep = "")
        class = ordered(class, levels = c("LL", "LH", "HL", "HH"))
        fit = survfit(y ~ class)
        print(fit)
        print(survdiff(y ~ class))
        pval = pchisq(survdiff(y ~ class)$chisq, 3, lower.tail = FALSE)
        pal = brewer.pal(4, "Set2")
        names(pal) = c("LL", "LH", "HL", "HH")
        plot(axis2 ~ axis1, xlab = "A1 activity", ylab = "A2 activity", col = pal[class], pch = 16, ...)
        abline(h = t2)
        abline(v = t1)
        #plot(fit, col = pal, lwd = 2, xlab = "Time from diagnosis (days)", ylab = "Fraction surviving",
        plot(fit$surv ~ fit$time, type = "n", ylim = c(0, 1), xlab = "Time from diagnosis (days)", ylab
        for (i in 1:length(fit$strata))
        {
                stratum_start = sum(fit$strata[1:i]) - fit$strata[i] + 1
```

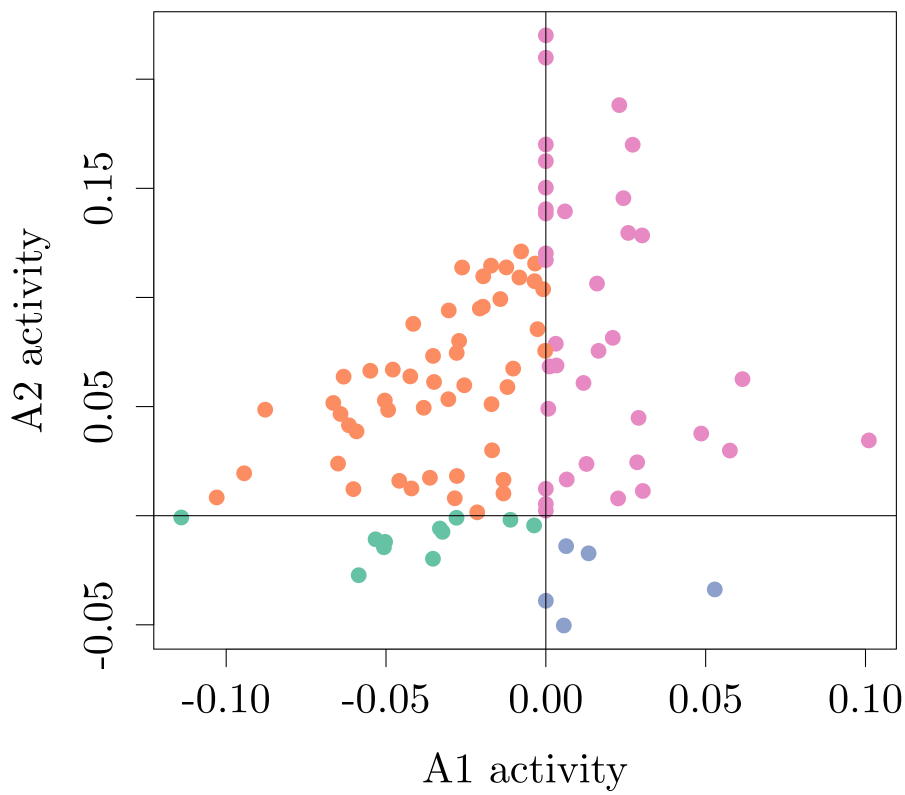
```
                stratum_end = stratum_start + fit$strata[i] - 1
                stratum_surv = c(1, fit$surv[stratum_start:stratum_end])
                stratum_time = c(0, fit$time[stratum_start:stratum_end])
                lines(stratum_surv ~ stratum_time, col = pal[i], type = "s", ...)
        }
}

plot_km_axes_tcga = function(code, mc, ...)
{
        if ("illuminahiseq_rnaseqv2" %in% names(data.merged[[code]]$gex))
        {
                temp.gex = data.merged[[code]]$gex$illuminahiseq_rnaseqv2
                temp.gex = temp.gex[!grepl("^\\?\\|", rownames(temp.gex)),]
                rownames(temp.gex) = gsub("\\|.*", "", rownames(temp.gex))
                temp.gex.axes = temp.gex[match(rownames(val.basis), rownames(temp.gex)),]
                temp.gex.axes[apply(is.na(temp.gex.axes), 1, all),] = 0
                temp.gex.axes = temp.gex.axes - apply(temp.gex.axes, 1, min, na.rm = TRUE)
                temp.gex.axes = temp.gex.axes / apply(temp.gex.axes, 1, max, na.rm = TRUE)
                temp.gex.axes[is.na(temp.gex.axes)] = 0
                temp.coefs = apply(temp.gex.axes, 2, function(xcol) nnls(val.basis, xcol)$x)
                temp.axis1 = temp.coefs[1,] - temp.coefs[5,]
                temp.axis2 = temp.coefs[6,] - temp.coefs[2,]

                temp.clin = data.merged[[code]]$clin
                temp.days_to_death = temp.clin$days_to_death
                temp.days_to_death[temp.days_to_death == "[Not Applicable]"] = NA
                temp.days_to_death = as.numeric(as.character(temp.days_to_death))
                temp.days_to_initial_pathologic_diagnosis = temp.clin$days_to_initial_pathologic_diagnos
                temp.days_to_initial_pathologic_diagnosis[temp.days_to_initial_pathologic_diagnosis == '
                temp.days_to_initial_pathologic_diagnosis = as.numeric(as.character(temp.days_to_initial
                temp.days_to_last_followup = temp.clin$days_to_last_followup
                temp.days_to_last_followup[temp.days_to_last_followup == "[Not Applicable]"] = NA
                temp.days_to_last_followup = as.numeric(as.character(temp.days_to_last_followup))
                temp.time_event = temp.days_to_death - temp.days_to_initial_pathologic_diagnosis
                temp.time_lfu = temp.days_to_last_followup - temp.days_to_initial_pathologic_diagnosis
                temp.time_obs = temp.time_event
                temp.time_obs[is.na(temp.time_obs)] = temp.time_lfu[is.na(temp.time_obs)]
                temp.time_obs[!is.na(temp.time_obs) & !is.na(temp.time_lfu)] = pmin(temp.time_obs[!is.na
                temp.event = (temp.time_event <= temp.time_lfu & !is.na(temp.time_event) & !is.na(temp.t
                temp.y = Surv(temp.time_obs, temp.event)

                plot_km_axes(temp.axis1, temp.axis2, temp.y, mc = FALSE, main = "", ...)
        }
}

plot_km_axes(axis_coefs.diag_dsd[,1], axis_coefs.diag_dsd[,2], y.diag_dsd, mc = FALSE, main = "", lwd =

## Call: survfit(formula = y ~ class)
##
##          records n.max n.start events median 0.95LCL 0.95UCL
## class=LL      38    38      38     14     NA     913      NA
## class=LH      21    21      21     16    641     482      NA
## class=HL      23    23      23     15    970     709      NA
## class=HH      28    28      28     25    345     280     495
```

```
## Call:
## survdiff(formula = y ~ class)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL 38       14    31.77     9.942    18.645
## class=LH 21       16    13.02     0.683     0.852
## class=HL 23       15    17.66     0.401     0.543
## class=HH 28       25     7.55    40.376    48.292
##
##  Chisq= 55.7  on 3 degrees of freedom, p= 4.98e-12
```

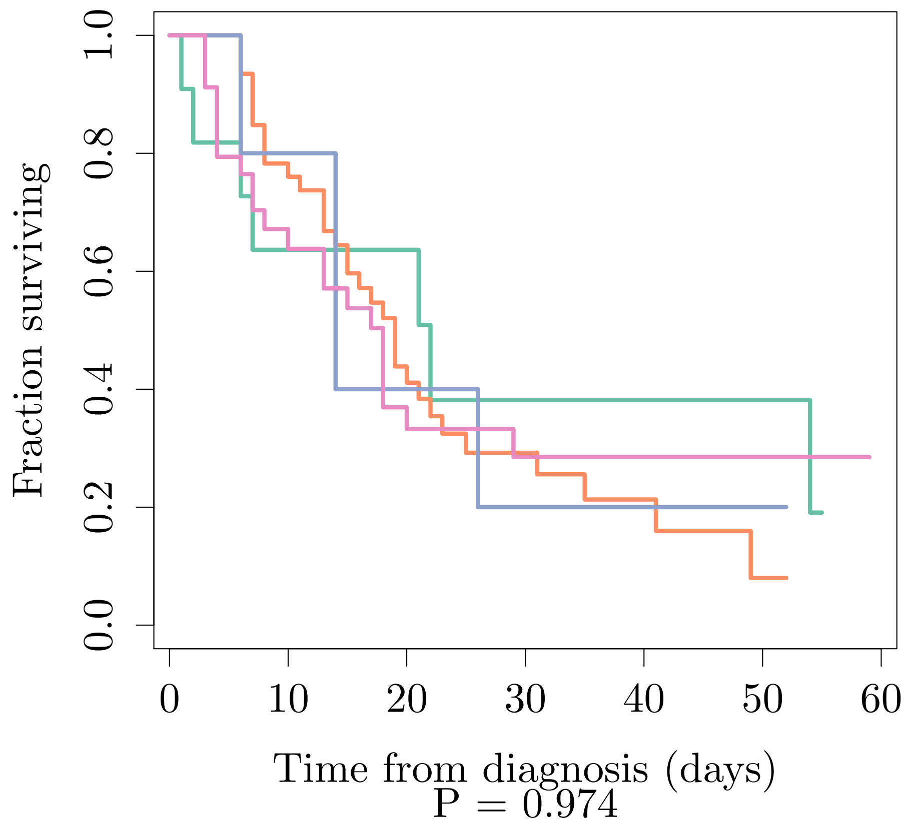




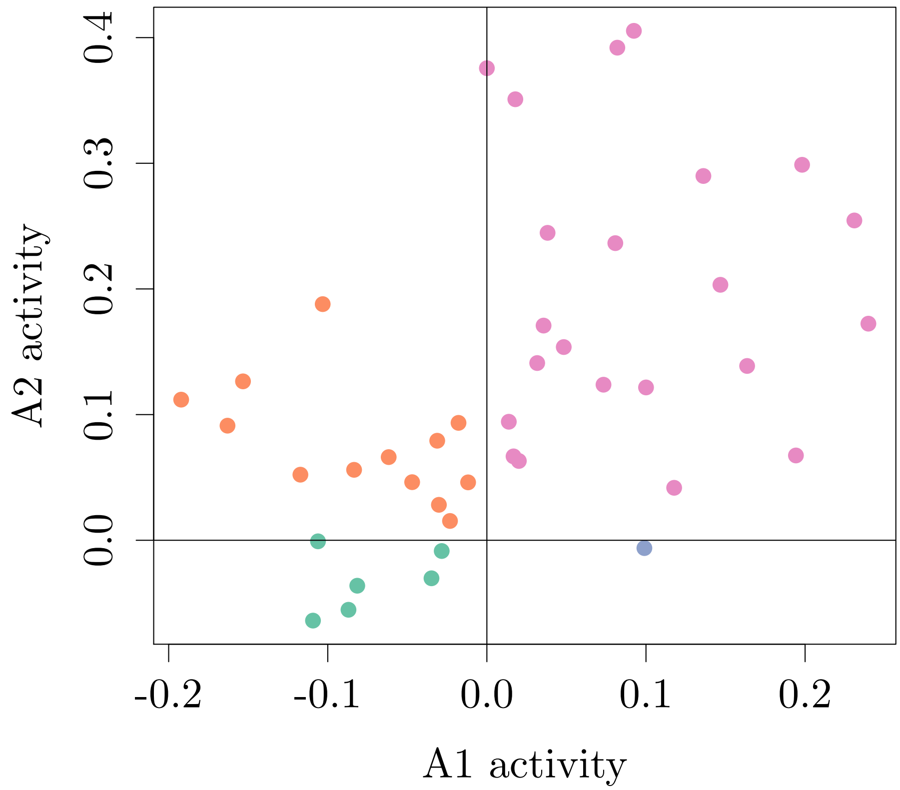

```
plot_km_axes(GSE21501.axis1, GSE21501.axis2, Surv(GSE21501.samp$time, GSE21501.samp$event), mc = FALSE,

