

SIS NMF diag rec

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
## 12804 196

mean(cpss.sis$sel)

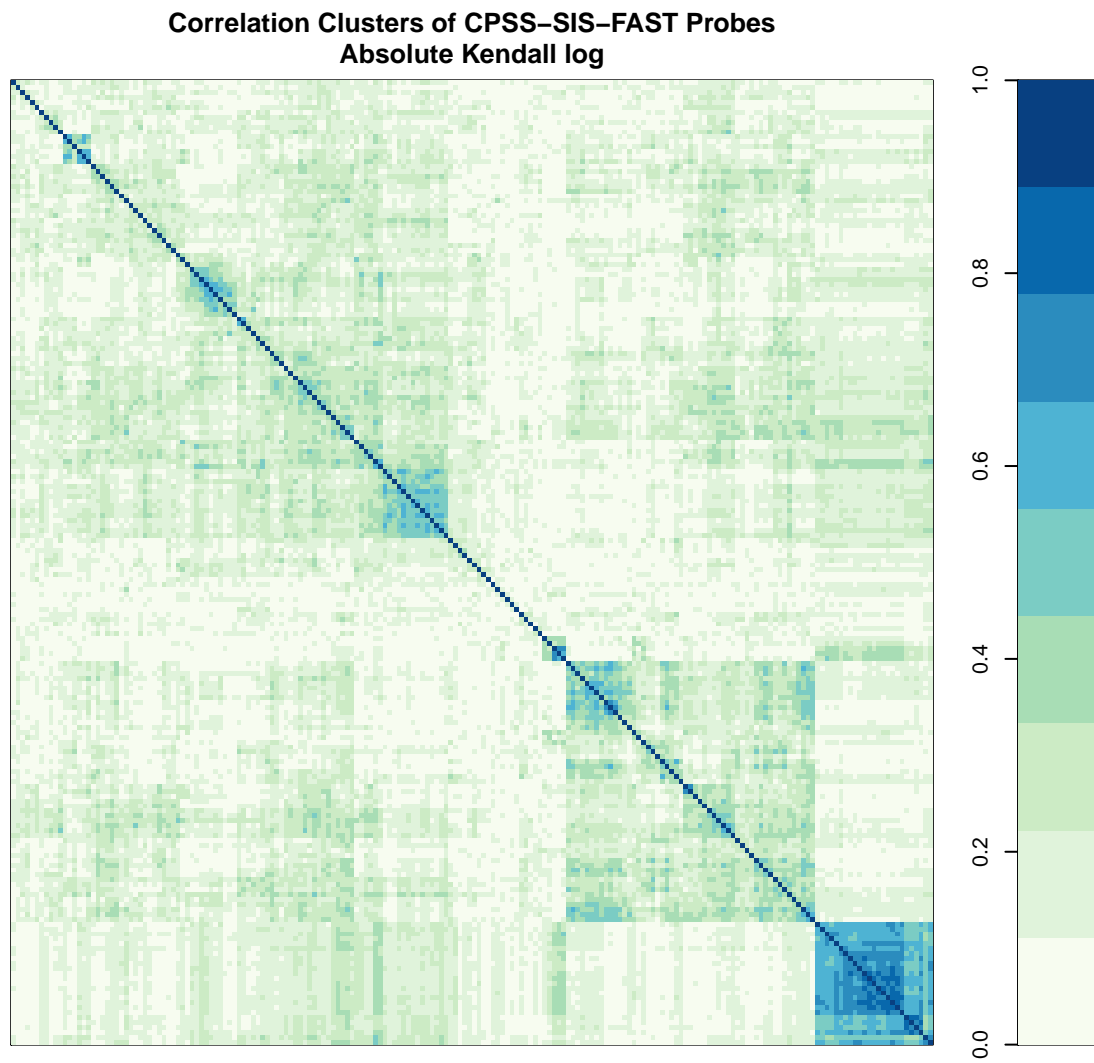
## [1] 0.01508
```

3 Expression correlation

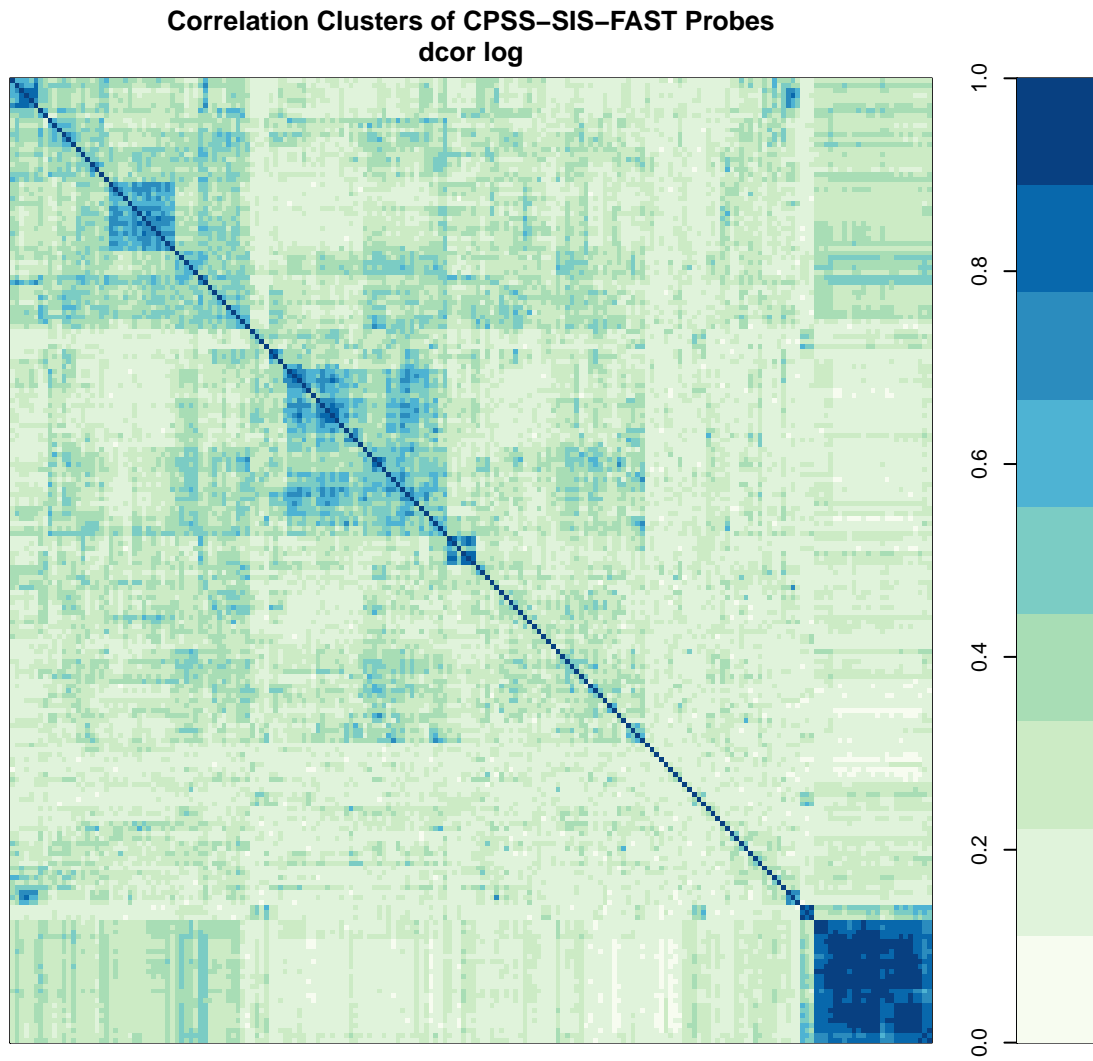
```
corPlot(x.sel.kcor, main = "Correlation Clusters of CPSS-SIS-FAST Probes\nKendall log",
        useRaster = FALSE)
```



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



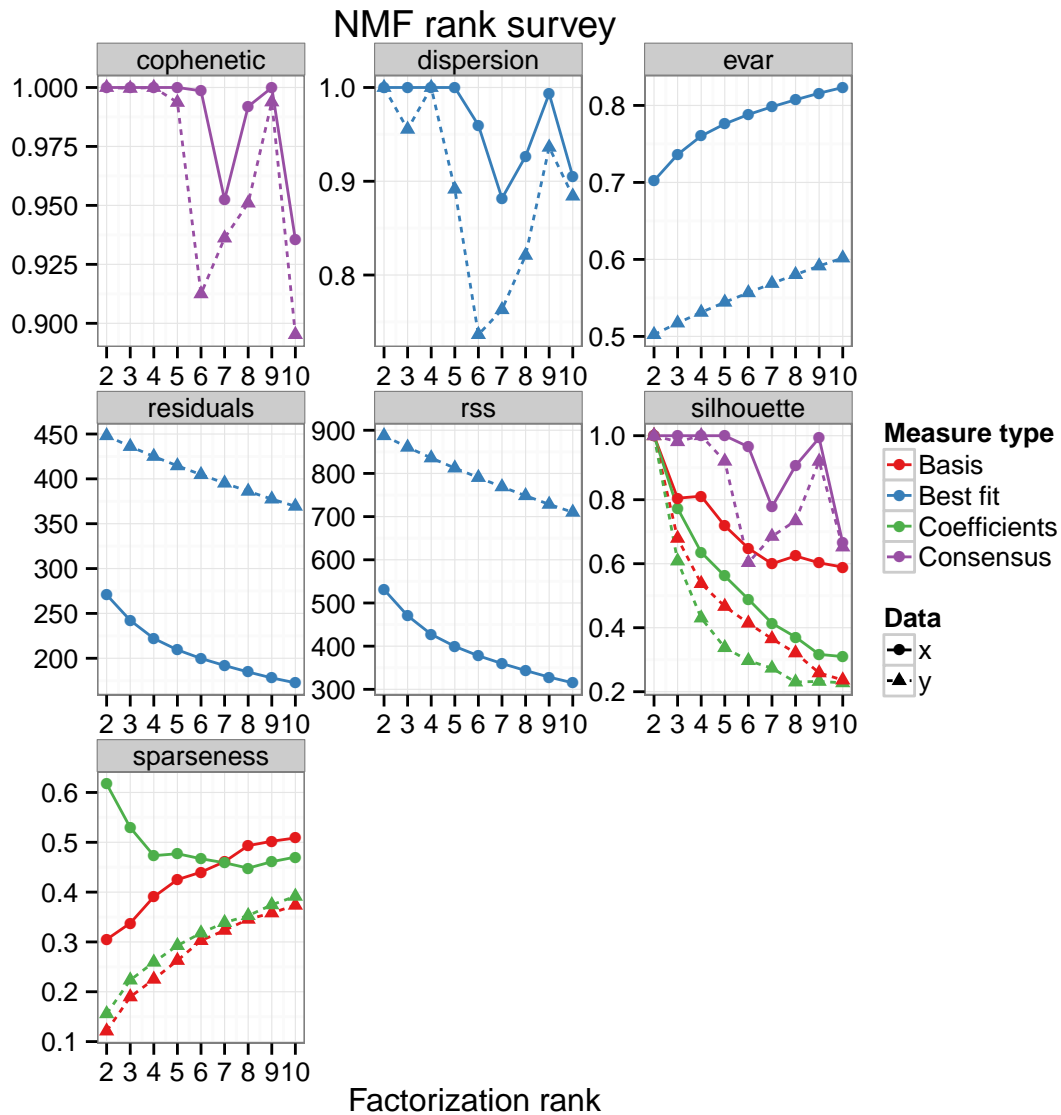
```
corPlot(x.sel.dcor, zlim = c(0, 1), pal = "GnBu", main = "Correlation Clusters of CPSS-SIS-FAST Probes",
        useRaster = FALSE)
```



4 Factorization

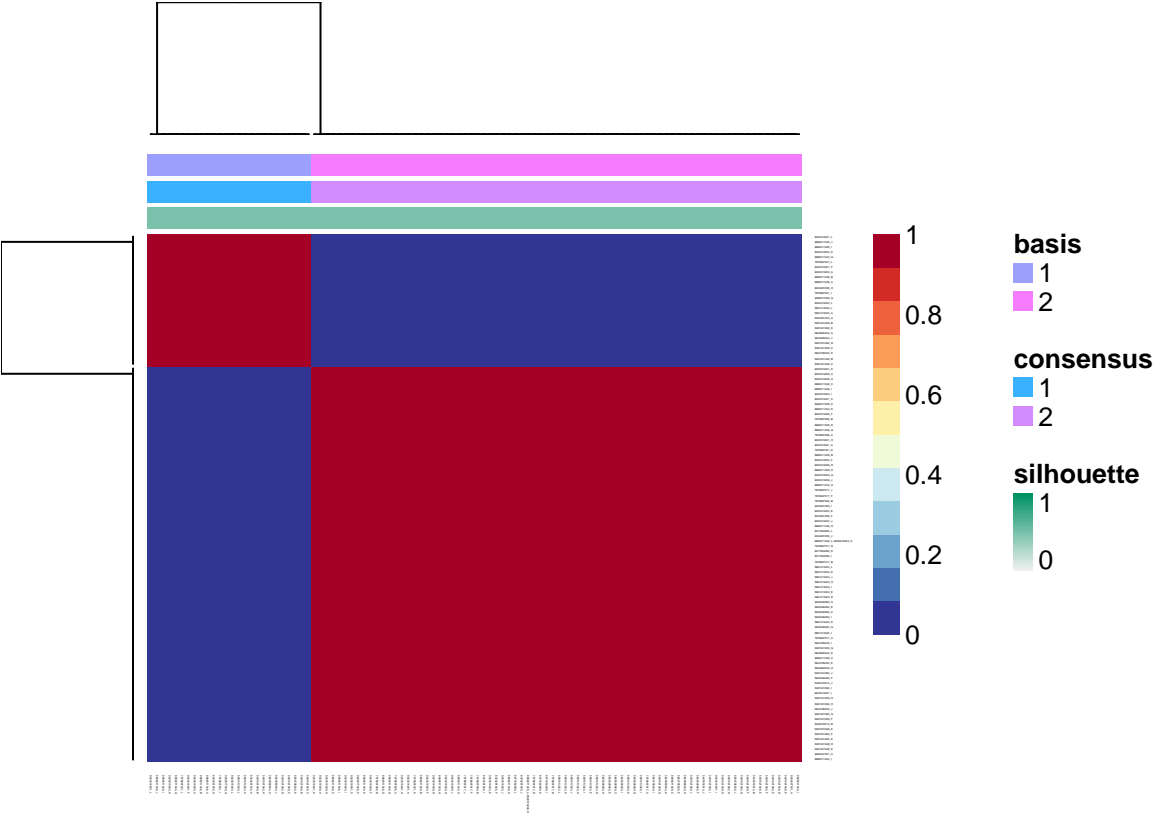
4.1 Rank estimation

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

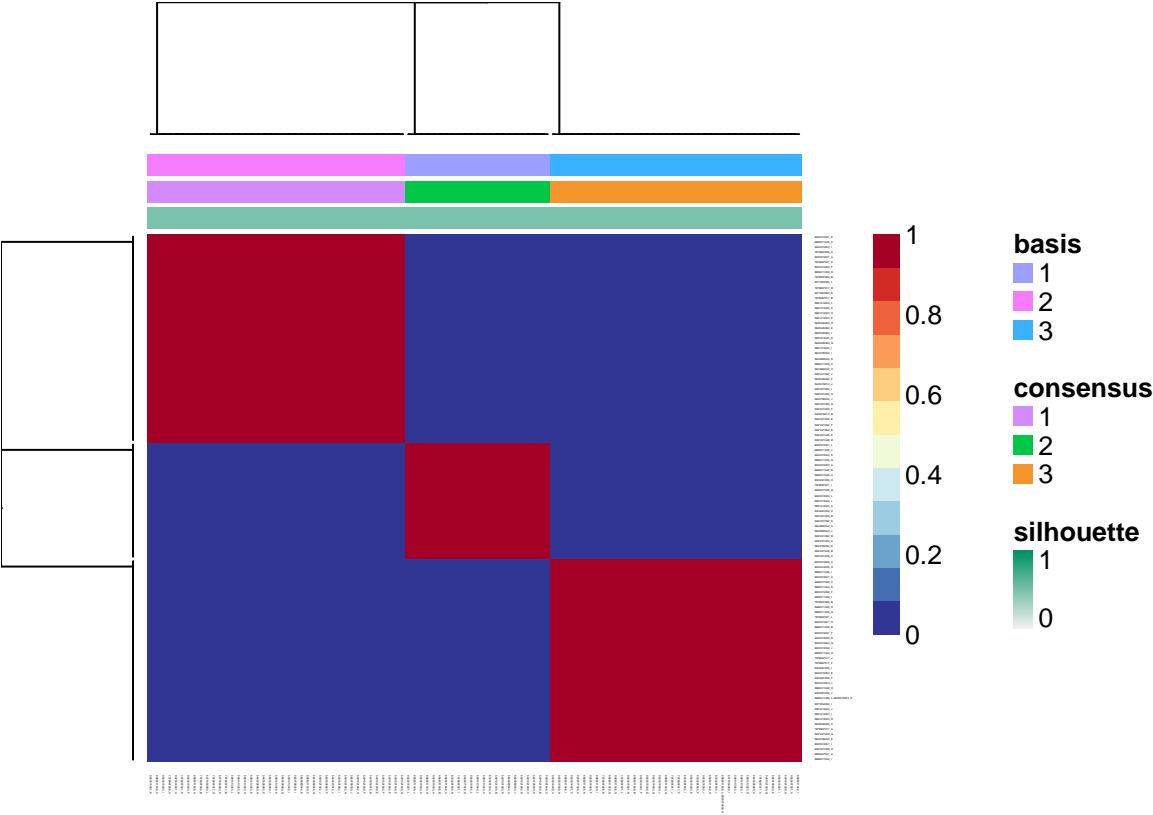


