**Smoking PheWAS Workspace information**

**Title: PheWAS Smoking with Focus in Respiratory system**

**Primary purpose of project:**

**Educational**

**Question: What are the specific scientific question(s) you intend to study, and why is the question important (i.e. relevance to science or public health)?**

This tutorial will present the results of Phenome-Wide Association Studies (PheWAS) to show how the various sources of data contained within All of Us research dataset can be used to inform scientific discovery. We will perform separate PheWAS studies with smoking status as the independent variable. Specific questions include: 1. How can one implement a PheWAS within the All of Us Researcher Workbench? 2. How can one use heterogeneous data sources within the All of Us dataset to explore disease associations using self-reported exposures (Survey, Participant Provided Information, or “PPI”) and exposures captured in the electronic medical record (EHR).

**Question: What are the scientific approaches you plan to use for your study? Describe the datasets, research methods, and tools you will use to answer your scientific question(s).**

As a method for assessing the health burden of smoking on potential observed phenotypes, we implement a Phenome-Wide Association study. A Phenome-wide association study consists of an array of association tests over an indexed representation of the human phenome. In this analysis, we will conduct PheWAS for EHR derived smoking and PPI derived smoking exposures included in the All of Us research dataset. We will be representing "Smoking Exposure” in three ways: EHR Smoking ICD Billing Codes Participant Provided Information (PPI) Smoking lifetime 100 cigarettes yes/no Participant Provided Information (PPI) Smoking lifetime smoking everyday To perform PheWAS, we will map ICD representations of disease to a common vocabulary of PheCodes. We then use Jupyter Notebooks to create reusable functions to perform PheWAS and generate Manhattan Plots to summarize associations.

**Question: What are the anticipated findings from the study? How would your findings contribute to the body of scientific knowledge in the field?**

This tutorial will replicate known respiratory disease associations with smoking exposure as a teaching experience. This will serve to demonstrate the power of gathering multiple data sources for a single phenotype, providing options for study design and validation. Importantly the entire pheWAS package is made available for reuse by researchers in the Workbench, for new hypothesis generation.