

# Chromosomal Instability Determines Taxane Response

Swanton *et. al.*

April 20, 2009

This R script performs the analysis of the OV01 ovarian cancer trial.

**OV01 trial** In vivo ovarian carcinoma data were obtained from Cambridge Translational Research Ovarian Cancer Study 01 (OV01) (GEO accession GSE15622) . Patients were randomised to receive 3 cycles of either Paclitaxel or Carboplatin. They were classed as sensitive or resistant to their respective treatment based on the change in serum CA125 level using the response criteria defined by Rustin *et. al.* Table below shows the number of samples used in this study. 1 pre-treatment tumour sample suspicious for contamination was excluded from further analysis. 3 patients had post-treatment samples taken following 3 cycles of Paclitaxel and a further 3 cycles of Paclitaxel/Carboplatin and were also excluded from further analysis.

|                       | Pre-treatment  | Post-treatment |
|-----------------------|----------------|----------------|
| Paclitaxel-sensitive  | 13             | 11             |
| Paclitaxel-resistant  | 7 (1 excluded) | 8 (3 excluded) |
| Carboplatin-sensitive | 9              | NA             |
| Carboplatin-resistant | 6              | NA             |

**Loading and pre-processing data** Expression data were acquired on HGU133Av2 Affymetrix arrays and were re-annotated using the Ensembl database (release 46) derived custom CDFs of Dai *et. al.* The data were processed using the RMA methodology without background correction. The annotation package can be downloaded at <http://brainarray.mbni.med.umich.edu/Brainarray/Data>

```
> options(width = 70)
> library(affy)
> library(hs133av2hsensgcdf)
> library(biomaRt)
```

*clinicalInfo* is the list of CEL files and the associated patient information. The code below assumes that all input files reside in a subdirectory "raw\_data". The Bioconductor library *biomaRt* is used to map gene symbols to array annotation.

```
> clinicalInfo <- read.delim("raw_data/OV01_targets.txt")

> data <- ReadAffy(filenamees = clinicalInfo$Target, celfile.path = "raw_data",
+   sampleNames = clinicalInfo$Target)
```

```
> data@cdfName <- "HS133Av2_HS_ENSG"
> data.norm <- expresso(data, normalize.method = "quantiles",
+   bg.correct = FALSE, pmcorrect.method = "pmonly",
+   summary.method = "medianpolish")
```

### CIN70 genes

```
> 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", "ZWILCH", "TOPK", "FLJ10036",
+   "CDCA3", "ECT2", "CDC6", "UNG", "MTCH2", "RAD21",
+   "ACTL6A", "GPI", "SFRS2", "HDGF", "NXT1", "NEK2",
+   "DHCR7", "NDUFAB1", "KIAA0286", "KIF4A")
> ensembl = useMart("ensembl", dataset = "hsapiens_gene_ensembl")
> ann <- getBM(attributes = c("ensembl_gene_id", "hgnc_symbol"),
+   filters = "hgnc_symbol", values = cin70, mart = ensembl)
> ann[, 1] <- paste(ann[, 1], "_at", sep = "")
> ord <- match(ann[, 1], featureNames(data.norm))
> cin70genesExprs <- exprs(data.norm)[ord, ]
> rownames(cin70genesExprs) <- ann[, 2]
> cin70median <- apply(cin70genesExprs, 2, median)
```

