

Analysis for Supplement Figure 5D

Charles Swanton Barbara Nicke Marion Schuett Aron C. Eklund
Charlotte Ng Qiyuan Li Thomas Hardcastle Alvin Lee Phil East
Maik Kschischo Paul Wylie Se Nyun Kim Jie-Guang Chen
Michael Howell Thomas Ried Jens K. Habermann Gert Auer
James D. Brenton Zoltan Szallasi Julian Downward

April 20, 2009

This is a supplementary document of supplement figure 5D, in the paper "Chromosomal Instability Determines Taxane Response". This document contains technical details of survival analysis based on expression of Taxane-suppressed genes in 3 cancer datasets with untreated Oestrogen receptor positive tumours.

1 Prerequisite for the Analysis

The entire survival analysis is conducted on the R, a language and environment for statistical computing. Several packages are required to process microarray data and generate numerical and graphical output for survival analysis.

```
> library(affydata)
> library(hgu133plus2)
> library(hgu133a)
> library(hgu133b)
> library(hgu95av2)
> library(survival)
> library(aronmisc)
```

2 Customized Functions

We use three customized functions to retrieve gene symbols from merged Affymetrix hgu133a, hgu133b array data, remove fixed block effects and conduct survival analysis. These two functions are defined as below:

```
> getu133symbol<-function(x) {
+   n1<-mget(x, hgu133aSYMBOL, ifnotfound=NA)
+   n2<-mget(x, hgu133bSYMBOL, ifnotfound=NA)
+   n1[is.na(n1) & !is.na(n2)]<-n2[is.na(n1) & !is.na(n2)]
+   return(n1)
+ }
> subgrpmean<-function(v, g) {
+   m <- mean(v, na.rm=T)
+   for (x in levels(as.factor(g[!is.na(g)]))) {
+     v[as.factor(g)==x & !is.na(g)]<-
+       as.vector(scale(v[as.factor(g)==x & !is.na(g)], scale=F))+m
+   }
+ }
```

```

+   v[is.na(g)]<-as.vector(scale(v[is.na(g)], scale=F))+ m
+   return(v)
+ }
> SurvAnal<-function(ETab=NULL, GES=NULL, S=NULL, PLOT=TRUE,
+                     title="Unknow Author",
+                     type="Unknown Surviving",details=FALSE) {
+   ETab <- ETab[rownames(ETab) %in% GES, ]
+   ETab<-ETab[order(apply(ETab, 1, function(x)
+     var(x, na.rm=T)),decreasing=T),]
+   ETab <- ETab[!duplicated(rownames(ETab)),]
+   logmean<-apply(ETab,1,function(x) mean(x, na.rm=T))
+   C<-apply(ETab,2, function(x)
+     ifelse(mean(x, na.rm=T)>mean(logmean),
+       "Expression", "Suppression"))
+   S<-data.frame(time=S[,1], event=S[,2], CIN_survival_genes=C)
+   #S <- S[which(S$time <= 10),]
+   C<-ifelse(C==1, "Expression", "Suppression")
+   score <- summary(coxph(Surv(time, event)~CIN_survival_genes, data=S))
+   if (PLOT) {
+     survplot(Surv(time, event)~CIN_survival_genes, data=S, col=grey(0:2 /3),
+       main=title, lwd=3, cex=1.8,
+       show.n.risk = FALSE,
+       xlab="Time (Years)",
+       ylab=paste("Fraction", type))
+   }
+   if (details) {
+     Detail<-data.frame(data=title,
+       HRratio=score$conf.int[1],
+       HR.95low=score$conf.int[3],
+       HR.95upper=score$conf.int[4],
+       logrankP=score$sctest[3])
+     rownames(Detail)<-NULL
+     Detail
+   }
+ }

```

3 Genes

22 CIN survival genes were obtained from Charles Swanton.

```

> GES<-as.matrix(read.table(
+   "/home/projects/qiyuan/R/slide/charles/CIN_survival.txt"))
> sort(GES)

