

HCC-biomarker Scripts

Qingyan(Lilly) Xie

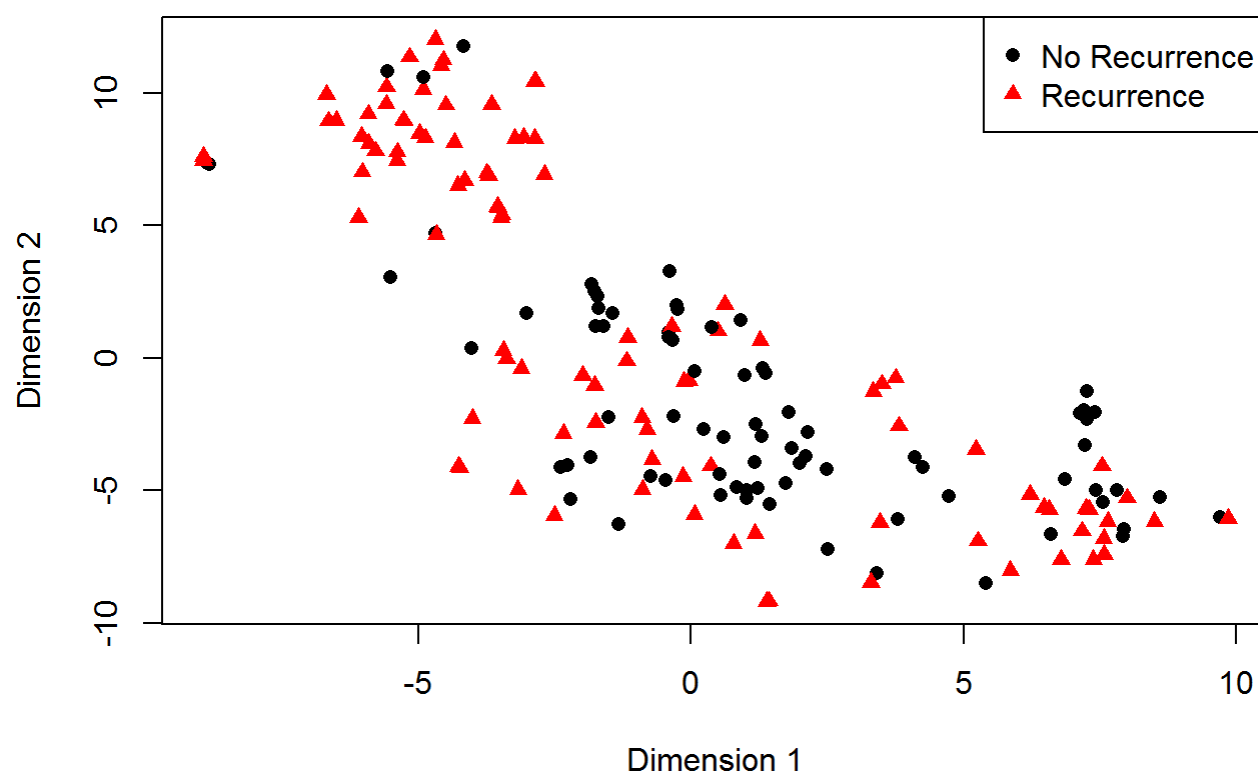
11/19/2015

Data analysis

The t-SNE algorithm is applied to 176 samples with 847 features. Require to load 'spall-dist-ptab.rdata' first. The data frame 'spall' includes all samples info. The colname 14-860 are all miRNA features (raw data)

```
library('Rtsne')
#pdf('Rtsne-176sample-recur.pdf')
load('spall-ptab.rdata')
rtsne_sp<-Rtsne(as.matrix(spall[,14:860]))
plot(rtsne_sp$Y,col=c('black','red')[as.factor(spall$recur)],pch=c(16,17)[as.factor(spall$recur)],main='t-SNE distribution for 176 Samples', xlab='Dimension 1', ylab='Dimension 2')
legend("topright",pch=c(16,17),col=c('black','red'),c('No Recurrence','Recurrence'))
```

t-SNE distribution for 176 Samples



Kmeans clustering

The kmeans clustering(K=2) is applied to 176 samples with 847 features. Require to load 'spall-dist-ptab.rdata' first. The

data frame 'spall' includes all samples info. The colname 14-860 are all miRNA features (raw data)

```
library(survival) # survial package
set.seed(55)
km2<-kmeans(spall[,14:860],2)
rownames(km2$center)<-c('c1','c2')
dist12<-spall[,14:860]
dist12<-rbind(dist12,km2$center)
dist12<-as.matrix(dist(dist12)) # computing the distances between 176 samples and two centers
dist12<-dist12[1:176,177:178] # extract the distances from 176 samples to the two cluster centers
dist12<-as.data.frame(dist12)
dist12$M<-dist12$c2-dist12$c1 # the difference between two distances
dist12$A<-(dist12$c2+dist12$c1)/2 # average of two distances
dist12$kclst<-km2$cluster # cluster assignment
dist12$recur<-spall$recur # clinical feature
dist12$batchid<-spall$batchid # clinical feature
dist12$vasc<-spall$vasc # clinical feature
dist12$focality<-spall$focality # clinical feature
dist12$age<-spall$age # clinical feature
dist12$ntumor<-spall$ntumor # clinical feature
dist12$milan<-spall$milan # clinical feature
dist12$rfsurv<-spall$rfsurv # clinical feature
dist12$pid<-spall$pid
dist12$batchid<-spall$batchid # clinical feature
```

All patients are categorized into three groups:

1. patients with samples all in cluster 1
2. patients with samples all in cluster 2
3. Patients with samples both in cluster 1 and 2

```
p1<-unique(dist12$pid[dist12$kclst==1])
pmix<-which( (dist12$kclst==2) & (dist12$pid %in% p1) )
pmix<-unique(dist12$pid[pmix])
p1<-p1[!p1%in%pmix]
p2<-unique(dist12$pid[dist12$kclst==2])
p2<-p2[!p2%in%pmix]
dist12$pclst<-NA
dist12$pclst[dist12$pid %in% p1]<-1
dist12$pclst[dist12$pid %in% p2]<-2
dist12$pclst[dist12$pid %in% pmix]<-3
```

The patient-level table is created

```
pdist12<-unique(dist12[,c('pid','pclst','rfsurv','recur','focality','ntumor','milan')])
pdist12$miclst<-NA # Milan and cluster features together
pdist12$miclst[which(pdist12$milan=='Outside' & pdist12$pclst==1)]<-'omc1' # outside milan & cluster 1
pdist12$miclst[which(pdist12$milan=='Outside' & pdist12$pclst==2)]<-'omc2' # outside milan & cluster 2
pdist12$miclst[which(pdist12$milan=='Outside' & pdist12$pclst==3)]<-'omc3' # outside milan & mixed
```

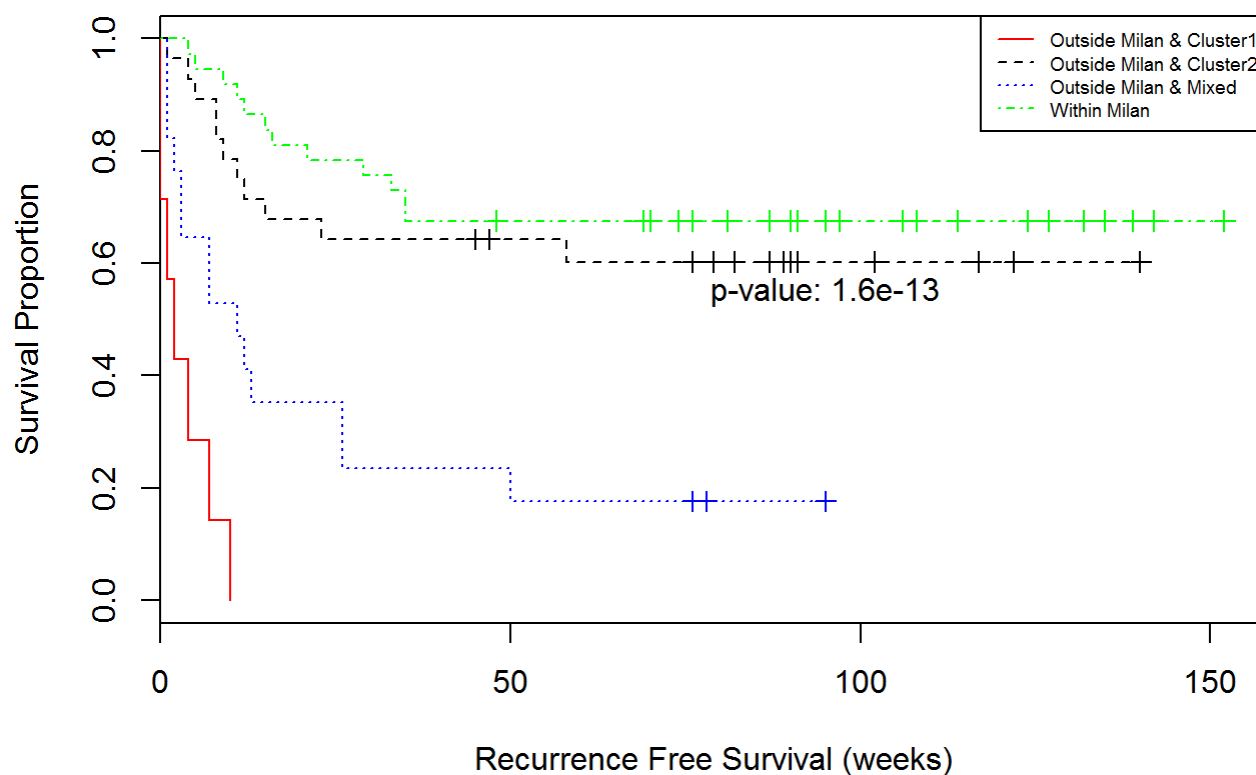
```
pdist12$miclst[which(pdist12$milan=="Within")]<-"wm" # within milan
```

Suverival curves

Plot the patient -level survival curves, using feature miclst as group factor

```
#pdf('Kaplan-meier-milan-cluster-patients.pdf')
fitmclst<-survfit(Surv(pdist12$rfsurv,as.numeric(as.factor(pdist12$recur))-1)~as.factor(pdist12$miclst))
plot(fitmclst,lty=1:4,col=c('red','black','blue','green'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="All Patients (n=89)")
text(x=95, y=0.55, paste0("p-value: ",round(pchisq(survdiff(Surv(pdist12$rfsurv,as.numeric(as.factor(pdist12$recur))-1)~as.factor(pdist12$miclst))$chisq,df=2,lower.tail=F),digits=14)))
legend("topright",cex=0.6,lty=1:4,col=c('red','black','blue','green'),c('Outside Milan & Cluster1','Outside Milan & Cluster2','Outside Milan & Mixed','Within Milan'))
```

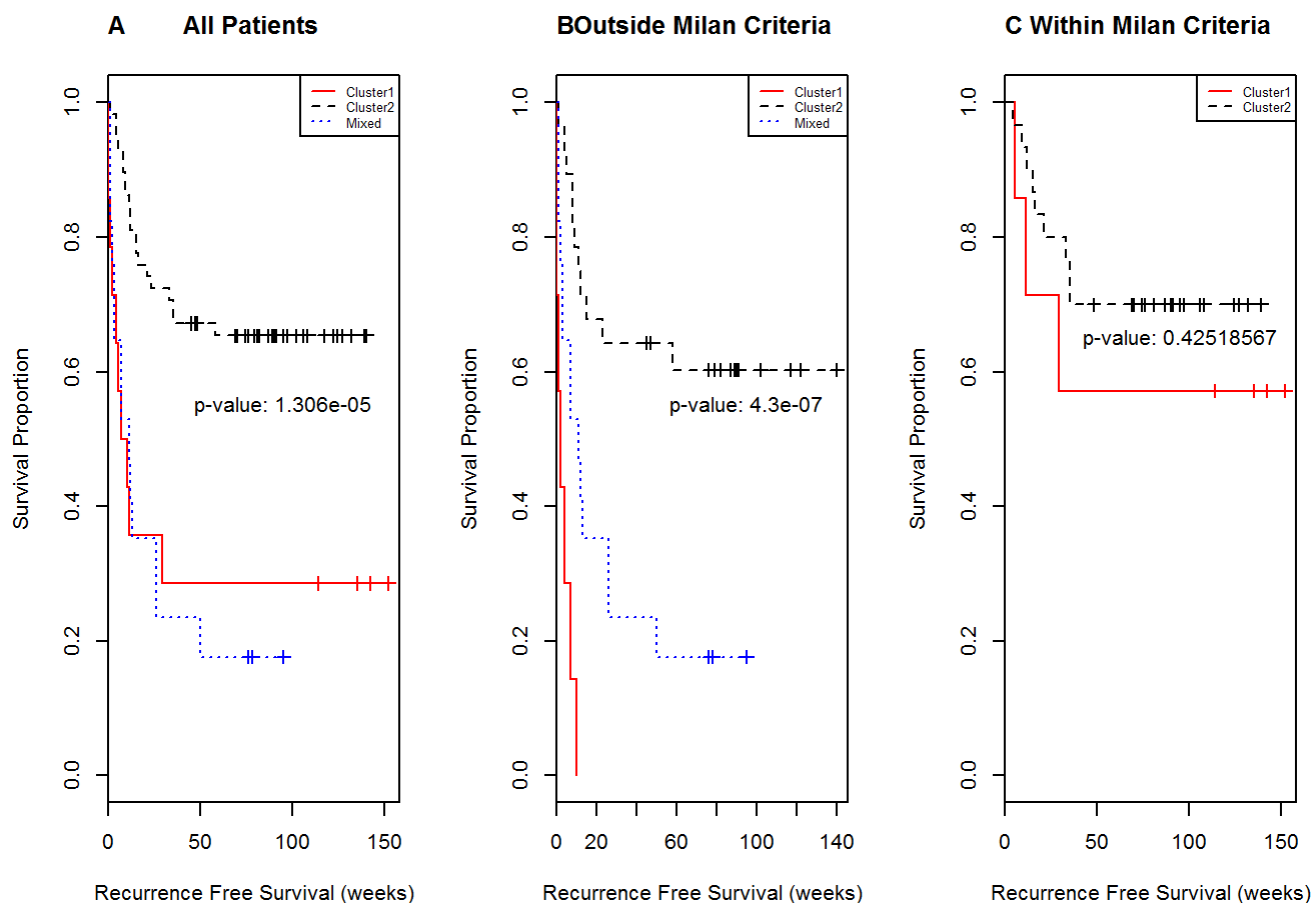
All Patients (n=89)



Plot patient-level survival curves, using feature pclst as group factor

```
poutside<-which(pdist12$milan=="Outside")
pallfit<-survfit(Surv(pdist12$rfsurv,as.numeric(as.factor(pdist12$recur))-1)~as.factor(pdist12$pclst))
moutfit<-survfit(Surv(pdist12$rfsurv[poutside],as.numeric(as.factor(pdist12$recur[poutside]))-1)~as.factor(pdist12$pclst[poutside]))
minfit<-survfit(Surv(pdist12$rfsurv[-poutside],as.numeric(as.factor(pdist12$recur[-poutside]))-1)~as.factor(pdist12$pclst[-poutside]))
```

```
#pdf('Kaplan-meier-km-milan-paitents.pdf')
par(mfrow=c(1,3))
plot(pallfit,lty=1:3,col=c('red','black','blue'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="All Patients ")
legend("topright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('Cluster1','Cluster2','Mixed'))
text(x=95, y=0.55, paste0("p-value: ",round(pchisq(survdif(Surv(pdist12$rfsurv,as.numeric(as.factor(pdist12$recur)))-1)~as.factor(pdist12$pclst))$chisq,df=2,lower.tail=F),digits=8)))
title(main='A', adj=0)
plot(moutfit,lty=1:3,col=c('red','black','blue'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Outside Milan Criteria")
legend("topright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('Cluster1','Cluster2','Mixed'))
text(x=95, y=0.55, paste0("p-value: ",round(pchisq(survdif(Surv(pdist12$rfsurv[poutside],as.numeric(as.factor(pdist12$recur[poutside])))-1)~as.factor(pdist12$pclst[poutside]))$chisq,df=2,lower.tail=F),digits=8)))
title(main='B', adj=0)
plot(minfit,lty=1:3,col=c('red','black','blue'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Within Milan Criteria")
legend("topright",cex=0.6,lty=1:2,col=c('red','black'),c('Cluster1','Cluster2'))
text(x=95, y=0.65, paste0("p-value: ",round(pchisq(survdif(Surv(pdist12$rfsurv[-poutside],as.numeric(as.factor(pdist12$recur[-poutside])))-1)~as.factor(pdist12$pclst[-poutside]))$chisq,df=1,lower.tail=F),digits=8)))
title(main='C', adj=0)
```



Plot survival curves based on batch id.

b12: samples from batch 1 and 2

b34: sample from batch3 and 4

b34pure: samples from batch 3 and 4 whose patients have not any other samples in batch 1 and 2

b3: samples from batch 3 b4: samples from batch 4

b3pure: only samples in batch 3 whose owners don't have any samples in other batch group

b4pure: only samples in batch 4 whose owners don't have any samples in other batch group

```
b12<-which(dist12$batchid %in% c(1,2))
b34<-which(dist12$batchid %in% c(3,4))
b3<-which(dist12$batchid==3)
b4<-which(dist12$batchid==4)
bmix<-which(dist12$pid %in% intersect(dist12$pid[b12],dist12$pid[b34]))
b34pure<-b34[!b34%in%bmix]
b3pure<-intersect(b34pure, which(dist12$batchid==3))
b4pure<-b34pure[!b34pure %in% b3]
```

