SIS NMF diag 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")
```
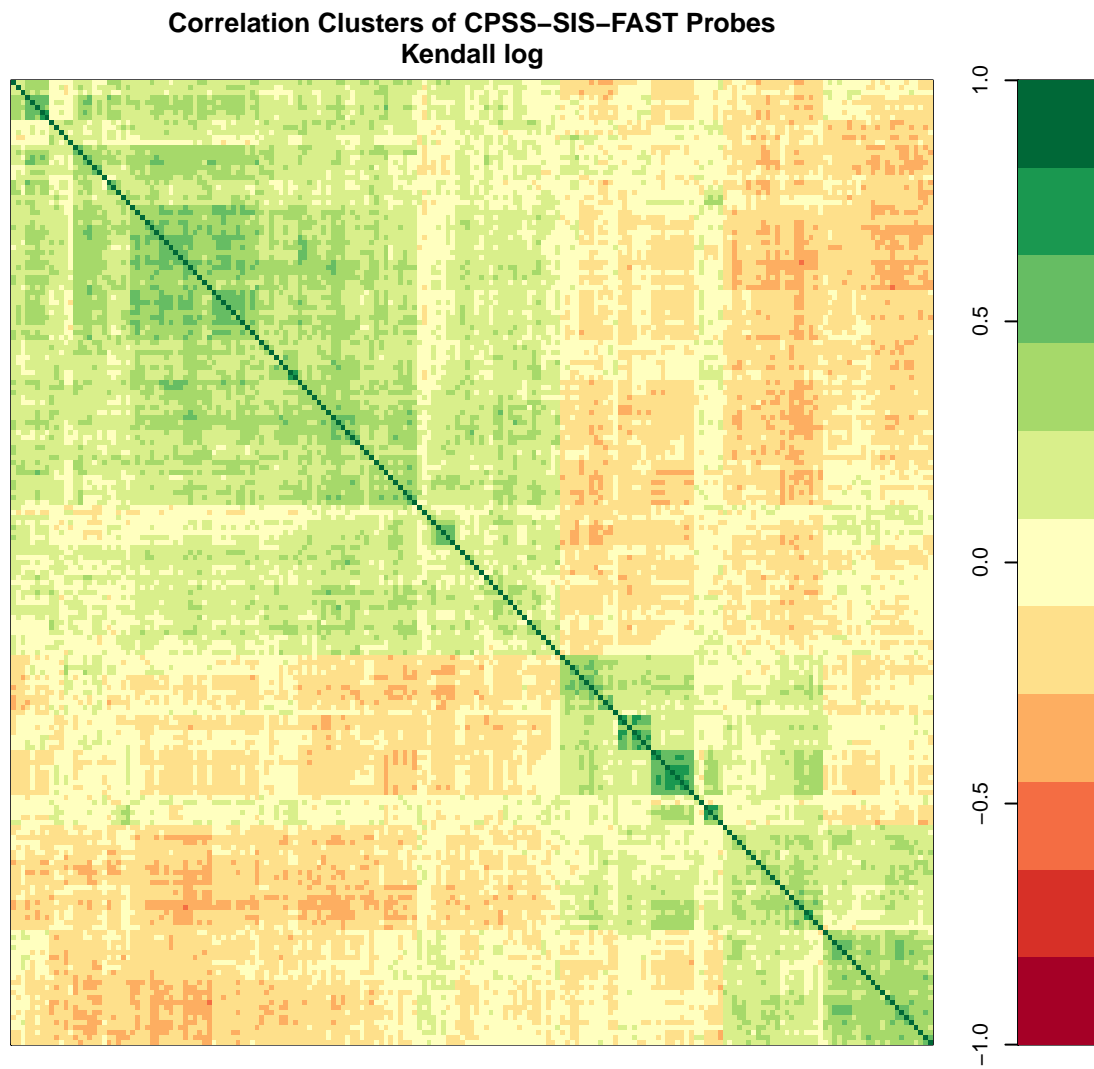
2 Probe selection

```
table(cpss.sis$sel)

##
## FALSE   TRUE
## 12807    193

mean(cpss.sis$sel)

