In our investigation for now, we have developed three main methods. These methods mainly focus on calculating the FDR of input dataset or random data generated from standard normal distribution, and also fitting data to the empirical null distribution.

The first method we have is called the 'Mootha_function' that was used to process the dataset that contains data measured from the paired groups. In this case, the data we now have is the Mootha data that contains genetic expression levels measured from diabetes patients and control groups. By dividing the theta into control and experimental group, we perform a two sample t test on them to check any significant data point. The first decision we had to make was choosing the thresholds for the determination of significance. We picked a set of thresholds based on the p-value and t-value distributions. By choosing the range of -5 to 5, we are able to cover all the p-values and t-values generated from the two samples t test for the Mootha dataset. With the thresholds, we calculate the expected count of t and p values that are greater than the threshold using the pnorm() function in R and the obtained count. As the final step, which is the main purpose of this method, we plot the FDR rate as a function of threshold that provides a reference for how the FDR changes with different thresholds.

The second method we have is called the 'Jongho_function' that was used to fit a dataset of z-scores with the empirical null distribution. The first step in this method is to overlap the theoretic distribution using the midpoint of cells by the dnorm() function on top of the density of z-scores. This allows us to see how well the theoretical distribution fits the actual z-score distributions. Then we use the lm(y~x + I(x^2)) where as x is the midpoints and y are the log density at the midpoints to make prediction of the density. Finally, we plot the empirical null distribution on top of the actual density to compare how it fits better than the theoretical distribution, and then saves the coefficients. As a result, we can compare and contrast the two plots on how the empirical distribution fits better. We have the Jongho dataset from the paper we studied for this method.

The third method we have is called 'Rates_function' that was used to study the correlation between TPR, FPR, and FDR of two distributions of randomly generated data from normal distribution. The thresholds we choose begin from the smaller mean - 2 to the bigger mean + 2 with steps of 0.1. Then we set anything greater than the threshold to be positive and negative otherwise to simulate model predictions. Eventually, we calculate the TPR, FPR, and FDR of each threshold chosen. Then we plot the TPR, FPR, and FDR together as a function of the threshold, and TPR as functions of FPR and FDR separately. These three plots allow us to check the relationship between the three more important rates, which helps us in determining the threshold for the replication project.