**Bioinformatics Application Form for Research Project (BINP37 and BINP39)**

|  |  |
| --- | --- |
| Name of student: | Felicia Schulz |
| Contact email: | fe5502sc-s@student.lu.se |

|  |  |
| --- | --- |
| Name of Supervisor: | Johan Staaf |
| Name of department: | Division of Translational Cancer Research, Department of Laboratory Medicine, Faculty of Medicine |
| Contact email: | johan.staaf@med.lu.se |

|  |  |
| --- | --- |
| Name of Supervisor: | Suze Julia Roostee |
| Name of department: | Division of Translational Cancer Research, Department of Laboratory Medicine, Faculty of Medicine |
| Contact email: | suze\_julia.roostee@med.lu.se |

|  |  |
| --- | --- |
| Course code (BINP37 or BINP39): | BINP37 |
| Start and end date (expected): | 29/03/23 - |

**Project Description**

Please complete the following(do it on this form).

Project title:

Abstract (200-300 words):

In this project, the role of the programmed death ligand 1 (PD-L1) biomarker in triple-negative breast cancer (TNBC) is investigated. Immunotherapy targeting the PD-L1 protein has recently been found to be effective in increasing survival time of PD-L1 positive patients. The expression of PD-L1 is assessed by expert pathologists, often from hematoxylin and eosin (H&E) stained images. However, there is often high inter- as well as intra-observer variability in the evaluation. Therefore, a deep learning neural network approach can be used to predict the expression of PD-L1 in breast cancer tissue with high accuracy. This deep learning model will be applied to a SCAN-B TNBC dataset throughout the course of this project to find whether the prediction accuracy is consistent with the prediction scored on the external datasets which the model was originally validated on. In addition to that, the results of the model will be compared to pathologist scores and automated cell counts, with the goal of identifying the most accurate and reliable approach for PD-L1 scoring in TNBC. Furthermore, a survival analysis will also be carried out in order to find whether there is a correlation between PD-L1 expression and patient outcomes in this data.

Project plan (1 page):

* Introduction, with the theoretical background to the project and key references
* The specific aim(s) of your project
* Methods (specify whether there are pre-existing pipelines for these in the group)
* Time plan (weekly planning of the project). Remember to include time for writing of the report and preparations for the seminar, and prepare to submit your project report and readme one full week before the presentation date.

Project Plan

Introduction

Breast cancer is a major global health concern, affecting millions of women each year. Among the different subtypes of breast cancer, triple-negative breast cancer (TNBC) is the most aggressive, with poor prognosis and a high risk of relapse. Of all women diagnosed with breast cancer, it is estimated that 12-17% have TNBC (Foulkes, Smith & Reis-Filho, 2010). Due to its lack of expression of the three receptors ER, PR and HER2, which are usually commonly found on breast cancer cells, TNBC does not respond to hormonal therapies. This calls for new treatment options being developed and researched (Staaf et al., 2019). Recent studies have highlighted the role of programmed death ligand 1 (PD-L1) and anti-PD-L1 in breast cancer, with immunotherapy showing promising results in patients with PD-L1-positive tumours (Shamai et al., 2022). PD-L1 is a protein present on the surface of some cancer cells, which binds to PD-1 on immune cells, thereby inhibiting the immune system from killing the cancerous cells. However, accurately scoring PD-L1 expression can be challenging due to inter- and intraobserver variability in pathologist scoring (Brunnström et al., 2017). Thus, a deep learning model was developed by Shamai et al. (2022) with the aim of predicting PD-L1 expression in breast cancer from H&E stained breast cancer tissue. This model was able to predict PD-L1 positivity with high accuracy (area under the curve (AUC) = 0.91-0.93).

Project aims

In this 10-week project, I will investigate the utility of a deep learning (DL) model for PD-L1 scoring in TNBC. Specifically, I will test the DL model developed by Shamai et al. (2022) trained on external data to assess its performance on our SCAN-B TNBC dataset. I will compare the DL scores generated by the model with pathologist scores and automated cell counts, with the goal of identifying the most accurate and reliable approach for PD-L1 scoring in TNBC.

In addition, I will perform a survival analysis on PD-L1 scores to determine whether PD-L1 expression is associated with patient outcomes in our cohort. The ultimate goal of this project is to improve our understanding of PD-L1 expression in TNBC and develop more reliable methods for PD-L1 scoring that can be applied in the clinic.

Methods

As stated above, the DL model developed by Shamai et al. (2022) will be adapted to our SCAN-N TNBC dataset. The pipelines for the other analyses in this project will mostly be created by me.

Project Timeplan

|  |  |
| --- | --- |
| Week | Task |
| 1-2 | Literature reading, establishing data structures / data, organisation of necessary existing code |
| 3-8 | Practical data analysis, interpretation, survival analysis |
| 9 | Documentation of code and report writing |
| 10 | Report writing and preparation of presentation |

References:

Brunnström, H., Johansson, A., Westbom-Fremer, S., Backman, M., Djureinovic, D., Patthey, A., ... & Micke, P. (2017). PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: inter-pathologist variability is higher than assay variability. *Modern Pathology*, *30*(10), 1411-1421.

Foulkes, W. D., Smith, I. E., & Reis-Filho, J. S. (2010). Triple-negative breast cancer. *New England journal of medicine*, *363*(20), 1938-1948.

Shamai, G., Livne, A., Polónia, A., Sabo, E., Cretu, A., Bar-Sela, G., & Kimmel, R. (2022). Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. Nature Communications, 13(1), 6753.

Staaf, J., Glodzik, D., Bosch, A., Vallon-Christersson, J., Reuterswärd, C., Häkkinen, J., ... & Nik-Zainal, S. (2019). Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nature medicine, 25(10), 1526-1533.

**Instruction for Supervisors**

The research project course is important and valuable for our MSc students in Bioinformatics. Therefore, we are  grateful for the support and mentorship that you, any co-supervisors, and other  members of your research team can provide throughout the different stages of  the project

Please read the information for supervisors and let us know if you have any questions. <https://canvas.education.lu.se/courses/8057/pages/template-for-email?module_item_id=609034>

I hereby agree to the conditions for the supervision of this research project in Bioinformatics

Date: Signature supervisor

Date: Signature supervisor