## [1] 0.01485
```
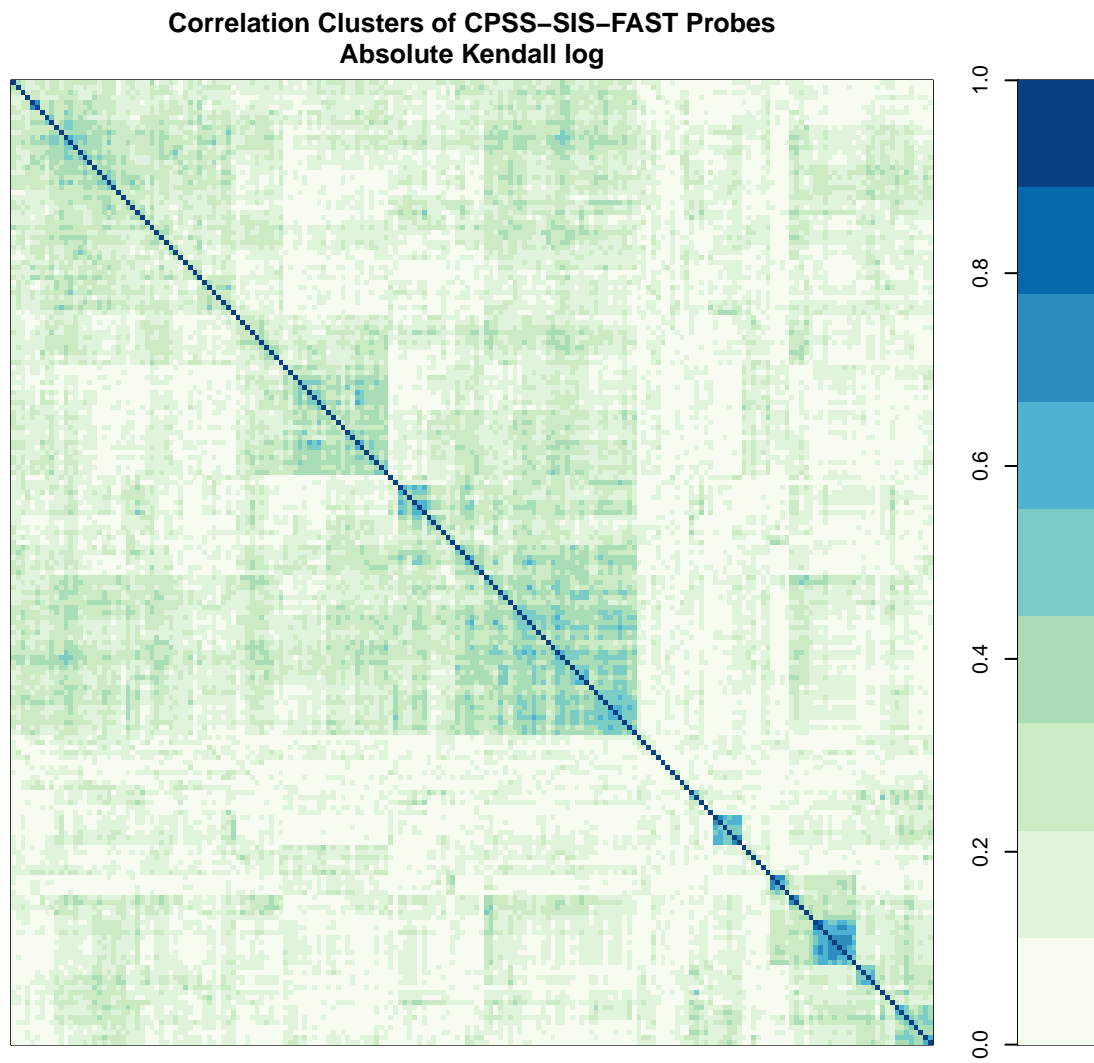
3 Expression correlation

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

Correlation Clusters of CPSS–SIS–FAST Probes
Kendall log


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

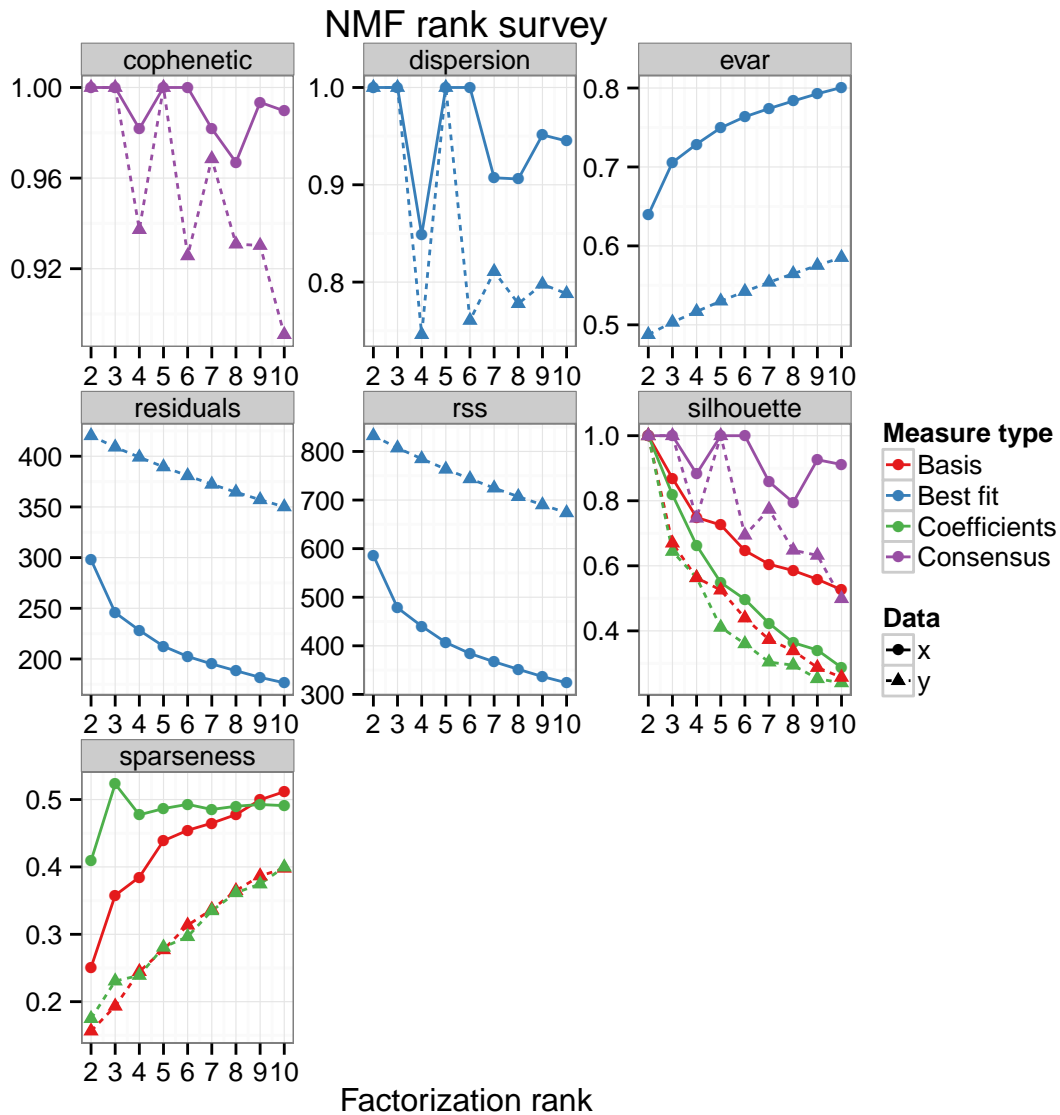## Correlation Clusters of CPSS–SIS–FAST Probes
Absolute Kendall log


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

Correlation Clusters of CPSS–SIS–FAST Probes
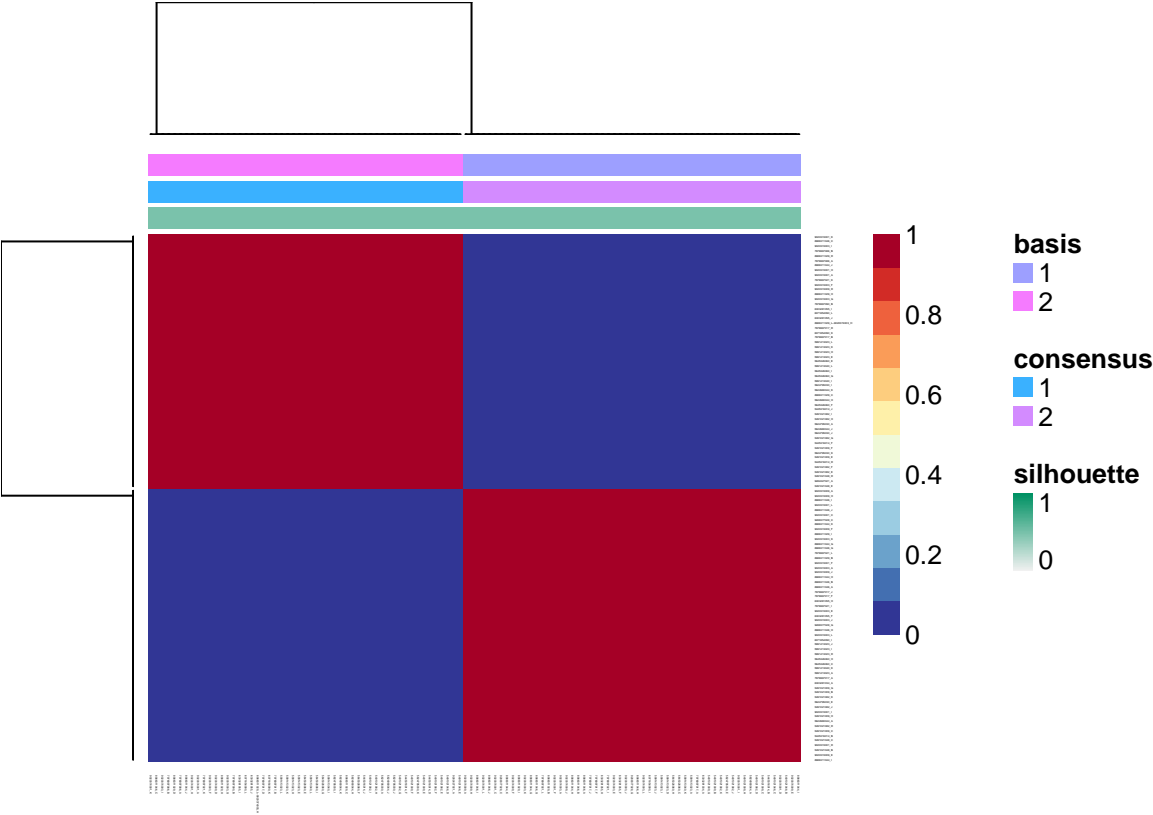**dcor log**
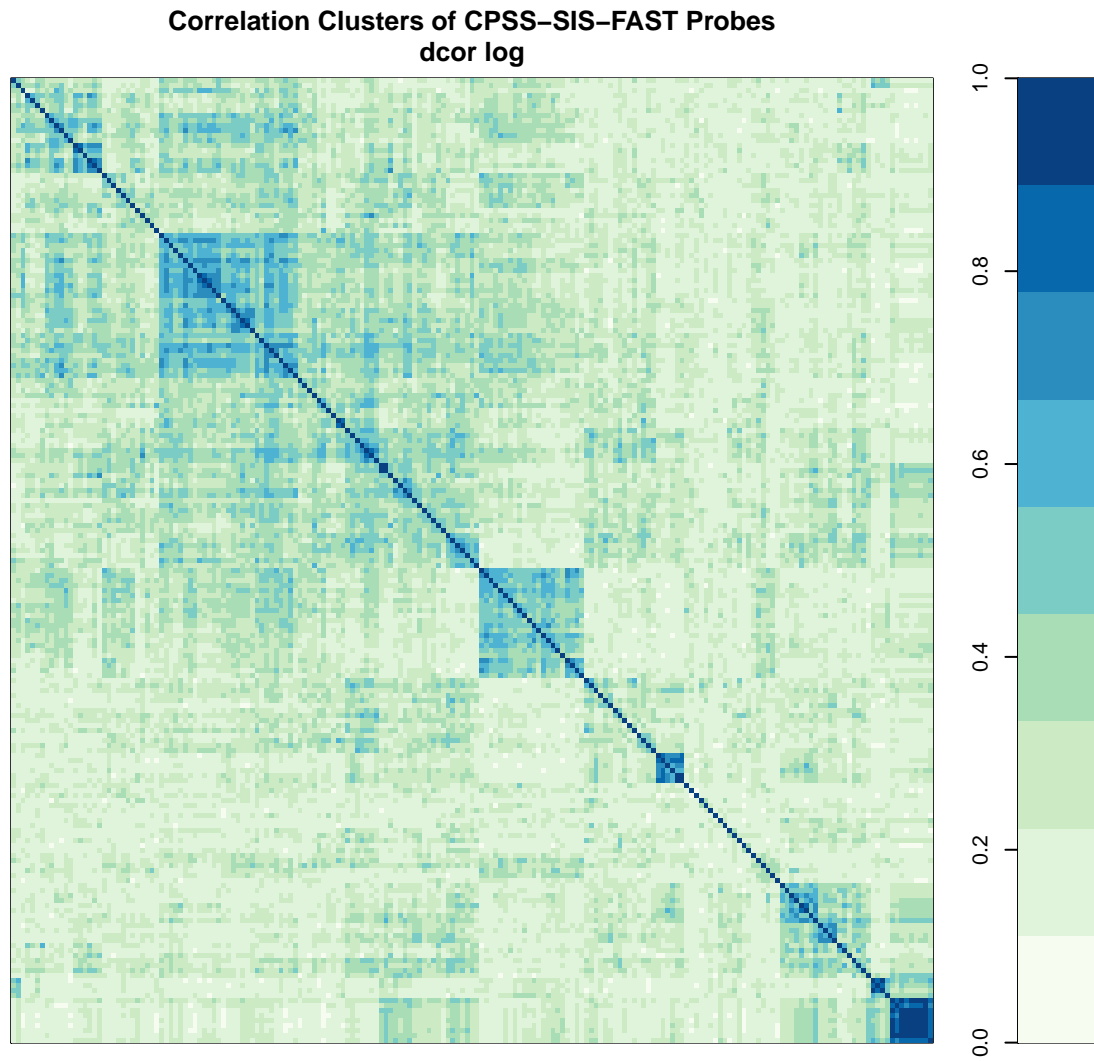
4 Factorization

4.1 Rank estimation

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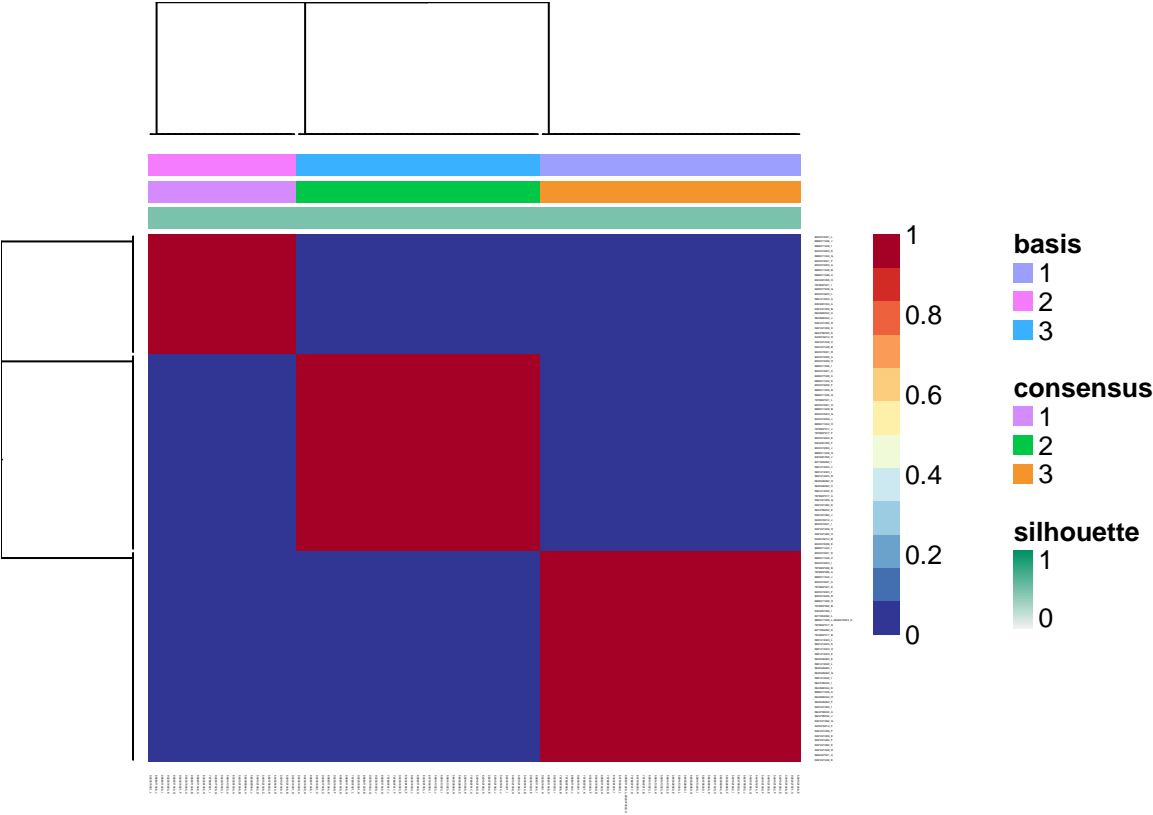
NMF rank survey

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

