NCI-DOE Pilot and Precision Medicine Initiative in Oncology

Warren Kibbe, PhD

NCI Center for Biomedical Informatics and Information Technology

NIH NATIONAL CANCER INSTITUTE

September 30th, 2015

1. *Precision Medicine*
2. *NIH HPC*
3. *NCI-DOE pilot*
4. *PMI-O Informatics Goals*
5. *PMI-O Genomic Data Commons*
6. *PMI-O Cloud Pilots*

Slides are from many sources, but special thanks to
Drs. Harold Varmus, Doug Lowy, Jim Doroshow, Lou Staudt

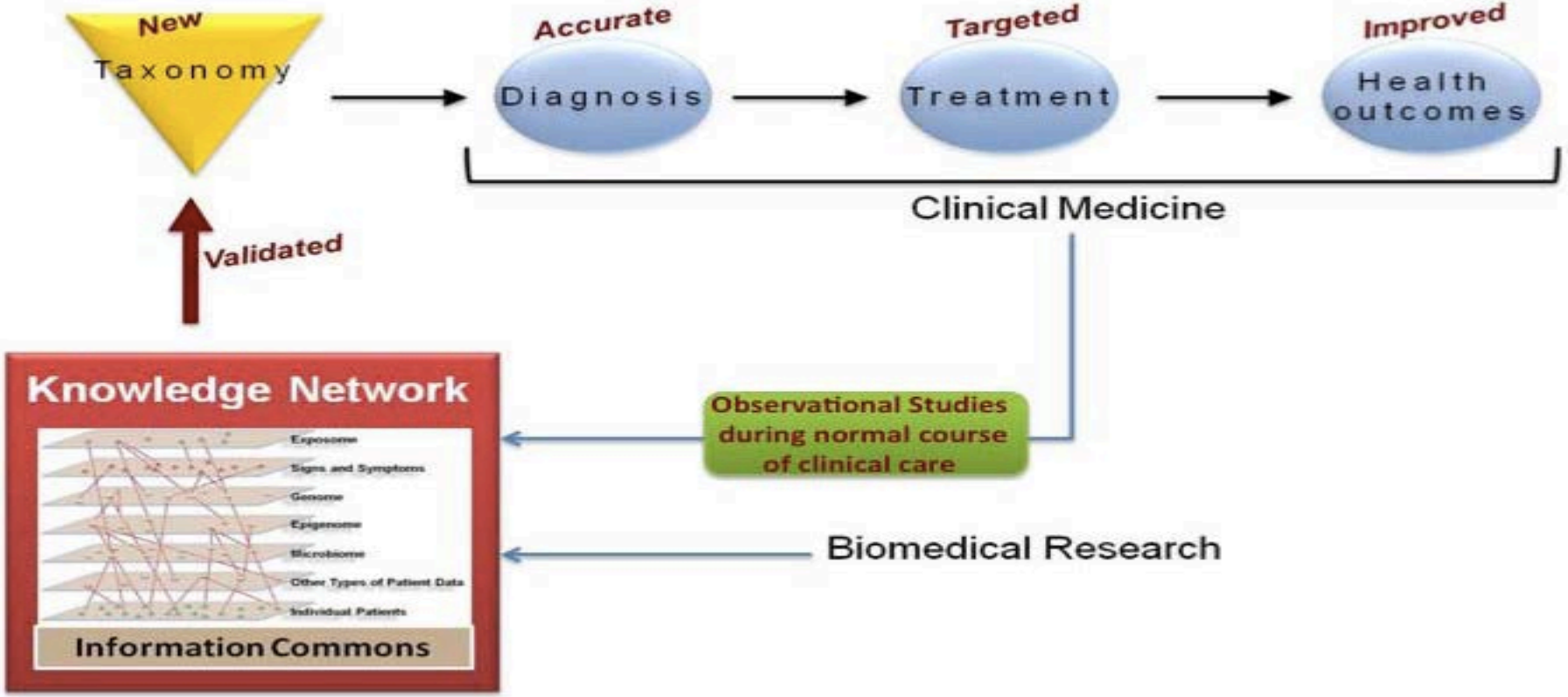
President Obama Announces the Precision Medicine Initiative

The East Room, January 30, 2015

Photo by F. Collins

TOWARDS PRECISION MEDICINE

(IoM REPORT, NOVEMBER 2011)



Definition of Precision Oncology

- Interventions to prevent, diagnose, or treat cancer, based on a molecular and/or mechanistic understanding of the causes, pathogenesis, and/or pathology of the disease. Where the **individual characteristics** of the patient are sufficiently distinct, interventions can be concentrated on those who will benefit, sparing expense and side effects for those who will not.

