

Researchers reveal a common deficiency in genetic prediction methods

August 29 2019



The study revealed a common deficiency in current enhancer-promoter interaction prediction methods . Credit: NUS

A study conducted by researchers from the Cancer Science Institute of Singapore (CSI Singapore) at the National University of Singapore and the School of Biological Sciences at Nanyang Technological University, Singapore (NTU Singapore) revealed a common deficiency in existing artificial intelligence methods used to predict enhancer-promoter

interactions, that may result in inflated performance measurements. The findings, published in scientific journal *Nature Genetics* in July 2019, provides an enhanced road map for the understanding of gene regulation.

An enhancer is a short sequence of DNA that works to speed up genetic transcription while a promoter is a piece of DNA which acts to initiate [gene transcription](#). Understanding the interactions between an enhancer and a promoter is critical for gene regulation studies as there is great scientific interest in whether interactions may be dysfunctional in [cancer cells](#), and present an opportunity for clinical intervention. In order to study enhancer-promoter interactions on a large scale and in a cost-effective manner, artificial intelligence methods for predicting such interactions are vital to facilitate researchers in their studies and enable them to extend the availability of such data to new cell types.

In the study conducted by Dr. Cao Fan, a research fellow at CSI Singapore, and Dr. Melissa J. Fullwood, Principal Investigator at CSI Singapore and a Nanyang Assistant Professor at NTU Singapore, the research team attempted to develop an enhancer-promoter interaction prediction method using existing datasets from TargetFinder, an advanced machine learning method that predicts enhancer-promoter interactions based on transcription factor and histone modification profiles in the window regions between enhancers and promoters. During then, the team observed that enhancer-promoter interactions were predicted at random DNA sequence features in the window regions, indicating [high performance](#).

However, upon careful examination of the TargetFinder datasets, the team realised the reported high performances could be attributed to the high overlap between window regions of positive samples in the datasets, affecting the predicted performance. To mitigate the issue of overlapping samples, the team then evaluated enhancer-promoter interaction methods using a chromosome-split strategy. TargetFinder

achieved significantly lower performance with the chromosome-split strategy, which proved that the performance measurements were indeed inflated in the earlier prediction.

The team also examined another method, JEME, a supervised machine learning method that makes use of datasets with significant differences in distance distributions between positive and negative samples to predict enhancer-promoter interactions. Their investigation revealed that JEME too, results in inflated performance measurements due to erroneous use of input data.

"Our study highlights the need for careful experimental design when applying machine learning to genomic research. It is key to properly evaluate an enhancer-promoter interaction method, and take into account the possibility of generating highly inflated performance measurement." said Dr. Cao.

"Accurate enhancer-promoter interactions prediction is essential in gene regulation studies in order to facilitate our ability to understand if there are any differences between [cancer](#) samples, such as different clinical subtypes of cancers, in order to better develop biomarkers and therapies for cancer in the future," said Dr. Fullwood.

Moving forward, the research team will be working on a new accurate machine learning approach for the prediction of enhancer-promoter interactions, and applying the method to the analysis of cancer cohorts in order to understand alterations in enhancer-[promoter](#) interactions in cancer.

Provided by National University of Singapore

Citation: Researchers reveal a common deficiency in genetic prediction methods (2019, August

29) retrieved 27 April 2024 from <https://phys.org/news/2019-08-reveal-common-deficiency-genetic-methods.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.