# Data Science Initiatives at the NCI

Norman E. Sharpless, M.D.

Director

*Computational Approaches for Cancer Workshop*
*November 13, 2020*

@NCIDirector
@TheNCI

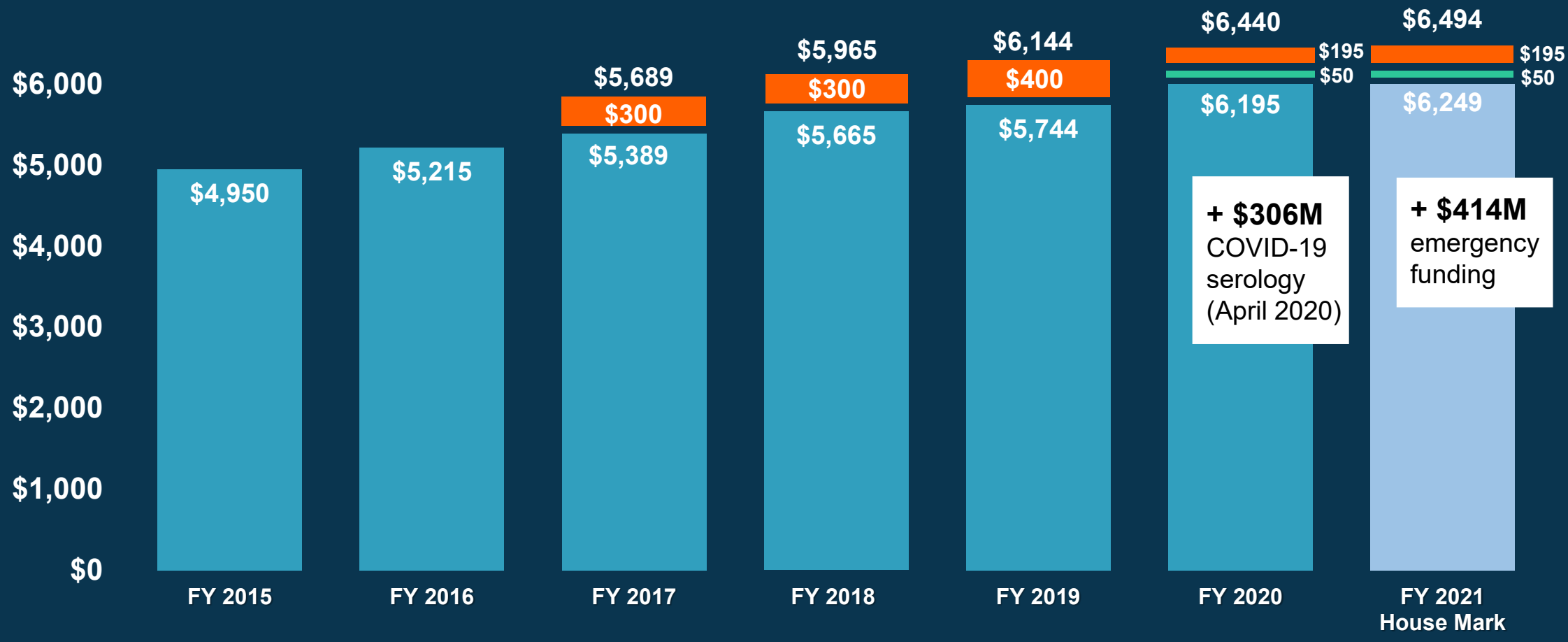
NIH NATIONAL CANCER INSTITUTE

# Today

- NCI Updates
- COVID-19 Research at NCI
- Select NCI Data Science Efforts
- A Look Ahead

# NCI Appropriations
## FY 2015 – 2020 (in millions)

21st Century Cures Act - orange
Childhood Cancer Initiative - green



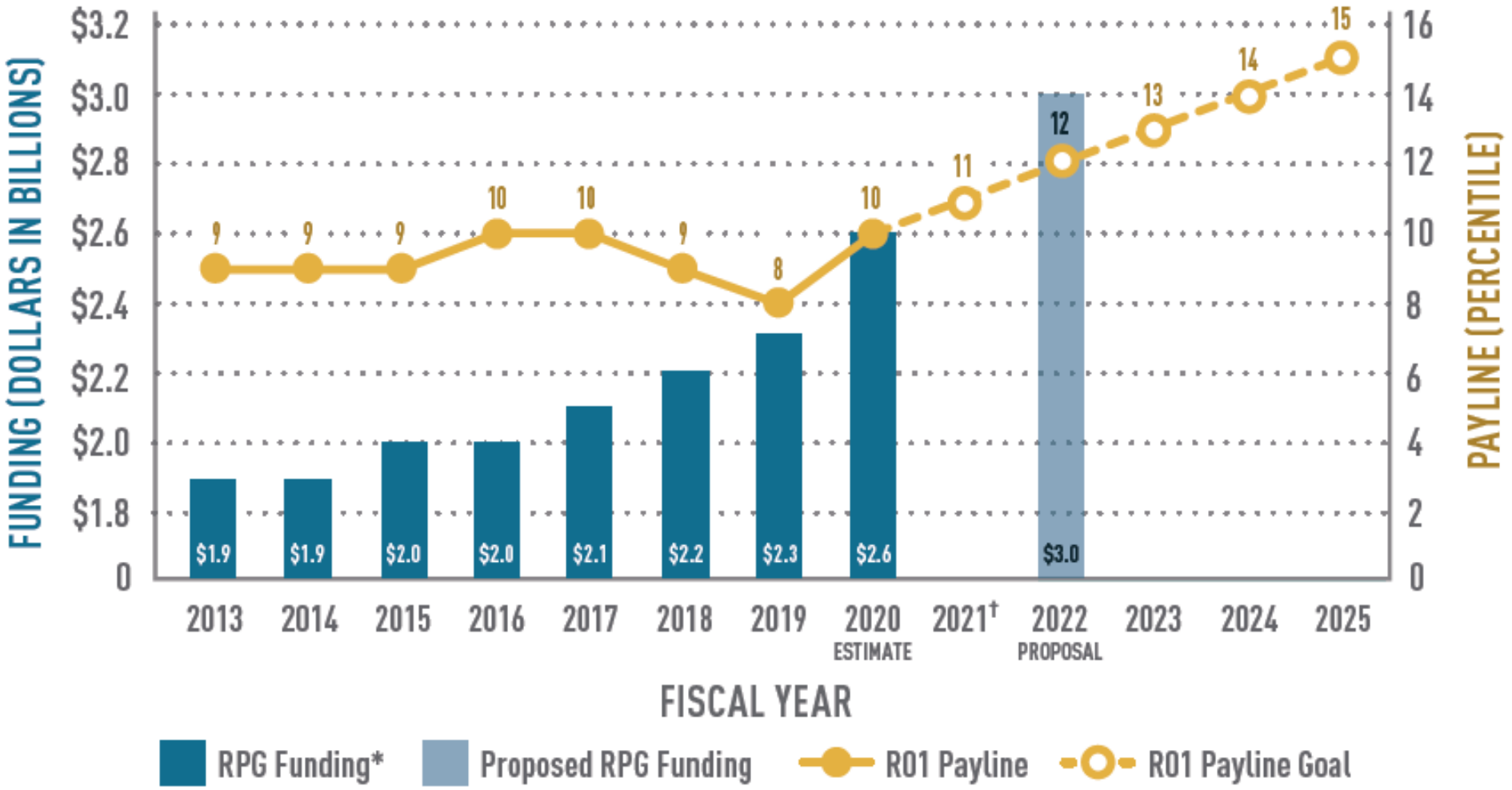

NIH NATIONAL CANCER INSTITUTE

# ANNUAL PLAN & BUDGET PROPOSAL

## Fiscal Year 2022

cancer.gov/research/annual-plan

# NCI Research Project Grants (RPG) Funding and R01 Paylines




* RPG funding levels exclude small business grant set-asides.
† FY 2021 appropriations not yet finalized.

