

SIS NMF Final: Diagnosis to DSD

November 28, 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 [NO: bigmemory] | Cores 31/32
## To enable shared memory capabilities, try: install.extras('
## NMF
## ')

library(glmulti)

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

library(glmnet)

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

library(RColorBrewer)
library(gplots)

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

```
library(xtable)
library(stargazer)

##
## Please cite as:
##
## Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression
and summary statistics tables.
## R package version 5.1. http://CRAN.R-project.org/package=stargazer

load("image.rda")
```

2 Probe selection

```
table(cpss.sis$sel)

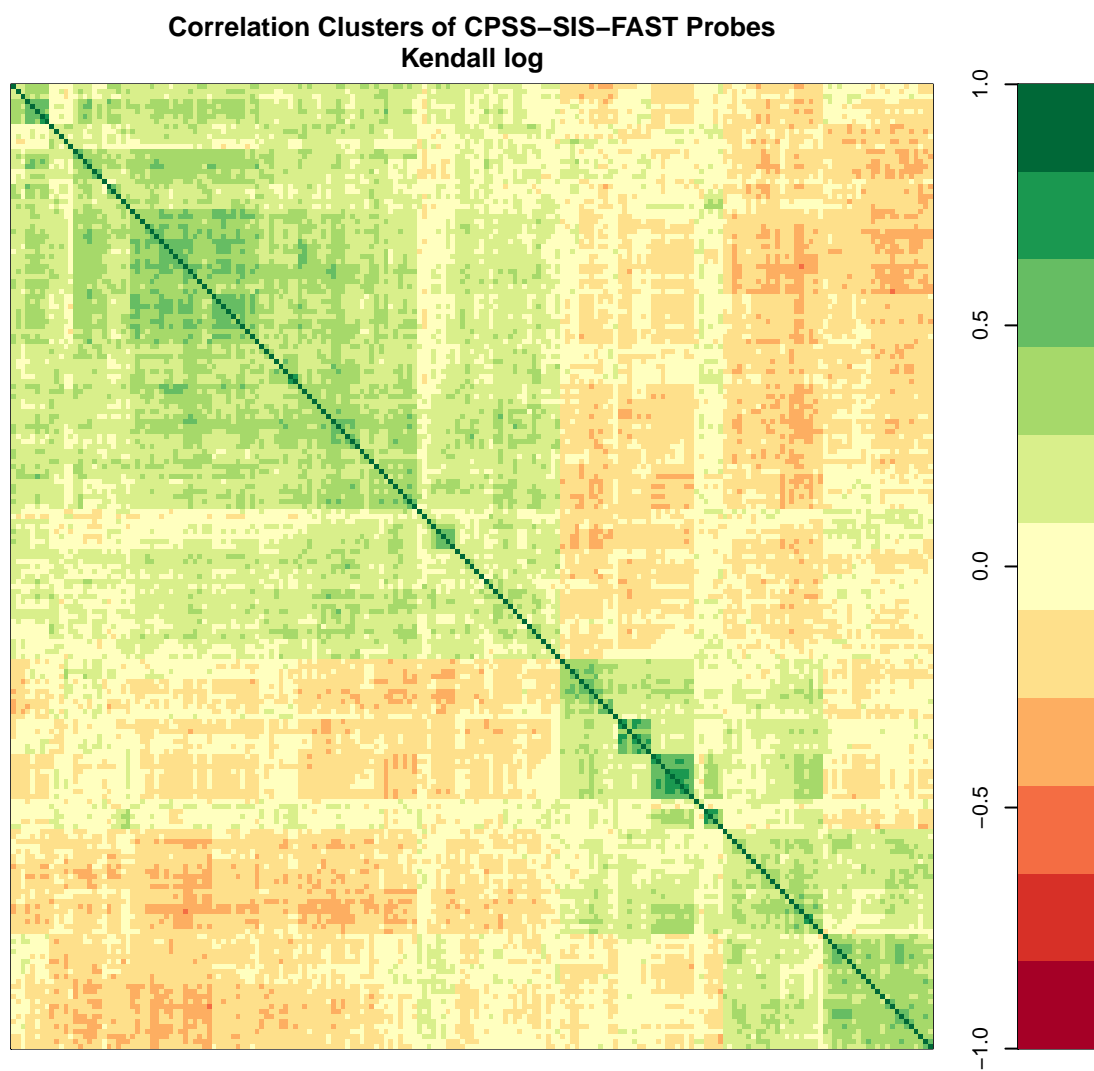
##
## FALSE TRUE
## 12807 193

mean(cpss.sis$sel)

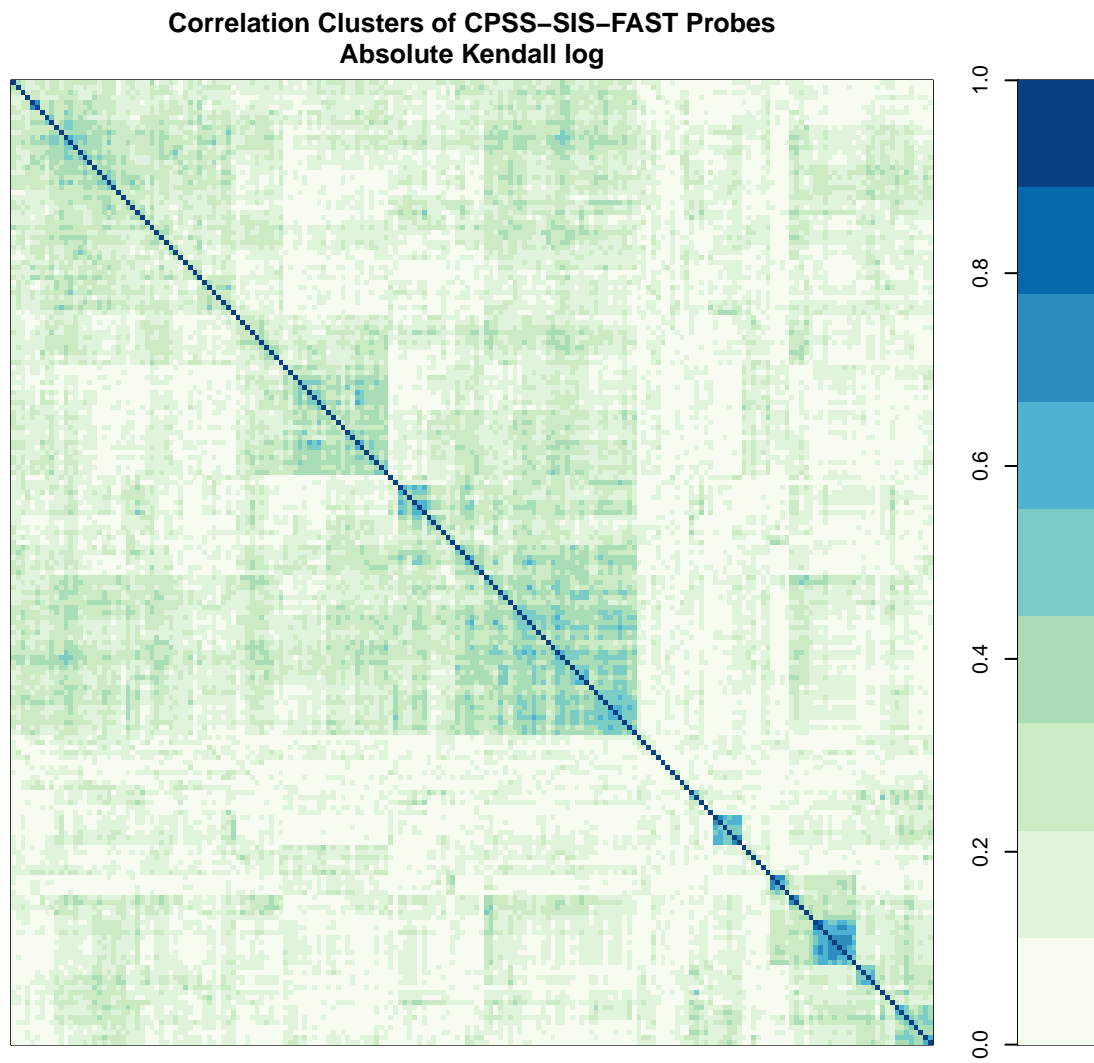
## [1] 0.01485
```

3 Expression correlation

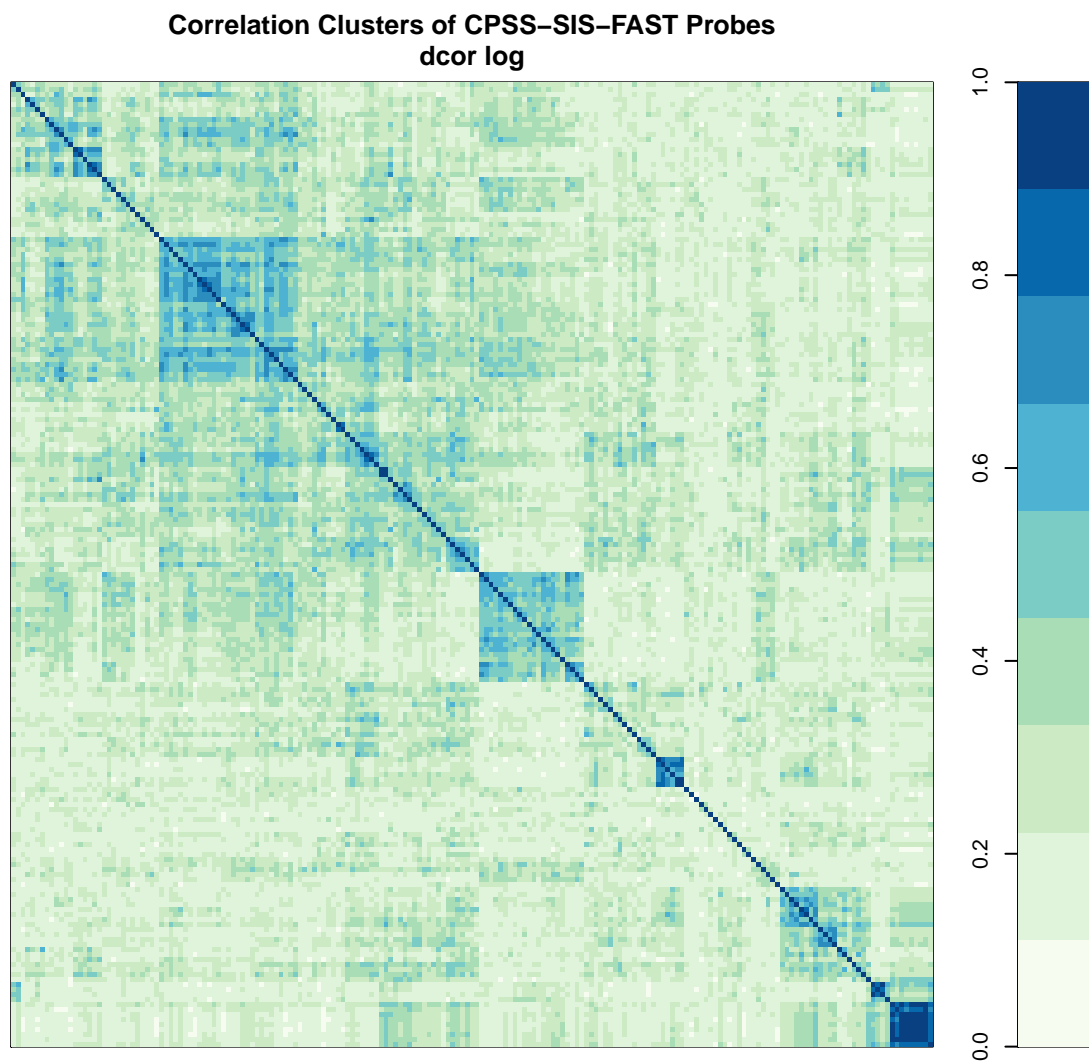
```
corPlot(x.sel.kcor, main = "Correlation Clusters of CPSS-SIS-FAST Probes\nKendall log",
        useRaster = FALSE)
```



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

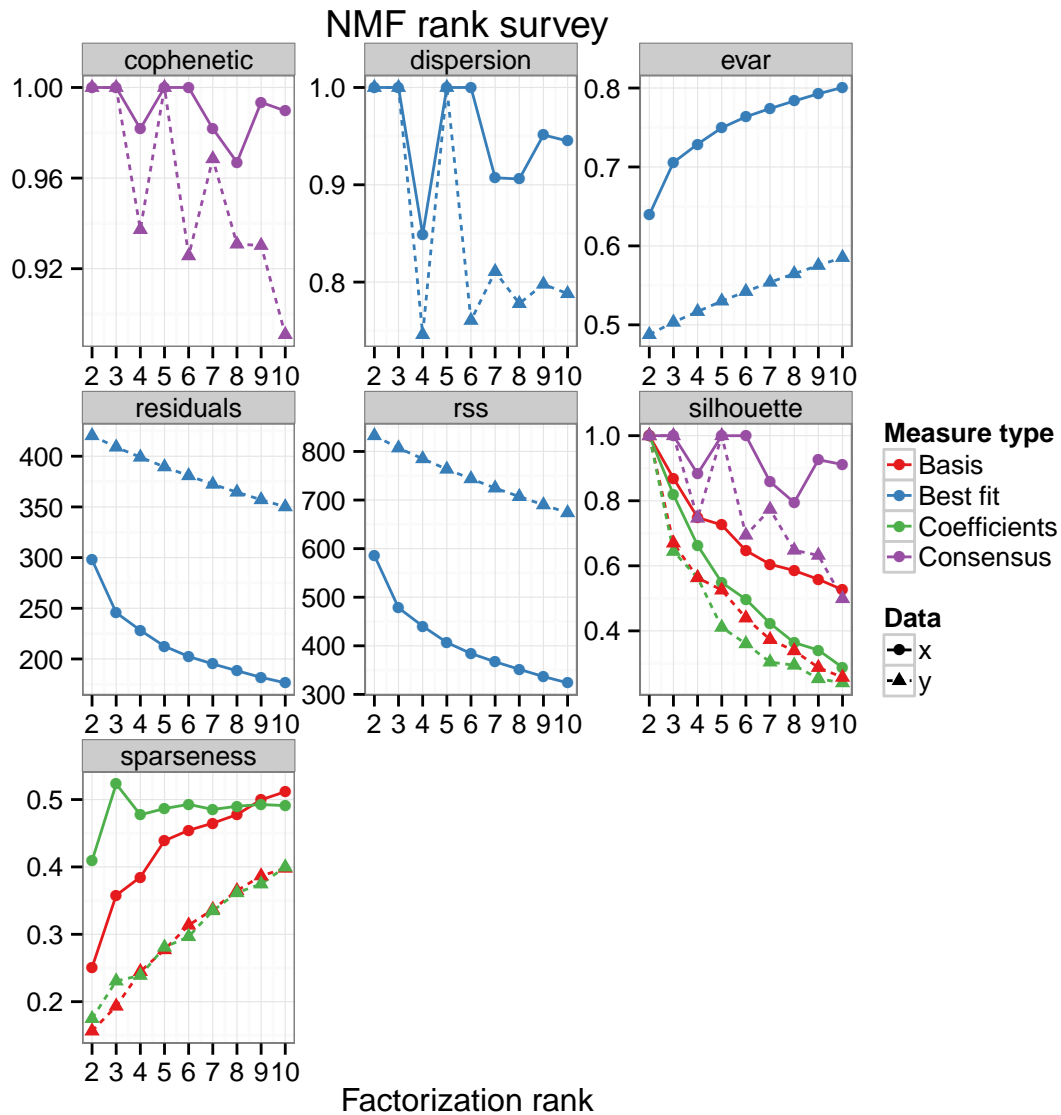


```
corPlot(x.sel.dcor, zlim = c(0, 1), pal = "GnBu", main = "Correlation Clusters of CPSS-SIS-FAST Probes",
        useRaster = FALSE)
```



