

Frederick National Laboratory for Cancer Research



Frontiers for Deep Learning and Cancer

CANcer Distributed Learning Environment (CANDLE) Workshop

February 21-22, 2018

Introduction



- Motivations
 - For workshop
 - For large scale computing in cancer research
- Joint Design for Advanced Computing Solutions for Cancer
- CANDLE and Deep Learning

Workshop Objectives



- Overview of NCI and DOE collaborative efforts
- Grow community around deep learning applied to cancer and CANDLE (CANcer Distributed Learning Environment)
- Identify priority areas to explore challenges and opportunities

NIH CANDLE Workshop 2017



- Participation included ~60 attendees spanning 13 NIH institutes, listed by breadth of attendance as follows:

- NCI (30)
- NHLBI (5)
- NLM (4)
- DRD (3)
- CIT (2)
- Netrias, LLC (1)
- NHGRI (1)
- NIA (1)
- NIDCD (1)
- NIDCR (1)
- NIDDK (1)
- NINDS (1)
- IRSB (1)



- Take Away Points
- Request to establish trans-NIH Deep Learning group
- New use cases identified - Applying deep learning to genomic data, imaging, data visualization, natural language processing
- Future CANDLE workshops are under development for anyone interested in deep/machine learning, or leveraging these to increase scientific productivity and accelerate cancer research

NIH 2018 CANDLE Workshop

February 21-22, 2018



2018 CANDLE Workshop registrants

(> 100 participants):

- NCI (77)
- NIAID (14)
- CC (13)
- NLM (12)
- NHLBI (10)
- NIMH (10)
- NIDDK (8)
- CIT (6)
- OD (6)
- NIA (5)
- NICHD (5)
- NHGRI (4)
- NINDS (3)
- NCATS (3)
- NEI (2)
- NIAAA (2)
- NIAMS (2)
- NIBIB (2)
- NIDA (2)
- NIEHS (2)
- NIGMS (2)
- NIDCR (1)
- CSR (1)
- FIC (1)

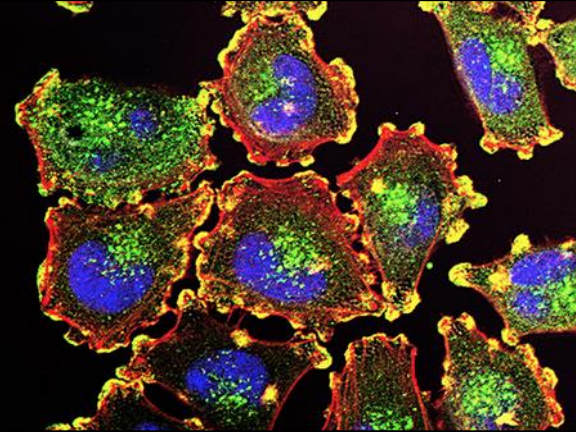
Machine Learning/Deep Learning topics of interest (top 5 survey responses) :

- Imaging
- Next-generation Sequencing
- Genomics and Genetics
- Text Analysis
- Big Data

APPROXIMATE Level of adoption (beginner, intermediate, advanced):

- 100 introductory users
- 48 beginner-level adoption
- 30 intermediate
- 15 advanced users (not including instructors)

Pushing the Frontiers of Cancer Research with Large Scale HPC



Challenge Problem

Need for Exascale

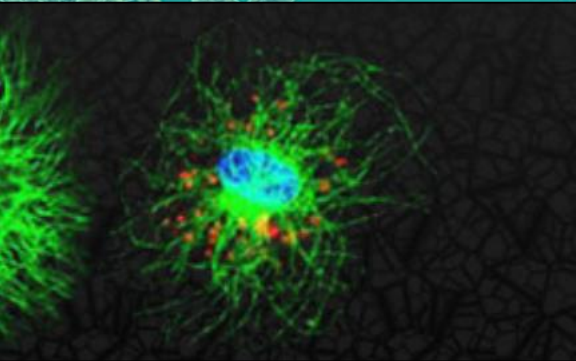
Mapping genetic susceptibility to cancer and its outcomes: investigate the contributions of genetic heritability to cancer and its outcomes, investigators must assess complex models of multiple gene-gene and gene-environmental interactions

Requires exascale computing to explore the interactions and develop subsequent robust models, including confirmation with large-scale permutation of parameters



Integrated genomic data analysis: numerous different types of genetic, genomic, and clinical data provides the insight to develop molecularly-targeted interventions

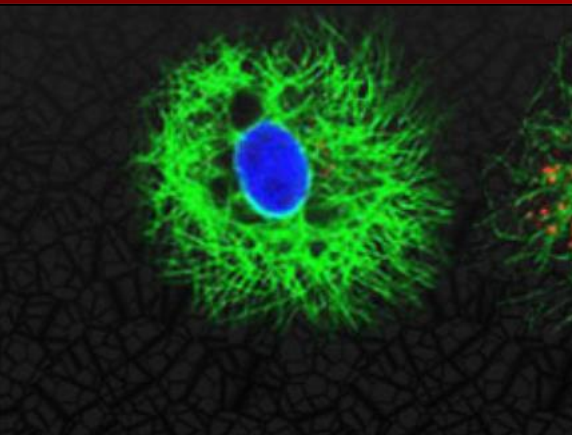
Requires exascale computing as combinatorial possibilities increase geometrically with the number of different genomic features considered



High-accuracy modeling of intracellular molecular signaling in complex mutational backgrounds: inform the critical pathways used, explore functionally-redundant signaling for inhibited pathways and bring greater awareness of complex interactions

Requires exascale computing to advance explorations of the large dynamic combinatorial multi-protein assemblies (membrane-associated and vary across cells, tissues and cancer types)

Pushing the Frontiers of Cancer Research with Large Scale HPC



Challenge Problem

High-accuracy modeling of cancer drug interactions and reaction dynamics in complex mutational backgrounds: allow the study and modeling of critical aspects of a drug's mechanism of action across the full spectrum of possible mutational backgrounds

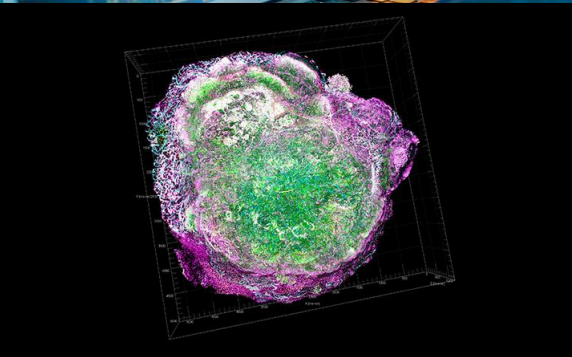
Need for Exascale

Requires exascale computing to evaluate reaction energies for extremely large systems in complex biological systems orders of magnitude greater than current capabilities



***In silico* characterization of 100 billion readily synthesizable potential cancer drugs:** provide knowledge base of cost-effective, producible molecules characterized for potential clinical use and for future studies focusing on the system biology and dynamics of cancer

Requires exascale computing to support the generation and *in silico* characterization of more than 100 billion candidate molecules important to clinical and research applications in cancer



Computing three-dimensional structures of RNA: critical first step to understanding their function and activity and ultimately to the development of therapeutic agents targeting RNA molecules

Requires exascale computing to enable data collected from instrument observations to be used as a basis for accurate determination of previously unknown 3D topological structures of RNA

Pushing the Frontiers of Cancer Research with Large Scale HPC

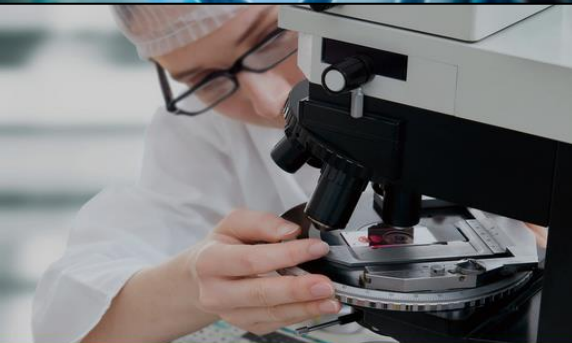


Challenge Problem

Need for Exascale

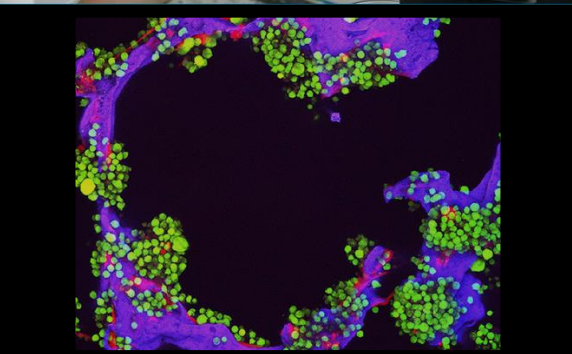
Accelerated characterization of biomolecular structures using advanced imaging techniques: ability to generate a more detailed biophysical profile for each patient will greatly improve diagnostic capabilities and expedite the delivery of precision medicine

Requires exascale computing to facilitate more rapid evaluation of higher-resolution images using significantly greater levels of optimal pattern registration and multi-modality data integration



Rapid characterization of bio-molecular dynamics and drug interactions from NMR data: provide essential knowledge of critical ligand interactions, biological complexes and interaction partners in cancer systems