# NCI Response to COVID-19

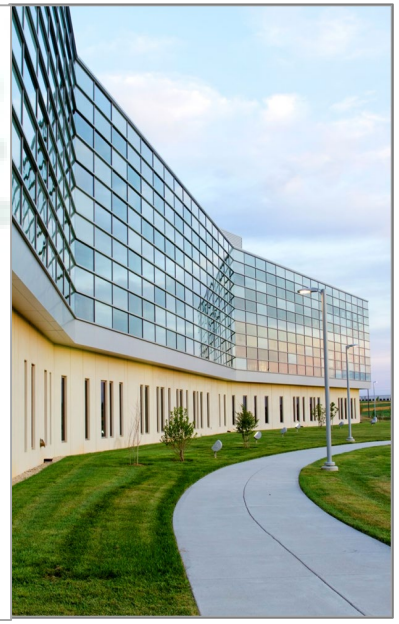
- SARS-CoV-2 serology research
- NCI COVID-19 in Cancer Patients Study (NCCAPS)
- Guidance and special procedures for cancer clinical trials
- Flexibilities and opportunities for grantees
- Genomic studies of COVID-19 outcomes

***cancer.gov/coronavirus-researchers***

# Why NCI?

Frederick National Laboratory
for Cancer Research
sponsored by the National Cancer Institute

The Frederick National Laboratory for Cancer Research and COVID-19 serology testing



134 STAT. 620 PUBLIC LAW 116–139—APR. 24, 2020

Public Law 116–139
116th Congress

An Act

Apr. 24, 2020
[H.R. 266]

Making appropriations for the Department of the Interior, environment, and related agencies for the fiscal year ending September 30, 2019, and for other purposes.

*Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,*

Paycheck Protection Program and Health Care Enhancement Act.
15 USC 9001 note.

**SECTION 1. SHORT TITLE.**

This Act may be cited as the "Paycheck Protection Program and Health Care Enhancement Act".

**SEC. 2. TABLE OF CONTENTS.**

The table of contents for this Act is as follows:

Sec. 1. Short title.
Sec. 2. Table of contents.
Sec. 3. References.

# SeroHub: SARS-CoV-2 Serology Study Dashboard

- Effort began in early June at the request of HHS, CDC, and NIAID to develop a data warehouse & dashboard for tracking SARS-CoV-2 seroprevalence and other US-based serology studies
- Builds on FNL expertise in warehouse and dashboard development for other NCI resources

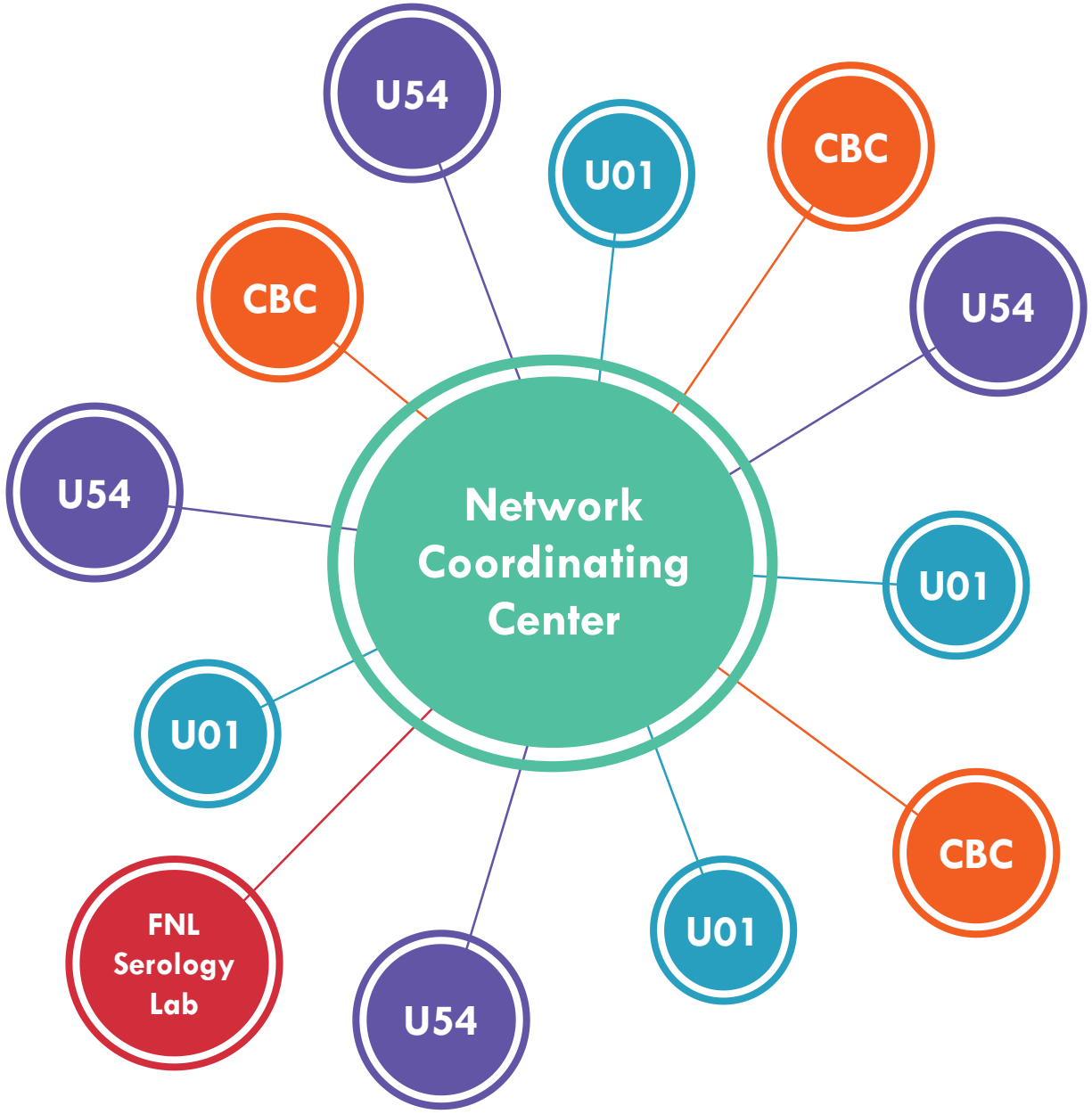
## Key features

- Publicly accessible data warehouse to systematically document and track SARS-Cov-2 serology studies and associated test results
- Tracking dashboard to visualize SARS-Cov-2 serology data and present results overall and by key strata

# SeroNet Serological Sciences Network

- Grants (U54s and U01s)
- Contracts (CBCs)
- National Laboratory

***Launched October 8***



# Geographical Distribution of SeroNet Sites


U54 Site

U01 Site

CBC Site

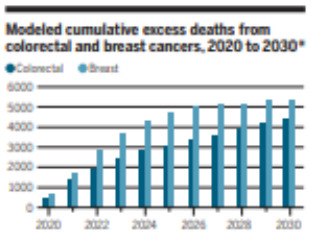
# Science 19 June 2020

## COVID-19 and cancer

**Norman E. Sharpless**
is director of the U.S. National Cancer Institute, Bethesda, MD, USA. norman.sharpless@nih.gov

With the spread of coronavirus disease 2019 (COVID-19), countries and states have instituted lockdowns. These decisions have been difficult and are sometimes described as benefiting the public health at the expense of the economy. Fear of contracting the coronavirus in health care settings has dissuaded people from screening, diagnosis, and treatment for non-COVID-19 diseases. The consequences for cancer outcomes, for example, could be substantial. What can be done to minimize this effect?

