Name: Bradley Webb, Ph.D. Email: btwebb1@geisinger.edu

Assembling a large genetically informed cohort for longitudinal analysis of smoking behavior

Background: Tobacco smoking and nicotine use continues to have significant negative impacts on both personal and public health despite decades of public health efforts. While discoveries of genetic loci influencing smoking initiation, quantity, and cessation have advanced, most genetic association results are based on lifetime status or cross-sectional phenotyping. There are significant opportunities to understand how genetic factors act over time and life-course. However, longitudinal analyses require large samples, repeated measures, and measured genotypes. Assembling this combination is challenging, both in scale and cost. Here we report a summary of a resource to address this challenge.
Methods: We curated self-report measures of tobacco use across ~230k exome sequenced individuals in Geisinger's biobank using the Phenomics Initiative Database (PIDB) which integrates data from Geisinger's data warehouse, including lab values, prescriptions, and surveys. Smoking class, quantity, change over time, and periods covered by repeated measures were systematically summarized and used to construct a longitudinal matrix.
Results: Self report smoking status was obtained for >1.9 million individuals which represented >41 million unique responses. For sequenced individuals, there were >16 million observations with a median of 56 observations covering 15.4 years. Four smoking status classes were identified, including persistent never (>105k), former (>29k), and current smokers (>15k) as well as individuals that changed class at least once (>79k). For persistent current smokers with repeated measures, quantity class (e.g. heavy to light) was stable with 89.5% (13,917/15,553) not changing class.
Discussion: This represents the largest dataset with repeated measures of self-report smoking status in genotyped individuals, to our knowledge. This resource will enable investigating the role of known genetic associations on smoking behavior stability over time and across a wide range of ages. Future analyses include growth curve mixture modeling with the inclusion of smoking and substance use disorders polygenic risk scores.

Authors:
Submitter: Bradley Webb, Senior Investigator, Geisinger, Danville
Donielle Beiler, M.S., Geisinger
Vanessa Troiani, Ph.D., Geisinger

DRAFT