

BDA211 of Spring 2022: Term Project



Diabetes and its Correlation with Genetic and Environmental Factors

by Lamia Merhy

202002614

Introduction

In this presentation, we will attempt to find out whether cholesterol, hba1c gene, and the gut microbiota are linked to an individual's risk to developing diabetes.

According to the governmental agency for disease control and prevention, the hba1c, or the glycated hemoglobin, is a form of protein that is chemically linked to a sugar

Microbiota, according to Medical News Today, is the a wide variety of bacteria, viruses, fungi, and other single-celled animals that live in the body.

Most of the gut microbiota, or the microbiota living in our guts, are useful bacteria that we live with symbiotically. That's why the presence and amount of these microbiota can be a strong indicator of many diseases in the body.

In this study, we attempt to find a relation between these concepts and type 2 diabetes. The levels of cholesterol and gut microbiota can be tested by the stool collected by the patients, and the samples we have collected are of diabetic, pre-diabetic, and normal individuals.

Literature Reviews

- According to [Qin, J., Li, Y., Cai, Z. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 \(2012\).](#), diabetic patients had only a moderate degree gut bacterial imbalance, where a decline in butyrate-producing bacteria was indicated, which may be beneficial. Such changes in the intestinal bacteria composition have recently been reported for cancer patients and ageing population. Thus, butyrate-producing bacteria seem to have a protective role against several types of diseases. Furthermore, given that other intestinal diseases show a loss of butyrate-producing bacteria, it is possible that imbalance that results in a disordered, rather than directional, alteration of gut microbial composition may itself have a role in increasing the susceptibility to a variety of diseases.
 - Additionally, there was an increase in functions relating to gut oxidative stress response is also of interest, given that previous studies have shown that a high oxidative stress level is related to a symptom for diabetic complications. Finally, gut metagenomic markers are able to differentiate between T2D cases and controls with a higher level of specificity than similar analyses based on human genome variation raises the possibility for a mode of monitoring gut health and a complementary approach for risk assessment of this common disorder.
- In addition, another study conducted by [Bennett, C.M., Guo, M. and Dharmage, S.C. \(2007\), HbA1c as a screening tool for detection of Type 2 diabetes: a systematic review. Diabetic Medicine, 24: 333-343.](#) that aimed to assess the validity of HBA1C as a screening tool for early detection of diabetes, by systematically reviewing of primary cross-sectioning studies of the accuracy of the HBA1C for the detection of diabetes using the oral glucose tolerance test as a reference standard. As a result, some studies met the inclusion criteria. HbA1c has slightly lower sensitivity than the

standard test in detecting diabetes, but slightly higher specificity. Both HbA1c and the standard test have low sensitivity for the detection of impaired glucose tolerance. In conclusion, HbA1c and the standard are equally effective screening tools for the detection of diabetes.

Data Description

Data Cleaning

For this study, the stool subset data set was used, but with some enhancements. Source: [Kaggle](#) Concerning the rows, I was only using the subjects that concerned type 2 diabetes. Whereas, concerning the columns, I used only the important variables, shrinking the number of variables 2339 variables to only 9, which are the basic variables, like gender and age, and the ones of interest, like the HBA1C gene, and the bacteria.

```
proj <- read.csv('C:\\Users\\User\\Downloads\\abundance_stoolsubset.csv')

proj1 <- proj[c(proj$dataset_Name == 't2dmeta_long' | proj$dataset_Name ==
't2dmeta_short' | proj$dataset_Name == 'WT2D'),
             c('dataset_Name', 'age', 'gender', 'disease',
'country', 'hba1c', 'cholesterol',
'k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.
g__Roseburia',
'k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Clostridiaceae.g
__Clostridium.s__Clostridium_bolteae')]

diabetes <- proj1[c(proj1$disease == 't2d'),] #223 subjects

pre_diabetes <- proj1[c(proj1$disease == 'impaired_glucose_tolerance'),] #49
subjects

no_diabetes <- proj1[c(proj1$disease == 'n'),] #217 subjects
```

