

Relationship of Extreme Chromosomal Instability with
Superior Long-term Survival in a Retrospective Analysis of
ER-Negative Primary Breast Cancer
Supplementary Document 1 - Data Analysis

Roylance et al.

January 31, 2011

This document describes the data analysis.
All analysis was performed in the R statistical environment.

1 Getting started

1.1 Required Packages

```
> library(fields)
> library(coxrobust)
> library(gregmisc)
```

```
gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
```

```
gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.
```

```
> library(beeswarm)
> library(geoR)
```

Analysis of geostatistical data

For an Introduction to geoR go to <http://www.leg.ufpr.br/geoR>
geoR version 1.6-32 (built on 2010-10--4) is now loaded

```
> library(multtest)
> library(survival)
> library(survplot)
> library(rmeta)
> library(affy)
> library(genefilter)
```

1.2 A few functions

```
> fun.deviationMod <- function(x) {
+   perc.mod2 <- c()
+   for (i in 1:ncol(x)) {
+     tmp <- which(x[, i] == max(x[, i], na.rm = T))[1]
+     perc.mod2 <- c(perc.mod2, sum(x[-tmp, i])/sum(x[, i]))
+   }
+   return(perc.mod2)
+ }
> fun.div <- function(x) {
+   div <- 0
+   for (i in which(x != 0)) div <- div + log(x[i]) * x[i]
+   return(-div)
+ }
> function.GNS.plot <- function(mnew1, fact, grade, node, size,
+   q, xnames, main) {
+   mat.grade <- matrix(NA, nr = 3, nc = 4)
+   mat.grade[1, 1] <- length(which(grade == 1 & mnew1 >= q[1] &
+     mnew1 < q[2]))
+   mat.grade[2, 1] <- length(which(grade == 2 & mnew1 >= q[1] &
+     mnew1 < q[2]))
+   mat.grade[3, 1] <- length(which(grade == 3 & mnew1 >= q[1] &
+     mnew1 < q[2]))
+   mat.grade[1, 2] <- length(which(grade == 1 & mnew1 >= q[2] &
+     mnew1 < q[3]))
+   mat.grade[2, 2] <- length(which(grade == 2 & mnew1 >= q[2] &
```

```

+     mnew1 < q[3]))
+ mat.grade[3, 2] <- length(which(grade == 3 & mnew1 >= q[2] &
+     mnew1 < q[3]))
+ mat.grade[1, 3] <- length(which(grade == 1 & mnew1 >= q[3] &
+     mnew1 < q[4]))
+ mat.grade[2, 3] <- length(which(grade == 2 & mnew1 >= q[3] &
+     mnew1 < q[4]))
+ mat.grade[3, 3] <- length(which(grade == 3 & mnew1 >= q[3] &
+     mnew1 < q[4]))
+ mat.grade[1, 4] <- length(which(grade == 1 & mnew1 >= q[4] &
+     mnew1 <= q[5]))
+ mat.grade[2, 4] <- length(which(grade == 2 & mnew1 >= q[4] &
+     mnew1 <= q[5]))
+ mat.grade[3, 4] <- length(which(grade == 3 & mnew1 >= q[4] &
+     mnew1 <= q[5]))
+ colnames(mat.grade) <- c("[0.0, 0.15)", "[0.15, 0.3)", "[0.3, 0.45)",
+     "[0.45, 0.62]")
+ rownames(mat.grade) <- c("Grade1", "Grade2", "Grade3")
+ mat.node <- matrix(NA, nr = 2, nc = 4)
+ mat.node[1, 1] <- length(which(node == "yes" & mnew1 >= q[1] &
+     mnew1 < q[2]))
+ mat.node[2, 1] <- length(which(node == "no" & mnew1 >= q[1] &
+     mnew1 < q[2]))
+ mat.node[1, 2] <- length(which(node == "yes" & mnew1 >= q[2] &
+     mnew1 < q[3]))
+ mat.node[2, 2] <- length(which(node == "no" & mnew1 >= q[2] &
+     mnew1 < q[3]))
+ mat.node[1, 3] <- length(which(node == "yes" & mnew1 >= q[3] &
+     mnew1 < q[4]))
+ mat.node[2, 3] <- length(which(node == "no" & mnew1 >= q[3] &
+     mnew1 < q[4]))
+ mat.node[1, 4] <- length(which(node == "yes" & mnew1 >= q[4] &
+     mnew1 <= q[5]))
+ mat.node[2, 4] <- length(which(node == "no" & mnew1 >= q[4] &
+     mnew1 <= q[5]))
+ mat.size <- matrix(NA, nr = 3, nc = 4)
+ mat.size[1, 1] <- length(which(size.fact == "0 - 2cm" & mnew1 >=
+     q[1] & mnew1 < q[2]))
+ mat.size[2, 1] <- length(which(size.fact == "2 - 5cm" & mnew1 >=
+     q[1] & mnew1 < q[2]))
+ mat.size[3, 1] <- length(which(size.fact == "5 - 8cm" & mnew1 >=
+     q[1] & mnew1 < q[2]))
+ mat.size[1, 2] <- length(which(size.fact == "0 - 2cm" & mnew1 >=
+     q[2] & mnew1 < q[3]))
+ mat.size[2, 2] <- length(which(size.fact == "2 - 5cm" & mnew1 >=
+     q[2] & mnew1 < q[3]))
+ mat.size[3, 2] <- length(which(size.fact == "5 - 8cm" & mnew1 >=
+     q[2] & mnew1 < q[3]))
+ mat.size[1, 3] <- length(which(size.fact == "0 - 2cm" & mnew1 >=
+     q[3] & mnew1 < q[4]))
+ mat.size[2, 3] <- length(which(size.fact == "2 - 5cm" & mnew1 >=
+     q[3] & mnew1 < q[4]))
+ mat.size[3, 3] <- length(which(size.fact == "5 - 8cm" & mnew1 >=
+     q[3] & mnew1 < q[4]))

```

```

+   mat.size[1, 4] <- length(which(size.fact == "0 - 2cm" & mnew1 >=
+     q[4] & mnew1 <= q[5]))
+   mat.size[2, 4] <- length(which(size.fact == "2 - 5cm" & mnew1 >=
+     q[4] & mnew1 <= q[5]))
+   mat.size[3, 4] <- length(which(size.fact == "5 - 8cm" & mnew1 >=
+     q[4] & mnew1 <= q[5]))
+   colnames(mat.size) <- c("[0.0, 0.15]", "[0.15, 0.3]", "[0.3, 0.45]",
+     "[0.45, 0.62]")
+   rownames(mat.size) <- c("0 - 2cm", "2 - 5cm", "5 - 8cm")
+   colnames(mat.node) <- c("[0.0, 0.15]", "[0.15, 0.3]", "[0.3, 0.45]",
+     "[0.45, 0.62]")
+   rownames(mat.node) <- c("node+", "node-")
+   print(fisher.test(matrix(c(mat.node[, 1] + mat.node[, 2],
+     mat.node[, 3] + mat.node[, 4]), 2, 2)))
+   mat.tmp <- rbind(t(t(mat.node)/colSums(mat.node)), rep(0,
+     4))
+   mat.tmp <- cbind(t(t(mat.size)/colSums(mat.size)), mat.tmp,
+     t(t(mat.grade)/colSums(mat.grade)))
+   b <- barplot(mat.tmp, main = main, col = c("darkred", "red",
+     "palevioletred3"), ylim = c(0, 1.5), ylab = NA, xaxt = "n",
+     yaxt = "n", space = c(rep(0.2, 4), 1, rep(0.2, 3), 1,
+     rep(0.2, 3)))
+   axis(1, at = b, labels = F)
+   text(b - 0.6, par("usr")[3] - 0.09, labels = rep(xnames,
+     4), srt = 45, pos = 1, xpd = TRUE, cex = 0.7)
+   axis(2, at = seq(0, 1, 0.2), paste(seq(0, 100, 20), "%",
+     sep = ""), las = 2, cex.axis = 0.7, mgp = c(3, 0.6, 0))
+   legend(0, 1, c("0 - 2", "2 - 5", "5 - 8"), pch = rep(15,
+     3), col = c("darkred", "red", "palevioletred3"), yjust = 0,
+     bty = "n", cex = 0.7, horiz = F, title = "Size (cm)")
+   legend(5.7, 1.087, c("positive", "negative"), pch = rep(15,
+     2), col = c("darkred", "red"), yjust = 0, bty = "n",
+     cex = 0.7, horiz = F, title = "Node")
+   legend(11.2, 1, c("G1", "G2", "G3"), pch = rep(15, 2), col = c("darkred",
+     "red", "palevioletred3"), yjust = 0, bty = "n", cex = 0.7,
+     horiz = F, title = "Grade")
+ }
> fun.Image <- function(chrom15.tmp, chrom2.tmp, data.ER.sort,
+   data.HER2.sort) {
+   image(t(sweep(chrom15.tmp[, order(mean15and2)], MARGIN = 2,
+     colSums(chrom15.tmp, na.rm = T), FUN = "/"))[, -c(1,
+     7, 8)], xlim = c(0, 3), ylim = c(0, 2.5), axes = F, col = "white",
+     xaxs = "r", yaxs = "r")
+   text(rep(0.1, 5), c(1.05, 0.84, 0.63, 0.42, 0.21) - 0.21,
+     5:1)
+   text(rep(0.1, 5), c(2.14, 1.93, 1.72, 1.51, 1.3), 5:1)
+   add.image(0.15, 1.095, as.matrix(data.ER.sort), col = c("red",
+     "black", "white"), image.width = 0.032, image.height = 0.8,
+     adj.x = 0, adj.y = 0)
+   add.image(0.15, 0.98, as.matrix(data.HER2.sort), col = c("red",
+     "black", "white"), image.width = 0.032, image.height = 0.8,
+     adj.x = 0, adj.y = 0)
+   add.image(0.15, -0.1, t(sweep(chrom15.tmp[, order(mean15and2)],
+     MARGIN = 2, colSums(chrom15.tmp, na.rm = T), FUN = "/"))[,

```

```

+       -c(1, 7, 8)], image.width = 0.39, image.height = 0.8,
+       adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+       1))
+   add.image(0.15, 1.21, t(sweep(chrom2.tmp[, order(mean15and2)],
+     MARGIN = 2, colSums(chrom2.tmp, na.rm = T), FUN = "/"))[,
+     -c(1, 7, 8, 9)], image.width = 0.39, image.height = 0.8,
+     adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+     1))
+   text(2.4, 1.75, "Chromosome 2", srt = 270)
+   text(2.4, 0.42, "Chromosome 15", srt = 270)
+   text(0.49, 2.35, "1")
+   text(1.17, 2.35, "2")
+   text(1.77, 2.35, "3")
+   text(2.18, 2.35, "4")
+   text(2.53, 1.14, "ER+/-", cex = 0.9)
+   text(2.61, 1.02, "HER2+/-", cex = 0.9)
+   text(1.3, 2.5, "MCD Cohorts")
+   lines(rep(0.15, 10), seq(-0.1, 2.4, length = 10), lwd = 1,
+     col = "black")
+   lines(rep(0.83, 10), seq(-0.1, 2.4, length = 10), lwd = 1,
+     col = "black")
+   lines(rep(1.51, 10), seq(-0.1, 2.4, length = 10), lwd = 1,
+     col = "black")
+   lines(rep(2.03, 10), seq(-0.1, 2.4, length = 10), lwd = 1,
+     col = "black")
+   lines(rep(2.33, 10), seq(-0.1, 2.4, length = 10), lwd = 1,
+     col = "black")
+   legend.krige(x.leg = c(2.55, 2.65), y.leg = c(0.6, 1.6) -
+     0.77, seq(0, 1, length = 100), vert = TRUE, off = 1.3,
+     cex = 1, col = bluered(70), scale.vals = c(0.2, 0.4,
+     0.6, 0.8))
+ }
> fun.Image2 <- function(chrom15.tmp, chrom2.tmp, HER2.neg, HER2.pos,
+   m.neg, m.pos) {
+   image(t(t(sweep(chrom15.tmp[, which(data.tmp2[, "ER.status.corr"] ==
+     "negative")], MARGIN = 2, colSums(chrom15.tmp[, which(data.tmp2[,
+     "ER.status.corr"] == "negative")], na.rm = T), FUN = "/"))[order(m.neg,
+     decreasing = T), -c(1, 7, 8)]), axes = F, col = "white",
+     main = NA, ylim = c(0, 3), xlim = c(0, 2), zlim = c(0,
+     1))
+   add.image(0.153, 0.15, t(t(sweep(chrom15.tmp[, which(data.tmp2[,
+     "ER.status.corr"] == "negative")], MARGIN = 2, colSums(chrom15.tmp[,
+     which(data.tmp2[, "ER.status.corr"] == "negative")],
+     na.rm = T), FUN = "/"))[order(m.neg, decreasing = T),
+     -c(1, 7, 8)]), image.width = 0.2, image.height = 0.4,
+     adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+     1))
+   add.image(0.88, 0.15, t(t(sweep(chrom2.tmp[, which(data.tmp2[,
+     "ER.status.corr"] == "negative")], MARGIN = 2, colSums(chrom2.tmp[,
+     which(data.tmp2[, "ER.status.corr"] == "negative")],
+     na.rm = T), FUN = "/"))[order(m.neg, decreasing = T),
+     -c(1, 7, 8, 9)]), image.width = 0.2, image.height = 0.4,
+     adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+     1))

