

# SIS NMF surg dsd

November 23, 2014

## 1 Preparation

```
##### LIBRARIES
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 63/64

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

setwd(data_dir)
load("image.rda")
```

## 2 Probe selection

```
table(cpss.sis$sel)

##
## FALSE  TRUE
## 12809   191

mean(cpss.sis$sel)

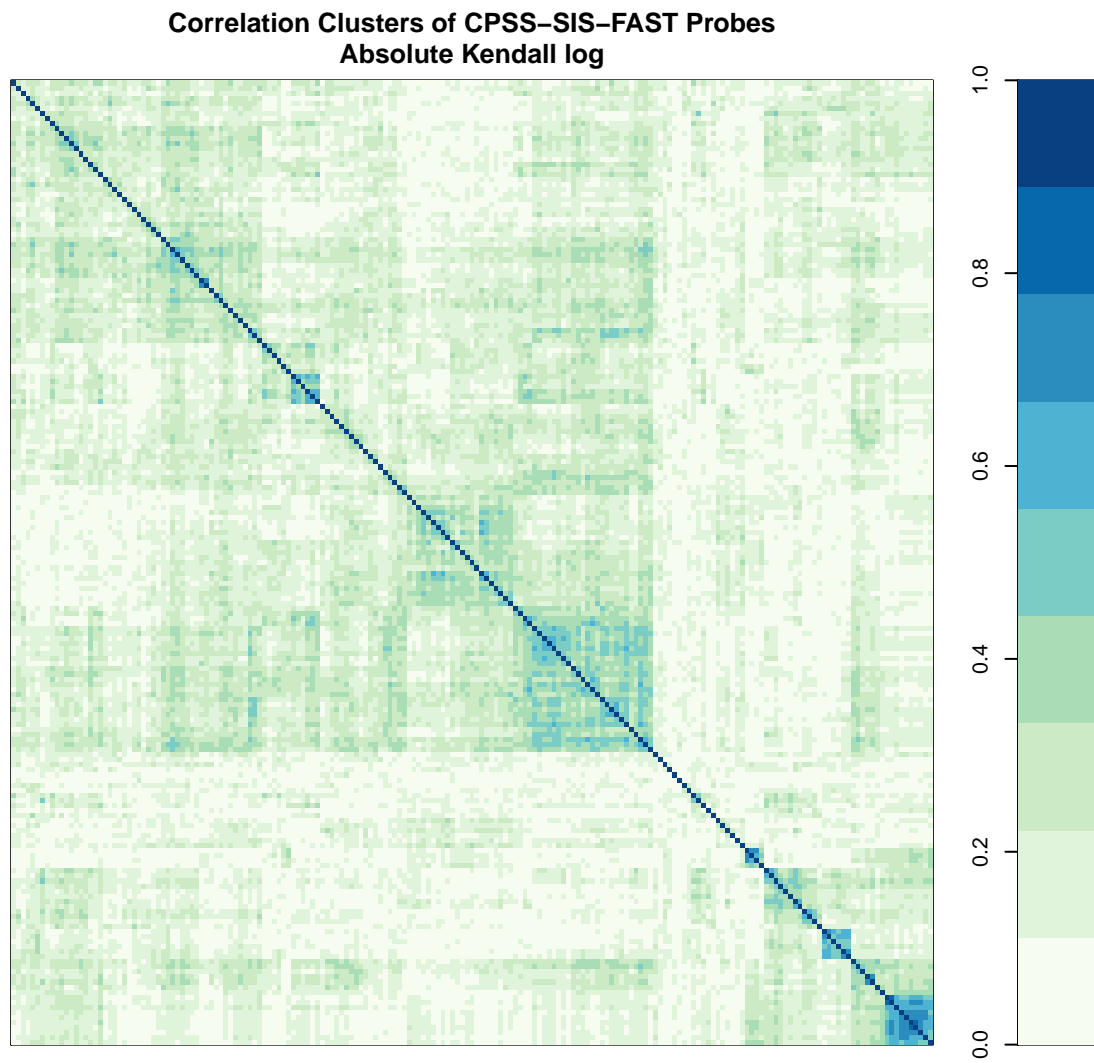
## [1] 0.01469
```

## 3 Expression correlation

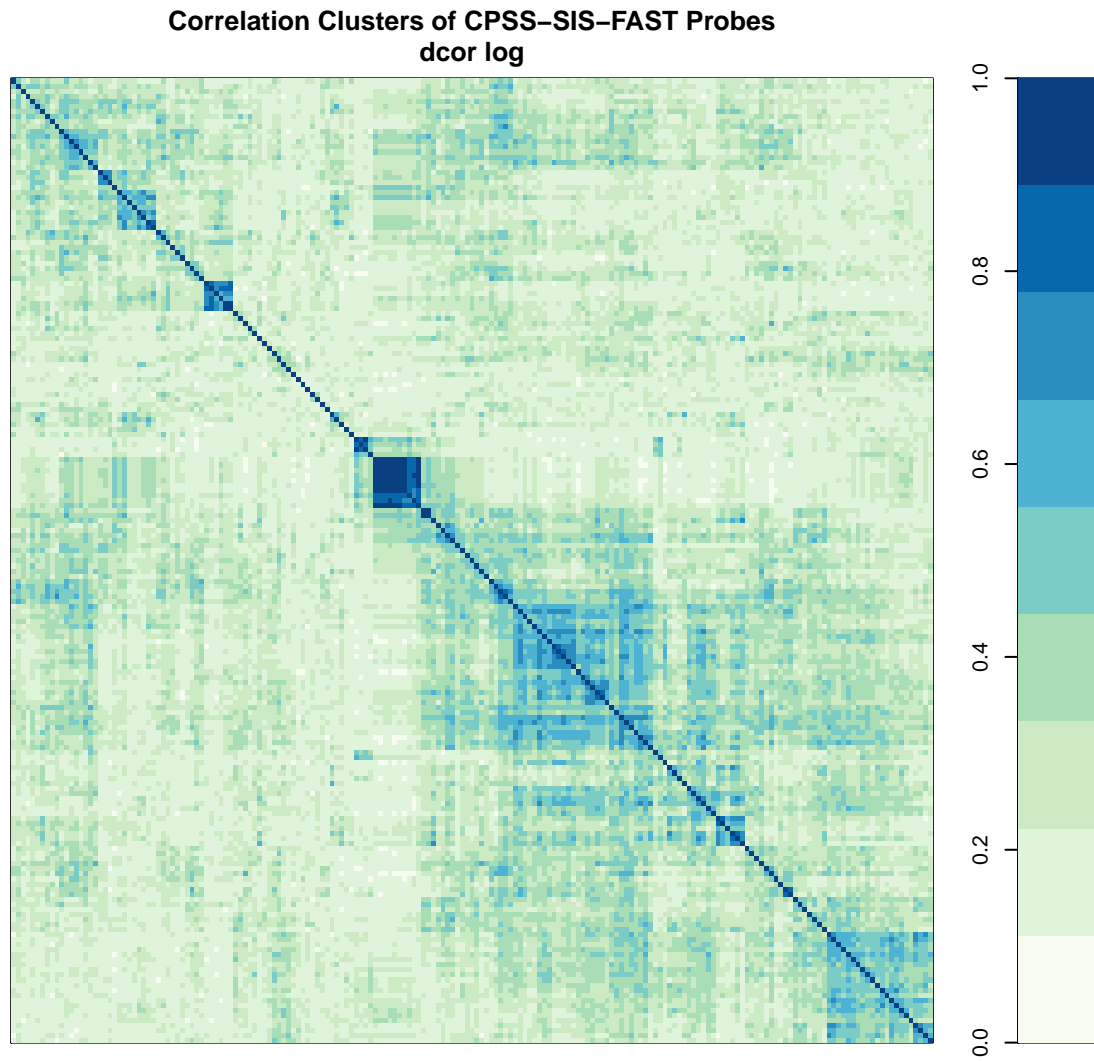
```
corPlot(x.sel.kcor, main = "Correlation Clusters of CPSS-SIS-FAST Probes\nKendall log",
        useRaster = FALSE)
```



```
corPlot(abs(x.sel.kcor), zlim = c(0, 1), pal = "GnBu", main = "Correlation Clusters of CPSS-SIS-FAST Probes",
        useRaster = FALSE)
```



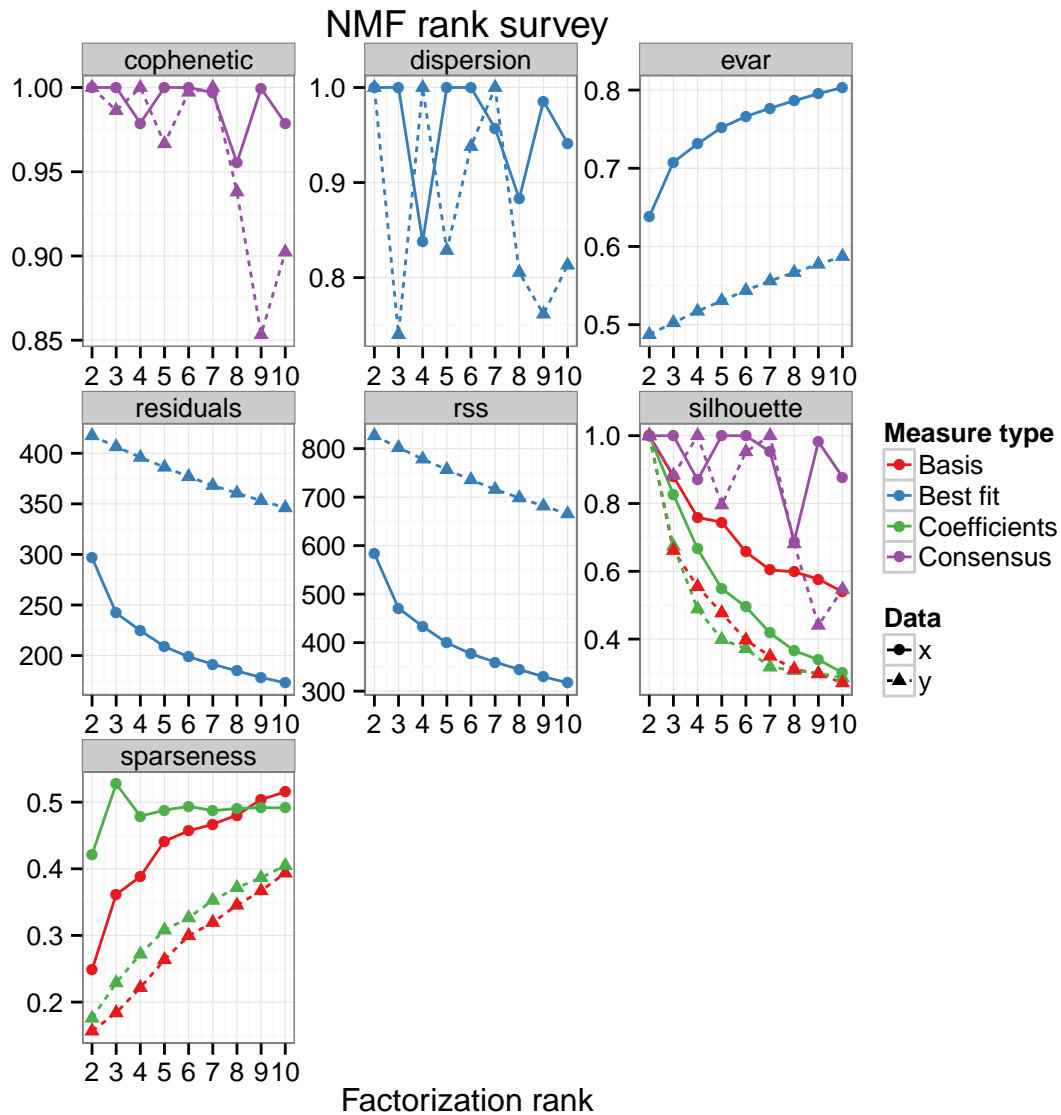
```
corPlot(x.sel.dcor, zlim = c(0, 1), pal = "GnBu", main = "Correlation Clusters of CPSS-SIS-FAST Probes",
        useRaster = FALSE)
```



## 4 Factorization

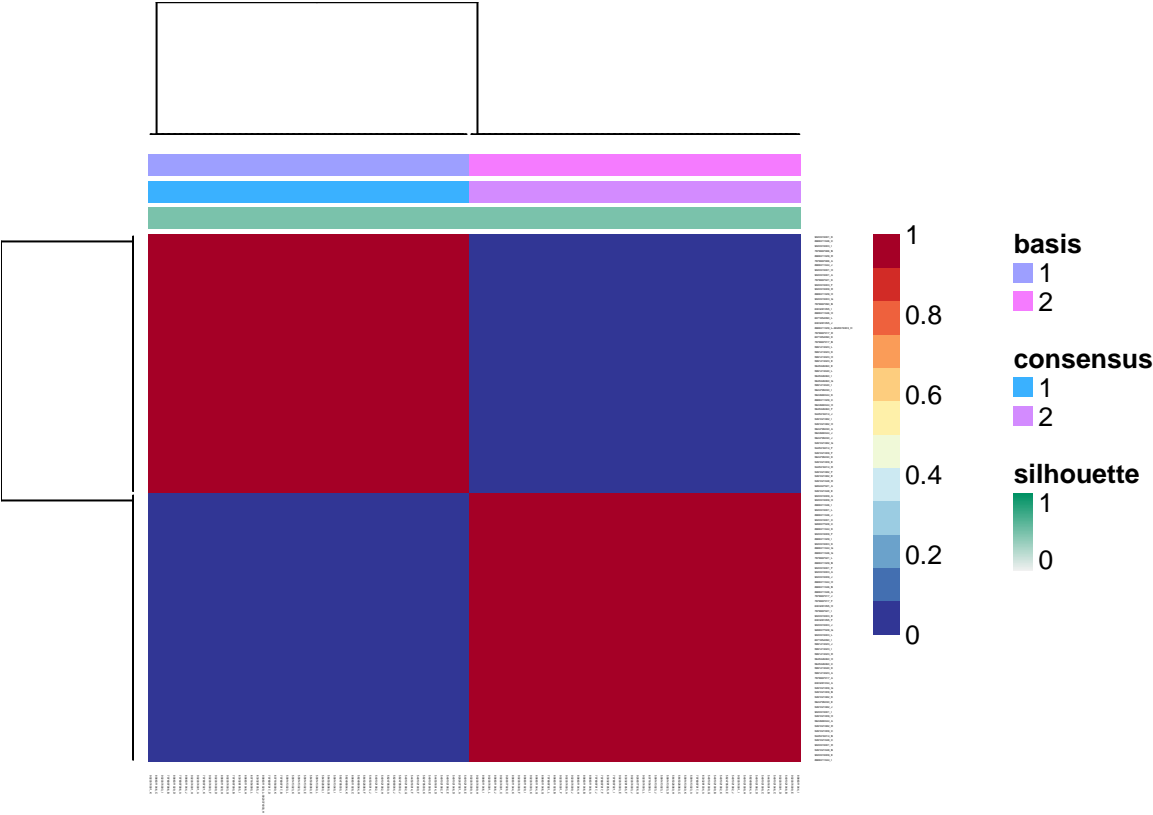
### 4.1 Rank estimation

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

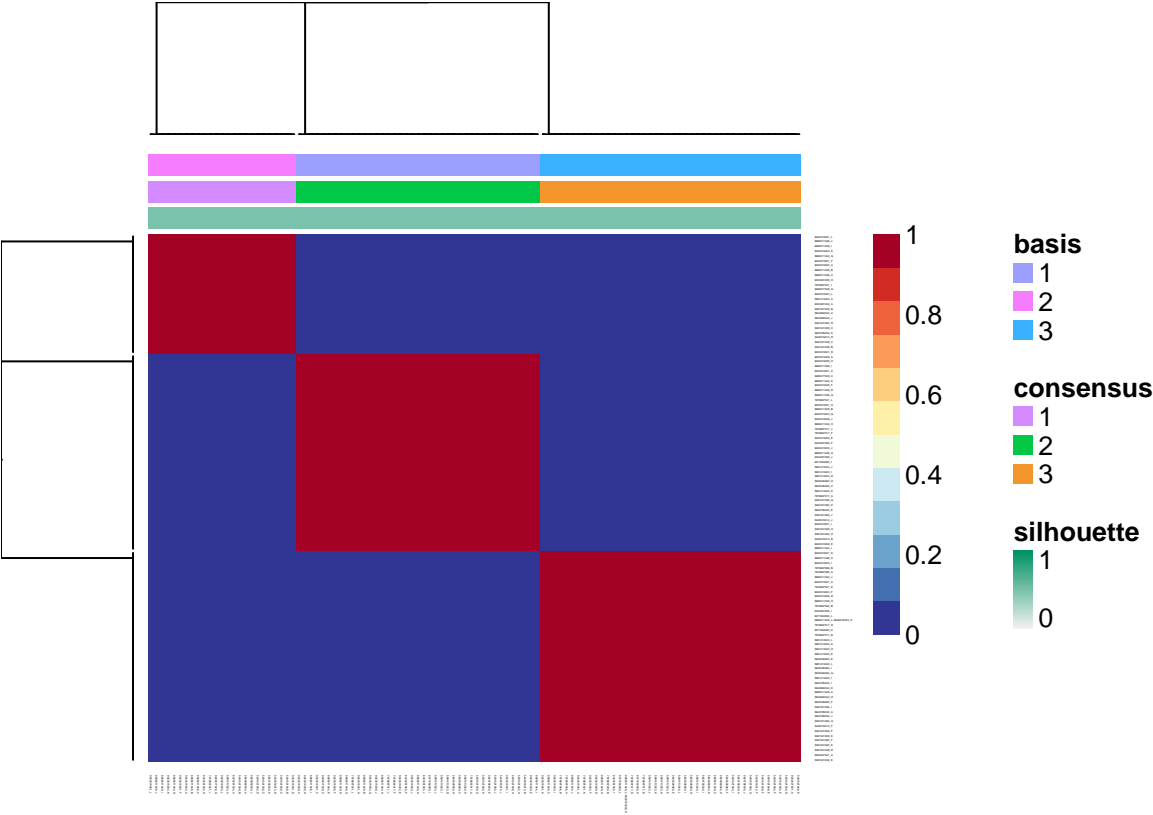


```
for (i in 1:length(temp.nmf.rank$fit)) {
  consensusmap(temp.nmf.rank$fit[[i]])
}
```

