Integrative Omics for Precision Medicine

Ugwu Paschal | ugwupaschal@gmail.com

November 16, 2022

Overview

In this project, I have 2 omics datasets (transcriptomics and epigenomics) from several asthma cases. The epigenomic data here is a measure of the DNA methylation of CpG dinucleotides.

Project Objective: DNA methylation is known to suppress gene expression. My task is to see if this is true for Asthma patients.

1. I will perform differential expression between the healthy and asthmatic patients to identify the highly expressed and lowly expressed genes.

2. I will visualize the methylation pattern of the highly expressed genes and lowly expressed genes.

3. I will check if methylation affected expression.

4. I will perform functional enrichment for genes with high expression and low methylation and vice versa to observe the type of cells that are enriched.

During this project, I will build an R script using R version 4.2.2 and RStudio version 2022.07.2+576. The live note is available in my github repository: https://github.com/paschalugwu/HB_Transcriptomics_FinalProject.R/raw/main/Final_Asthma.R

I seek to answer the four questions above. For any section when any of the questions is answered, it will be well indicated.

Prior to this project, I ensured I successfully installed R, RStudio, and all packages needed for this project. Next I loaded the libraries.

```
## Load required libraries
library(readr)
library(ggplot2)
library(FactoMineR)
library(factoextra)
library(devtools)
library(ggpubr)
library(tidyr)
library(dplyr)
library(data.table)
library(tidyverse)
library (useful)
```

The first data I will be importing is the expression datasets that will be used to perform differential expression between the healthy and asthmatic patients

```
## Import .csv file in R
asthma <- read_delim("Asthma.txt", delim = "\t",
                     escape_double = FALSE, trim_ws = TRUE)
```

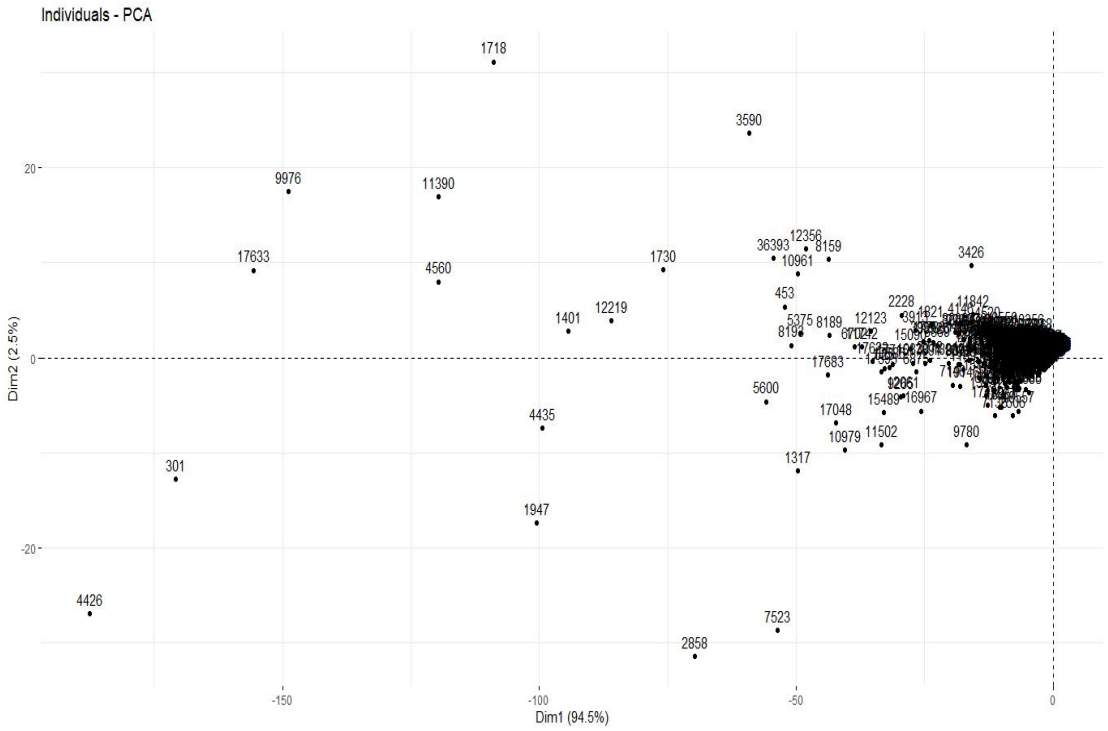
1. I will perform differential expression between the healthy and asthmatic patients to identify the highly expressed and lowly expressed genes

