UC DAVIS HEALTH

Project: UC Health cBioPortal

Submitter: Cy Huynh, MPH, Data Curation Supervisor, UC Davis Health, cvhuynh@ucdavis.edu, (916) 798-8744

I. Project Team

The UC Davis Health Data Provisioning Core (DPC) in IT Health Informatics promotes an interdisciplinary organizational structure that includes clinical expertise, clinical IT Informatics and Research Informatics to advance applied and high-quality domain-focused applications within the UC Davis Health (UCDH) learning healthcare system. One of the key successes of this project was the formation of the oncology domain specific Cancer Data Informatics Integration Initiative (CDI3) team at UC Davis Health to ensure partnerships with collaborators within UC Davis and across other UC campuses.

UC Davis Health Team	Implementation Partners
Project Leadership Kent Anderson, M.S. Nicholas Anderson, Ph.D. Jason Yeates Adams, M.D. Sharon L. Myers, Ph.D. **Project Sponsors** John McPherson, Ph.D. Primo N. Lara, Jr., MD, Director, University of California Davis Comprehensive Cancer Center **Project Team** Cy Huynh, M.S., MPH Matthew Renquist Jared Cobabe Joseph Cawood, MHA Albert William Riedl, M.S. Christopher Lambertus	**University of California Comprehensive Cancer Center San Diego** Olivier Harismendy, Ph.D. **University of California Comprehensive Cancer Center San Francisco** Brenda Venkatesh, Ph.D. Michelle Turski, Ph.D. Gundolf Schenk, Ph.D.

II. Problem statement and Summary

The 5 UC NCI-designated comprehensive cancer centers each serve a diverse and unique regional patient population, and collectively facilitate a network of cancer treatment and support that spans the UC Health system and State of California. Each of the centers leverages a multitude of commercial genomic lab vendors and specialist expertise to assess genetic variants in cancer patients as to inform protocols for determining the best course of treatment for patients diagnosed with cancer. To address the collaborative potential of the full UC CCC's oncology and treatment expertise, and to aid in providing the full range of information and knowledge contained within these results to the clinicians and researchers across the system, we developed a collaborative environment that seeks to empower clinicians and researchers in discovery and leverage of genomic data annotated with with clinical attributes. As part of the UC Davis Cancer Data Informatics Integration Initiative (CDI3) and IT Health Informatics, UC Davis Health adapted and implemented a research and clinical focused tool

developed by Memorial Sloan Kettering to provide an interactive exploration view of UC curated cancer genomics data sets. This builds towards further partnership with oncologists and researchers across the UC, and targets future support by the health system and extension of this tool through a focus on reusability and extensibility. The initiative:

1) Built data pipelines that ingested UCDH cancer genomic lab results into a secure cancer informatics data store,

2) Shared our genomics data model and code to the UC Health Data Warehouse team to build out the UC Data Warehouse OMOP genomic extension and

3) Pushed genomic results in flat files into the cBioPortal environment for data exploration by principal investigators and clinical oncologists across multiple studies, genes, samples, and clinical data.

The technical data provisioning pipeline is described in Figure 1.