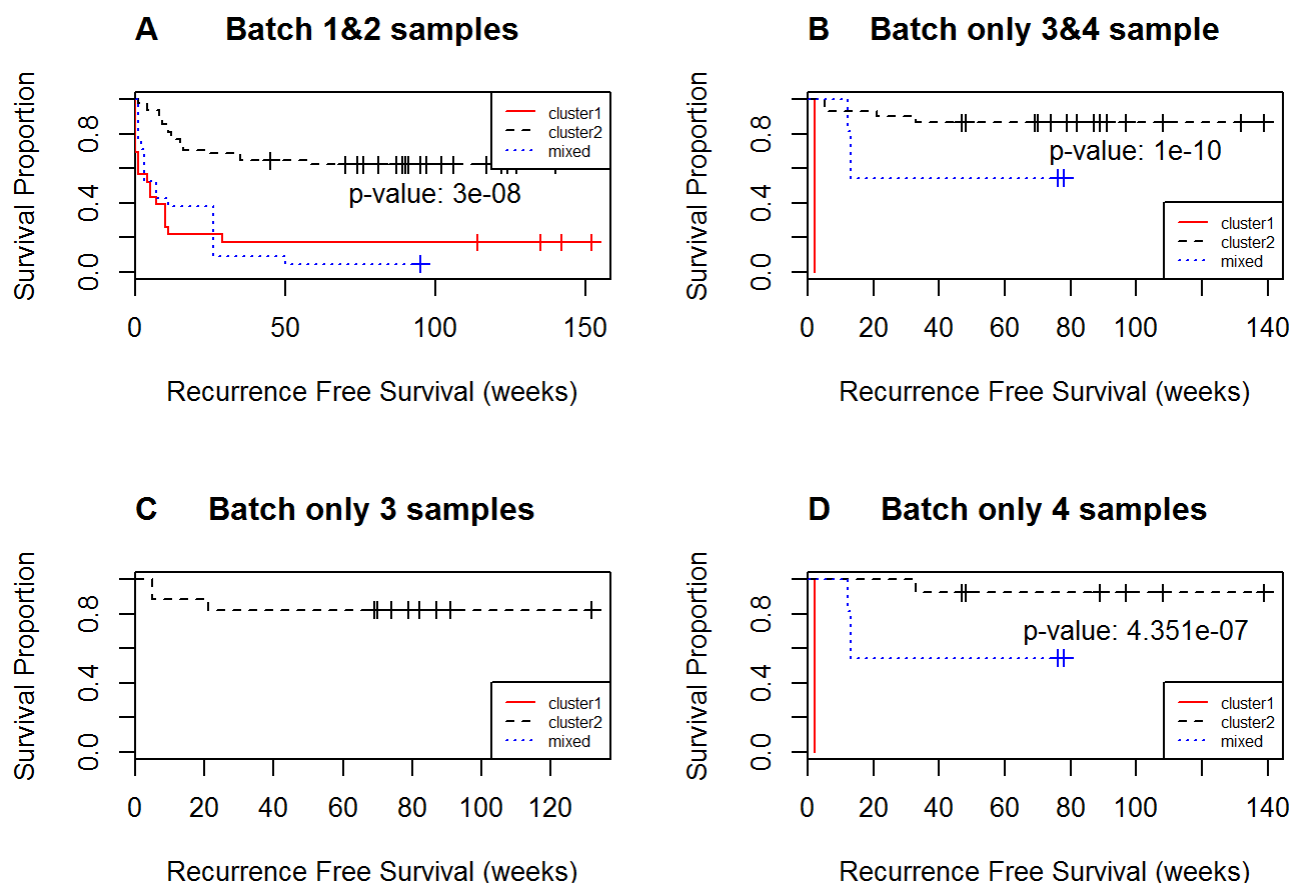
Survival objects

```
fitb12<-survfit(Surv(dist12$rfsurv[b12],as.numeric(factor(dist12$recur[b12],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b12],levels=c(1,2,3)))
fitb34<-survfit(Surv(dist12$rfsurv[b34],as.numeric(factor(dist12$recur[b34],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b34],levels=c(1,2,3)))
fitb3<-survfit(Surv(dist12$rfsurv[b3],as.numeric(factor(dist12$recur[b3],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b3],levels=c(1,2,3)))
fitb4<-survfit(Surv(dist12$rfsurv[b4],as.numeric(factor(dist12$recur[b4],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b4],levels=c(1,2,3)))
fitb34pure<-survfit(Surv(dist12$rfsurv[b34pure],as.numeric(factor(dist12$recur[b34pure],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b34pure],levels=c(1,2,3)))
fitb3pure<-survfit(Surv(dist12$rfsurv[b3pure],as.numeric(factor(dist12$recur[b3pure],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b3pure],levels=c(1,2,3)))
fitb4pure<-survfit(Surv(dist12$rfsurv[b4pure],as.numeric(factor(dist12$recur[b4pure],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b4pure],levels=c(1,2,3)))
```

Plot 1: Kaplan-meier survival curves: batch 1&2 batch 3&4 (pure), batch 3 batch 4.

```
#pdf('Kaplan-meier-kmclst-batchpure.pdf')
par(mfrow=c(2,2))
plot(fitb12,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch 1&2 samples")
legend("topright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdiff(Surv(dist12$rfsurv[b12],as.numeric(factor(dist12$recur[b12],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b12],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=8)))
title(main='A', adj=0)
plot(fitb34pure,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch only 3&4 sample")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.7, paste0("p-value: ",round(pchisq(survdiff(Surv(dist12$rfsurv[b34pure],as.numeric(factor(dist12$recur[b34pure],levels=c("No Recurrence", "Recurrence")))-1)~factor(dist12$pclst[b34pure],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=10)))
title(main='B', adj=0)
```

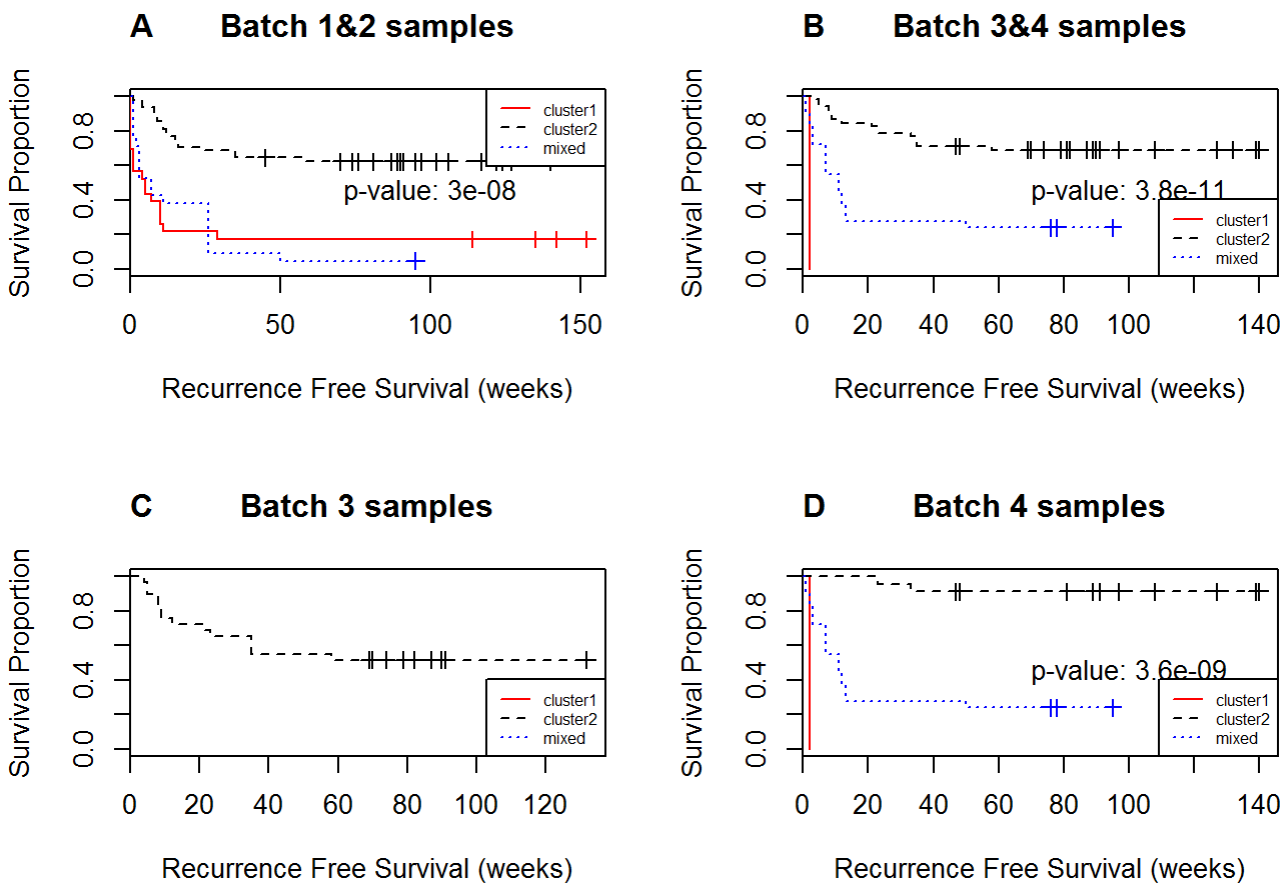
```
plot(fitb3pure,conf=F,lty=2,col='black', xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch only 3 samples")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
title(main='C', adj=0)
plot(fitb4pure,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch only 4 samples")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.7, paste0("p-value: ",round(pchisq(survdif(Surv(dist12$rfsurv[b4pure],as.numeric(factor(dist12$recur[b4pure],levels=c("No Recurrence","Recurrence")))-1)~factor(dist12$pc1st[b4pure],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=10)))
title(main='D', adj=0)
```



Plot 2: Kaplan-meier survival curves: batch 1&2, batch 3&4, batch 3 batch 4.

```
#pdf('kaplan-meier-kmclst-batch.pdf')
par(mfrow=c(2,2))
plot(fitb12,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch 1&2 samples")
legend("topright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdif(Surv(dist12$rfsurv[b12],as.numeric(factor(dist12$recur[b12],levels=c("No Recurrence","Recurrence")))-1)~factor(dist12$pc1st[b12],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=8)))
title(main='A', adj=0)
```

```
plot(fitb34,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch 3&4 samples")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdif(Surv(dist12$rfsurv[b34],as.numeric(factor(dist12$recur[b34],levels=c("No Recurrence","Recurrence")))-1~factor(dist12$pclst[b34],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=12)))
title(main='B', adj=0)
plot(fitb3,conf=F,lty=2,col='black', xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch 3 samples")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
title(main='C', adj=0)
plot(fitb4,lty=1:3,col=c('red','black','blue'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Batch 4 samples")
legend("bottomright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('cluster1','cluster2','mixed'))
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdif(Surv(dist12$rfsurv[b4],as.numeric(factor(dist12$recur[b4],levels=c("No Recurrence","Recurrence")))-1~factor(dist12$pclst[b4],levels=c(1,2,3)))$chisq,df=2,lower.tail=F),digits=10)))
title(main='D', adj=0)
```



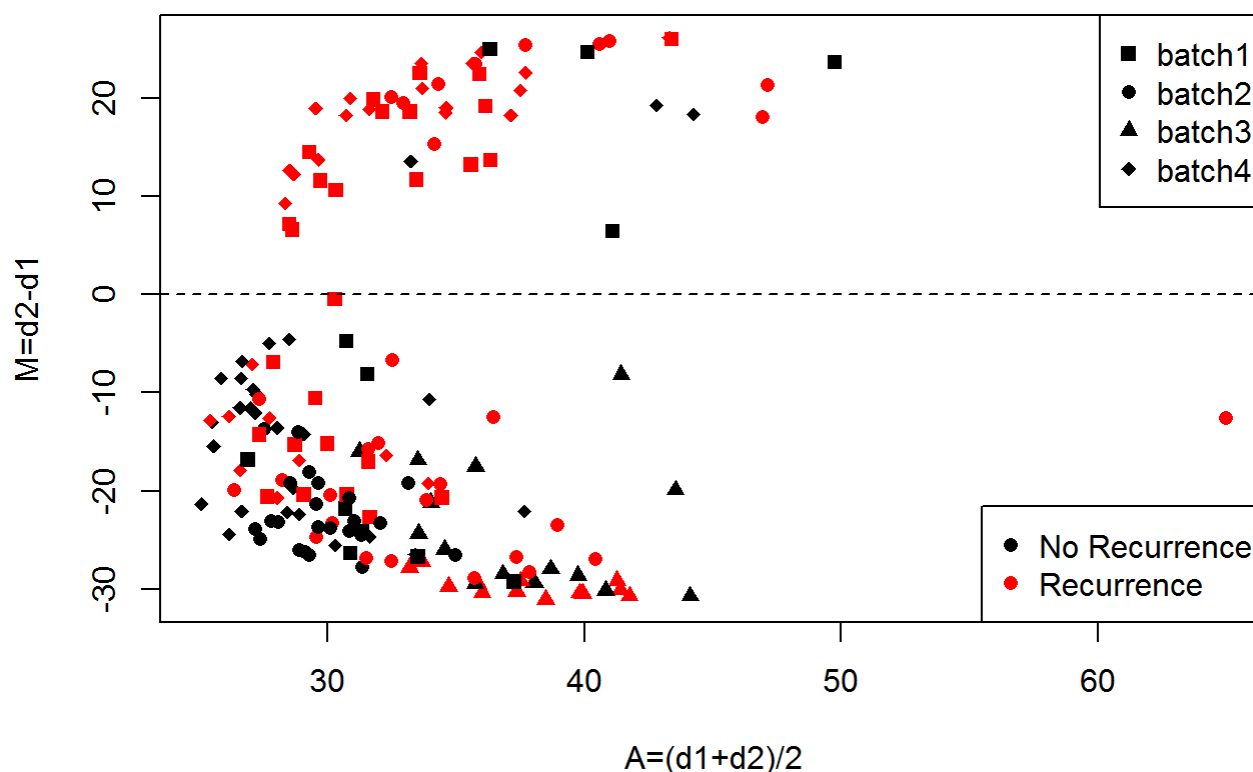
##Plot

sample VS clinical variables.
Plot 176 samples vs recur(color), batchid(symbol)

```
#pdf("km2_nonscale_176_dist_recur.pdf")
```

```
plot(dist12$A,dist12$M,pch=c(15,16,17,18)[as.numeric(as.factor(dist12$batchid))],col=c("black",
"red")[as.numeric(as.factor(dist12$recur))],xlab="A=(d1+d2)/2",ylab="M=d2-d1",main="176 sam
ples distance distribution (KMeans=2)")
legend("topright",c("batch1","batch2","batch3","batch4"), pch=c(15,16,17,18))
legend("bottomright",c("No Recurrence","Recurrence"),col=c("black","red"),pch=16)
abline(h=0,lty=2) # Optional: separate two clusters
```

176 samples distance distribution (KMeans=2)



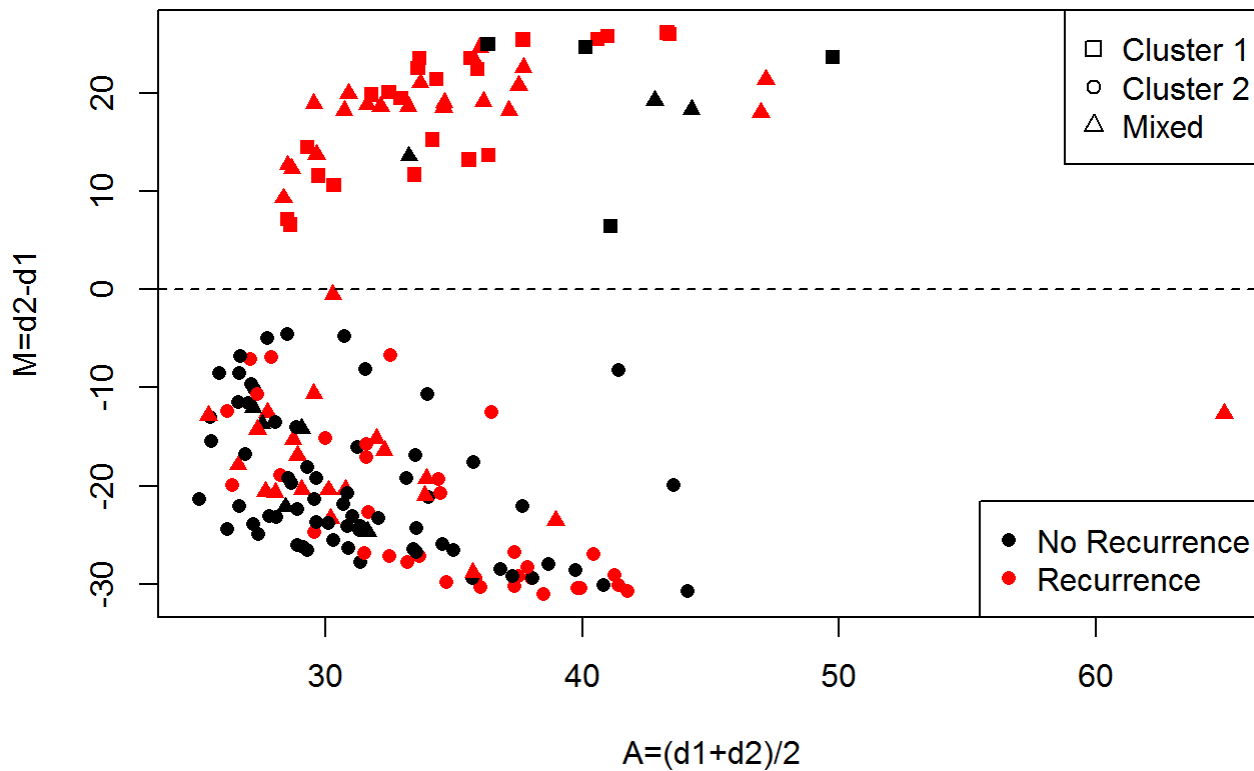
Plot 176 samples vs recur(Using color to mark), pclst(Using symbol to mark).

patient level cluster:

- 1.samples from patients whose samples all from cluster 1
- 2.samples from patients whose sample all from cluster 2
- 3.samples from patients whose sample both from cluster 1 and 2

```
#pdf("km2_nonscale_176_dist_cluster.pdf")
plot(dist12$A,dist12$M,pch=c(15,16,17)[as.factor(dist12$pclst)],col=c("black","red")[as.numer
ic(as.factor(dist12$recur))],xlab="A=(d1+d2)/2",ylab="M=d2-d1",main="176 Samples Distance Dis
tribution (KMeans=2)")
#legend("topright",c("Cluster 1","Cluster 2","Mixed"), pch=c(15,16,17))
legend("topright",c("Cluster 1","Cluster 2","Mixed"), pch=c(0,1,2)) # optional
legend("bottomright",c("No Recurrence","Recurrence"),col=c("black","red"),pch=16)
abline(h=0,lty=2)
```

