

## DATA SCIENCE, CANCER AND SEQUENCING ANALYSES

Mehmet BAYSAN<sup>1</sup>

<sup>1</sup>Istanbul Technical University, Computer Engineering Department, Istanbul, Türkiye

---

### ABSTRACT

Recent developments in technology led to exponential growth in processing speed, storage capacities and communication rates. These available resources allowed the creation of thousands of applications that run on smart devices and the world wide web. Today we generate lots of data using these applications and the availability of vast digital data created a new field: data science. Data science aims to utilize digital data to develop solutions to improve human life. Next-generation sequencing technologies enabled the generation of high-throughput genomic data from DNA and RNA. Considering the great potential for genomic data to understand biological functions and the substantial drop in sequencing costs; the amount of available data increased exponentially over time. Sequencing technologies are based on the shredding of the DNA and capture of short reads. These reads are then combined to identify the genomic map of the biological sample. Some of our studies are based on extraction of multiple samples from a tumor and application of sequencing on these samples to understand the tumor development, progression and drug response. Unfortunately, all the frequently used analysis algorithms are based on optimistic assumptions. High level of disparity is observed in literature and in our studies when we compare different algorithms. Currently, we do extensive testing on realistic scenarios to understand the sources of variation in the results. Finally, we develop a user-friendly analysis platform for comparative sequencing analyses to enhance the cooperation among scientists with different backgrounds on sequencing projects.

**Keywords:** Data Science, Genomic Data, Sequencing Technologies