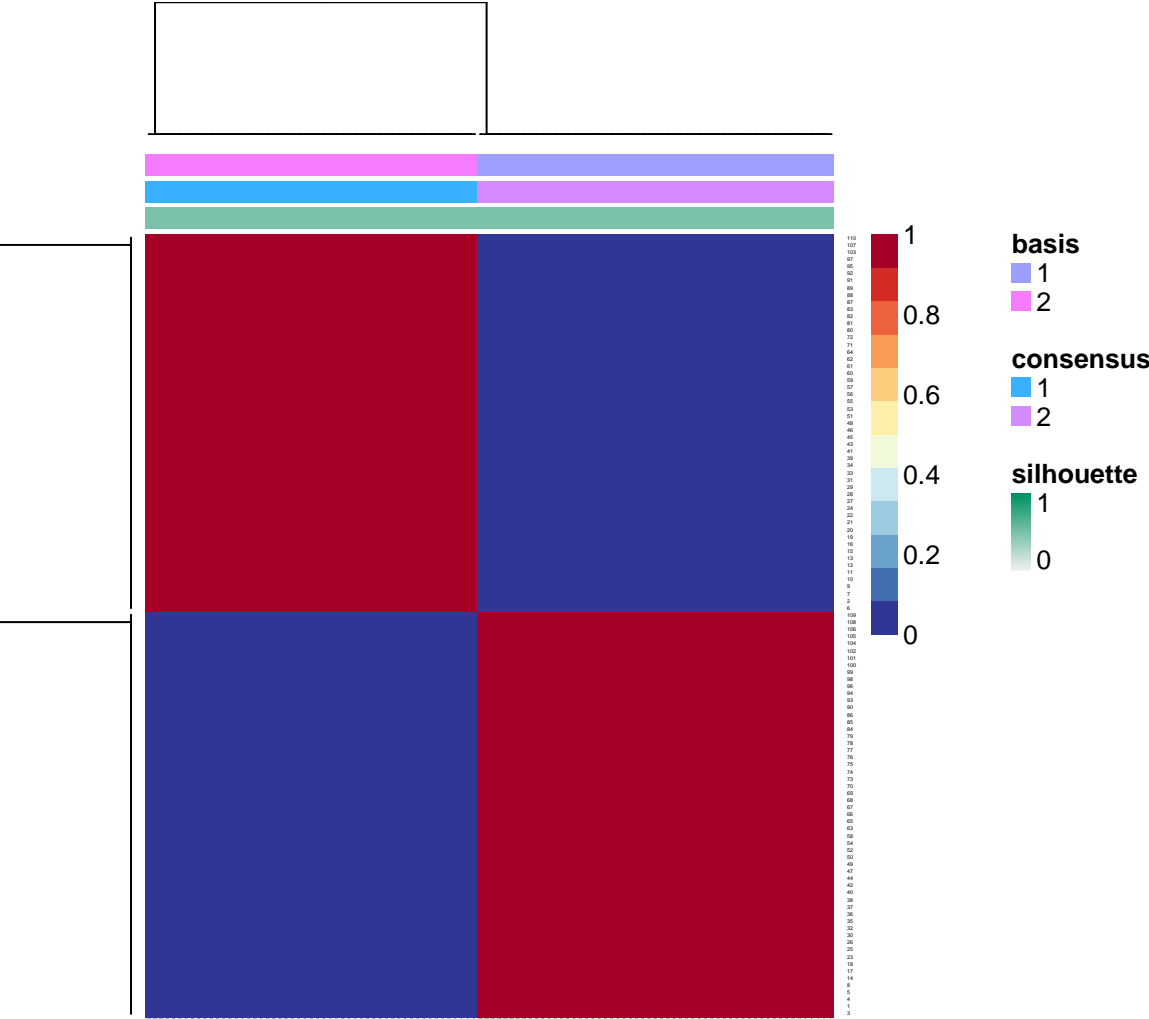
4 Factorization

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

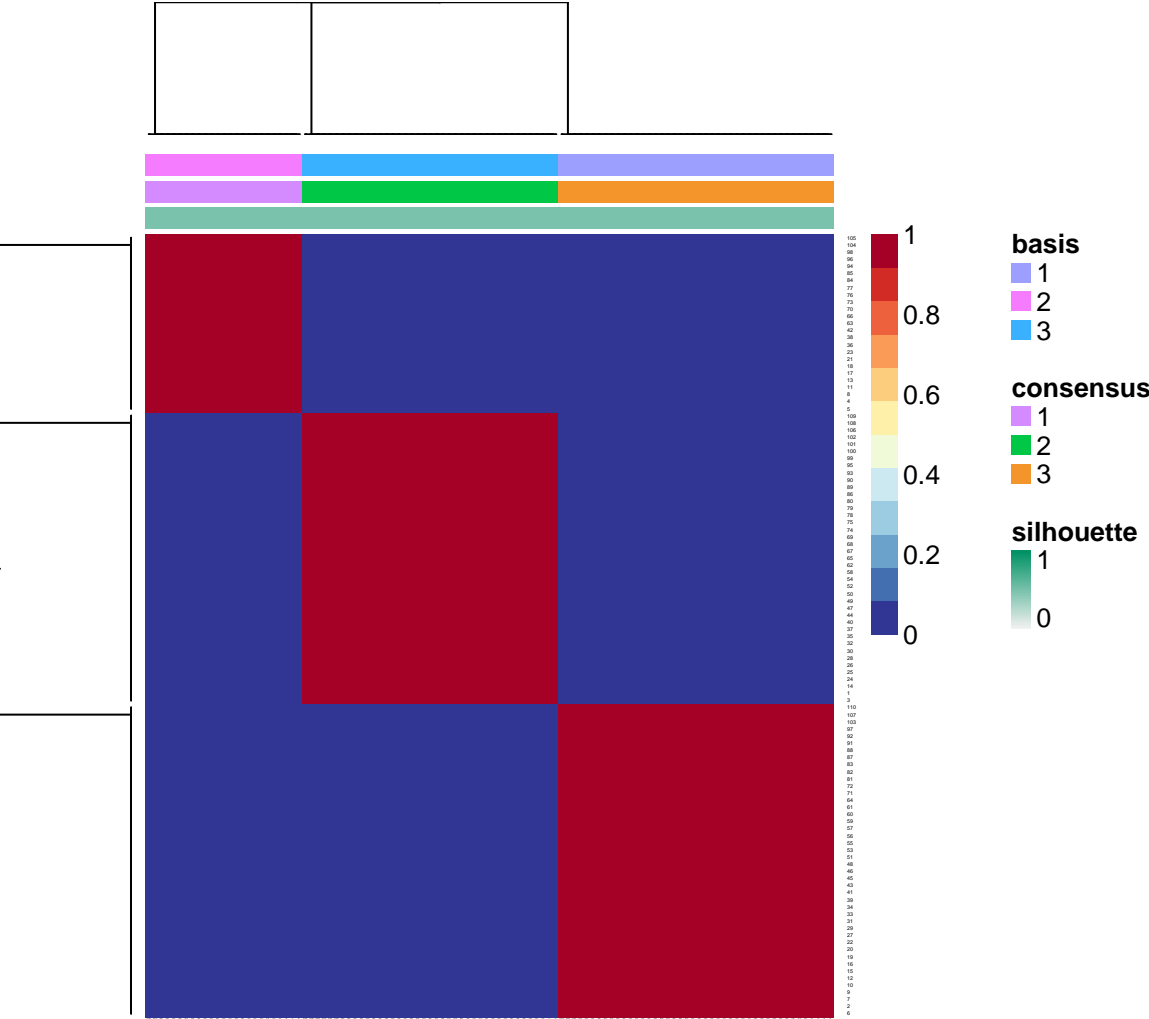


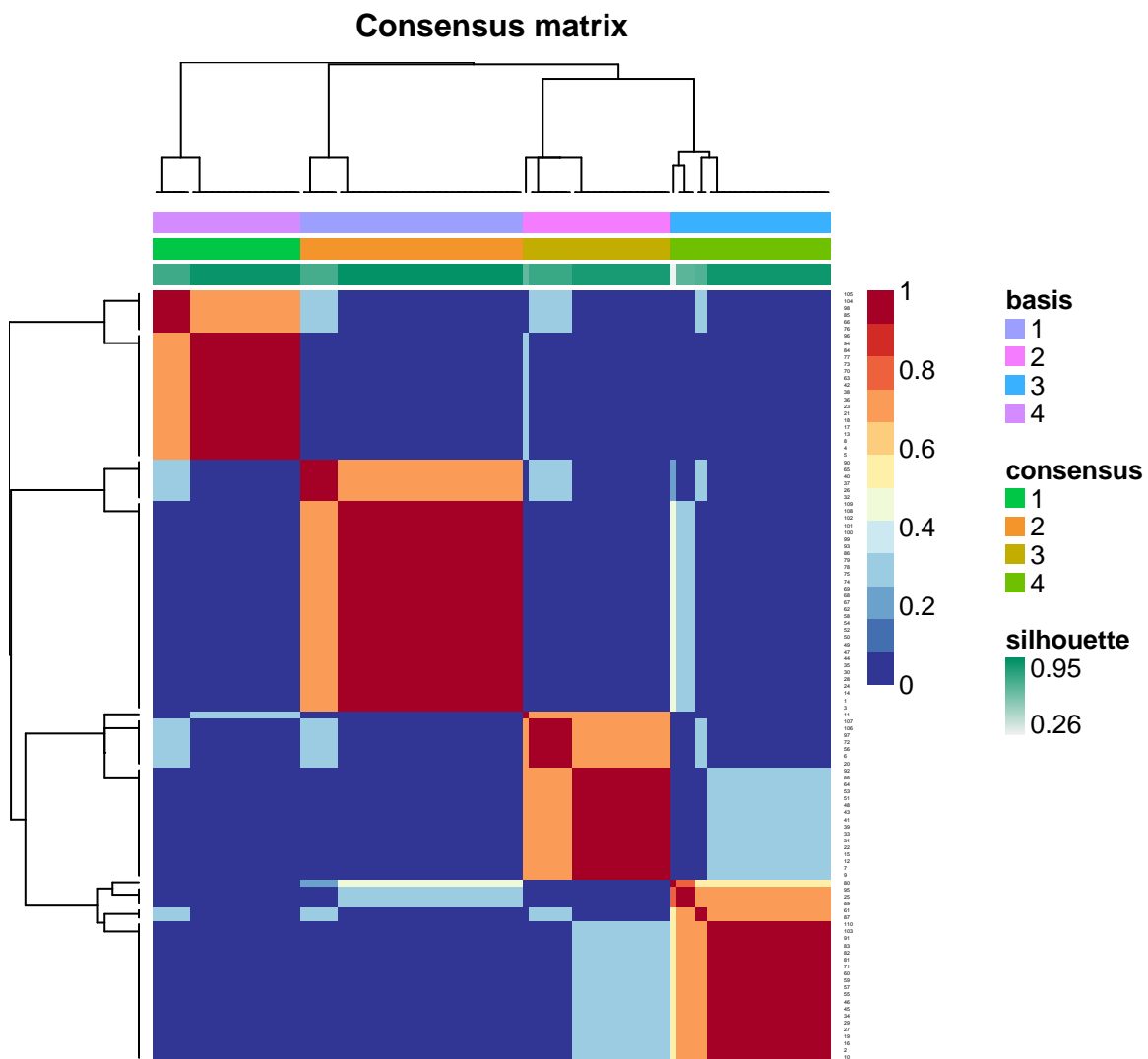
```
for (i in nmf.runs.rank$fit) {
  consensusmap(i)
}
```