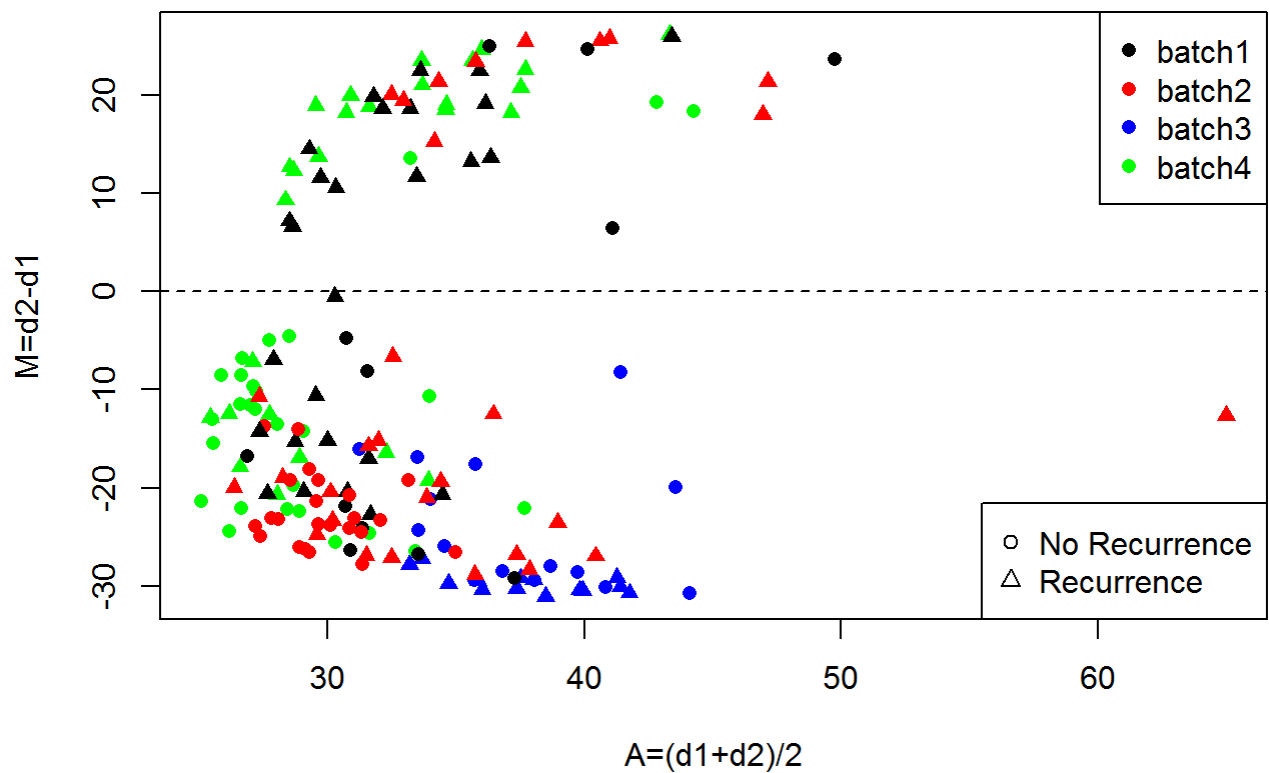

176 Samples Distance Distribution (KMeans=2)



Plot 176 samples vs batchid (color), recur(symbol)

```
#pdf("km2_nonscale_176_dist_batch.pdf")
plot(dist12$A,dist12$M,col=c("black","red","blue","green")[as.numeric(as.factor(dist12$batchid))],pch=c(16,17)[as.numeric(as.factor(dist12$recur))],xlab="A=(d1+d2)/2",ylab="M=d2-d1",main="176 samples distance distribution (KMeans=2)")
legend("topright",c("batch1","batch2","batch3","batch4"),col=c("black","red","blue","green"),pch=16)
legend("bottomright",c("No Recurrence","Recurrence"),pch=c(1,2))
abline(h=0,lty=2)
```

176 samples distance distribution (KMeans=2)

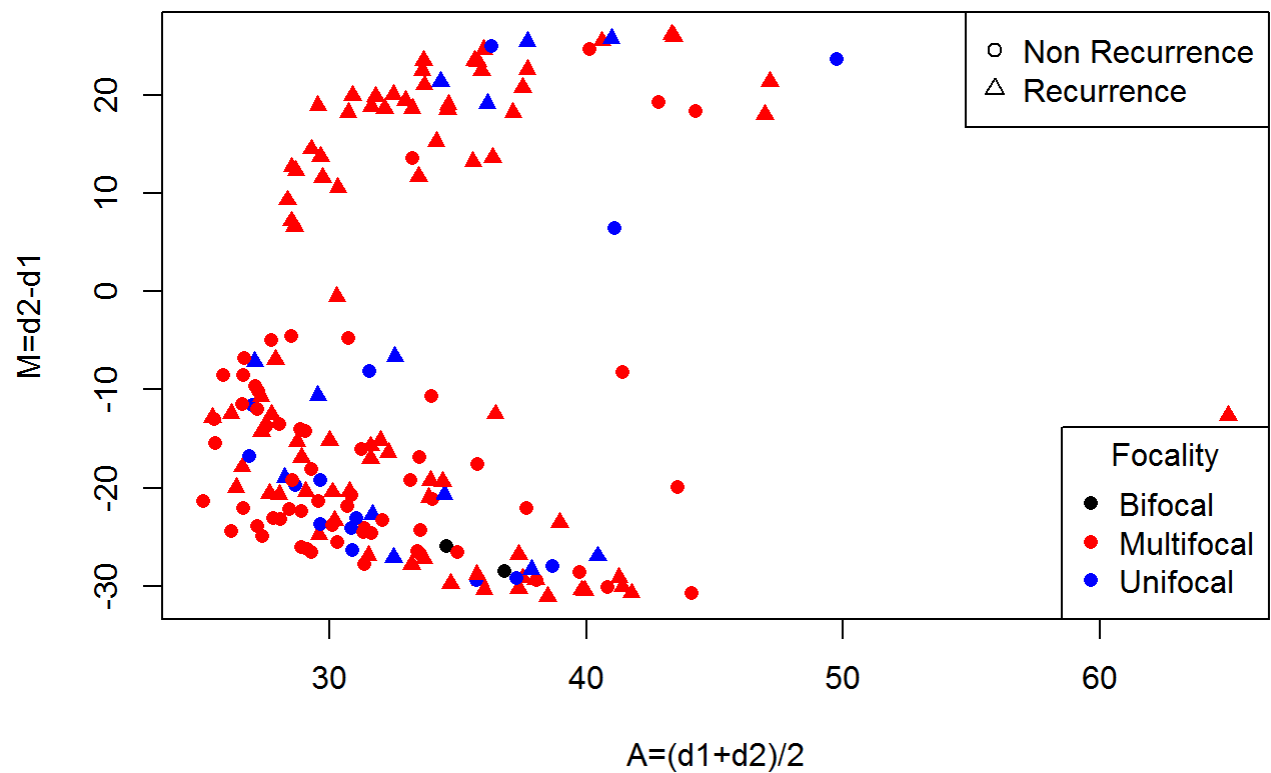


Plot 176 samples vs clinical feature focality

```
#pdf("km2_nonscale_176_focality.pdf")
plot(dist12$A,dist12$M,col=c('black','red','blue')[as.numeric(as.factor(dist12$focality))],pch=
c(16,17)[as.numeric(as.factor(dist12$recur))],main="Focality & Recur",xlab="A=(d1+d2)/2",yl
ab="M=d2-d1")
legend("bottomright",pch=16,col=c('black','red','blue'),c('Bifocal','Multifocal','Unifocal'),
title="Focality")
legend("topright",pch=c(1,2),c('Non Recurrence','Recurrence'))
```

Plot 176

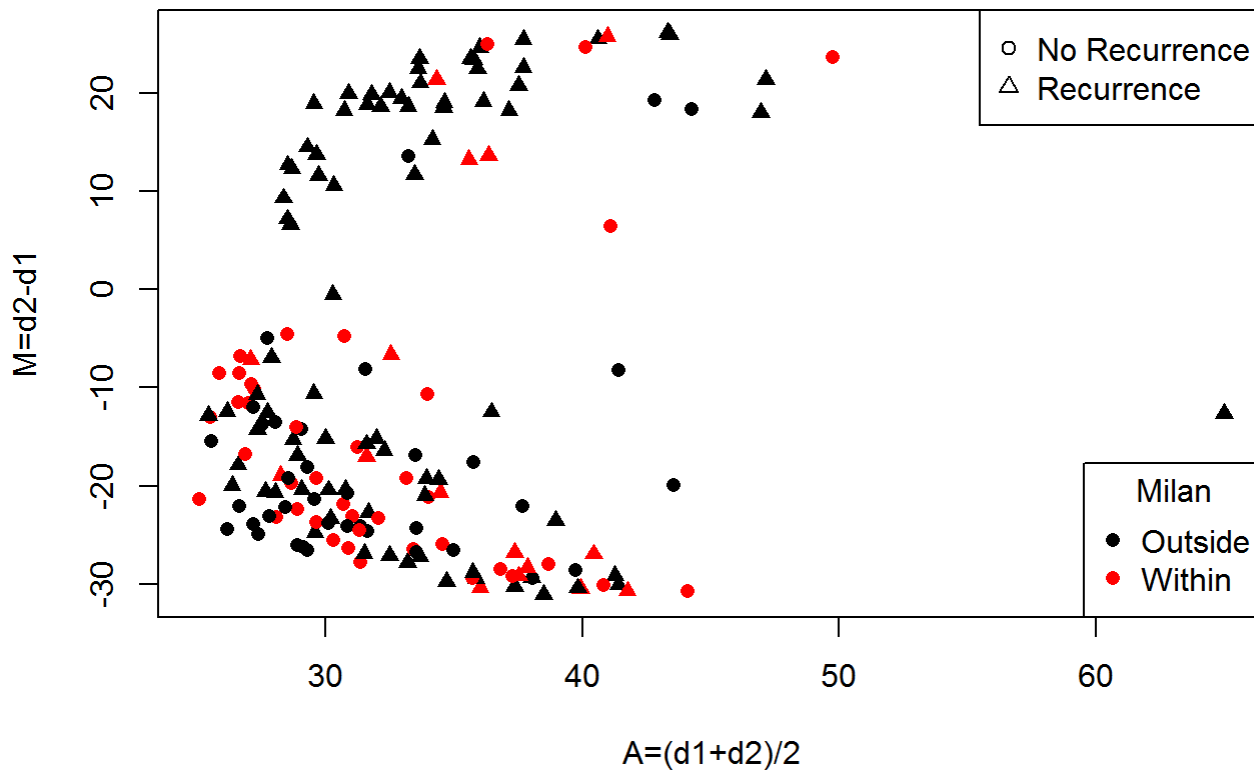
Focality & Recur



samples vs clinical feature milan

```
#pdf("km2_nonscale_176_milan.pdf")
plot(dist12$A,dist12$M,col=c('black','red')[as.numeric(as.factor(dist12$milan))],pch=c(16,17)
[as.numeric(as.factor(dist12$recur))],main="Milan & Recur",xlab="A=(d1+d2)/2",ylab="M=d2-d1")
legend("bottomright",pch=16,col=c('black','red'),c('Outside','Within'),title="Milan")
legend("topright",pch=c(1,2),c('No Recurrence','Recurrence'))
```

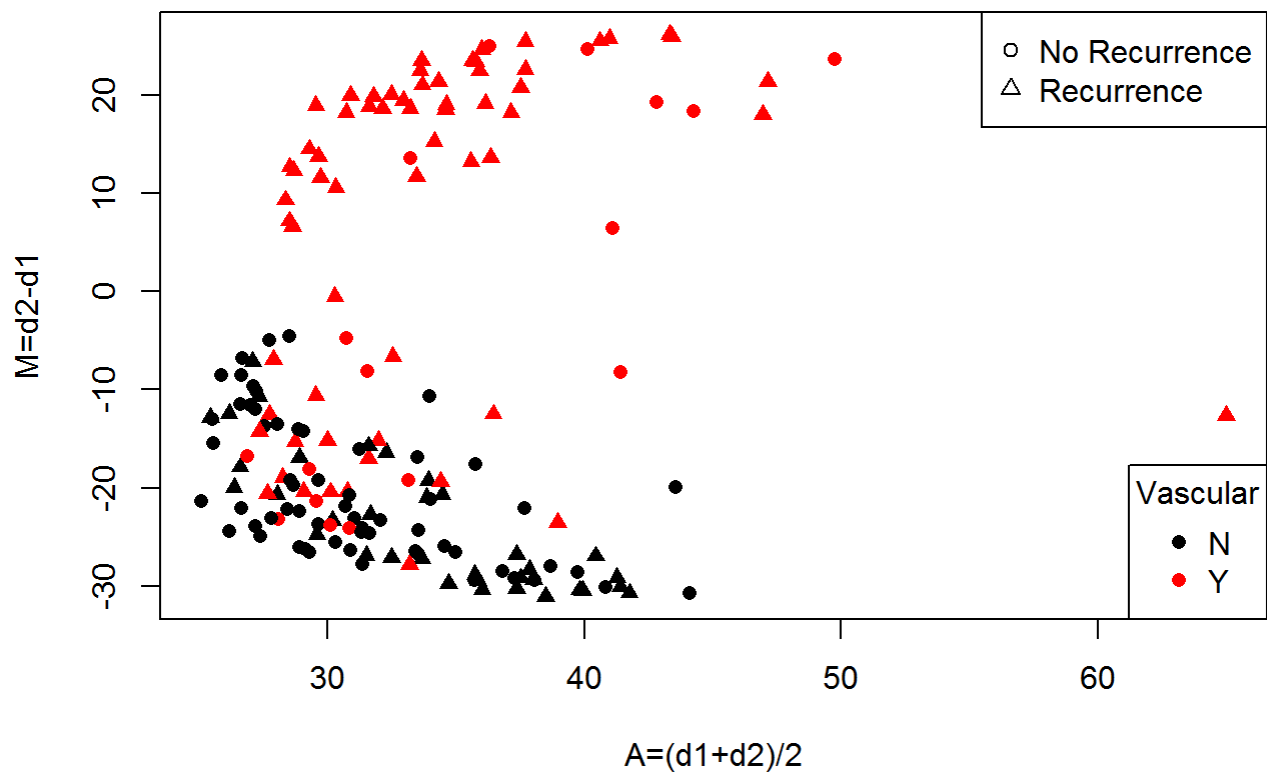
Milan & Recur



Plot 176 samples vs clinical feature Vasc

```
#pdf("km2_nonscale_176_vasc.pdf")
plot(dist12$A,dist12$M,col=c('black','red')[as.numeric(as.factor(dist12$vasc))],pch=c(16,17)[
as.numeric(as.factor(dist12$recur))],main="Vascular & Recur",xlab="A=(d1+d2)/2",ylab="M=d2-d1")
legend("bottomright",pch=16,col=c('black','red'),c('N','Y'),title="Vascular")
legend("topright",pch=c(1,2),c('No Recurrence','Recurrence'))
```

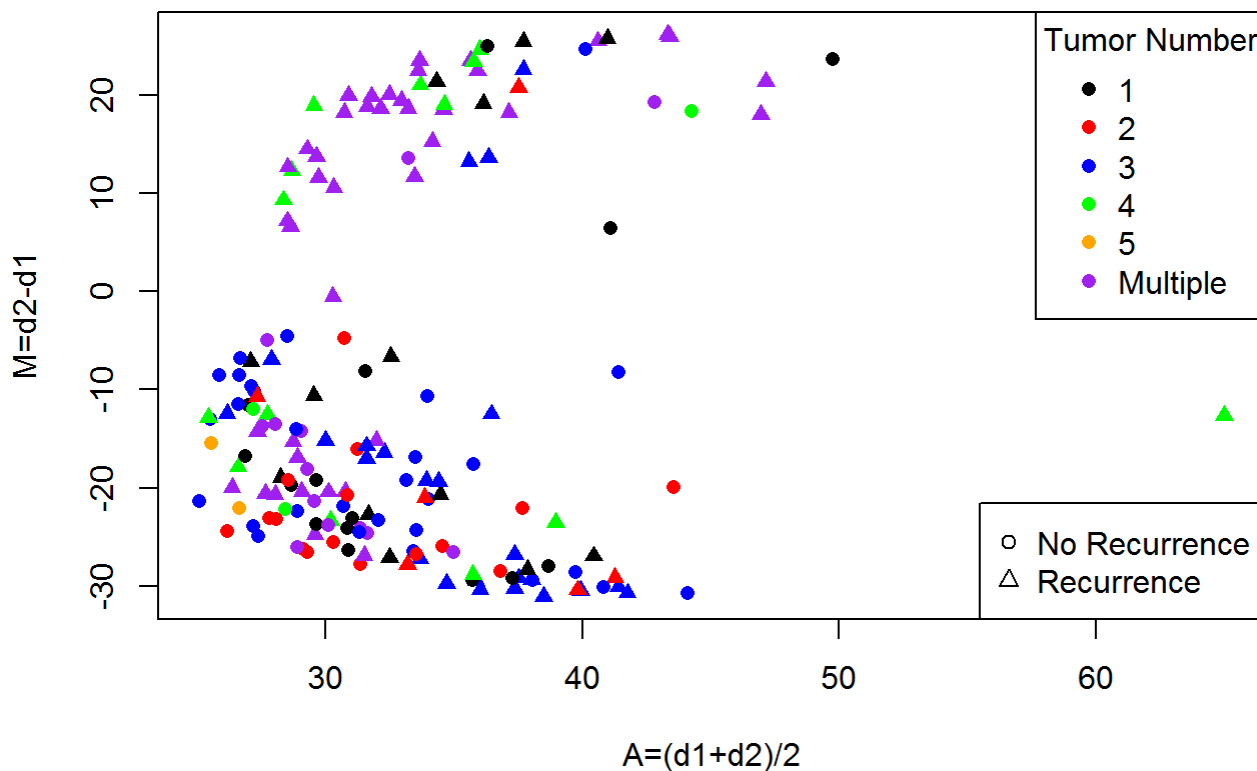
Vascular & Recur



Plot 176 samples vs clinical feature number tumors

```
#pdf("km2_nonscale_176_ntumor.pdf")
plot(dist12$A,dist12$M,col=c('black','red','blue','green','orange','purple')[as.numeric(as.factor(dist12$ntumor))],pch=c(16,17)[as.numeric(as.factor(dist12$recur))],main="Tumor Number & Recur",xlab="A=(d1+d2)/2",ylab="M=d2-d1")
legend("topright",pch=16,col=c('black','red','blue','green','orange','purple'),c('1','2','3','4','5','Multiple'),title="Tumor Number")
legend("bottomright",pch=c(1,2),c('No Recurrence','Recurrence'))
```

