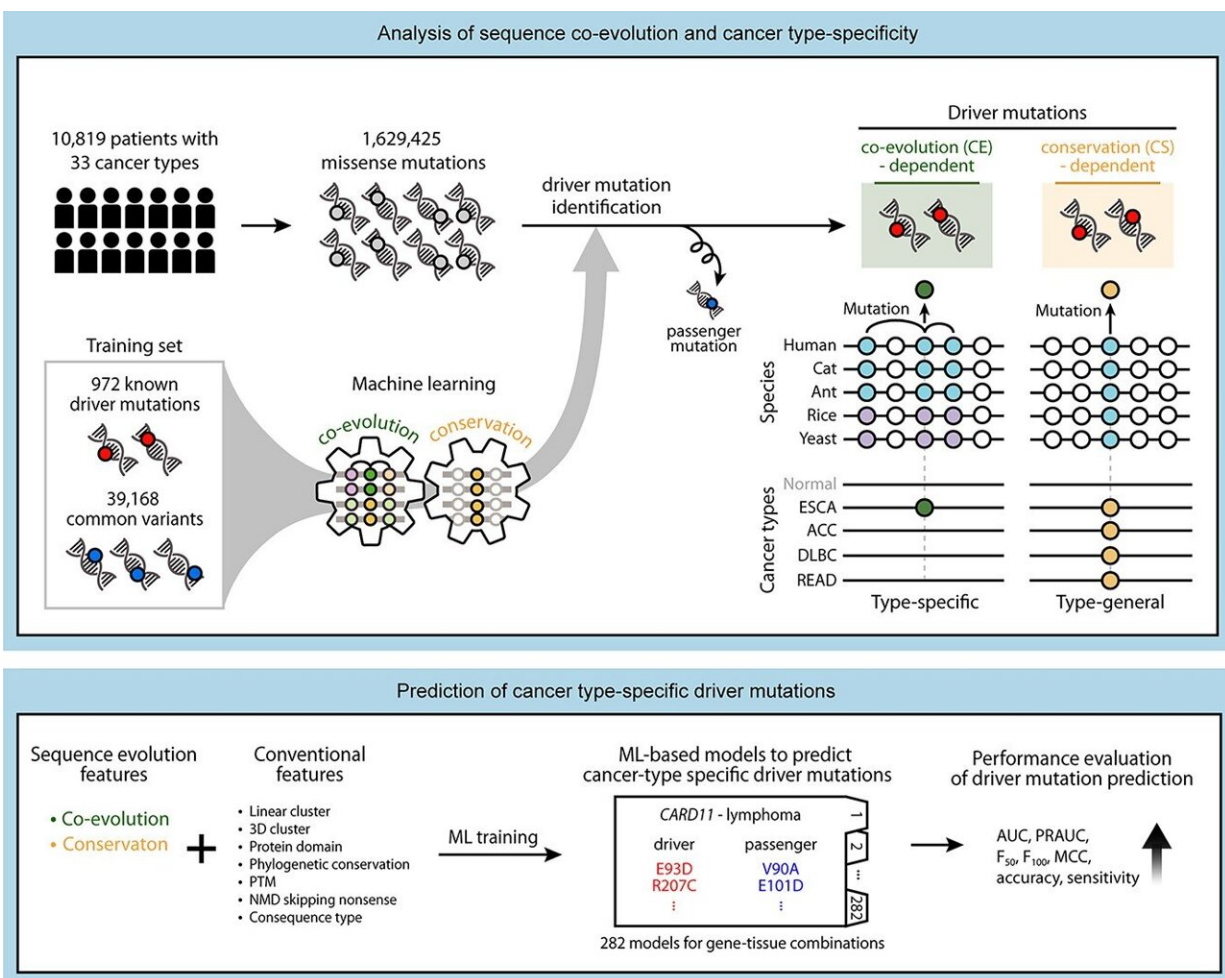# Machine learning to identify cancer type-specific driver mutations for the development of new drug targets

March 2 2023



Credit: POSTECH

According to Statistics Korea, cancer is the top cause of death in 2021, accounting for 26% of deaths. Most cancer patients miss the golden window for treatment since symptoms only develop after cancer progresses. The World Health Organization reports that more than 30% of patients can be in complete remission if cancer is detected and treated early. For early diagnosis of cancer, it is necessary to predict the driver mutations in tissues and identify if they are cancer-causing.

Recently, a POSTECH research team led by Professor Sanguk Kim, Dr. Donghyo Kim, and Dr. Doyeon Ha (Department of Life Sciences) developed a machine learning model that can accurately predict whether tissue-specific mutations in patients' genes could cause cancer. The findings from the study were published in *Briefings in Bioinformatics*.

Identifying the cancer type-specific mutations (driver mutations) is pivotal to shedding light on the distinct pathological mechanisms across various tumors and to provide each patient with opportunities for treatment.

The research team devised a novel feature based on sequence co-evolution analysis to identify cancer type-specific driver mutations and constructed a machine learning model with state-of-the-art performance. The team's ML framework outperformed current leading methods of detection as it collected data from 28,000 tumor samples across 66 cancer types.

The researchers developed a machine learning model that predicts the oncogenicity of driver mutations, using protein sequencing. The model has better accuracy and sensitivity compared to pre-existing models. Also, they successfully identified protein residues or mutations that may cause specific cancers by devising a novel feature based on sequence co-evolution analysis for machine learning.

The cancer mutations in the study have been confirmed to shape specific oncogenesis by mediating networks of tissue-specific protein interactions. These results show promise to lead to the effective prevention and treatment of cancer, combining early detection diagnostic technologies and the identification of new treatments.

"This technology can identify novel oncogenic driver mutations—that were previously undetectable—to help design distinct strategies for cancer diagnosis and treatment that are different from conventional methods," explained Professor Sanguk Kim.