upf. Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2024-2025

Personal Information

Supervisor	Francisco Martínez-Jiménez
Email	fmartinez@vhio.net
Institution	VHIO
Website	https://www.franmartinezlab.com/ https://vhio.net/pf/computational-immunogenomics-group/
Group	Computational Immunogenomics Group

Project

Computational genomics

Project Title:

A predictive machine-learning model for inferring tumoral HLA-II expression

Keywords:

Cancer, Immunogenomics, Immunology, Machine Learning, Computer simulations

Summary:

The Human Leukocyte Antigen II locus (HLA-II) encompasses several genes (i.e., HLA-DR, HLA-DP, HLA-DQ) that collectively orchestrate the presentation of extracellular antigens, including bacterial and viral antigens, following cellular intake. HLA-II presents non-self-antigens that may trigger the activation of CD4+ T cells, prompting their differentiation into diverse T helper cell subsets. This activation bears critical importance as CD4+ T cells substantially enhance tumor immunogenicity and are strongly associated with better outcomes in diverse anti-cancer immunotherapies (Tay et al., 2021). HLA-II genes are primarily expressed by professional antigen presenting cells (APCs). However, recent studies leveraging technological advancements in single-cell sequencing revealed variable degrees of HLA-II expression in malignant cells from renal clear cell, breast and lung carcinomas (Axelrod et al., 2019). This observation defies the existing dogma, prompting new avenues of investigation into the landscape of cancer-cell HLA-II expression and their influence in tumor evolution and responses to cancer immunotherapies. Single-cell sequencing techniques can directly discriminate the HLA-II expression arising from immune cells or tumoral cells in each sample, but the cost of these techniques limits their broad application. Due to this limitation, the implications of HLA-II expression in tumor cells across patients and cancer types at a large-scale level cannot be addressed using such techniques. Alternatively, bulk RNAseq sequencing methods have been applied to large cohorts of samples (both public and private) but the contribution of different cell types to total HLA-II expression in each sample remains unknown. In this project we aim to generate a machine-learning model, capable of predicting the fraction of HLA-II expression attributable to cancer cells from bulk RNA sequencing data. To this aim, we will first create a training set from single-cell datasets (in which the fraction of tumoral HLA-II expression is known), using computer-generated simulations, to obtain synthetic data with different degrees of tumoral HLA-II expression (Dietrich et al., 2022). Afterwards, multiple machine learning algorithms will be tested and fine-tuned to produce a definitive model with optimal performance. The determination of tumoral HLA-II expression in bulk sequencing data from large-scale genomic studies will allow us to interrogate its association with multiple facts of the tumor microenvironment composition, cancer immune escape mechanisms and its influence on patient responses to anti-cancer immunotherapies. To address these questions, we will make use of the Hartwig medical foundation cohort (Martínez-Jiménez, Movasati, et al., 2023; Martínez-Jiménez, Priestley, et al., 2023), the largest cohort of tumors treated with immunotherapy with available RNASeq and WGS data, as well as other publicly cohorts of immunotherapy-treated patients.

References:

Axelrod, M. L., Cook, R. S., Johnson, D. B., & Balko, J. M. (2019). Biological Consequences of MHC-II Expression by Tumor Cells in Cancer. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 25(8), 2392–2402. https://doi.org/10.1158/1078-0432.CCR-18-3200 Dietrich, A., Sturm, G., Merotto, L., Marini, F., Finotello, F., & List, M. (2022). SimBu: bias-aware simulation of bulk RNA-seq data with variable cell-type composition. Bioinformatics (Oxford, England), 38(Suppl_2), II141–II147. https://doi.org/10.1093/BIOINFORMATICS/BTAC499 Martínez-Jiménez, F., Movasati, A., Brunner, S. R., Nguyen, L., Priestley, P., Cuppen, E., & Van Hoeck, A. (2023). Pan-cancer whole-genome comparison of primary and metastatic solid tumours. Nature, 618(7964), 333–341. https://doi.org/10.1038/S41586-023-06054-Z Martínez-Jiménez, F., Priestley, P., Shale, C., Baber, J., Rozemuller, E., & Cuppen, E. (2023). Genetic immune escape landscape in primary and metastatic cancer. Nature Genetics, 55(5), 820–831. https://doi.org/10.1038/S41588-023-01367-1 Tay, R. E., Richardson, E. K., & Toh, H. C. (2021). Revisiting the role of CD4+ T cells in cancer immunotherapy-new insights into old paradigms. Cancer Gene Therapy, 28(1–2), 5–17. https://doi.org/10.1038/S41417-020-0183-X

Expected skills:

Basic understanding of cancer and general immunology, basic shell scripting, Python/R programming, Motivation to learn!

Possibility of funding:

To be discussed

Possible continuity with PhD:

To be discussed