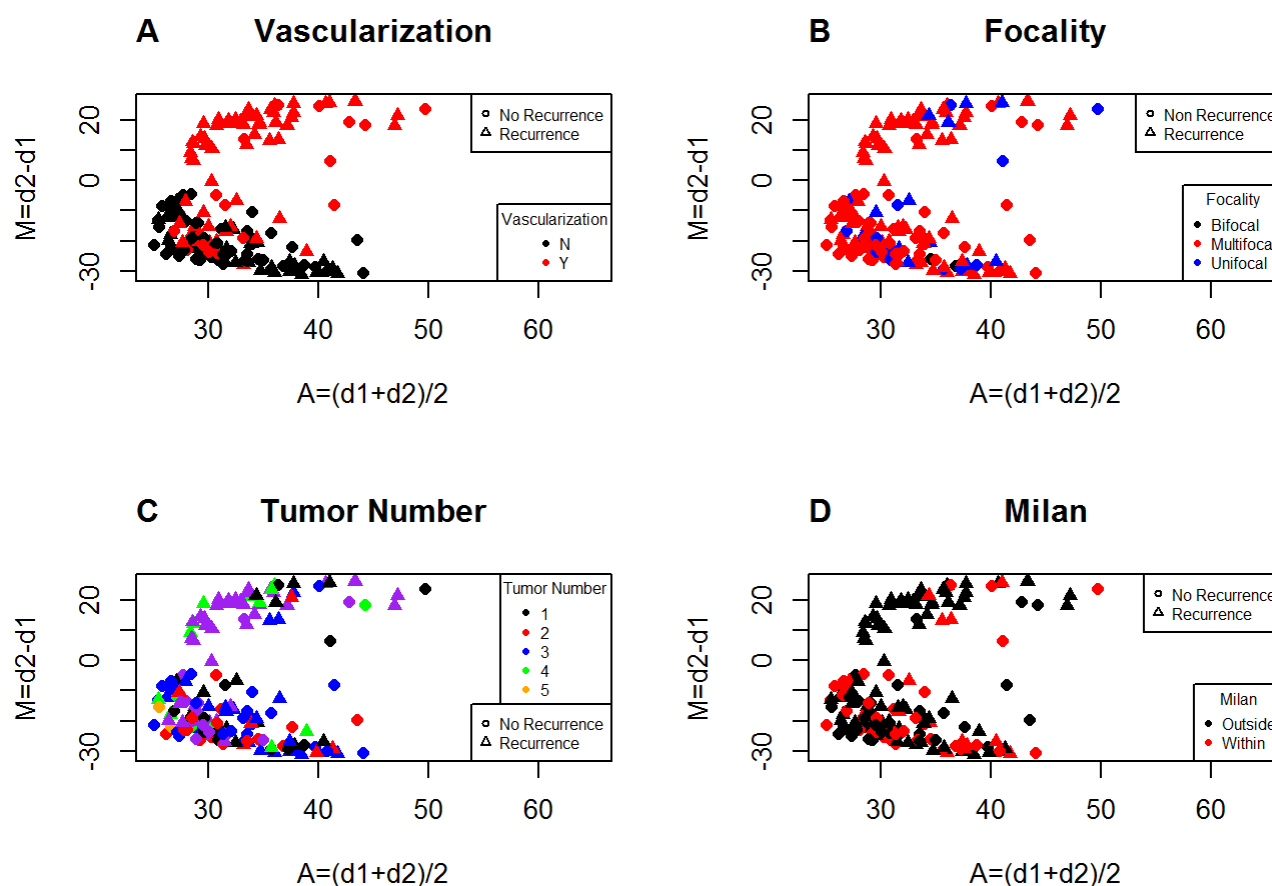
Tumor Number & Recur



Plot 176 sample vs four clinical features

```
#pdf('km2-noscale-176-4variates.pdf')
par(mfrow=c(2,2))
plot(dist12$A,dist12$M,col=c('black','red')[as.numeric(as.factor(dist12$vasc))],pch=c(16,17)[
as.numeric(as.factor(dist12$recur))],main="Vascularization",xlab="A=(d1+d2)/2",ylab="M=d2-d1"
);
legend("bottomright",cex=0.6, pch=16,col=c('black','red'),c('N','Y'),title="Vascularization")
;
legend("topright",cex=0.6,pch=c(1,2),c('No Recurrence','Recurrence'))
title(main='A', adj=0)
plot(dist12$A,dist12$M,col=c('black','red','blue')[as.numeric(as.factor(dist12$focality))],pc
h=c(16,17)[as.numeric(as.factor(dist12$recur))],main="Focality",xlab="A=(d1+d2)/2",ylab="M=d2
-d1");
legend("bottomright",cex=0.6,pch=16,col=c('black','red','blue'),c('Bifocal','Multifocal','Uni
focal'),title="Focality");
legend("topright",cex=0.6,pch=c(1,2),c('Non Recurrence','Recurrence'))
title(main='B', adj=0)
plot(dist12$A,dist12$M,col=c('black','red','blue','green','orange','purple')[as.numeric(as.fa
ctor(dist12$ntumor))],pch=c(16,17)[as.numeric(as.factor(dist12$recur))],main="Tumor Number",x
lab="A=(d1+d2)/2",ylab="M=d2-d1")
legend("topright",cex=0.6,pch=16,col=c('black','red','blue','green','orange','purple'),c('1',
'2','3','4','5','Multiple'),title="Tumor Number")
legend("bottomright",cex=0.6,pch=c(1,2),c('No Recurrence','Recurrence'))
title(main='C', adj=0)
plot(dist12$A,dist12$M,col=c('black','red')[as.numeric(as.factor(dist12$milan))],pch=c(16,17)
```

```
[as.numeric(as.factor(dist12$recur))],main="Milan",xlab="A=(d1+d2)/2",ylab="M=d2-d1");
legend("bottomright",cex=0.6,pch=16,col=c('black','red'),c('Outside','Within'),title="Milan")
;
title(main='D', adj=0)
legend("topright",cex=0.6,pch=c(1,2),c('No Recurrence','Recurrence'));
```

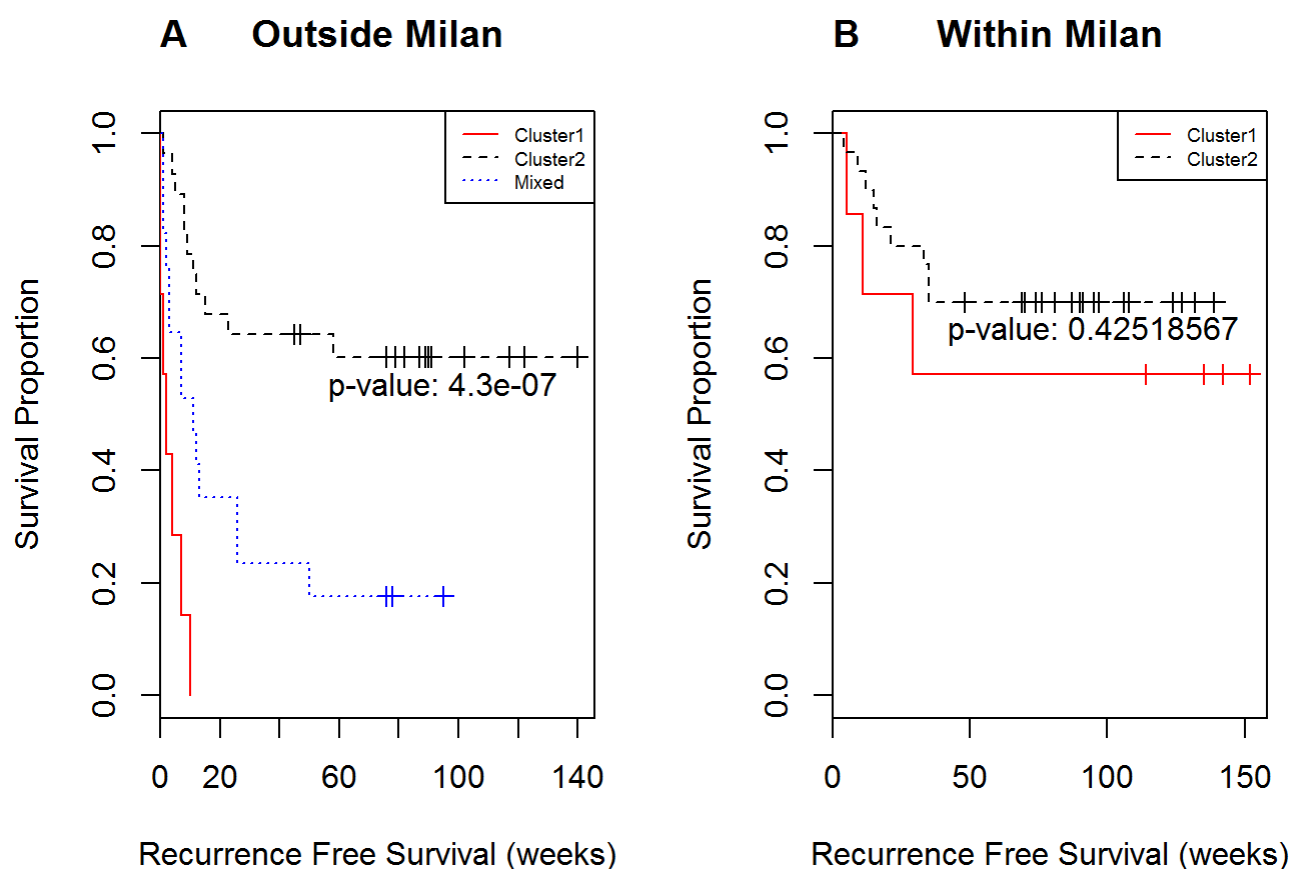


##Patient-

level survival curves

pclst as groups: two figures: outside and within milan

```
#pdf('Kaplan-meier-km-milan-paitents.pdf')
par(mfrow=c(1,2))
plot(moutfit,lty=1:3,col=c('red','black','blue'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Outside Milan")
legend("topright",cex=0.6,lty=1:3,col=c('red','black','blue'),c('Cluster1','Cluster2','Mixed'))
text(x=95, y=0.55, paste0("p-value: ",round(pchisq(survdif(Surv(pdist12$rfsurv[poutside],as.numeric(as.factor(pdist12$recur[poutside]))-1)~as.factor(pdist12$pclst[poutside]))$chisq,df=2,lower.tail=F),digits=8)))
title(main='A', adj=0)
plot(minfit,lty=1:3,col=c('red','black','blue'),xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Within Milan")
legend("topright",cex=0.6,lty=1:2,col=c('red','black'),c('Cluster1','Cluster2'))
text(x=95, y=0.65, paste0("p-value: ",round(pchisq(survdif(Surv(pdist12$rfsurv[-poutside],as.numeric(as.factor(pdist12$recur[-poutside]))-1)~as.factor(pdist12$pclst[-poutside]))$chisq,df=1,lower.tail=F),digits=8)))
title(main='B', adj=0)
```



Data from AJT paper

60 Min-Max features in table 2 of ATJ paper

```
hsamin<-read.csv(file='hsaMIN.txt',header=FALSE,colClasses = "character")
hsamax<-read.csv(file="hsaMAX.txt",header=FALSE,colClasses = "character")
```

Min-MAX for 847 features

```
pmax<-as.data.frame(matrix(nrow=89,ncol=0))
for(i in 14:860)
{
  cn<-colnames(spall)[i]
  cp<-tapply(spall[,i],as.factor(spall$pid),max)
  pmax<-cbind(pmax,cp=cp)
  colnames(pmax)[colnames(pmax)=='cp']<-paste0(cn,"_MAX")
}

pmin<-as.data.frame(matrix(nrow=89,ncol=0))
for(i in 14:860)
{
  cn<-colnames(spall)[i];
  cp<-tapply(spall[,i],as.factor(spall$pid),min);
  pmin<-cbind(pmin,cp=cp);
  colnames(pmin)[colnames(pmin)=='cp']<-paste0(cn,"_MIN")
}
```



```
hmin<-pmin[,hsamin$V1]
hmax<-pmax[,hsamax$V1]
```

Top 60 min-max features + clinical variates

```
p60new<-cbind(hmin,hmax)
p60new<-merge(ptab,p60new,by.x='pid',by.y='row.names')
#write.csv(p60new,file='p60new.csv')
```

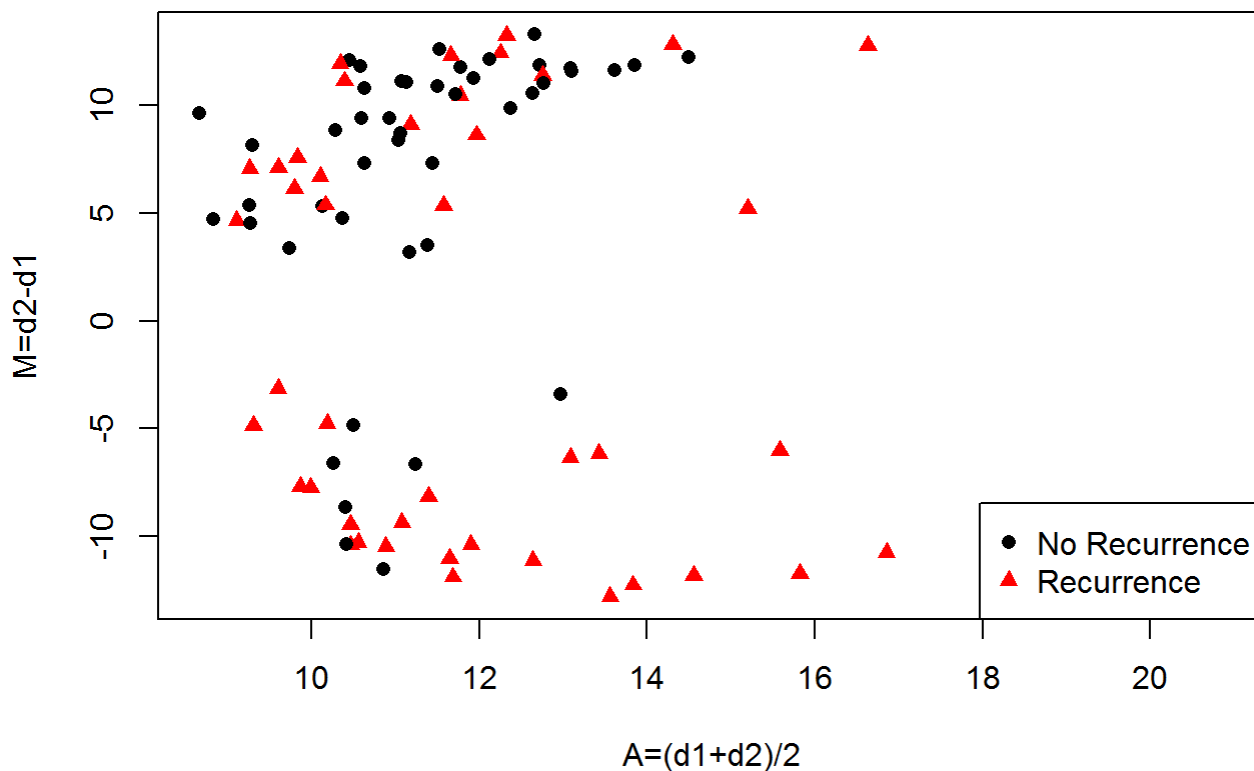
kmeans for 89 patients with 60 features.

```
set.seed(55)
p60km<-kmeans(p60new[,9:68],2)
rownames(p60km$centers)<-c('c1','c2')
pdist60<-rbind(p60new[,9:68],p60km$centers)
pdist60<-as.matrix(dist(pdist60))
pdist60<-as.data.frame(pdist60[1:89,90:91])
pdist60$M<-pdist60$c2-pdist60$c1
pdist60$A<-(pdist60$c2+pdist60$c1)/2
p60new$kclst<-p60km$cluster
p60new<-cbind(p60new,pdist60)
#write.csv(p60new,file='p60new.csv')
```

Plot the distribution of 89 patients by 60 min-mdax features

```
#pdf('min-max-60f-89p-distribution.pdf')
plot(p60new$A,p60new$M,col=c('black','red')[as.numeric(as.factor(p60new$recur))],pch=c(16,17)
[as.numeric(as.factor(p60new$recur))],main="60 Min-Max Features & 89 Patients",xlab="A=(d1+d2)
)/2",ylab="M=d2-d1")
legend("bottomright",pch=c(16,17),col=c('black','red'),c('No Recurrence','Recurrence'))
```

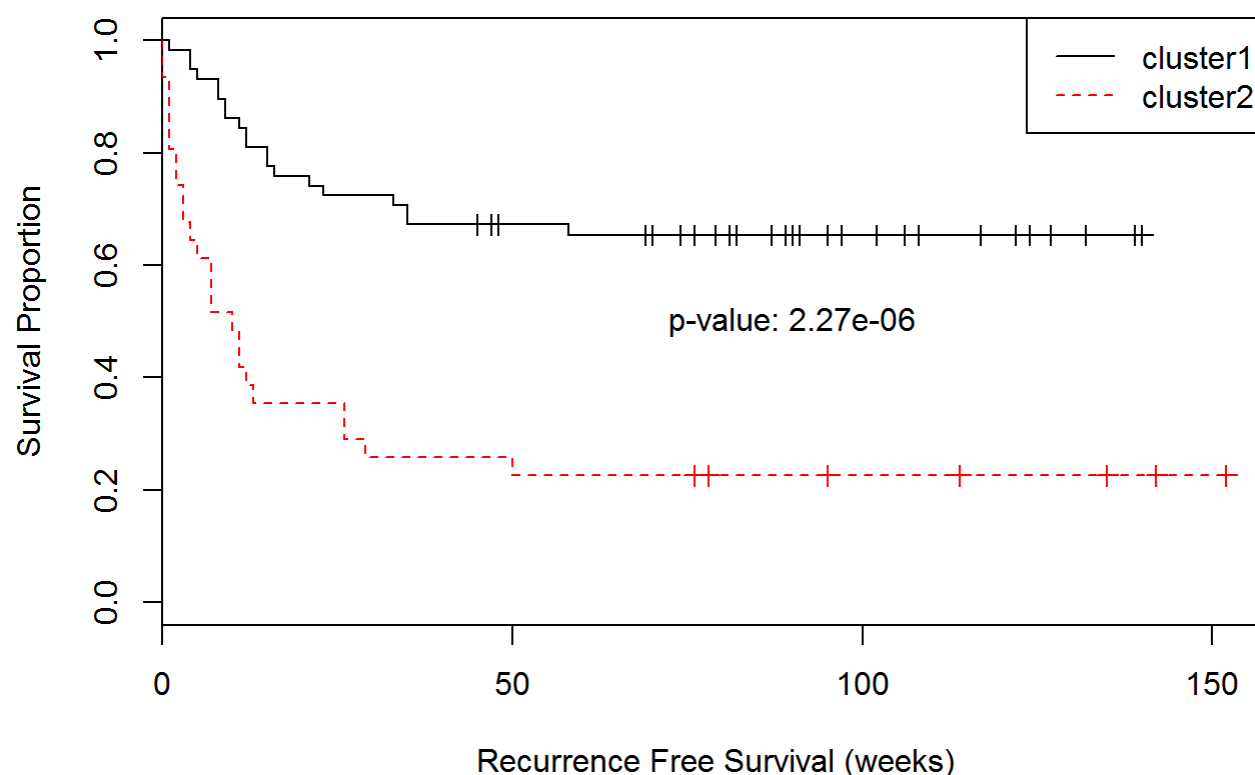
60 Min-Max Features & 89 Patients



Plot the survival curve

```
#library('survival')
p60fit<-survfit(Surv(p60new$rfsurv,as.numeric(as.factor(p60new$recur))-1)~as.factor(p60new$kc
lst))
#pdf('min-max-60f-89p-survival.pdf')
plot(p60fit,lty=1:2,col=c('black','red'), xlab="Recurrence Free Survival (weeks)", ylab="Surv
ival Proportion",main="60 Min-Max Features & 89 Patients")
legend("topright",lty=1:2,col=c('black','red'),c('cluster1','cluster2'))
text(x=90, y=0.5, paste0("p-value: ",round(pchisq(survdiff(Surv(p60new$rfsurv,as.numeric(as.f
actor(p60new$recur))-1)~as.factor(p60new$kc1st))$chisq,df=1,lower.tail=F),digits=8)))
```