```
for (i in 1:length(temp.nmf.rank$fit)) {
  consensusmap(temp.nmf.rank$fit[[i]])
}
```

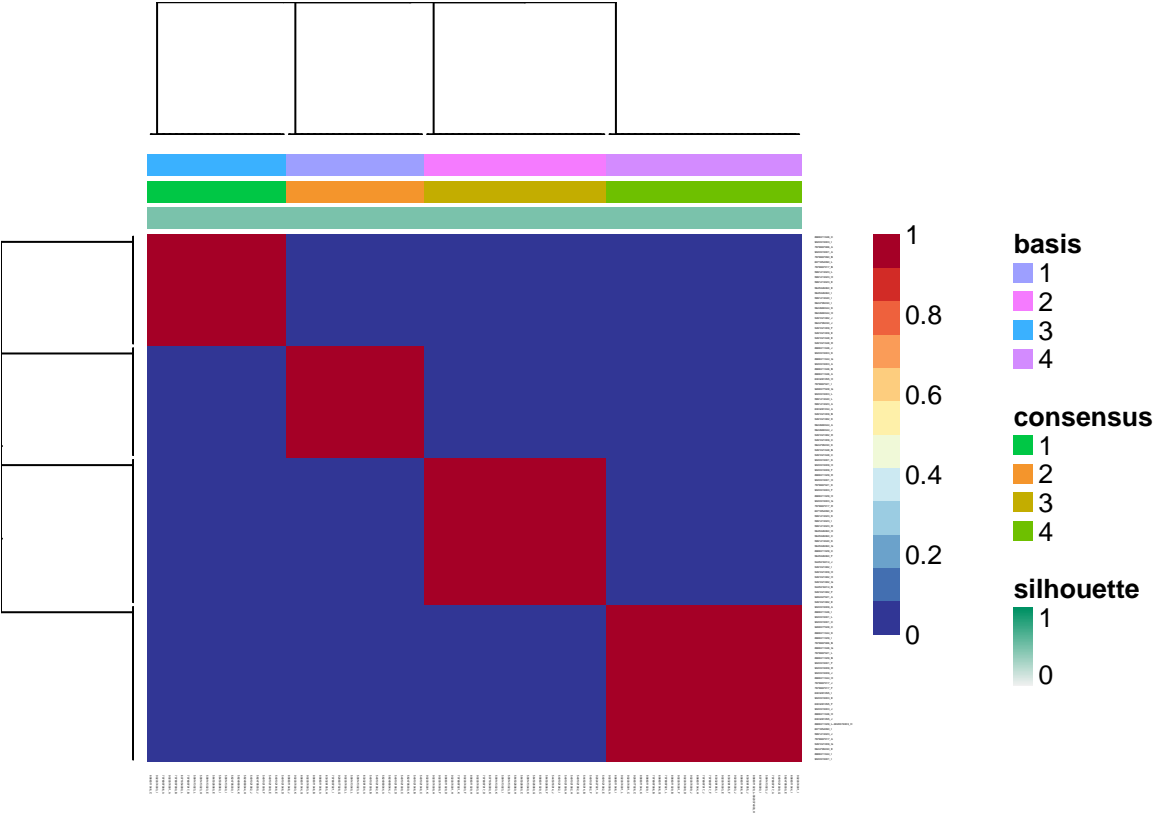
Consensus matrix

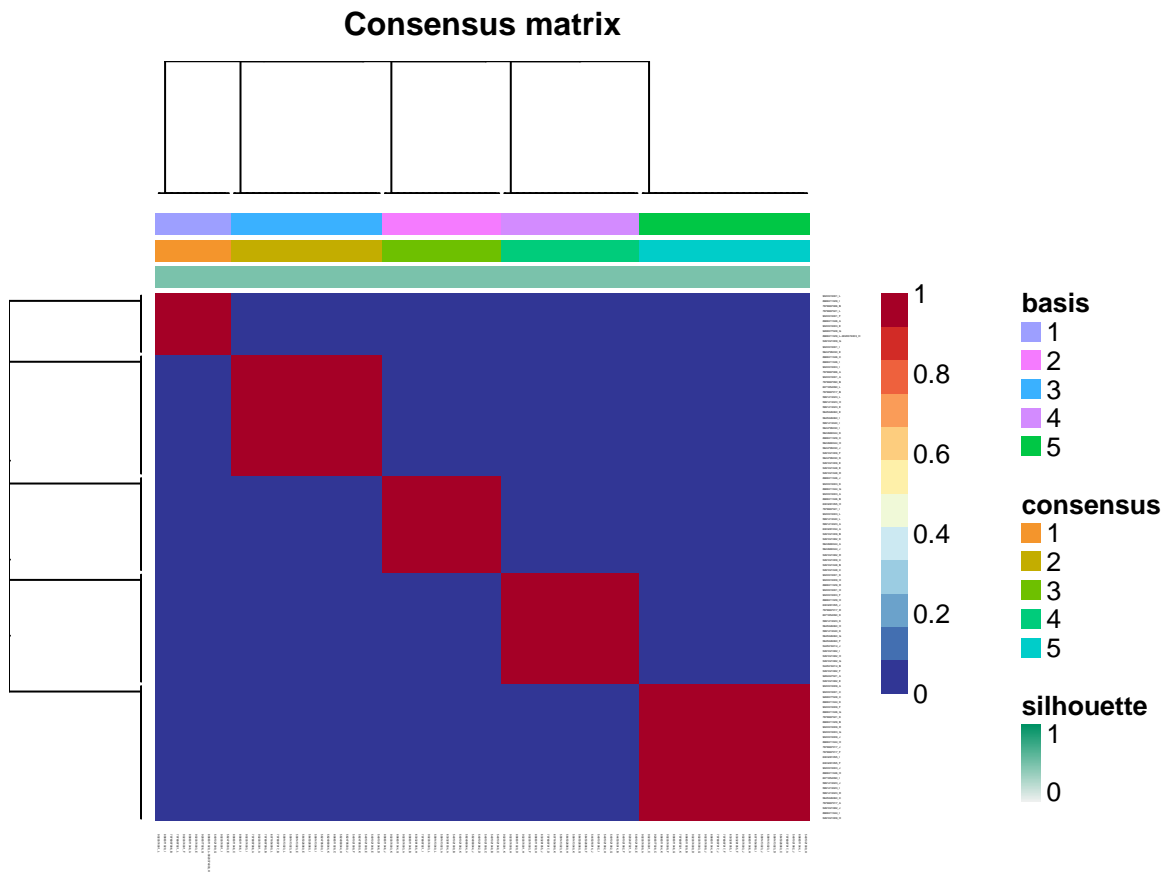


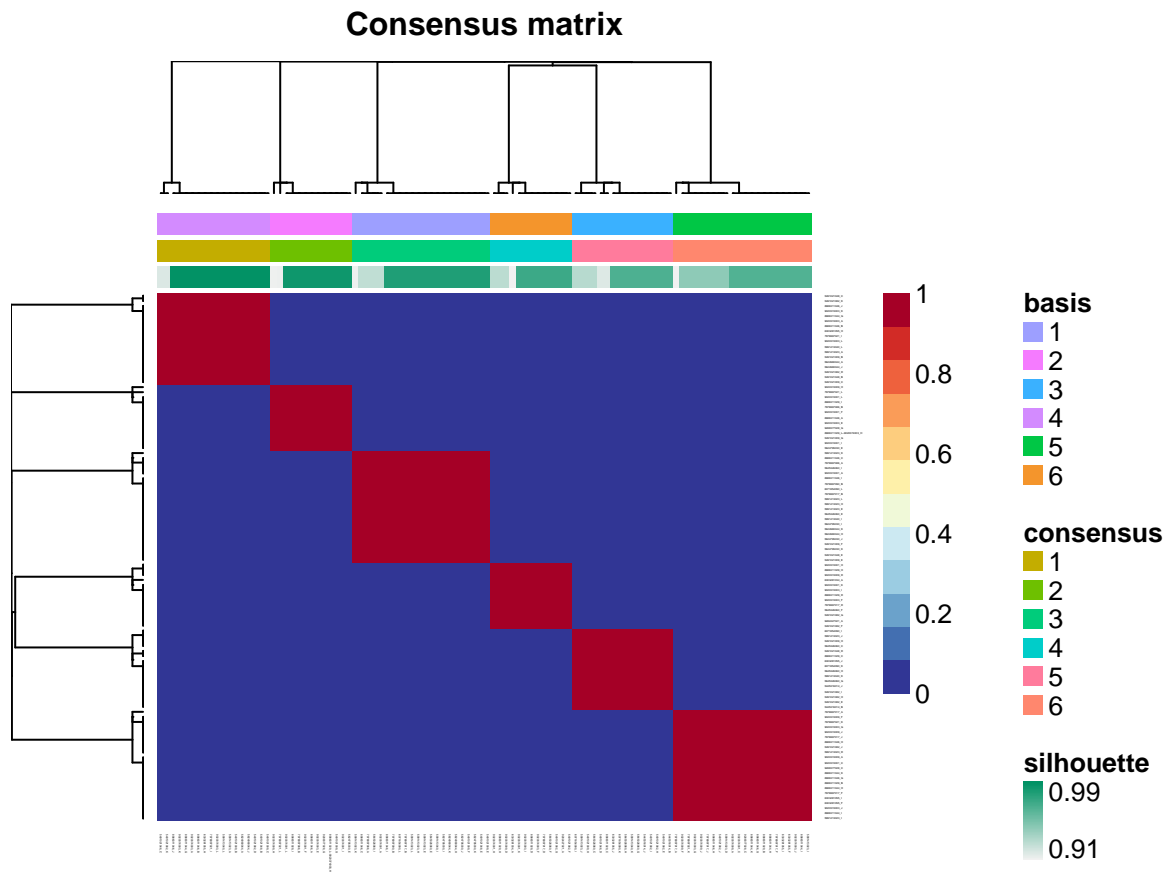
Consensus matrix

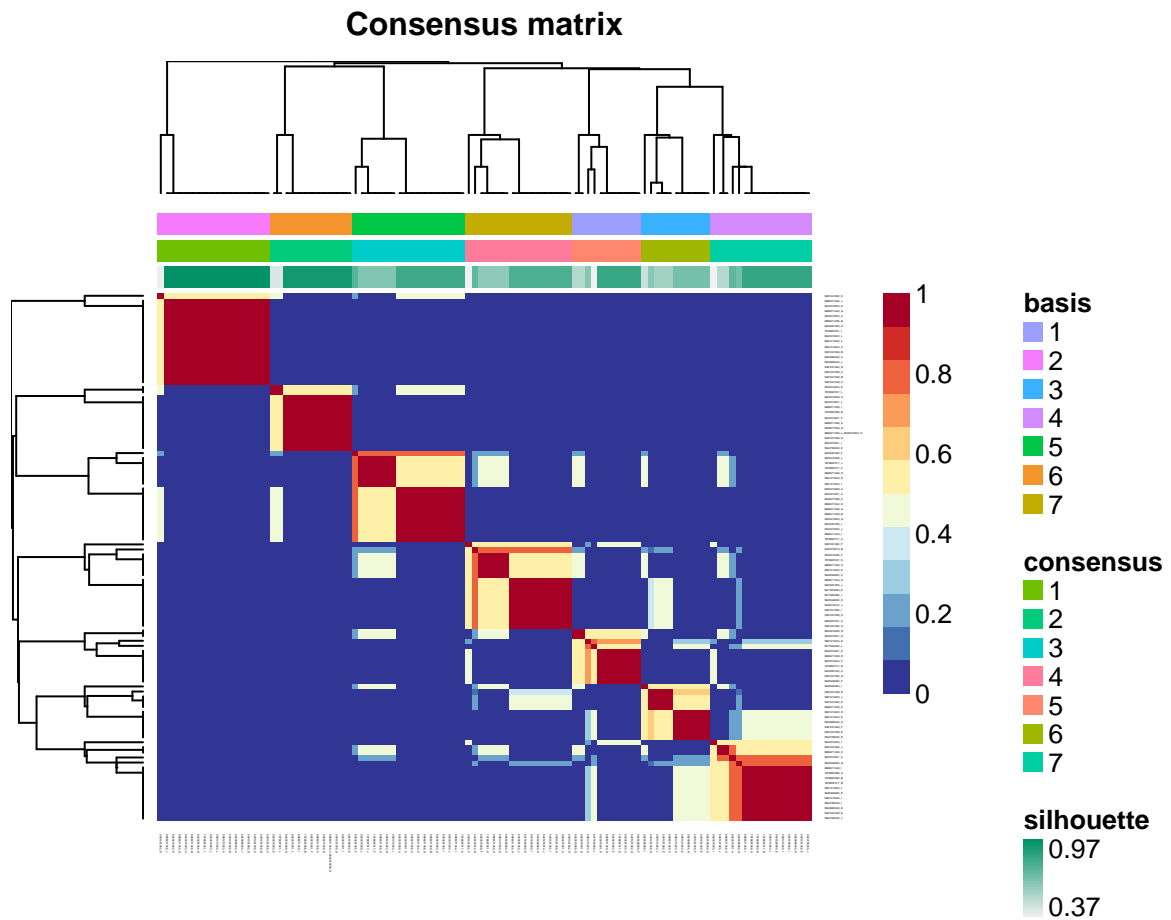


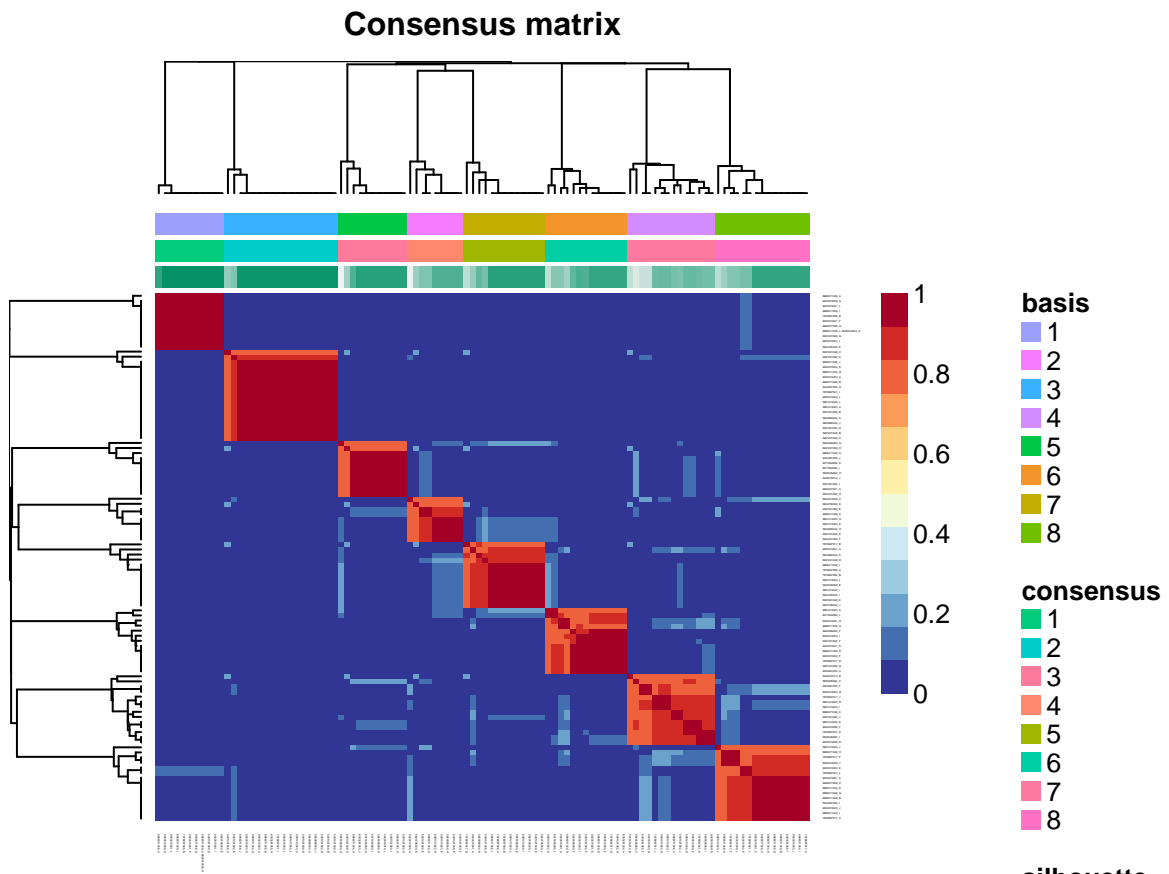
Consensus matrix

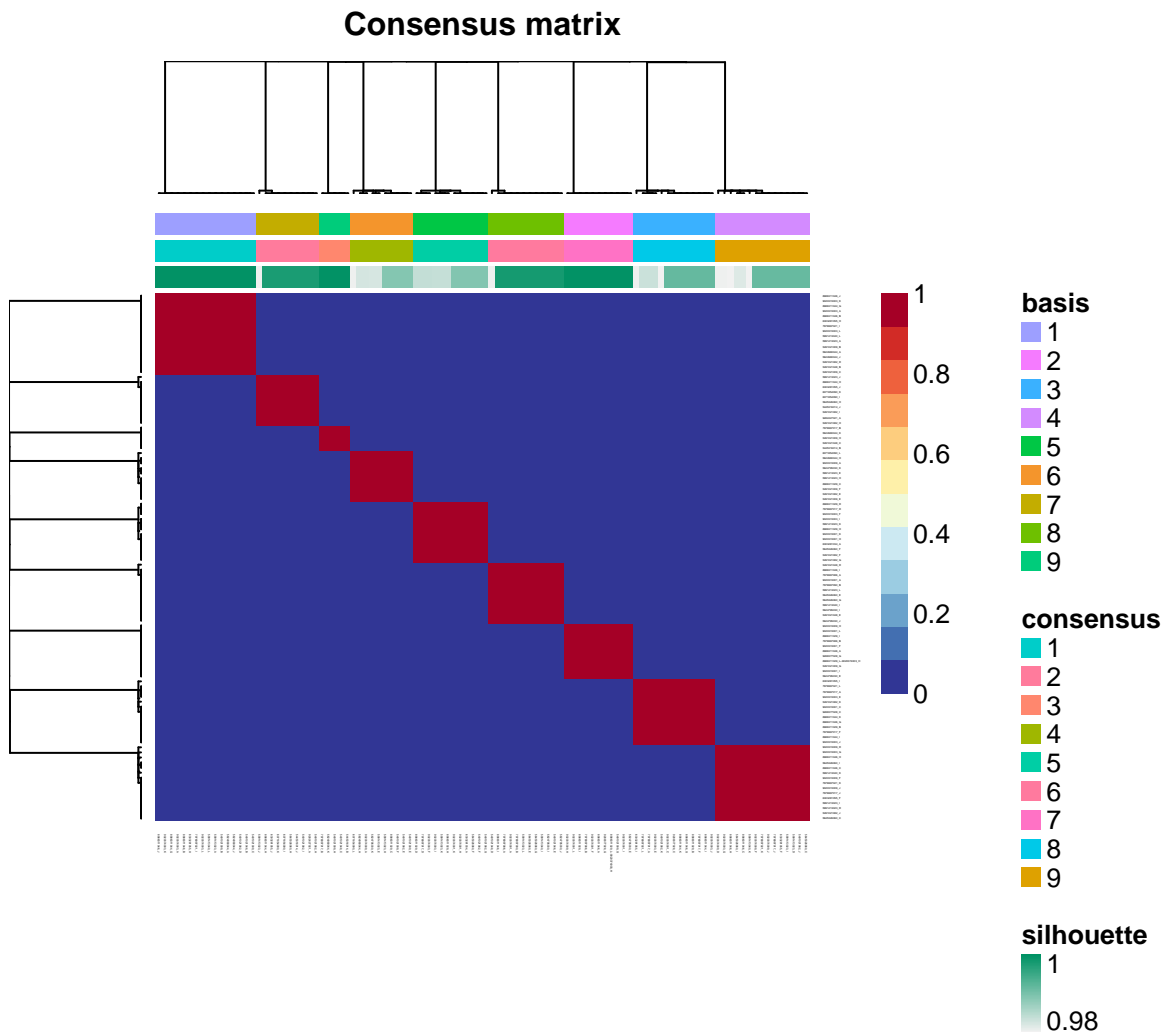


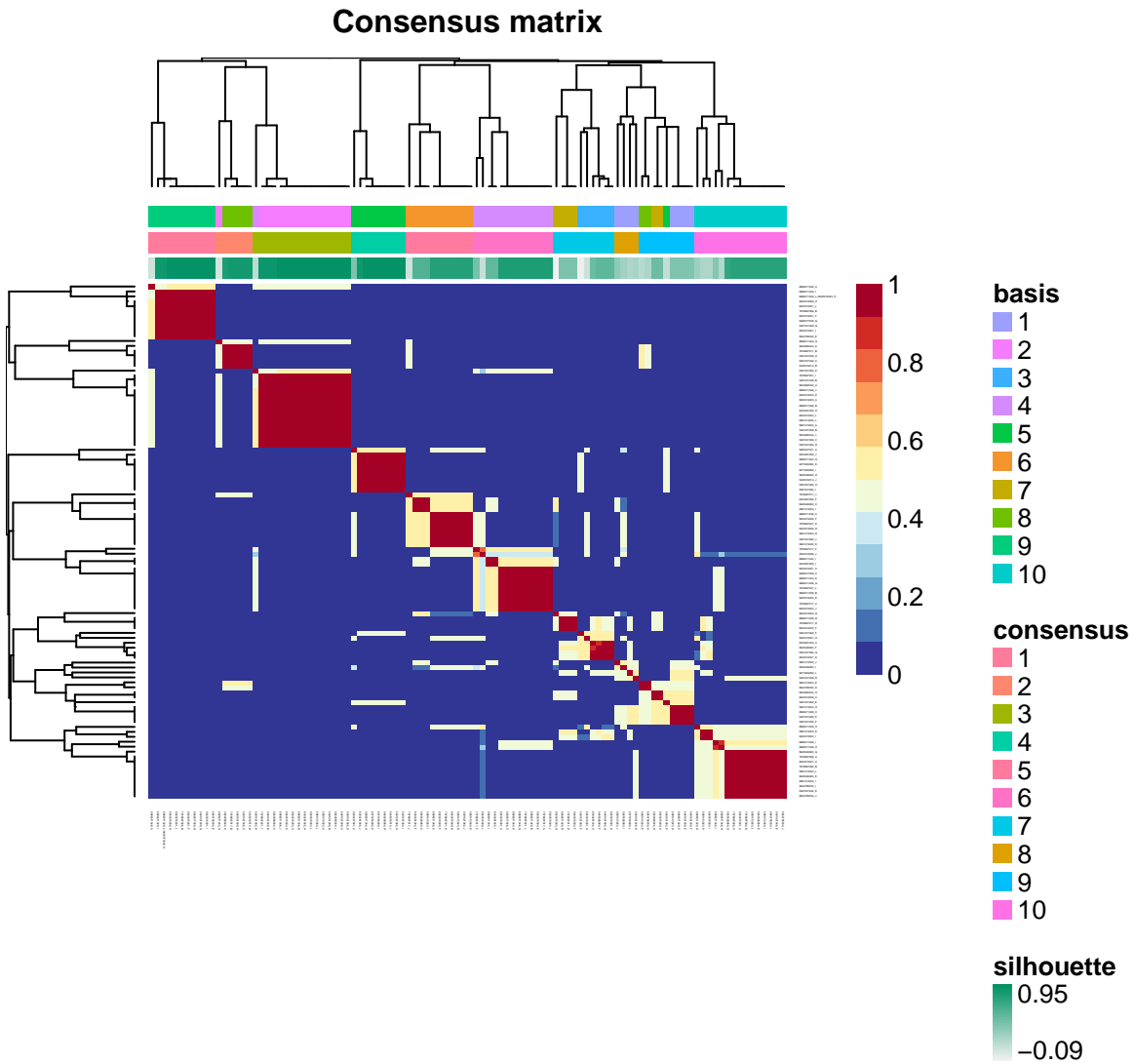




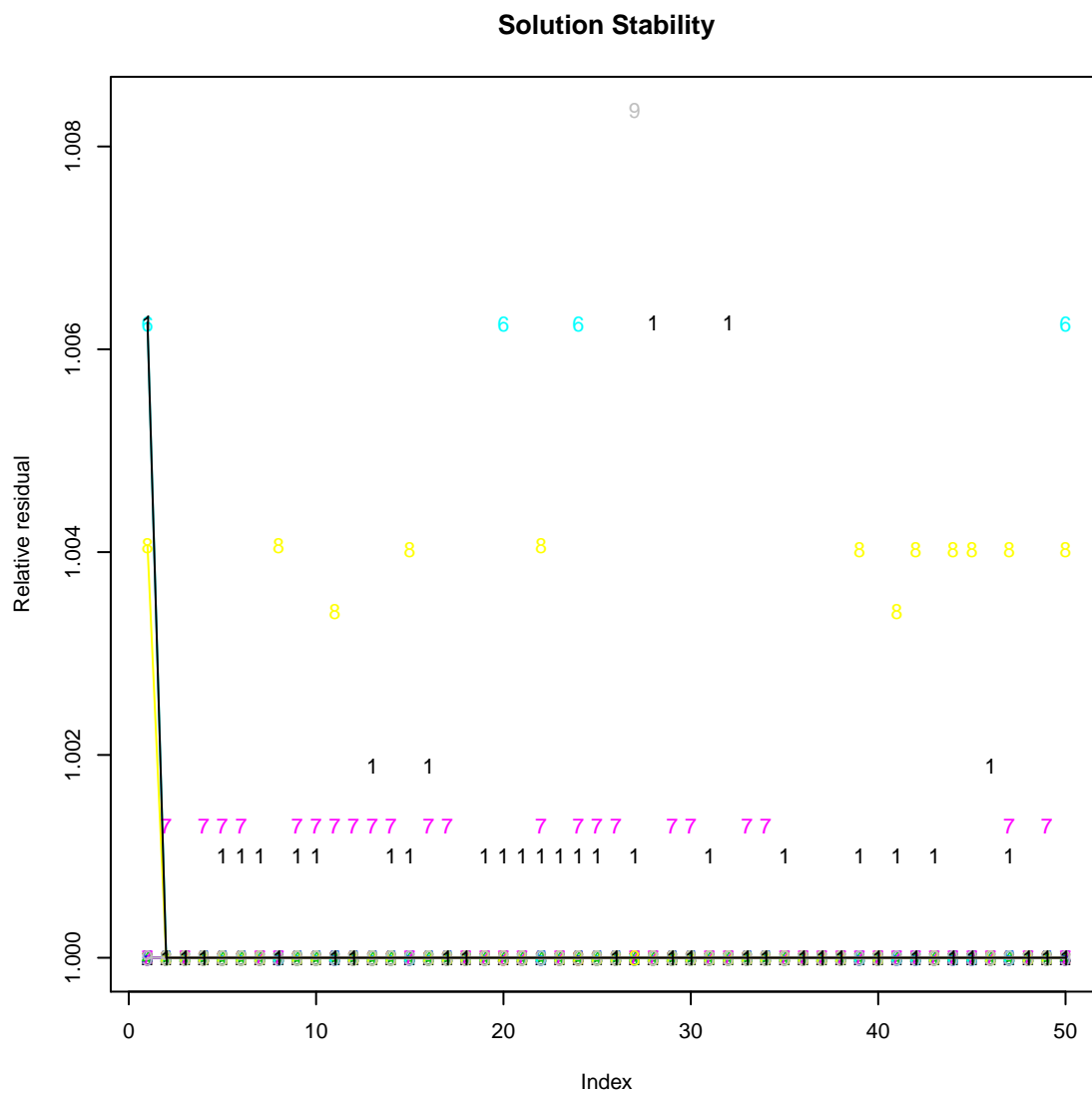




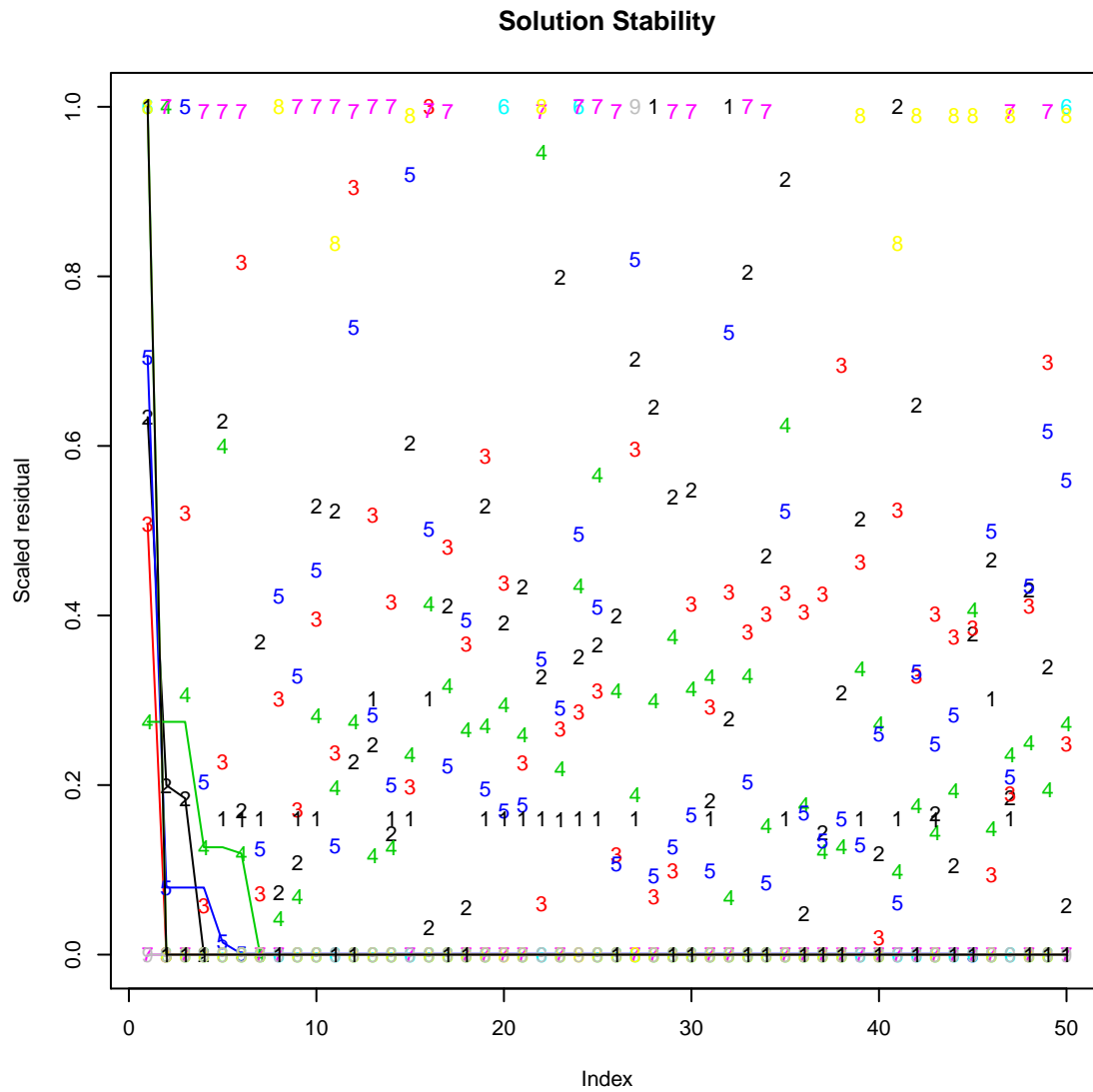