```

```

+   add.image(0.835, 0.15, t(as.matrix(HER2.neg)), col = c("red",
+     "black"), image.width = 0.2, image.height = 0.02, adj.x = 0,
+     adj.y = 0)
+   add.image(0.153, 0.8, t(t(sweep(chrom15.tmp[, which(data.tmp2[,
+     "ER.status.corr"] == "positive")], MARGIN = 2, colSums(chrom15.tmp[,
+     which(data.tmp2[, "ER.status.corr"] == "positive")],
+     na.rm = T), FUN = "/"))[order(m.pos, decreasing = T),
+     -c(1, 7, 8)]), image.width = 0.7, image.height = 0.4,
+     adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+     1))
+   add.image(0.835, 0.8, t(as.matrix(HER2.pos)), col = c("red",
+     "black"), image.width = 0.7, image.height = 0.02, adj.x = 0,
+     adj.y = 0)
+   add.image(0.88, 0.8, t(t(sweep(chrom2.tmp[, which(data.tmp2[,
+     "ER.status.corr"] == "positive")], MARGIN = 2, colSums(chrom2.tmp[,
+     which(data.tmp2[, "ER.status.corr"] == "positive")],
+     na.rm = T), FUN = "/"))[order(m.pos, decreasing = T),
+     -c(1, 7, 8, 9)]), image.width = 0.7, image.height = 0.4,
+     adj.x = 0, adj.y = 0, col = bluered(70), zlim = c(0,
+     1))
+   lines(seq(0.153, 1.65, length = 100), rep(0.15, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(0.276, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(0.47, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(0.605, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(0.75, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(0.8, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(1.04, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(1.5, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(2.21, 100), lwd = 1)
+   lines(seq(0.153, 1.65, length = 100), rep(2.9, 100), lwd = 1)
+   text(0.08, 0.45, "ER-", srt = 270)
+   text(0.08, 1.95, "ER+", srt = 270)
+   text(1.7, 1.5, "MCD Cohorts", srt = 270)
+   text(1.6, 0.21, "4", srt = 270)
+   text(1.6, 0.38, "3", srt = 270)
+   text(1.6, 0.535, "2", srt = 270)
+   text(1.6, 0.68, "1", srt = 270)
+   text(1.6, 0.92, "4", srt = 270)
+   text(1.6, 1.27, "3", srt = 270)
+   text(1.6, 1.85, "2", srt = 270)
+   text(1.6, 2.55, "1", srt = 270)
+   text(0.45 + 0.85 - 0.085, 0.08, "Chromosome 2")
+   text(0.48, 0.08, "Chromosome 15")
+   text(c(0.165, 0.3, 0.435, 0.57, 0.705) + 0.066, rep(2.95,
+     5), 1:5, srt = 270)
+   text(c(0.165, 0.3, 0.435, 0.57, 0.705) + 0.88 - 0.1, rep(2.95,
+     5), 1:5, srt = 270)
+   legend.krige(x.leg = c(1.7, 1.75), y.leg = c(0.75, 1.65) +
+     1.3, seq(0, 1, length = 100), vert = TRUE, off = 1.4,
+     cex = 1, col = bluered(70))
+ }
> no.tel <- function(seg.out, chrominfo, min.size = 1e+06, min.probes = 1,
+   max.size = 1e+09) {
+   if (class(seg.out) == "DNAcopy") {

```

```

+     seg.out <- seg.out$output
+   }
+   tmp.segs <- seg.out
+   tmp.segs <- tmp.segs[!tmp.segs[, 5] < min.probes, ]
+   tmp.segs[, 2] <- as.character(tmp.segs[, 2])
+   if (nrow(tmp.segs[tmp.segs[, 2] == "X", ]) > 0) {
+     tmp.segs[tmp.segs[, 2] == "X", 2] <- rep(23, nrow(tmp.segs[tmp.segs[,
+     2] == "X", ]))
+   }
+   seg.out <- seg.out[!seg.out[, 2] %in% c("MT", "Y", "24"),
+   ]
+   tmp.segs[tmp.segs[, 6] == 1, 6] <- 2
+   for (j in 1:length(unique(tmp.segs[, 1]))) {
+     tmp.sample <- tmp.segs[tmp.segs[, 1] == unique(tmp.segs[,
+     1])[j], ]
+     for (i in 1:23) {
+       tmp1 <- tmp.sample[tmp.sample[, 2] == i, , drop = F]
+       if (tmp1[1, 6] == 2 & nrow(tmp1) != 1 & tmp1[1, 4] <
+       (chrominfo[i, 3] * 1000)) {
+         tmp.sample[tmp.sample[, 2] == i, 6][1] <- 1
+       }
+       if (tmp1[nrow(tmp1), 6] == 2 & nrow(tmp1) != 1 &
+       tmp1[nrow(tmp1), 3] > (chrominfo[i, 3] * 1000)) {
+         tmp.sample[tmp.sample[, 2] == i, 6][nrow(tmp.sample[tmp.sample[,
+         2] == i, ])] <- 1
+       }
+       if (nrow(tmp.sample[tmp.sample[, 2] == i, ]) == 1 &
+       tmp.sample[tmp.sample[, 2] == i, 6][1] != 0) {
+         tmp.sample[tmp.sample[, 2] == i, 6][1] <- 3
+       }
+     }
+     tmp.segs[tmp.segs[, 1] == unique(tmp.segs[, 1])[j], ] <- tmp.sample
+   }
+   no.events <- matrix(0, nrow = length(unique(tmp.segs[, 1])),
+   ncol = 28)
+   rownames(no.events) <- unique(tmp.segs[, 1])
+   colnames(no.events) <- c("Telomeric AI", "Mean size", "Interstitial AI",
+   "Mean Size", "Wholo chr AI", 1:23)
+   for (i in unique(tmp.segs[, 1])) {
+     tmp <- tmp.segs[tmp.segs[, 1] == i, ]
+     tmp <- tmp[(tmp[, 4] - tmp[, 3]) > min.size, ]
+     tmp <- tmp[(tmp[, 4] - tmp[, 3]) < max.size, ]
+     no.events[i, 1] <- nrow(tmp[tmp[, 6] == 1, ])
+     no.events[i, 2] <- mean(tmp[tmp[, 6] == 1, 4] - tmp[tmp[,
+     6] == 1, 3])
+     no.events[i, 3] <- nrow(tmp[tmp[, 6] == 2, ])
+     no.events[i, 4] <- mean(tmp[tmp[, 6] == 2, 4] - tmp[tmp[,
+     6] == 2, 3])
+     no.events[i, 5] <- nrow(tmp[tmp[, 6] == 3, ])
+     no.events[i, tmp[tmp[, 6] == 3, 2]] <- 1
+   }
+   return(no.events)
+ }
> gii.fn <- function(seg.out, loss = -0.3, gain = 0.3, seg.size = 0,

```

```

+   min.length = 3) {
+   if (class(seg.out) == "DNACopy") {
+     seg.out <- seg.out$output
+   }
+   seg.out[, 2] <- as.character(seg.out[, 2])
+   if (nrow(seg.out[seg.out[, 2] == "X", ]) > 0) {
+     seg.out[seg.out[, 2] == "X", 2] <- rep(23, nrow(seg.out[seg.out[,
+     2] == "X", ]))
+   }
+   seg.out <- seg.out[!seg.out[, 2] %in% c("MT", "Y", "24"),
+   ]
+   if (length(gain) == 1) {
+     gain <- rep(gain, length(unique(seg.out[, 1])))
+   }
+   if (length(loss) == 1) {
+     loss <- rep(loss, length(unique(seg.out[, 1])))
+   }
+   stopifnot(length(gain) == length(unique(seg.out[, 1])))
+   stopifnot(length(loss) == length(unique(seg.out[, 1])))
+   seg.out <- seg.out[!(seg.out[, 5] <= min.length), ]
+   GII <- matrix(NA, nrow = length(unique(seg.out[, 1])), ncol = 7)
+   rownames(GII) <- unique(seg.out[, 1])
+   colnames(GII) <- c("GII", "No. breakpoints", "Mean Length",
+   "No. gain", "mean gain length", "No. loss", "Mean loss length")
+   for (i in 1:nrow(GII)) {
+     a <- seg.out[seg.out[, 1] == rownames(GII)[i], ]
+     b <- sum(a[, 4] - a[, 3])
+     a <- a[a[, 6] >= gain[i] | a[, 6] <= loss[i], ]
+     GII[i, 1] <- sum(a[, 4] - a[, 3])/b
+     GII[i, 2] <- nrow(a)
+     GII[i, 3] <- mean(a[, 4] - a[, 3])
+     b <- a[a[, 6] >= gain[i], ]
+     GII[i, 4] <- nrow(b)
+     GII[i, 5] <- mean(b[, 4] - b[, 3])
+     b <- a[a[, 6] <= loss[i], ]
+     GII[i, 6] <- nrow(b)
+     GII[i, 7] <- mean(b[, 4] - b[, 3])
+   }
+   return(GII)
+ }
> library(genefilter)
> loadIntoList <- function(filenames, pre = "", post = "") {
+   filenames <- paste(pre, filenames, post, sep = "")
+   e <- new.env()
+   for (f in filenames) {
+     load(f, envir = e)
+   }
+   rev(as.list(e))
+ }
> var.probes <- function(genes, data.rma) {
+   genes <- genes[names(genes) %in% rownames(exprs(data.rma))]
+   tmp <- rowSds(exprs(data.rma)[names(genes), ])
+   genes <- genes[order(tmp, decreasing = T)]
+   genes[!duplicated(genes)]

```



```
+ }
```

1.3 Data Preparation

```
> load("/media/MyBook/LeedsNew13112010/NicolaiData/all.er.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/all.pam.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/all.cin.scale.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/all.surv.rfs.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/datasets.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/gastric.cin.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/gastric.surv.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/lung.s.cin.scale.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/lung.surv.cens.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/surv.rfs.vec.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/tcga.duke.dressman.cin70.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/tcga.duke.dressman.surv.cens.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/zwang.cn.seg.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/zwang.mcp.seg.RData")
> load("/media/MyBook/LeedsNew13112010/NicolaiData/chrominfo.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/tcga.duke.dressman.cin70.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/lung.surv.cens.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/lung.s.cin.scale.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/gastric.surv.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/gastric.cin.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/tcga.duke.dressman.surv.cens.RData")
> load("/home/endesf01/genesymb.RData")
> load("/home/endesf01/windows2/LeedsNew13112010/NicolaiMetaplotData/sigs_SC_ACE-hg17.RData")
> load("/home/endesf01/dataFISHchromosome15.Rdata")
> load("/home/endesf01/LeedsdataFISHchromosome2.Rdata")
> load("/home/endesf01/data.RData")
> load("/home/endesf01/treatmentData.Rdata")
> load("/home/endesf01/Chromosome2counts.Rdata")
> load("/home/endesf01/Chromosome15counts.Rdata")
> mm2 <- match(colnames(chrom15.tmp), rownames(data))
> data.tmp2 <- data[mm2, ]
> data.tmp2 <- as.matrix(data.tmp2)
> s.15 <- chrom15.tmp[-c(1, 2), ]
> s.2 <- chrom2.tmp[-c(1, 2), ]
> dev2.15 <- fun.deviationMod(s.15)
> dev2.2 <- fun.deviationMod(s.2)
> dev2.15[is.na(dev2.15)] <- 0
> mean15and2 <- (dev2.15 + dev2.2)/2
> fact <- rep(NA, length(mean15and2))
> fact[mean15and2 >= 0 & mean15and2 < 0.15] <- "MCD1"
> fact[mean15and2 >= 0.15 & mean15and2 < 0.3] <- "MCD2"
> fact[mean15and2 >= 0.3 & mean15and2 < 0.45] <- "MCD3"
> fact[mean15and2 >= 0.45 & mean15and2 <= max(mean15and2)] <- "MCD4"
> perc.2 <- sweep(chrom2.tmp, MARGIN = 2, FUN = "/", colSums(chrom2.tmp,
+   na.rm = T))
> perc.15 <- sweep(chrom15.tmp, MARGIN = 2, FUN = "/", colSums(chrom15.tmp,
+   na.rm = T))
> div.2 <- apply(perc.2, 2, fun.div)
> div.15 <- apply(perc.15, 2, fun.div)
> mat.Subtypes <- matrix(NA, nr = 3, nc = 4)
```

```

> colnames(mat.Subtypes) <- c("[0.0, 0.15)", "[0.15, 0.3)", "[0.3, 0.45)",
+   "[0.45, 0.62]")
> rownames(mat.Subtypes) <- c("ER+", "ER-", "HER2+")
> mat.Subtypes[1, 1] <- length(which(mean15and2 >= 0 & mean15and2 <
+   0.15 & data.tmp2[, "ER.status.corr"] == "positive"))
> mat.Subtypes[1, 2] <- length(which(mean15and2 >= 0.15 & mean15and2 <
+   0.3 & data.tmp2[, "ER.status.corr"] == "positive"))
> mat.Subtypes[1, 3] <- length(which(mean15and2 >= 0.3 & mean15and2 <
+   0.45 & data.tmp2[, "ER.status.corr"] == "positive"))
> mat.Subtypes[1, 4] <- length(which(mean15and2 >= 0.45 & mean15and2 <=
+   max(mean15and2) & data.tmp2[, "ER.status.corr"] == "positive"))
> mat.Subtypes[2, 1] <- length(which(mean15and2 >= 0 & mean15and2 <
+   0.15 & data.tmp2[, "ER.status.corr"] == "negative"))
> mat.Subtypes[2, 2] <- length(which(mean15and2 >= 0.15 & mean15and2 <
+   0.3 & data.tmp2[, "ER.status.corr"] == "negative"))
> mat.Subtypes[2, 3] <- length(which(mean15and2 >= 0.3 & mean15and2 <
+   0.45 & data.tmp2[, "ER.status.corr"] == "negative"))
> mat.Subtypes[2, 4] <- length(which(mean15and2 >= 0.45 & mean15and2 <=
+   max(mean15and2) & data.tmp2[, "ER.status.corr"] == "negative"))
> mat.Subtypes[3, 1] <- length(which(mean15and2 >= 0 & mean15and2 <
+   0.15 & data.tmp2[, "HER2.status"] == "positive"))
> mat.Subtypes[3, 2] <- length(which(mean15and2 >= 0.15 & mean15and2 <
+   0.3 & data.tmp2[, "HER2.status"] == "positive"))
> mat.Subtypes[3, 3] <- length(which(mean15and2 >= 0.3 & mean15and2 <
+   0.45 & data.tmp2[, "HER2.status"] == "positive"))
> mat.Subtypes[3, 4] <- length(which(mean15and2 >= 0.45 & mean15and2 <=
+   max(mean15and2) & data.tmp2[, "HER2.status"] == "positive"))

```

1.4 Figures

Figure 1 B

```

> hist(mean15and2, breaks = 70, xlab = "Mean Percentage Modal \n Centromere Deviation",
+   main = NA, ylim = c(0, 13), col = "lightblue", xaxt = "n",
+   yaxt = "n", mgp = c(2.4, 1, 0), ylab = NA)
> text(c(0.075, 0.225, 0.375, 0.5375), par("usr")[3] + 13, labels = c("MCD1",
+   "MCD2", "MCD3", "MCD4"))
> abline(v = c(0, 0.15, 0.3, 0.45, max(mean15and2) + 0.005), col = "red")
> axis(1, at = c(0, 0.15, 0.3, 0.45, max(mean15and2) + 0.005),
+   labels = c("0%", "15%", "30%", "45%", "62%"), mgp = c(3,
+   0.5, 0))
> axis(2, at = c(0, 2, 4, 6, 8, 10, 12), c(0, 2, 4, 6, 8, 10, 12),
+   las = 2, mgp = c(3, 0.6, 0))
> mtext(side = 2, text = "Frequency", line = 1.6)