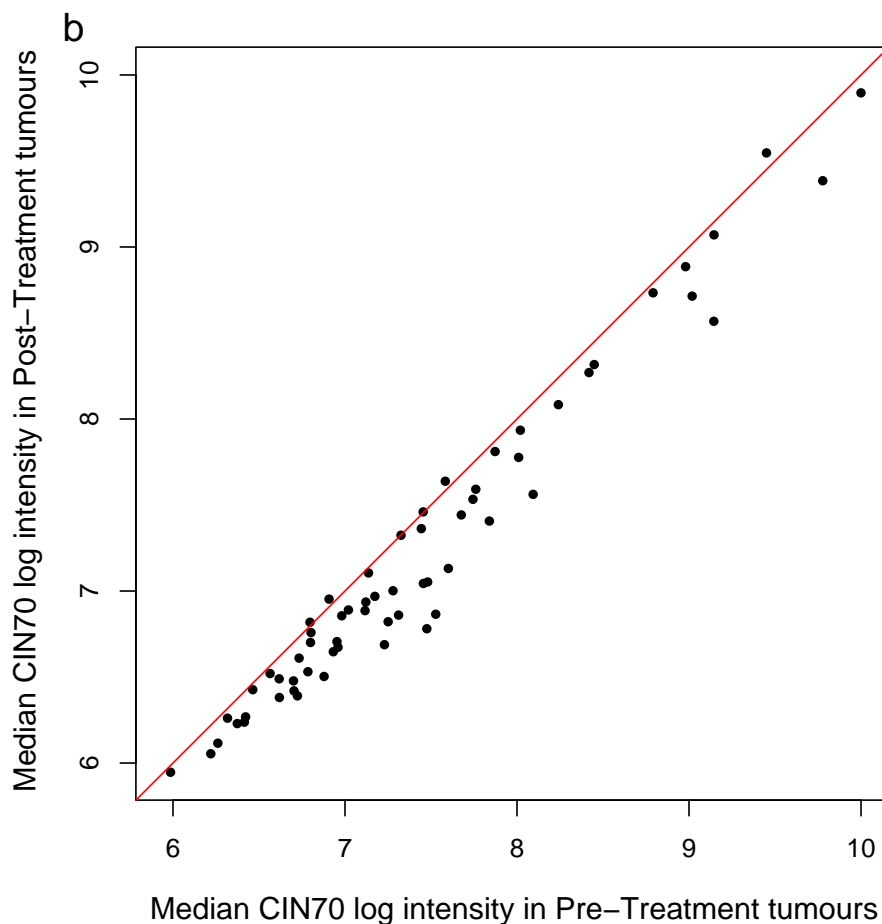
**Figure 1b. Genes over-expressed in tumours with CIN are repressed by MTS treatment in vitro and in vivo** As part of the OV01 clinical trial, expression profiling was performed on ovarian carcinomas before and after 3 cycles of paclitaxel (Px) treatment. For each gene in the CIN70 signature, the median expression amongst patients is compared between the pre-treatment and post-treatment tumours, with the red line indicating equivalence.

```
> par(pch=20, cex.axis=1, cex=1, cex.lab=1.2, mar=c(4.5,4.5,1.5,1.5))
> list1b1 <- list(prePacli=apply(cin70genesExprs[,intersect(
+   grep("pre_pacl",clinicalInfo$Type),
+   which(clinicalInfo$Target!="33a2.CEL"))],1,median),
+   postPacli=apply(cin70genesExprs[,intersect(
+   grep("post_pacl",clinicalInfo$Type),
+   which(clinicalInfo$Drug=="Pacli"))],1,median))
> max <- max(unlist(list1b1))
> min <- min(unlist(list1b1))
> plot(
+   list1b1$prePacli, list1b1$postPacli,
+   xlab="Median CIN70 log intensity in Pre-Treatment tumours",
```

```

+       ylab="Median CIN70 log intensity in Post-Treatment tumours",
+       xlim=c(min,max), ylim=c(min,max))
> abline(a=0,b=1,col="red")
> mtext("b", side=3, line=0, adj=-0.1, cex=1.5)

```



**Figure 4a. Expression of CIN70 genes determines sensitivity to Paclitaxel and Carboplatin** Figure contrasts basal median gene expression for each CIN70 gene amongst tumours with differing response to paclitaxel and carboplatin. Paclitaxel-resistant tumours have higher median log intensity of the CIN70 signature compared to paclitaxel-sensitive tumours ( $p=0.043$ ). CIN70 gene expression differs significantly between tumours subsequently resistant to paclitaxel and tumours resistant to carboplatin ( $p = 0.044$ ). Statistical tests: Student two-sided T-tests.

```

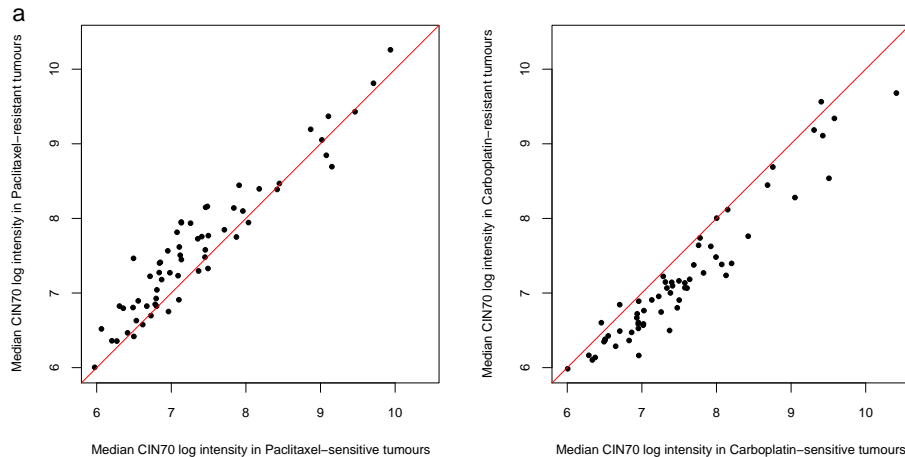
> par(pch=20, cex.axis=0.85, cex=0.85, cex.lab=0.85,
+     mfrow=c(1,2), mar=c(4.5,4.5,1.5,1.5))
> list4a <- list(Psens=apply(cin70genesExprs[,
+     which(clinicalInfo$Type=="pre_paclisens")],1,median),
+     Presist=apply(cin70genesExprs[,
+     which(clinicalInfo$Type=="pre_pacliresist")],
+     [,which(colnames(cin70genesExprs[,