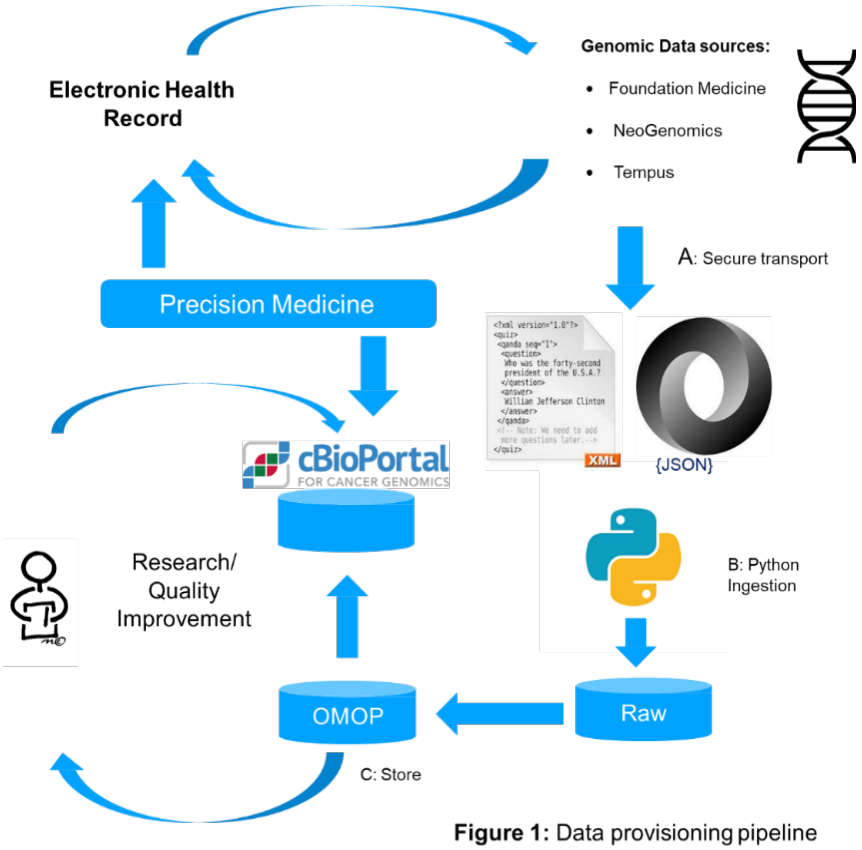

Figure 1: Data provisioning pipeline

III. Project Narrative and Problem Statement

Healthcare and linkage to genomic data continues to expand in complexity, volume and heterogeneity. At the UC Davis Comprehensive Cancer Center and across all UC Comprehensive Cancer Centers, a variety of genomic lab tests from a diverse set of vendors are now leveraged as part of the standard of care for assessing treatment paths for cancer and other chronic patients. Lab results of these assays are still often manually scanned into the EHR with content that is not searchable or shareable in any forms, thereby resulting in usage that is one patient, one report at a time, and a constrained group of physicians or researchers. This single-use and limited view of the resulting data and metadata continues to create barriers to clinicians and researchers in accessing the molecular profiles and clinical attributes of patients as part of the full potential of treatment outcomes. Developing institutional

data curation, integration between clinical data and genomic laboratory data results and provisioning these data within a secure data platform is still isolated within indivudal health systems, and particularly lacking in cancer networks or large health systems. Multi-disciplinary access to common and curated platforms for these reviews are central to supporting precision and population health and improving the care of cancer patients.

Solution

Through close partnership of the clinical expertise from oncology subject matter experts and the informatics research data provisioning services, we sought to evalute and leverage a cancer genomics EHR data workflow that links the UCDH cancer registry with clinical patient cohort data from the Epic EHR. The resulting data products are comprehensive and curated cancer patient data enriched with genomics data and outcomes that is advancing the measurement of somatic gene variation for clinical, quality improvement and cancer research needs. The EHR workflow established an iterative and repeatable process that enables genomic data ingestion from and linkage to EHR data from a variety of vendors.

One of the consumers of this data curation process is cBioPortal (cBP), an open-source data visualization and exploration platform for cancer molecular profiles alongside patient clinical data. UCDH's cBioPortal implementation leveraged the genomic profiling assay results from Foundation Medicine integrated with data curated from the local cancer registry and Epic to provide additional characteristics about the disease populations available to the platform. The infrastructure, tooling and methodology from the UC Davis Health implementation of cBioPortal resulted in a new UC-wide solution that has standardized and generalized approaches for use by other UC sites, and enables data sharing and cancer cohort discovery across the network of UC medical and cancer centers

Our informatics project champion (Nick Anderson) and project sponsor (John McPherson), along with support from our Cancer Center director (Dr. Primo Lara, Jr.) and the UC Cancer Consortium, gained critical collaboration from UC San Diego and UC San Francisco as initial partners to build out a central UC cBioPortal environment. The UC Davis Health team spearheaded the implementation of the central UC cBioPortal environment, provided project management and cancer-informatics technical expertise to guide our collaboration partners in their implementation and sharing of deidentified Foundation Medicine genomic data into the platform. We also collaboratively developed and managed the roadmap for implementation with our partners, including defining short- and long-term strategies for maintenance and further support.

Through monthly working meetings with UC San Diego, UC San Francisco, and UCLA, we were able to fine tune the implementation framework, develop pilot evaluation, and address technical hurdles in data curation, ingestion and provisionig. The result is a novel views of deidentified pan-Cancer study linking Foundation Medicine genomics data from UC Davis Health and UC San Diego, loaded onto the cBioPortal analytics platform, and providing the capability to explore data on various cancer types with specific genetic molecular profiles. Figure 2 provides a UI view of the combined study platform.