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

Consensus matrix

basis
1
2
3
4
5
6
7
8
9
10

consensus
1
2
3
4
5
6
7
8
9
10

silhouette
1
0.65
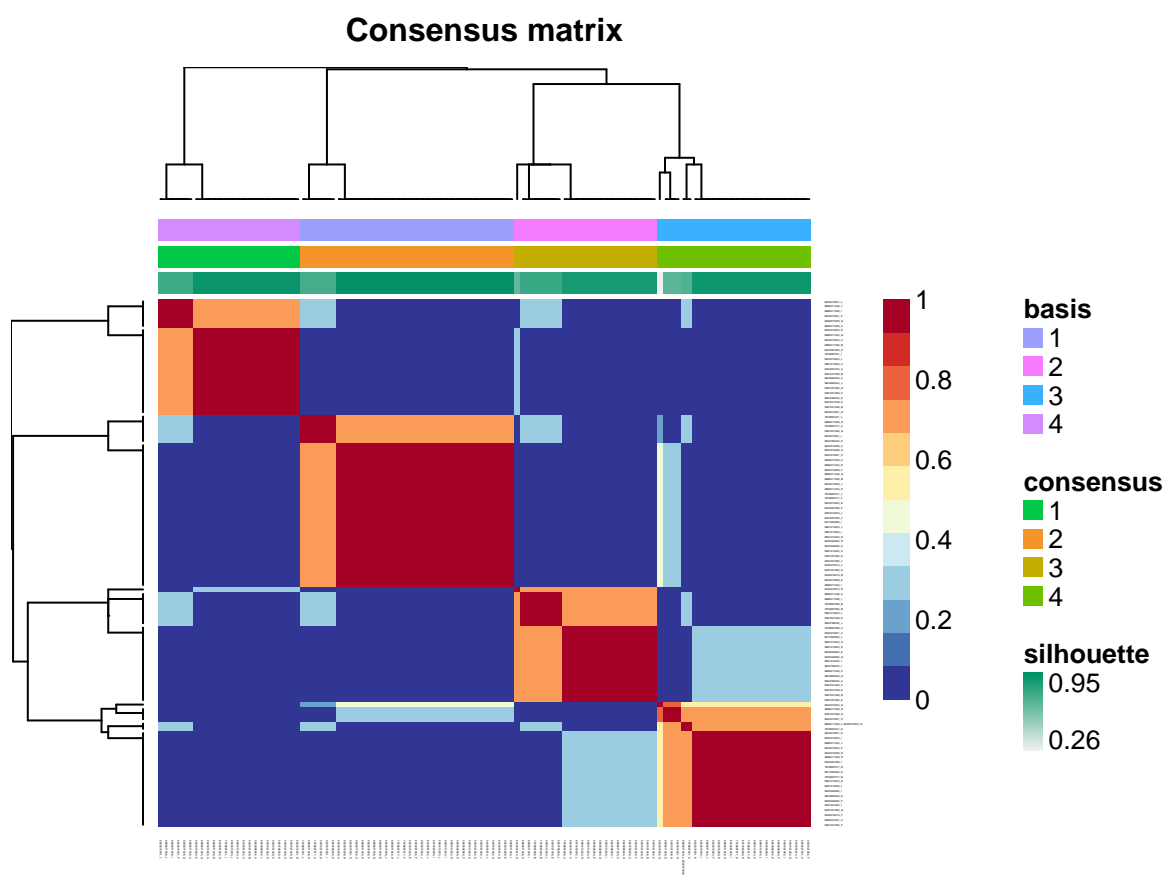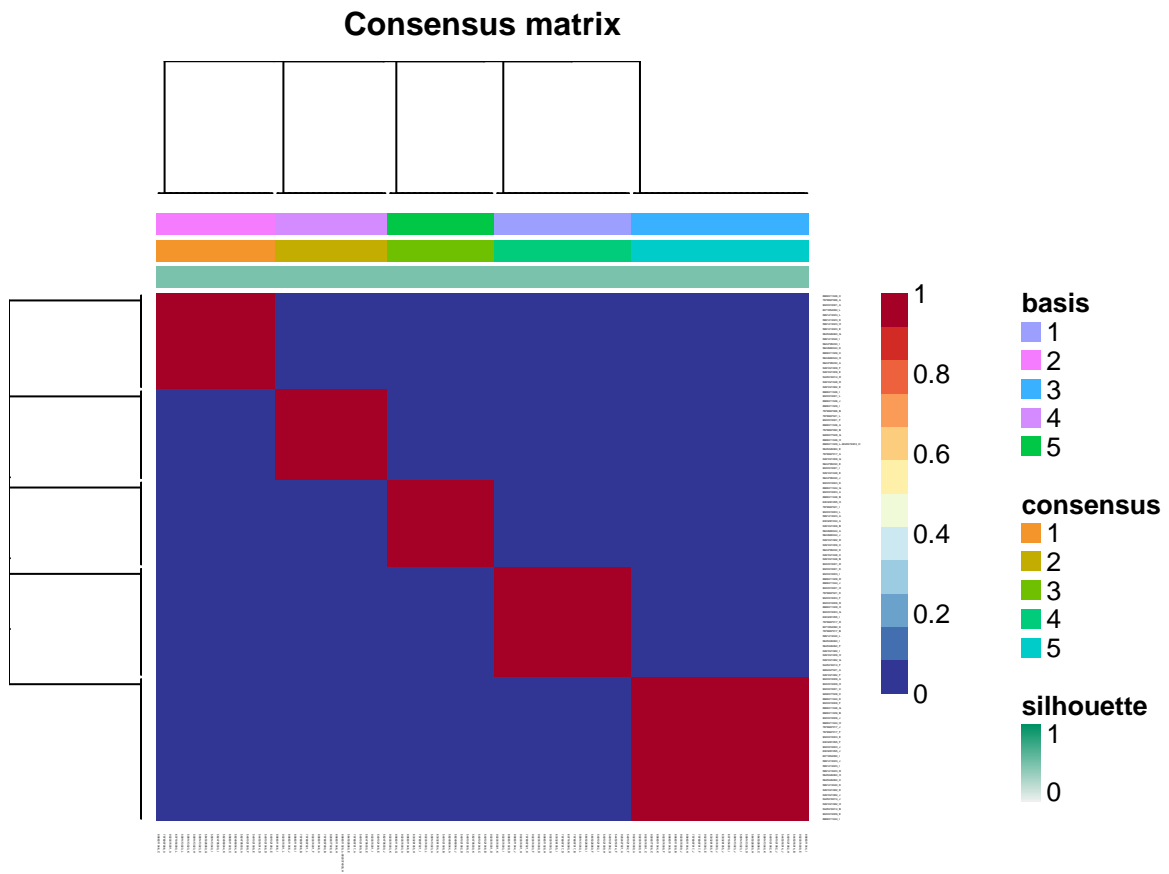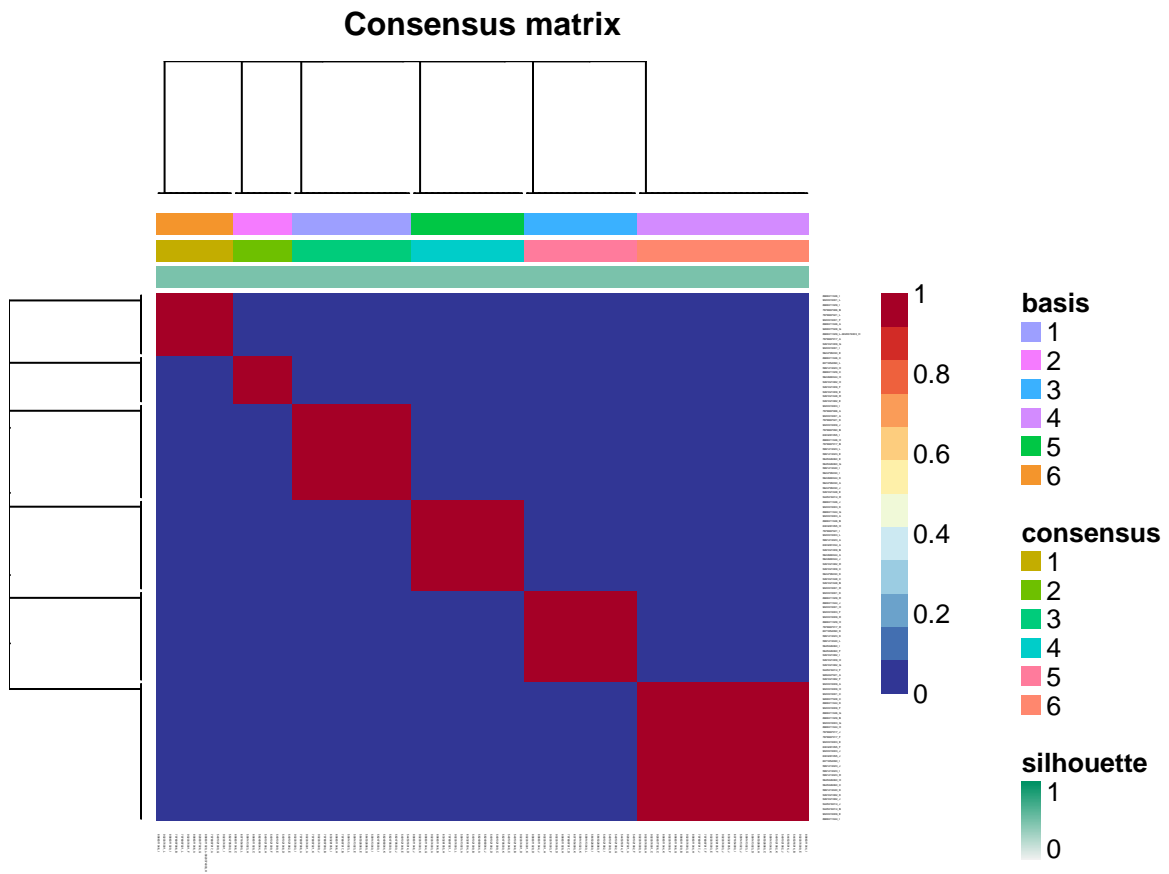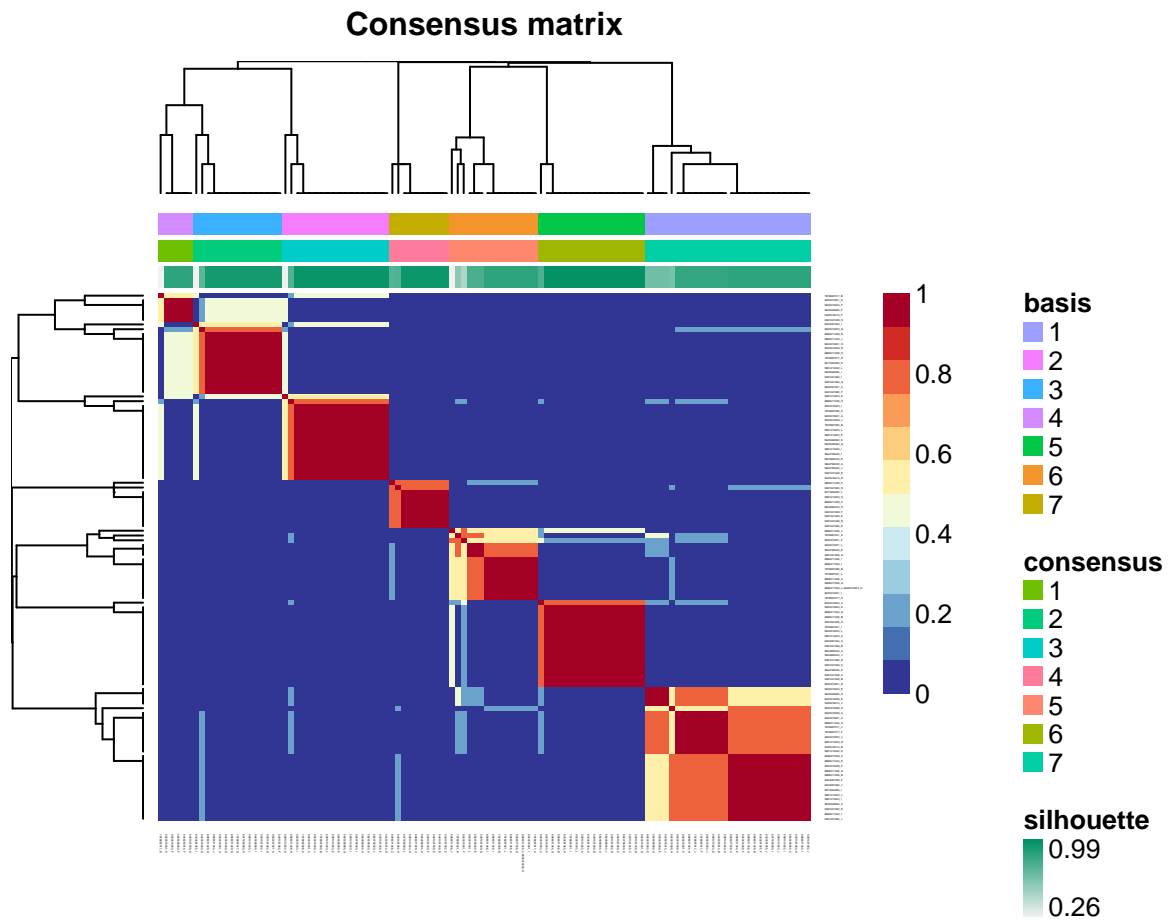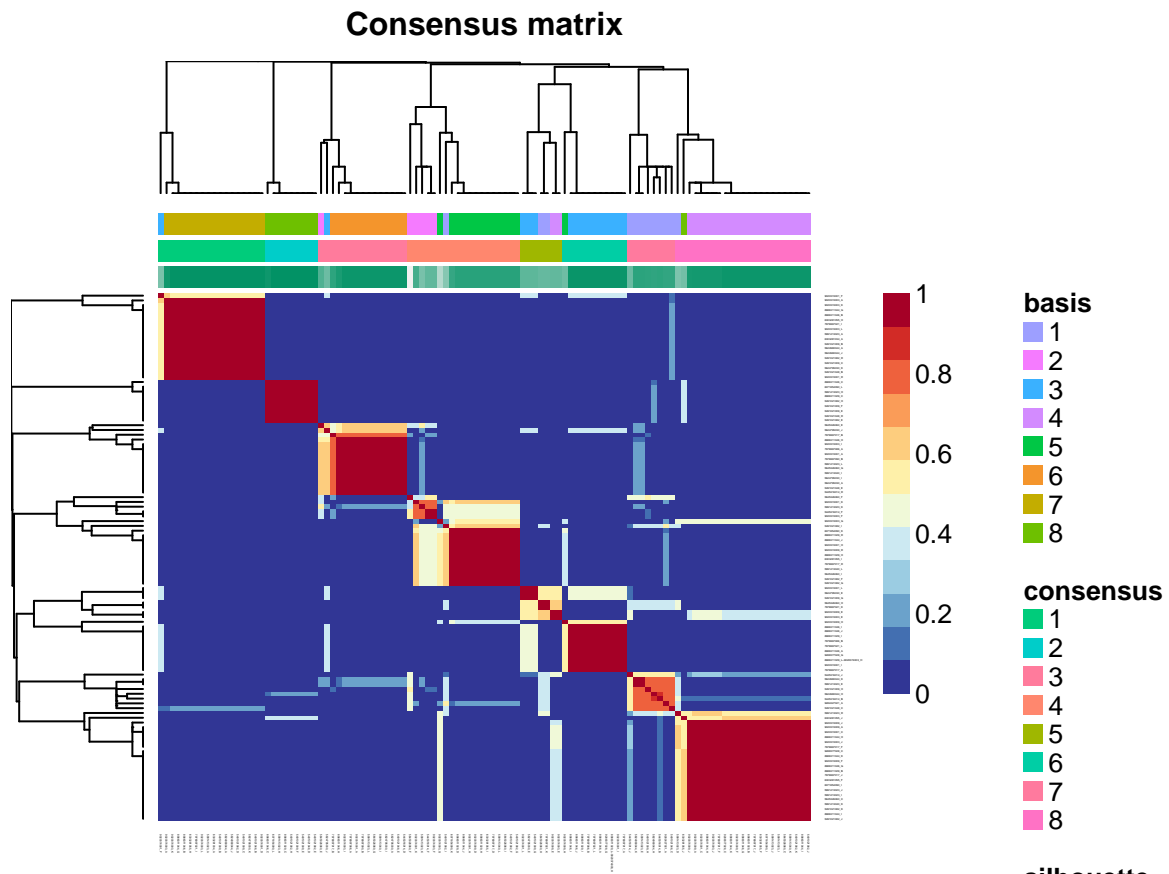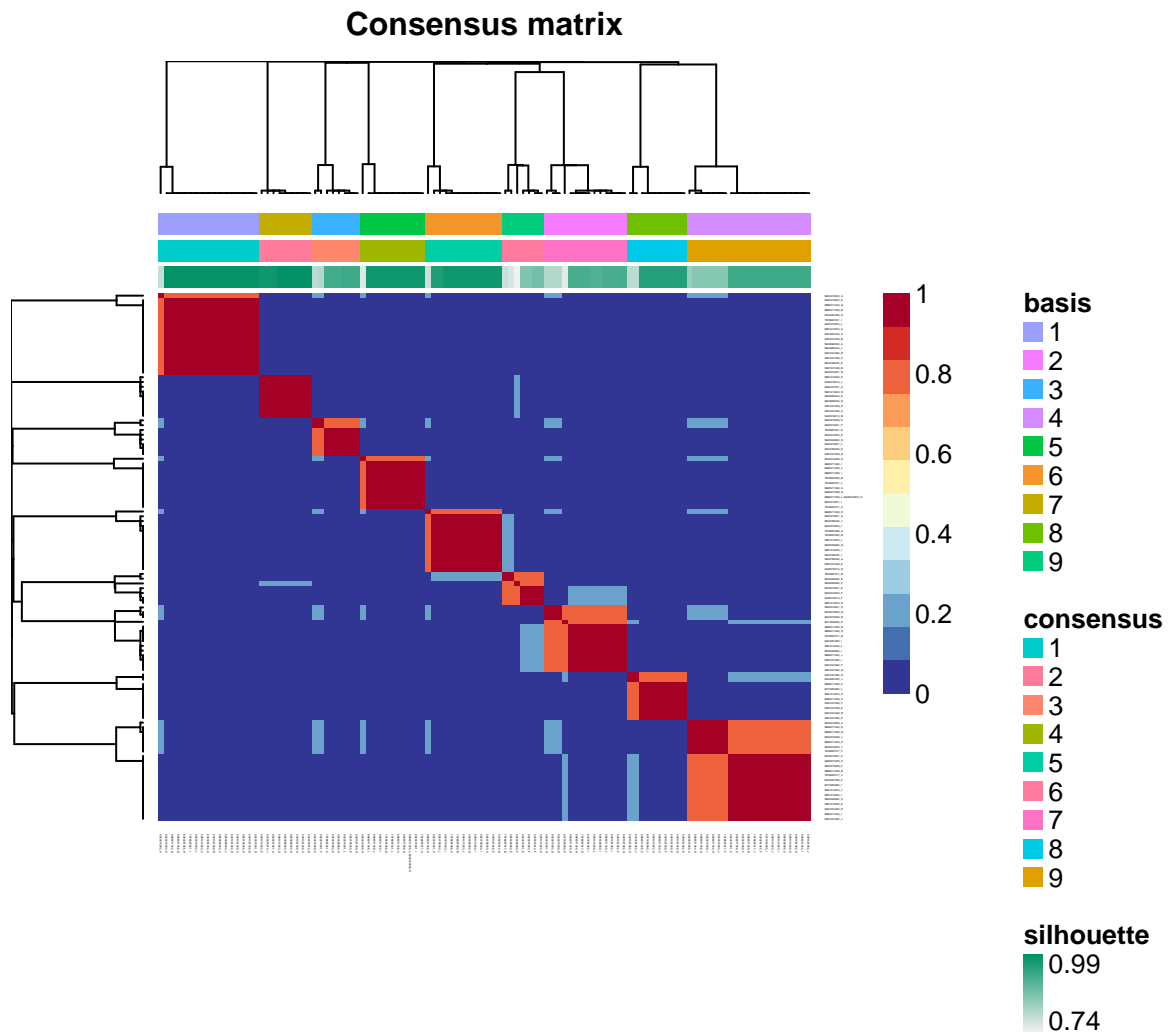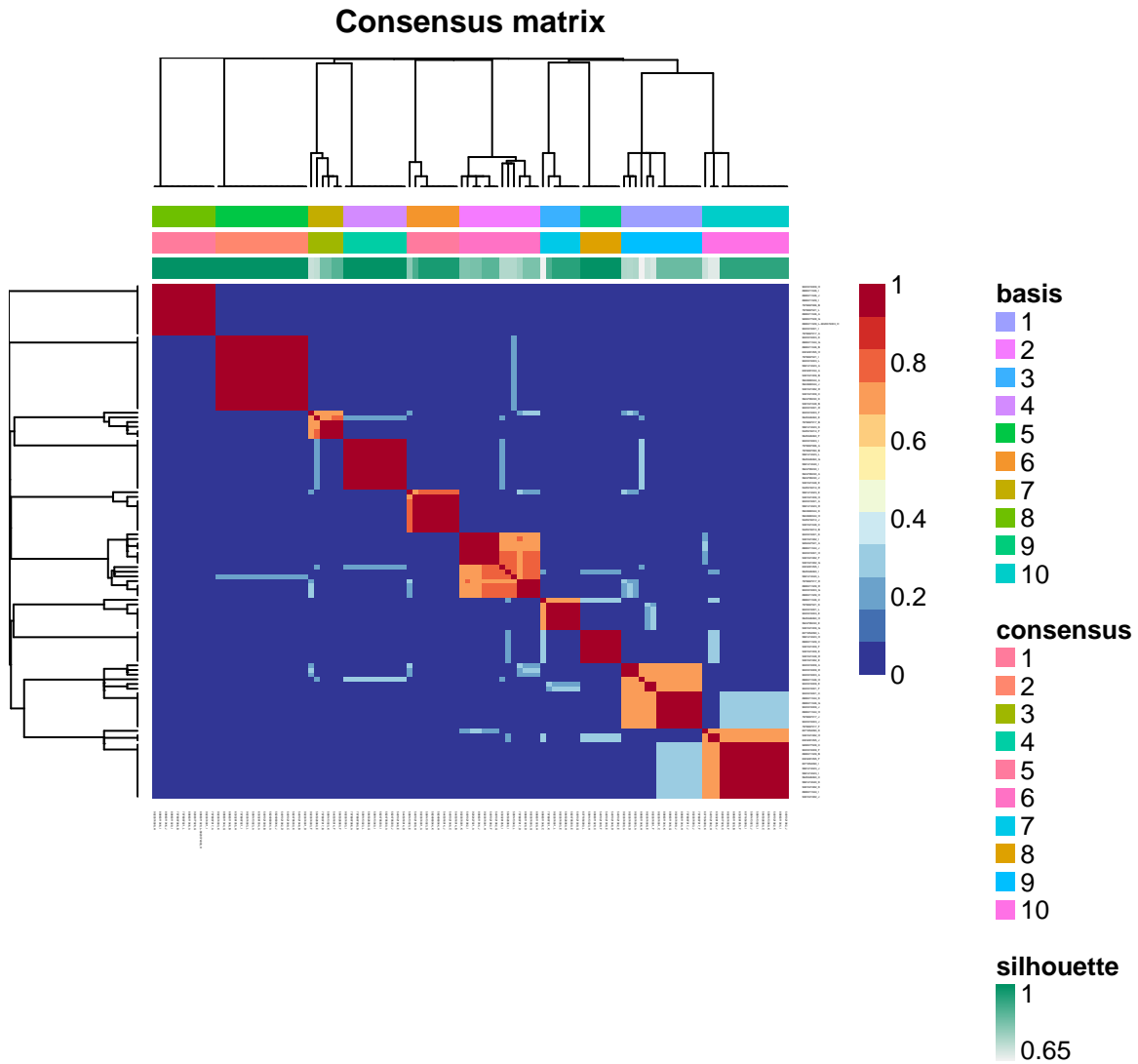
```
plot(0 ~ 0, type = "n", xlim = c(1, nrow(temp.resids)), ylim = range(temp.resids_rel),
    ylab = "Relative residual", main = "Solution Stability")
for (i in 1:ncol(temp.resids)) {
    points(temp.resids_rel[, i], col = i, pch = colnames(temp.resids)[i])
    lines(cummin(temp.resids_rel[, i]), col = i)
}
```

Solution Stability
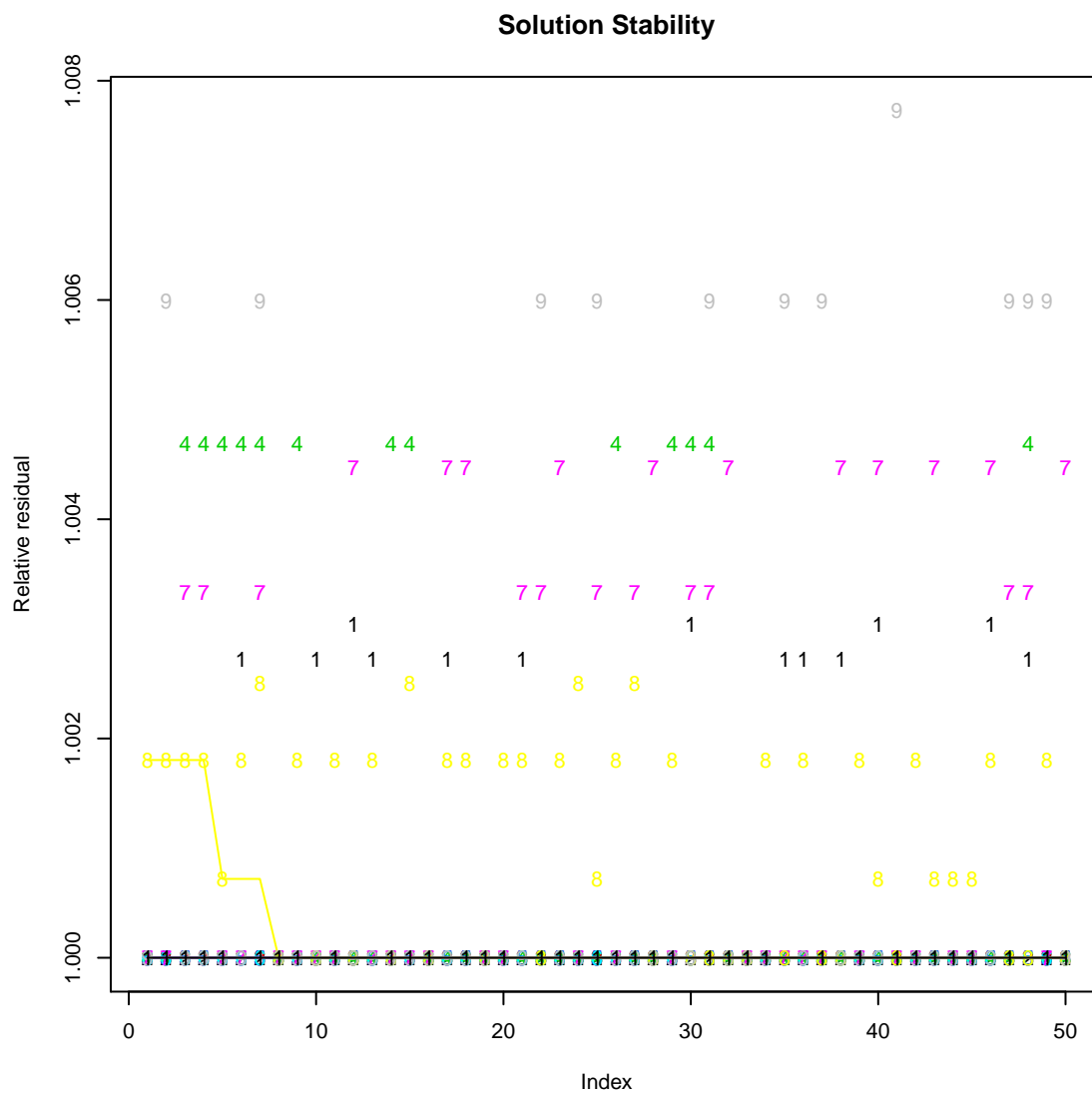