Consensus matrix

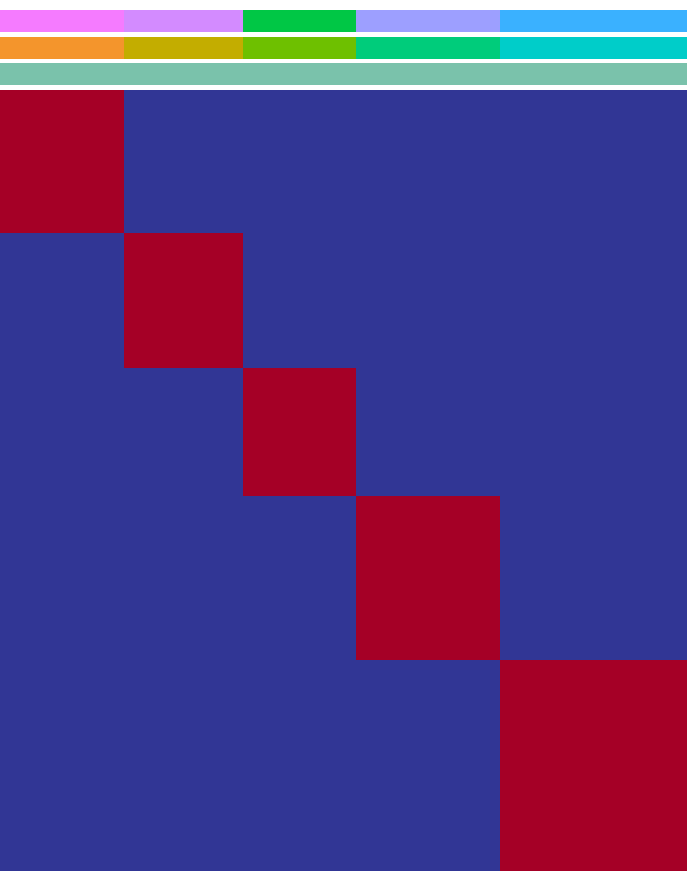


Consensus matrix

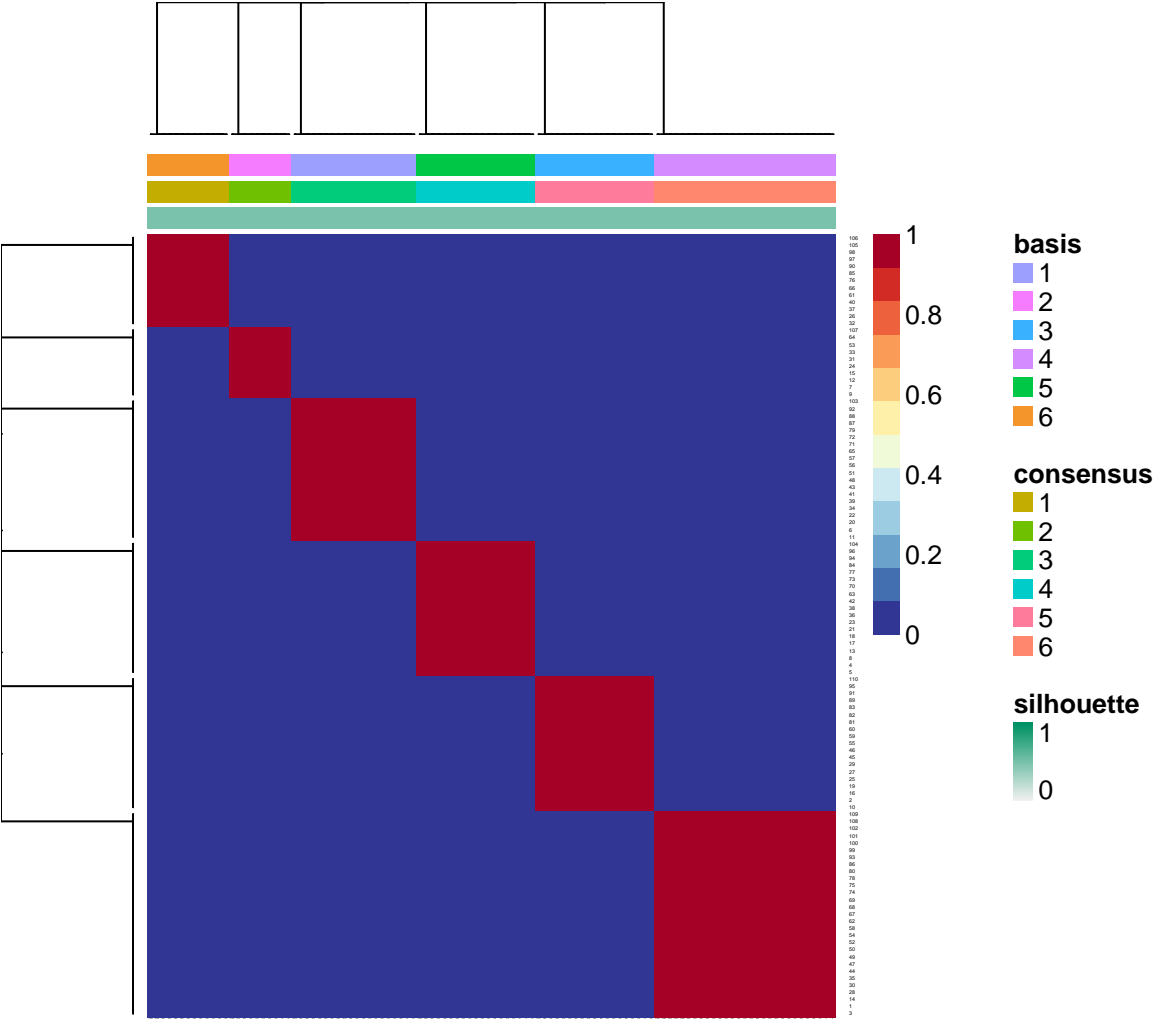


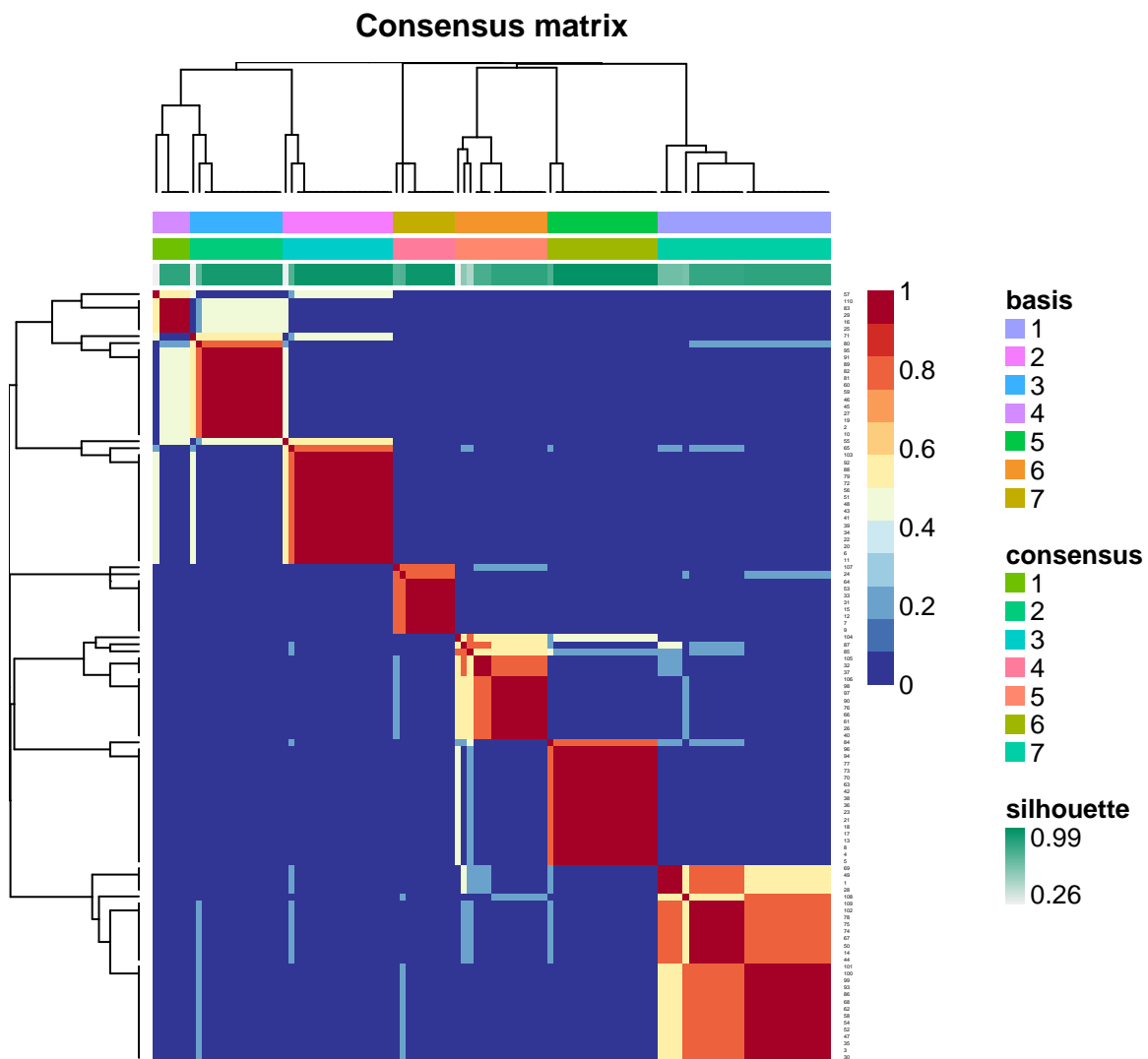


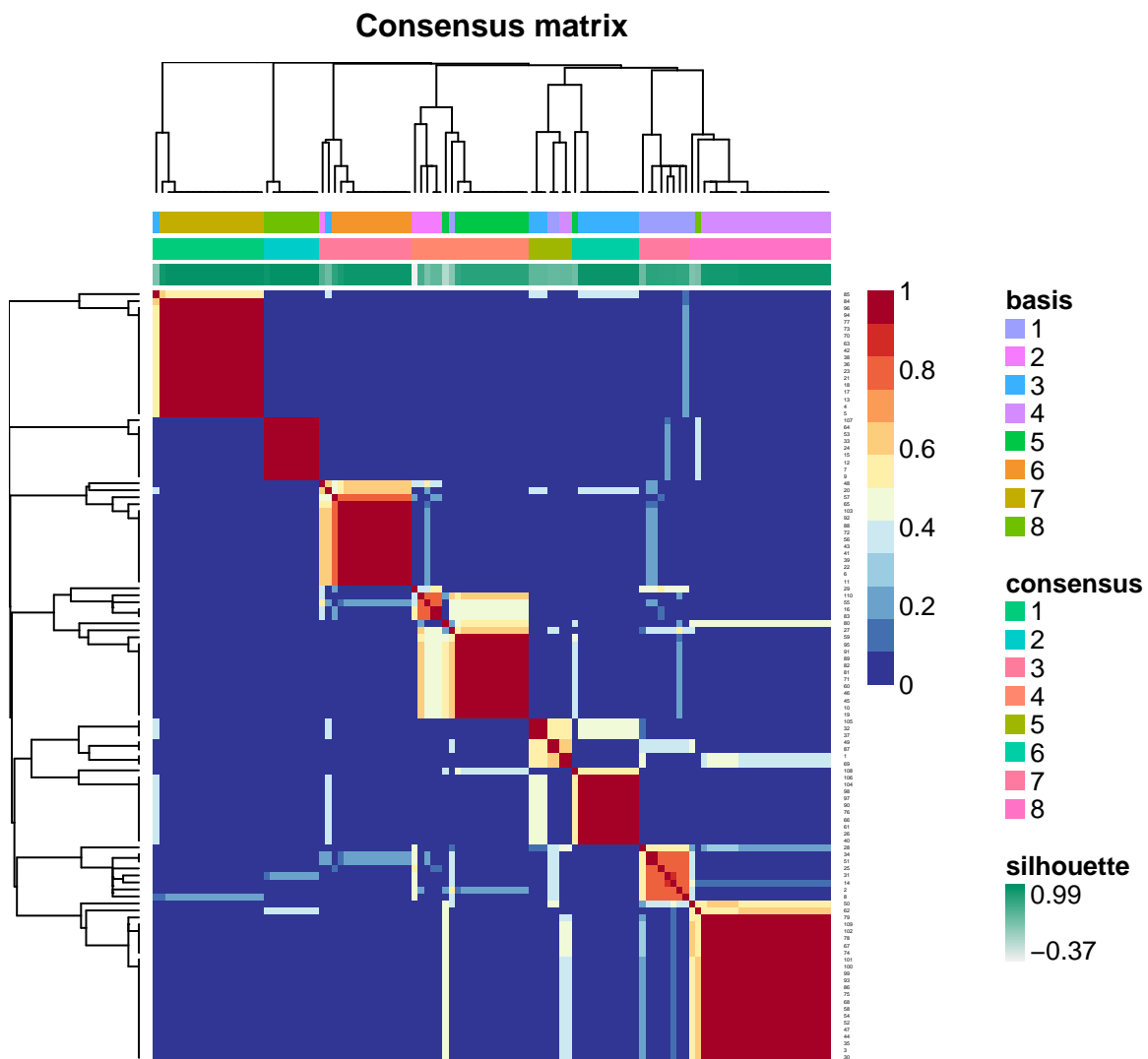
| | | | |
|--|--|--|--|
| | | | |
|--|--|--|--|

