

SIS NMF Final: Diagnosis to DSD

December 5, 2014

1 Preparation

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

library(survival)

## Loading required package: splines

library(energy)
library(NMF)

## Loading required package: methods
## Loading required package: pkgmaker
## Loading required package: registry
## Loading required package: rngtools
## Loading required package: cluster
## NMF - BioConductor layer [OK] | Shared memory capabilities [OK] | Cores 7/8

library(nnlsl)

library(glmulti)

## Loading required package: rJava
##
## Attaching package: 'glmulti'
##
## The following object is masked from 'package:NMF':
##
##     consensus

library(glmnet)

## Loading required package: Matrix
## Loaded glmnet 1.9-8

library(RColorBrewer)
library(gplots)

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

```

library(xtable)
library(stargazer)

##
## Please cite as:
##
## Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression
and summary statistics tables.
## R package version 5.1. http://CRAN.R-project.org/package=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.d
cpvs.diag_dsd$Treat.Surgery.ExcisionStatus.Coarse = ordered(ifelse(cpvs.diag_dsd$Treat.Surgery.ExcisionS
"R0", "Clear", "Involved"), levels = c("Clear", "Involved"))
cpvs.diag_dsd$Path.Grade.Coarse = ordered(ifelse(cpvs.diag_dsd$Path.Grade %in%
c("1", "2"), "1or2", "3or4"), levels = c("1or2", "3or4"))
cpvs.diag_dsd$Path.TumourLocation.Coarse = factor(ifelse(cpvs.diag_dsd$Path.TumourLocation %in%
c("Head", "Head (Uncinate)", "Head", "Other"))

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

```



```

##      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

mean(cpss.sis$sel)

## [1] 0.02777

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

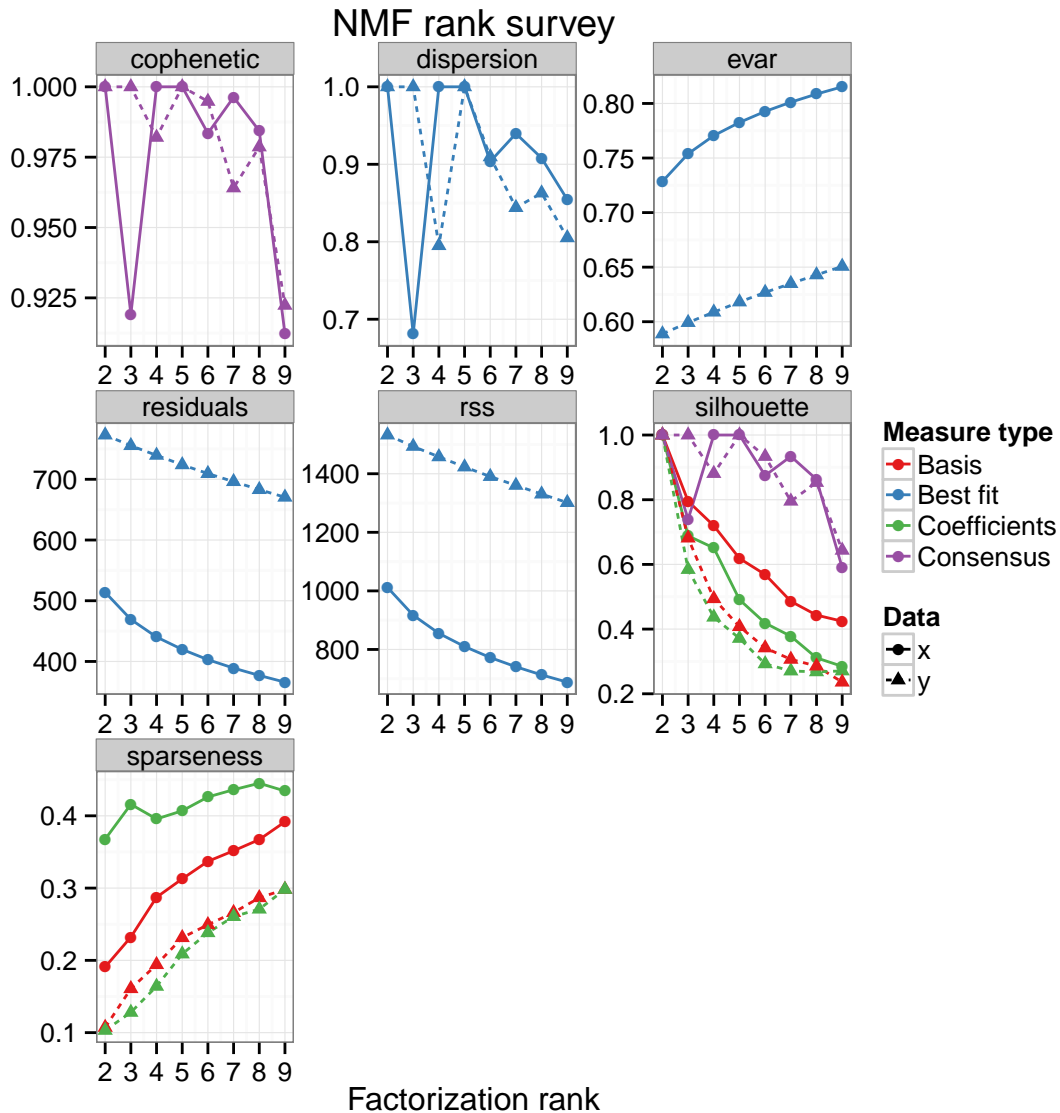
## [1] 37 175 92 32 298 49 47 138 43 173 98 86 207 102 147 41 28
## [18] 160 75 273 154 124 415 109 41 141 50 63 107 63 64 237 84 52
## [35] 40 203 88 55 98 87 57 231 54 48 81 186 114 43 58 347

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

## [1] 87.5
```

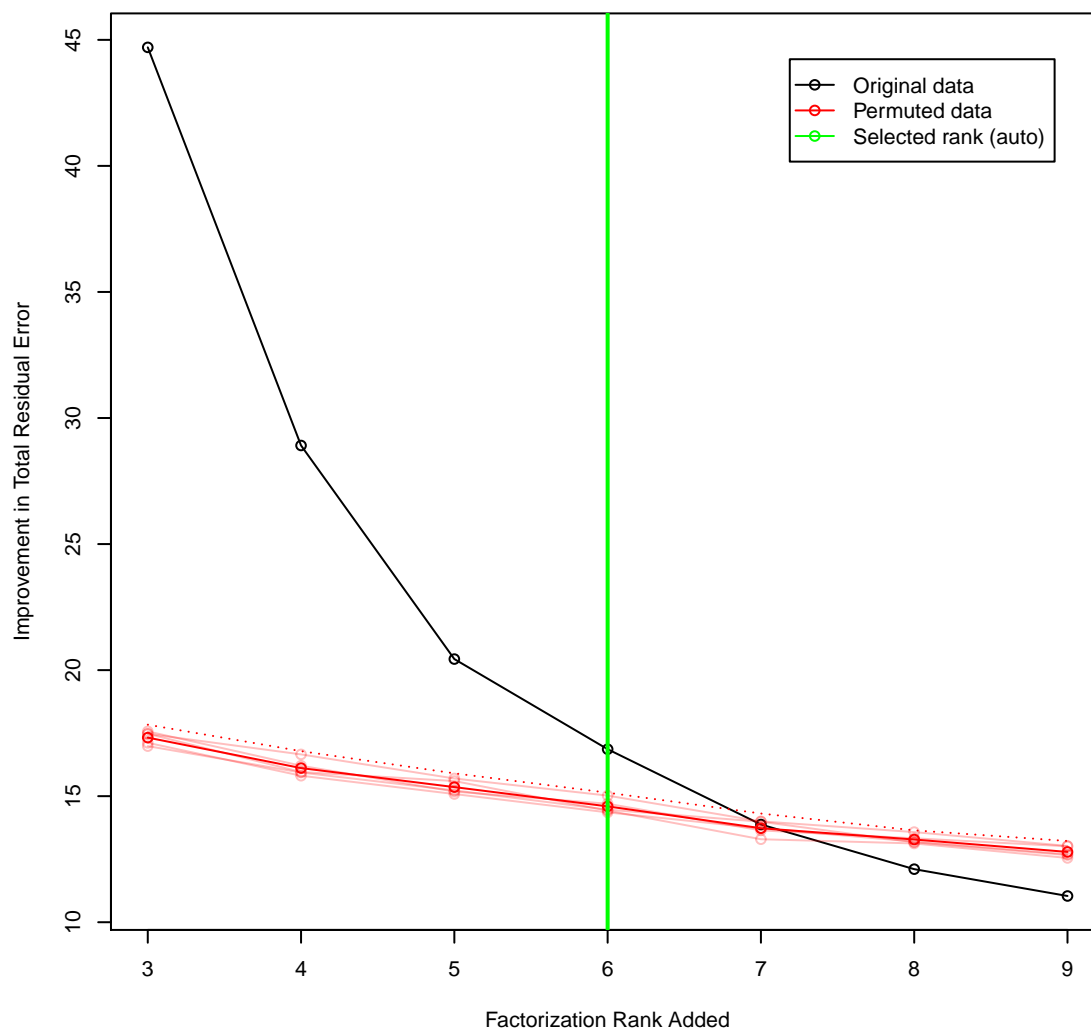
4 Factorization

```
plot(nmf.runs.rank, nmf.runs.rank.random[[1]])
```



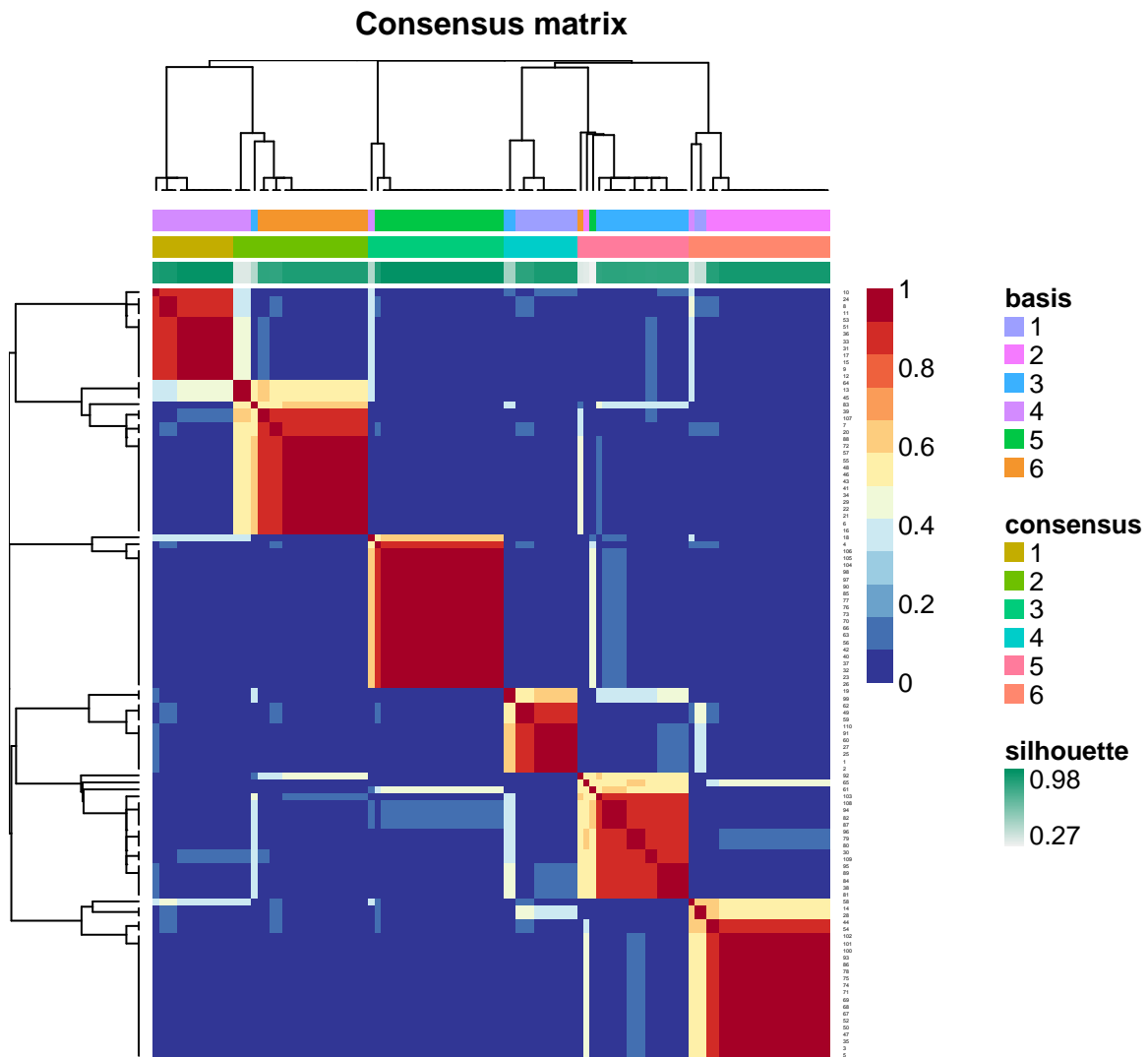
```
plot(nmf.rankrange[-1], -temp.orig_resids.delta, type = "o", col = "black",
     pch = 21, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.delta.mean)),
     xlab = "Factorization Rank Added", ylab = "Improvement in Total Residual Error")
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean, col = "red", type = "o",
      pch = 21, lwd = 1)
for (i in 1:ncol(temp.perm_resids)) {
  lines(nmf.rankrange[-1], -temp.perm_resids.delta[, i], type = "o", col = rgb(1,
    0, 0, 0.25))
}
lines(nmf.rankrange[-1], -temp.perm_resids.delta.threshold, col = "red", lty = "dotted")
if (nmf.rank.wasauto == TRUE) {
  temp.col = "green"
} else {
  temp.col = "blue"
}
```

```
abline(v = nmf.rank, col = temp.col, lwd = 2)
legend("topright", legend = c("Original data", "Permuted data", sprintf("Selected rank (%s)",
  ifelse(nmf.rank.wasauto == TRUE, "auto", "fixed"))), col = c("black", "red",
  temp.col), lty = "solid", pch = 21, inset = 0.05)
```

