GENERAL

In Milestone 1.3 our goal was to show how the dosage of the medical treatment affects the treatment results.

ANALYSIS OF THE RESULTS AND THE PROCEDURE

At the beginning of our analysis for this milestone, we reduced the prism data frame only to the ovary cell lines and reversed the x-axis with the y-axis. Then we merged the reduced data frame with prism.treat dataset, thereby deleting all not needed columns, and created a new data frame 'df_ovary_DepMI_prism_treat' which we worked with in our further analysis.

According to the experimental methods, implied in the scientific paper (Corsello et al. 2020), we used the single dose 2.5 micromolar of the drugs as a marker to divide them into subgroups.

Then we created a vector containing only the most effective drugs from milestones 1.1 and 1.2 'most_effective_moas_drugs'. In order to create a data frame with the most effective drugs from the previous milestones, a reduction of the 'df_ovary_DepMI_prism_treat' data frame to the most effective moa drugs from the mentioned milestones was conducted.

Proceeding to the further analysis, a ggplot and a boxplot were implemented to plot the dosage against the ovary cell lines proliferation values and demonstrate how the proliferation values change with an increasing or decreasing dosage of the medicaments.
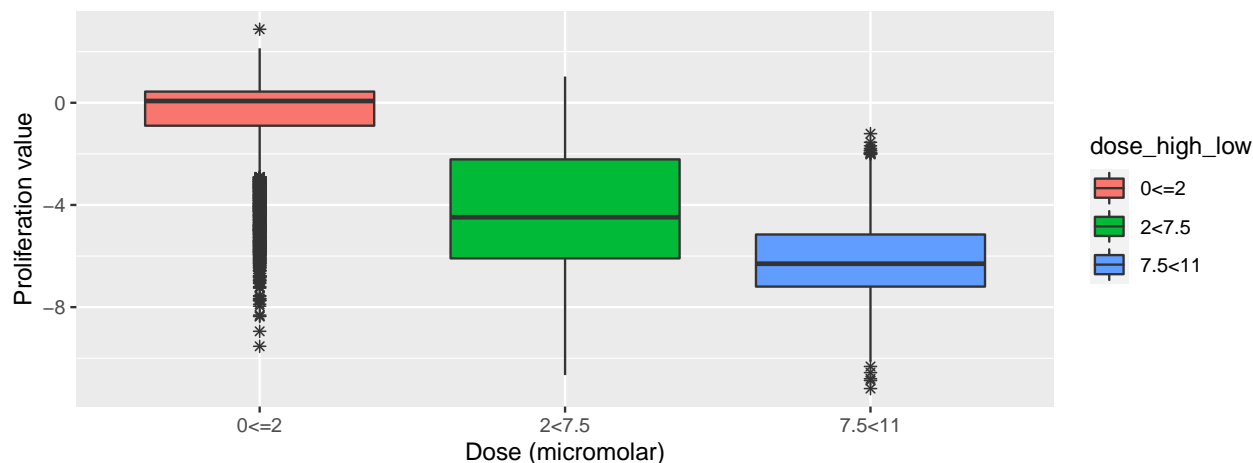
VISUALIZATION - BOXPLOT

Figure 1: **1.** Boxplot of the distribution of the proliferation rates of ovary cell lines according to the drug dosage

Therefore, a visualization with a boxplot to show the distribution of the proliferation rates of the ovary cell lines, regarding the drug dose was performed. Unfortunately, we could not divide the boxplot according to the 2.5 micromolar dosage. However, the boxplot was divided in three dose-stages: low dose, median dose and high dose. The highest median proliferation value was observed by the lowest dosage 0<=2 micromolar, by dosage 2<7.5 micromolar there was an increase of the median, whereas by the highest dosage 7.5<11 micromolar the corresponding median was at the lowest. This means that with an increasing dose of the medicaments, the proliferation rates are reduced.

PEARSON CORRELATION

In order to show how the dosage of the medical treatment affects the proliferation and viability of the ovary cell lines, a Pearson correlation between the dose of the drugs and the proliferation value was performed. Almost all Pearson correlation coefficients have negative values, except for one between the cell line 'ACH-000713'

and the drug 'exatecan-mesylate'. In comparison with the other coefficients, this has a positive value and can be considered as an outlier. A reason for this could be a measurement error, because there is only one cell line with such value. We implemented specifically a ggplot which demonstrated that there is an outlier. Consequently,there is an inverse negative correlation. Some medicaments show a very strong correlation with the proliferation of the cell lines, while others almost do not affect it.

VISUALIZATION - GGPLOT

Furthermore, we created the correlation matrix from the dosages of the different medicaments per cell line and to display graphically the Pearson correlation, the function ggcorrplot() on the created correlation matrix was applied. As you can see, the blue and the purple squares which indicate higher correlation, ranging from 0.0 to -1.0 are in prevalence, meaning that the majority of the drugs reduce the cell proliferation. A Pearson correlation histogram showed the frequency of the Pearson correlation values, with most frequent values between -0.9 and -1.0.
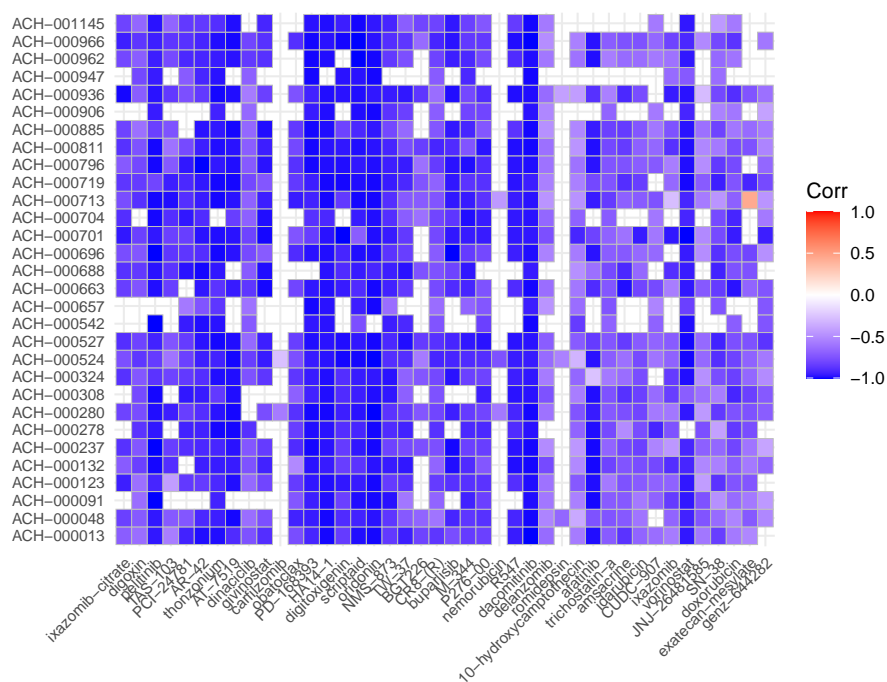

Figure 2: **2.** ggplot of the correlation between the proliferation rates of ovary cell lines and the dosage of the applied medical treatment