Since I have widely spread values, I will set scale to true, so that I can capture all my points in the plot

```
# Set scale
# before the PCA analysis
> res.pca <- prcomp(asthma,  scale = TRUE)
```

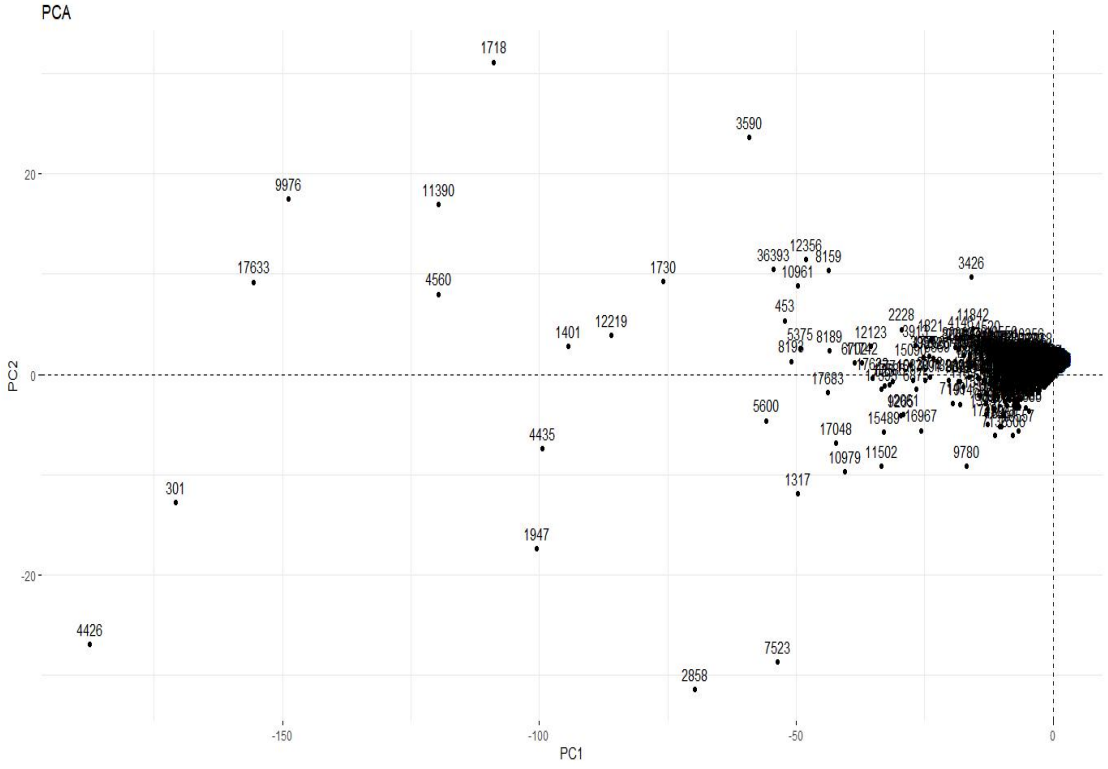
Default plot: here I have each point representing the gene that occupied the position indicated by the number it posses. For now, I do not know where these genes are from. But I already can see the up regulated genes (upwards) and down regulated genes (downwards).

```
> fviz_pca_ind(res.pca)
```



