

76552 | Workshop in Computational BioSkills | Final Project

Guy Lutsker 207029448

Introduction

The project I chose was the project of Shimon Balsam, which was titled: “Resolving the Tissue Topology of the Alzheimer’s Brain”. The idea was to explore the brain tissue of Alzheimer(AD) patients VS healthy(WT) brain tissue by examining 3 genes which have been associated with Alzheimer’s disease - Mfge8, Slc38a1, Myoc. The data I was given was a CSV file containing data on pictures taken of the brain tissue. These are the pictures the data was created from:

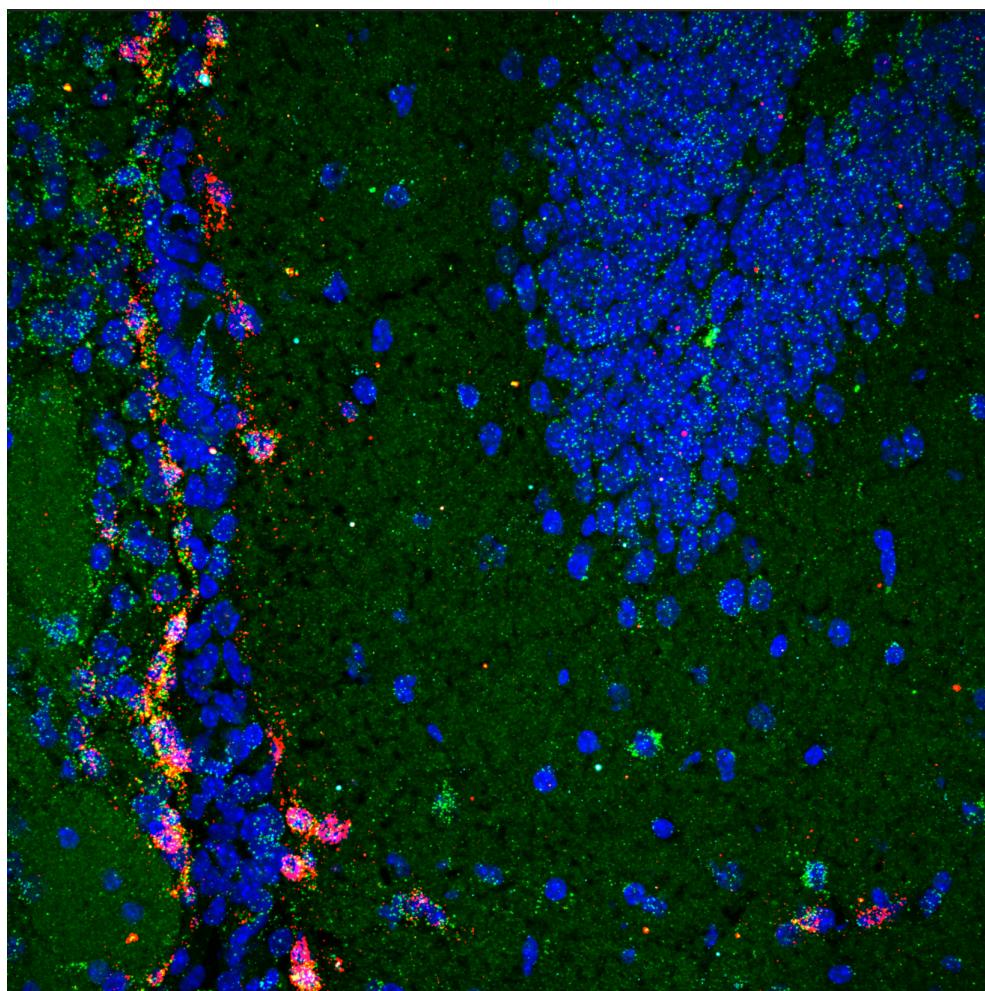


Figure 1: Picture of WT cells with fluorescent coloring for the 3 markers

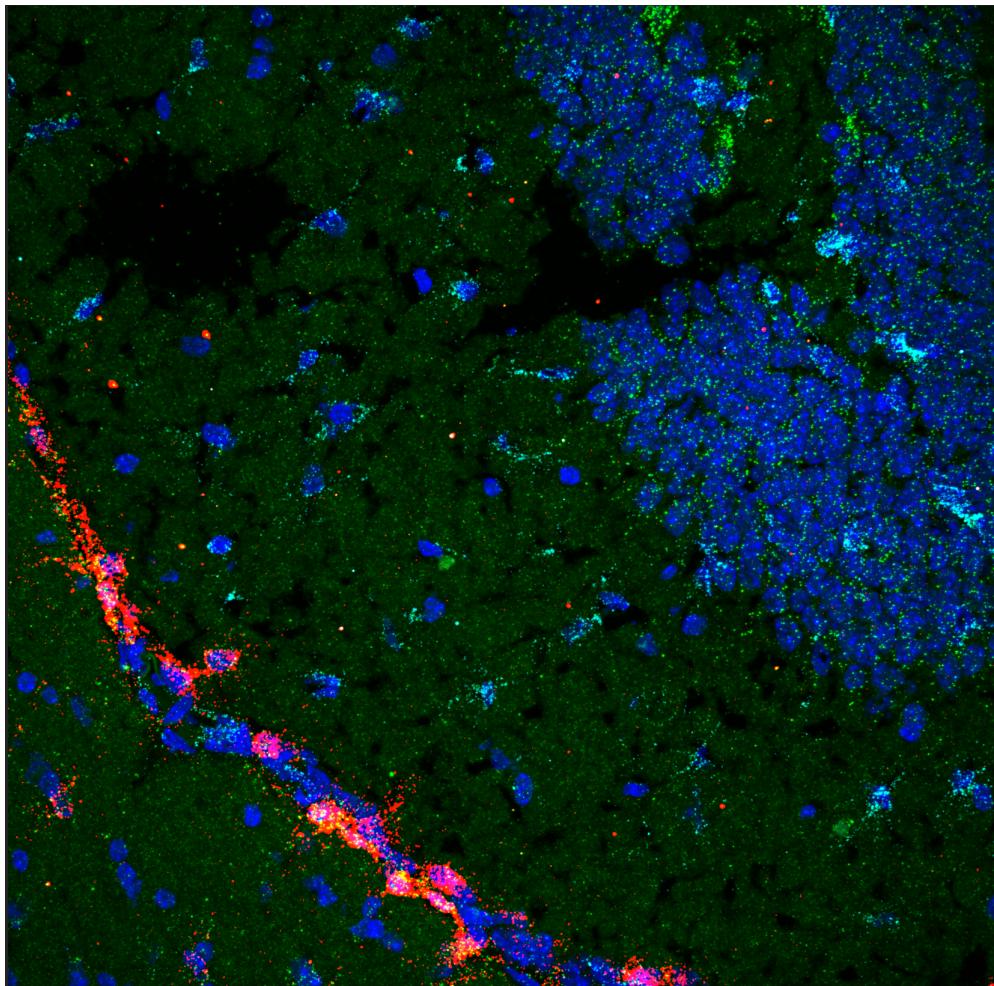


Figure 2: Picture of AD cells with fluorescent coloring for the 3 markers

As we can see there is a clear difference in the tissues. For example we can clearly see distinct “groups” of cyan colored markers in the AD brain tissue sample, whereas in the WT such groups are not visible. But to learn exactly how AD brain tissue differs from WT brain tissue, we need to analyze the data. So let's go ahead and make some graphs!

