Name: Jiangwen Sun Email: jsun@odu.edu

Training Deep Neural Networks Predicting DNA Methylation to Elucidate Genetic Associations in Substance Use Disorders

Houyi Du, Sanjeeva Dodlapati, Jin Lu, Jiangwen Sun

University of Michigan, Dearborn, Old Dominion University

DNA methylation, a fundamental biological process in human genome, plays critical roles in the regulation of both genome stability and gene expression. Alterations in usual methylation patterns may lead to disruption of normal cellular functions and disease conditions including substance use disorders. In a given cell, the methylation state of a cytosine is largely determined by sequence patterns in its surrounding DNA fragment. As a result, perturbations in the sequence patterns caused by genetic variants may lead to alternations in methylation states. Deep neural networks (DNNs) excelling in characterizing sequence patterns has emerged as potential tools for evaluating functional impact of genetic variants. In this study, we attempted to train DNNs to predict DNA methylation in adult brain tissues at eight different locations and two fetal brains. Specifically, the networks were trained to predict methylation at individual CpG sites from their surrounding DNA sequence of the reference genome. We explored the utility of different loss functions as the training objective: logistic and hinge losses. Our results indicate that there is relatively small difference in prediction performance of trained networks between the two loss functions but with large discrepancy in learned sequence patterns. We applied the trained networks to evaluate the impact of every individual variant in the human genome (over 700 millions) by calculating the difference in the predicted methylation state between before and after introducing the variant into the reference genome. The results indicate these networks are valuable assets for elucidating functional mechanisms underlying genetic associations in substance use disorders.