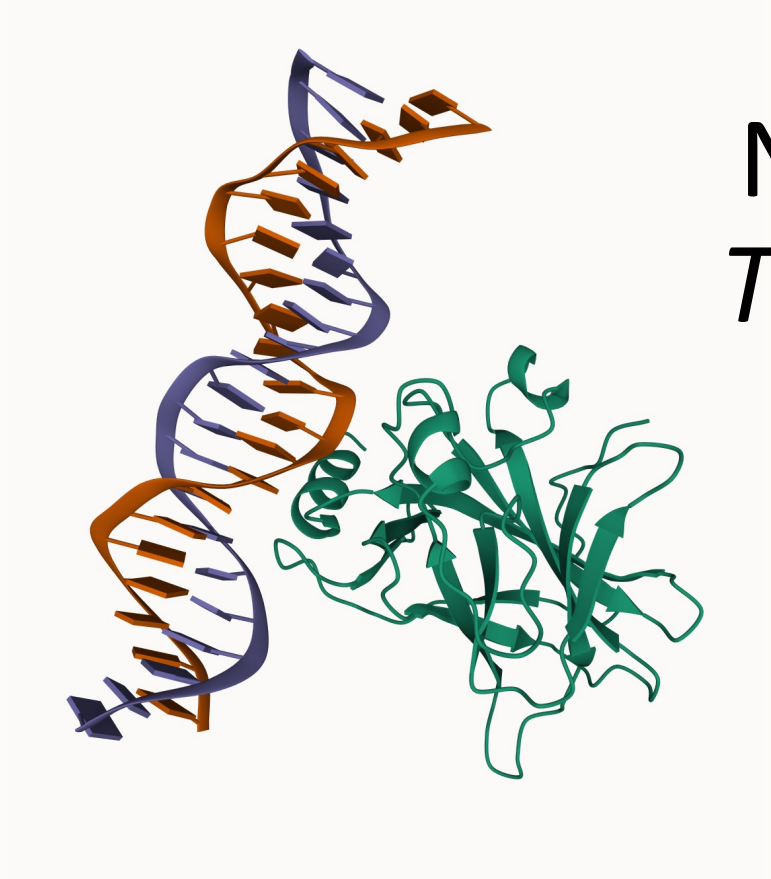
Breakout Session 1: Track B

NCI CRDC Cloud Transfer of TP53 Website and Database

Mr. William Longabaugh
Senior Software Engineer, Institute for Systems Biology



NCI CRDC Cloud Transfer of *TP53* Website and Database

William Longabaugh

Senior Software Engineer, Institute for Systems Biology

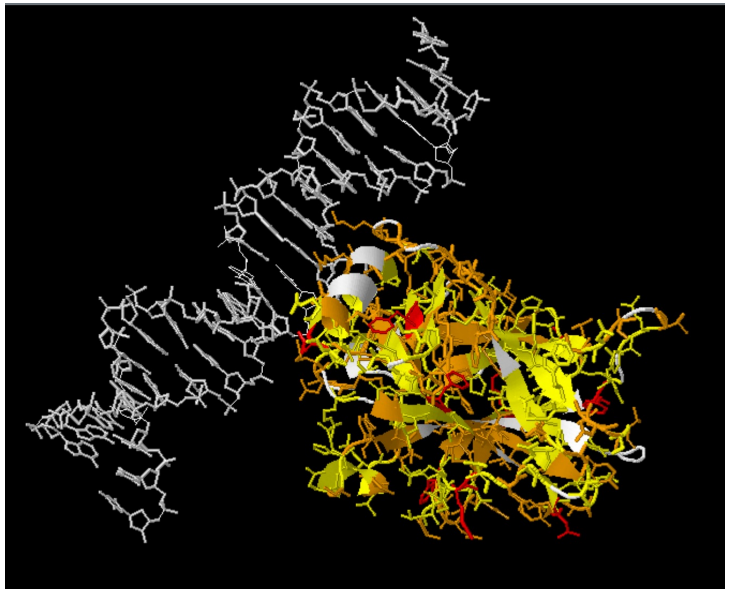
Jan 17 2024

Funding

- We received funds from *"FY2021 Request for ODSS Funds to Catalyze Migration to and Usage of the Cloud via the STRIDES Initiative (HVD 21)"*
- Google cloud credits were provided to us to support cloud operations underlying our migration of the IARC WHO TP53 database (now retired) to become part of the ISB-CGC Cloud Resource, a component of the Cancer Research Data Commons (CRDC)
- Additionally, the credits covered cloud operation costs of our development, test, and production tier Google cloud projects until September 2023

