

Submitter Name: James W. Baurley  
Submitted email: [baurley@biorealm.ai](mailto:baurley@biorealm.ai)

### **Prediction of nicotine biomarkers from genotypes**

James W. Baurley<sup>1</sup>, Carolyn Ervin<sup>1</sup>, Christopher McMahan<sup>2</sup>, Andrew W Bergen<sup>1,3</sup>

<sup>1</sup>BioRealm, LLC; <sup>2</sup>School of Mathematical and Statistical Sciences, Clemson University;  
<sup>3</sup>Oregon Research Institute

Tobacco-attributable disease remains the largest potentially modifiable cause of mortality. Nicotine metabolism and dependence are predictors of smoking behaviors, including the development of dependence, and response to smoking cessation treatments. In our NIH Small Business Innovation Research (SBIR) project, we are developing prediction models of nicotine biomarkers (nicotine metabolite ratio; NMR and total nicotine equivalents; TNE) and smoking cessation from genomic, metabolic and behavioral data. We will perform initial clinical validation in collaboration with ongoing tobacco dependence treatment programs. Our translational goal is to provide an optimized set of models to implement in a prediction service where clinicians and researchers could upload genomic and clinical data and obtain predictions. In the first phase of the project, we applied a series of machine learning algorithms to nicotine metabolism data to develop a prediction ensemble containing 70 variants in the *CYP2A6* and *CYP2A7* regions as well as regions that regulate *CYP2A6* transcription and metabolic activity. We have since validated the models in a large cohort of smokers with genotypes and the NMR, and the genomic signature with cigarettes per day in UK Biobank smokers. We are now planning to apply our validated approach to predict the NMR and TNE biomarkers from genotypes and relate them to smoking outcomes in new datasets.