Exploratory Graph : HeatMap of Cells with Respect to Genes

In this Section I decided to create a HeatMap describing the changes in cells of Alzheimer's brain tissue, VS healthy cells. Specifically I tried to show the difference in the way the 3 gene markers appear. The Graph:

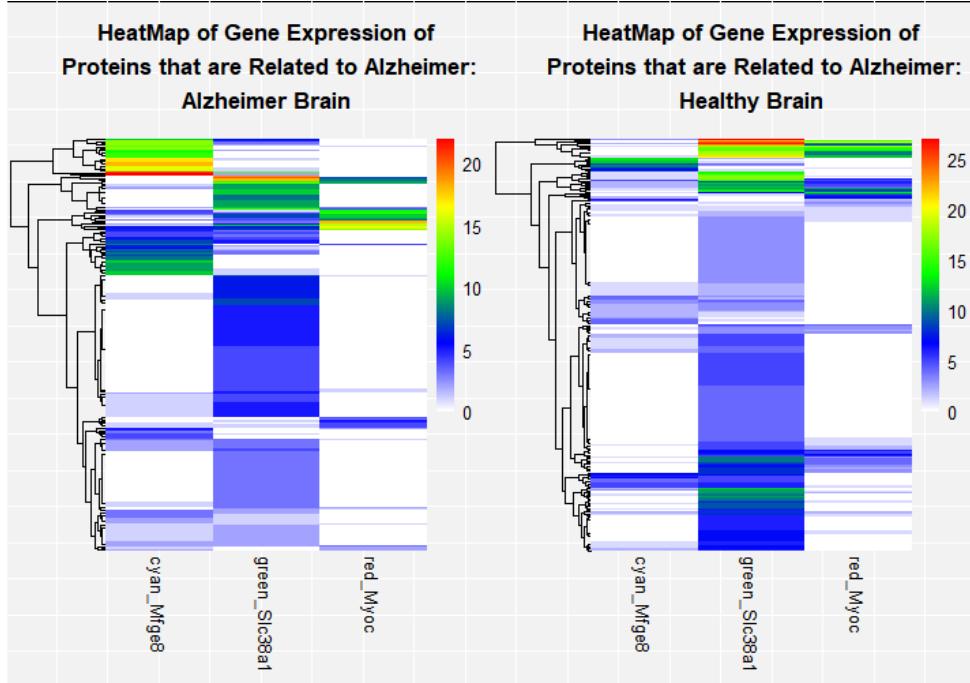


Figure 3: HeatMap of WT/AD

Explanation of results:

In this graph we can look at the data “from a birds eye view” and see what are the largest differences between the topology of the brain tissue of healthy WT brains VS an Alzheimer’s brain tissue. Since the data itself was rather sparse, I chose to filter some of the data in the following manner: I chose to show in both HeatMaps only cells that expressed at least 2 genes - i.e. I took the sum of the gene expression in each cell and filtered out these with less than 3 genes expressed. As we can see the main result from this graph is that Alzheimer’s brain tissue expresses Mfge8, and Myoc in larger amount, whereas it expresses Slc38a1 in a smaller amount than a WT brain tissue. To explore these differences more closely, we can take a look at the Explanatory Graph in section 2.

