

Submitter Name: Christina Markunas  
Submitted email: [cmarkunas@rti.org](mailto:cmarkunas@rti.org)

## Epigenome-wide analysis uncovers a blood-based DNA methylation biomarker of lifetime cannabis use.

Christina A. Markunas<sup>1</sup>, Dana B. Hancock<sup>1</sup>, Zongli Xu<sup>2</sup>, Bryan Quach<sup>1</sup>, Dale P. Sandler<sup>2</sup>, Eric O. Johnson<sup>1,3</sup>, and Jack A. Taylor<sup>2,4</sup>

<sup>1</sup>Center for Omics Discovery and Epidemiology, RTI International; <sup>2</sup>Epidemiology Branch, National Institute of Environmental Health Sciences (NIEHS); <sup>3</sup>Fellow Program, RTI International; <sup>4</sup>Epigenetic and Stem Cell Biology Laboratory, NIEHS

**Background/Significance:** Cannabis, the most commonly used illicit drug in the US, is associated with adverse and beneficial health effects. To understand the full spectrum of health effects, biomarkers that accurately quantify lifetime cannabis use are needed. DNA methylation (DNAm) is an excellent biomarker candidate, yet no epigenome-wide association studies (EWAS) of cannabis exposure in humans have been published.

**Methods:** We conducted an EWAS using blood-based Illumina 450K data from the Sister Study, a prospective cohort of women (Discovery N=1,730; Replication N=853). We ran robust linear regression models of DNAm by lifetime cannabis use, adjusting for age, breast cancer, tobacco and alcohol use, technical factors, and cell-type proportions. Additionally, a multi-CpG predictor of lifetime cannabis use was developed using penalized regression of top EWAS CpGs and validated in the replication sample.

**Results:** We identified and replicated one significant association (false discovery rate<0.10) at cg15973234 (*CEMIP*): meta  $P=3.3\times 10^{-8}$ . Significant *cis*-meQTLs for cg15973234 (BIOS QTL browser) were not associated with lifetime cannabis use in the largest GWAS to date (N=184,765; Pasman *et al.* 2018), suggesting an exposure rather than genetically-driven cg15973234-cannabis association. Within the replication sample, our multi-CpG biomarker's predictive performance was fair with an area under the receiver operating characteristic curve of 0.75 ( $P=2.2\times 10^{-16}$ ).

**Discussion:** Our findings add evidence that blood-based DNAm can inform cannabis use histories and suggest that cannabis exposure itself underlies the cg15973234 association. Expanded sample size and additional phenotype data are needed to uncover additional DNAm changes and evaluate secondary applications of DNAm biomarkers (e.g., predicting cannabis-related effects).