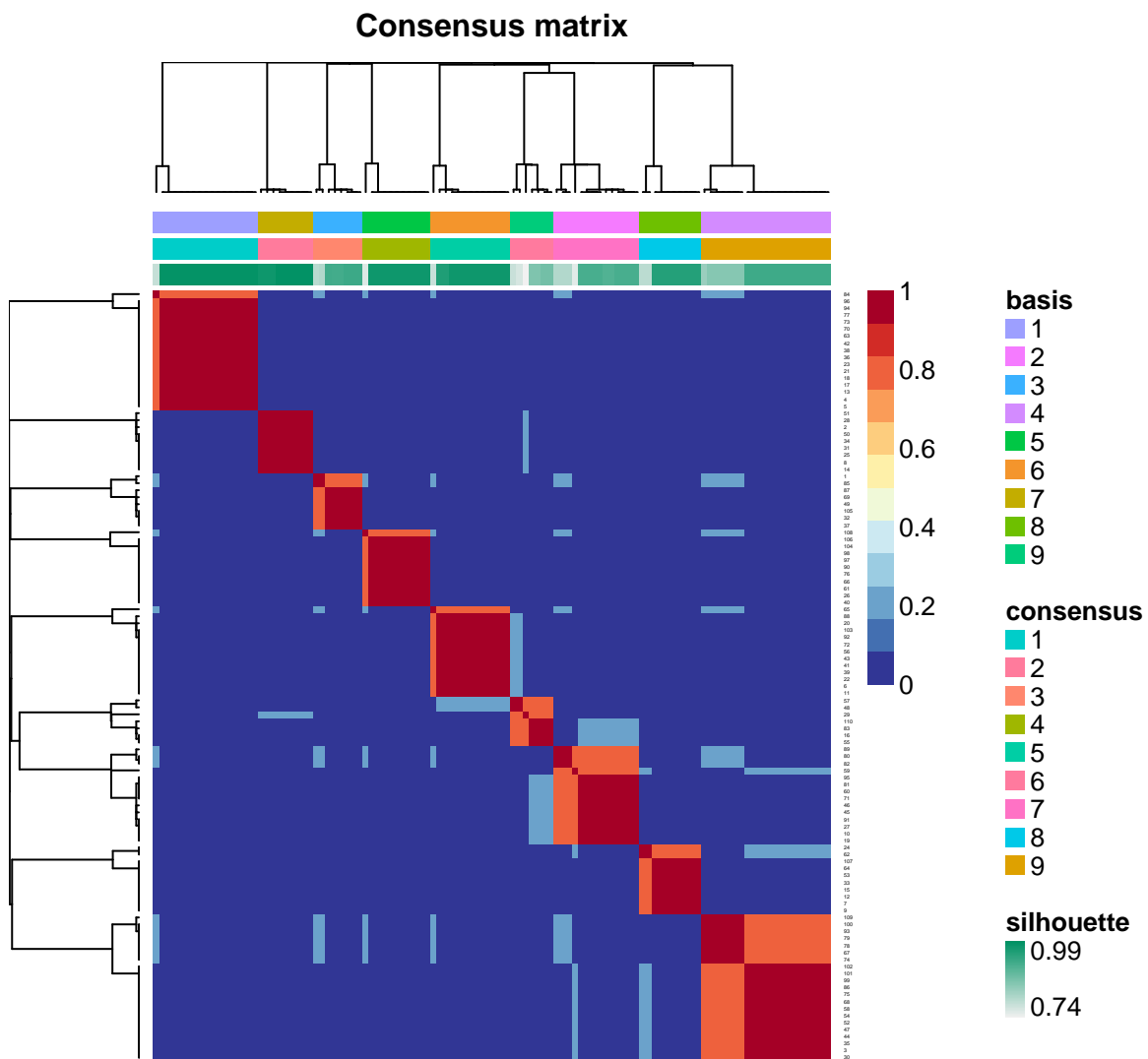


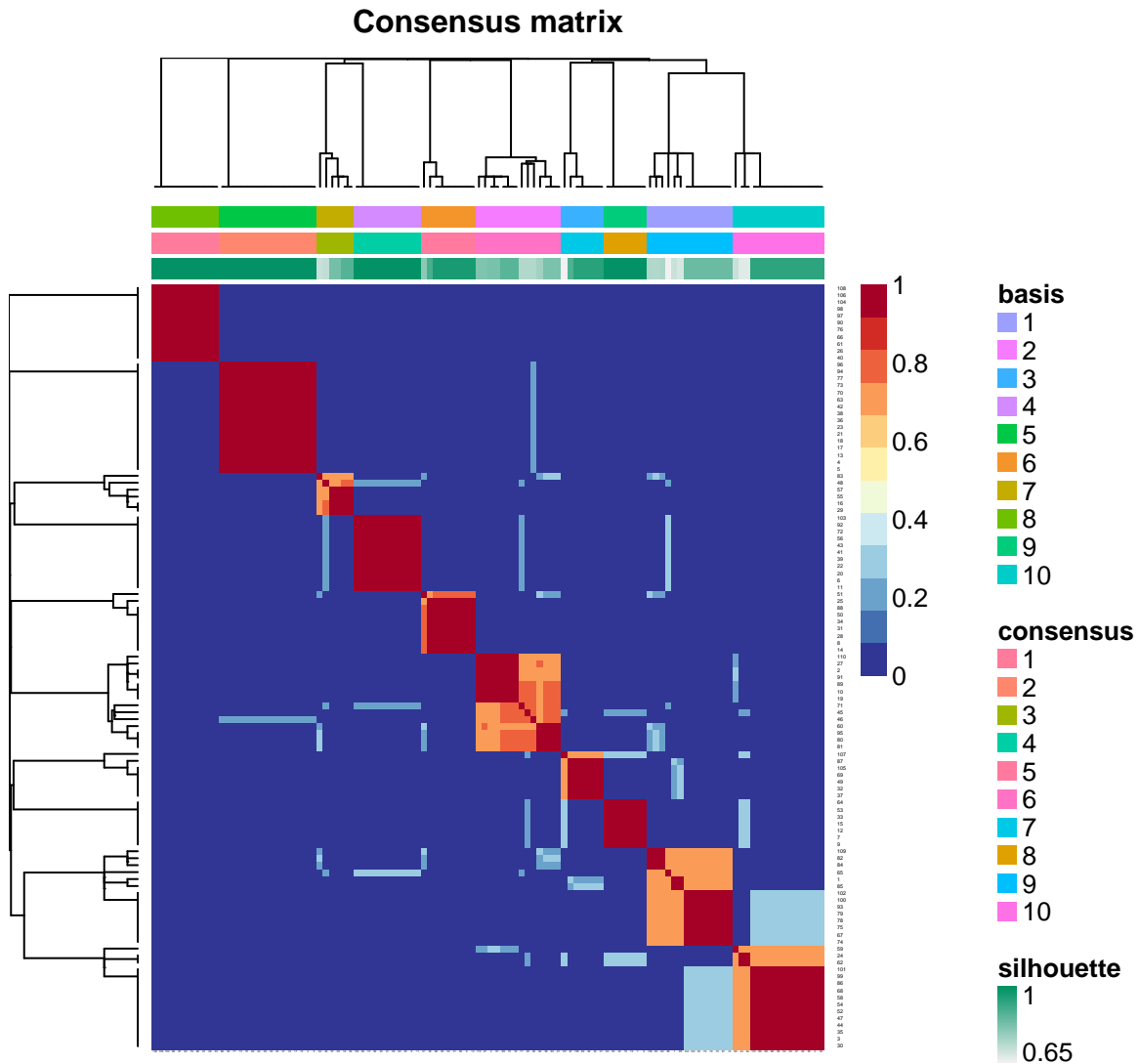
Consensus matrix











```

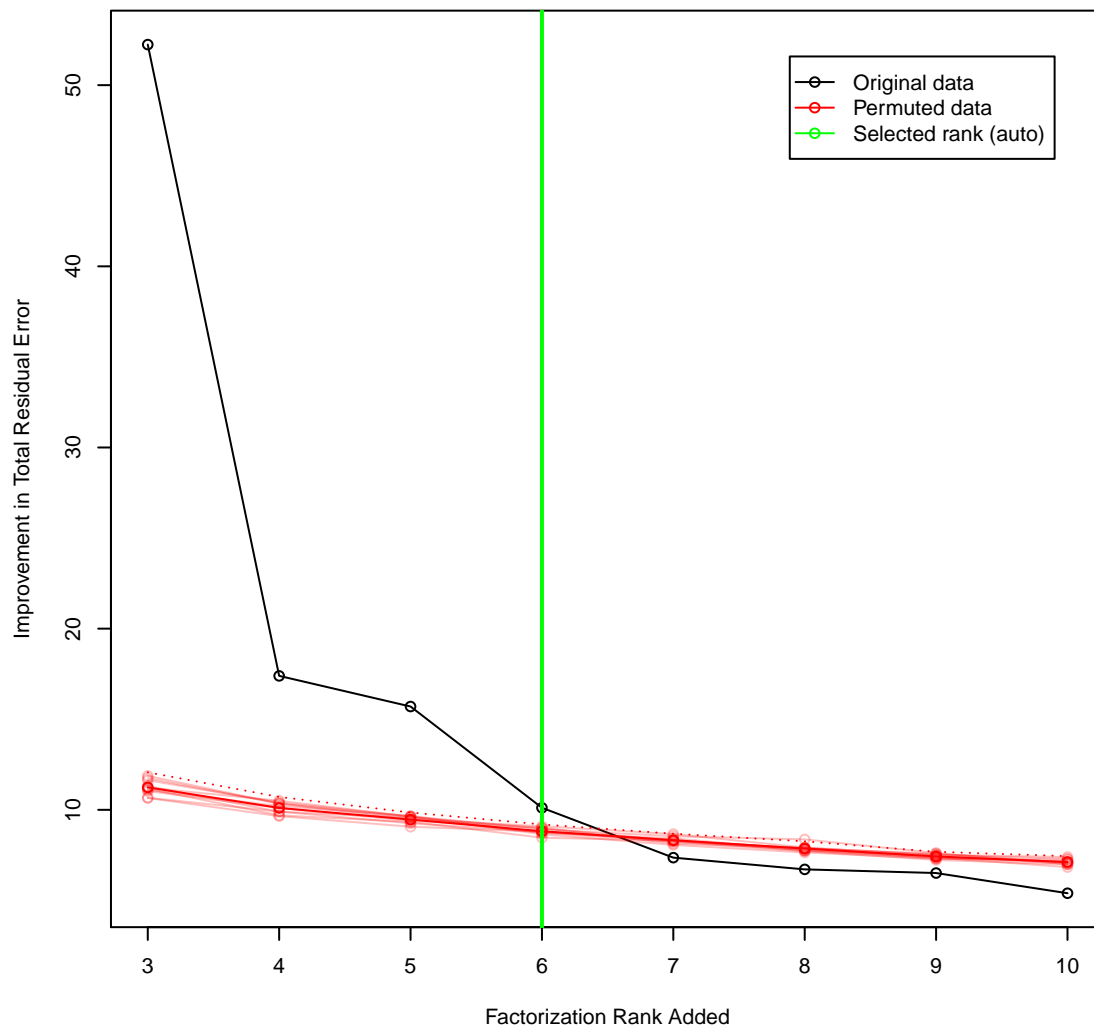
plot(nmf.rankrange[-1], -temp.orig_resids.delta, type = "o", col = "black",
     pch = 21, ylim = range(-c(temp.orig_resids.delta, temp.perm_resids.delta.mean)),
     xlab = "Factorization Rank Added", ylab = "Improvement in Total Residual Error")
lines(nmf.rankrange[-1], -temp.perm_resids.delta.mean, col = "red", type = "o",
      pch = 21, lwd = 1)
for (i in 1:ncol(temp.perm_resids)) {
  lines(nmf.rankrange[-1], -temp.perm_resids.delta[, i], type = "o", col = rgb(1,
    0, 0, 0.25))
}
lines(nmf.rankrange[-1], -temp.perm_resids.delta.threshold, col = "red", lty = "dotted")
if (nmf.rank.wasauto == TRUE) {
  temp.col = "green"
} else {
  temp.col = "blue"
}
abline(v = nmf.rank, col = temp.col, lwd = 2)
legend("topright", legend = c("Original data", "Permuted data", sprintf("Selected rank (%s)",

```

```

ifelse(nmf.rank.wasauto == TRUE, "auto", "fixed")), col = c("black", "red",
temp.col), lty = "solid", pch = 21, inset = 0.05)

