

PCA with PERMANOVA

Data import

```
set.seed(27)
data_read<-read.table("Merged_Fold_change_all_values.txt",sep="\t",header=TRUE)
data_read$SampleType<-as.numeric(data_read$SampleType)
data_read$SampleID<-as.numeric(data_read$SampleID)
data_read$Drug<-as.numeric(data_read$Drug)
```

PCA and PerMANOVA on all data

```
iris_c <- scale(data_read[, -1:-4])
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-6
```

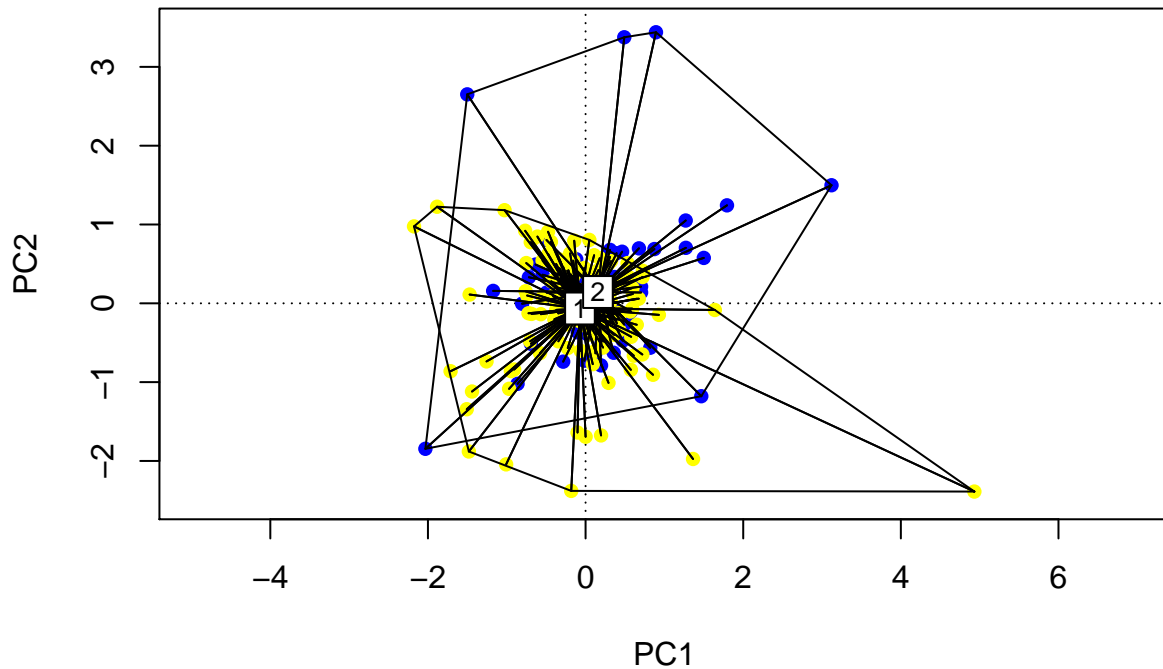
```
pca <- rda(iris_c)

# plot
plot(pca, type = 'n', display = 'sites', xlim = c(-4,6), ylim=c(-2.5,3.5))
cols <- c('yellow', 'blue')
points(pca, display='sites', col = cols[data_read$SampleType], pch = 16)
ordihull(pca, groups=data_read$SampleType)
ordispider(pca, groups = data_read$SampleType, label = TRUE)

# PerMANOVA - partitioning the euclidean distance matrix by species
adonis(iris_c ~ SampleType, data = data_read, method='eu')
```

```
##
## Call:
## adonis(formula = iris_c ~ SampleType, data = data_read, method = "eu")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## SampleType    1      245.1  245.131  4.7843 0.01856 0.001 ***
## Residuals   253    12962.9   51.237      0.98144
## Total       254    13208.0           1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Coordinates of centroids
a<-ordispider(pca, groups = data_read$SampleType, label = TRUE)
```



```
a[1,]
```

```
##      PC1      PC2
## 0.1565998 0.1436174
```

```
a[255,]
```

```
##      PC1      PC2
## -0.07158849 -0.06565365
```

```
#Euclidean distance between points
sqrt(((0.1436174 -0.06565365)^2) + ((0.1565998 -0.07158849)^2))
```

```
## [1] 0.1153485
```

The centroid of the patient samples (blue) is (0.1565998,0.1436174). The centroid of the cell line samples (yellow) is (-0.07158849,-0.06565365). The Euclidean distance between these centroids is 0.1153485.