Modified by D. Lowy, M.D., from IoM's Toward Precision Medicine report, 2011

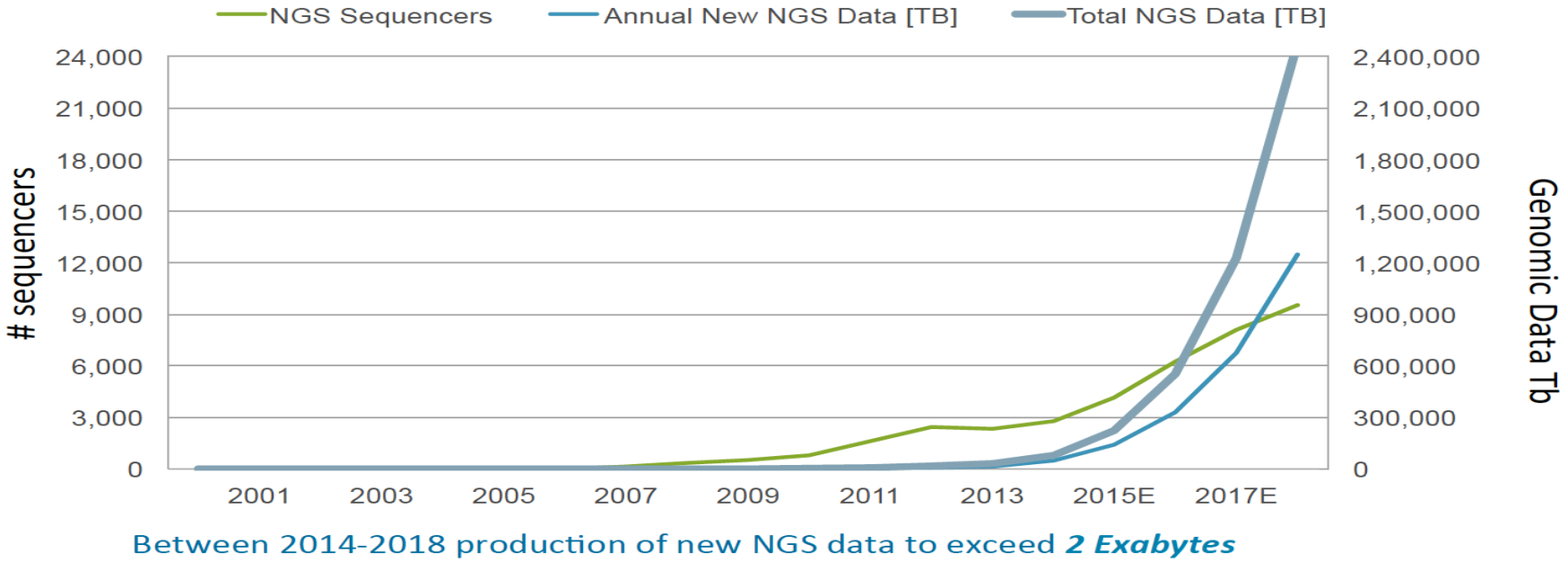
Understanding Cancer

- **Precision medicine** will lead to **fundamental understanding** of the complex interplay between genetics, epigenetics, nutrition, environment and clinical presentation and **direct effective, evidence-based prevention and treatment**.

Drivers for High Performance Computing and Computational Modeling

We need sophisticated computational models to understand patient response, methods of resistance, and to integrate pre-clinical model data

Amount of genomic data will exceed available resources



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.

Slide courtesy of Deniz Kural, Seven Bridges Genomics

NIH High Performance Computing Working Group

- Andy Baxevanis, NHGRI, Chair of the Working Group
- From the Biowulf team, Steve Bailey and Steve Fellini
- Vivien Bonazzi, OD/ADDS
- Bernie Brooks, NHLBI
- Sean Davis, NCI
- Yang Fann, NINDS
- Susan Gregurich, NIGMS
- Warren Kibbe, NCI
- Don Preuss, NCBI
- Mike Tartakovsky, NIAID
- Andrea Norris and Renita Anderson, Office of the CIO