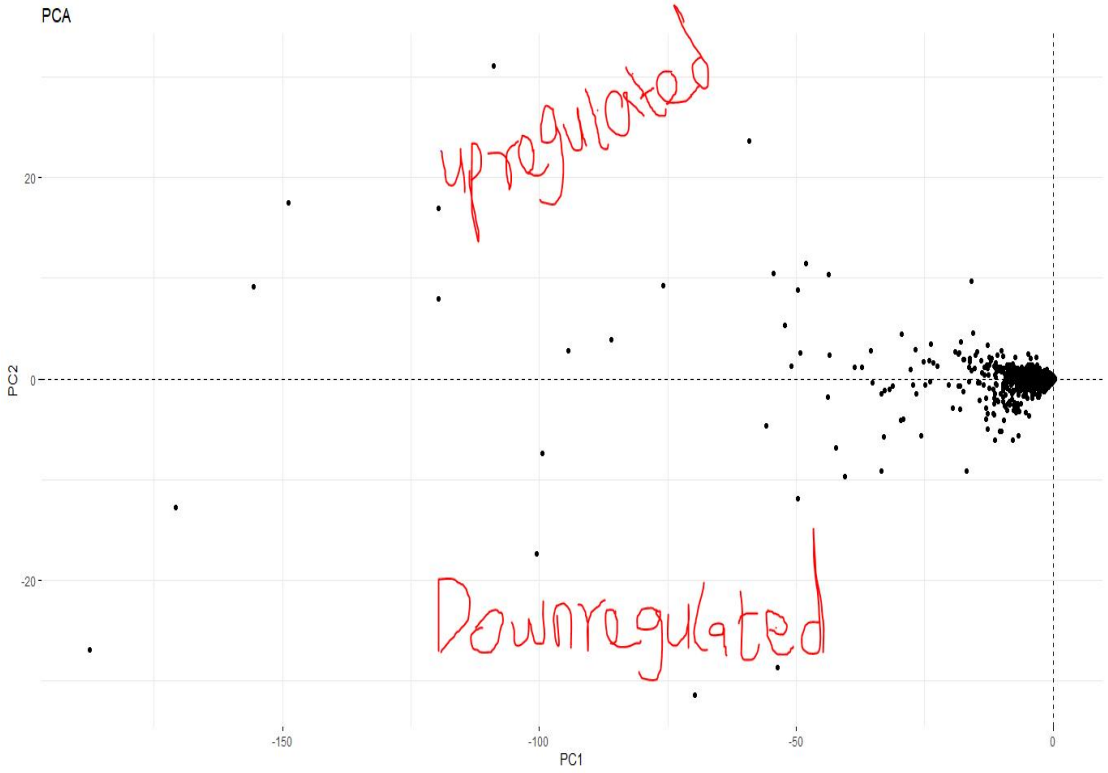
Let me change the title and axis labels.

```
> fviz_pca_ind(res.pca) +
  labs(title ="PCA", x = "PC1", y = "PC2")
```



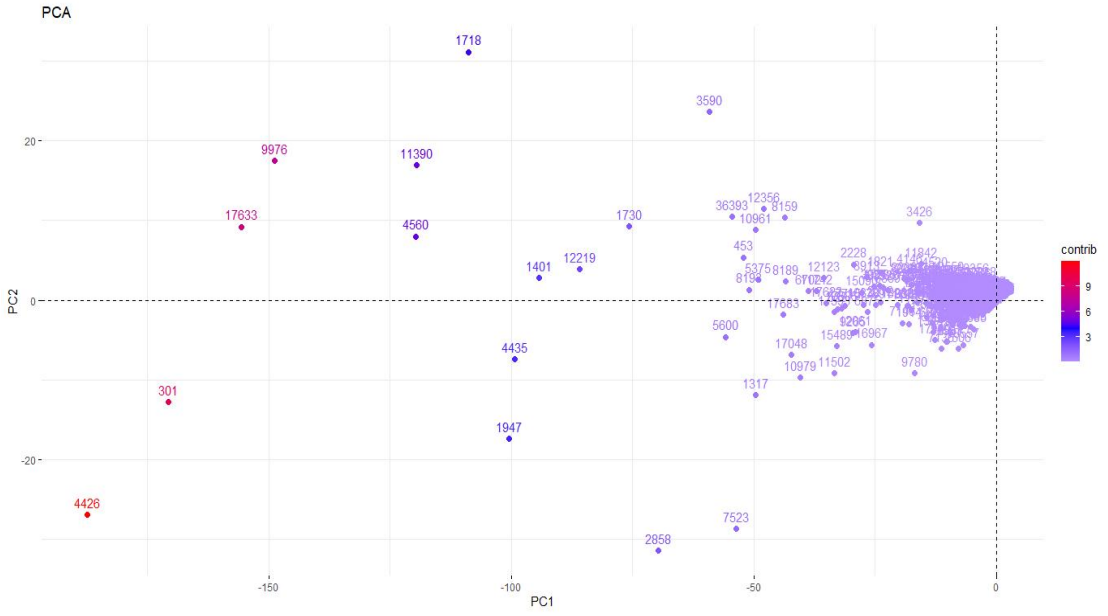
I will use points only to see clearly the regulation of the genes. I have also indicated the position of my up and down regulated genes.

```
> fviz_pca_ind(res.pca, geom="point") +
  labs(title ="PCA", x = "PC1", y = "PC2")
```



