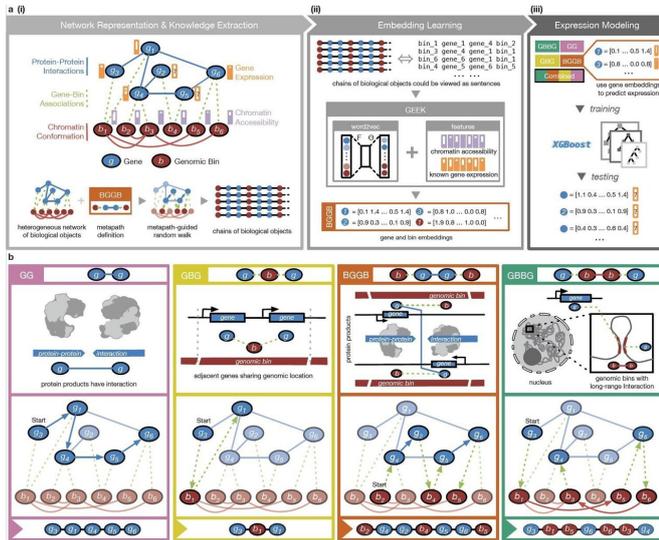# New AI approach investigates multiple gene regulatory mechanisms

31 August 2020

A schematic showing how to use machine learning and natural language processing investigate multiple gene regulatory mechanisms. Credit: CUHK

A research team from the Department of Computer Science and Engineering (CSE) at The Chinese University of Hong Kong (CUHK) has developed a new Gene Expression Embedding frameworK (GEEK), which uses artificial intelligence technologies in machine learning and natural language processing to study the regulation of gene expression. In contrast to previous works that focused on one or a few regulatory mechanisms at a time, this new framework can study the joint effects of many mechanisms simultaneously. A research article describing this new study has been published in the renowned international science journal *Nature Machine Intelligence*. The framework may help study the causes of cancers and treatment methods.

Each human body contains tens of trillions of cells. While they mostly share the same DNA sequences, their gene activities can be markedly different. Such activities, referred to as "gene expression," are affected by many regulatory mechanisms, such as transcription factor binding and protein interactions. In 2017, Prof. Kevin Yip from CUHK CSE and his research team studied one of the mechanisms that involves regulatory elements called enhancers. They investigated how enhancers are related to gene expression, and applied the results to discover three genes potentially related to liver cancer. This and other similar studies considered only individual gene regulatory mechanisms, and therefore could not fully understand the complex interplay between different mechanisms.

Prof. Yip used a metaphor to explain the intricate relationships among gene regulatory mechanisms. He said, "If you fail to turn on an electronic appliance using a remote controller, it seems like there is a problem with the controller, but the problem may also lie with the receiver or compatibility issues between the two. If we have a tool that can analyze the different components at the same time, it would be much easier to identify the root cause of the problem."

The GEEK framework proposed by Prof. Yip's team makes use of machine learning and natural language processing methods, treating genes as "words" to capture their relationships in "sentences." In the published study, GEEK was used to study several diverse gene regulatory mechanisms, including contacts in three-dimensional genome architecture, protein interactions, genomic neighborhoods and broad chromatin accessibility domains. The results showed that gene expression could be better explained when these mechanisms were modeled together than when they were considered separately.

Cancer is caused by mutations that lead to abnormal cell proliferation. "GEEK represents a novel way to study gene expression in different

types of cells, including cancer cells," says Prof Yip. "We will work closely with medical experts to try explaining some causes of [liver cancer](link) using GEEK. In the long run, we hope to extend our research to other cancer types and contribute to the development of new prevention and treatment methods."

Among cancer treatments, immunotherapies are receiving a lot of attention due to their much greater efficacy in some cancer types. Yet the treatment outcome varies from patient to patient. Prof. Yip hopes that artificial intelligence can be used in the future to predict patients' responses to immunotherapies, which would improve treatment precision and reduce the burden on patients.

Provided by The Chinese University of Hong Kong