**Name: Tolu and Shoma**

**Frequency Distribution Analysis of the Nuclear Areas of tumor cells expressing ITGA1 and PLK1 proteins in pancreatic cancer patient samples using AI tools and R.**

**​**ITGA1 is a transmembrane protein that Plays a crucial role in mediating interactions between cells and their extracellular environment. Whereas PLK1 plays a different role in the biological process. It acts as a master regulator of various cell cycle stages. Despite having different biological roles for these two proteins, many research has shown that the expression level of these proteins goes up in the Pancreatic Ductal Adenocarcinoma (PDAC) samples compared to Normal pancreas sample.

For this case study we wanted to see the correlation between these proteins in pancreatic tumor cells using some of the data from Albert et al 2024 paper (Aquino et al 2024). This paper established that there were some both ITGA1 and PLK1 proteins play an important role in the progression of PDAC.

To carry out our case study, we pulled some of the Immunofluorescence staining data and used ChatGPT and R language to look for frequency distribution of the tumor cells expressing ITGA1 and PLK1 and the correlation between these proteins in the tumors. The data sheet we used (A1) contained information from PDAC tissue sample from the tissue microarray slide.

We supplied prompts to ChatGPT to enable us carry out this analysis. The prompt ranged from simple prompts like how to upload our data on ChatGPT to how to creating a path that ChatGPT and R can use for the analysis. We also gave more complex prompts to ChatGPT to generate the necessary codes to run the frequency distribution and correlation analysis on R. ChatGPT was able to help with this, whenever the codes ChatGPT gave came back with an error on R, we simply copied this error prompts to ChatGPT and it was able to debug the error and solve the problem. It gave us codes to use for R studio packages like readxl, (ggplot2, and tidyverse to make the graphs. We documented this process each step of the way from the beginning, to the generation of successful codes and all the errors in between.

Not surprisingly, with few tries, ChatGPT was successfully able to generate codes that ran on R and gave the same results as was documented in the paper we used. It is unsurprising because the tasks we gave ChatGPT were relatively simple and previous findings have made it known that ChatGPT is useful in elementary/beginners coding. It runs into issue with more sophisticated coding tasks (Atkinson 2023).

The biggest challenge we had with this assignment was when we tried to rename a column in the data sheet and ChatGPT gave codes that were unsuccessful in R. This could probably be because the data file we were using had multiple sheets and we only changed the names in the columns of one sheet. However, we specified the sheet we wanted to rename for ChatGPT but it still did not give a successful run in R studio. After multiple tries, we bypassed this issue by creating a new data file of the sheet we were working with and with proper titles for the columns we were interested in. We used this as the new path for R and we were able to successfully get our correlation graph.