Thank you to the Office of Data Science Strategy

The *TP53* Database: Aim and Scope



Database compiles *TP53* variant data from 1989

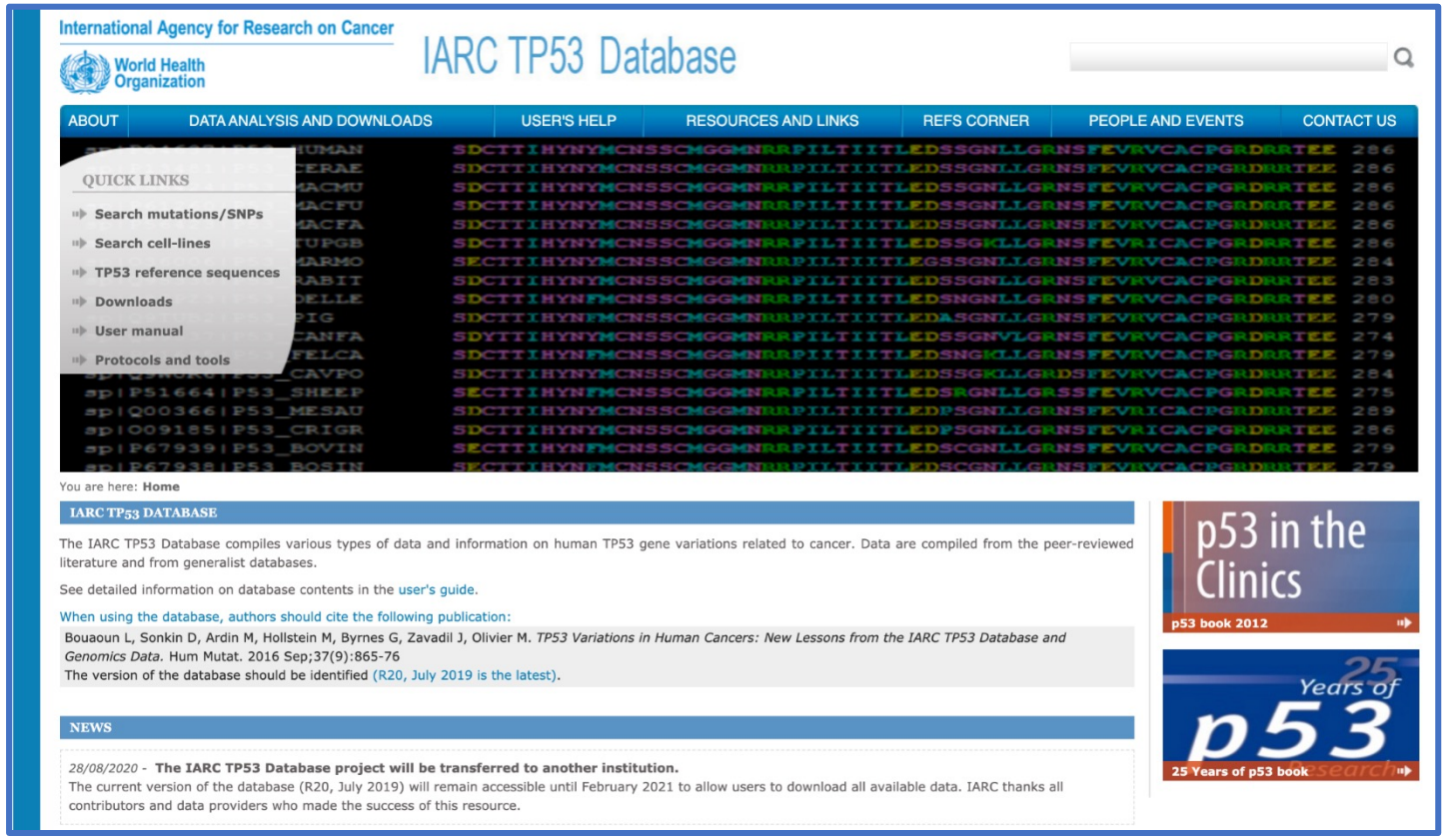
Currently holds information on 24,547 *TP53* variants