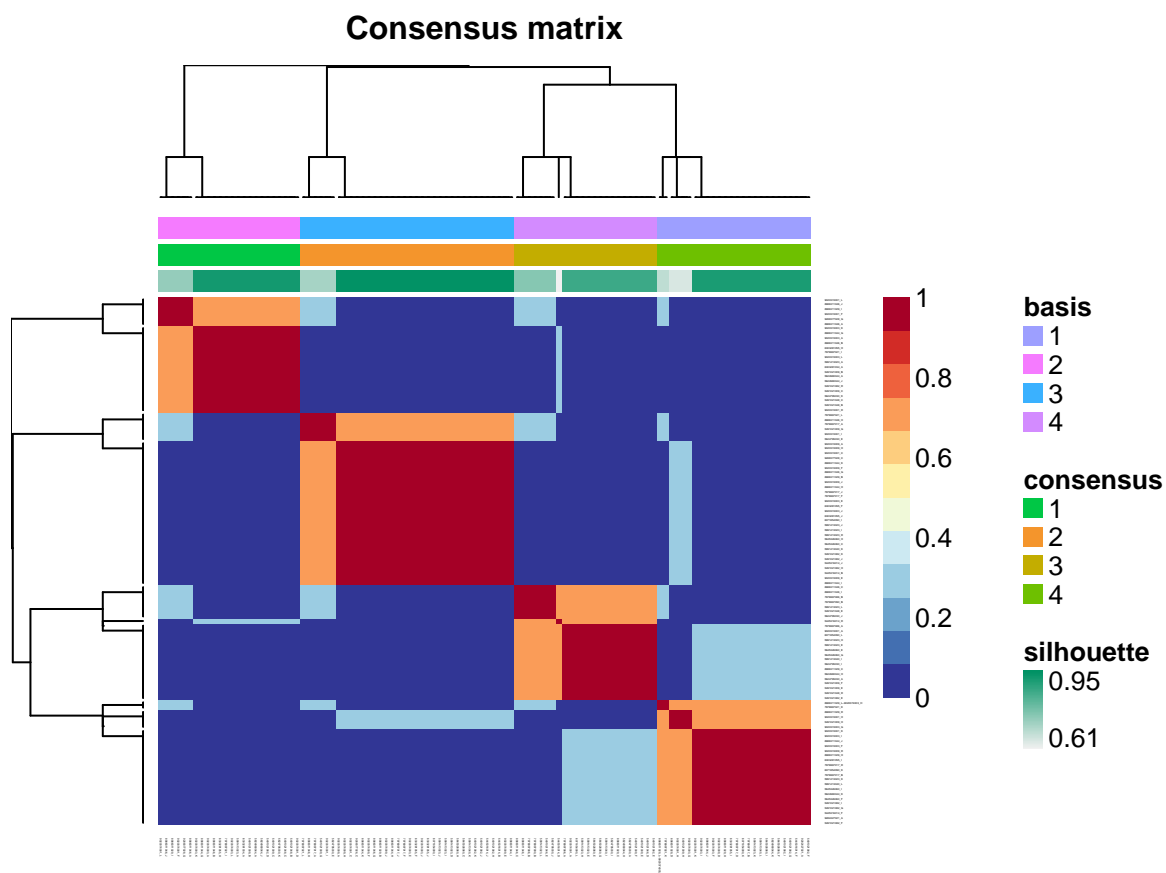
Consensus matrix

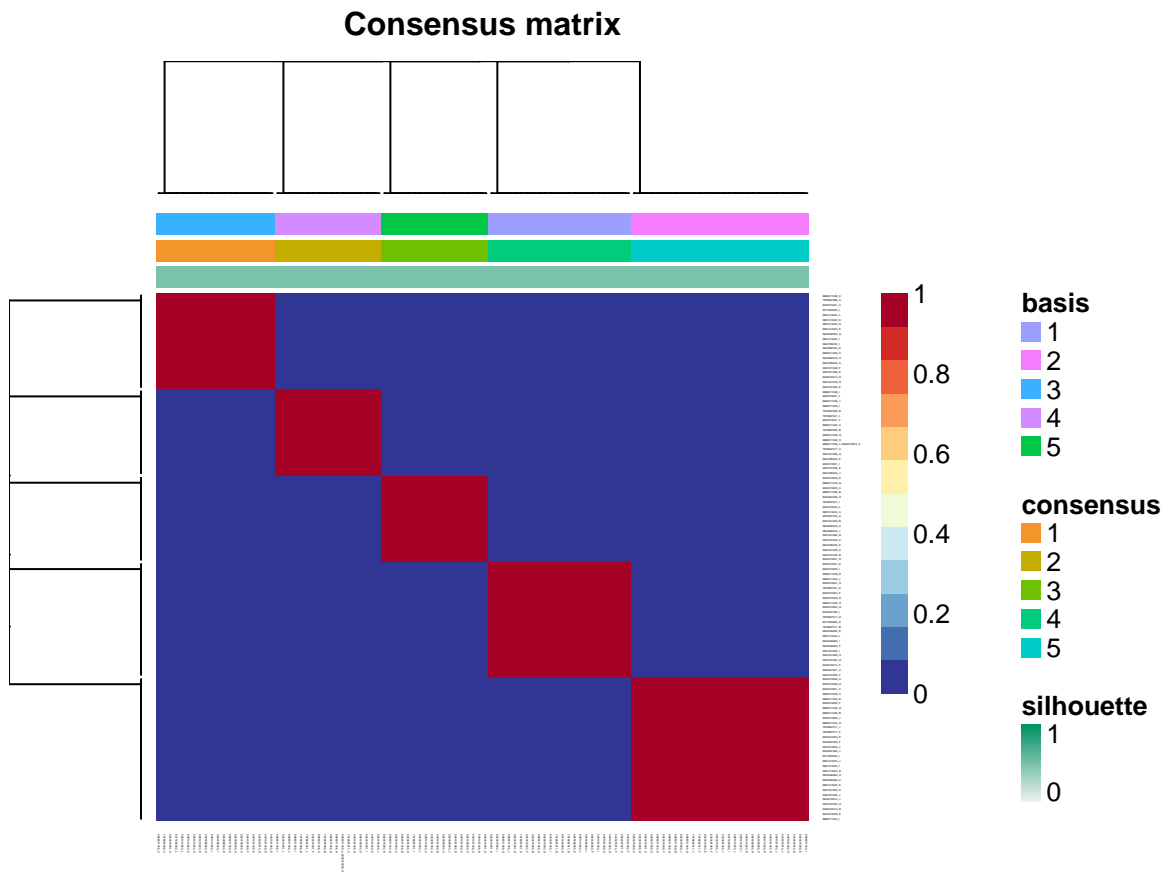


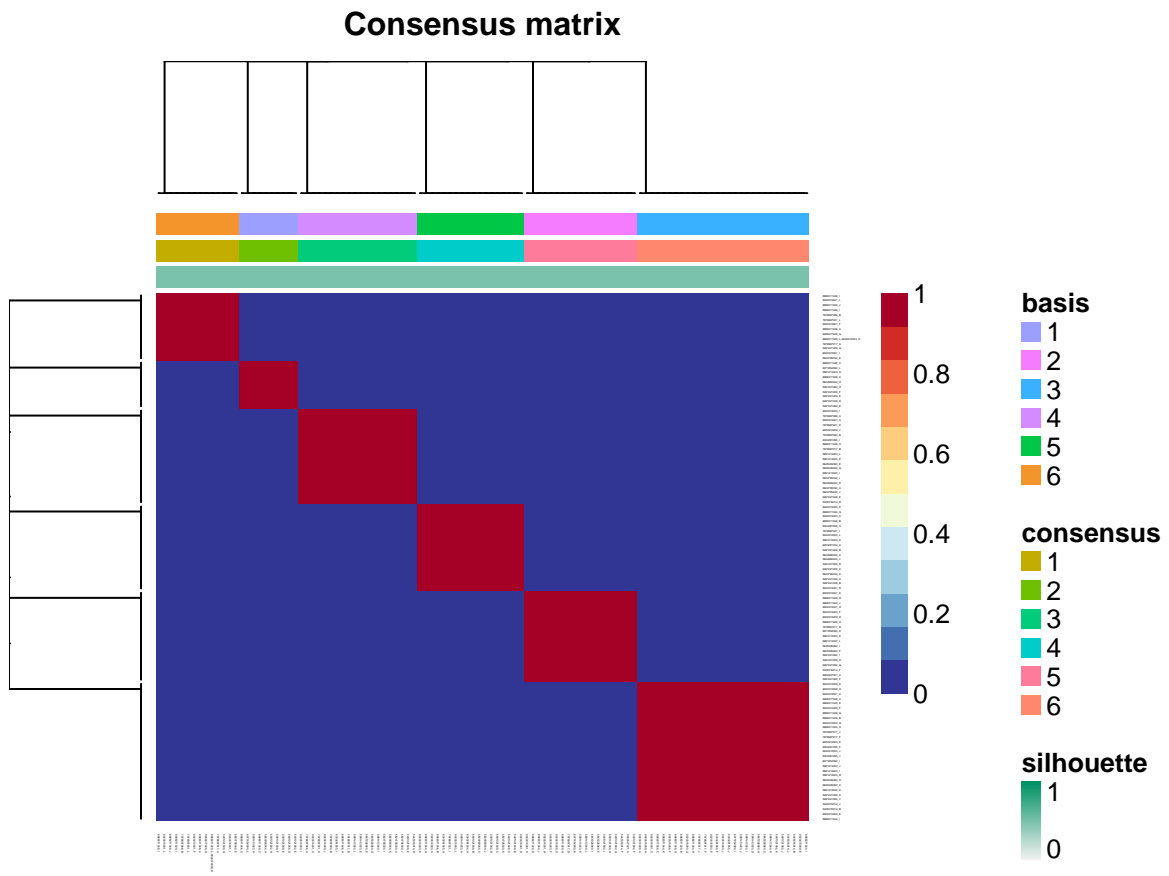
Consensus matrix

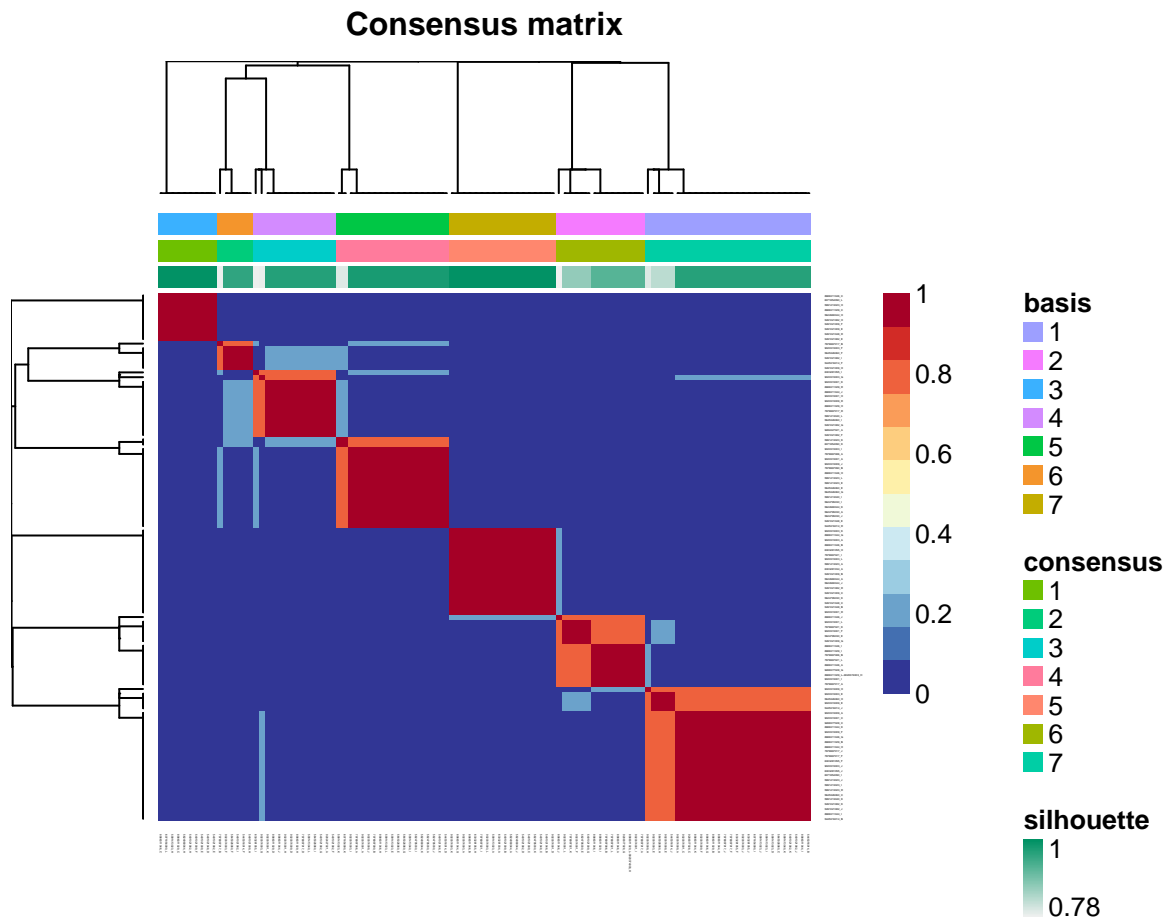


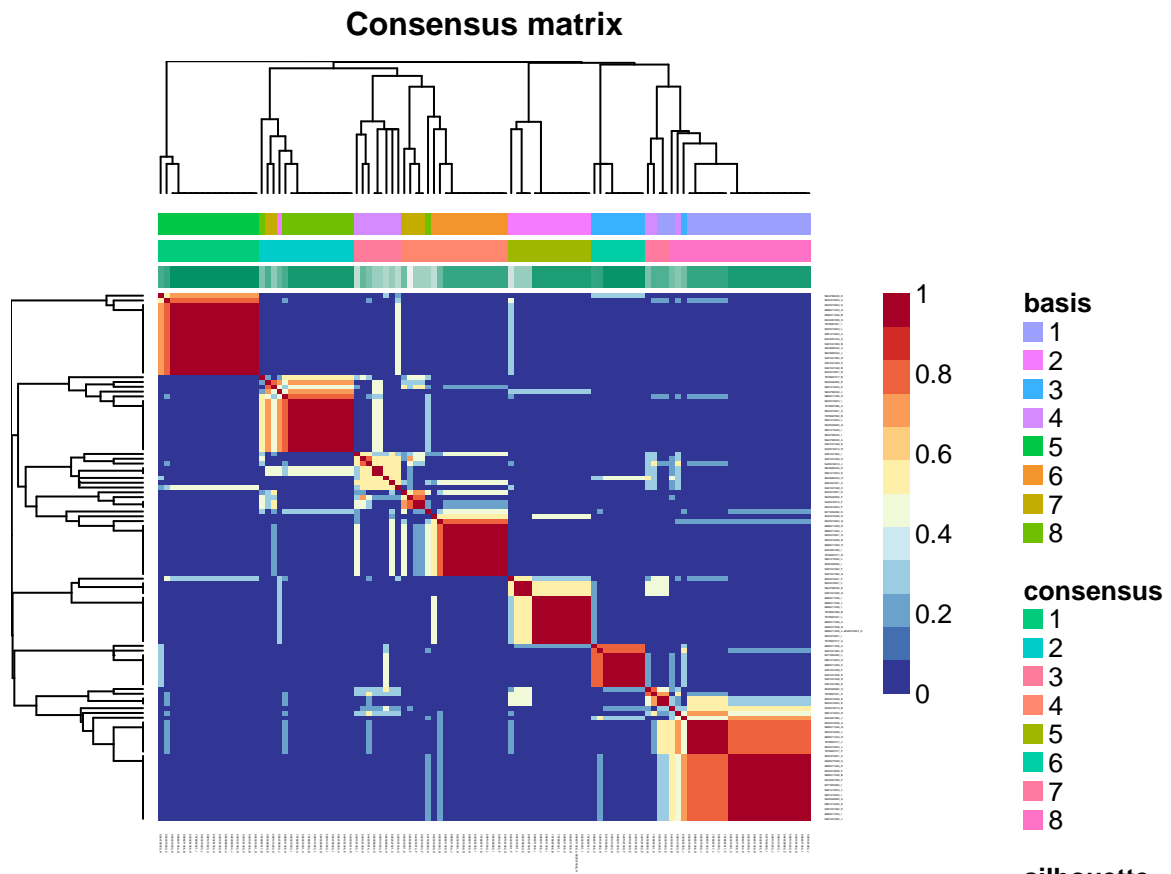


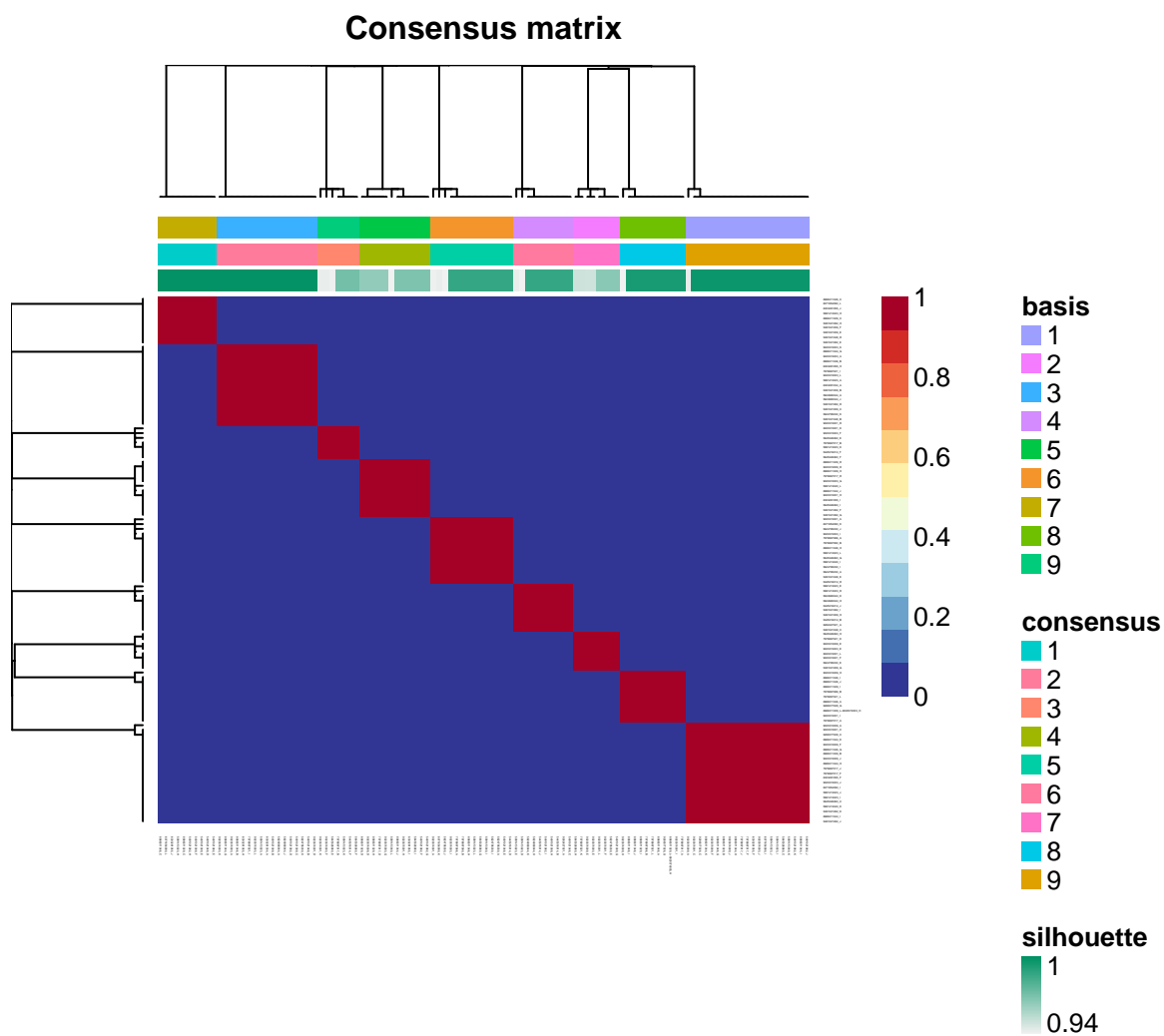


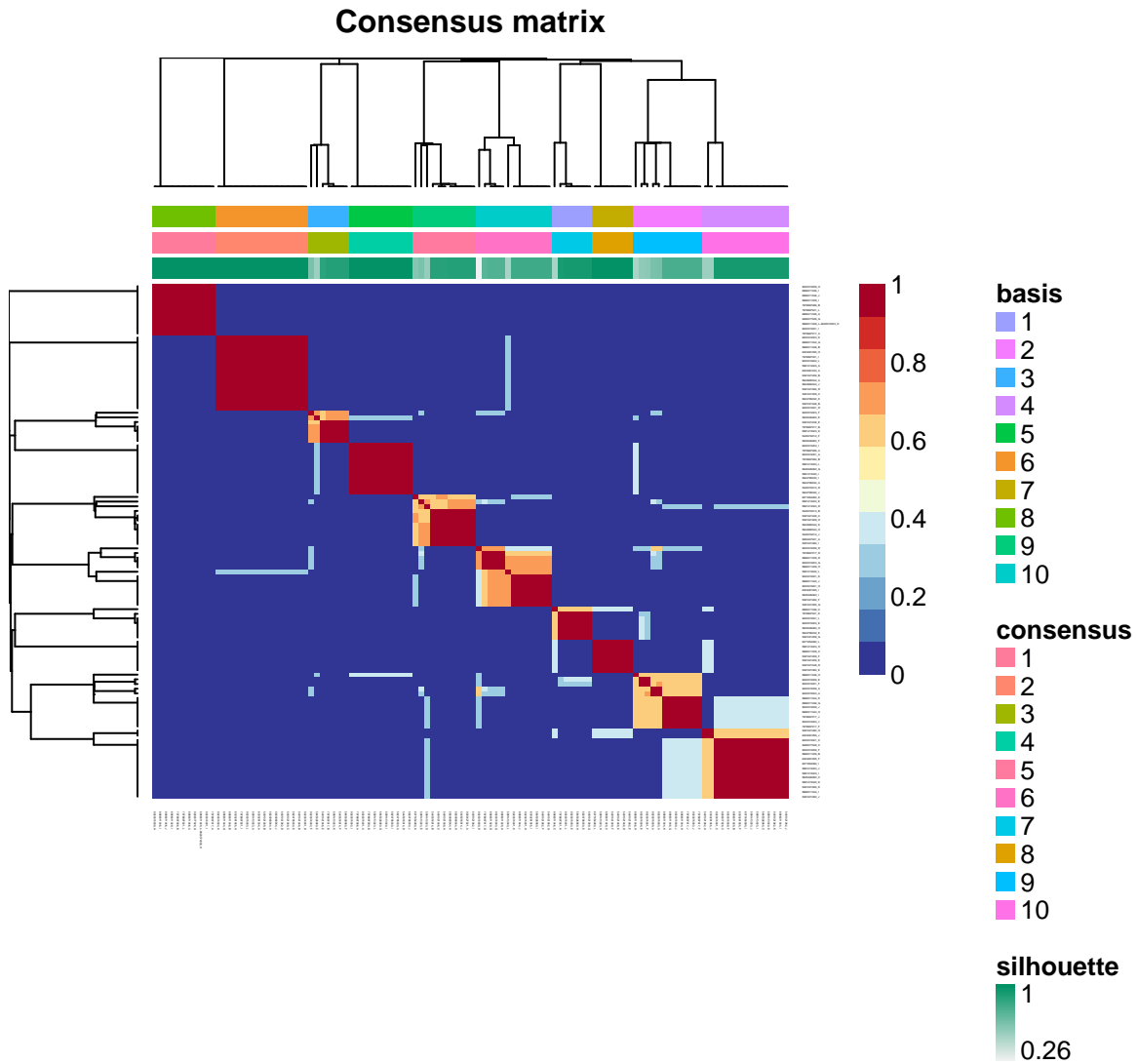








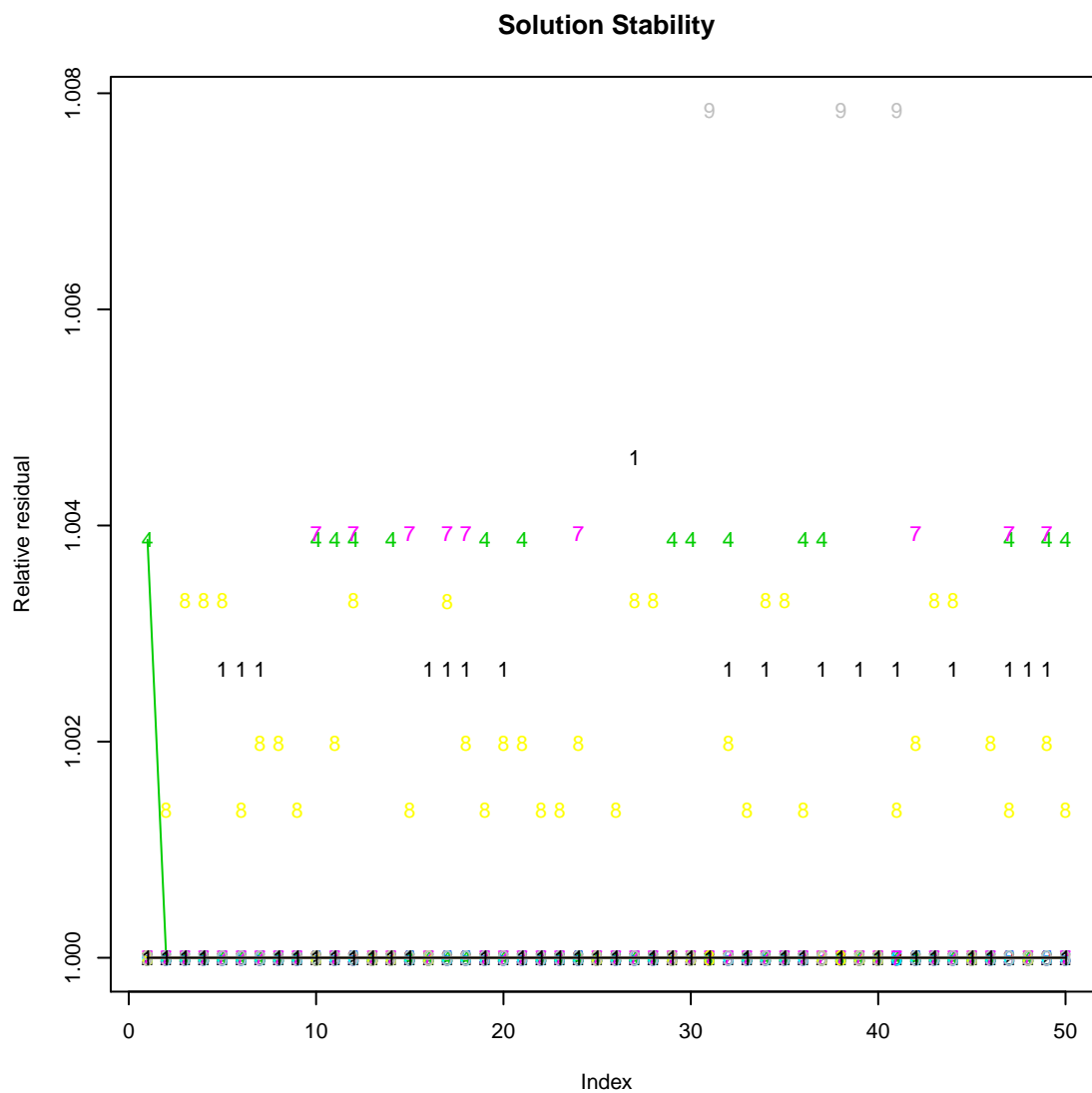




```

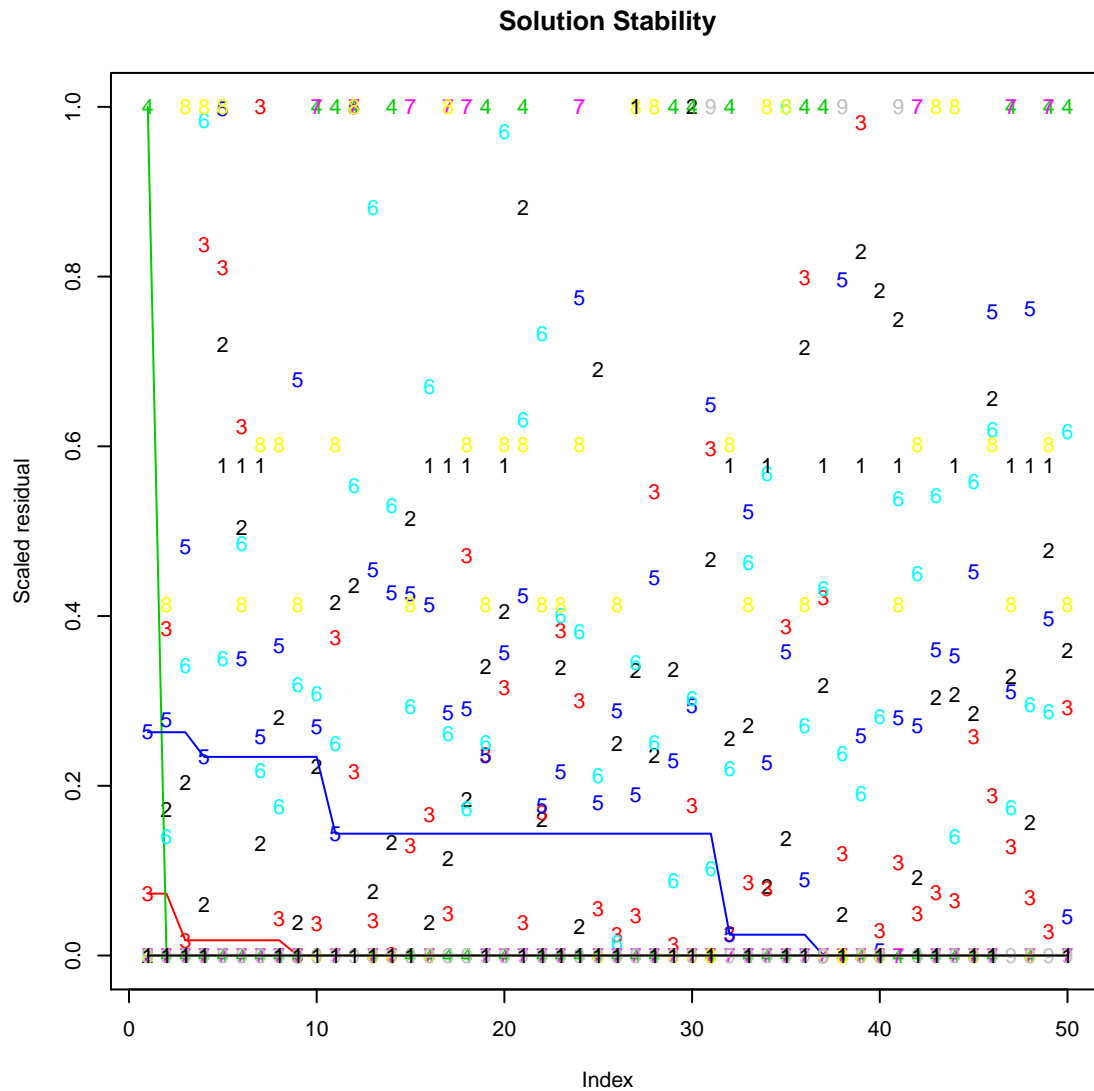
plot(0 ~ 0, type = "n", xlim = c(1, nrow(temp.resids)), ylim = range(temp.resids_rel),
     ylab = "Relative residual", main = "Solution Stability")
for (i in 1:ncol(temp.resids)) {
  points(temp.resids_rel[, i], col = i, pch = colnames(temp.resids)[i])
  lines(cummin(temp.resids_rel[, i]), col = i)
}

```



```
plot(0 ~ 0, type = "n", xlim = c(1, nrow(temp.resids)), ylim = range(temp.resids_scaled),