```

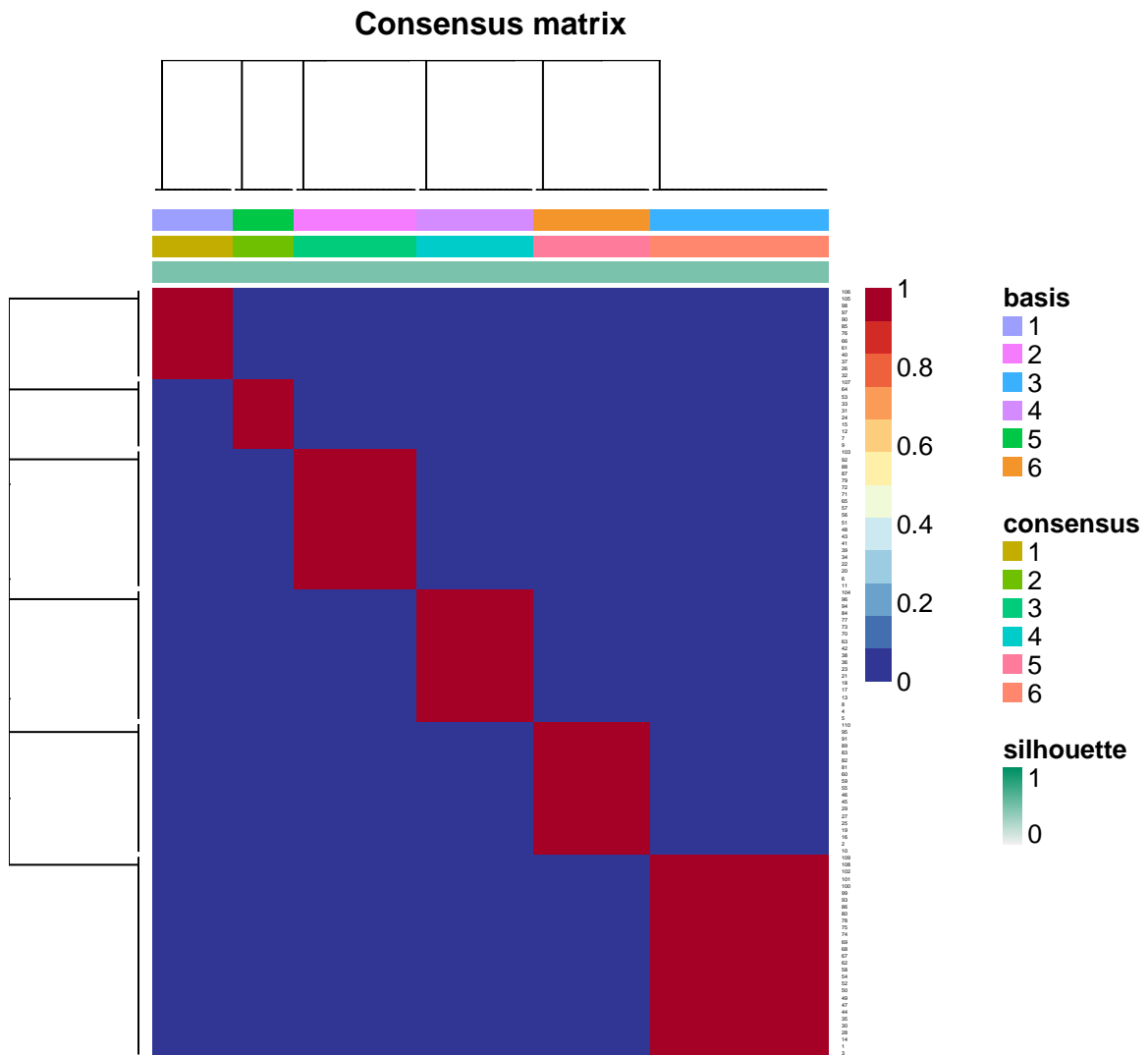


4.1 Fit

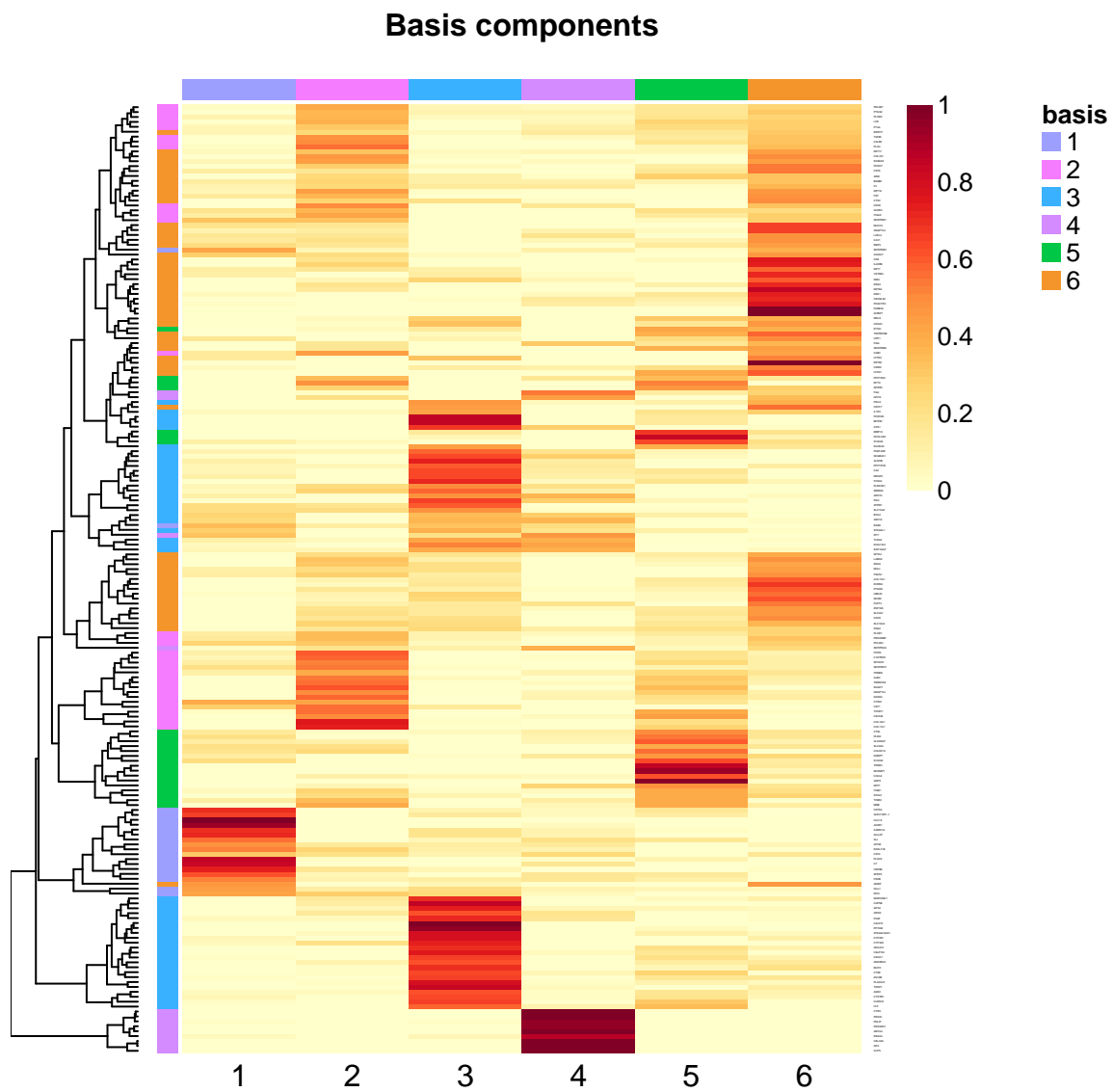
```

consensusmap(nmf.final)

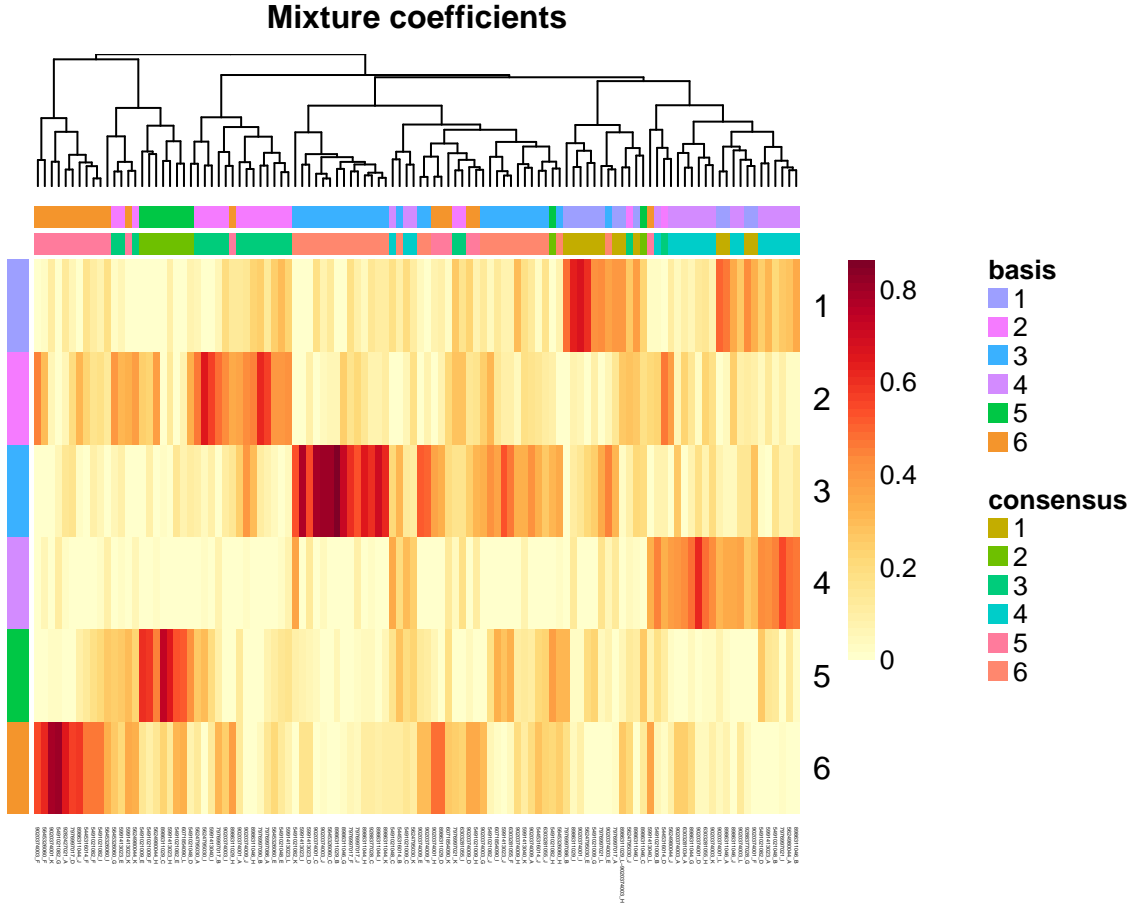
```

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



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



```

coefs.diag_dsd = t(xlin.diag_dsd.sel) %*% basis(nmf.final)
coefs.diag_rec = t(xlin.diag_rec.sel) %*% basis(nmf.final)
coefs.recr_dsd = t(xlin.recr_dsd.sel) %*% basis(nmf.final)
coefs.pdac_au = t(xlin.pdac_au.sel) %*% basis(nmf.final)

```

4.2 MTC P-values

```

nmf.final.cpv.pvals = data.frame(surv.diag_rec.p = apply(coefs.diag_rec, 2,
function(xc) pchisq(2 * diff(coxph(y.diag_rec ~ xc)$loglik), df = 1, lower.tail = FALSE)),
surv.diag_rec.c = apply(coefs.diag_rec, 2, function(xc) coef(coxph(y.diag_rec ~
xc))), surv.diag_dsd.p = apply(coefs.diag_dsd, 2, function(xc) pchisq(2 *
diff(coxph(y.diag_dsd ~ xc)$loglik), df = 1, lower.tail = FALSE)), surv.diag_dsd.c = apply(coefs
2, function(xc) coef(coxph(y.diag_dsd ~ xc))), surv.recr_dsd.p = apply(coefs.recr_dsd,
2, function(xc) pchisq(2 * diff(coxph(y.recr_dsd ~ xc)$loglik), df = 1,
lower.tail = FALSE)), surv.recr_dsd.c = apply(coefs.recr_dsd, 2,

```

```

function(xc) coef(coxph(y.recr_dsd ~ xc)), pure.p = apply(coefs.pdac_au,
2, function(xc) cor.test(samps.pdac_au$purity_qpure, xc, method = "kendall")$p.value),
pure.s = apply(coefs.pdac_au, 2, function(xc) cor.test(samps.pdac_au$purity_qpure,
xc, method = "kendall")$statistic))
temp.pvals = as.matrix(nmf.final.cpv.pvals[, grepl("\\.p$", colnames(nmf.final.cpv.pvals))])
temp.pvals.FWER = matrix(p.adjust(as.vector(temp.pvals), "holm"), nrow = nrow(temp.pvals))
colnames(temp.pvals.FWER) = paste(colnames(temp.pvals), "Holm", sep = ".")
temp.pvals.BY = matrix(p.adjust(as.vector(temp.pvals), "BY"), nrow = nrow(temp.pvals))
colnames(temp.pvals.BY) = paste(colnames(temp.pvals), "BY", sep = ".")
nmf.final.cpv.pvals = cbind(nmf.final.cpv.pvals, temp.pvals.FWER, temp.pvals.BY)
nmf.final.cpv.pvals = nmf.final.cpv.pvals[, order(colnames(nmf.final.cpv.pvals))]
nmf.final.cpv.pvals

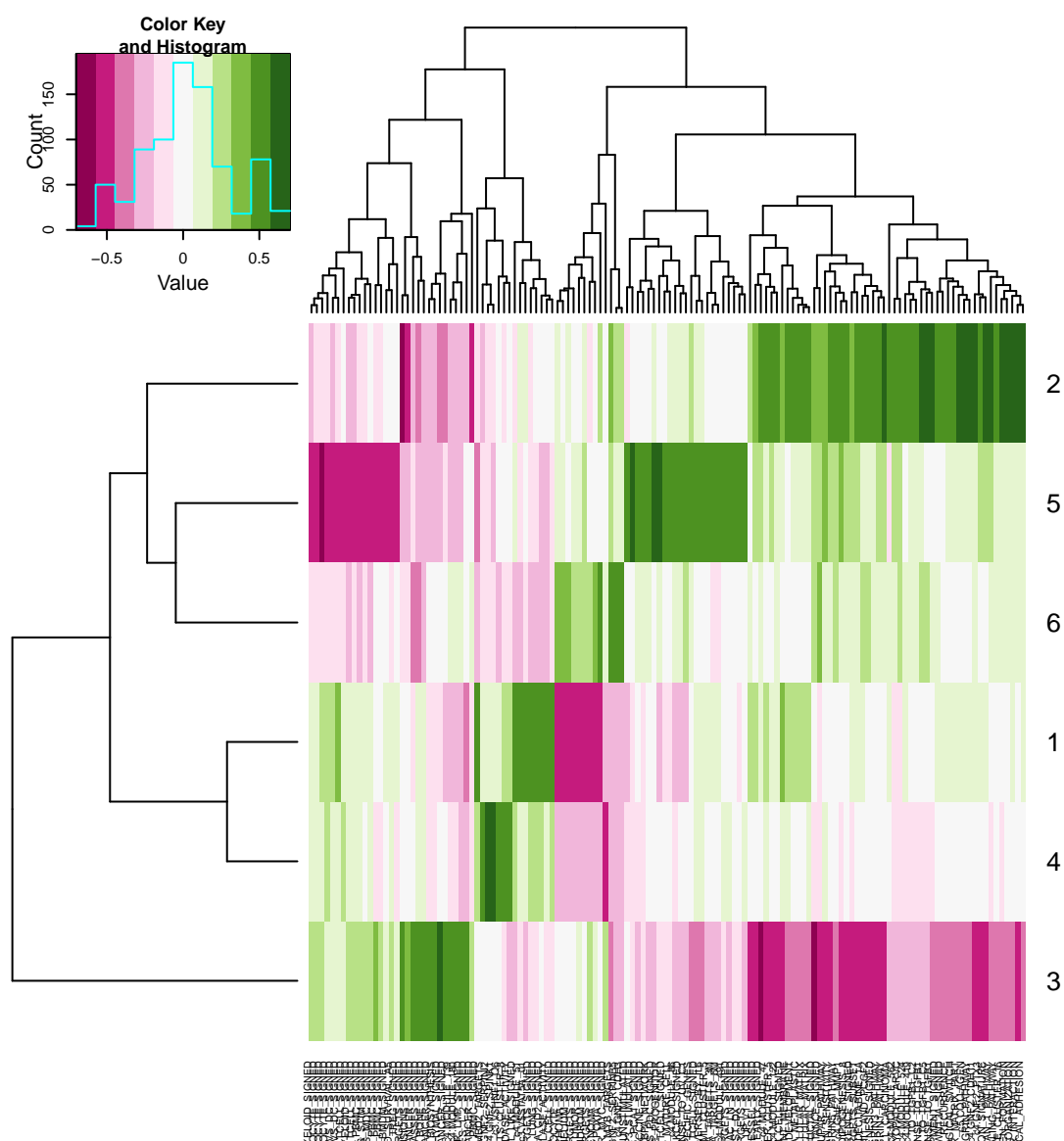
##      pure.p pure.p.BY pure.p.Holm pure.s surv.diag_dsd.c surv.diag_dsd.p
## 1 1.599e-07 7.245e-06 3.677e-06 -5.2408 -0.06024 7.029e-02
## 2 2.463e-02 1.717e-01 2.955e-01 -2.2472 0.07627 2.258e-05
## 3 8.582e-05 1.555e-03 1.716e-03 3.9275 -0.05558 1.474e-03
## 4 1.183e-01 6.305e-01 9.462e-01 -1.5620 -0.01401 3.410e-01
## 5 2.258e-01 1.000e+00 1.000e+00 -1.2113 0.08344 1.583e-04
## 6 6.804e-01 1.000e+00 1.000e+00 -0.4119 0.09724 2.271e-08
##      surv.diag_dsd.p.BY surv.diag_dsd.p.Holm surv.diag_rec.c surv.diag_rec.p
## 1 4.247e-01 7.029e-01 -0.05245 8.551e-02
## 2 5.117e-04 4.743e-04 0.05918 4.485e-04
## 3 1.336e-02 2.211e-02 -0.02194 1.410e-01
## 4 1.000e+00 1.000e+00 -0.02095 1.406e-01
## 5 2.391e-03 3.007e-03 0.07636 2.492e-04
## 6 2.058e-06 5.450e-07 0.08127 6.338e-07
##      surv.diag_rec.p.BY surv.diag_rec.p.Holm surv.recr_dsd.c surv.recr_dsd.p
## 1 4.843e-01 7.696e-01 -0.022044 0.5322169
## 2 5.081e-03 7.624e-03 0.053432 0.0052978
## 3 6.724e-01 9.841e-01 -0.051641 0.0084798
## 4 6.724e-01 9.841e-01 0.004636 0.7700435
## 5 3.226e-03 4.485e-03 0.049978 0.0426355
## 6 1.915e-05 1.394e-05 0.063927 0.0009377
##      surv.recr_dsd.p.BY surv.recr_dsd.p.Holm
## 1 1.000000 1.000000
## 2 0.043646 0.07417
## 3 0.064039 0.11024
## 4 1.000000 1.000000
## 5 0.275983 0.46899
## 6 0.009442 0.01500