```

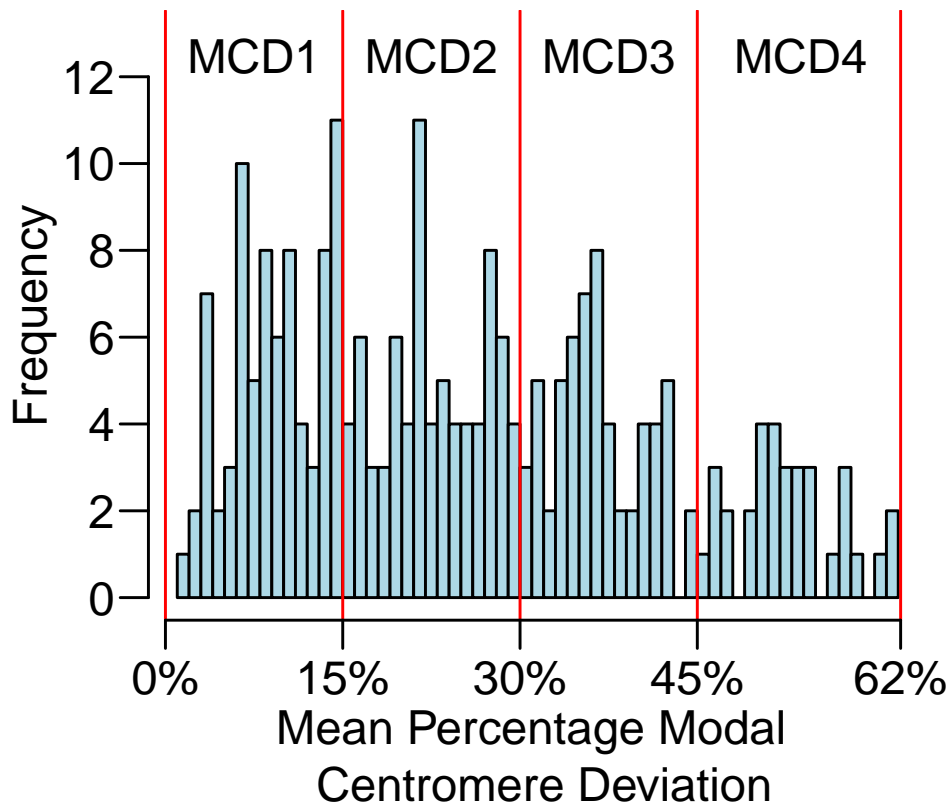


Figure 1 C

```

> data.sort <- data.tmp2[order(mean15and2), ]
> data.ER.sort <- rep(NA, nrow(data.sort))
> data.ER.sort[which(data.sort[, "ER.status.corr"] == "negative")] <- 1
> data.ER.sort[which(data.sort[, "ER.status.corr"] == "positive")] <- 2
> data.ER.sort[which(is.na(data.sort[, "ER.status.corr"]))] <- 3
> data.HER2.sort <- rep(NA, nrow(data.sort))
> data.HER2.sort[which(data.sort[, "HER2.status"] == "negative")] <- 1
> data.HER2.sort[which(data.sort[, "HER2.status"] == "positive")] <- 2
> data.HER2.sort[which(is.na(data.sort[, "HER2.status"]))] <- 3
> fun.Image(chrom15.tmp, chrom2.tmp, data.ER.sort, data.HER2.sort)

```

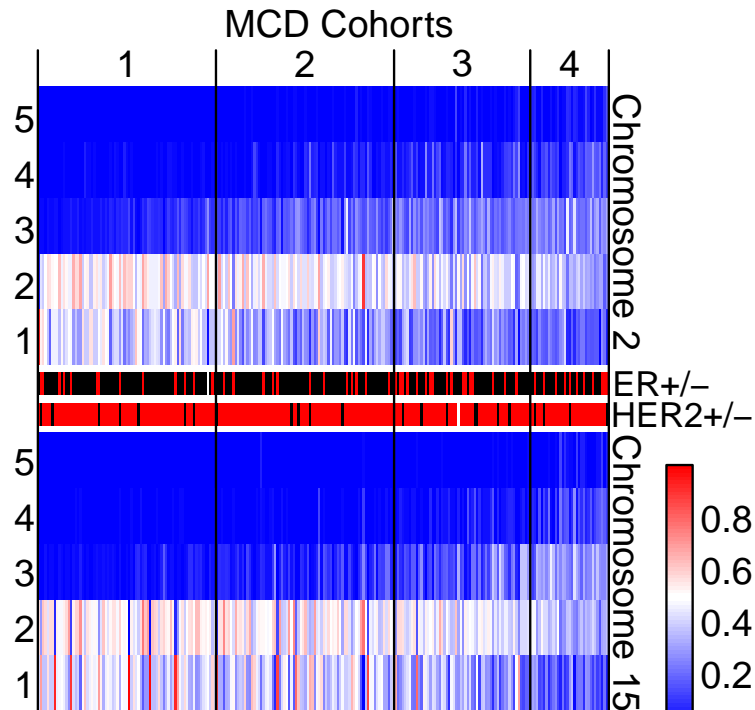


Figure 2 A

```

> plot(dev2.2, div.2, main = NA, ylab = NA, xlab = "Percentage Modal Centromere \n Deviation (Ch
+   xaxt = "n", las = 2, mgp = c(2.5, 0.6, 0))
> axis(1, at = c(0, 0.15, 0.3, 0.45, 0.62), labels = c("0%", "15%",
+   "30%", "45%", "62%"), mgp = c(3, 0.5, 0))
> abline(lm(div.2 ~ dev2.2)$coefficients, col = "red")
> legend("bottomright", "r = 0.941", col = "red", lty = 1, bty = "n")
> mtext(side = 2, text = "Diversity Index", line = 1.8)
> print(cor(dev2.2, div.2))

[1] 0.9410414

> print(cor.test(dev2.2, div.2))

Pearson's product-moment correlation

data: dev2.2 and div.2
t = 43.4523, df = 244, p-value < 2.2e-16

```

```

alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9248184 0.9538476
sample estimates:
      cor
0.9410414

```

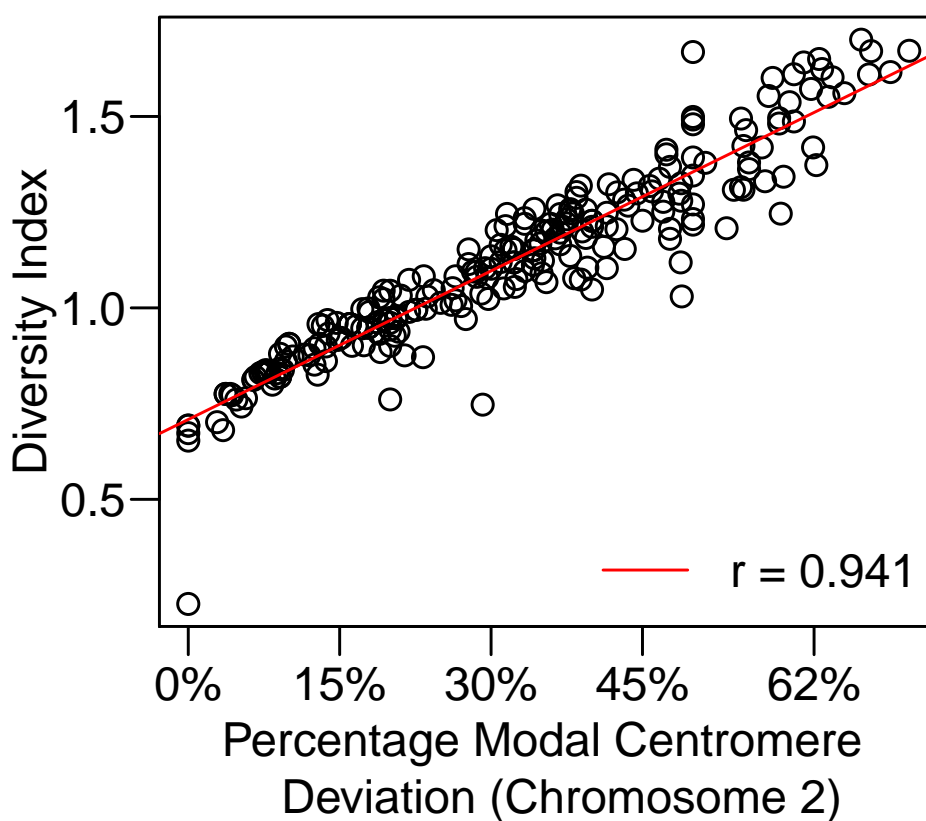


Figure 2 B

```

> plot(dev2.15, div.15, main = NA, ylab = NA, xlab = "Percentage Modal Centromere \n Deviation (Chromosome 2)",
+       xaxt = "n", las = 2, mgp = c(2.5, 0.6, 0))
> axis(1, at = c(0, 0.15, 0.3, 0.45, 0.62), labels = c("0%", "15%",
+ "30%", "45%", "62%"), mgp = c(3, 0.5, 0))
> abline(lm(div.15 ~ dev2.15)$coefficients, col = "red")
> legend("bottomright", "r = 0.902", col = "red", lty = 1, bty = "n")
> mtext(side = 2, text = "Diversity Index", line = 1.8)
> print(cor(dev2.15, div.15))

```

```
[1] 0.9019995
```

```
> print(cor.test(dev2.15, div.15))
```

Pearson's product-moment correlation

data: dev2.15 and div.15

t = 32.6349, df = 244, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.8757216 0.9229493

sample estimates:

cor

0.9019995

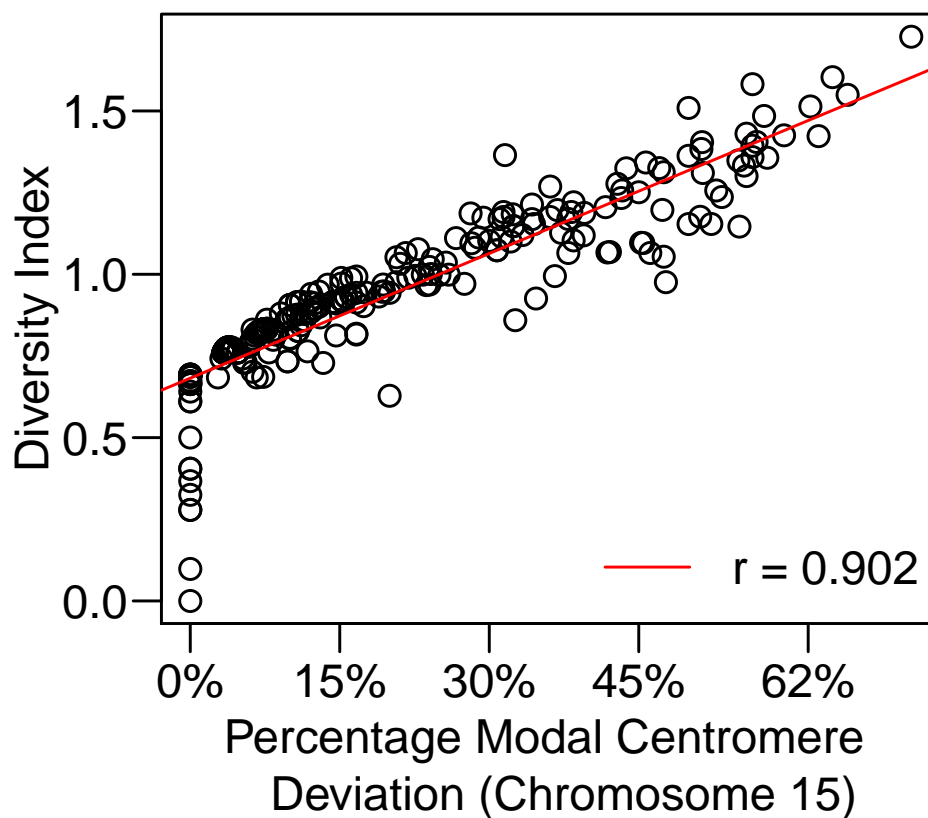


Figure 2 C

```
> boxplot((div.2 + div.15)/2 ~ fact, ylab = NA, xaxt = "n", las = 2,  
+         outline = F, yaxt = "n")  
> axis(2, c(0.8, 1, 1.2, 1.4), as.character(c(0.8, 1, 1.2, 1.4)),  
+      las = 2, mgp = c(3, 0.6, 0))
```

```

> mtext(side = 2, text = "Mean Diversity Index", line = 2)
> beeswarm((div.2 + div.15)/2 ~ fact, add = T, cex = 0.5)
> axis(1, at = c(1, 2, 3, 4), labels = F)
> text(c(1, 2, 3, 4) - 0.245, par("usr")[3] - 0.09, labels = c("MCD1",
+ "MCD2", "MCD3", "MCD4"), srt = 45, pos = 1, xpd = TRUE)
> print(t.test((div.2 + div.15)/2 ~ fact == "MCD4"))

```

Welch Two Sample t-test

```

data: (div.2 + div.15)/2 by fact == "MCD4"
t = -18.1313, df = 56.506, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4489748 -0.3596511
sample estimates:
mean in group FALSE mean in group TRUE
 0.9728719          1.3771848

```

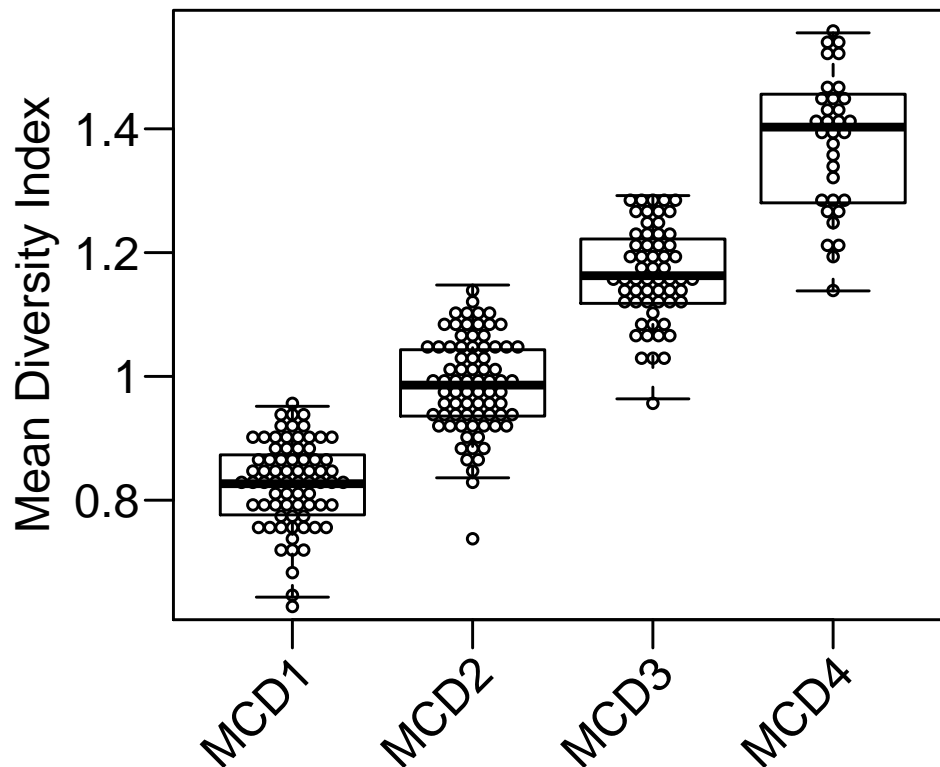


Figure 3 A

```

> lablist <- c("MCD1", "MCD2", "MCD3", "MCD4")
> b = barplot(mat.Subtypes/rowSums(mat.Subtypes), beside = T, col = c("blue",
+   "red", "magenta"), yaxt = "n", ylab = NA, xlab = NA, xaxt = "n",
+   main = NA, ylim = c(0, 0.5))
> axis(1, at = b[c(2, 5, 8, 11)], labels = F)
> text(b[c(2, 5, 8, 11)] - 0.88, par("usr")[3] - 0.042, labels = lablist,
+   srt = 45, pos = 1, xpd = TRUE)
> axis(2, seq(0, 0.8, 0.1), paste(seq(0, 80, 10), "%", sep = ""),
+   las = 2, mgp = c(3, 0.6, 0))
> legend("top", c("ER+", "ER-", "HER2+"), col = c("blue", "red",
+   "magenta"), pch = rep(15, 3), bty = "n", horiz = T, cex = 0.9)