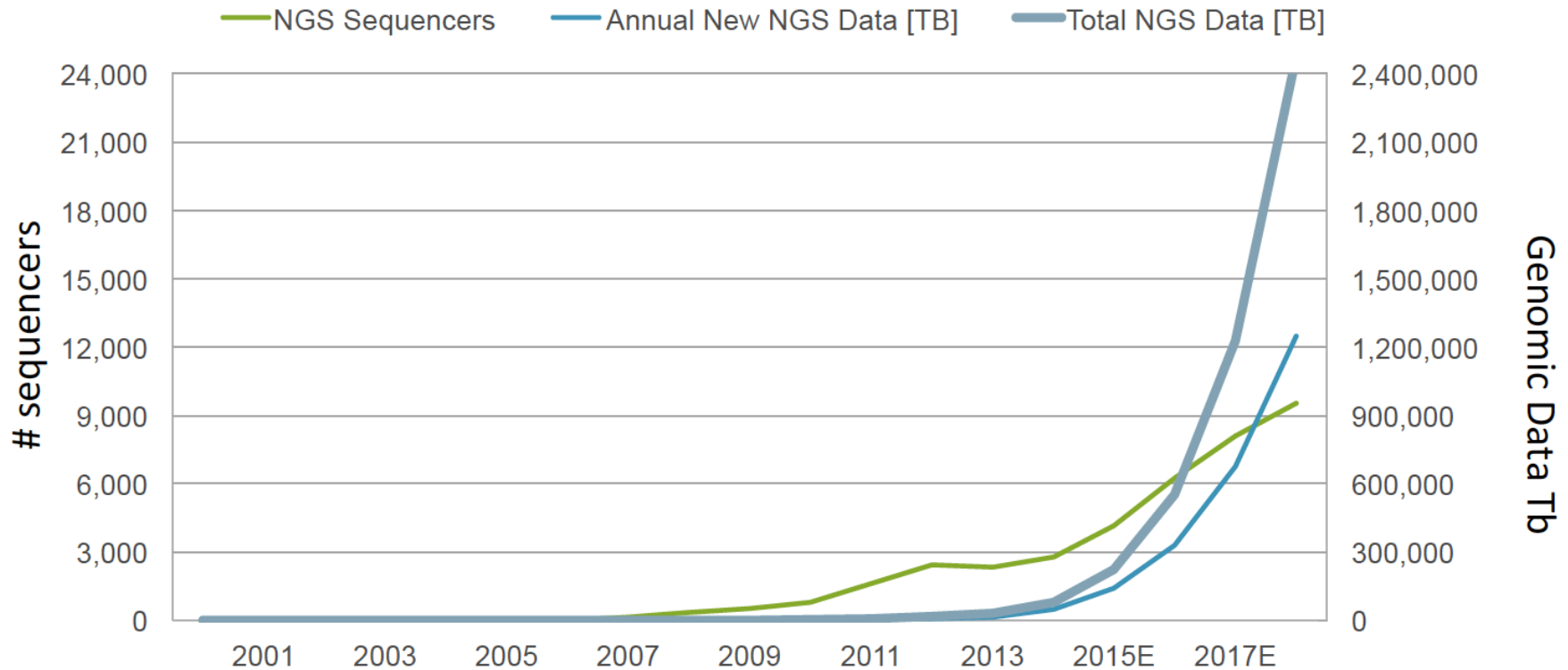
Requires exascale computing to move this challenging manual problem into a computational domain to reduce the time required to interpret data from months to minutes



Integration of multi-modal experimental data to model macromolecular complexes: blend detailed physical studies with molecular and cellular biology to provide critical molecular insight in the development of new treatment modalities

Requires exascale computing to explore all possible inter-relationships and enable validation of the correlations through back-simulation of experimental data

Amount of genomic data will exceed available resources



Between 2014-2018 production of new NGS data to exceed **2 Exabytes**

NGS: Next Generation Sequencing

NGS sequencers include machines from Illumina, Life Technologies, and Pacific Biosciences. Human genome data based on estimates of whole human genomes sequenced

Sources: Financial reports of Illumina, Life Technologies, Pacific Biosciences; revenue guidances; JP Morgan; The Economist; Seven Bridges Analysis.

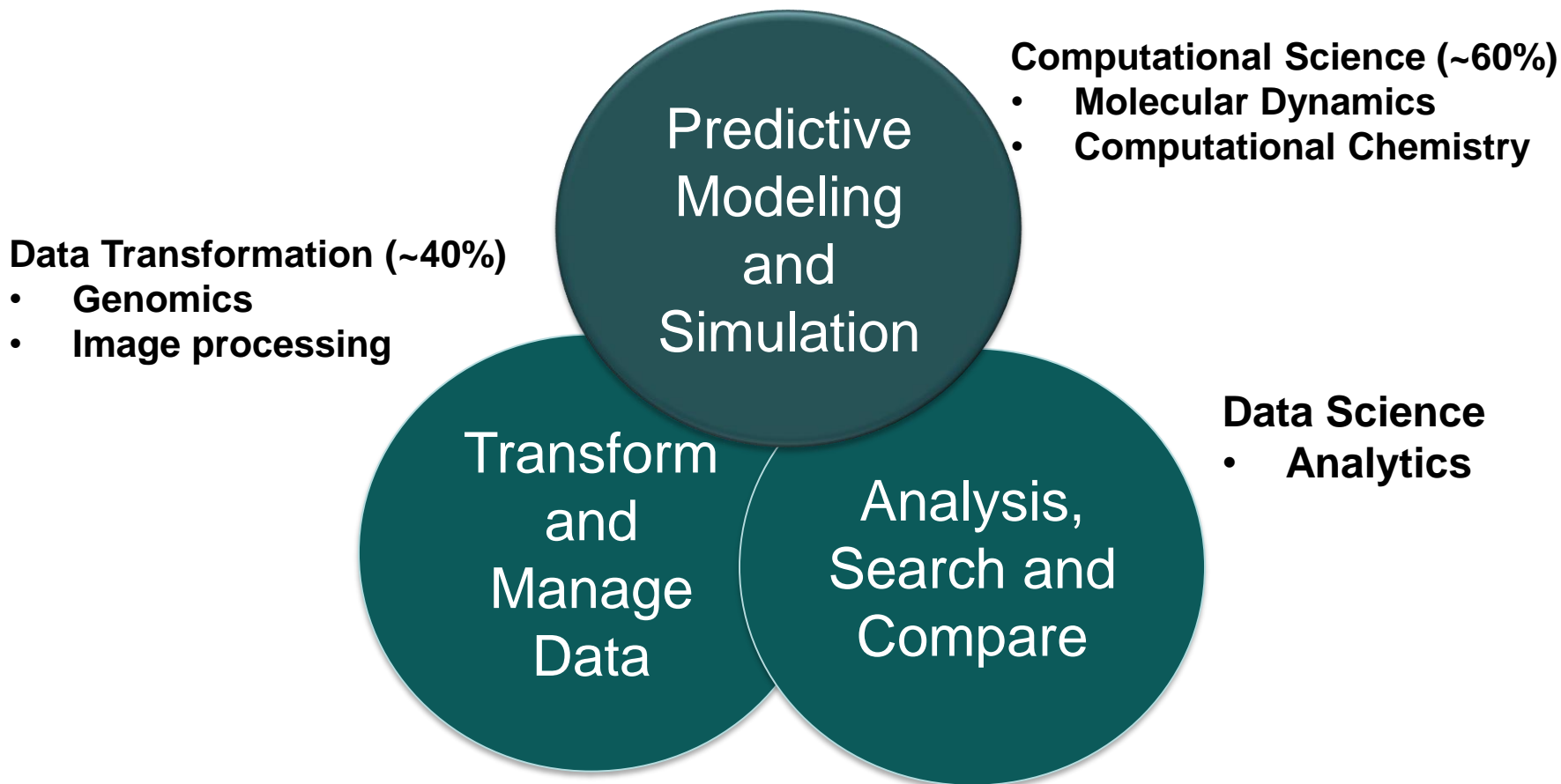
Context for Advanced Computing



- Tools and cultures of HPC and Big Data have diverged to the detriment of both; unification essential to address major challenges
- The challenges of scale tax ability to transmit data, compute complicated functions, or store substantive portions; new approaches are needed
- International nature of science demands further development of advanced computer architectures and global standards for processing data, even as international competitiveness complicates the openness of the scientific process

Source –Reed and Dongarra – Communications of the ACM, July 2015 (vol 57, no 7)

Expanding Role of HPC in Cancer



(Estimated usage numbers for provided by Sean Davis, NCI Center for Cancer Research)

Advancing Precision Oncology




NATIONAL CANCER INSTITUTE
ADVANCING PRECISION ONCOLOGY
UNDER THE NATIONAL PRECISION MEDICINE INITIATIVE

Precision oncology: using molecular information about a patient's cancer to inform treatment


To make precision oncology a reality in everyday clinical practice, NCI is leading research to:

EXPAND PRECISION MEDICINE CLINICAL STUDIES TO ADULTS AND CHILDREN IN THEIR COMMUNITIES



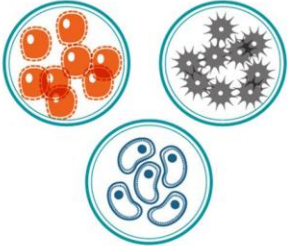
to test new cancer treatments

OVERCOME DRUG RESISTANCE




to learn why cancer treatments stop working in many patients

INCREASE THE NUMBER OF LABORATORY MODELS OF HUMAN CANCER



to test potential treatments and learn more about cell changes that drive cancer

BUILD A KNOWLEDGE NETWORK THAT INTEGRATES CANCER GENOMIC INFORMATION WITH CLINICAL INFORMATION



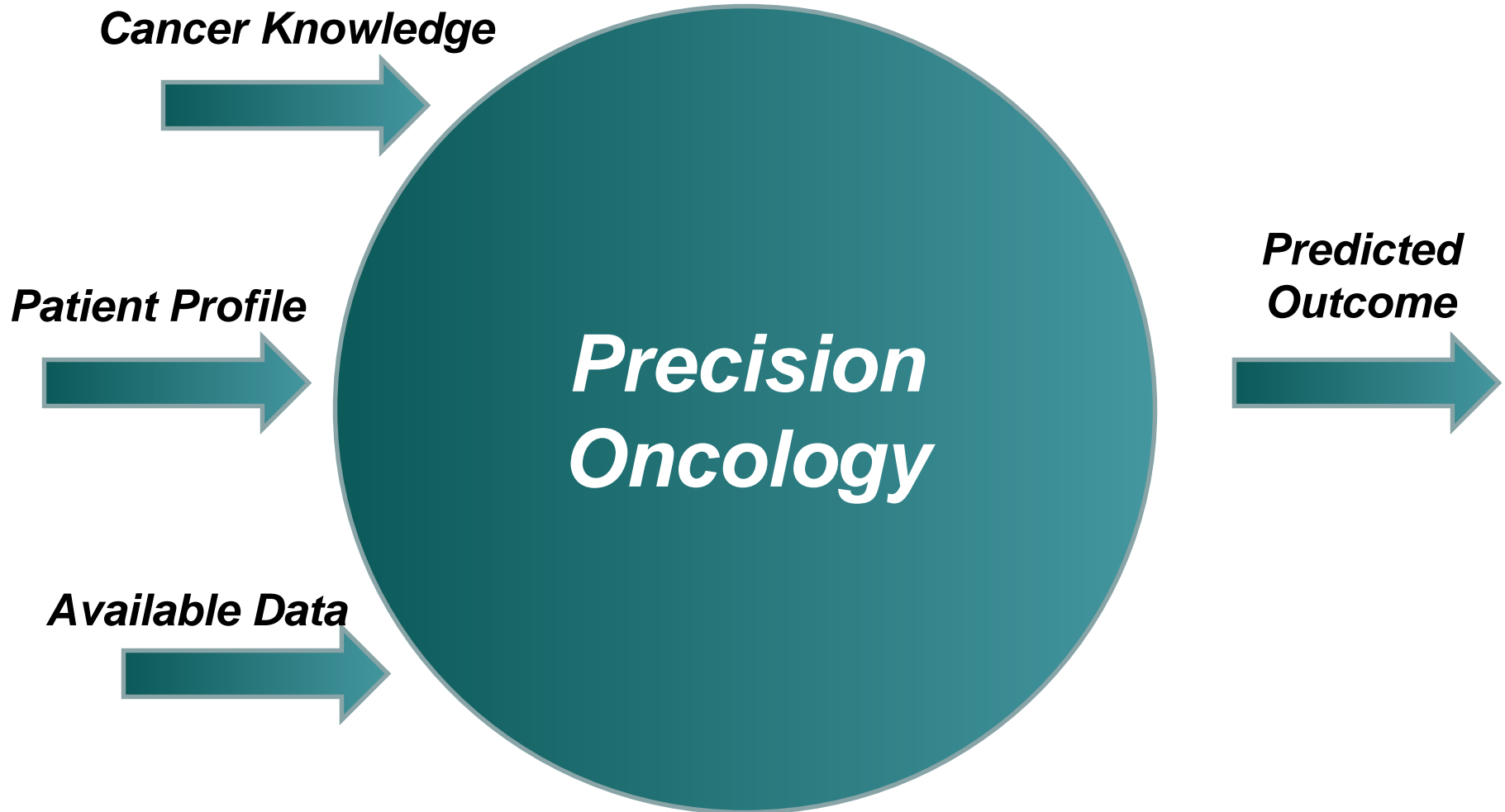
to serve as a resource for scientists, health care professionals, and patients

