Combined Analysis of SIS-NMF Metagenes

November 24, 2014

1 Data used

Diagnosis and surgery timepoints were determined to be very similar, and therefore surgery times will not be examined further. The intervals of interest are then:

- Diagnosis to DSD
- Diagnosis to Recurrence
- Recurrence to DSD

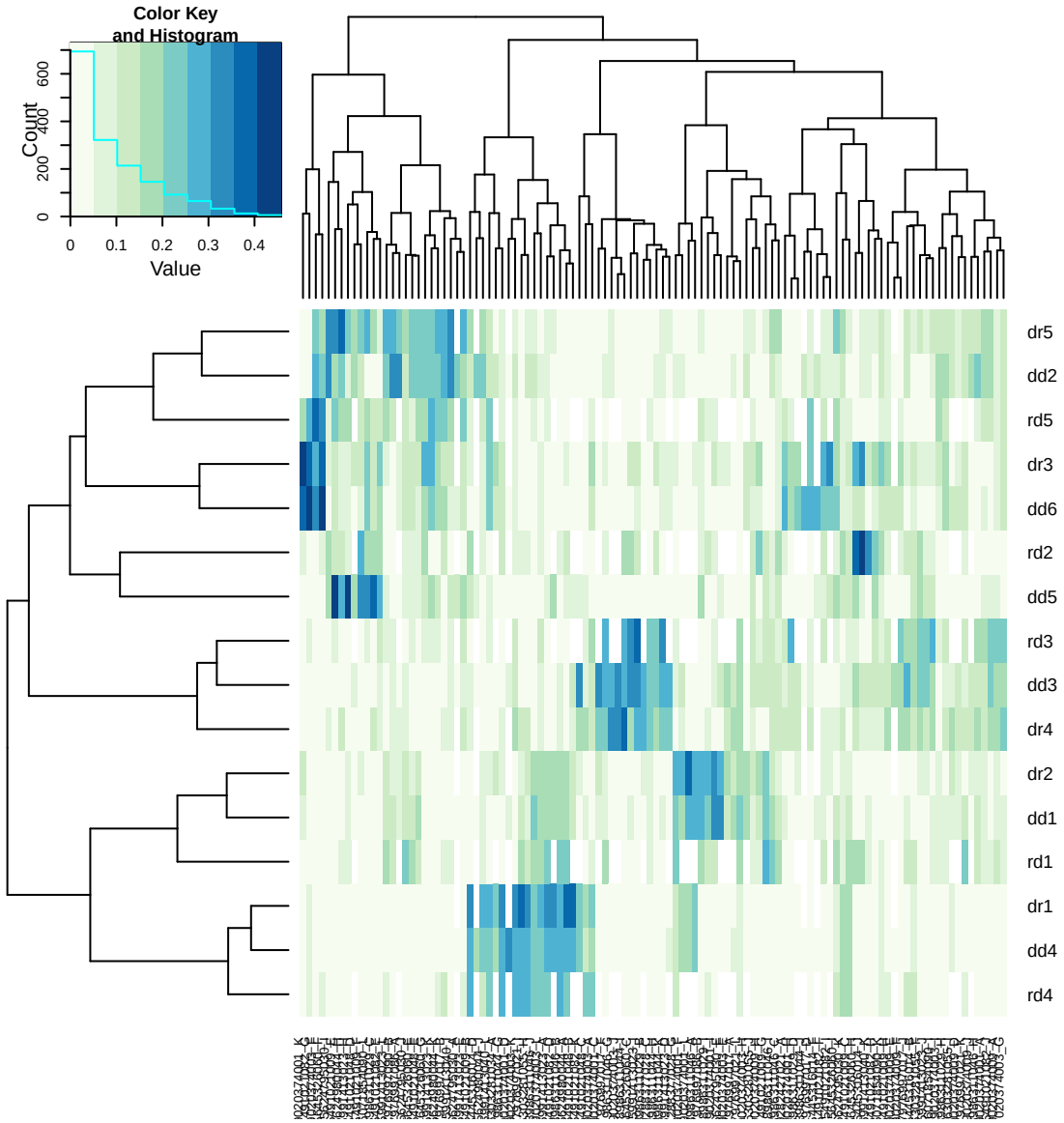
2 Preparation and data loading

```
fits = list()
load("../09_SIS_diag_dsd/image.rda")
fits$diag_dsd = xlin.scaled.sel.nmf
load("../11_SIS_recr_dsd/image.rda")
fits$recr_dsd = xlin.scaled.sel.nmf
load("../12_SIS_diag_rec/image.rda")
fits$diag_rec = xlin.scaled.sel.nmf
temp = NA
temp = ls()
rm(list = temp[!(temp %in% c("fits", "cpvs", "samples", "features"))])
```