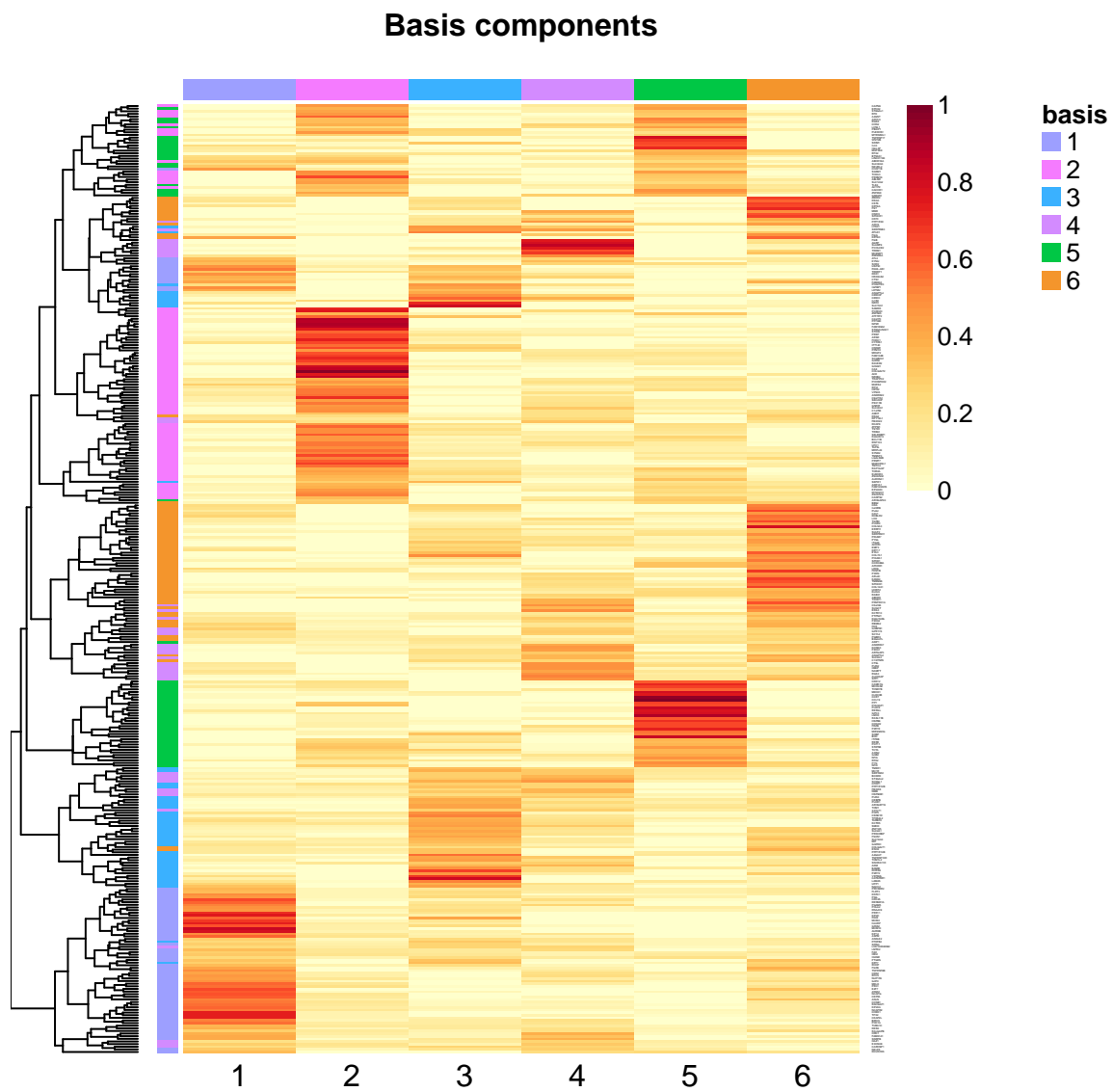


4.1 Fit

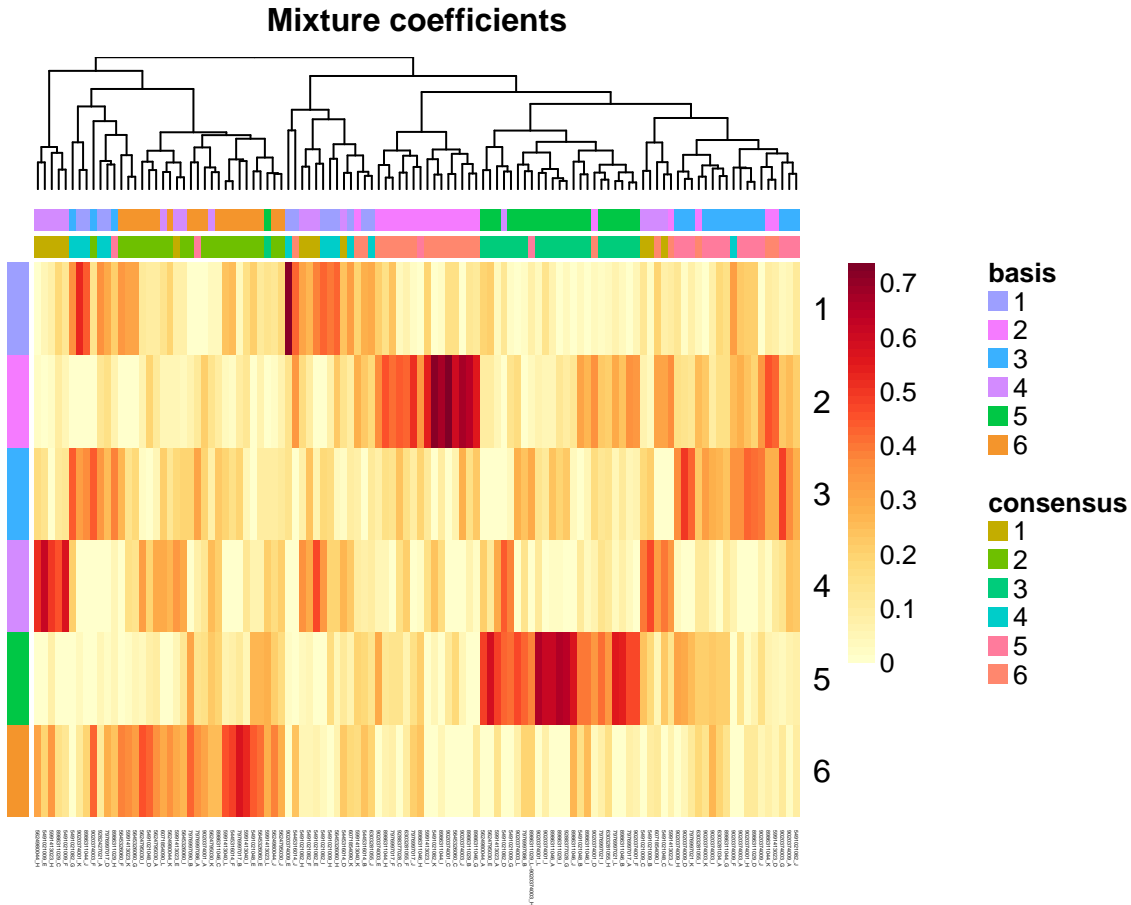
```
consensusmap(nmf.final)
```



```
basimap(nmf.final)
```



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



```

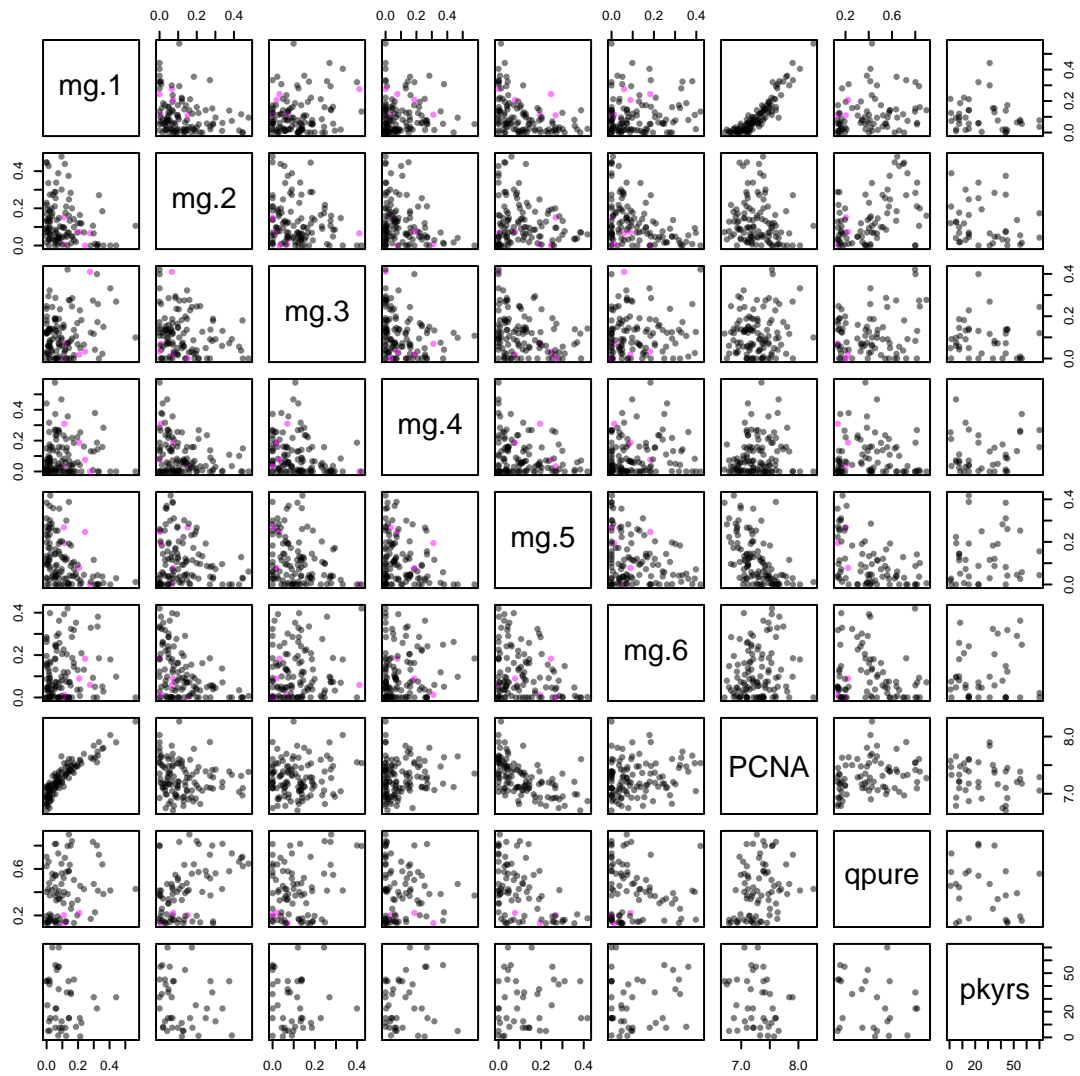
coefs.diag_dsd = apply(xlin.diag_dsd.sel, 2, function(xcol) nnls(basis(nmf.final),
  xcol)$x)
coefs.diag_rec = apply(xlin.diag_rec.sel, 2, function(xcol) nnls(basis(nmf.final),
  xcol)$x)
coefs.recr_dsd = apply(xlin.recr_dsd.sel, 2, function(xcol) nnls(basis(nmf.final),
  xcol)$x)
coefs.pdac_au = apply(xlin.pdac_au.sel, 2, function(xcol) nnls(basis(nmf.final),
  xcol)$x)

```

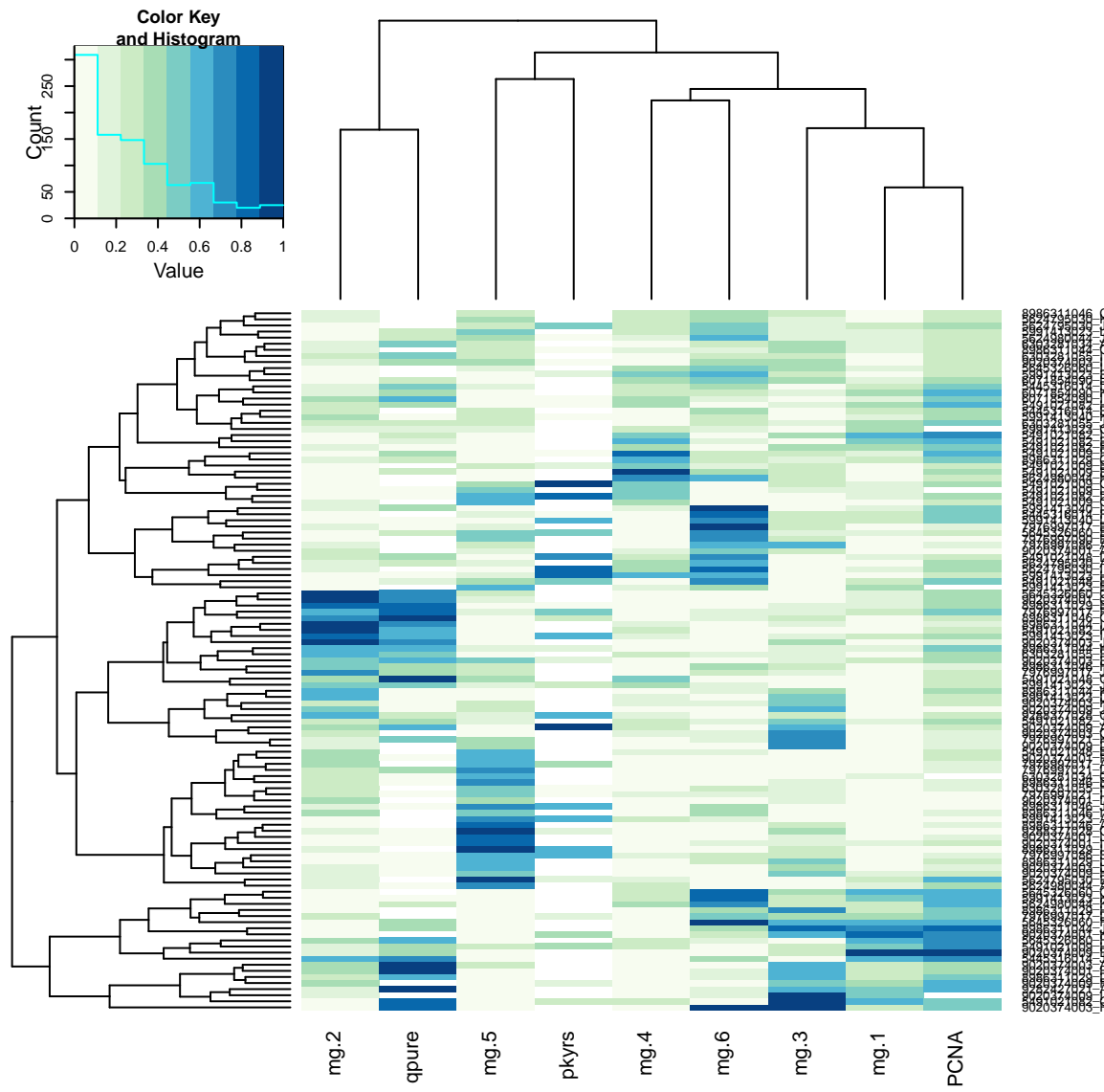
```

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 = samp.pdac_au$purity_qpure,
  pkyrs = cpvs.pdac_au$History.Smoking.PackYears)
pairs(temp.pred.pairs, pch = 16, cex = 1, col = ifelse(rownames(temp.pred.pairs) %in%
  colnames(xlin.diag_dsd.sel), rgb(0, 0, 0, 0.5), rgb(1, 0, 1, 0.5)))

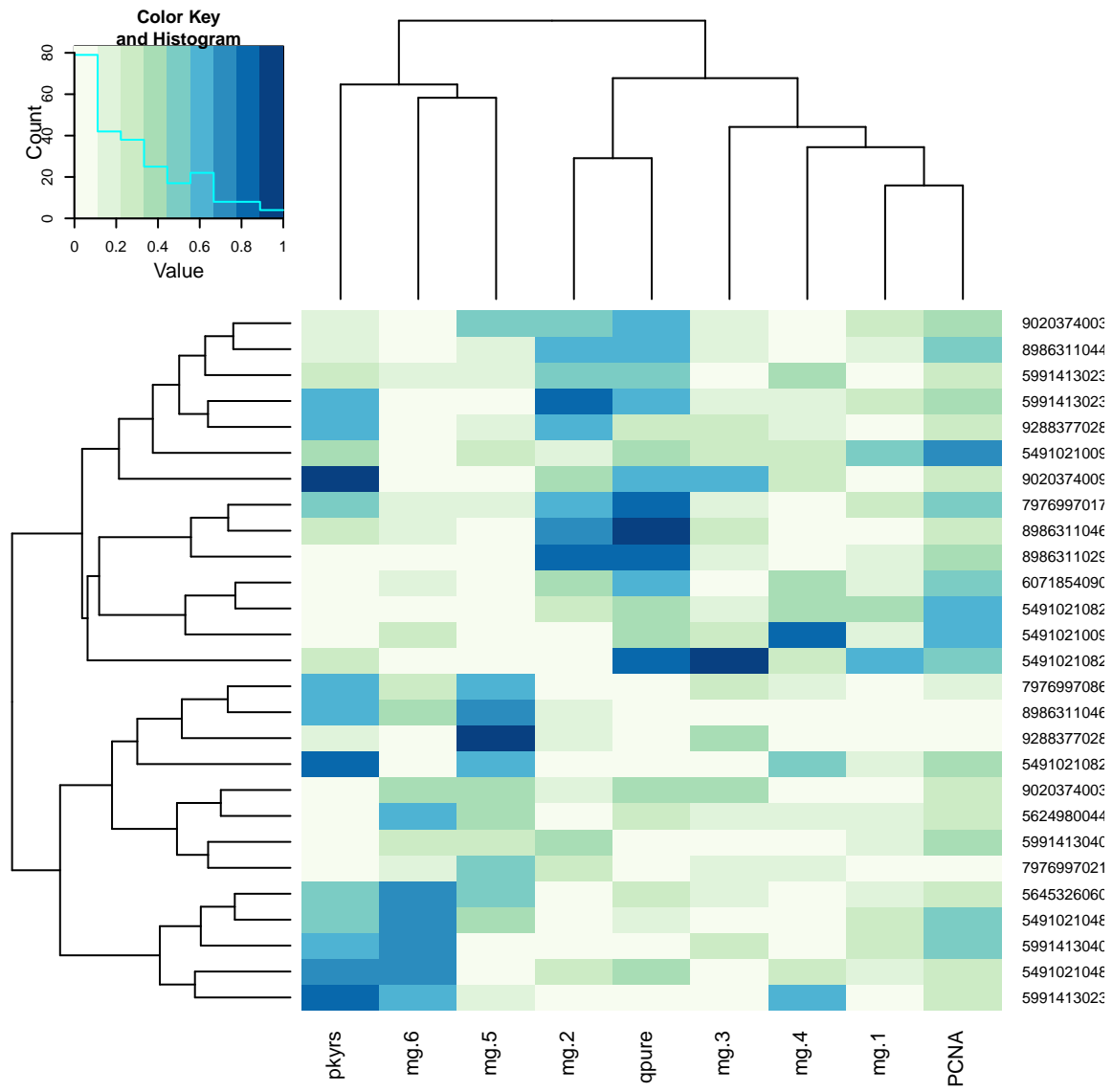
```