www.cancer.gov/precision-medicine

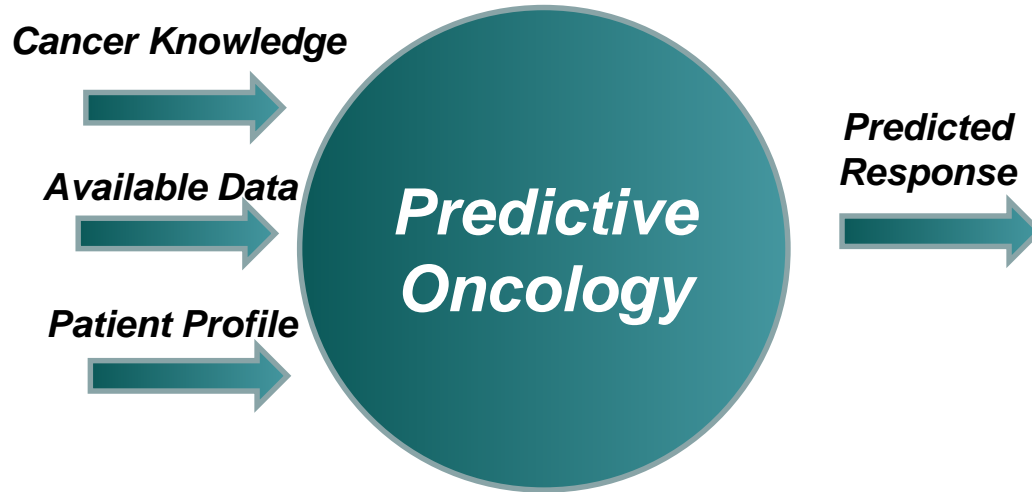
NCI Mission (under National Precision Medicine Initiative)

- *Accelerating development of new treatment options*
- *Overcome drug resistance*
- *Increase number of laboratory models*
- *Build an integrated knowledge network*

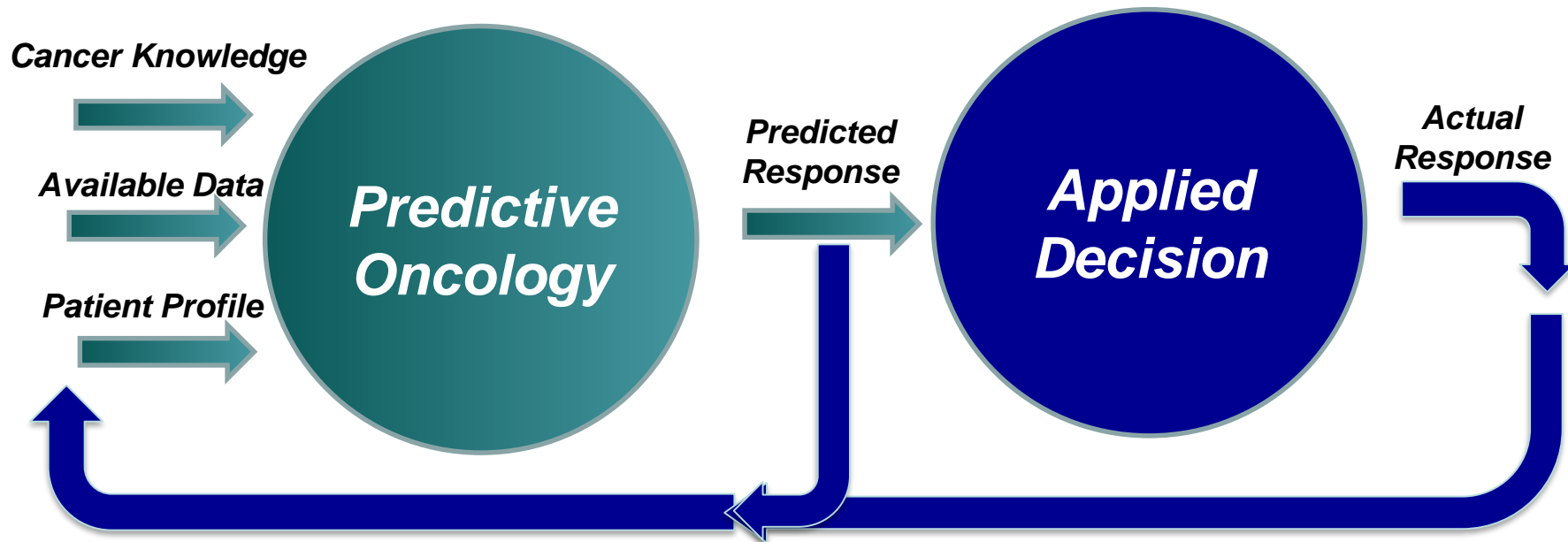
HPC Enabling Precision Medicine



Predictive Oncology



Oncology Learning System

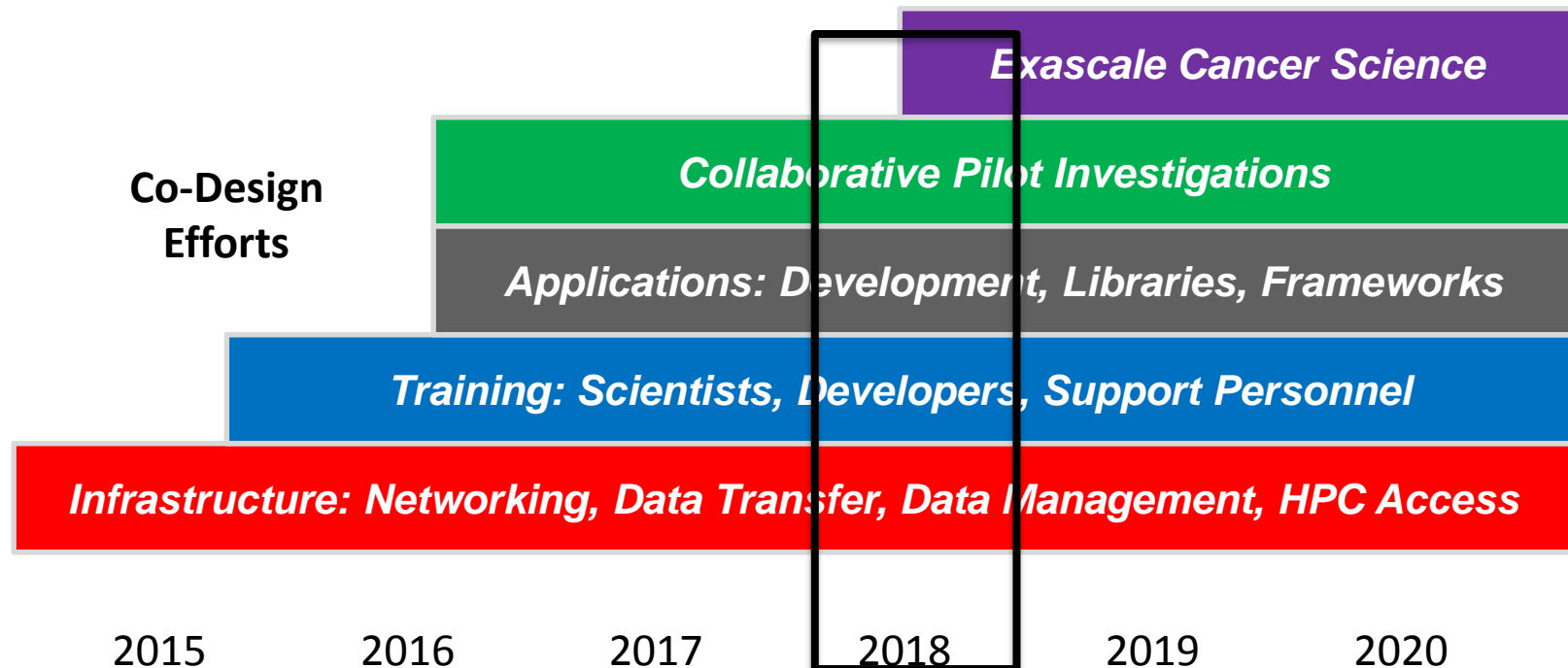


Exascale Cancer Science



Exascale in a nutshell:

- Millions of CPU cores contributing to a single task
- Nearly 1000 times faster than fastest computer today
- Focus of DOE Advanced Strategic Computing



NCI-DOE Collaboration Pilots: Advancing Precision Oncology





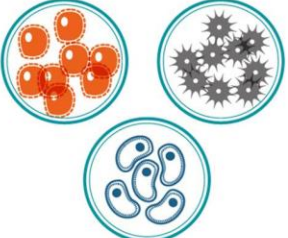

NCI Mission Impact:
*Accelerating
development of new
treatment options for
precision cohorts*

Pilot 1:
Pre-clinical Models
*Predictive patient drug
response models with
advanced computing*

**NATIONAL CANCER INSTITUTE
ADVANCING PRECISION ONCOLOGY**
UNDER THE NATIONAL PRECISION MEDICINE INITIATIVE

Precision oncology: using molecular information about a patient's cancer to inform treatment

