Identification of putative driver genes and their role based on patient profile.

PID: 2021\DSAI\007

Global efforts in cancer research, such as TCGA, lead to the identifications of many driver mutations and genes in different cancer types. Various resources like COSMIC and CGC exist to document the mutations, and high frequency mutated genes found across tumour samples. While a large majority of samples show the presence of a mutation in at least one known driver gene, a small subset of tumours does not contain any known driver gene mutation. The absence of known driver genes mutated elucidates the gap in our knowledge of all possible driver genes. Computational methods using genomic, proteomic, network information exist for prediction of driver genes. However, most of these methods are cohort-based. While a cohort-based study helps understand disease and mechanism, it is difficult to translate these results to a single patient, especially when the patient does not have known mutated driver genes. Cancer is a disease of accumulated mutations, with multiple mutation events triggering the progression of the tumour. The study aims to train a model to identify driver genes mutated during the tumour evolution, which helps maintain tumour, using the multi-omic data available for the patient. A tumour level prediction of driver genes helps identify potential targets for personalised therapy for the given patient.

Task to be assigned to the intern: The project comprises pre-processing, analysis and building machine learning models for driver gene prediction and functional impact.

Learning outcomes: The project will help familiarise with various types of omic datasets and steps to analyse them. The project allows an understanding of available cancer biology tools and the challenges faced when applying machine learning algorithms to biological datasets.

Duration: 6 months

Skills required: Python

Pre-requisite courses: Machine learning

No. of interns required: 1