Cancer is a complex set of diseases whose prognoses are influenced by the timing of diagnosis and intervention. In general, the earlier one receives cancer treatment, the better the results. There already has been a steep drop in cancer diagnoses in the United States since the start of the pandemic, but there is no reason to believe the actual incidence of cancer has dropped. Cancers being missed now will still come to light eventually, but at a later stage ("upstaging") and with worse prognoses. At many hospitals, so-called "elective" cancer treatments and surgeries have been deprioritized to preserve clinical capacity for COVID-19 patients. For example, some patients are receiving less intense chemotherapy and/or radiotherapy, and in other cases, patients' operations to remove a newly detected tumor are being delayed. There can be no doubt that the COVID-19 pandemic is causing delayed diagnosis and suboptimal care for people with cancer.




What will be the likely impact of the pandemic on cancer mortality in the United States? Modeling the effect of COVID-19 on cancer screening and treatment for breast and colorectal cancer (which together account for about one-sixth of all cancer deaths) over the next decade suggests almost 10,000 excess deaths from breast and colorectal cancer deaths; that is, a ~1% increase in deaths from these tumor types during a period when we would expect to see almost 1,000,000 deaths from these two diseases types.* The number of excess deaths per year would peak in the next year or two. This analysis is conservative, as it does not consider other cancer types, it does not account for the additional nonlethal morbidity from upstaging, and it assumes a moderate disruption in care that completely resolves after 6 months. It also does not account for regional variations in the response to the pandemic, and these effects may be less severe in parts of the country with shorter or less severe lockdowns.

Beyond clinical care, the COVID-19 pandemic has caused an unprecedented disruption throughout the cancer research community, shuttering many labs and slowing down cancer clinical trial operations. Many scientists and clinicians are pivoting their cancer research activities to study the impact of SARS-CoV-2 on cancer. The scientific community must ensure that this pause is only temporary, because trials are the only way to make progress in developing new therapies for cancer. Given the long timeline between basic cancer research and changes to cancer care, the effects of pausing research today may lead to slowdowns in cancer progress for many years to come.

Collective action by the clinical and research communities and by governmental agencies can mitigate this potentially substantial impact. The U.S. National Cancer Institute (NCI), for example, has started to address this challenge (see www.cancer.gov). The NCI has worked with the U.S. Food and Drug Administration to increase flexibility and support for clinical trials during the pandemic. For example, allowances have been made to accept "remote" informed consent, and other protocol deviations. In addition, the NCI has announced several new clinical trials and funding opportunities aimed at addressing the relationship between COVID-19 and cancer. Of particular note is the NCI COVID-19 in Cancer Patients Study, a prospective longitudinal study that will collect blood samples, imaging, and other data to understand how COVID-19 affects cancer patients.

Clearly, postponing procedures and deferring care as a result of the pandemic was prudent at one time, but the spread, duration, and future peaks of COVID-19 remain unclear. However, ignoring life-threatening non-COVID-19 conditions such as cancer for too long may turn one public health crisis into many others. Let's avoid that outcome.

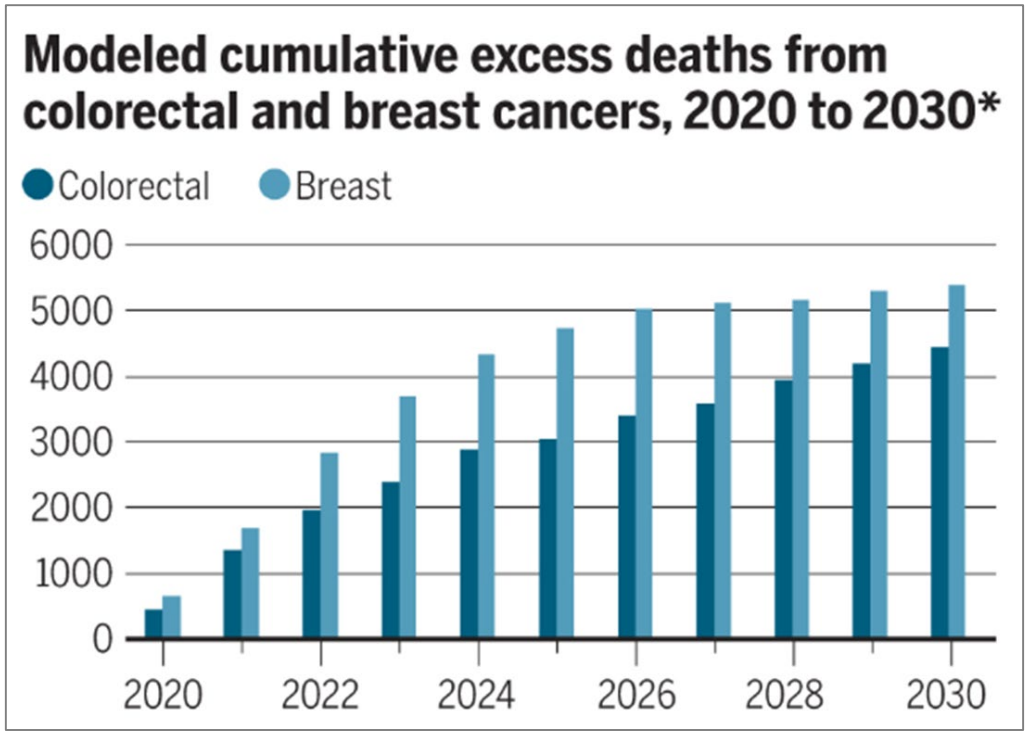
–Norman E. Sharpless

*See supplementary materials (science.sciencemag.org/content/368/6497/1290/suppl/DC1).

10.1126/science.abd3377








The Washington Post

By Laurie McGinley

June 18, 2020 at 7:30 p.m. EDT

...chief warns delays in ... likely to result in ...xtra deaths in coming years

STAT

**Ignoring cancer care now may trade one public health crisis — Covid-19 — for another, NCI chief warns**

By ELIZABETH COONEY @cooney_liz / JUNE 19, 2020

# Supporting Cancer Research during the Pandemic

- Impacts on cancer care delivery and outcomes
- Studying COVID-19 in people with cancer
- Adjusting clinical trial protocols
- Cushioning the blow for cancer researchers
- Outlook for research funding beyond the pandemic

# Flexibilities to support grantees during the pandemic

- Extending deadlines for applications
- Allowing institutions to use NCI grant funds to maintain salaries and stipends
- Extending project timelines and reporting requirements
- Extending eligibility periods for early-stage investigators and trainees
- Carryover for institutional training grants (T35, T32, K12) with prior approval

# Key Focus Areas

## BASIC SCIENCE

Reaffirm our commitment to basic science to drive novel approaches and technologies

## WORKFORCE DEVELOPMENT

Support the cancer research enterprise by focusing on the workforce of cancer investigators

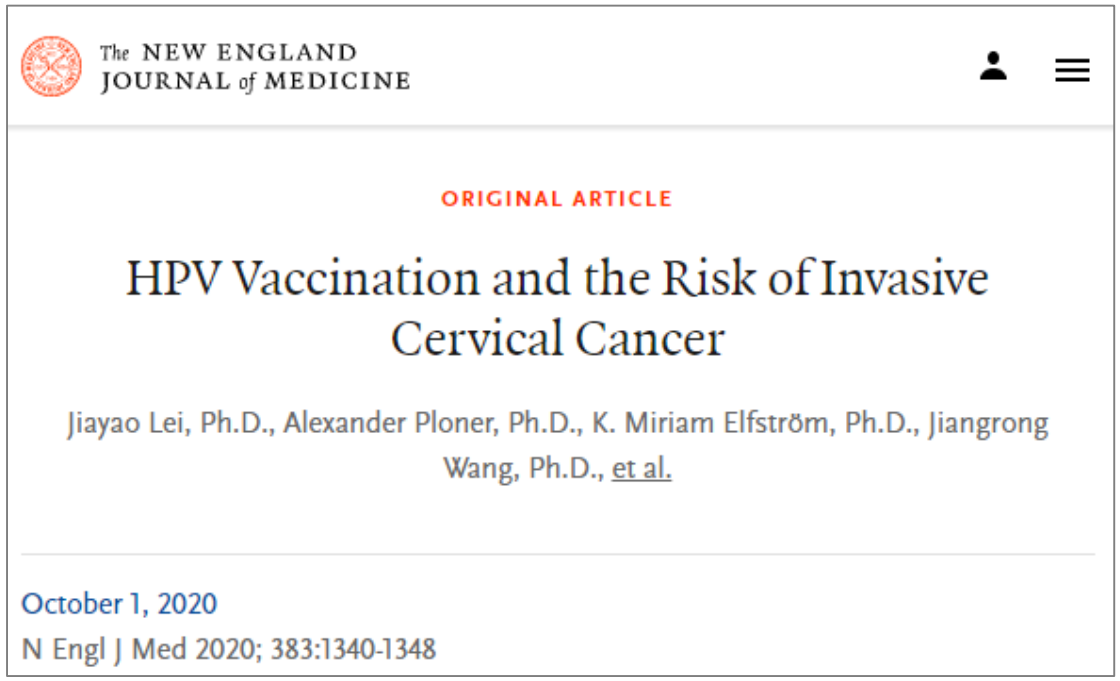
## BIG DATA

Increase data aggregation and interpretation to speed our work across the cancer enterprise