```

```

+       which(clinicalInfo$Type=="pre_pacl_i_resist"))!="33a2.CEL")],1,median),
+       Csens=apply(cin70genesExprs[,
+       which(clinicalInfo$Type=="pre_carbo_sens")],1,median),
+       Cresist=apply(cin70genesExprs[,
+       which(clinicalInfo$Type=="pre_carbo_resist")],1,median))
> max <- max(unlist(list4a))
> min <- min(unlist(list4a))
> plot(
+       list4a$Psens, list4a$Presist,
+       xlim=c(min,max), ylim=c(min,max),
+       xlab="Median CIN70 log intensity in Paclitaxel-sensitive tumours",
+       ylab="Median CIN70 log intensity in Paclitaxel-resistant tumours")
> abline(a=0,b=1,col="red")
> mtext("a", side=3, line=0, adj=-0.2, cex=1.5)
> plot(
+       list4a$Csens, list4a$Cresist,
+       xlim=c(min,max), ylim=c(min,max),
+       xlab="Median CIN70 log intensity in Carboplatin-sensitive tumours",
+       ylab="Median CIN70 log intensity in Carboplatin-resistant tumours")
> abline(a=0,b=1,col="red")

```



**Supplementary Figure 2b-d. Genes over-expressed in tumours with CIN are repressed by MTS treatment in vitro and in vivo** As part of the OV01 clinical trial, expression profiling was performed on ovarian carcinomas before and after 3 cycles of paclitaxel (Px) treatment. (b) To quantify tumor response, an exponential model was fitted for each patient to the CA125 tumour marker over three cycles of treatment. The resulting coefficient is concordant with the response criteria defined by Rustin et. al. (24) and was used to identify the most paclitaxel-sensitive patients (red box). (c) The median expression of the CIN70 genes is decreased in post-treatment tumours relative to pre-treatment tumours, (d) especially in those patients with the most paclitaxel-sensitive disease (boxed in red in (b)). Statistics shown are one-sided t-tests. Horizontal bars indicate mean values.

```

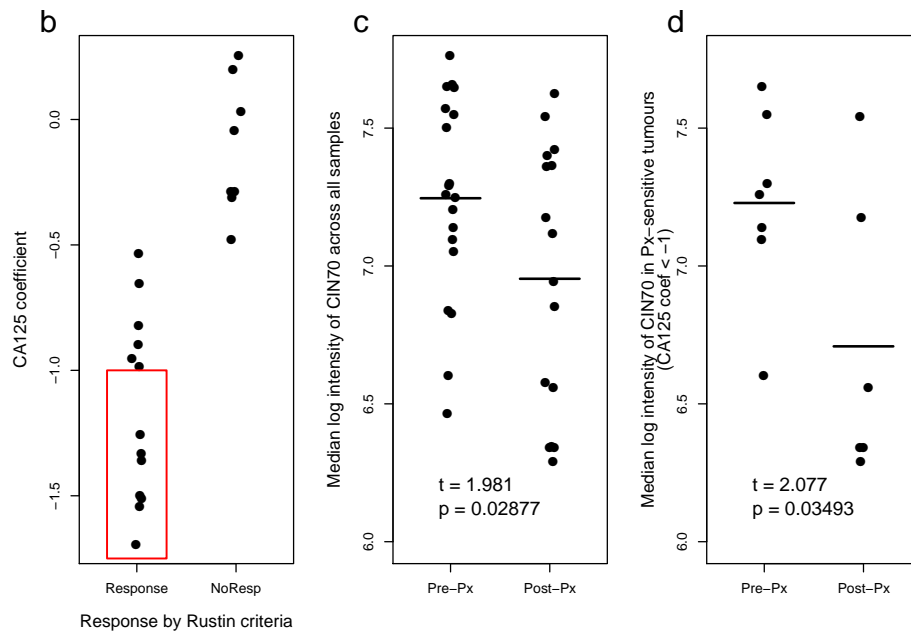
> par(pch=20, cex.axis=1, cex=0.85, cex.lab=1.2, mar=c(4.5,4.5,1.5,1.5), mfrow=c(1,3))
> stripchart(list(Response=unique(
+   clinicalInfo$CA125.coef[grep("pacli_sens",clinicalInfo$Type)]),