```

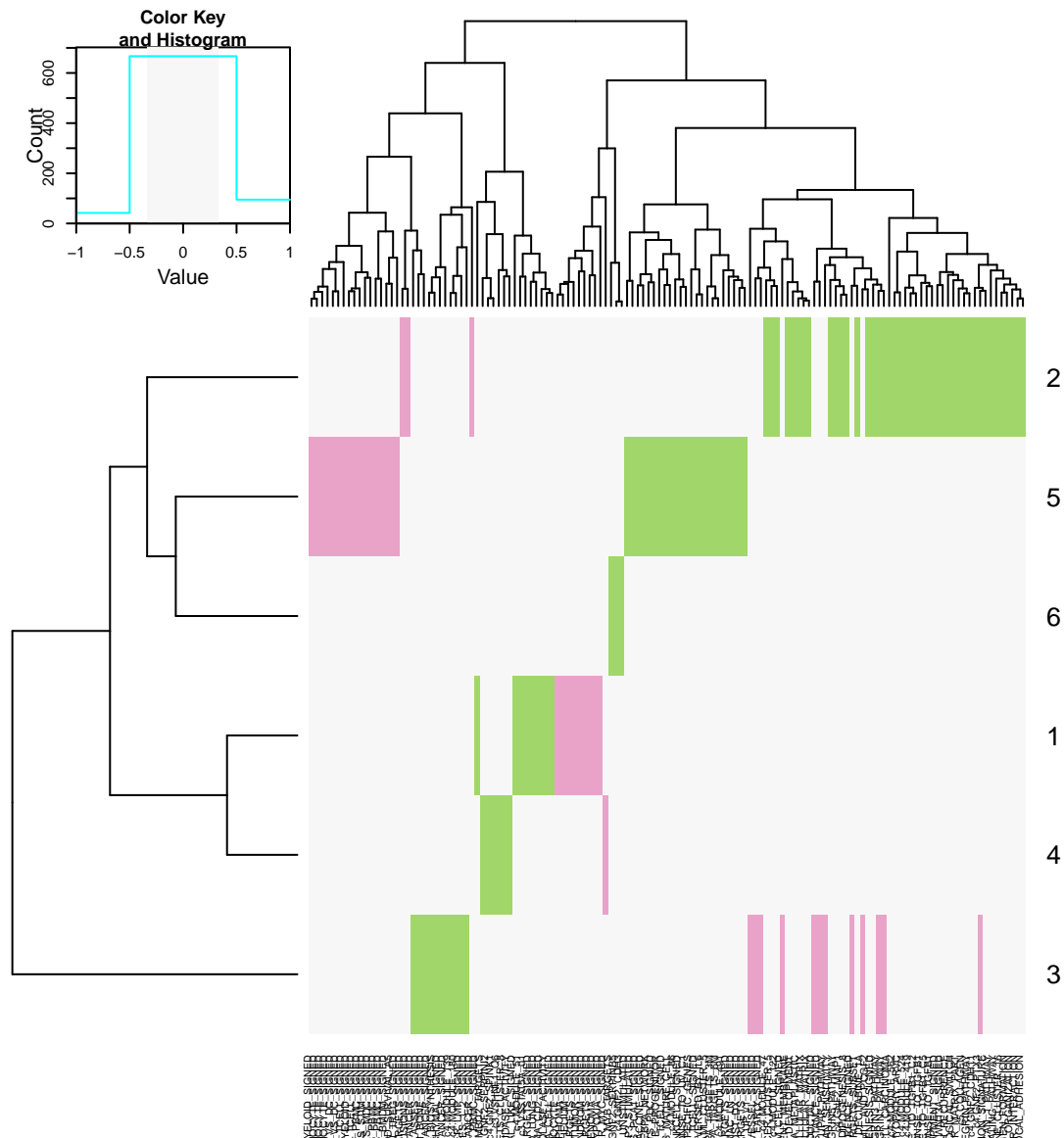
| | pure.p.Holm | pure.s | surv.diag_dsd.c | surv.diag_dsd.p.Holm | surv.diag_rec.c | surv.diag_rec.p.Holm | surv.recr_dsd.c | surv.recr_dsd.p.Holm |
|---|-------------|--------|-----------------|----------------------|-----------------|----------------------|-----------------|----------------------|
| 1 | 0.00 | -5.24 | -0.06 | 0.70 | -0.05 | 0.77 | -0.02 | 1.00 |
| 2 | 0.30 | -2.25 | 0.08 | 0.00 | 0.06 | 0.01 | 0.05 | 0.07 |
| 3 | 0.00 | 3.93 | -0.06 | 0.02 | -0.02 | 0.98 | -0.05 | 0.11 |
| 4 | 0.95 | -1.56 | -0.01 | 1.00 | -0.02 | 0.98 | 0.00 | 1.00 |
| 5 | 1.00 | -1.21 | 0.08 | 0.00 | 0.08 | 0.00 | 0.05 | 0.47 |
| 6 | 1.00 | -0.41 | 0.10 | 0.00 | 0.08 | 0.00 | 0.06 | 0.02 |

4.3 MSigDB score correlation thresholding

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



```
heatmap.2(nmf.final.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(nmf.final.msigdb.corr)
temp.sig_class = gsub("\\\\.*", "", temp.sig_id)
temp.nsigs = length(temp.sig_id)
temp.nmeta = nrow(nmf.final.msigdb.corr)
tables = lapply(1:temp.nmeta, function(metagene_i) {
  tapply(1:temp.nsigs, temp.sig_class, function(sig_class_is) {
    all_cors = nmf.final.msigdb.corr[, sig_class_is]
    this_cors = all_cors[metagene_i, ]
    this_ids = temp.sig_id[sig_class_is]

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

    sigs_to_report = which(this_sig_cors)

    if (length(sigs_to_report) == 0) {
      table = data.frame(GeneSet = c(), Correlation = c(), Metagenes = c())
    }
  })
})
```

```

    } else {
      table = data.frame(GeneSet = this_ids[sigs_to_report], Correlation = this_cors[sigs_to_report],
        Metagenes = apply(all_cors[, sigs_to_report, drop = FALSE],
          2, function(cors) {
            sel = abs(cors) >= sig.corr.threshold
            # A positive number implies that positive GSVA signal is associated with
            # worse prognosis
            paste(which(sel) * sign(cors[which(sel)]) * sign(nmf.final.cpv.pvals$surv.diag_dsd.
              collapse = ","))
          })
      table = table[order(-(table$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.KATSANOUELA_VL1_TARGETS_SIGNED      0.5096      -1
## 2      c2.ZHAN_MULTIPLE_MYELOMA_CD2_SIGNED      0.5086      -1
## 3      c2.SMID_BREAST_CANCER_NORMAL_LIKE_SIGNED      0.5080      -1
## 4      c2.GREENBAUM_E2A_TARGETS_SIGNED     -0.5009       1
## 5 c2.MARTORIATI_MDM4_TARGETS_NEUROEPITHELIUM_SIGNED     -0.5012       1
## 6      c2.YU_MYC_TARGETS_SIGNED     -0.5029       1
## 7      c2.SABATES_COLORECTAL_ADENOMA_SIGNED     -0.5036       1
## 8      c2.WINTER_HYPOXIA_SIGNED     -0.5241       1
## 9      c2.WEST_ADRENOCORTICAL_TUMOR_SIGNED     -0.5416       1
## 10     c2.SHIPP_DLBCL_VS_FOLLICULAR_LYMPHOMA_SIGNED     -0.5500       1
## 11     c2.LEE_EARLY_T_LYMPHOCYTE_SIGNED     -0.5517       1
## 12     c2.HAHTOLA_SEZARY_SYNDROM_SIGNED     -0.5641       1
##
## [[1]]$c3
##           GeneSet Correlation Metagenes
## 1 c3.V$STAT5A_01      0.5234      -1
##
## [[1]]$c4
##           GeneSet Correlation Metagenes
## 1 c4.MODULE_51      0.5399      -1
##
## [[1]]$c5
##           GeneSet Correlation Metagenes
## 1 c5.PHOSPHORIC_DIESTER_HYDROLASE_ACTIVITY      0.5113      -1
##
## [[1]]$c6
## data frame with 0 columns and 0 rows
##
## [[1]]$c7
##
##           GeneSet Correlation

```

```

## 1      c7.GSE20715_OH_VS_48H_OZONE_TLR4_KO_LUNG_SIGNED      0.5160
## 2 c7.GSE22886_IGM_MEMORY_BCELL_VS_BLOOD_PLASMA_CELL_SIGNED      0.5019
## 3      c7.GSE34205_HEALTHY_VS_RSV_INF_INFANT_PBMK_SIGNED      0.5002
## Metagenes
## 1      -1
## 2      -1
## 3      -1
##
##
## [[2]]
## [[2]]$c1
## data frame with 0 columns and 0 rows
##
## [[2]]$c2
##
##                                     GeneSet
## 1                                     c2.REACTOME_COLLAGEN_FORMATION
## 2                                c2.REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
## 3                                     c2.PID_SYNDECAN_1_PATHWAY
## 4                                c2.VERRECCHIA_DELAYED_RESPONSE_TO_TGFB1
## 5                                     c2.PID_INTEGRIN1_PATHWAY
## 6                                c2.PID_AVB3_INTEGRIN_PATHWAY
## 7                                c2.YAO_TEMPORAL_RESPONSE_TO_PROGESTERONE_CLUSTER_16
## 8                                c2.ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE
## 9                                     c2.KEGG_ECM_RECEPTOR_INTERACTION
## 10                                c2.VERRECCHIA_RESPONSE_TO_TGFB1_C5
## 11                                c2.VERRECCHIA_EARLY_RESPONSE_TO_TGFB1
## 12                                     c2.KEGG_FOCAL_ADHESION
## 13                                c2.FARMER_BREAST_CANCER_CLUSTER_5
## 14                                c2.MAHADEVAN_GIST_MORPHOLOGICAL_SWITCH
## 15                                c2.CAIRO_LIVER_DEVELOPMENT_SIGNED
## 16                                     c2.PID_INTEGRIN3_PATHWAY
## 17                                c2.KEGG_BASAL_CELL_CARCINOMA
## 18                                     c2.BURTON_ADIPOGENESIS_8
## 19                                c2.VERRECCHIA_RESPONSE_TO_TGFB1_C2
## 20                                     c2.CROMER_TUMORIGENESIS_SIGNED
## 21                                c2.WIEDERSCHAIN_TARGETS_OF_BMI1_AND_PCGF2
## 22                                c2.ROZANOV_MMP14_TARGETS_SUBSET
## 23                                c2.PID_WNT_SIGNALING_PATHWAY
## 24 c2.KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC
## 25                                c2.LABBE_TARGETS_OF_TGFB1_AND_WNT3A_SIGNED
## 26                                c2.LIEN_BREAST_CARCINOMA_METAPLASTIC
## 27                                     c2.PID_INTEGRIN5_PATHWAY
## 28                                c2.LINDGREN_BLADDER_CANCER_HIGH_RECURRENCE
## 29                                     c2.POTTI_TOPOTECAN_SENSITIVITY
## 30                                c2.MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSIONS_SIGNED
## 31                                c2.BERTUCCI_MEDULLARY_VS_DUCTAL_BREAST_CANCER_SIGNED
## 32                                c2.PASINI_SUZ12_TARGETS_SIGNED
## Correlation Metagenes
## 1      0.6567      2
## 2      0.6365      2
## 3      0.6355      2
## 4      0.6178      2
## 5      0.6067      2,-3
## 6      0.6020      2

```



```

## 7      0.5963      2
## 8      0.5960      2
## 9      0.5953      2
## 10     0.5849      2
## 11     0.5829      2
## 12     0.5758      2
## 13     0.5634      2
## 14     0.5587      2
## 15     0.5429      2
## 16     0.5409      2,-3
## 17     0.5396      2
## 18     0.5346      2
## 19     0.5312      2
## 20     0.5258      2
## 21     0.5242      2
## 22     0.5228      2
## 23     0.5171      2
## 24     0.5114      2
## 25     0.5081      2
## 26     0.5077      2
## 27     0.5074      2
## 28     0.5047      2
## 29     0.5017      2
## 30     -0.5087     -2
## 31     -0.5436     -2
## 32     -0.5916     -2
##
## [[2]]$c3
## data frame with 0 columns and 0 rows
##
## [[2]]$c4
##      GeneSet Correlation Metagenes
## 1 c4.GNF2_CDH11      0.6295      2
## 2 c4.GNF2_PTX3      0.5533      2,-3
## 3 c4.MODULE_122      0.5369      2
## 4 c4.GNF2_MMP1      0.5366      2
## 5 c4.MODULE_562      0.5178      2
## 6 c4.MODULE_524      0.5144      2
## 7 c4.MODULE_419      0.5087      2
## 8 c4.MODULE_47       0.5003      2
##
## [[2]]$c5
##      GeneSet Correlation Metagenes
## 1 c5.COLLAGEN      0.6496      2
## 2 c5.EXTRACELLULAR_MATRIX_PART      0.6157      2
## 3 c5.PROTEINACEOUS_EXTRACELLULAR_MATRIX      0.5339      2
## 4 c5.EXTRACELLULAR_MATRIX      0.5235      2
## 5 c5.BASEMENT_MEMBRANE      0.5148      2
## 6 c5.SKELETAL_DEVELOPMENT      0.5101      2
##
## [[2]]$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      0.5882
## 2                c2.REACTOME_GLYCEROPHOSPHOLIPID_BIOSYNTHESIS  0.5269
## 3                c2.SMID_BREAST_CANCER_RELAPSE_IN_BONE_SIGNED  0.5215
## 4                        c2.LIU_PROSTATE_CANCER_SIGNED         0.5202
## 5                c2.WAMUNYOKOLI_OVARIAN_CANCER_GRADES_1_2_SIGNED 0.5178
## 6  c2.WANG_BARRETTES_ESOPHAGUS_AND_ESOPHAGUS_CANCER_SIGNED  0.5175
## 7                c2.WAMUNYOKOLI_OVARIAN_CANCER_LMP_SIGNED      0.5165
## 8                        c2.WALLACE_PROSTATE_CANCER_SIGNED     0.5155
## 9                c2.DOANE_BREAST_CANCER_CLASSES_SIGNED        0.5111
## 10               c2.WOO_LIVER_CANCER_RECURRENCE_SIGNED        -0.5000
## 11                        c2.PID_UPA_UPAR_PATHWAY              -0.5011
## 12               c2.SUZUKI_RESPONSE_TO_TSA_AND_DECITABINE_1A   -0.5141
## 13               c2.HUANG_DASATINIB_RESISTANCE_SIGNED          -0.5145
## 14                       c2.LIM_MAMMARY_STEM_CELL_SIGNED       -0.5175
## 15                       c2.PID_INTEGRIN3_PATHWAY              -0.5175
## 16               c2.ROY_WOUND_BLOOD_VESSEL_SIGNED             -0.5235
## 17                       c2.PID_INTEGRIN1_PATHWAY              -0.5248
## 18  c2.LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_SIGNED     -0.6110
## 19               c2.VECCHI_GASTRIC_CANCER_ADVANCED_VS_EARLY_SIGNED -0.6217
##      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          -2,3
## 16           3
## 17          -2,3
## 18           3
## 19           3
##
## [[3]]$c3
## data frame with 0 columns and 0 rows
##
## [[3]]$c4
##      GeneSet Correlation Metagenes

