NCI-DOE Micro Lab

Toward Building a Cancer Patient "Digital Twin": A Preview of the July Ideas Lab

Thursday, April 23, 2020, 1:00 pm-2:30 pm EST via Zoom

1

What is an Ideas Lab?

- 5-day residential experience
- Teams and Ideas bubble up from participants
- Unintended consequences: Community
- Intended as a face-to-face event but can go virtual
- Application and selection process

Micro Lab Origins

Emily Greenspan
Center for Biomedical Informatics and Information Technology
National Cancer Institute

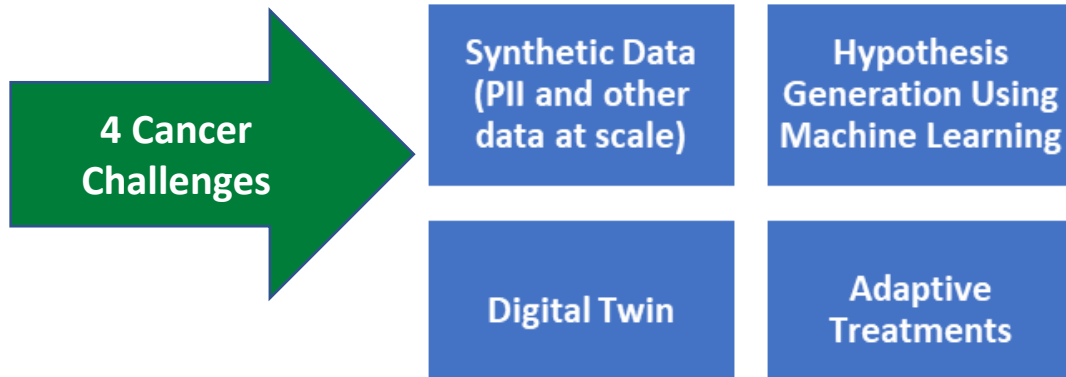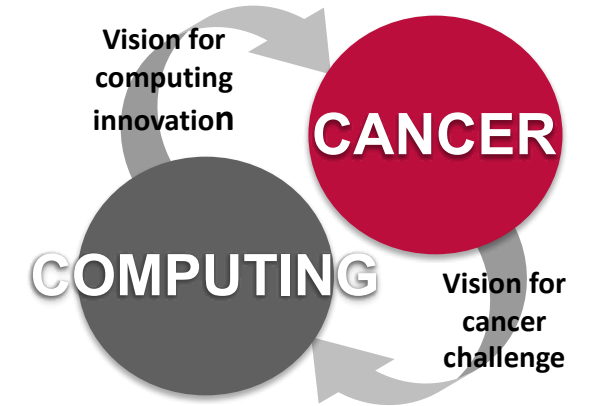# Multidisciplinary Engagement Across Cancer Research, HPC and AI
Envisioning Computational Innovations for Cancer Challenges (ECICC) Community

PURPOSE: Build a community, Multidisciplinary engagement, and collaboration among cancer, data, and computational scientists to create transformative impact

ORIGIN: Outgrowth of **NCI DOE Collaboration** Joint Design of Advanced Computing Solutions for Cancer (JDACS4C**)**

ECICC SCOPING MEETING – March 25-27, 2019
- Multidisciplinary meeting
- Held at DOE's Lawrence Livermore National Lab
- *74 computational scientists & cancer researchers from all career stages participated*
- *Identified over 200 cancer challenge ideas*

Vision for computing innovation

CANCER

COMPUTING

Vision for cancer challenge

4 Cancer Challenges →

| Synthetic Data (PII and other data at scale) | Hypothesis Generation Using Machine Learning |
| Digital Twin | Adaptive Treatments |

Scoping Meeting Follow-up: Community Building
- Ongoing leadership from challenge area team leads
- Meeting Report
- Virtual, interactive "Micro Labs"
- NCI Hub site created for community member engagement
- Plans underway for a 5-day Ideas Lab in July 2020

Multidisciplinary Engagement Across Cancer Research, HPC and AI
Envisioning Computational Innovations for Cancer Challenges (ECICC) Community