Database includes:

- *TP53* **functional** and **structural** data
- *TP53* **tumor** variants in sporadic cancer
- *TP53* **germline** variants in cancer patients, families with cancers
- *TP53* gene status in human **cell-lines**
- **Mouse models** with engineered *p53*
- **Experimentally-induced** *TP53* variants

Holds information on *TP53* variants for a broad range of scientists and clinicians who work in different research areas

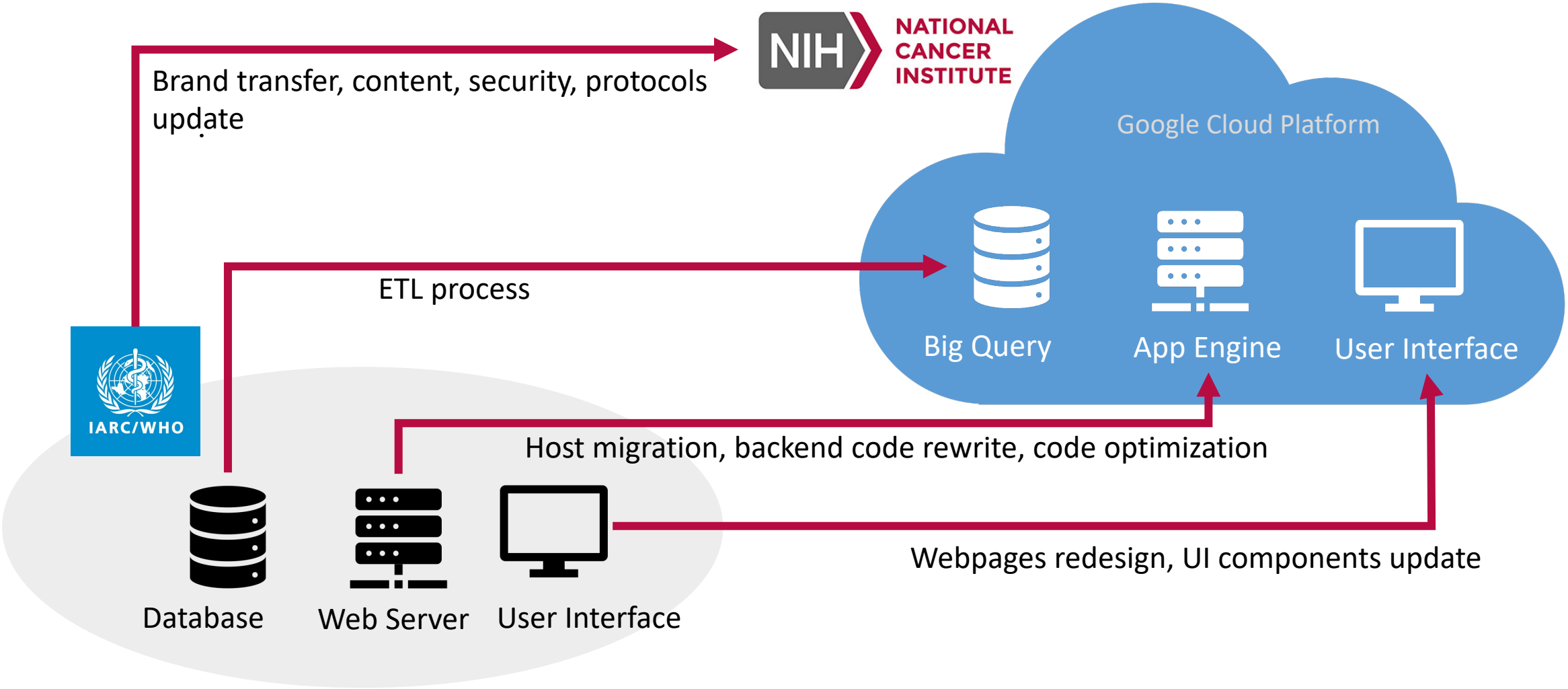
IARC *TP53* Database



IARC TP53 Database Website in 2020

The original ***TP53* database** was initiated in 1991, further developed and maintained by WHO's **International Agency for Research on Cancer** until 2021.