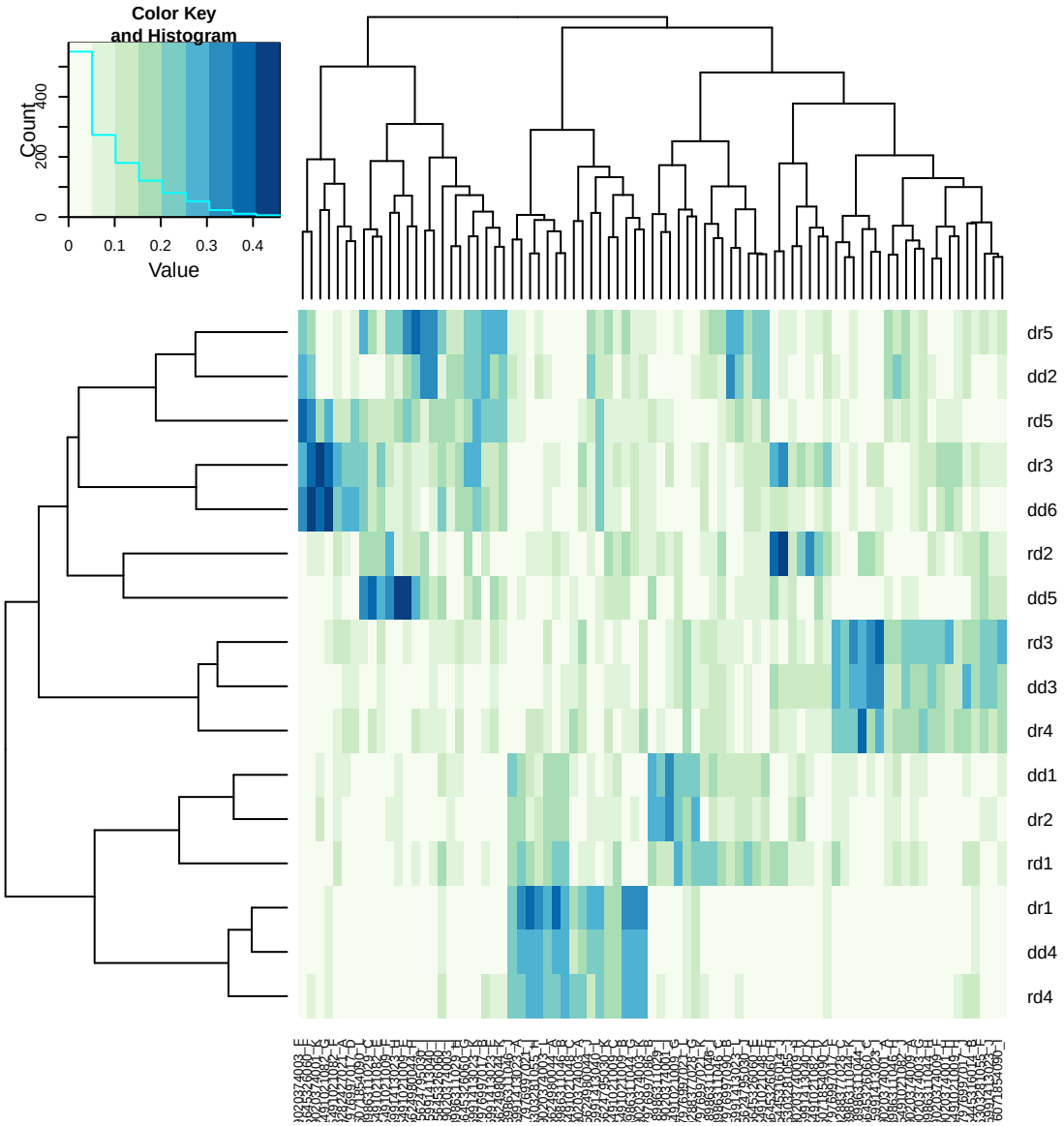
3 Coefficient merging

```
library(NMF)
patient_ids = sort(unique(unlist(lapply(fits, function(x) colnames(coef(x))))))
merged_coefs = do.call(rbind, lapply(names(fits), function(fit_name) {
    coefs = coef(fits[[fit_name]])
    coefs = coefs[, match(patient_ids, colnames(coefs))]
    rownames(coefs) = paste(substr(fit_name, 1, 1), substr(fit_name, 6, 6),
        1:nrow(coefs), sep = "")
    coefs
}))
merged_coefs_allpresent = merged_coefs[, apply(!is.na(merged_coefs), 2, all)]
```

```
library(RColorBrewer)
library(gplots)

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

heatmap.2(merged_coefs, scale = "none", trace = "none", col = brewer.pal(9,
    "GnBu"))
```

