GENE EXPRESSION ANALYSIS

BACKGROUND FUNCTIONS

```
read.pcl <- function(filename,na.type = "",make.rownames = TRUE,Nrows=
-1,Comment.char="",...) {
     x.df <- read.table(paste(filename,"pcl",sep="."),header=TRUE,sep="\t",
     quote="\"",as.is = rep(T,3),
na.strings=na.type,skip=0,nrows=Nrows,comment.char=Comment.char,...);
     v <- x.df[[1]];
     ifelse(make.rownames,rownames(x.df)<-v,rownames(x.df)<-(1:{length(v)}));
return(x.df)};

write.pcl <- function(df,dataname,fileaddress="") {
      dir.address <- paste(fileaddress,dataname,".pcl",sep="");
      X <- write.table(df,file=dir.address,

append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.names=FALSE,col.names=TR
UE);
return(X)};

l1vec <- function(vec, na.rm = TRUE) {junk1 <- sum(abs(vec), na.rm =
na.rm);return(junk1)};
```

MODIFIED SURVIVAL FUNCTIONS:

```
# group = FACTOR of the different classes of patients (for distinct KM curves
# survival.time & censoring.status = vectors
# survival.type = string for the title e.g. "metastasis" or "death" etc...
# p-value is obtained using log-rank test


m.plotsurvival <- function (group, survival.time, censoring.status,mark.time
= TRUE, my.colors,my.title = "",time.units = "",survival.type =
"",class.names = as.character(1:length(unique(group))),pv.legend = 0.25,
groups.legend = 0.2, x.legend = 0.1)
{
    require(survival)
    n.class <- length(unique(group))
    junk <- survfit(Surv(survival.time, censoring.status) ~ as.factor(group))
    junk2 <- coxph(Surv(survival.time, censoring.status) ~ as.factor(group))
    pv <- 1 - pchisq(2 * (junk2$loglik[2] - junk2$loglik[1]),
        df = n.class - 1)
    plot(junk, mark.time = mark.time, col = my.colors, xlab =
paste("Time",time.units), ylab = paste("Probability of
survival",survival.type),main = paste("KM survival",my.title))
    legend(x.legend * max(survival.time, na.rm = TRUE), groups.legend, col =
my.colors,
```

```
            lty = rep(1, n.class), legend = class.names, bty = "n")
    text(x.legend * max(survival.time, na.rm = TRUE), pv.legend,
paste("pvalue=", as.character(round(pv,
        4))))
    return()
}

m.PVAL.survival <- function (group, survival.time, censoring.status)
{
    require(survival)
    n.class <- length(unique(group))
    junk2 <- coxph(Surv(survival.time, censoring.status) ~ as.factor(group))
    pv <- 1 - pchisq(2 * (junk2$loglik[2] - junk2$loglik[1]), df = n.class -
1)
    return(pv)
}

m.coxphCOEF.survival <- function (group, survival.time, censoring.status)
{
    require(survival)
    n.class <- length(unique(group))
    junk2 <- coxph(Surv(survival.time, censoring.status) ~ as.factor(group))
      junk3 <- junk2$coef;
    return(junk3)
}

sign0.mat <- function(mat, thr){junk0 <- mat; junk0[abs(junk0) < thr] <- NA;
junk1 = junk0>0; junk2 = 2*junk1 - 1; junk2[is.na(junk2)] <- 0;
return(junk2)};
```

```
# NKI ANALYSIS ON DSGA TRANSFORMED DATA.
```

```
# PERFORM THE DSGA ANALYSIS, ON THE BREAST CANCER DATA
(NKI tumor and BCN normal) AND PLACE THE OUTPUT IN A
DIRECTORY CALLED:
#             work/DSGA.NKI
#    MAKE "work" the working directory.
```

```
# read non-DSGA-transformed gene expression data for
NKI
torg.pcl <- read.pcl("DSGA.NKI/NKI.normMesh");
norg.pcl <- read.pcl("DSGA.NKI/BCN.normMesh");

torg.mat <- as.matrix(torg.pcl[-(1:3)][-1,]);
norg.mat <- as.matrix(norg.pcl[-(1:3)][-1,]);
```

```
tumor.names <- colnames(torg.mat);

# read DSGA-transformed gene expression data for NKI
tdc.pcl <- read.pcl("DSGA.NKI/NKI.Tdis");
ndc.pcl <- read.pcl("DSGA.NKI/BCN.L1out");
tnc.pcl <- read.pcl("DSGA.NKI/NKI.Tnorm");
nnc.pcl <- read.pcl("DSGA.NKI/BCN.Nnorm");


tdc.mat <- as.matrix(tdc.pcl[-(1:3)][-1,]);
ndc.mat <- as.matrix(ndc.pcl[-(1:3)][-1,]);
tnc.mat <- as.matrix(tnc.pcl[-(1:3)][-1,]);
nnc.mat <- as.matrix(nnc.pcl[-(1:3)][-1,]);


# read basement-membrane (BM) gene list with Unigene
cluster ID from current build and with NKI build 219
bm.df <-
read.table("List_2_BM_protein.modifying_enzymes_list.tx
t", header=TRUE, sep="\t", quote="\"", as.is =
rep(T,3), na.strings=c("NA", "na", ""), skip=0);

v.hs <- as.vector(bm.df[["Hs.ID_219.NKI"]], mode =
"character");
bm.all.df <- bm.df[!is.na(v.hs),];
bm.true.df <- bm.all.df[bm.all.df[["BMgene.TF"]],];

id.vec <- rownames(tdc.mat);

# identify id's for ESR1 and ERBB2 for tumor placement
into molecular subtypes

esr1.id <- "Hs.208124"
erbb2.id <- "Hs.446352";

# tumor subtypes (ER-status and ERBB2 or her2 status)
are based on DSGA-transformed data: Disease component
for nki
esr.tdc.vec <- tdc.mat[esr1.id,];
```

```
erbb2.tdc.vec <- tdc.mat[erbb2.id,];

# NKI stratification based on ESR1 status and her2
status in Disease component

hist(esr.tdc.vec, breaks=100, col="red", border="red3",
main="Histogram of all tumors ESR1\nNKI Disease
component", xlab="ESR1 Dc all tumors");
abline(v = -4, lty = "dashed", col = "red4");
legend(x = -4.4, y = 10, legend = "-4", bty = "n");
legend(x = -6.5, y = 9, legend = "ER negative", bty =
"n");
legend(x = -4.3, y = 9, legend = "ER positive", bty =
"n");

esr1.cut = -4

hist(erbb2.tdc.vec, breaks=100, col="purple",
border="purple3", main="Histogram of all tumors
her2\nNKI Disease component", xlab="her2 Dc all
tumors");
abline(v = 0.5, lty = "dashed", col = "purple4");
legend(x = 0, y = 19.5, legend = "0.5", bty = "n");
legend(x = -2.5, y = 17, legend = "her2 negative", bty
= "n");
legend(x = 0.5, y = 17, legend = "her2 positive", bty =
"n");

her2.cut = 0.5
```

```
# tumor stratification based on Disease component of
DSGA-transformed NKI data

basal.names <- tumor.names[{esr.tdc.vec < esr1.cut} &
{erbb2.tdc.vec < her2.cut}];
length(basal.names)
```

```
#[1] 49

her2.names <- tumor.names[{erbb2.tdc.vec > her2.cut}];
length(her2.names)
#[1] 52

luminal.A.names <- tumor.names[{esr.tdc.vec > esr1.cut}
& {erbb2.tdc.vec < her2.cut}];
length(luminal.A.names)
#[1] 194

luminal.B.names <- tumor.names[{esr.tdc.vec > esr1.cut}
& {erbb2.tdc.vec > her2.cut}];
length(luminal.B.names)
#[1] 30

her2.erN.names <- tumor.names[{esr.tdc.vec < esr1.cut}
& {erbb2.tdc.vec > her2.cut}];
length(her2.erN.names)
#[1] 22

erP.names <- tumor.names[{esr.tdc.vec > esr1.cut}];
length(erP.names)
#[1] 224

erN.names <- tumor.names[{esr.tdc.vec < esr1.cut}];
length(erN.names)
#[1] 71
```

```
# extract Basement Membrane genes only

rownames(bm.true.df) <-
as.vector(bm.true.df[["Hs.ID_219.NKI"]], mode =
"character");
V.hs <- rownames(bm.true.df);
V.gene <- as.vector(bm.true.df[["gene.symbol"]], mode =
```

```
"character");
names(V.hs) <- V.gene;

tdc.T.mat <- tdc.mat[V.hs,];
ndc.T.mat <- ndc.mat[V.hs,];
tnc.T.mat <- tnc.mat[V.hs,];
nnc.T.mat <- nnc.mat[V.hs,];
torg.T.mat <- torg.mat[V.hs,];
norg.T.mat <- norg.mat[V.hs,];
```

```
# separate BM data matrices by breast cancer type:

tdc.basal.mat <- tdc.T.mat[,paste(basal.names,"Dis",sep
= ".")];
tdc.luminal.A.mat <-
tdc.T.mat[,paste(luminal.A.names,"Dis",sep = ".")];
tdc.luminal.B.mat <-
tdc.T.mat[,paste(luminal.B.names,"Dis",sep = ".")];
tdc.her2.mat <- tdc.T.mat[,paste(her2.names,"Dis",sep =
".")];