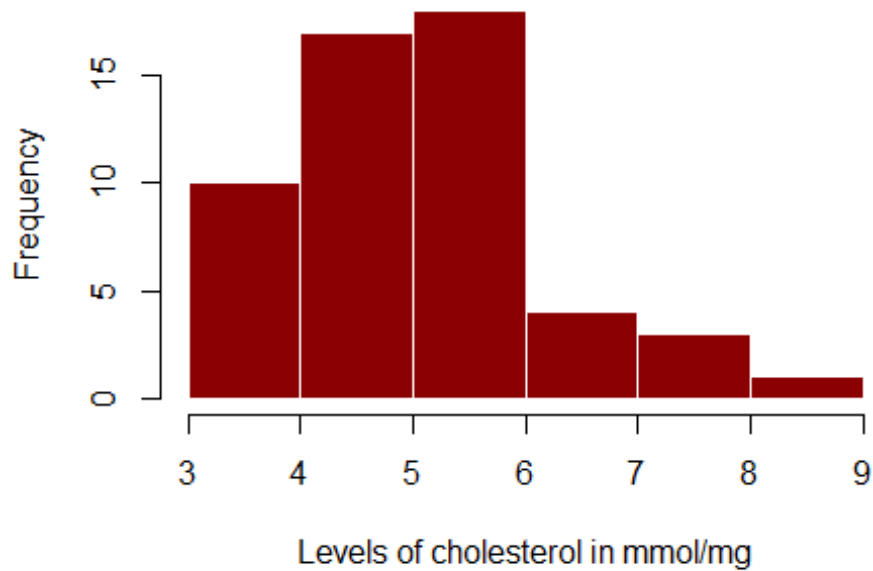
Data Exploration

After being done with data preparation, we now move on to exploring our data through plots to understand better that data collected.

The Relation of Cholesterol with Diabetes

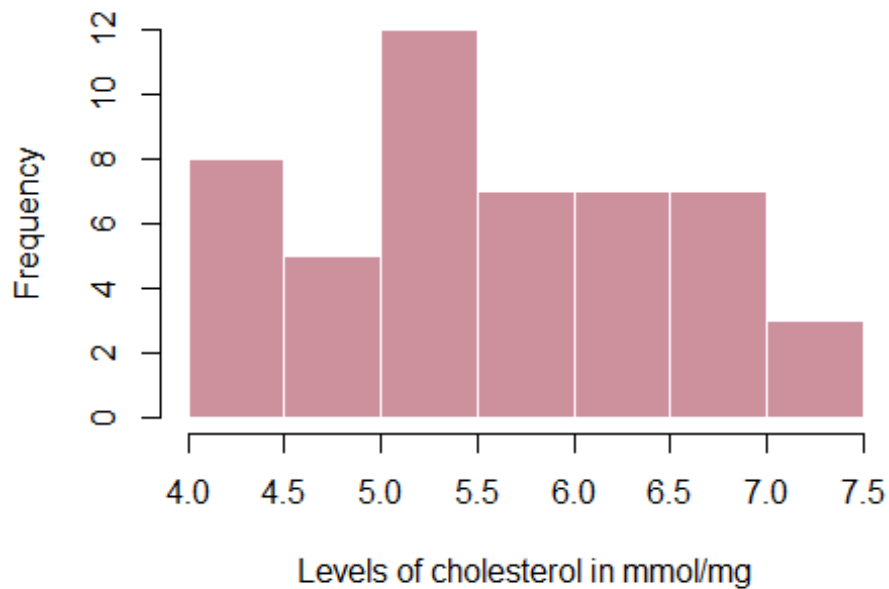
```
hist(x=diabetes$cholesterol, main = 'Histogram of the levels of cholesterol
in diabetic individuals', xlab = 'Levels of cholesterol in mmol/mg', col =
'darkred', border = 'white')
```

Histogram of the levels of cholesterol in diabetic individuals



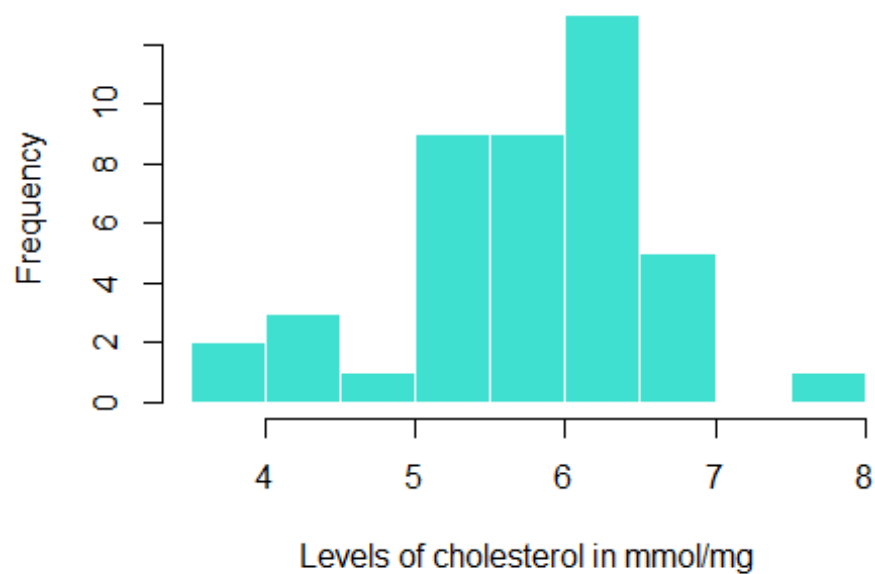
```
hist(x=pre_diabetes$cholesterol, main = 'Histogram of the levels of  
cholesterol in pre-diabetic individuals', xlab = 'Levels of cholesterol in  
mmol/mg', col = 'pink3', border = 'white')
```