     ylab = "Scaled residual", main = "Solution Stability")
for (i in 1:ncol(temp.resids)) {
  points(temp.resids_scaled[, i], col = i, pch = colnames(temp.resids)[i])
  lines(cummin(temp.resids_scaled[, i]), col = i)
}
```





```

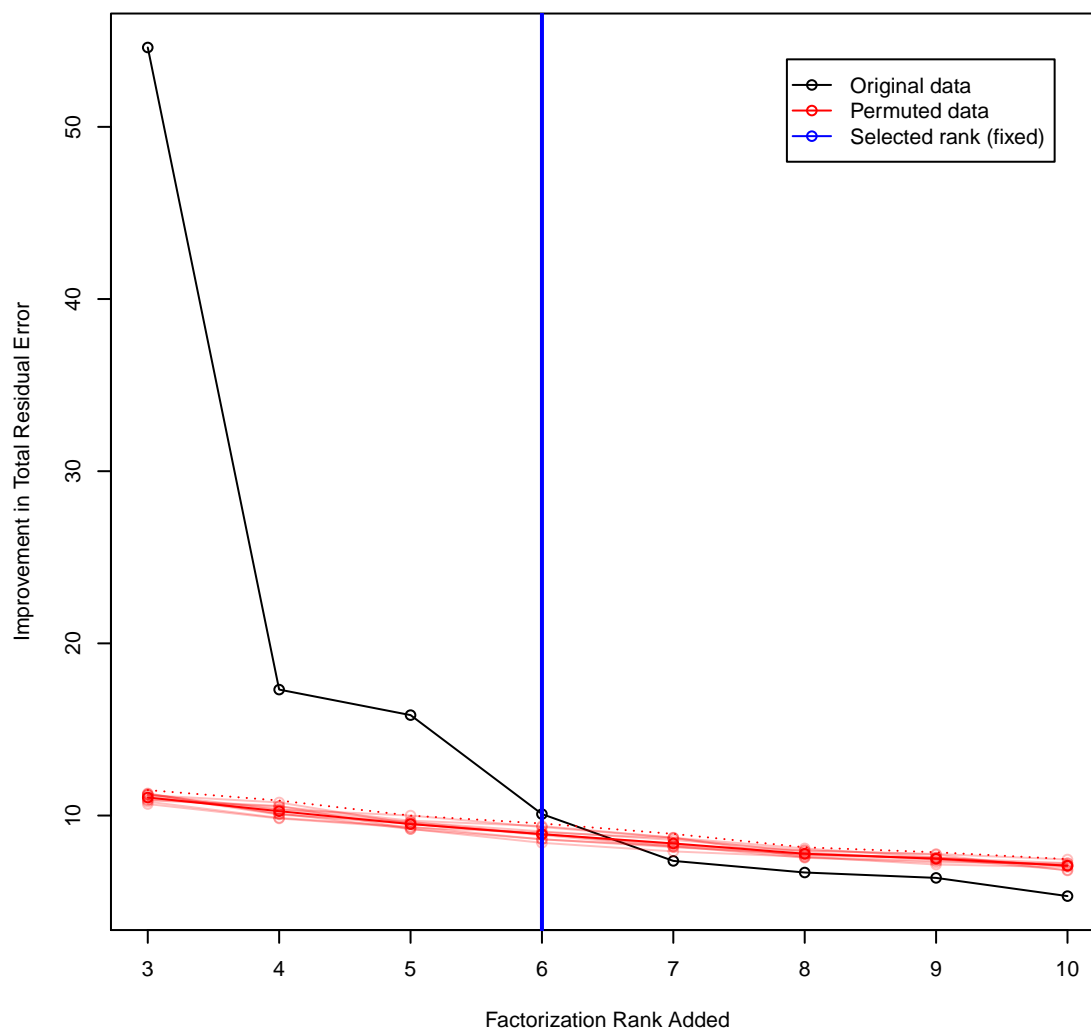
plot(nmf.rankrange[-1], -temp.orig_resids.delta, type = "o", col = "black",
     pch = 21, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.delta.mean)),
     xlab = "Factorization Rank Added", ylab = "Improvement in Total Residual Error")
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean, 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 == "auto") {
  temp.col = "green"
  nmf.rank = nmf.rank.auto
} 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(temp.col == "green", "auto", "fixed"))), col = c("black", "red",
temp.col), lty = "solid", pch = 21, inset = 0.05)