tdc.her2.erN.mat <-
tdc.T.mat[,paste(her2.erN.names,"Dis",sep = ".")];
tdc.erP.mat <- tdc.T.mat[,paste(erP.names,"Dis",sep =
".")];
tdc.erN.mat <- tdc.T.mat[,paste(erN.names,"Dis",sep =
".")];

tnc.basal.mat <-
tnc.T.mat[,paste(basal.names,"Norm",sep = ".")];
tnc.luminal.A.mat <-
tnc.T.mat[,paste(luminal.A.names,"Norm",sep = ".")];
tnc.luminal.B.mat <-
tnc.T.mat[,paste(luminal.B.names,"Norm",sep = ".")];
tnc.her2.mat <- tnc.T.mat[,paste(her2.names,"Norm",sep
= ".")];
tnc.her2.erN.mat <-
```

```
tnc.T.mat[,paste(her2.erN.names,"Norm",sep = ".")];
tnc.erP.mat <- tnc.T.mat[,paste(erP.names,"Norm",sep =
".")];
tnc.erN.mat <- tnc.T.mat[,paste(erN.names,"Norm",sep =
".")];

torg.basal.mat <- torg.T.mat[,basal.names];
torg.luminal.A.mat <- torg.T.mat[,luminal.A.names];
torg.luminal.B.mat <- torg.T.mat[,luminal.B.names];
torg.her2.mat <- torg.T.mat[,her2.names];
torg.her2.erN.mat <- torg.T.mat[,her2.erN.names];
torg.erP.mat <- torg.T.mat[,erP.names];
torg.erN.mat <- torg.T.mat[,erN.names];
```

```
# read in clinical data

clin.df <-
read.table("NKI.Clinical_Data_Supplement.txt",header=TR
UE,sep="\t",quote="\"");
rownames(clin.df) <- paste("SAMPLE",clin.df[["ID"]],sep
= ".")

surv.time.death <-
as.vector(clin.df[["survival.death."]], mode =
"numeric");
names(surv.time.death) <- rownames(clin.df);
censor.death <- as.vector(clin.df[["event_death"]],
mode = "numeric");
names(censor.death) <- rownames(clin.df);
surv.time.met <-
as.vector(clin.df[["Follow_up_time_or_metastasis"]],
mode = "numeric");
names(surv.time.met) <- rownames(clin.df);
censor.met <- as.vector(clin.df[["event_metastasis"]],
mode = "numeric");
names(censor.met) <- rownames(clin.df);
```

```
list.names <- list(basal.names, luminal.A.names,
luminal.B.names, her2.names, her2.erN.names, erP.names,
erN.names);
names(list.names) <- c("Basal","Lum.A", "Lum.B",
"Her2", "Her2.ESRneg", "ERpos", "ERneg");

add.dis <- function(vec){return(paste(vec, "Dis",sep =
"."))};
add.norm <- function(vec){return(paste(vec, "Norm",sep
= "."))};

list.dc.names <- lapply(list.names,add.dis);
names(list.dc.names) <- paste(names(list.names),"Dc",
sep = ".");

list.nc.names <- lapply(list.names,add.norm);
names(list.nc.names) <- paste(names(list.names),"Nc",
sep = ".");

list.org.names <- list.names;
names(list.org.names) <- names(list.names);
```

```
library(survival)

pval.death.Dc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
pval.death.Nc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
pval.death.Org.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));

coef.death.Dc.mat <- matrix(data = NA, nrow =
```

```
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
coef.death.Nc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
coef.death.Org.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));

pval.met.Dc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
pval.met.Nc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
pval.met.Org.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));

coef.met.Dc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
coef.met.Nc.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));
coef.met.Org.mat <- matrix(data = NA, nrow =
length(V.gene), ncol = 7, dimnames =
list(V.gene,names(list.names)));

for(i in (1:length(V.gene))){
    for(j in (1:7)){
        G.dc.vec <- tdc.T.mat[i,list.dc.names[[j]]];
    q.G.dc.33.67 <- quantile(G.dc.vec, probs = c(0,
0.33,0.67,1));
group3.dc.33.67 <- cut(G.dc.vec, breaks = q.G.dc.33.67,
labels = (1:3), include.lowest = TRUE);
v.G.dc.33.67.tmp <- as.vector(group3.dc.33.67, mode =
"character")
```

```
group.dc.33.67.f <- as.factor(v.G.dc.33.67.tmp[!
{v.G.dc.33.67.tmp %in% c(2)}]);

        G.nc.vec <- tnc.T.mat[i,list.nc.names[[j]]];
    q.G.nc.33.67 <- quantile(G.nc.vec, probs = c(0,
0.33,0.67,1));
group3.nc.33.67 <- cut(G.nc.vec, breaks = q.G.nc.33.67,
labels = (1:3), include.lowest = TRUE);
v.G.nc.33.67.tmp <- as.vector(group3.nc.33.67, mode =
"character")
group.nc.33.67.f <- as.factor(v.G.nc.33.67.tmp[!
