Submitter Name: Bryan Quach
Submitted email: bquach@rti.org
PI Name (if different): Dana Hancock
email (if different): dhancock@rti.org

Smoking-dependent expression QTL evidence in human nucleus accumbens among variants identified by the GSCAN consortium

Bryan C. Quach[1], Christina A. Markunas[1], Ran Tao[2], Amy Deep-Soboslay[2], Laura J. Bierut[3], Thomas M. Hyde[2], Joel E. Kleinman[2], Eric O. Johnson[1], Andrew E. Jaffe[2], Dana B. Hancock[1]

[1]Center for Omics Discovery and Epidemiology, RTI International; [2]Lieber Institute for Brain Development (LIBD); and [3]Department of Psychiatry, Washington University in St. Louis

The GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) identified hundreds of genome-wide significant variants associated with smoking initiation, initiation age, and cessation. Most associated variants are noncoding: their neurobiological effects and target genes largely unknown. To evaluate their regulatory potential, we used genome-wide genotypes and RNA-sequencing data in postmortem human nucleus accumbens from the LIBD Human Brain Repository, defining active smokers (N=57) and nonsmokers (N=168) based on detectable cotinine and next-of-kin report of smoking. We ran parallel *cis*-expression quantitative trait loci (*cis*-eQTL) models, with and without variant-by-smoking interaction. Our baseline model (without interaction) resembles the Genotype-Tissue Expression (GTEx) model for mapping *cis*-eQTLs, where smoking information is unavailable. For the model with interaction, we used the joint 2df method to simultaneously test the variant main and interaction effects on gene expression. Under the baseline model, we identified 37 significant (Bonferroni corrected) *cis*-eQTL variants among the 455 GSCAN variants, and another 10 variants identified only when taking variant-by-smoking interaction into account: smallest $P_{baseline}=4.9\times10^{-23}$ vs. smallest $P_{2df}=8\times10^{-29}$. Examples of variants with no *cis*-eQTL evidence in nucleus accumbens from our baseline model or GTEx that were revealed by interaction with smoking include: rs6874731 on *ARSB* expression ($P_{baseline}=0.29$ [GTEx P=0.24] vs. $P_{2df}=4.9\times10^{-5}$) and rs17197663 on *CCDC169* expression ($P_{baseline}=0.41$ [GTEx P=0.8] vs. $P_{2df}=5.0\times10^{-5}$). Besides highlighting target genes for robustly associated variants for smoking from GSCAN, our findings emphasize that active smoking is an important context for *cis*-eQTL mapping and that smoking-dependent QTLs are missed in agnostic data that do not account for smoking.