## CLINICAL TRIALS

Fully realize the power of clinical trials through innovative design, administration, and analyses

# Real World Evidence



The NEW ENGLAND JOURNAL of MEDICINE

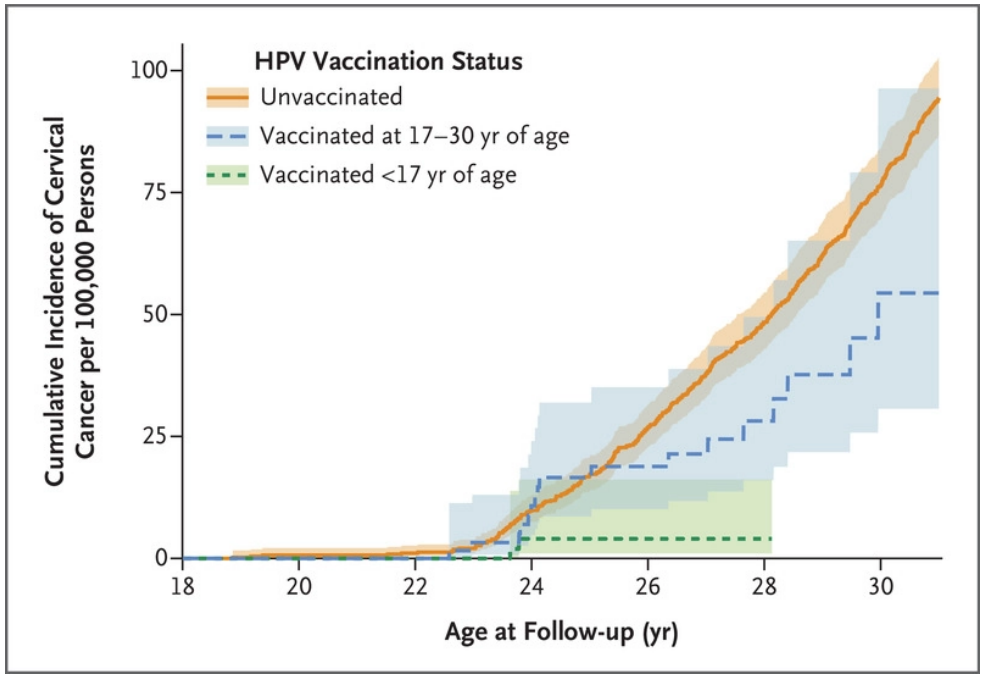
ORIGINAL ARTICLE

## HPV Vaccination and the Risk of Invasive Cervical Cancer

Jiayao Lei, Ph.D., Alexander Ploner, Ph.D., K. Miriam Elfström, Ph.D., Jiangrong Wang, Ph.D., et al.

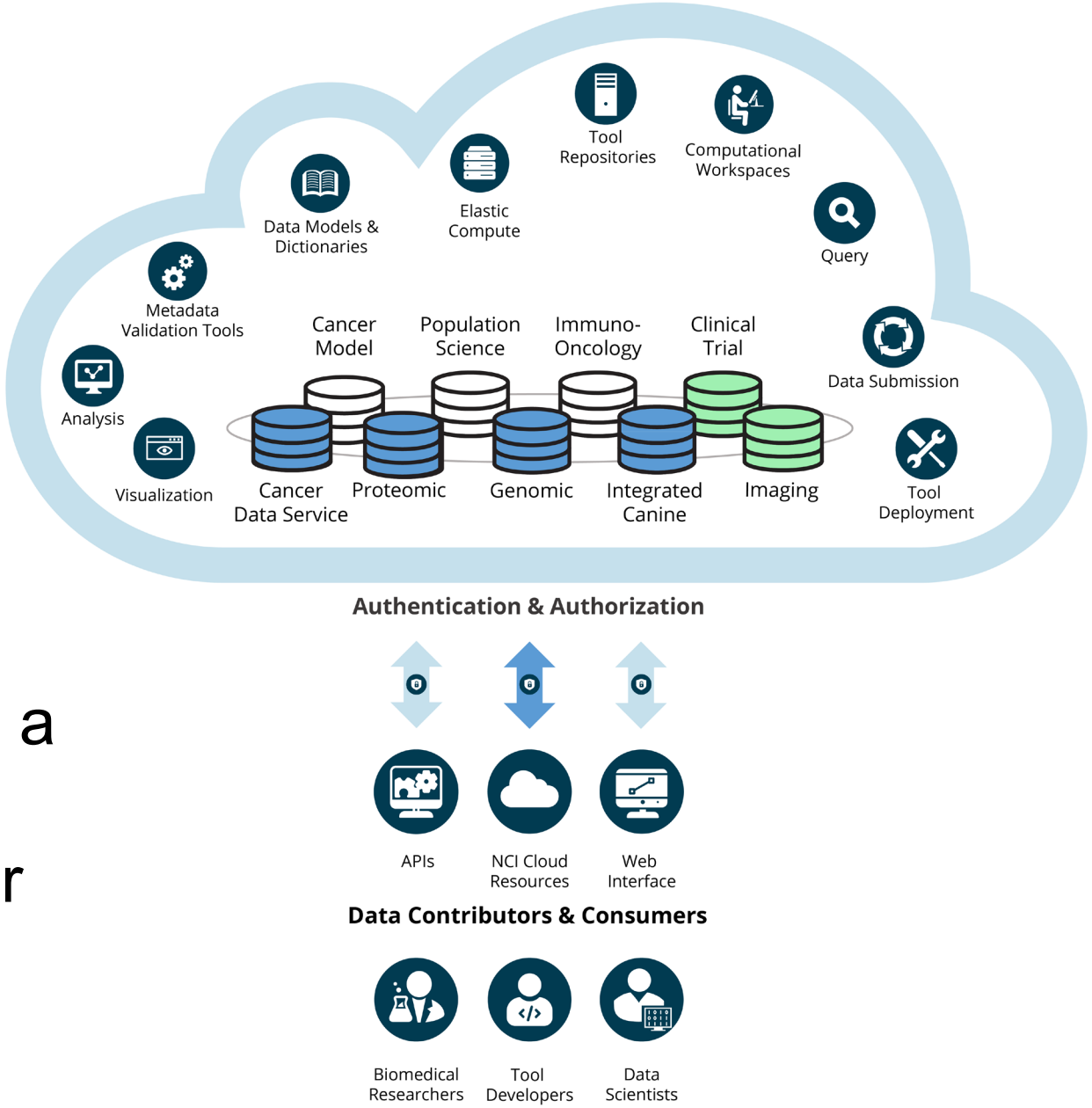
October 1, 2020

N Engl J Med 2020; 383:1340-1348



# NCI Cancer Research Data Commons

Cloud-based **data science infrastructure** that provides secure access to a **large, comprehensive, and expanding** collection of cancer research data



