Deep learning-based integration of circulating and cellular miRNAs expression of pancreatic cancer.

Pancreatic cancer is one of the most deadly oncological diseases. According to official statistics, in Poland, it is diagnosed in over 4000 patients yearly and out of those 4000 only about 200 survives 5 years of the diseases. This high mortality is associated with late presentation of symptoms and lack of reliable blood-based test. Early detection of pancreatic cancer is a key for efficient treatment. This is why blood-based biomarkers are so extensively searched in almost every type of cancer.

During last years, researchers have pointed their attention on special kind particles that are present in blood. Over 1500 types of those particles, called microRNAs (miRNAs) can be detected in serum and although we know that they are small particles of ribonucleic acid interfering with intracellular protein creation, the reason of their presence in blood is not well understood. Some researchers suggest that via blood they transform some kind of instructions for different cells. Other researcher, however, suggest that they are just a waste (byproducts) released to the blood after cell death. Either way, both concepts could explain why we expect the profile of miRNAs in blood to be different in oncological diseases. However, it also seems obvious that this blend of different miRNAs from many sources cannot be simple in interpretation.

Classic statistics as a tool of miRNAs profile analysis is highly unreliable because of multiple comparison problem. An aid in this process is delivered by rapidly developing methods of artificial intelligence (data mining) which are designed to find patterns in huge datasets. Although these methods can model very complex problems they sometimes are prone to overfitting. Overfitting in developed classification model makes them ungeneralizable. Just due to the chance and random sampling, in analysis of over 1500 potential miRNA biomarkers, we can propose a lot of promising sets of few miRNAs. Without further external validation it remains very hard to find correct set and propose a solution.

We hope that breakthrough technology of artificial intelligence, called deep learning, will help us to solve this problem. Deep learning is an extension of artificial neural networks which consist of artificial mathematical neurons that can be trained to perform for example a task of classification, i.e. distinguishing pancreatic cancer from healthy individuals. This technology has recently revolutionized such areas as speech recognition, translations or computer vision because of its unique ability to discover deep features. In this project, despite deep feature extraction, we would also like to readapt one of the hallmarks of human intelligence, called transfer learning. In transfer learning the knowledge from one task can be readapted to different task, facilitating learning process in the second task.

Hence, in this project we aim to design and validate a novel approach of identification of pancreatic cancer signature in the profile of circulating miRNAs by proposing integration of miRNAs cellular profile. We hypothesize that using cutting edge artificial intelligence as well as the concept of transfer learning and deep feature extraction we will be able to integrate the knowledge from cellular miRNAs expression into development of deep learning model utilizing circulating miRNAs. First of all we will recruit 44 patients with pancreatic cancer and conduct next-generation sequencing of miRNAs in their serum to extend our previously recruited cohort. Then, by testing 3 integration approaches we would like to proof their abilities to maximize the resilience to overfitting of developed integrated models as well as of their ability to pointing model training towards most biologically-accurate solutions. This, hopefully, will provide new insights into biomarker discovery and will be the first paper describing the integration of cellular and circulating profile though advances artificial intelligence methods. Positive results would be a breakthrough in biomarker discover based on circulating miRNAs.