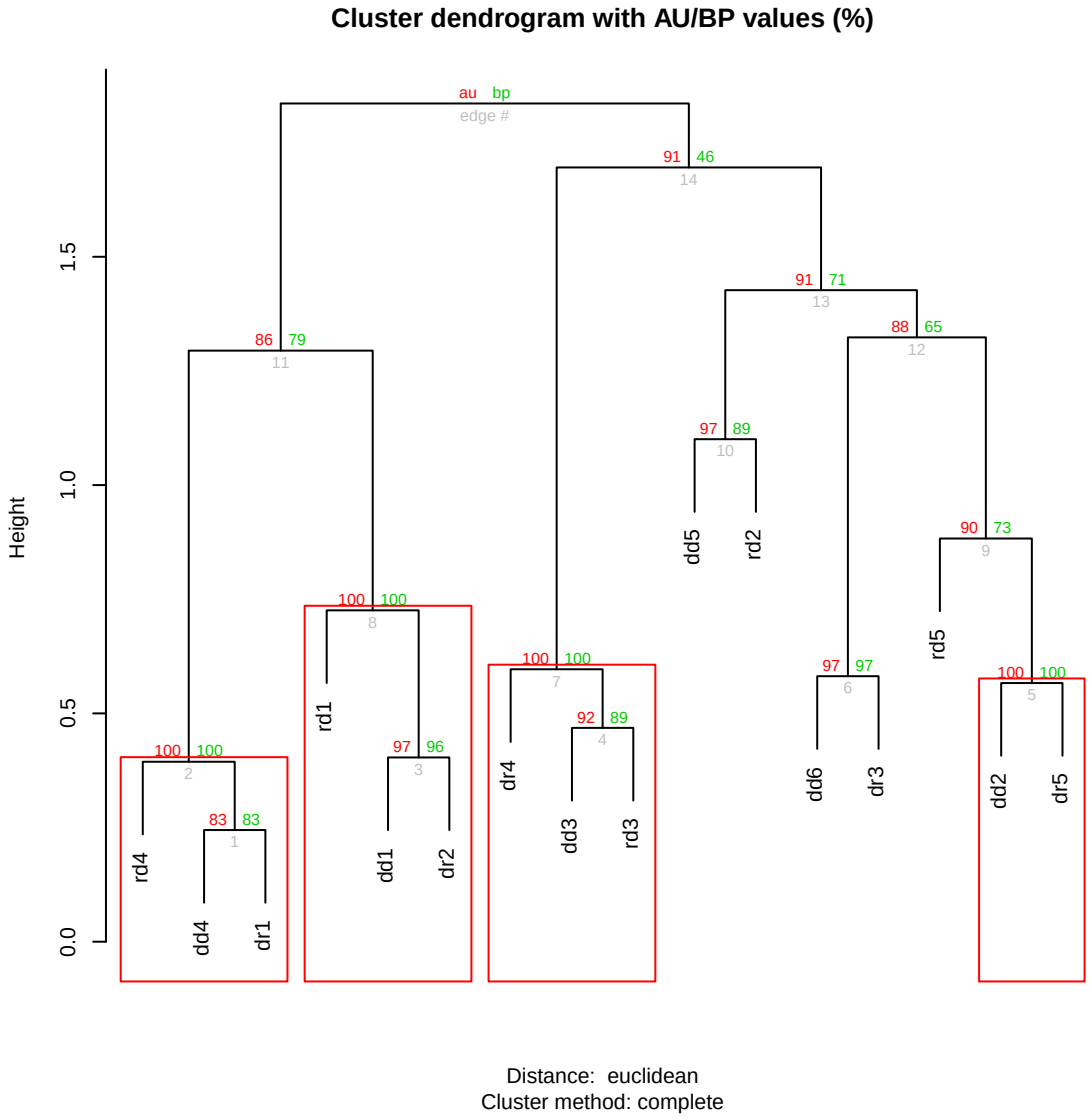



```
heatmap.2(merged_coefs_allpresent, scale = "none", trace = "none", col = brewer.pal(9,
    "GnBu"))
```