NCI High Performance Computing Group

Cross-organizational participation for HPC directions

CBIIT
Warren Kibbe
Carl McCabe
Kelly Lawhead

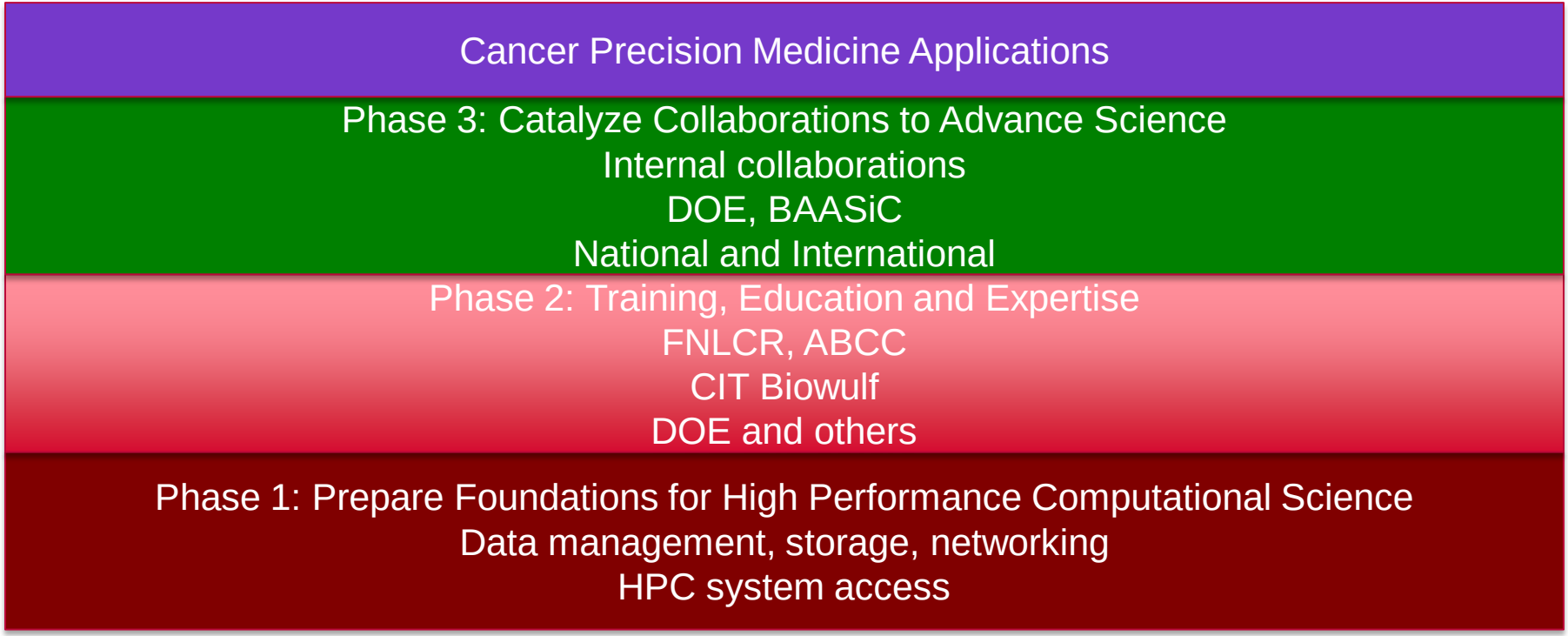
NCI-OSO
Dianna Kelly
Jim Cherry

NCI-CCR
Sean Davis

FNLCR
Nathan Cole (DCEG)
Jack Collins (ABCC)
Xinyu Wen (CCR)
Greg Warth (ITOG)
Eric Stahlberg (CBIIT)