1st Micro Lab (June 2019):
- Over 200 participants spanning 50 organizations
- Breakout on the 4 major themes developed at March ECICC scoping meeting

2nd Micro Lab (Sept 2019):
- Over 100 clinicians, researchers, and academics in cancer and computational science
- Participants developed 13 use cases

Scoping Meeting March 25-27, 2019	1st Micro Lab June 2019	Scoping Meeting Report Released August 2019	2nd Micro Lab September 2019	3rd Micro Lab April 2020	Ideas Lab July 6-10, 2020

3rd Micro Lab (Today)
- Promote the Ideas Lab
- Develop questions to be addressed at the Ideas Lab

Ideas Lab (July 2020)
- 5-day, Multidisciplinary workshop to guide participants in developing pre-project proposals
- Focus on Digital Twin is being developed with a Steering Committee from NCI, DOE, & FNL
- *Applications* **are due April 30** (https://events.cancer.gov/cbiit/dtwin2020/registration)

NCI-DOE Collaboration Lab: Toward Building a Cancer Patient "Digital Twin"

An Ideas Lab to Shape the Future of Predictive Modeling Across Scales from Basic Biology to Clinical Care

What is an Ideas Lab: an intensive 5-day program for 30 participants with a range of expertise and experience

What you will do: Form new, multidisciplinary teams and receive guidance from mentors to develop innovative research projects.

- Projects will focus on a digital twin component that, within the next 2-3 years, will advance the development of a model of an individual cancer patient.

Outcomes:

- Innovative research projects to advance the development of a human digital twin; can be submitted for seed funding

- Each team will write a 3-5-page research concept and deliver a 10-minute presentation describing their concept on the last day of the event.

Digital Twin Technology

Sean Hanlon

Associate Director of the Center for Strategic Scientific Initiatives
*National Cancer Institute*

What is *predictive* medicine?

- *Precision* **medicine** is generally focused on stratification to match cancer patients to the "right" intervention, based on individual profiling.
  - Which prior patients does this patient best match?
  - What worked best for those best matched patients?
  - *Entirely based on observables and prior measurements.*

- *Predictive* **medicine** aims to predict individual patient trajectories, based on patient-centric calibration.
  - What is the expected disease course without intervention?
  - What is the expected response to a proposed intervention scheme?
 - *Integrates observables and dynamical theory.*

What is a digital twin?

- A ***digital twin is synchronized digital replica*** of a physical system. The digital twin is used to ***monitor, model, and fine-tune performance*** of processes, people, places, systems and devices.

- Digital twins can be used for ***virtual experiments*:**
  - *What if* I run the engine hotter?
  - *What if* there are 100x as many people working remotely?

- In medicine, a digital twin is a ***patient-tailored model*** that can incorporate genetic, molecular, clinical, environmental, and social factors to ***predict individual patient trajectories*** to:
  - Evaluate potential ***screening, preventive, or therapeutic*** plans;
  - Help prioritize a plan to meet ***personalized objectives***;
  - Continuously ***integrate new data and knowledge*** to refine intervention plans.
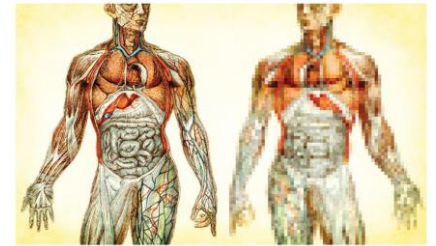
Digital twins could help us plan cancer care