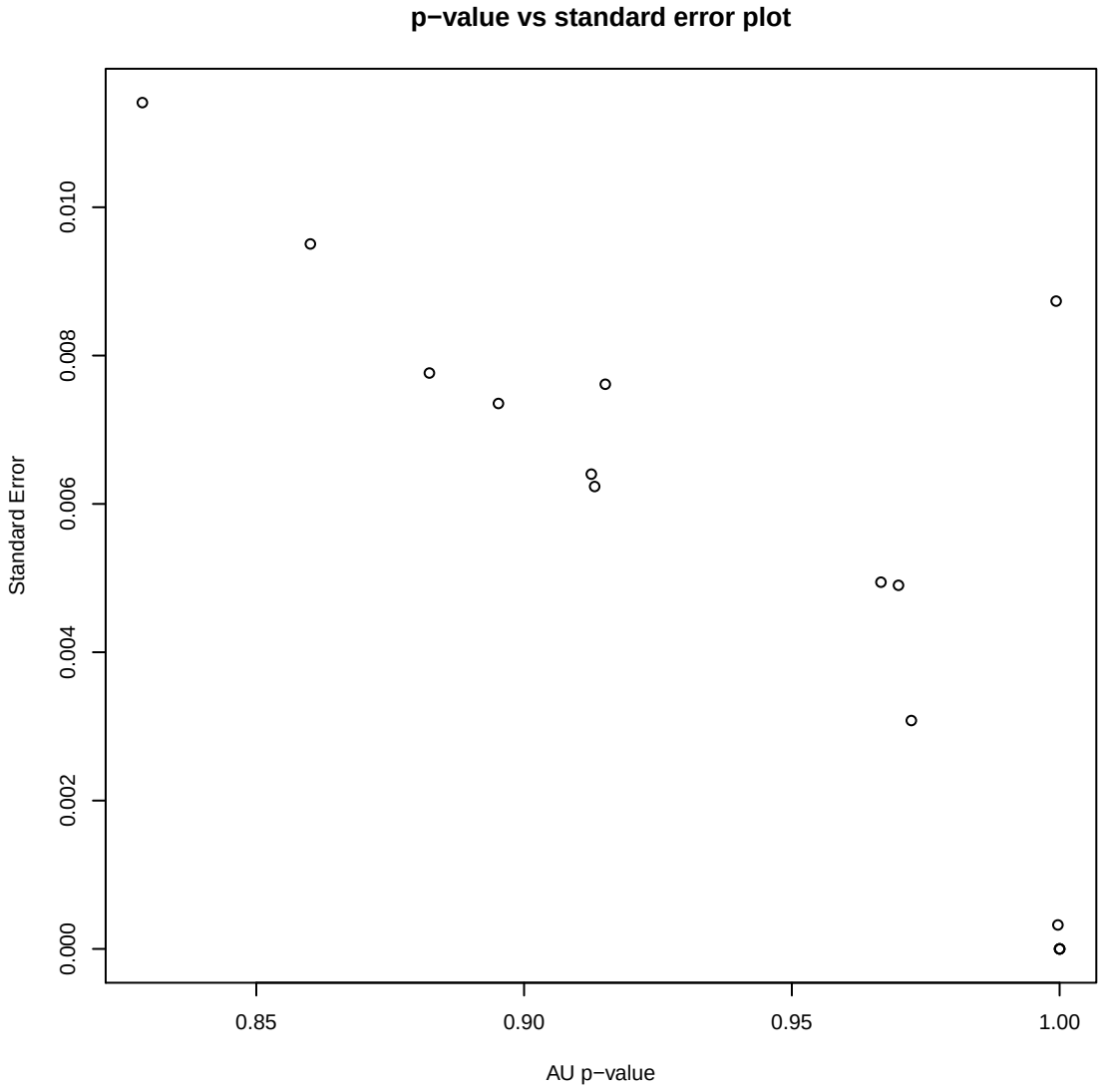



```
library(pvclust)
set.seed(1234567890)
cons_clust = pvclust(t(merged_coefs), method.hclust = "complete", method.dist = "euclidean",
    nboot = 4000)

## Bootstrap (r = 0.5)... Done.
## Bootstrap (r = 0.6)... Done.
## Bootstrap (r = 0.7)... Done.
## Bootstrap (r = 0.8)... Done.
## Bootstrap (r = 0.9)... Done.
## Bootstrap (r = 1.0)... Done.
## Bootstrap (r = 1.1)... Done.
## Bootstrap (r = 1.2)... Done.
## Bootstrap (r = 1.3)... Done.
## Bootstrap (r = 1.4)... Done.
```

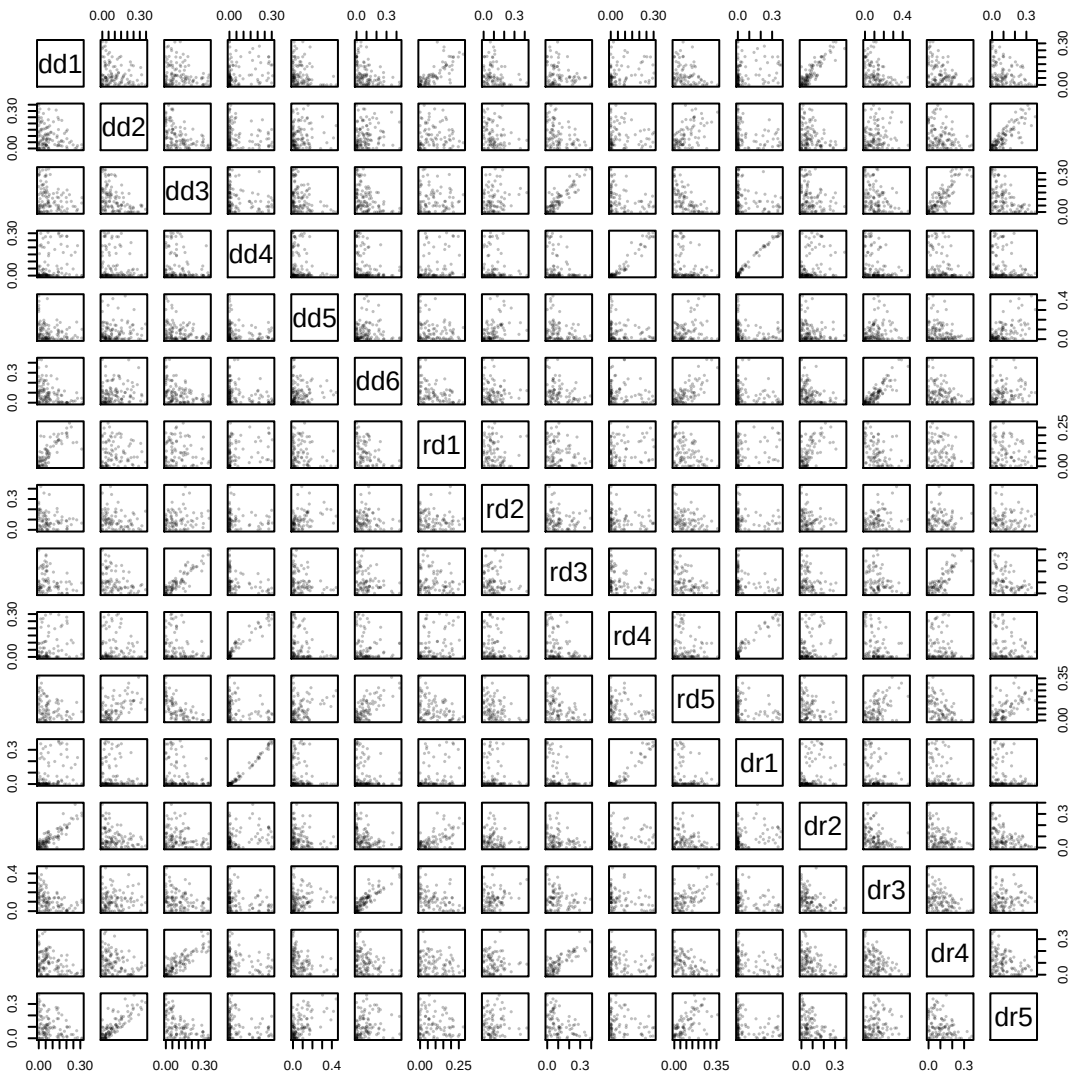
```
library(pvclust)
plot(cons_clust)
pvrect(cons_clust, alpha = 0.99)
```




```
seplot(cons_clust)
```