```r
plot(0 ~ 0, type = "n", xlim = c(1, nrow(temp.resids)), ylim = range(temp.resids_scaled),
    ylab = "Scaled residual", main = "Solution Stability")
for (i in 1:ncol(temp.resids)) {
    points(temp.resids_scaled[, i], col = i, pch = colnames(temp.resids)[i])
    lines(cummin(temp.resids_scaled[,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

Solution Stability
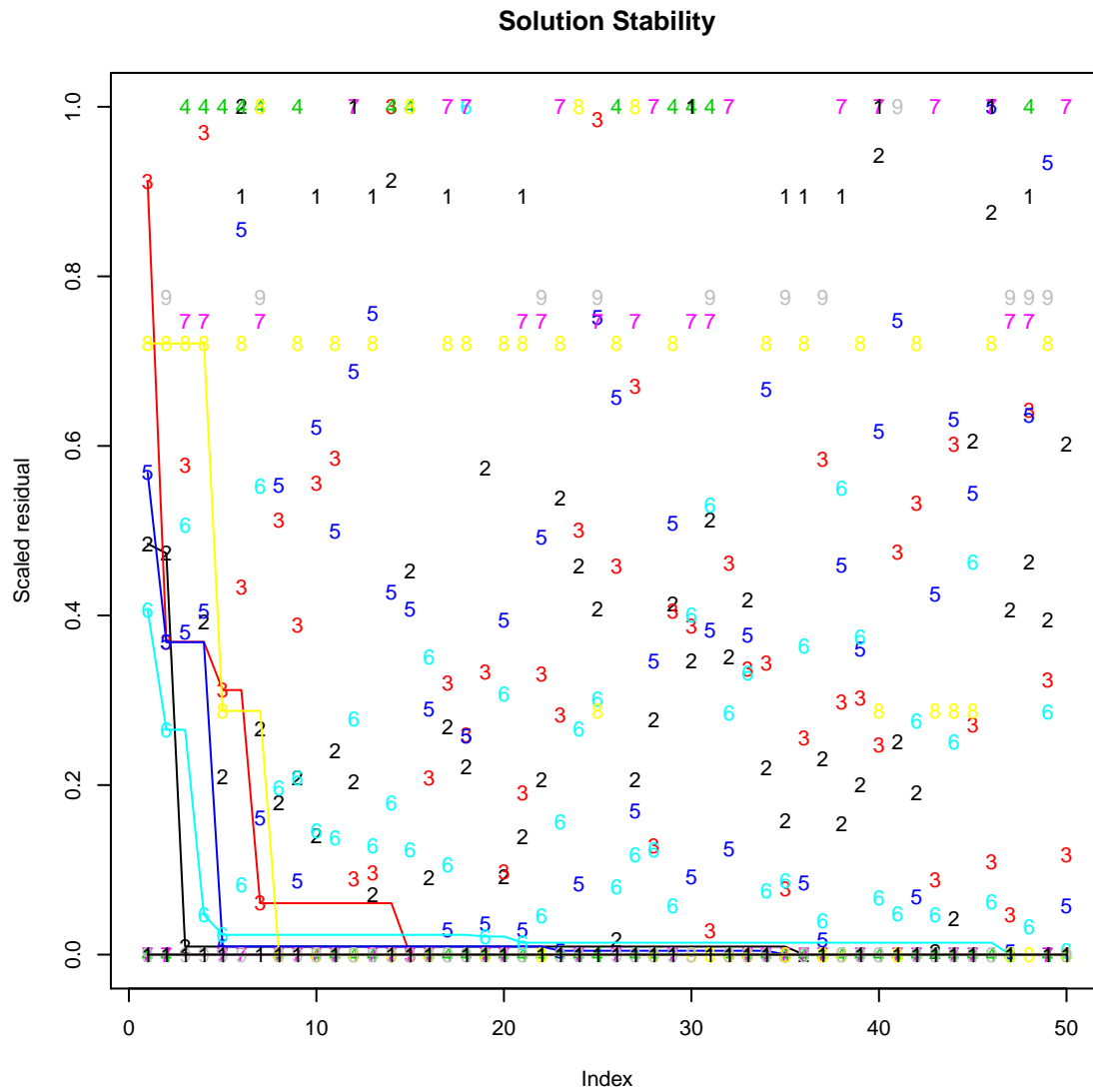
```r
plot(nmf.rankrange[-1], -temp.orig_resids.delta, type = "o", col = "black",
    pch = 21, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.delta.mean)),
    xlab = "Factorization Rank Added", ylab = "Improvement in Total Residual Error")
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean, 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)
```
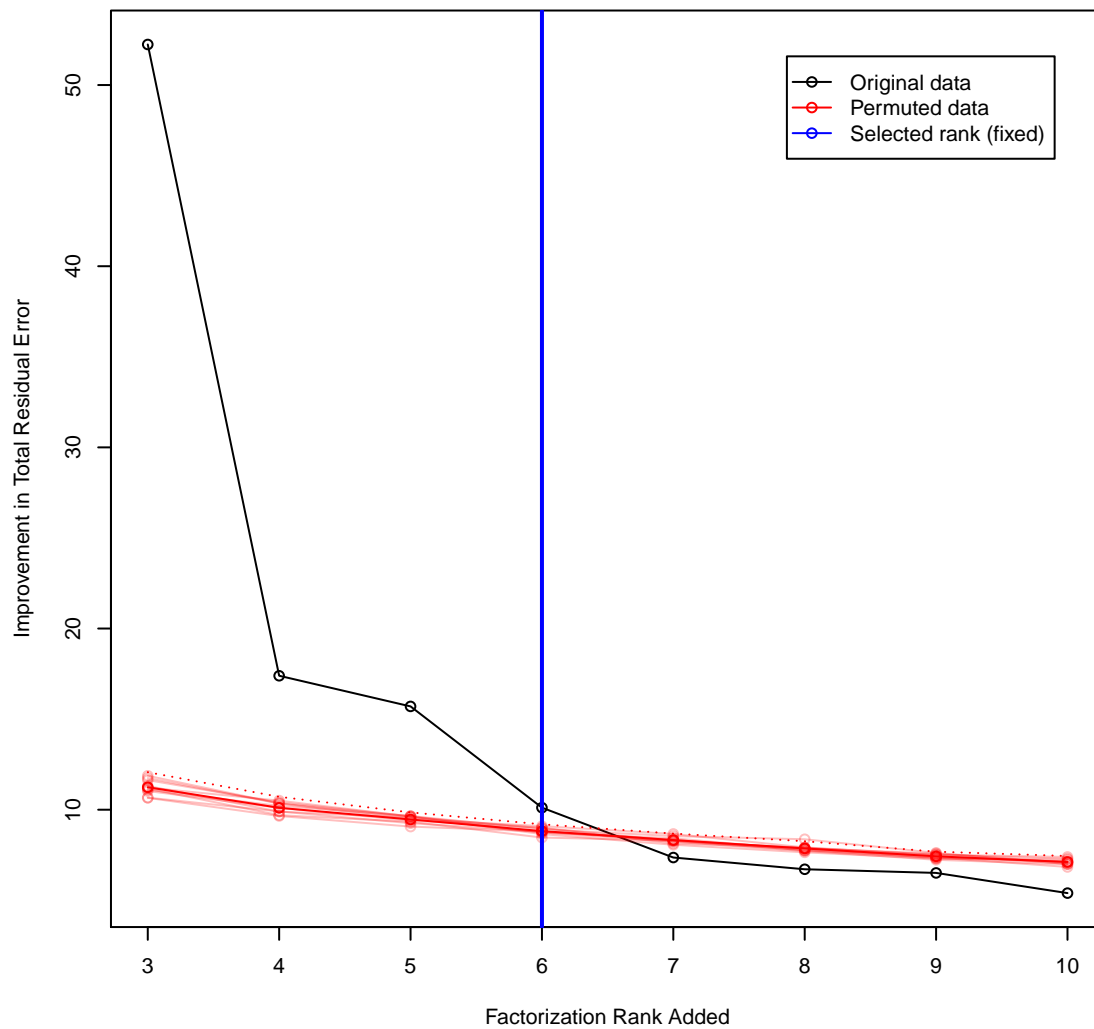

4.2 Fit

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

Consensus matrix


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

Basis components
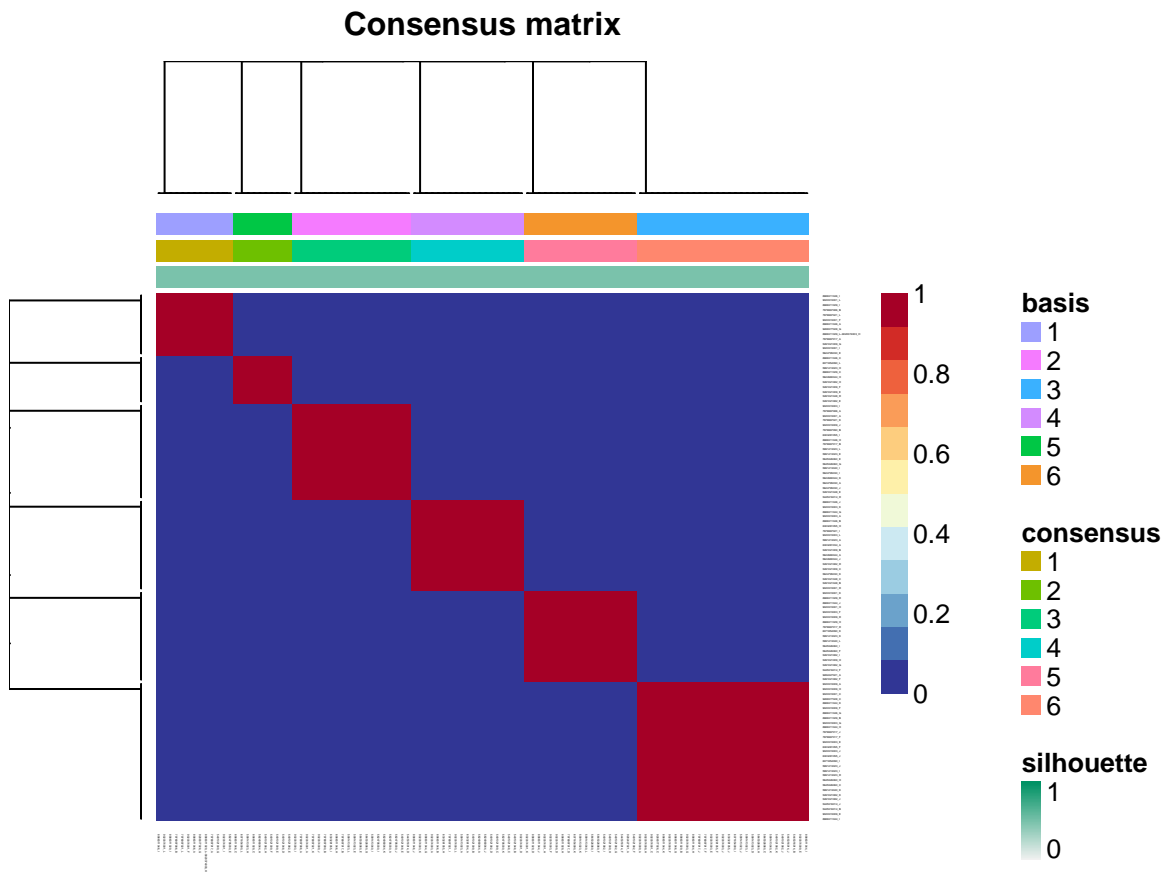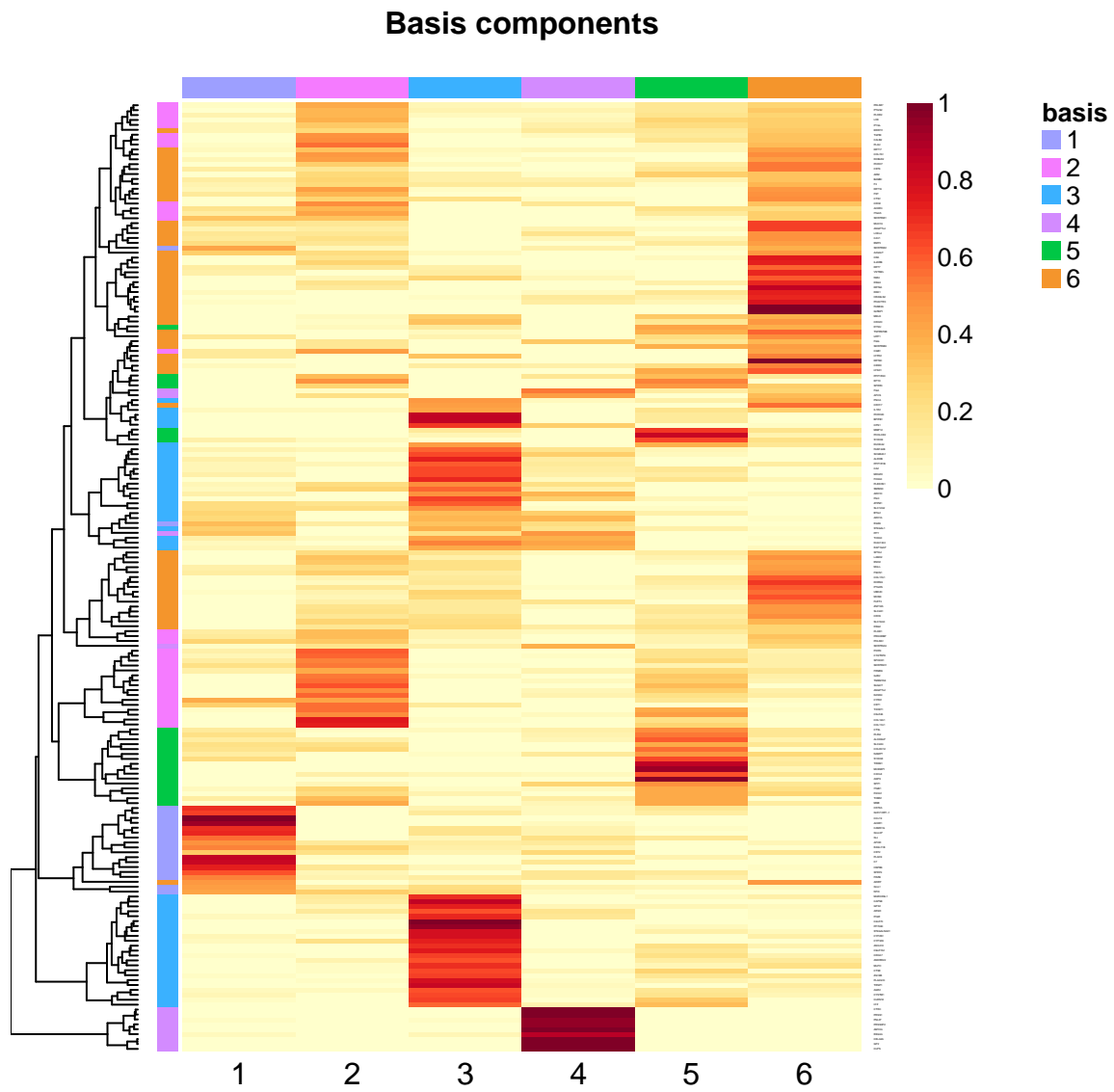

```
coefmap(xlin.scaled.sel.nmf)
```

Mixture coefficients
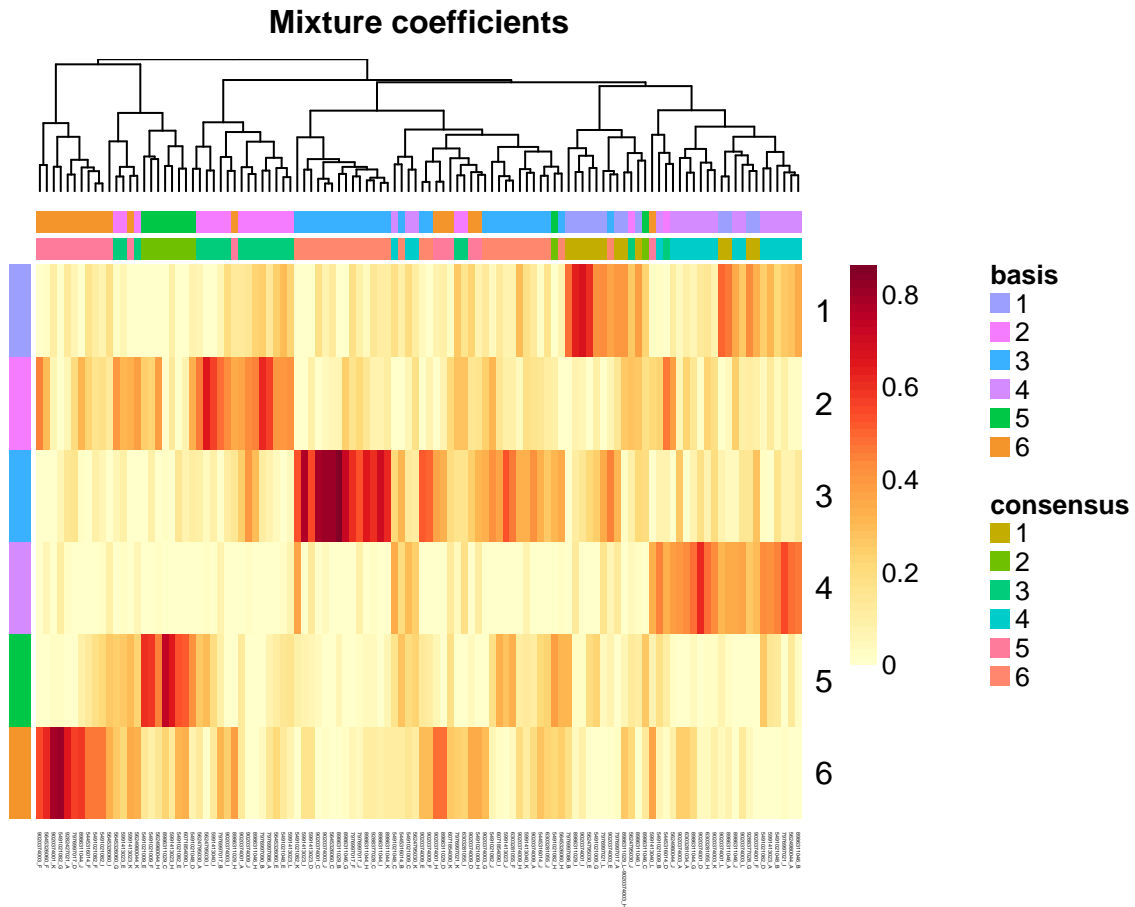

```
temp.resids = sapply(xlin.scaled.sel.nmf, residuals)
plot(1:length(temp.resids), temp.resids, ylab = "Residual", main = "Solution Stability")
lines(1:length(temp.resids), cummin(temp.resids))
```