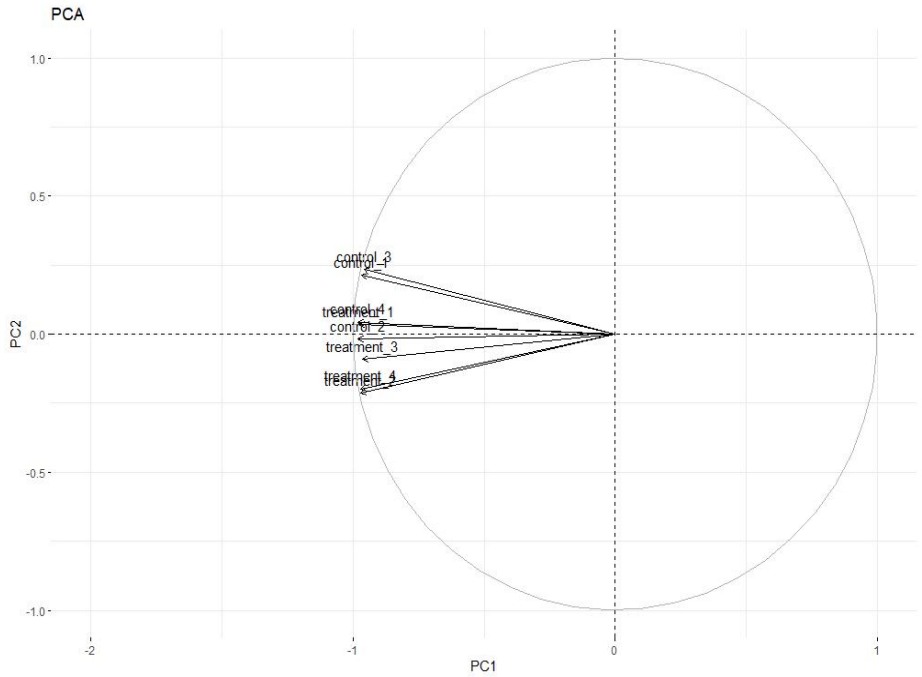
I want to use contributions to view my plot. This will provide me the quality of the individuals on the factor map. This factor map shows us how much the genes are expressed based on the contribution level. The most expressed posses the “red” coloration in contrast to the lowly expressed with “light blue” coloration.

```
> fviz_pca_ind(res.pca, col.ind="contrib", pointsize = 2) +
  labs(title ="PCA", x = "PC1", y = "PC2") +
  scale_color_gradient2(low="white", mid="blue",
                        high="red", midpoint=4)
```



I will now use “fviz_pca_var()” function to plot few graphs of variables. This will aid my inference.