Histogram of the levels of cholesterol in pre-diabetic individuals



```
hist(x=no_diabetes$cholesterol, main = 'Histogram of the levels of
cholesterol in normal individuals', xlab = 'Levels of cholesterol in
mmol/mg', col = 'turquoise', border = 'white')
```

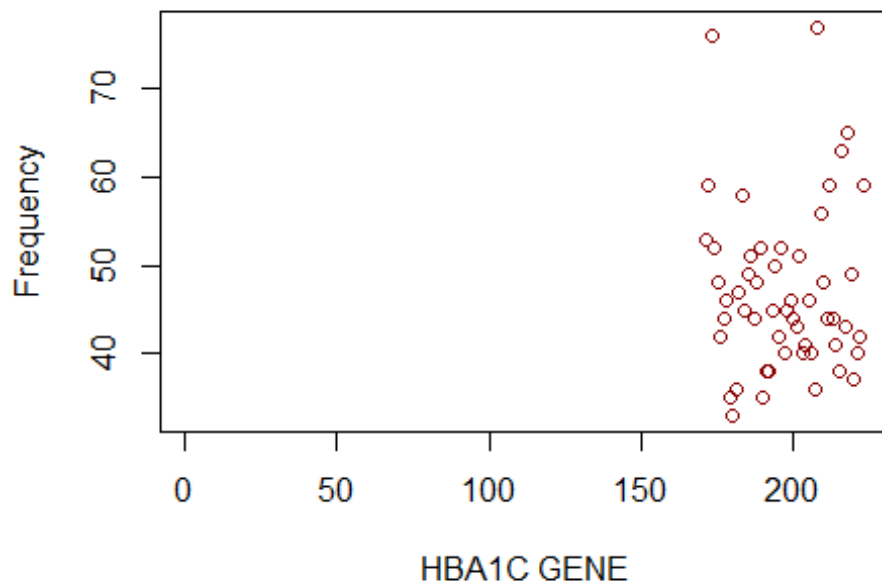
istogram of the levels of cholesterol in normal indivi



The Relation of the HBA1C gene with Diabetes

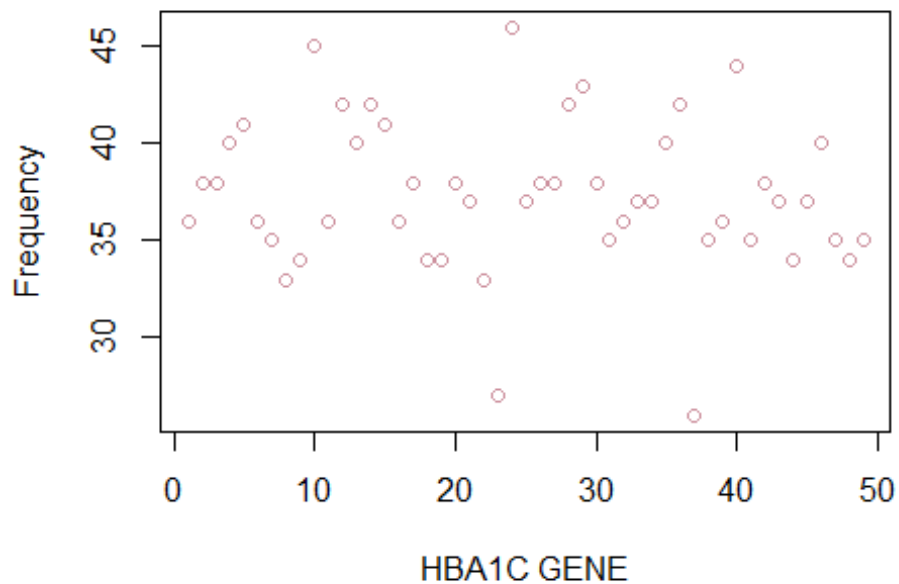
```
plot(x=diabetes$hba1c, main = 'Scatterplot of the levels of HBA1C in diabetic
individuals', xlab = 'HBA1C GENE', ylab='Frequency', col = 'darkred')
```