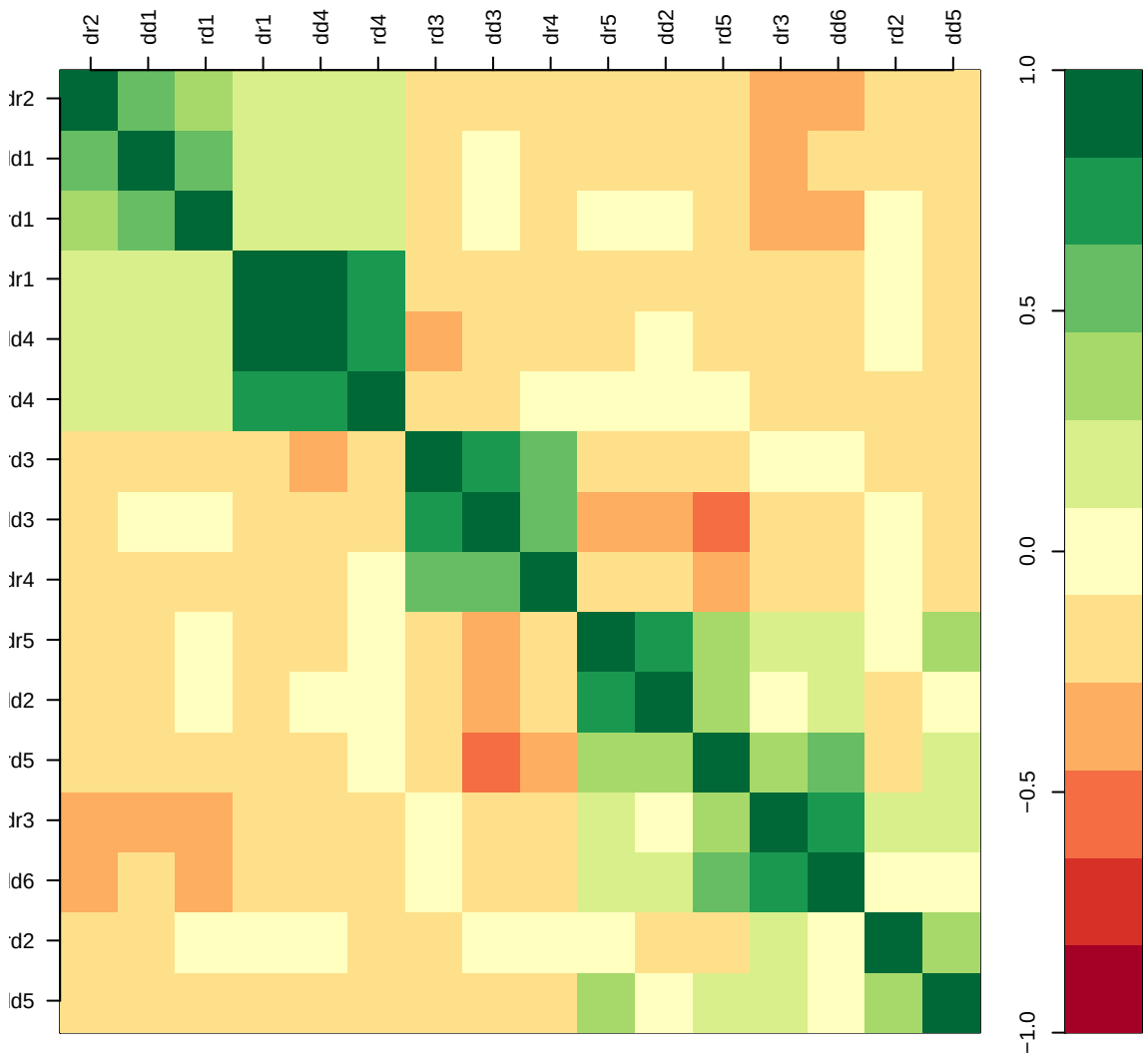



```
pairs(t(merged_coefs), pch = 16, cex = 0.5, col = rgb(0, 0, 0, 0.25))
```