## Call: survfit(formula = y ~ class)
##
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL       11    11      11      7     22       7      NA
## class=LH       50    50      50     33     19      15      25
## class=HL        5     5       5      4     14      14      NA
## class=HH       36    36      36     22     18      10      NA
## Call:
## survdiff(formula = y ~ class)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL 11        7     8.12  1.56e-01  1.95e-01
## class=LH 50       33    32.96  3.96e-05  8.43e-05
## class=HL  5        4     3.91  2.27e-03  2.53e-03
## class=HH 36       22    21.01  4.71e-02  7.23e-02
##
##  Chisq= 0.2  on 3 degrees of freedom, p= 0.974
```

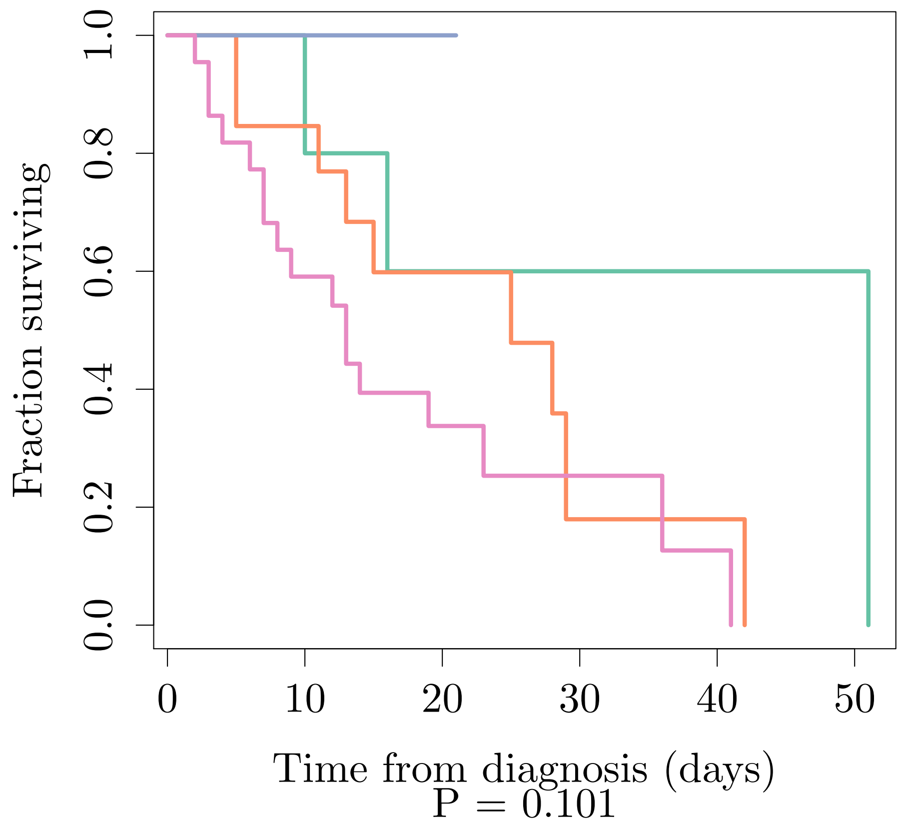


A2 activity
-0.05
0.05
0.15
-0.10
-0.05
0.00
0.05
0.10
A1 activity

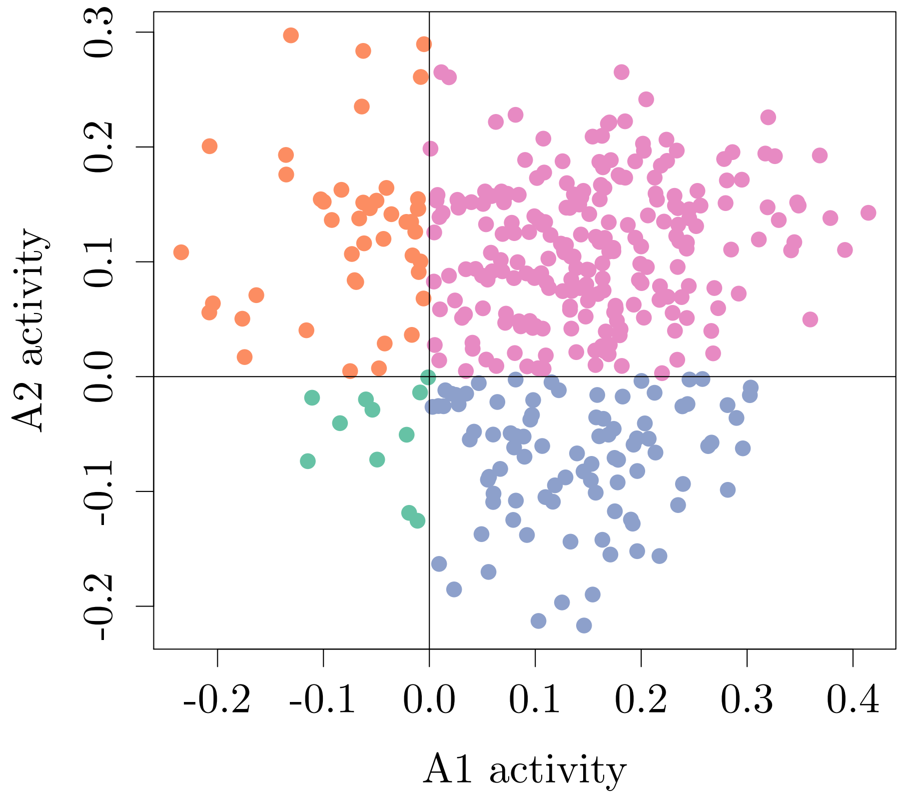


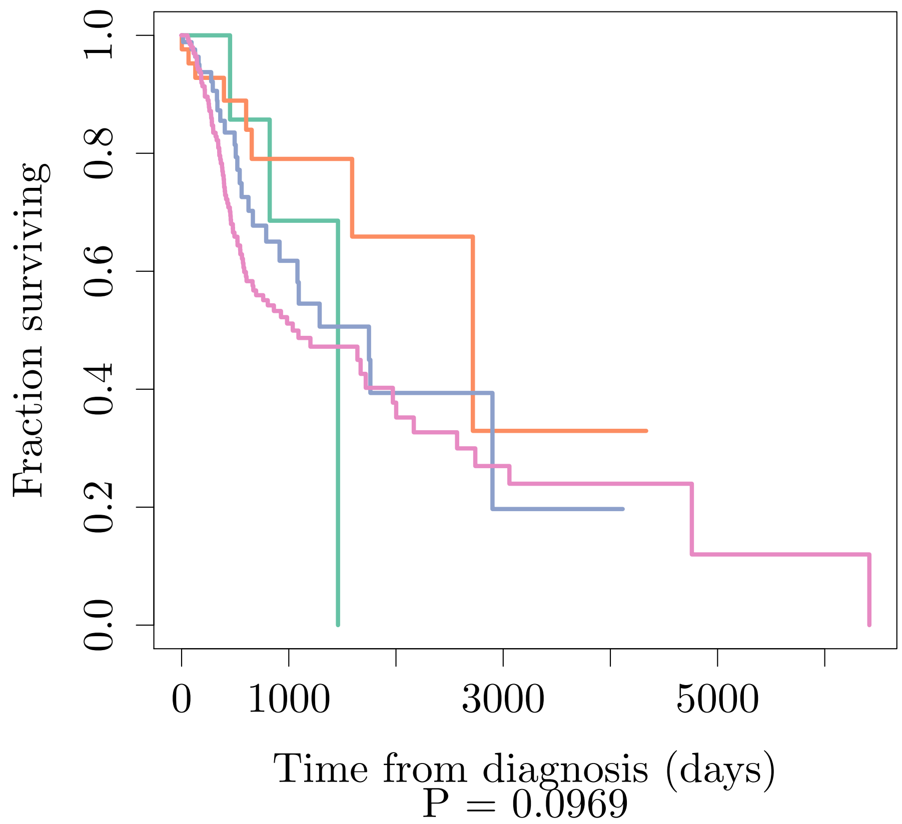

```
plot_km_axes(GSE28735.axis1, GSE28735.axis2, Surv(GSE28735.samp$time, GSE28735.samp$event), mc = FALSE,

## Call: survfit(formula = y ~ class)
##
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL        6     6       6      3     51      16      NA
## class=LH       13    13      13      9     25      13      NA
## class=HL        1     1       1      0     NA      NA      NA
## class=HH       22    22      22     17     13       8      NA
## Call:
## survdiff(formula = y ~ class)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL  6        3    6.651     2.004     3.168
## class=LH 13        9   10.078     0.115     0.186
## class=HL  1        0    0.735     0.735     0.779
## class=HH 22       17   11.536     2.588     4.733
##
##  Chisq= 6.2  on 3 degrees of freedom, p= 0.101
```


A2 activity
0.4
0.3
0.2
0.1
0.0
-0.2
-0.1
0.0
0.1
0.2
A1 activity



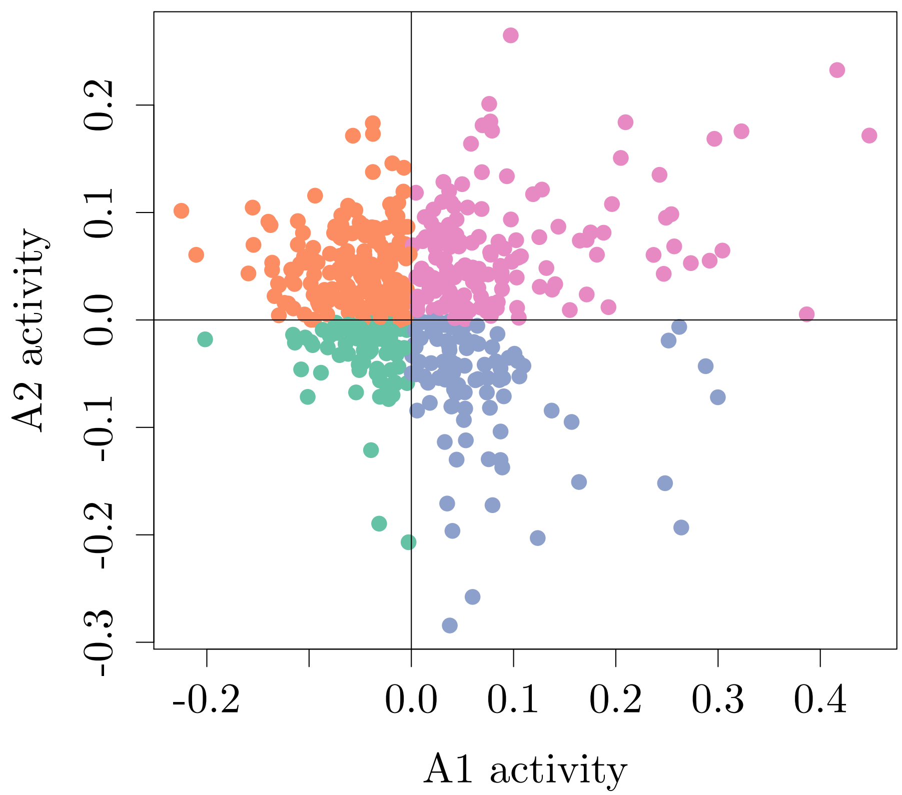

```
plot_km_axes_tcga("hnsc", mc = FALSE, lwd = 4, cex = 1.2)

## Call: survfit(formula = y ~ class)
##
##    1 observation deleted due to missingness
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL       11    11      11      3   1459     822      NA
## class=LH       43    43      43      8   2717    1591      NA
## class=HL       90    90      90     26   1748     914      NA
## class=HH      223   223     223     87   1037     669    2002
## Call:
## survdiff(formula = y ~ class)
##
## n=367, 1 observation deleted due to missingness.
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL   11        3     3.77     0.156     0.162
## class=LH   43        8    15.60     3.700     4.249
## class=HL   90       26    29.68     0.457     0.606
## class=HH  223       87    74.95     1.936     4.969
##
##  Chisq= 6.3  on 3 degrees of freedom, p= 0.0969
```

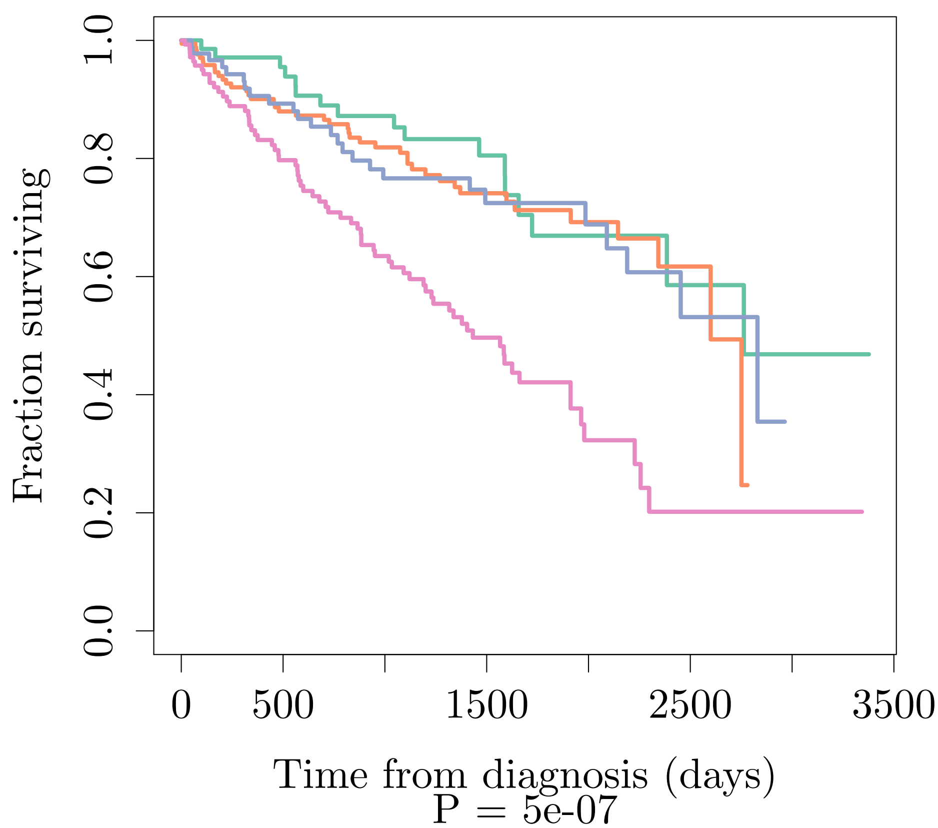

0.3
0.2
0.1
0.0
-0.1
-0.2
A2 activity
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
A1 activity



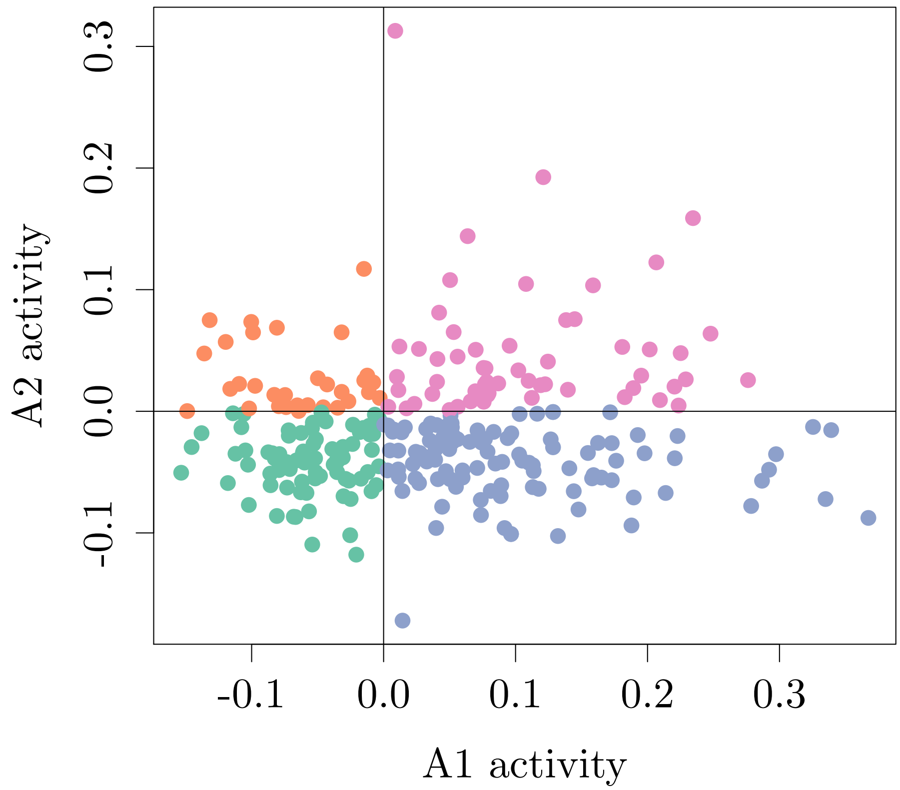