+   NoResp=unique(
+   clinicalInfo$CA125.coef[intersect(intersect(
+   grep("pacli_resist",clinicalInfo$Type), which(clinicalInfo$Drug=="Pacli")),
+   which(clinicalInfo$Target!="33a2.CEL"))])),
+   vert=T, xlim=c(0.5,2.5), ylab="CA125 coefficient",
+   xlab="Response by Rustin criteria",
+   pch=16,cex=1.5,method="j", jitter=0.05)
> rect(0.7,-1.75,1.3,-1, border="red", lwd=1.5)
> mtext("b", side=3, line=0, adj=-0.2, cex=1.5)
> lists2c <- list(prePacli=cin70median[intersect(
+   grep("pre_pacli",clinicalInfo$Type),
+   which(clinicalInfo$Target!="33a2.CEL"))],
+   postPacli=cin70median[intersect(
+   grep("post_pacli",clinicalInfo$Type),
+   which(clinicalInfo$Drug=="Pacli"))])
> max <- max(unlist(lists2c))
> min <- min(unlist(lists2c))
> stripchart(list(lists2c$prePacli, lists2c$postPacli),
+   vertical=T,method="j", jitter=0.05, pch=16, cex=1.5,
+   group.names=c("Pre-Px", "Post-Px"),
+   ylab="Median log intensity of CIN70 across all samples",
+   xlim=c(0.5,2.5),ylim=c(min-0.3,max))
> segments(0.7,mean(lists2c$prePacli),
+   1.3,mean(lists2c$postPacli),lwd=2)
> segments(1.7,mean(lists2c$postPacli),
+   2.3,mean(lists2c$postPacli),lwd=2)
> t <- t.test(lists2c$prePacli, lists2c$postPacli,alternative="greater")
> legend(0.6, min, legend=c(paste("t =", format(t$stat,digits=4)),
+   paste("p =",format(t$p.val,digits=4))),
+   bty='n', x.intersp=0.6, cex=1.4)
> mtext("c", side=3, line=0, adj=-0.2, cex=1.5)
> lists2d <- list(prePacli=cin70median[intersect(
+   grep("pre_pacli",clinicalInfo$Type),
+   which(clinicalInfo$CA125.coef < -1))],
+   postPacli=cin70median[intersect(intersect(
+   grep("post_pacli", clinicalInfo$Type),
+   which(clinicalInfo$CA125.coef < -1)),
+   which(clinicalInfo$Drug=="Pacli"))])
> stripchart(list(lists2d$prePacli, lists2d$postPacli),
+   vertical=T,method="j", jitter=0.05, pch=16, cex=1.5,
+   group.names=c("Pre-Px", "Post-Px"),
+   ylab=c("Median log intensity of CIN70 in Px-sensitive tumours",
+   "(CA125 coef < -1)"),xlim=c(0.5,2.5),ylim=c(min-0.3,max))
> segments(0.7,mean(lists2d$prePacli),