60 Min-Max Features & 89 Patients



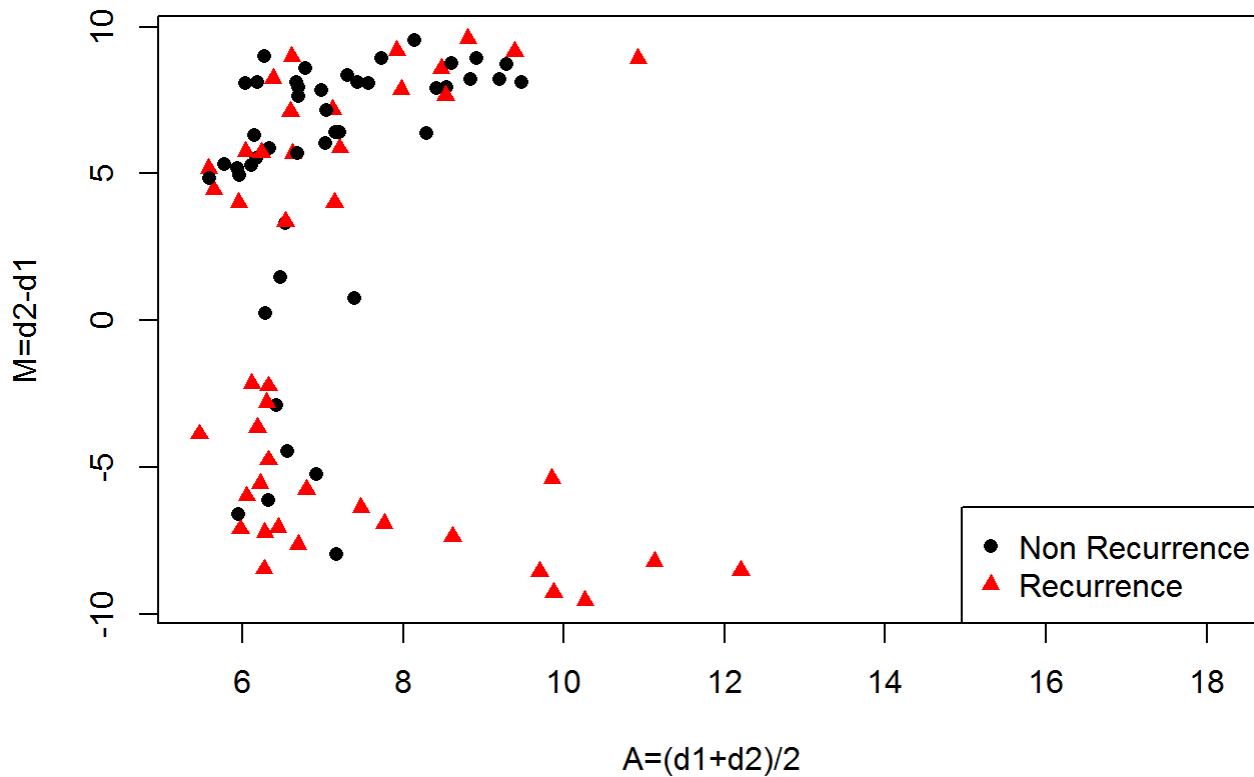
The 16 min-max features

```
hsa16<-read.csv("hsaHCClit.txt",header=F,stringsAsFactor=F)
phsa16<-p60new[,hsa16$V1]
set.seed(55)
phsa16km<-kmeans(phsa16,2)
rownames(phsa16km$centers)<-c('c1','c2')
phsa16dist<-rbind(phsa16,phsa16km$centers)
phsa16dist<-as.matrix(dist(phsa16dist))
phsa16dist<-phsa16dist[1:89,90:91]
phsa16<-cbind(phsa16,phsa16dist)
phsa16$M<-phsa16$c2-phsa16$c1
phsa16$A<-(phsa16$c2+phsa16$c1)/2
phsa16$kc1st<-ifelse(phsa16$M>0,1,2)
phsa16<-cbind(phsa16,p60new[,1:8])
#write.csv(phsa16,file='phsa16.csv')
```

plot the distribution of 89 patients by 16 features

```
#pdf('min-max-16f-89p-distribution.pdf')
plot(phsa16$A,phsa16$M,col=c('black','red')[as.numeric(as.factor(phsa16$recur))],pch=c(16,17)
[as.numeric(as.factor(phsa16$recur))],main=" 16 Min-Max Features & 89 Patients",xlab="A=(d1+d
2)/2",ylab="M=d2-d1")
legend("bottomright",pch=c(16,17),col=c('black','red'),c('Non Recurrence','Recurrence'))
```

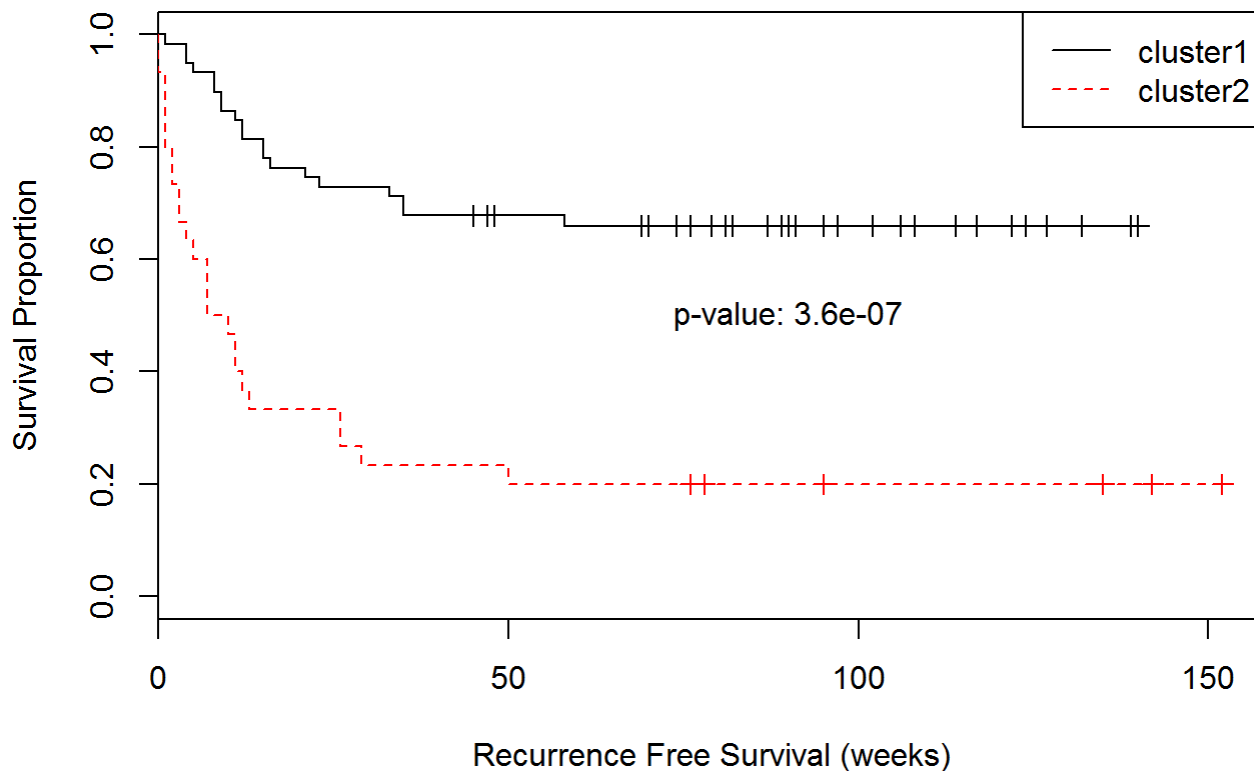
16 Min-Max Features & 89 Patients



Survival curve (16 min-max features)

```
p16fit<-survfit(Surv(phsa16$rfsurv,as.numeric(as.factor(phsa16$recur))-1)~as.factor(phsa16$kclst))
#pdf('min-max-16f-89p-survival.pdf')
plot(p16fit,lty=1:2,col=c('black','red'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="16 Min-Max Features & 89 Patients")
legend("topright",lty=1:2,col=c('black','red'),c('cluster1','cluster2'))
text(x=90, y=0.5, paste0("p-value: ",round(pchisq(survdiff(Surv(phsa16$rfsurv,as.numeric(as.factor(phsa16$recur))-1)~as.factor(phsa16$kclst))$chisq,df=1,lower.tail=F),digits=8)))
```

16 Min-Max Features & 89 Patients



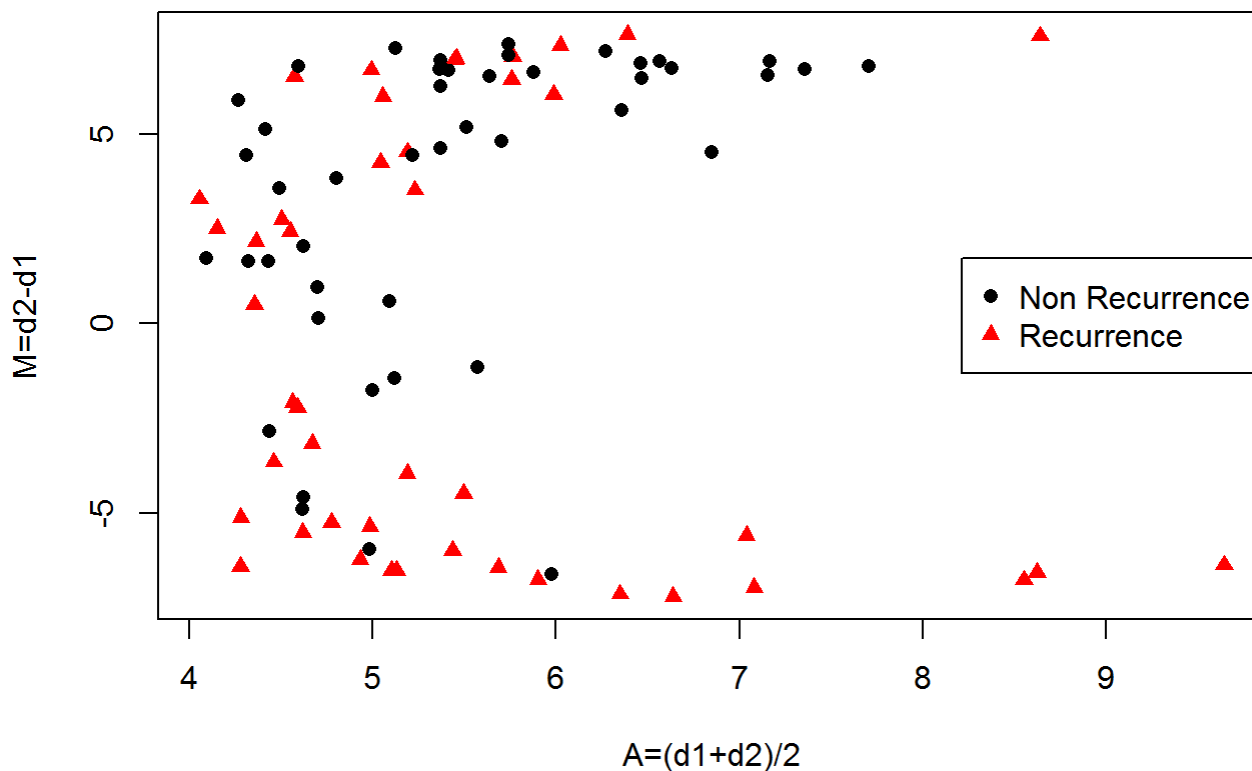
The 6 min-max features

```
hsa6<-read.csv("hsabold6.txt",header=F,stringsAsFactor=F)
phsa6<-p60new[,hsa6$V1]
set.seed(55)
phsa6km<-kmeans(phsa6,2)
rownames(phsa6km$centers)<-c('c1','c2')
phsa6dist<-rbind(phsa6,phsa6km$centers)
phsa6dist<-as.matrix(dist(phsa6dist))
phsa6dist<-phsa6dist[1:89,90:91]
phsa6<-cbind(phsa6,phsa6dist)
phsa6$M<-phsa6$c2-phsa6$c1
phsa6$A<-(phsa6$c2+phsa6$c1)/2
phsa6$kc1st<-ifelse(phsa6$M>0,1,2)
phsa6<-cbind(phsa6,p60new[,1:8])
#write.csv(phsa6,file='phsa6.csv')
```

Plot the distribution of 89 patients by 6 features

```
#pdf('min-max-6f-89p-distribution.pdf')
plot(phsa6$A,phsa6$M,col=c('black','red')[as.numeric(as.factor(phsa6$recur))],pch=c(16,17)[as.numeric(as.factor(phsa6$recur))],main=" 6 Min-Max Features & 89 Patients",xlab="A=(d1+d2)/2",ylab="M=d2-d1")
legend("right",pch=c(16,17),col=c('black','red'),c('Non Recurrence','Recurrence'))
```

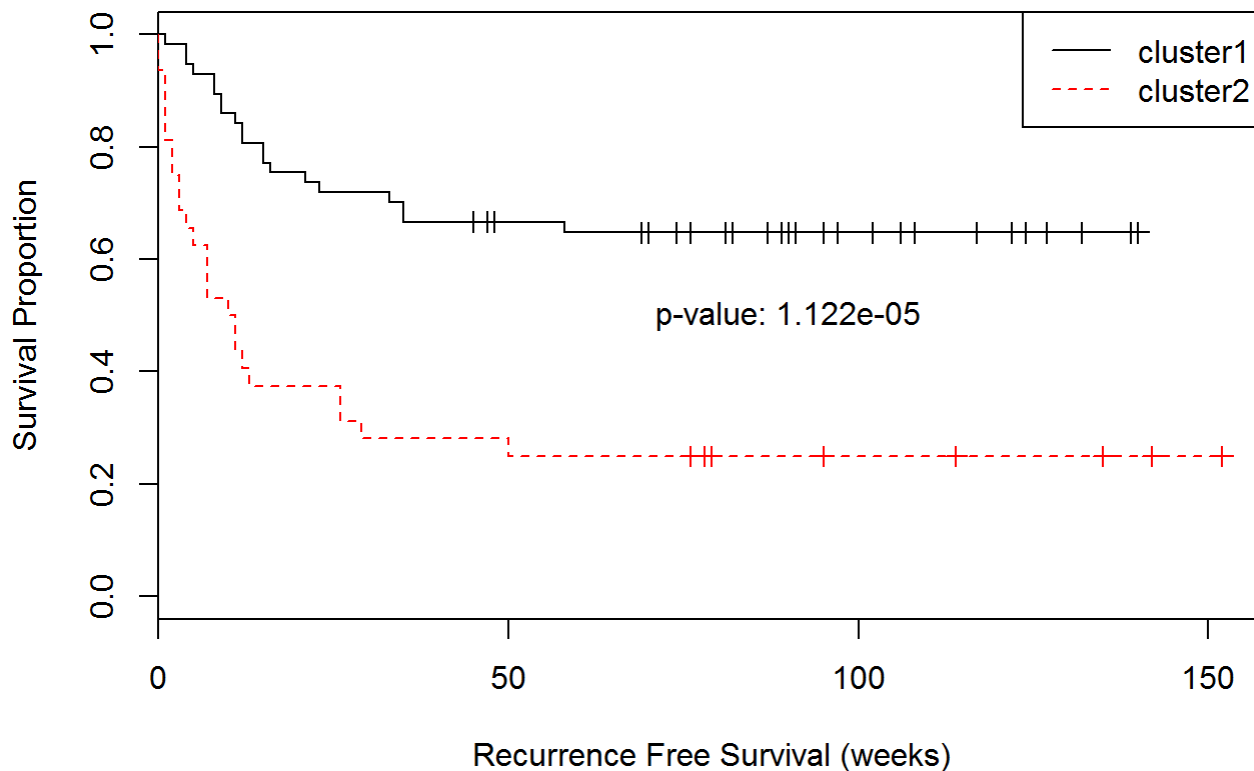
6 Min-Max Features & 89 Patients



Survival curve (6 min-max features)

```
p6fit<-survfit(Surv(phsa6$rfsurv,as.numeric(as.factor(phsa6$recur))-1)~as.factor(phsa6$kclst))
#pdf('min-max-6f-89p-survival.pdf')
plot(p6fit,lty=1:2,col=c('black','red'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="6 Min-Max Features & 89 Patients")
legend("topright",lty=1:2,col=c('black','red'),c('cluster1','cluster2'))
text(x=90, y=0.5, paste0("p-value: ",round(pchisq(survdiff(Surv(phsa6$rfsurv,as.numeric(as.factor(phsa6$recur))-1)~as.factor(phsa6$kclst))$chisq,df=1,lower.tail=F),digits=8)))
```

6 Min-Max Features & 89 Patients



Mutual information

This section calculate Mutual Information and use Naive Bayes and Support vector machine to validate the 5mirna. Require to load 'spall-dist-ptab.rdata' first.