```
plot(0 ~ 0, type = "n", xlim = c(1, nrow(temp.resids)), ylim = range(temp.resids_rel),
     ylab = "Relative residual", main = "Solution Stability")
for (i in 1:ncol(temp.resids)) {
  points(temp.resids_rel[, i], col = i, pch = colnames(temp.resids)[i])
  lines(cummin(temp.resids_rel[, i]), col =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

```

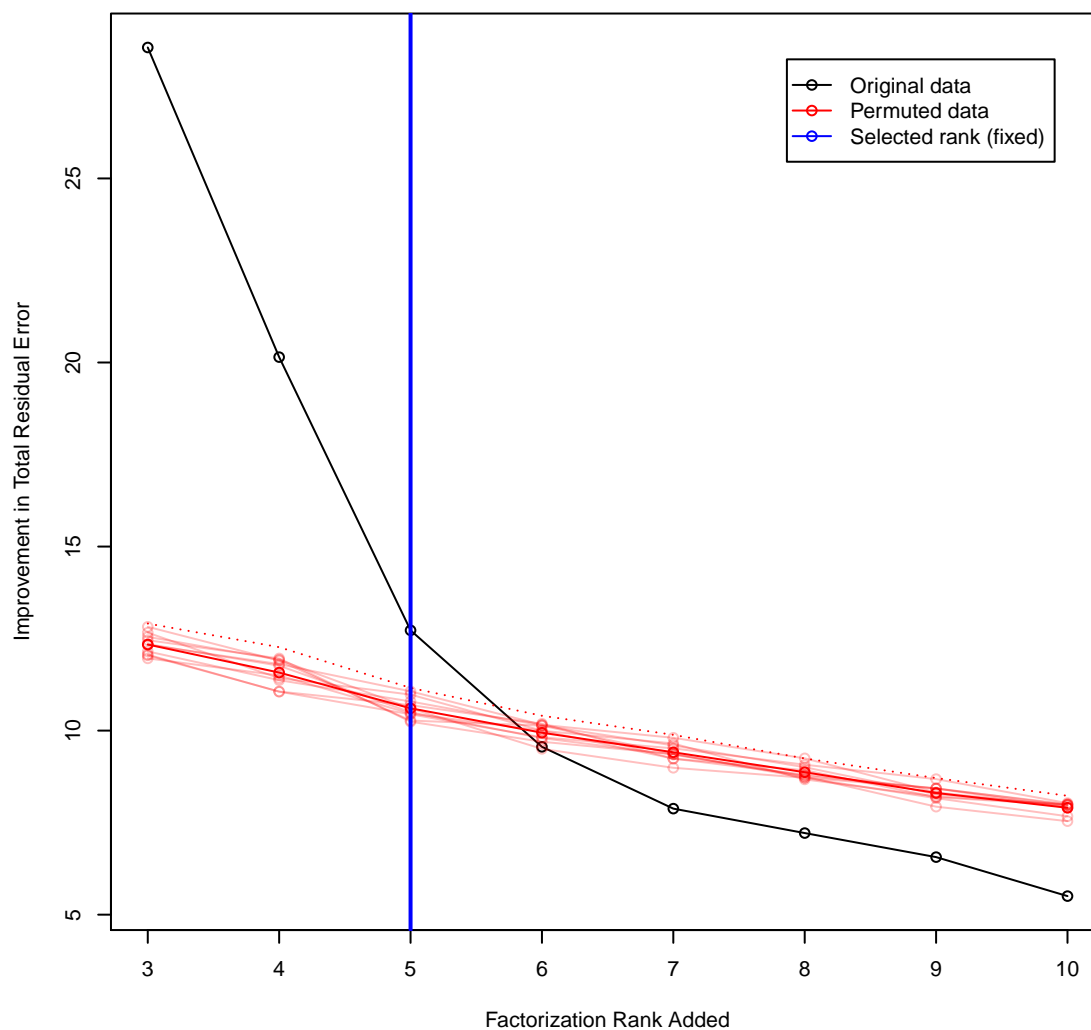
plot(nmf.rankrange[-1], -temp.orig_resids.delta, type = "o", col = "black",
     pch = 21, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.delta.mean)),
     xlab = "Factorization Rank Added", ylab = "Improvement in Total Residual Error")
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean,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