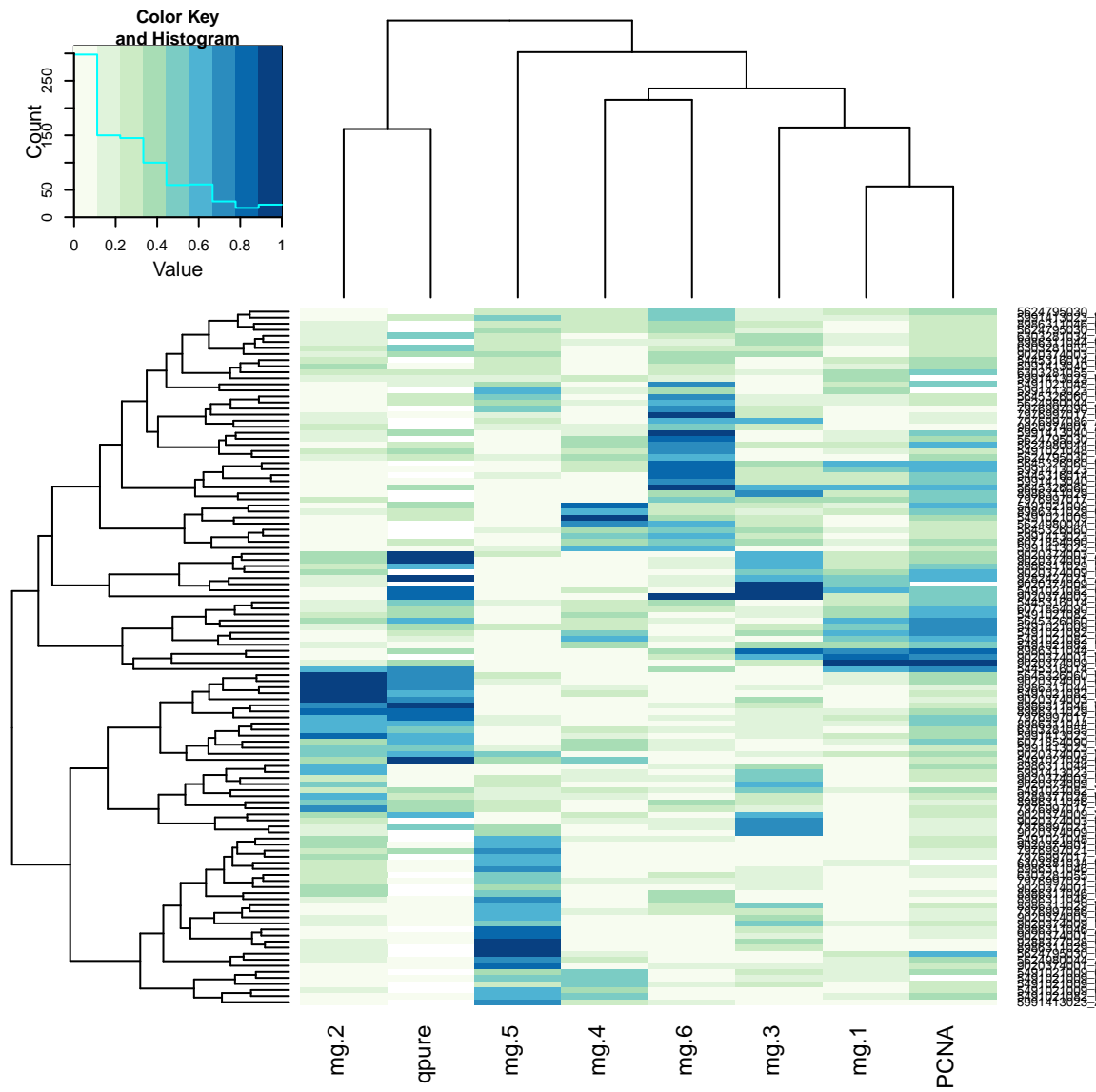
```
temp.pred.pairs.rescaled = t((t(temp.pred.pairs) - apply(temp.pred.pairs, 2,
  min, na.rm = TRUE))/(apply(temp.pred.pairs, 2, function(x) diff(range(x,
  na.rm = TRUE))))))
heatmap.2(temp.pred.pairs.rescaled, trace = "none", scale = "none", col = brewer.pal(9,
  "GnBu"))
```

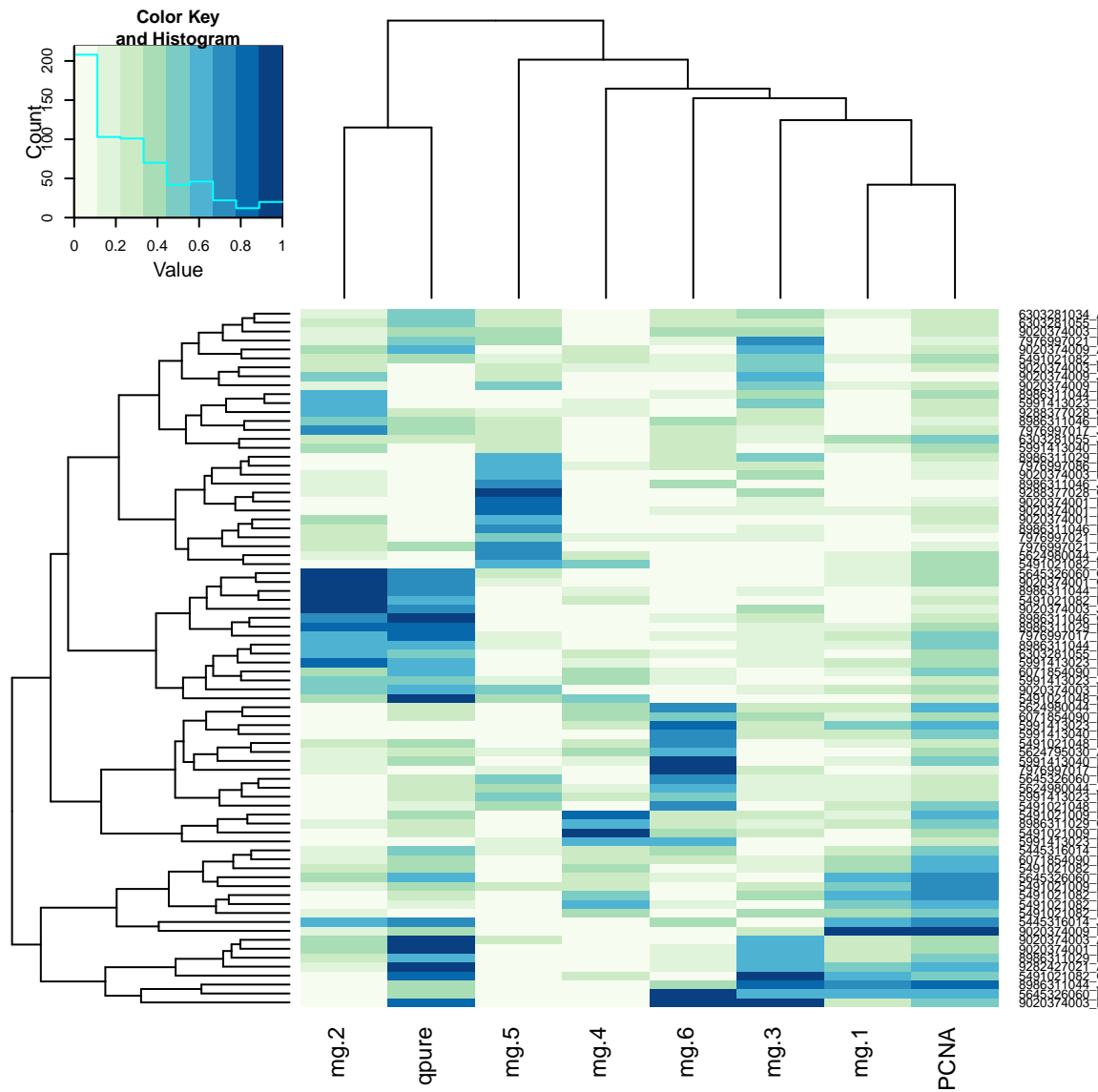
```
heatmap.2(temp.pred.pairs.rescaled[apply(!is.na(temp.pred.pairs.rescaled), 1,
all), ], trace = "none", scale = "none", col = brewer.pal(9, "GnBu"))
```



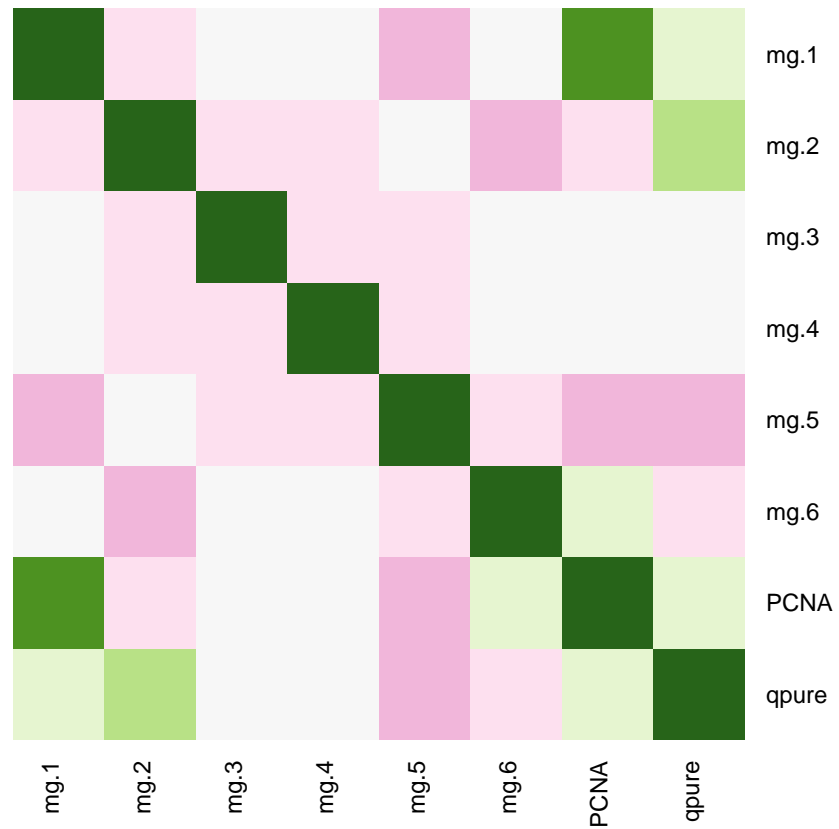
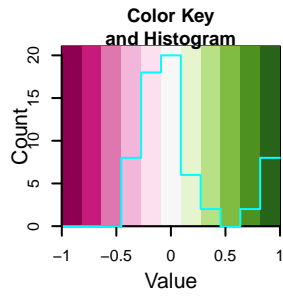
```
temp.pred.pairs.rescaled2 = temp.pred.pairs.rescaled[, colnames(temp.pred.pairs.rescaled) !=
"pkys"]
heatmap.2(temp.pred.pairs.rescaled2, trace = "none", scale = "none", col = brewer.pal(9,
"GnBu"))
```



```
heatmap.2(temp.pred.pairs.rescaled2[apply(!is.na(temp.pred.pairs.rescaled2),
1, all), ], trace = "none", scale = "none", col = brewer.pal(9, "GnBu"))
```



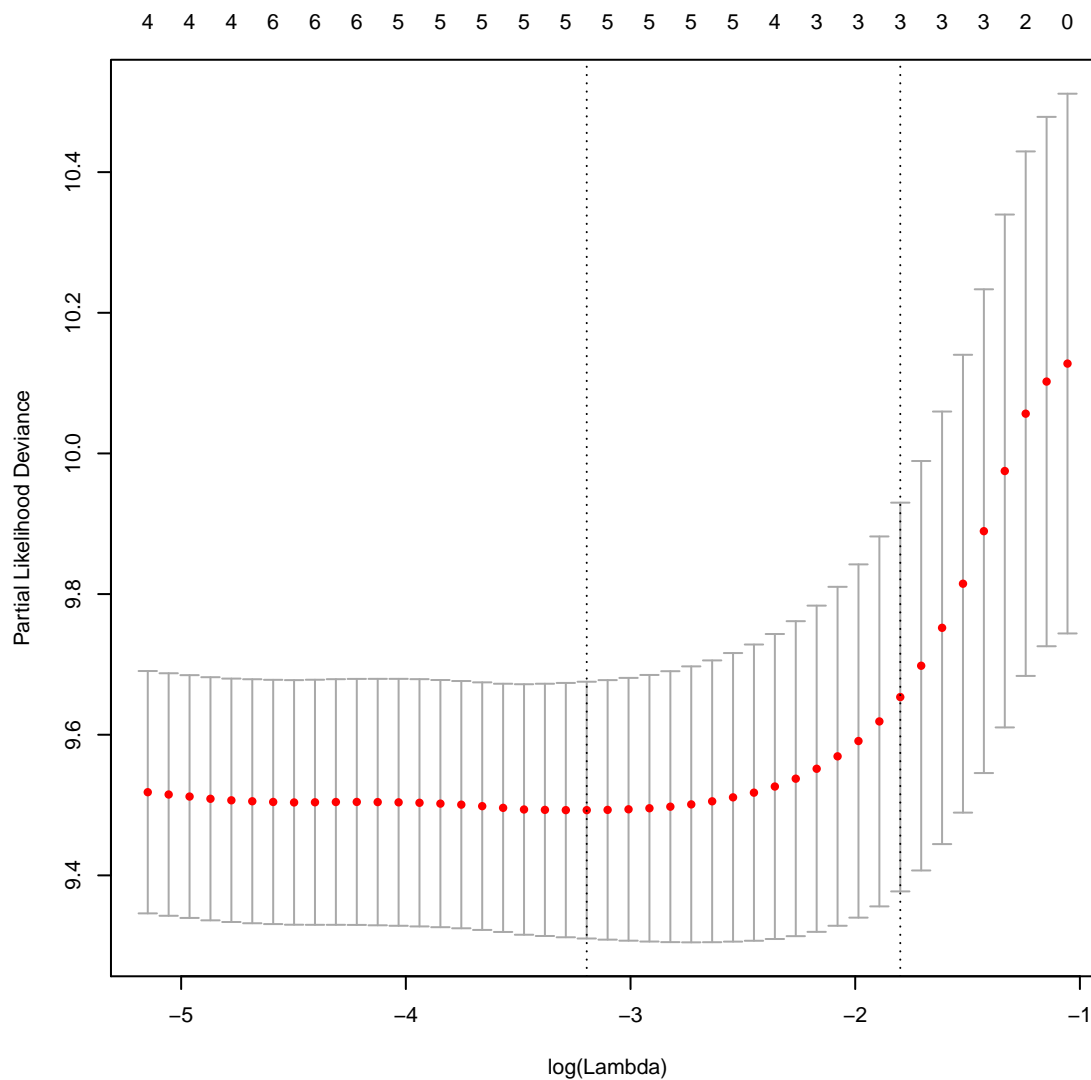
```
temp.cors = apply(temp.pred.pairs[, colnames(temp.pred.pairs) != "pkysr"], 2,
  function(x) apply(temp.pred.pairs[, colnames(temp.pred.pairs) != "pkysr"],
    2, function(y) {
      sel = !(is.na(x) | is.na(y))
      cor(x[sel], y[sel], method = "kendall")
    })
# diag(temp.cors) = NA
heatmap.2(temp.cors, trace = "none", Rowv = FALSE, Colv = FALSE, col = brewer.pal(11,
  "PiYG"), dendrogram = "none", scale = "none")
```