```

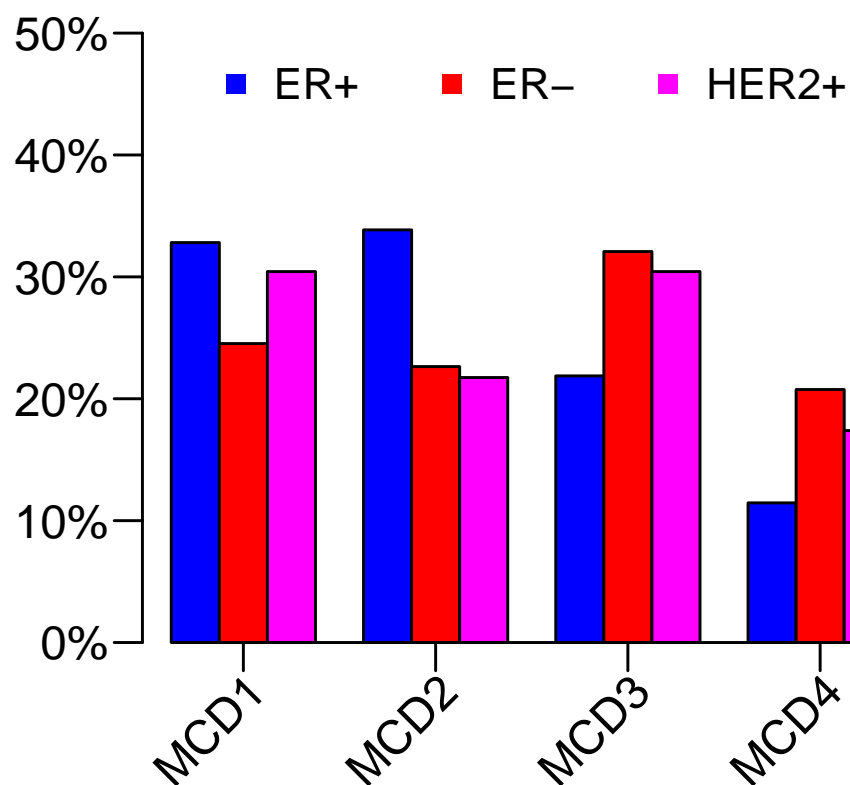


Figure 3 B

```

> size <- as.numeric(data.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "negative"), "Size.in.mm.Corrected"])
> mnew1 <- mean15and2[which(data.tmp2[, "ER.status.corr"] == "negative")]
> q = c(0, 0.15, 0.3, 0.45, max(mean15and2) + 0.01)
> fact.size <- rep(NA, length(size))

```



```

> fact.tmp <- rep(NA, length(mnew1))
> fact.tmp[mnew1 >= 0 & mnew1 < 0.15] <- "MCD1"
> fact.tmp[mnew1 >= 0.15 & mnew1 < 0.3] <- "MCD2"
> fact.tmp[mnew1 >= 0.3 & mnew1 < 0.45] <- "MCD3"
> fact.tmp[mnew1 >= 0.45 & mnew1 <= max(mean15and2)] <- "MCD4"
> grade <- as.numeric(data.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "negative"), "Grade"])
> node <- data.tmp2[which(data.tmp2[, "ER.status.corr"] == "negative"),
+   "Node.positive"]
> q = c(0, 0.15, 0.3, 0.45, max(mean15and2) + 0.01)
> size.fact <- rep(NA, length(fact.tmp))
> size.fact[size < 20 & !is.na(size)] <- "0 - 2cm"
> size.fact[size >= 20 & size < 50] <- "2 - 5cm"
> size.fact[size >= 50 & !is.na(size)] <- "5 - 8cm"
> function.GNS.plot(mnew1, fact.tmp, grade, node, size.fact, q,
+   xnames = c("MCD1", "MCD2", "MCD3", "MCD4"), main = NA)

```

Fisher's Exact Test for Count Data

```

data:
p-value = 0.004319
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.557351 25.176968
sample estimates:
odds ratio
 5.873173

```

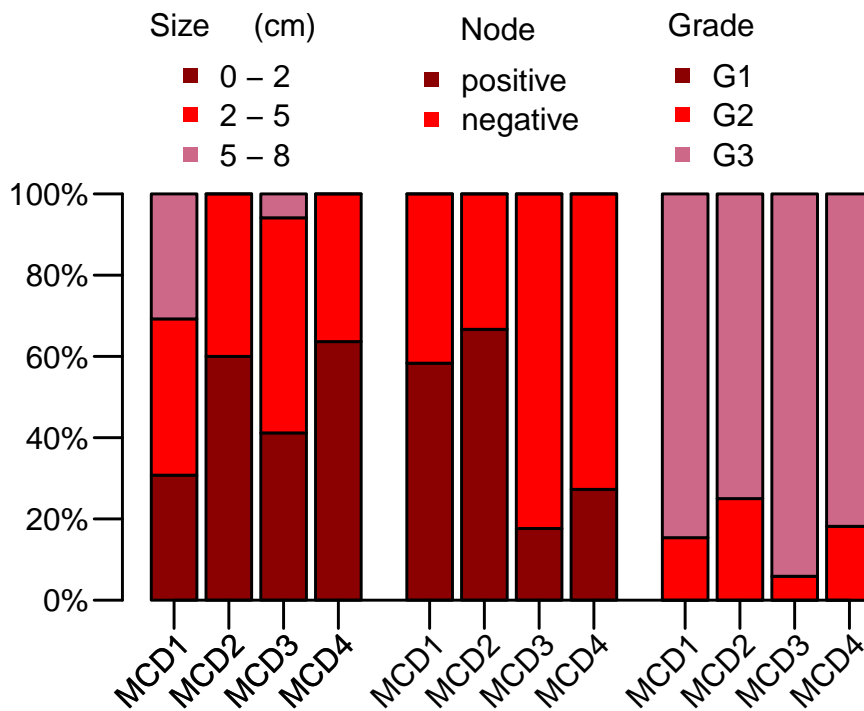


Figure 4

```

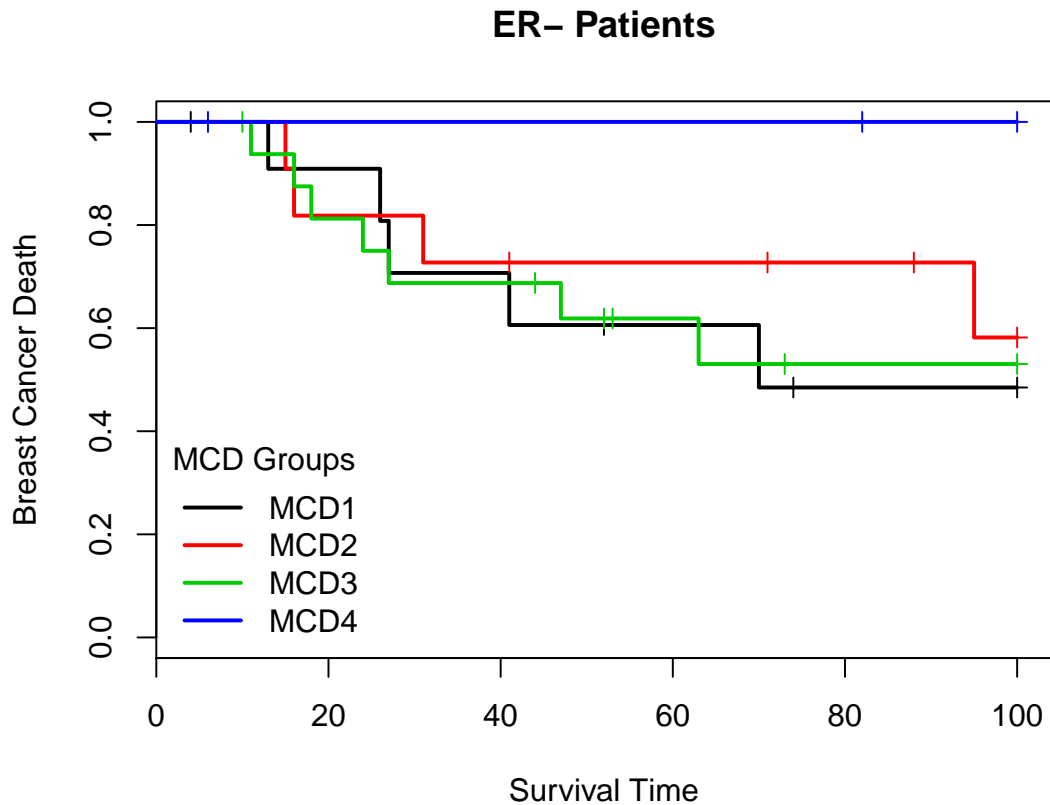
> surv_matrix2 <- surv_matrix
> surv_matrix2[surv_matrix2[, 1] > 100 & !is.na(surv_matrix2[,
+ 1]), 1] <- 100
> surv_matrix2[surv_matrix2[, 1] == 100 & !is.na(surv_matrix2[,
+ 1]), 2] <- 1
> mnew1 <- mean15and2[which(data.tmp2[, "ER.status.corr"] == "negative")]
> mm2 <- match(colnames(chrom2.tmp), rownames(surv_matrix))
> surv_matrix.tmp2 <- surv_matrix2[mm2, ]
> surv_matrix.tmp2 <- as.matrix(surv_matrix.tmp2)[which(data.tmp2[,
+ "ER.status.corr"] == "negative"), ]
> mnew1 <- mnew1[-which(is.na(surv_matrix.tmp2[, 1]))]
> surv_matrix.tmp2 <- surv_matrix.tmp2[-which(is.na(surv_matrix.tmp2[,
+ 1])), ]
> surv.obj <- Surv(surv_matrix.tmp2[, 1], surv_matrix.tmp2[, 2])
> fact.ER <- rep(NA, length(mnew1))
> fact.ER[mnew1 >= 0 & mnew1 < 0.15] <- "[0.0, 0.15)"
> fact.ER[mnew1 >= 0.15 & mnew1 < 0.3] <- "[0.15, 0.3)"

```

```

> fact.ER[mnew1 >= 0.3 & mnew1 < 0.45] <- "[0.3, 0.45)"
> fact.ER[mnew1 >= 0.45 & mnew1 <= max(mean15and2)] <- "[0.45, 0.62]"
> survplot(surv.obj ~ fact.ER, lwd = 2, snames = c("MCD1", "MCD2",
+ "MCD3", "MCD4"), stitle = "MCD Groups", ylab = "Breast Cancer Death",
+ main = "ER- Patients", xlab = "Survival Time")

```



number at risk					
12	9	7	5	3	3
11	9	8	7	6	4
17	13	11	7	5	5
11	10	10	10	10	9

1.5 Figure 5 Preparation

```

> prefix <- "/home/endesf01/windows2/AronDataSets/"
> postfix <- ".rma.RData"
> datasets <- unlist(list(wang = "Wang", chin = "Chin", sotiriou = "Sotiriou",
+ gse2603 = "Minn", gse1456.u133a = "Pawitan", desmedt = "Desmedt",
+ gse5460 = "Andrea"))
> all.rma <- loadIntoList(paste(prefix, names(datasets), postfix,
+ sep = ""))
> names(all.rma) <- datasets
> pool.surv.rfs <- as.matrix(all.surv.rfs[[1]])
> for (i in c(2:6, 11)) {

```

```

+   pool.surv.rfs <- rbind(pool.surv.rfs, all.surv.rfs[[i]])
+ }
> pool.surv.rfs <- Surv(pool.surv.rfs[, 1], pool.surv.rfs[, 2])
> pool.surv.rfs.cens <- censor(pool.surv.rfs, 5)
> combined.cin <- unlist(all.cin.scale)
> combined.subtype <- unlist(all.er)
> cin.quartile.value <- quantile(combined.cin, probs = seq(0, 1,
+   0.25))
> cin.quartile <- cut(combined.cin, cin.quartile.value, labels = c("0-25%",
+   "25-50%", "50-75%", "75-100%"), include.lowest = T)
> dataset.vec <- vector()
> for (i in 1:length(all.pam)) {
+   dataset.vec <- c(dataset.vec, rep(i, length(all.pam[[i]])))
+ }
> genesymb.var <- var.probes(genesymb, all.rma[[1]])
> all.ggi <- lapply(all.rma, function(x) {
+   scale(colMeans(rbind(exprs(x)[names(genesymb.var[genesymb.var %in%
+     sigs$sotiriu_grade[sigs$sotiriu_grade[, 2] == 1, 1]]),
+     ], (-1 * exprs(x)[names(genesymb.var[genesymb.var %in%
+     sigs$sotiriu_grade[sigs$sotiriu_grade[, 2] == -1, 1]]),
+     ])))
+ })
> all.mamma <- lapply(all.rma, function(x) {
+   scale(colMeans(rbind(exprs(x)[names(genesymb.var[genesymb.var %in%
+     sigs$vantveer_naive[sigs$vantveer_naive[, 2] == 1, 1]]),
+     ], (-1 * exprs(x)[names(genesymb.var[genesymb.var %in%
+     sigs$vantveer_naive[sigs$vantveer_naive[, 2] == -1, 1]]),
+     ])))
+ })
> a <- unlist(all.mamma)
> b <- unlist(all.er[surv.rfs.vec])
> mam.quartile.value <- quantile(a, probs = seq(0, 1, 0.25))
> mam.quartile <- cut(a, mam.quartile.value, labels = c("0-25%",
+   "25-50%", "50-75%", "75-100%"), include.lowest = T)[dataset.vec %in%
+   surv.rfs.vec]
> a <- unlist(all.ggi)
> ggi.quartile.value <- quantile(a, probs = seq(0, 1, 0.25))
> ggi.quartile <- cut(a, ggi.quartile.value, labels = c("0-25%",
+   "25-50%", "50-75%", "75-100%"), include.lowest = T)[dataset.vec %in%
+   surv.rfs.vec]
> combined.cin.rfs <- unlist(all.cin.scale[surv.rfs.vec])
> combined.subtype.rfs <- unlist(all.er[surv.rfs.vec])
> cin.quartile.rfs <- cin.quartile[dataset.vec %in% surv.rfs.vec]
> cin.hr.list <- list()
> cin.se.list <- list()
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs ~ cin.quartile.rfs ==
+     tmp[i], subset = combined.subtype.rfs == 0))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs ~ cin.quartile.rfs ==
+     tmp[i], subset = combined.subtype.rfs == 0))$coef[, "se(coef)"]
+ }