```

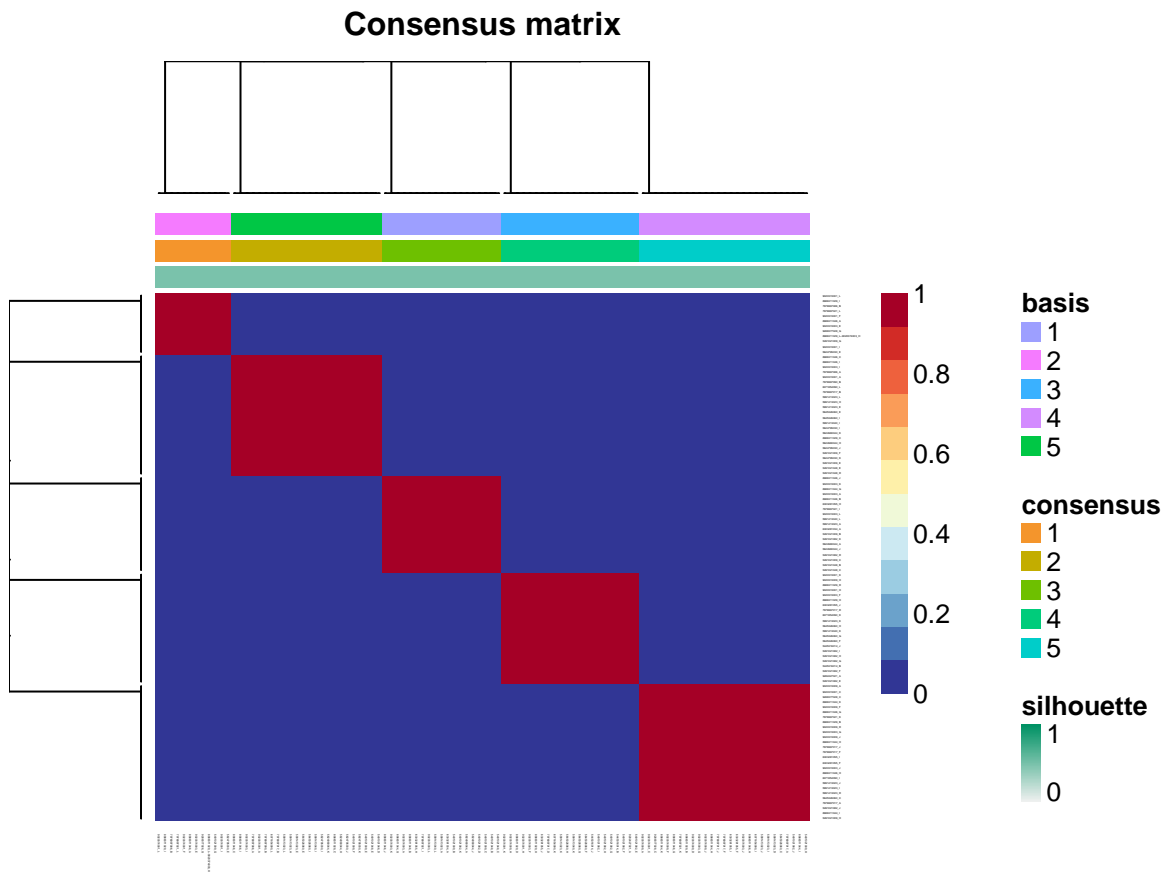


4.2 Fit

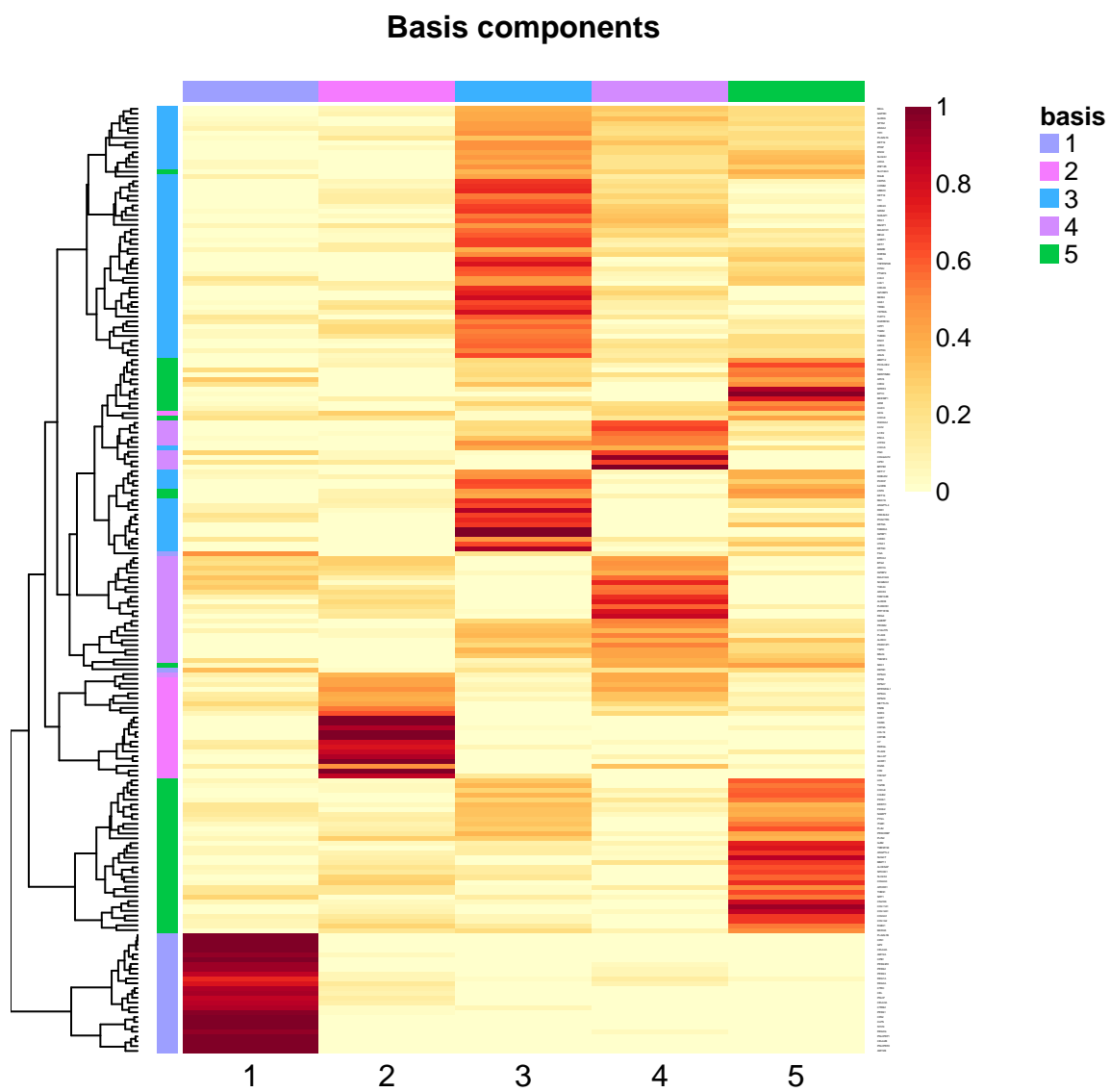
```

consensusmap(xlin.scaled.sel.nmf)

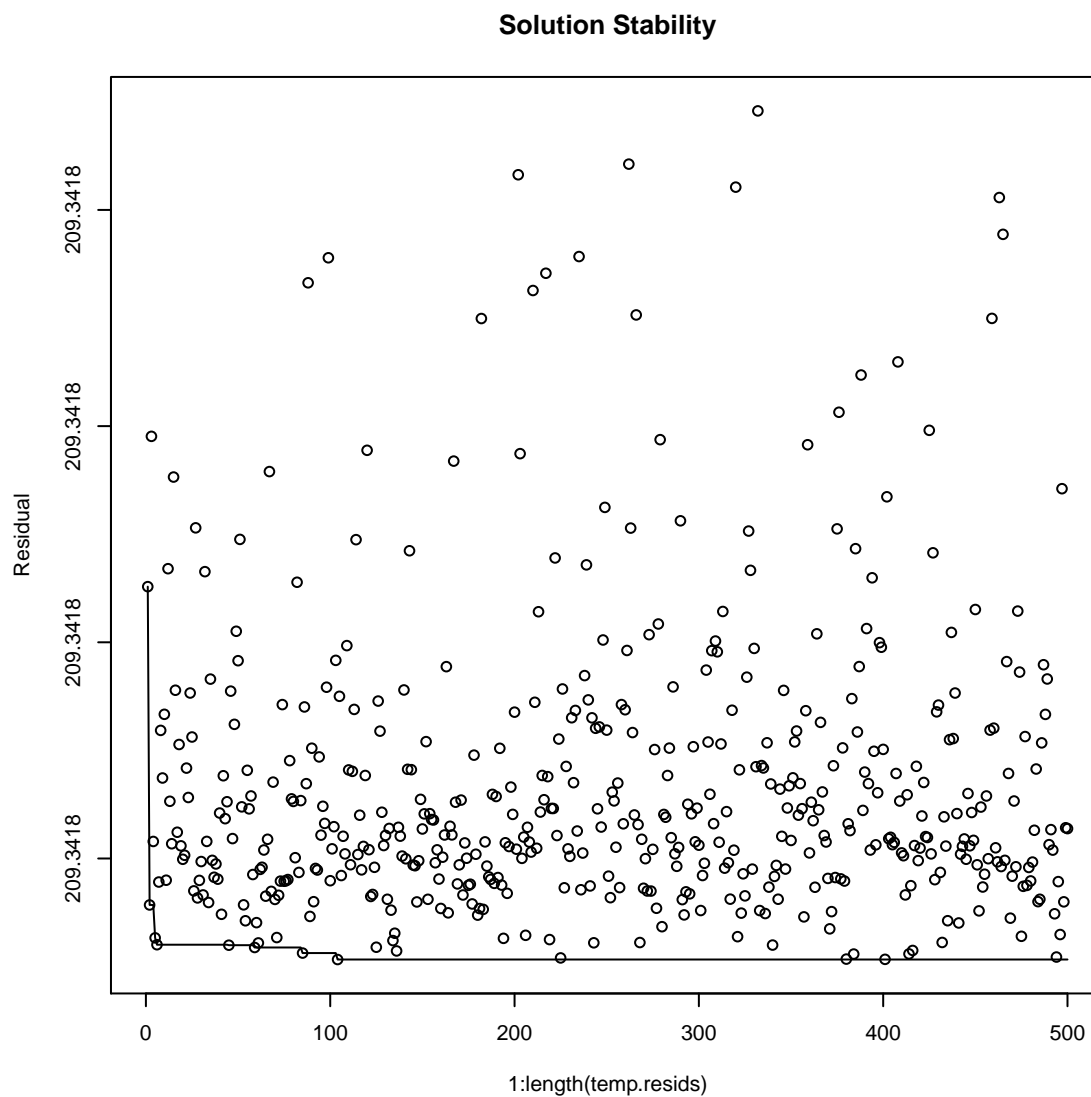
```



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



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

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= 104, number of events= 77

##

	coef	exp(coef)	se(coef)	z	Pr(> z)
## coef(xlin.scaled.sel.nmf)[i,]	-1.214	0.297	1.048	-1.16	0.25

##

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

Table 1:

<i>Dependent variable:</i>	
	y
coef(xlin.scaled.sel.nmf)[i,]	-1.214 (1.048)
Observations	104
R ²	0.014
Max. Possible R ²	0.997
Log Likelihood	-307.300
Wald Test	1.340 (df = 1)
LR Test	1.440 (df = 1)
Score (Logrank) Test	1.353 (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,]	-6.251*** (1.702)
Observations	104
R ²	0.148
Max. Possible R ²	0.997
Log Likelihood	-299.700
Wald Test	13.490*** (df = 1)
LR Test	16.660*** (df = 1)
Score (Logrank) Test	14.060*** (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,]	7.852*** (1.302)
Observations	104
R ²	0.265
Max. Possible R ²	0.997
Log Likelihood	-292.000
Wald Test	36.380*** (df = 1)
LR Test	32.030*** (df = 1)
Score (Logrank) Test	38.560*** (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,]	-1.481 (1.298)
Observations	104
R ²	0.013
Max. Possible R ²	0.997
Log Likelihood	-307.400
Wald Test	1.300 (df = 1)
LR Test	1.351 (df = 1)
Score (Logrank) Test	1.306 (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,]	4.076*** (1.172)
Observations	104
R ²	0.102
Max. Possible R ²	0.997
Log Likelihood	-302.500
Wald Test	12.100*** (df = 1)
LR Test	11.160*** (df = 1)
Score (Logrank) Test	12.500*** (df = 1)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