The code that generated the exploratory Graph:

```
#                                         Graph One:  
  
data$sum = data$cyan_Mfge8+data$green_Slc38a1+data$red_Myoc temp<-data [ data$sum > 2 ,]  
  
# split table for sick samples and healthy ones  
data1<-temp [ temp$imageNumber=="AD" ,]  
data2<-temp [ temp$imageNumber=="WT" ,]  
  
# nice palette:  
my_palette<-  
colorRampPalette(c("white","blue","green","yellow","red"))(n = 367)  
  
# adjust the data as matrix for heatmap  
s1 = as.matrix(data1[,c("cyan_Mfge8","green_Slc38a1","red_Myoc")])  
s2 = as.matrix(data2[,c("cyan_Mfge8","green_Slc38a1","red_Myoc")])  
  
# creating heat map for sick and one for healthy  
a <- list(pheatmap(s1,  
                    main = "HeatMap of Gene Expression of\nProteins that are Related to Alzheimer:\\nAlzheimer Brain",  
                    # heat map title  
                    notecol="black", # change font color of cell labels to black  
                    #density.info="none", # turns off density plot inside color legend  
                    trace="none", # turns off trace lines inside the heat map  
                    margins =c(14,2), # widens margins around plot  
                    col=my_palette, # use on color palette defined earlier  
                    #breaks=col_breaks, # enable color transition at specified limits  
                    dendrogram="row", # only draw a row dendrogram  
                    cluster_cols = FALSE,  
                    show_rownames = FALSE,  
                    Colv="NA") # turn off column clustering  
[[4]])  
  
a [[2]] <- pheatmap(s2,  
                    main = "HeatMap of Gene Expression of\nProteins that are Related to Alzheimer:\\nHealthy Brain",  
                    # heat map title  
                    notecol="black", # change font color of cell labels to black  
                    #density.info="none", # turns off density plot inside color legend  
                    trace="none", # turns off trace lines inside the heat map  
                    margins =c(14,2), # widens margins around plot  
                    col=my_palette, # use on color palette defined earlier  
                    #breaks=col_breaks, # enable color transition at specified limits  
                    dendrogram="row", # only draw a row dendrogram  
                    cluster_cols = FALSE,  
                    show_rownames = FALSE,  
                    Colv="NA")[[4]] # turn off column clustering  
merged_heatmaps <- grid.arrange(arrangeGrob(grobs= a ,nrow =1)) plot(merged_heatmaps)
```

Explanatory Graph : Distribution of Gene Markers in WT/AD

In this section I have decided to explore the distribution of the 3 gene markers Mfge8, Slc38a1, Myoc.
The idea is to count the number of cells in the pictures taken of the tissue,
and count how many proteins of each marker appear in each cell in the picture.
We can take this data and plot the different distributions of the proteins in WT tissue VS AD tissue.

The Graph:

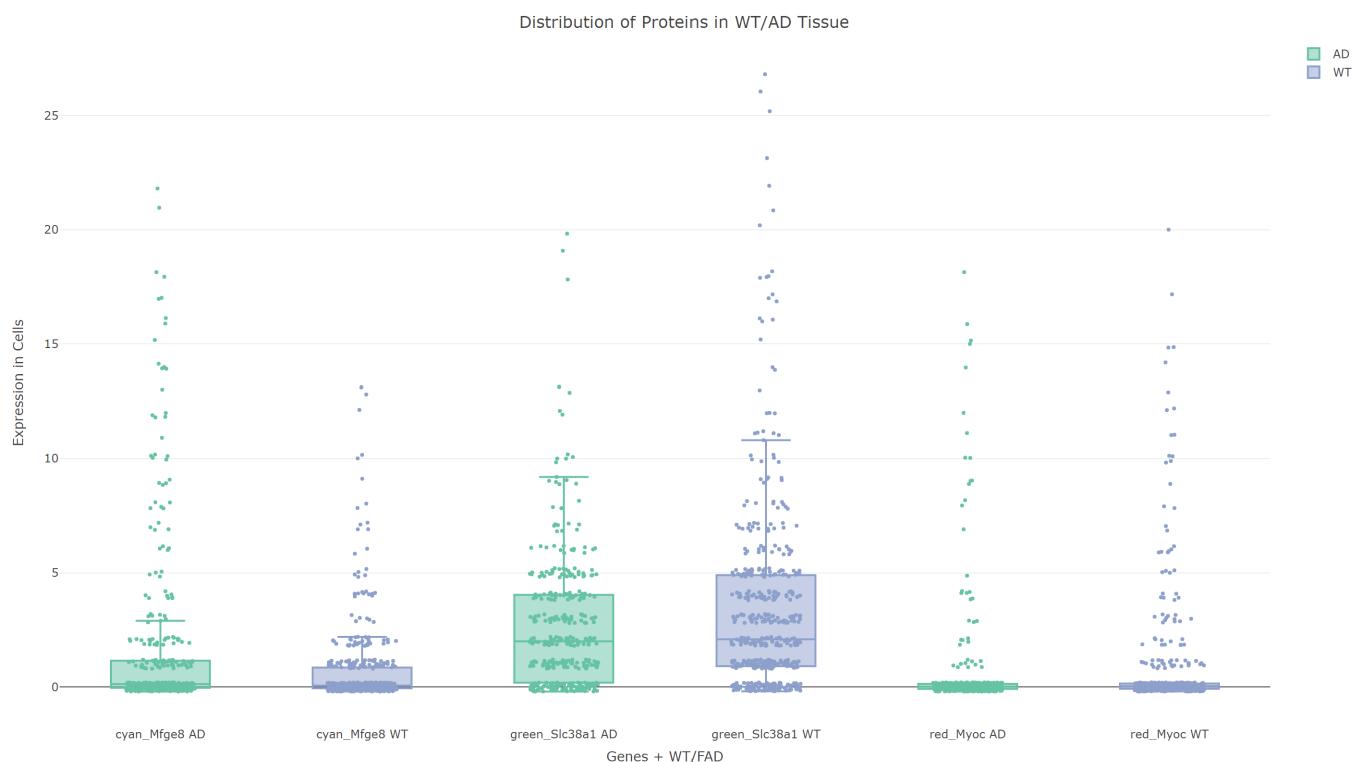


Figure 4: Box Plot of Distribution of Protein Markers in Brain tissue: WT/AD

Explanation of results:

In this graph we can actually look at the individual data points and get some additional information.

Here I have a direct comparison between the 3 protein markers with WT VS AD.

Since this is an interactive plot we can slide our mouse on the graph and get additional information regarding each element in the graph - such as the mean of the expression of each marker in WT or in AD, the max value, the min value, etc.

And so, we can explore the data in a more meaningful way, and understand which parts of the data require additional analysis, and which are ready to reach conclusions from.

the plot can also be found here: https://www.cse.huji.ac.il/~guy_lutsker/plot.html

The code that generated the explanatory Graph:

```
#                                         Graph Two:  
  
# melted data to tranfoms it , thusly it will be easier to graph later  
melted_data<-melt(data = data , id.vars = c("id","ImageNumber") ,  
measure.vars = c("cyan_Mfge8","green_Slc38a1","red_Myoc"))  
  
# adding column for graph  
melted_data$new <- paste(melted_data$variable ,melted_data$ImageNumber)  
  
# title ="Expression in Cells" #type = "log")  
  
jittered_vals<-jitter(as.numeric(melted_data$value) , factor=1 , amount = NULL)  
# creating the plotly graph  
p2<-melted_data %>%  
group_by(ImageNumber) %>%  
plot_ly(x=~new ,y=jittered_vals ,color = ~ImageNumber ,  
text = ~paste(" Id:",id,"<br>" ,  
"State:",ImageNumber,"<br>" ,  
"Value:",value,"<br>") ,  
hoverinfo="text" ,  
boxpoints = "all" , pointpos = 0 , jitter = 0.7 , marker = list(size = 4))%>%  
#add_trace(x = ~new,y = ~value , color = ~ImageNumber , type = "violin") %>%  
add_boxplot() %>%  
layout(xaxis= list(title ="Genes + WT/FAD") ,  
yaxis= list(title ="Expression in Cells") ,  
title ="Distribution of Proteins in WT/AD Tissue")  
  
p2
```