```

```

> cin.hr.list[[1]] <- cin.hr
> cin.se.list[[1]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs ~ cin.quartile.rfs ==
+     tmp[i], subset = combined.subtype.rfs == 1))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs ~ cin.quartile.rfs ==
+     tmp[i], subset = combined.subtype.rfs == 1))$coef[, "se(coef)"]
+ }
> cin.hr.list[[2]] <- cin.hr
> cin.se.list[[2]] <- cin.se
> gii.wang <- gii.fn(zwang.cn.seg, min.length = 6)
> no.tel.zwang.6p <- no.tel(zwang.mcp.seg, chrominfo = chrominfo,
+   min.probes = 6)
> s <- rownames(gii.wang)
> s <- substr(s, 2, nchar(s))
> s2 <- names(all.pam$Wang)
> s2 <- substr(s2, 1, nchar(s2) - 7)
> pool.surv.rfs <- as.matrix(all.surv.rfs[[1]])
> rownames(pool.surv.rfs) <- s2
> subtype <- all.er[[1]]
> names(subtype) <- s2
> rownames(gii.wang) <- s
> tmp <- intersect(rownames(gii.wang), names(subtype))
> subtype.cn <- subtype[tmp]
> tmp1 <- gii.wang[tmp, ]
> no.tel.zwang.6p <- no.tel.zwang.6p[-c(313:322), ]
> rownames(no.tel.zwang.6p)[1:272] <- substr(rownames(no.tel.zwang.6p)[1:272],
+   2, nchar(rownames(no.tel.zwang.6p)[1:272]))
> comb.ab <- tmp1[, 1]/max(tmp1[, 1]) + tmp1[, 2]/max(tmp1[, 2]) +
+   (no.tel.zwang.6p[tmp, 3] + no.tel.zwang.6p[tmp, 1])/max(no.tel.zwang.6p[tmp,
+   3] + no.tel.zwang.6p[tmp, 1])
> comb.ab <- comb.ab/max(comb.ab)
> surv.obj <- Surv(pool.surv.rfs[tmp, 1], pool.surv.rfs[tmp, 2])
> cin.quartile.value <- quantile(comb.ab, probs = seq(0, 1, 0.25))
> cin.quartile <- cut(comb.ab, cin.quartile.value, labels = c("0-25%",
+   "25-50%", "50-75%", "75-100%"), include.lowest = T)
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(surv.obj ~ cin.quartile == tmp[i],
+     subset = subtype.cn == 0))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(surv.obj ~ cin.quartile == tmp[i],
+     subset = subtype.cn == 0))$coef[, "se(coef)"]
+ }
> cin.hr.list[[3]] <- cin.hr
> cin.se.list[[3]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {

```

```

+   cin.hr[i] <- summary(coxph(surv.obj ~ cin.quartile == tmp[i],
+     subset = subtype.cn == 1))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(surv.obj ~ cin.quartile == tmp[i],
+     subset = subtype.cn == 1))$coef[, "se(coef)"]
+ }
> cin.hr.list[[4]] <- cin.hr
> cin.se.list[[4]] <- cin.se
> mm2 <- match(colnames(chrom2.11), rownames(surv_matrix))
> surv_matrix.tmp2 <- surv_matrix[mm2, ]
> survERneg <- surv_matrix.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "negative"), ]
> factERneg <- fact[which(data.tmp2[, "ER.status.corr"] == "negative")]
> surv.objERneg <- Surv(survERneg[, 1], survERneg[, 2])
> cin.hr <- vector()
> cin.se <- vector()
> tmp2 <- c("MCD1", "MCD2", "MCD3", "MCD4")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(surv.objERneg ~ factERneg == tmp2[i]))$coefficients[,
+     "coef"]
+   cin.se[i] <- summary(coxph(surv.objERneg ~ factERneg == tmp2[i]))$coef[,
+     "se(coef)"]
+ }
> cin.hr.list[[5]] <- cin.hr
> cin.se.list[[5]] <- cin.se
> survERpos <- surv_matrix.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "positive"), ]
> factERpos <- fact[which(data.tmp2[, "ER.status.corr"] == "positive")]
> surv.objERpos <- Surv(survERpos[, 1], survERpos[, 2])
> cin.hr <- vector()
> cin.se <- vector()
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(surv.objERpos ~ factERpos == tmp2[i]))$coefficients[,
+     "coef"]
+   cin.se[i] <- summary(coxph(surv.objERpos ~ factERpos == tmp2[i]))$coef[,
+     "se(coef)"]
+ }
> cin.hr.list[[6]] <- cin.hr
> cin.se.list[[6]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs.cens ~ mam.quartile ==
+     tmp[i], subset = b == 0))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs.cens ~ mam.quartile ==
+     tmp[i], subset = b == 0))$coef[, "se(coef)"]
+ }
> cin.hr.list[[7]] <- cin.hr
> cin.se.list[[7]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs.cens ~ ggi.quartile ==

```

```

+       tmp[i], subset = b == 0))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs.cens ~ ggi.quartile ==
+       tmp[i], subset = b == 0))$coef[, "se(coef)"]
+ }
> cin.hr.list[[8]] <- cin.hr
> cin.se.list[[8]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs.cens ~ mam.quartile ==
+       tmp[i], subset = b == 1))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs.cens ~ mam.quartile ==
+       tmp[i], subset = b == 1))$coef[, "se(coef)"]
+ }
> cin.hr.list[[9]] <- cin.hr
> cin.se.list[[9]] <- cin.se
> cin.hr <- vector()
> cin.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   cin.hr[i] <- summary(coxph(pool.surv.rfs.cens ~ ggi.quartile ==
+       tmp[i], subset = b == 1))$coef[, "coef"]
+   cin.se[i] <- summary(coxph(pool.surv.rfs.cens ~ ggi.quartile ==
+       tmp[i], subset = b == 1))$coef[, "se(coef)"]
+ }
> cin.hr.list[[10]] <- cin.hr
> cin.se.list[[10]] <- cin.se

```

Figure 5A

```

> par(mfrow = c(2, 2))
> metaplot(cin.hr.list[[1]], cin.se.list[[1]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "CIN70 \n(ER negative)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[3]], cin.se.list[[3]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "aCGH \n(ER negative)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[5]], cin.se.list[[5]], labels = tmp2, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "CEP2+15 \n(ER negative)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))

```

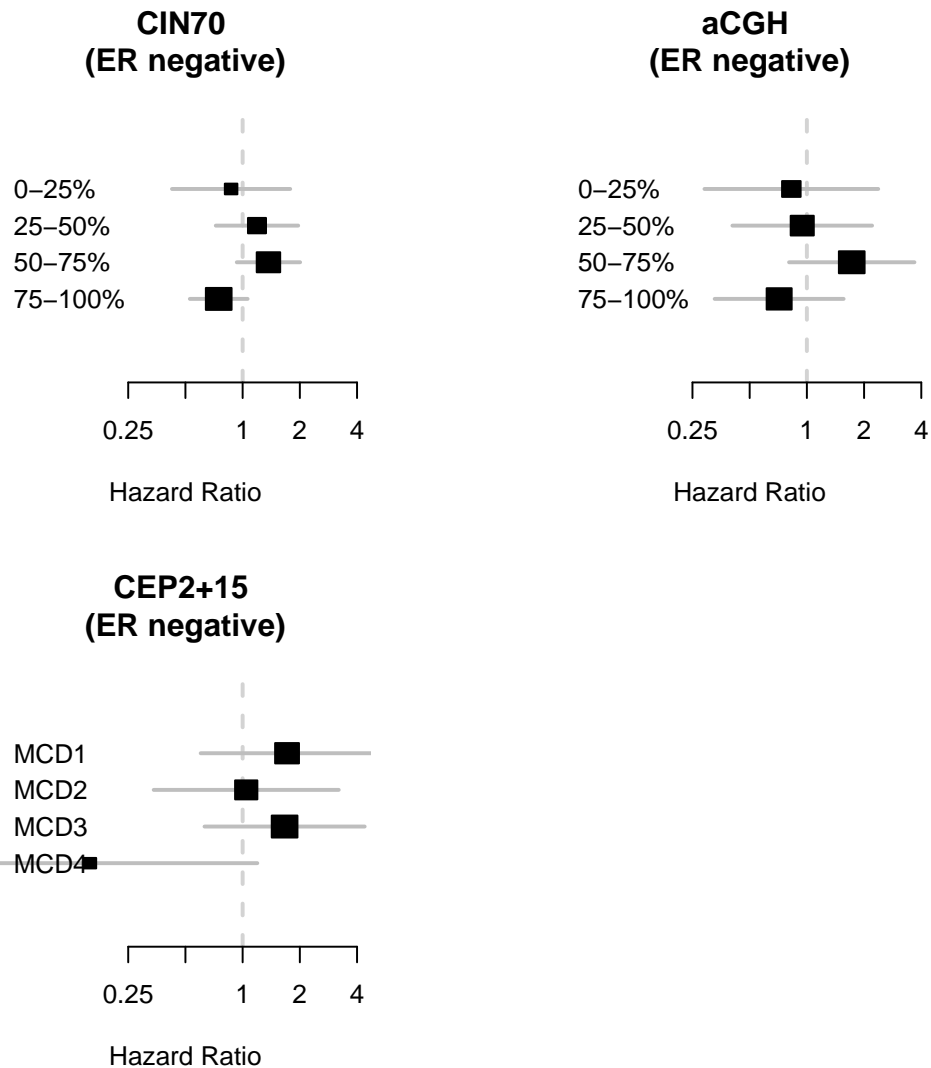


Figure 5B

```

> par(mfrow = c(1, 2))
> metaplot(cin.hr.list[[7]], cin.se.list[[7]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "Mammaprint \n(ER negative)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[8]], cin.se.list[[8]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "GGI \n(ER negative)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))

```

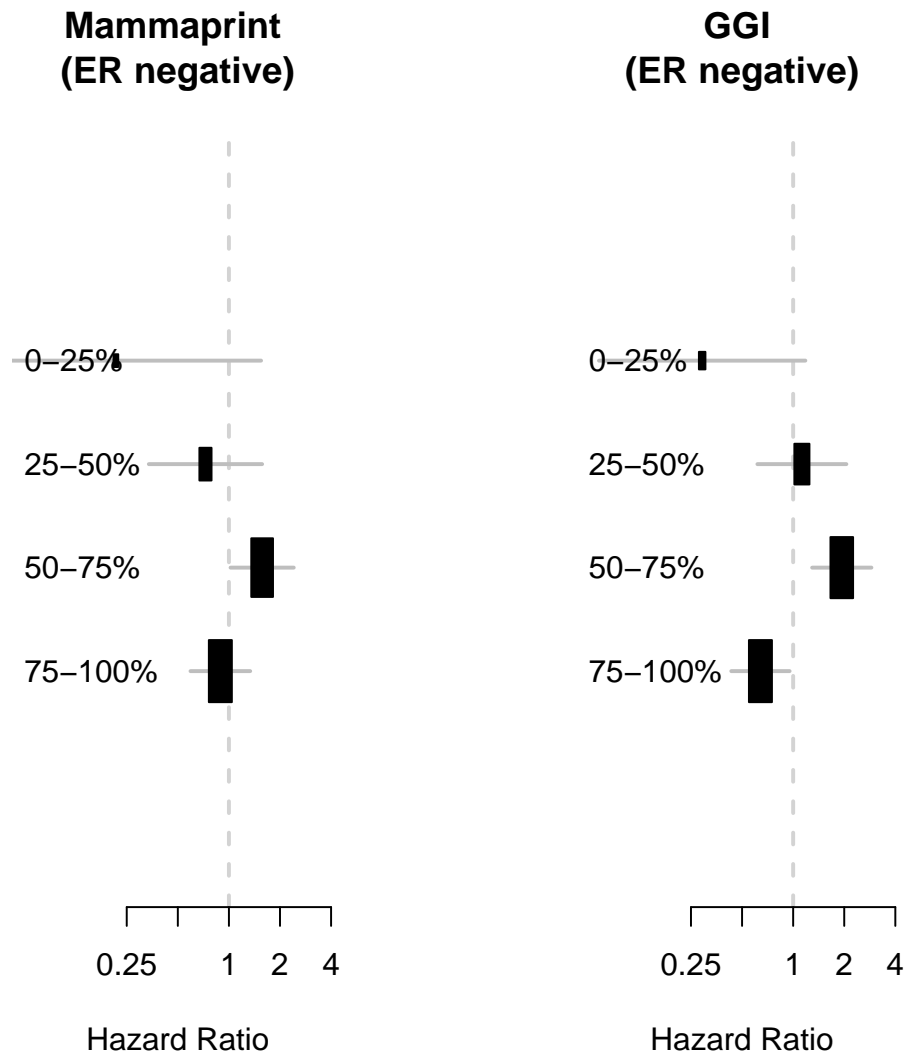



Figure 5C

```

> par(mfrow = c(2, 2))
> metaplot(cin.hr.list[[2]], cin.se.list[[2]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "CIN70 \n(ER positive)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[4]], cin.se.list[[4]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "aCGH \n(ER positive)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[6]], cin.se.list[[6]], labels = tmp2, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "CEP2+15 \n(ER positive)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))