Solution Stability

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, ] -4.65441   0.00952  1.72954 -2.69   0.0071
##
##                                 exp(coef) exp(-coef) lower .95 upper .95
```

Table 1:

	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	−4.654***
	(1.730)
Observations	110
R^2	0.074
Max. Possible R^2	0.995
Log Likelihood	−284.300
Wald Test	7.240*** (df = 1)
LR Test	8.445*** (df = 1)
Score (Logrank) Test	7.472*** (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2:

	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	3.203**
	(1.295)
Observations	110
R^2	0.050
Max. Possible R^2	0.995
Log Likelihood	−285.700
Wald Test	6.110** (df = 1)
LR Test	5.597** (df = 1)
Score (Logrank) Test	6.233** (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3:

	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	−5.529***
	(1.558)
Observations	110
R^2	0.127
Max. Possible R^2	0.995
Log Likelihood	−281.000
Wald Test	12.600*** (df = 1)
LR Test	14.950*** (df = 1)
Score (Logrank) Test	13.160*** (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4:

	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	−0.970
	(1.283)
Observations	110
R^2	0.005
Max. Possible R^2	0.995
Log Likelihood	−288.200
Wald Test	0.570 (df = 1)
LR Test	0.594 (df = 1)
Score (Logrank) Test	0.573 (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5:

	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	2.937***
	(1.110)
Observations	110
R^2	0.053
Max. Possible R^2	0.995
Log Likelihood	−285.500
Wald Test	7.000*** (df = 1)
LR Test	6.014** (df = 1)
Score (Logrank) Test	7.178*** (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6:

|	Dependent variable:
	y
coef(xlin.scaled.sel.nmf)[i,]	6.743***
	(1.141)
Observations	110
R^2	0.225
Max. Possible R^2	0.995
Log Likelihood	−274.500
Wald Test	34.930*** (df = 1)
LR Test	28.090*** (df = 1)
Score (Logrank) Test	38.050*** (df = 1)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
## coef(xlin.scaled.sel.nmf)[i, ]   0.00952        105  0.000321      0.282
##
## Concordance= 0.603  (se = 0.038 )
## Rsquare= 0.074   (max possible= 0.995 )
## Likelihood ratio test= 8.45  on 1 df,   p=0.00366
## Wald test            = 7.24  on 1 df,   p=0.00712
## Score (logrank) test = 7.47  on 1 df,   p=0.00627
##
## 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      24.6      1.3 2.47    0.013
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]      24.6     0.0407      1.94       312
##
## Concordance= 0.578  (se = 0.038 )
## Rsquare= 0.05   (max possible= 0.995 )
## Likelihood ratio test= 5.6  on 1 df,   p=0.018
## Wald test            = 6.11  on 1 df,   p=0.0134
## Score (logrank) test = 6.23  on 1 df,   p=0.0125
##
## 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, ] -5.52933   0.00397  1.55793 -3.55  0.00039
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]   0.00397        252  0.000187    0.0841
##
## Concordance= 0.636  (se = 0.038 )
## Rsquare= 0.127   (max possible= 0.995 )
## Likelihood ratio test= 14.9  on 1 df,   p=0.00011
## Wald test            = 12.6  on 1 df,   p=0.000386
## Score (logrank) test = 13.2  on 1 df,   p=0.000286
##
## 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, ] -0.970     0.379    1.283 -0.76     0.45
##
##                                exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]     0.379       2.64    0.0307      4.69
##
```

```
## Concordance= 0.553  (se = 0.037 )
## Rsquare= 0.005   (max possible= 0.995 )
## Likelihood ratio test= 0.59  on 1 df,    p=0.441
## Wald test             = 0.57  on 1 df,    p=0.45
## Score (logrank) test = 0.57  on 1 df,    p=0.449
##
## 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, ]  2.94     18.86     1.11 2.65   0.0081
##
##                                  exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]       18.9      0.053      2.14       166
##
## Concordance= 0.556  (se = 0.037 )
## Rsquare= 0.053   (max possible= 0.995 )
## Likelihood ratio test= 6.01  on 1 df,    p=0.0142
## Wald test             = 7  on 1 df,   p=0.00814
## Score (logrank) test = 7.18  on 1 df,    p=0.00738
##
## 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, ]  6.74    848.38     1.14 5.91 3.4e-09
##
##                                  exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]       848    0.00118      90.6      7940
##
## Concordance= 0.701  (se = 0.037 )
## Rsquare= 0.225   (max possible= 0.995 )
## Likelihood ratio test= 28.1  on 1 df,    p=1.16e-07
## Wald test             = 34.9  on 1 df,   p=3.43e-09
## Score (logrank) test = 38  on 1 df,    p=6.89e-10
```

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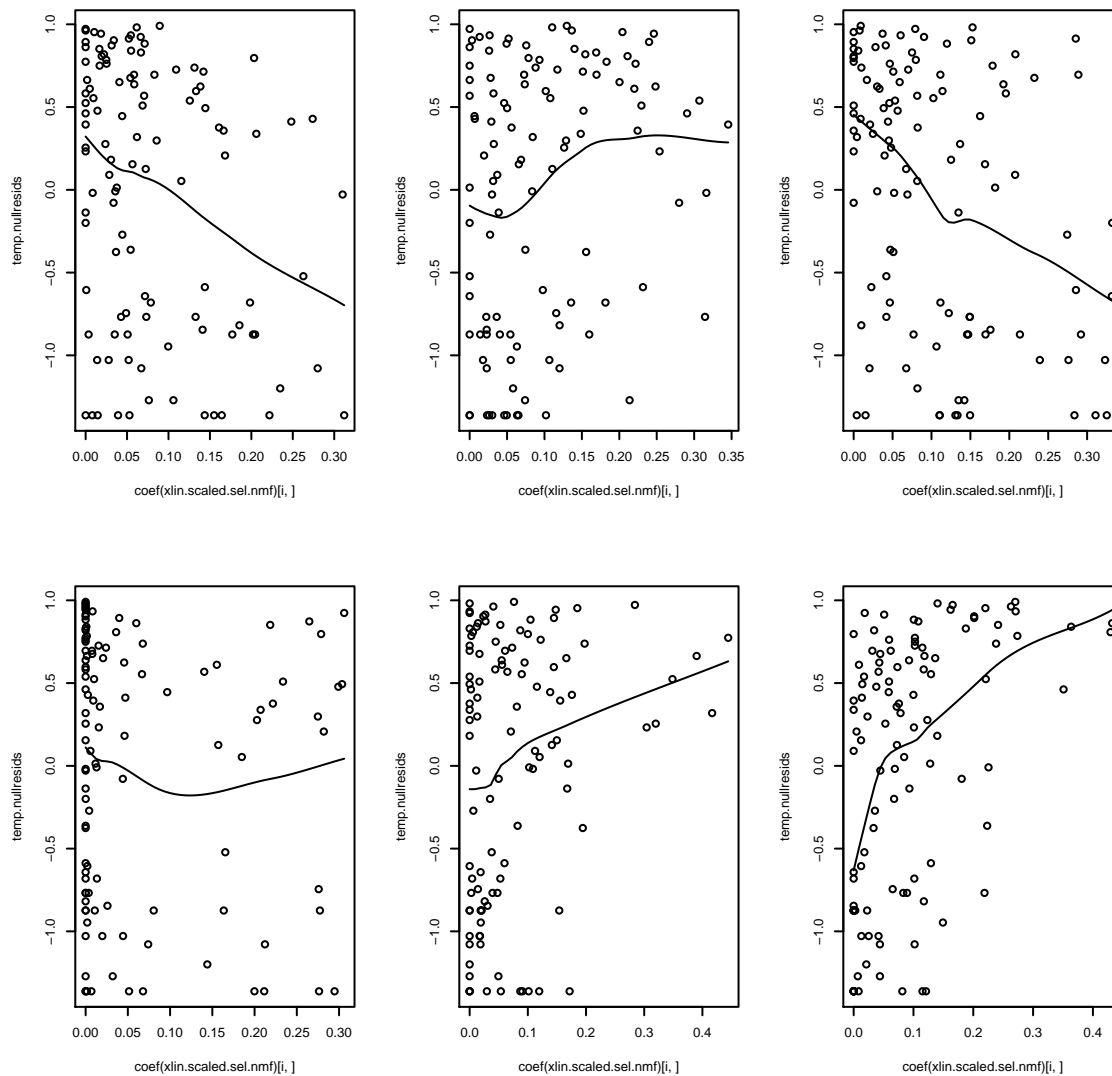
```
par(mfrow = c(1, 1))
```

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 = -3.211, p-value = 0.001324
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
```

```
##      tau
## -0.2498
##
##
## [[2]]
##
##  Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -2.26, p-value = 0.02382
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1758
##
##
## [[3]]
##
##  Kendall's rank correlation tau
##
## data:  samps$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
##
##
## [[4]]
##
##  Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -1.844, p-value = 0.06518
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1502
##
##
## [[5]]
##
##  Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -1.727, p-value = 0.08422
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1357
##
##
## [[6]]
##
```

```
##  Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = 0.1944, p-value = 0.8459
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##    tau
## 0.0151
```

```r
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, ])
}
```
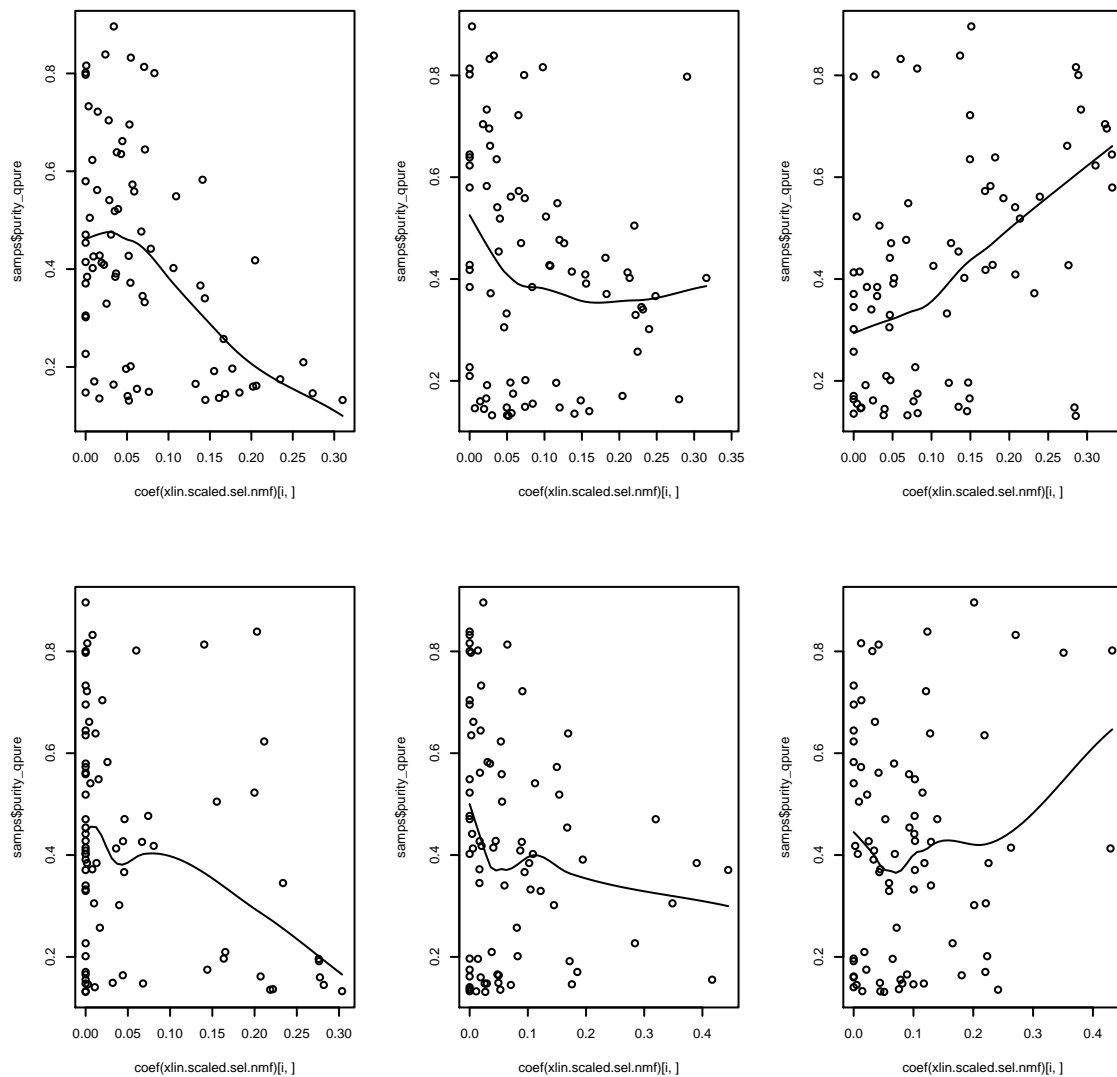


```r
par(mfrow = c(1, 1))
```

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 3.660e-03 0.0013244     -1     -1   2.928e-02     0.01192 2.726e-02
## 2 1.799e-02 0.0238226      1     -1   1.079e-01     0.11911 9.569e-02
## 3 1.103e-04 0.0001054     -1      1   1.160e-03     0.00116 1.370e-03
## 4 4.408e-01 0.0651752     -1     -1   8.816e-01     0.26070 1.000e+00
## 5 1.419e-02 0.0842243      1     -1   9.935e-02     0.26070 8.809e-02
## 6 1.161e-07 0.8458778      1      1   1.393e-06     0.88162 4.324e-06
##   q.pure.BY
## 1   0.01233
## 2   0.11089
## 3   0.00137
## 4   0.26967
## 5   0.31364
## 6   1.00000
```

