

Name: Fang Fang

Email: ffang@rti.org

A DNA methylation-based classifier for lifetime cannabis use

Fang Fang¹, Linran Zhou¹, Bryan C. Quach¹, Kaitlyn G. Lawrence², Jenny van Dongen³, Jesse A. Marks¹, Zongli Xu², Laura J Bierut⁴, Miina Ollikainen⁵, Jack A. Taylor², Jordana T. Bell⁶, Jaakko Kaprio⁷, Dorret I. Boomsma³, Ke Xu^{8,9}, Dale P. Sandler², Dana B. Hancock¹, Eric O. Johnson^{1,10}

¹Omics, Epidemiology, and Analytics Program, RTI International, Research Triangle Park, NC, USA.

²Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ³Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ⁴Department of Psychiatry, Washington University in Saint Louis School of Medicine, St. Louis, Missouri, USA.

⁵Minerva Foundation Institute for Medical Research, and Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland. ⁶Department of Twin Research & Genetic Epidemiology, King's College London, London, UK. ⁷Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland. ⁸Department of Psychiatry, Yale School of Medicine, West Haven, CT, USA.

⁹VA Connecticut Healthcare System, West Haven, CT, USA. ¹⁰Fellow Program, RTI International, Research Triangle Park, NC, USA.

Cannabis use may result in persistent changes across the DNA methylome, suggesting the DNA methylome has the potential to serve as a reliable biomarker for cannabis exposure. In this study, we developed a classifier to predict lifetime cannabis use (ever vs. never) using DNA methylation data from peripheral blood samples. We identified 108 significant CpGs ($p < 1E-4$) for lifetime cannabis use in a cohort of 2,073 (1,009 ever users) participants, all of European ancestry, within the Sister Study, a prospective cohort of women at risk of developing breast cancer. Using penalized regression with 10-fold cross-validation, 77 CpGs were selected to build a model that predicted lifetime cannabis use in a testing sample of 517 participants from the Sister Study, achieving an AUC (area under curve) of 0.67 (95% CI 0.62~0.71, $p = 1.36E-6$). We validated the classifier's performance in an independent sample from the Gulf Long-Term Follow-Up Study (GuLF), which included 1,195 individuals (665 ever users). Stratified by ancestry, the classifier was predictive of lifetime cannabis use in both European (AUC=0.62, 95%CI 0.58~0.66, $p = 3.13E-4$) and African (AUC=0.62, 95%CI 0.57~0.66, $p = 1.29E-3$) ancestry groups. These findings suggest that a peripheral blood-based biomarker may be developed to identify lifetime cannabis use and enable research into its health effects.