1. Patient and clinicians discuss *intervention goals* and preferences
2. Use *patient data* to build a *digital twin*
3. Use HPC to explore *thousands of intervention plans* on the virtual twin
4. Patient and clinicians *explore the results*:
   - Predicted response
   - Adverse effects
   - Long-term implications
5. *Choose* a plan
6. *Benchmark* progress against digital twin
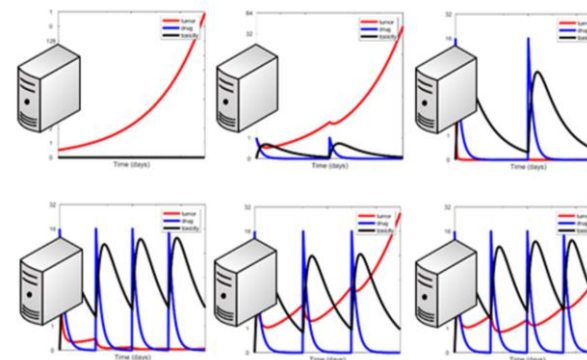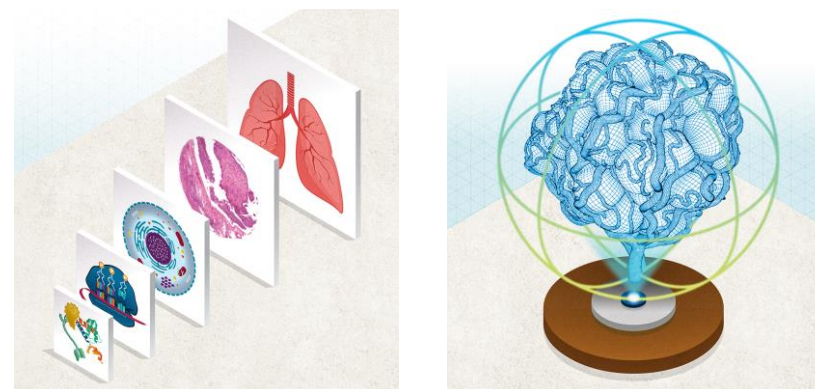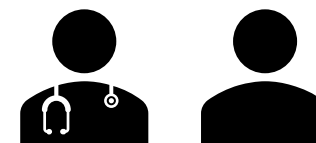
# What are the barriers and how have we made progress?

- **Data sandboxes and platforms**
  - Assembly of coordinated *data across scales per patient*
  - Development of *data commons* to support the analysis
- **Analytics**
  - *Multi-scale analysis* and coupling of data
  - Integration of *dynamical models with AI*
- **HPC-driven healthcare informatics**
  - *Simulation and virtual evaluation* of care pathways
  - *Assess simulated data* for therapies and other interventions to assist decision making and development healthcare policies

NCI Human Tumor Atlas Network is generating highly multi-scale and multi-modal datasets

Host level (microbiome, circulating fact... ...ical data)

Medical imaging modalities (radio... ...)

Histology (Highly multiplexed... ...D imaging)

Metabolomics (Mass Sp... ...I Imaging)

Proteomics (Mass... ...ging Approaches)

Transcriptomi... ...Seq, in situ/FISH, etc.)

DNA-se... ...enome, whole exome) and Epigenetics

EM imag... ...2D, 3D)

Many at single-cell resolution, longitudinal, & with spatial information

https://humantumoratlas.org/
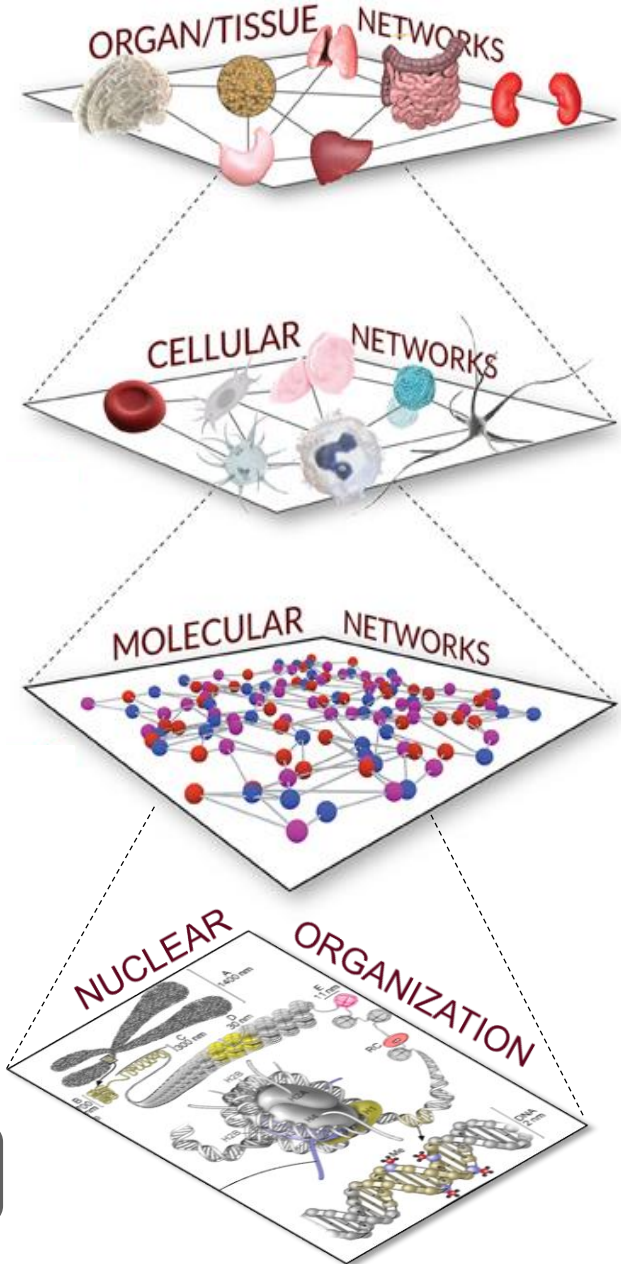
Multi-scale modeling!