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)
```
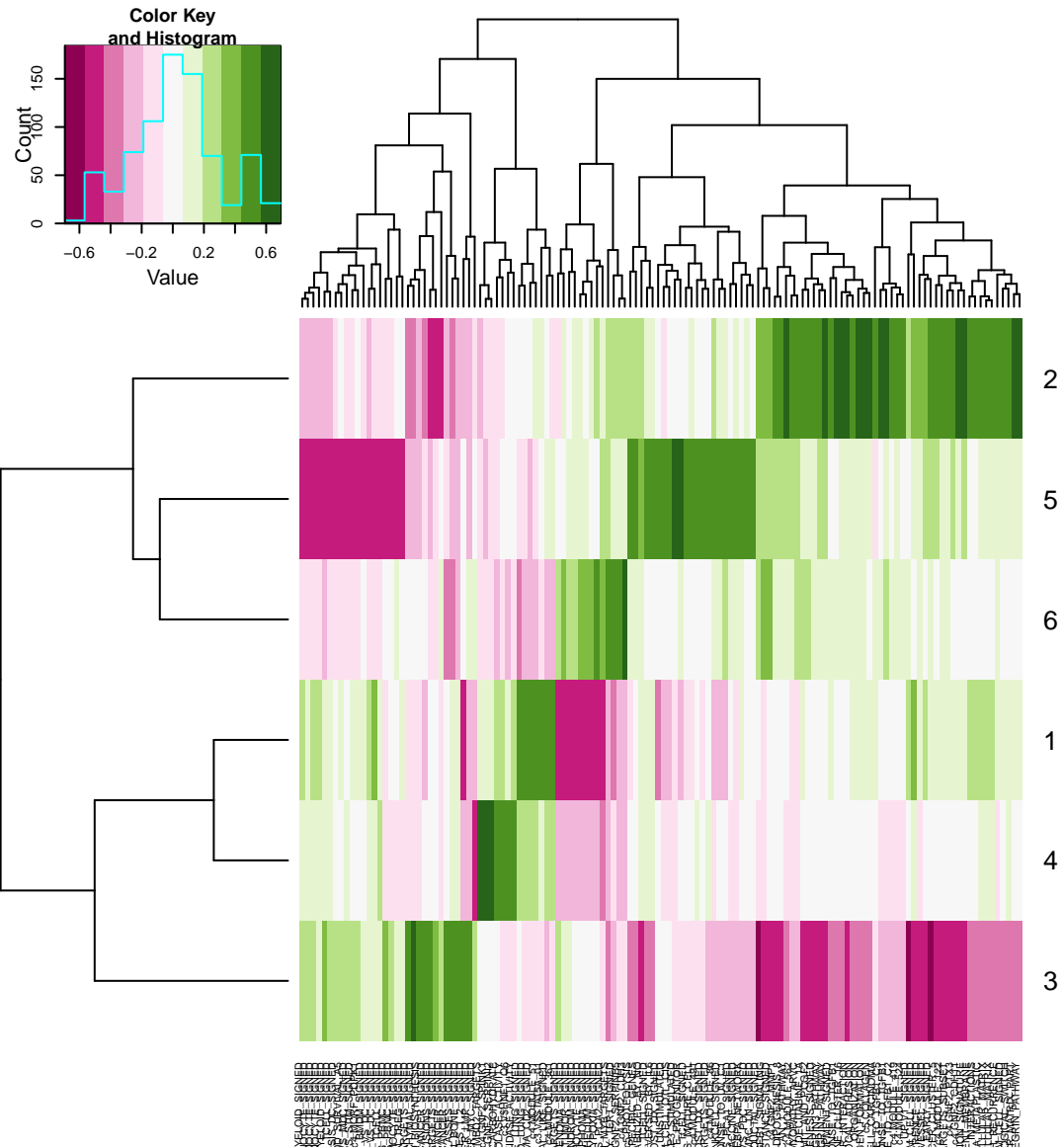
```
heatmap.2(xlin.scaled.sel.nmf.msigdb.corr[, temp.sel_cols], trace = "none",
    scale = "none", useRaster = TRUE, col = brewer.pal(3, "PiYG"), breaks = c(-1,
        -sig.corr.threshold, sig.corr.threshold, 1))
```
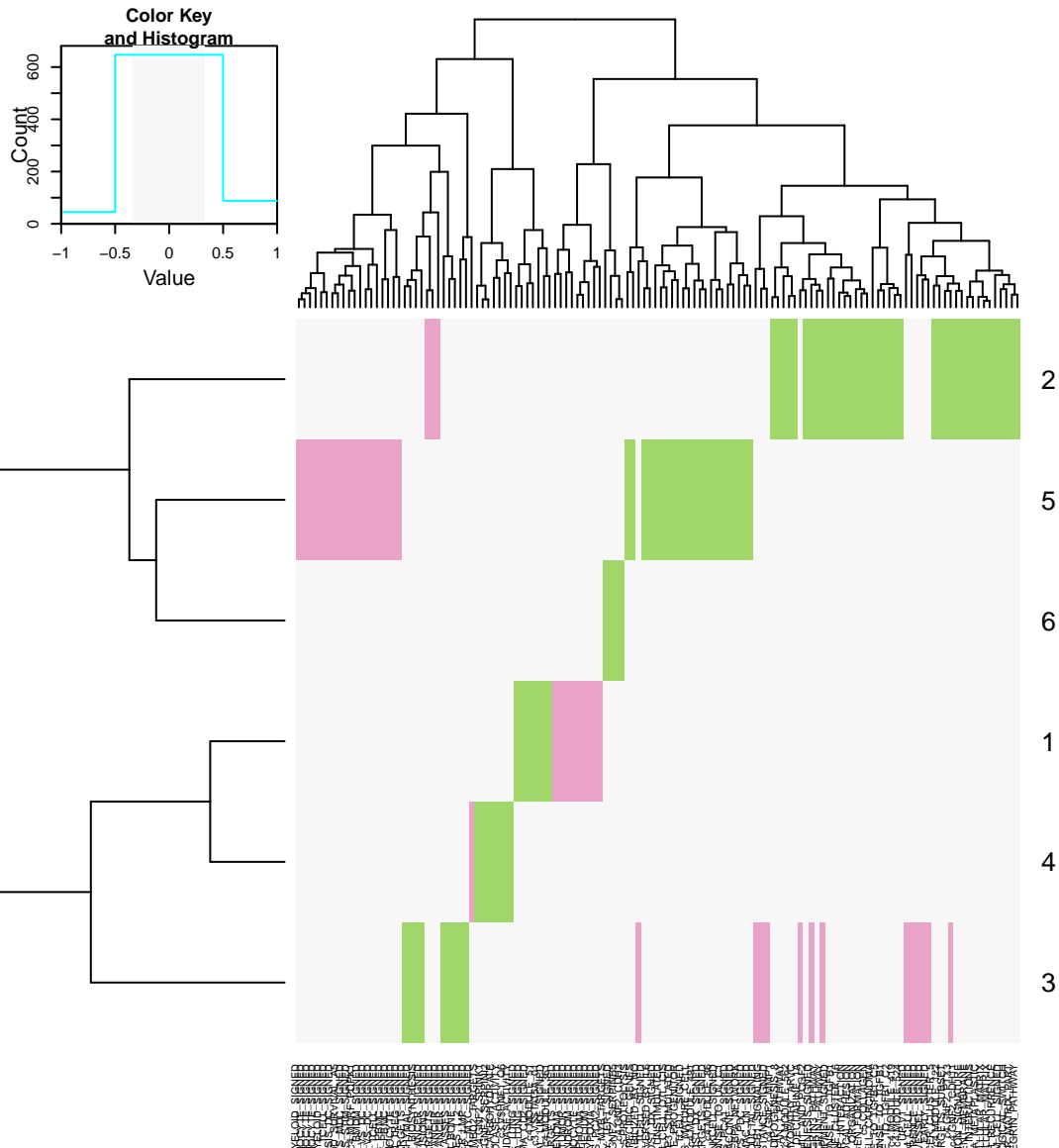
```
temp.sig_id = colnames(xlin.scaled.sel.nmf.msigdb.corr)
temp.sig_class = gsub("\\..*", "", temp.sig_id)
temp.nsigs = length(temp.sig_id)
temp.nmeta = nrow(xlin.scaled.sel.nmf.msigdb.corr)
tables = lapply(1:temp.nmeta, function(metagene_i) {
    tapply(1:temp.nsigs, temp.sig_class, function(sig_class_is) {
        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_repor
                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$d.sur
                      collapse = ",")
                  }))
            table = table[order(-(table$Correlation)), ]
            rownames(table) <- NULL
        }
        table
    }, simplify = FALSE)
})
tables

## [[1]]
## [[1]]$c1
## data frame with 0 columns and 0 rows
##
## [[1]]$c2
##                                              GeneSet Correlation Metagenes
## 1              c2.ZHAN_MULTIPLE_MYELOMA_CD2_SIGNED      0.5063        -1
## 2            c2.SMID_BREAST_CANCER_NORMAL_LIKE_SIGNED    0.5029        -1
## 3                   c2.GREENBAUM_E2A_TARGETS_SIGNED    -0.5026         1
## 4          c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_SIGNED   -0.5049         1
## 5   c2.MARTORIATI_MDM4_TARGETS_NEUROEPITHELIUM_SIGNED   -0.5090         1
## 6               c2.SABATES_COLORECTAL_ADENOMA_SIGNED    -0.5164         1
## 7                          c2.WINTER_HYPOXIA_SIGNED    -0.5164         1
## 8                 c2.LEE_EARLY_T_LYMPHOCYTE_SIGNED     -0.5490         1
## 9               c2.WEST_ADRENOCORTICAL_TUMOR_SIGNED    -0.5533         1
## 10      c2.SHIPP_DLBCL_VS_FOLLICULAR_LYMPHOMA_SIGNED   -0.5543         1
## 11                c2.HAHTOLA_SEZARY_SYNDROM_SIGNED     -0.5631         1
##
## [[1]]$c3
##          GeneSet Correlation Metagenes
## 1 c3.V$STAT5A_01      0.5227        -1
##
## [[1]]$c4
##         GeneSet Correlation Metagenes
## 1  c4.MODULE_51     0.5443        -1
## 2 c4.MODULE_361     0.5160        -1
##
## [[1]]$c5
## data frame with 0 columns and 0 rows
##
## [[1]]$c6
## data frame with 0 columns and 0 rows
##
## [[1]]$c7
##                                            GeneSet Correlation Metagenes
## 1 c7.GSE20715_0H_VS_48H_OZONE_TLR4_KO_LUNG_SIGNED      0.5224        -1
```

```
## 2            c7.GSE20715_0H_VS_48H_OZONE_LUNG_SIGNED      0.5083        -1
## 
## 
## [[2]]
## [[2]]$c1
## data frame with 0 columns and 0 rows
## 
## [[2]]$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.FARMER_BREAST_CANCER_CLUSTER_5
## 14 c2.KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
## 15                               c2.BURTON_ADIPOGENESIS_8
## 16                    c2.MAHADEVAN_GIST_MORPHOLOGICAL_SWITCH
## 17                          c2.KEGG_BASAL_CELL_CARCINOMA
## 18                        c2.CAIRO_LIVER_DEVELOPMENT_SIGNED
## 19                         c2.ROZANOV_MMP14_TARGETS_SUBSET
## 20                               c2.PID_INTEGRIN3_PATHWAY
## 21                      c2.VERRECCHIA_RESPONSE_TO_TGFB1_C2
## 22                    c2.LIEN_BREAST_CARCINOMA_METAPLASTIC
## 23                        c2.CROMER_TUMORIGENESIS_SIGNED
## 24              c2.WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2
## 25              c2.LINDGREN_BLADDER_CANCER_HIGH_RECURRENCE
## 26                         c2.REACTOME_NCAM1_INTERACTIONS
## 27                           c2.BOQUEST_STEM_CELL_SIGNED
## 28            c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_SIGNED
## 29        c2.BERTUCCI_MEDULLARY_VS_DUCTAL_BREAST_CANCER_SIGNED
## 30                          c2.PASINI_SUZ12_TARGETS_SIGNED
##    Correlation Metagenes
## 1       0.6503        2
## 2       0.6355        2
## 3       0.6188        2
## 4       0.6090        2
## 5       0.6060        2
## 6       0.5983      2,-3
## 7       0.5949        2
## 8       0.5916        2
## 9       0.5829        2
## 10      0.5775        2
## 11      0.5688        2
## 12      0.5544        2
## 13      0.5433        2
## 14      0.5399        2
```

34

```
## 15        0.5383         2
## 16        0.5352         2
## 17        0.5305         2
## 18        0.5268         2
## 19        0.5232         2
## 20        0.5195         2
## 21        0.5195         2
## 22        0.5178         2
## 23        0.5138       2,-3
## 24        0.5091         2
## 25        0.5081         2
## 26        0.5013         2
## 27        0.5007         2
## 28       -0.5084        -2
## 29       -0.5362        -2
## 30       -0.5557        -2
##
## [[2]]$c3
## data frame with 0 columns and 0 rows
##
## [[2]]$c4
##         GeneSet Correlation Metagenes
## 1 c4.GNF2_CDH11      0.6271         2
## 2   c4.GNF2_PTX3      0.5450      2,-3
## 3   c4.GNF2_MMP1      0.5369         2
## 4 c4.MODULE_122      0.5242         2
## 5 c4.MODULE_524      0.5141         2
## 6 c4.MODULE_419      0.5121         2
## 7 c4.MODULE_562      0.5074         2
##
## [[2]]$c5
##                               GeneSet Correlation Metagenes
## 1                         c5.COLLAGEN      0.6547         2
## 2          c5.EXTRACELLULAR_MATRIX_PART      0.6110         2
## 3 c5.PROTEINACEOUS_EXTRACELLULAR_MATRIX      0.5352         2
## 4              c5.EXTRACELLULAR_MATRIX      0.5195         2
## 5                 c5.BASEMENT_MEMBRANE      0.5067         2
##
## [[2]]$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
## 1  c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_SIGNED      0.5842
## 2                        c2.LUCAS_HNF4A_TARGETS_SIGNED      0.5232
## 3           c2.REACTOME_GLYCEROPHOSPHOLIPID_BIOSYNTHESIS      0.5175
```

```
## 4                      c2.DOANE_BREAST_CANCER_CLASSES_SIGNED      0.5161
## 5                         c2.LIU_PROSTATE_CANCER_SIGNED           0.5158
## 6               c2.WAMUNYOKOLI_OVARIAN_CANCER_LMP_SIGNED           0.5155
## 7           c2.WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_SIGNED        0.5141
## 8              c2.SMID_BREAST_CANCER_RELAPSE_IN_BONE_SIGNED        0.5068
## 9   c2.WANG_BARRETTS_ESOPHAGUS_AND_ESOPHAGUS_CANCER_SIGNED        0.5041
## 10                 c2.WOO_LIVER_CANCER_RECURRENCE_SIGNED          -0.5014
## 11             c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABINE_1A         -0.5027
## 12                         c2.KARAKAS_TGFB1_SIGNALING              -0.5051
## 13                      c2.CROMER_TUMORIGENESIS_SIGNED             -0.5054
## 14                      c2.KUWANO_RNA_STABILIZED_BY_NO             -0.5141
## 15                   c2.LIM_MAMMARY_STEM_CELL_SIGNED               -0.5151
## 16                         c2.PID_INTEGRIN1_PATHWAY                -0.5175
## 17                c2.HUANG_DASATINIB_RESISTANCE_SIGNED             -0.5181
## 18                    c2.ROY_WOUND_BLOOD_VESSEL_SIGNED             -0.5202
## 19      c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SIGNED         -0.6103
## 20  c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SIGNED         -0.6190
##     Metagenes
## 1         -3
## 2         -3
## 3         -3
## 4         -3
## 5         -3
## 6         -3
## 7         -3
## 8         -3
## 9         -3
## 10         3
## 11         3
## 12         3
## 13      -2,3
## 14         3
## 15         3
## 16      -2,3
## 17         3
## 18         3
## 19         3
## 20         3
##
## [[3]]$c3
## data frame with 0 columns and 0 rows
##
## [[3]]$c4
##         GeneSet Correlation Metagenes
## 1 c4.GNF2_PTX3     -0.5111       -2,3
##
## [[3]]$c5
## data frame with 0 columns and 0 rows
##
## [[3]]$c6
##                 GeneSet Correlation Metagenes
## 1 c6.LEF1_UP.V1_SIGNED     -0.5842          3
##
## [[3]]$c7
```

```
## data frame with 0 columns and 0 rows
##
##
## [[4]]
## [[4]]$c1
## data frame with 0 columns and 0 rows
##
## [[4]]$c2
##                                        GeneSet Correlation Metagenes
## 1                       c2.BERGER_MBD2_TARGETS      0.6061        -4
## 2 c2.REACTOME_DIGESTION_OF_DIETARY_CARBOHYDRATE      0.5185        -4
## 3                 c2.LEE_LIVER_CANCER_MYC_SIGNED     -0.5178         4
##
## [[4]]$c3
##         GeneSet Correlation Metagenes
## 1 c3.V$HNF1_Q6      0.5199        -4
##
## [[4]]$c4
##           GeneSet Correlation Metagenes
## 1   c4.GNF2_SPINK1      0.6927        -4
## 2 c4.GNF2_SERPINI2      0.6791        -4
##
## [[4]]$c5
##                      GeneSet Correlation Metagenes
## 1 c5.CARBOXYPEPTIDASE_ACTIVITY      0.5342        -4
## 2 c5.SERINE_HYDROLASE_ACTIVITY      0.5002        -4
##
## [[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      0.6131         5
## 2           c2.MARSON_BOUND_BY_FOXP3_STIMULATED      0.5920         5
## 3                   c2.SESTO_RESPONSE_TO_UV_C1      0.5525         5
## 4     c2.MARTINEZ_RB1_AND_TP53_TARGETS_SIGNED      0.5291         5
## 5         c2.IVANOVA_HEMATOPOIESIS_MATURE_CELL      0.5267         5
## 6   c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA1_SIGNED      0.5236         5
## 7         c2.MARSON_BOUND_BY_FOXP3_UNSTIMULATED      0.5212         5
## 8   c2.BYSTRYKH_HEMATOPOIESIS_STEM_CELL_QTL_CIS      0.5175         5
## 9       c2.RAGHAVACHARI_PLATELET_SPECIFIC_GENES      0.5168         5
## 10            c2.KAMIKUBO_MYELOID_CEBPA_NETWORK      0.5161         5
## 11           c2.BROCKE_APOPTOSIS_REVERSED_BY_IL6      0.5049         5
## 12     c2.SHEDDEN_LUNG_CANCER_GOOD_SURVIVAL_A5     -0.5059        -5
##
## [[5]]$c3
```