```

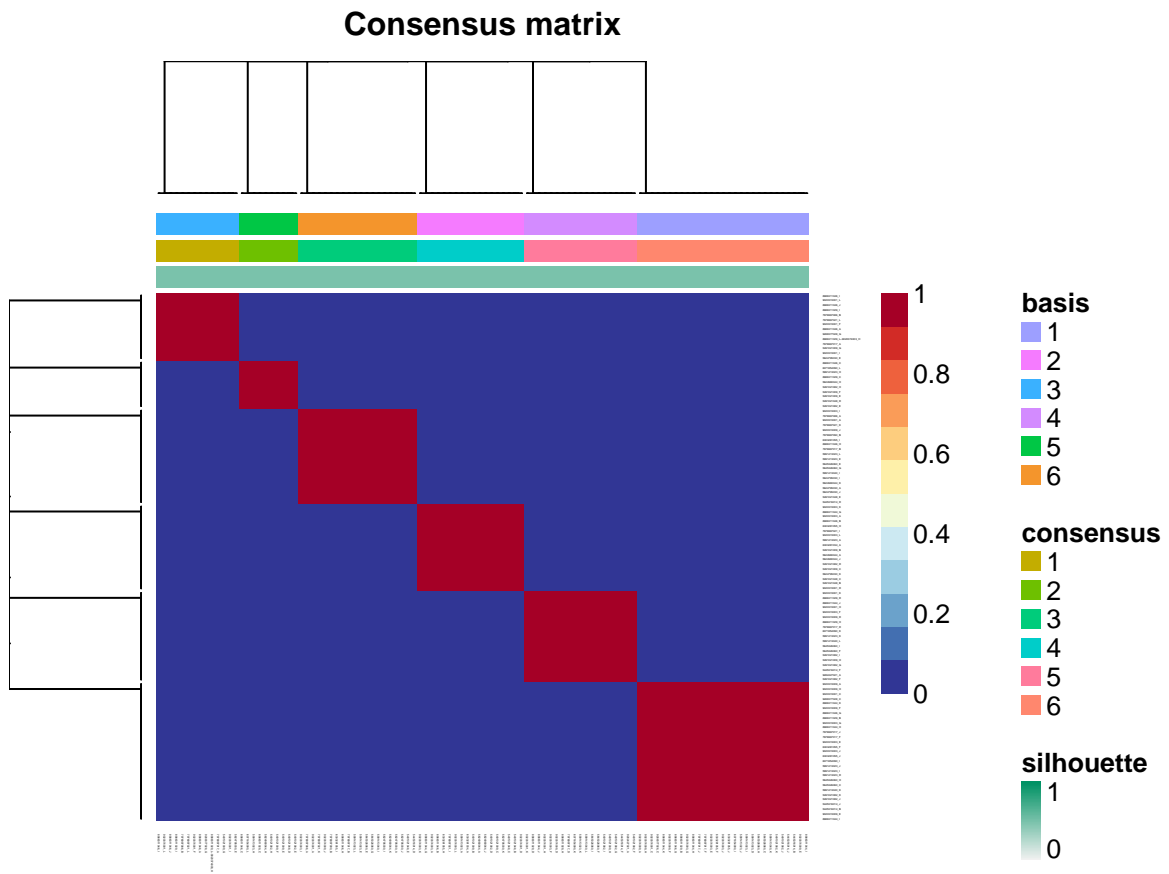


## 4.2 Fit

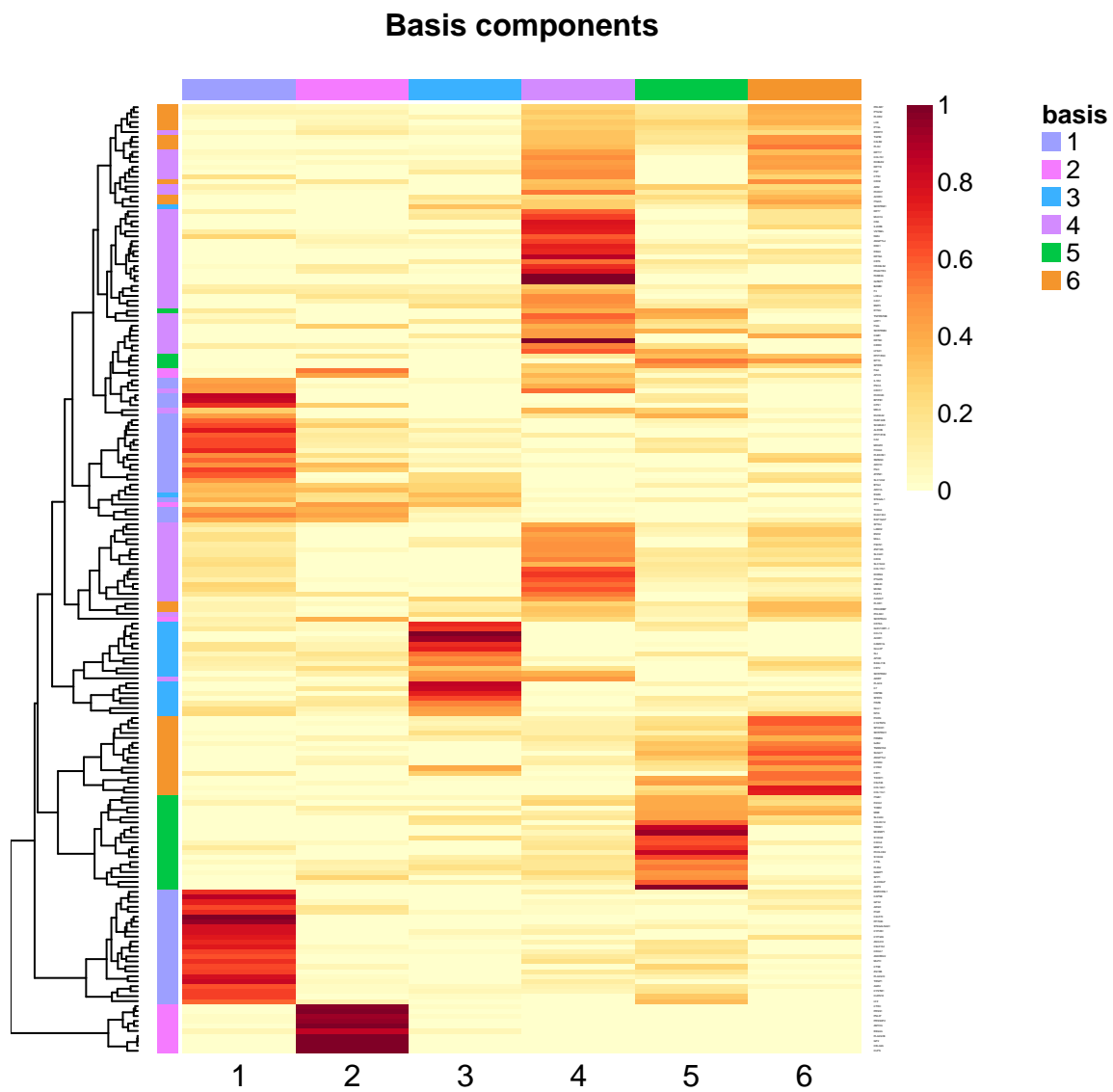
```

consensusmap(xlin.scaled.sel.nmf)

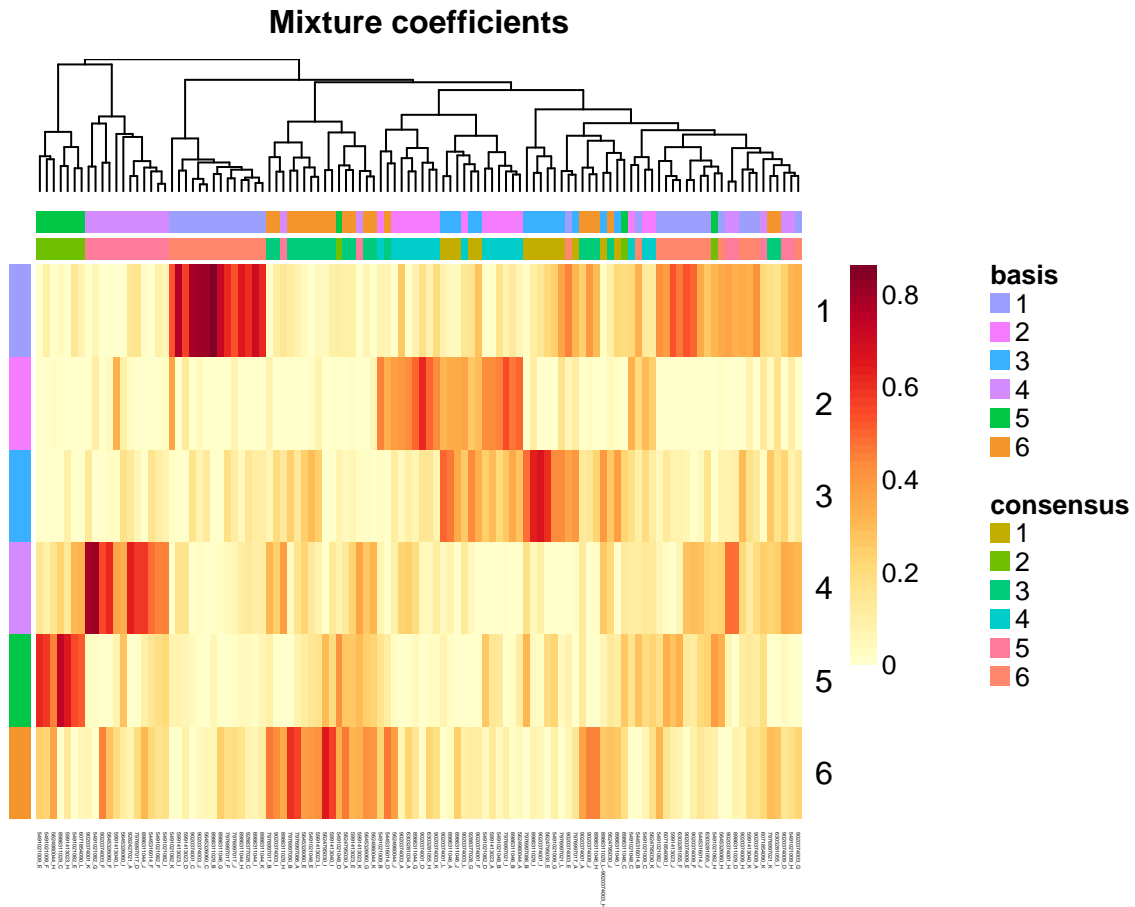
```



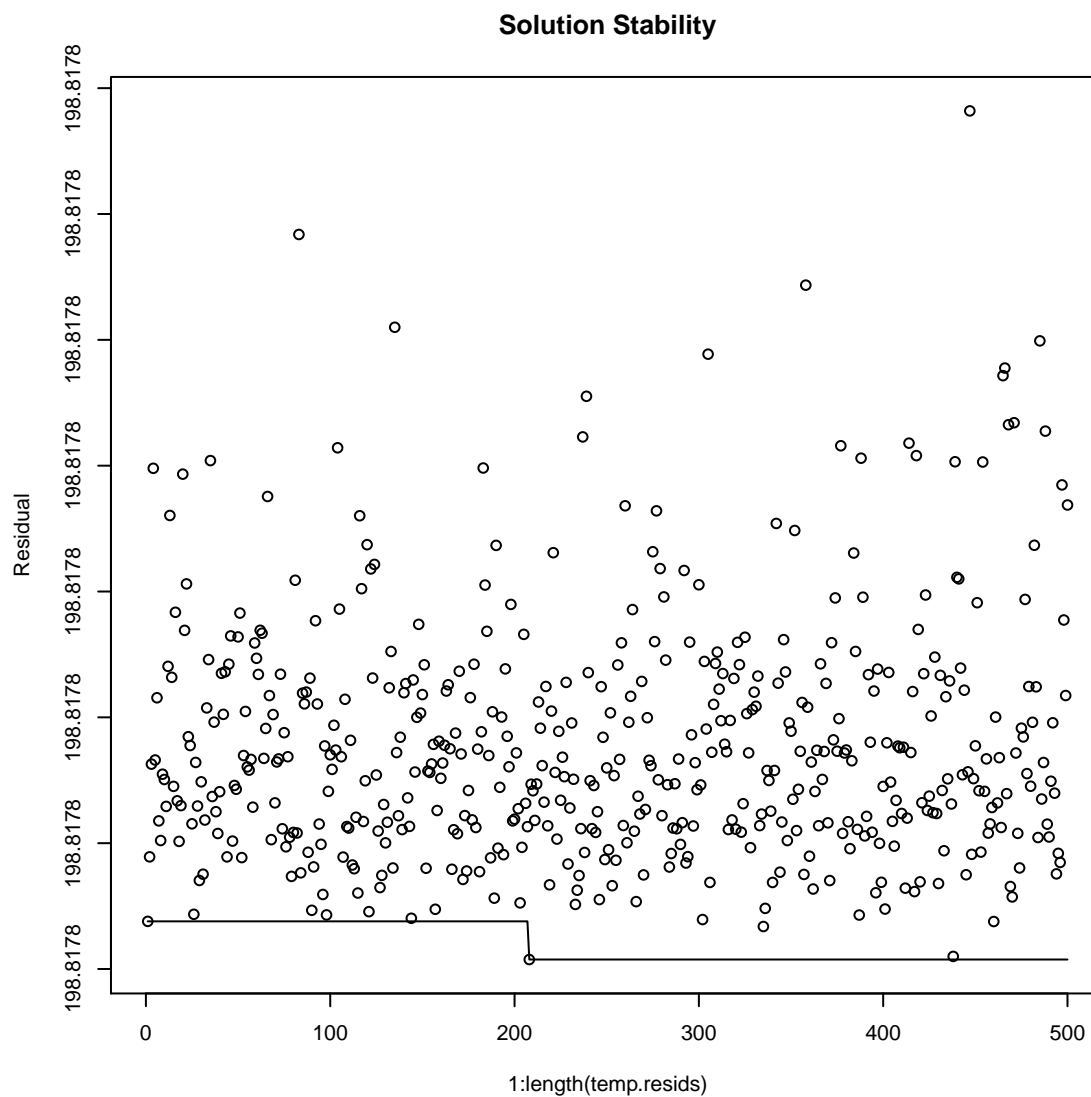
```
basimap(xlin.scaled.sel.nmf)
```



```
coefmap(xlin.scaled.sel.nmf)
```



```
temp.resids = sapply(xlin.scaled.sel.nmf, residuals)
plot(1:length(temp.resids), temp.resids, ylab = "Residual", main = "Solution Stability")
lines(1:length(temp.resids), cummin(temp.resids))
```



### 4.3 Component CPV associations

#### 4.3.1 Survival: Diagnosis to disease-specific death

```
for (i in 1:nrow(coef(xlin.scaled.sel.nmf))) {
  print(summary(coxph(y ~ coef(xlin.scaled.sel.nmf)[i, ])))
}
```

## Call:  
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])  
##  
## n= 110, number of events= 70  
##

	coef	exp(coef)	se(coef)	z	Pr(> z )
## coef(xlin.scaled.sel.nmf)[i, ]	-5.61713	0.00364	1.57210	-3.57	0.00035

##

	exp(coef)	exp(-coef)	lower .95	upper .95
##				

Table 1:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	-5.617*** (1.572)
Observations	110
R <sup>2</sup>	0.129
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-280.800
Wald Test	12.770*** (df = 1)
LR Test	15.170*** (df = 1)
Score (Logrank) Test	13.340*** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 2:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	-0.883 (1.262)
Observations	110
R <sup>2</sup>	0.005
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-288.100
Wald Test	0.490 (df = 1)
LR Test	0.509 (df = 1)
Score (Logrank) Test	0.491 (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	-4.729*** (1.740)
Observations	110
R <sup>2</sup>	0.076
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-284.000
Wald Test	7.390*** (df = 1)
LR Test	8.636*** (df = 1)
Score (Logrank) Test	7.631*** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	6.752*** (1.151)
Observations	110
R <sup>2</sup>	0.222
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-274.500
Wald Test	34.390*** (df = 1)
LR Test	27.660*** (df = 1)
Score (Logrank) Test	37.400*** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 5:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	2.962*** (1.116)
Observations	110
R <sup>2</sup>	0.054
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-285.300
Wald Test	7.050*** (df = 1)
LR Test	6.052** (df = 1)
Score (Logrank) Test	7.229*** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 6:

	<i>Dependent variable:</i>
	y
coef(xlin.scaled.sel.nmf)[i, ]	3.199** (1.305)
Observations	110
R <sup>2</sup>	0.049
Max. Possible R <sup>2</sup>	0.995
Log Likelihood	-285.600
Wald Test	6.010** (df = 1)
LR Test	5.509** (df = 1)
Score (Logrank) Test	6.128** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01