```

```

## 1 c4.MODULE_180      0.5148      -3
## 2 c4.MODULE_139      0.5031      -3
## 3 c4.GNF2_PTX3      -0.5155      -2,3
##
## [[3]]$c5
## data frame with 0 columns and 0 rows
##
## [[3]]$c6
##           GeneSet Correlation Metagenes
## 1 c6.LEF1_UP.V1_SIGNED      -0.5597      3
##
## [[3]]$c7
## data frame with 0 columns and 0 rows
##
##
## [[4]]
## [[4]]$c1
## data frame with 0 columns and 0 rows
##
## [[4]]$c2
##           GeneSet Correlation Metagenes
## 1          c2.BERGER_MBD2_TARGETS      0.5646      -4
## 2 c2.TERAMOTO_OPN_TARGETS_CLUSTER_8      0.5274      -4
## 3    c2.LEE_LIVER_CANCER_MYC_SIGNED     -0.5203      4
##
## [[4]]$c3
##           GeneSet Correlation Metagenes
## 1 c3.V$HNF1_Q6      0.5124      -4
##
## [[4]]$c4
##           GeneSet Correlation Metagenes
## 1    c4.GNF2_SPINK1      0.7013      -4
## 2 c4.GNF2_SERPINI2      0.6798      -4
##
## [[4]]$c5
##           GeneSet Correlation Metagenes
## 1 c5.CARBOXYPEPTIDASE_ACTIVITY      0.5342      -4
##
## [[4]]$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.IVANOV_HEMATOPOIESIS_LATE_PROGENITOR      0.6114      5
## 2          c2.MARSON_BOUND_BY_FOXP3_STIMULATED      0.5798      5
## 3          c2.SESTO_RESPONSE_TO_UV_C1      0.5491      5

```

```

## 4 c2.NAKAYAMA_SOFT_TISSUE_TUMORS_PCA1_SIGNED 0.5413 5
## 5 c2.IVANOV_HEMATOPOIESIS_MATURE_CELL 0.5410 5
## 6 c2.MARTINEZ_RB1_AND_TP53_TARGETS_SIGNED 0.5369 5
## 7 c2.KAMIKUBO_MYELOID_CEBPA_NETWORK 0.5280 5
## 8 c2.MARSON_BOUND_BY_FOXP3_UNSTIMULATED 0.5154 5
## 9 c2.VALK_AML_CLUSTER_5 0.5134 5
## 10 c2.RAGHAVACHARI_PLATELET_SPECIFIC_GENES 0.5124 5
## 11 c2.MARTINEZ_TP53_TARGETS_SIGNED 0.5066 5
## 12 c2.LIAN_LIPA_TARGETS_6M 0.5008 5
## 13 c2.BROCKE_APOPTOSIS_REVERSED_BY_IL6 0.5001 5
## 14 c2.LIAN_LIPA_TARGETS_3M 0.5001 5
## 15 c2.SHEDDEN_LUNG_CANCER_GOOD_SURVIVAL_A5 -0.5080 -5
##
## [[5]]$c3
## data frame with 0 columns and 0 rows
##
## [[5]]$c4
## GeneSet Correlation Metagenes
## 1 c4.MODULE_86 0.5240 5
## 2 c4.MODULE_491 0.5223 5
##
## [[5]]$c5
## data frame with 0 columns and 0 rows
##
## [[5]]$c6
## data frame with 0 columns and 0 rows
##
## [[5]]$c7
## GeneSet
## 1 c7.GSE29618_MONOCYTE_VS_PDC_SIGNED
## 2 c7.GSE29618_MONOCYTE_VS_PDC_DAY7_FLU_VACCINE_SIGNED
## 3 c7.GSE3982_MAC_VS_NKCELL_SIGNED
## 4 c7.GSE15767_MED_VS_SCS_MAC_LN_SIGNED
## 5 c7.GSE3982_BASOPHIL_VS_CENT_MEMORY_CD4_TCELL_SIGNED
## 6 c7.GSE11057_PBMC_VS_MEM_CD4_TCELL_SIGNED
## 7 c7.GSE9006_HEALTHY_VS_TYPE_1_DIABETES_PBMC_4MONTH_POST_DX_SIGNED
## 8 c7.GSE11057_NAIVE_CD4_VS_PBMC_CD4_TCELL_SIGNED
## 9 c7.GSE3982_DC_VS_MAC_LPS_STIM_SIGNED
## 10 c7.GSE17721_POLYIC_VS_GARDIQUIMOD_1H_BMDM_SIGNED
## 11 c7.GSE29618_PDC_VS_MDC_SIGNED
## 12 c7.GSE6269_HEALTHY_VS_STREP_AUREUS_INF_PBMC_SIGNED
## 13 c7.GSE20366_CD103_KLRG1_DP_VS_DN_TREG_SIGNED
## 14 c7.GSE22886_NAIVE_CD8_TCELL_VS_DC_SIGNED
## 15 c7.GSE22886_NAIVE_CD8_TCELL_VS_MONOCYTE_SIGNED
## 16 c7.GSE11057_CD4_CENT_MEM_VS_PBMC_SIGNED
## 17 c7.GSE22886_NAIVE_CD4_TCELL_VS_DC_SIGNED
## 18 c7.GSE11057_NAIVE_VS_MEMORY_CD4_TCELL_SIGNED
## 19 c7.GSE11057_CD4_EFF_MEM_VS_PBMC_SIGNED
## 20 c7.GSE10325_BCELL_VS_MYELOID_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 0.5760 5

```

```

## 2      0.5712      5
## 3      0.5573      5
## 4      0.5502      5
## 5      0.5352      5
## 6      0.5314      5
## 7      0.5209      5
## 8     -0.5042     -5
## 9     -0.5042     -5
## 10     -0.5076     -5
## 11     -0.5086     -5
## 12     -0.5086     -5
## 13     -0.5233     -5
## 14     -0.5267     -5
## 15     -0.5274     -5
## 16     -0.5352     -5
## 17     -0.5355     -5
## 18     -0.5379     -5
## 19     -0.5420     -5
## 20     -0.5519     -5
## 21     -0.5610     -5
## 22     -0.5699     -5
## 23     -0.5825     -5
##
##
## [[6]]
## [[6]]$c1
## data frame with 0 columns and 0 rows
##
## [[6]]$c2
##           GeneSet Correlation Metagenes
## 1 c2.LEI_MYB_TARGETS      0.509      6
##
## [[6]]$c3
## data frame with 0 columns and 0 rows
##
## [[6]]$c4
##           GeneSet Correlation Metagenes
## 1      c4.GNF2_CDH3      0.5630      6
## 2 c4.GNF2_SERPINB5      0.5488      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

```

4.3.1 Outcome: Diagnosis to disease-specific death

```

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

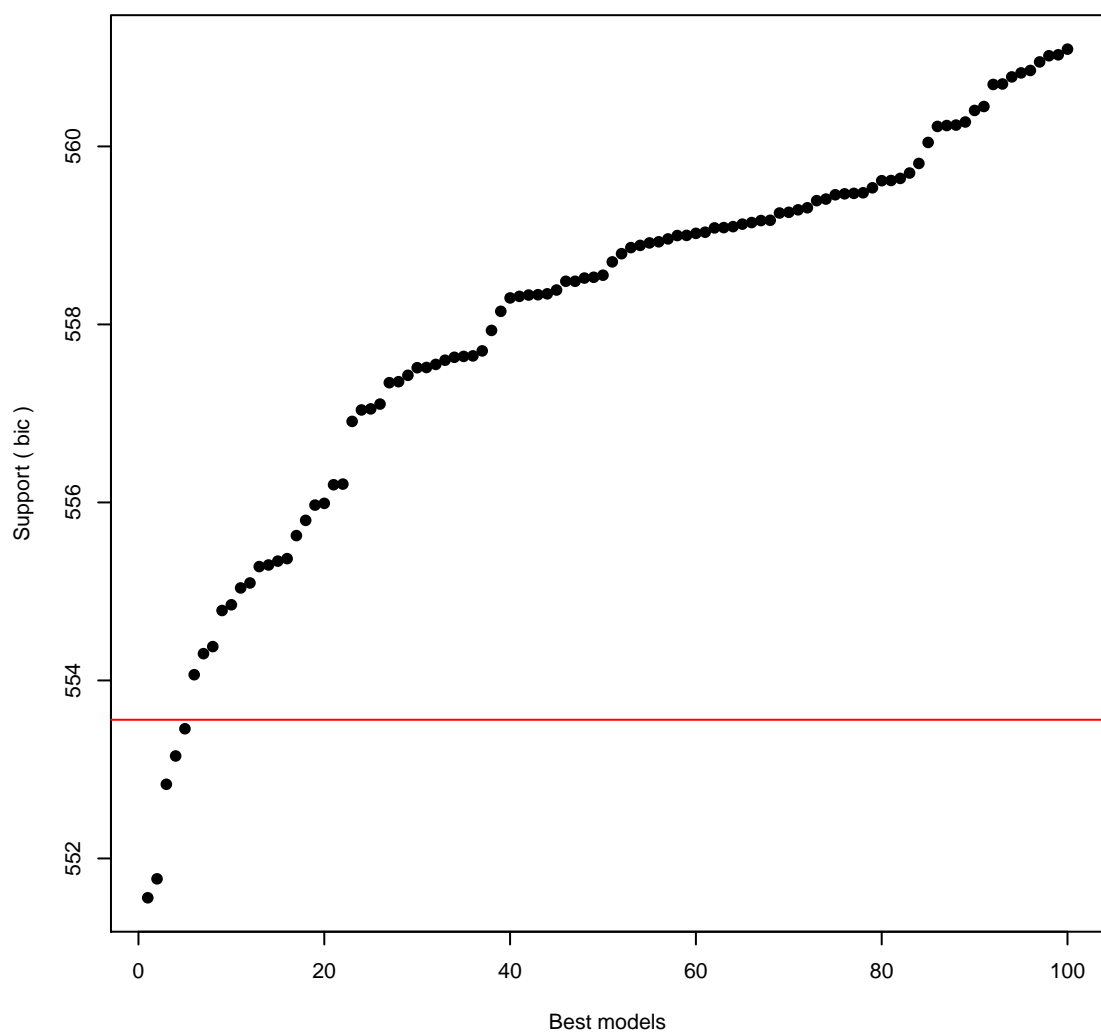
```

```
## 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
```