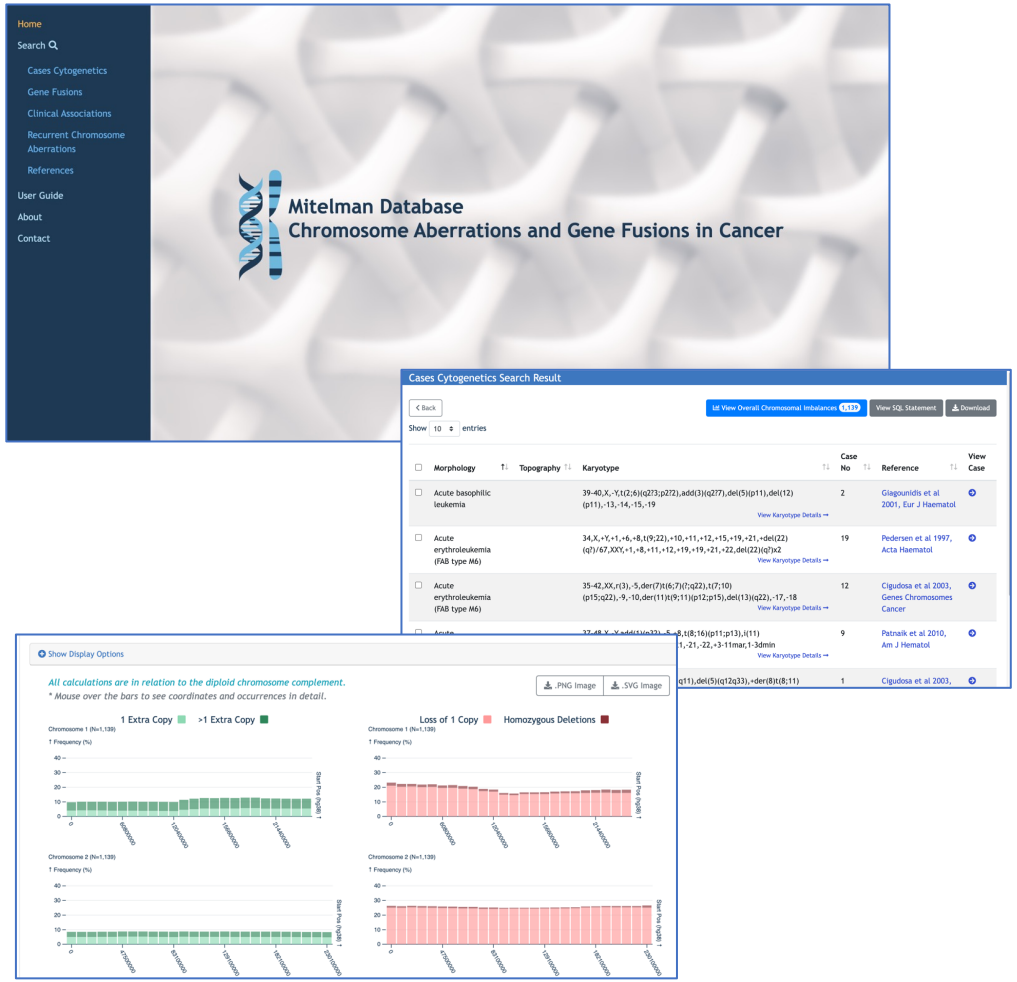
Transfer of Website and Database into the Cloud