```r
plot_km_axes_tcga("kirc", mc = FALSE, lwd = 4, cex = 1.2)
```

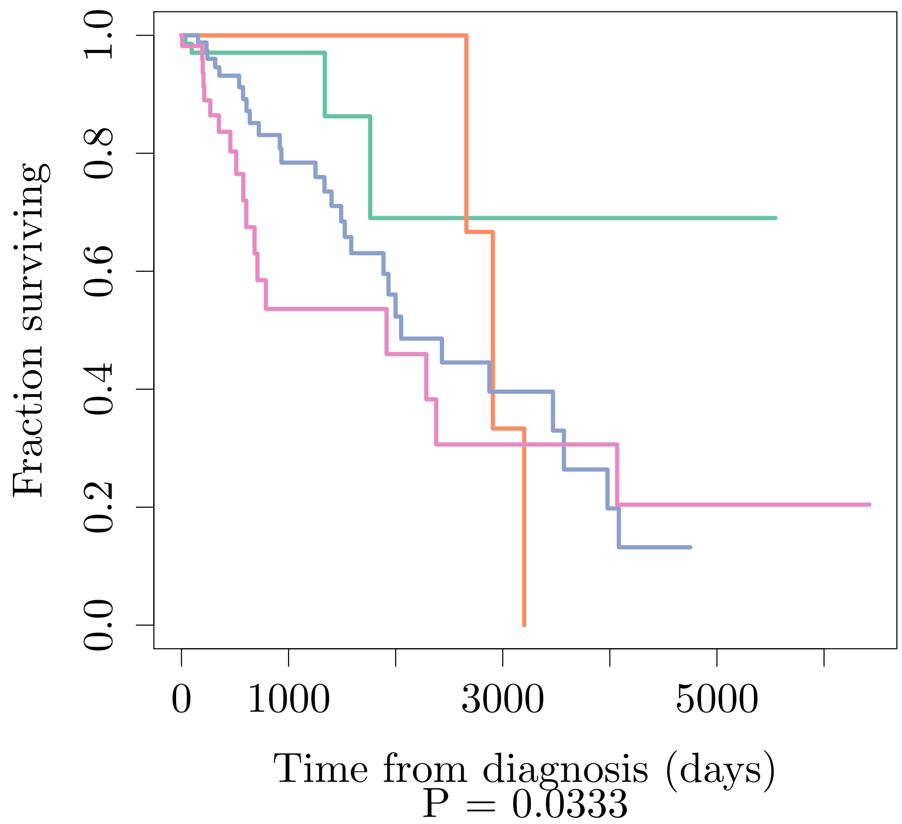
```
## Call: survfit(formula = y ~ class)
##
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL       77    77      77     17   2763    2385      NA
## class=LH      179   179     179     42   2600    2343      NA
## class=HL       96    96      96     25   2830    2190      NA
## class=HH      146   146     146     69   1432    1200    1964
## Call:
## survdiff(formula = y ~ class)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL  77       17     26.8      3.59      4.39
## class=LH 179       42     55.3      3.21      5.08
## class=HL  96       25     32.0      1.54      1.96
## class=HH 146       69     38.8     23.44     31.71
##
##  Chisq= 32.1  on 3 degrees of freedom, p= 5e-07
```


A2 activity
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.2
0.0
0.1
0.2
0.3
0.4
A1 activity



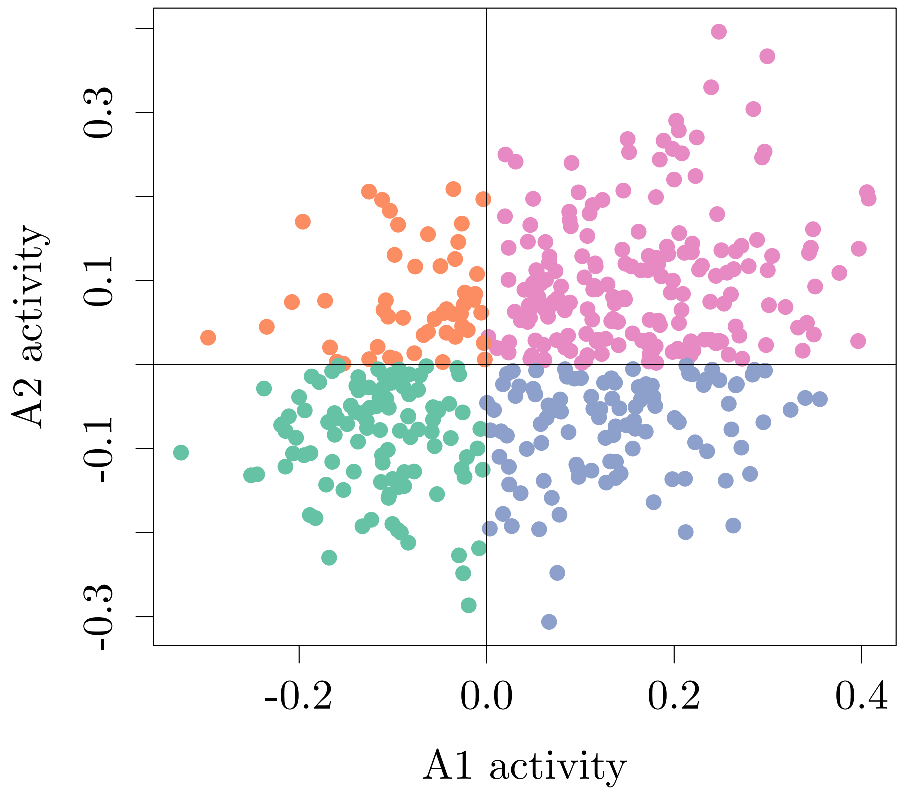

```
plot_km_axes_tcga("lgg", mc = FALSE, lwd = 4, cex = 1.2)

## Warning in plot_km_axes_tcga("lgg", mc = FALSE, lwd = 4, cex = 1.2):  NAs introduced by coercion

## Call: survfit(formula = y ~ class)
##
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL       77    77      77      4     NA    1762      NA
## class=LH       32    32      32      3   2907    2660      NA
## class=HL      106   106     106     28   2051    1886    3978
## class=HH       57    57      57     18   1915     682      NA
## Call:
## survdiff(formula = y ~ class)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL   77        4     9.86     3.479     4.389
## class=LH   32        3     5.38     1.055     1.191
## class=HL  106       28    26.36     0.102     0.207
## class=HH   57       18    11.40     3.815     4.950
##
##  Chisq= 8.7  on 3 degrees of freedom, p= 0.0333
```

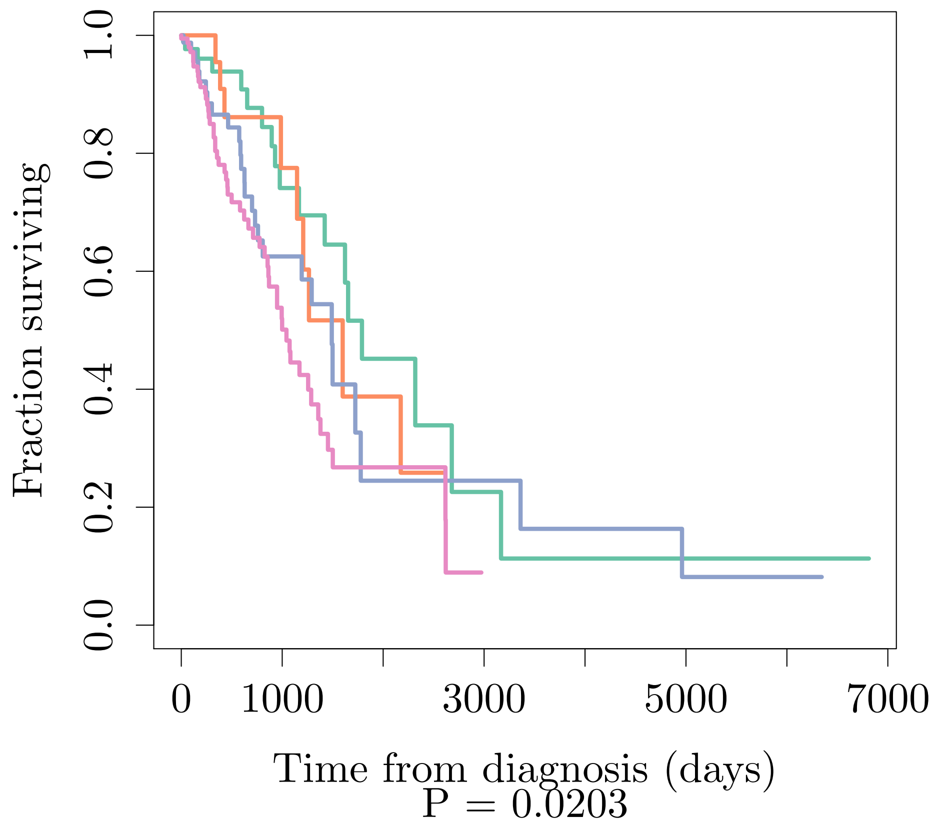

0.3
0.2
0.1
0.0
-0.1
A2 activity
-0.1
0.0
0.1
0.2
0.3
A1 activity



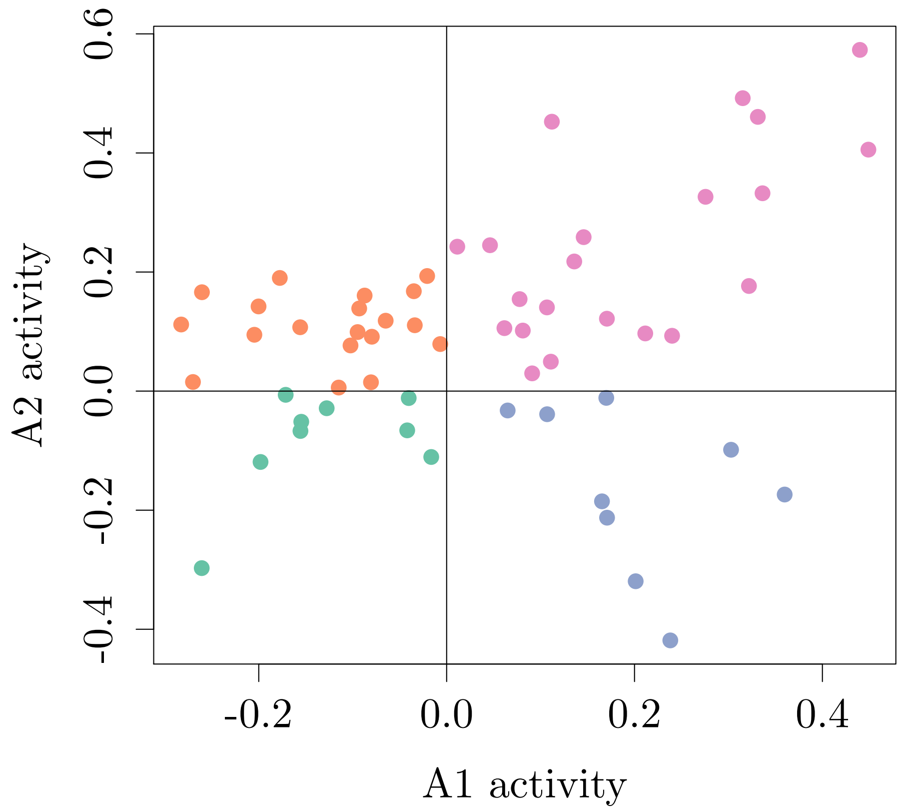

```
plot_km_axes_tcga("luad", mc = FALSE, lwd = 4, cex = 1.2)

## Warning in plot_km_axes_tcga("luad", mc = FALSE, lwd = 4, cex = 1.2):  NAs introduced by
coercion

## Call: survfit(formula = y ~ class)
##
##    19 observations deleted due to missingness
##          records n.max n.start events median 0.95LCL 0.95UCL
## class=LL     102   102     102     18   1790    1421      NA
## class=LH      49    49      49      9   1599    1147      NA
## class=HL      98    98      98     26   1491     807      NA
## class=HH     182   182     182     53   1042     863    1379
## Call:
## survdiff(formula = y ~ class)
##
## n=431, 19 observations deleted due to missingness.
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL 102       18     28.0    3.5523     4.875
## class=LH  49        9     11.9    0.6911     0.786
## class=HL  98       26     27.5    0.0801     0.112
## class=HH 182       53     38.7    5.2967     8.567
##
```

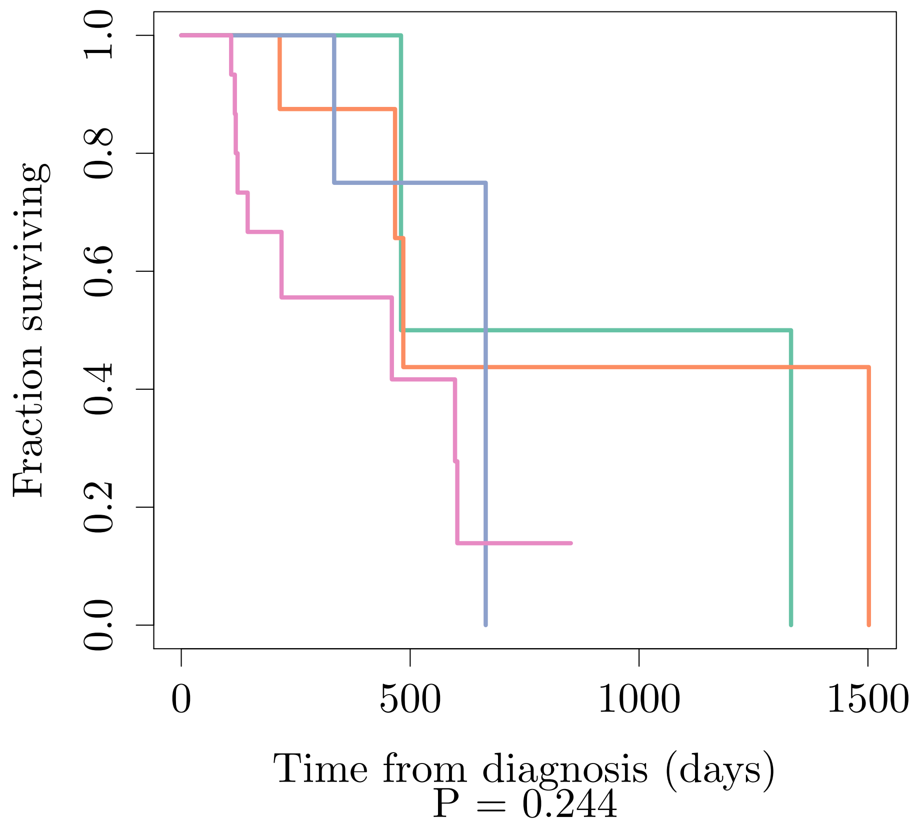
```
##  Chisq= 9.8  on 3 degrees of freedom, p= 0.0203
```






```
plot_km_axes_tcga("paad", mc = FALSE, lwd = 4, cex = 1.2)

## Call: survfit(formula = y ~ class)
##
##           records n.max n.start events median 0.95LCL 0.95UCL
## class=LL        9     9       9      2    906     480      NA
## class=LH       19    19      19      4    485     467      NA
## class=HL        9     9       9      2    665     334      NA
## class=HH       21    21      21      9    460     145      NA
## Call:
## survdiff(formula = y ~ class)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## class=LL  9        2     3.13     0.408     0.536
## class=LH 19        4     5.86     0.592     1.052
## class=HL  9        2     2.71     0.187     0.233
## class=HH 21        9     5.29     2.593     4.105
##
##  Chisq= 4.2  on 3 degrees of freedom, p= 0.244
```


0.6
0.4
0.2
0.0
-0.2
-0.4
A2 activity
-0.2
0.0
0.2
0.4
A1 activity