Code

The entire code of this project:

```
# clear all
clear all rm(list=ls())

# set directory
setwd("/Users/guylu/Downloads/")

# all libs:
library(ggplot2)
library(plotly)
library("tidyverse")
library(stringr)
library(dplyr)
library(reshape2)
library(gplots)
library(gridGraphics)
library(grid)
library(gplots)
library(pheatmap)
library(gridExtra)

#           abit of preprocessing:

#read data:
data <- read.csv("MyExpt_IdentifySecondaryObjectsBlue.csv")

# add id column
data <- tibble::rowid_to_column(data, "id")
# change values for readability 1 - sick, 2- healthy
data$ImageNumber <- sub(1, "AD", data$ImageNumber)
data$ImageNumber <- sub(2, "WT", data$ImageNumber)

# change column names for markers in image to gene names
names(data)[names(data) == 'Children_IdentifyPrimaryObjectsCyon_Count'] <-
'cyan_Mfge8'
names(data)[names(data) == 'Children_IdentifyPrimaryObjectsGreen_Count'] <-
'green_Slc38a1'

names(data)[names(data) == 'Children_IdentifyPrimaryObjectsRed_Count'] <- 'red_Myoc'

#           Graph One:

data$sum = data$cyan_Mfge8+data$green_Slc38a1+data$red_Myoc temp<-data[data$sum > 2 ,]

# split table for sick samples and healthy ones
data1<-temp[temp$imageNumber=="AD",]
data2<-temp[temp$imageNumber=="WT",]
```

```

# nice palette:
my_palette<-
colorRampPalette(c("white","blue","green","yellow","red"))(n = 367)

# adjust the data as matrix for heatmap
s1 = as.matrix(data1[,c("cyan_Mfge8","green_Slc38a1","red_Myoc")])
s2 = as.matrix(data2[,c("cyan_Mfge8","green_Slc38a1","red_Myoc")])

# creating heat map for sick and one for healthy
a <- list(pheatmap(s1,
  main = "HeatMap of Gene Expression of\nProteins that are Related to Alzheimer:\\nAlzheimer Brain",
  # heat map title
  notecol="black", # change font color of cell labels to black
  #density.info="none", # turns off density plot inside color legend
  trace="none", # turns off trace lines inside the heat map
  margins =c(14,2), # widens margins around plot
  col=my_palette, # use on color palette defined earlier
  #breaks=col_breaks, # enable color transition at specified limits
  dendrogram="row", # only draw a row dendrogram
  cluster_cols = FALSE,
  show_rownames = FALSE,
  Colv="NA") # turn off column clustering
[[4]])

a[[2]] <- pheatmap(s2,
  main = "HeatMap of Gene Expression of\nProteins that are Related to Alzheimer:\\nHealthy Brain",
  # heat map title
  notecol="black", # change font color of cell labels to black
  #density.info="none", # turns off density plot inside color legend
  trace="none", # turns off trace lines inside the heat map
  margins =c(14,2), # widens margins around plot
  col=my_palette, # use on color palette defined earlier
  #breaks=col_breaks, # enable color transition at specified limits
  dendrogram="row", # only draw a row dendrogram
  cluster_cols = FALSE,
  show_rownames = FALSE,
  Colv="NA")[[4]] # turn off column clustering
merged_heatmaps <- grid.arrange(arrangeGrob(grobs= a , nrow=1)) plot(merged_heatmaps)

#
# Graph Two:

# melted data to tranfroms it, thusly it will be easier to graph later
melted_data<-melt(data = data , id.vars = c("id","ImageNumber"),
measure.vars = c("cyan_Mfge8","green_Slc38a1","red_Myoc"))

# adding column for graph
melted_data$new <- paste(melted_data$variable ,melted_data$ImageNumber)

# title ="Expression in Cells" #type = "log")

jittered_vals<-jitter(as.numeric(melted_data$value), factor=1, amount = NULL)
# creating the plotly graph
p2<-melted_data %>%

```

```
group_by(ImageNumber) %>%
plot_ly(x=~new ,y=jittered_vals ,color = ~ImageNumber ,
text = ~paste(" Id:", id,"<br>",
"State:", ImageNumber,"<br>",
"Value:", value,"<br>"),
hoverinfo="text",
boxpoints = "all", pointpos = 0,jitter=0.7,marker = list(size = 4))%>%
#add_trace(x = ~new,y = ~value , color = ~ImageNumber , type = "violin") %>%
add_boxplot() %>%
layout(xaxis= list(title ="Genes + WT/FAD",),
yaxis= list(title ="Expression in Cells"),
title="Distribution of Proteins in WT/AD Tissue")
```

p2