```

```

+       1.3,mean(lists2d$prePacli),lwd=2)
> segments(1.7,mean(lists2d$postPacli),
+       2.3,mean(lists2d$postPacli),lwd=2)
> t <- t.test(lists2d$prePacli, lists2d$postPacli,alternative="greater")
> legend(0.6, min, legend=c(paste("t =", format(t$stat,digits=4)),
+       paste("p =",format(t$p.val,digits=4))),
+       bty='n', x.intersp=0.6,cex=1.4)
> mtext("d", side=3, line=0, adj=-0.2, cex=1.5)

```



**Supplementary Figure 5c. Residual CIN expression post-treatment is higher in paclitaxel-resistant compared to paclitaxel-sensitive tumours.** (Top) Sensitivity defined by the Rustin criteria: The distribution of log2 intensities of CIN70 genes in residual tumours following paclitaxel across all Rustin-resistant (black line) and Rustin-sensitive (red line) patients. (Kolmogorov-Smirnov test,  $D = 0.11$ ,  $p = 0.01$ ). (Bottom) Sensitivity defined by the CA125 coefficient: The same analysis was applied to taxane-resistant (black line: CA125 coefficient  $> -0.5$ ) and patients most sensitive to paclitaxel (red line: CA125 coefficient  $< -1$ , see Supplementary Figure 2b) (Kolmogorov-Smirnov test,  $D = 0.23$ ,  $p = 3.5e-8$ ).

```

> par(pch=20, cex.axis=0.85, cex=0.85, cex.lab=1,
+     mfrow=c(2,1), mar=c(4,4,1,1))
> density.sens <- density(cin70genesExprs[,which(clinicalInfo$Type=="post_pacl_i_sens")])
> density.resist <- density(cin70genesExprs[,intersect(which(
+     clinicalInfo$Type=="post_pacl_i_resist"), which(clinicalInfo$Drug=="Pacli"))])
> density.msens <- density(cin70genesExprs[,intersect(
+     grep("post_pacl_i",clinicalInfo$Type),which(clinicalInfo$CA125.coef <= -1))])
> xmax <- max(max(density.sens$x),max(density.resist$x), max(density.msens$x))
> xmin <- min(min(density.sens$x),min(density.resist$x),min(density.msens$x))
> ymax <- max(max(density.sens$y),max(density.resist$y), max(density.msens$y))

```

```

> plot(density.resist, xlab="Median CIN70 log2 intensity in post-treatment samples",
+       ylab="Density",main="", xlim=c(xmin,xmax), ylim=c(0,ymax),lwd=2)
> lines(density.sens, col="red", lwd=2)
> legend(8,0.5,legend=c("Sensitive","Resistant"),col=c("red","black"),lwd=2,bty='n')
> mtext("c", side=3, line=0, adj=-0.2, cex=1.5)
> plot(density.resist, xlab="Median CIN70 log2 intensity in post-treatment samples",
+       ylab="Density",main="", xlim=c(xmin,xmax), ylim=c(0,ymax),lwd=2)
> lines(density.msens, col="red", lwd=2)
> legend(8,0.5,legend=c("Most sensitive","Resistant"),col=c("red","black"),lwd=2,bty='n')
> options(width=70)
> ks.test(cin70genesExprs[,intersect(which(
+       clinicalInfo$Type=="post_pacl_i_resist"), which(clinicalInfo$Drug=="Pacli"))],
+       cin70genesExprs[,which(clinicalInfo$Type=="post_pacl_i_sens")])

```

Two-sample Kolmogorov-Smirnov test

```

data: cin70genesExprs[, intersect(which(clinicalInfo$Type == "post_pacl_i_resist"), and c
D = 0.1114, p-value = 0.01005
alternative hypothesis: two-sided