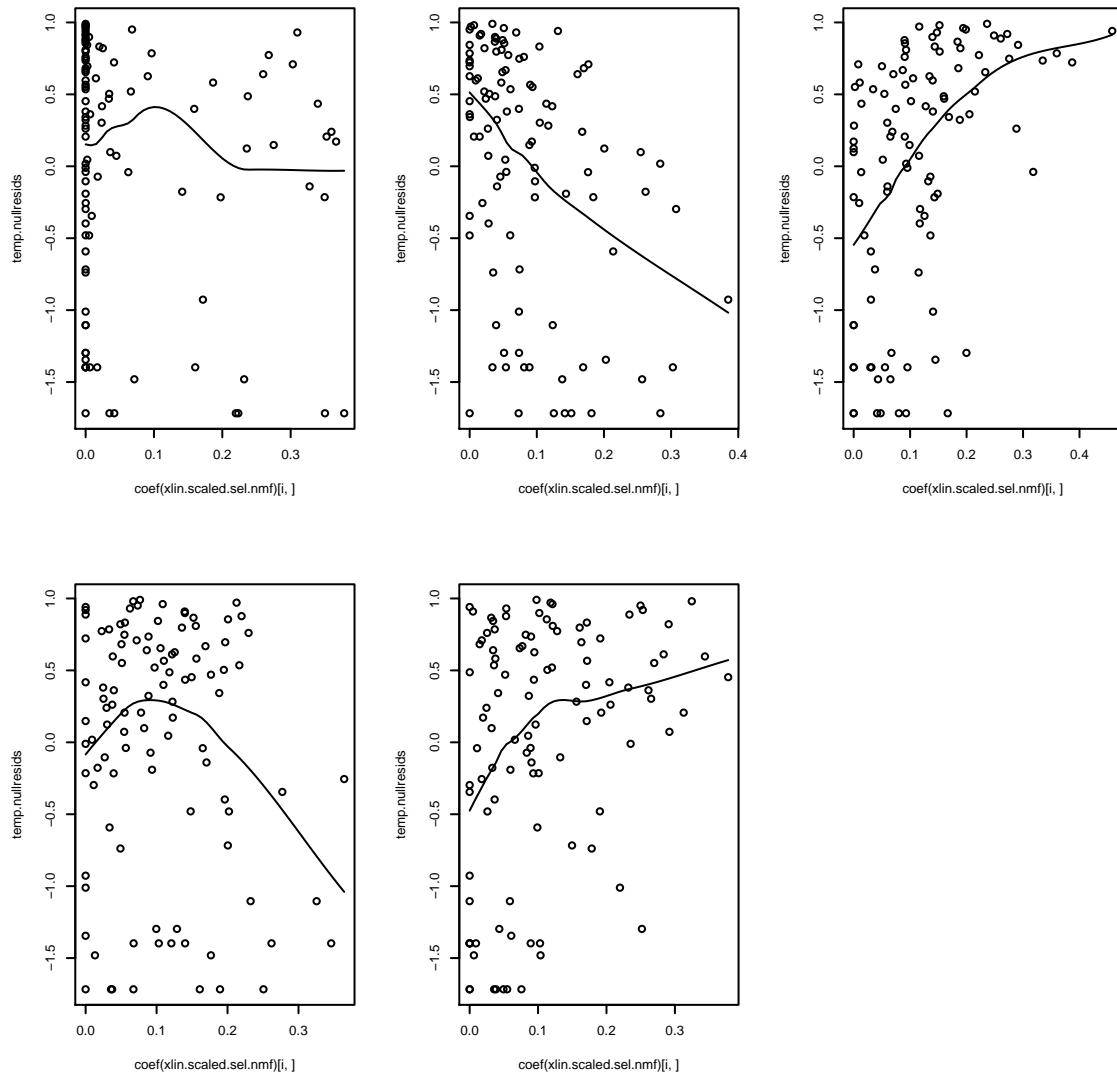

```

## coef(xlin.scaled.sel.nmf)[i, ]      0.297      3.37      0.0381      2.32
##
## Concordance= 0.548 (se = 0.035 )
## Rsquare= 0.014 (max possible= 0.997 )
## Likelihood ratio test= 1.44 on 1 df, p=0.23
## Wald test          = 1.34 on 1 df, p=0.247
## Score (logrank) test = 1.35 on 1 df, p=0.245
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 104, number of events= 77
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ] -6.25093  0.00193  1.70215 -3.67  0.00024
##
##              exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]  0.00193      518 6.86e-05  0.0542
##
## Concordance= 0.652 (se = 0.036 )
## Rsquare= 0.148 (max possible= 0.997 )
## Likelihood ratio test= 16.7 on 1 df, p=4.47e-05
## Wald test          = 13.5 on 1 df, p=0.00024
## Score (logrank) test = 14.1 on 1 df, p=0.000177
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 104, number of events= 77
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ]  7.85  2570.58      1.30 6.03  1.6e-09
##
##              exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]  2571  0.000389      200  32965
##
## Concordance= 0.694 (se = 0.036 )
## Rsquare= 0.265 (max possible= 0.997 )
## Likelihood ratio test= 32 on 1 df, p=1.51e-08
## Wald test          = 36.4 on 1 df, p=1.62e-09
## Score (logrank) test = 38.6 on 1 df, p=5.31e-10
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 104, number of events= 77
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ] -1.481  0.227  1.298 -1.14  0.25
##
##              exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]  0.227      4.4  0.0178  2.9
##

```

```
## Concordance= 0.51 (se = 0.036 )
## Rsquare= 0.013 (max possible= 0.997 )
## Likelihood ratio test= 1.35 on 1 df, p=0.245
## Wald test = 1.3 on 1 df, p=0.254
## Score (logrank) test = 1.31 on 1 df, p=0.253
##
## Call:
## coxph(formula = y ~ coef(xlin.scaled.sel.nmf)[i, ])
##
## n= 104, number of events= 77
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## coef(xlin.scaled.sel.nmf)[i, ]  4.08    58.92    1.17 3.48    5e-04
##
##               exp(coef) exp(-coef) lower .95 upper .95
## coef(xlin.scaled.sel.nmf)[i, ]    58.9    0.017    5.93    586
##
## Concordance= 0.605 (se = 0.036 )
## Rsquare= 0.102 (max possible= 0.997 )
## Likelihood ratio test= 11.2 on 1 df, p=0.000835
## Wald test = 12.1 on 1 df, p=0.000504
## Score (logrank) test = 12.5 on 1 df, p=0.000407
```

```
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"))

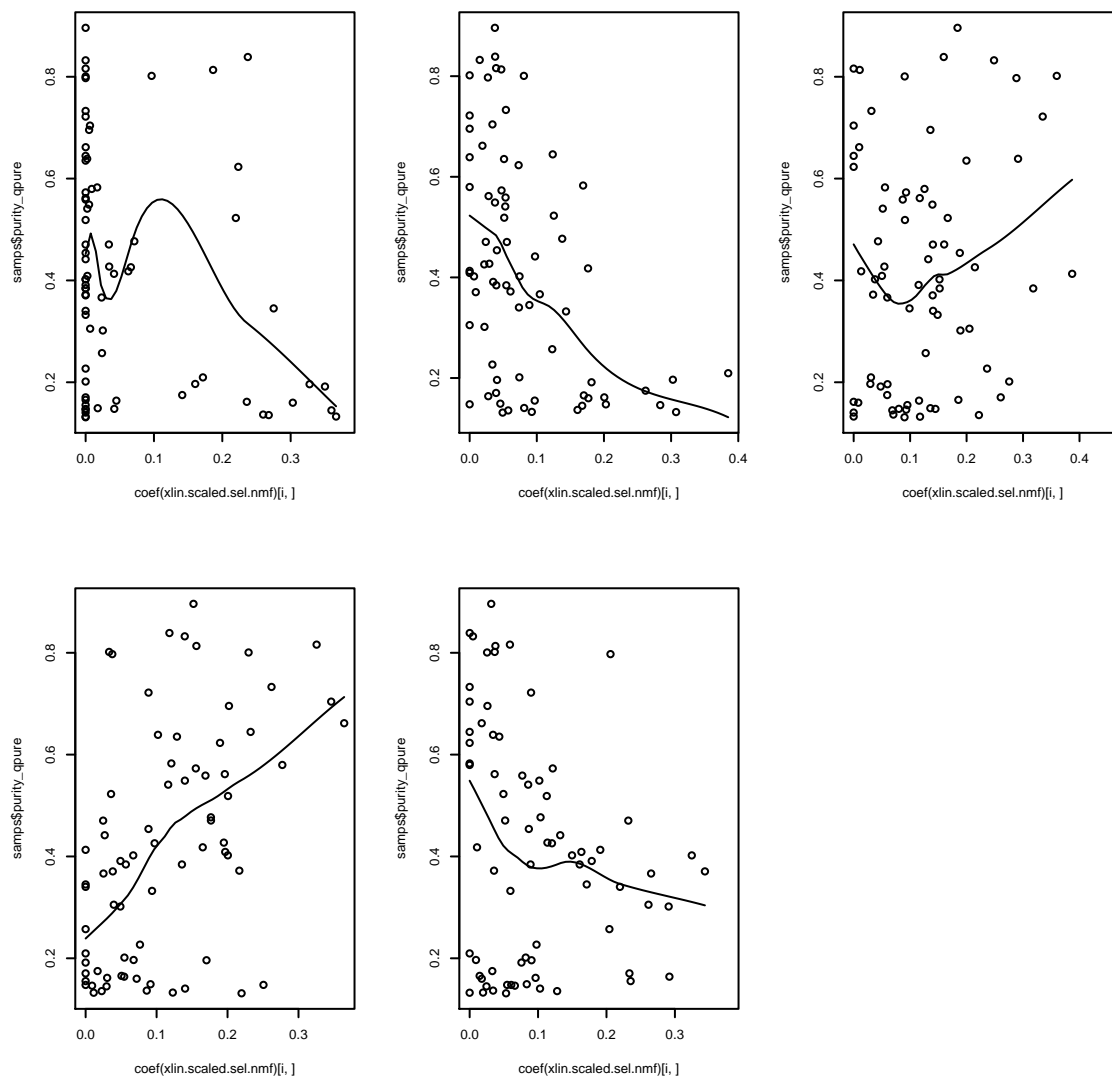
## [[1]]
##
## Kendall's rank correlation tau
##
## data: samps$purity_qpure and xc
## z = -1.799, p-value = 0.07208
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1526
##
```

```

##
## [[2]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -3.877, p-value = 0.0001056
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.3094
##
##
## [[3]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = 0.7237, p-value = 0.4692
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.05762
##
##
## [[4]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = 4.158, p-value = 3.216e-05
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.3318
##
##
## [[5]]
##
## Kendall's rank correlation tau
##
## data:  samps$purity_qpure and xc
## z = -2.139, p-value = 0.03244
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1705

par(mfrow = c(2, 3))
for (i in 1:nrow(coef(xlin.scaled.sel.nmf))) {
  scatter.smooth(samps$purty_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 2.302e-01 7.208e-02    -1    -1  6.906e-01  0.2883331 7.976e-01
## 2 4.471e-05 1.056e-04    -1    -1  3.577e-04  0.0007393 4.365e-04
## 3 1.515e-08 4.692e-01     1     1  1.515e-07  0.6905629 4.437e-07
## 4 2.451e-01 3.216e-05    -1     1  6.906e-01  0.0002894 7.976e-01
## 5 8.347e-04 3.244e-02     1    -1  5.008e-03  0.1622146 4.890e-03
##      q.pure.BY
## 1 0.3016137
## 2 0.0007733
## 3 1.0000000
## 4 0.0004365
## 5 0.1583738

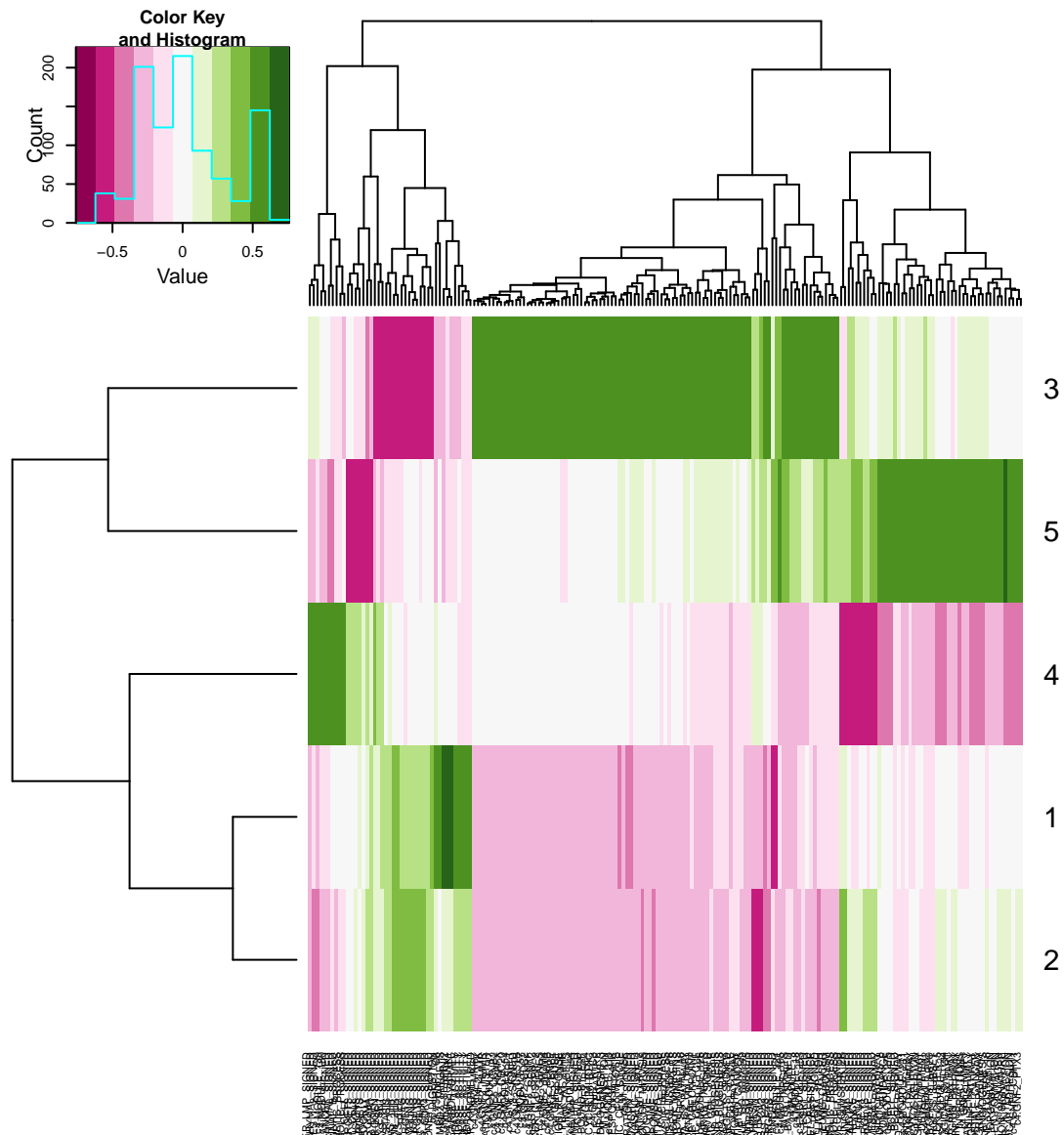
```