{v.G.nc.33.67.tmp %in% c(2)}]);

        G.org.vec <- torg.T.mat[i,list.names[[j]]];
    q.G.org.33.67 <- quantile(G.org.vec, probs = c(0,
0.33,0.67,1));
group3.org.33.67 <- cut(G.org.vec, breaks =
q.G.org.33.67, labels = (1:3), include.lowest = TRUE);
v.G.org.33.67.tmp <- as.vector(group3.org.33.67, mode =
"character")
group.org.33.67.f <- as.factor(v.G.org.33.67.tmp[!
{v.G.org.33.67.tmp %in% c(2)}]);

surv.time.death.G <- surv.time.death[list.names[[j]]];
SURV.time.death.dc_3367 <- surv.time.death.G[!
{v.G.dc.33.67.tmp %in% c(2)}];
SURV.time.death.nc_3367 <- surv.time.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];
SURV.time.death.org_3367 <- surv.time.death.G[!
{v.G.org.33.67.tmp %in% c(2)}];

censor.death.G <- censor.death[list.names[[j]]]
CENSOR.death.dc_3367 <- censor.death.G[!
{v.G.dc.33.67.tmp %in% c(2)}];
CENSOR.death.nc_3367 <- censor.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];
CENSOR.death.org_3367 <- censor.death.G[!
{v.G.org.33.67.tmp %in% c(2)}];
```

```
pval.death.Dc.mat[i,j] <- m.PVAL.survival(group =
group.dc.33.67.f, survival.time =
SURV.time.death.dc_3367, censoring.status =
CENSOR.death.dc_3367)
pval.death.Nc.mat[i,j] <- m.PVAL.survival(group =
group.nc.33.67.f, survival.time =
SURV.time.death.nc_3367, censoring.status =
CENSOR.death.nc_3367)
pval.death.Org.mat[i,j] <- m.PVAL.survival(group =
group.org.33.67.f, survival.time =
SURV.time.death.org_3367, censoring.status =
CENSOR.death.org_3367)

coef.death.Dc.mat[i,j] <- m.coxphCOEF.survival(group =
group.dc.33.67.f, survival.time =
SURV.time.death.dc_3367, censoring.status =
CENSOR.death.dc_3367)
coef.death.Nc.mat[i,j] <- m.coxphCOEF.survival(group =
group.nc.33.67.f, survival.time =
SURV.time.death.nc_3367, censoring.status =
CENSOR.death.nc_3367)
coef.death.Org.mat[i,j] <- m.coxphCOEF.survival(group =
group.org.33.67.f, survival.time =
SURV.time.death.org_3367, censoring.status =
CENSOR.death.org_3367)


surv.time.met.G <- surv.time.met[list.names[[j]]];
SURV.time.met.dc_3367 <- surv.time.met.G[!
{v.G.dc.33.67.tmp %in% c(2)}];
SURV.time.met.nc_3367 <- surv.time.met.G[!
{v.G.nc.33.67.tmp %in% c(2)}];
SURV.time.met.org_3367 <- surv.time.met.G[!
{v.G.org.33.67.tmp %in% c(2)}];

censor.met.G <- censor.met[list.names[[j]]]
CENSOR.met.dc_3367 <- censor.met.G[!{v.G.dc.33.67.tmp
%in% c(2)}];
CENSOR.met.nc_3367 <- censor.met.G[!{v.G.nc.33.67.tmp
```

```
%in% c(2)}];
CENSOR.met.org_3367 <- censor.met.G[!{v.G.org.33.67.tmp
%in% c(2)}];

pval.met.Dc.mat[i,j] <- m.PVAL.survival(group =
group.dc.33.67.f, survival.time =
SURV.time.met.dc_3367, censoring.status =
CENSOR.met.dc_3367)
pval.met.Nc.mat[i,j] <- m.PVAL.survival(group =
group.nc.33.67.f, survival.time =
SURV.time.met.nc_3367, censoring.status =
CENSOR.met.nc_3367)
pval.met.Org.mat[i,j] <- m.PVAL.survival(group =
group.org.33.67.f, survival.time =
SURV.time.met.org_3367, censoring.status =
CENSOR.met.org_3367)

coef.met.Dc.mat[i,j] <- m.coxphCOEF.survival(group =
group.dc.33.67.f, survival.time =
SURV.time.met.dc_3367, censoring.status =
CENSOR.met.dc_3367)
coef.met.Nc.mat[i,j] <- m.coxphCOEF.survival(group =
group.nc.33.67.f, survival.time =
SURV.time.met.nc_3367, censoring.status =
CENSOR.met.nc_3367)
coef.met.Org.mat[i,j] <- m.coxphCOEF.survival(group =
group.org.33.67.f, survival.time =
SURV.time.met.org_3367, censoring.status =
CENSOR.met.org_3367)
}};

# Figure 1b & 1c

# dir.create("extras")

i = 23; # NTN4
```

```
j = 6; # ER positive

        G.nc.vec <- tnc.T.mat[i,list.nc.names[[j]]];
    q.G.nc.33.67 <- quantile(G.nc.vec, probs = c(0,
0.33,0.67,1));
group3.nc.33.67 <- cut(G.nc.vec, breaks = q.G.nc.33.67,
labels = (1:3), include.lowest = TRUE);
v.G.nc.33.67.tmp <- as.vector(group3.nc.33.67, mode =
"character")
group.nc.33.67.f <- as.factor(v.G.nc.33.67.tmp[!