Scatterplot of the levels of HBA1C in diabetic individ



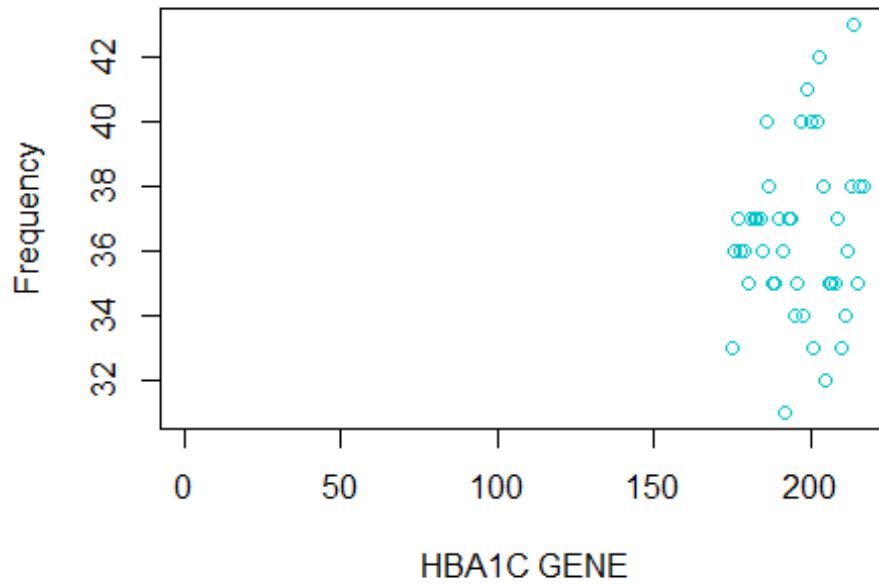
```
plot(x=pre_diabetes$hba1c, main = 'Scatterplot of the levels of HBA1C in pre-diabetic individuals', xlab = 'HBA1C GENE', ylab='Frequency', col = 'pink3')
```

Scatterplot of the levels of HBA1C in pre-diabetic individ



```
plot(x=no_diabetes$hba1c, main = 'Scatterplot of the levels of HBA1C in normal individuals', xlab = 'HBA1C GENE', ylab='Frequency', col = 'turquoise3')
```

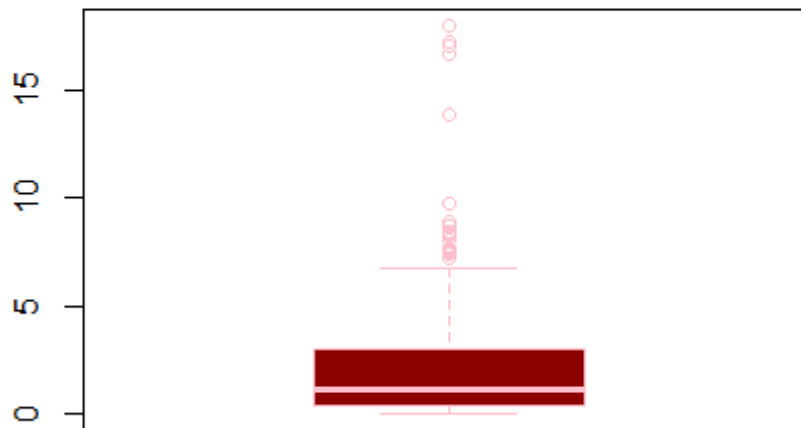
Scatterplot of the levels of HBA1C in normal individuals



The Relation of Rosebiura Bacteria with Diabetes

```
boxplot(x=diabetes$k__Bacteria.p__Firmicutes.c__Clostridia.o__Clostridiales.f__Lachnospiraceae.g__Roseburia, main = 'Boxplot of the levels of Rosebiura in diabetic individuals', xlab = 'Levels of Rosebiura', col = 'darkred', border = 'pink')
```

