Submitter Name: Fang Fang Submitter Email: ffang@rti.org

Exposure- and genetically driven epigenetic associations with lifetime cannabis use.

Stephanie N. Giamberardino[1], Linran Zhou[1], Jesse A. Marks[1], Kaitlyn G. Lawrence[2], Zongli Xu[2], Jack A. Taylor[2], Dale P. Sandler[2], Aino E. Heikkinen[3], Mikaela Hukkanen[3], Miina Ollikainen[3], Jaakko Kaprio[3], Laura J. Bierut[4], Jenny van Dongen[5], Dorret I. Boomsma[5], Youshu Cheng[6], Ke Xu[6,7], Sergio Villicaña Muñoz[8], Jordana T. Bell[8], Dana B. Hancock[1], Eric O. Johnson[1,9], Fang Fang[1]

[1]GenOmics and Translational Research Center, RTI International;
[2]Epidemiology Branch, National Institute of Environmental Health Sciences;
[3]Institute for Molecular Medicine Finland, University of Helsinki;
[4]Department of Psychiatry, Washington University School of Medicine;
[5]Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam;
[6]Department of Psychiatry, Yale School of Medicine;
[7]VA Connecticut Healthcare System;
[8]Department of Twin Research & Genetic Epidemiology, King's College London;
[9]Fellow Program, RTI International

Cannabis use, though prevalent, may be poorly reported, making it important to find a reliable biomarker for quantifying lifetime use. Leukocyte DNA methylation (DNAm) may serve as a viable and sensitive biomarker of cannabis use. Previously, we identified DNAm associations with cannabis use, but DNAm levels can also be influenced by genetic variation. Here, we identify genetically-associated DNAm signals, with the aim to exclude these effects when characterizing DNAm biomarkers of cannabis use. We identified *cis*-methylation quantitative trait loci (*cis*-meQTL) for cannabis-associated CpGs identified in our prior EWAS. We conducted within-ancestry (African and European) inverse variance-weighted meta-analyses of SNP-CpG associations generated within cohort using the Matrix eQTL software. Amongst the N=1,353 CpGs meta-analyzed in either ancestry group, we identified *cis*-meQTL (random effects p-value < 1×10^{-5}) for N_{Afr}=244 and N_{Eur}=386 CpGs, with N=154 CpGs having *cis*-meQTL identified in both groups. Using location-based enrichment testing, we found that CpGs with *cis*-meQTL were significantly (FDR adjusted p-value<0.05) enriched in CpG shore regions (FDR p_{Afr}=0.001, FDR $p_{Eur}=8.36\times10^{-7}$) and depleted in open sea regions (FDR p_{Afr}=0.001, FDR p_{Eur}=0.043), compared to the total set of meta-analyzed cannabis-associated CpGs. For the CpGs with *cis*-meQTL identified in the African- and European-ancestry meta-analyses, 89% and 95%, respectively, of these CpGs had previously detected *cis*-meQTL in the GoDMC study, the largest meQTL database with predominantly European ancestry samples. These results suggest an influence of unaccounted-for genetic variation in prior cannabis CpG associations. Excluding these genotype-related CpGs will enhance the power to identify cannabis exposure-associated DNAm biomarkers.