```
sig.emt_groger_pos = c("ADAM12", "COL1A1", "COL3A1", "COL5A1", "COL6A1", "COL6A3", "CTGF", "CYP1B1", "DI
sig.emt_groger_neg = c("CD24","CXCL16","DSG3","ELF3","EPCAM","EPHA1","JUP","MPZL2","OVOL2","PLXNB1","S10
x.medcent = x - apply(x, 1, median)
emt.scores.pos = apply(x.medcent[rownames(x.medcent) %in% sig.emt_groger_pos,], 2, median)
emt.scores.neg = apply(x.medcent[rownames(x.medcent) %in% sig.emt_groger_neg,], 2, median)
emt.scores = emt.scores.pos - emt.scores.neg
```

4.4 MSigDB score correlation thresholding

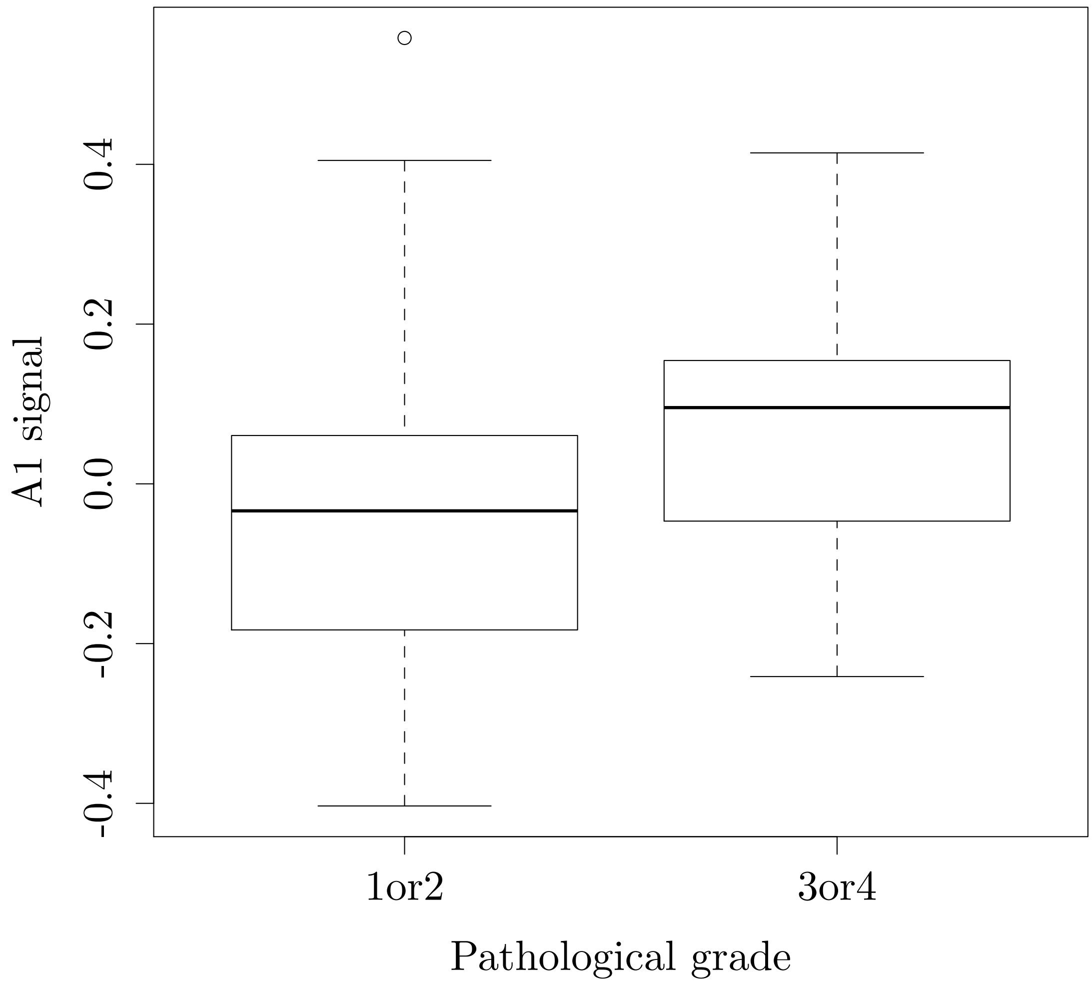
```
axis_coefs.msigdb.corr = cor(axis_coefs.diag_dsd, t(sigs), method = "kendall")

temp.sel_cols = apply(abs(axis_coefs.msigdb.corr) >= sig.corr.threshold, 2, any)
#heatmap.2(axis_coefs.msigdb.corr[, temp.sel_cols], trace = "none", scale = "none", useRaster = TRUE, co
#heatmap.2(axis_coefs.msigdb.corr[, temp.sel_cols], trace = "none", scale = "none", useRaster = TRUE, co

cpv.pvals = apply(axis_coefs.diag_dsd, 2, function(mg) sapply(cbind(cpvs.diag_dsd, purity = samps.diag_c
        s = !is.na(mg) & !is.na(x)
        x = x[s]
        mg = mg[s]
        if (any(c("numeric", "integer") %in% class(x)))
        {
```

```r
                return(cor.test(x, mg, method = "pearson")$p.value)
        }
        else if (any(c("factor", "ordered", "logical") %in% class(x)) && length(unique(x)) > 1)
        {
                return(anova(lm(mg ~ x))[,"Pr(>F)"][1])
        }
        NA
}))
cpv.pvals = cpv.pvals[!apply(is.na(cpv.pvals), 1, all),]
cpv.pvals = cpv.pvals[!grepl("^Surv\\.", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^Treat\\.", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^Path\\.Nodes", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^Staging\\.Version", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^History\\.Recurrence$", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^History\\.Status$", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^History\\.Death\\.Cause$", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^Path\\.Grade$", rownames(cpv.pvals)),]
cpv.pvals = cpv.pvals[!grepl("^Path\\.TumourLocation$", rownames(cpv.pvals)),]

temp = as.vector(cpv.pvals)
temp = p.adjust(temp, "holm")
cpv.qvals = matrix(temp, nrow = nrow(cpv.pvals))
rownames(cpv.qvals) = rownames(cpv.pvals)
colnames(cpv.qvals) = colnames(cpv.pvals)

cpv.pvals

##                                          axis1     axis2
## Patient.Gender                       0.1581541 0.0098535
## Patient.Ethnicity                    0.7711156 0.1130046
## History.Smoking.PackYears            0.3562152 0.2753851
## History.Diagnosis.AgeAtYears         0.9250804 0.6658699
## Path.HistoType.Subtype               0.6966533 0.1569139
## Path.TumourSizeMm                    0.8438715 0.1709600
## Path.Invasion.PN                     0.0951996 0.2251091
## Path.Invasion.VS                     0.6500594 0.0707968
## Staging.pM                           0.4414498 0.4245233
## Staging.pN                           0.2524195 0.2629997
## Staging.pT                           0.2640385 0.4273685
## Staging.Stage                        0.0605854 0.2355348
## History.Recurrence.Site.Peritoneum   0.9162045 0.0149891
## History.Recurrence.Site.PancRemnant 0.5341395 0.1839586
## History.Recurrence.Site.PancBed      0.8869735 0.5303110
## History.Recurrence.Site.Other        0.1930828 0.1614602
## History.Recurrence.Site.Omentum      0.1388378 0.0820434
## History.Recurrence.Site.Mesentery    0.9326763 0.1206991
## History.Recurrence.Site.LymphNodes   0.9332622 0.8703023
## History.Recurrence.Site.Lung         0.3900712 0.7130517
## History.Recurrence.Site.Liver        0.1596616 0.1046158
## History.Recurrence.Site.Brain        0.4296978 0.0621650
## History.Recurrence.Site.Bone         0.7889803 0.4128670
## Path.Grade.Coarse                    0.0023854 0.0001297
## Path.TumourLocation.Coarse           0.1767526 0.1392750
## purity                               0.0002129 0.0004113
```

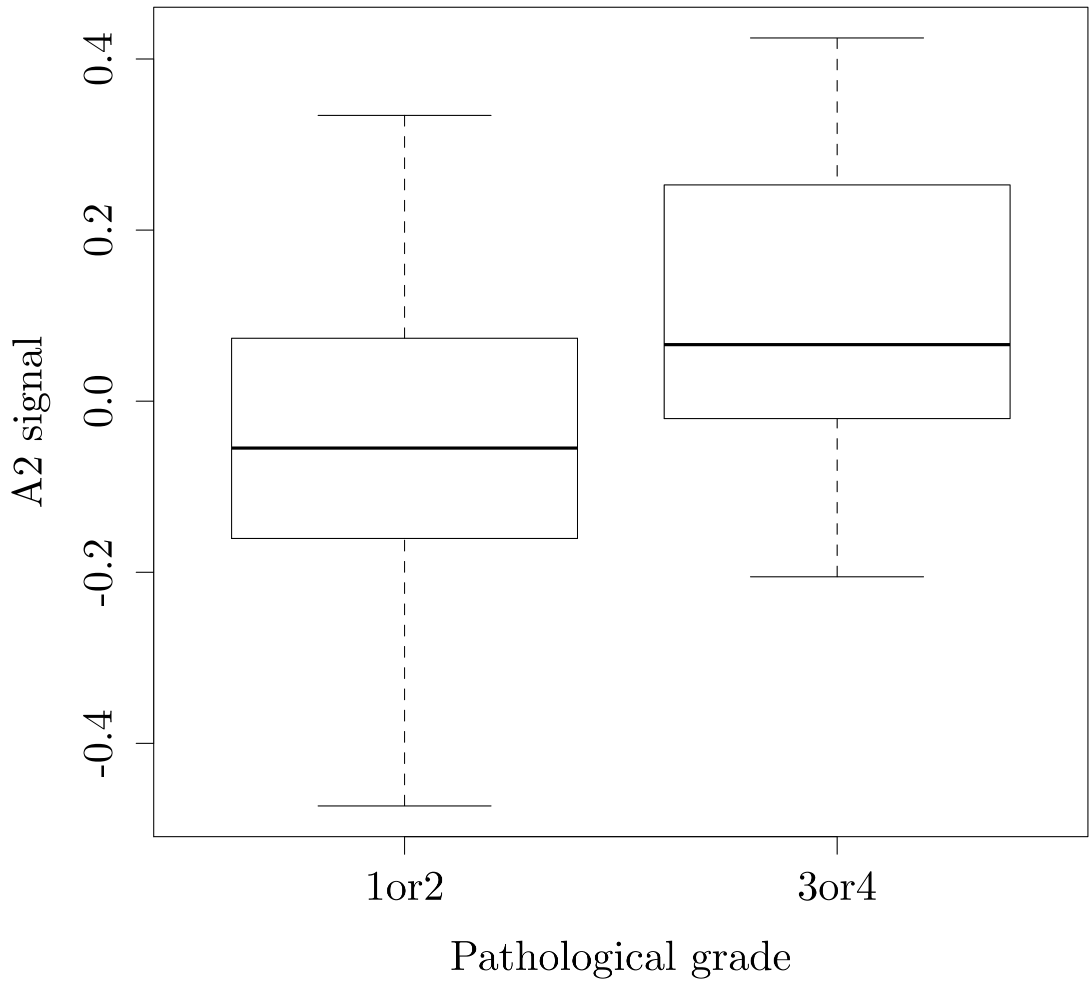
```r
cpv.qvals
```

```
##                                        axis1    axis2
## Patient.Gender                       1.00000 0.472968
## Patient.Ethnicity                    1.00000 1.000000
## History.Smoking.PackYears            1.00000 1.000000
## History.Diagnosis.AgeAtYears         1.00000 1.000000
## Path.HistoType.Subtype               1.00000 1.000000
## Path.TumourSizeMm                    1.00000 1.000000
## Path.Invasion.PN                     1.00000 1.000000
## Path.Invasion.VS                     1.00000 1.000000
## Staging.pM                           1.00000 1.000000
## Staging.pN                           1.00000 1.000000
## Staging.pT                           1.00000 1.000000
## Staging.Stage                        1.00000 1.000000
## History.Recurrence.Site.Peritoneum   1.00000 0.704486
## History.Recurrence.Site.PancRemnant 1.00000 1.000000
## History.Recurrence.Site.PancBed      1.00000 1.000000
## History.Recurrence.Site.Other        1.00000 1.000000
## History.Recurrence.Site.Omentum      1.00000 1.000000
## History.Recurrence.Site.Mesentery    1.00000 1.000000
## History.Recurrence.Site.LymphNodes   1.00000 1.000000
## History.Recurrence.Site.Lung         1.00000 1.000000
## History.Recurrence.Site.Liver        1.00000 1.000000
## History.Recurrence.Site.Brain        1.00000 1.000000
## History.Recurrence.Site.Bone         1.00000 1.000000
## Path.Grade.Coarse                    0.11688 0.006743
## Path.TumourLocation.Coarse           1.00000 1.000000
## purity                               0.01086 0.020564
```

```r
boxplot(axis_coefs.diag_dsd[,1] ~ cpvs.diag_dsd$Path.Grade.Coarse, xlab = "Pathological grade", ylab = '
```




```
boxplot(axis_coefs.diag_dsd[,2] ~ cpvs.diag_dsd$Path.Grade.Coarse, xlab = "Pathological grade", ylab = '
```



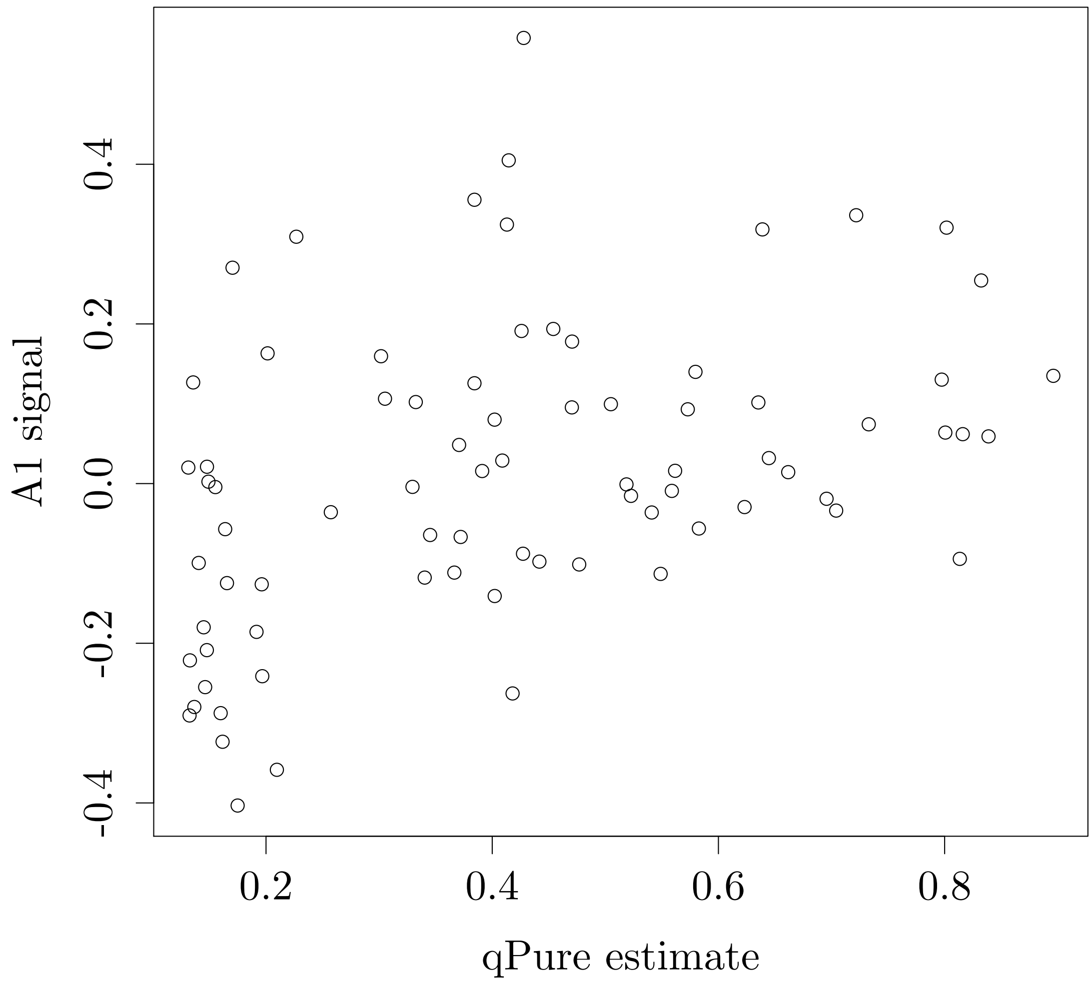

