

Chromosomal Instability Determines Taxane Response

Swanton et al.

April 20, 2009

This R script generates Figure 4B, which features box plots indicating the distribution of three gene expression signatures in a cohort of breast tumors.

1 The tumor data

```
> library(Biobase)

> load("habermann.RData")
> str(exprs(habermann))

num [1:7657, 1:44] -0.997 -0.239 2.547 0.336 -1.205 ...
- attr(*, "dimnames")=List of 2
..$ : chr [1:7657] "x953484" "x1805089" "x2240502" "x3357825" ...
..$ : chr [1:44] "JD01" "JD02_14" "JD03" "JD05" ...
```

These tumor samples were classified as diploid genome stable (dGS), aneuploid genome stable (aGS), or aneuploid genome unstable (aGU).

```
> ploidy <- factor(habermann$Ploidie, levels = c("dGS", "aGS",
+       "aGU"))
> table(ploidy)
```

```
ploidy
dGS aGS aGU
 14  14  16
```

We are most interested in the comparison between broad-peak (genome unstable) and narrow-peak (genome stable) tumors.

```
> gu <- factor(ifelse(ploidy == "aGU", "GU", "GS"))
> table(gu)
```

```
gu
GS GU
28 16
```

2 Gene signatures

```
> cin70 <- c("TPX2", "PRC1", "FOXN1", "CDC2", "C20orf24", "TGIF2",
+ "MCM2", "H2AFZ", "TOP2A", "PCNA", "UBE2C", "MELK", "TRIP13",
+ "CNAP1", "MCM7", "RNASEH2A", "RAD51AP1", "KIF20A", "CDC45L",
+ "MAD2L1", "ESPL1", "CCNB2", "FEN1", "TTK", "CCT5", "RFC4",
+ "ATAD2", "ch-TOG", "NUP205", "ELAVL1", "CDC20", "CCNB1",
+ "RRM1", "AURKB", "MSH6", "EZH2", "CTPS", "DKC1", "OIP5",
+ "CDCA8", "PTTG1", "CEP55", "H2AFX", "CMAS", "BRRN1", "NCAPH",
+ "MCM10", "LSM4", "NCAPG2", "ASF1B", "ZWINT", "TOPK", "FLJ10036",
+ "CDCA3", "ECT2", "CDC6", "UNG", "MTCH2", "RAD21", "ACTL6A",
+ "GPI", "SFRS2", "HDGF", "NXT1", "NEK2", "DHCR7", "NDUFAB1",
+ "KIAA0286", "KIF4A")
> cin27WP <- c("CDC2", "C20orf24", "H2AFZ", "MCM7", "RNASEH2A",
+ "KIF20A", "CCT5", "NUP205", "ELAVL1", "MSH6", "CTPS", "OIP5",
+ "CMAS", "NCAPH", "MCM10", "LSM4", "NCAPG2", "ZWILCH", "MTCH2",
+ "ACTL6A", "GPI", "SFRS2", "HDGF", "NXT1", "DHCR7", "NDUFAB1",
+ "KIF4A")
> cinsurvival <- c("BRCA1", "CCT4", "CDC2", "CDC6", "CHEK1", "DCLRE1A",
+ "FDFT1", "GART", "GPI", "H2AFX", "IMPA2", "MCM3", "MSH6",
+ "NUP153", "NUP155", "NUP205", "RCN2", "RFC3", "RPA1", "SAE2",
+ "TOP2A", "TOPBP1", "XP01")
```

3 Prepare data

Calculate signature scores (simply the normalized mean expression of all genes in the signature):

```
> sigs <- list(CIN70 = cin70, CIN27WP = cin27WP, `CIN-survival` = cinsurvival)
> sigscores <- lapply(sigs, function(x) {
+   wh <- featureData(habermann)$symb2 %in% x
+   m <- exprs(habermann)[wh, , drop = FALSE]
+   colMeans(m - rowMeans(m))
+ })
> str(sigscores)
```