```

[1]	"BRCA1"	"CDC2"	"CDC6"	"CHEK1"	"DCLRE1A"	"FDFT1"	"GART"
[8]	"GART"	"GPI"	"H2AFX"	"IMPA2"	"MCM3"	"MSH6"	"NUP153"
[15]	"NUP155"	"NUP205"	"RCN2"	"RFC3"	"RPA1"	"SAE2"	"TOP2A"
[22]	"TOPBP1"	"XP01"					

4 Survival Analysis for Individual Dataset

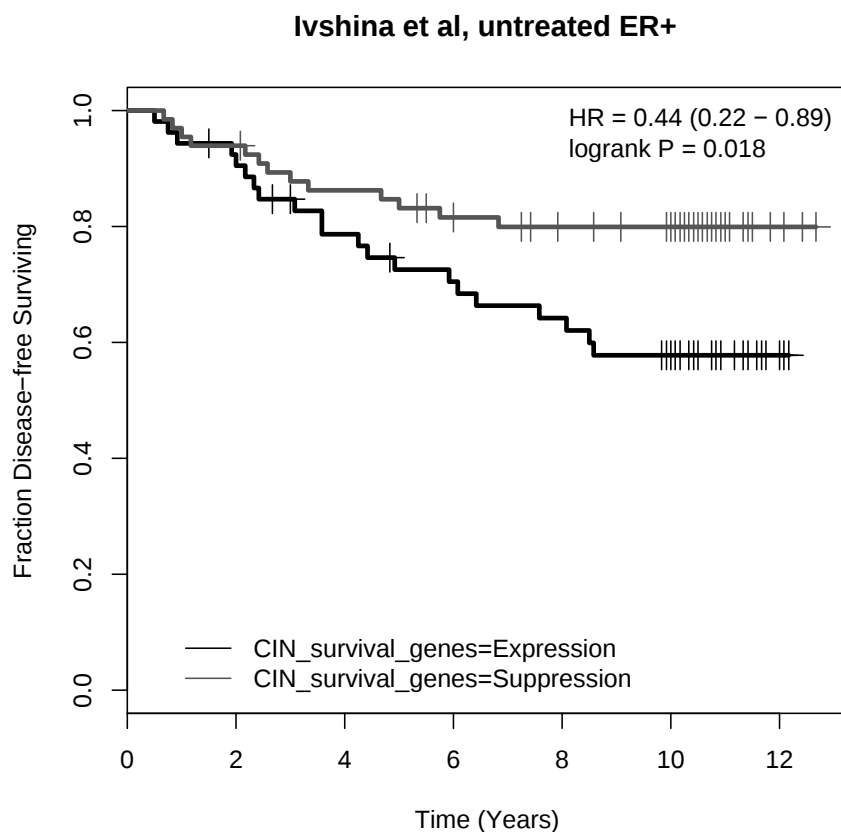
Ivshina et al data was processed by RMA.

```

> load("/home/projects/qiyuan/data/gse4922.rma.RData")
> ETab<-exprs(gse4922.rma)
> rownames(ETab) <- as.vector(unlist(getu133symbol(rownames(ETab))))
> Ivshina.erpath <- ifelse(pData(gse4922.rma)[,13]=="ER+", 1, 0)
> Ivshina.erpath[pData(gse4922.rma)[,13]=="ER?"] <- NA
> Ivshina.erpath[pData(gse4922.rma)[,13]==""] <- NA
> Ivshina.treat <- ifelse(pData(gse4922.rma)[,12]==1, 0, 1)
> S<-pData(gse4922.rma)[,8:9]
> SurvAnal(ETab=ETab[,which(Ivshina.erpath==1 & Ivshina.treat==0)],
+         GES=GES,
+         S=S[which(Ivshina.erpath==1 & Ivshina.treat==0),], PLOT=T,
+         title="Ivshina et al, untreated ER+",
+         type="Disease-free Surviving", details=T)