```
## data frame with 0 columns and 0 rows
##
## [[5]]$c4
##        GeneSet Correlation Metagenes
## 1  c4.MODULE_86      0.5212          5
## 2 c4.MODULE_491      0.5120          5
## 3 c4.MORF_PDPK1     -0.5124         -5
##
## [[5]]$c5
##         GeneSet Correlation Metagenes
## 1 c5.PROTEOLYSIS      0.5151          5
##
## [[5]]$c6
## data frame with 0 columns and 0 rows
##
## [[5]]$c7
##                                                             GeneSet
## 1                            c7.GSE29618_MONOCYTE_VS_PDC_SIGNED
## 2                      c7.GSE11057_PBMC_VS_MEM_CD4_TCELL_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_1MONTH_POST_DX_SIGNED
## 8  c7.GSE9006_HEALTHY_VS_TYPE_1_DIABETES_PBMC_4MONTH_POST_DX_SIGNED
## 9                    c7.GSE20366_CD103_KLRG1_DP_VS_DN_TREG_SIGNED
## 10                 c7.GSE22886_NAIVE_CD8_TCELL_VS_MONOCYTE_SIGNED
## 11                 c7.GSE11057_NAIVE_CD4_VS_PBMC_CD4_TCELL_SIGNED
## 12                c7.GSE17721_POLYIC_VS_GARDIQUIMOD_1H_BMDM_SIGNED
## 13                              c7.GSE29618_PDC_VS_MDC_SIGNED
## 14                c7.GSE360_T_GONDII_VS_M_TUBERCULOSIS_DC_SIGNED
## 15            c7.GSE1448_CTRL_VS_ANTI_VALPHA2_DP_THYMOCYTE_SIGNED
## 16                       c7.GSE3982_DC_VS_MAC_LPS_STIM_SIGNED
## 17                   c7.GSE22886_NAIVE_CD8_TCELL_VS_DC_SIGNED
## 18               c7.GSE11057_NAIVE_VS_MEMORY_CD4_TCELL_SIGNED
## 19                  c7.GSE11057_CD4_CENT_MEM_VS_PBMC_SIGNED
## 20                      c7.GSE10325_BCELL_VS_MYELOID_SIGNED
## 21                   c7.GSE11057_CD4_EFF_MEM_VS_PBMC_SIGNED
## 22                   c7.GSE22886_NAIVE_CD4_TCELL_VS_DC_SIGNED
## 23                 c7.GSE22886_NAIVE_TCELL_VS_MONOCYTE_SIGNED
## 24            c7.GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_MYELOID_SIGNED
## 25               c7.GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_SIGNED
##    Correlation Metagenes
## 1       0.5604          5
## 2       0.5600          5
## 3       0.5437          5
## 4       0.5430          5
## 5       0.5416          5
## 6       0.5362          5
## 7       0.5049          5
## 8       0.5015          5
## 9      -0.5042         -5
## 10     -0.5042         -5
## 11     -0.5046         -5
```

```
## 12      -0.5073         -5
## 13      -0.5080         -5
## 14      -0.5097         -5
## 15      -0.5117         -5
## 16      -0.5165         -5
## 17      -0.5199         -5
## 18      -0.5250         -5
## 19      -0.5257         -5
## 20      -0.5352         -5
## 21      -0.5372         -5
## 22      -0.5372         -5
## 23      -0.5406         -5
## 24      -0.5532         -5
## 25      -0.5583         -5
##
##
## [[6]]
## [[6]]$c1
## data frame with 0 columns and 0 rows
##
## [[6]]$c2
##                                 GeneSet Correlation Metagenes
## 1 c2.KANG_IMMORTALIZED_BY_TERT_SIGNED      0.5107         6
## 2                   c2.LEI_MYB_TARGETS      0.5050         6
##
## [[6]]$c3
## data frame with 0 columns and 0 rows
##
## [[6]]$c4
##            GeneSet Correlation Metagenes
## 1    c4.GNF2_CDH3      0.5700         6
## 2 c4.GNF2_SERPINB5      0.5448         6
##
## [[6]]$c5
## data frame with 0 columns and 0 rows
##
## [[6]]$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: 551.558245978867
## Best model:
## [1] "Surv(time, event) ~ 1 + mg.5 + mg.6"
## Evidence weight: 0.121525761025609
## Worst IC: 561.093163081812
```

```
## 5 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.1:mg.4 -0.00342         0.001098         1    0.001595            0.06569
## mg.2:mg.4  0.06847         0.024924         2    0.003030            0.31297
## mg.4:mg.5 -0.05064         0.019737         1    0.003679            0.27850
## mg.4:mg.6  0.02227         0.015093         2    0.004699            0.24355
## mg.2:mg.3  0.08587         0.057059         3    0.004885            0.47354
## mg.1:mg.2  0.10761         0.069012         2    0.004997            0.52078
## mg.1:mg.5 -0.01322         0.015809         2    0.005116            0.24925
## mg.3:mg.4 -0.25668         0.347670         4    0.008845            1.16890
## mg.3:mg.5  0.01013         0.047858         3    0.009870            0.43368
## mg.2:mg.5  0.09367         0.064793         4    0.010069            0.50461
## mg.5:mg.6  0.10404         0.413552         8    0.036678            1.27485
## mg.1:mg.6  0.75549         3.893974         8    0.039107            3.91192
## mg.1:mg.3 -1.05886         7.166918         9    0.043683            5.30712
## mg.2:mg.6 -1.53508         9.333667        13    0.074749            6.05646
## mg.3:mg.6  2.60281        29.616218        14    0.097762           10.78842
## mg.4      0.09036          0.113089        33    0.136430            0.66666
## mg.2      0.33327          0.778397        42    0.219557            1.74902
## mg.1     -1.85625          6.374836        54    0.439670            5.00527
## mg.5      1.08749          2.127350        49    0.444550            2.89143
## mg.3     -2.66620          7.511723        65    0.592820            5.43329
## mg.6      5.31298          5.463222        90    0.932557            4.63359
```

```
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.5    2.81     16.65     1.08 2.60   0.0093
## mg.6    6.99   1089.19     1.19 5.87 4.4e-09
##
##      exp(coef) exp(-coef) lower .95 upper .95
## mg.5      16.7   0.060050         2       139
## mg.6    1089.2   0.000918       105     11264
##
## Concordance= 0.702  (se = 0.038 )
## Rsquare= 0.265   (max possible= 0.995 )
## Likelihood ratio test= 33.9  on 2 df,   p=4.29e-08
## Wald test            = 38.8  on 2 df,   p=3.77e-09
## Score (logrank) test = 42.9  on 2 df,   p=4.87e-10
```
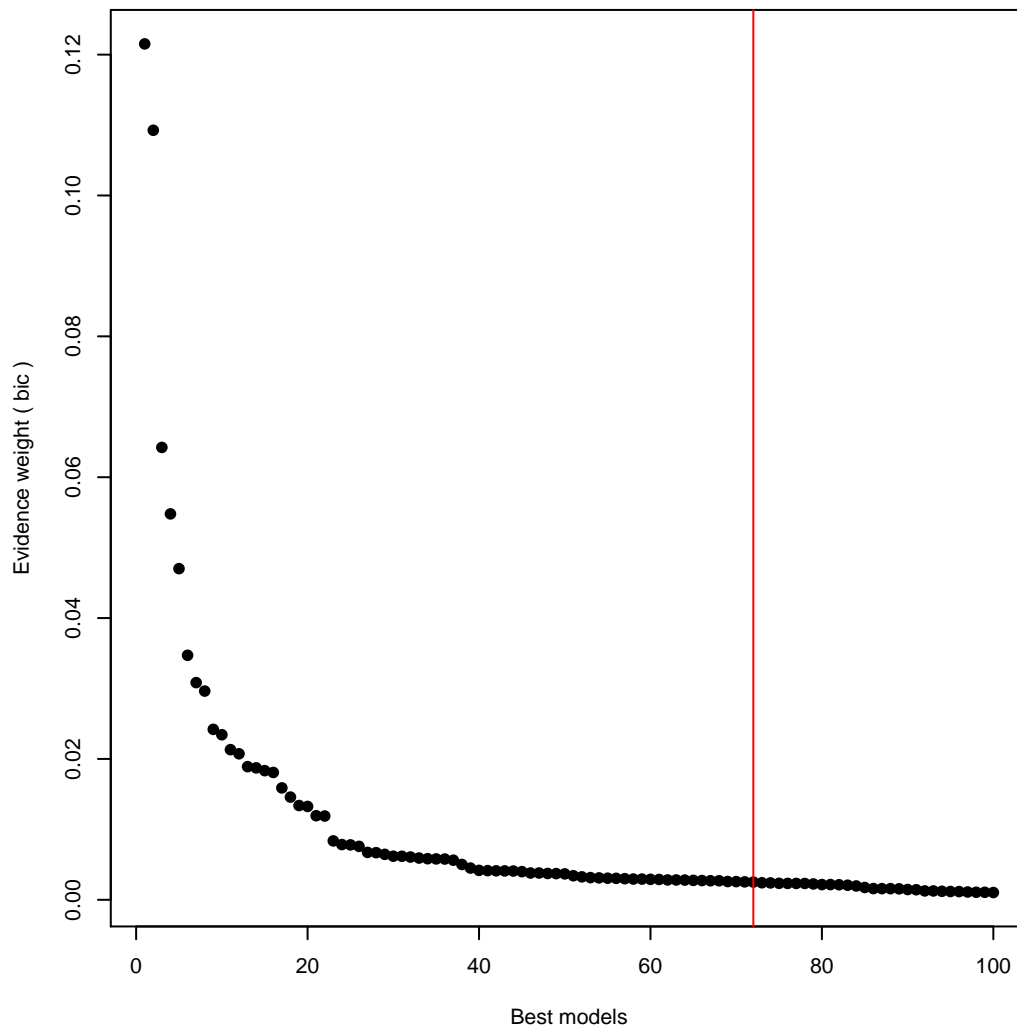
```
plot(asreg.result, type = "p")
```

IC profile


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

41
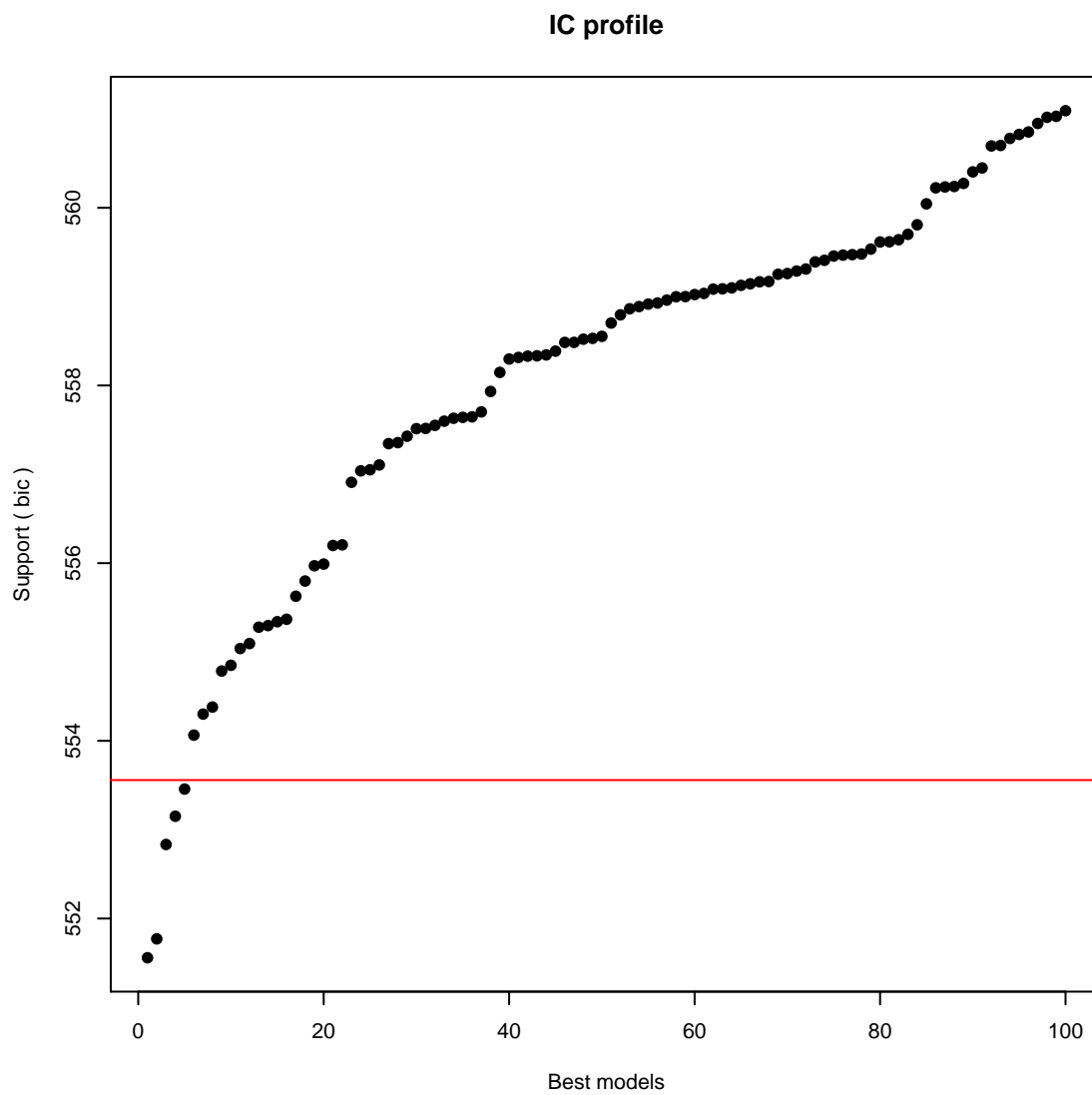
Model−averaged importance of terms
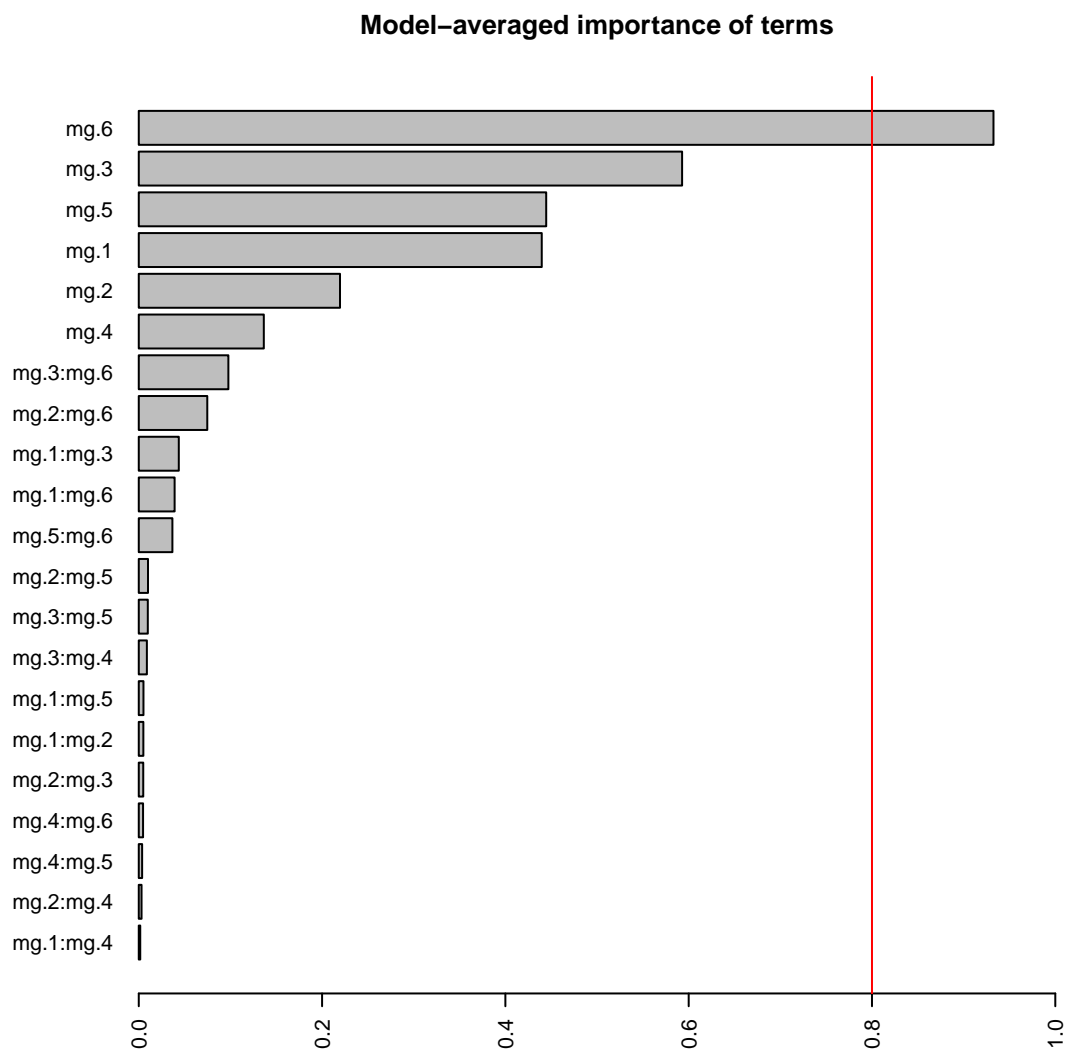