4.6 MSigDB score correlation thresholding

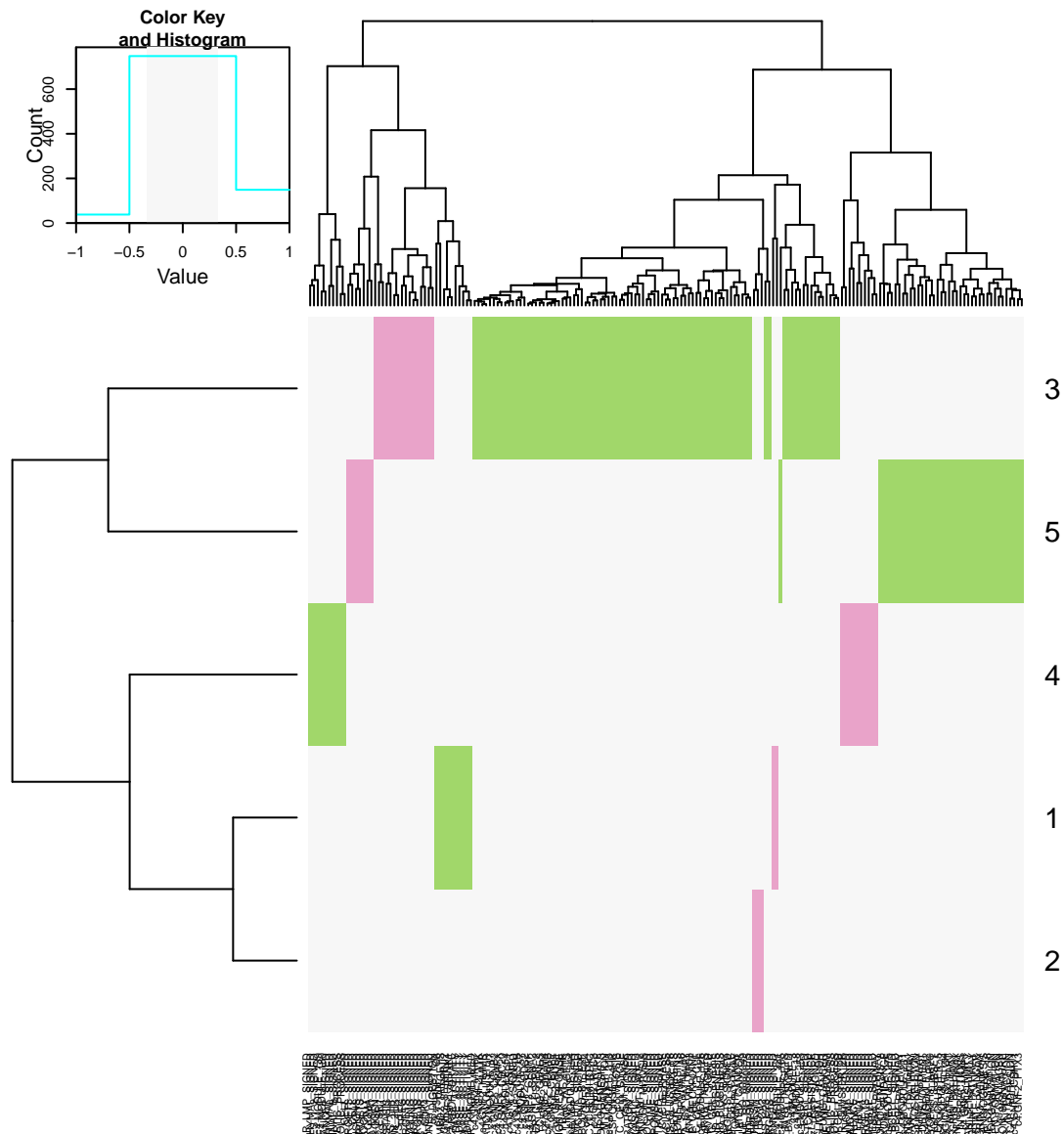
```

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$d.survival),
              collapse = ",")
          })
      table = table[order(-(table$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.BERGER_MBD2_TARGETS      0.6812      -1
## 2 c2.REACTOME_DIGESTION_OF_DIETARY_CARBOHYDRATE      0.5101      -1
## 3      c2.GRUETZMANN_PANCREATIC_CANCER_SIGNED     -0.5020       1
## 4      c2.IGLESIAS_E2F_TARGETS_SIGNED     -0.5260       1
##
## [[1]]$c3
##      GeneSet Correlation Metagenes
## 1 c3.V$HNF1_Q6      0.5106      -1
##
## [[1]]$c4
##      GeneSet Correlation Metagenes
## 1  c4.GNF2_SPINK1      0.760      -1
## 2  c4.GNF2_SERPINI2      0.742      -1
##
## [[1]]$c5
##      GeneSet Correlation Metagenes
## 1  c5.SERINE_TYPE_PEPTIDASE_ACTIVITY      0.5230      -1
## 2      c5.CARBOXYPEPTIDASE_ACTIVITY      0.5195      -1
## 3      c5.SERINE_HYDROLASE_ACTIVITY      0.5166      -1
## 4 c5.SERINE_TYPE_ENDOPEPTIDASE_ACTIVITY      0.5166      -1
## 5      c5.DIGESTION      0.5144      -1
##
## [[1]]$c6
## data frame with 0 columns and 0 rows
##
## [[1]]$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.HAHTOLA_SEZARY_SYNDROM_SIGNED -0.5039 2
## 2 c2.BOQUEST_STEM_CELL_CULTURED_VS_FRESH_SIGNED -0.5087 2
## 3 c2.WINTER_HYPOXIA_SIGNED -0.5230 2
##
## [[2]]$c3
## data frame with 0 columns and 0 rows
##
## [[2]]$c4
## data frame with 0 columns and 0 rows
##
## [[2]]$c5
## data frame with 0 columns and 0 rows
##
## [[2]]$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
## 1 c2.FOURNIER_ACINAR_DEVELOPMENT_LATE_2
## 2 c2.SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6
## 3 c2.WHITEFORD_PEDIATRIC_CANCER_MARKERS
## 4 c2.BASAKI_YBX1_TARGETS_SIGNED
## 5 c2.WU_APOPTOSIS_BY_CDKN1A_VIA_TP53
## 6 c2.GAVIN_FOXP3_TARGETS_CLUSTER_P6
## 7 c2.CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_SIGNED
## 8 c2.RHODES_UNDIFFERENTIATED_CANCER
## 9 c2.SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_SIGNED
## 10 c2.REACTOME_KINESINS
## 11 c2.BIOCARTA_G2_PATHWAY
## 12 c2.ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER
## 13 c2.SMID_BREAST_CANCER_BASAL_SIGNED
## 14 c2.RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_SIGNED
## 15 c2.BURTON_ADIPOGENESIS_PEAK_AT_24HR
## 16 c2.REACTOME_PYRIMIDINE_METABOLISM
## 17 c2.OHASHI_AURKB_TARGETS
## 18 c2.HOFFMANN_LARGE_TO_SMALL_PRE_BII_LYMPHOCYTE_SIGNED
## 19 c2.ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR
## 20 c2.LOPEZ_MESOTELIOMA_SURVIVAL_TIME_SIGNED
## 21 c2.WANG_METASTASIS_OF_BREAST_CANCER_ESR1_SIGNED
## 22 c2.FARMER_BREAST_CANCER_CLUSTER_2

```

```

## 23          c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_SIGNED
## 24          c2.ZHAN_MULTIPLE_MYELOMA_PR_SIGNED
## 25          c2.FINETTI_BREAST_CANCER_KINOME_RED
## 26          c2.BURTON_ADIPOGENESIS_3
## 27          c2.REN_BOUND_BY_E2F
## 28          c2.PID_FOXM1PATHWAY
## 29          c2.CHANG_CYCLING_GENES
## 30          c2.AMUNDSON_GAMMA_RADIATION_RESPONSE
## 31          c2.CHANG_CORE_SERUM_RESPONSE_SIGNED
## 32          c2.LEI_MYB_TARGETS
## 33          c2.MONTERO_THYROID_CANCER_POOR_SURVIVAL_SIGNED
## 34          c2.KANG_IMMORTALIZED_BY_TERT_SIGNED
## 35          c2.BOYVAULT_LIVER_CANCER_SUBCLASS_G23_SIGNED
## 36          c2.HONMA_DOCETAXEL_RESISTANCE
## 37          c2.LI_LUNG_CANCER
## 38          c2.WHITFIELD_CELL_CYCLE_LITERATURE
## 39          c2.REACTOME_CELL_CYCLE_MITOTIC
## 40          c2.NGO_MALIGNANT_GLIOMA_1P_LOH
## 41 c2.REACTOME_CYCLIN_A_B1_ASSOCIATED_EVENTS_DURING_G2_M_TRANSITION
## 42          c2.REICHERT_MITOSIS_LIN9_TARGETS
## 43          c2.CHUNG_BLISTER_CYTOTOXICITY_SIGNED
## 44          c2.KEGG_CELL_CYCLE
## 45          c2.RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_SIGNED
## 46          c2.HONRADO_BREAST_CANCER_BRCA1_VS_BRCA2
## 47          c2.DOANE_BREAST_CANCER_ESR1_SIGNED
## 48          c2.VANTVEER_BREAST_CANCER_ESR1_SIGNED
## 49          c2.LI_WILMS_TUMOR_VS_FETAL_KIDNEY_1_SIGNED
## 50          c2.BROWNE_HCMV_INFECTION_2HR_SIGNED
## 51          c2.MOLENAAR_TARGETS_OF_CCND1_AND_CDK4_SIGNED
## 52          c2.RUIZ_TNC_TARGETS_SIGNED
## 53          c2.KOBAYASHI_EGFR_SIGNALING_24HR_SIGNED
## 54          c2.STEIN_ESRRA_TARGETS_RESPONSIVE_TO_ESTROGEN_SIGNED
## 55          c2.VANTVEER_BREAST_CANCER_METASTASIS_SIGNED
## 56          c2.TANG_SENESCENCE_TP53_TARGETS_SIGNED
##      Correlation Metagenes
## 1      0.6043      3
## 2      0.5930      3
## 3      0.5897      3
## 4      0.5807      3
## 5      0.5758      3
## 6      0.5683      3
## 7      0.5597      3
## 8      0.5548      3
## 9      0.5533      3
## 10     0.5522      3
## 11     0.5409      3
## 12     0.5402      3
## 13     0.5387      3
## 14     0.5349      3
## 15     0.5338      3
## 16     0.5334      3
## 17     0.5301      3
## 18     0.5274      3
## 19     0.5252      3

```

```

## 20      0.5237      3
## 21      0.5233      3
## 22      0.5214      3
## 23      0.5196      3
## 24      0.5192      3
## 25      0.5188      3
## 26      0.5177      3
## 27      0.5169      3
## 28      0.5154      3
## 29      0.5132      3
## 30      0.5124      3
## 31      0.5117      3
## 32      0.5113      3
## 33      0.5106      3
## 34      0.5083      3
## 35      0.5079      3
## 36      0.5072      3
## 37      0.5064      3
## 38      0.5053      3
## 39      0.5042      3
## 40      0.5038      3
## 41      0.5027      3
## 42      0.5019      3
## 43      0.5016      3
## 44      0.5016      3
## 45      0.5012      3
## 46      0.5012      3
## 47     -0.5042     -3
## 48     -0.5083     -3
## 49     -0.5106     -3
## 50     -0.5128     -3
## 51     -0.5293     -3
## 52     -0.5391     -3
## 53     -0.5593     -3
## 54     -0.5720     -3
## 55     -0.5837     -3
## 56     -0.6013     -3
##
## [[3]]$c3
## data frame with 0 columns and 0 rows
##
## [[3]]$c4
##      GeneSet Correlation Metagenes
## 1  c4.MODULE_54      0.5649      3
## 2  c4.GNF2_ESPL1      0.5593      3
## 3  c4.GNF2_CDC20      0.5518      3
## 4  c4.MODULE_105      0.5432      3
## 5  c4.GNF2_CKS1B      0.5417      3
## 6  c4.GNF2_H2AFX      0.5402      3
## 7   c4.MODULE_18      0.5368      3
## 8  c4.GNF2_MKI67      0.5364      3
## 9  c4.MODULE_126      0.5361      3
## 10 c4.MODULE_57      0.5334      3
## 11 c4.GNF2_HMMR      0.5327      3

```

```

## 12 c4.MODULE_253      0.5304      3
## 13 c4.GNF2_CCNA2      0.5293      3
## 14 c4.GNF2_CCNB2      0.5282      3
## 15   c4.GNF2_TTK      0.5282      3
## 16 c4.GNF2_CENPF      0.5271      3
## 17 c4.GNF2_CDC2      0.5248      3
## 18 c4.GNF2_RFC4      0.5211      3
## 19 c4.MODULE_17      0.5211      3
## 20   c4.MODULE_3      0.5162      3
## 21 c4.GNF2_BUB1      0.5147      3
## 22 c4.GNF2_RFC3      0.5094      3
## 23 c4.GNF2_MCM4      0.5076      3
## 24 c4.MODULE_244      0.5053      3
## 25 c4.GNF2_BUB1B      0.5038      3
## 26 c4.GNF2_CENPE      0.5008      3
## 27 c4.MODULE_315      0.5001      3
##
## [[3]]$c5
##
##                                     GeneSet Correlation
## 1                                c5.M_PHASE_OF_MITOTIC_CELL_CYCLE      0.5717
## 2                                c5.CELL_CYCLE_GO_0007049      0.5694
## 3                                c5.MITOSIS      0.5679
## 4                                c5.M_PHASE      0.5672
## 5                                c5.MITOTIC_CELL_CYCLE      0.5627
## 6                                c5.CELL_CYCLE_CHECKPOINT_GO_0000075      0.5499
## 7                                c5.REGULATION_OF_MITOSIS      0.5484
## 8                                c5.CELL_CYCLE_PROCESS      0.5424
## 9                                c5.CELL_CYCLE_PHASE      0.5379
## 10                               c5.NUCLEOTIDE_METABOLIC_PROCESS      0.5241
## 11                               c5.SPINDLE_ORGANIZATION_AND_BIOGENESIS      0.5226
## 12                               c5.MITOTIC_CELL_CYCLE_CHECKPOINT      0.5192
## 13                               c5.REGULATION_OF_CELL_CYCLE      0.5188
## 14 c5.NUCLEOBASENUCLEOSIDE_AND_NUCLEOTIDE_METABOLIC_PROCESS      0.5109
## 15                               c5.MITOTIC_SPINDLE_ORGANIZATION_AND_BIOGENESIS      0.5083
##
##      Metagenes
## 1           3
## 2           3
## 3           3
## 4           3
## 5           3
## 6           3
## 7           3
## 8           3
## 9           3
## 10          3
## 11          3
## 12          3
## 13          3
## 14          3
## 15          3
##
## [[3]]$c6
## data frame with 0 columns and 0 rows
##

```

```

## [[3]]$c7
##
##                                     GeneSet
## 1 c7.GSE30962_PRIMARY_VS_SECONDARY_ACUTE_LCMV_INF_CD8_TCELL_SIGNED
## 2                                     c7.GOLDRATH_EFF_VS_MEMORY_CD8_TCELL_SIGNED
## 3                                     c7.GSE24634_NAIVE_CD4_TCELL_VS_DAY3_IL4_CONV_TREG_SIGNED
## 4                                     c7.GSE3982_CENT_MEMORY_CD4_TCELL_VS_TH2_SIGNED
## 5 c7.GSE36476_CTRL_VS_TSST_ACT_72H_MEMORY_CD4_TCELL_YOUNG_SIGNED
## 6                                     c7.GSE3982_MEMORY_CD4_TCELL_VS_TH1_SIGNED
## 7                                     c7.GSE3982_BCELL_VS_TH1_SIGNED
## 8 c7.GSE36476_CTRL_VS_TSST_ACT_40H_MEMORY_CD4_TCELL_YOUNG_SIGNED
## Correlation Metagenes
## 1      0.5615      3
## 2      0.5019      3
## 3     -0.5031     -3
## 4     -0.5049     -3
## 5     -0.5188     -3
## 6     -0.5259     -3
## 7     -0.5417     -3
## 8     -0.5473     -3
##
##
## [[4]]
## [[4]]$c1
## data frame with 0 columns and 0 rows
##
## [[4]]$c2
##
##                                     GeneSet Correlation
## 1 c2.CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_SIGNED      0.5873
## 2                                     c2.SMID_BREAST_CANCER_LUMINAL_B_SIGNED      0.5685
## 3                                     c2.WAMUNYOKOLI_OVARIAN_CANCER_LMP_SIGNED      0.5159
## 4 c2.CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_SIGNED      0.5133
## 5                                     c2.LIU_PROSTATE_CANCER_SIGNED      0.5061
## 6 c2.WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_SIGNED      0.5057
## 7                                     c2.LINDGREN_BLADDER_CANCER_CLUSTER_2B     -0.5057
## 8 c2.SENGUPTA_NASOPHARYNGEAL_CARCINOMA_SIGNED     -0.5080
## 9                                     c2.WOO_LIVER_CANCER_RECURRENCE_SIGNED     -0.5185
## 10 c2.SCHUETZ_BREAST_CANCER_DUCTAL_INVASIVE_SIGNED     -0.5396
## 11                                     c2.LIM_MAMMARY_STEM_CELL_SIGNED     -0.5414
## 12 c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SIGNED     -0.5553
## 13                                     c2.ROY_WOUND_BLOOD_VESSEL_SIGNED     -0.5557
## 14 c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SIGNED     -0.5689
## 15 c2.SERVITJA_ISLET_HNF1A_TARGETS_SIGNED     -0.5719
## Metagenes
## 1      -4
## 2      -4
## 3      -4
## 4      -4
## 5      -4
## 6      -4
## 7       4
## 8       4
## 9       4
## 10      4