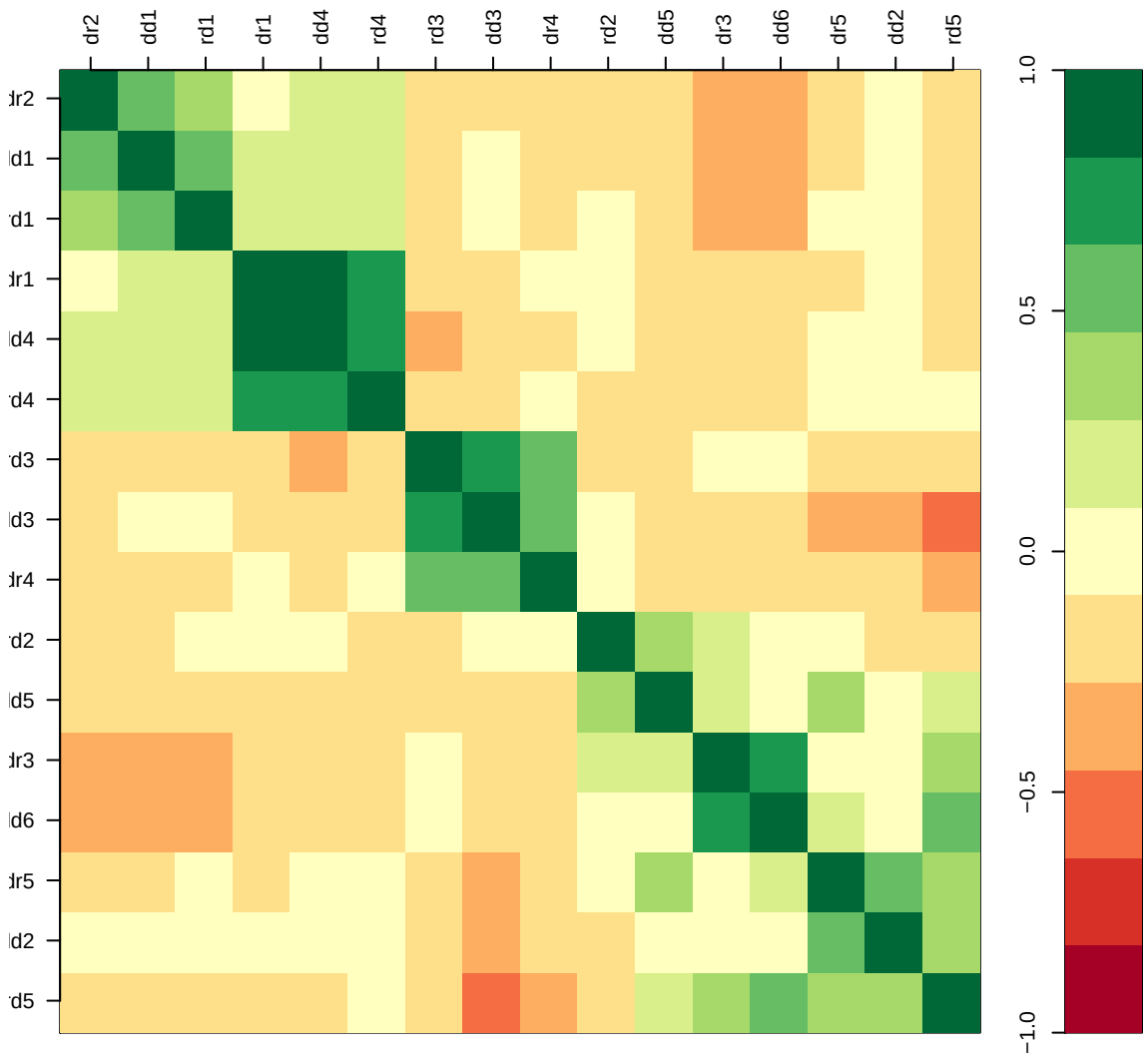



```
library(RColorBrewer)

corPlot = function(cors, zlim = c(-1, 1), pal = "RdYlGn", scores = NULL, ...) {
    clust = hclust(dist(cors))
    thepal = brewer.pal(brewer.pal.info[pal, ]$maxcolors, pal)

    z = cors[rev(clust$order), clust$order]
    if (!is.null(scores)) {
        scores = t(scores)
        scores = (scores - apply(scores, 1, min))/as.vector(diff(apply(scores,
            1, range)))
        scores = t(apply(scores, 1, function(x) {
            if (mean(x) < 0.5)
                x else 1 - x
        }))
        scores = scores * (zlim[2] - zlim[1]) + zlim[1]
        scores = t(scores)
```