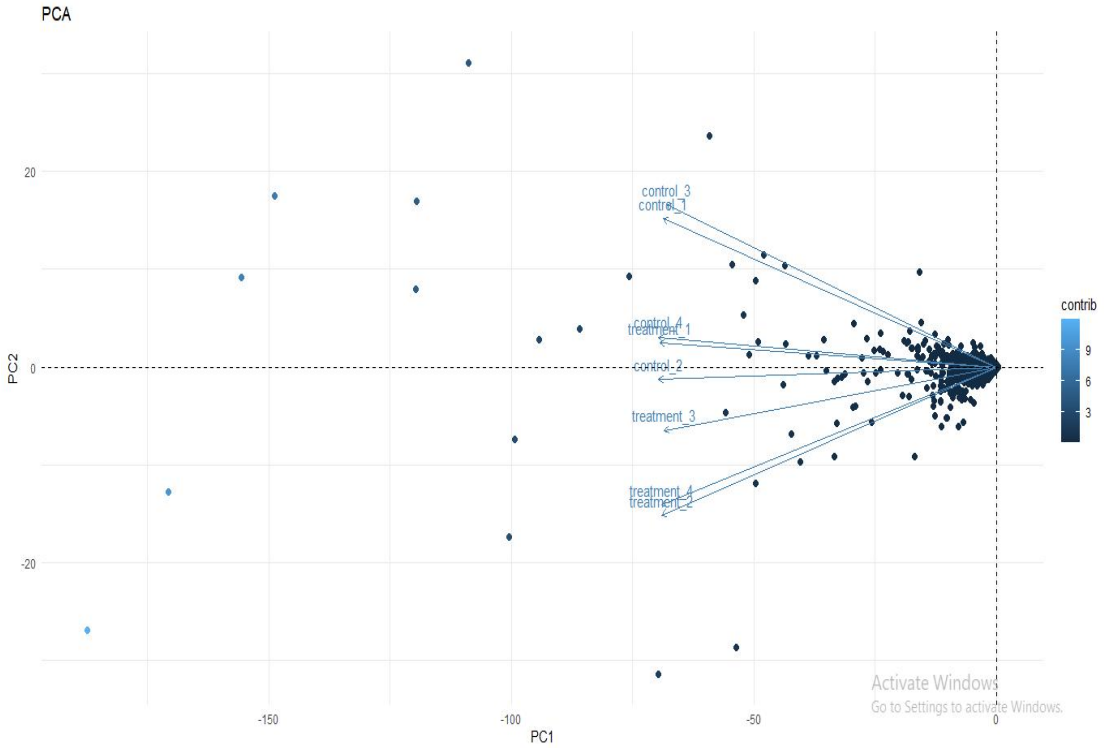
```
> fviz_pca_var(res.pca) +
  labs(title ="PCA", x = "PC1", y = "PC2") +
  xlim(-2, 1)
```



Note: from the graph above, we can deduce that three of all the experimental controls contained genes that were upregulated. The first treated sample genes were also upregulated. While other three treated samples were down regulated. Control_2 sample was with a bit more downregulated genes.

I will use “fviz_pca_biplot()” for Biplot of individuals of variables. This will aid visualization.

```
>fviz_pca_biplot(res.pca, label ="var", col.ind="contrib", pointsize = 2) +
  labs(title ="PCA", x = "PC1", y = "PC2") +
  theme_minimal()
```



NOTE: the above visualizations were helpful for me to solve the first problem. I have performed differential expression between the healthy and asthmatic patients to identify the highly expressed and lowly expressed genes

METHYLATION ANALYSIS

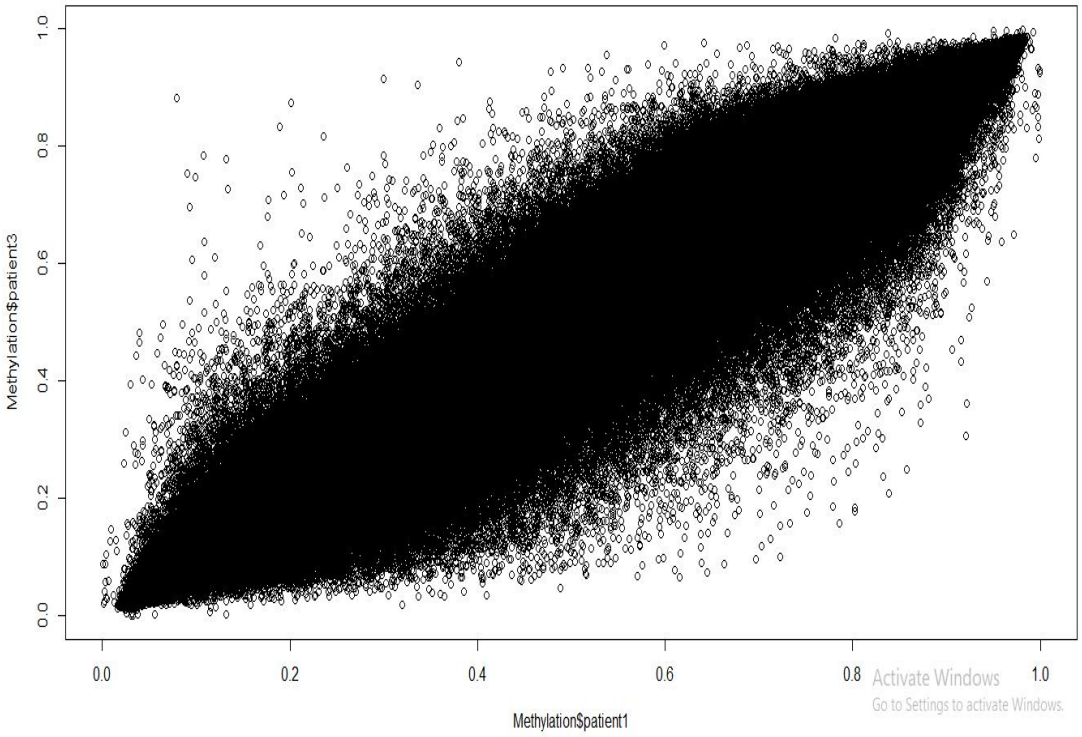
```
## Import .csv file in R
Methylation <- read_delim("Asthma_Methylation.txt",
                          delim = "\t", escape_double = FALSE,
                          trim_ws = TRUE)
```