```

```

> ks.test(cin70genesExprs[,intersect(which(
+       clinicalInfo$Type=="post_pacl_i_resist"), which(clinicalInfo$Drug=="Pacli"))],
+       cin70genesExprs[,intersect(grep("post_pacl_i",clinicalInfo$Type),
+       which(clinicalInfo$CA125.coef <= -1))])

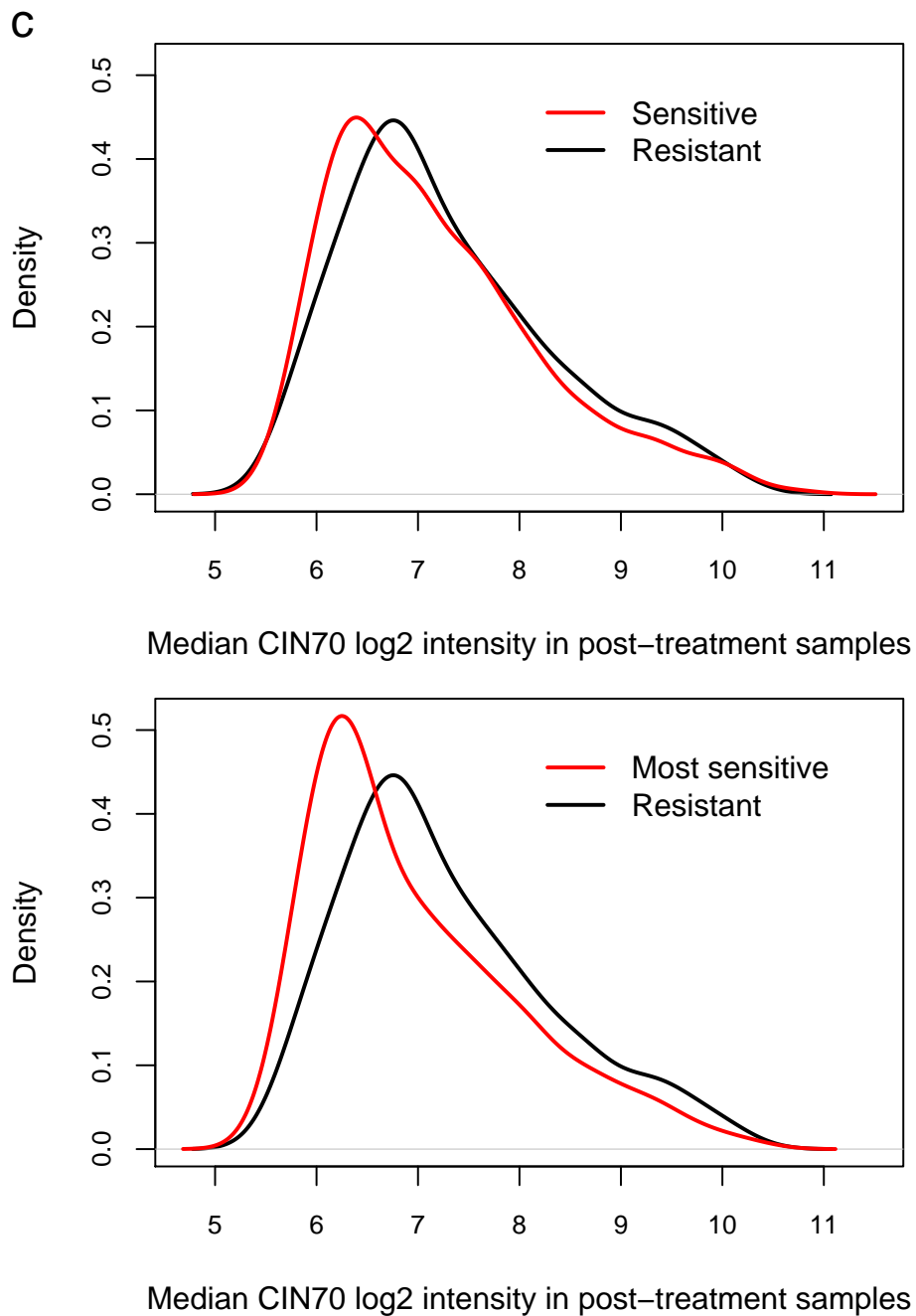
```

Two-sample Kolmogorov-Smirnov test

```

data: cin70genesExprs[, intersect(which(clinicalInfo$Type == "post_pacl_i_resist"), and c
D = 0.2242, p-value = 8.297e-08
alternative hypothesis: two-sided

```



`sessionInfo()` The results in this file are generated using the following packages.

```
> sessionInfo()
```

R version 2.7.1 (2008-06-23)

i386-apple-darwin8.10.1

locale:

en\_GB.UTF-8/en\_GB.UTF-8/C/C/en\_GB.UTF-8/en\_GB.UTF-8



attached base packages:

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

other attached packages:

```
[1] biomaRt_1.14.0      RCurl_0.9-4
[3] hs133av2hsensgcdf_10.0.0 affy_1.18.2
[5] preprocessCore_1.2.1 affyio_1.8.1
[7] Biobase_2.0.1
```

loaded via a namespace (and not attached):

```
[1] XML_1.96-0
```