Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2024-2025

Personal Information

Supervisor	Marc Weber
Email	marc.weber@flomics.com
Institution	Flomics Biotech
Website	www.flomics.com
Group	Bioinformatics team

Project

Computational genomics

Project Title:

Metagenomics of plasma samples for early cancer detection and diagnosis

Keywords:

microbiome, cell free RNA, genomics, cancer, machine learning, RNA-Seq, NGS

Summary:

INTRODUCTION We are a dynamic international team of researchers, with access to a cutting-edge computing cloud infrastructure. Here's your opportunity to learn hands-on, and participate in a cool, applied scientific project, while contributing to a better future for all of us! We're looking forward to welcoming excellent, enthusiastic and intellectually curious students. Do not hesitate to contact us directly if you need further details. PROJECT DESCRIPTION Early detection of cancer could significantly improve the likelihood of treatment efficacy and increase the overall survival of cancer patients. Tissue-based analyses using tumor biopsies are invasive, and limited by their availability and narrow scope of sampling. Recent technology uses circulating tumor DNA, termed cell-free tumor DNA (cfDNA), and circulating cell free RNA (cfRNA), that is present in body fluids (blood plasma, urine, spinal fluid) from dead cancer cells to detect and characterize cancer. The advantage of liquid biopsies is that the cfDNA or cfRNA is reflective of both local tumors and distant metastatic sites, and that multiple samples can be taken non-invasively. Thus, the detection of biomarkers in peripheral fluids by liquid biopsy is one of the most promising solutions to improve early detection and diagnosis of cancer. cfDNA and cfRNA are molecules used to detect and quantify the presence of cancer, but are often not sufficiently informative to distinguish all types of cancer. Thanks to high-throughput sequencing techniques, bacterial- and fungal-derived nucleic acids have been detected in human plasma. Additionally, it has been found that the composition of microbially derived nucleic acids in blood can change in the event of disease, reflecting changes in the microbiomes of the affected tissues. Several studies have found that bacteria are involved in the development and progression of cancer and that the composition of the microbiota is unique to each type of cancer. Recently, Poore et al. published a manuscript that addressed the possibility of using circulating microbial DNA (c-mbDNA) to discriminate between cancer patients and healthy donors. First, by analyzing transcriptomics data from The Cancer Genome Atlas (TCGA), the authors found unique microbial fingerprints in each tumor tissue type. Subsequently, they showed that it is possible to discriminate samples from healthy controls and multiple cancer types (prostate, lung and melanoma) using only c-mbDNA. These results suggest that microbial fingerprints in plasma, both of bacterial and fungal origin, may have diagnostic value in the detection and discrimination of different types of cancer. In this project, we look for a motivated candidate to analyze the microbial cfRNA in plasma samples of cancer patients, and develop a machine learning approach to identify a biomarker signature capable of classifying multiple cancer types. The project roadmap includes the analysis of RNA-seq data from plasma samples (>300 samples), the taxonomical characterization of short RNA fragments using fast alignment tools such as kraken2, the identification of contaminants cfRNAs, the application of normalization methods, and the development of machine learning models to classify between healthy donors and cancer types.

References:

Cancer type classification using plasma cell-free RNAs derived from human and microbes | eLife, https://elifesciences.org/articles/75181 Microbiome-Derived Liquid Biopsy: New Hope for Cancer Screening? | Clinical Chemistry | Oxford Academic, https://academic.oup.com/clinchem/article/67/3/463/6024633

Expected skills:

- Basic knowledge in bioinformatics, RNA-seq. - Basic knowledge in machine learning. - Intermediate level in at least one programming language (Python, R, C++). - Good command of the Linux terminal. - English communication skills. - Desirable skill: Nextflow.

Possibility of funding:

To be discussed

Possible continuity with PhD:

To be discussed

Comments:

allowance of 500€ per month, hybrid work model with 2 days per week remote work, wonderful location in Poblenou (Barcelona), brand new offices (Cosymbio labs).