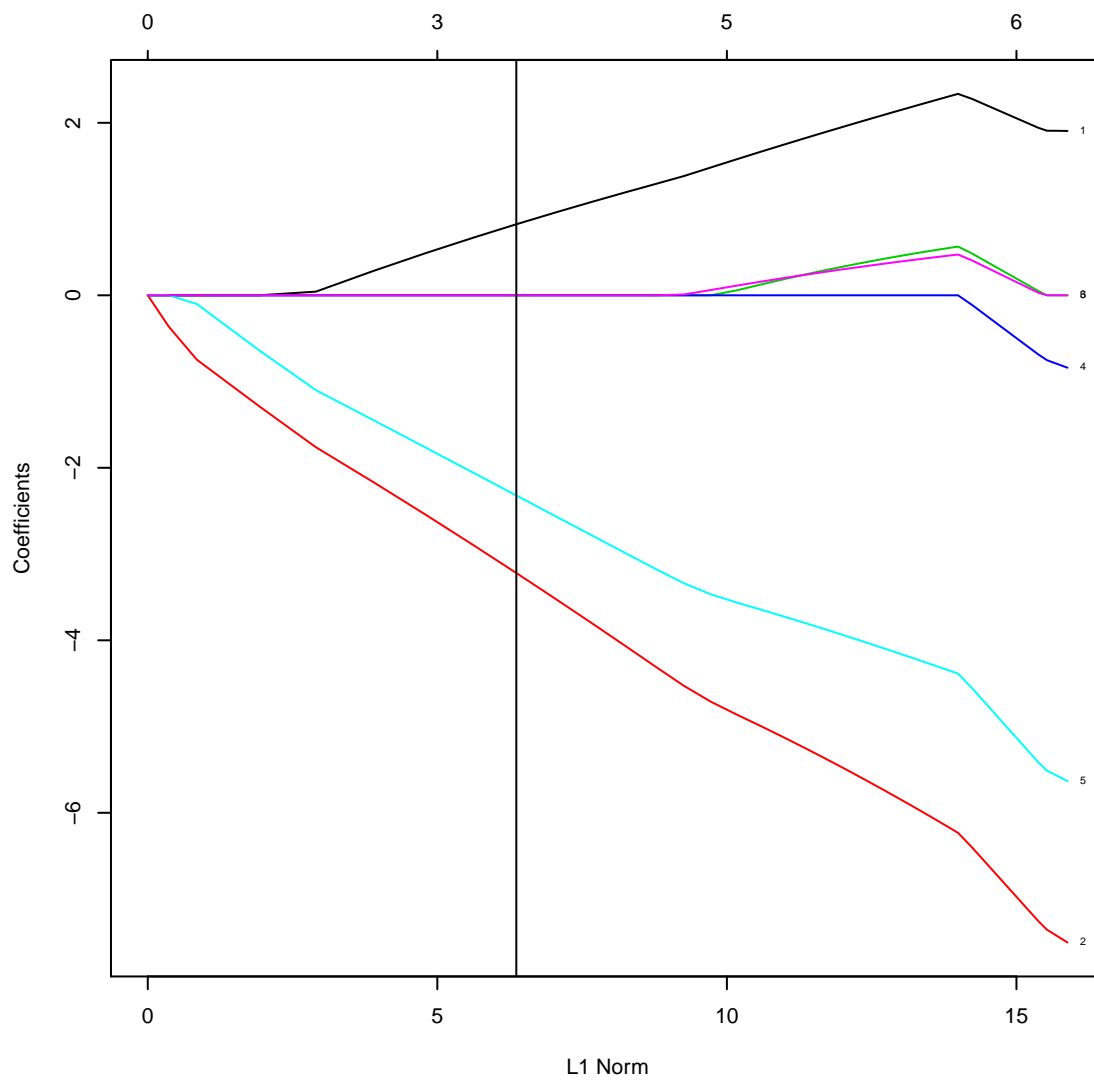
4.2 LASSO on training set

```
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)
```

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



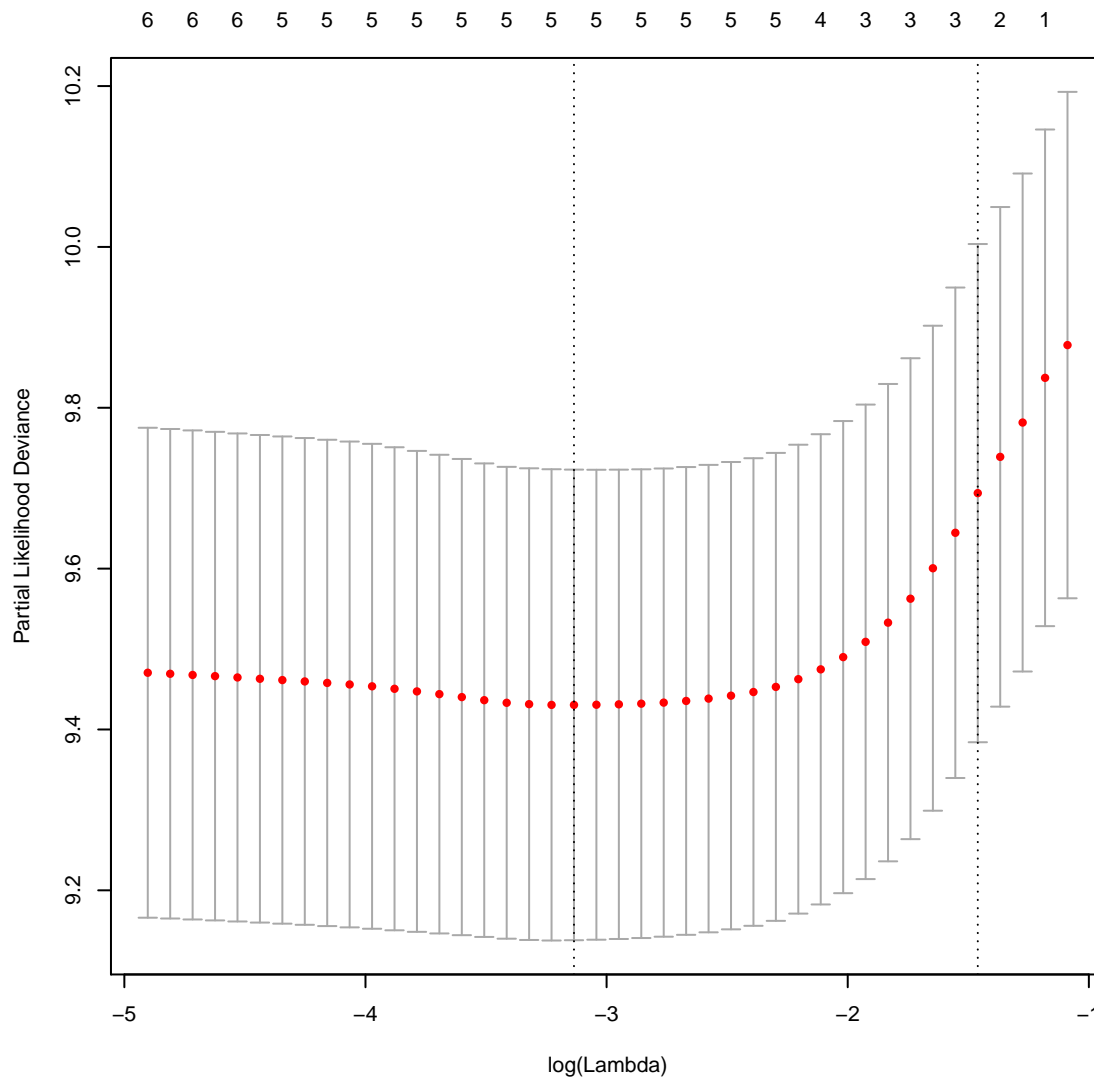
```
plot(glmnet.fit.cv.diag_dsd$glmnet.fit, label = TRUE)
abline(v = sum(abs(coef(glmnet.fit.cv.diag_dsd$glmnet.fit, s = glmnet.fit.cv.diag_dsd$lambda.1se))))
```



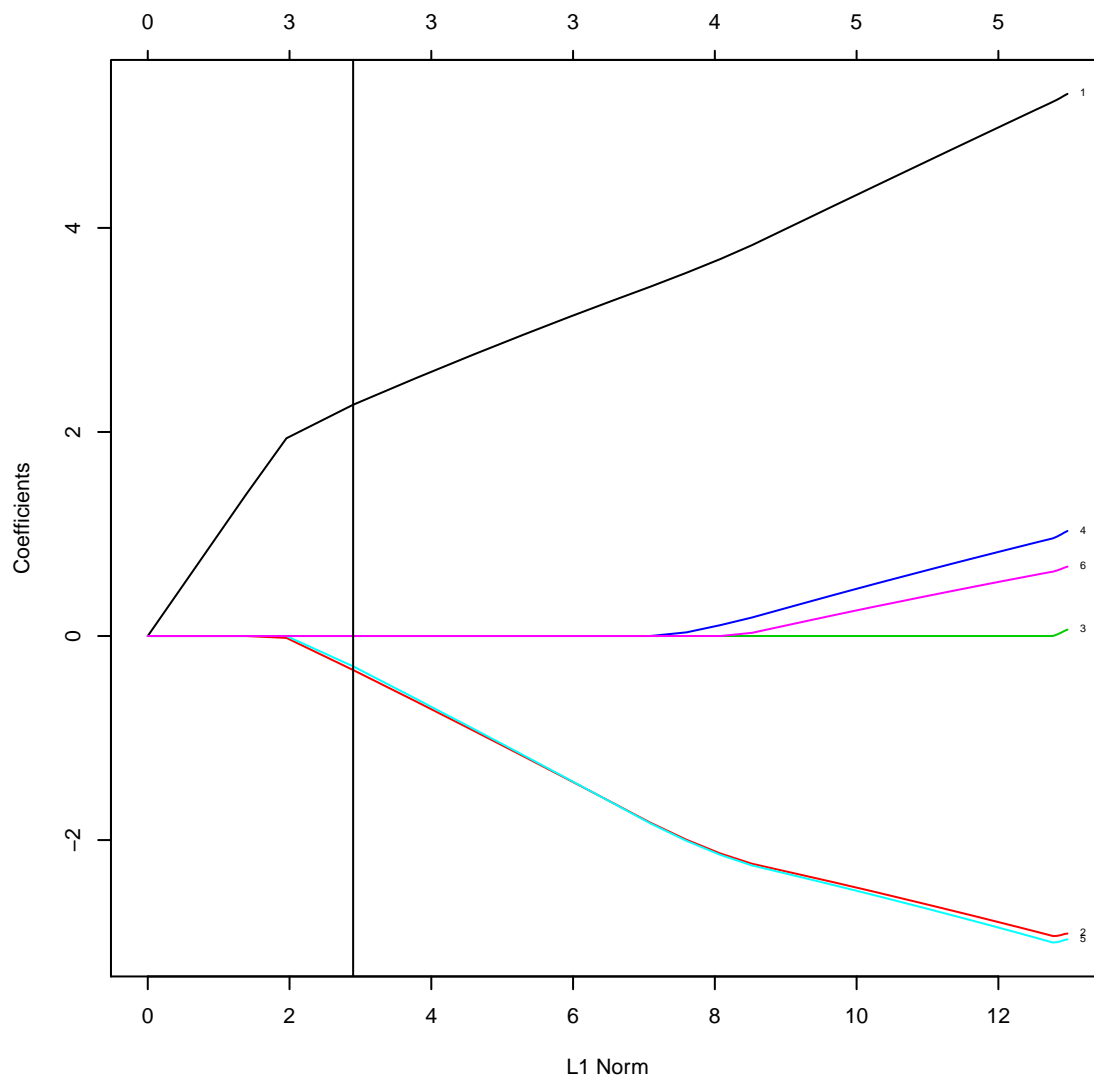
```
# abline(v = sum(abs(coef(glmnet.fit.cv.diag_dsd$glmnet.fit, s =
# glmnet.fit.cv.diag_dsd$lambda.min))))
coef(glmnet.fit.cv.diag_dsd$glmnet.fit, s = glmnet.fit.cv.diag_dsd$lambda.1se)

## 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)
```



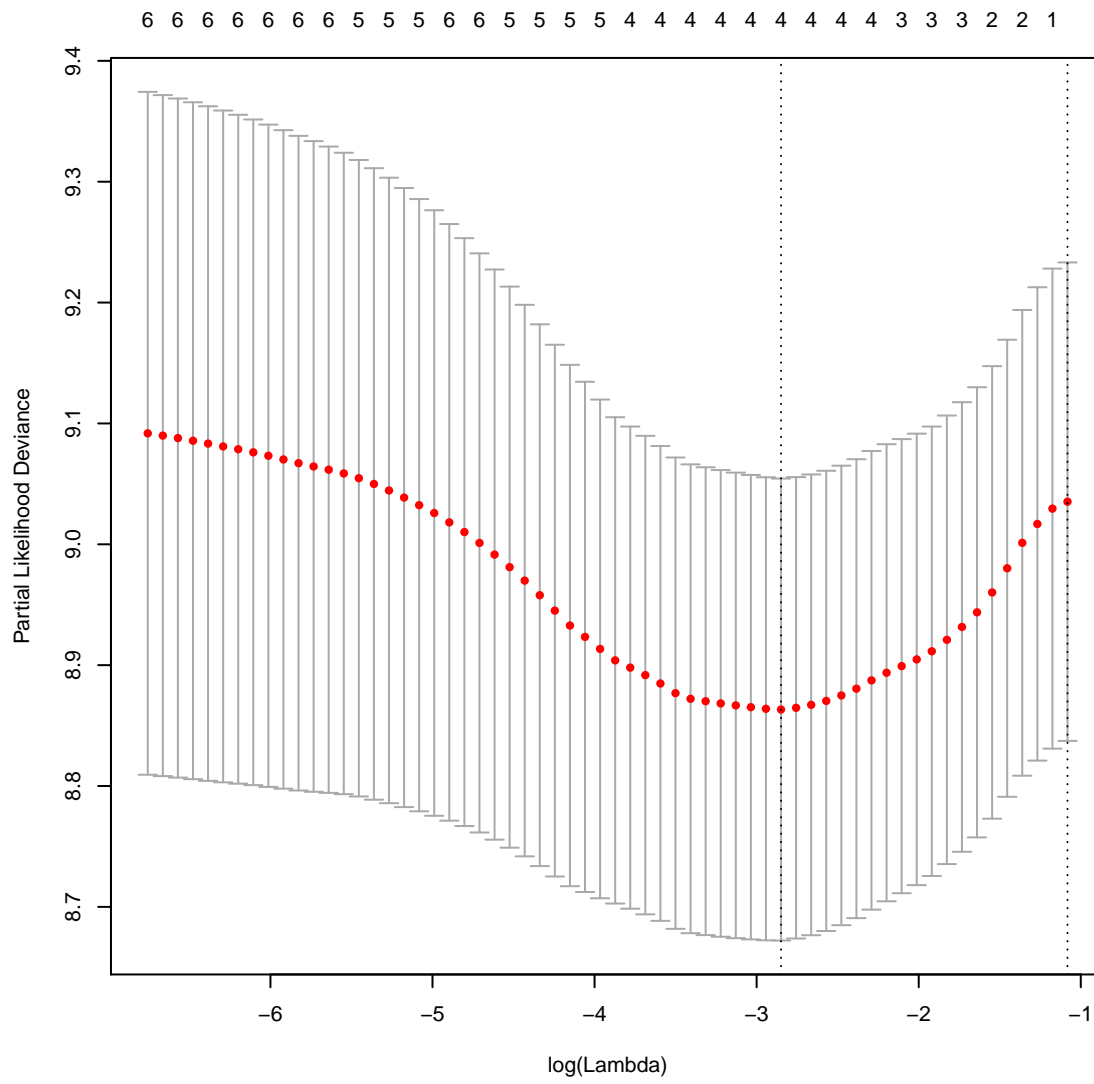
```
plot(glmnet.fit.cv.diag_rec$glmnet.fit, label = TRUE)
abline(v = sum(abs(coef(glmnet.fit.cv.diag_rec$glmnet.fit, s = glmnet.fit.cv.diag_rec$lambda.1se))))
```