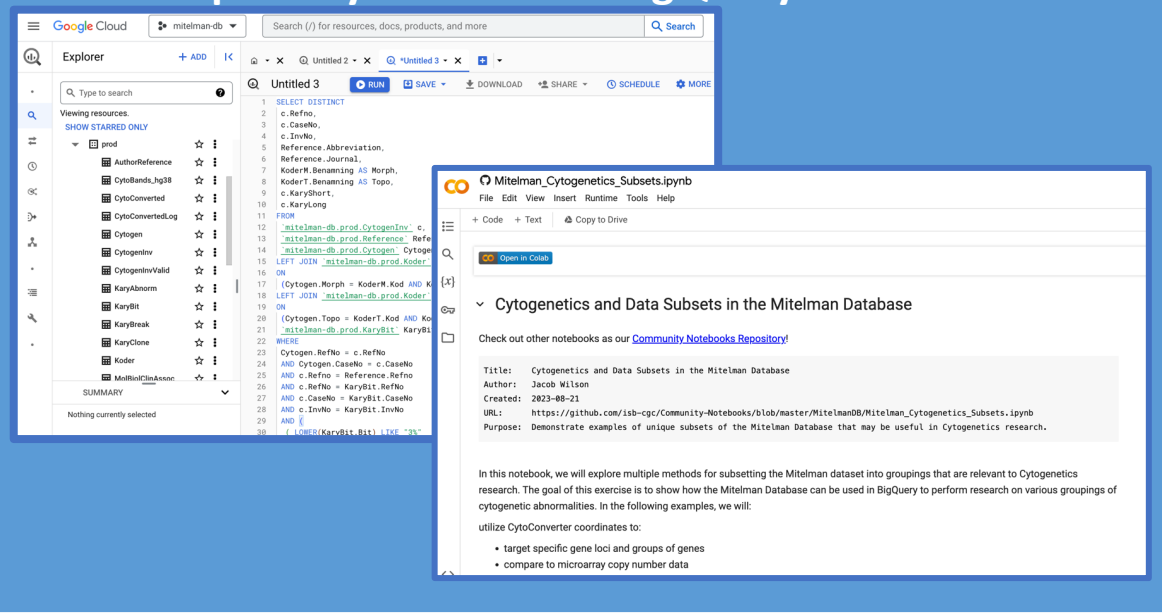
Transfer of Website and Database into the Clouds: Mitelman Database

- The **Mitelman Database** was part of CGAP (Cancer Genome Anatomy Project, NCI)
- That website was retired on 2019
- ISB-CGC was responsible for transferring all web components to the Google Cloud Platform
- The application has been further developed for advanced queries and additional features