CIT
Steve Fellini

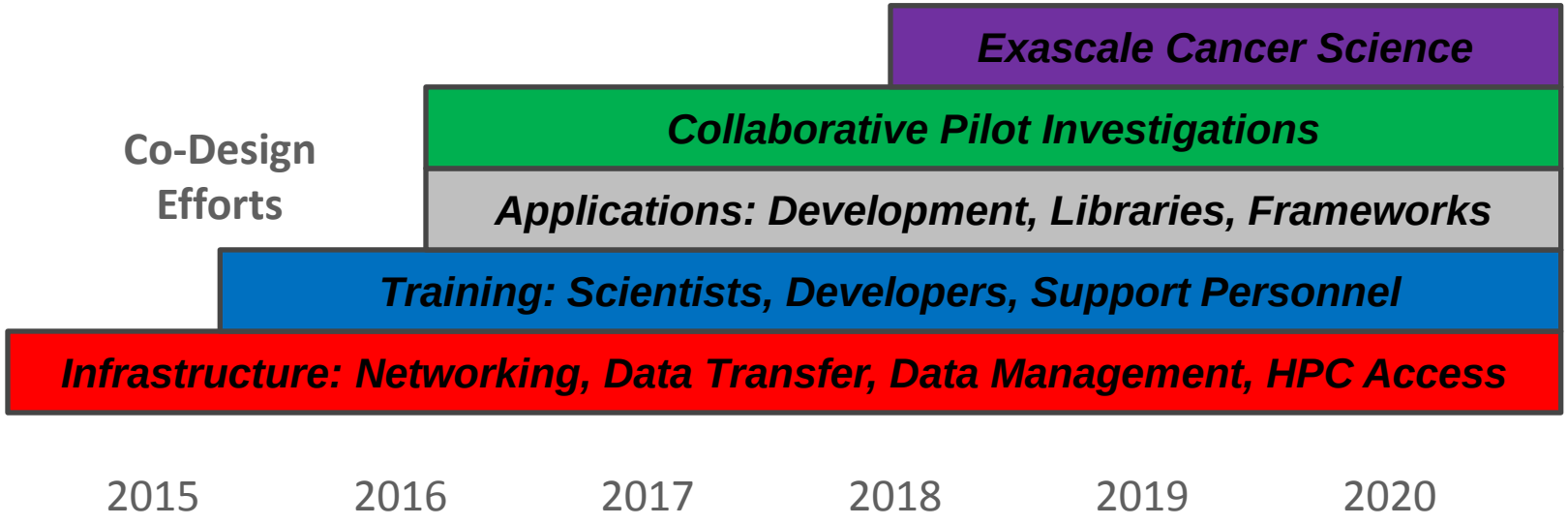
Key Points for NCI



Preparing for Exascale Cancer Science

Exascale in a nutshell:

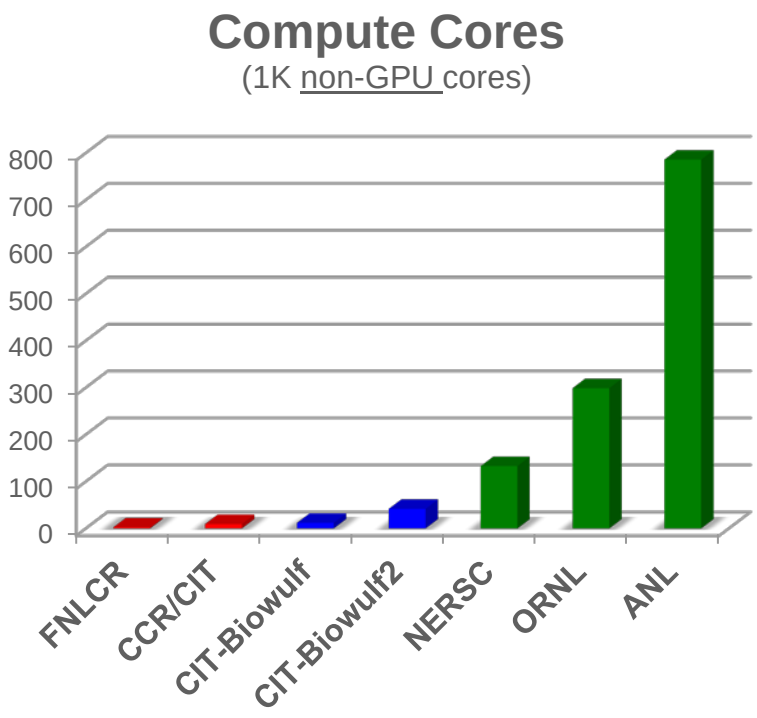
- Millions of CPU cores contributing to a single task
- Nearly 1000 times faster than fastest computer today
- Focus of DOE Advanced Strategic Computing



Cancer and Exascale Computing

- Motivations
  - Expanding options for cancer precision medicine
  - Promising new computing technologies
  - Understanding basic mechanisms of cancer
- NCI
  - Extensive domain experience in cancer
  - Vast amounts of new data to provide key insights
- DOE
  - World leading HPC systems
  - Extensive experience in complex predictive modeling

US Department of Energy – Leaders in Computing



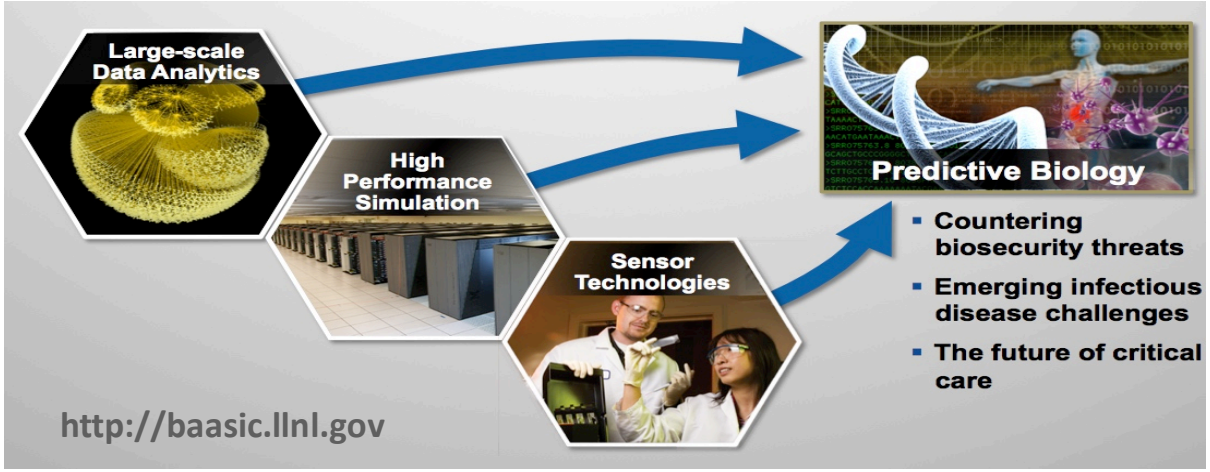
US Department of Energy

- Extreme scale systems
- Network Innovations
- $478M in FY14 into advanced computing research
- Lead for Exascale Computing Initiative

