Submitter Name: Greg Keele
PI Name: Dan Jacobson

Submitted Email: gkeele@rti.org
PI Email: jacobsonda@ornl.gov

Network biology algorithms identify biological pathways underlying cigarette smoking behaviors

Gregory R. Keele[1], Kyle A. Sullivan[2*], Alice Townsend[3], J. Izaak Miller[2], David Kainer[2], Matthew Lane[3], Mikaela Cashman[2], Michael R. Garvin[2], Peter Kruse[3], Bryan C. Quach[1], Caryn Willis[1], Ke Xu[4,5], Bradley E. Aouizerat[6], Eric O. Johnson[1,7], Dana B. Hancock[1], and Daniel A. Jacobson[2]

[1]GenOmics and Translational Research Center, RTI International, Research Triangle Park, NC;
[2]Computational and Predictive Biology Group, Oak Ridge National Laboratory, Oak Ridge, TN;
[3]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee-Knoxville, Knoxville, TN;
[4]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA;
[5] Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA;
[6]Translational Research Center, College of Dentistry, New York University, New York, NY, USA;
[7]Fellow Program, RTI International, Research Triangle Park, NC

Recent genome-wide association studies (GWAS) from the GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) have identified hundreds of genome-wide significant loci contributing to the heritability of cigarette smoking behaviors. However, the biological pathways reflected by the full repertoire of smoking-associated variants are poorly understood. To better understand the biological pathways underlying smoking behaviors, we applied the network biology algorithms GRIN and MENTOR to GSCAN GWAS summary statistics from European ancestry subjects for Smoking Cessation (SmkCess; 388,313 individuals) and Cigarettes Per Day (CigsPerDay; 326,497 individuals). We assigned genes from genome-wide significant ($p < 5e^{-8}$) single nucleotide polymorphisms (SNPs) by using multiple methods, including SNP-nearest gene and H-MAGMA. Using a multiplex network of 10 gene-gene network layers from distinct types of biological experimental evidence including two dlPFC-specific layers, we applied GRIN to remove false positive genes based on the premise that correctly assigned genes would be highly interconnected in the networks, as determined by the random walk with restart (RWR) network traversal algorithm. We assigned 526 unique genes targeted by SNPs associated with CigsPerDay, and GRIN retained 235 genes based on high network interconnectivity. We assigned 310 unique genes targeted by SmkCess-associated SNPs, and GRIN retained 149 highly interconnected genes. We then used MENTOR to identify functional groupings of smoking-associated genes based on RWR-based network topology, followed by gene set enrichment of biological pathways. These systems biology approaches link these different GWAS-based genetic architectures to fundamental biological mechanisms, including fatty acid metabolism and microtubule dynamics.