Data Integration

ORGAN/TISSUE NETWORKS

CELLULAR NETWORKS

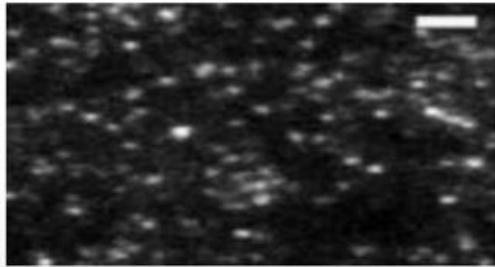MOLECULAR NETWORKS

NUCLEAR ORGANIZATION

Digital Twin Technology

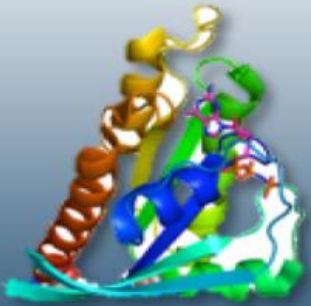**Eric Stahlberg**
Director of Biomedical Informatics and Data Science
Frederick National Laboratory for Cancer Research

JDACS4C – Precursors to Cancer Patient Digital Twins

Inhibiting the RAS cancer pathway

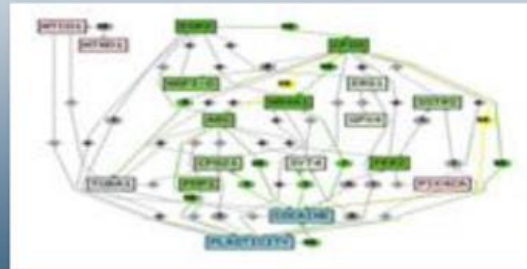
Experimental views of RAS dynamics

Atomic-resolution molecular-dynamics simulations

Predicting drug response in preclinical models

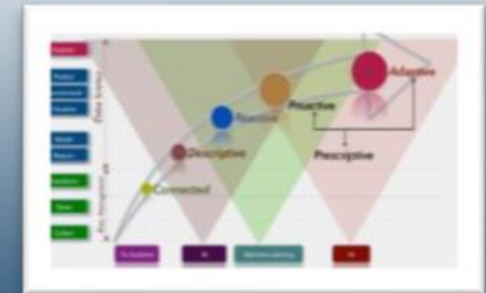
Cell line and PDX database with patient-derived samples