{v.G.nc.33.67.tmp %in% c(2)}]);

surv.time.death.G <- surv.time.death[list.names[[j]]];
SURV.time.death.nc_3367 <- surv.time.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];

censor.death.G <- censor.death[list.names[[j]]]
CENSOR.death.nc_3367 <- censor.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];



m.plotsurvival(group = group.nc.33.67.f, survival.time
= SURV.time.death.nc_3367,
censoring.status = CENSOR.death.nc_3367, my.colors =
c("blue", "red"), survival.type = "death", class.names
= c("low NTN4", "high NTN4"),
my.title = paste(names(list.nc.names)[[j]]," group\nNKI
33% 67% separation of", V.gene[[i]], "levels"));

group.nc.33.67.vec <- as.vector(group.nc.33.67.f, mode
= "character");
ERp.survival.grouplo.NTN4 <-
SURV.time.death.nc_3367[group.nc.33.67.vec %in%
c("1")];
ERp.survival.grouphi.NTN4 <-
SURV.time.death.nc_3367[group.nc.33.67.vec %in%
c("3")];
```

```
ERp.censor.grouplo.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("1")];
ERp.censor.grouphi.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("3")];


km.ERp.df <- data.frame("Sample" =
c(names(ERp.survival.grouplo.NTN4),names(ERp.survival.g
rouphi.NTN4)),
"group" = c(rep("lo.NTN4",
length(ERp.survival.grouplo.NTN4)), rep("hi.NTN4",
length(ERp.survival.grouphi.NTN4))),
"time" = c(ERp.survival.grouplo.NTN4,
ERp.survival.grouphi.NTN4),
"censor" = c(ERp.censor.grouplo.NTN4,
ERp.censor.grouphi.NTN4));

# write.table(km.ERp.df,file="extras/figure1b.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);

j = 7; # ER negative

        G.nc.vec <- tnc.T.mat[i,list.nc.names[[j]]];
    q.G.nc.33.67 <- quantile(G.nc.vec, probs = c(0,
0.33,0.67,1));
group3.nc.33.67 <- cut(G.nc.vec, breaks = q.G.nc.33.67,
labels = (1:3), include.lowest = TRUE);
v.G.nc.33.67.tmp <- as.vector(group3.nc.33.67, mode =
"character")
group.nc.33.67.f <- as.factor(v.G.nc.33.67.tmp[!
{v.G.nc.33.67.tmp %in% c(2)}]);

surv.time.death.G <- surv.time.death[list.names[[j]]];
SURV.time.death.nc_3367 <- surv.time.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];
```

```
censor.death.G <- censor.death[list.names[[j]]]
CENSOR.death.nc_3367 <- censor.death.G[!
{v.G.nc.33.67.tmp %in% c(2)}];

m.plotsurvival(group = group.nc.33.67.f, survival.time
= SURV.time.death.nc_3367,
censoring.status = CENSOR.death.nc_3367, my.colors =
c("blue", "red"), survival.type = "death", class.names
= c("low NTN4", "high NTN4"),
my.title = paste(names(list.nc.names)[[j]]," group\nNKI
33% 67% separation of", V.gene[[i]], "levels"));

group.nc.33.67.vec <- as.vector(group.nc.33.67.f, mode
= "character");
ERn.survival.grouplo.NTN4 <-
SURV.time.death.nc_3367[group.nc.33.67.vec %in%
c("1")];
ERn.survival.grouphi.NTN4 <-
SURV.time.death.nc_3367[group.nc.33.67.vec %in%
c("3")];

ERn.censor.grouplo.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("1")];
ERn.censor.grouphi.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("3")];

km.ERn.df <- data.frame("Sample" =
c(names(ERn.survival.grouplo.NTN4),names(ERn.survival.g
rouphi.NTN4)),
"group" = c(rep("lo.NTN4",
length(ERn.survival.grouplo.NTN4)), rep("hi.NTN4",
length(ERn.survival.grouphi.NTN4))),