To make precision oncology a reality in everyday clinical practice, NCI is leading research to:

<p>EXPAND PRECISION MEDICINE CLINICAL STUDIES TO ADULTS AND CHILDREN IN THEIR COMMUNITIES</p>  <p>to test new cancer treatments</p>	<p>OVERCOME DRUG RESISTANCE</p>  <p>to learn why cancer treatments stop working in many patients</p>
<p>INCREASE THE NUMBER OF LABORATORY MODELS OF HUMAN CANCER</p>  <p>to test potential treatments and learn more about cell changes that drive cancer</p>	<p>BUILD A KNOWLEDGE NETWORK THAT INTEGRATES CANCER GENOMIC INFORMATION WITH CLINICAL INFORMATION</p>  <p>to serve as a resource for scientists, health care professionals, and patients</p>

www.cancer.gov/precision-medicine

Pilot 2:
Biological Models
*Multi-scale
computational biological
models*

Pilot 3:
Cancer Surveillance
*Computational insight
into factors impacting
clinical response*

Integrated Precision Oncology



Pilot 1 Pre-clinical Model Development

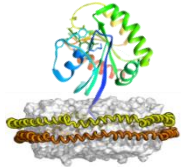


Aim 1: Predictive Models of Drug Response (signatures)

Aim 2: Uncertainty Quantification and Improved Experimental Design

Aim 3: Develop Hybrid Predictive Models

Pilot 2 RAS Therapeutic Targets

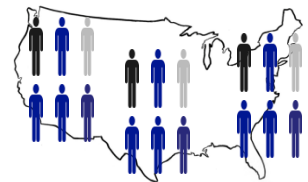


Aim 1: Adaptive time and length scaling in dynamic multi-scale simulations

Aim 2: Validated model for Extended RAS/RAS-complex interactions

Aim 3: Development of machine learning for dynamic model validation

Pilot 3 Precision Oncology Surveillance



Aim 1: Information Capture Using NLP and Deep Learning Algorithms

Aim 2: Information Integration and Analysis for extreme scale heterogeneous data

Aim 3: Modeling for patient health trajectories

Crosscut: CANDLE exascale technologies, uncertainty quantification

Integrated Precision Oncology



Crosscut: Integrated Precision and Predictive Oncology

Pilot 1 Pre-clinical Model Development

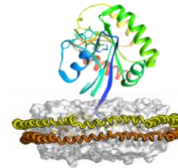


Aim 1: Predictive Models
of Drug Response
(signatures)

Aim 2: Uncertainty
Quantification and
Improved Experimental
Design

Aim 3: Develop Hybrid
Predictive Models

Pilot 2 RAS Therapeutic Targets

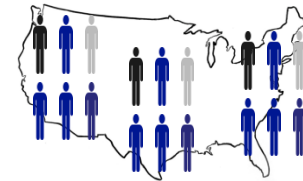


Aim 1: Adaptive time and
length scaling in dynamic
multi-scale simulations

Aim 2: Validated model for
Extended RAS/RAS-
complex interactions

Aim 3: Development of
machine learning for
dynamic model validation

Pilot 3 Precision Oncology Surveillance



Aim 1: Information Capture
Using NLP and Deep
Learning Algorithms

Aim 2: Information
Integration and Analysis for
extreme scale
heterogeneous data

Aim 3: Modeling for patient
health trajectories

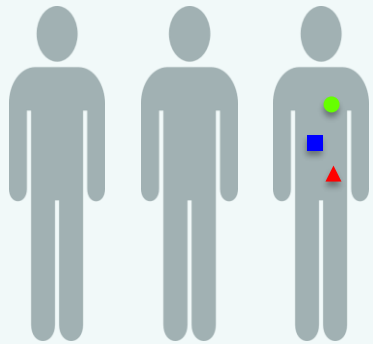
Crosscut: CANDLE exascale technologies, uncertainty quantification

Molecular Scale Pilot



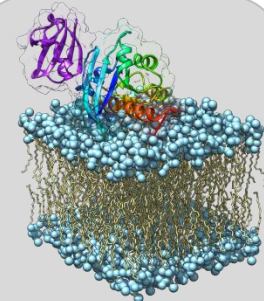
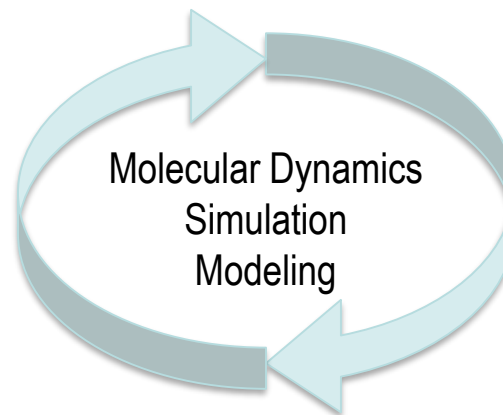
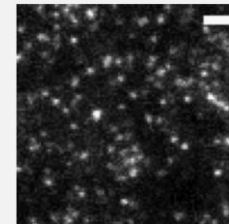
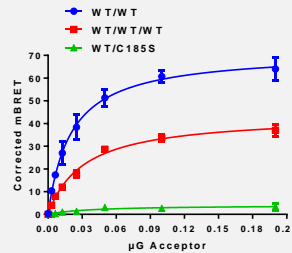
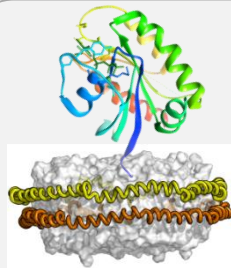
30% of cancers have mutated RAS

~1M deaths/year



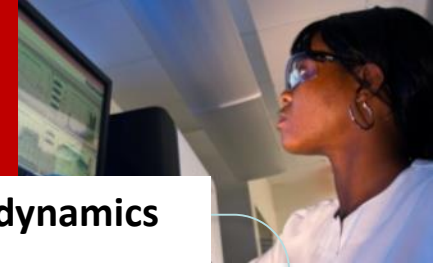
Current therapies ineffective against RAS-driven cancer

Facilitate discovery and development of novel therapeutics



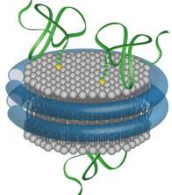
RAS biology
ID targets
New inhibitors



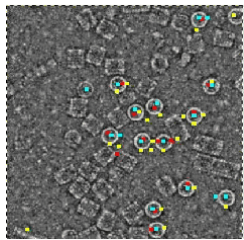


RAS proteins in membranes

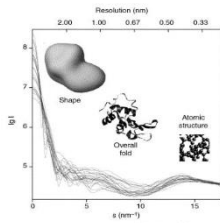
RAS activation experiments at NCI/FNL



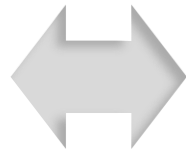
Experiments on nanodisc



CryoEM imaging



X-ray/neutron scattering



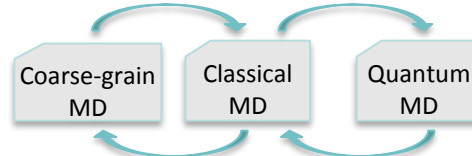
Multi-modal experimental data, image reconstruction, analytics

Protein structure databases



New adaptive sampling molecular dynamics simulation codes

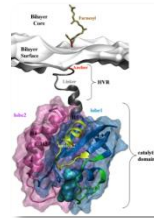
Adaptive time stepping



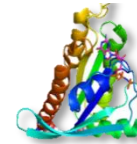
Adaptive spatial resolution

High-fidelity subgrid modeling

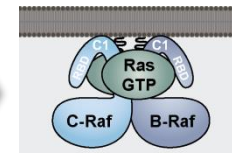
Predictive simulation and analysis of RAS activation



Granular RAS membrane interaction simulations

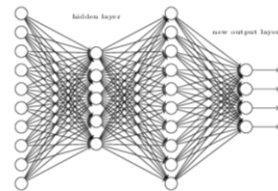


Atomic resolution sim of RAS-RAF interaction

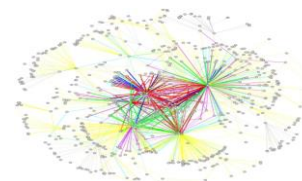


Inhibitor target discovery

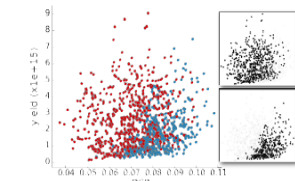
Machine learning guided dynamic validation