```
lm(axis_coefs.diag_dsd[,2] ~ cpvs.diag_dsd$Path.Grade.Coarse)

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 2] ~ cpvs.diag_dsd$Path.Grade.Coarse)
##
## Coefficients:
##                      (Intercept)  cpvs.diag_dsd$Path.Grade.Coarse.L
##                           0.0261                             0.1103

summary(lm(axis_coefs.diag_dsd[,2] ~ cpvs.diag_dsd$Path.Grade.Coarse))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 2] ~ cpvs.diag_dsd$Path.Grade.Coarse)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.4212 -0.1130 -0.0137  0.1372  0.3860
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.0261     0.0197    1.33  0.18771
## cpvs.diag_dsd$Path.Grade.Coarse.L   0.1103     0.0278    3.97  0.00013
##
## Residual standard error: 0.185 on 108 degrees of freedom
## Multiple R-squared:  0.127,Adjusted R-squared:  0.119
## F-statistic: 15.8 on 1 and 108 DF,  p-value: 0.00013
```

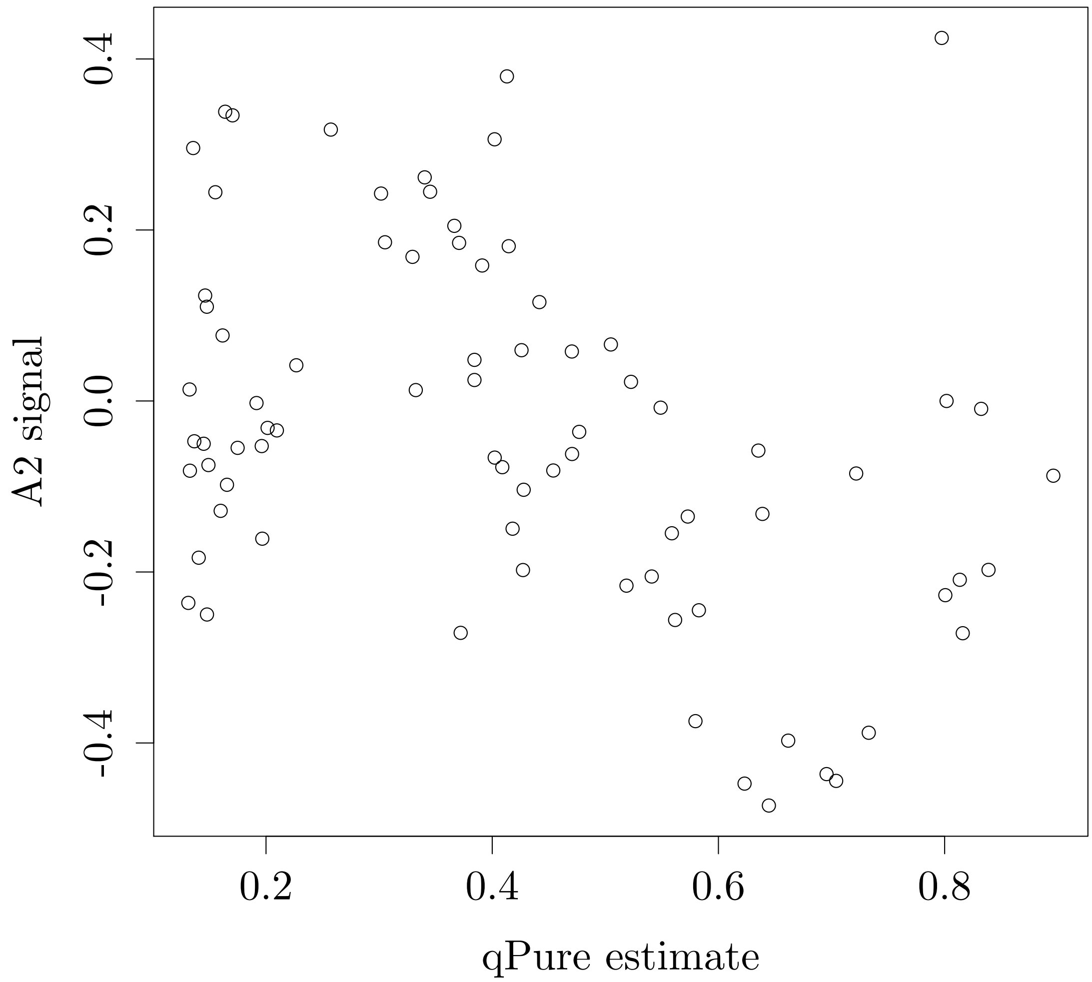
```
anova(lm(axis_coefs.diag_dsd[,2] ~ cpvs.diag_dsd$Path.Grade.Coarse))
```

```
## Analysis of Variance Table
##
## Response: axis_coefs.diag_dsd[, 2]
##                                  Df Sum Sq Mean Sq F value  Pr(>F)
## cpvs.diag_dsd$Path.Grade.Coarse   1   0.54   0.542    15.8 0.00013
## Residuals                       108   3.71   0.034
```

```
plot(axis_coefs.diag_dsd[,1] ~ samps$purity_qpure, xlab = "qPure estimate", ylab = "A1 signal")
```

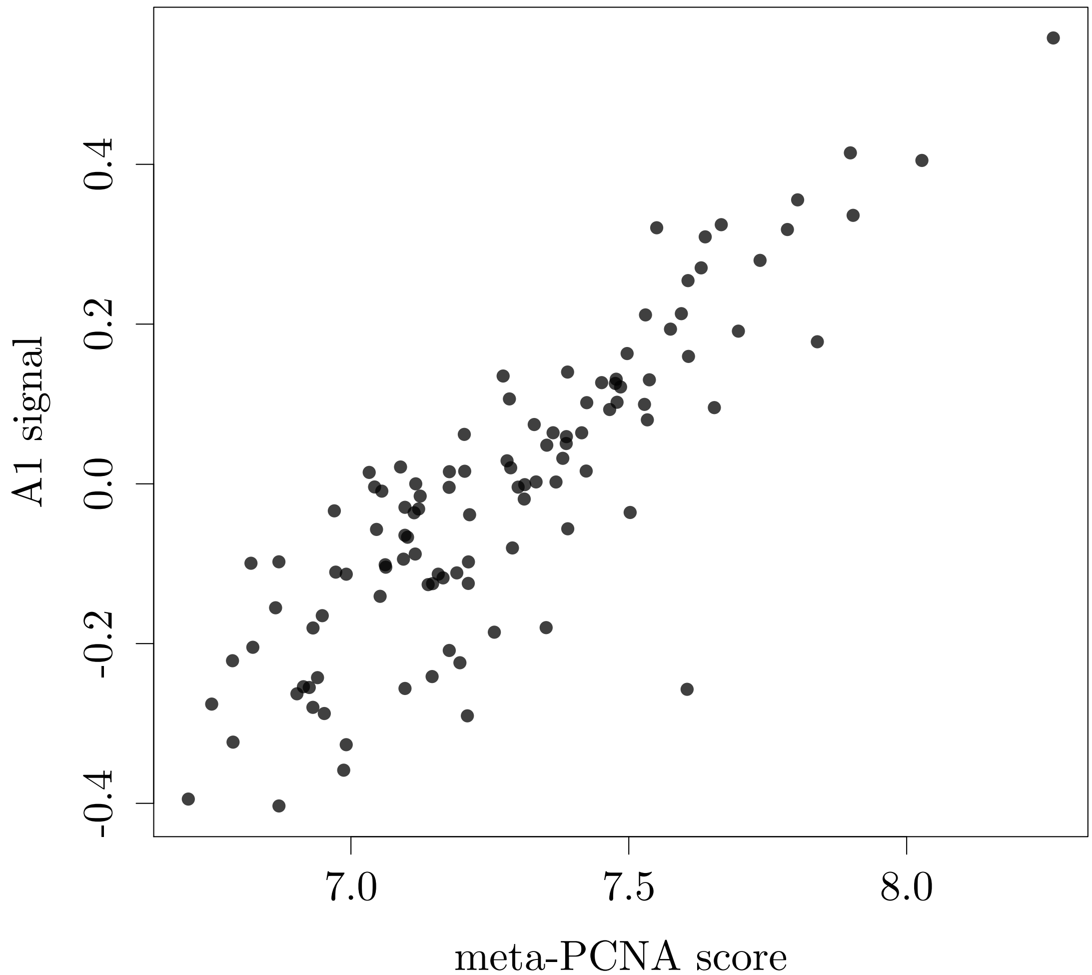



```
plot(axis_coefs.diag_dsd[,2] ~ samps$purity_qpure, xlab = "qPure estimate", ylab = "A2 signal")
```

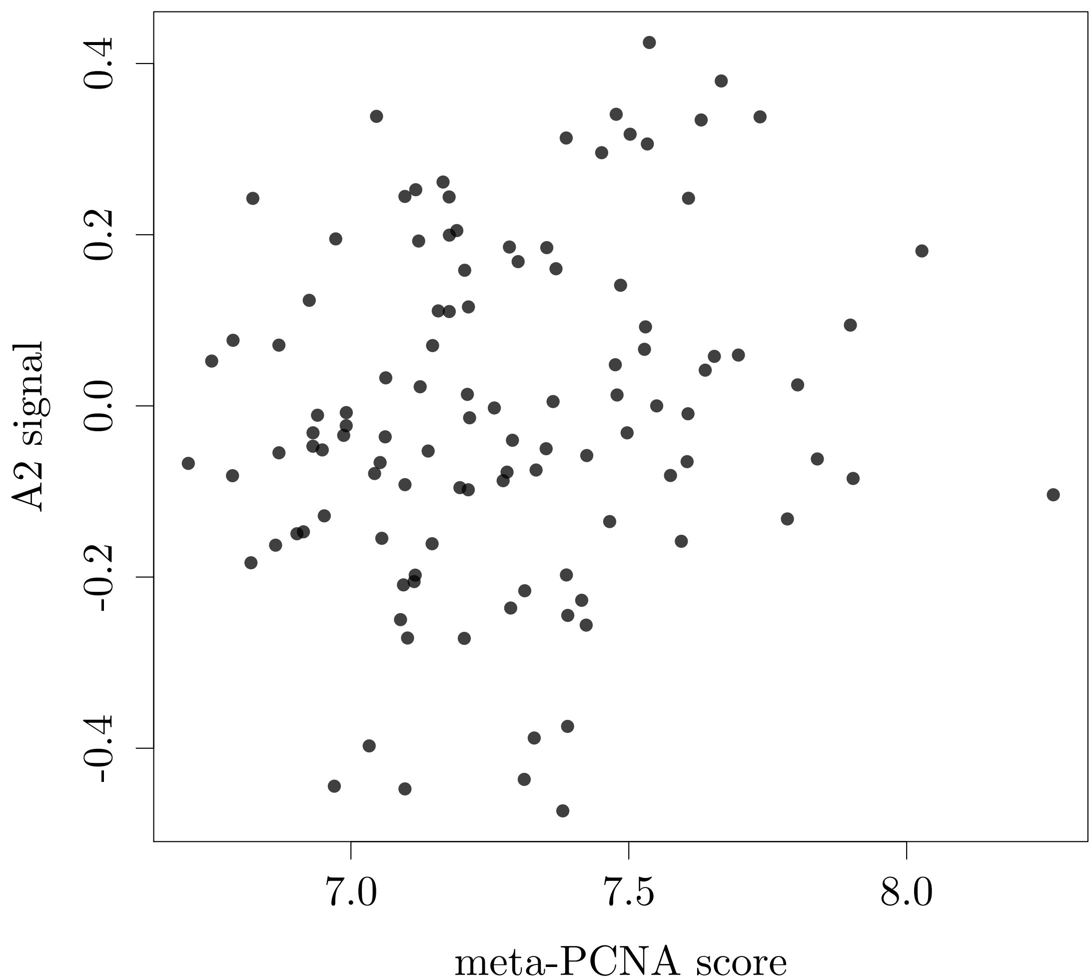



```
cor.test(axis_coefs.diag_dsd[,1], samps$purity_qpure, method = "kendall")

##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 1] and samps$purity_qpure
## z = 3.676, p-value = 0.0002369
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##    tau
## 0.2838

cor.test(axis_coefs.diag_dsd[,2], samps$purity_qpure, method = "kendall")

##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 2] and samps$purity_qpure
```

```
## z = -3.598, p-value = 0.0003203
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##     tau
## -0.2778

summary(lm(axis_coefs.diag_dsd[,1] ~ samps$purity_qpure))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 1] ~ samps$purity_qpure)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3318 -0.1172 -0.0469  0.1011  0.5422
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.132      0.042   -3.14  0.00240
## samps$purity_qpure    0.346      0.089    3.89  0.00021
##
## Residual standard error: 0.173 on 76 degrees of freedom
##   (32 observations deleted due to missingness)
## Multiple R-squared:  0.166,Adjusted R-squared:  0.155
## F-statistic: 15.1 on 1 and 76 DF,  p-value: 0.000213

summary(lm(axis_coefs.diag_dsd[,2] ~ samps$purity_qpure))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 2] ~ samps$purity_qpure)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3541 -0.1356 -0.0213  0.1531  0.6002
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.1195     0.0473    2.53  0.01363
## samps$purity_qpure  -0.3701     0.1001   -3.70  0.00041
##
## Residual standard error: 0.195 on 76 degrees of freedom
##   (32 observations deleted due to missingness)
## Multiple R-squared:  0.152,Adjusted R-squared:  0.141
## F-statistic: 13.7 on 1 and 76 DF,  p-value: 0.000411

plot(axis_coefs.diag_dsd[,1] ~ metapcna.scores, xlab = "meta-PCNA score", ylab = "A1 signal", pch = 16,
```




