

## Personal Reflection

Ioanna Nika

Student ID: 4788311

The main goal of this project was to reproduce figures from the Lindel paper: “Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair”. The paper essentially proposes machine learning methods, specifically logistic model regression models for the prediction of the outcomes (deletion events, insertion events, indel ratio & frameshift) of CRISPR/Cas9-mediated double-strand break repair.

Our team started by reading the corresponding reference paper, and lantern on the corresponding code-base and data given on Teams to get a good understanding of the topic. I and the rest of the team encountered challenges in understanding the exact methodology and data used in the beginning but through in-depth investigation of their code-base, teamwork, and occasional help from the teacher assistants we managed to overcome the challenges. In general, at the end of every work session, we would summarize what we did shorty in text and write what needs to be done till and for the next meeting. In that way, we would steadily progress through the project.

Everyone on the team had a good understanding of the project, and I would be the one to code mostly during meetings. Everybody collaborated in order to get a base understanding of the assignment, then, I, Gijs, and Connor managed to train the logistic regression classifiers while Bianca developed the aggregate model and further documented the code base. Then the whole team had to collaborate again, in order to overcome some data challenges in running the test set on the trained models to obtain the predictions. Once the predictions were ready I took the initiative and finalized the last bits of the project by plotting the figures for the aggregate model and the Lindel model as well as by calculating the predicted frameshift and plotting it against the actual frameshift. Bianca also went ahead and created some analysis for our results, comparing the weights of our trained model with those of the pre-trained model given by the authors of the Lindel paper as part of their code base.

I believe our project was successful since we managed to reproduce the figures, as well as collaborate effectively as a team. Throughout the whole quarter, I was very satisfied and happy to work with my teammates. Gijs helped a lot in understanding the code base and training the models, Connor helped us with the biological context, set up our repository, and helped us overcome some of the obstacles we faced. Bianca helped too throughout the project and provided additional insights into our results.

Throughout this project, I learned about gene repair and how such mechanisms can be manipulated for gene editing, and the importance of being able to predict the outcomes of such interventions. Moreover, I learned about the applications of such methods in performing gene knockouts or even correcting harmful mutations.