Genetic architecture of four smoking behaviors using partitioned SNP heritability.

Luke M. Evans\textsuperscript{1,2}, Seonkyeong Jang\textsuperscript{3}, Dana B. Hancock\textsuperscript{4}, Marissa A. Ehringer\textsuperscript{1,5}, Jacqueline M. Otto\textsuperscript{3}, Scott I. Vrieze\textsuperscript{3}, Matthew C. Keller\textsuperscript{1,6}

\textsuperscript{1}Institute for Behavioral Genetics, University of Colorado; \textsuperscript{2}Department of Ecology & Evolutionary Biology, University of Colorado; \textsuperscript{3}Department of Psychology, University of Minnesota; \textsuperscript{4}GenOmic, Bioinformatics, and Translational Research Center, RTI International; \textsuperscript{5}Department of Integrative Physiology, University of Colorado; \textsuperscript{6}Department of Psychology and Neuroscience, University of Colorado

Smoking is a leading cause of premature death. Although genome-wide association studies have identified many loci that influence smoking behaviors, much of the genetic variance remains unexplained. We characterized the genetic architecture of four smoking behaviors (smoking initiation, age of initiation, cigarettes per day, and smoking cessation) through SNP-based heritability ($h^2_{SNP}$), an estimate of narrow-sense heritability specifically estimating the proportion of phenotypic variation due to causal variants tagged by SNPs partitioned among frequency, LD, or functional annotation categories. When applied to smoking traits assessed in UK Biobank (N=54,792 to 323,068), we estimated that roughly 18% and 12% of the phenotypic variance in smoking initiation and smoking cessation, respectively, was captured by imputed SNPs ($h^2_{SNP}(SE)=0.18(0.01)$ and $0.12(0.02)$), both of which were more than twice as large as previously reported estimates. Age of initiation $h^2_{SNP}=0.05(0.01)$ and binned CPD $h^2_{SNP}=0.1(0.01)$ were similar to previous reports and substantially less than published twin-based $h^2=50\%$. CPD encoding influenced estimates, with dichotomized CPD $h^2_{SNP}=0.28$. Functional annotations related to LD, allele frequency, sequence conservation, and genes specifically expressed within individual brain regions contributed significantly to partitioned heritability. We found no evidence of dominance genetic variance for any trait. The phenotypic variance captured by genetic variants, $h^2_{SNP}$, of these four specific smoking behaviors is modest overall. The patterns of partitioned $h^2_{SNP}$ for these highly polygenic traits is consistent with negative selection and a key role of several brain regions. Deep sequencing of large samples and/or improved imputation will be required to fully assess the role of rare variants.