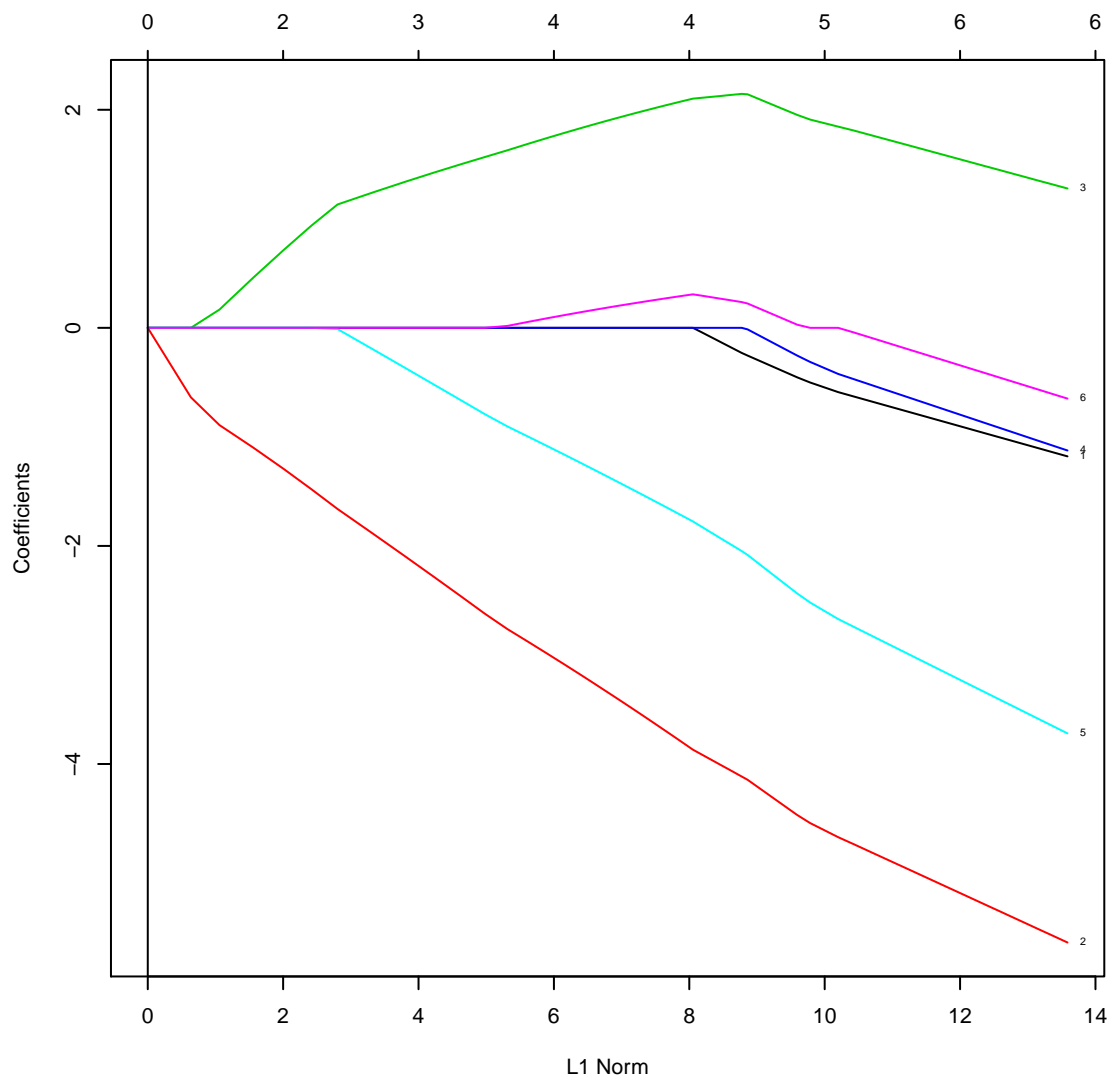
```
# abline(v = sum(abs(coef(glmnet.fit.cv.diag_rec$glmnet.fit, s =
# glmnet.fit.cv.diag_rec$lambda.min))))
coef(glmnet.fit.cv.diag_rec$glmnet.fit, s = glmnet.fit.cv.diag_rec$lambda.1se)

## 6 x 1 sparse Matrix of class "dgCMatrix"
##      1
## V1  2.2667
## V2 -0.3332
## V3   .
## V4   .
## V5 -0.2974
## 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 .
## V3 .
## V4 .
## V5 .
## V6 .
```

4.3 Prediction on 10-fold CV

```
cv_preds = readRDS("../analysis/14_SIS_NMF_CV_results.rds")
```

```
summary(coxph(y.diag_dsd ~ cv_preds["lasso.1se", ]))

## Call:
## coxph(formula = y.diag_dsd ~ cv_preds["lasso.1se", ])
##
##      n= 110, number of events= 70
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## cv_preds["lasso.1se", ] 0.606      1.833      0.200 3.03   0.0024
##
##              exp(coef) exp(-coef) lower .95 upper .95
## cv_preds["lasso.1se", ]      1.83      0.545      1.24      2.71
##
## Concordance= 0.594 (se = 0.038 )
## Rsquare= 0.078 (max possible= 0.995 )
## Likelihood ratio test= 8.91 on 1 df,  p=0.00284
## Wald test               = 9.18 on 1 df,  p=0.00244
## Score (logrank) test = 9.07 on 1 df,  p=0.0026
```

4.4 Prediction on validation sets

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

```
val.basis = basis(nmf.final)
rownames(GSE21501.lingex) = GSE21501.feats$Gene.symbol
rownames(GSE28735.lingex) = GSE28735.feats$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.pcna = apply(GSE21501.gex[match(metapcna.sig, GSE21501.feats$Gene.symbol),
], 2, median, na.rm = TRUE)
GSE28735.pcna = apply(GSE28735.gex[match(metapcna.sig, GSE28735.feats$Gene.symbol),
], 2, median, na.rm = TRUE)
```

```
temp = coxph(Surv(GSE21501.samps$time, GSE21501.samps$event) ~ predict(glmnet.fit.cv.diag_dsd$glmnet.fit,
newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se))
summary(temp)
```

```

## Call:
## coxph(formula = Surv(GSE21501.samp$time, GSE21501.samp$event) ~
##       predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs),
##       s = glmnet.fit.cv.recr_dsd$lambda.1se))
##
## n= 102, number of events= 66
##
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE21501.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## Concordance= 0.499 (se = 0.039 )
## Rsquare= 0 (max possible= 0.993 )
## Likelihood ratio test= 0.04 on 1 df, p=0.849
## Wald test = 0.04 on 1 df, p=0.849
## Score (logrank) test = 0.04 on 1 df, p=0.849

temp = coxph(Surv(GSE28735.samp$time, GSE28735.samp$event) ~ predict(glmnet.fit.cv.diag_dsd$glmnet.fit,
newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se))
summary(temp)

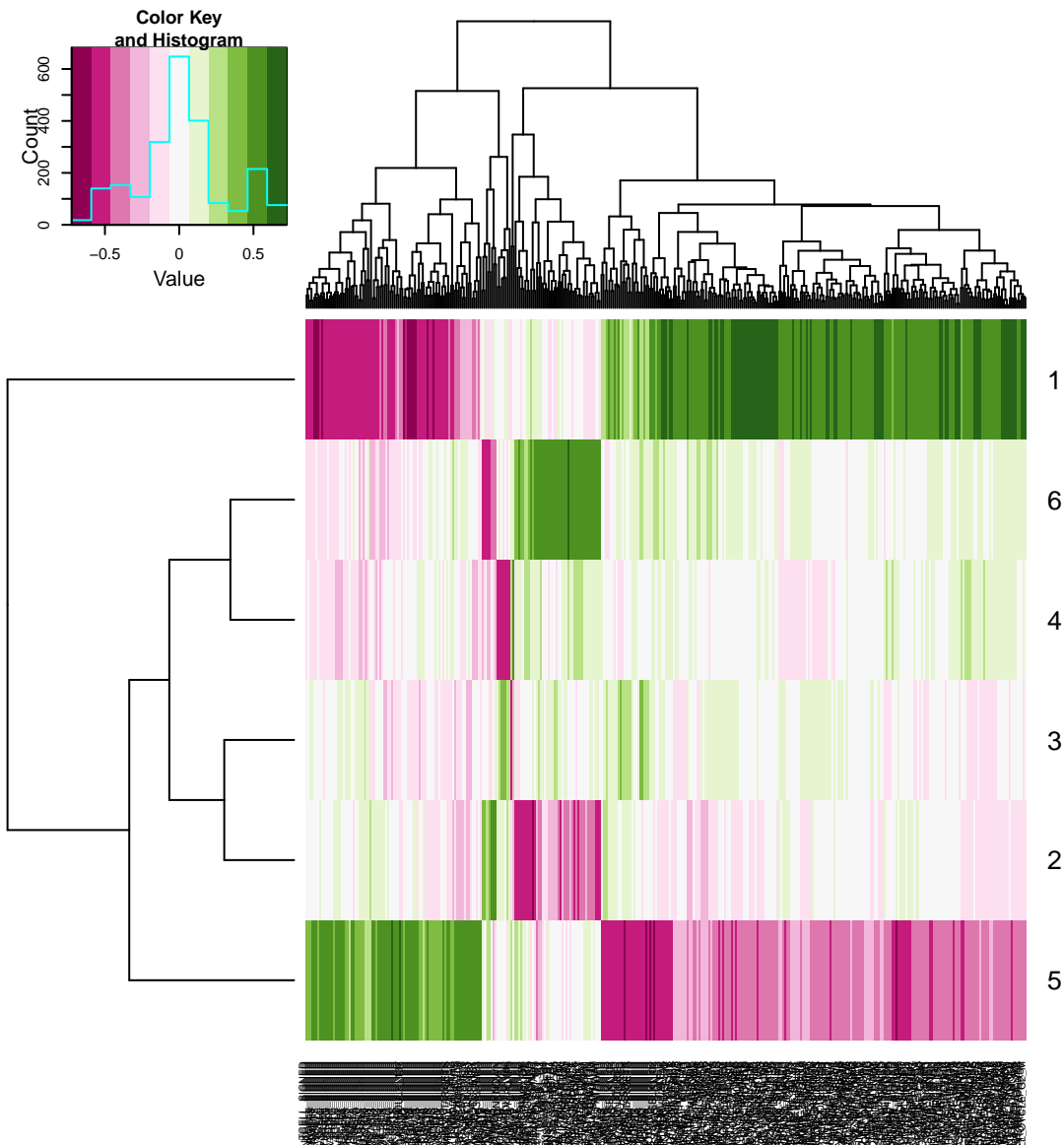
## Call:
## coxph(formula = Surv(GSE28735.samp$time, GSE28735.samp$event) ~
##       predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs),
##       s = glmnet.fit.cv.recr_dsd$lambda.1se))
##
## n= 42, number of events= 29
##
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda.1se)

```

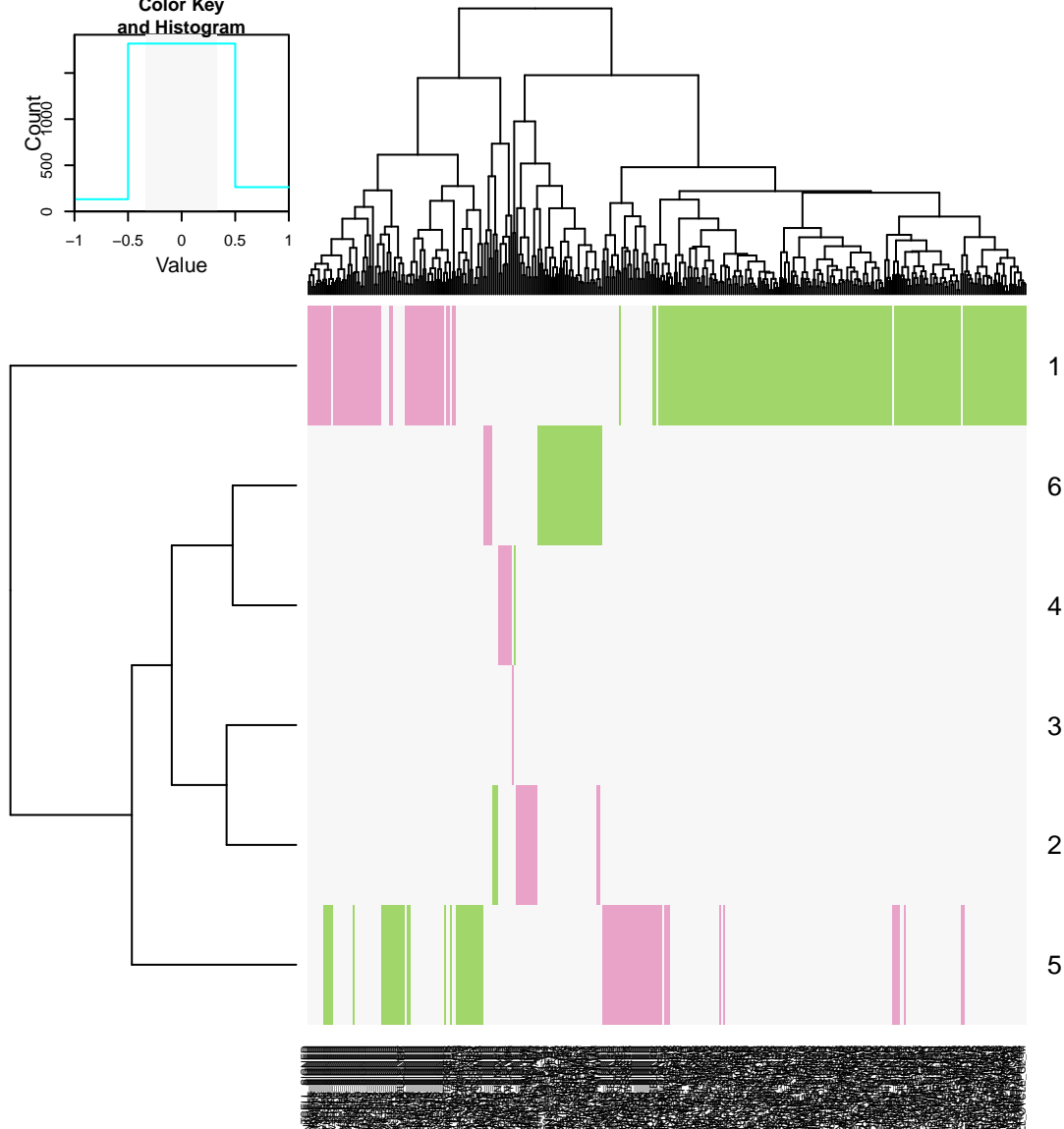
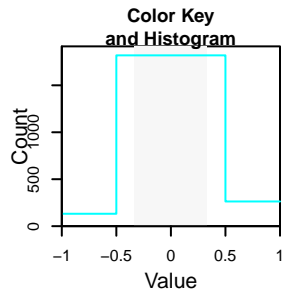
```
##
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda)
##
## predict(glmnet.fit.cv.diag_dsd$glmnet.fit, newx = t(GSE28735.coefs), s = glmnet.fit.cv.recr_dsd$lambda)
##
## Concordance= 0.693 (se = 0.064 )
## Rsquare= 0.134 (max possible= 0.981 )
## Likelihood ratio test= 6.06 on 1 df, p=0.0139
## Wald test = 5.71 on 1 df, p=0.0169
## Score (logrank) test = 6.04 on 1 df, p=0.014
```