```

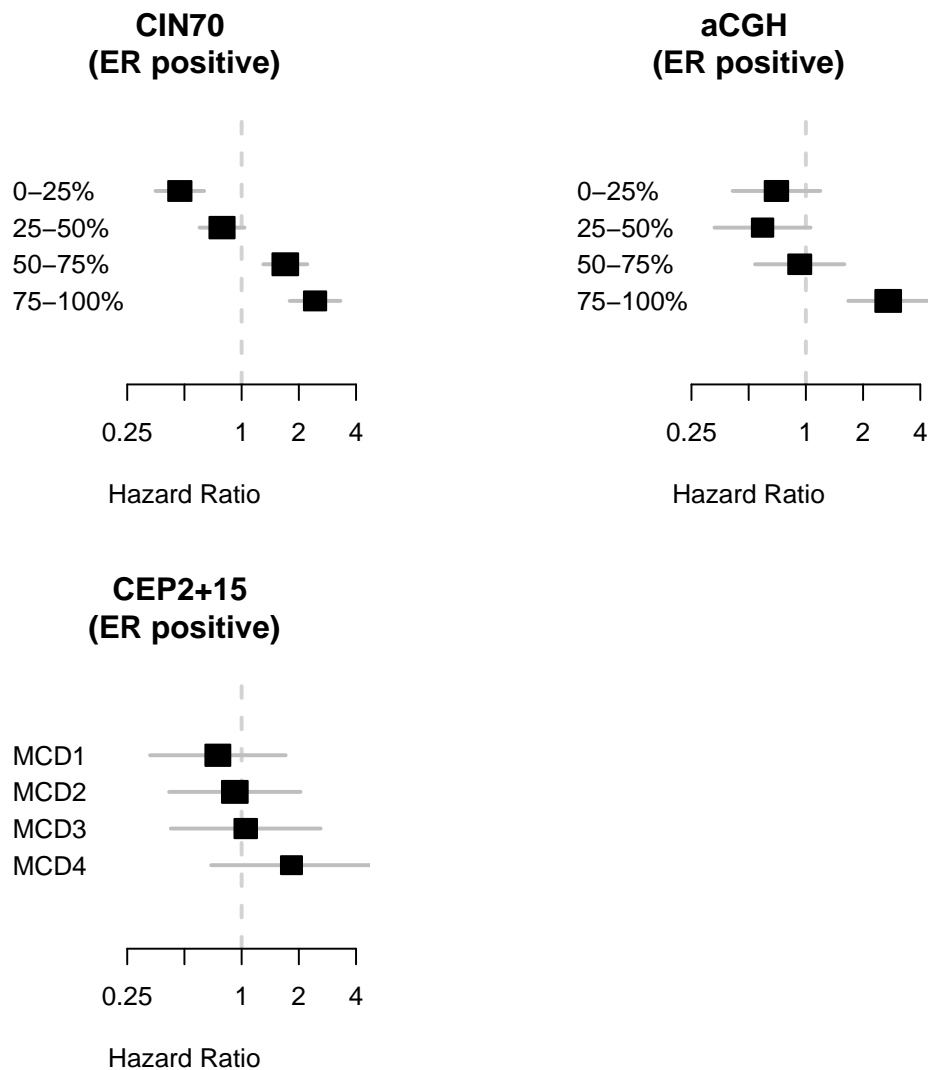
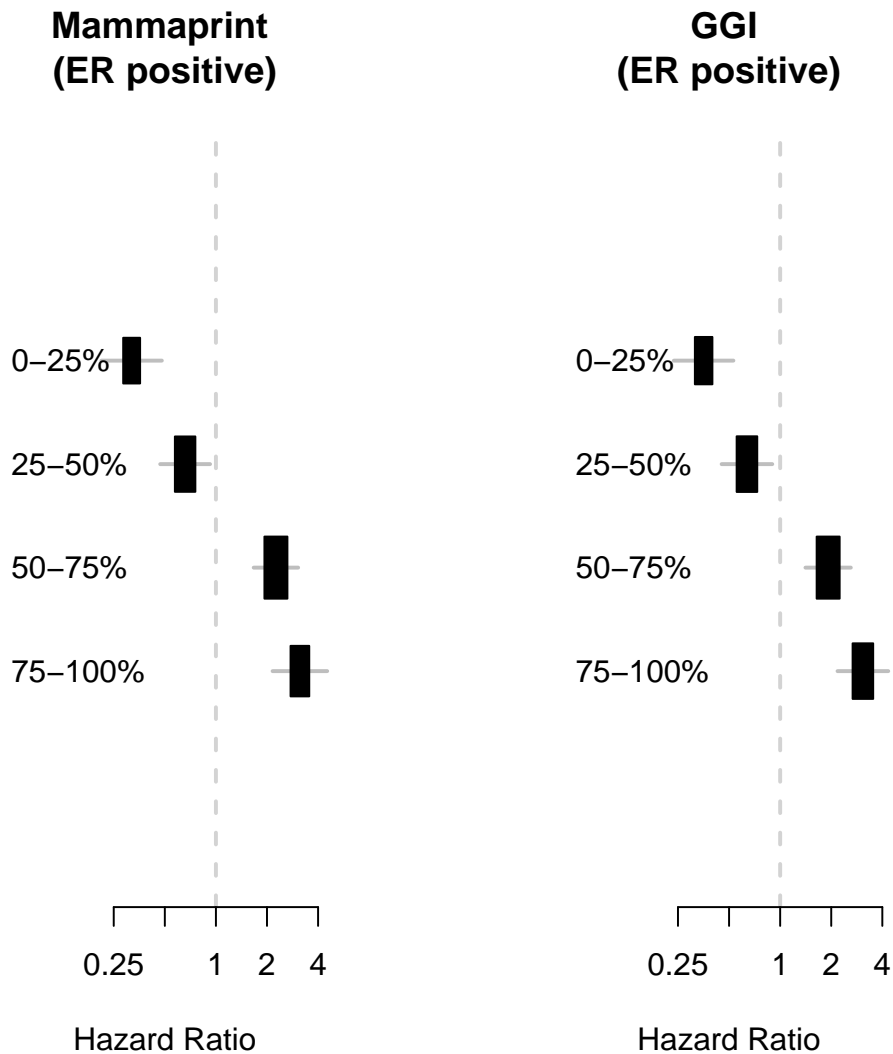


Figure 5D

```

> par(mfrow = c(1, 2))
> metaplot(cin.hr.list[[9]], cin.se.list[[9]], labels = tmp, xlim = c(log(0.25),
+   log(4)), xlab = "Hazard Ratio", ylab = "", main = "Mammaprint \n(ER positive)",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))
> metaplot(cin.hr.list[[10]], cin.se.list[[10]], labels = tmp,
+   xlim = c(log(0.25), log(4)), xlab = "Hazard Ratio", ylab = "",
+   main = "GGI \n(ER positive)", logeffect = F, xaxt = "n")
> at <- log(c(0.25, 0.5, 1, 2, 4))
> axis(1, at = at, labels = c(0.25, 0.5, 1, 2, 4))

```

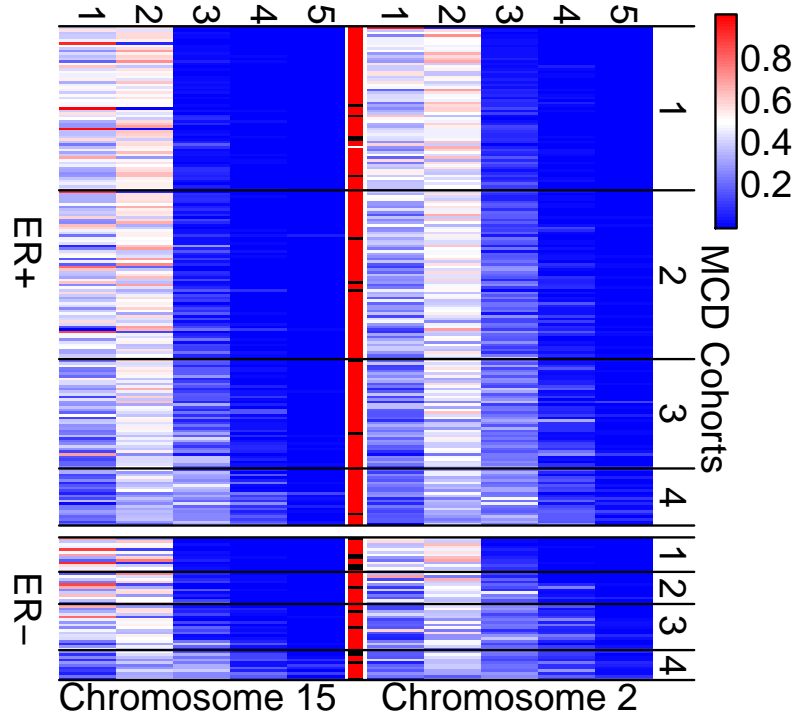


Supplementary Figure 1

```

> m.neg <- mean15and2[which(data.tmp2[, "ER.status.corr"] == "negative")]
> HER2.tmp <- data.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "negative"), ][order(m.neg, decreasing = T), ]
> HER2.neg <- rep(NA, length(m.neg))
> HER2.neg[which(HER2.tmp[, "HER2.status"] == "negative")] <- 1
> HER2.neg[which(HER2.tmp[, "HER2.status"] == "positive")] <- 2
> m.pos <- mean15and2[which(data.tmp2[, "ER.status.corr"] == "positive")]
> HER2.tmp <- data.tmp2[which(data.tmp2[, "ER.status.corr"] ==
+   "positive"), ][order(m.pos), ]
> HER2.pos <- rep(NA, length(m.pos))
> HER2.pos[which(HER2.tmp[, "HER2.status"] == "negative")] <- 1
> HER2.pos[which(HER2.tmp[, "HER2.status"] == "positive")] <- 2
> fun.Image2(chrom15.tmp, chrom2.tmp, HER2.neg, HER2.pos, m.neg,
+   m.pos)

```



Supplementary Figure 2A

```

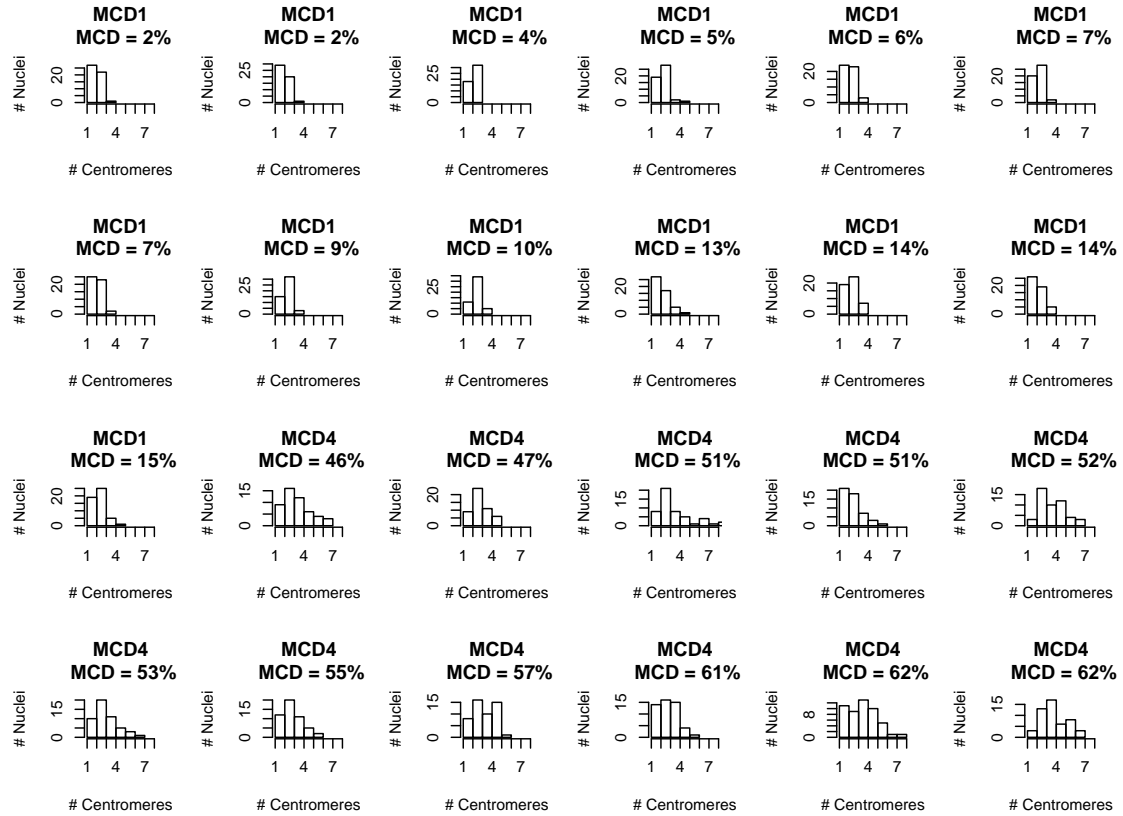
> chrom2.ERneg <- chrom2.11[, which(data.tmp2[, "ER.status.corr"] ==
+   "negative")]
> m.ER.neg <- mean15and2[which(data.tmp2[, "ER.status.corr"] ==
+   "negative")]
> fact.ER.neg <- rep(NA, length(m.ER.neg))
> fact.ER.neg[m.ER.neg >= 0 & m.ER.neg < 0.15] <- "MCD1"
> fact.ER.neg[m.ER.neg >= 0.15 & m.ER.neg < 0.3] <- "MCD2"
> fact.ER.neg[m.ER.neg >= 0.3 & m.ER.neg < 0.45] <- "MCD3"
> fact.ER.neg[m.ER.neg >= 0.45 & m.ER.neg <= max(m.ER.neg)] <- "MCD4"
> chrom2.ERneg <- chrom2.ERneg[, order(m.ER.neg)]
> fact.ERneg <- fact.ER.neg[order(m.ER.neg)]
> chrom2.ERneg.tmp <- cbind(chrom2.ERneg[, fact.ERneg == "MCD1"],
+   chrom2.ERneg[, fact.ERneg == "MCD4"])
> fact.ERneg.tmp <- c(fact.ERneg[fact.ERneg == "MCD1"], fact.ERneg[fact.ERneg ==
+   "MCD4"])
> m.ER.neg.tmp <- c(sort(m.ER.neg)[fact.ERneg == "MCD1"], sort(m.ER.neg)[fact.ERneg ==
+   "MCD4"])

```

```

> par(mfrow = c(4, 6))
> for (i in 1:24) hist(chrom2.ERneg.tmp[, i], xlim = c(1, 8), breaks = 1:(length(unique(chrom2.ER
+   i])) + 1), right = F, main = paste(fact.ERneg.tmp[i], "\n MCD = ",
+   round(100 * m.ER.neg.tmp[i]), "%", sep = ""), xlab = "# Centromeres",
+   ylab = "# Nuclei")

