Submitter Name: An Hoang
Submitted email: anhoang@wi.mit.edu
PI Name (if different): Olivia Corradin
PI email (if different): corradin@wi.mit.edu

Epigenomic and machine learning strategies reveal key pathways that distinguish the opioid dependent brain

An Hoang[1], Schahram Akbarian[2], Cynthia Bartels[3], Olivia Corradin[1], Dana Hancock[4], Eric Johnson[4], Katreya Lovrnert[3], Deborah Mash[5], Bryan Quach[4], Richard Sallari[6], Peter Scacheri[3];

[1]Whitehead Institute Biomedical Research; [2] Department of Neuroscience, Icahn School of Medicine at Mount Sinai; [3]Department of Genetics and Genome Sciences, Case Western Reserve University; [4]Center for Omics Discovery and Epidemiology, RTI International; [5]Dr. Kiran C. Patel College of Allopathic Medicine, Nova South Eastern University; [6]Tempus Precision Medicine Company

While opioid use disorder is estimated to be 60% heritable, the genomic regions that define this heritability remain largely unknown. ChIP-seq profiling provides an opportunity to identify epigenetic variation that distinguish opioid dependent brains. In contrast to traditional approaches, which evaluate each regulatory element independently, machine learning enables identification of sets of elements that together can distinguish opioid cases from controls. Here, we fit gradient boosted tree models on H3K27ac ChIP-seq data of neuronal nuclei isolated from 51 opioid cases and 51 accidental death controls. We identified a model that required only six ChIP-seq peaks to accurately differentiate opioid cases from controls (5-fold cross-validated AUC = 0.98). This model outperformed models based on randomly selected peak sets or peaks identified through linear regression ($P<0.001$). We reasoned that there may be additional peak sets capable of achieving similar predictive accuracy. To evaluate this hypothesis, we developed a step-wise strategy which iteratively removes peaks and then identifies new models from the remaining peaks. In total, we identified 43 independent models ranging from 4 to 43 peaks with an AUC >0.9. Using promoter-captured Hi-C data, we identified putative gene targets of model peaks. While individual peaks were unique to each model, we found several genes that were incorporated across multiple models including DUSP4, GABBR2 and KCNMA1. Gene targets of model peaks were found in inter-connected gene networks involving MAPK and GABA signaling Altogether, these results provide new insights into the key epigenetic differences and pathways that distinguish opioid dependent brains.