# Mapping Multi-layered Oncology Data in OMOP

**John Methot[1], Sherry Lee[1]**
**[1]Dana-Farber Cancer Institute, Informatics & Analytics Department**

## Background

OMOP has long been used for the study of chronic diseases, but observational research in oncology using OMOP is relatively new and conventions are still maturing. Oncology research requires significantly more details in modeling diagnosis, treatment, and outcomes. Unlike non-oncology disease areas, oncology also requires mapping data from additional sources beyond EHRs or claims. Required data is found in cancer registries, genomic databases, and manually curated disease-specific databases. Here we describe our experience mapping oncology data from multiple sources to our OMOP CDM.

## Methods

We wished to build a deep repository of data for our cancer patients from the following sources:

- Our EHR for the traditional OMOP domains of PATIENT, VISIT, CONDITION, DRUG EXPOSURE, OBSERVATION, and MEASUREMENT (N=300,000)
- The Dana-Farber Cancer Registry, which contains highly structured diagnosis data needed to accurately phenotype patients, including molecular biomarkers not available in structured form elsewhere (N=60,000)
- Dana-Farber's Genomic Database, which contains structured somatic DNA variant calls for tumors dating to 2011 from both internal and commercial tumor mutation assays (N=55,000)
- Dana-Farber's Enterprise Curation Platform, a nascent system for curating patient diagnosis, treatment, and outcome data to fill gaps in the sources described above (N=2,000)

## Results

Many of the necessary oncology data elements contained in the sources described above have not previous;y been mapped to OMOP for a large patient population. Our mapping process is currently underway (June 2023). In our poster we will report 1) the quantitative details of the mapping including characterization statistics; 2) the challenges we faced mapping new data types to OMOP in a way that supports valuable oncology phenotyping; 3) lessons learned to assist in other cancer centers mapping data to OMOP; 4) example phenotypes that are only possible with mapped data of the depth described

## Conclusion

Our poster described the relatively novel exercise of mapping deep cancer patient data to the OMOP CDM to support meaningful observational research in oncology. We share our experience to help subsequent efforts at other cancer institutions and to promote OMOP-enabled cancer research.