"time" = c(ERn.survival.grouplo.NTN4,
ERn.survival.grouphi.NTN4),
"censor" = c(ERn.censor.grouplo.NTN4,
ERn.censor.grouphi.NTN4));
```

```
# write.table(km.ERn.df,file="extras/figure1c.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);
```

```
# compute statistics and consensus statistics

# take sign of the coefficient matrix, with anything
whose abs.val < 0.1 --> 0
sign0.death.Dc.1.mat <- sign0.mat(mat =
coef.death.Dc.mat, thr =  0.01);
sign0.death.Nc.1.mat <- sign0.mat(mat =
coef.death.Nc.mat, thr =  0.01);
sign0.death.Org.1.mat <- sign0.mat(mat =
coef.death.Nc.mat, thr =  0.01);

sign0.met.Dc.1.mat <- sign0.mat(mat = coef.met.Dc.mat,
thr =  0.01);
sign0.met.Nc.1.mat <- sign0.mat(mat = coef.met.Nc.mat,
thr =  0.01);
sign0.met.Org.1.mat <- sign0.mat(mat = coef.met.Nc.mat,
thr =  0.01);
```

```
# log10, then turn 0 anything with pval > threshold

sign0.death.Dc.mat <- sign0.death.Dc.1.mat *
log10(pval.death.Dc.mat);
sign0.death.Nc.mat <- sign0.death.Nc.1.mat *
log10(pval.death.Nc.mat);
sign0.death.Org.mat <- sign0.death.Org.1.mat *
log10(pval.death.Org.mat);

sign0.met.Dc.mat <- sign0.met.Dc.1.mat *
log10(pval.met.Dc.mat);
```

```
sign0.met.Nc.mat <- sign0.met.Nc.1.mat *
log10(pval.met.Nc.mat);
sign0.met.Org.mat <- sign0.met.Org.1.mat *
log10(pval.met.Org.mat);

signed.sep.mean.Nc.vec <- apply(X =
cbind(sign0.death.Nc.mat,sign0.met.Nc.mat),
MARGIN=1,FUN=mean);
signed.sep.mean.Dc.vec <- apply(X =
cbind(sign0.death.Dc.mat,sign0.met.Dc.mat),
MARGIN=1,FUN=mean);
signed.sep.mean.Org.vec <- apply(X =
cbind(sign0.death.Org.mat,sign0.met.Org.mat),
MARGIN=1,FUN=mean);

signed.sep.death.mean.Nc.vec <- apply(X =
sign0.death.Nc.mat, MARGIN=1,FUN=mean);

#o1.separation.vec gives L1-norm of log10 p-values for
all molecular subtypes & death & Disease comp. & Normal
comp & Original data

o1.separation.death.Nc.vec <-
apply(log10(pval.death.Nc.mat), 1, l1vec);
    names(o1.separation.death.Nc.vec) <- V.gene;

o1.separation.death.Dc.vec <-
apply(log10(pval.death.Dc.mat), 1, l1vec);
    names(o1.separation.death.Dc.vec) <- V.gene;

o1.separation.death.Org.vec <-
apply(log10(pval.death.Org.mat), 1, l1vec);
    names(o1.separation.death.Org.vec) <- V.gene;

o1.separation.met.Nc.vec <-
apply(log10(pval.met.Nc.mat), 1, l1vec);
    names(o1.separation.met.Nc.vec) <- V.gene;
```

```
o1.separation.met.Dc.vec <-
apply(log10(pval.met.Dc.mat), 1, l1vec);
    names(o1.separation.met.Dc.vec) <- V.gene;

o1.separation.met.Org.vec <-
apply(log10(pval.met.Org.mat), 1, l1vec);
    names(o1.separation.met.Org.vec) <- V.gene;

o1.separation.Nc.vec <- o1.separation.death.Nc.vec +
o1.separation.met.Nc.vec;

o1.separation.Dc.vec <- o1.separation.death.Dc.vec +
o1.separation.met.Dc.vec;

o1.separation.Org.vec <- o1.separation.death.Org.vec +
o1.separation.met.Org.vec;

# extended data figure 1a

plot(x =
c(signed.sep.mean.Nc.vec,signed.sep.mean.Dc.vec,signed.
sep.mean.Org.vec, rep(-2.8, 3)),
y = c(o1.separation.Nc.vec,
o1.separation.Dc.vec,o1.separation.Org.vec, 10,8,6),
main = "Comparison of overall association (L1) vs.
consensus signed association (mean)\nNormal comp. vs.