https://mitelmandatabase.isb-cgc.org

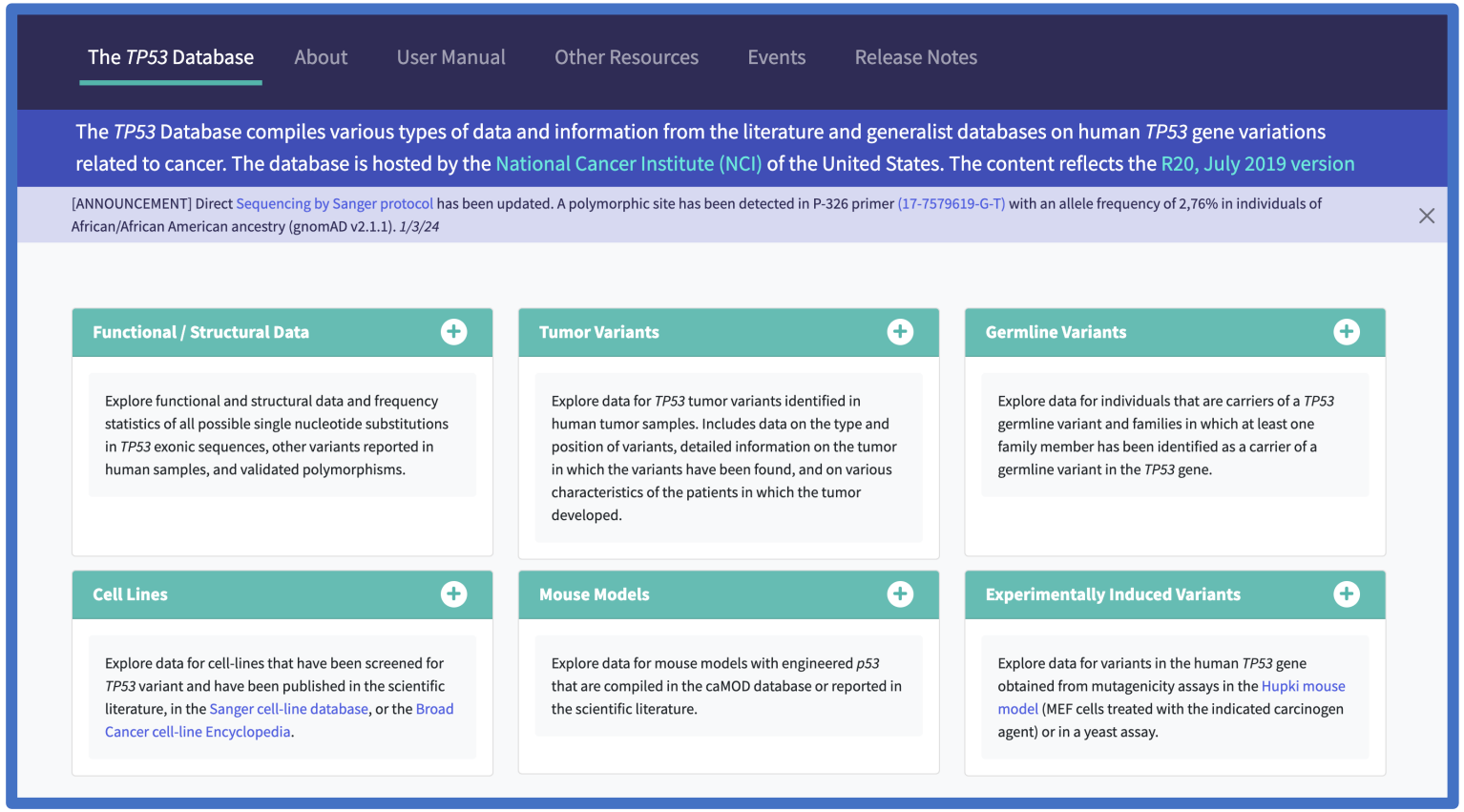


All data is **publicly available in BigQuery**.



Transfer of Website and Database into the Clouds: The *TP53* Database

https://tp53.isb-cgc.org

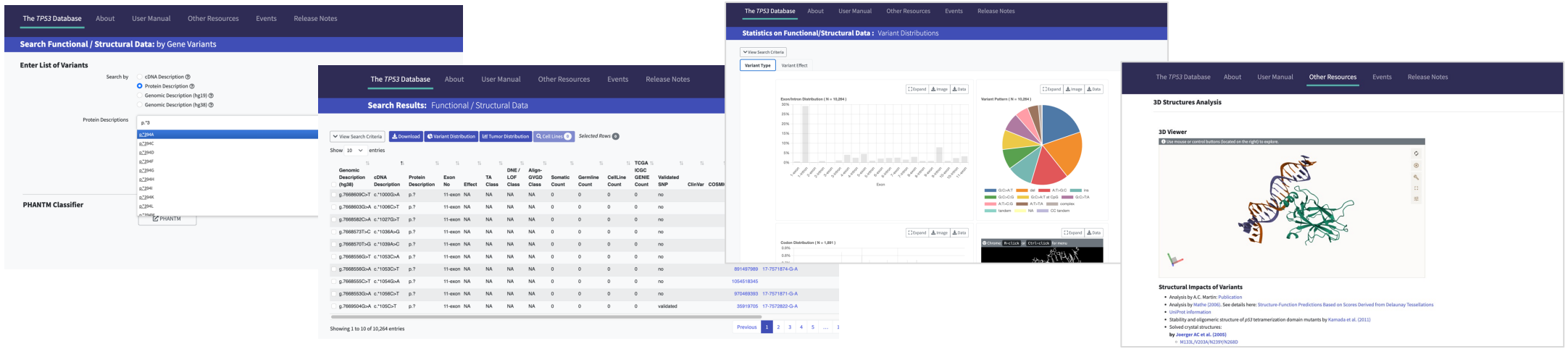


The TP53 Database of NCI was launched in 2021 with all of its web components operating under **Google Cloud Platform**.
All web queries are directly run in **BigQuery.**