Select 88 samples: 22 from Recurrence, 66 from No recurrence

```
clrecur<-which(dist12$pclst==1&dist12$recur=='Recurrence')
c2nonrecur<-which(dist12$pclst==2&dist12$recur=='No Recurrence')
train88<-c(clrecur,c2nonrecur)
strain88<-spall[train88,]
```

Calculate Q1,Q2,Q3,MAX for intervals [Min,Q1],[Q1,Q2],[Q2,Q3],[Q3,Max]

```
i1<-apply(strain88[,14:860],2,function(x)quantile(x,probs=0.25))
i2<-apply(strain88[,14:860],2,function(x)quantile(x,probs=0.50))
i3<-apply(strain88[,14:860],2,function(x)quantile(x,probs=0.75))
i4<-apply(strain88[,14:860],2,function(x)quantile(x,probs=1))
intv88<-rbind(i1=i1,i2=i2,i3=i3,i4=i4)
#write.csv(intv88,'intv88.csv')
```

The 88 samples fall into four bins.

```
sint88<-strain88[,14:860]
for( i in 1:88)
```

```
{ for(j in 1:847)
  { if(sint88[i,j]-intv88[1,j]<0) {sint88[i,j]<-1}
    else if(sint88[i,j]-intv88[2,j]<0) {sint88[i,j]<-2}
    else if(sint88[i,j]-intv88[3,j]<0) {sint88[i,j]<-3}
    else {sint88[i,j]<-4}
  }
}
sint88$recur<-strain88$recur
```

Mifun function: caculate MI value

```
MIfun<-function(t,value) { N1a<-22
N0a<-66
N<-88
N11<-t[ 'Recurrence' ,value ]
N10<-N1a-N11
N01<-t[ 'No Recurrence' ,value ]
N00<-N0a-N01
Na1<-N01+N11
Na0<-N00+N10
Mi<-0;
if(N11!=0) {Mi<-Mi+(N11/N)*log2(N*N11/(N1a*Na1)) }
if(N01!=0) {Mi<-Mi+(N01/N)*log2(N*N01/(N0a*Na1)) }
if(N10!=0) {Mi<-Mi+(N10/N)*log2(N*N10/(N1a*Na0)) }
if(N00!=0) {Mi<-Mi+(N00/N)*log2(N*N00/(N0a*Na0)) }
return(Mi)
}
```

Caculate 847*4 MI matrix

```
MI88<-data.frame(row.names=c('m1','m2','m3','m4'))
for ( i in 1:847)
{ nc<-colnames(sint88)[i];
t<-table(sint88$recur, sint88[,i]);
m1<-MIfun(t,'1');
m2<-MIfun(t,'2');
m3<-MIfun(t,'3');
m4<-MIfun(t,'4');
MI88[,nc]<-c(m1,m2,m3,m4)
}
#write.csv(MI88,'MI88.csv')
```

Top 50 MI values and their miRNA and interval (table S1)

```
allMI88<-as.vector(as.matrix(MI88))
allMI88sort<-sort(allMI88,decreasing=TRUE)
top50<-unique(allMI88sort[1:50])
mrna<-c();
miv<-c();
interval<-c();
intvalue<-c();
```



```
for ( i in 1:length(top50) ) # cacuplate table S1
{
  tl<-which(MI88==top50[i],arr.ind=TRUE);
  tr<-dim(tl)[1];
  for( j in 1:tr)
  {
    mrna<-c(mrna, colnames(intv88)[tl[j,2]]);
    miv<-c(miv,top50[i]);
    interval<-c(interval,rownames(tl)[j]);
    intvalue<-c(intvalue, intv88[tl[j,1],tl[j,2]]);
  }
}
mitop50<-cbind(miRNA=mrna,MI=miv,Interval=interval, Intvalue=intvalue); # matrix for table S1
```

176 samples with 847*4 bins (binary matrix).

```
sint176<-spall[,14:860]
for( i in 1:176)
{ for(j in 1:847)
  { if(sint176[i,j]-intv88[1,j]<0) {sint176[i,j]<-1}
    else if(sint176[i,j]-intv88[2,j]<0) {sint176[i,j]<-2}
    else if(sint176[i,j]-intv88[3,j]<0) {sint176[i,j]<-3}
    else {sint176[i,j]<-4}
  }
}
```

Top 5 miRNAs in [min,Q1] interval

```
sint5ml<-sint176[,mitop50[1:5,1]]
sint5ml<-ifelse(sint5ml!=1,0,1)
```

Naive Bayes: using sint5ml binary data.

- * training set: 88 samples;
- * testing data: the left 88 samples

```
library(e1071)
nb88<-naiveBayes(sint5ml[train88,],as.factor(spall$recur[train88]))
preb88<-predict(nb88,sint5ml)
nbsint5<-cbind(kclst=dist12$kclst,recur=dist12$recur,preb88)
#write.csv(nbsint5,'nbsint5.csv')
```

Support vector machine:

using sint5ml binary data.

- * training set: 88 samples;
- * testing data: the left 88 samples

```
svm88<-svm(sint5ml[train88,],as.factor(spall$recur[train88]),cost=1000)
psvm88<-predict(svm88,sint5ml)
svmsint5<-cbind(kclst=dist12$kclst,recur=dist12$recur,psvm88)
#write.csv(svmsint5,'svmsint5.csv')
```

Top 5 miRNA raw data

```

raw5mRna<-spall[,mitop50[1:5,1]]
# naive Bayes: using 5 miRNAs' rawa data: training set: 88 samples; testing data: the left 88
samples
rawnb88<-naiveBayes(raw5mRna[train88,],as.factor(spall$recur[train88]))
rawpreb88<-predict(rawnb88,raw5mRna)
nbrawmRNA5<-cbind(kclst=dist12$kclst,recur=dist12$recur,rawpreb88)
#write.csv(nbrawmRNA5,'nbrawmRNA5.csv')

```

Support vector machine: using 5 miRNAs' rawa data.

* training set: 88 samples; * testing data: the left 88 samples

```

rawsvm88<-svm(raw5mRna[train88,],as.factor(spall$recur[train88]),cost=1000)
rawpsvm88<-predict(rawsvm88,raw5mRna)
svmraw5mRNA<-cbind(kclst=dist12$kclst,recur=dist12$recur,rawpsvm88)
write.csv(svmraw5mRNA,'svmraw5mRNA.csv')

```

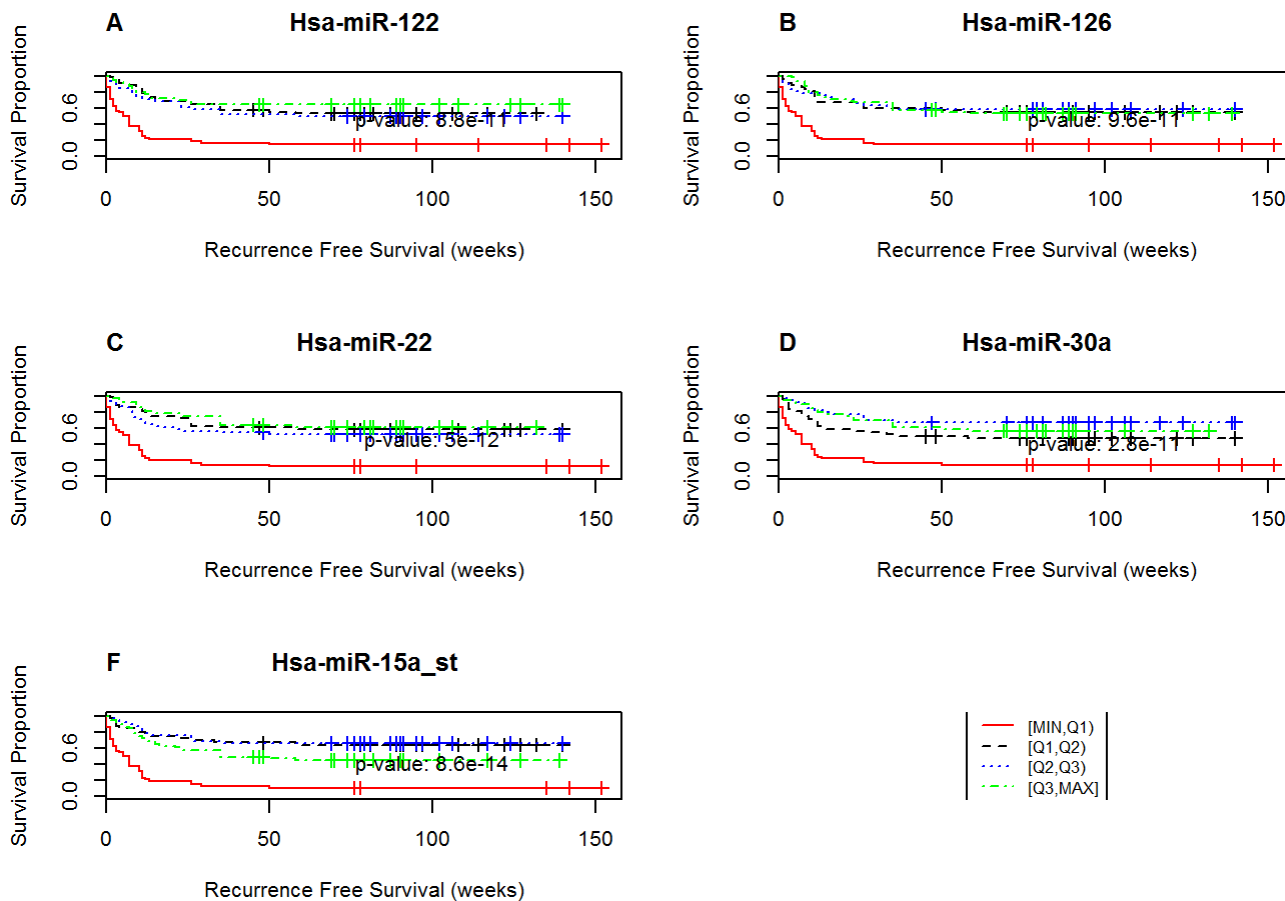
Kaplan-Meier survival curve for 5 selected miRNAs

```

#library('survival')
fit122<-survfit(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-
miR-122_st']))
fit126<-survfit(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-
miR-126_st']))
fit22<-survfit(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-
miR-22_st']))
fit15a<-survfit(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-
miR-15a_st']))
fit30a<-survfit(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-
miR-30a_st']))
#pdf('f5miRNA-survival-curvs.pdf')
par(mfrow=c(3,2))
plot(fit122,lty=1:4,col=c('red','black','blue','green'), xlab="Recurrence Free Survival (week
s)", ylab="Survival Proportion",main="Hsa-miR-122")
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdiff(Surv(spall$rfsurv,as.numeric(as.
factor(spall$recur)))~as.factor(sint176[, 'hsa-miR-122_st']))$chisq,df=3,lower.tail=F),digits=
12)))
title(main='A', adj=0)
plot(fit126,lty=1:4,col=c('red','black','blue','green'), xlab="Recurrence Free Survival (week
s)", ylab="Survival Proportion",main="Hsa-miR-126")
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdiff(Surv(spall$rfsurv,as.numeric(as.
factor(spall$recur)))~as.factor(sint176[, 'hsa-miR-126_st']))$chisq,df=3,lower.tail=F),digits=
12)))
title(main='B', adj=0)
plot(fit22,lty=1:4,col=c('red','black','blue','green'), xlab="Recurrence Free Survival (weeks
)", ylab="Survival Proportion",main="Hsa-miR-22")
text(x=100, y=0.45, paste0("p-value: ",round(pchisq(survdiff(Surv(spall$rfsurv,as.numeric(as.
factor(spall$recur)))~as.factor(sint176[, 'hsa-miR-22_st']))$chisq,df=3,lower.tail=F),digits=1
2)))
title(main='C', adj=0)
plot(fit30a,lty=1:4,col=c('red','black','blue','green'), xlab="Recurrence Free Survival (week
s)", ylab="Survival Proportion",main="Hsa-miR-30a")

```

```
text(x=100, y=0.4, paste0("p-value: ",round(pchisq(survdiff(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-miR-30a_st']))$chisq,df=3,lower.tail=F),digits=12)))
title(main='D', adj=0)
plot(fit15a,lty=1:4,col=c('red','black','blue','green'), xlab="Recurrence Free Survival (weeks)", ylab="Survival Proportion",main="Hsa-miR-15a_st")
text(x=100, y=0.4, paste0("p-value: ",round(pchisq(survdiff(Surv(spall$rfsurv,as.numeric(as.factor(spall$recur)))~as.factor(sint176[, 'hsa-miR-15a_st']))$chisq,df=3,lower.tail=F),digits=15)))
title(main='F', adj=0)
plot.new()
legend('center',cex=0.8,lty=1:4,col=c('red','black','blue','green'),c('[MIN,Q1)','[Q1,Q2)','[Q2,Q3)','[Q3,MAX]'))
```