4.5 MSigDB score correlation thresholding

```
temp.sel_cols = apply(abs(nmf.final.msigdb.corr) >= sig.corr.threshold, 2, any)
heatmap.2(nmf.final.msigdb.corr[, temp.sel_cols], trace = "none", scale = "none",
          useRaster = TRUE, col = brewer.pal(11, "PiYG"), symbreaks = TRUE)
```



```
heatmap.2(nmf.final.msigsdb.corr[, temp.sel_cols], trace = "none", scale = "none",
  useRaster = TRUE, col = brewer.pal(3, "PiYG"), breaks = c(-1, -sig.corr.threshold,
    sig.corr.threshold, 1))
```



```

cpv.pvals = apply(coef(nmf.final), 1, function(mg) sapply(cbind(cpv.diag_dsd,
  purity = samp$purity_qpure), function(x) {
    s = !is.na(mg) & !is.na(x)
    x = x[s]
    mg = mg[s]
    if (any(c("numeric", "integer") %in% class(x))) {
      return(cor.test(x, mg, method = "pearson")$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

##              [,1]      [,2]      [,3]      [,4]
## Patient.Gender      0.138407 2.069e-02 0.72732 0.135009
## Patient.Ethnicity    0.622781 7.244e-01 0.18976 0.765995
## History.Smoking.PackYears 0.348090 2.604e-01 0.73136 0.321908
## History.Diagnosis.AgeAtYears 0.881590 6.816e-01 0.49230 0.914690
## Path.HistoType.Subtype 0.893359 5.620e-01 0.22914 0.104016
## Path.TumourSizeMm     0.940836 1.352e-01 0.70907 0.215133
## Path.Invasion.PN      0.326270 6.452e-02 0.89694 0.509692
## Path.Invasion.VS      0.626581 1.745e-01 0.93410 0.631215
## Staging.pM            0.396481 3.624e-01 0.85889 0.135211
## Staging.pN            0.731239 1.811e-01 0.62403 0.751373
## Staging.pT            0.137691 5.530e-01 0.34901 0.064967
## Staging.Stage         0.038185 2.195e-01 0.60239 0.089545
## History.Recurrence.Site.Peritoneum 0.958146 5.285e-02 0.61695 0.281081
## History.Recurrence.Site.PancRemnant 0.551611 2.399e-01 0.01780 0.914636
## History.Recurrence.Site.PancBed    0.869281 3.251e-01 0.66994 0.475247
## History.Recurrence.Site.Other      0.682468 2.960e-01 0.70889 0.003592
## History.Recurrence.Site.Omentum    0.120811 2.790e-01 0.95417 0.784937
## History.Recurrence.Site.Mesentery  0.502274 2.771e-01 0.78581 0.707436
## History.Recurrence.Site.LymphNodes 0.331799 9.865e-01 0.63649 0.014826
## History.Recurrence.Site.Lung       0.488661 2.979e-01 0.55622 0.748563
## History.Recurrence.Site.Liver      0.129400 4.693e-02 0.06894 0.547538
## History.Recurrence.Site.Brain      0.774503 2.987e-02 0.63502 0.971582
## History.Recurrence.Site.Bone       0.782178 8.047e-01 0.98003 0.854252
## Path.Grade.Coarse      0.006438 5.655e-03 0.36879 0.129514
## Path.TumourLocation.Coarse 0.469811 2.301e-01 0.07124 0.567640
## purity                0.047335 1.512e-05 0.09329 0.463260
##              [,5]      [,6]
## Patient.Gender      3.733e-01 0.0387623
## Patient.Ethnicity    9.939e-01 0.0191721
## History.Smoking.PackYears 5.061e-01 0.4738379
## History.Diagnosis.AgeAtYears 7.512e-01 0.7515763
## Path.HistoType.Subtype 5.660e-01 0.0595163
## Path.TumourSizeMm     8.154e-01 0.4037090
## Path.Invasion.PN      6.422e-02 0.8347923
## Path.Invasion.VS      7.806e-01 0.0862062
## Staging.pM           7.803e-01 0.2488073

```

```

## Staging.pN      1.124e-01 0.5892122
## Staging.pT      6.051e-01 0.4778634
## Staging.Stage   1.922e-01 0.4591937
## History.Recurrence.Site.Peritoneum 8.085e-01 0.0275532
## History.Recurrence.Site.PancRemnant 6.621e-01 0.2693233
## History.Recurrence.Site.PancBed    9.389e-01 0.9235058
## History.Recurrence.Site.Other       6.915e-02 0.1710832
## History.Recurrence.Site.Omentum    3.392e-01 0.0596320
## History.Recurrence.Site.Mesentery   4.057e-01 0.1203379
## History.Recurrence.Site.LymphNodes  3.963e-01 0.7818857
## History.Recurrence.Site.Lung        2.854e-02 0.0896567
## History.Recurrence.Site.Liver       3.725e-01 0.4265668
## History.Recurrence.Site.Brain       2.810e-01 0.3088652
## History.Recurrence.Site.Bone        4.605e-01 0.0952863
## Path.Grade.Coarse 1.468e-02 0.0001869
## Path.TumourLocation.Coarse 1.092e-01 0.1813727
## purity          1.375e-05 0.1063724

```

cpv.qvals

```

##           [,1]      [,2] [,3]      [,4]      [,5]
## Patient.Gender 1.0000 1.000000 1 1.0000 1.000000
## Patient.Ethnicity 1.0000 1.000000 1 1.0000 1.000000
## History.Smoking.PackYears 1.0000 1.000000 1 1.0000 1.000000
## History.Diagnosis.AgeAtYears 1.0000 1.000000 1 1.0000 1.000000
## Path.HistoType.Subtype 1.0000 1.000000 1 1.0000 1.000000
## Path.TumourSizeMm 1.0000 1.000000 1 1.0000 1.000000
## Path.Invasion.PN 1.0000 1.000000 1 1.0000 1.000000
## Path.Invasion.VS 1.0000 1.000000 1 1.0000 1.000000
## Staging.pM      1.0000 1.000000 1 1.0000 1.000000
## Staging.pN      1.0000 1.000000 1 1.0000 1.000000
## Staging.pT      1.0000 1.000000 1 1.0000 1.000000
## Staging.Stage   1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Peritoneum 1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.PancRemnant 1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.PancBed    1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Other       1.0000 1.000000 1 0.5496 1.000000
## History.Recurrence.Site.Omentum    1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Mesentery   1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.LymphNodes  1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Lung        1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Liver       1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Brain       1.0000 1.000000 1 1.0000 1.000000
## History.Recurrence.Site.Bone        1.0000 1.000000 1 1.0000 1.000000
## Path.Grade.Coarse 0.9721 0.859520 1 1.0000 1.000000
## Path.TumourLocation.Coarse 1.0000 1.000000 1 1.0000 1.000000
## purity          1.0000 0.002343 1 1.0000 0.002145
##           [,6]
## Patient.Gender 1.00000
## Patient.Ethnicity 1.00000
## History.Smoking.PackYears 1.00000
## History.Diagnosis.AgeAtYears 1.00000
## Path.HistoType.Subtype 1.00000
## Path.TumourSizeMm 1.00000

```

```
## Path.Invasion.PN 1.00000
## Path.Invasion.VS 1.00000
## Staging.pM 1.00000
## Staging.pN 1.00000
## Staging.pT 1.00000
## Staging.Stage 1.00000
## History.Recurrence.Site.Peritoneum 1.00000
## History.Recurrence.Site.PancRemnant 1.00000
## History.Recurrence.Site.PancBed 1.00000
## History.Recurrence.Site.Other 1.00000
## History.Recurrence.Site.Omentum 1.00000
## History.Recurrence.Site.Mesentery 1.00000
## History.Recurrence.Site.LymphNodes 1.00000
## History.Recurrence.Site.Lung 1.00000
## History.Recurrence.Site.Liver 1.00000
## History.Recurrence.Site.Brain 1.00000
## History.Recurrence.Site.Bone 1.00000
## Path.Grade.Coarse 0.02879
## Path.TumourLocation.Coarse 1.00000
## purity 1.00000
```

```
cpv.pvals = cpv.pvals[, c(1, 2, 5, 6)]
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
```

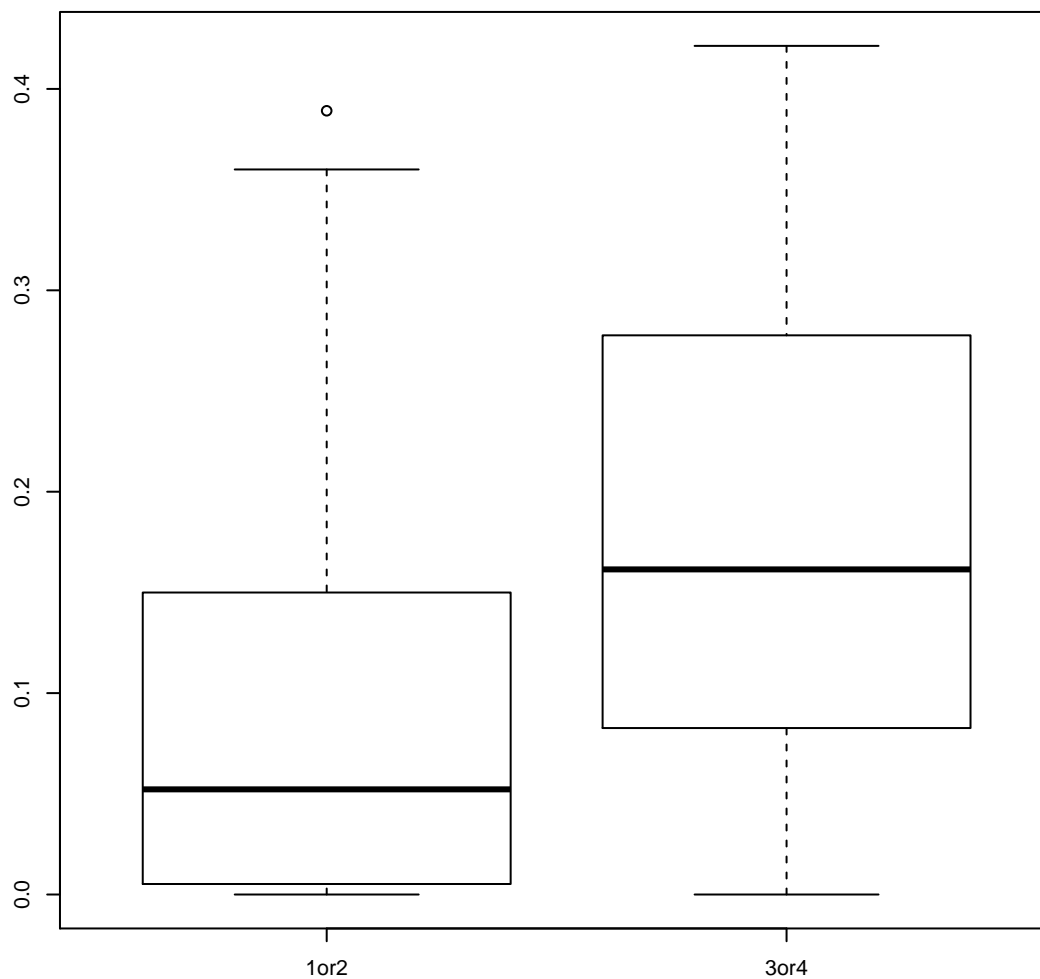
##	[,1]	[,2]	[,3]	[,4]
## Patient.Gender	0.138407	2.069e-02	3.733e-01	0.0387623
## Patient.Ethnicity	0.622781	7.244e-01	9.939e-01	0.0191721
## History.Smoking.PackYears	0.348090	2.604e-01	5.061e-01	0.4738379
## History.Diagnosis.AgeAtYears	0.881590	6.816e-01	7.512e-01	0.7515763
## Path.HistoType.Subtype	0.893359	5.620e-01	5.660e-01	0.0595163
## Path.TumourSizeMm	0.940836	1.352e-01	8.154e-01	0.4037090
## Path.Invasion.PN	0.326270	6.452e-02	6.422e-02	0.8347923
## Path.Invasion.VS	0.626581	1.745e-01	7.806e-01	0.0862062
## Staging.pM	0.396481	3.624e-01	7.803e-01	0.2488073
## Staging.pN	0.731239	1.811e-01	1.124e-01	0.5892122
## Staging.pT	0.137691	5.530e-01	6.051e-01	0.4778634
## Staging.Stage	0.038185	2.195e-01	1.922e-01	0.4591937
## History.Recurrence.Site.Peritoneum	0.958146	5.285e-02	8.085e-01	0.0275532
## History.Recurrence.Site.PancRemnant	0.551611	2.399e-01	6.621e-01	0.2693233
## History.Recurrence.Site.PancBed	0.869281	3.251e-01	9.389e-01	0.9235058
## History.Recurrence.Site.Other	0.682468	2.960e-01	6.915e-02	0.1710832
## History.Recurrence.Site.Omentum	0.120811	2.790e-01	3.392e-01	0.0596320
## History.Recurrence.Site.Mesentery	0.502274	2.771e-01	4.057e-01	0.1203379
## History.Recurrence.Site.LymphNodes	0.331799	9.865e-01	3.963e-01	0.7818857
## History.Recurrence.Site.Lung	0.488661	2.979e-01	2.854e-02	0.0896567
## History.Recurrence.Site.Liver	0.129400	4.693e-02	3.725e-01	0.4265668
## History.Recurrence.Site.Brain	0.774503	2.987e-02	2.810e-01	0.3088652
## History.Recurrence.Site.Bone	0.782178	8.047e-01	4.605e-01	0.0952863
## Path.Grade.Coarse	0.006438	5.655e-03	1.468e-02	0.0001869

```
## Path.TumourLocation.Coarse      0.469811 2.301e-01 1.092e-01 0.1813727
## purity                          0.047335 1.512e-05 1.375e-05 0.1063724
```

```
cpv.qvals
```

```
##           [,1]      [,2]      [,3]      [,4]
## Patient.Gender      1.0000 1.000000 1.00000 1.00000
## Patient.Ethnicity    1.0000 1.000000 1.00000 1.00000
## History.Smoking.PackYears 1.0000 1.000000 1.00000 1.00000
## History.Diagnosis.AgeAtYears 1.0000 1.000000 1.00000 1.00000
## Path.HistoType.Subtype 1.0000 1.000000 1.00000 1.00000
## Path.TumourSizeMm     1.0000 1.000000 1.00000 1.00000
## Path.Invasion.PN      1.0000 1.000000 1.00000 1.00000
## Path.Invasion.VS      1.0000 1.000000 1.00000 1.00000
## Staging.pM            1.0000 1.000000 1.00000 1.00000
## Staging.pN            1.0000 1.000000 1.00000 1.00000
## Staging.pT            1.0000 1.000000 1.00000 1.00000
## Staging.Stage         1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Peritoneum 1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.PancRemnant 1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.PancBed    1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Other      1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Omentum    1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Mesentery  1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.LymphNodes 1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Lung       1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Liver      1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Brain      1.0000 1.000000 1.00000 1.00000
## History.Recurrence.Site.Bone       1.0000 1.000000 1.00000 1.00000
## Path.Grade.Coarse      0.6438 0.571128 1.00000 0.01907
## Path.TumourLocation.Coarse 1.0000 1.000000 1.00000 1.00000
## purity                  1.0000 0.001557 0.00143 1.00000
```

```
boxplot(coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse)
```



```
# plot(coef(nmf.final)[1,] ~ samps$purity_qpure) plot(coef(nmf.final)[2,] ~
# samps$purity_qpure) plot(coef(nmf.final)[3,] ~ samps$purity_qpure)
# plot(coef(nmf.final)[4,] ~ samps$purity_qpure) plot(coef(nmf.final)[5,] ~
# samps$purity_qpure) plot(coef(nmf.final)[6,] ~ samps$purity_qpure)

apply(coef(nmf.final), 1, function(c1) cor.test(c1, samps$purity_qpure, method = "kendall"))

## [[1]]
##
## Kendall's rank correlation tau
##
## data: c1 and samps$purity_qpure
## z = 2.503, p-value = 0.01233
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## 0.1934
```

```

##
##
## [[2]]
##
## Kendall's rank correlation tau
##
## data: c1 and samps$purity_qpure
## z = 3.915, p-value = 9.044e-05
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.3031
##
##
## [[3]]
##
## Kendall's rank correlation tau
##
## data: c1 and samps$purity_qpure
## z = 0.2677, p-value = 0.789
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.02075
##
##
## [[4]]
##
## Kendall's rank correlation tau
##
## data: c1 and samps$purity_qpure
## z = -0.7204, p-value = 0.4713
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.0572
##
##
## [[5]]
##
## Kendall's rank correlation tau
##
## data: c1 and samps$purity_qpure
## z = -3.714, p-value = 0.0002043
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.2882
##
##
## [[6]]
##
## Kendall's rank correlation tau
##