```

```

## 11      4
## 12      4
## 13      4
## 14      4
## 15      4
##
## [[4]]$c3
## data frame with 0 columns and 0 rows
##
## [[4]]$c4
##      GeneSet Correlation Metagenes
## 1 c4.MODULE_180      0.5760      -4
## 2 c4.MODULE_139      0.5317      -4
##
## [[4]]$c5
##      GeneSet Correlation Metagenes
## 1 c5.GLYCOPROTEIN_BIOSYNTHETIC_PROCESS      0.5223      -4
## 2 c5.GLYCOPROTEIN_METABOLIC_PROCESS      0.5110      -4
##
## [[4]]$c6
##      GeneSet Correlation Metagenes
## 1 c6.LEF1_UP.V1_SIGNED      -0.5907      4
##
## [[4]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[5]]
## [[5]]$c1
## data frame with 0 columns and 0 rows
##
## [[5]]$c2
##      GeneSet
## 1 c2.ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE
## 2 c2.PID_INTEGRIN5_PATHWAY
## 3 c2.PID_SYNDECAN_1_PATHWAY
## 4 c2.BURTON_ADIPOGENESIS_8
## 5 c2.PID_INTEGRIN1_PATHWAY
## 6 c2.KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
## 7 c2.POTTI_TOPOTECAN_SENSITIVITY
## 8 c2.KEGG_ECM_RECEPTOR_INTERACTION
## 9 c2.PID_INTEGRIN3_PATHWAY
## 10 c2.REACTOME_COLLAGEN_FORMATION
## 11 c2.PID_UPA_UPAR_PATHWAY
## 12 c2.KEGG_FOCAL_ADHESION
## 13 c2.REACTOME_CELL_EXTRACELLULAR_MATRIX_INTERACTIONS
## 14 c2.YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_16
## 15 c2.VERRECCHIA_EARLY_RESPONSE_TO_TGFB1
## 16 c2.REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
## 17 c2.HELLEBREKERS_SILENCED_DURING_TUMOR_ANGIOGENESIS
## 18 c2.PID_AVB3_INTEGRIN_PATHWAY
## 19 c2.FARMER_BREAST_CANCER_CLUSTER_5
## 20 c2.PID_INTEGRIN_CS_PATHWAY
## 21 c2.KEGG_REGULATION_OF_ACTIN_CYTOSKELETON

```

```

## 22          c2.YAMASHITA_METHYLATED_IN_PROSTATE_CANCER
## 23          c2.MAHADEVAN_GIST_MORPHOLOGICAL_SWITCH
## 24          c2.ROZANOV_MMP14_TARGETS_SUBSET
## 25          c2.PID_FAK_PATHWAY
## 26          c2.REN_ALVEOLAR_RHABDOMYOSARCOMA_SIGNED
## 27          c2.ELVIDGE_HIF1A_AND_HIF2A_TARGETS_SIGNED
## 28          c2.PASINI_SUZ12_TARGETS_SIGNED
## 29          c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_SIGNED
## 30          c2.RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY_SIGNED
##      Correlation Metagenes
## 1      0.6253      5
## 2      0.6185      5
## 3      0.6174      5
## 4      0.6069      5
## 5      0.5897      5
## 6      0.5852      5
## 7      0.5653      5
## 8      0.5642      5
## 9      0.5642      5
## 10     0.5612      5
## 11     0.5533      5
## 12     0.5514      5
## 13     0.5469      5
## 14     0.5443      5
## 15     0.5379      5
## 16     0.5297      5
## 17     0.5271      5
## 18     0.5199      5
## 19     0.5188      5
## 20     0.5181      5
## 21     0.5121      5
## 22     0.5117      5
## 23     0.5038      5
## 24     0.5008      5
## 25     0.5001      5
## 26     -0.5132     -5
## 27     -0.5151     -5
## 28     -0.5574     -5
## 29     -0.5773     -5
## 30     -0.5810     -5
##
## [[5]]$c3
## data frame with 0 columns and 0 rows
##
## [[5]]$c4
##      GeneSet Correlation Metagenes
## 1  c4.GNF2_MMP1      0.6088      5
## 2  c4.GNF2_CDH11      0.5638      5
## 3  c4.GNF2_PTX3      0.5571      5
## 4  c4.MODULE_562      0.5503      5
## 5  c4.MODULE_321      0.5484      5
## 6  c4.MODULE_122      0.5102      5
## 7  c4.MODULE_275      0.5057      5
## 8  c4.MODULE_211      0.5023      5

```



```
##
## [[5]]$c5
##               GeneSet Correlation Metagenes
## 1             c5.COLLAGEN      0.5466      5
## 2             c5.TISSUE_DEVELOPMENT 0.5154      5
## 3             c5.AXON_GUIDANCE 0.5083      5
## 4 c5.EXTRACELLULAR_MATRIX_PART 0.5034      5
##
## [[5]]$c6
## data frame with 0 columns and 0 rows
##
## [[5]]$c7
##               GeneSet
## 1             c7.GSE3982_DC_VS_BCELL_SIGNED
## 2             c7.GSE3982_MAC_VS_BCELL_SIGNED
## 3             c7.GSE1460_CD4_THYMOCYTE_VS_THYMIC_STROMAL_CELL_SIGNED
## 4 c7.GSE1460_INTRATHYMIC_T_PROGENITOR_VS_THYMIC_STROMAL_CELL_SIGNED
## Correlation Metagenes
## 1      0.5327      5
## 2      0.5113      5
## 3     -0.5511     -5
## 4     -0.6159     -5
```

```
print(asreg.result)

## glmulti.analysis
## Method: h / Fitting: coxph / IC used: bic
## Level: 2 / Marginality: TRUE
## From 100 models:
## Best IC: 585.934888469493
## Best model:
## [1] "Surv(time, event) ~ 1 + mg.3 + mg.5"
## Evidence weight: 0.18524553830014
## Worst IC: 597.912592370947
## 4 models within 2 IC units.
## 45 models to reach 95% of evidence weight.

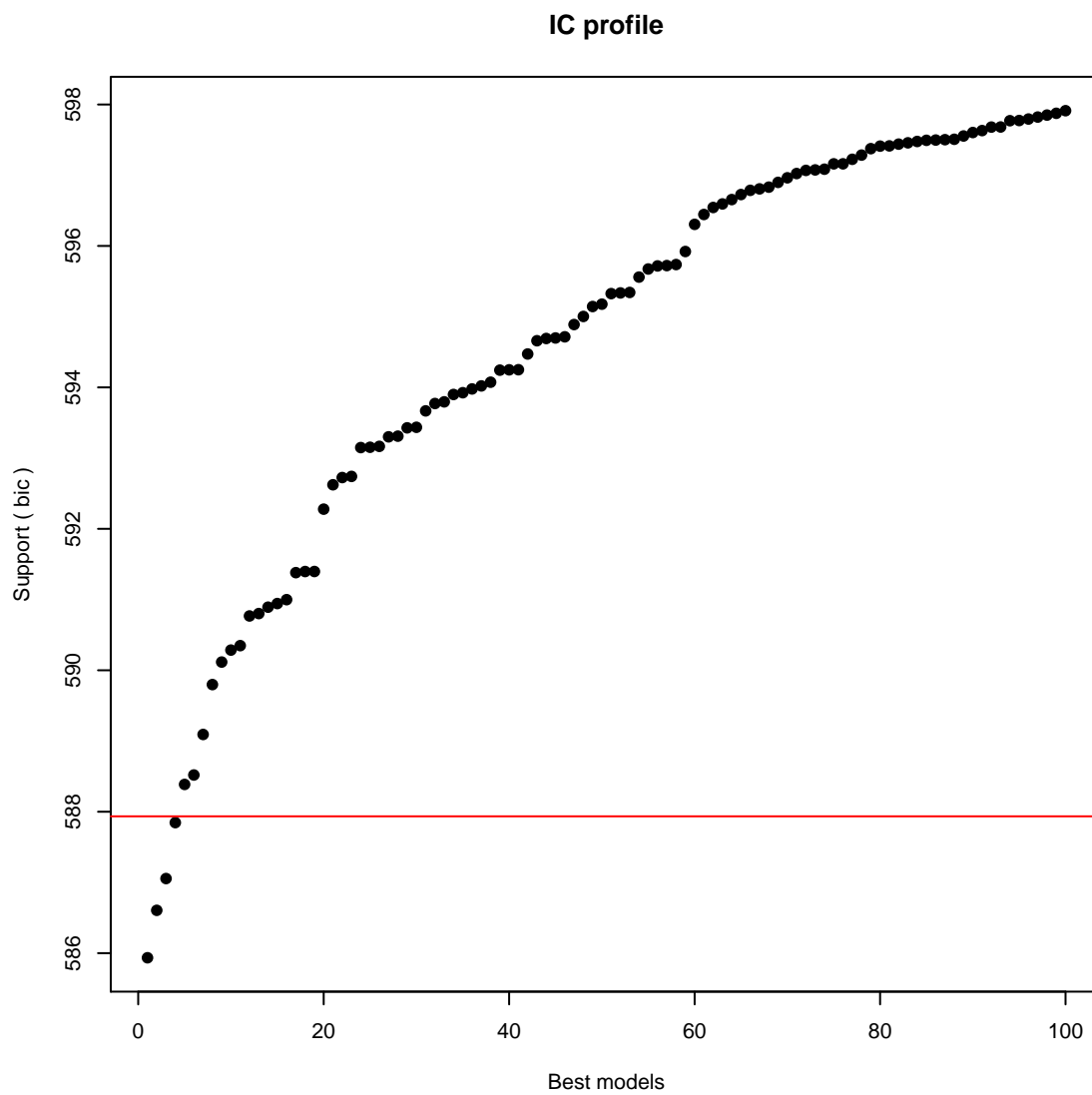
coef(asreg.result)

##               Estimate Uncond. variance Nb models Importance +/- (alpha=0.05)
## mg.1:mg.4  0.065698      0.02929      5  0.005136      0.3395
## mg.1:mg.2 -0.058005      0.02662      4  0.005555      0.3236
## mg.1:mg.5  0.038208      0.03253      6  0.008336      0.3578
## mg.2:mg.4  0.147005      0.20932      8  0.010285      0.9076
## mg.1:mg.3 -0.230410      0.35202     11  0.016488      1.1770
## mg.2:mg.5 -0.002633      0.24840      7  0.020826      0.9887
## mg.4:mg.5  0.791112      3.21226     16  0.037646      3.5554
## mg.2:mg.3 -0.121791      1.34565     13  0.044405      2.3012
## mg.3:mg.4  1.162125      6.73045     20  0.052598      5.1464
## mg.3:mg.5 -2.702005     29.68599     26  0.140337     10.8083
## mg.1      0.131635      0.12555     54  0.154944      0.7029
## mg.4      0.150908      1.04859     65  0.298478      2.0313
## mg.2     -1.823729      5.66012     60  0.471721      4.7195
## mg.5      2.544507      5.07447     71  0.687623      4.4686
## mg.3      7.468709      3.80552     96  0.993506      3.8698
```

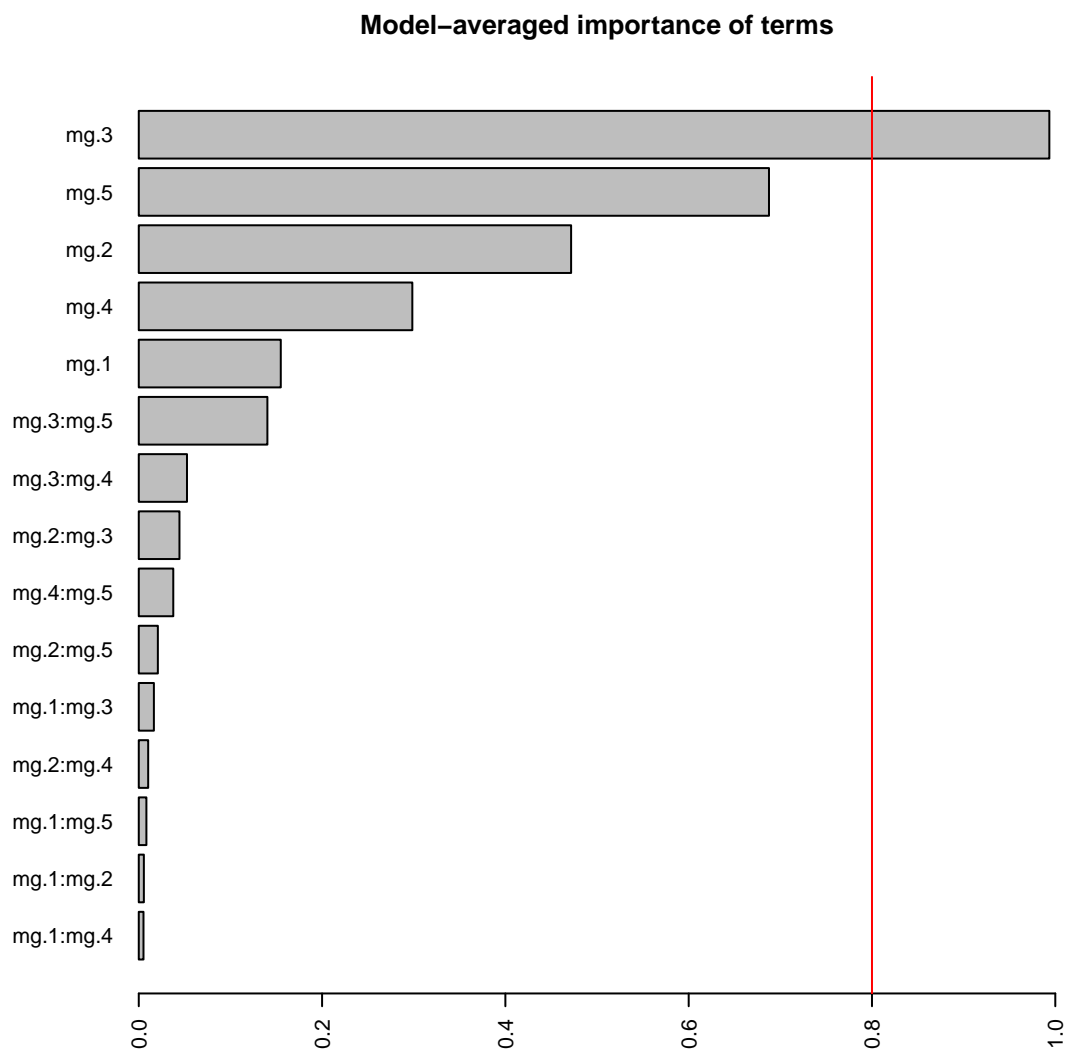
```
summary(asreg.result@objects[[1]])

## Call:
## fitfunc(formula = as.formula(x), data = data)
##
## n= 104, number of events= 77
##
##      coef exp(coef) se(coef)    z Pr(>|z|)
## mg.3   7.50  1803.67    1.33 5.64  1.7e-08
## mg.5   3.32   27.63    1.24 2.69   0.0072
##
##      exp(coef) exp(-coef) lower .95 upper .95
## mg.3   1803.7   0.000554   133.44   24379
## mg.5    27.6   0.036190    2.45     311
##
## Concordance= 0.707 (se = 0.036 )
## Rsquare= 0.312 (max possible= 0.997 )
## Likelihood ratio test= 38.8 on 2 df,  p=3.7e-09
## Wald test               = 41.7 on 2 df,  p=8.9e-10
## Score (logrank) test = 44.2 on 2 df,  p=2.57e-10

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

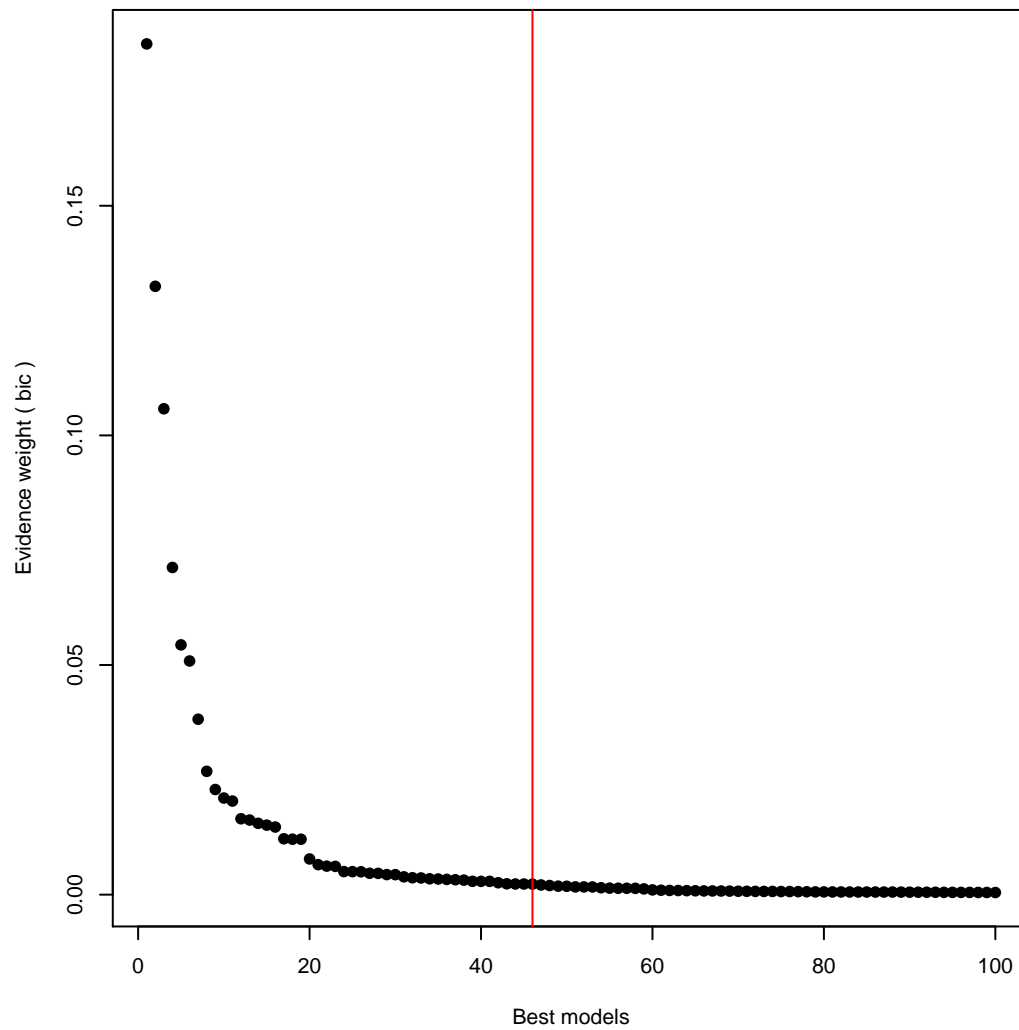