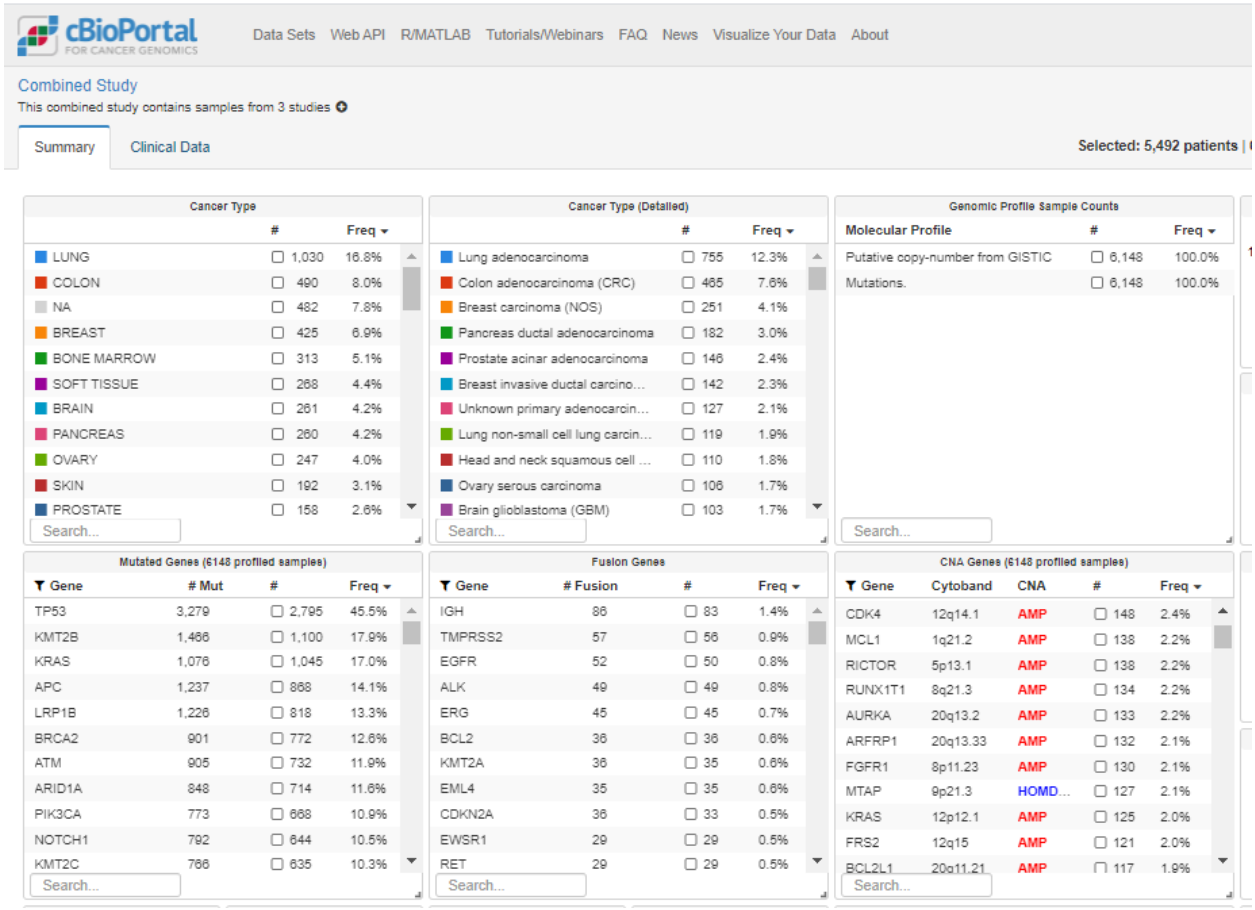

Figure 2. UC cBioPortal Pan-Cancer Study

Impact and Project Success of the UC cBioPortal

The joint effort between UC Davis Health and UC San Diego combined roughly 6,000 studies across both health systems and molecularly characterized over 50 cancer types including the commonly occurring cancers such as lung, colon, breast, prostate and melanoma. Locally, cBioPortal has been used as cohort discovery and characterization tool to identify populations for research resulting in two research papers currently in the works on lung cancer.

