Silesian University of Technology
Faculty of Automatic Control, Electronics and Computer Science

Models of cancer genome evolution used to evaluate the role of selection and occurrence of new mutations

Doctoral thesis

Paweł Kuś

Supervisor:
Prof. dr hab. inż. Marek Kimmel
Co-supervisor:
Dr inż. Roman Jaksik

Gliwice 2023

Dissertation abstract

Models of cancer genome evolution used to evaluate the role of selection and occurrence of new mutations

mgr inż. Paweł Kuś

Cancer is one of the leading causes of death worldwide. Risk factors are often tied to lifestyle changes in developed countries, making cancer a significant research focus. The advent of Next Generation Sequencing (NGS) made molecular cancer data more available than ever and allowed scientists to unravel many molecular mechanisms that characterize cancer cells. Despite these advancements, effective anti-cancer therapy preventing the evolution towards drug resistance and relapse remains a challenge. Understanding the mechanisms that drive tumor evolution, mutagenesis, and selection may bring us closer to effective anti-cancer treatments.

Bulk DNA sequencing allows us to identify variants in tumor genomes and measure their allelic frequencies (VAF). It has been shown that processes of mutagenesis and selection shape the distribution of VAFs in the sample. Models were proposed that fit the VAF distribution with a mixture of power-law-shaped and binomial distributions. The power-law component models the neutral tail of variants, containing primarily neutral variants occurring in all cells, while the binomial components model the clones and selectively advantageous subclones. The parameters of these components reflect the evolutionary dynamics of the tumor.

We developed a new R package cevomod capable of fitting the mixture of the power-law and binomial components to the whole exome sequencing data, which previously could not be analyzed with other well-known algorithms due to the strict data quality requirements. cevomod allows one to choose between two types of models, a neutral-like one with the power-law exponent equal to 2 and an optimized model, in which the exponent is optimized to fit the data best. While the first model uses the assumptions of exponential tumor growth and constant mutation rate, the second one allows for validating these assumptions.

Using our new package and the collected data from 4 cancer types, we show that bulk DNA sequencing can be used to quantify the changes in the evolutionary dynamics of cancer upon progression, metastasis, and relapse. To prove that, we analyzed the DNA sequencing data from patients with Acute Myeloid Leukaemia, including samples from the time-points of diagnosis and relapse, patients with Breast Cancer and Laryngeal Cancer, including samples from the primary tumors and lymph node metastases, and two whole

organ maps of Bladder Cancer, including the urothelial cancer samples along with the pre-malignant samples with different stage of disease progression.

We found significant differences in the evolutionary parameters between samples from the same tumor, such as the predominant increase of the mutation rate in lymph node metastases of laryngeal cancers, compared to the primary tumors or common upward and downward changes of mutation rate in the recurrent leukaemias.

Finally, we show that the assumptions underlying the most frequently used models used to estimate the parameters of tumor evolution may be violated in many cancers. We identified significant deviations of the neutral tail power-law exponent from the expected value of 2 that may indicate the non-exponential tumor growth, changing mutation rate, or presence of selectively advantageous micro-clones. We proposed a mathematical explanation for the observed phenomena, relating the deviations to the non-constant mutation rate.

We believe that our results can contribute to the understanding of processes responsible for the evolution of cancer.