# Datasets in the CRDC

GECCO

FOUNDATION MEDICINE


CBTTC
Children's Brain Tumor Tissue Consortium

MMRF MULTIPLE MYELOMA Research Foundation

CPTAC
National Cancer Institute
at the National Institutes of Health


CGCI
CANCER GENOME CHARACTERIZATION INITIATIVE

TARGET

ICPC
INTERNATIONAL CANCER PROTEOGENOME CONSORTIUM

PROJECTGENIE
Genomics Evidence Neoplasia Information Exchange

And many more…

# Clinical Trials Data

## NCTN/NCORP Data Archive - Overview

**A centralized, controlled-access database that contains individual, patient-level data for all phase 3 & phase 2/3 clinical trials with**:

- Primary Results publications available on or after January 1, 2015 and selected non-primary publications available on or after April 1, 2018
- Other data (e.g., phase 2 trials or legacy phase 2 & phase 3 trials published before 2015, etc.) are included on a case-by-case basis.
- Submissions are due within 6 months of publication date with all data elements presented in the publication included
- All data available within the Archive are also accessible via Project Data Sphere (PDS) within 24 hours

**Complements other NCI data sharing activities that focus on genomic or imaging data as well as sharing of biospecimens from NCTN trials for approved correlative science proposals (i.e., Navigator).**

# **Current NCTN Navigator Biospecimen Inventory** (Oct 2020)

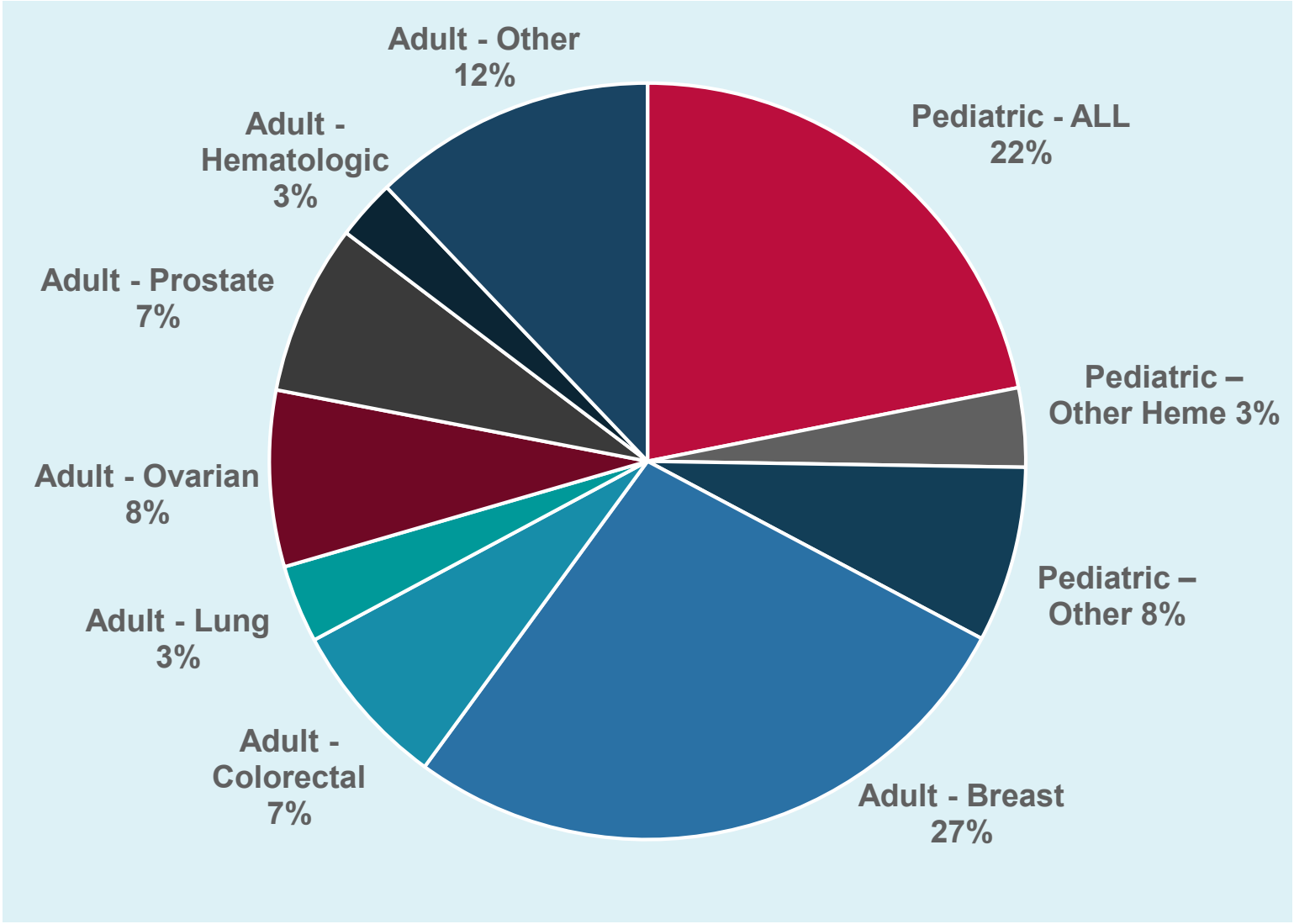
from completed Phase 3 trials; available for approved research proposals

**1999211**
Specimens

**146562**
Patients

**231**
Trials

**navigator.ctsu.org**




NIH NATIONAL CANCER INSTITUTE

# NCTN/NCORP Data Archive – *New Data Integration*

- For RTOG-0617*, a data integration is in place with The Cancer Imaging Archive, NCI's official imaging repository for NCTN trials, that enables users to request both clinical and imaging data.
- Images from 3 additional trials are being de-identified.


cancerimagingarchive.net

**High-Dose or Standard-Dose Radiation Therapy and Chemotherapy With or Without Cetuximab in Treating Patients With Newly Diagnosed Stage III Non-Small Cell Lung Cancer That Cannot Be Removed by Surgery*

# Molecular Profiling to Predict Response and Treatment (MP2PRT)

**Challenge**

Gathering information from clinical trials to inform new trial design takes too long, particularly in rare cancer types or in specific populations

Through retrospective analysis of archival specimens collected through clinical trials, **develop predictive models that will identify what patients might benefit from standard therapy or who might require additional or novel interventions**

# Molecular Profiling to Predict Response and Treatment (MP2PRT)

| | | | |
|---|---|---|---|
| Children's Oncology Group (COG) | **Acute Lymphoblastic Leukemia (ALL)** | Comprehensive Genomic Profiling to Identify Alterations Associated with Relapse for NCI Standard-Risk B-lineage ALL and NCI High-Risk B-lineage ALL with Favorable Genetic Features | **Mignon L. Loh, M.D.** Benioff Children's Hospital, University of California SF, CA |
| Children's Oncology Group (COG) | **Wilms Tumor** | Identification of Genetic Changes Associated with Relapse and/or Adaptive Resistance in Patients Registered with Favorable Histology Wilms Tumor on AREN03B2 (Renal Tumors Classification, Biology, and Banking Study) | **Elizabeth Perlman, M.D.** Anne & Robert H. Lurie Children's Hospital of Chicago, IL |
| NRG Oncology Group | **Cervical Cancer** | Genomic and molecular characterization of biomarkers associated with tumor angiogenesis, DNA repair, and immunologic tolerance among exceptional responders and long-term survivors in NRG /GOG protocol 240 | **Krishnasu Tewari, M.D.** UC Irvine Orange, CA |
| NRG Oncology Group | **Meningioma** | Identifying novel molecular markers of response to radiotherapy in meningiomas using samples from the RTOG-0539 | **Kenneth Aldape, M.D.** Center for Cancer Research, NCI |