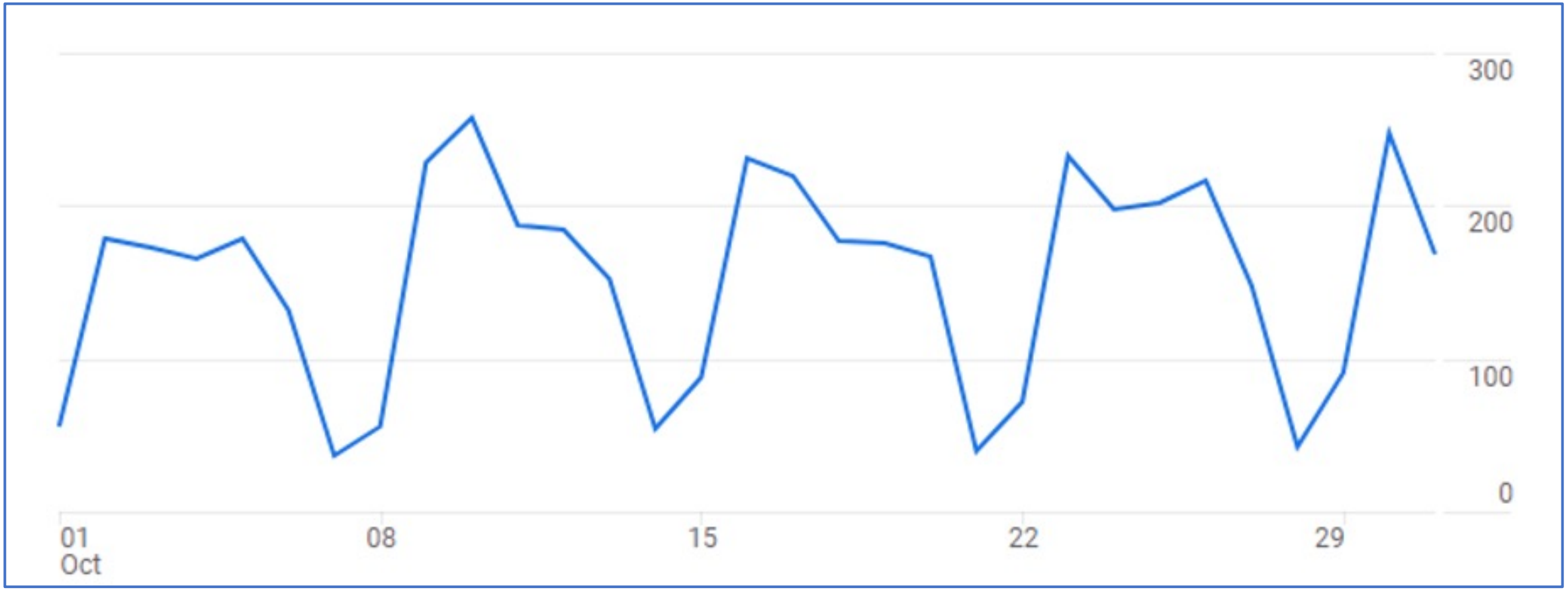
The *TP53* Database of NCI

Application is now

- Faster to search or run analyses
- Easier to navigate
- Secure
- Shares the same development, deployment, hosting, testing, and security framework with other ISB-CGC components