```

```

## data: c1 and samps$spurity_qpure
## z = -2.15, p-value = 0.03156
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1686

p.adjust(apply(coef(nmf.final), 1, function(c1) cor.test(c1, samps$spurity_qpure,
  method = "kendall")$p.value), "BY")

## [1] 0.060403 0.001329 1.000000 1.000000 0.001501 0.115993

# termplot(lm(coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse), se =
# TRUE, rug = TRUE)
lm(coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse)

##
## Call:
## lm(formula = coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse)
##
## Coefficients:
##              (Intercept)  cpvs.diag_dsd$Path.Grade.Coarse.L
##                   0.1353                      0.0623

summary(lm(coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse))

##
## Call:
## lm(formula = coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.1793 -0.0878 -0.0365  0.0655  0.2979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         0.1353     0.0114   11.89 < 2e-16
## cpvs.diag_dsd$Path.Grade.Coarse.L  0.0623     0.0161    3.87 0.00019
##
## Residual standard error: 0.107 on 108 degrees of freedom
## Multiple R-squared:  0.122, Adjusted R-squared:  0.114
## F-statistic: 15 on 1 and 108 DF, p-value: 0.000187

anova(lm(coef(nmf.final)[6, ] ~ cpvs.diag_dsd$Path.Grade.Coarse))

## Analysis of Variance Table
##
## Response: coef(nmf.final)[6, ]
##              Df Sum Sq Mean Sq F value Pr(>F)
## cpvs.diag_dsd$Path.Grade.Coarse  1  0.173  0.1727     15 0.00019
## Residuals                108  1.245  0.0115

```

```

temp.sig_id = colnames(nmf.final.msigdb.corr)
temp.sig_class = gsub("\\.*", "", temp.sig_id)

```

```

temp.nsig = length(temp.sig_id)
temp.nmeta = nrow(nmf.final.msigdb.corr)
tables = lapply(1:temp.nmeta, function(metagene_i) {
  tapply(1:temp.nsig, temp.sig_class, function(sig_class_is) {
    all_cors = nmf.final.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 associated with
            # worse prognosis
            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
## 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
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70

```
## 71
## 72
## 73
## 74
## 75
## 76
## 77
## 78
## 79 c2.REACTOME_CYCLIN_E_ASSOCIATED_EVENTS_DURING_G1_S_TRANSITION_/c2.REACTOME_REGULATION_OF_MITOTIC_
## 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
##      Correlation Metagenes
## 1      0.7237      1
## 2      0.7139      1
## 3      0.7106      1
## 4      0.7103      1
## 5      0.7019      1
## 6      0.7002      1
## 7      0.6945      1
## 8      0.6858      1
## 9      0.6828      1
## 10     0.6825      1
## 11     0.6818      1
## 12     0.6808      1
## 13     0.6784      1
## 14     0.6748      1,-5
## 15     0.6691      1
## 16     0.6654      1
## 17     0.6597      1
## 18     0.6567      1
## 19     0.6537      1

```

## 20	0.6523	1
## 21	0.6500	1
## 22	0.6476	1
## 23	0.6473	1
## 24	0.6436	1,-5
## 25	0.6399	1,-5
## 26	0.6366	1
## 27	0.6359	1
## 28	0.6309	1
## 29	0.6188	1
## 30	0.6175	1
## 31	0.6172	1
## 32	0.6158	1
## 33	0.6158	1
## 34	0.6128	1
## 35	0.6128	1
## 36	0.6121	1
## 37	0.6115	1
## 38	0.6101	1
## 39	0.6095	1
## 40	0.6088	1
## 41	0.6088	1
## 42	0.6081	1
## 43	0.6068	1
## 44	0.6065	1
## 45	0.6048	1
## 46	0.6014	1
## 47	0.5981	1
## 48	0.5967	1
## 49	0.5964	1
## 50	0.5957	1
## 51	0.5941	1
## 52	0.5937	1
## 53	0.5937	1
## 54	0.5914	1
## 55	0.5904	1
## 56	0.5887	1
## 57	0.5854	1
## 58	0.5850	1
## 59	0.5820	1
## 60	0.5817	1
## 61	0.5813	1
## 62	0.5807	1
## 63	0.5800	1
## 64	0.5770	1
## 65	0.5766	1
## 66	0.5753	1,-5
## 67	0.5753	1
## 68	0.5746	1
## 69	0.5743	1,-5
## 70	0.5720	1
## 71	0.5713	1
## 72	0.5706	1
## 73	0.5679	1

## 74	0.5673	1
## 75	0.5666	1
## 76	0.5663	1
## 77	0.5612	1,-5
## 78	0.5612	1
## 79	0.5606	1
## 80	0.5602	1
## 81	0.5582	1
## 82	0.5552	1,-5
## 83	0.5552	1
## 84	0.5545	1
## 85	0.5529	1
## 86	0.5502	1
## 87	0.5489	1
## 88	0.5472	1
## 89	0.5468	1
## 90	0.5455	1
## 91	0.5425	1
## 92	0.5408	1
## 93	0.5385	1
## 94	0.5381	1
## 95	0.5375	1
## 96	0.5361	1
## 97	0.5348	1
## 98	0.5338	1
## 99	0.5318	1,-5
## 100	0.5311	1
## 101	0.5301	1
## 102	0.5291	1,-5
## 103	0.5281	1
## 104	0.5274	1
## 105	0.5271	1
## 106	0.5241	1,-5
## 107	0.5224	1
## 108	0.5221	1
## 109	0.5217	1
## 110	0.5187	1
## 111	0.5184	1
## 112	0.5177	1,-5
## 113	0.5174	1
## 114	0.5170	1
## 115	0.5167	1
## 116	0.5160	1
## 117	0.5150	1
## 118	0.5120	1
## 119	0.5113	1,-5
## 120	0.5113	1,-5
## 121	0.5113	1
## 122	0.5110	1
## 123	0.5100	1
## 124	0.5097	1
## 125	0.5090	1
## 126	0.5083	1
## 127	-0.5053	-1

```

## 128      -0.5080      -1
## 129      -0.5090      -1
## 130      -0.5100      -1
## 131      -0.5134      -1
## 132      -0.5134      -1
## 133      -0.5150     -1,5
## 134      -0.5187      -1
## 135      -0.5201      -1
## 136      -0.5274      -1
## 137      -0.5318      -1
## 138      -0.5351      -1
## 139      -0.5475      -1
## 140      -0.5495      -1
## 141      -0.5556      -1
## 142      -0.5582      -1
## 143      -0.5586      -1
## 144      -0.5699     -1,5
## 145      -0.5780      -1
## 146      -0.5877      -1
## 147      -0.5884      -1
## 148      -0.5954      -1
## 149      -0.5961      -1
## 150      -0.5994      -1
## 151      -0.6105      -1
## 152      -0.6128      -1
## 153      -0.6138      -1
## 154      -0.6165      -1
## 155      -0.6570      -1
## 156      -0.6594      -1
## 157      -0.6979      -1
##
## [[1]]$c3
##
##           GeneSet Correlation Metagenes
## 1           c3.V$E2F_Q4_01      0.5401      1
## 2           c3.V$E2F_Q6_01      0.5284      1
## 3 c3.V$E2F_Q3_01/c3.V$E2F1_Q4_01  0.5167      1
## 4           c3.SGCGSSAAA_V$E2F1DP2_01  0.5093      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.MORF_BUB1B
## 3                                     c4.MODULE_403
## 4                                     c4.MORF_FEN1
## 5 c4.MODULE_125/c4.MODULE_158
## 6                                     c4.MODULE_17
## 7                                     c4.MODULE_320
## 8                                     c4.MODULE_126
## 9                                     c4.MORF_ESPL1
## 10                                    c4.MODULE_315
## 11                                    c4.MODULE_124
## 12                                    c4.MODULE_244
## 13                                    c4.GNF2_MSH2
## 14 c4.MODULE_98/c4.MODULE_198/c4.MODULE_252

```

```

## 15 c4.GNF2_MCM5
## 16 c4.MODULE_451
## 17 c4.MORF_BUB1
## 18 c4.MODULE_278
## 19 c4.MORF_CCNF
## 20 c4.GNF2_PA2G4/c4.GNF2_RAN
## 21 c4.MORF_RFC4/c4.MORF_RRM1
## 22 c4.GNF2_MSH6
## 23 c4.MORF_UNG
## 24 c4.MORF_DNMT1
## 25 c4.MORF_BUB3/c4.MORF_RAD23A
## 26 c4.MORF_PCNA
## 27 c4.MODULE_337
## 28 c4.MODULE_8
## Correlation Metagenes
## 1 0.7032 1
## 2 0.6517 1
## 3 0.6245 1
## 4 0.6239 1
## 5 0.6212 1
## 6 0.6175 1
## 7 0.6078 1
## 8 0.6061 1
## 9 0.6048 1
## 10 0.5998 1
## 11 0.5904 1
## 12 0.5904 1
## 13 0.5820 1
## 14 0.5787 1
## 15 0.5713 1
## 16 0.5643 1
## 17 0.5602 1
## 18 0.5545 1
## 19 0.5425 1
## 20 0.5348 1
## 21 0.5278 1
## 22 0.5244 1
## 23 0.5154 1
## 24 0.5117 1
## 25 0.5093 1
## 26 0.5063 1
## 27 0.5030 1,-5
## 28 0.5006 1
##
## [[1]]$c5
##
## GeneSet
## 1 c5.M_PHASE/c5.MITOSIS/c5.M_PHASE_OF_MITOTIC_CELL_CYCLE
## 2 c5.CELL_CYCLE_PROCESS/c5.MITOTIC_CELL_CYCLE/c5.CELL_CYCLE_PHASE
## 3 c5.REGULATION_OF_MITOSIS
## 4 c5.SPINDLE
## 5 c5.CELL_CYCLE_GO_0007049
## 6 c5.CELL_CYCLE_CHECKPOINT_GO_0000075
## 7 c5.MITOTIC_SPINDLE_ORGANIZATION_AND_BIOGENESIS/c5.SPINDLE_ORGANIZATION_AND_BIOGENESIS
## 8 c5.MITOTIC_CELL_CYCLE_CHECKPOINT

```

```

## 9 c5.SPINDLE_POLE
## 10 c5.CHROMOSOMAL_PART/c5.CHROMOSOME
## 11 c5.DNA_METABOLIC_PROCESS
## 12 c5.SPINDLE_MICROTUBULE
## 13 c5.REGULATION_OF_CELL_CYCLE
## 14 c5.ORGANELLE_PART/c5.INTRACELLULAR_ORGANELLE_PART
## 15 c5.CHROMOSOMEPERICENTRIC_REGION/c5.KINETOCHORE
## 16 c5.MICROTUBULE_CYTOSKELETON_ORGANIZATION_AND_BIOGENESIS
## 17 c5.DNA_REPLICATION
## 18 c5.MITOTIC_SISTER_CHROMATID_SEGREGATION/c5.SISTER_CHROMATID_SEGREGATION
## 19 c5.INTERPHASE/c5.INTERPHASE_OF_MITOTIC_CELL_CYCLE
## 20 c5.DNA_POLYMERASE_ACTIVITY/c5.DNA_DIRECTED_DNA_POLYMERASE_ACTIVITY
## 21 c5.RESPONSE_TO_ENDOGENOUS_STIMULUS/c5.RESPONSE_TO_DNA_DAMAGE_STIMULUS
## 22 c5.CHROMOSOME_SEGREGATION
## Correlation Metagenes
## 1 0.6808 1
## 2 0.6798 1
## 3 0.6543 1
## 4 0.6496 1
## 5 0.6493 1
## 6 0.6316 1
## 7 0.6195 1
## 8 0.6162 1
## 9 0.5827 1
## 10 0.5740 1
## 11 0.5706 1
## 12 0.5612 1
## 13 0.5458 1
## 14 0.5432 1
## 15 0.5301 1
## 16 0.5284 1
## 17 0.5281 1
## 18 0.5207 1
## 19 0.5201 1
## 20 0.5157 1
## 21 0.5137 1
## 22 0.5124 1
##
## [[1]]$c6
## GeneSet Correlation Metagenes
## 1 c6.CSR_LATE_UP.V1_SIGNED 0.5612 1
## 2 c6.E2F1_UP.V1_SIGNED 0.5274 1
##
## [[1]]$c7
##
## 1
## 2 c7.GSE15750_DAY6_VS_I
## 3
## 4
## 5 c7.GSE24634_TEFF_VS_TCONV
## 6
## 7
## 8
## 9

```



```

## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27 c7.GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_CD8_TCELL_SIGNED/c7.GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_IL1
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38 c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_T
## 39
## 40 c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_T
## Correlation Metagenes
## 1 0.6664 1
## 2 0.6131 1
## 3 0.5847 1
## 4 0.5710 1
## 5 0.5468 1
## 6 0.5224 1
## 7 0.5217 1,-5
## 8 0.5204 1
## 9 -0.5057 -1
## 10 -0.5060 -1
## 11 -0.5080 -1
## 12 -0.5083 -1
## 13 -0.5097 -1
## 14 -0.5097 -1
## 15 -0.5130 -1
## 16 -0.5134 -1
## 17 -0.5147 -1
## 18 -0.5170 -1
## 19 -0.5234 -1
## 20 -0.5254 -1
## 21 -0.5314 -1,5
## 22 -0.5358 -1

```

```

## 23      -0.5361      -1,5
## 24      -0.5391      -1,5
## 25      -0.5415      -1
## 26      -0.5445      -1
## 27      -0.5458      -1
## 28      -0.5489      -1
## 29      -0.5505      -1
## 30      -0.5509      -1
## 31      -0.5519      -1
## 32      -0.5582      -1,5
## 33      -0.5639      -1
## 34      -0.5649      -1
## 35      -0.5656      -1
## 36      -0.5696      -1
## 37      -0.5854      -1
## 38      -0.5900      -1,5
## 39      -0.5964      -1,5
## 40      -0.6051      -1,5
##
##
## [[2]]
## [[2]]$c1
## data frame with 0 columns and 0 rows
##
## [[2]]$c2
##
##                                     GeneSet Correlation
## 1  c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_SIGNED      0.5520
## 2                                c2.LIU_PROSTATE_CANCER_SIGNED      0.5175
## 3  c2.WANG_BARRETTS_ESOPHAGUS_AND_ESOPHAGUS_CANCER_SIGNED      0.5105
## 4                                c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABINE_1A -0.5008
## 5                                c2.PID_INTEGRIN1_PATHWAY -0.5014
## 6                                c2.SERVITJA_ISLET_HNF1A_TARGETS_SIGNED -0.5021
## 7  c2.SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_SIGNED -0.5074
## 8                                c2.ROY_WOUND_BLOOD_VESSEL_SIGNED -0.5332
## 9  c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SIGNED -0.5352
## 10                                c2.KARAKAS_TGFB1_SIGNALING -0.5493
## 11                                c2.HUANG_DASATINIB_RESISTANCE_SIGNED -0.5523
## 12                                c2.KAN_RESPONSE_TO_ARSENIC_TRIOXIDE -0.5546
## 13  c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SIGNED -0.6035
##      Metagenes
## 1           2
## 2           2
## 3           2
## 4        -2,6
## 5        -2,6
## 6          -2
## 7          -2
## 8          -2
## 9          -2
## 10         -2
## 11         -2
## 12         -2
## 13         -2
##

```