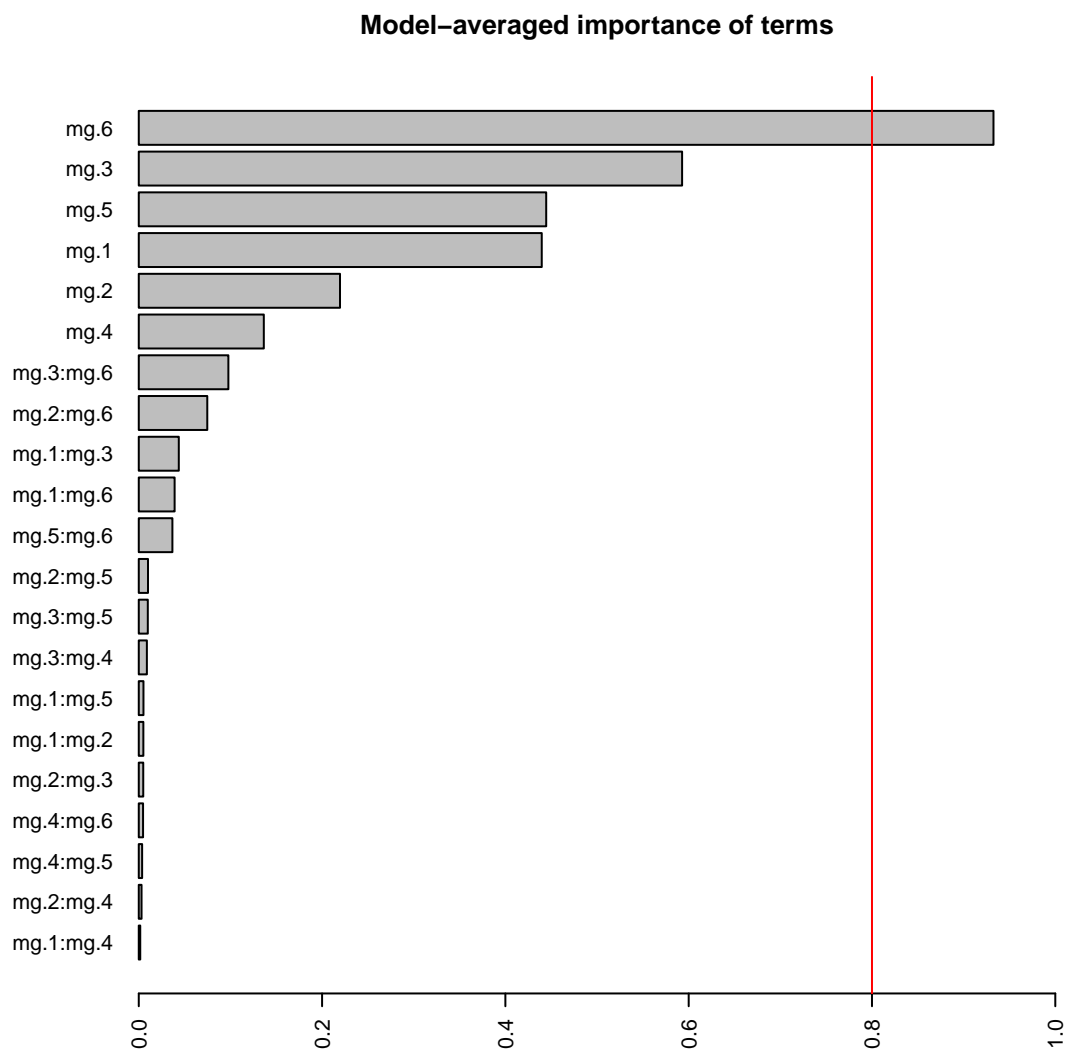
```
plot(asreg.result, type = "p")
```

All-subsets regression

IC profile

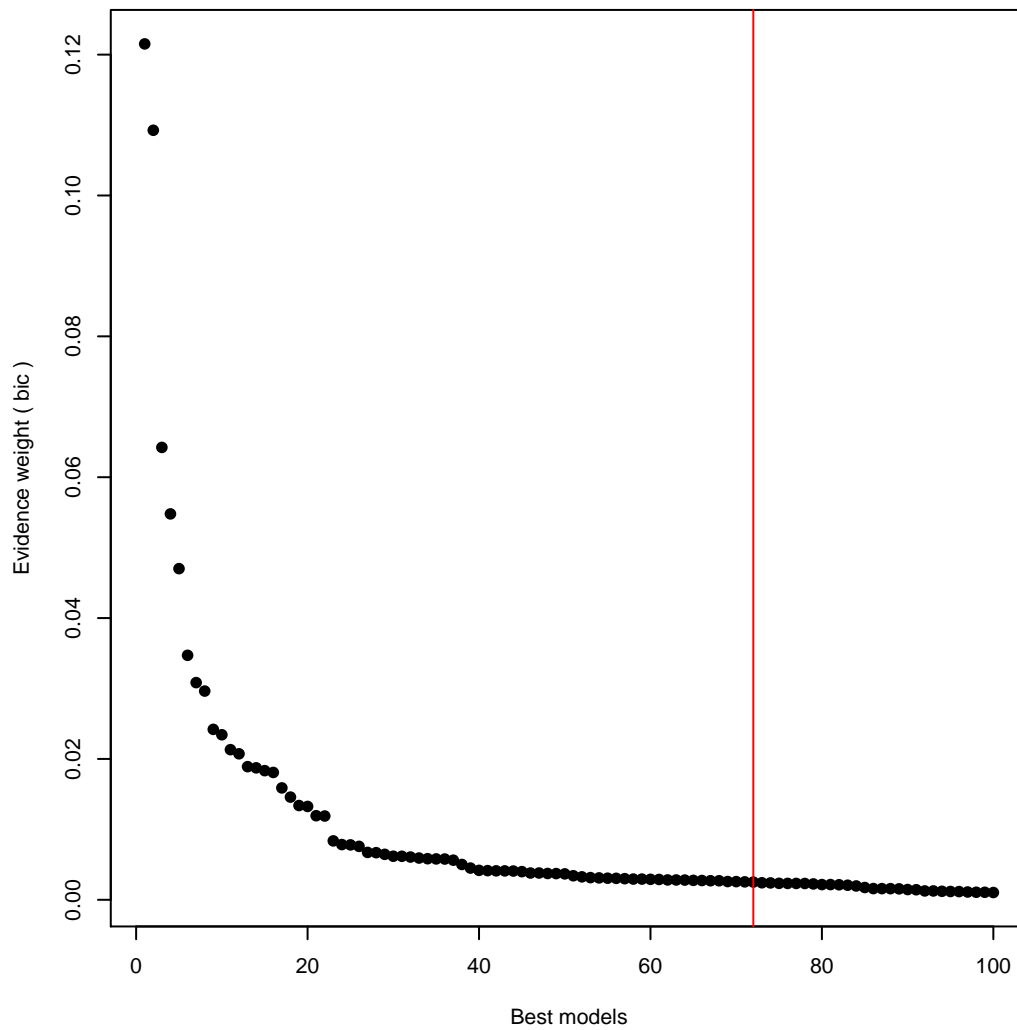


```
plot(asreg.result, type = "s")
```



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


Profile of model weights



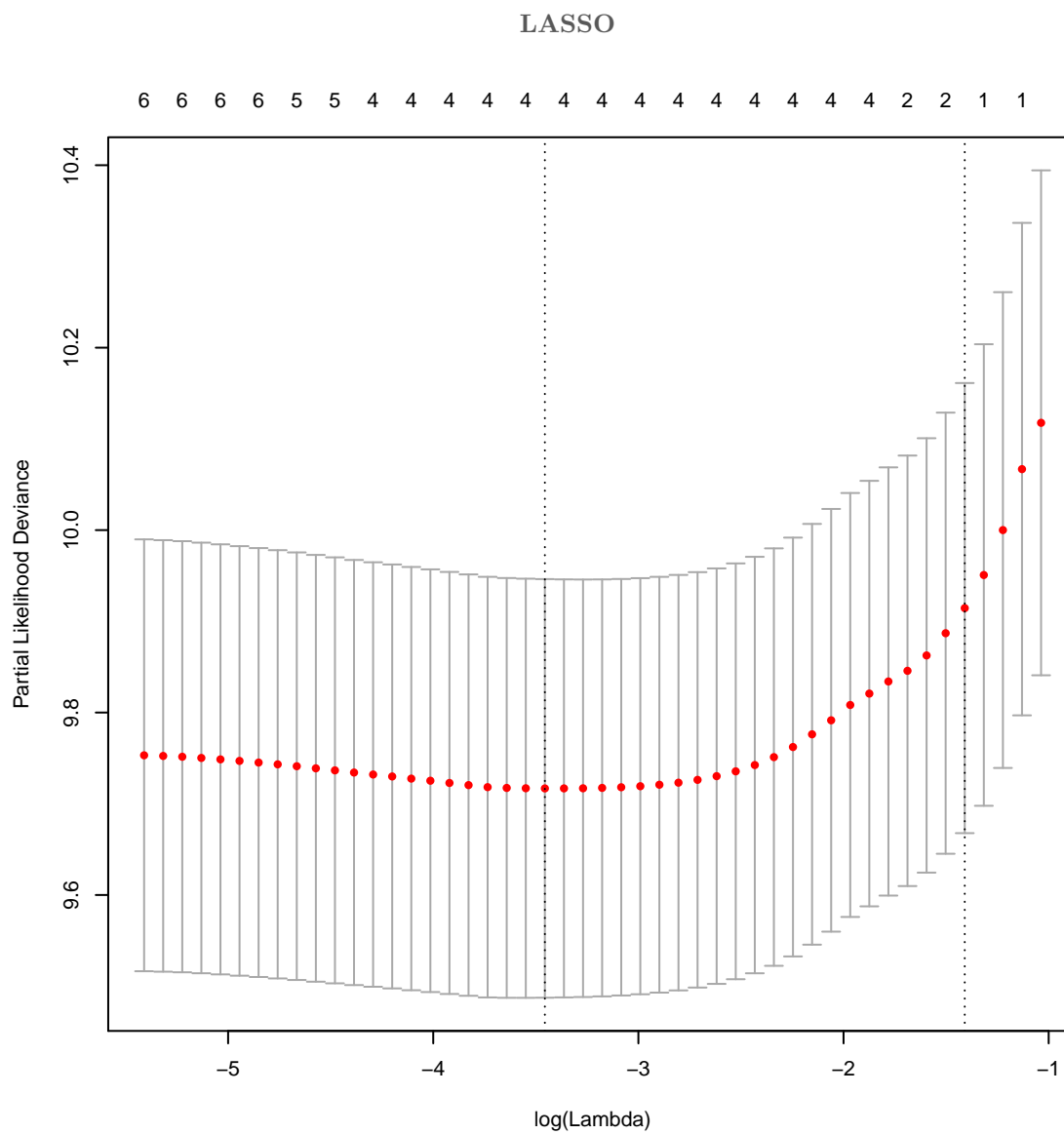
```
glmnet.coef.1se

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

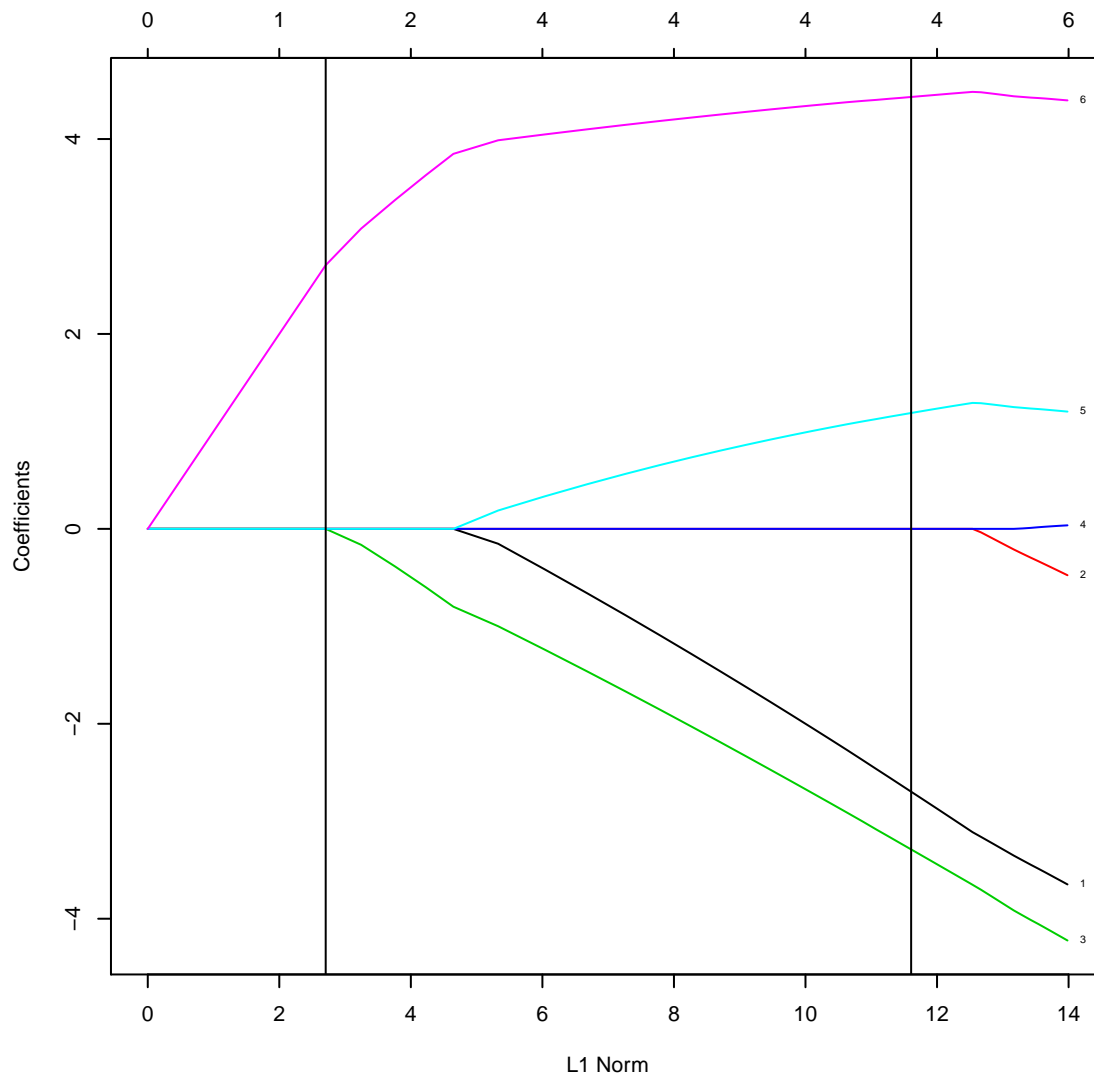
glmnet.coef.min
```

```
## 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)))
```



5 Session information

```
session_info
```

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

```
## [1] splines parallel methods stats graphics grDevices utils
## [8] datasets base
##
## other attached packages:
## [1] doParallel_1.0.8 iterators_1.0.7 foreach_1.4.2
## [4] ahaz_1.14 survival_2.37-7 NMF_0.20.5
## [7] Biobase_2.26.0 BiocGenerics_0.12.1 cluster_1.15.3
## [10] rngtools_1.2.4 pkgmaker_0.22 registry_0.2
## [13] energy_1.6.2 glmnet_1.9-8 Matrix_1.1-4
## [16] glmulti_1.0.7 rJava_0.9-6
##
## loaded via a namespace (and not attached):
## [1] boot_1.3-13 codetools_0.2-9 colorspace_1.2-4
## [4] compiler_3.1.1 digest_0.6.4 ggplot2_1.0.0
## [7] grid_3.1.1 gridBase_0.4-7 gtable_0.1.2
## [10] lattice_0.20-29 MASS_7.3-35 munsell_0.4.2
## [13] plyr_1.8.1 proto_0.3-10 RColorBrewer_1.0-5
## [16] Rcpp_0.11.3 reshape2_1.4 scales_0.2.4
## [19] stringr_0.6.2 tools_3.1.1 xtable_1.7-4

sessionInfo()

## R version 3.1.1 (2014-07-10)
## Platform: x86_64-unknown-linux-gnu (64-bit)
##
## locale:
## [1] LC_CTYPE=en_AU.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_AU.UTF-8 LC_COLLATE=en_AU.UTF-8
## [5] LC_MONETARY=en_AU.UTF-8 LC_MESSAGES=en_AU.UTF-8
## [7] LC_PAPER=en_AU.UTF-8 LC_NAME=en_AU.UTF-8
## [9] LC_ADDRESS=en_AU.UTF-8 LC_TELEPHONE=en_AU.UTF-8
## [11] LC_MEASUREMENT=en_AU.UTF-8 LC_IDENTIFICATION=en_AU.UTF-8
##
## attached base packages:
## [1] 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_0.20.5
## [10] Biobase_2.26.0 BiocGenerics_0.12.1 cluster_1.15.3
## [13] rngtools_1.2.4 pkgmaker_0.22 registry_0.2
## [16] energy_1.6.2 survival_2.37-7 knitr_1.8
##
## loaded via a namespace (and not attached):
## [1] bitops_1.0-6 boot_1.3-13 caTools_1.17.1
## [4] codetools_0.2-9 colorspace_1.2-4 digest_0.6.4
## [7] doParallel_1.0.8 evaluate_0.5.5 foreach_1.4.2
## [10] formatR_1.0 gdata_2.13.3 ggplot2_1.0.0
## [13] grid_3.1.1 gridBase_0.4-7 gtable_0.1.2
## [16] gtools_3.4.1 highr_0.4 iterators_1.0.7
## [19] KernSmooth_2.23-13 labeling_0.3 lattice_0.20-29
## [22] MASS_7.3-35 munsell_0.4.2 plyr_1.8.1
```

```
## [25] proto_0.3-10      Rcpp_0.11.3      reshape2_1.4
## [28] scales_0.2.4      stringr_0.6.2    tools_3.1.1
```