```

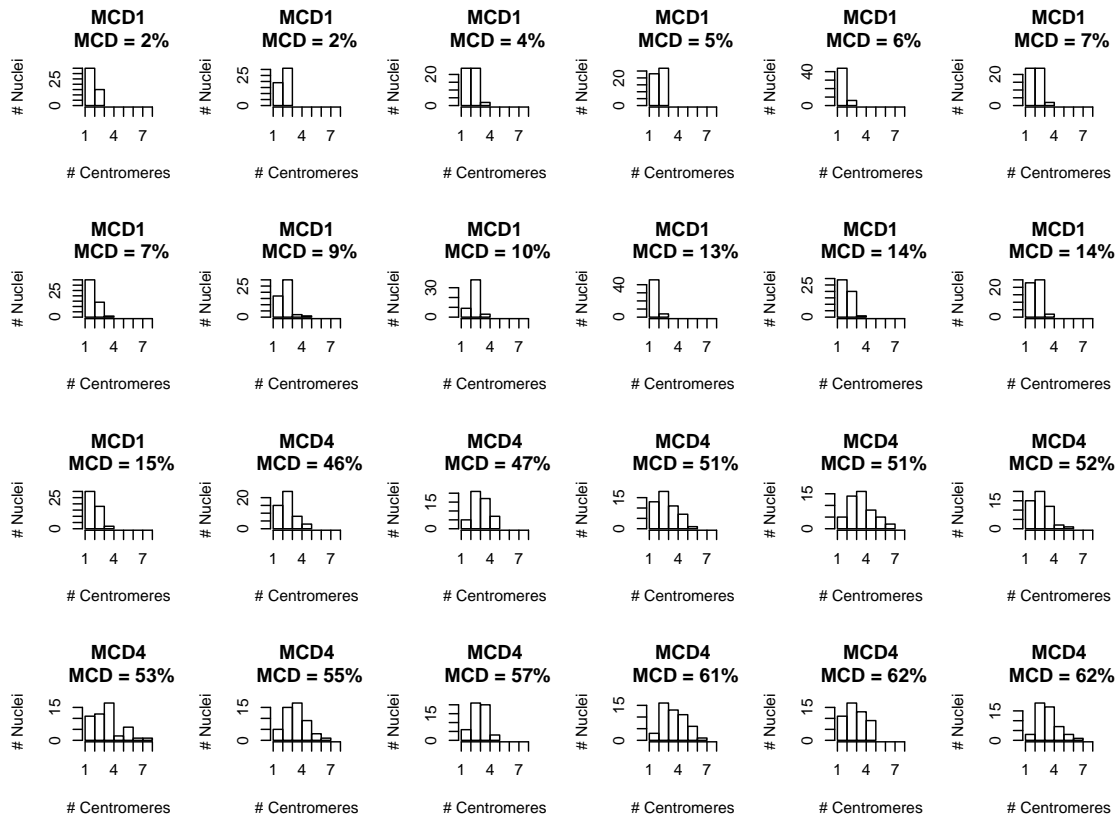


Supplementary Figure 2A

```

> chrom15.ERneg <- chrom15.11[, which(data.tmp2[, "ER.status.corr"] ==
+   "negative")]
> chrom15.ERneg <- chrom15.ERneg[, order(m.ER.neg)]
> chrom15.ERneg.tmp <- cbind(chrom15.ERneg[, fact.ERneg == "MCD1"],
+   chrom15.ERneg[, fact.ERneg == "MCD4"])
> par(mfrow = c(4, 6))
> for (i in 1:24) hist(chrom15.ERneg.tmp[, i], xlim = c(1, 8),
+   breaks = 1:(length(unique(chrom15.ERneg.tmp[, i])) + 1),
+   right = F, main = paste(fact.ERneg.tmp[i], "\n MCD = ", round(100 *
+   m.ER.neg.tmp[i]), "%", sep = ""), xlab = "# Centromeres",
+   ylab = "# Nuclei")

```



Supplementary Figure 3

```

> treat.ER <- read.csv("/home/endesf01/Leeds/TreatmentERneg.csv")
> treat.ER <- as.matrix(treat.ER)
> treat.ER[treat.ER == "ND"] <- NA
> mm2 <- match(colnames(chrom2.tmp), rownames(surv_matrix2))
> surv_matrix.tmp2 <- surv_matrix2[mm2, ]
> surv_matrix.tmp2 <- as.matrix(surv_matrix.tmp2)
> tmp <- mean15and2[match(as.numeric(treat.ER[, 4]), as.numeric(rownames(data.tmp2)))]
> data.ER.neg <- data.tmp2[match(as.numeric(treat.ER[, 4]), as.numeric(rownames(data.tmp2))),
+ ]
> surv.ER <- surv_matrix.tmp2[match(as.numeric(treat.ER[, 4]),
+ as.numeric(rownames(data.tmp2))), ]
> fact.ER <- fact[match(as.numeric(treat.ER[, 4]), as.numeric(rownames(data.tmp2)))]
> tmp2 <- cbind(tmp, treat.ER[, c(6, 7)])
> data.ER.neg <- data.ER.neg[-which(is.na(tmp2[, 2]) & is.na(tmp2[,
+ 3])), ]
> fact.ER <- fact.ER[-which(is.na(tmp2[, 2]) & is.na(tmp2[, 3]))]
> surv.ER <- surv.ER[-which(is.na(tmp2[, 2]) & is.na(tmp2[, 3])),
+ ]
> tmp2 <- tmp2[-which(is.na(tmp2[, 2]) & is.na(tmp2[, 3])), ]
> tmp3 <- vector()
> tmp3[as.numeric(tmp2[, 2]) > 0 | as.numeric(tmp2[, 3]) > 0] <- "Chemo"
> tmp3[is.na(tmp3)] <- "No Chemo"
> surv.obj <- Surv(surv.ER[, 1], surv.ER[, 2])
> cc <- cbind(surv.ER, tmp3, fact.ER)
> fact.help <- rep(NA, nrow(cc))

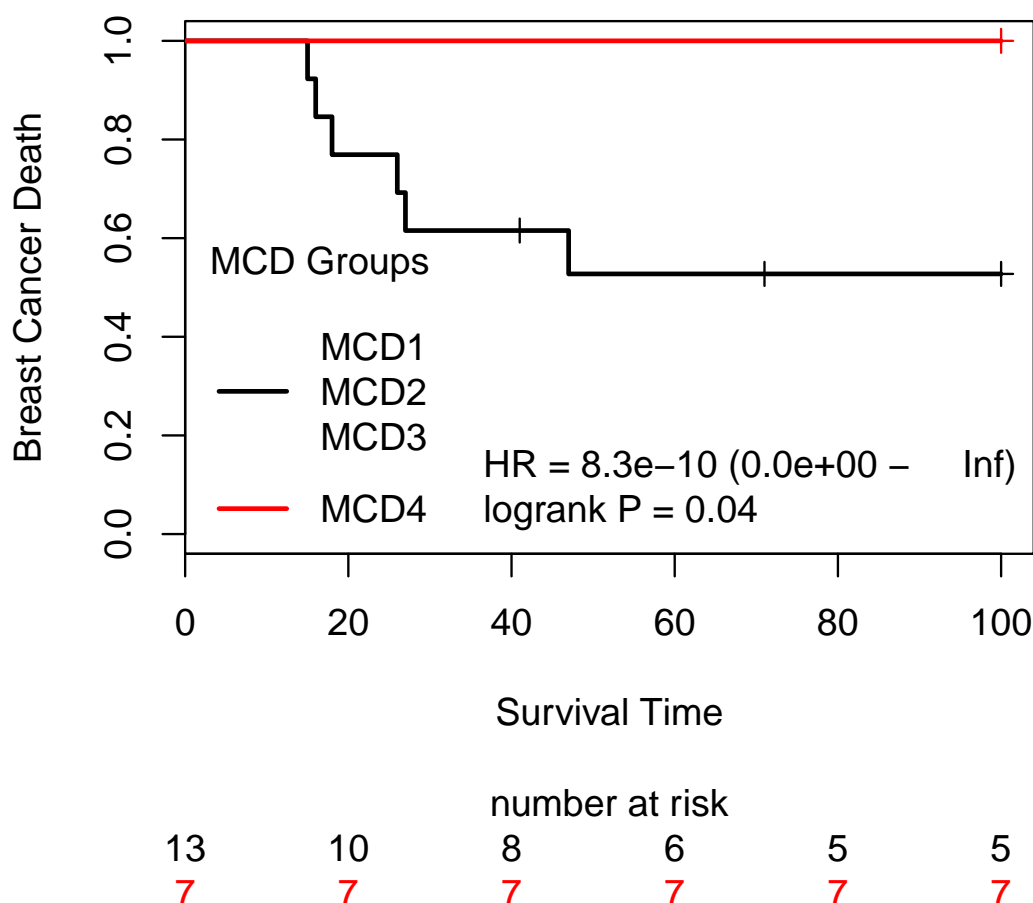
```

```

> fact.help[cc[, 3] == "Chemo" & cc[, 4] != "MCD4"] <- "Chemo1-3"
> fact.help[cc[, 3] == "No Chemo" & cc[, 4] != "MCD4"] <- "NoChemo1-3"
> fact.help[cc[, 3] == "Chemo" & cc[, 4] == "MCD4"] <- "Chemo4"
> fact.help[cc[, 3] == "No Chemo" & cc[, 4] == "MCD4"] <- "NoChemo4"
> cc.help <- cc[cc[, 3] == "Chemo", ]
> surv.obj <- Surv(as.numeric(cc.help[, 1]), as.numeric(cc.help[,
+ 2]))
> survplot(surv.obj ~ cc.help[, 4] == "MCD4", stitle = "MCD Groups",
+ snames = c("MCD1\nMCD2\nMCD3", "MCD4"), main = "ER- Patients with Chemotherapy",
+ lwd = 2, hr.pos = "bottomright", ylab = "Breast Cancer Death",
+ xlab = "Survival Time")

```

ER- Patients with Chemotherapy



Supplementary Figure 4

```

> tcga.duke.dressman.cin.quartiles <- cut(tcga.duke.dressman.cin70,
+ quantile(tcga.duke.dressman.cin70), labels = c("0-25%", "25-50%",
+ "50-75%", "75-100%"), include.lowest = T)
> lung.cin.quartile <- cut(lung.s.cin.scale, quantile(lung.s.cin.scale),
+ labels = c("0-25%", "25-50%", "50-75%", "75-100%"), include.lowest = T)

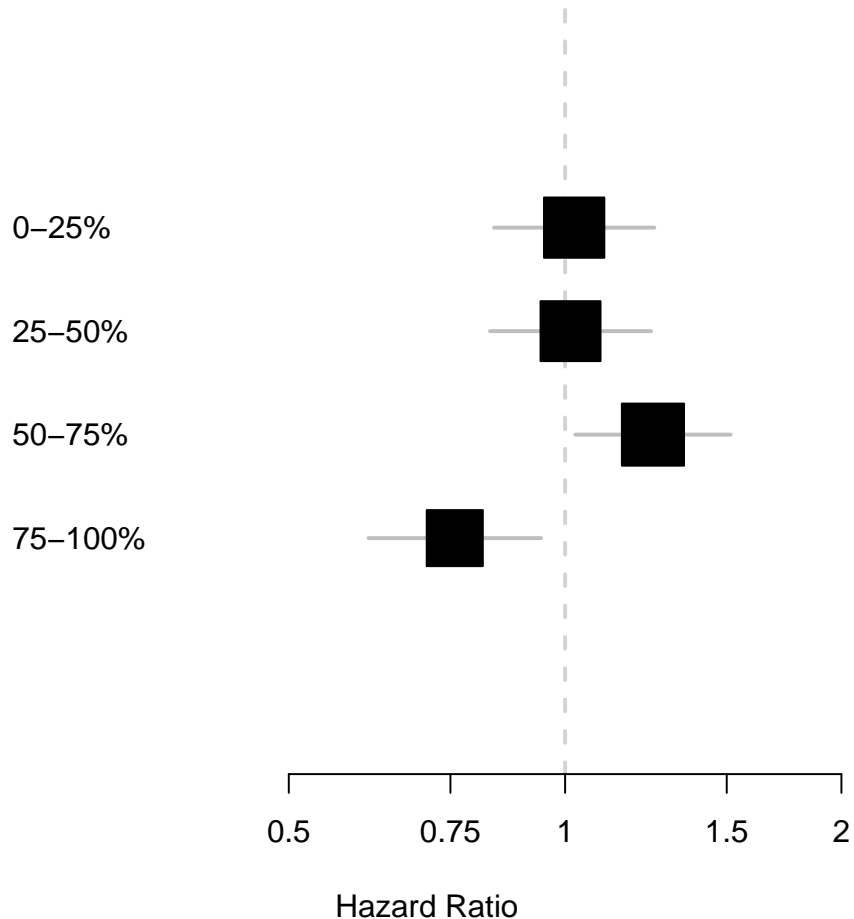
```

```

> gastric.cin.quartile <- cut(gastric.cin, quantile(gastric.cin),
+   labels = c("0-25%", "25-50%", "50-75%", "75-100%"), include.lowest = T)
> tmp.quart <- factor(c(as.character(tcga.duke.dressman.cin.quartiles),
+   as.character(lung.cin.quartile), as.character(gastric.cin.quartile)))
> tmp.surv <- Surv(rbind(tcga.duke.dressman.surv.cens, lung.surv.cens,
+   gastric.surv))
> tmp.hr <- vector()
> tmp.se <- vector()
> tmp <- c("0-25%", "25-50%", "50-75%", "75-100%")
> for (i in 1:4) {
+   tmp.hr[i] <- summary(coxph(tmp.surv ~ tmp.quart == tmp[i]))$coef[,
+     "coef"]
+   tmp.se[i] <- summary(coxph(tmp.surv ~ tmp.quart == tmp[i]))$coef[,
+     "se(coef)"]
+ }
> metaplot(tmp.hr, tmp.se, labels = tmp, xlim = c(log(0.5), log(2)),
+   xlab = "Hazard Ratio", ylab = "", main = "Combined Ovary, NSCLC, Gastric",
+   logeffect = F, xaxt = "n")
> at <- log(c(0.5, 0.75, 1, 1.5, 2))
> axis(1, at = at, labels = c(0.5, 0.75, 1, 1.5, 2))

```


Combined Ovary, NSCLC, Gastric



1.6 Multivariate Analysis Mean Percentage Modal Centromere Deviation

```
> treat <- as.matrix(treat)
> treat <- treat[match(as.numeric(rownames(data.tmp2)), as.numeric(treat[,
+ 4])), ]
> treat <- apply(treat[, c(4, 7, 8)], 2, as.numeric, na.rm = T)
> treat.tmp <- rep(NA, nrow(treat))
> treat.tmp[which(treat[, 2] != 0 | treat[, 3] != 0)] <- "Chemo"
> treat.tmp[which(treat[, 2] == 0 & treat[, 3] == 0)] <- "No Chemo"
> mm2 <- match(colnames(chrom2.tmp), rownames(surv_matrix))
> data.tmp <- cbind(surv_matrix[mm2, ], data.tmp2[, c("ER.status.corr",
+ "HER2.status", "Size.in.mm.Corrected", "Grade", "Node.positive")],
+ fact, mean15and2, treat.tmp)
> data.tmp <- na.omit(data.tmp)
> data.tmp.ER <- data.tmp[which(data.tmp[, "ER.status.corr"] ==
+ "negative"), ]
```

```

> surv.obj <- Surv(as.numeric(data.tmp.ER[, 1]), as.numeric(data.tmp.ER[,
+ 2]))
> library(coxrobust)
> multi_coxph <- coxr(surv.obj ~ as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"]) +
+ as.factor(data.tmp.ER[, "fact"] == "MCD4") + as.factor(data.tmp.ER[,
+ "Grade"]) + as.factor(data.tmp.ER[, "Node.positive"]) + as.factor(data.tmp.ER[,
+ "HER2.status"]) + as.factor(data.tmp.ER[, "treat.tmp"]))
> print(multi_coxph)

