Background: Strategies to reduce smoking prevalence include developing more effective smoking cessation treatments, including biomarker assessment of nicotine metabolism to personalize treatment. We and others have identified multiple CYP2A6 SNPs associated with the nicotine metabolite ratio, a measure of nicotine metabolism, at genome-wide significance. We present our work developing prediction models of nicotine metabolism activity in multiple ethnicities, and plans to translate this research into clinical genomics to support smoking cessation.

Methods: We used existing Smokescreen genomewide and novel CYP2A6 genotype and CNV data and the biochemically measured nicotine metabolite ratio (NMR) from laboratory studies from African, Asian and European ancestry individuals, to model nicotine metabolism activity. We apply existing CYP2A6 knowledge-bases, an existing CYP2A6-centered model, and polygenic approaches to incorporate CYP2A6 with other genes involved in the nicotine metabolic pathway. We evaluate potential utility of these models for translational research.

Results: Quality control yielded Smokescreen and NMR data from N=312 individuals. We identified PharmGKB-annotated CYP2A6 *alleles, constructed diplotypes, and estimated correlations of PharmGKB annotation and the Bloom et al CYP2A6 activity model with study NMR. In our modeling of the nicotine metabolism pathway, we detect associations in multiple genes, including the {EGLN2, CYP2A6, CYP2A7, CYP2G1P, CYP2B7P1, CYP2B6} complex, UGT2B10, FMO3, AOX1, POR, HNF4A, NFE2L2, and the {UGT1A4, UGT1A9} complex.

Conclusions: The NMR has analytic validity, clinical validity and clinical utility as a peripheral biomarker for treating tobacco use disorders. Estimating nicotine metabolism activity through a genomic approach has been effective in the context of clinical research of nicotine dependence, and in the retrospective analysis of response to nicotine replacement therapy in clinical trials. We plan to validate novel modeling approaches in additional independent datasets and to extend this approach to model response to smoking cessation treatments. We envision translation of these models into custom arrays and algorithms for personalized treatment of tobacco dependence.