For the purpose of visualization, I will like to use any two columns that correlate well.

```
> cor(Methylation)
```

NOTE: since, patient1 and patient3 correlate highest (0.9737972). I will use both for my plot. The plot that was obtained has all the genes clustered together. This indicates similarity I the genes. Another important thing to note about this plot is that it steeps from down upward. This is a clear indication that the genes present in patients whose DNA were methylated wereupregulated.

```
> plot (Methylation$patient1, Methylation$patient3)
```



So we can have same results

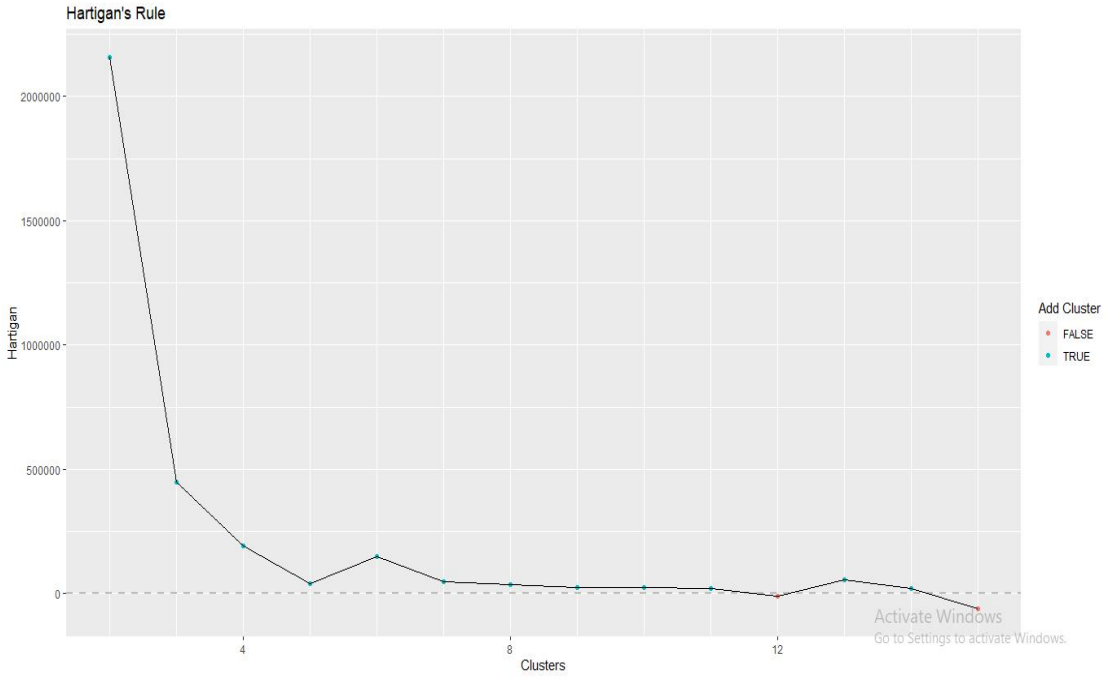
```
> set.seed(102)
```

Choosing the right number of clusters

```
> MethylationBEST <- FitKMeans(Methylation, max.clusters=15)
```