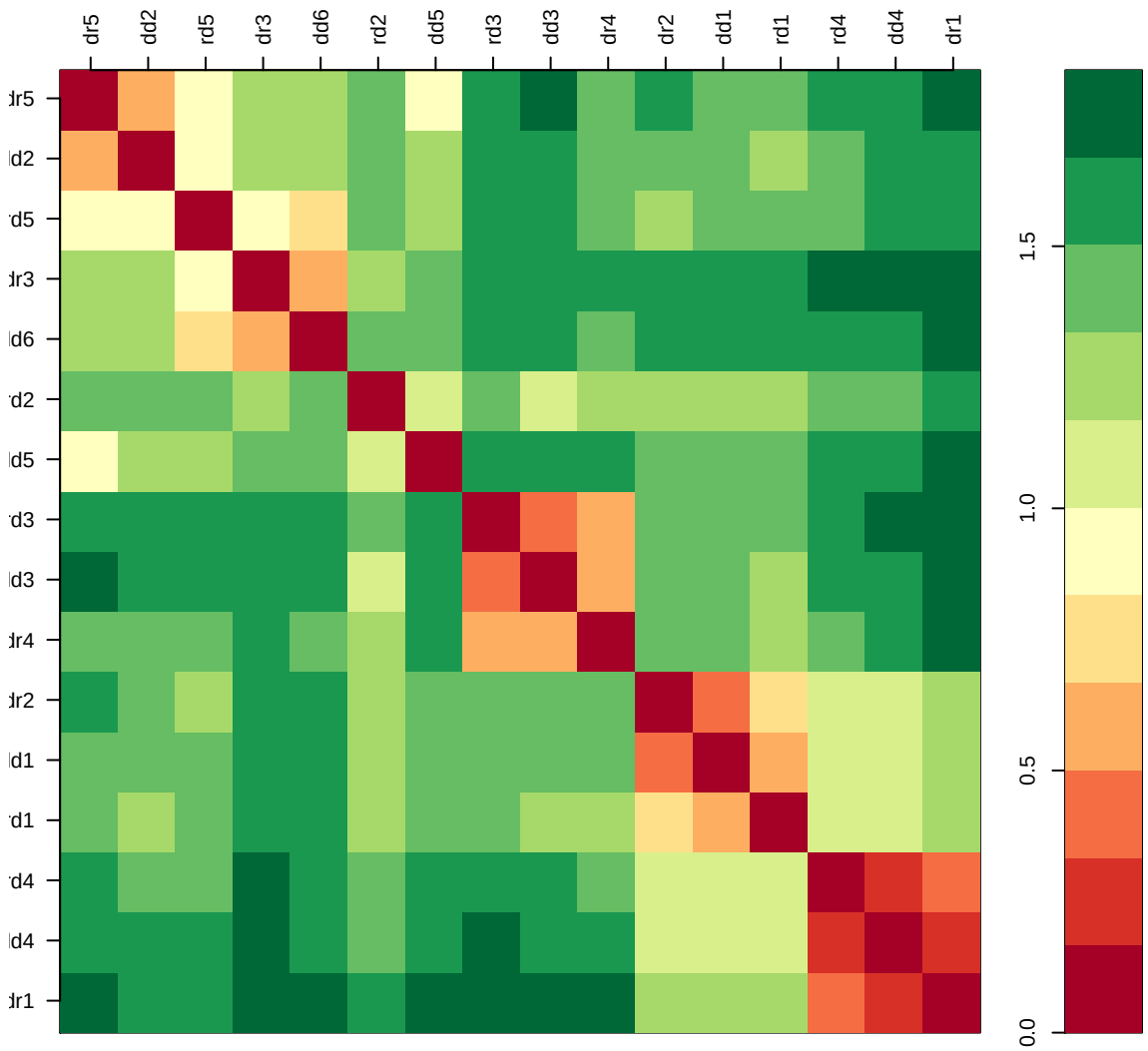
```
        scores = scores[, ncol(scores):1]
        z = cbind(z, scores[rev(clust$order), ])
    }

    pars = par(no.readonly = TRUE)
    par(mar = c(6, 3, 5, 3)/1.5)
    layout(matrix(c(1, 2), nrow = 1), widths = c(8, 1))
    image(z = z, zlim = zlim, col = thepal, xaxt = "n", yaxt = "n", ...)
    axis(3, (0:(nrow(z) - 1) + 0.5)/(nrow(z) - 1) - (0.5/nrow(z)), labels = rownames(z),
        las = 2)
    axis(2, (0:(ncol(z) - 1) + 0.5)/(ncol(z) - 1) - (0.5/ncol(z)), labels = colnames(z),
        las = 2)
    par(mar = c(6, 2, 5, 1)/1.5)
    image(x = c(0, 1), y = seq(zlim[1], zlim[2], length.out = 100), z = matrix(seq(zlim[1],
        zlim[2], length.out = 99), nrow = 1), col = thepal, xaxt = "n", xlab = "",
        ylab = "", useRaster = TRUE)
    par(pars)
}

general_corfun = function(x, ...) {
    cors = sapply(1:nrow(x), function(i1) {
        sapply(1:nrow(x), function(i2) {
            x1 = x[i1, ]
            x2 = x[i2, ]
            s = !is.na(x1) & !is.na(x2)
            if (sum(s) == 0) {
                return(0)
            }
            x1 = x1[s]
            x2 = x2[s]
            return(cor(x1, x2, ...))
        })
    })
    rownames(cors) = rownames(x)
    colnames(cors) = rownames(x)
    cors
}

corPlot(general_corfun(merged_coefs, method = "kendall"))
```