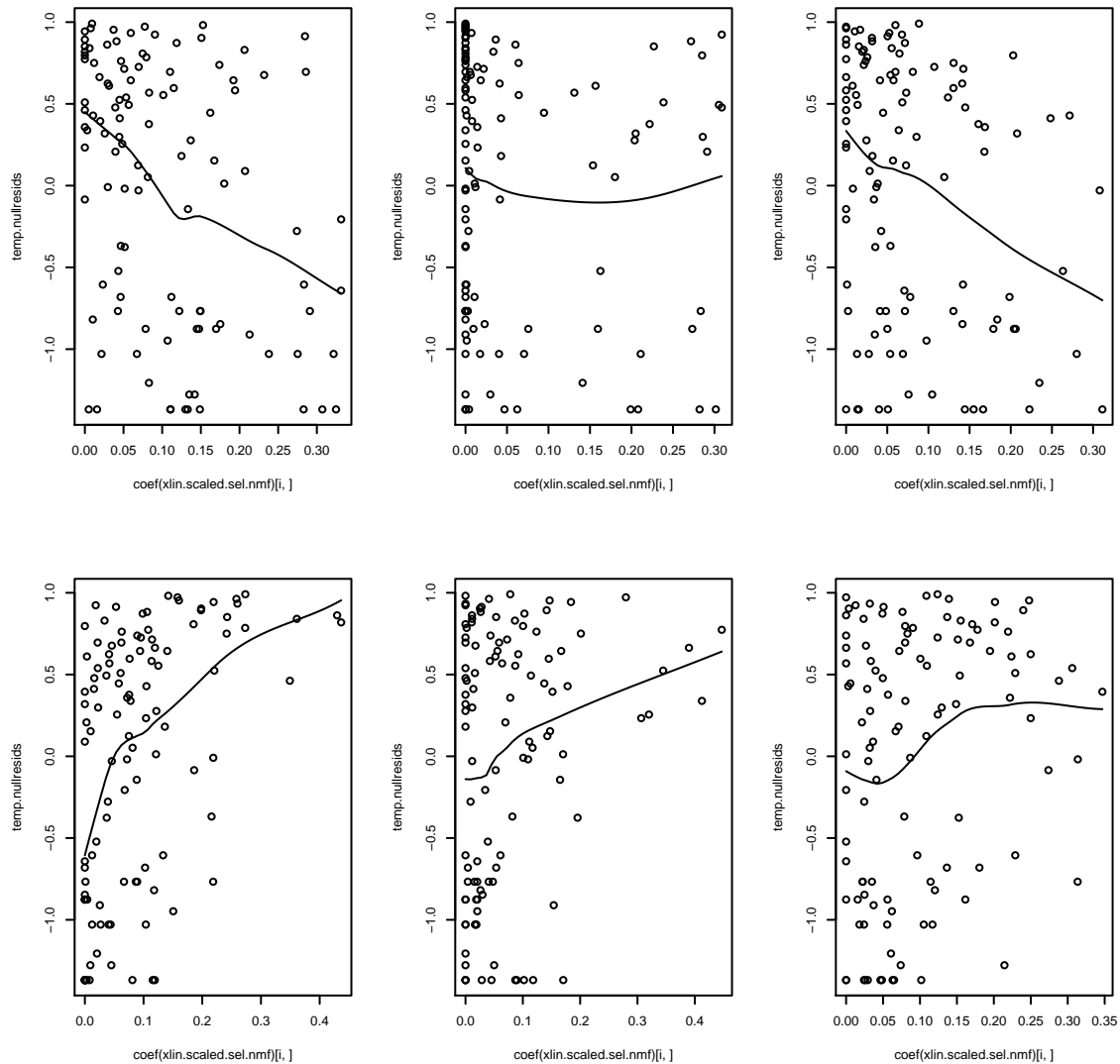
```

## coef(xlin.scaled.sel.nmf)[i, ] 0.00364      275 0.000167 0.0792
##
## Concordance= 0.636 (se = 0.038 )
## Rsquare= 0.129 (max possible= 0.995 )
## Likelihood ratio test= 15.2 on 1 df, p=9.83e-05
## Wald test = 12.8 on 1 df, p=0.000353
## Score (logrank) test = 13.3 on 1 df, p=0.00026
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 110, number of events= 70
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ] -0.883    0.413    1.262 -0.7    0.48
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]    0.413    2.42    0.0349    4.9
##
## Concordance= 0.552 (se = 0.037 )
## Rsquare= 0.005 (max possible= 0.995 )
## Likelihood ratio test= 0.51 on 1 df, p=0.476
## Wald test = 0.49 on 1 df, p=0.484
## Score (logrank) test = 0.49 on 1 df, p=0.483
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 110, number of events= 70
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ] -4.72921 0.00883 1.73989 -2.72 0.0066
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ] 0.00883    113 0.000292    0.267
##
## Concordance= 0.606 (se = 0.038 )
## Rsquare= 0.076 (max possible= 0.995 )
## Likelihood ratio test= 8.64 on 1 df, p=0.0033
## Wald test = 7.39 on 1 df, p=0.00657
## Score (logrank) test = 7.63 on 1 df, p=0.00574
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 110, number of events= 70
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ] 6.75    855.55    1.15 5.86 4.5e-09
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]    856    0.00117    89.6    8171
##

```

```
## Concordance= 0.699 (se = 0.037 )
## Rsquare= 0.222 (max possible= 0.995 )
## Likelihood ratio test= 27.7 on 1 df, p=1.45e-07
## Wald test = 34.4 on 1 df, p=4.51e-09
## Score (logrank) test = 37.4 on 1 df, p=9.61e-10
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 110, number of events= 70
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ]  2.96      19.33      1.12 2.65   0.0079
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]    19.3      0.0517      2.17      172
##
## Concordance= 0.554 (se = 0.037 )
## Rsquare= 0.054 (max possible= 0.995 )
## Likelihood ratio test= 6.05 on 1 df, p=0.0139
## Wald test = 7.05 on 1 df, p=0.00794
## Score (logrank) test = 7.23 on 1 df, p=0.00717
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 110, number of events= 70
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ]  3.2      24.5      1.3 2.45   0.014
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]    24.5      0.0408      1.9      316
##
## Concordance= 0.58 (se = 0.038 )
## Rsquare= 0.049 (max possible= 0.995 )
## Likelihood ratio test= 5.51 on 1 df, p=0.0189
## Wald test = 6.01 on 1 df, p=0.0142
## Score (logrank) test = 6.13 on 1 df, p=0.0133
```

```
temp.nullfit = coxph(y ~ 1)
temp.nullresids = residuals(temp.nullfit, type = "martingale")
par(mfrow = c(2, 3))
for (i in 1:nrow(coef(xlin.scaled.sel.nmf))) {
  scatter.smooth(temp.nullresids ~ coef(xlin.scaled.sel.nmf)[i, ])
}
```



```
par(mfrow = c(1, 1))
```

## 4.4 Purity

```
apply(coef(xlin.scaled.sel.nmf), 1, function(xc) cor.test(samps$purity_qpure,
  xc, method = "kendall"))

## [[1]]
##
## Kendall's rank correlation tau
##
## data: samps$purity_qpure and xc
## z = 3.878, p-value = 0.0001054
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
```

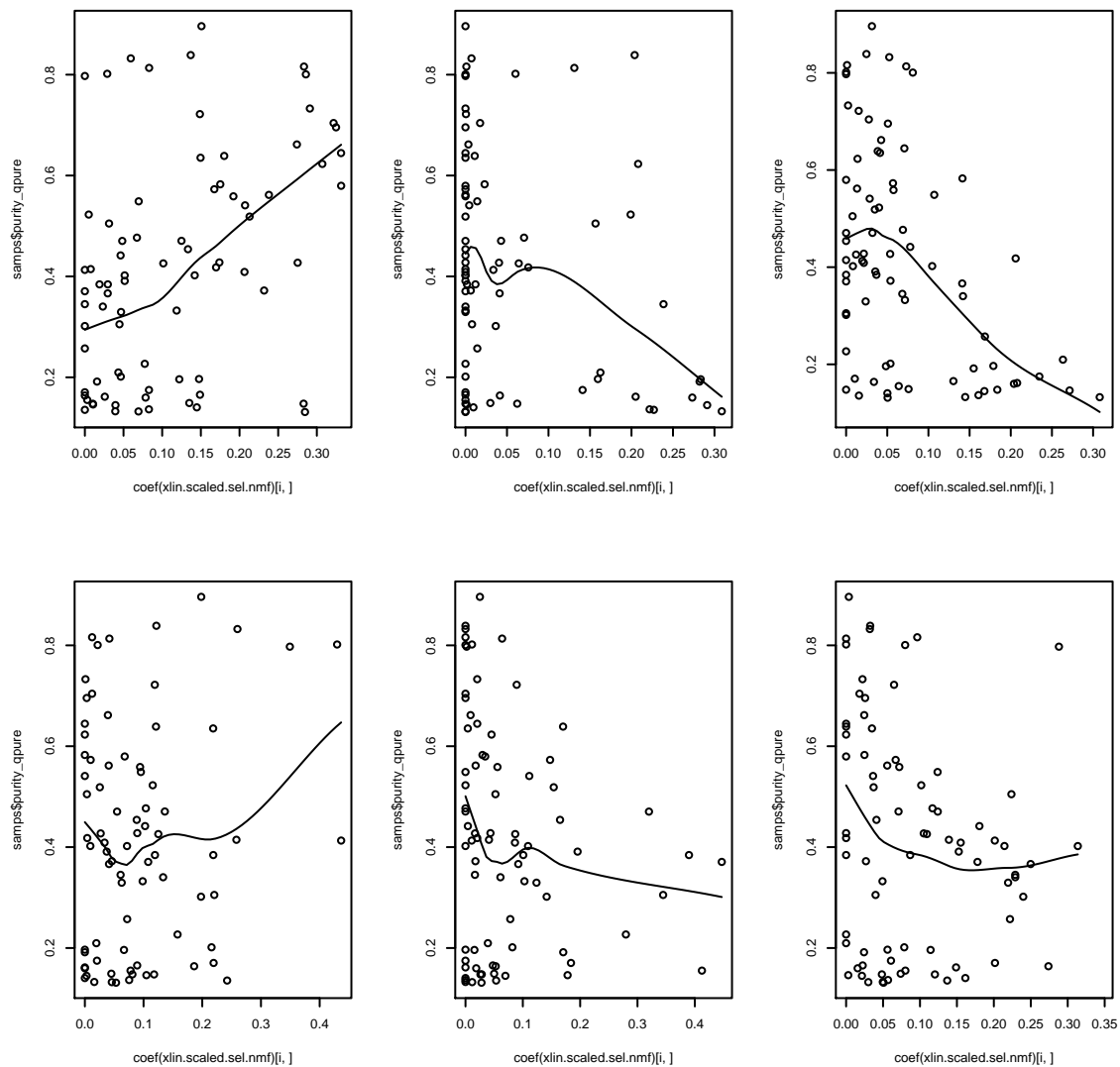
```

##      tau
## 0.3009
##
##
## [[2]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -1.923, p-value = 0.05452
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.157
##
##
## [[3]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -3.156, p-value = 0.0016
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.2458
##
##
## [[4]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = 0.1641, p-value = 0.8697
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.01273
##
##
## [[5]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -1.787, p-value = 0.07387
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1404
##
##
## [[6]]
##

```

```
## Kendall's rank correlation tau
##
## data: samps$purity_qpure and xc
## z = -2.148, p-value = 0.03174
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1671

par(mfrow = c(2, 3))
for (i in 1:nrow(coef(xlin.scaled.sel.nmf))) {
  scatter.smooth(samps$purity_qpure ~ coef(xlin.scaled.sel.nmf)[i, ])
}
```



```
par(mfrow = c(1, 1))
```

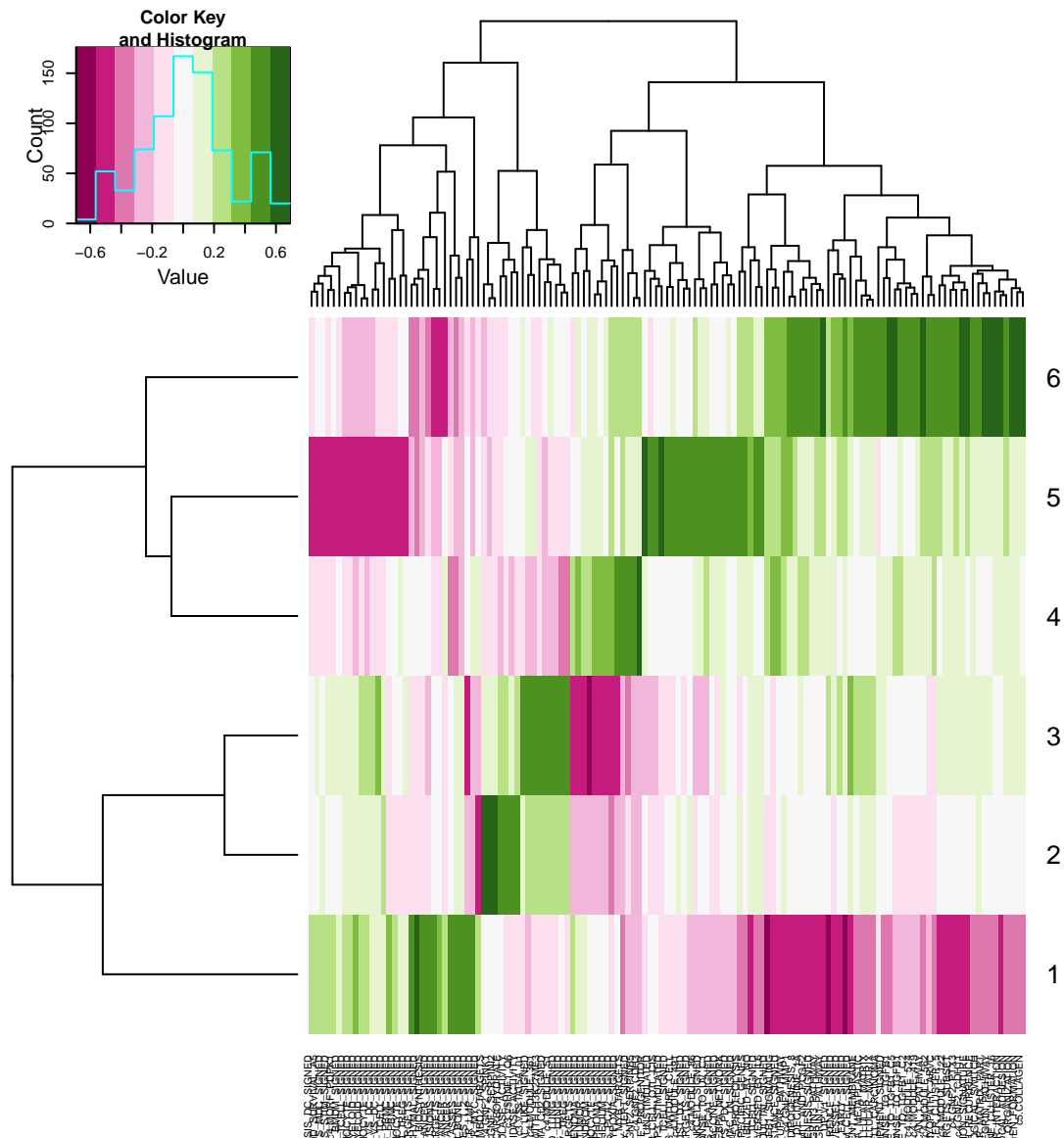
## 4.5 MTC P-values

```
xlin.scaled.sel.nmf.cpv.pvals = data.frame(p.surv = apply(coef(xlin.scaled.sel.nmf),
  1, function(xc) pchisq(2 * diff(coxph(y ~ xc)$loglik), df = 1, lower.tail = FALSE)),
  p.pure = apply(coef(xlin.scaled.sel.nmf), 1, function(xc) cor.test(samps$purity_qpure,
    xc, method = "kendall")$p.value), d.surv = apply(coef(xlin.scaled.sel.nmf),
    1, function(xc) sign(coef(coxph(y ~ xc)))), d.pure = apply(coef(xlin.scaled.sel.nmf),
    1, function(xc) sign(cor.test(samps$purity_qpure, xc, method = "kendall")$statistic)))
temp.pvals.FWER = p.adjust(c(xlin.scaled.sel.nmf.cpv.pvals$p.surv, xlin.scaled.sel.nmf.cpv.pvals$p.pure),
  "holm")
temp.qvals.BY = p.adjust(c(xlin.scaled.sel.nmf.cpv.pvals$p.surv, xlin.scaled.sel.nmf.cpv.pvals$p.pure),
  "BY")
xlin.scaled.sel.nmf.cpv.pvals$p.surv.FWER = temp.pvals.FWER[1:(length(temp.pvals.FWER)/2)]
xlin.scaled.sel.nmf.cpv.pvals$p.pure.FWER = temp.pvals.FWER[(length(temp.pvals.FWER)/2 +
  1):length(temp.pvals.FWER)]
xlin.scaled.sel.nmf.cpv.pvals$q.surv.BY = temp.qvals.BY[1:(length(temp.qvals.BY)/2)]
xlin.scaled.sel.nmf.cpv.pvals$q.pure.BY = temp.qvals.BY[(length(temp.qvals.BY)/2 +
  1):length(temp.qvals.BY)]
xlin.scaled.sel.nmf.cpv.pvals

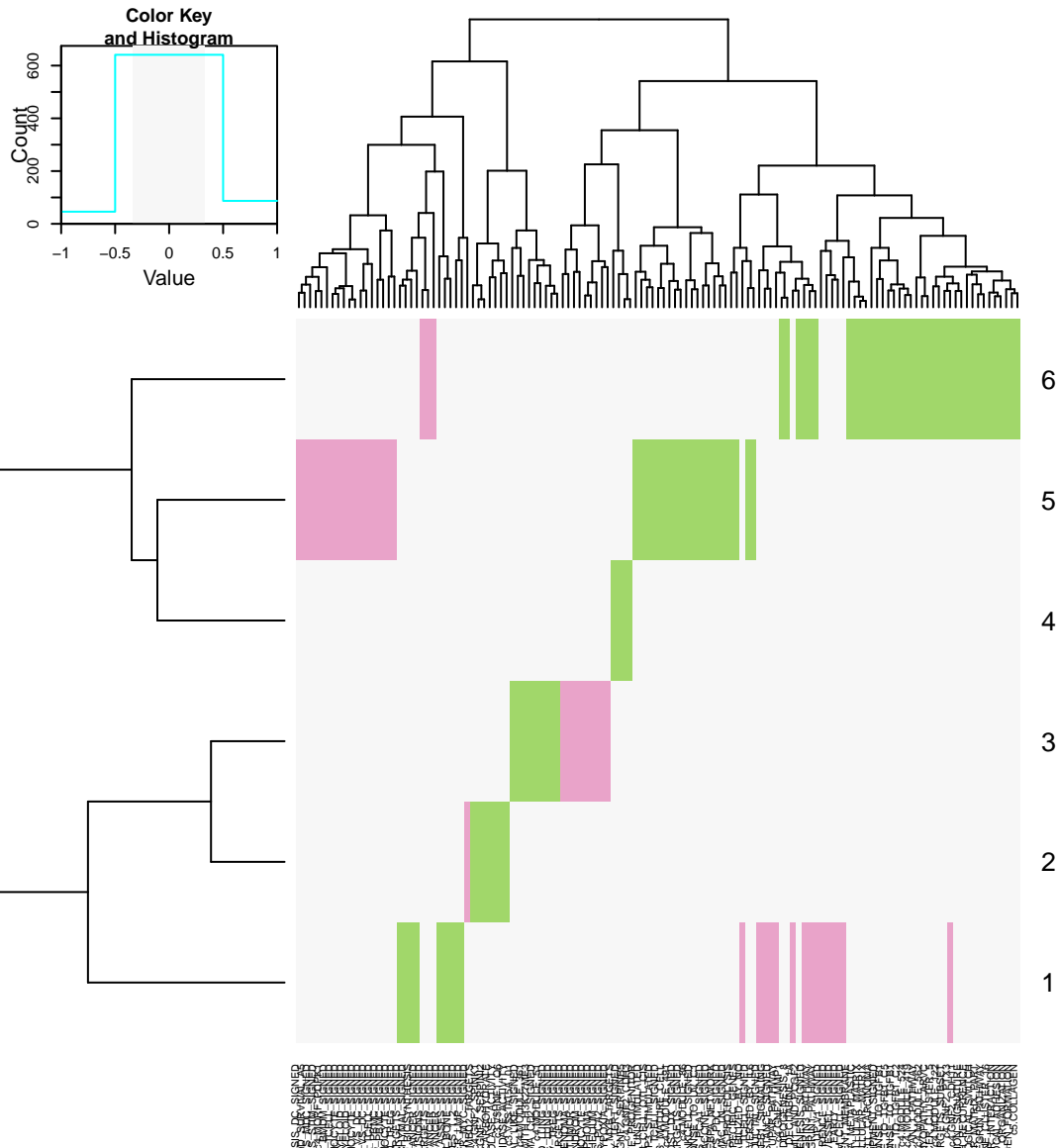
##      p.surv    p.pure d.surv d.pure p.surv.FWER p.pure.FWER q.surv.BY
## 1 9.829e-05 0.0001054    -1      1  1.081e-03  0.001081 1.308e-03
## 2 4.755e-01 0.0545221    -1     -1  9.511e-01  0.218088 1.000e+00
## 3 3.295e-03 0.0016002    -1     -1  2.636e-02  0.014402 2.454e-02
## 4 1.449e-07 0.8696578      1      1  1.739e-06  0.951062 5.395e-06
## 5 1.389e-02 0.0738693      1     -1  9.725e-02  0.221608 8.622e-02
## 6 1.892e-02 0.0317433      1     -1  1.135e-01  0.158717 1.006e-01
##    q.pure.BY
## 1  0.001308
## 2  0.225592
## 3  0.014897
## 4  1.000000
## 5  0.275078
## 6  0.147759
```

