Submitter Name: Xinyu Zhang
Submitted email: xinyu.zhang@yale.edu
PI Name: Ke Xu
PI email: ke.xu@yale.edu

Feature Selection of Smoking-Associated DNA methylation for Prediction on HIV Prognosis and Mortality

Xinyu Zhang[1,2], Ying Hu[3], Bradley E. Aouizerat[4], Ke Xu[1,2]

[1]Department of Psychiatry, Yale School of Medicine; [2]VA Connecticut Healthcare System; [3]Center for Biomedical Bioinformatics, National Cancer Institute, [4]Bluestone Center for Clinical Research

Background: The effects of smoking upon DNA methylation in blood from persons living with HIV may have important implications for their immune-related outcomes, including frailty and mortality. The machine learning approach enables the selection of phenotypically relevant features from high-dimensional epigenome-wide DNA methylation. Here, we report that a set of smoking-associated DNA-methylated CpGs predicts HIV prognosis and mortality in an HIV-positive veteran population.

Results: We first identified 137 significant CpGs for smoking from an epigenome-wide association study (EWAS) in 1,137 HIV-positive samples ($p < 1.70E-07$). To examine whether smoking-associated CpGs were predictive of HIV frailty and mortality, we then applied an ensemble-based machine learning from four base models [Lasso and Elastic-Net Regularized Generalized Linear Model (GLMNET), Support Vector Method (SVM), Random Forest (RF), and XGBoot] to select features in a training sample. The 698-CpG from EWAS was selected and was predictive of high HIV frailty in a testing sample [(Area Under Curve (AUC) = 0.73, (95%CI: 0.63 ~ 0.83)] and was replicated in an independent sample [(AUC = 0.78, 95%CI: 0.73 ~ 0.83)]. We further found an association of a DNA methylation index from the 698-CpGs that were associated with a 5-year survival rate [HR =1.46; 95%CI: 1.06 ~ 2.02, $p = 0.02$]. Interestingly, the 698-CpGs located on 445 genes were enriched on the integrin signaling pathway ($p=9.55E-05$), which is responsible for regulation of the cell cycle, differentiation, and adhesion.

Conclusion: We demonstrated that smoking-associated DNA methylation features in blood predict clinic outcomes in a population living with HIV.