```

	data	HRatio	HR.95low	HR.95upper	logrankP
1	Ivshina et al, untreated ER+	0.4442181	0.2223074	0.8876438	0.01824392



Wang et al data was processed by RMA.

```

> load("/home/projects/qiyuan/data/wang.rma.Rd")
> ETab<-exprs(wang.rma)
> rownames(ETab) <- as.vector(unlist(mget(rownames(ETab), hgu133aSYMBOL)))
> Wang.erpath <- ifelse(pData(wang.rma)[,7]=="pos", 1, 0)
> S<-data.frame(pData(wang.rma)[,5]/12, pData(wang.rma)[,6])
> SurvAnal(ETab=ETab[,which(Wang.erpath==1)], GES=GES, S=S[which(Wang.erpath==1),], PLOT=T,

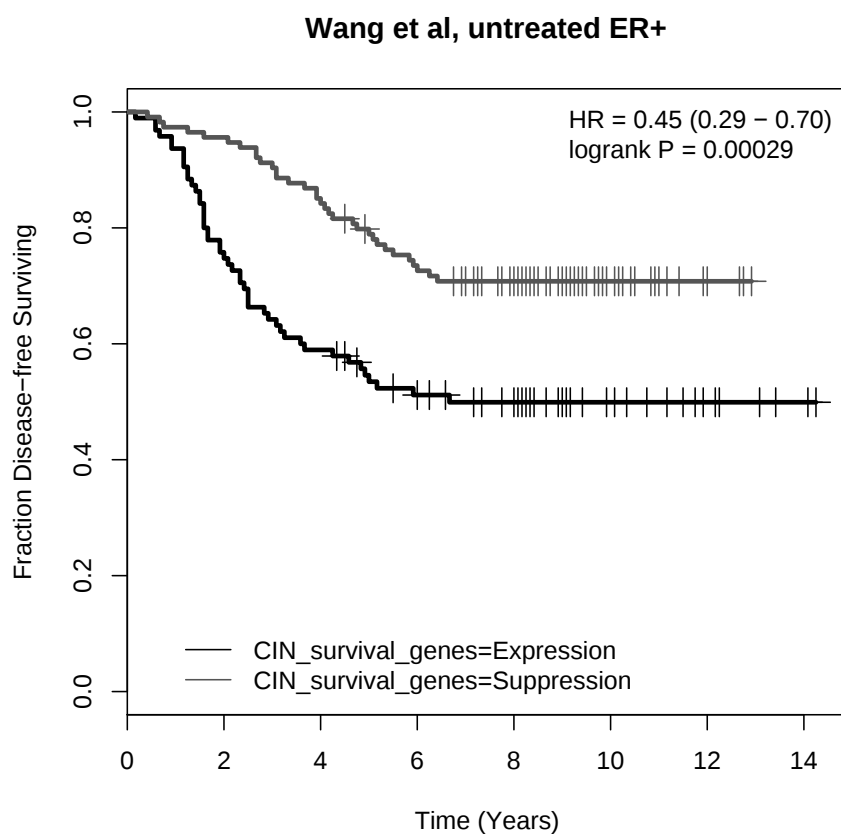
```

```

+         title="Wang et al, untreated ER+",
+         type="Disease-free Surviving", details=T)

      data      HRatio  HR.95low HR.95upper      logrankP
1 Wang et al, untreated ER+ 0.4476984 0.2865461    0.699482 0.0002898918