## 4.6 MSigDB score correlation thresholding

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



```
heatmap.2(xlin.scaled.sel.nmf.msigdb.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))
```



```
temp.sig_id = colnames(xlin.scaled.sel.nmf.msigdb.corr)
temp.sig_class = gsub("\\.*", "", temp.sig_id)
temp.nsig = length(temp.sig_id)
temp.nmeta = nrow(xlin.scaled.sel.nmf.msigdb.corr)
tables = lapply(1:temp.nmeta, function(metagene_i) {
  tapply(1:temp.nsig, temp.sig_class, function(sig_class_is) {
    all_cors = xlin.scaled.sel.nmf.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)]) * sign(xlin.scaled.sel.nmf.cpv.pvals$d.surrogate),
              collapse = ",")
          })
      table = table[order(-(table$Correlation)), ]
      rownames(table) <- NULL
    }
  }
  table
}, simplify = FALSE)
})
tables

## [[1]]
## [[1]]$c1
## data frame with 0 columns and 0 rows
##
## [[1]]$c2
##
##                                     GeneSet Correlation
## 1  c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_SIGNED      0.5832
## 2                                     c2.LUCAS_HNF4A_TARGETS_SIGNED      0.5235
## 3                                     c2.DOANE_BREAST_CANCER_CLASSES_SIGNED      0.5205
## 4                                     c2.LIU_PROSTATE_CANCER_SIGNED      0.5141
## 5                                     c2.WAMUNYOKOLI_OVARIAN_CANCER_LMP_SIGNED      0.5131
## 6                                     c2.REACTOME_GLYCEROPHOSPHOLIPID_BIOSYNTHESIS      0.5131
## 7                                     c2.WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_SIGNED      0.5111
## 8                                     c2.SMID_BREAST_CANCER_RELAPSE_IN_BONE_SIGNED      0.5084
## 9  c2.WANG_BARRETTS_ESOPHAGUS_AND_ESOPHAGUS_CANCER_SIGNED      0.5044
## 10                                    c2.PID_UPA_UPAR_PATHWAY      -0.5011
## 11                                    c2.PID_INTEGRIN3_PATHWAY      -0.5021
## 12                                    c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABINE_1A      -0.5031
## 13                                    c2.WOO_LIVER_CANCER_RECURRENCE_SIGNED      -0.5057
## 14                                    c2.CROMER_TUMORIGENESIS_SIGNED      -0.5084
## 15                                    c2.KARAKAS_TGFB1_SIGNALING      -0.5088
## 16                                    c2.LIM_MAMMARY_STEM_CELL_SIGNED      -0.5135
## 17                                    c2.KUWANO_RNA_STABILIZED_BY_NO      -0.5145
## 18                                    c2.HUANG_DASATINIB_RESISTANCE_SIGNED      -0.5192
## 19                                    c2.ROY_WOUND_BLOOD_VESSEL_SIGNED      -0.5198
## 20                                    c2.PID_INTEGRIN1_PATHWAY      -0.5205
## 21                                    c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SIGNED      -0.6133
## 22  c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SIGNED      -0.6180
##      Metagenes
## 1          -1
## 2          -1
## 3          -1
## 4          -1
## 5          -1
## 6          -1
## 7          -1

```

```

## 8      -1
## 9      -1
## 10     1
## 11     1,-6
## 12     1
## 13     1
## 14     1,-6
## 15     1
## 16     1
## 17     1
## 18     1
## 19     1
## 20     1,-6
## 21     1
## 22     1
##
## [[1]]$c3
## data frame with 0 columns and 0 rows
##
## [[1]]$c4
##      GeneSet Correlation Metagenes
## 1 c4.GNF2_PTX3      -0.5135      1,-6
##
## [[1]]$c5
## data frame with 0 columns and 0 rows
##
## [[1]]$c6
##      GeneSet Correlation Metagenes
## 1 c6.LEF1_UP.V1_SIGNED      -0.5825      1
##
## [[1]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[2]]
## [[2]]$c1
## data frame with 0 columns and 0 rows
##
## [[2]]$c2
##
##      GeneSet Correlation Metagenes
## 1      c2.BERGER_MBD2_TARGETS      0.6024      -2
## 2 c2.REACTOME_DIGESTION_OF_DIETARY_CARBOHYDRATE      0.5220      -2
## 3      c2.LEE_LIVER_CANCER_MYC_SIGNED      -0.5188      2
##
## [[2]]$c3
##      GeneSet Correlation Metagenes
## 1 c3.V$HNF1_Q6      0.5202      -2
##
## [[2]]$c4
##      GeneSet Correlation Metagenes
## 1 c4.GNF2_SPINK1      0.6907      -2
## 2 c4.GNF2_SERPINI2      0.6774      -2
##
## [[2]]$c5

```

```

##                               GeneSet Correlation Metagenes
## 1 c5.CARBOXYPEPTIDASE_ACTIVITY      0.5332      -2
## 2 c5.SERINE_HYDROLASE_ACTIVITY      0.5083      -2
##
## [[2]]$c6
## data frame with 0 columns and 0 rows
##
## [[2]]$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.ZHAN_MULTIPLE_MYELOMA_CD2_SIGNED      0.5079      -3
## 2          c2.SMID_BREAST_CANCER_NORMAL_LIKE_SIGNED      0.5022      -3
## 3          c2.MIKKELSEN_MCV6_ICP_WITH_H3K27ME3      0.5022      -3
## 4          c2.GREENBAUM_E2A_TARGETS_SIGNED      -0.5035       3
## 5          c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_SIGNED      -0.5092       3
## 6 c2.MARTORIATI_MDM4_TARGETS_NEUROEPITHELIUM_SIGNED      -0.5129       3
## 7          c2.WINTER_HYPOXIA_SIGNED      -0.5213       3
## 8          c2.SABATES_COLORECTAL_ADENOMA_SIGNED      -0.5217       3
## 9          c2.LEE_EARLY_T_LYMPHOCYTE_SIGNED      -0.5506       3
## 10         c2.WEST_ADRENOCORTICAL_TUMOR_SIGNED      -0.5553       3
## 11         c2.SHIPP_DLBCL_VS_FOLLICULAR_LYMPHOMA_SIGNED      -0.5594       3
## 12         c2.HAHTOLA_SEZARY_SYNDROM_SIGNED      -0.5654       3
##
## [[3]]$c3
##                               GeneSet Correlation Metagenes
## 1 c3.V$STAT5A_01      0.5224      -3
##
## [[3]]$c4
##                               GeneSet Correlation Metagenes
## 1 c4.MODULE_51      0.5439      -3
## 2 c4.MODULE_361      0.5160      -3
##
## [[3]]$c5
## data frame with 0 columns and 0 rows
##
## [[3]]$c6
## data frame with 0 columns and 0 rows
##
## [[3]]$c7
##                               GeneSet Correlation
## 1          c7.GSE20715_OH_VS_48H_OZONE_TLR4_KO_LUNG_SIGNED      0.5277
## 2          c7.GSE20715_OH_VS_48H_OZONE_LUNG_SIGNED      0.5123
## 3 c7.GSE24634_NAIVE_CD4_TCELL_VS_DAY3_IL4_CONV_TREG_SIGNED      0.5015
## Metagenes
## 1      -3
## 2      -3
## 3      -3

```

```
##
##
## [[4]]
## [[4]]$c1
## data frame with 0 columns and 0 rows
##
## [[4]]$c2
##
##           GeneSet Correlation Metagenes
## 1           c2.LEI_MYB_TARGETS      0.5146      4
## 2 c2.KANG_IMMORTALIZED_BY_TERT_SIGNED      0.5049      4
##
## [[4]]$c3
## data frame with 0 columns and 0 rows
##
## [[4]]$c4
##
##           GeneSet Correlation Metagenes
## 1           c4.GNF2_CDH3      0.5809      4
## 2 c4.GNF2_SERPINB5      0.5560      4
##
## [[4]]$c5
## data frame with 0 columns and 0 rows
##
## [[4]]$c6
## data frame with 0 columns and 0 rows
##
## [[4]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[5]]
## [[5]]$c1
## data frame with 0 columns and 0 rows
##
## [[5]]$c2
##
##           GeneSet Correlation Metagenes
## 1           c2.IVANOVA_HEMATOPOIESIS_LATE_PROGENITOR      0.6073      5
## 2           c2.MARSON_BOUND_BY_FOXP3_STIMULATED      0.5876      5
## 3           c2.SESTO_RESPONSE_TO_UV_C1      0.5488      5
## 4           c2.MARTINEZ_RB1_AND_TP53_TARGETS_SIGNED      0.5308      5
## 5           c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA1_SIGNED      0.5267      5
## 6           c2.RAGHAVACHARI_PLATELET_SPECIFIC_GENES      0.5219      5
## 7           c2.IVANOVA_HEMATOPOIESIS_MATURE_CELL      0.5216      5
## 8           c2.KAMIKUBO_MYELOID_CEBPA_NETWORK      0.5192      5
## 9           c2.MARSON_BOUND_BY_FOXP3_UNSTIMULATED      0.5189      5
## 10          c2.BYSTRYKH_HEMATOPOIESIS_STEM_CELL_QTL_CIS      0.5131      5
## 11           c2.BROCKE_APOPTOSIS_REVERSED_BY_IL6      0.5080      5
## 12          c2.SHEDDEN_LUNG_CANCER_GOOD_SURVIVAL_A5      -0.5042     -5
##
## [[5]]$c3
## data frame with 0 columns and 0 rows
##
## [[5]]$c4
##
##           GeneSet Correlation Metagenes
## 1           c4.MODULE_86      0.5209      5
```

```

## 2 c4.MODULE_491      0.5117      5
## 3 c4.MORF_PDPK1     -0.5107     -5
##
## [[5]]$c5
##      GeneSet Correlation Metagenes
## 1 c5.PROTEOLYSIS      0.5134      5
##
## [[5]]$c6
## data frame with 0 columns and 0 rows
##
## [[5]]$c7
##                                     GeneSet
## 1 c7.GSE11057_PBMC_VS_MEM_CD4_TCELL_SIGNED
## 2 c7.GSE29618_MONOCYTE_VS_PDC_SIGNED
## 3 c7.GSE15767_MED_VS_SCS_MAC_LN_SIGNED
## 4 c7.GSE3982_MAC_VS_NKCELL_SIGNED
## 5 c7.GSE3982_BASOPHIL_VS_CENT_MEMORY_CD4_TCELL_SIGNED
## 6 c7.GSE29618_MONOCYTE_VS_PDC_DAY7_FLU_VACCINE_SIGNED
## 7 c7.GSE9006_HEALTHY_VS_TYPE_1_DIABETES_PBMC_4MONTH_POST_DX_SIGNED
## 8 c7.GSE20366_CD103_KLRG1_DP_VS_DN_TREG_SIGNED
## 9 c7.GSE11057_NAIVE_CD4_VS_PBMC_CD4_TCELL_SIGNED
## 10 c7.GSE1448_CTRL_VS_ANTI_VALPHA2_DP_THYMOCYTE_SIGNED
## 11 c7.GSE360_T_GONDII_VS_M_TUBERCULOSIS_DC_SIGNED
## 12 c7.GSE17721_POLYIC_VS_GARDIQUIMOD_1H_BMDM_SIGNED
## 13 c7.GSE29618_PDC_VS_MDC_SIGNED
## 14 c7.GSE22886_NAIVE_CD8_TCELL_VS_DC_SIGNED
## 15 c7.GSE3982_DC_VS_MAC_LPS_STIM_SIGNED
## 16 c7.GSE11057_CD4_CENT_MEM_VS_PBMC_SIGNED
## 17 c7.GSE11057_NAIVE_VS_MEMORY_CD4_TCELL_SIGNED
## 18 c7.GSE22886_NAIVE_CD4_TCELL_VS_DC_SIGNED
## 19 c7.GSE10325_BCELL_VS_MYELOID_SIGNED
## 20 c7.GSE11057_CD4_EFF_MEM_VS_PBMC_SIGNED
## 21 c7.GSE22886_NAIVE_TCELL_VS_MONOCYTE_SIGNED
## 22 c7.GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_MYELOID_SIGNED
## 23 c7.GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_SIGNED
##      Correlation Metagenes
## 1      0.5617      5
## 2      0.5607      5
## 3      0.5481      5
## 4      0.5399      5
## 5      0.5386      5
## 6      0.5352      5
## 7      0.5018      5
## 8     -0.5005     -5
## 9     -0.5035     -5
## 10    -0.5080     -5
## 11    -0.5080     -5
## 12    -0.5090     -5
## 13    -0.5124     -5
## 14    -0.5168     -5
## 15    -0.5182     -5
## 16    -0.5267     -5
## 17    -0.5308     -5
## 18    -0.5314     -5

```

```

## 19      -0.5348      -5
## 20      -0.5396      -5
## 21      -0.5396      -5
## 22      -0.5529      -5
## 23      -0.5553      -5
##
##
## [[6]]
## [[6]]$c1
## data frame with 0 columns and 0 rows
##
## [[6]]$c2
##
##                                     GeneSet
## 1                                c2.REACTOME_COLLAGEN_FORMATION
## 2                                c2.PID_SYNDECAN_1_PATHWAY
## 3                   c2.REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
## 4                   c2.VERRECCHIA_DELAYED_RESPONSE_TO_TGFB1
## 5      c2.ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE
## 6                                c2.PID_INTEGRIN1_PATHWAY
## 7                                c2.PID_AVB3_INTEGRIN_PATHWAY
## 8                                c2.KEGG_ECM_RECEPTOR_INTERACTION
## 9                                c2.VERRECCHIA_RESPONSE_TO_TGFB1_C5
## 10                   c2.YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_16
## 11                   c2.VERRECCHIA_EARLY_RESPONSE_TO_TGFB1
## 12                   c2.KEGG_FOCAL_ADHESION
## 13 c2.KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
## 14                   c2.FARMER_BREAST_CANCER_CLUSTER_5
## 15                   c2.BURTON_ADIPOGENESIS_8
## 16                   c2.KEGG_BASAL_CELL_CARCINOMA
## 17                   c2.MAHADEVAN_GIST_MORPHOLOGICAL_SWITCH
## 18                   c2.CAIRO_LIVER_DEVELOPMENT_SIGNED
## 19                   c2.ROZANOV_MMP14_TARGETS_SUBSET
## 20                   c2.LIEN_BREAST_CARCINOMA_METAPLASTIC
## 21                   c2.VERRECCHIA_RESPONSE_TO_TGFB1_C2
## 22                   c2.PID_INTEGRIN3_PATHWAY
## 23                   c2.CROMER_TUMORIGENESIS_SIGNED
## 24                   c2.LINDGREN_BLADDER_CANCER_HIGH_RECURRENCE
## 25                   c2.WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2
## 26                   c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_SIGNED
## 27                   c2.BERTUCCI_MEDULLARY_VS_DUCTAL_BREAST_CANCER_SIGNED
## 28                   c2.PASINI_SUZ12_TARGETS_SIGNED
##      Correlation Metagenes
## 1      0.6446      6
## 2      0.6278      6
## 3      0.6124      6
## 4      0.6067      6
## 5      0.6023      6
## 6      0.5919     -1,6
## 7      0.5913      6
## 8      0.5872      6
## 9      0.5778      6
## 10     0.5758      6
## 11     0.5637      6
## 12     0.5513      6