```
plot(asreg.result, type = "w")
```
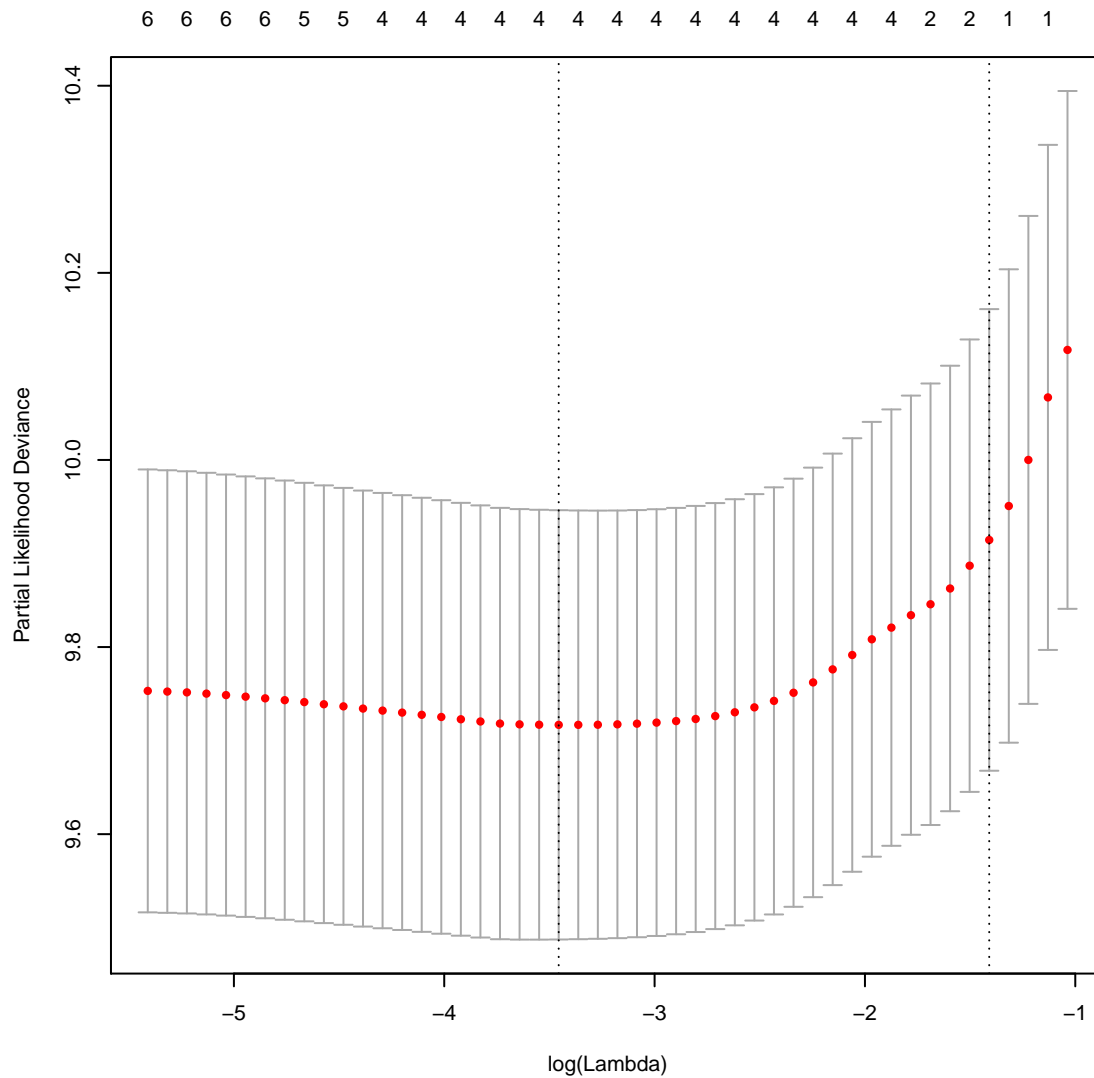
Profile of model weights


```
glmnet.coef.1se

## 6 x 1 sparse Matrix of class "dgCMatrix"
##          1
## mg.1 .
## mg.2 .
## mg.3 .
## mg.4 .
## mg.5 .
## mg.6 2.706

glmnet.coef.min

## 6 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1 -2.697
## mg.2   .
## mg.3 -3.289
```
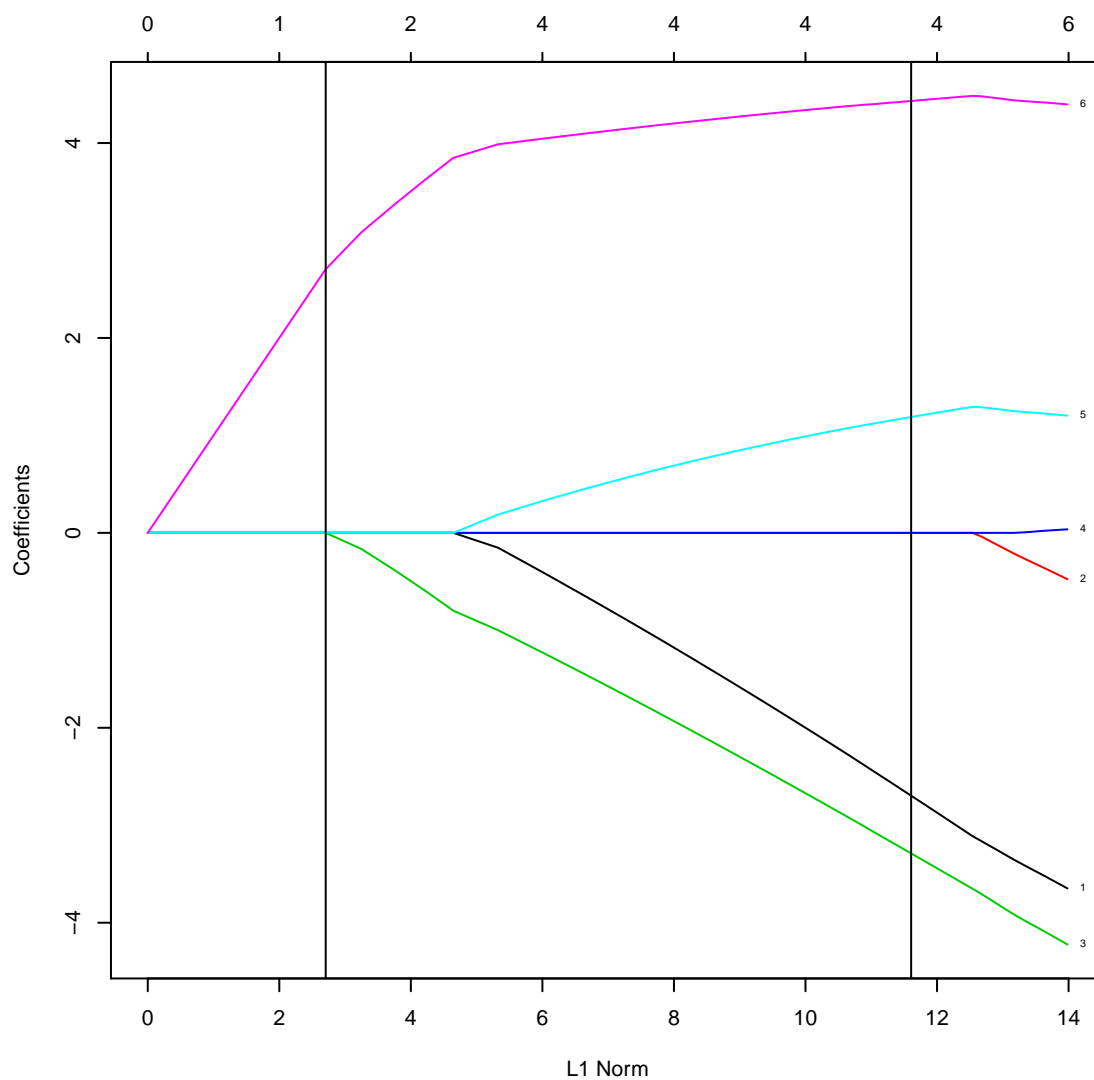
```
## mg.4  .
## mg.5  1.189
## mg.6  4.431
```

```
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  1.078
## mg.3  .
## mg.4  .
## mg.5  2.404
## mg.6 33.683

adaglmnet.coef.min/adaglmnet.weights

## 6 x 1 sparse Matrix of class "dgCMatrix"
##              1
## mg.1 .
## mg.2 9.983e-01
## mg.3 .
```
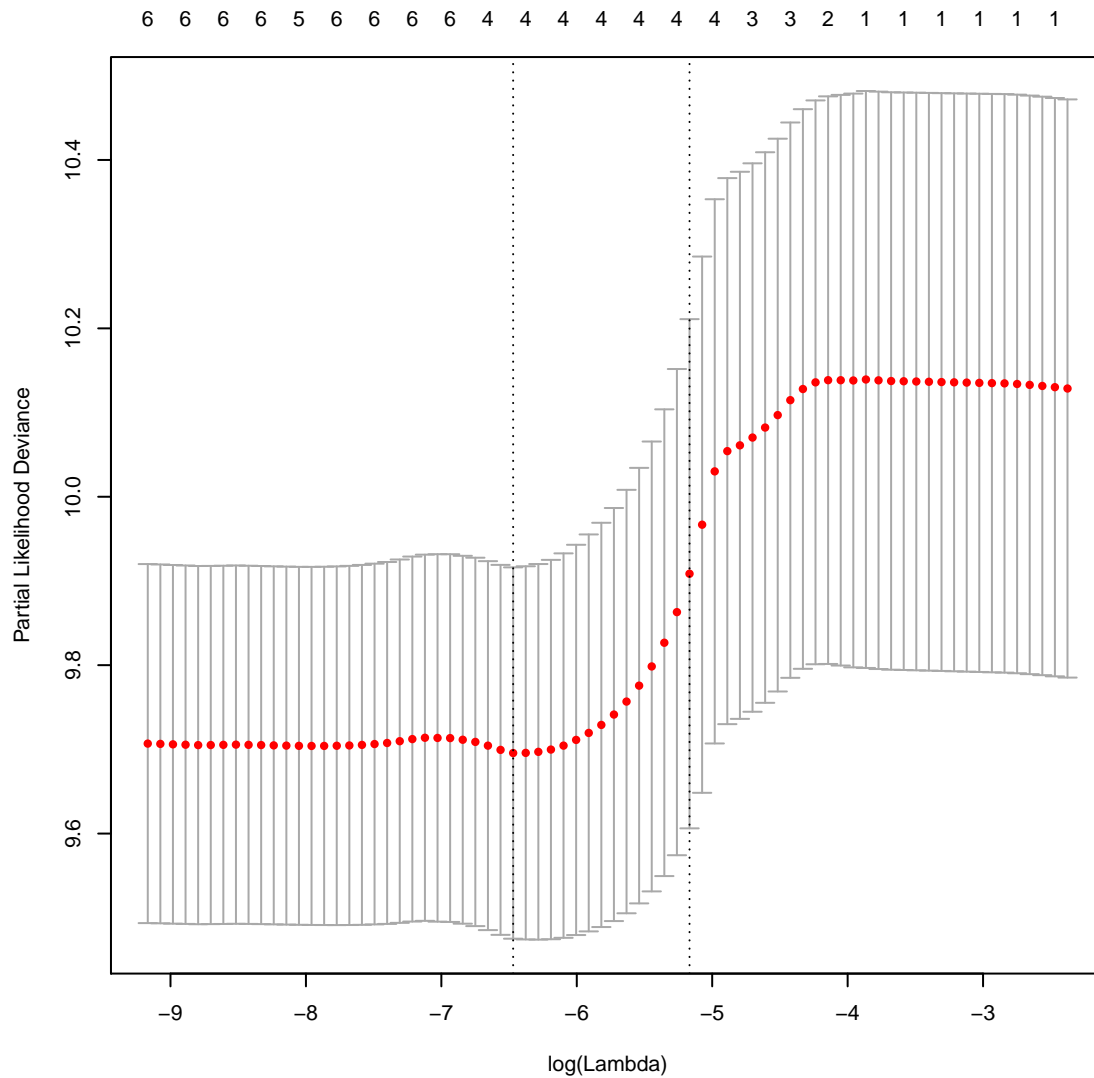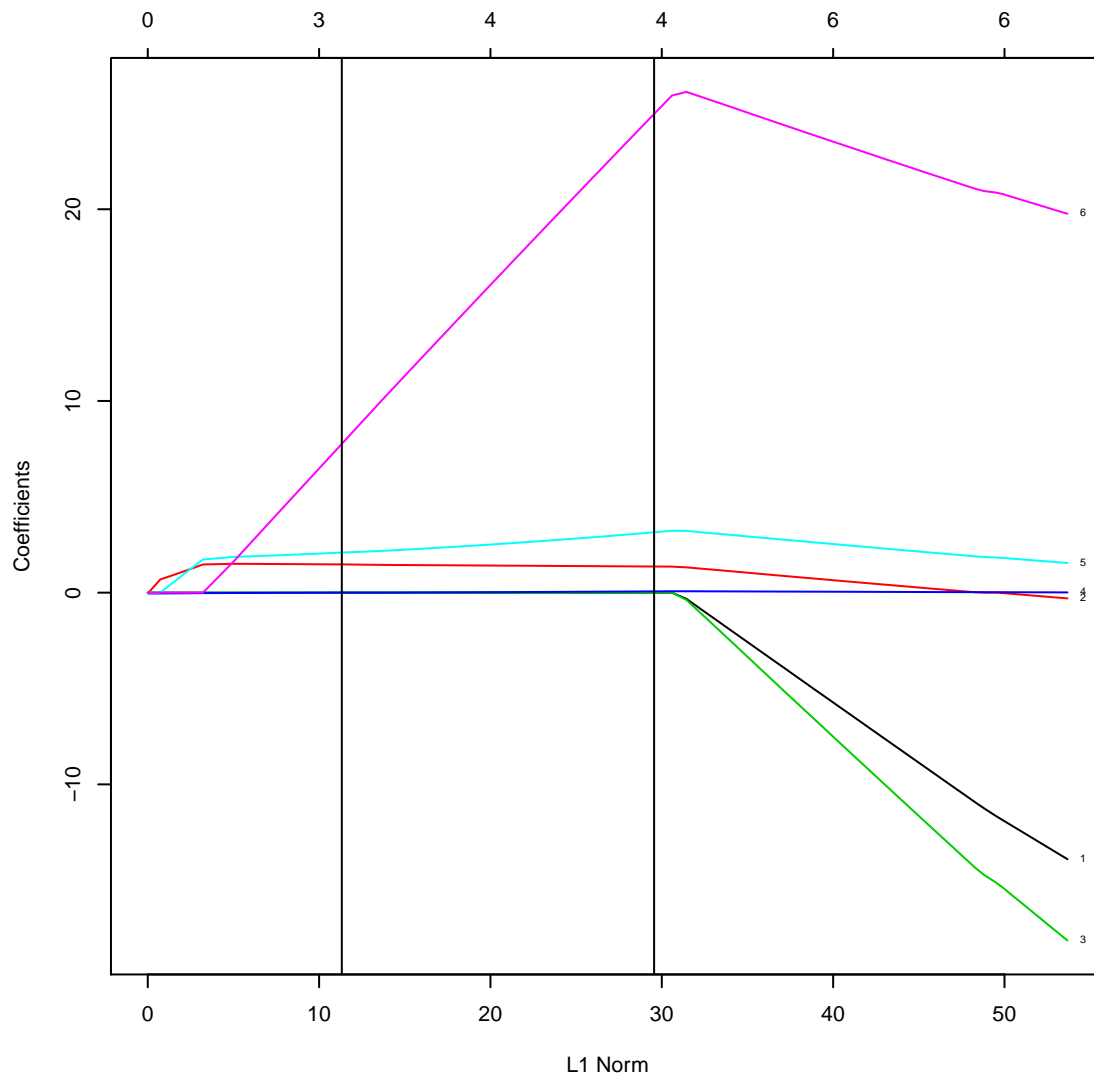
```
## mg.4 4.073e-03
## mg.5 3.620e+00
## mg.6 1.084e+02
```

```
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
```