```

Call:

```
coxr(formula = surv.obj ~ as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"]) + as.factor(data.
```

Partial likelihood estimator

	coef	exp(coef)	se(coef)
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.0217	1.022	0.0216
as.factor(data.tmp.ER[, "fact"] == "MCD4")TRUE	-1.7963	0.166	1.1516
as.factor(data.tmp.ER[, "Grade"])3	1.5393	4.661	1.0992
as.factor(data.tmp.ER[, "Node.positive"])yes	0.5800	1.786	0.6078
as.factor(data.tmp.ER[, "HER2.status"])positive	0.6432	1.903	0.9229
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.1143	1.121	0.6785
	p		
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.316		
as.factor(data.tmp.ER[, "fact"] == "MCD4")TRUE	0.119		
as.factor(data.tmp.ER[, "Grade"])3	0.161		
as.factor(data.tmp.ER[, "Node.positive"])yes	0.340		
as.factor(data.tmp.ER[, "HER2.status"])positive	0.486		
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.866		

Wald test=11.8 on 6 df, p=0.0656

Robust estimator

	coef	exp(coef)	se(coef)
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.100	1.1051	0.0390
as.factor(data.tmp.ER[, "fact"] == "MCD4")TRUE	-2.492	0.0827	1.0945
as.factor(data.tmp.ER[, "Grade"])3	2.694	14.7902	1.0866
as.factor(data.tmp.ER[, "Node.positive"])yes	0.471	1.6023	0.6287
as.factor(data.tmp.ER[, "HER2.status"])positive	1.966	7.1390	0.8475
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.596	1.8154	0.5872
	p		
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.0103		
as.factor(data.tmp.ER[, "fact"] == "MCD4")TRUE	0.0228		
as.factor(data.tmp.ER[, "Grade"])3	0.0132		
as.factor(data.tmp.ER[, "Node.positive"])yes	0.4533		
as.factor(data.tmp.ER[, "HER2.status"])positive	0.0204		
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.3099		

Extended Wald test=11185 on 6 df, p=0

1.7 Multivariate Analysis Shannon Diversity Index

```

> div15and2 <- (div.2 + div.15)/2
> tt <- (max(div15and2) - min(div15and2))/4
> thresh <- c(0, min(div15and2) + tt, min(div15and2) + 2 * tt,
+ min(div15and2) + 3 * tt, max(div15and2))

```

```

> fact.div <- rep(NA, length(div15and2))
> fact.div[div15and2 >= thresh[1] & div15and2 < thresh[2]] <- "SDI1"
> fact.div[div15and2 >= thresh[2] & div15and2 < thresh[3]] <- "SDI2"
> fact.div[div15and2 >= thresh[3] & div15and2 < thresh[4]] <- "SDI3"
> fact.div[div15and2 >= thresh[4] & div15and2 <= max(div15and2)] <- "SDI4"
> mm2 <- match(colnames(chrom2.tmp), rownames(surv_matrix))
> data.tmp <- cbind(surv_matrix[mm2, ], data.tmp2[, c("ER.status.corr",
+ "HER2.status", "Size.in.mm.Corrected", "Grade", "Node.positive")],
+ fact.div, div15and2, treat.tmp)
> data.tmp <- na.omit(data.tmp)
> data.tmp.ER <- data.tmp[which(data.tmp[, "ER.status.corr"] ==
+ "negative"), ]
> surv.obj <- Surv(as.numeric(data.tmp.ER[, 1]), as.numeric(data.tmp.ER[,
+ 2]))
> library(coxrobust)
> multi_coxph <- coxr(surv.obj ~ as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"]) +
+ as.factor(data.tmp.ER[, "fact.div"] == "SDI4") + as.factor(data.tmp.ER[,
+ "Grade"]) + as.factor(data.tmp.ER[, "Node.positive"]) + as.factor(data.tmp.ER[,
+ "HER2.status"]) + as.factor(data.tmp.ER[, "treat.tmp"]))
> print(multi_coxph)

```

Call:

```
coxr(formula = surv.obj ~ as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"]) + as.factor(data.
```

Partial likelihood estimator

	coef	exp(coef)	se(coef)
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.0256	1.026	0.0219
as.factor(data.tmp.ER[, "fact.div"] == "SDI4")TRUE	-1.6824	0.186	1.1070
as.factor(data.tmp.ER[, "Grade"])3	1.7612	5.819	1.1478
as.factor(data.tmp.ER[, "Node.positive"])yes	0.5806	1.787	0.6131
as.factor(data.tmp.ER[, "HER2.status"])positive	0.3852	1.470	0.9580
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.1737	1.190	0.6626
	p		
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.241		
as.factor(data.tmp.ER[, "fact.div"] == "SDI4")TRUE	0.129		
as.factor(data.tmp.ER[, "Grade"])3	0.125		
as.factor(data.tmp.ER[, "Node.positive"])yes	0.344		
as.factor(data.tmp.ER[, "HER2.status"])positive	0.688		
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.793		

Wald test=11.7 on 6 df, p=0.0694

Robust estimator

	coef	exp(coef)	se(coef)
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.131	1.1398	0.060
as.factor(data.tmp.ER[, "fact.div"] == "SDI4")TRUE	-2.580	0.0758	0.917
as.factor(data.tmp.ER[, "Grade"])3	4.132	62.3061	1.345
as.factor(data.tmp.ER[, "Node.positive"])yes	0.869	2.3842	0.732
as.factor(data.tmp.ER[, "HER2.status"])positive	1.214	3.3676	0.940
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo	0.711	2.0365	0.558
	p		
as.numeric(data.tmp.ER[, "Size.in.mm.Corrected"])	0.02915		
as.factor(data.tmp.ER[, "fact.div"] == "SDI4")TRUE	0.00489		
as.factor(data.tmp.ER[, "Grade"])3	0.00212		

```

as.factor(data.tmp.ER[, "Node.positive"])yes      0.23554
as.factor(data.tmp.ER[, "HER2.status"])positive  0.19662
as.factor(data.tmp.ER[, "treat.tmp"])No Chemo    0.20267

```

Extended Wald test=60.7 on 6 df, p=3.22e-11

1.8 Statistical Analysis for Table 1

```

> age.high <- c(length(which(as.numeric(data.tmp2[, 1]) > 50 &
+   fact == "MCD1")), length(which(as.numeric(data.tmp2[, 1]) >
+   50 & fact == "MCD2")), length(which(as.numeric(data.tmp2[,
+   1]) > 50 & fact == "MCD3")), length(which(as.numeric(data.tmp2[,
+   1]) > 50 & fact == "MCD4")))
> age.low <- c(length(which(as.numeric(data.tmp2[, 1]) <= 50 &
+   fact == "MCD1")), length(which(as.numeric(data.tmp2[, 1]) <=
+   50 & fact == "MCD2")), length(which(as.numeric(data.tmp2[,
+   1]) <= 50 & fact == "MCD3")), length(which(as.numeric(data.tmp2[,
+   1]) <= 50 & fact == "MCD4")))
> print(prop.trend.test(age.high, age.high + age.low))

```

Chi-squared Test for Trend in Proportions

```

data: age.high out of age.high + age.low ,
      using scores: 1 2 3 4
X-squared = 2.757, df = 1, p-value = 0.09683

```

```

> grade1 <- c(length(which(as.numeric(data.tmp2[, "Grade"]) ==
+   1 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD1")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 1 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD2")),
+   length(which(as.numeric(data.tmp2[, "Grade"]) == 1 & !is.na(data.tmp2[,
+   "Grade"]) & fact == "MCD3")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 1 & !is.na(data.tmp2[, "Grade"]) & fact ==
+   "MCD4")))
> grade2 <- c(length(which(as.numeric(data.tmp2[, "Grade"]) ==
+   2 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD1")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 2 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD2")),
+   length(which(as.numeric(data.tmp2[, "Grade"]) == 2 & !is.na(data.tmp2[,
+   "Grade"]) & fact == "MCD3")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 2 & !is.na(data.tmp2[, "Grade"]) & fact ==
+   "MCD4")))
> grade3 <- c(length(which(as.numeric(data.tmp2[, "Grade"]) ==
+   3 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD1")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 3 & !is.na(data.tmp2[, "Grade"]) & fact == "MCD2")),
+   length(which(as.numeric(data.tmp2[, "Grade"]) == 3 & !is.na(data.tmp2[,
+   "Grade"]) & fact == "MCD3")), length(which(as.numeric(data.tmp2[,
+   "Grade"]) == 3 & !is.na(data.tmp2[, "Grade"]) & fact ==
+   "MCD4")))
> print(prop.trend.test(grade1, grade1 + grade2 + grade3))

```

Chi-squared Test for Trend in Proportions

```

data: grade1 out of grade1 + grade2 + grade3 ,
      using scores: 1 2 3 4
X-squared = 4.9787, df = 1, p-value = 0.02566

```

```

> print(prop.trend.test(grade2, grade1 + grade2 + grade3))

      Chi-squared Test for Trend in Proportions

data:  grade2 out of grade1 + grade2 + grade3 ,
      using scores: 1 2 3 4
X-squared = 2.4522, df = 1, p-value = 0.1174

> print(prop.trend.test(grade3, grade1 + grade2 + grade3))

      Chi-squared Test for Trend in Proportions

data:  grade3 out of grade1 + grade2 + grade3 ,
      using scores: 1 2 3 4
X-squared = 10.8411, df = 1, p-value = 0.0009927

> ERneg <- c(length(which(data.tmp2[, "ER.status.corr"] == "negative" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD1")),
+   length(which(data.tmp2[, "ER.status.corr"] == "negative" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD2")),
+   length(which(data.tmp2[, "ER.status.corr"] == "negative" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD3")),
+   length(which(data.tmp2[, "ER.status.corr"] == "negative" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD4")))
> ERpos <- c(length(which(data.tmp2[, "ER.status.corr"] == "positive" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD1")),
+   length(which(data.tmp2[, "ER.status.corr"] == "positive" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD2")),
+   length(which(data.tmp2[, "ER.status.corr"] == "positive" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD3")),
+   length(which(data.tmp2[, "ER.status.corr"] == "positive" &
+   !is.na(data.tmp2[, "ER.status.corr"]) & fact == "MCD4")))
> print(prop.trend.test(ERneg, ERneg + ERpos))

      Chi-squared Test for Trend in Proportions

data:  ERneg out of ERneg + ERpos ,
      using scores: 1 2 3 4
X-squared = 5.4391, df = 1, p-value = 0.01969

> NodePos <- c(length(which(data.tmp2[, "Node.positive"] == "yes" &
+   !is.na(data.tmp2[, "Node.positive"]) & fact == "MCD1")),
+   length(which(data.tmp2[, "Node.positive"] == "yes" & !is.na(data.tmp2[,
+   "Node.positive"]) & fact == "MCD2")), length(which(data.tmp2[,
+   "Node.positive"] == "yes" & !is.na(data.tmp2[, "Node.positive"]) &
+   fact == "MCD3")), length(which(data.tmp2[, "Node.positive"] ==
+   "yes" & !is.na(data.tmp2[, "Node.positive"]) & fact ==
+   "MCD4")))
> NodeNEg <- c(length(which(data.tmp2[, "Node.positive"] == "no" &
+   !is.na(data.tmp2[, "Node.positive"]) & fact == "MCD1")),
+   length(which(data.tmp2[, "Node.positive"] == "no" & !is.na(data.tmp2[,
+   "Node.positive"]) & fact == "MCD2")), length(which(data.tmp2[,
+   "Node.positive"] == "no" & !is.na(data.tmp2[, "Node.positive"]) &
+   fact == "MCD3")), length(which(data.tmp2[, "Node.positive"] ==
+   "no" & !is.na(data.tmp2[, "Node.positive"]) & fact ==
+   "MCD4")))
> print(prop.trend.test(NodeNEg, NodeNEg + NodePos))

```

Chi-squared Test for Trend in Proportions

```
data: NodeNEg out of NodeNEg + NodePos ,
      using scores: 1 2 3 4
X-squared = 0.1644, df = 1, p-value = 0.6851
```

```
> Size.high <- c(length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) >=
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD1")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) >=
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD2")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) >=
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD3")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) >=
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD4")))
> Size.low <- c(length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) <
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD1")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) <
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD2")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) <
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD3")), length(which(as.numeric(data.tmp2[, "Size.in.mm.Corrected"]) <
+ 20 & !is.na(data.tmp2[, "Size.in.mm.Corrected"]) & fact ==
+ "MCD4")))
> print(prop.trend.test(Size.high, Size.high + Size.low))
```

Chi-squared Test for Trend in Proportions

```
data: Size.high out of Size.high + Size.low ,
      using scores: 1 2 3 4
X-squared = 0.2184, df = 1, p-value = 0.6403
```

```
> rawp <- c(prop.trend.test(age.high, age.high + age.low)$p.value,
+ prop.trend.test(grade1, grade1 + grade2 + grade3)$p.value,
+ prop.trend.test(grade2, grade1 + grade2 + grade3)$p.value,
+ prop.trend.test(grade3, grade1 + grade2 + grade3)$p.value,
+ prop.trend.test(ERneg, ERneg + ERpos)$p.value, prop.trend.test(NodeNEg,
+ NodeNEg + NodePos)$p.value, prop.trend.test(Size.high,
+ Size.high + Size.low)$p.value)
> names(rawp) <- c("Age", "Grade1", "Grade2", "Grade3", "ERStatus",
+ "Node", "Size")
> library(multtest)
> mt.rawp2adjp(rawp, proc = "Holm")
```

```
$adjp
      rawp      Holm
[1,] 0.0009927051 0.006948936
[2,] 0.0196910241 0.118146145
[3,] 0.0256614164 0.128307082
[4,] 0.0968273491 0.387309396
[5,] 0.1173604887 0.387309396
[6,] 0.6402752377 1.000000000
[7,] 0.6851465850 1.000000000
```

```
$index
```

[1] 4 5 2 1 3 7 6

\$h0.ABH
NULL

\$h0.TSBH
NULL