```

```

## 13      0.5389      6
## 14      0.5389      6
## 15      0.5366      6
## 16      0.5356      6
## 17      0.5322      6
## 18      0.5238      6
## 19      0.5195      6
## 20      0.5181      6
## 21      0.5171      6
## 22      0.5151     -1,6
## 23      0.5107     -1,6
## 24      0.5097      6
## 25      0.5040      6
## 26     -0.5034     -6
## 27     -0.5299     -6
## 28     -0.5493     -6
##
## [[6]]$c3
## data frame with 0 columns and 0 rows
##
## [[6]]$c4
##      GeneSet Correlation Metagenes
## 1 c4.GNF2_CDH11      0.6214      6
## 2 c4.GNF2_PTX3      0.5406     -1,6
## 3 c4.GNF2_MMP1      0.5325      6
## 4 c4.MODULE_122      0.5211      6
## 5 c4.MODULE_524      0.5144      6
## 6 c4.MODULE_419      0.5124      6
## 7 c4.MODULE_562      0.5037      6
##
## [[6]]$c5
##      GeneSet Correlation Metagenes
## 1      c5.COLLAGEN      0.6510      6
## 2 c5.EXTRACELLULAR_MATRIX_PART      0.6074      6
## 3 c5.PROTEINACEOUS_EXTRACELLULAR_MATRIX      0.5315      6
## 4      c5.EXTRACELLULAR_MATRIX      0.5158      6
## 5      c5.BASEMENT_MEMBRANE      0.5030      6
##
## [[6]]$c6
## data frame with 0 columns and 0 rows
##
## [[6]]$c7
## data frame with 0 columns and 0 rows

```

```

print(asreg.result)

## glmulti.analysis
## Method: h / Fitting: coxph / IC used: bic
## Level: 2 / Marginality: TRUE
## From 100 models:
## Best IC: 551.639174852458
## Best model:
## [1] "Surv(time, event) ~ 1 + mg.1 + mg.3 + mg.4"

```

```
## Evidence weight: 0.117360168471618
## Worst IC: 561.061609642289
## 6 models within 2 IC units.
## 71 models to reach 95% of evidence weight.

coef(asreg.result)

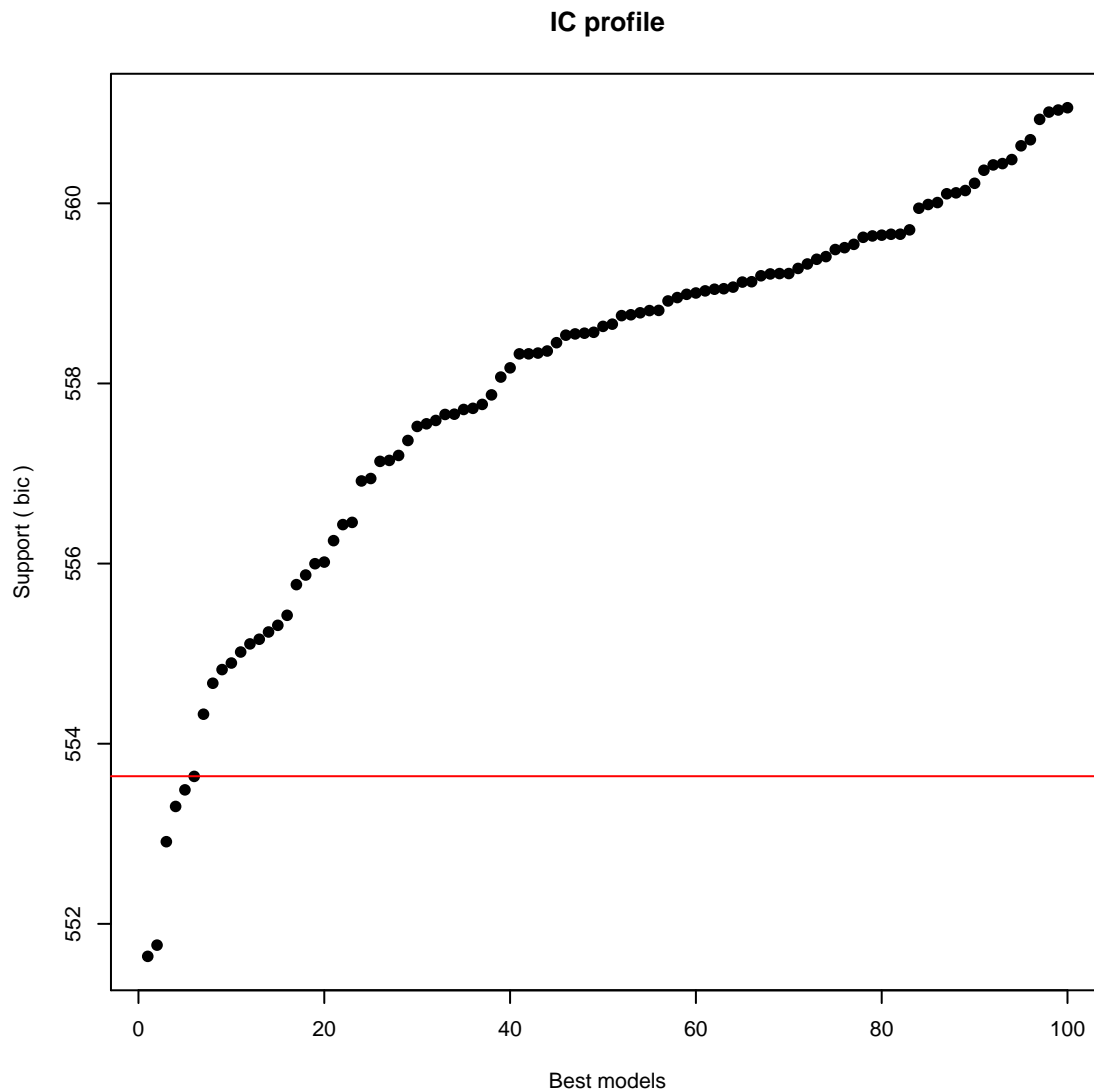
##           Estimate Uncond. variance Nb models Importance +/- (alpha=0.05)
## mg.2:mg.3 -0.003903      0.001261         1    0.001702      0.0704
## mg.2:mg.6  0.063780      0.021627         2    0.002862      0.2915
## mg.2:mg.5 -0.050020      0.018803         1    0.003553      0.2718
## mg.2:mg.4  0.019628      0.013704         2    0.004645      0.2321
## mg.3:mg.5 -0.007807      0.013838         2    0.005025      0.2332
## mg.1:mg.6  0.080632      0.062185         3    0.005594      0.4944
## mg.3:mg.6  0.123712      0.093139         2    0.005882      0.6050
## mg.5:mg.6  0.049049      0.022109         2    0.006588      0.2948
## mg.1:mg.2 -0.255147      0.347058         4    0.009010      1.1679
## mg.1:mg.5  0.022082      0.049319         3    0.009350      0.4403
## mg.4:mg.5  0.110325      0.384665         8    0.033979      1.2295
## mg.3:mg.4  0.798068      4.337448         8    0.041161      4.1287
## mg.1:mg.3 -1.417167     11.904678        10    0.052791      6.8399
## mg.4:mg.6 -1.565596      9.702828        14    0.073923      6.1751
## mg.1:mg.4  2.406234     26.096330        14    0.092711     10.1271
## mg.2        0.087386      0.108511        32    0.135941      0.6530
## mg.6        0.277767      0.725540        42    0.218103      1.6886
## mg.5        0.994225      1.963498        47    0.413344      2.7779
## mg.3       -2.153903      7.468908        57    0.485242      5.4178
## mg.1       -2.925961      7.990738        66    0.624675      5.6039
## mg.4        5.059734      6.107143        88    0.906422      4.8991

summary(asreg.result@objects[[1]])

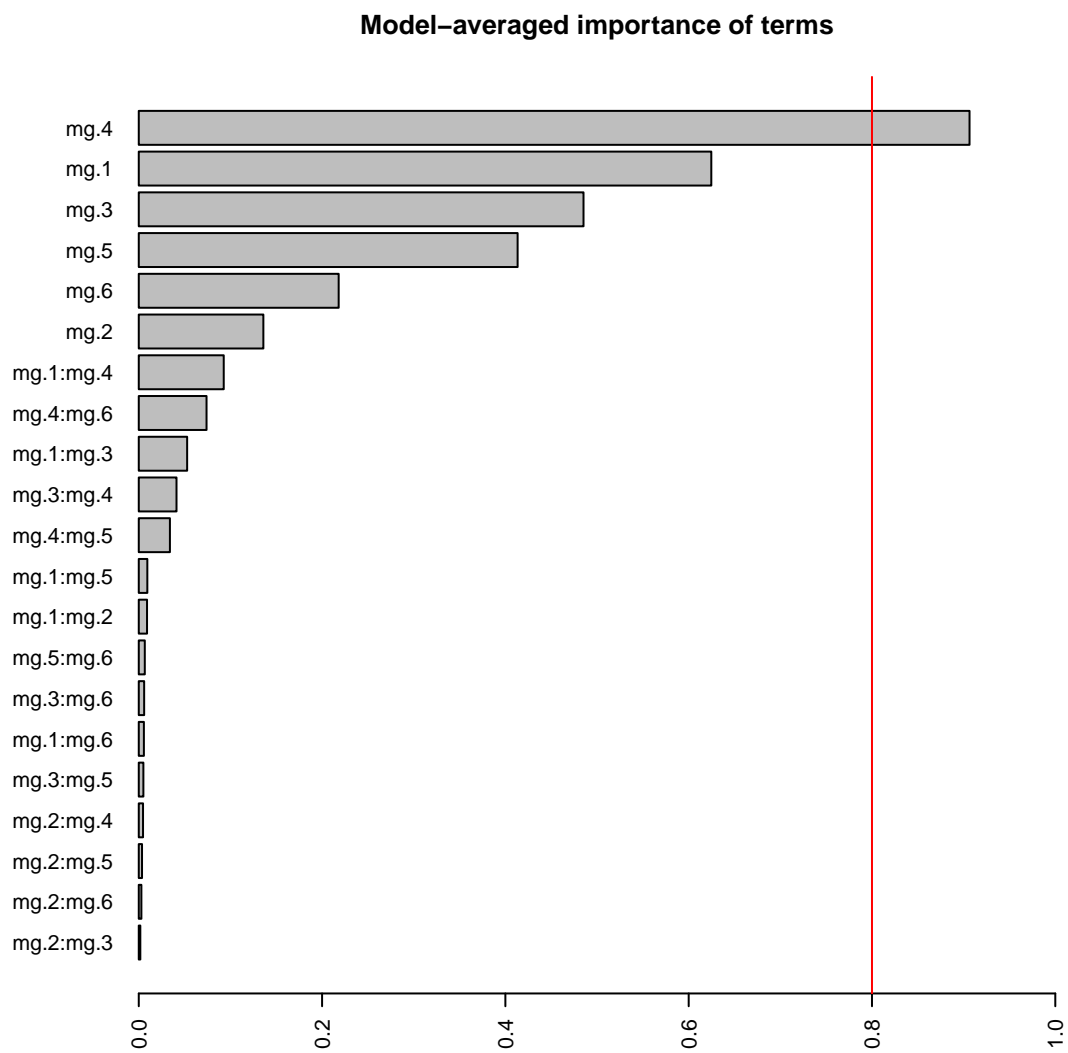
## Call:
## fitfunc(formula = as.formula(x), data = data)
##
##      n= 110, number of events= 70
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## mg.1 -4.78541   0.00835  1.66389 -2.88   0.004
## mg.3 -4.48037   0.01133  1.98473 -2.26   0.024
## mg.4  3.75726  42.83085  1.47929  2.54   0.011
##
##           exp(coef) exp(-coef) lower .95 upper .95
## mg.1   0.00835    119.7509  0.000320    0.218
## mg.3   0.01133    88.2673  0.000232    0.554
## mg.4  42.83085     0.0233  2.358218  777.910
##
## Concordance= 0.724 (se = 0.038 )
## Rsquare= 0.291 (max possible= 0.995 )
## Likelihood ratio test= 37.8 on 3 df, p=3.09e-08
## Wald test = 39.7 on 3 df, p=1.25e-08
## Score (logrank) test = 44 on 3 df, p=1.53e-09
```



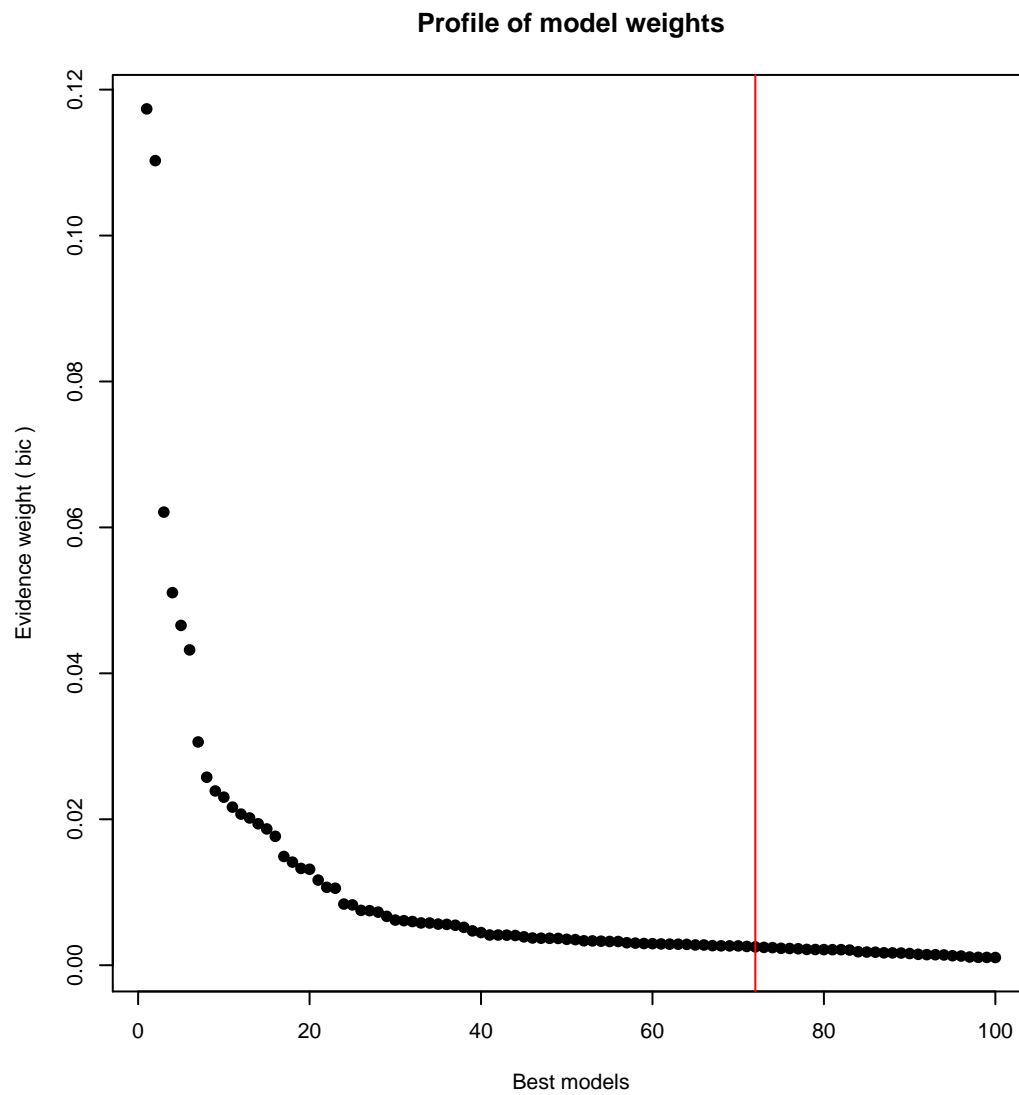
```
plot(asreg.result, type = "p")
```



```
plot(asreg.result, type = "s")
```



```
plot(asreg.result, type = "w")
```