```



van de Vijver et al data was from previous study [1], effects of treatment (chemotherapy, hormone therapy or both) are removed.

```

> annotate <- as.matrix(read.delim(
+   "/home/projects/qiyuan/Bayesian/data/Vandevijver/RAW_DATA/dataset_annot.csv",
+   sep = "\t"))
> ETab <- read.delim(
+   "/home/projects/qiyuan/Bayesian/data/Vandevijver/RAW_DATA/Vandevijver_Breast_ACE-ncbi35",
+   sep = "\t")
> ETab <- ETab[, order(colnames(ETab))]
> annotate <- annotate[order(annotate[, 1]), ]
> vdv.erpath <- ifelse(annotate[,20]=="Positive", 1, 0)
> vdv.treat <- ifelse(annotate[,23]=="No" & annotate[,24]=="No", 0, 1)
> ETab<-ETab[rownames(ETab) %in% GES,]
> ETab <- t(apply(ETab, 1, function(x)
+   subgrpmean(x, as.factor(annotate[,24])))) )
> ETab <- t(apply(ETab, 1, function(x)
+   subgrpmean(x, as.factor(annotate[,23])))) )
> S<-data.frame(as.numeric(annotate[,7])/12, as.numeric(annotate[,5]))

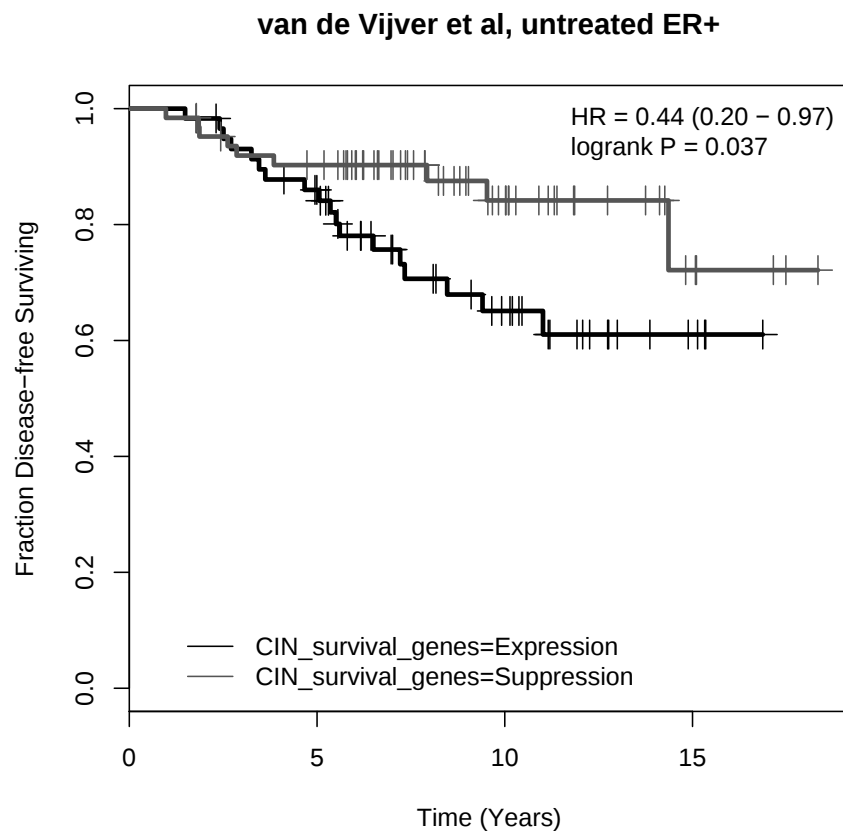
```

```

> SurvAnal(ETab=ETab[,which(vdv.erpath==1 & vdv.treat==0)],
+          GES=GES,
+          S=S[which(vdv.erpath==1 & vdv.treat==0),], PLOT=T,
+          title="van de Vijver et al, untreated ER+",
+          type="Disease-free Surviving", details=T)

      data      HRatio  HR.95low HR.95upper  logrankP
1 van de Vijver et al, untreated ER+ 0.4363525 0.1958818 0.9720328 0.03682814

```



5 Reference

- [1] Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet.* 2006 Sep;38(9):1043-8