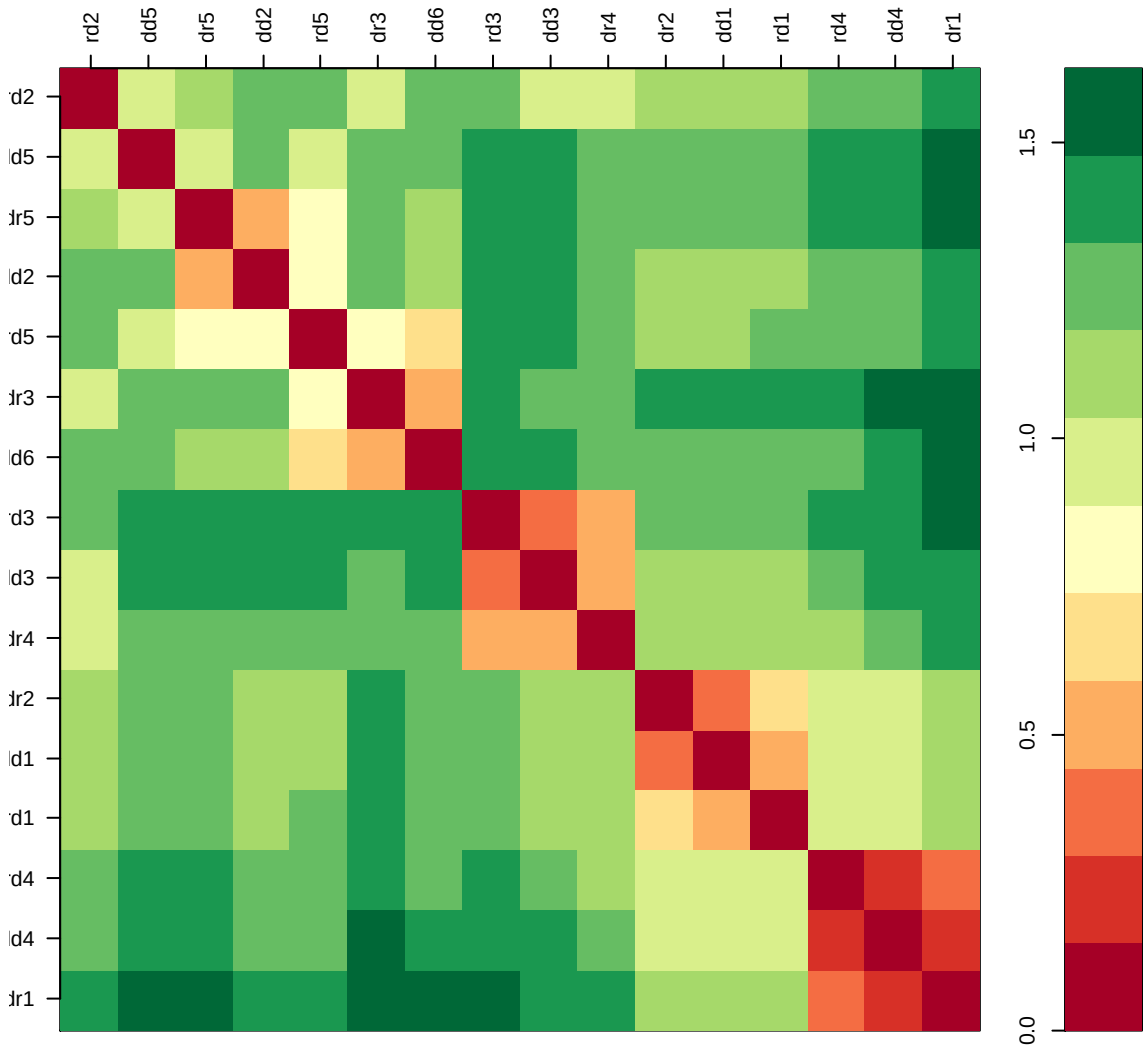

```
corPlot(general_corfun(merged_coefs_allpresent, method = "kendall"))
```




```
corPlot(as.matrix(dist(merged_coefs)), zlim = c(0, max(dist(merged_coefs))))
```




```
corPlot(as.matrix(dist(merged_coefs_allpresent)), zlim = c(0, max(dist(merged_coefs_allpresent))))
```



4 Session information

```
sessionInfo()
```

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

```
## [1] stats     graphics  grDevices utils     datasets  base
##
## other attached packages:
## [1] pvclust_1.3-0      gplots_2.14.2      RColorBrewer_1.0-5
## [4] knitr_1.8
##
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-6       caTools_1.17.1     digest_0.6.4
##  [4] evaluate_0.5.5     formatR_1.0        gdata_2.13.3
##  [7] gtools_3.4.1       highr_0.4          KernSmooth_2.23-13
## [10] stringr_0.6.2      tools_3.1.1
```