PCA and PerMANOVA - outlying cell line and patient samples removed

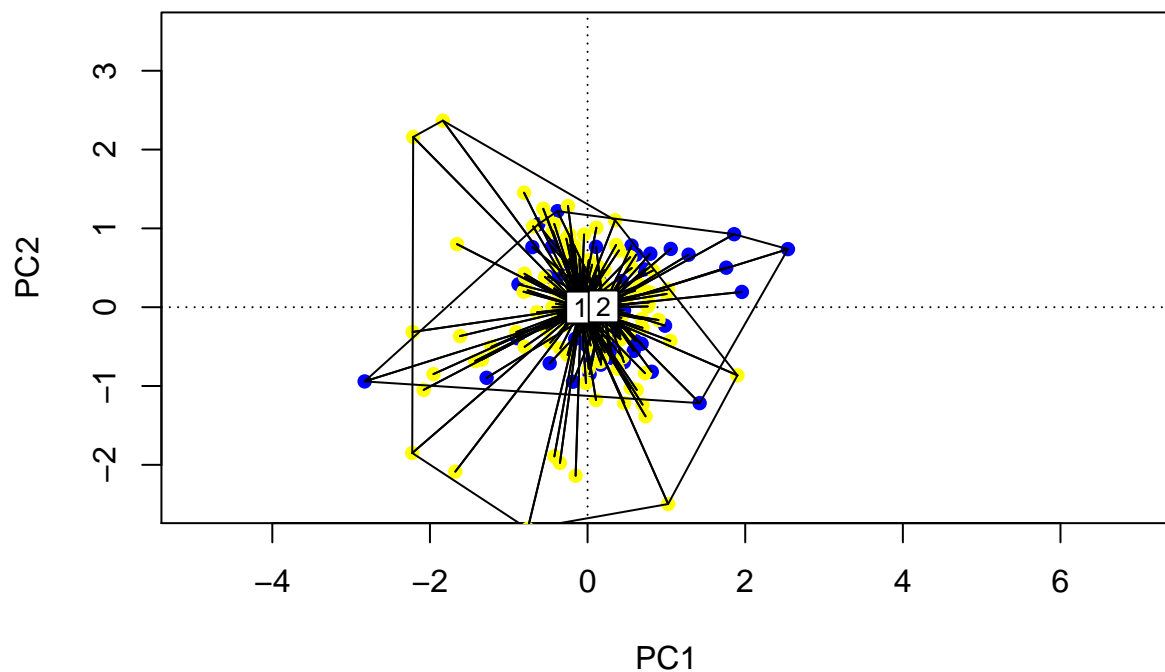
```
set.seed(27)
data_read<-read.table("Merged_Fold_change_all_values.txt",sep="\t",header=TRUE)
data_read$SampleType<-as.numeric(data_read$SampleType)
data_read$SampleID<-as.numeric(data_read$SampleID)
data_read$Drug<-as.numeric(data_read$Drug)
data_read<-subset(data_read, data_read$SampleID!=32)
data_read<-subset(data_read, data_read$SampleID!=37)
data_read<-subset(data_read, data_read$SampleID!=39)
iris_c <- scale(data_read[, -1:-4])
library(vegan)
pca <- rda(iris_c)

# plot
plot(pca, type = 'n', display = 'sites', xlim = c(-4,6), ylim=c(-2.5,3.5))
cols <- c('yellow', 'blue')
points(pca, display='sites', col = cols[data_read$SampleType], pch = 16)
ordihull(pca, groups=data_read$SampleType)
ordispider(pca, groups = data_read$SampleType, label = TRUE)

# PerMANOVA - partitioning the euclidean distance matrix by species
adonis(iris_c ~ SampleType, data = data_read, method='eu')
```

```
##
## Call:
## adonis(formula = iris_c ~ SampleType, data = data_read, method = "eu")
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##          Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## SampleType  1      236.6 236.601  4.6189 0.01904 0.001 ***
## Residuals 238    12191.4  51.224      0.98096
## Total      239    12428.0      1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Coordinates of centroids
a<-ordispider(pca, groups = data_read$SampleType, label = TRUE)
```



```
a[1,]
```

```
##          PC1          PC2
## 0.20285899 0.01003338
```

```
a[240,]
```

```
##          PC1          PC2
## -0.083530172 -0.004131391
```

```
#Euclidean distance between points
sqrt(((0.20285899 -0.083530172)^2) + ((0.01003338 -0.004131391)^2))
```

```
## [1] 0.1194747
```

The centroid of the patient samples (blue) is (0.20285899, 0.010033377). The centroid of the cell line samples (yellow) is (-0.08353017, -0.004131391). The Euclidean distance between these points is 0.119474 - removing the outliers moves the centroids slightly further apart.