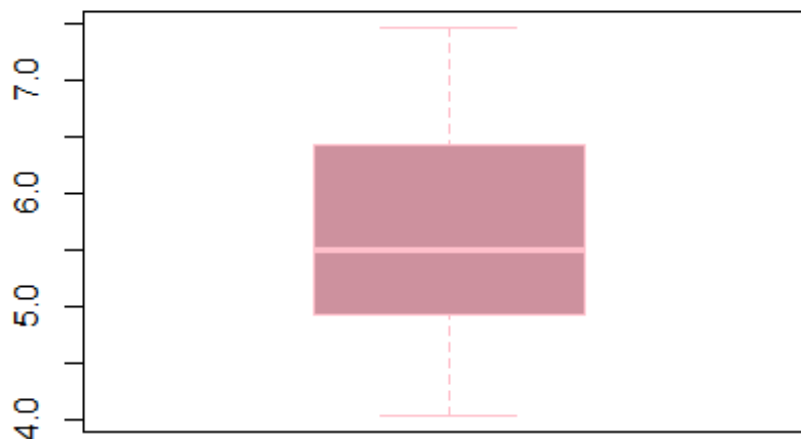
Boxplot of the levels of Rosebiura in diabetic individ



Levels of Rosebiura

```
boxplot(x=pre_diabetes$cholesterol, main = 'Boxplot of the levels of  
Rosebiura in pre-diabetic individuals', xlab = 'Levels of Rosebiura', col =  
'pink3', border = 'pink')
```

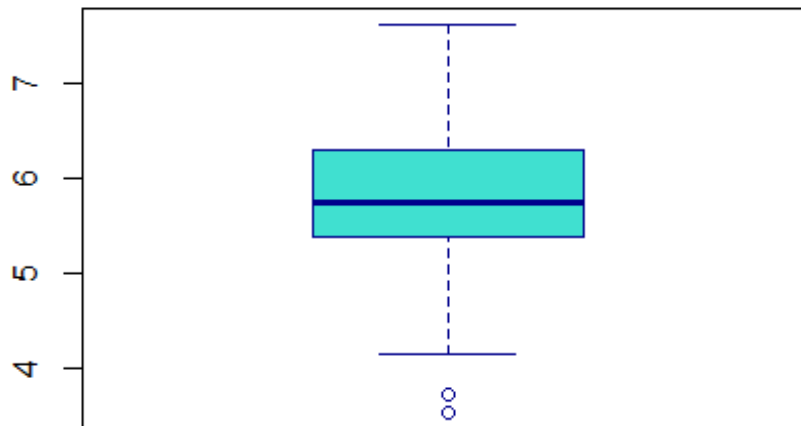
Boxplot of the levels of Rosebiura in pre-diabetic indiv



Levels of Rosebiura


```
boxplot(x=no_diabetes$cholesterol, main = 'Boxplot of the levels of Rosebiura  
in normal individuals', xlab = 'Levels of Rosebiura', col = 'turquoise',  
border = 'darkblue')
```