Testing hypotheses generated by ML models

Cancer trajectories from population models
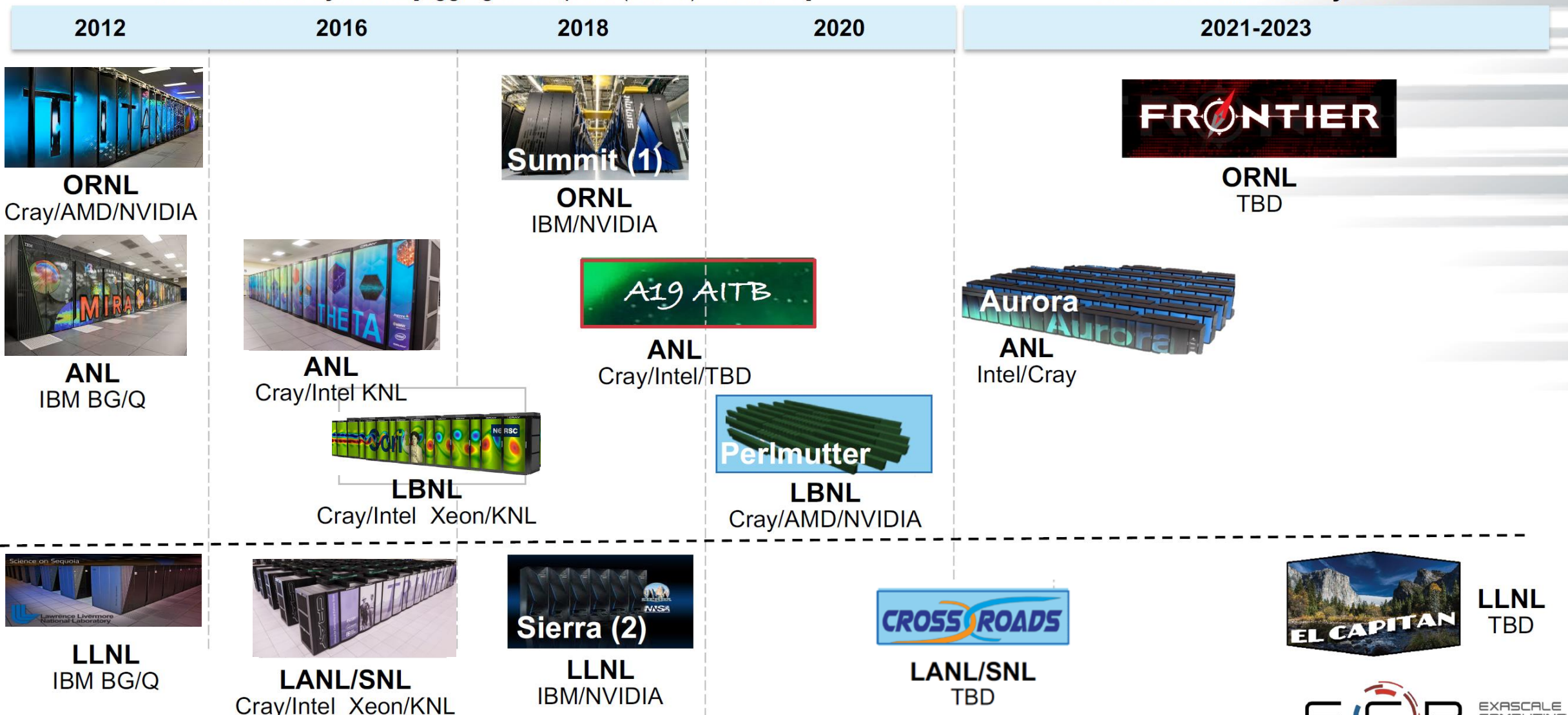

Enhancing the national database

Learning patterns of treatment and outcomes

DEPARTMENT OF ENERGY ROADMAP TO EXASCALE SYSTEMS

Pre-Exascale Systems [Aggregate Linpack (Rmax) = 323 PF!]

First U.S. Exascale Systems

2012	2016	2018	2020	2021-2023

ORNL
Cray/AMD/NVIDIA

Summit (1)
ORNL
IBM/NVIDIA

FRONTIER
ORNL
TBD

ANL
IBM BG/Q

ANL
Cray/Intel KNL

A19 AITB
ANL
Cray/Intel/TBD

Aurora
ANL
Intel/Cray

LBNL
Cray/Intel Xeon/KNL

Perlmutter
LBNL
Cray/AMD/NVIDIA

LLNL
IBM BG/Q

LANL/SNL
Cray/Intel Xeon/KNL

Sierra (2)
LLNL
IBM/NVIDIA

CROSSROADS
LANL/SNL
TBD

EL CAPITAN
LLNL
TBD

CANDLE – Pushing the Limits for AI Applied to Cancer

Project focuses on machine learning and building a scalable distributed learning environment called CANDLE (CANcer Distributed Learning Environment) used to help address the three challenges listed above.