BAASiC

Biological Applications of Advanced Strategic Computing

- **Predictive**
- **Physiology**
- **Pharmacology**
- **Pathophysiology**
- **Pathogen biology**



Lawrence Livermore National Laboratory

Sandia National Laboratories

Argonne NATIONAL LABORATORY

NCI

THE UNIVERSITY OF CHICAGO MEDICINE & BIOLOGICAL SCIENCES

EMORY UNIVERSITY SCHOOL OF MEDICINE

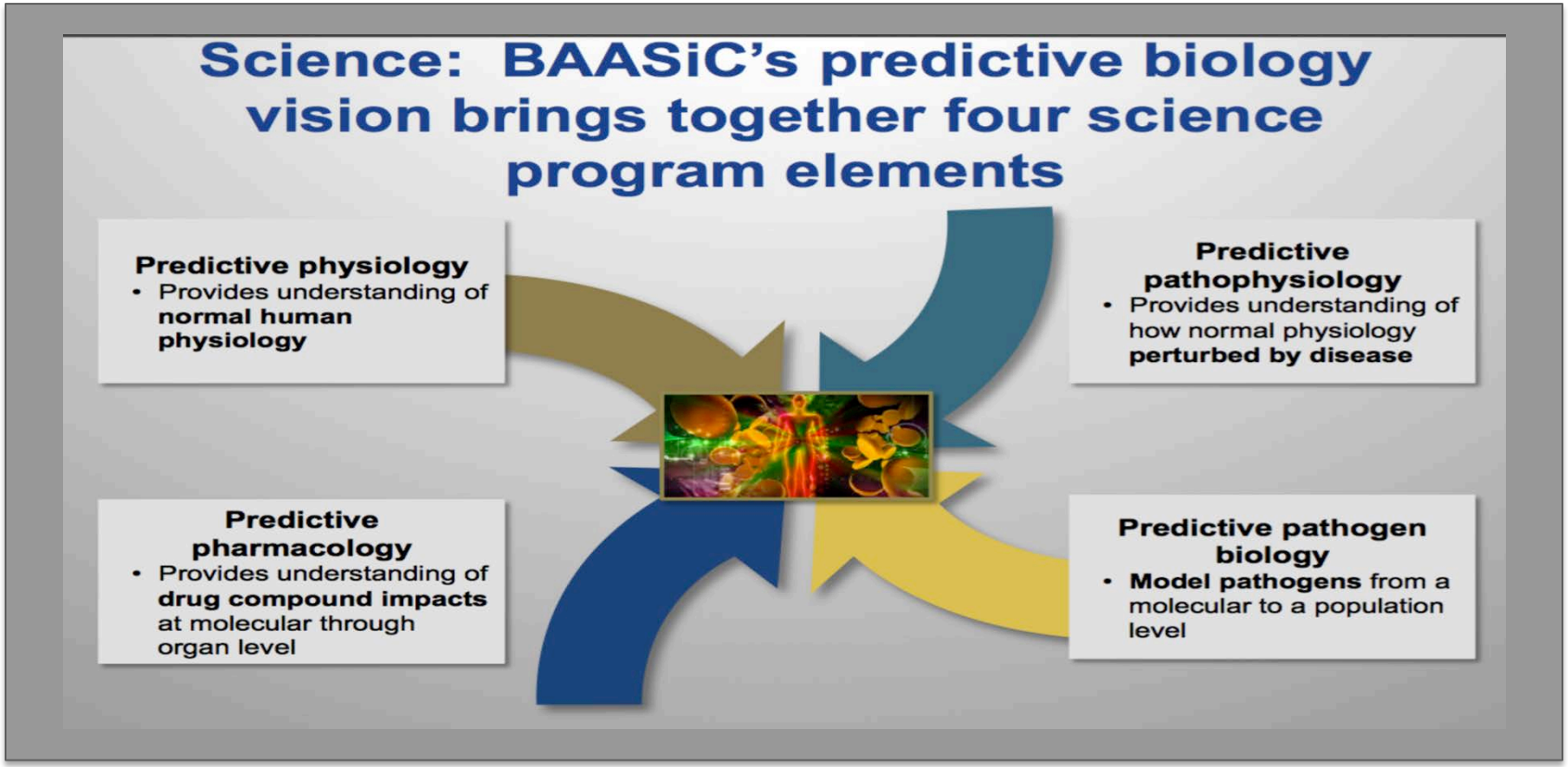
HARVARD MEDICAL SCHOOL

UCSF

- Livermore led consortium
- Driving DOE Exascale advances in computing
- Specifically interested in cancer applications