```
plot(axis_coefs.diag_dsd[,2] ~ metapcna.scores, xlab = "meta-PCNA score", ylab = "A2 signal", pch = 16,
```



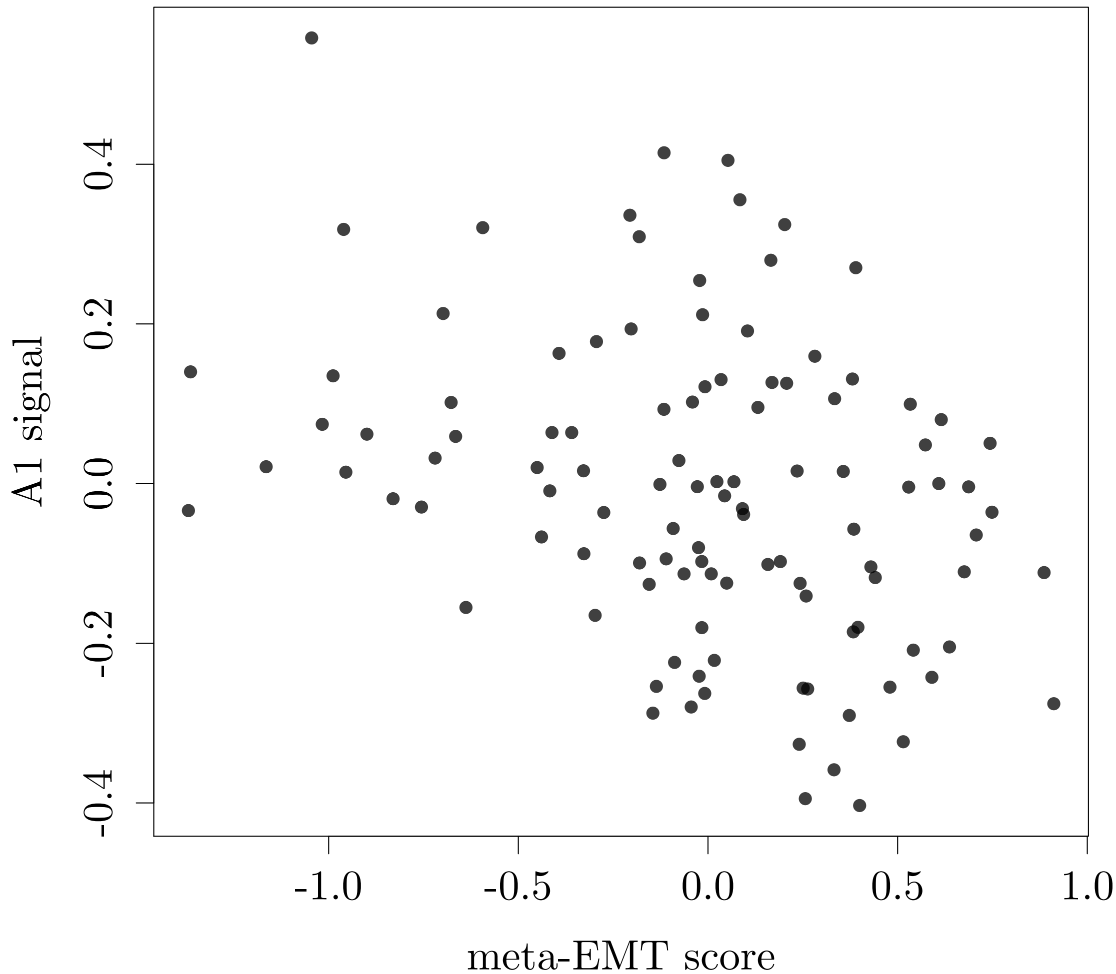

```
cor.test(axis_coefs.diag_dsd[,1], metapcna.scores, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 1] and metapcna.scores
## z = 10.27, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##    tau
## 0.6634
```

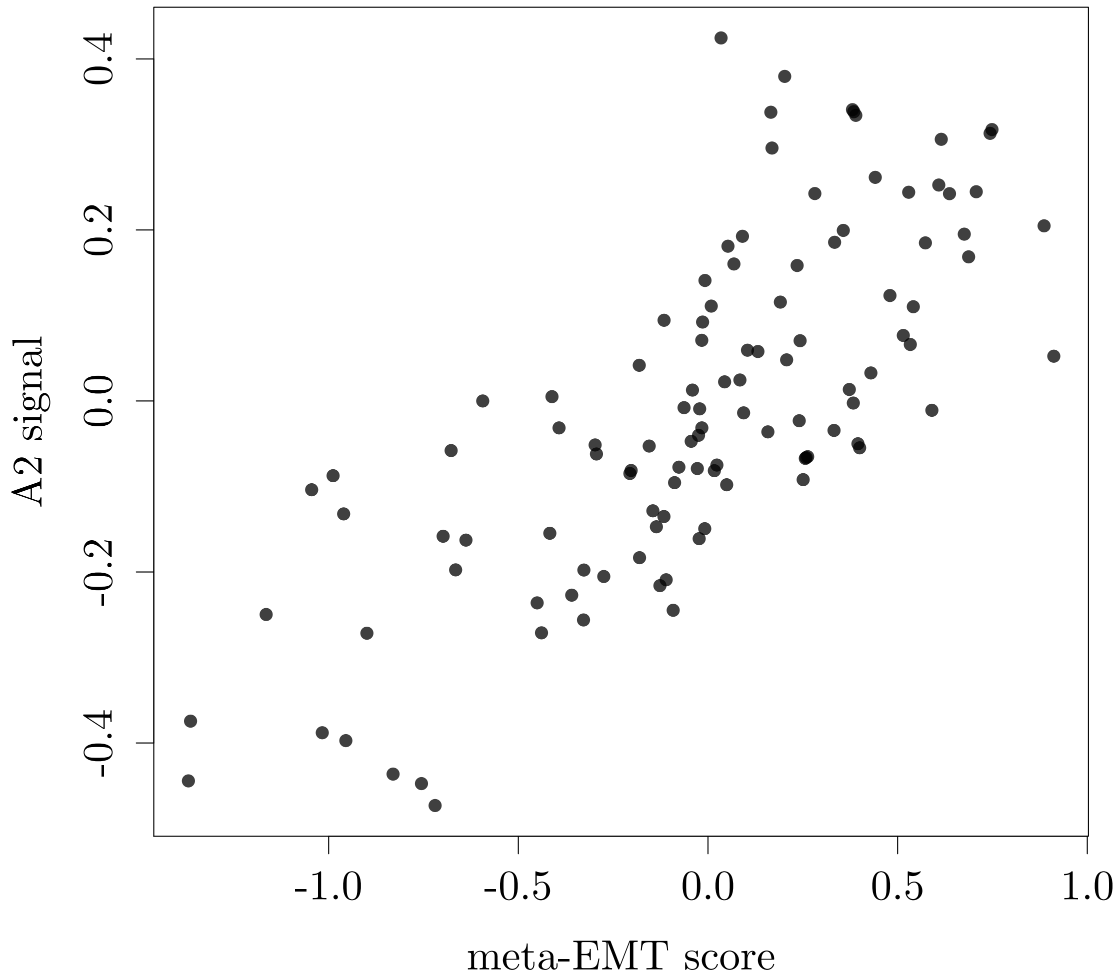
```
cor.test(axis_coefs.diag_dsd[,2], metapcna.scores, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 2] and metapcna.scores
```

```
## z = 1.899, p-value = 0.05762
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##    tau
## 0.1226

summary(lm(axis_coefs.diag_dsd[,1] ~ metapcna.scores))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 1] ~ metapcna.scores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4295 -0.0477  0.0151  0.0622  0.1785
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -4.0135     0.2274   -17.6   <2e-16
## metapcna.scores   0.5504     0.0312    17.6   <2e-16
##
## Residual standard error: 0.0971 on 108 degrees of freedom
## Multiple R-squared:  0.742,Adjusted R-squared:  0.74
## F-statistic:  311 on 1 and 108 DF,  p-value: <2e-16

summary(lm(axis_coefs.diag_dsd[,2] ~ metapcna.scores))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 2] ~ metapcna.scores)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.478 -0.117  0.000  0.132  0.402
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.8487     0.4577   -1.85    0.066
## metapcna.scores   0.1156     0.0629    1.84    0.069
##
## Residual standard error: 0.195 on 108 degrees of freedom
## Multiple R-squared:  0.0303,Adjusted R-squared:  0.0214
## F-statistic: 3.38 on 1 and 108 DF,  p-value: 0.0688

plot(axis_coefs.diag_dsd[,1] ~ emt.scores, xlab = "meta-EMT score", ylab = "A1 signal", pch = 16, col =
```




```
plot(axis_coefs.diag_dsd[,2] ~ emt.scores, xlab = "meta-EMT score", ylab = "A2 signal", pch = 16, col =
```




```
cor.test(axis_coefs.diag_dsd[,1], emt.scores, method = "kendall")

##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 1] and emt.scores
## z = -3.371, p-value = 0.0007492
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##     tau
## -0.2177

cor.test(axis_coefs.diag_dsd[,2], emt.scores, method = "kendall")

##
##  Kendall's rank correlation tau
##
## data:  axis_coefs.diag_dsd[, 2] and emt.scores
```

```
## z = 8.795, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##   tau
## 0.568

summary(lm(axis_coefs.diag_dsd[,1] ~ emt.scores))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 1] ~ emt.scores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3487 -0.1156 -0.0137  0.1312  0.4365
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0129     0.0172   -0.75  0.45511
## emt.scores   -0.1289     0.0351   -3.68  0.00037
##
## Residual standard error: 0.18 on 108 degrees of freedom
## Multiple R-squared:  0.111,Adjusted R-squared:  0.103
## F-statistic: 13.5 on 1 and 108 DF,  p-value: 0.000372

summary(lm(axis_coefs.diag_dsd[,2] ~ emt.scores))

##
## Call:
## lm(formula = axis_coefs.diag_dsd[, 2] ~ emt.scores)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.2545 -0.0977 -0.0088  0.0914  0.4166
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.00235    0.01255   -0.19     0.85
## emt.scores   0.30063    0.02560   11.74   <2e-16
##
## Residual standard error: 0.132 on 108 degrees of freedom
## Multiple R-squared:  0.561,Adjusted R-squared:  0.557
## F-statistic:  138 on 1 and 108 DF,  p-value: <2e-16

anova(lm(axis_coefs.diag_dsd[,1] ~ samps$purity_qpure + emt.scores))

## Analysis of Variance Table
##
## Response: axis_coefs.diag_dsd[, 1]
##                    Df Sum Sq Mean Sq F value  Pr(>F)
## samps$purity_qpure  1  0.453   0.453   15.20 0.00021
## emt.scores          1  0.039   0.039    1.32 0.25381
## Residuals          75  2.236   0.030

anova(lm(axis_coefs.diag_dsd[,2] ~ samps$purity_qpure + emt.scores))
```

```
## Analysis of Variance Table
##
## Response: axis_coefs.diag_dsd[, 2]
##                    Df Sum Sq Mean Sq F value  Pr(>F)
## samps$purity_qpure  1  0.518   0.518    28.4 1.0e-06
## emt.scores          1  1.513   1.513    82.9 9.4e-14
## Residuals          75  1.369   0.018
```

```
temp.sig_id = colnames(axis_coefs.msigdb.corr)
temp.sig_class = gsub("\\..*", "", temp.sig_id)
temp.nsigs = length(temp.sig_id)
temp.nmeta = nrow(axis_coefs.msigdb.corr)
tables = lapply(1:temp.nmeta, function(metagene_i) {
        tapply(1:temp.nsigs, temp.sig_class, function(sig_class_is) {
                all_cors = axis_coefs.msigdb.corr[, sig_class_is]
                this_cors = all_cors[metagene_i, ]
                this_ids = temp.sig_id[sig_class_is]

                all_sig_cors = abs(all_cors) >= sig.corr.threshold
                this_sig_cors = all_sig_cors[metagene_i, ]

                sigs_to_report = which(this_sig_cors)

                if (length(sigs_to_report) == 0)
                {
                        table = data.frame(GeneSet = c(), Correlation = c(), Metagenes = c())
                }
                else
                {
                        table = data.frame(
                                GeneSet = this_ids[sigs_to_report],
                                Correlation = this_cors[sigs_to_report],
                                Metagenes = apply(all_cors[,sigs_to_report,drop=FALSE], 2, function(cors
                                        sel = abs(cors) >= sig.corr.threshold
                                        # A positive number implies that positive GSVA signal is associa
                                        paste(which(sel) * sign(cors[which(sel)]), collapse = ",")
                                }))
                        table = table[order(-(table$Correlation)),]
                        rownames(table) <- NULL
                }
                table
        }, simplify = FALSE)
})
tables

## [[1]]
## [[1]]$c1
## data frame with 0 columns and 0 rows
##
## [[1]]$c2
##
## 1
## 2
```

```
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10                                                         c2.AMUNDSON_GAMMA_RADIATION_RESPONSE/c4.GNF2_CDC2
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19                                                          c2.EGUCHI_CELL_CYCLE_RB1_TARGETS/c2.ROSTY_CERVIC
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44
## 45
## 46
## 47
## 48
## 49
## 50
## 51
## 52
## 53
## 54
## 55
## 56
```

```
## 57                                                                                   c2.REACTOME_ACTIVATION_OF_THE_PRE_R
## 58
## 59
## 60
## 61
## 62
## 63
## 64
## 65
## 66
## 67
## 68
## 69  c2.REACTOME_CELL_CYCLE_CHECKPOINTS/c2.REACTOME_G1_S_TRANSITION/c2.REACTOME_SYNTHESIS_OF_DNA/c2.RE
## 70
## 71
## 72
## 73
## 74
## 75
## 76
## 77
## 78
## 79
## 80
## 81
## 82
## 83
## 84
## 85
## 86
## 87
## 88
## 89
## 90
## 91
## 92
## 93
## 94
## 95
## 96
## 97
## 98
## 99
## 100
## 101
## 102
## 103
## 104
## 105
## 106
## 107
## 108
## 109
```

```
## 110
## 111
## 112
## 113
## 114
## 115
## 116
## 117
## 118
## 119
## 120
## 121
## 122
## 123
## 124
## 125
## 126
## 127
## 128
## 129
## 130
## 131
## 132
## 133
## 134
## 135
## 136
## 137
## 138
## 139
## 140
## 141
## 142
## 143
## 144
## 145
## 146
## 147
## 148
## 149
## 150
## 151
## 152
## 153
## 154
## 155
## 156
## 157
## 158
## 159
## 160
## 161
## 162
## 163
```

```
## 164
## 165
## 166
## 167
## 168
## 169
## 170
## 171
## 172
## 173
## 174
## 175
##     Correlation Metagenes
## 1        0.7141         1
## 2        0.6991         1
## 3        0.6981         1
## 4        0.6981         1
## 5        0.6964         1
## 6        0.6911         1
## 7        0.6894         1
## 8        0.6767         1
## 9        0.6767         1
## 10       0.6737         1
## 11       0.6711         1
## 12       0.6694         1
## 13       0.6617         1
## 14       0.6601         1
## 15       0.6584         1
## 16       0.6474         1
## 17       0.6470         1
## 18       0.6464         1
## 19       0.6460         1
## 20       0.6444         1
## 21       0.6440         1
## 22       0.6434         1
## 23       0.6404         1
## 24       0.6400         1
## 25       0.6344         1
## 26       0.6317         1
## 27       0.6294         1
## 28       0.6290         1
## 29       0.6287         1
## 30       0.6264         1
## 31       0.6257         1
## 32       0.6254         1
## 33       0.6230         1
## 34       0.6224         1
## 35       0.6180         1
## 36       0.6153         1
## 37       0.6143         1
## 38       0.6140         1
## 39       0.6123         1
## 40       0.6117         1
## 41       0.6073         1
```