```
plot(asreg.result, type = "s")
```



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

Profile of model weights

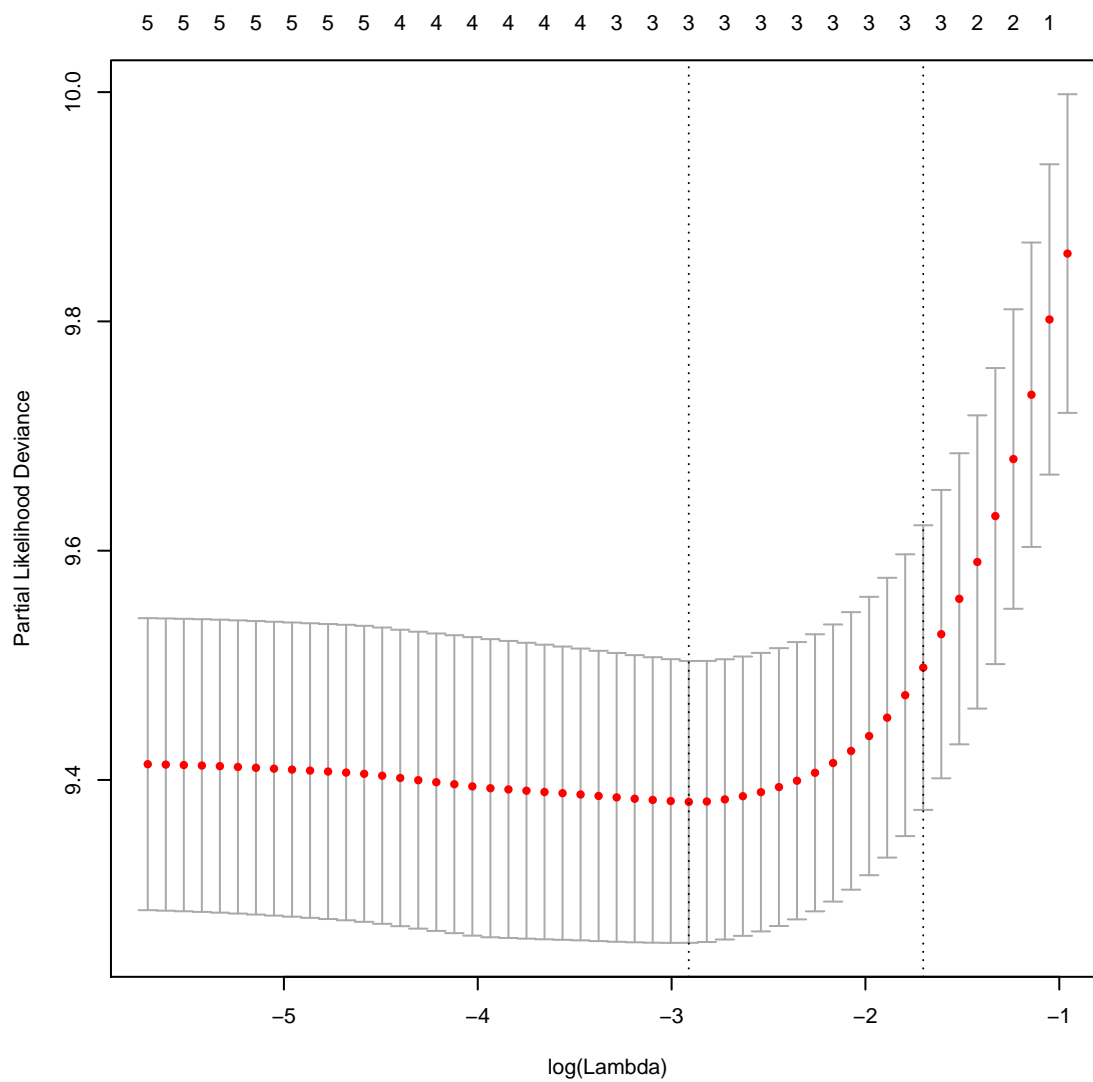


```
glmnet.coef.1se
## 5 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1  .
## mg.2 -1.2588
## mg.3  4.0834
## mg.4  .
## mg.5  0.4581

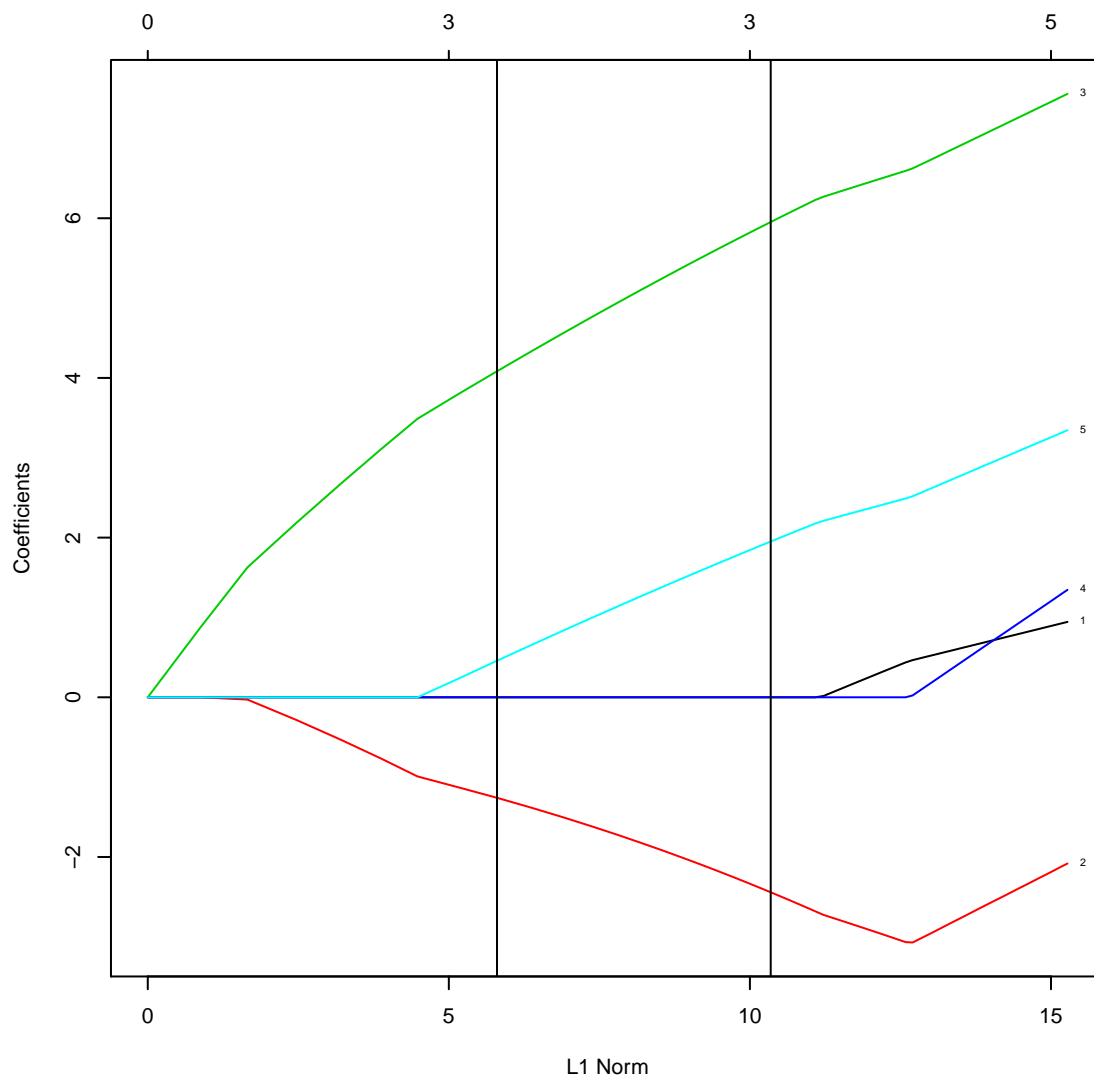
glmnet.coef.min
## 5 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1  .
## mg.2 -2.442
## mg.3  5.953
## mg.4  .
```

```
## mg.5 1.951
```

```
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

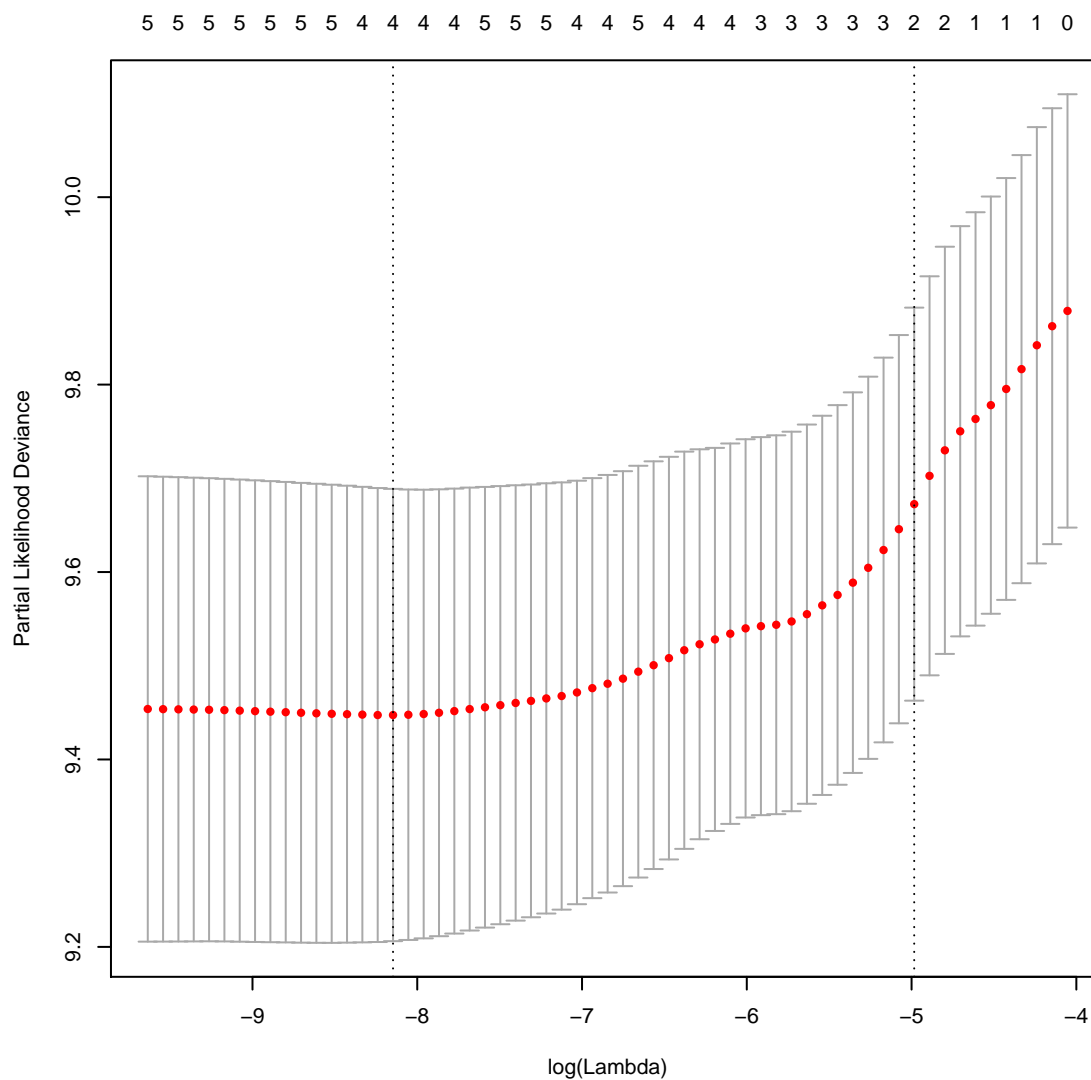
## 5 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1      .
## mg.2 -10.806
## mg.3      .
## mg.4  -4.482
## mg.5      .

adaglmnet.coef.min/adaglmnet.weights

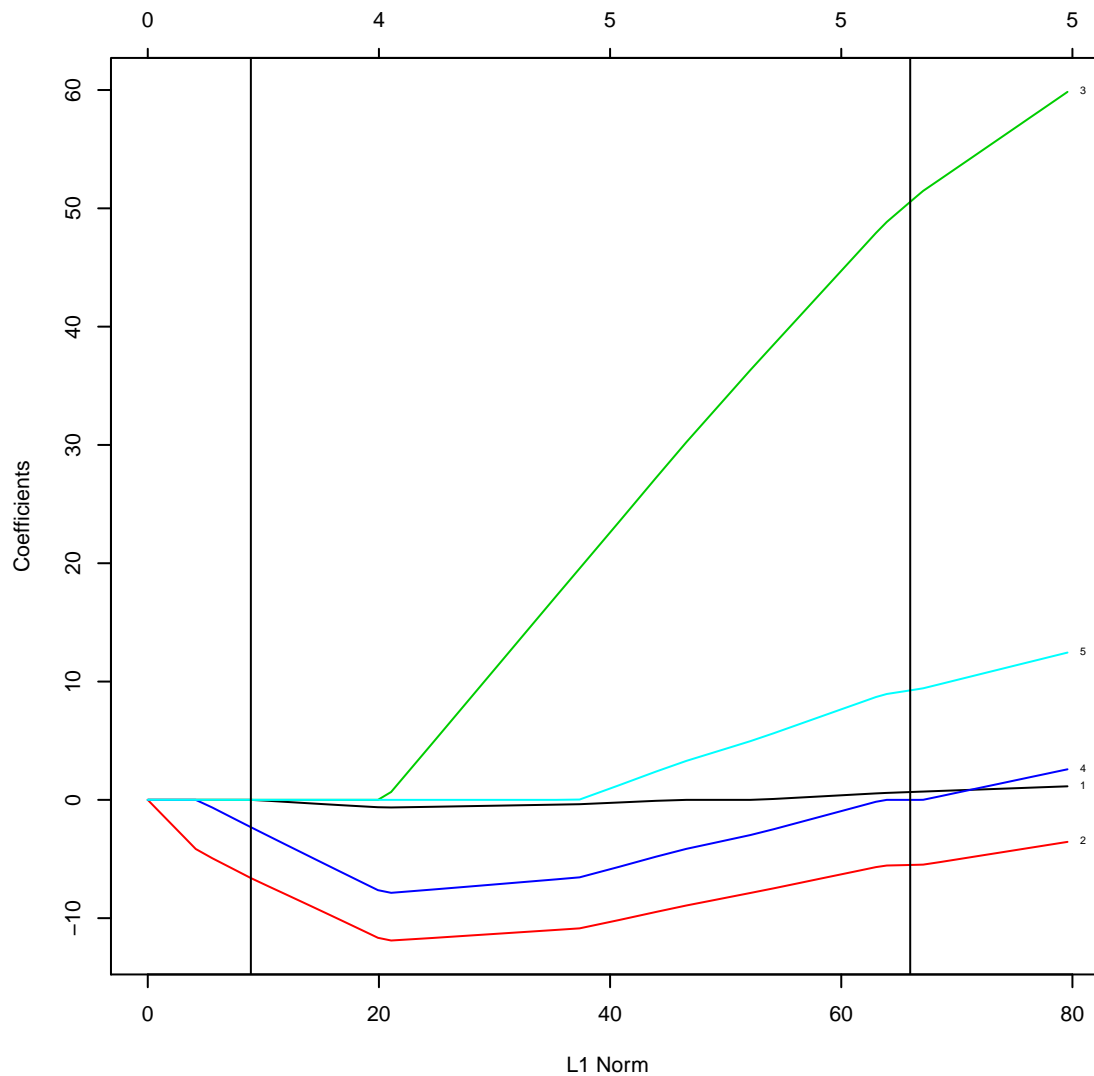
## 5 x 1 sparse Matrix of class "dgCMatrix"
##           1
## mg.1    0.770
## mg.2  -8.996
## mg.3 403.209
## mg.4      .
```

```
## mg.5 34.457
```

```
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
```