TP53 Database Usage

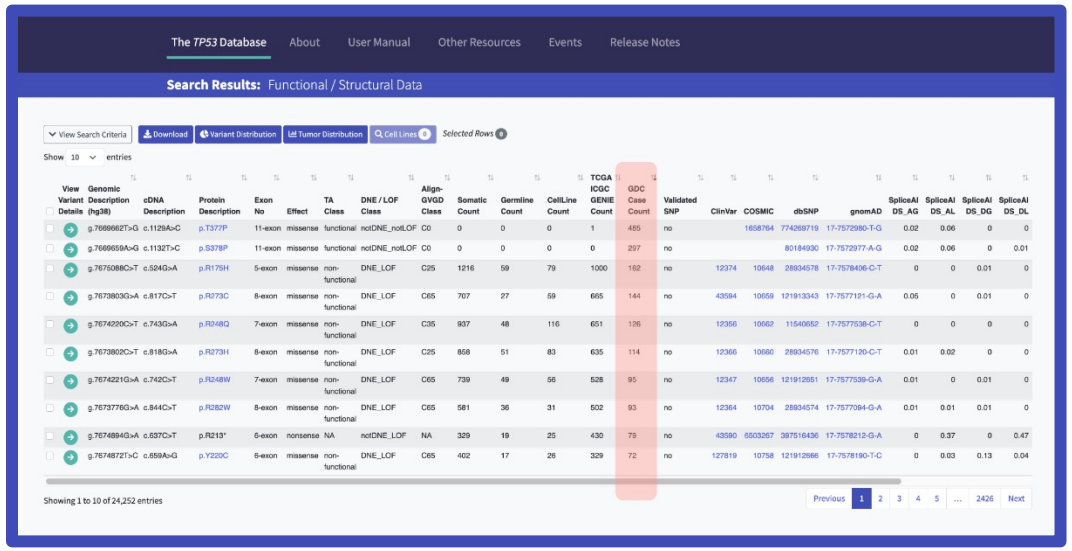


Future Development:
Easy Access to *TP53* dataset in BigQuery

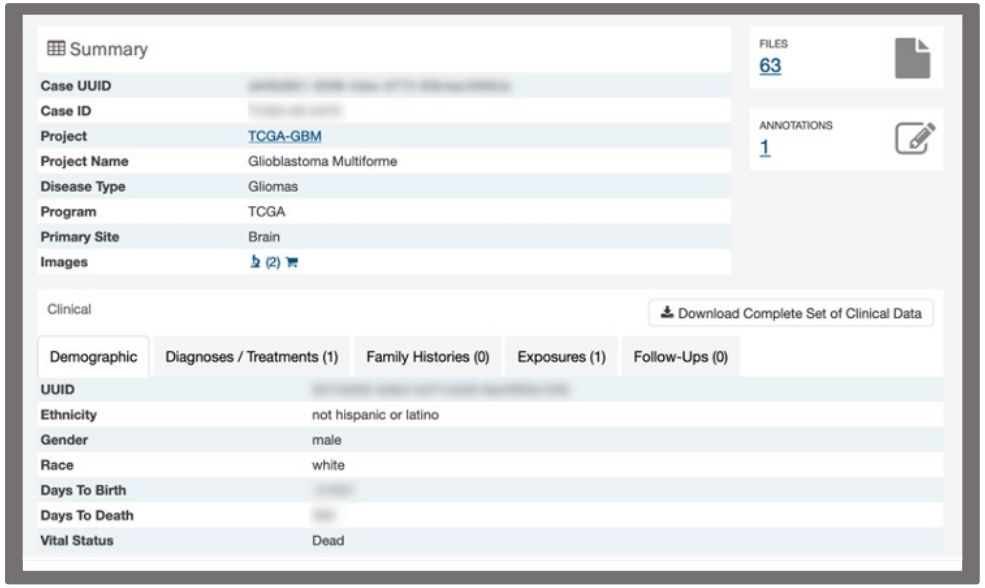
- The current BigQuery tables are not yet public (*cf.* Mitelman Database)
- The current data tables are too complex
  - The data is extracted from 70 tables, which have over 500 columns all together
  - Need to optimize the data by trimming fields that are not related to *TP53* variants
  - Need to remove extraneous columns that were never exposed
- Making the data in BigQuery public will make it easily accessible to any researcher or clinician
- The field of the data analysis can then be easily expanded with arbitrary queries

Future Development:
Linking *TP53* variant data with GDC case data

With TP53 now part of the CRDC, we can use the data to inform analyses of CRDC data



Prototype: TP53 variant search results with GDC case info



Genomic Data Common case page

ISB-CGC

ISB

GENERAL DYNAMICS
Information Technology

Elaine Lee
William Longabaugh
Boris Aguilar
Lauren Hagen
Lauren Wolfe
Mi Tian
Suzanne Paquette
Ilya Shmulevich

David Pot
Danna Huffman
Deena Bleich
Fabian Seidl
Jacob Wilson
Poojitha Gundluru
Prema Venkatesan
Owais Shahzada

DCEG

Division of Cancer Epidemiology & Genetics at the National Cancer Institute

Kelvin de Andrade
Sharon Savage

Original Team and IARC

Monica Hollstein
Curt C. Harris
Pierre Hainaut
Magali Olivier
Lucile Alteyrac
Jiri Zavadil

Plus…

Elise Tookmanian, Chimene Kesserwan, James Manfredi, Jessica Hatton, Jennifer Loukissas, Lei Zhou, Megan Frone, Christian Kratz, David Malkin, Pierre Hainaut

https://tp53.isb-cgc.org/