NOTE: from the output, the maximum number of clusters is 11

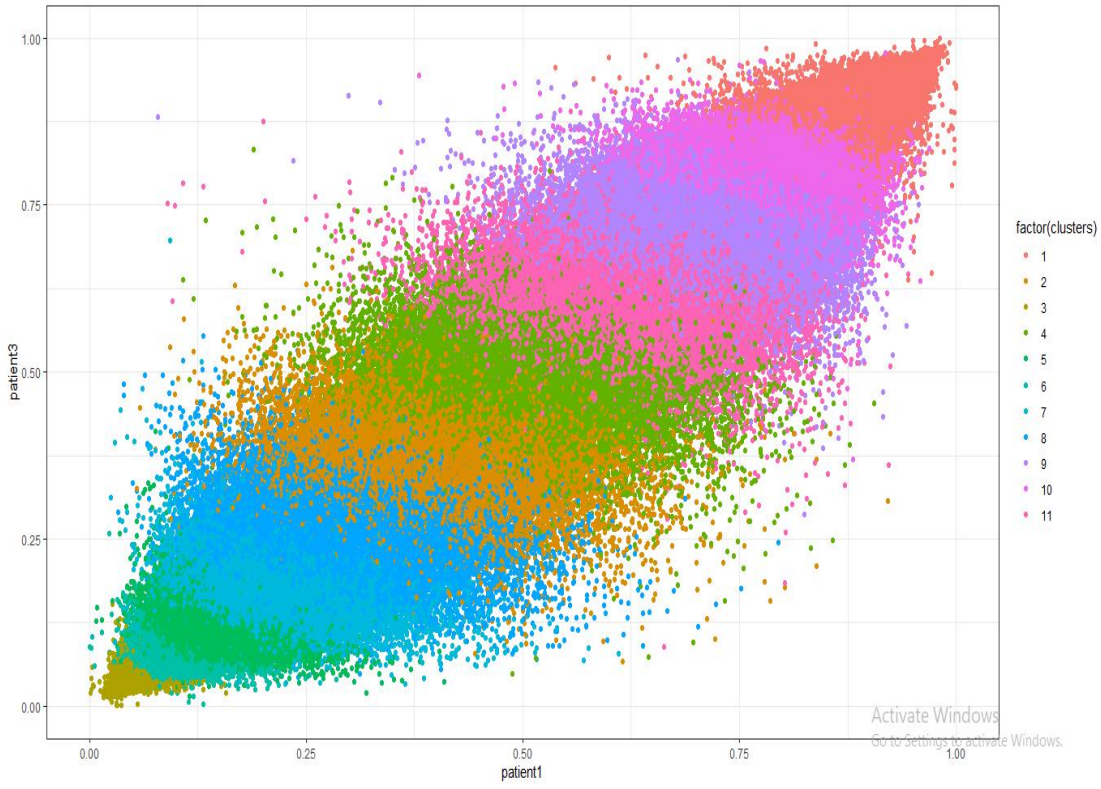
```
> PlotHartigan(MethylationBEST)
```



Perform KMEAN Clustering
> MethylationK3 <- kmeans (x = Methylation, centers = 11)

Well, I can also add this cluster information to my dataset
Methylation$clusters <- c(MethylationK3$cluster)

# To visualize this cluster information and compare patients
ggplot(Methylation, aes(x = patient1, y = patient3, color = factor(clusters))) + geom_point() + theme_bw()



Answer2: We have visualized our genes in the plots above.

Answer3: Methylation led to high expression of genes as compared to the treatment experiment performed previously.

Answer4: From the correlation and clustering we got it our plots. It is evident that from patients 1,2, and 3 had most of of their genes up regulated by DNA methylation they underwent. Up and down regulation is the process where the number of receptors is dependent upon receptor bonding.

In pharmacodynamics, upregulation refers to a system where we have more and more receptors being expressed in a cell surface because there is very little activity. This net result is that if they have low concentration of the drug, the cell becomes sensitive to the drug because it has more receptors.

Downregulation refers to the opposite. There is fewer receptors because they are always used up and broken down. The effect on the drug is reduced drug effect over time.