Boxplot of the levels of Rosebiura in normal individuals



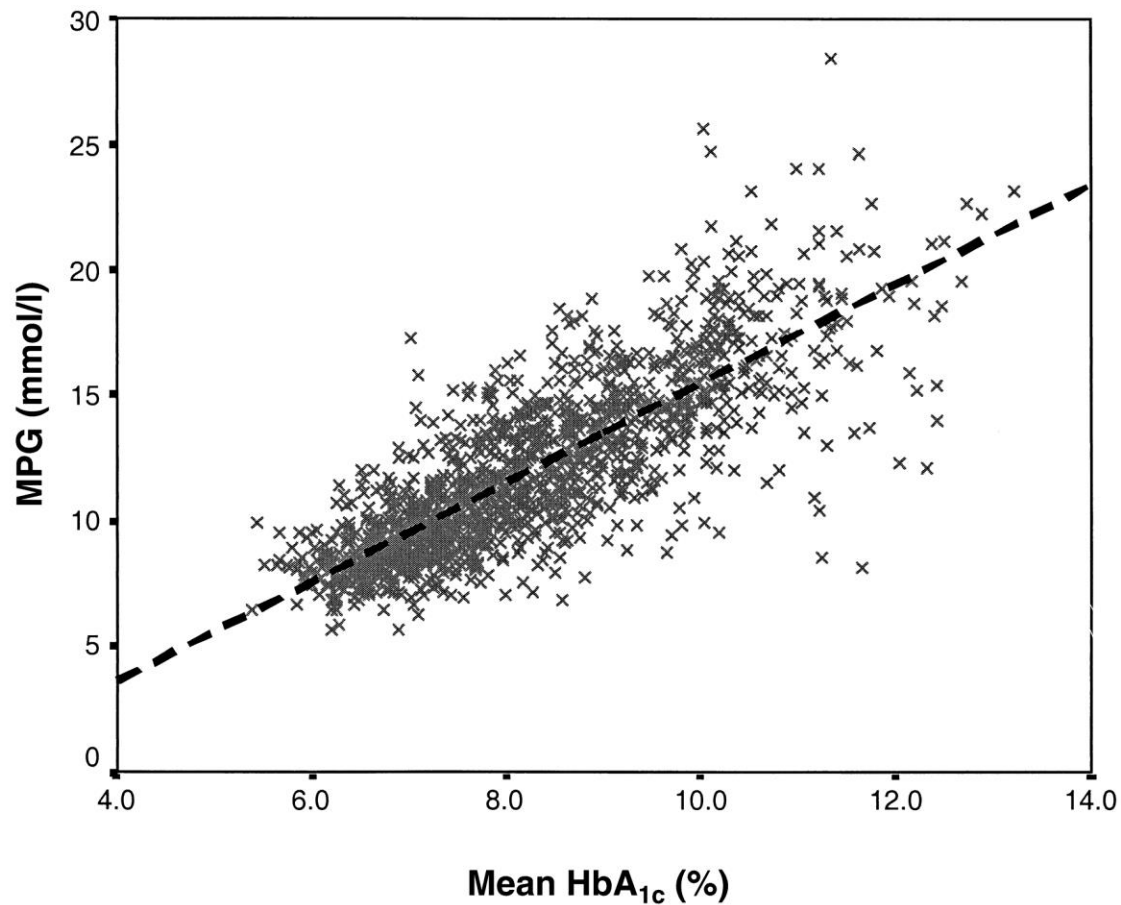
Levels of Rosebiura

Equations Related to the Work

Linear Regression

In the previous study, a linear regression was used to prove a positive linear correlation between using HBA1C and the standard tool to test early for diabetes

$$Y_i = f(X_i, b) + e_i$$



Cholesterol Average

We calculate the average levels of cholesterol for diabetic, pre-diabetic, and normal patients:

$$mean = \frac{\sum_{i=1}^n x_i}{n}$$

```
mean(na.omit(diabetes$cholesterol))
## [1] 5.081509
mean(na.omit(pre_diabetes$cholesterol))
## [1] 5.600612
mean(na.omit(no_diabetes$cholesterol))
## [1] 5.735349
```

Descriptive Analytics

- Cholesterol:
 - According to the histograms, it looks like total cholesterol has a normal distribution across all disease categories, making the relationship between it

and diabetes rather weak. We conclude that cholesterol is not an indicator for diabetes.

- HBA1C:
 - According to the scatter plots, there is a clear clustering in the plots referring to normal patients and diabetics, whereas, surprisingly, there is no correlation whatsoever in the plot referring to pre-diabetic individuals. We can see very high levels of HBA1C in the plot referring to diabetics, which supports the claim of the second study we have seen the HBA1C is a strong indicator of diabetes, and that the more the HBA1C protein in the blood, the higher chance it is due to diabetes.
 - As for the unusual result of pre-diabetic plot, it might be the result of scarce subjects of study that were diabetic, as the number of the did not even reach 50.
- Rosebiura
 - The boxplots seem to show patients upper quartile of non diabetics have slightly more Roseburia inulinivorans in their stool, which supports the study we discussed earlier, in which diseases do make clear imbalances in the gut microbiota. So, we can say that microbiota are potential strong indicators for diabetes, once enough studies have been conducted on them.

Conclusion

There has been various studies being conducted on whether different components can be indicative diseases as an attempt to prevent the diseases from being more severe. One of these diseases are diabetes, and it has been subject to many studies for it is due to both genetic and environmental factors. Seeing that we have found that cholesterol shows little correspondence to diabetes, HBA1C is an example of one of the genetic factors that has a strong correlation to diabetes, as we have found out an increase in HBA1C is strongly related to presence of diabetes. In addition, the gut microbiota is an example of the environmental factors that play a role in determining whether the patient is diagnosed with diabetes or not. Much more studies can be done in this field, with much more subjects, to show the significance and potential of the gut microbiota in showing whether diseases exist.

Thank you!