The Foundation Medicine genomic data pipeline the UC Davis Health team established has also been leveraged to create the UC Health OMOP genomic data extension. This expanded the footprint of collaboration with UC Davis, UC San Francisco, UCLA, and UC San Diego all contributing data to the OMOP Genomic-Common Data Model (CDM), with UC Irvine currently in process to contribute to this data feed and resource. This has resulted in the centralization of over 22,000 samples and results across these 4 medical centers that are directly queryable and searchable. With data now streaming from 4 UCs into the OMOP genomic extension, UC Davis is developing the next stages of the cBioPortal instance to point to the shared OMOP common data model (see Figure 3).

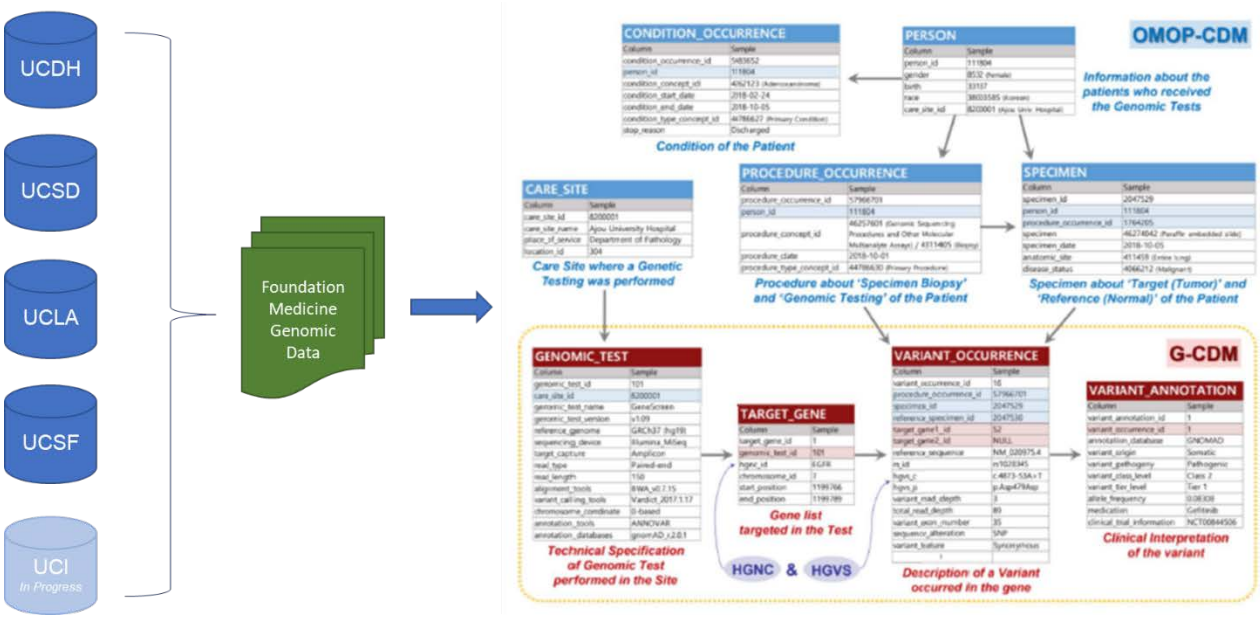

Figure 3. OMOP Genomic Common Data Model

Technology

While the core of the UC Health instance of cBioPortal (cbp.ucdmc.ucdavis.edu) is powered by the open source cbioportal.org software, the genomics report ingestion pipeline, flat file pre- and post-processing, and the data sharing workflow are based on UC Davis Health developed set of automation tools. The UC Davis tooling leverages many open source projects and the shared informatics development team including Python (python.org), Pandas (pandas.pydata.org), and Docker (docker.com), as to make use of best of class infrastructure software while ensuring domain specific code quality, extensibility, and maintainability. In addition, we launched a novel cross-campus collaborative development and project coordination environment through a health and bioportal-oriented software git repository (gitlab.com/uc-health-cbp/cbp).

This project, its grounding partnership with UC Davis IT and cancer informat research, its impact across multiple UC sites and cancer centers, and the secondary products are a testament to the capabilities of the UC to share expertise and resources in novel and response ways. The timeline of the project is described below.

Timeframe