Disease comp. vs. Original data, death, met",
xlab = "consensus signed association (mean)", ylab =
"overall association L1", pch =
c(rep(19,35),rep(20,70),19,20,20),
col = c(rep("black",35), rep("red", 35), rep("blue",
35), "black","red", "blue"))


legend(x = -2.8, y = 11, legend = "Normal component",
cex = 0.7,bty = "n");
```

```
legend(x = -2.8, y = 9, legend = "Disease component",
cex = 0.7, bty = "n");
legend(x = -2.8, y = 7, legend = "Original data, no
DSGA", cex = 0.7, bty = "n");
```

```
e_figure1a.df <- data.frame("GENE" =
rownames(e_figure1a.mat),
 "consensus.signed.association" =
c(signed.sep.mean.Nc.vec,signed.sep.mean.Dc.vec,signed.
sep.mean.Org.vec),
 "overall.association.L1" = c(o1.separation.Nc.vec,
o1.separation.Dc.vec,o1.separation.Org.vec));

# write.table(e_figure1a.df,file="extras/
Efigure1a.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);
```

```
# Figure 1a
```

```
signed.sep.Nc.Dc.vec <- apply(X =
cbind(sign0.death.Nc.mat,sign0.death.Dc.mat,sign0.met.N
c.mat,sign0.met.Dc.mat), MARGIN=1,FUN=mean);
o1.separation.Nc.Dc.vec <- o1.separation.death.Nc.vec +
o1.separation.met.Nc.vec + o1.separation.death.Dc.vec +
o1.separation.met.Dc.vec

plot(x = signed.sep.Nc.Dc.vec, y =
o1.separation.Nc.Dc.vec,
main = "Comparison of overall association (L1) vs.
consensus signed association (mean)\ncombined Normal
comp. Disease comp. death, met",
xlab = "consensus signed association (mean)", ylab =
```

```
"overall association L1", pch = 19,
col = "gray");


figure1a.df <- data.frame("GENE" =
names(signed.sep.Nc.Dc.vec),
 "consensus.signed.association_NcDc.met.death" =
signed.sep.Nc.Dc.vec,
 "overall.association.L1_NcDc.met.death" =
o1.separation.Nc.Dc.vec);

# write.table(figure1a.df,file="extras/figure1a.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);
```

```
# COMPUTING CORRELATIONS OF BASEMENT MEMBRANE GENES TO
NTN4 IN BREAST, AND COMPARING TO OVARIAN CANCER
```

```
# compute correlation to NTN4 of all BM genes r.value
and p.value

ntnc.T.mat <- cbind(nnc.T.mat, tnc.T.mat);

# correlation plot between NKI.Nc basement membrane
genes correlation to NTN4 and OC basement membrane
correlation to NTN4

get.cor.p <- function(x,y){j = cor.test(x,y,
alternative = "two.sided", method = "pearson"); CC =
j$estimate; PP = j$p.value;
J <- c(CC,PP); names(J) = c("cor", "p.val");
return(J)};

ntn4.id <- "Hs.201034";
```

```
cor.NTN4.ntnc <- t(apply(ntnc.T.mat,MARGIN = 1, FUN =
get.cor.p, y = ntnc.T.mat[ntn4.id,]));
rownames(cor.NTN4.ntnc) <- names(V.hs);

cor2ntn4.bc.oc.df <-
read.table("SupplementaryTable.5_N31_BM.mechanics.txt",
header=TRUE, sep="\t", quote="\"", na.strings=c("NA"),
skip=0);
rownames(cor2ntn4.bc.oc.df) <-
as.vector(cor2ntn4.bc.oc.df[[1]], mode = "character");

cor2ntn4.bc.oc.mat <- as.matrix(cor2ntn4.bc.oc.df[-
c(1)]);

tf.vec <- {!is.na(cor2ntn4.bc.oc.mat[,1])} & {!
is.na(cor2ntn4.bc.oc.mat[,3])}

small.bc.oc.mat <- cor2ntn4.bc.oc.mat[tf.vec,]

library("GmicR");
library("WGCNA")

# EXTENDED DATA FIGURE 9 d

verboseScatterplot(x = small.bc.oc.mat[,3], y =
small.bc.oc.mat[,1], xlab = "OC cor to NTN4", ylab =
"NKI.Nc cor to NTN4", abline = TRUE,
xlim = range(-1,1), ylim = range(-1,1), pch = 20, col =
"blue", main = "Correlation of BM genes to NTN4 in
\novarian cancer vs. breast cancer normal component\n",
cex.axis = 1, cex.lab = 1, cex.main = 1);

e_figure_9d.df <- data.frame("Gene" =
rownames(small.bc.oc.mat), "cor.NKI.ntNc" =
small.bc.oc.mat[,1], "cor.OC" = small.bc.oc.mat[,3]);
```

```
# write.table(e_figure_9d.df,file="extras/
Efigure9d.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);
```

```
    #. DONE  NKI
```

```
# RENAL CANCER ANALYSIS OF NTN4 ASSOCIATION
WITH SURVIVAL
```

```
# RENAL CANCER ANALYSIS ON DSGA TRANSFORMED DATA.
```

```
# PERFORM THE DSGA ANALYSIS, ON THE RENAL CANCER DATA
(RTum tumor data & RNorm normal data) AND PLACE THE
OUTPUT IN A DIRECTORY CALLED:
#            work/DSGA.Renal
#   MAKE "work" the working directory.
```

```
# read DSGA-transformed gene expression data for NKI
tdc.pcl <- read.pcl("DSGA.Renal/RTum.Tdis");
ndc.pcl <- read.pcl("DSGA.Renal/RNorm.L1out");
tnc.pcl <- read.pcl("DSGA.Renal/RTum.Tnorm");
nnc.pcl <- read.pcl("DSGA.Renal/RNorm.Nnorm");


tnc.mat <- as.matrix(tnc.pcl[-(1:3)]);

tumor.names <- sub( pattern = ".Norm", replacement =
```

```
"", x = colnames(tnc.mat))


NTN4.id <- "IMAGE:143661"

# read in clinical data

clin.df <-
read.table("cRCC.clinical.information.clean.txt",header
=TRUE,sep="\t",quote="\"");
rownames(clin.df) <- as.vector(clin.df[[1]], mode =
"character");
Clin.df <- clin.df[tumor.names,];

surv.time <- as.vector(Clin.df[["SurvivalMonths"]],
mode = "numeric");
names(surv.time) <- rownames(Clin.df);
censor.death <- as.vector(Clin.df[["Event.Death"]],
mode = "numeric");
names(censor.death) <- rownames(Clin.df);
censor.met <- as.vector(Clin.df[["EventRecur"]], mode =
"numeric");
names(censor.met) <- rownames(Clin.df);

library(survival)


        G.nc.vec <- tnc.mat[NTN4.id,];
    q.G.nc.33.67 <- quantile(G.nc.vec, probs = c(0,
0.33,0.67,1));
group3.nc.33.67 <- cut(G.nc.vec, breaks = q.G.nc.33.67,
labels = (1:3), include.lowest = TRUE);
v.G.nc.33.67.tmp <- as.vector(group3.nc.33.67, mode =
"character")
group.nc.33.67.f <- as.factor(v.G.nc.33.67.tmp[!
{v.G.nc.33.67.tmp %in% c(2)}]);

SURV.time_3367 <- surv.time[!{v.G.nc.33.67.tmp %in%
```

```
c(2)}];
CENSOR.death.nc_3367 <- censor.death[!{v.G.nc.33.67.tmp
%in% c(2)}];
CENSOR.met.nc_3367 <- censor.met[!{v.G.nc.33.67.tmp
%in% c(2)}];

m.plotsurvival(group = group.nc.33.67.f, survival.time
= SURV.time_3367,
censoring.status = CENSOR.death.nc_3367, my.colors =
c("blue", "red"), survival.type = "death", class.names
= c("low NTN4", "high NTN4"),
my.title = paste("Renal cancer death\nNKI 33% 67%
separation of", "NTN4 in Normal component", "levels"));


m.plotsurvival(group = group.nc.33.67.f, survival.time
= SURV.time_3367,
censoring.status = CENSOR.met.nc_3367, my.colors =
c("blue", "red"), survival.type = "met", class.names =
c("low NTN4", "high NTN4"),
my.title = paste("Renal cancer metastasis\nNKI 33% 67%
separation of", "NTN4 in Normal component", "levels"));

group.nc.33.67.vec <- as.vector(group.nc.33.67.f, mode
= "character");
Renal.survival.grouplo.NTN4 <-
SURV.time_3367[group.nc.33.67.vec %in% c("1")];
Renal.survival.grouphi.NTN4 <-
SURV.time_3367[group.nc.33.67.vec %in% c("3")];

Renal.censor.death.grouplo.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("1")];
Renal.censor.death.grouphi.NTN4 <-
CENSOR.death.nc_3367[group.nc.33.67.vec %in% c("3")];

Renal.censor.met.grouplo.NTN4 <-
CENSOR.met.nc_3367[group.nc.33.67.vec %in% c("1")];
Renal.censor.met.grouphi.NTN4 <-
```

```
CENSOR.met.nc_3367[group.nc.33.67.vec %in% c("3")];

km.Renal.df <- data.frame("Sample" =
c(names(Renal.survival.grouplo.NTN4),names(Renal.surviv
al.grouphi.NTN4)),
"group" = c(rep("lo.NTN4",
length(Renal.survival.grouplo.NTN4)), rep("hi.NTN4",
length(Renal.survival.grouphi.NTN4))),
"time" = c(Renal.survival.grouplo.NTN4,
Renal.survival.grouphi.NTN4),
"censor.death" = c(Renal.censor.death.grouplo.NTN4,
Renal.censor.death.grouphi.NTN4),
"censor.met" = c(Renal.censor.met.grouplo.NTN4,
Renal.censor.met.grouphi.NTN4));

# write.table(km.Renal.df,file="extras/
Efigure1.bc.txt",
#
append=FALSE,quote=FALSE,sep="\t",eol="\n",na="",row.na
mes=FALSE,col.names=TRUE);
```

```
    #. DONE  RenalCancer
```