Michael A Beer

Department of Biomedical Engineering and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA

on

Using Machine Learning to Detect Enhancer Variation, Enhancer-Promoter Interactions, and Regulatory Conservation

Monday, February 18 at 4 p.m. s.t

Lecture Hall, ground floor, CECAD Research Center, Joseph-Stelzmann-Str. 26

Host: Andreas Beyer

Systems Biology

Abstract

Most SNPs associated with common human disease are intergenic and contribute to disease susceptibility by altering enhancer activity. Several machine learning methods have recently been proposed to detect the disrupted regulatory sequences and to predict the impact of regulatory mutations, based on SVMs and Deep Neural Networks (DNN). While these methods have similar cross-fold validation rates, variability in feature detection and importance can lead to significantly different predictions for mutation impact. Here I will show that most human population variation arises from weak transcription factor binding site disruption, and that differences in machine learning approaches can dramatically affect the accuracy of the predictions. I will then discuss serious limitations in most current machine learning formulations to predict enhancer promoter interactions and potential improvements. Finally, I will describe how these methods can identify functionally conserved regulatory elements missed by conventional sequence alignment methods. Together, these results show that statistical learning from large functional datasets can more accurately determine the quantitative contribution of weak binding sites to enhancer function in their native cellular contexts.