- NCI/LBR target roles
- Cancer expertise and essential data
- Models, frameworks, “collaboratorium”

BAASiC R&D Framework



National Strategic Computing Initiative

Backdrop for the NCI and DOE pilot

DOE National Labs and FNLCR

National Strategic Computing Initiative

- Executive Order announced July 29, 2015
- Create a cohesive, multi-agency strategic vision and Federal investment strategy in high-performance computing (HPC)
- Lead agencies
  - **DOE**, DoD and NSF
- Deployment agencies
  - NASA, **NIH**, DHS, and NOAA
  - Participate in shaping future HPC systems to meet aims of respective missions and support workforce development needs
- Implications for **NCI**
  - Work cross agency with DOE and others to expand use of HPC to advance research and clinical applications impacting cancer

Status of NCI-DOE efforts aligned with NSCI

Three candidate pilot projects identified:

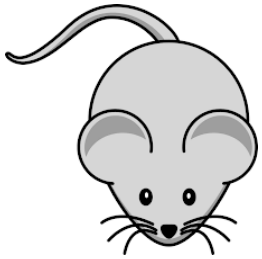
- Pre-clinical Model Development and Therapeutic Evaluation (Doroshow)
- Improving Outcomes for RAS Related Cancers (McCormick)
- Information Integration for Evidence-based Cancer Precision Medicine (Penberthy)

Collaboratively developing project plans with DOE computational scientists

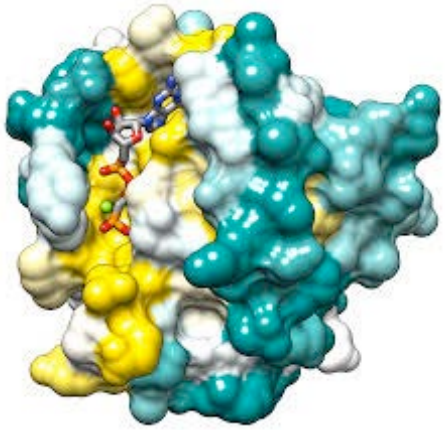
Plan definitions targeted by mid October 2015

Pilot Project 1: Pre-clinical Models

- Pre-clinical Model Development and Therapeutic Evaluation
- Scientific lead: Dr. James Doroshow
- Key points:
  - Rapid evaluation of large arrays of small compounds for impact on cancer
  - Deep understanding of cancer biology
  - Development of *in silico* models of biology and predictive models capable of evaluating therapeutic potential of billions of compounds

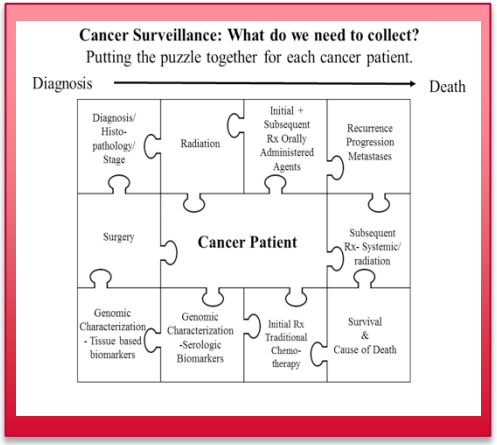


Pilot Project 2: RAS Related Cancers



- Improving Outcomes for RAS Related Cancers
- Scientific lead: Dr. Frank McCormick
- Key points:
  - Mutated RAS is found in nearly one-third of cancers, yet remains untargeted with known drugs
  - Advanced multi-modality data integration is required for model development
  - Simulation and predictive models for RAS related molecular species and key interactions
  - Provide insight into potential drugs and assays