```
## 42        0.6060         1
## 43        0.6053         1
## 44        0.6050         1
## 45        0.6047         1
## 46        0.6037         1
## 47        0.5993         1
## 48        0.5990         1
## 49        0.5963         1
## 50        0.5960         1
## 51        0.5950         1
## 52        0.5883         1
## 53        0.5870         1
## 54        0.5857         1
## 55        0.5853         1
## 56        0.5837         1
## 57        0.5833         1
## 58        0.5807         1
## 59        0.5807         1
## 60        0.5770         1
## 61        0.5756         1
## 62        0.5753         1
## 63        0.5750         1
## 64        0.5733         1
## 65        0.5730         1
## 66        0.5713         1
## 67        0.5683         1
## 68        0.5680         1
## 69        0.5660         1
## 70        0.5653         1
## 71        0.5650         1
## 72        0.5640         1
## 73        0.5620         1
## 74        0.5586         1
## 75        0.5586         1
## 76        0.5560         1
## 77        0.5536         1
## 78        0.5513         1
## 79        0.5500         1
## 80        0.5473         1
## 81        0.5466         1
## 82        0.5466         1
## 83        0.5463         1
## 84        0.5463         1
## 85        0.5456         1
## 86        0.5453         1
## 87        0.5336         1
## 88        0.5319         1
## 89        0.5313         1
## 90        0.5309         1
## 91        0.5306         1
## 92        0.5296         1
## 93        0.5279         1
## 94        0.5276         1
## 95        0.5273         1
```

```
## 96       0.5273          1
## 97       0.5246          1
## 98       0.5243          1
## 99       0.5239          1
## 100      0.5233          1
## 101      0.5226          1
## 102      0.5223          1
## 103      0.5199          1
## 104      0.5199          1
## 105      0.5196          1
## 106      0.5179          1
## 107      0.5179          1
## 108      0.5173          1
## 109      0.5173          1
## 110      0.5159          1
## 111      0.5146          1
## 112      0.5129          1
## 113      0.5103          1
## 114      0.5103          1
## 115      0.5099          1
## 116      0.5086          1
## 117      0.5073          1
## 118      0.5063          1
## 119      0.5059          1
## 120      0.5056          1
## 121      0.5056          1
## 122      0.5043          1
## 123      0.5029          1
## 124      0.5019          1
## 125      0.5019          1
## 126      0.5016          1
## 127      0.5013          1
## 128      0.5009          1
## 129      0.5003          1
## 130      0.5003          1
## 131     -0.5009         -1
## 132     -0.5033         -1
## 133     -0.5043         -1
## 134     -0.5056         -1
## 135     -0.5083         -1
## 136     -0.5089         -1
## 137     -0.5096         -1
## 138     -0.5243         -1
## 139     -0.5289         -1
## 140     -0.5316         -1
## 141     -0.5319         -1
## 142     -0.5393         -1
## 143     -0.5399         -1
## 144     -0.5416         -1
## 145     -0.5433         -1
## 146     -0.5516         -1
## 147     -0.5520         -1
## 148     -0.5570         -1
## 149     -0.5580         -1
```

```
## 150     -0.5583        -1
## 151     -0.5640        -1
## 152     -0.5646        -1
## 153     -0.5730        -1
## 154     -0.5733        -1
## 155     -0.5750        -1
## 156     -0.5893        -1
## 157     -0.5900        -1
## 158     -0.5913        -1
## 159     -0.5940        -1
## 160     -0.6047        -1
## 161     -0.6063        -1
## 162     -0.6147        -1
## 163     -0.6153        -1
## 164     -0.6217        -1
## 165     -0.6247        -1
## 166     -0.6260        -1
## 167     -0.6310        -1
## 168     -0.6347        -1
## 169     -0.6357        -1
## 170     -0.6370        -1
## 171     -0.6387        -1
## 172     -0.6454        -1
## 173     -0.6791        -1
## 174     -0.6894        -1
## 175     -0.6951        -1
##
## [[1]]$c3
##                  GeneSet Correlation Metagenes
## 1          c3.V$ELK1_02      0.5740         1
## 2 c3.SCGGAAGY_V$ELK1_02      0.5580         1
## 3    c3.CTGCAGY_UNKNOWN     -0.5046        -1
## 4          c3.V$OCT1_01     -0.5089        -1
## 5           c3.V$GATA_Q6     -0.5153        -1
## 6          c3.V$OCT1_04     -0.5313        -1
## 7            c3.V$OCT_C     -0.5436        -1
##
## [[1]]$c4
##                                                                          GeneSet
## 1  c4.GNF2_RFC3/c4.GNF2_RFC4/c4.GNF2_SMC2L1/c4.GNF2_CKS1B/c4.GNF2_CKS2/c4.GNF2_TTK
## 2                                                                     c4.MODULE_17
## 3                                                                    c4.MODULE_315
## 4                                                                    c4.MORF_BUB1B
## 5                                                                    c4.MODULE_244
## 6                                                                    c4.MODULE_337
## 7                                                                     c4.MORF_FEN1
## 8                                                                    c4.MODULE_126
## 9                                                                    c4.MODULE_124
## 10                                                                   c4.MORF_ESPL1
## 11                                                                    c4.MORF_BUB1
## 12                                                                   c4.MODULE_403
## 13                                                    c4.MORF_BUB3/c4.MORF_RAD23A
## 14                                                      c4.MORF_RFC4/c4.MORF_RRM1
## 15                                         c4.MODULE_98/c4.MODULE_198/c4.MODULE_252
```

```
## 16                                                        c4.MODULE_125/c4.MODULE_158
## 17                                                                        c4.MORF_UNG
## 18                                                                      c4.MODULE_278
## 19                                                                      c4.MORF_GSPT1
## 20                                                                      c4.MODULE_320
## 21                                                                        c4.MODULE_8
## 22                                                                       c4.MORF_CCNF
## 23                                                                       c4.MORF_EI24
## 24                                                          c4.GNF2_PA2G4/c4.GNF2_RAN
## 25                                                                      c4.MORF_PRKDC
## 26                                                                       c4.MORF_GMPS
## 27                                                                      c4.MODULE_219
## 28                                                                       c4.GNF2_MCM5
## 29                                                                      c4.MORF_DNMT1
## 30                                                                       c4.GNF2_MSH2
## 31                                                                     c4.MORF_CSNK2B
## 32                                                                     c4.MORF_PTPN11
## 33                                                                     c4.MORF_PPP1CC
## 34                                                         c4.MORF_XRCC5/c4.MORF_GNB1
## 35                                                                      c4.MODULE_451
## 36                                                                       c4.MORF_SOD1
## 37                                                                      c4.MORF_HDAC1
## 38                                                                       c4.MODULE_51
## 39                                                                       c4.GNF2_MAPT
## 40                                                                       c4.MODULE_19
## 41                           c4.MODULE_11/c4.MODULE_66/c4.MODULE_100/c4.MODULE_137
##    Correlation Metagenes
## 1       0.6637         1
## 2       0.6510         1
## 3       0.6324         1
## 4       0.6307         1
## 5       0.6294         1
## 6       0.6244         1
## 7       0.5860         1
## 8       0.5817         1
## 9       0.5813         1
## 10      0.5656         1
## 11      0.5650         1
## 12      0.5640         1
## 13      0.5633         1
## 14      0.5606         1
## 15      0.5586         1
## 16      0.5586         1
## 17      0.5536         1
## 18      0.5536         1
## 19      0.5503         1
## 20      0.5490         1
## 21      0.5480         1
## 22      0.5436         1
## 23      0.5379         1
## 24      0.5313         1
## 25      0.5279         1
## 26      0.5279         1
```

```
## 27      0.5266         1
## 28      0.5249         1
## 29      0.5243         1
## 30      0.5206         1
## 31      0.5203         1
## 32      0.5163         1
## 33      0.5089         1
## 34      0.5039         1
## 35      0.5026         1
## 36      0.5019         1
## 37      0.5009         1
## 38     -0.5066        -1
## 39     -0.5259        -1
## 40     -0.5656        -1
## 41     -0.5967        -1
##
## [[1]]$c5
##                                                                                         GeneSet
## 1                                 c5.M_PHASE/c5.MITOSIS/c5.M_PHASE_OF_MITOTIC_CELL_CYCLE
## 2                                                                 c5.REGULATION_OF_MITOSIS
## 3                           c5.CELL_CYCLE_PROCESS/c5.MITOTIC_CELL_CYCLE/c5.CELL_CYCLE_PHASE
## 4                                                                               c5.SPINDLE
## 5                                                                          c5.SPINDLE_POLE
## 6                                     c5.ORGANELLE_PART/c5.INTRACELLULAR_ORGANELLE_PART
## 7                                                                c5.CHROMOSOME_SEGREGATION
## 8                                                                 c5.CELL_CYCLE_GO_0007049
## 9                                                                   c5.SPINDLE_MICROTUBULE
## 10                                                          c5.MITOTIC_CELL_CYCLE_CHECKPOINT
## 11                                                                 c5.CONDENSED_CHROMOSOME
## 12                 c5.MITOTIC_SISTER_CHROMATID_SEGREGATION/c5.SISTER_CHROMATID_SEGREGATION
## 13                                                           c5.CELL_CYCLE_CHECKPOINT_GO_0000075
## 14 c5.MITOTIC_SPINDLE_ORGANIZATION_AND_BIOGENESIS/c5.SPINDLE_ORGANIZATION_AND_BIOGENESIS
## 15                                                             c5.DOUBLE_STRAND_BREAK_REPAIR
## 16                                                                 c5.DNA_METABOLIC_PROCESS
## 17                                                           c5.REGULATION_OF_MITOTIC_CELL_CYCLE
## 18                 c5.RESPONSE_TO_ENDOGENOUS_STIMULUS/c5.RESPONSE_TO_DNA_DAMAGE_STIMULUS
## 19                                       c5.CHROMOSOMEPERICENTRIC_REGION/c5.KINETOCHORE
## 20                                                           c5.PORE_COMPLEX/c5.NUCLEAR_PORE
## 21                                                                            c5.DNA_REPAIR
## 22                                         c5.MACROMOLECULAR_COMPLEX/c5.PROTEIN_COMPLEX
## 23                                    c5.INTERPHASE/c5.INTERPHASE_OF_MITOTIC_CELL_CYCLE
## 24        c5.NON_MEMBRANE_BOUND_ORGANELLE/c5.INTRACELLULAR_NON_MEMBRANE_BOUND_ORGANELLE
## 25                                         c5.NUCLEAR_MEMBRANE/c5.NUCLEAR_MEMBRANE_PART
## 26                                                      c5.CHROMOSOMAL_PART/c5.CHROMOSOME
## 27                                            c5.PHOSPHORIC_DIESTER_HYDROLASE_ACTIVITY
## 28                        c5.CELL_SURFACE_RECEPTOR_LINKED_SIGNAL_TRANSDUCTION_GO_0007166
##    Correlation Metagenes
## 1       0.6894         1
## 2       0.6821         1
## 3       0.6527         1
## 4       0.6437         1
## 5       0.6280         1
## 6       0.6244         1
## 7       0.5883         1
```

```
## 8       0.5760         1
## 9       0.5726         1
## 10      0.5690         1
## 11      0.5620         1
## 12      0.5546         1
## 13      0.5426         1
## 14      0.5420         1
## 15      0.5369         1
## 16      0.5166         1
## 17      0.5156         1
## 18      0.5146         1
## 19      0.5136         1
## 20      0.5083         1
## 21      0.5063         1
## 22      0.5059         1
## 23      0.5033         1
## 24      0.5029         1
## 25      0.5013         1
## 26      0.5003         1
## 27     -0.5023        -1
## 28     -0.5176        -1
##
## [[1]]$c6
##                          GeneSet Correlation Metagenes
## 1       c6.CSR_LATE_UP.V1_SIGNED      0.6297         1
## 2           c6.MTOR_UP.V1_SIGNED      0.5123         1
## 3 c6.GCNP_SHH_UP_EARLY.V1_SIGNED      0.5026         1
##
## [[1]]$c7
##
## 1                                                                                  c7.GSE15750_DAY6_VS_I
## 2
## 3
## 4
## 5
## 6
## 7
## 8                                                                             c7.GSE24634_TEFF_VS_TCONV_
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
```

```
## 25
## 26
## 27
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44
## 45
## 46
## 47
## 48
## 49
## 50
## 51
## 52
## 53
## 54 c7.GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_CD8_TCELL_SIGNED/c7.GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_IL1
## 55
## 56
## 57
## 58
## 59
## 60
## 61
## 62
## 63
## 64
## 65
## 66
## 67                                                           c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_C
## 68
## 69
## 70
## 71                                                       c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_T
##    Correlation Metagenes
## 1       0.6187         1
## 2       0.6160         1
## 3       0.6143         1
## 4       0.5880         1
## 5       0.5857         1
## 6       0.5756         1
```

```
## 7        0.5696         1
## 8        0.5653         1
## 9        0.5630         1
## 10       0.5623         1
## 11       0.5580         1
## 12       0.5553         1
## 13       0.5546         1
## 14       0.5476         1
## 15       0.5466         1
## 16       0.5423         1
## 17       0.5349         1
## 18       0.5336         1
## 19       0.5276         1
## 20       0.5186         1
## 21       0.5186         1
## 22       0.5036         1
## 23      -0.5039        -1
## 24      -0.5086        -1
## 25      -0.5109        -1
## 26      -0.5119        -1
## 27      -0.5119        -1
## 28      -0.5149        -1
## 29      -0.5179        -1
## 30      -0.5183        -1
## 31      -0.5223        -1
## 32      -0.5239        -1
## 33      -0.5269        -1
## 34      -0.5303        -1
## 35      -0.5316        -1
## 36      -0.5336        -1
## 37      -0.5343        -1
## 38      -0.5343        -1
## 39      -0.5426        -1
## 40      -0.5516        -1
## 41      -0.5520        -1
## 42      -0.5543        -1
## 43      -0.5560        -1
## 44      -0.5603        -1
## 45      -0.5603        -1
## 46      -0.5613        -1
## 47      -0.5630        -1
## 48      -0.5636        -1
## 49      -0.5650        -1
## 50      -0.5716        -1
## 51      -0.5743        -1
## 52      -0.5786        -1
## 53      -0.5830        -1
## 54      -0.5853        -1
## 55      -0.5860        -1
## 56      -0.5867        -1
## 57      -0.5920        -1
## 58      -0.5950        -1
## 59      -0.5953        -1
## 60      -0.6007        -1
```