5 selected miRNA biomarkers and their sample distribution

The ggplot2 and Rmisc packages are required. Dataset spall and dist12 are required.

```
library(lattice)
library(plyr)
library(ggplot2)
library(Rmisc)
```

```

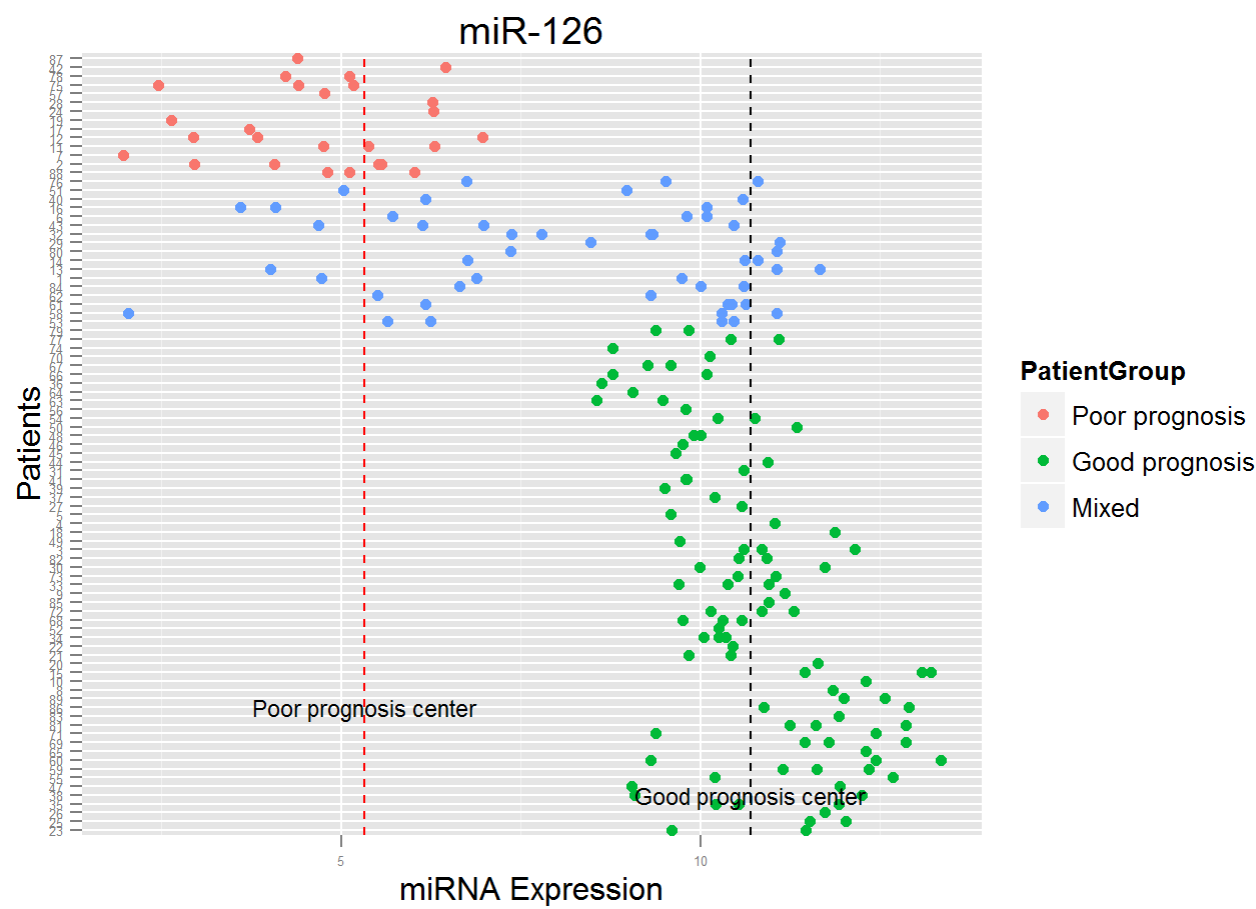
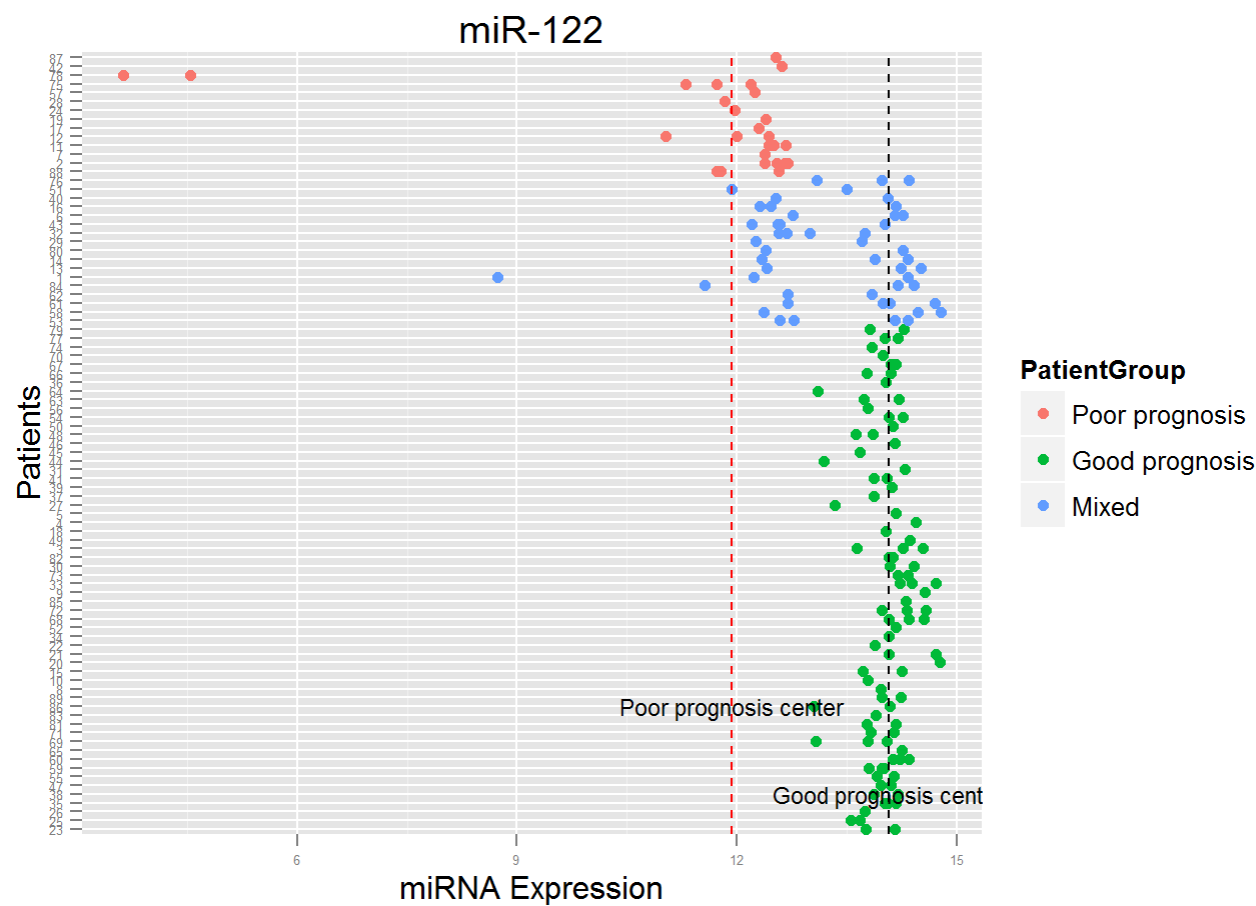
# kmeans on 847 features.
set.seed(55)
km2<-kmeans(spall[,14:860],2)

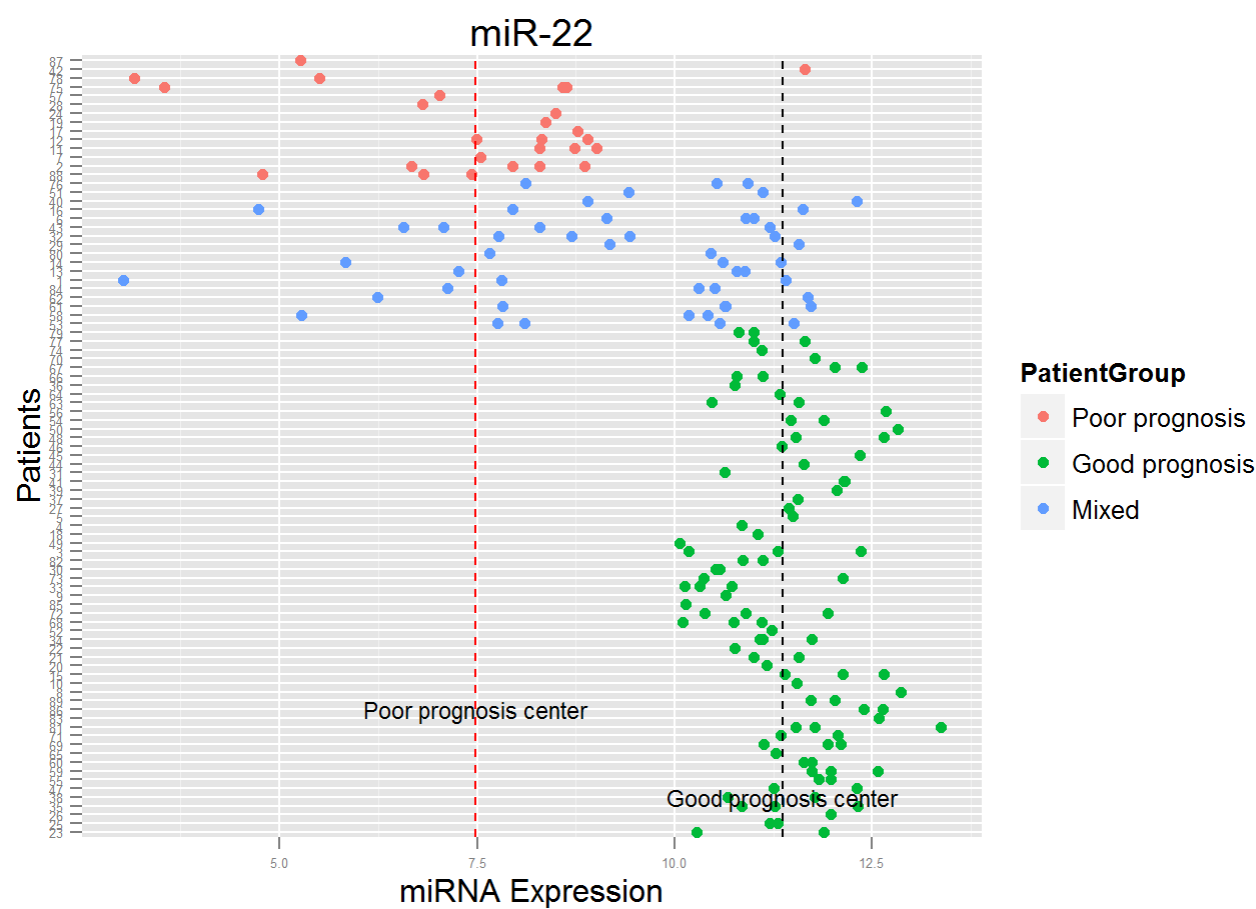
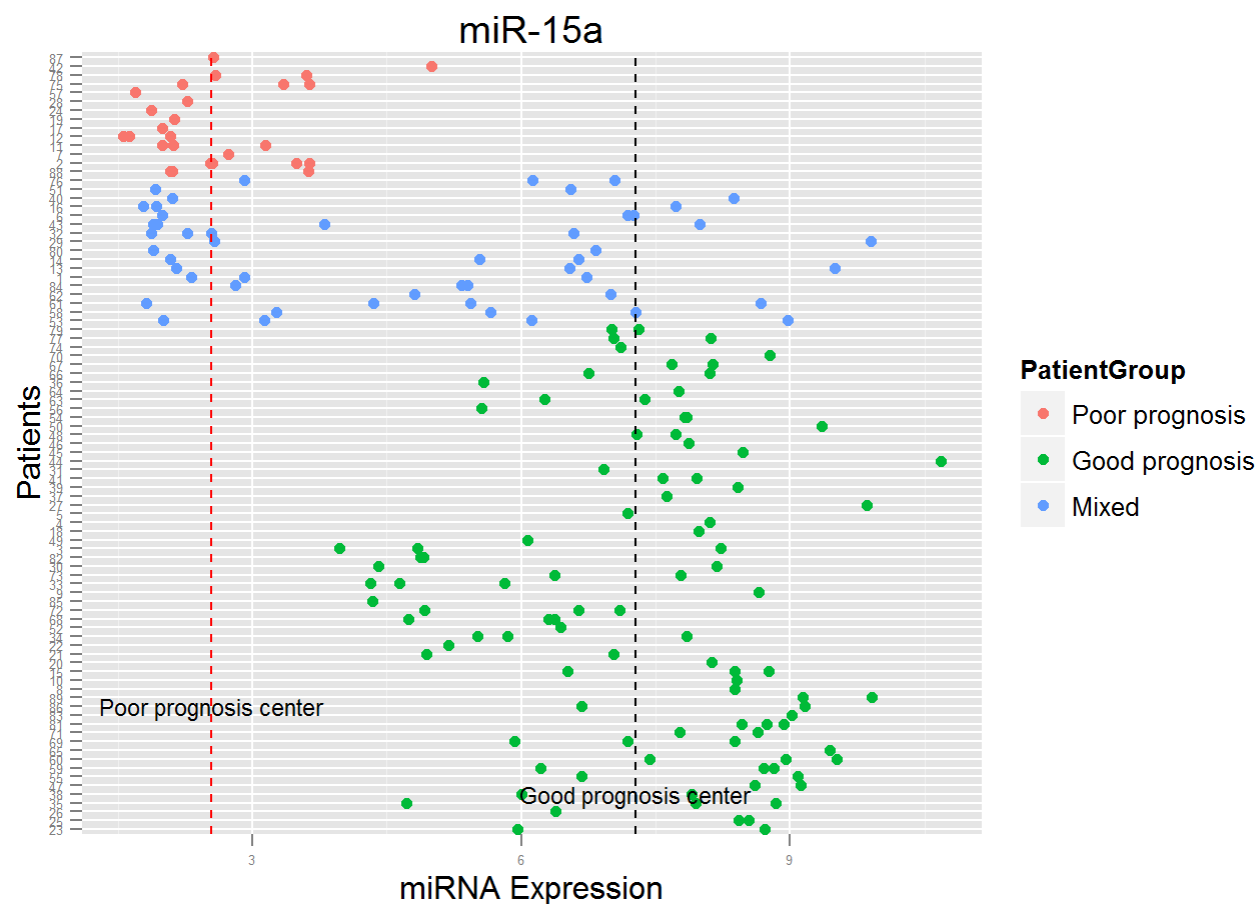
# set the data in order
t<-spall
t$pclst<-dist12$pclst
t<-t[order(t$pclst),]
# set levels in order : good prognosis group, mixed group and poor prognosis group
pidN<-as.numeric(as.factor(t$pid))
lels<-c(unique(pidN[t$pclst==2]),unique(pidN[t$pclst==3]),unique(pidN[t$pclst==1]))
t$pidN<-factor(pidN,lels)
PatientGroup<-factor(t$pclst,labels=c('Poor prognosis','Good prognosis','Mixed'))
# 5 miRNA selected from mutual information.
mir5<-c(
  "hsa-miR-122_st",
  "hsa-miR-126_st",
  "hsa-miR-22_st",
  "hsa-miR-15a_st",
  "hsa-miR-30a_st"
)
m5<-which(colnames(t) %in% mir5)

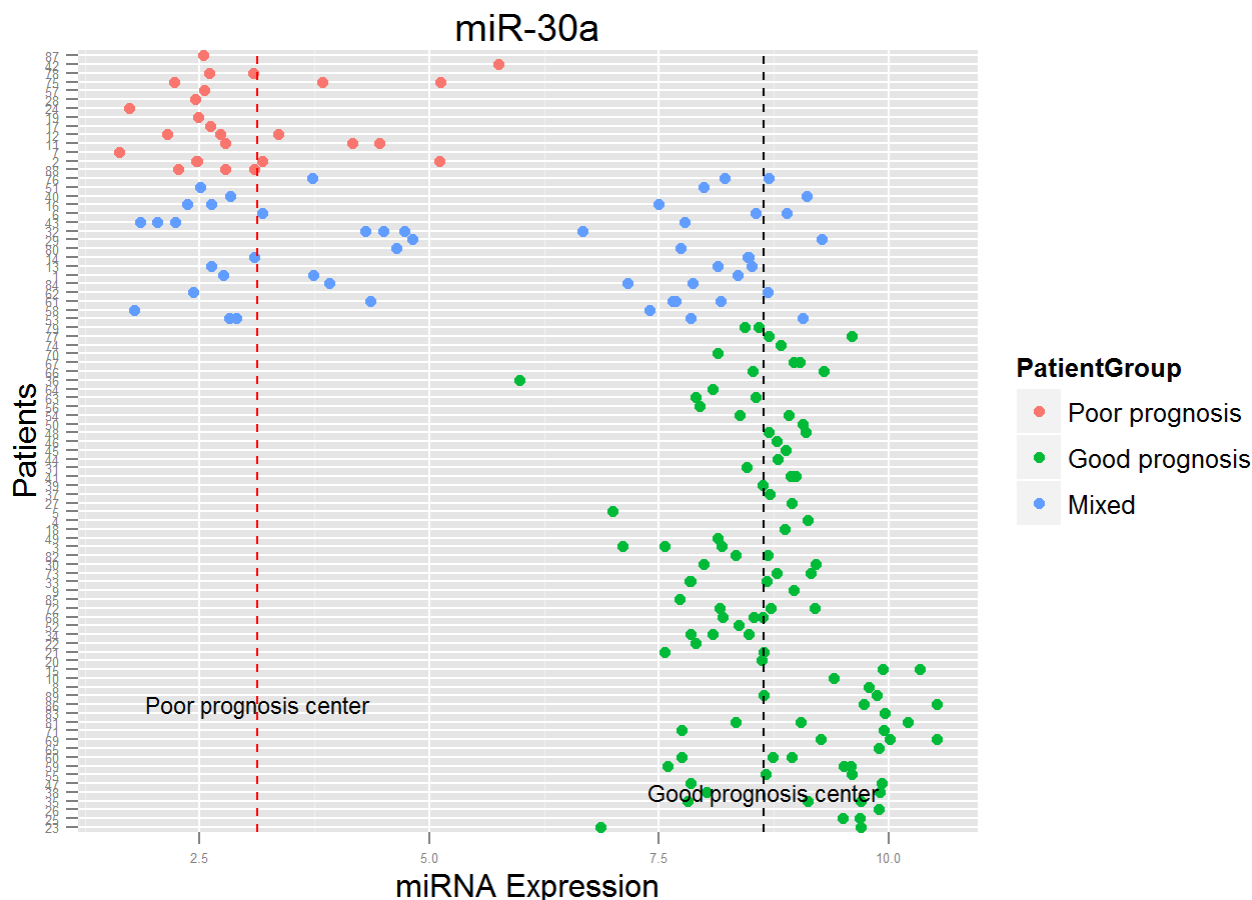
### for-loop each individual miRNA
pm<-list()
for( i in 1:5) {

p<-qplot(factor(pidN,lels),t[,m5[i]],color=PatientGroup,xlab='Patients',ylab='miRNA Expression',main=gsub("hsa-|_st","",colnames(t)[m5[i]]))+theme(axis.text=element_text(size=5))+geom_hline(aes(yintercept=c(km2$center[1,m5[i]-13],km2$center[2,m5[i]-13])),color=c('red','black'),linetype='dashed')
p<-p+annotate("text", 5, km2$center[2,m5[i]-13], label = "Good prognosis center",size=3)+annotate("text", 15, km2$center[1,m5[i]-13], label = "Poor prognosis center",size=3)+coord_flip()
pm[[i]]<-p
#pdf(paste0(gsub("hsa-|_st","",colnames(t)[m5[i]]),".pdf"))
plot(p)
}

```





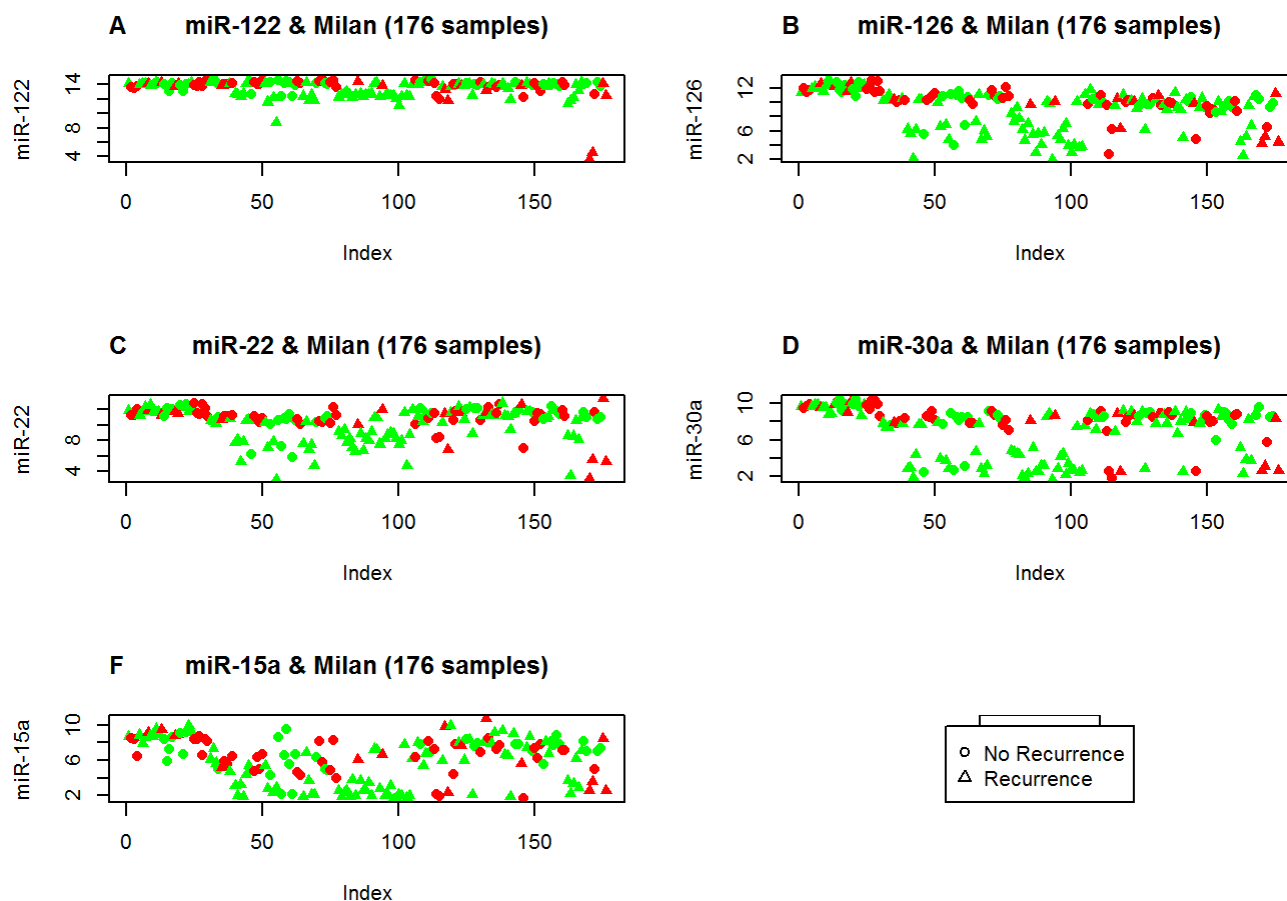


Plot 5 selected miRNAs biomarkers and cilnical variates

Require to load 'spall-dist-ptab.rdata' first.

```
#pdf('f5-milan.pdf')
par(mfrow=c(3,2))
plot(spall[, 'hsa-miR-122_st'], col=c('green', 'red')[as.numeric(as.factor(spall$milan))], pch=c(16, 17)[as.numeric(as.factor(spall$recur))], main='miR-122 & Milan (176 samples)', xlab='Index', ylab='miR-122')
title(main='A', adj=0)
plot(spall[, 'hsa-miR-126_st'], col=c('green', 'red')[as.numeric(as.factor(spall$milan))], pch=c(16, 17)[as.numeric(as.factor(spall$recur))], main='miR-126 & Milan (176 samples)', xlab='Index', ylab='miR-126')
title(main='B', adj=0)
plot(spall[, 'hsa-miR-22_st'], col=c('green', 'red')[as.numeric(as.factor(spall$milan))], pch=c(16, 17)[as.numeric(as.factor(spall$recur))], main='miR-22 & Milan (176 samples)', xlab='Index', ylab='miR-22')
title(main='C', adj=0)
plot(spall[, 'hsa-miR-30a_st'], col=c('green', 'red')[as.numeric(as.factor(spall$milan))], pch=c(16, 17)[as.numeric(as.factor(spall$recur))], main='miR-30a & Milan (176 samples)', xlab='Index', ylab='miR-30a')
title(main='D', adj=0)
plot(spall[, 'hsa-miR-15a_st'], col=c('green', 'red')[as.numeric(as.factor(spall$milan))], pch=c(16, 17)[as.numeric(as.factor(spall$recur))], main='miR-15a & Milan (176 samples)', xlab='Index', ylab='miR-15a')
title(main='F', adj=0)
plot.new()
```