Treatment Optimization - ATOM Active Learning
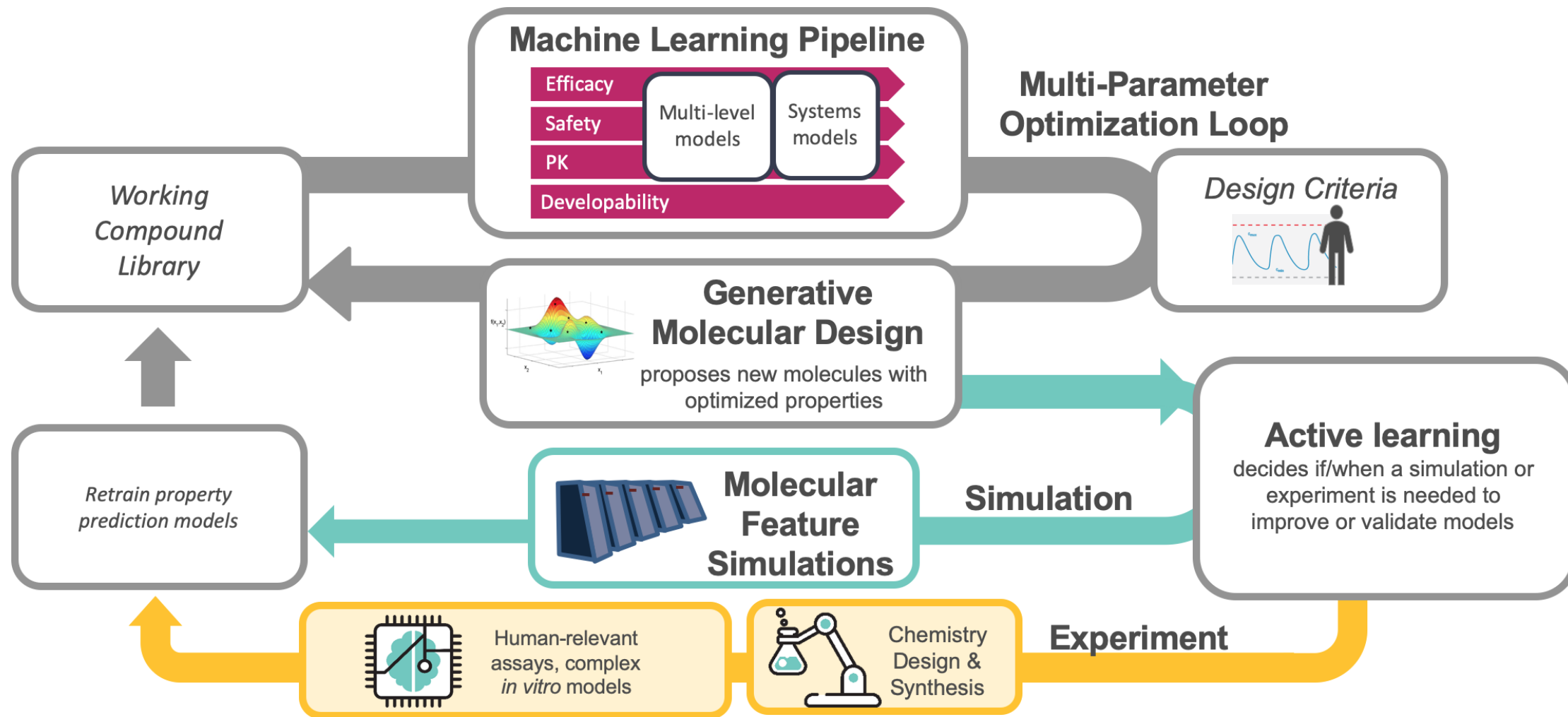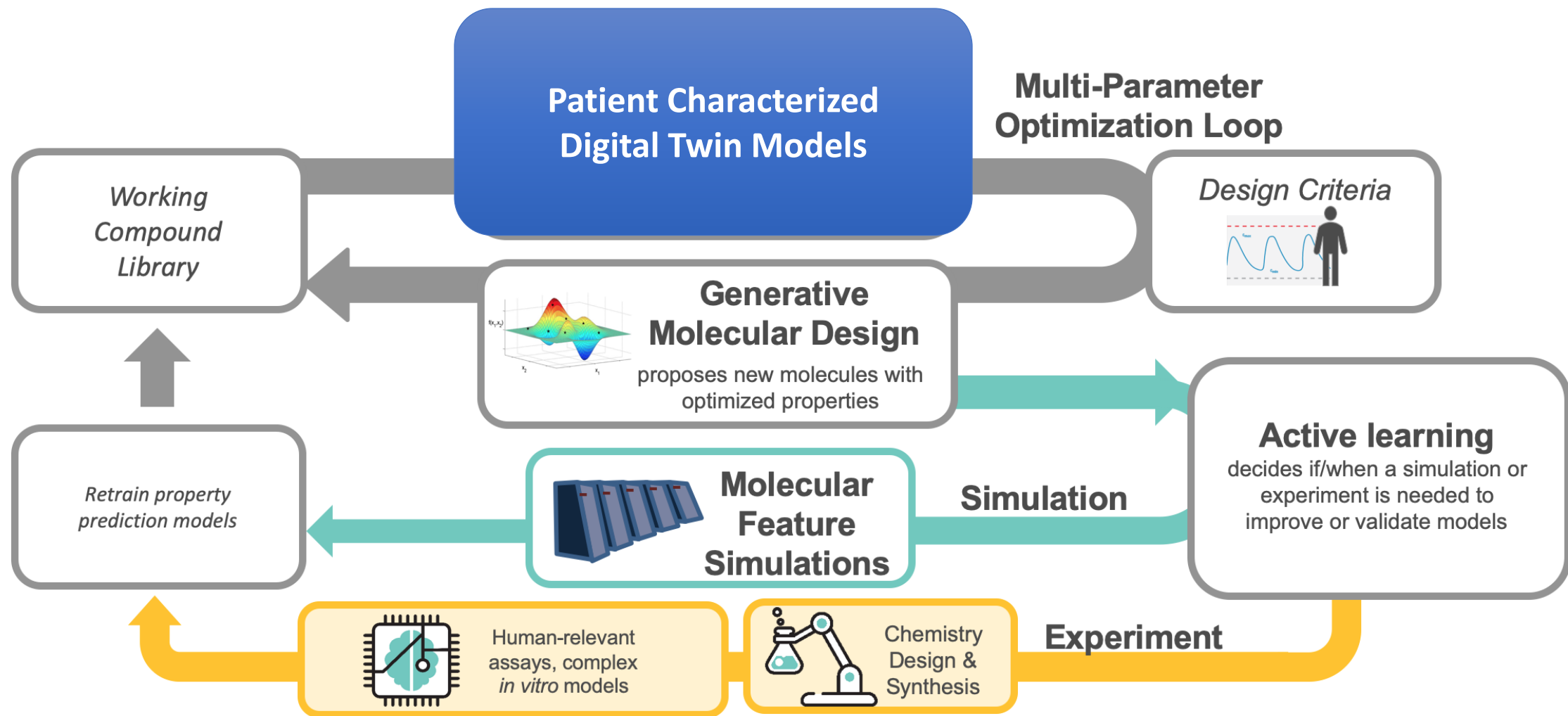