```
## 61     -0.6010        -1
## 62     -0.6090        -1
## 63     -0.6190        -1
## 64     -0.6193        -1
## 65     -0.6254        -1
## 66     -0.6417        -1
## 67     -0.6500        -1
## 68     -0.6530        -1
## 69     -0.6637        -1
## 70     -0.6654        -1
## 71     -0.6667        -1
##
##
## [[2]]
## [[2]]$c1
## data frame with 0 columns and 0 rows
##
## [[2]]$c2
##                                                                                                    Ger
## 1                                                                                c2.PID_INTEGRIN1_PAT
## 2                                                       c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SI
## 3                                                   c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SI
## 4                                                                                c2.PID_INTEGRIN3_PAT
## 5                                                             c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABIN
## 6                                                                    c2.HUANG_DASATINIB_RESISTANCE_SI
## 7                                                               c2.WIEDERSCHAIN_TARGETS_OF_BMI1_AND_P
## 8                                                                                c2.BURTON_ADIPOGENES
## 9                                                                          c2.POTTI_TOPOTECAN_SENSITI
## 10                                                                             c2.KARAKAS_TGFB1_SIGNA
## 11                                                                                c2.PID_UPA_UPAR_PAT
## 12                                                                         c2.CROMER_TUMORIGENESIS_SI
## 13                                                                        c2.ROZANOV_MMP14_TARGETS_SU
## 14                                                                  c2.WOO_LIVER_CANCER_RECURRENCE_SI
## 15                                                                       c2.KEGG_ECM_RECEPTOR_INTERAC
## 16 c2.FARMER_BREAST_CANCER_CLUSTER_5/c2.ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE/c4.GNF2_C
## 17                                                                               c2.PID_INTEGRIN5_PAT
## 18                           c2.REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION/c2.REACTOME_COLLAGEN_FORMA
## 19                                                                       c2.ROY_WOUND_BLOOD_VESSEL_SI
## 20                                                                                 c2.KEGG_FOCAL_ADHE
## 21                                                                        c2.CHICAS_RB1_TARGETS_CONFI
## 22                                                                         c2.YIH_RESPONSE_TO_ARSENIT
## 23                                                                  c2.PHONG_TNF_RESPONSE_VIA_P38_PAR
## 24                                                                 c2.SERVITJA_ISLET_HNF1A_TARGETS_SI
## 25                                                                       c2.LI_PROSTATE_CANCER_EPIGEN
## 26                                                                              c2.PID_SYNDECAN_1_PAT
## 27                                                                         c2.AGARWAL_AKT_PATHWAY_TAR
## 28                                                     c2.REACTOME_CELL_EXTRACELLULAR_MATRIX_INTERACT
## 29                                                                             c2.PID_INTEGRIN_CS_PAT
## 30                                                     c2.HELLEBREKERS_SILENCED_DURING_TUMOR_ANGIOGEN
## 31                                                                                c2.MATTHEWS_AP1_TAR
## 32                                                                c2.RODWELL_AGING_KIDNEY_NO_BLOOD_SI
## 33                                                    c2.YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTE
## 34                                                                                c2.WESTON_VEGFA_TAR
## 35                                                                                   c2.WU_CELL_MIGRA
```

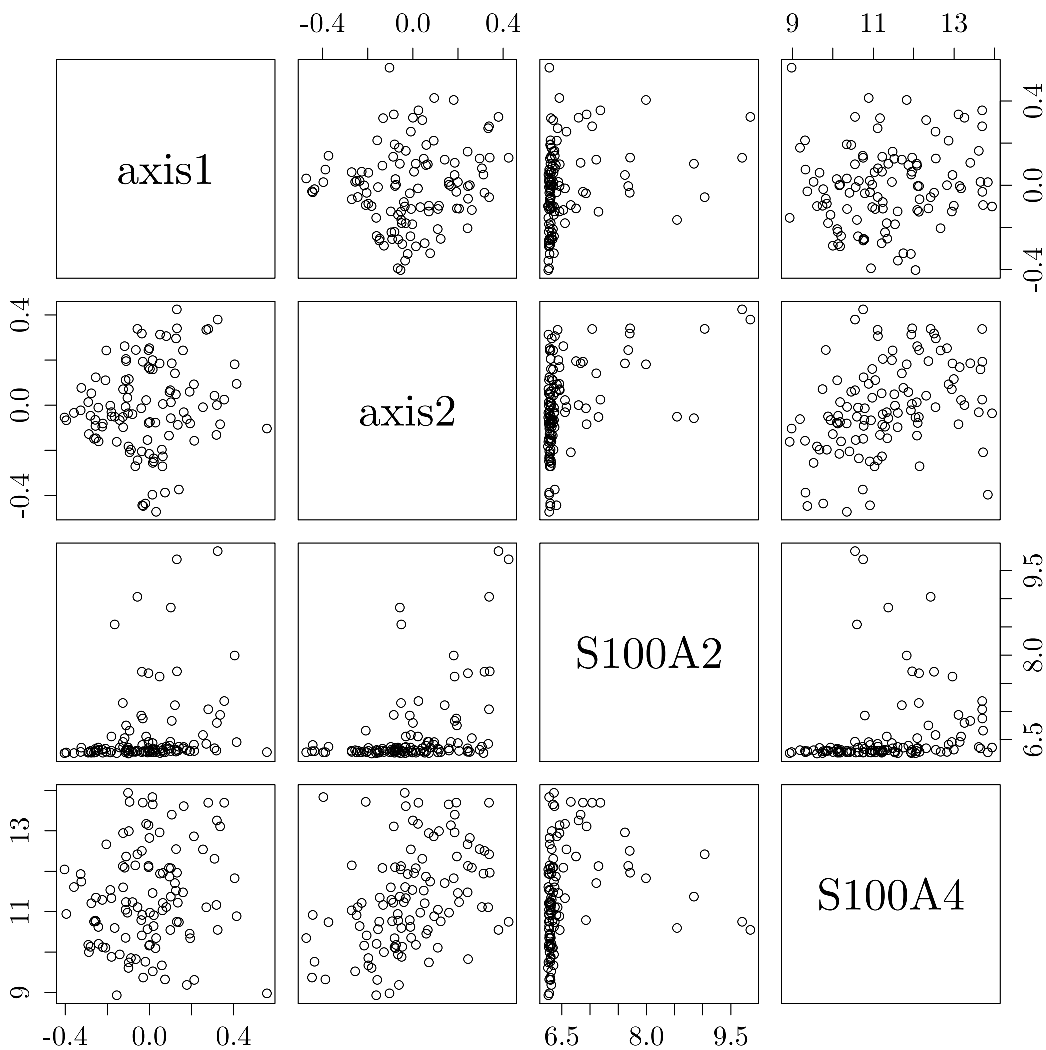
```
## 36                                                   c2.SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_SI
## 37                                                               c2.KAN_RESPONSE_TO_ARSENIC_TRIO
## 38                                                                    c2.VERHAAK_GLIOBLASTOMA_NE
## 39                                                    c2.REACTOME_INTEGRIN_CELL_SURFACE_INTERACT
## 40                                                                             c2.GILDEA_METAST
## 41                                     c2.TURASHVILI_BREAST_LOBULAR_CARCINOMA_VS_DUCTAL_NORMAL_SI
## 42                                                      c2.HARRIS_HYPOXIA/c2.WINTER_HYPOXIA_META
## 43                                                                     c2.BIOCARTA_PLATELETAPP_PAT
## 44                                                                 c2.WANG_METHYLATED_IN_BREAST_CA
## 45                                               c2.RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY_SI
## 46                                                    c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_SI
## 47                                                                   c2.WALLACE_PROSTATE_CANCER_SI
## 48                                              c2.WANG_BARRETTS_ESOPHAGUS_AND_ESOPHAGUS_CANCER_SI
## 49                                                        c2.SMID_BREAST_CANCER_RELAPSE_IN_BONE_SI
## 50                                                     c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_SI
## 51                                                                       c2.LIU_PROSTATE_CANCER_SI
## 52                                                                      c2.PASINI_SUZ12_TARGETS_SI
## 53                                                                c2.NAKAMURA_ADIPOGENESIS_LATE_SI
## 54                                                                 c2.DOANE_BREAST_CANCER_CLASSES_SI
## 55                                              c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_SI
##    Correlation Metagenes
## 1       0.6544         2
## 2       0.6514         2
## 3       0.6454         2
## 4       0.6374         2
## 5       0.6370         2
## 6       0.6350         2
## 7       0.6250         2
## 8       0.6103         2
## 9       0.6017         2
## 10      0.5997         2
## 11      0.5970         2
## 12      0.5910         2
## 13      0.5907         2
## 14      0.5883         2
## 15      0.5817         2
## 16      0.5783         2
## 17      0.5766         2
## 18      0.5723         2
## 19      0.5676         2
## 20      0.5666         2
## 21      0.5643         2
## 22      0.5623         2
## 23      0.5590         2
## 24      0.5586         2
## 25      0.5543         2
## 26      0.5516         2
## 27      0.5486         2
## 28      0.5379         2
## 29      0.5363         2
## 30      0.5289         2
## 31      0.5279         2
## 32      0.5256         2
## 33      0.5239         2
```

```
## 34      0.5229         2
## 35      0.5209         2
## 36      0.5206         2
## 37      0.5206         2
## 38      0.5203         2
## 39      0.5179         2
## 40      0.5179         2
## 41      0.5146         2
## 42      0.5119         2
## 43      0.5056         2
## 44      0.5009         2
## 45     -0.5043        -2
## 46     -0.5209        -2
## 47     -0.5209        -2
## 48     -0.5443        -2
## 49     -0.5536        -2
## 50     -0.5563        -2
## 51     -0.5643        -2
## 52     -0.5663        -2
## 53     -0.5680        -2
## 54     -0.6010        -2
## 55     -0.6097        -2
##
## [[2]]$c3
## data frame with 0 columns and 0 rows
##
## [[2]]$c4
##         GeneSet Correlation Metagenes
## 1  c4.GNF2_PTX3      0.5933         2
## 2  c4.GNF2_MMP1      0.5750         2
## 3 c4.MODULE_412      0.5670         2
## 4 c4.MODULE_122      0.5600         2
## 5  c4.MODULE_47      0.5463         2
## 6 c4.MODULE_153      0.5426         2
## 7 c4.MODULE_321      0.5309         2
## 8 c4.MODULE_275      0.5253         2
## 9 c4.MODULE_562      0.5066         2
##
## [[2]]$c5
##                                                           GeneSet
## 1                                                c5.AXON_GUIDANCE
## 2                                           c5.TISSUE_DEVELOPMENT
## 3                                                     c5.COLLAGEN
## 4 c5.AXONOGENESIS/c5.CELLULAR_MORPHOGENESIS_DURING_DIFFERENTIATION
##   Correlation Metagenes
## 1      0.5710         2
## 2      0.5363         2
## 3      0.5313         2
## 4      0.5146         2
##
## [[2]]$c6
##                                 GeneSet Correlation Metagenes
## 1 c6.CORDENONSI_YAP_CONSERVED_SIGNATURE      0.5256         2
## 2                  c6.LEF1_UP.V1_SIGNED      0.5193         2
```

```
## 3                   c6.STK33_NOMO_SIGNED      0.5073          2
##
## [[2]]$c7
##                                                                    GeneSet
## 1                              c7.GSE17721_CTRL_VS_CPG_12H_BMDM_SIGNED
## 2 c7.GSE1460_INTRATHYMIC_T_PROGENITOR_VS_THYMIC_STROMAL_CELL_SIGNED
##   Correlation Metagenes
## 1     -0.5076        -2
## 2     -0.5079        -2
```

```
for (subtable_index in 1:length(tables))
{
        write.csv(do.call(rbind, tables[[subtable_index]]), file = sprintf("A%d_corrs.csv", subtable_inc
}
```

```
pairs(cbind(axis_coefs.diag_dsd, t(x.diag_dsd[c("S100A2", "S100A4"),])))
```



Underwhelming, but the poor detection rate of A2 and A4 probes is a likely culprit. When I get APGI scores from DC, I can directly compare the staining patterns to A1 and A2 signals – this would be a better comparison to the work of chapter 2 anyway.

5 Session information

```
session_info

## R version 3.1.1 (2014-07-10)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_AU.UTF-8          LC_NUMERIC=C
##  [3] LC_TIME=en_AU.UTF-8           LC_COLLATE=en_AU.UTF-8
##  [5] LC_MONETARY=en_AU.UTF-8       LC_MESSAGES=en_AU.UTF-8
##  [7] LC_PAPER=en_AU.UTF-8          LC_NAME=en_AU.UTF-8
##  [9] LC_ADDRESS=en_AU.UTF-8        LC_TELEPHONE=en_AU.UTF-8
## [11] LC_MEASUREMENT=en_AU.UTF-8    LC_IDENTIFICATION=en_AU.UTF-8
##
## attached base packages:
## [1] splines   parallel  methods   stats     graphics  grDevices utils
## [8] datasets  base
##
## other attached packages:
##  [1] doParallel_1.0.8    iterators_1.0.7     foreach_1.4.2
##  [4] ahaz_1.14           survival_2.37-7     NMF_0.20.5
##  [7] Biobase_2.26.0      BiocGenerics_0.12.1 cluster_1.15.3
## [10] rngtools_1.2.4      pkgmaker_0.22       registry_0.2
## [13] energy_1.6.2        glmnet_1.9-8        Matrix_1.1-4
## [16] glmulti_1.0.7       rJava_0.9-6
##
## loaded via a namespace (and not attached):
##  [1] boot_1.3-13        codetools_0.2-9    colorspace_1.2-4
##  [4] compiler_3.1.1     digest_0.6.4       ggplot2_1.0.0
##  [7] grid_3.1.1         gridBase_0.4-7     gtable_0.1.2
## [10] lattice_0.20-29    MASS_7.3-35        munsell_0.4.2
## [13] plyr_1.8.1         proto_0.3-10       RColorBrewer_1.0-5
## [16] Rcpp_0.11.3        reshape2_1.4       scales_0.2.4
## [19] stringr_0.6.2      tools_3.1.1        xtable_1.7-4

sessionInfo()

## R version 3.1.1 (2014-07-10)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8          LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8           LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8       LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8          LC_NAME=en_US.UTF-8
##  [9] LC_ADDRESS=en_US.UTF-8        LC_TELEPHONE=en_US.UTF-8
## [11] LC_MEASUREMENT=en_US.UTF-8    LC_IDENTIFICATION=en_US.UTF-8
##
```

```
## attached base packages:
## [1] parallel  methods   splines   stats     graphics  grDevices utils
## [8] datasets  base
##
## other attached packages:
##  [1] MASS_7.3-39         stargazer_5.1       xtable_1.7-4
##  [4] gplots_2.16.0       RColorBrewer_1.1-2  glmnet_1.9-8
##  [7] Matrix_1.1-5        glmulti_1.0.7       rJava_0.9-6
## [10] bnlearn_3.7.1       nnls_1.4            NMF_0.20.5
## [13] synchronicity_1.1.4 bigmemory_4.4.6     BH_1.55.0-3
## [16] bigmemory.sri_0.1.3 Biobase_2.26.0      BiocGenerics_0.12.1
## [19] cluster_2.0.1       rngtools_1.2.4      pkgmaker_0.22
## [22] registry_0.2        energy_1.6.2        survival_2.37-7
## [25] tikzDevice_0.8.1    knitr_1.9
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-6       boot_1.3-15        caTools_1.17.1
##  [4] codetools_0.2-10   colorspace_1.2-4   digest_0.6.8
##  [7] doParallel_1.0.8   evaluate_0.5.5     filehash_2.2-2
## [10] foreach_1.4.2      formatR_1.0        gdata_2.13.3
## [13] ggplot2_1.0.0      grid_3.1.1         gridBase_0.4-7
## [16] gtable_0.1.2       gtools_3.4.1       highr_0.4
## [19] iterators_1.0.7    KernSmooth_2.23-14 labeling_0.3
## [22] lattice_0.20-30    munsell_0.4.2      plyr_1.8.1
## [25] proto_0.3-10       Rcpp_0.11.4        reshape2_1.4.1
## [28] scales_0.2.4       stringr_0.6.2      tools_3.1.1
```