# NCI Cloud Resources

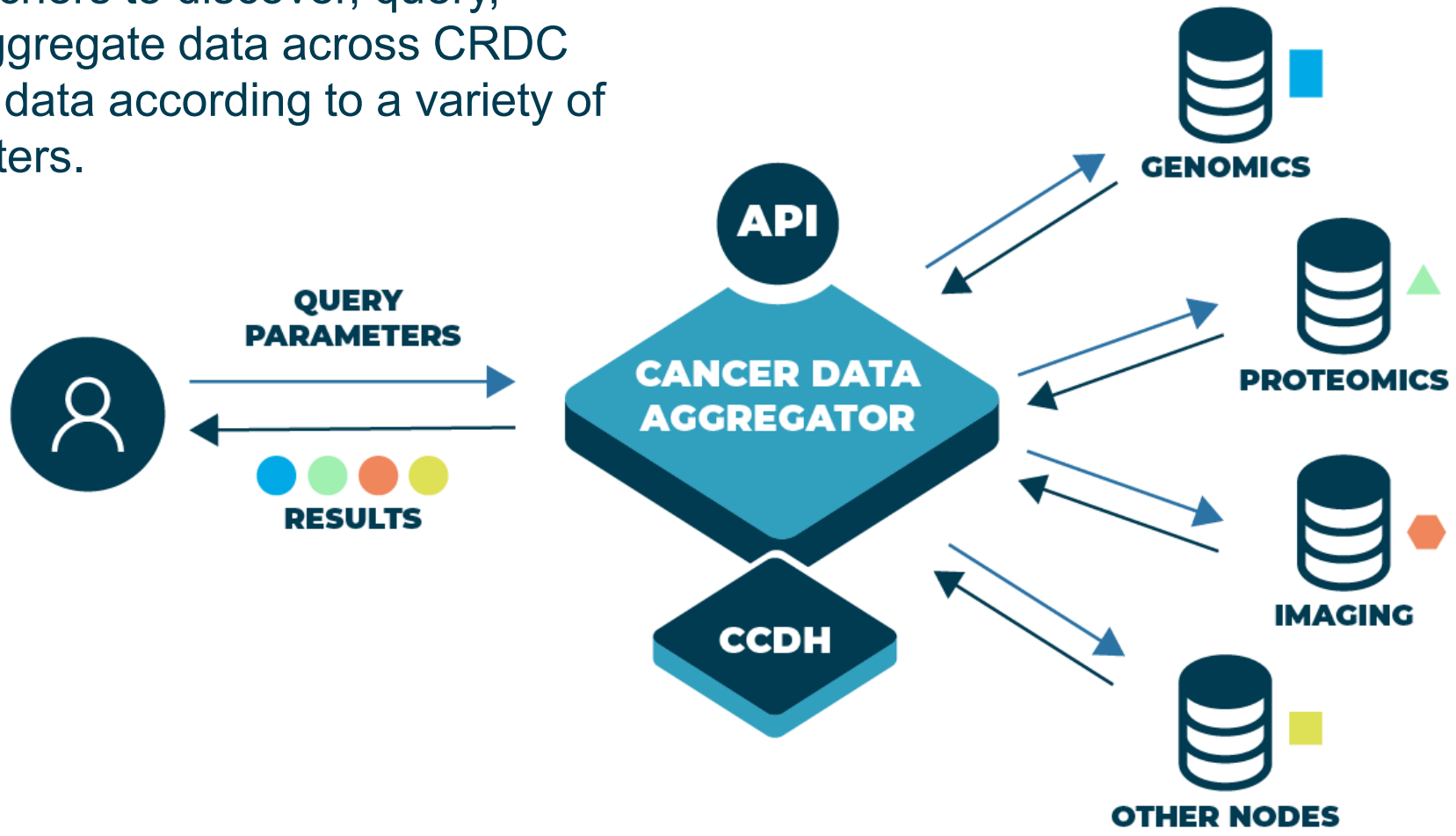
- The Institute for Systems Biology, Seven Bridges Genomics, and Broad Institute are referred to as **NCI Cloud Resources (CRs)**
- 3 CRs connect NCI data and compute in the cloud
- Access to data, workspaces, analysis tools, and pipelines
- Ability for researchers to bring their own data and tools
- Steady increase in active community of cancer researchers adopting the Cloud Resources for their data analysis
  - **Average of 3000 active users per month across the CRs**
  - **4 Million workflows run during 2020**


- 23+ Datasets
- Elastic Compute
- FISMA moderate



# Cancer Data Aggregator

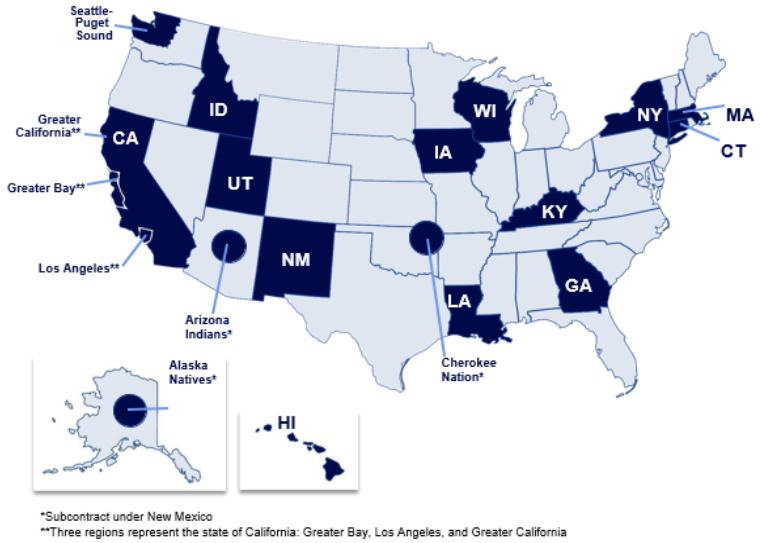
Enables researchers to discover, query, retrieve, and aggregate data across CRDC and NCI DCCs data according to a variety of search parameters.



# Surveillance, Epidemiology, and End Results (SEER) Program

- Recent RFP to expand the SEER program further (contract January 2021)
- With new registries−550,000 incident cases received annually
  - Approximately 85% of cases with real time electronic pathology (e-path) reporting
  - Facilitates rapid case identification supporting research
- All registries will be on a common data platform (SEER DMS) that permits
  - central linkages with external partners
  - facilitates scaling of new initiatives across all registries simultaneously

**16 population-based registries *now* covering 35% of the US population**




*Subcontract under New Mexico
**Three regions represent the state of California: Greater Bay, Los Angeles, and Greater California

# Partnerships and linkages to enhance SEER data

## Partnering to acquire source data

### Genomic/genetic testing

**GHI, Caris LS, Decipher, Castle Life Sciences, Myriad, Ambry,** etc.

### Claims data

**Unlimited Systems** - oncology claims processor ~15% of cases in SEER

Large insurers (**United Health Care**)

Exploring *All Payer All Claims* (6 SEER registries have state-wide APAC)