Machine Learning Pipeline

Efficacy
Safety
PK
Developability

Multi-level models

Systems models

Multi-Parameter Optimization Loop

Design Criteria

Working Compound Library

Generative Molecular Design
proposes new molecules with optimized properties

Active learning
decides if/when a simulation or experiment is needed to improve or validate models

Retrain property prediction models

Molecular Feature Simulations

Simulation

Human-relevant assays, complex *in vitro* models

Chemistry Design & Synthesis

Experiment

Treatment Optimization - ATOM Active Learning

Cancer Surveillance Big Data Challenges

Around 34% of US Population

Around 550,000 new cases per year

Multiple clinical reports per case

Over 360 reporting laboratories

High security and data confidentiality requirements

High data quality requirements

22-month delay in cancer reporting[1]

Heterogeneous data from acquisitions and linkages

Volume

Velocity

Big Data

Veracity

Variety

Patient/ Pathologist → Surgical pathology → Certified Tumor Registrar → Regional cancer registries → Integration with structured data from patient EMRs → NCI SEER database

SEER

A Digital Twin in a Learning Cancer Health System

Cancer Knowledge

Available Data

Patient Profile

Cancer Patient Digital Twin

Intervention Choice

Patient Care Decision
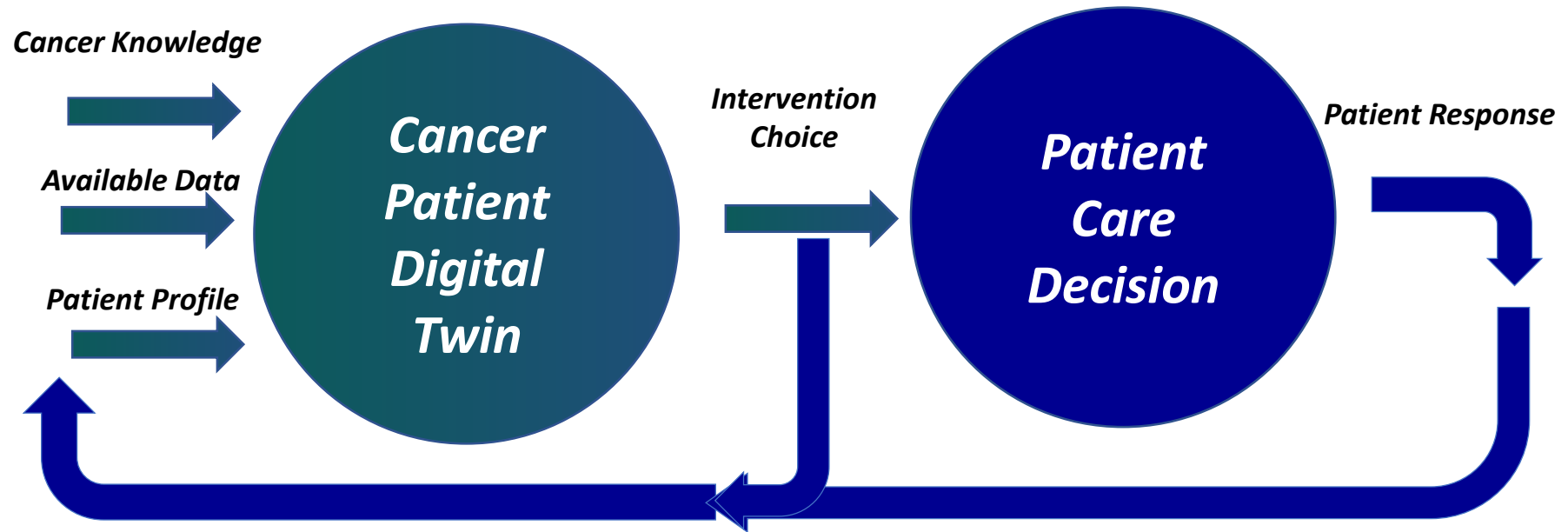
Patient Response

- Iterate with the CPDT to identify the most suitable treatment
- Apply the selected treatment as part of the patient care decision
- Capture the response of the patient to the decision

Multiple cycles compose the cancer patient health trajectory

So What's Next?

- ***Complete the application for the Ideas Lab***
  - ***Let us know you are interested***

- ***Micro Lab Discussion - Ask questions***
  - ***Quick questions first***
  - ***Longer scientific discussions for breakout sessions***

- ***Send follow-up questions to ECICC_Community@nih.gov***

- **Thank you!**