Unsupervised deep feature learning



Mechanistic network models



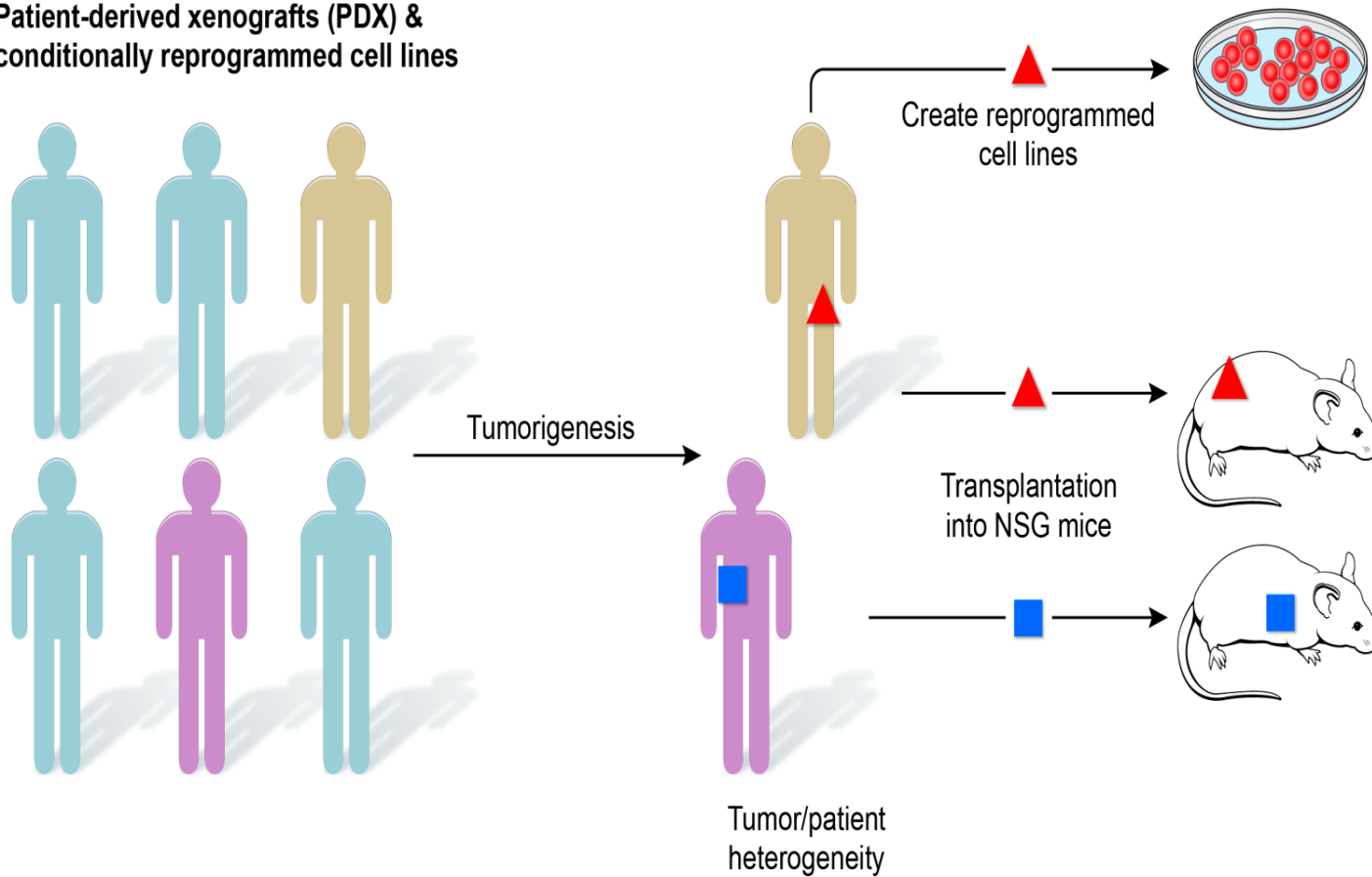
Uncertainty quantification



Pre-clinical Focused Pilot



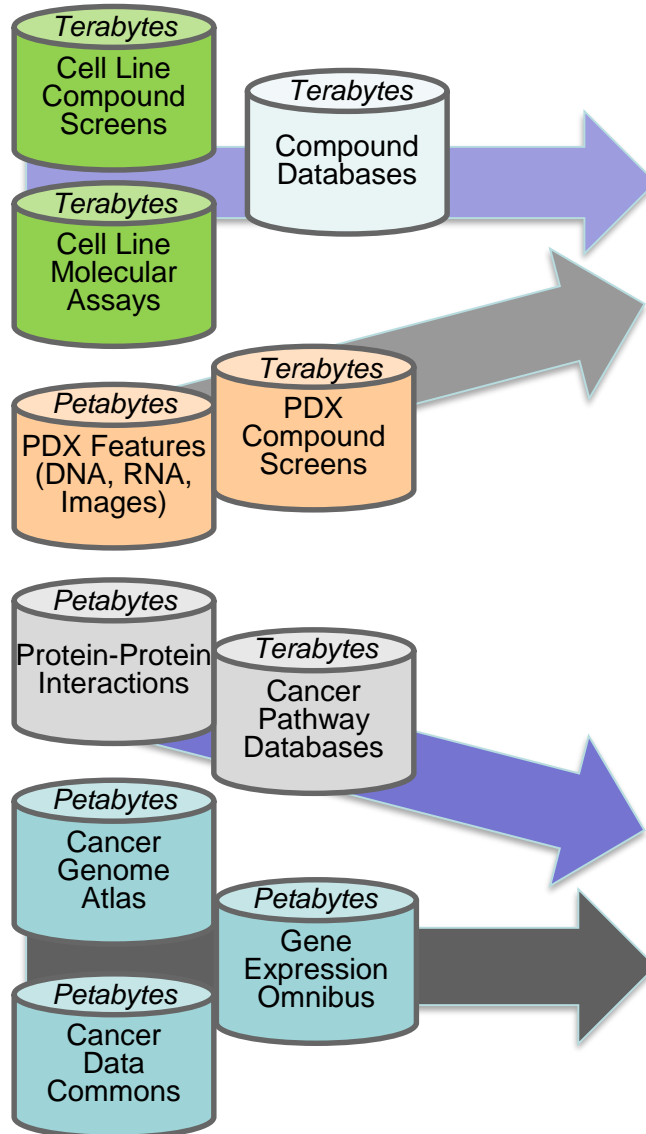
Patient-derived xenografts (PDX) & conditionally reprogrammed cell lines



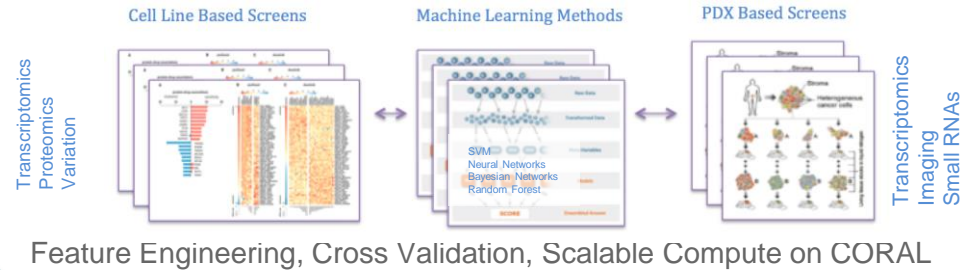
Molecularly characterize, treat/screen mice bearing transplants & cells with relevant drugs.

“Pre-clinical clinical trials”

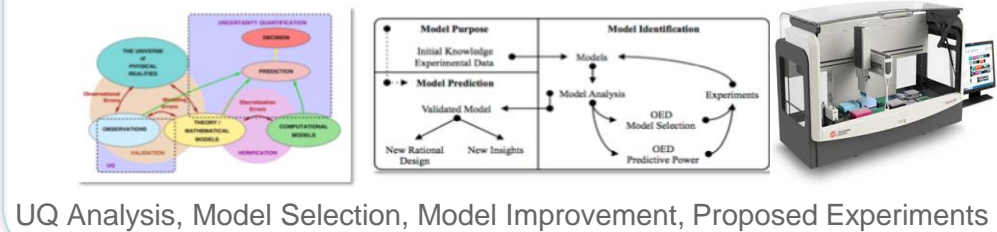
Predictive models for pre-clinical screening



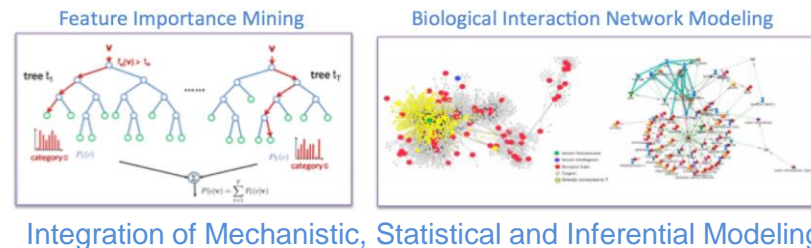
Machine Learning Based Predictive Models



Uncertainty and Optimal Experiment Design



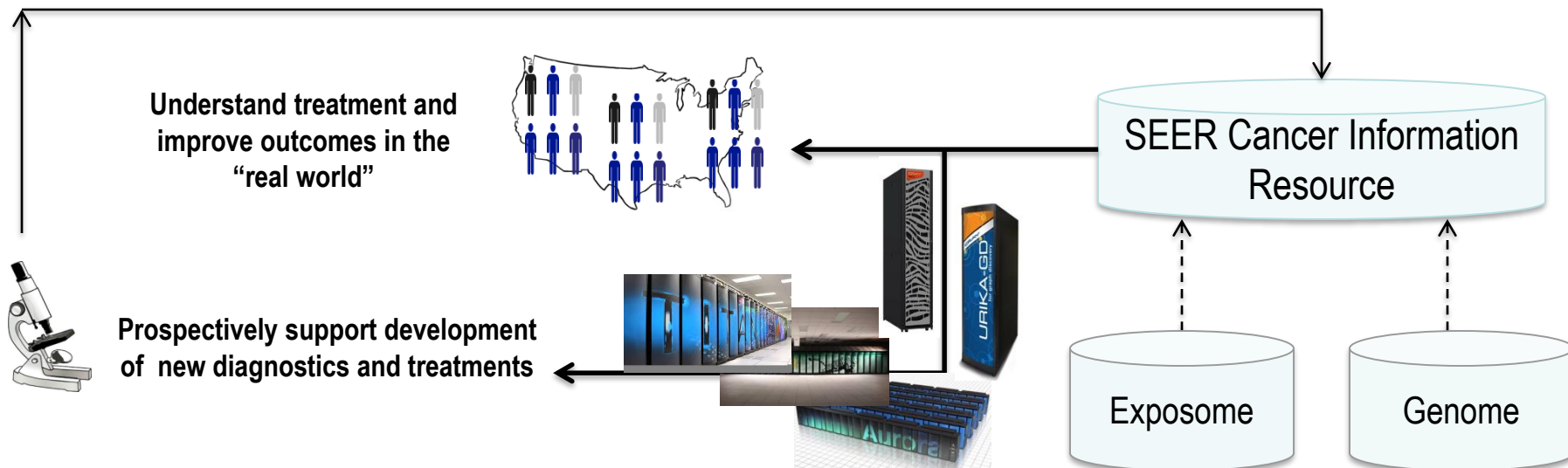
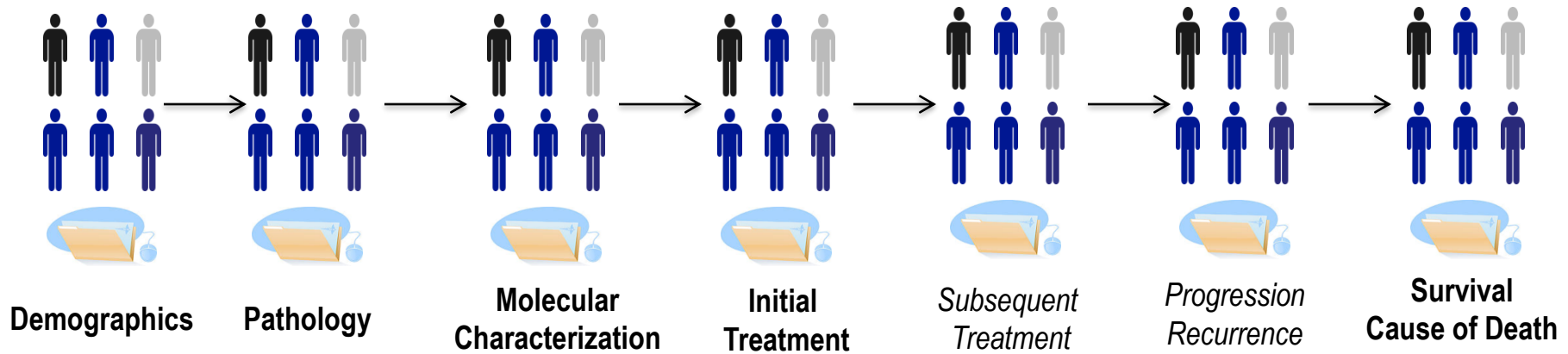
Hypotheses Formation and Mixed Modeling