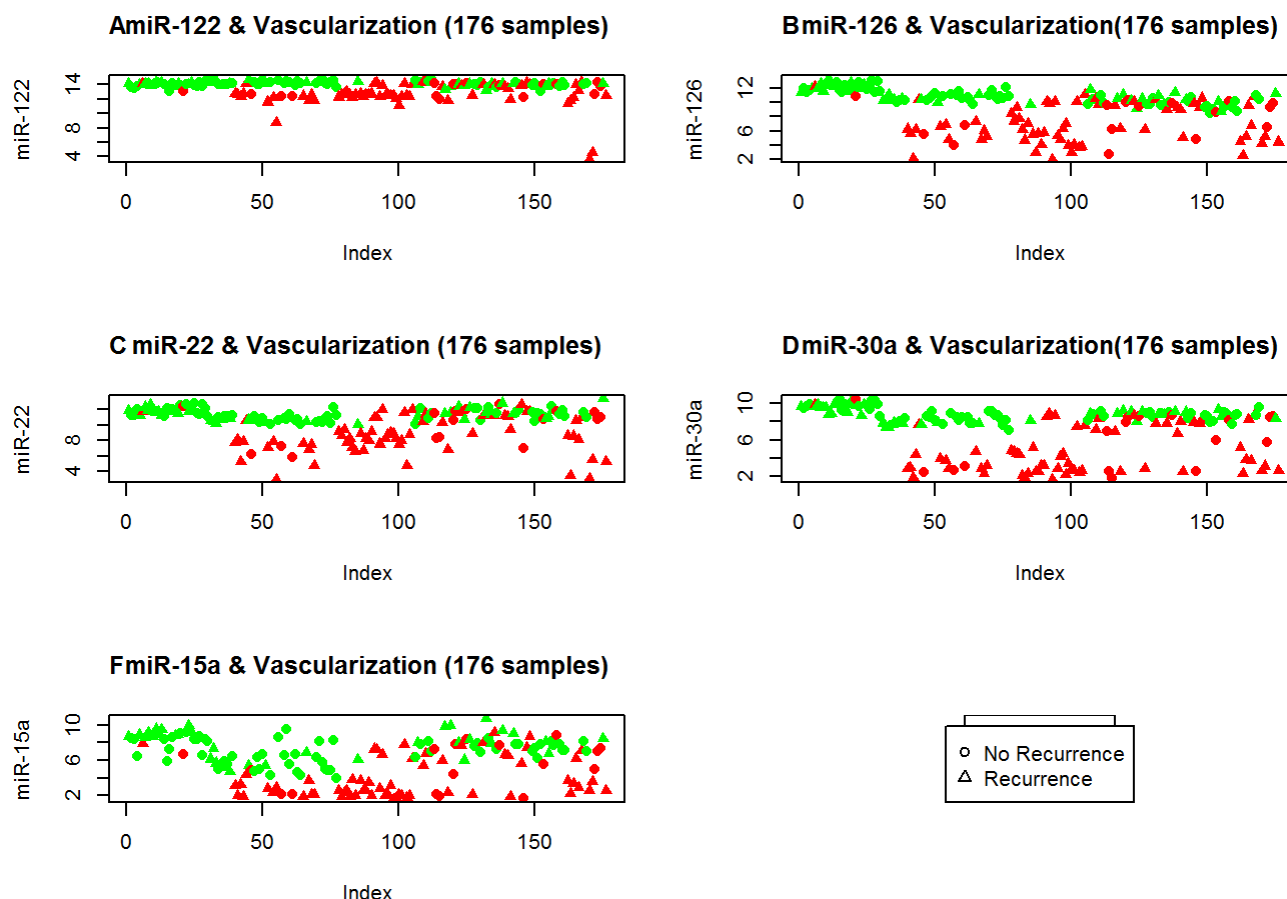
```
legend('top', col=c('green','red'),pch=16,c('Outside','Within'),title='Milan')
legend('bottom',pch=c(1,2),c('No Recurrence','Recurrence'))
```



```
#pdf('f5-vasc.pdf')
par(mfrow=c(3,2))
plot(spall[, 'hsa-miR-122_st'],col=c('green','red')[as.numeric(as.factor(spall$vasc))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-122 & Vascularization (176 samples)',xla
b='Index',ylab='miR-122')
title(main='A', adj=0)
plot(spall[, 'hsa-miR-126_st'],col=c('green','red')[as.numeric(as.factor(spall$vasc))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-126 & Vascularization(176 samples)',xlab
='Index',ylab='miR-126')
title(main='B', adj=0)
plot(spall[, 'hsa-miR-22_st'],col=c('green','red')[as.numeric(as.factor(spall$vasc))],pch=c(1
6,17)[as.numeric(as.factor(spall$recur))],main='miR-22 & Vascularization (176 samples)',xlab=
'Index',ylab='miR-22')
title(main='C', adj=0)
plot(spall[, 'hsa-miR-30a_st'],col=c('green','red')[as.numeric(as.factor(spall$vasc))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-30a & Vascularization(176 samples)',xlab
='Index',ylab='miR-30a')
title(main='D', adj=0)
plot(spall[, 'hsa-miR-15a_st'],col=c('green','red')[as.numeric(as.factor(spall$vasc))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-15a & Vascularization (176 samples)',xla
b='Index',ylab='miR-15a')
title(main='F', adj=0)
plot.new()
```

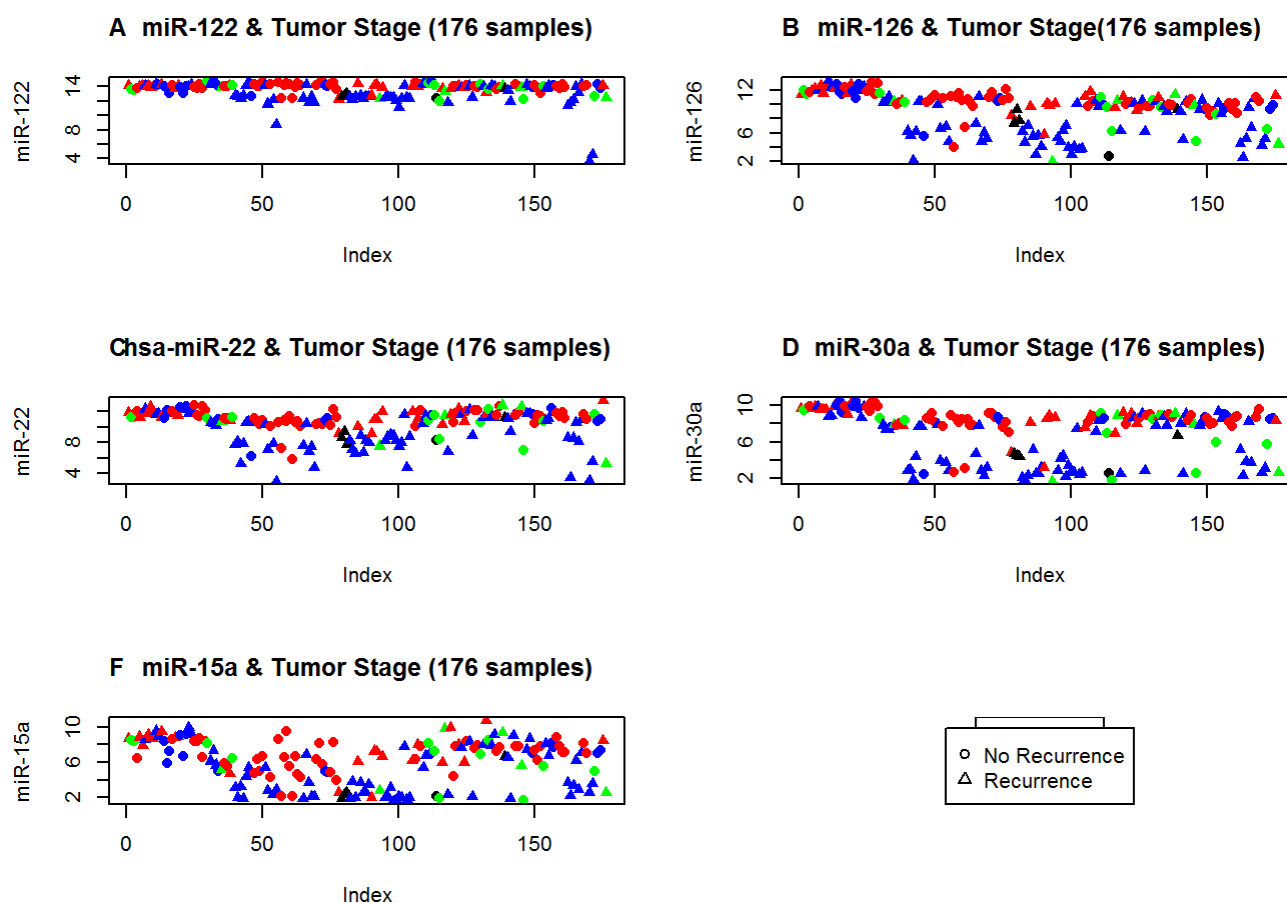


```
legend('top', col=c('green','red'),pch=16,c('N','Y'),title='Vascularization')
legend('bottom',pch=c(1,2),c('No Recurrence','Recurrence'))
```



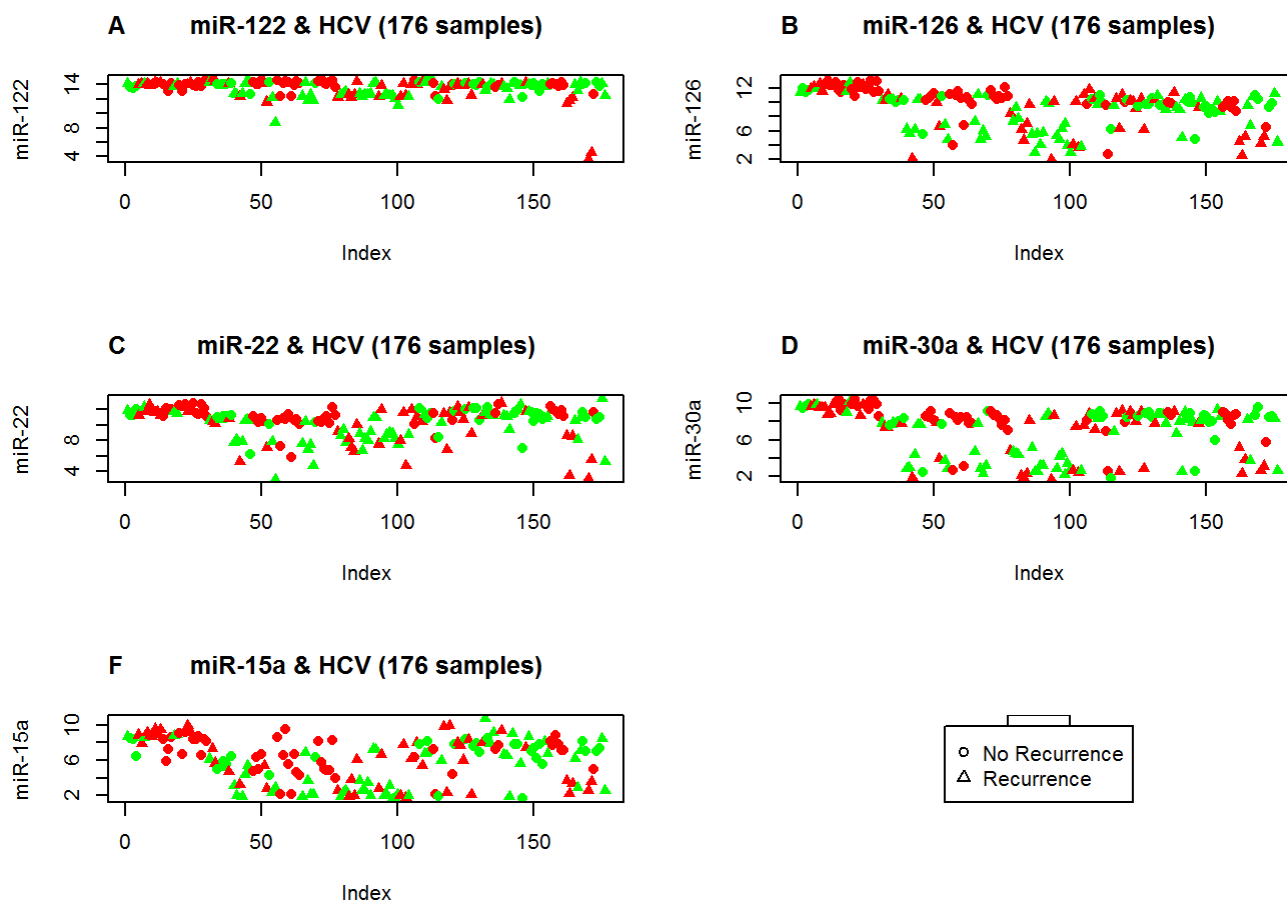
```
#pdf('f5-tumorStage.pdf')
par(mfrow=c(3,2))
plot(spall[, 'hsa-miR-122_st'],col=c('green','red','blue','black')[as.numeric(spall$Tumor.Stage)],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-122 & Tumor Stage (176 samples)',xlab='Index',ylab='miR-122')
title(main='A', adj=0)
plot(spall[, 'hsa-miR-126_st'],col=c('green','red','blue','black')[as.numeric(spall$Tumor.Stage)],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-126 & Tumor Stage(176 samples)',xlab='Index',ylab='miR-126')
title(main='B', adj=0)
plot(spall[, 'hsa-miR-22_st'],col=c('green','red','blue','black')[as.numeric(spall$Tumor.Stage)],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='hsa-miR-22 & Tumor Stage (176 samples)',xlab='Index',ylab='miR-22')
title(main='C', adj=0)
plot(spall[, 'hsa-miR-30a_st'],col=c('green','red','blue','black')[as.numeric(spall$Tumor.Stage)],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-30a & Tumor Stage (176 samples)',xlab='Index',ylab='miR-30a')
title(main='D', adj=0)
plot(spall[, 'hsa-miR-15a_st'],col=c('green','red','blue','black')[as.numeric(spall$Tumor.Stage)],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-15a & Tumor Stage (176 samples)',xlab='Index',ylab='miR-15a')
title(main='F', adj=0)
plot.new()
```

```
legend('top', col=c('green','red','blue','black'),pch=16,c('I','II','IIIA','IIIB'),title='Tumor Stage')
legend('bottom',pch=c(1,2),c('No Recurrence','Recurrence'))
```



```
#pdf('f5-HCV.pdf')
par(mfrow=c(3,2))
plot(spall[, 'hsa-miR-122_st'],col=c('green','red')[as.numeric(as.factor(spall$HCV))],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-122 & HCV (176 samples)',xlab='Index',ylab='miR-122')
title(main='A', adj=0)
plot(spall[, 'hsa-miR-126_st'],col=c('green','red')[as.numeric(as.factor(spall$HCV))],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-126 & HCV (176 samples)',xlab='Index',ylab='miR-126')
title(main='B', adj=0)
plot(spall[, 'hsa-miR-22_st'],col=c('green','red')[as.numeric(as.factor(spall$HCV))],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-22 & HCV (176 samples)',xlab='Index',ylab='miR-22')
title(main='C', adj=0)
plot(spall[, 'hsa-miR-30a_st'],col=c('green','red')[as.numeric(as.factor(spall$HCV))],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-30a & HCV (176 samples)',xlab='Index',ylab='miR-30a')
title(main='D', adj=0)
plot(spall[, 'hsa-miR-15a_st'],col=c('green','red')[as.numeric(as.factor(spall$HCV))],pch=c(16,17)[as.numeric(as.factor(spall$recur))],main='miR-15a & HCV (176 samples)',xlab='Index',ylab='miR-15a')
title(main='E', adj=0)
```

```
plot.new()
legend('top', col=c('green','red'),pch=16,c('N','Y'),title='HCV')
legend('bottom',pch=c(1,2),c('No Recurrence','Recurrence'))
```



```
#pdf('f5-recur.pdf')
par(mfrow=c(3,2))
plot(spall[, 'hsa-miR-122_st'],col=c('green','red')[as.numeric(as.factor(spall$recur))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-122 & Recurrence (176 samples)',xlab='
Index',ylab='miR-122')
title(main='A', adj=0)
plot(spall[, 'hsa-miR-126_st'],col=c('green','red')[as.numeric(as.factor(spall$recur))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-126 & Recurrence (176 samples)',xlab='
Index',ylab='miR-126')
title(main='B', adj=0)
plot(spall[, 'hsa-miR-22_st'],col=c('green','red')[as.numeric(as.factor(spall$recur))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-22 & Recurrence (176 samples)',xlab='
Index',ylab='miR-22')
title(main='C', adj=0)
plot(spall[, 'hsa-miR-30a_st'],col=c('green','red')[as.numeric(as.factor(spall$recur))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-30a & Recurrence (176 samples)',xlab='
Index',ylab='miR-30a')
title(main='D', adj=0)
plot(spall[, 'hsa-miR-15a_st'],col=c('green','red')[as.numeric(as.factor(spall$recur))],pch=c(
16,17)[as.numeric(as.factor(spall$recur))],main='miR-15a & Recurrence (176 samples)',xlab='
Index',ylab='miR-15a')
title(main='E', adj=0)
```

```
plot.new()  
legend('top', col=c('green','red'),pch=c(16,17),c('No Recurrence','Recurrence'),title='HCC Re  
currence')
```