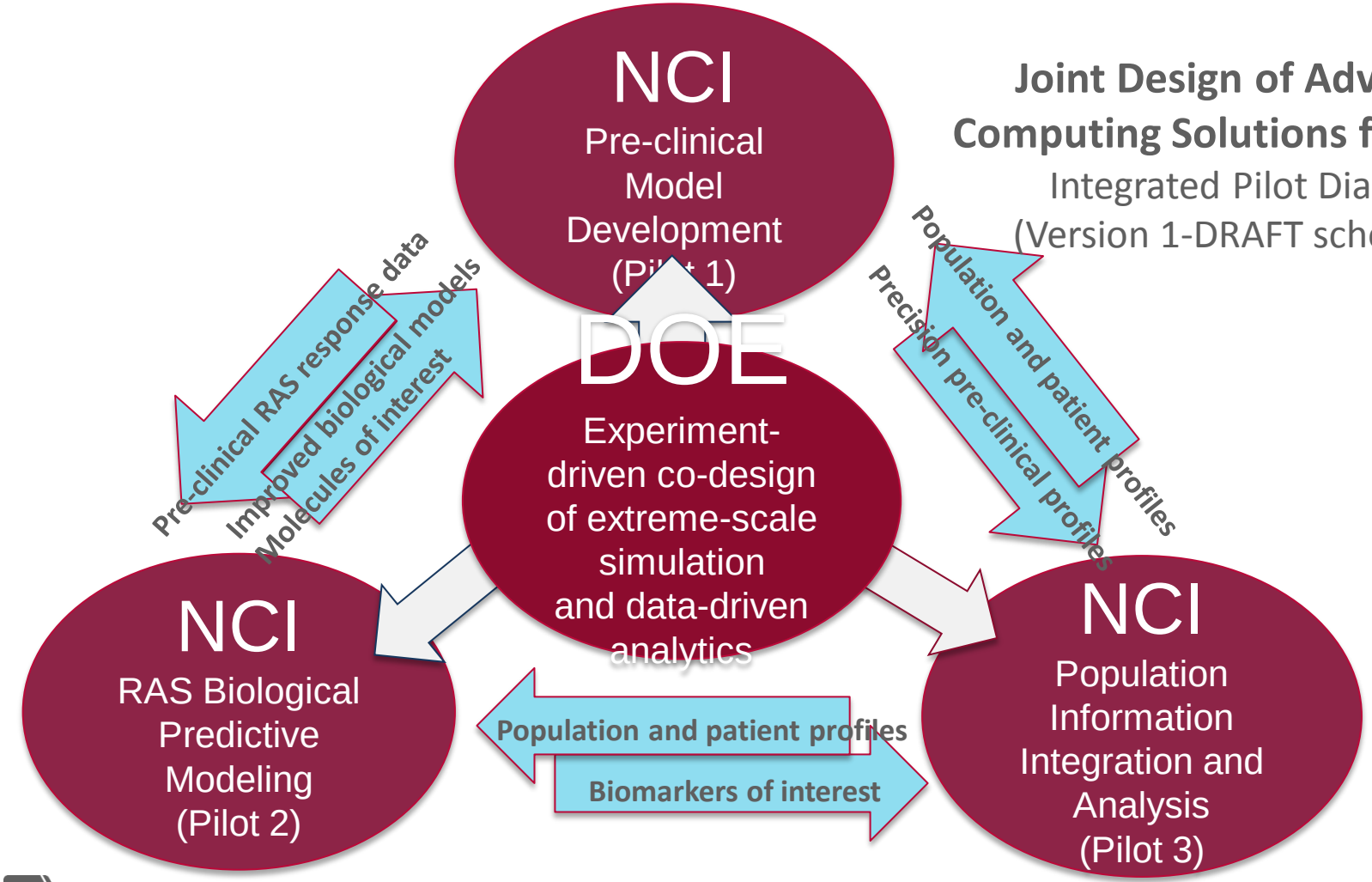
Pilot Project 3: Evidence-based Precision Medicine