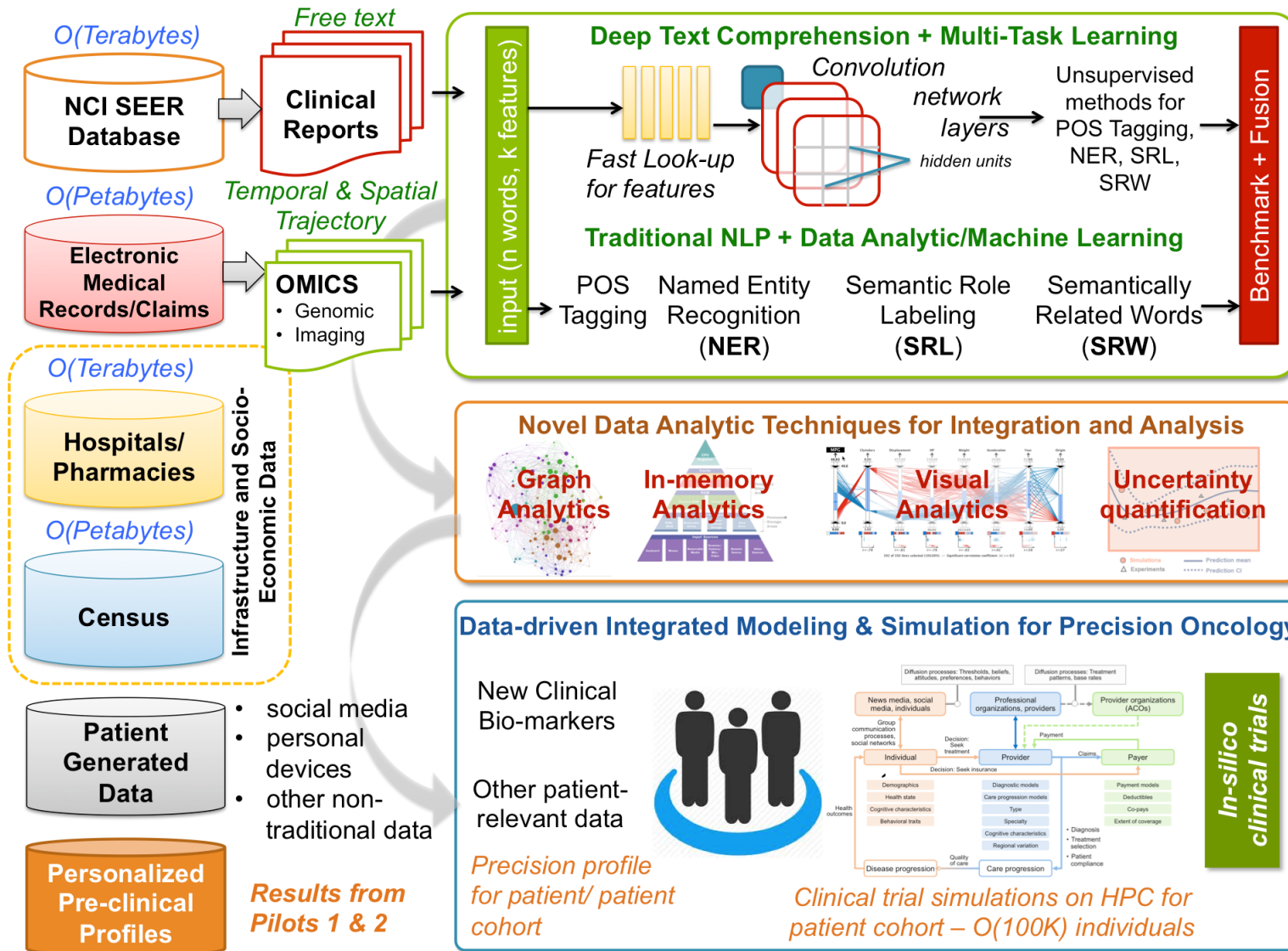
Population Focused Pilot



Surveillance data captured on each cancer patient for the entire population



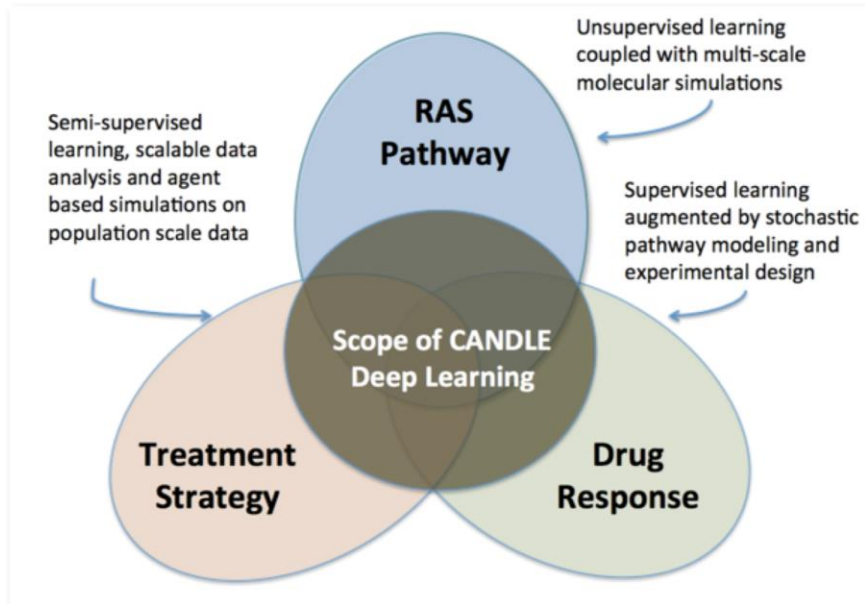
Population information integration, analysis and modeling



CANDLE – Deep Learning Across JDACS4C



ECP-CANDLE Project : CANcer Deep Learning Environment



- ### CANDLE Goals
- Develop an exscale deep learning environment for cancer
 - Building on open source Deep learning frameworks
 - Optimization for CORAL and exascale platforms
 - Support all three pilot project needs for deep learning
 - Collaborate with DOE computing centers, HPC vendors and ECP co-design and software technology projects



CANDLE – CANcer Distributed Learning Environment

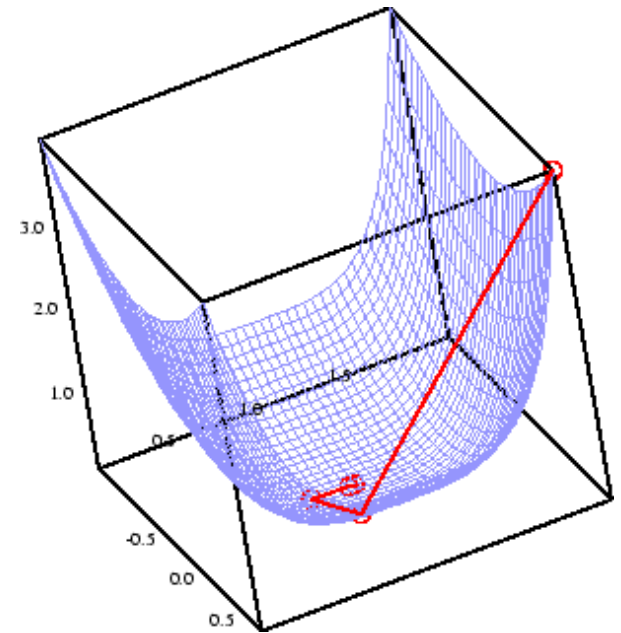


- CANDLE is DOE Funded contribution to JDACS4C
- Four year project (now in year 2)
- Focuses on creating scalable, open and portable Deep Learning framework
- Supports Deep Learning needs for all JDACS4C pilots
 - DOE scientific leads bring pilot-specific deep learning challenges
- Open source software release
- FNL brings NCI connection to CANDLE
 - Translating computational environment to broader cancer research community
 - Portability and standardization of model representations
 - Conventions and methods for model validation and evaluation in cancer

Optimization



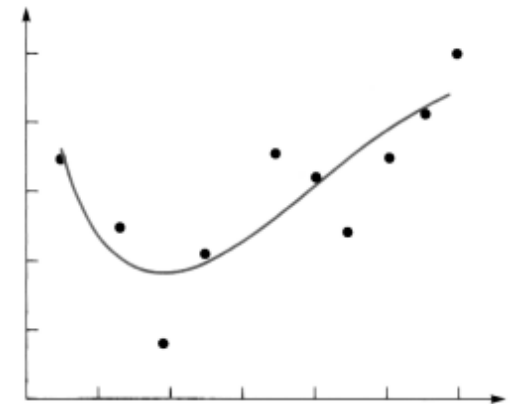
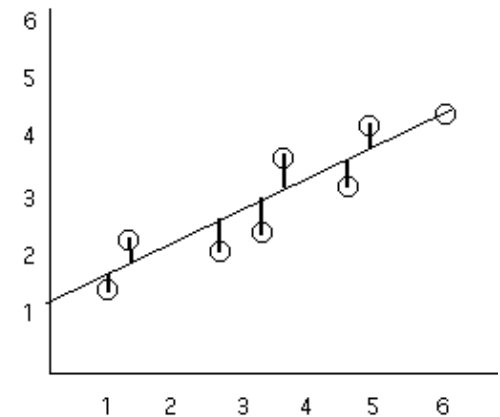
- Decision variables
 - Elements that can be adjusted to change the output of the objective functions
 - For example: $x_1, x_2, x_3, x_4, \dots, x_n$
- Objective functions
 - Functions that determine improved combination of decision variables
 - Objective = $F(x_1, x_2, x_3, x_4, \dots, x_n)$
- Constraints
 - Limits on possible values of decision variables



Least Squares Optimization



- Decision variables
 - Elements that can be adjusted to change the output of the objective functions
 - Choice of function
 - Linear: $y = ax + b$
 - Quadratic: $y = ax^2 + bx + c$
- Objective functions
 - Functions that determine improved combination of decision variables
 - Minimizing distance between $y_{\text{pred}}(x_i)$ and y_{actual} across all points