```

## [[2]]$c3
## data frame with 0 columns and 0 rows
##
## [[2]]$c4
##           GeneSet Correlation Metagenes
## 1 c4.MODULE_488      -0.5085      -2
##
## [[2]]$c5
## data frame with 0 columns and 0 rows
##
## [[2]]$c6
##           GeneSet Correlation Metagenes
## 1           c6.LEF1_UP.V1_SIGNED      -0.5259      -2
## 2 c6.CORDENONSI_YAP_CONSERVED_SIGNATURE      -0.5319      -2
##
## [[2]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[3]]
## [[3]]$c1
## data frame with 0 columns and 0 rows
##
## [[3]]$c2
##           GeneSet Correlation Metagenes
## 1 c2.ZHENG_BOUND_BY_FOXP3      -0.5047      -3
##
## [[3]]$c3
## data frame with 0 columns and 0 rows
##
## [[3]]$c4
## data frame with 0 columns and 0 rows
##
## [[3]]$c5
## data frame with 0 columns and 0 rows
##
## [[3]]$c6
## data frame with 0 columns and 0 rows
##
## [[3]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[4]]
## [[4]]$c1
## data frame with 0 columns and 0 rows
##
## [[4]]$c2
##           GeneSet
## 1 c2.MARTINEZ_RB1_AND_TP53_TARGETS_SIGNED/c2.MARTINEZ_TP53_TARGETS_SIGNED
## 2           c2.FLECHNER_PBL_KIDNEY_TRANSPLANT_OK_VS_DONOR_SIGNED
## 3           c2.DASU_IL6_SIGNALING_SCAR_SIGNED
## 4           c2.FAELT_B_CLL_WITH_VH3_21_SIGNED

```

```

## Correlation Metagenes
## 1 0.5081 4
## 2 -0.5020 -4
## 3 -0.5050 -4
## 4 -0.5146 -4
##
## [[4]]$c3
## GeneSet Correlation Metagenes
## 1 c3.GATGKMRCG_UNKNOWN -0.5321 -4
##
## [[4]]$c4
## GeneSet Correlation Metagenes
## 1 c4.MODULE_486 -0.5016 -4
##
## [[4]]$c5
## data frame with 0 columns and 0 rows
##
## [[4]]$c6
## data frame with 0 columns and 0 rows
##
## [[4]]$c7
## GeneSet Correlation
## 1 c7.GSE20366_CD103_KLRG1_DP_VS_DN_TREG_SIGNED -0.5057
## 2 c7.GSE1448_CTRL_VS_ANTI_VALPHA2_DP_THYMOCYTE_SIGNED -0.5109
## Metagenes
## 1 -4
## 2 -4
##
## [[5]]
## [[5]]$c1
## data frame with 0 columns and 0 rows
##
## [[5]]$c2
##
## 1 c2.REACTOME_G_ALPHA_S_SIGNALING
## 2 c2.ZHAN_MULTIPLE_MYELOMA_C
## 3 c2.GENTLES_LEUKEMIC_STEM_C
## 4 c2.SMID_BREAST_CANCER_NORMAL_L
## 5 c2.BROWNE_HCMV_INFECTION_2
## 6 c2.TERAMOTO_OPN_TARGETS_
## 7 c2.LE_NEURONAL_DIFFERENTIATI
## 8 c2.TARTE_PLASMA_CELL_VS_PLASMA
## 9 c2.KATSANOUE_ELAVL1_TARGE
## 10 c2.MIKKELSEN_MCV6_ICP_WITH
## 11 c2.BENPORATH_SUZ12_TARGETS/c2.BENPORATH_EED_TARGETS/c2.BENPORATH_ES_WITH_H3K27ME3/c2.BENPORATH_PRO
## 12 c2.ONDER_CD11_TARGETS
## 13 c2.AIGNER_ZEE
## 14 c2.WONG_ENDOMETRIUM_CANC
## 15 c2.SHEDDEN_LUNG_CANCER_POOR_SU
## 16 c2.TURASHVILI_BREAST_DUCTAL_CARCINOMA_VS_LOBULAR_NORM
## 17 c2.ZHAN_MULTIPLE_MYELOMA_
## 18 c2.CHANG_CORE_SERUM_RESPON
## 19 c2.LI_AMPLIFIED_IN_LU

```

## 20			c2.AKL_HTLV1_INFECTION
## 21			c2.FOURNIER_ACINAR_DEVELOPMENT
## 22			c2.CAIRO_HEPATOBLASTOMA_CLASSIFICATION
## 23			c2.NADERI_BREAST_CANCER_PROGNOSIS
## 24			c2.DELYS_THYROID_CANCER
## 25			c2.RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED
## 26			c2.BOYVAULT_LIVER_CANCER_SUBCLASSIFICATION
## 27			c2.BOQUEST_STEM_CELL_CULTURED_VS_FRESH
## 28			c2.LI_WILMS_TUMOR_ANAPLASTIC
## 29			c2.YU_MYC_TARGETS
## 30			c2.RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC
## 31			c2.REACTOME_METABOLISM_OF_NUCLEOTIDES
## 32			c2.MILICIC_FAMILIAL_ADENOMATOUS_POLYPOSIS
## 33			c2.GRADE_COLON_AND_RECTAL_CANCER
## 34			c2.BOYVAULT_LIVER_CANCER_SUBCLASSIFICATION
## 35			c2.CHIARADONNA_NEOPLASTIC_TRANSFORMATION_KRAS_CDKN2A
## 36			c2.GREENBAUM_E2A_TARGETS
## 37			c2.INAMURA_LUNG_CANCER_SIGNATURE
## 38			c2.VECCHI_GASTRIC_CANCER_EXPRESSION
## 39			c2.LEE_EARLY_T_LYMPHOCYTES
## 40			c2.SWEET_LUNG_CANCER_KINASE
## 41			c2.HOELZEL_NF1_TARGETS
## 42			c2.WINTER_HYPONATREMIA
## 43			c2.HAHTOLA_SEZARY_SYNDROME
## 44			c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PATHWAYS
## 45			c2.SABATES_COLORECTAL_ADENOCARCINOMA
## 46			c2.MARTORIATI_MDM4_TARGETS_NEUROEPITHELIAL
## 47			c2.WEST_ADRENOCORTICAL_TUMORS
## 48			c2.SHIPP_DLBCL_VS_FOLLICULAR_LYMPHOMA
##	Correlation	Metagenes	
## 1	0.5493	5	
## 2	0.5456	5	
## 3	0.5423	5	
## 4	0.5409	5	
## 5	0.5369	5	
## 6	0.5359	5	
## 7	0.5339	-1,5	
## 8	0.5262	-1,5	
## 9	0.5208	5	
## 10	0.5161	5	
## 11	0.5151	5	
## 12	0.5014	5	
## 13	-0.5004	-5	
## 14	-0.5014	-5	
## 15	-0.5084	1,-5	
## 16	-0.5088	-5	
## 17	-0.5111	1,-5	
## 18	-0.5151	1,-5	
## 19	-0.5158	-5	
## 20	-0.5168	-5	
## 21	-0.5171	1,-5	
## 22	-0.5181	-5	
## 23	-0.5198	-5	
## 24	-0.5215	-5	

```

## 25      -0.5225      1,-5
## 26      -0.5238      1,-5
## 27      -0.5252      -5
## 28      -0.5269      -5
## 29      -0.5292      -5
## 30      -0.5336      1,-5
## 31      -0.5376      -5
## 32      -0.5419      -5
## 33      -0.5500      1,-5
## 34      -0.5560      1,-5
## 35      -0.5564      -5
## 36      -0.5657      -5
## 37      -0.5694      -5
## 38      -0.5728      1,-5
## 39      -0.5734      1,-5
## 40      -0.5734      -5
## 41      -0.5838      -5
## 42      -0.5885      -5
## 43      -0.5905      -5
## 44      -0.5925      1,-5
## 45      -0.5959      -5
## 46      -0.6053      -5
## 47      -0.6140      1,-5
## 48      -0.6324      -5
##
## [[5]]$c3
##           GeneSet Correlation Metagenes
## 1      c3.V$STAT5A_01      0.5181      5
## 2      c3.V$ELK1_02      -0.5007      -5
## 3 c3.SCGGAAGY_V$ELK1_02      -0.5181      -5
##
## [[5]]$c4
##                                     GeneSet Correlation
## 1 c4.MODULE_11/c4.MODULE_66/c4.MODULE_100/c4.MODULE_137      0.5530
## 2                                     c4.MODULE_51      0.5403
## 3                                     c4.MODULE_19      0.5232
## 4                                     c4.MODULE_361      0.5091
## 5                                     c4.MODULE_200      0.5027
## 6                                     c4.MODULE_337      -0.5215
##
## Metagenes
## 1      5
## 2      5
## 3      5
## 4      5
## 5      5
## 6      1,-5
##
## [[5]]$c5
##                                     GeneSet
## 1 c5.3_5_CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY/c5.CYCLIC_NUCLEOTIDE_PHOSPHODIESTERASE_ACTIVITY
## 2                                     c5.PHOSPHORIC_DIESTER_HYDROLASE_ACTIVITY
##
## Correlation Metagenes
## 1      0.5165      5
## 2      0.5161      5

```

```

##
## [[5]]$c6
##                               GeneSet Correlation Metagenes
## 1 c6.SINGH_KRAS_DEPENDENCY_SIGNATURE_      -0.5051      -5
## 2                c6.AKT_UP.V1_SIGNED      -0.5218      -5
##
## [[5]]$c7
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14 c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_TCELL_YOUNG_SIGNED/c7.GSE36476_CTRL_VS_TSST_ACT_72H_ME
## 15
## 16 c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_TCELL_OLD_SIGNED/c7.GSE36476_CTRL_VS_TSST_ACT_72H_ME
## 17
## 18
## Correlation Metagenes
## 1      0.6033      5
## 2      0.5956      5
## 3      0.5503     -1,5
## 4      0.5383     -1,5
## 5      0.5339      5
## 6      0.5285      5
## 7      0.5282      5
## 8      0.5188      5
## 9      0.5185      5
## 10     0.5181     -1,5
## 11     0.5151      5
## 12     0.5141     -1,5
## 13     0.5098      5
## 14     0.5078     -1,5
## 15     0.5024     -1,5
## 16     0.5000     -1,5
## 17    -0.5175     1,-5
## 18    -0.5309     -5
##
##
## [[6]]
## [[6]]$c1
## data frame with 0 columns and 0 rows
##
## [[6]]$c2
##
## 1

```

## 2		c2.VERRECCHIA_EARLY_RESPONSE_TO_T
## 3		c2.PID_UPA_UPAR_PAT
## 4		c2.WIEDERSCHAIN_TARGETS_OF_BMI1_AND_P
## 5		c2.PID_SYNDECAN_1_PAT
## 6		c2.VERRECCHIA_RESPONSE_TO_TGFE
## 7		c2.PID_INTEGRIN3_PAT
## 8		c2.PID_INTEGRIN1_PAT
## 9		c2.VERRECCHIA_DELAYED_RESPONSE_TO_T
## 10	c2.REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION/c2.REACTOME_COLLAGEN_FORMA	
## 11		c2.BURTON_ADIPOGENES
## 12		c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABIN
## 13		c2.VERRECCHIA_RESPONSE_TO_TGFE
## 14		c2.KEGG_FOCAL_ADHE
## 15		c2.PID_AVB3_INTEGRIN_PAT
## 16		c2.KOINUMA_TARGETS_OF_SMAD2_OR_S
## 17		c2.SIMBULAN_UV_RESPONSE_IMMORTALIZE
## 18		c2.WU_CELL_MIGRA
## 19		c2.KEGG_ECM_RECEPTOR_INTERAC
## 20		c2.POTTI_TOPOTECAN_SENSITI
## 21		c2.YIH_RESPONSE_TO_ARSENIT
## 22	c2.FARMER_BREAST_CANCER_CLUSTER_5/c2.ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE/c4.GNF2_C	
## 23		c2.PHONG_TNF_RESPONSE_VIA_P38_PAT
## 24		c2.BERENJENO_TRANSFORMED_BY_RHOA_REVERSIBLY_S
## 25		c2.RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY_S
## 26		c2.REN_ALVEOLAR_RHABDOMYOSARCOMA_S
## 27		c2.PASINI_SUZ12_TARGETS_S
## 28		c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_S
##	Correlation	Metagenes
## 1	0.5952	6
## 2	0.5667	6
## 3	0.5606	6
## 4	0.5555	6
## 5	0.5521	6
## 6	0.5518	6
## 7	0.5504	6
## 8	0.5450	-2,6
## 9	0.5402	6
## 10	0.5318	6
## 11	0.5291	6
## 12	0.5257	-2,6
## 13	0.5253	6
## 14	0.5250	6
## 15	0.5185	6
## 16	0.5175	6
## 17	0.5148	6
## 18	0.5107	6
## 19	0.5087	6
## 20	0.5080	6
## 21	0.5067	6
## 22	0.5053	6
## 23	0.5013	6
## 24	-0.5141	-6
## 25	-0.5152	-6
## 26	-0.5165	-6


```
## 27      -0.5419      -6
## 28      -0.5887      -6
##
## [[6]]$c3
##           GeneSet Correlation Metagenes
## 1 c3.TGANTCA_V$AP1_C      0.5891      6
## 2      c3.V$AP1_Q4      0.5379      6
## 3      c3.V$AP1_Q6      0.5365      6
## 4      c3.V$AP1_Q6_01      0.5023      6
##
## [[6]]$c4
##           GeneSet Correlation Metagenes
## 1 c4.MODULE_321      0.5616      6
## 2 c4.MODULE_562      0.5301      6
## 3 c4.MODULE_153      0.5287      6
## 4 c4.GNF2_MMP1      0.5257      6
##
## [[6]]$c5
##           GeneSet Correlation Metagenes
## 1      c5.COLLAGEN      0.5138      6
## 2 c5.AXON_GUIDANCE      0.5101      6
##
## [[6]]$c6
## data frame with 0 columns and 0 rows
##
## [[6]]$c7
## data frame with 0 columns and 0 rows
```

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] stargazer_5.1      xtable_1.7-4      gplots_2.14.2
##  [4] RColorBrewer_1.0-5 glmnet_1.9-8      Matrix_1.1-4
##  [7] glmulti_1.0.7      rJava_0.9-6      nnls_1.4
## [10] NMF_0.20.5         synchronicity_1.1.4 bigmemory_4.4.6
## [13] BH_1.54.0-5        bigmemory.sri_0.1.3 Biobase_2.26.0
## [16] BiocGenerics_0.12.1 cluster_1.15.3    rngtools_1.2.4
## [19] pkgmaker_0.22      registry_0.2      energy_1.6.2
## [22] survival_2.37-7    knitr_1.8
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-6       boot_1.3-13       caTools_1.17.1
##  [4] codetools_0.2-9    colorspace_1.2-4  digest_0.6.4
##  [7] doParallel_1.0.8   evaluate_0.5.5    foreach_1.4.2
## [10] formatR_1.0        gdata_2.13.3      ggplot2_1.0.0
## [13] grid_3.1.1         gridBase_0.4-7    gtable_0.1.2
## [16] gtools_3.4.1       highr_0.4         iterators_1.0.7
## [19] KernSmooth_2.23-13 labeling_0.3       lattice_0.20-29
## [22] MASS_7.3-35        munsell_0.4.2     plyr_1.8.1
## [25] proto_0.3-10       Rcpp_0.11.3       reshape2_1.4
## [28] scales_0.2.4       stringr_0.6.2     tools_3.1.1
```