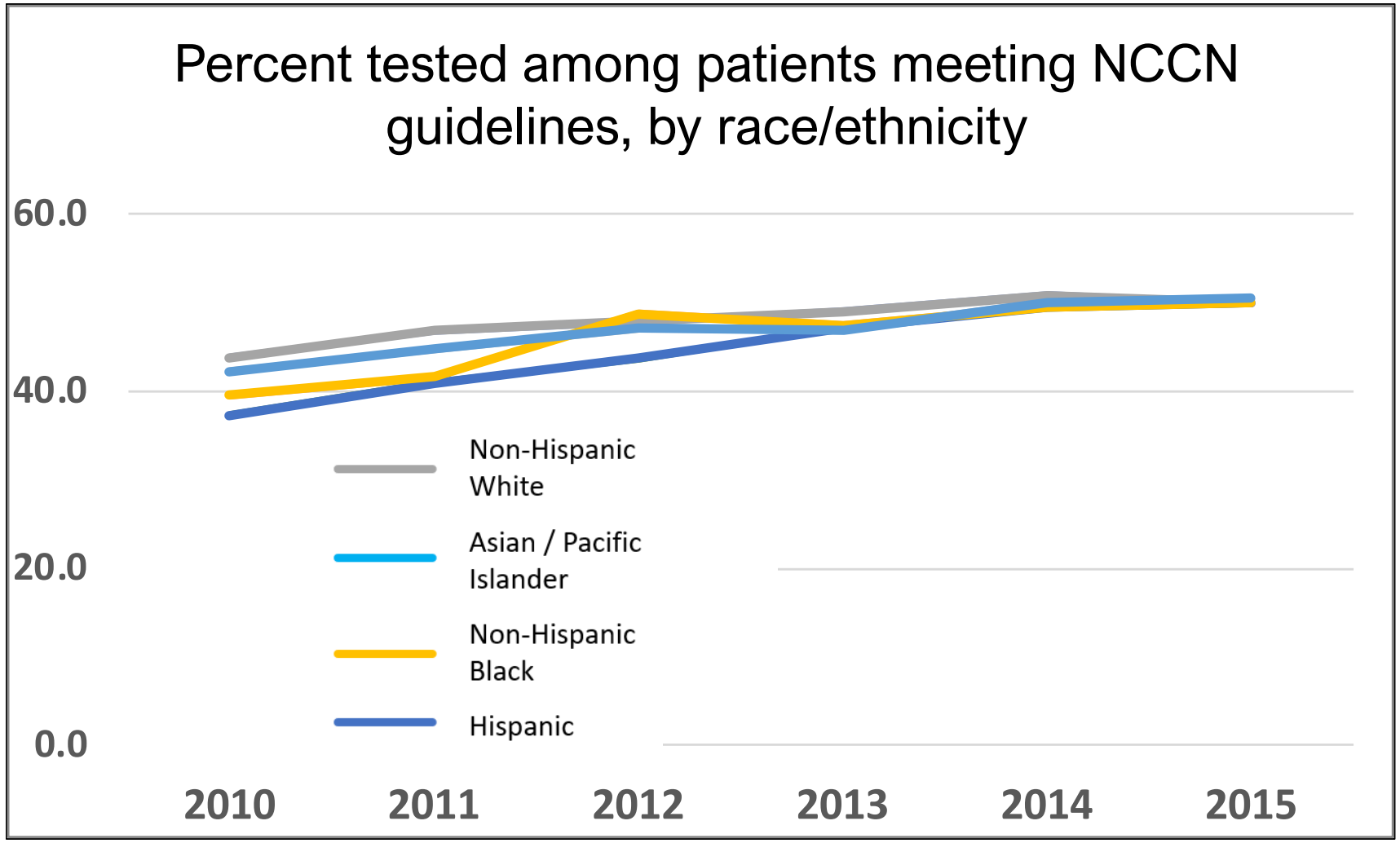
Pharmacy (**CVS and Walgreens/RiteAid**)

*Data received for all registries- 2014+*

## Partnerships with technology companies aggregating and using clinical data

- **CancerLinQ, Syapse, Tempus** (provider EMR data)
- **Varian/Elekta** (radiation oncology)
- **Ambra Health** (radiology)

# Example: Evaluating trends in standards of care- disparities in Oncotype DX testing rates




During the initial years (2010-2012), there was some evidence of differential testing by race and ethnicity. Recent data suggests disparities are disappearing.

# CCDI

BUILDING A COMMUNITY CENTERED AROUND CHILDHOOD CANCER CARE AND RESEARCH DATA

## DATA TYPES:

LEARN FROM EVERY CHILD

- CLINICAL
- TREATMENT
- OUTCOME
- MOLECULAR
- BIOSPECIMEN
- LONGITUDINAL
- POPULATION

## PROGRAM GOALS:

- DEVELOP NEW TREATMENTS
- IMPROVE EXISTING TREATMENTS
- IMPROVE OUTCOMES, QUALITY OF LIFE, AND SURVIVORSHIP

IMPROVING THE QUALITY, CONSISTENCY, ACCESS, AND ACCESSIBILITY OF DATA FROM EVERY CHILD WILL ALLOW US TO ACHIEVE OUR GOALS.

# National Childhood Cancer Registry
# Initial Registry Participation (70% of US childhood cancers)

## 7 NPCR Registries

Florida
Illinois
New Jersey
Ohio
Pennsylvania
Tennessee
Texas

2020 - submission of de-identified NAACCR data in NCCR

2021 - full registry submission with PII to DMS*Lite as a repository for linkages and submission to the NCCR

## SEER Registries

Georgia
Los Angeles
Greater CA
Greater Bay Area
Iowa
Connecticut
Kentucky
Louisiana
Seattle
Idaho
New York
Massachusetts

