Exploratory Data Analysis

Formative Assessment 1
DSC1105

Zion John Yousef T. Ramilo
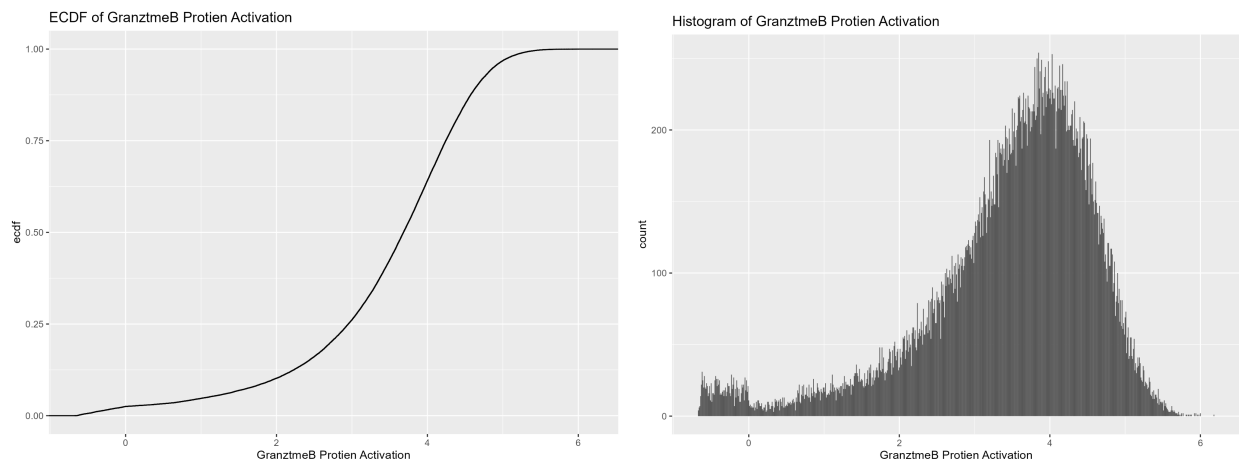
Figure 1: (a) Empirical cumulative distribution function of GranztmeB (left); (b) Histogram of GranztmeB (right)

Figure 1.a represents the data of the GranztmeB protein of a cell to which the Empirical Cumulative Distribution Function plot (ECDF plot) was used. This plot shows the range of values most cells can have for the measured Granzyme B protein. ECDF plot shows that 25% of the cells took a value of less than 3, which indicates that most of the cells that are measured are not active in targeting foreign entities. In the context of testing a new drug, this data suggests that the host's immune system is not responding as expected to the drug, deeming improvements to the drug's effectiveness necessary.

Figure 1.b is where the data is represented through a histogram where it represents the frequency of cells that have activated their Granzyme B protein and their corresponding level. The histogram shows that the distribution is skewed to the left, indicating that there is a high frequency of cells that have high activation of the Granzyme B protein.

Exploratory Data Analysis

Formative Assessment 1
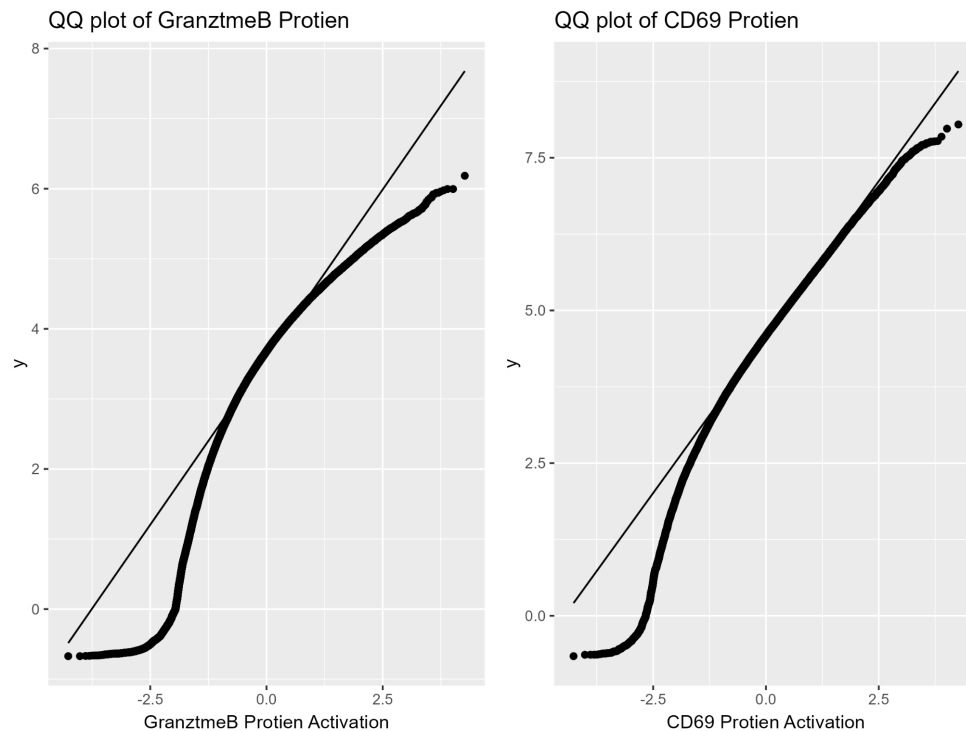DSC1105

Zion John Yousef T. Ramilo

Figure 2: GranztmeB vs CD69 - QQ plot

QQ plots were used to determine how much the sample to which in this case the values of the CD69 protein and GranztmeB protein were able to create a normal distribution. Figure 2 shows two plots on the left for the GranztmeB protein which shows a large divergence from the theoretical values of the normal distribution, this is apparent within the lower and higher theoretical values. It is also notable that only a small interval was able to align itself with the theoretical values. This could not be the same with the QQ plot for the CD69 protein where there is a considerable interval of the data close to the theoretical values, indicating that their values are more normally distributed than the GranztmeB protein. However it is still notable that smaller values still diverge from the theoretical values which is a key similarity of both proteins.