AFM-based G-quadruplex Detection Using Machine Learning

Chika Sugimoto[1]
Yoichiro Tanaka and Toshio Ogino
[1] Yokohama National University

Detecting the structures of G4 DNA associated with human disease has become important to analyze and understand quadruplexes biological functions and properties. It is expected to develop a quick method to identify the structure of G4 DNA. We have developed a molecular structure-based G4 identification method using atomic force microscopy (AFM) and machine learning technologies. The classifiers built using random forest and Mask R-CNN could identify seven molecules that are known to form either single-stranded, double-stranded, or quadruplex structure with high accuracy and segment the particles on AFM images. Our experiments also suggested that G4s are actually formed in DNA molecules derived from E. coli genome.