SEER will contractually require submission to the NCCR

## Goal

**100%**

coverage of all pediatric patients over the next few years

# NCI-DOE Collaboration: Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)

DOE-NCI partnership to advance exascale development through cancer research

NIH NATIONAL CANCER INSTITUTE

Argonne NATIONAL LABORATORY

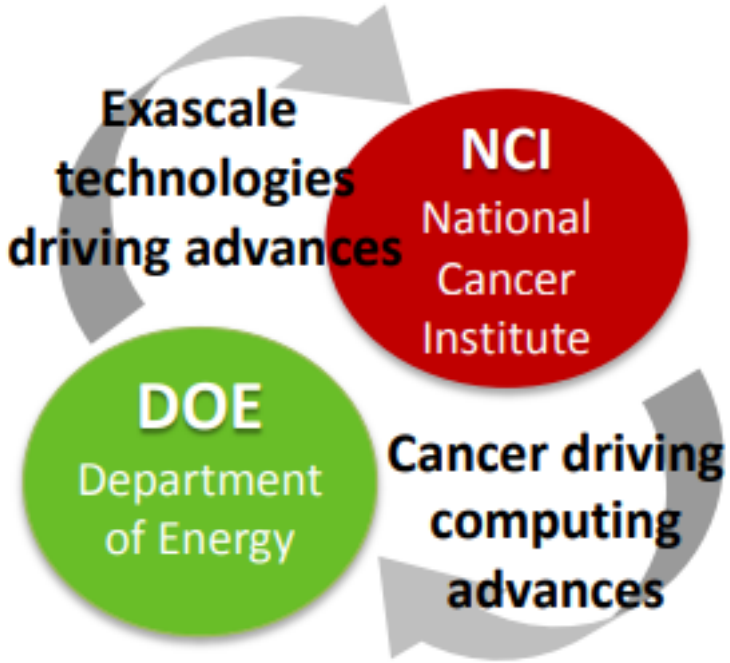
OAK RIDGE

Lawrence Livermore National Laboratory

Los Alamos NATIONAL LABORATORY EST. 1943

Frederick National Laboratory for Cancer Research

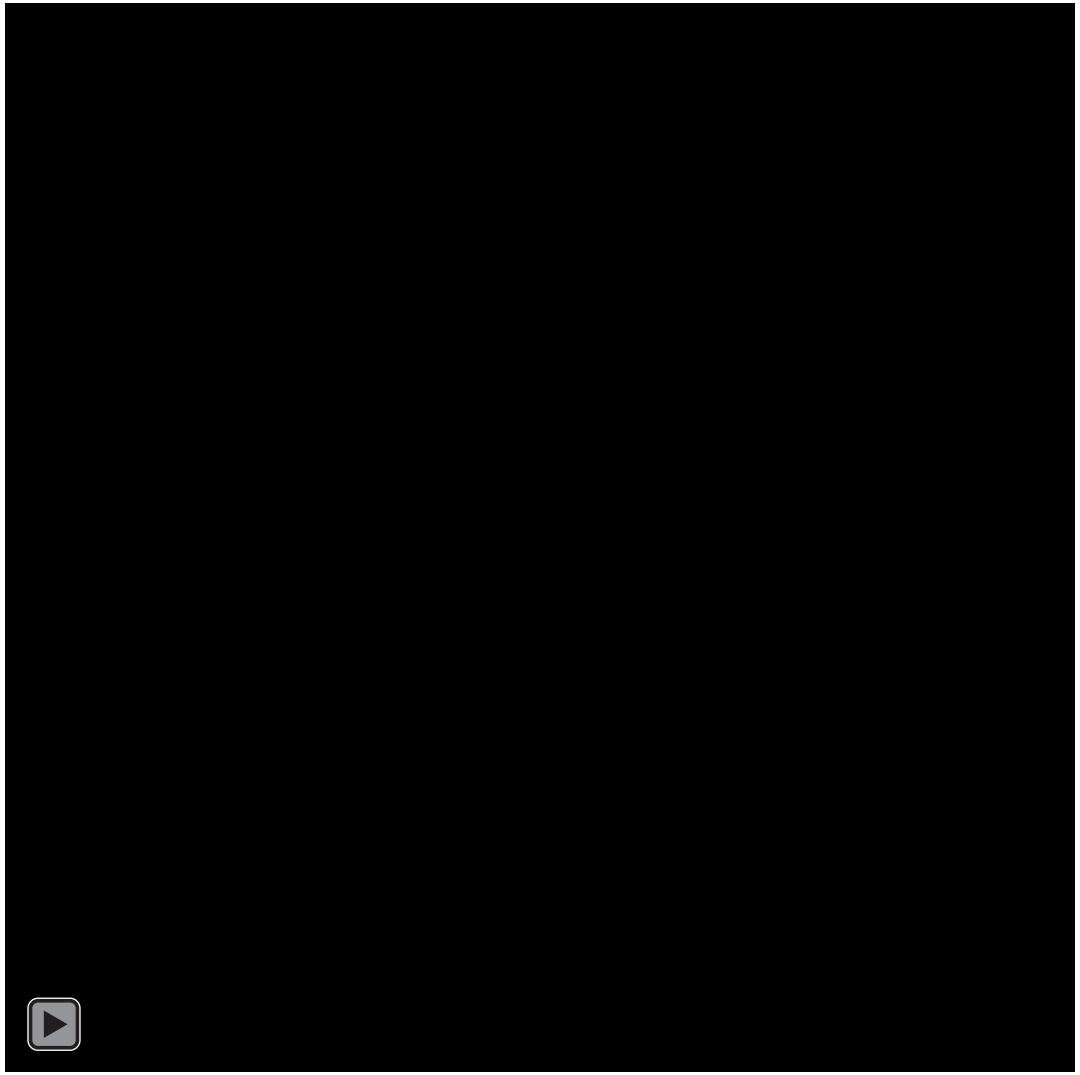
*sponsored by the National Cancer Institute*



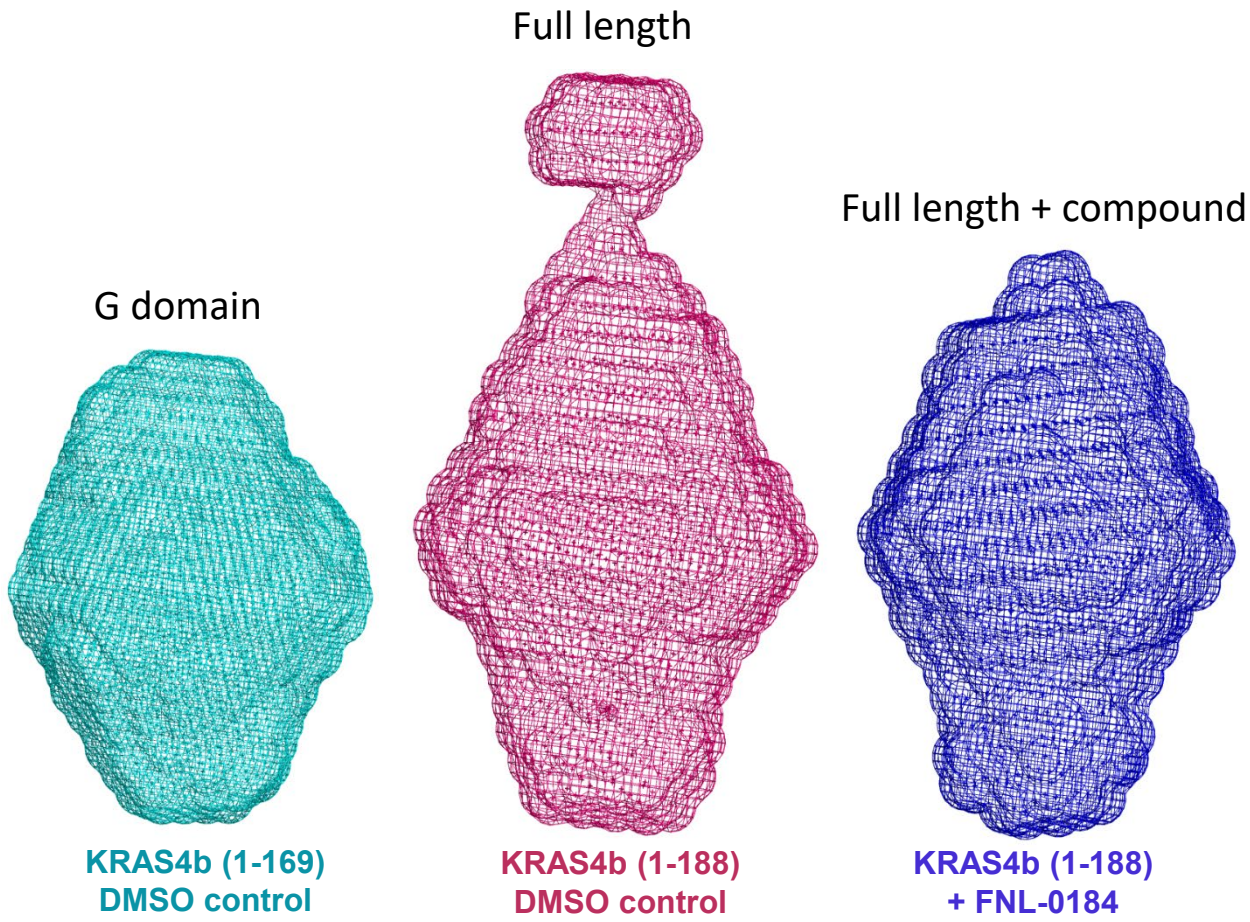
# NCI-DOE Collaboration (JDACS4C)

| | |
|---|---|
| **PILOT 1**<br>Predictive Modeling for Pre-Clinical Screening<br>Develop reliable machine-learning based predictive models of anti-cancer drug response | **CANDLE** (CANcer Distributed Learning Environment)<br>An exascale computing project to develop machine learning framework for cancer |
| **PILOT 2**<br>RAS Biology in Membranes<br>Develop multiscale modeling capabilities to investigate RAS dynamics on cell membranes | **ATOM** (Accelerating Therapeutics for Opportunities in Medicine)<br>Developing a pre-clinical drug design and optimization platform that leads with computation to help shorten the drug discovery timeline |
| **PILOT 3**<br>Population Information Integration, Analysis, and Modeling Information; capture of unstructured clinical text using Natural Language Processing (NLP) and Deep Learning algorithms | **UNCERTAINTY QUANTIFICATION**<br>General methods to improve confidence/level of certainty of results from predictive computational models from JDACS4C pilots |

# MD simulation of KRAS4b suggests the hypervariable region (HVR) forms binding pocket with the G domain



Yue Yang and Felice Lightstone, LLNL – collaboration with the DOE




Lixin Fan, The Small-Angle X-ray Scattering Core Facility, NCI

# What Lies Ahead? Unresolved questions in cancer research where high-performance computing may be critical

**Structural Biology**

*predicting T cell receptor binding to MHC + antigen, predicting Ab structure*

**Medicinal Chemistry and Drug Discovery**

*in silico docking studies, predicting in vivo toxicity, etc.*

**Novel Analytics of Large Complex Datasets**

*proteomic-genomic-clinical data; large deidentified sets of clinical and personal data linked via unique identifier technologies*

**Modeling of complex, multi-parameter systems within a large search space**

*models of hematopoiesis, drug resistance in population of inter-dependent tumor cells, etc.*

# Discussion