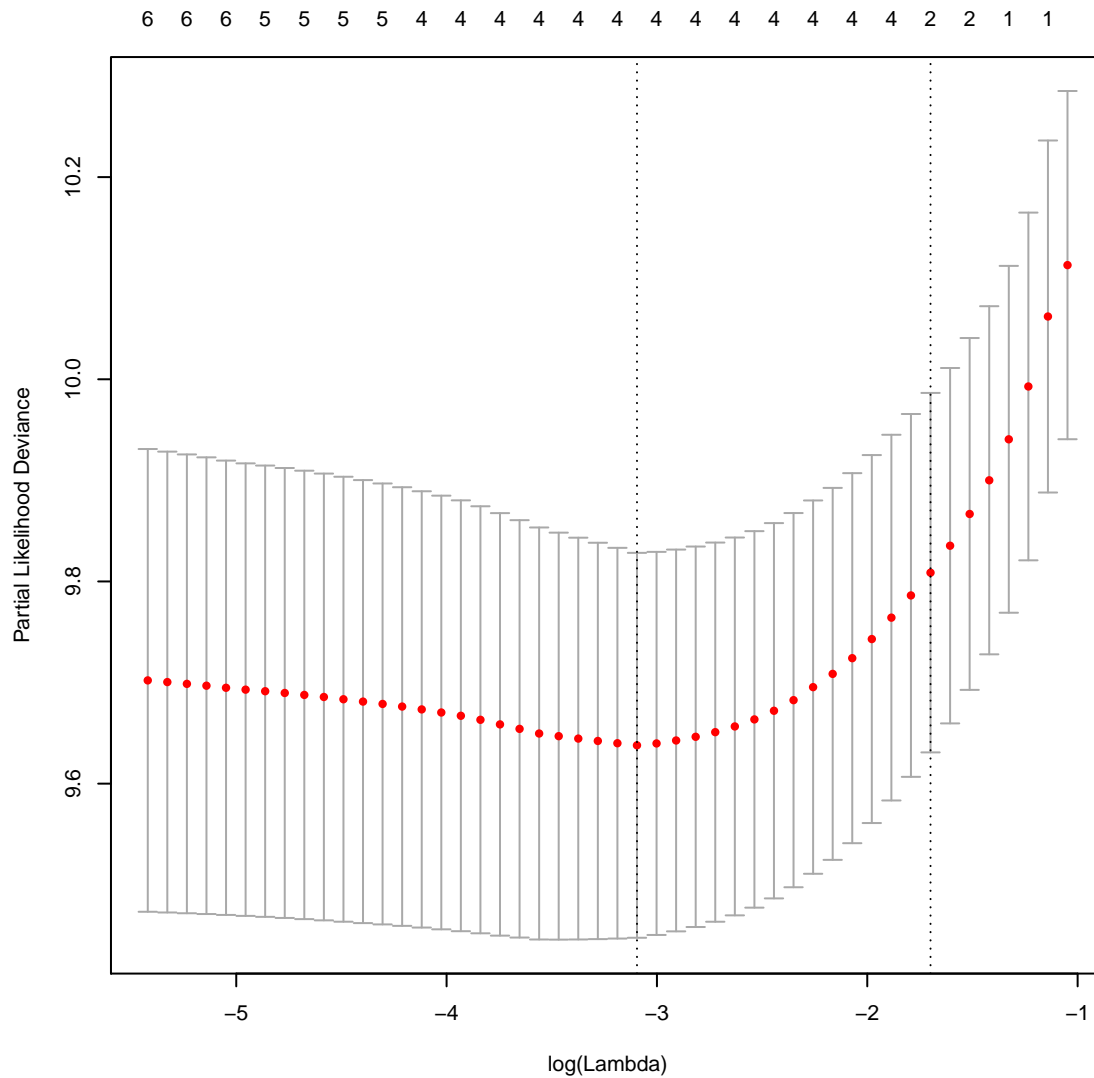


```
glmnet.coef.1se
## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1 -0.6543
## mg.2 .
## mg.3 .
## mg.4 3.6068
## mg.5 .
## mg.6 .

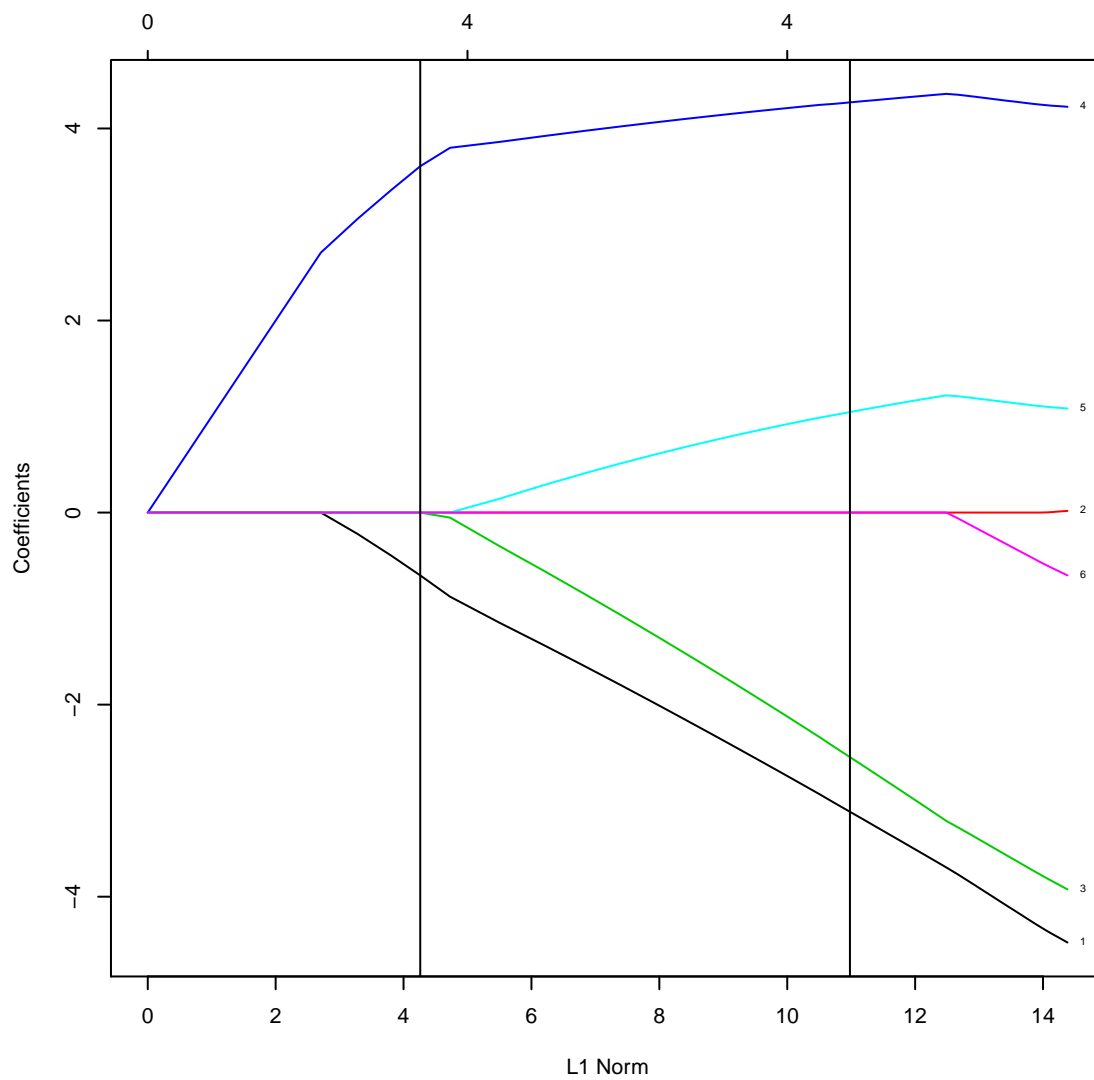
glmnet.coef.min
## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1 -3.116
## mg.2 .
## mg.3 -2.547
```

```
## mg.4 4.272
## mg.5 1.047
## mg.6 .
```

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



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

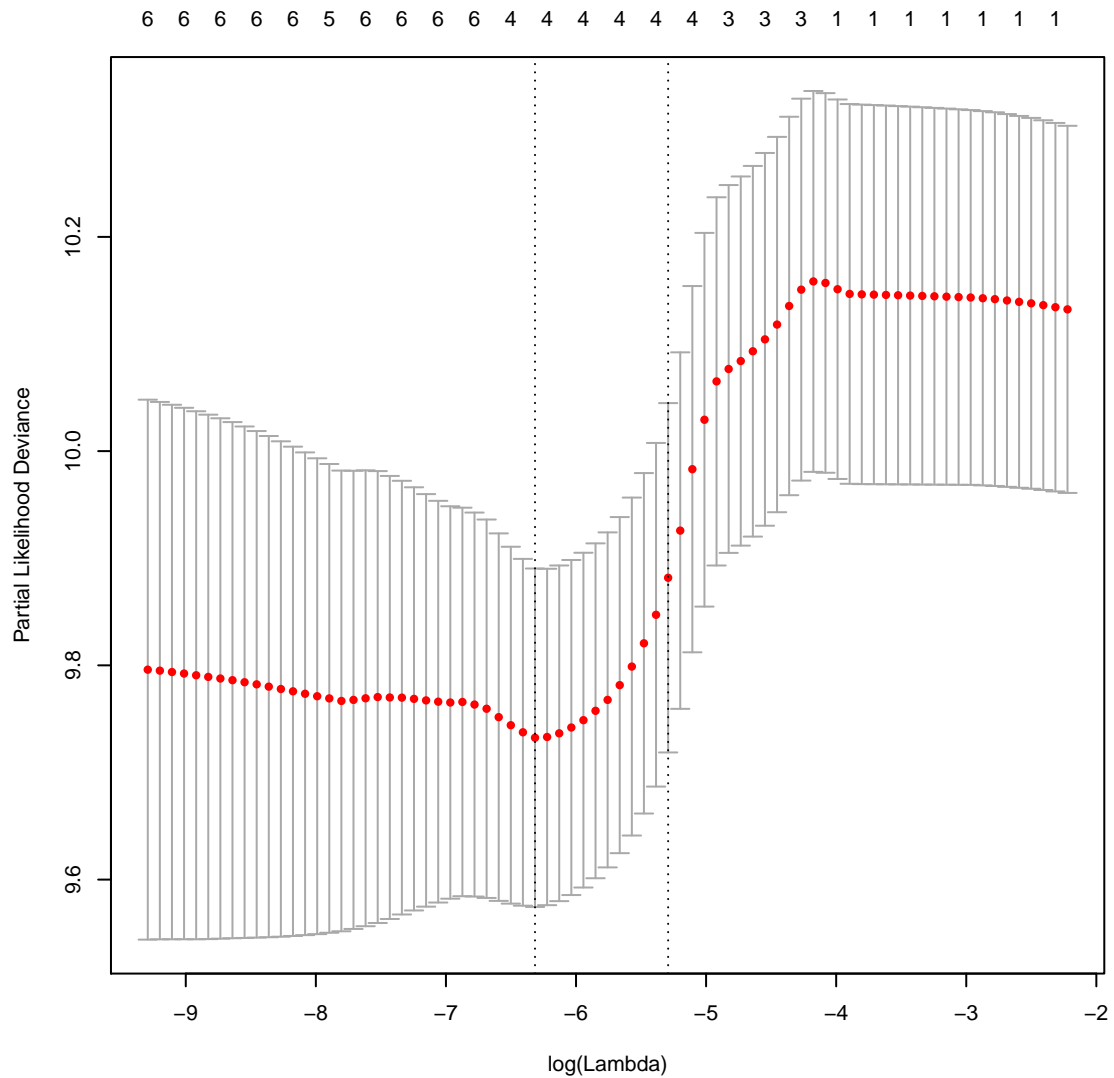


```
adaglmnet.coef.1se/adaglmnet.weights
## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1 .
## mg.2 3.314e-04
## mg.3 .
## mg.4 4.617e+01
## mg.5 2.202e+00
## mg.6 1.414e+00

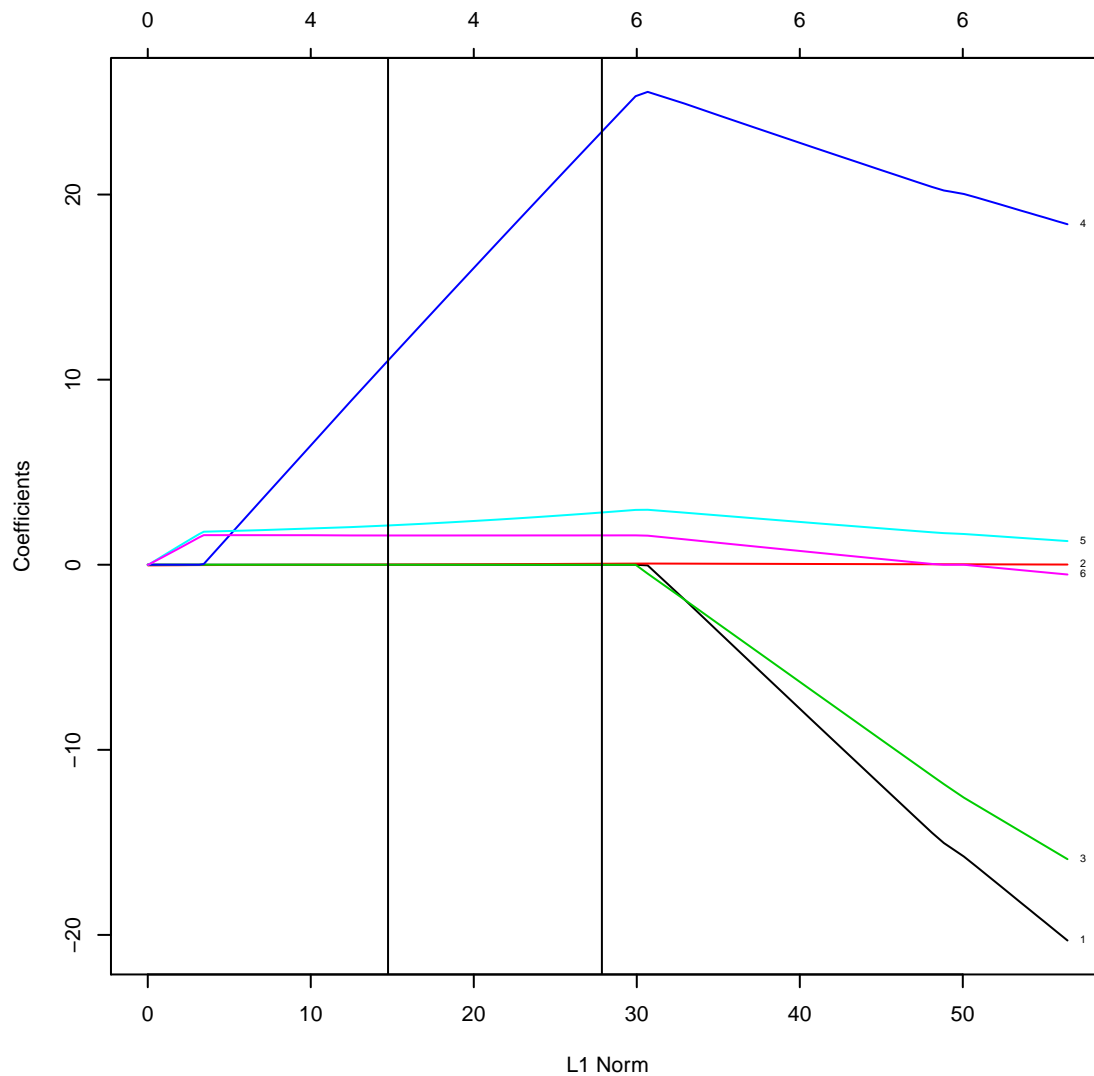
adaglmnet.coef.min/adaglmnet.weights
## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1 .
## mg.2 0.002602
## mg.3 .
```

```
## mg.4 97.923262
## mg.5 2.930487
## mg.6 1.418089
```

```
plot(adaglmnet.fit.cv)
```



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



## 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_US.iso885915      LC_NUMERIC=C
##  [3] LC_TIME=en_US.iso885915      LC_COLLATE=en_US.iso885915
##  [5] LC_MONETARY=en_US.iso885915  LC_MESSAGES=en_US.iso885915
##  [7] LC_PAPER=en_US.iso885915     LC_NAME=en_US.iso885915
##  [9] LC_ADDRESS=en_US.iso885915   LC_TELEPHONE=en_US.iso885915
## [11] LC_MEASUREMENT=en_US.iso885915 LC_IDENTIFICATION=en_US.iso885915
##
## 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      stargazer_5.1
## [7] gplots_2.14.2       NMF_0.20.4           synchronicity_1.1.4
## [10] bigmemory_4.4.6     BH_1.54.0-5          bigmemory.sri_0.1.3
## [13] Biobase_2.26.0      BiocGenerics_0.12.1  cluster_1.15.3
## [16] rngtools_1.2.4      pkgmaker_0.22        registry_0.2
## [19] RColorBrewer_1.0-5  energy_1.6.2         glmnet_1.9-8
## [22] Matrix_1.1-4        glmulti_1.0.7        rJava_0.9-6
## [25] xtable_1.7-4
##
## 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     compiler_3.1.1
## [7] digest_0.6.4        gdata_2.13.3         ggplot2_1.0.0
## [10] grid_3.1.1          gridBase_0.4-7       gtable_0.1.2
## [13] gtools_3.4.1        KernSmooth_2.23-13   lattice_0.20-29
## [16] MASS_7.3-35         munsell_0.4.2        plyr_1.8.1
## [19] proto_0.3-10        Rcpp_0.11.3          reshape2_1.4
## [22] scales_0.2.4        stringr_0.6.2        tools_3.1.1

sessionInfo()

## R version 3.1.1 (2014-07-10)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
## [1] LC_CTYPE=en_US.iso885915      LC_NUMERIC=C
## [3] LC_TIME=en_US.iso885915      LC_COLLATE=en_US.iso885915
## [5] LC_MONETARY=en_US.iso885915  LC_MESSAGES=en_US.iso885915
## [7] LC_PAPER=en_US.iso885915     LC_NAME=en_US.iso885915
## [9] LC_ADDRESS=en_US.iso885915   LC_TELEPHONE=en_US.iso885915
## [11] LC_MEASUREMENT=en_US.iso885915 LC_IDENTIFICATION=en_US.iso885915
##
## 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         NMF_0.20.4
## [10] synchronicity_1.1.4 bigmemory_4.4.6     BH_1.54.0-5
## [13] bigmemory.sri_0.1.3 Biobase_2.26.0      BiocGenerics_0.12.1
## [16] cluster_1.15.3     rngtools_1.2.4      pkgmaker_0.22
## [19] registry_0.2        energy_1.6.2         survival_2.37-7
## [22] 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
```