List of 3

```
$ CIN70      : Named num [1:44] 0.35752 0.00387 -0.01799 0.18988 -0.06362 ...
..- attr(*, "names")= chr [1:44] "JD01" "JD02_14" "JD03" "JD05" ...
$ CIN27WP    : Named num [1:44] 0.4764 0.0865 -0.1268 0.1101 -0.2515 ...
..- attr(*, "names")= chr [1:44] "JD01" "JD02_14" "JD03" "JD05" ...
$ CIN-survival: Named num [1:44] 0.3302 -0.0639 -0.0503 0.037 -0.1573 ...
..- attr(*, "names")= chr [1:44] "JD01" "JD02_14" "JD03" "JD05" ...
```

For each signature, perform t-tests between the genome-unstable and the genome-stable groups.

```
> pval <- sapply(sigscores, function(x) {
+   p <- t.test(x ~ gu)$p.value
+   paste("P =", format(p, digits = 2, scientific = FALSE))
+ })
> pval
```

```
      CIN70      CIN27WP CIN-survival
"P = 0.0004"  "P = 0.027" "P = 0.0048"
```

Split sigscores according to ploidy:

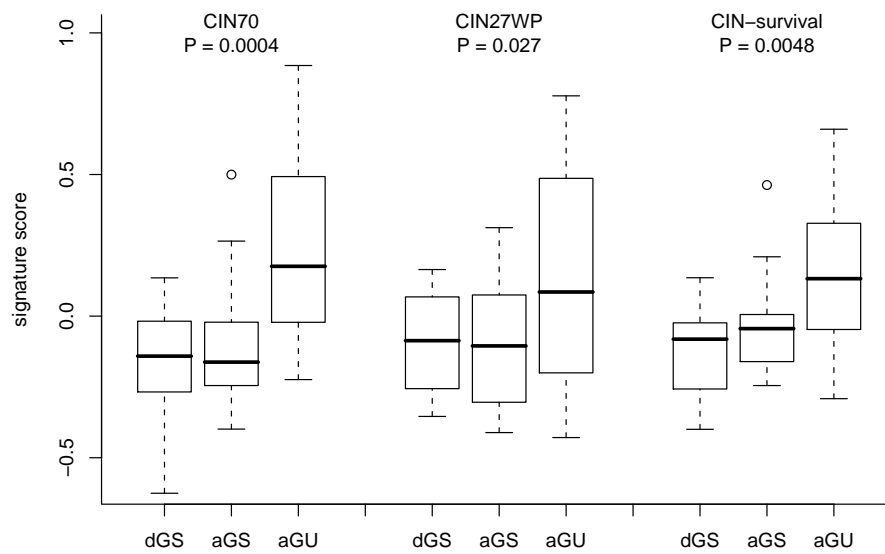
```
> sigscores.ploidy <- lapply(sigscores, split, ploidy)
```

Data for the box-and-whiskers plot (leaving spaces between groups):

```
> boxvals <- c(sigscores.ploidy$CIN70, `` = NA, sigscores.ploidy$CIN27WP,
+   `` = NA, sigscores.ploidy$`CIN-survival`)
```

4 Draw the figure

```
> par(bty = "l", xpd = NA)
> boxplot(boxvals, ylab = "signature score", ylim = c(-0.6, 1))
> for (i in 1:3) {
+   txt <- paste(names(sigs)[i], pval[i], sep = "\n")
+   text((4 * i) - 2, 1, txt)
+ }
```



5 sessionInfo

The results in this file are generated using the following packages:

```
> sessionInfo()
```

```
R version 2.8.1 (2008-12-22)
```

```
i386-apple-darwin8.11.1
```

```
locale:
```

```
C
```

```
attached base packages:
```

```
[1] tools      stats      graphics  grDevices  utils      datasets  methods
```

```
[8] base
```

```
other attached packages:
```

```
[1] Biobase_2.2.1
```