Machine Learning Models



- Prototype: Artificial Neural Networks
- Each node employs inputs, weights, and activation function to deliver outputs
- Employ multiple nodes and functions
- Creating a network among nodes

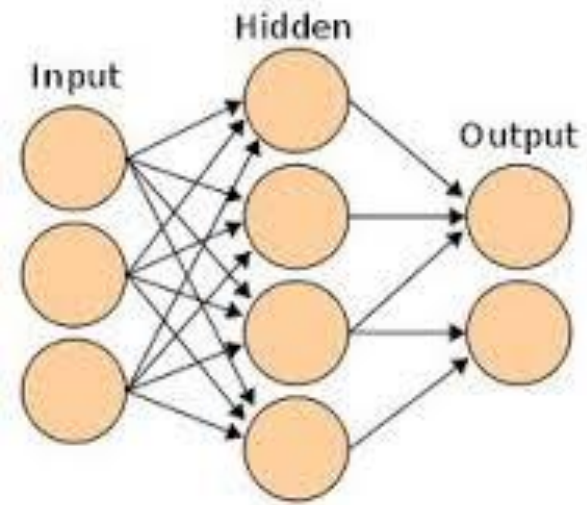
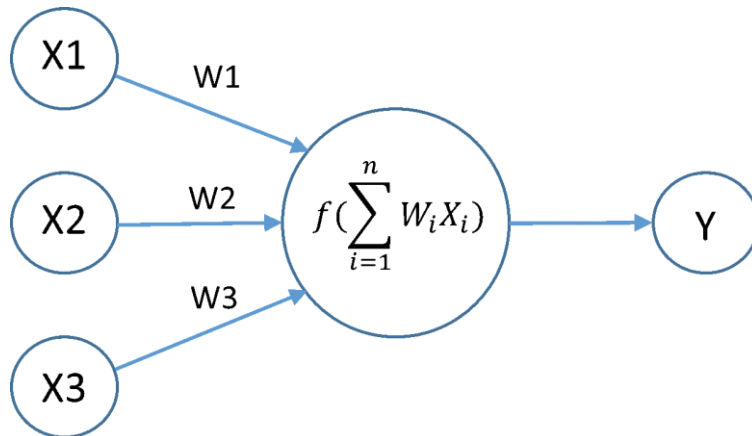
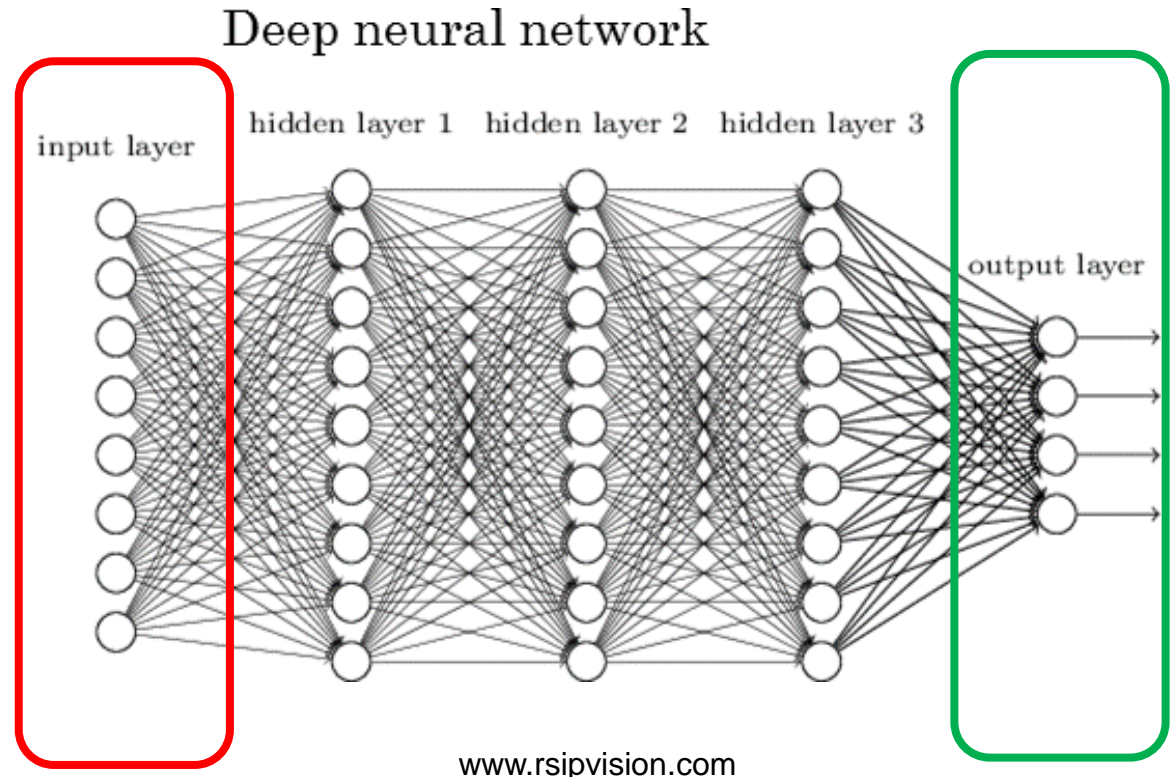


Image from mathworks.com

Deep Learning Models



- Large number of nodes
- Multiple layer configuration
- Enables combinations of non-linear functions
- Provides flexibility to detect subtle features



Deep Learning and Optimization

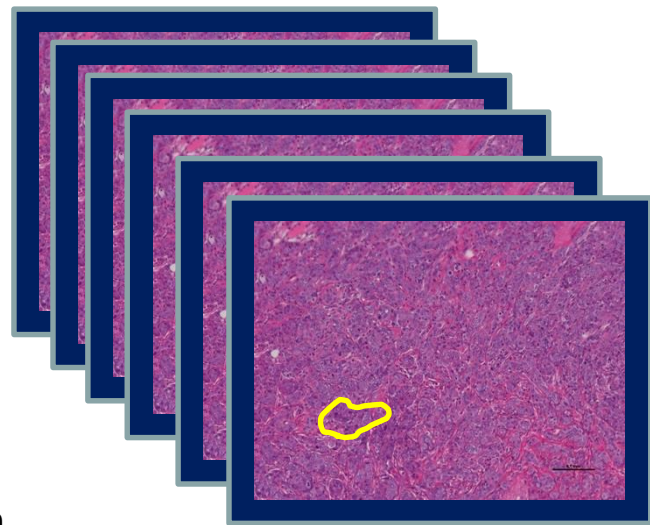


- Decision variables include
 - Connections within the network (topology)
 - Mathematical functions used in the network
 - Weights and parameters
 - Objective function
 - Minimize differences between actual outcomes and predicted outcomes
 - $Y_{\text{pred}} = \text{DL}(\text{input values})$
 - $\Sigma(y_{\text{pred}} - y_{\text{actual}})^2$
 - Constraints
 - Specific ranges and relationships enforced among decision variables
- } hyper parameters

Deep Learning



- Applications
 - Classification
 - Clustering
 - Prediction
- Basic data requirements
 - Relevant data
 - Consistent formatting – vector of input data
 - Well annotated/labeled for outcome
 - Accessible quantity
- Basic development process
 - Train – optimize decision variables to minimize differences between DL outcomes and actual outcomes across large datasets
 - Validate – evaluate relative to chosen data



Deep learning identifying skin cancer



- Paper in February 2, 2017 issue of Nature
- Matched performance of dermatologists
- Data
 - 130,000 visual images of skin cancer
 - Processed to harmonize data available from multiple sources
 - Developed annotation to differentiate types of cancer
 - Fed as raw pixels
- Leveraged algorithm used by Google to efficiently train
- Validated
 - using only high-quality, biopsy confirmed images
 - Compared results to human expert analysis
- Potential for algorithm to be deployed on smart phones

Recent applications of deep learning applied to cancer



Deep Learning in Drug Discovery
Wiley online, December 30, 2015

Deep learning algorithm does as well as dermatologists in identifying skin cancer
-Stanford News, January 25, 2017

Detecting Cancer Metastases on Gigapixel Pathology Images
-Cornell University Library, March 3, 2017

Deep Learning Drops Error Rate for Breast Cancer Diagnoses by 85%
-2016 Camelyon Grand Challenge

Accelerating cancer research with deep learning
-Phys.org, November 9, 2016

H&E-stained Whole Slide Image Deep Learning Predicts SPOP Mutation States in Prostate Cancer
bioRxiv, March 27, 2017

Deep learning application trial to lung cancer diagnosis for medical sensor systems
-IEEE 2016 International SoC Design Conference (ISOCC)

Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer
-Chemical Science, 2017

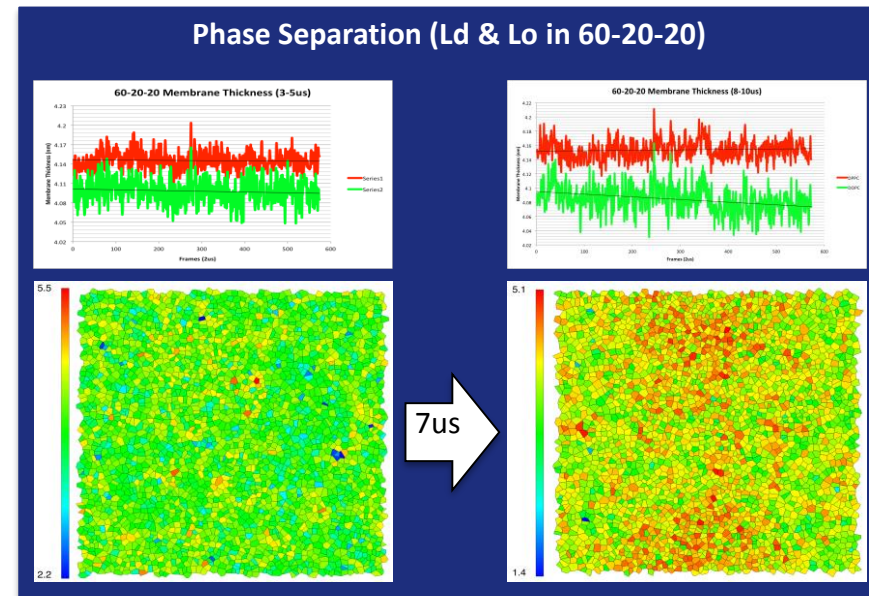
Challenge 2: RAS/RAF Pathway Prediction



Develop and test mechanistic models informed by cancer biology, structure and functional data that predict biophysical properties of normal and mutant proteins

Requires large-scale networks to train, and the generated models run on large-scale input data to produce actionable results:

- After completion of simulation scenarios, the DNN would be re-trained to select/determine the next set of exascale MD simulations to be run (considering $\sim 10^9$ alternatives)
- Neighborhoods in MD data sets will be defined to include $\sim 10^3$ particles and cover $\sim 10^3$ time steps producing an input vector dimension of $\sim 10^6$
- Large-scale MD simulations will provide a few billions of these neighborhood-based vectors as a training set for feature learning
- An MD data set with 10^6 particles and 10^9 time steps will provide $\sim 10^9$ neighborhood training vectors



JDACS4C collaboration

Frederick National Laboratory for Cancer Research

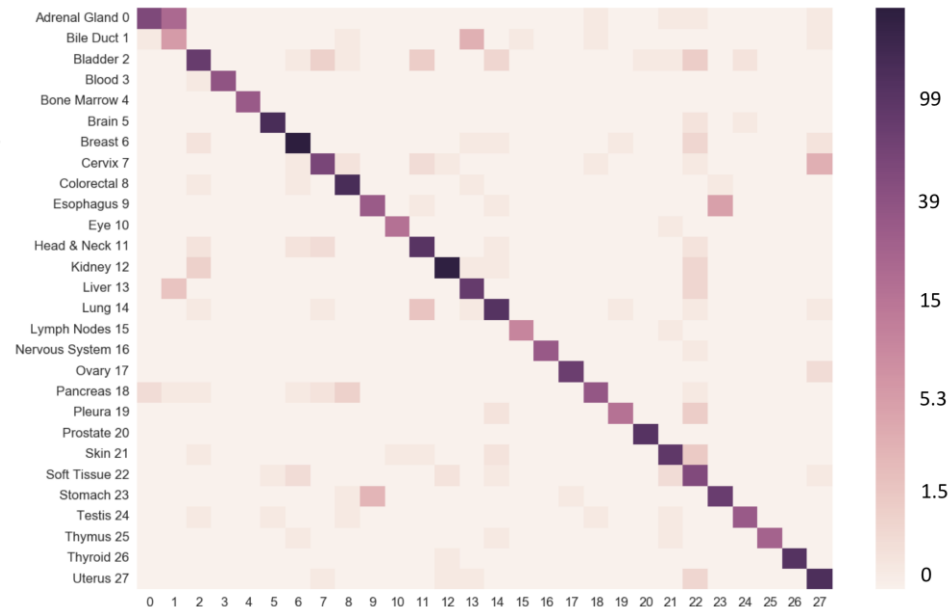
Challenge 1: Drug Response Prediction



Develop and test predictive models for patient outcomes based on clinical and pre-clinical genomic and molecular signatures

Requires exascale class simulation runs and data analysis coupled with the deep learning:

- After training the network on millions of previous screen results, the DNN would be used to search through up to 10^9 combinations of drugs to find those predicted to maximally inhibit a given tumor, or to search through 10^9 hypothetical compounds to identify those with potential as new drug development candidates
- For a given drug-tumor pair in a treatment, at least 10^7 features describe the input data; these features include 10^6 properties of the drug and 10^7 measurements of tumor molecular characteristics
- Access to $\sim 10^7$ samples of drug screening results, input data for supervised deep learning is a $10^7 \times 10^7$ mostly dense matrix (~ 1 PB) with redundancy



JDACS4C collaboration

Acknowledgements



- NCI CBIIT
 - Warren Kibbe, Tony Kerlavage, Carl McCabe
 - NCIP Cloud Pilot Team
 - Cancer Informatics Branch
 - Many more
- NCI Division of Cancer Control and Population Studies
 - Lynne Penberthy, Paul Fearn
- NCI Division of Cancer Therapeutics and Diagnostics
 - Jim Doroshow
- FNLCR
 - Data Science and Information Technology Program
 - Jack Collins (ABCC), Greg Warth (ITOG), Braulio Cabral (CBIIT Support), Megan Kaminiski (PMO)
 - Dwight Nissley, Yvonne Evrard, Frank McCormick(UCSF)
- Department of Energy
 - Dimitri Kusnezov
- Lawrence Livermore National Laboratory
 - Amy Gryshuk, Jim Brase, Jason Paragas, David Rakestraw, Fred Streitz, Felice Lightstone, Ken Turteltaub, Brian van Essen
- Argonne National Laboratory
 - Rick Stevens
- Los Alamos National Laboratory
 - Frank Alexander
- Oak Ridge National Laboratory
 - Gina Tourassi
 - Gil Weigand
- And many, many more

Workshop Aims



- Overview of NCI and DOE collaborative efforts
- Grow community around deep learning applied to cancer and CANDLE (CANcer Distributed Learning Environment)
- Identify priority areas to explore challenges and opportunities
 - Deep learning in cancer research today
 - Sources of cancer data today and the future
 - Validating, sharing and extending deep learning models

Questions



- Discussion.
- Contact info:
 - Eric.Stahlberg@nih.gov

Presentation Overview



- **Common Thread**

- Support efforts to establish high-performance computing foundations for accelerated predictive oncology and cancer precision medicine

- **Context**

- Frederick National Laboratory
 - Data Science and IT Program
 - Precision Oncology
 - Computational and Data Challenges

- **NCI-DOE JDACS4C Pilots**

- Motivations
 - Pilot 1 – Molecular domain
 - Pilot 2 – Preclinical domain
 - Pilot 3 – Population domain

- **Broader Engagement**

- CANDLE – CANcer Distributed Learning Environment
 - Workshops

Overview of Frederick National Laboratory for Cancer Research (FNLCR)



- **FNLCR is the only Federally Funded Research and Development Center (FFRDC) dedicated exclusively to biomedical research**
 - Operated in the public interest by **Leidos Biomedical Research, Inc** (formerly SAIC-Frederick) on behalf of the National Cancer Institute
- **Main campus located on 70 acres at Ft. Detrick, MD**
 - Leidos Biomed employees co-located with NCI researchers and other contractors on the NCI Campus at Frederick
 - Additional Leidos Biomed scientists at Bethesda and Rockville sites



Mission

Provide a unique national resource for the development of new technologies and the translation of basic science discoveries into novel agents for the prevention, diagnosis and treatment of cancer and AIDS.

Research & Development at FNLCR



- **Research & Development**

- **Basic Research:** New knowledge about AIDS and cancer
- **Applied R&D:** New diagnostics and therapeutics
- **Clinical Research:** Clinical trials and laboratory analysis
- **cGMP manufacturing:** Biologicals and vaccine production



- **Specialties**

- Genomics, proteomics, and metabolomics
- Bioinformatics and imaging
- Nanotechnology
- Animal models
- Tumor cell biology and virology
- Immunology and inflammation



- **Data Science and Information Technology**
underpin and support all R&D activities and specialties

Data Science and Information Technology Program



Mission: Leverage leading edge data science and information technology skills, tools, and capabilities to accelerate translation of biomedical data to scientific discoveries, medical treatments, diagnostic and prevention tools for cancer and AIDS patients.



Data

Analyze



Insight

Decide



Action

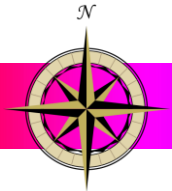
Enabling physicians, scientists, and patients to make critical decisions based on knowledge gained from all and not only a fraction of data and information available to them.



A New Era for Data and Computational Science Applied to Cancer

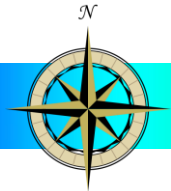


Pilot 2 - RAS



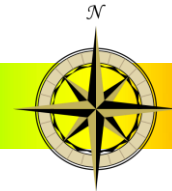
Fundamental
Biology

Pilot 1 - PDX



Pre-clinical

Pilot 3 - SEER



Clinical Use

- Three pilots in 3 priority areas
- Efforts inform future data and computing science needs
- Efforts beyond pilots needed to extend and accelerate across the complex cancer research and clinical spectrum
 - New technologies to probe and observe cancer at all scales, conditions, types, contexts
 - Much more larger, different and newer data

Getting Connected



- CANDLE Workshop at NIH
 - April 18-19, 2018
- GPU Technology Conference
 - May 9, 2017 – CANDLE workshop
- Precision Medicine Workshop
 - June 22, 2017 - First workshop scheduled for ISC17
- Computational Approaches for Cancer Workshop (CAFCW)
 - November 2017 Third workshop proposed for SC17

NCI Precision Oncology Extending the Frontiers



- Identify promising new treatment options through the use of advanced computation to rapidly develop, test and validate predictive pre-clinical models for precision oncology.
- Deepen understanding of cancer biology and identify new drugs through the integrated development and use of new simulations, predictive models and next-generation experimental data.
- Transform cancer care by applying advanced computational capabilities to population-based cancer data to understand the impact of new diagnostics, treatments and patient factors in real world patients.

Predictive Models



Precision Oncology and Computing



- **Motivations**

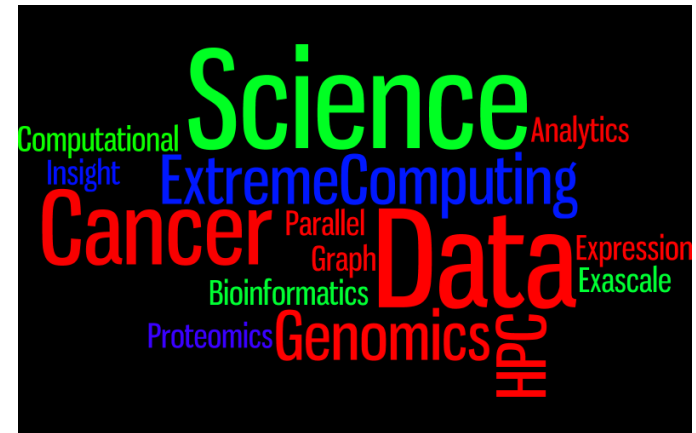
- Derive value from growing volumes, types and sources of data and information
- Expanding support for growing data and computational science needs and opportunities
- Essential to future mission success

- **Challenges**

- Limited resources
- Limited expertise in new areas
- Rapidly growing data requirements

- **Aims**

Support efforts to establish high-performance computing foundations for predictive oncology and cancer precision medicine



Emerging Opportunities



- Drug and treatment discovery
 - ATOM - Accelerating Therapeutics for Opportunities in Medicine
- Increased use of imaging based observations in research
 - Molecular scale observations
 - Pre-clinical domain
- Increased use of imaging in clinical settings
 - Diagnostics
 - Non or minimally invasive post-treatment monitoring
- Integration of imaging modalities and fusion of information sources
- Predictive/analytic model verification, evaluation and validation