Identification of cancer-driver mutations in non-coding parts of protein-coding genes and long non-coding RNA genes

Cancer encompasses a broad spectrum of heterogeneous diseases whose development (i.e., initiation, promotion, and progression) is associated with the accumulation of numerous genetic alterations in the cancer genome, which is the hallmark of all cancers. Numerous large cancer genome sequencing studies (mostly whole-exome sequencing, WES) have been performed, and hundreds of cancer-driving genes and thousands of cancer-driving mutations have been detected. Some of these genes/mutations, e.g., *EGFR*, *BRAF*, and *JAK2*, serve as important biomarkers for cancer-targeted therapies. As the overwhelming majority of the cancer genome studies have focused on protein-coding exons of protein-coding genes, the overwhelming majority of identified cancer-driver mutations are in protein-coding sequences, which encompass barely 2% of the genome. To date, however, very little (close to nothing) is known about somatic mutations in the vast area of the non-coding part of the genome. Therefore, in this project, we are going to explore the hypothesis that there are numerous cancer-driving mutations in the non-coding part of the genome particularly in sequences coding for non-coding parts of RNA and non-coding RNAs. In this project in a large panel of cancer (predominantly lung cancer) samples, we will look for and characterize somatic mutations in untranslated regions of protein-coding genes (5'UTRs, 3'UTRs, and introns) and long non-coding RNAs (lncRNAs), to identify potential non-coding cancer drivers. Some of these drivers may become cancer biomarkers or event targets of cancer therapies.