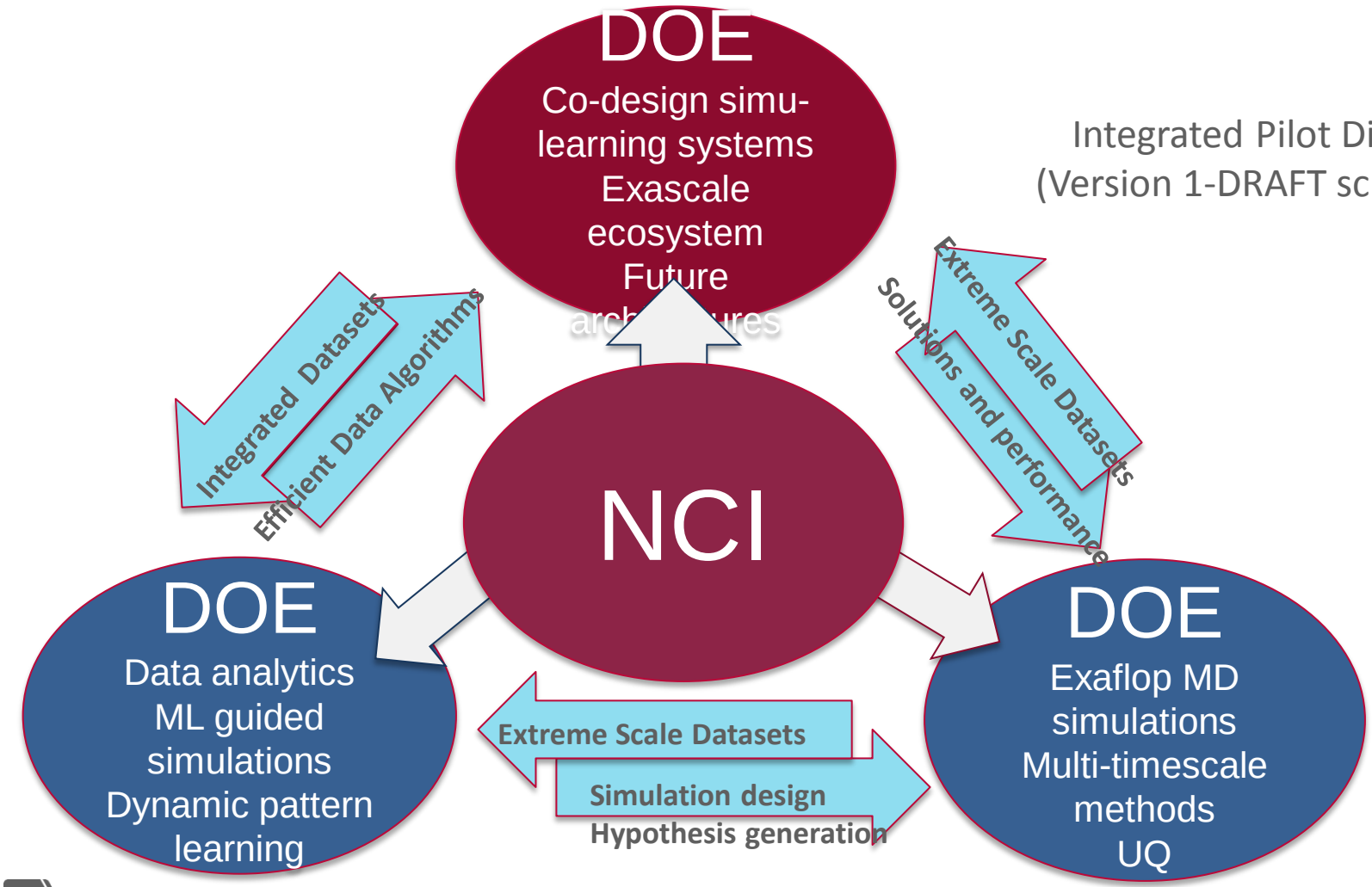


- Information Integration for Evidence-based Cancer Precision Medicine
- Scientific lead: Dr. Lynne Penberthy
- Key points:
  - Integrates population and citizen science into improving understanding of cancer and patient response
  - Gather key population-wide data on treatment, response and outcomes
  - Leverages existing SEER and tumor registry resources
  - Novel avenues for patient consent, data sharing and participation

Joint Design of Advanced Computing Solutions for Cancer

Integrated Pilot Diagram
(Version 1-DRAFT schematic)





Integrated Pilot Diagram
(Version 1-DRAFT schematic)

Pilot 1: Predictive Models for Preclinical Screening

Modeling framework established

Machine learning based predictors

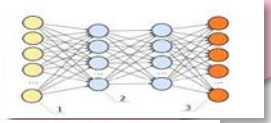
Patient derived cell lines and xenographs

Populating CL and PDX database with 1000's of samples

Project CL screen results to PDX models

Optimal predictor design using CORAL

Coupling ML models to biological interactions



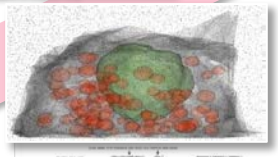
Testing hypotheses generated by models

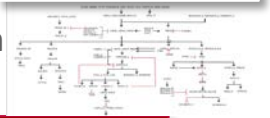
Adding imaging data to molecular assays

Coupling machine learning and mechanism

Automated Imaging based phenotype assessment

Multiscale cancer pathway models with inhibitors





Modeling supported N of 1 trial demo

Patient specific inhibitor proposed

Exascale search for optimal models enabling N of 1 trials

DOE

NCI

FY16 FY17 FY18

Pilot 2: RAS proteins in membranes

Multi-scale MD methods in time and space

Extreme scale MD/QMD targeting CORAL

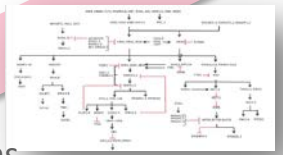
Predictive simulations driven by machine learning

Exascale class molecular simulations of cancer mechanisms and inhibitors

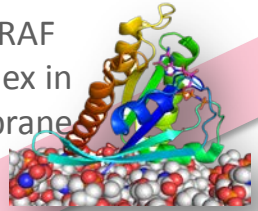
Context sensitive resolution switching

Mulit-modal inference tools

Extended protein complex interactions



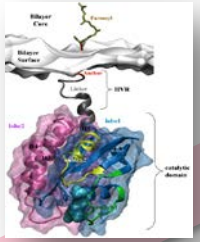
RAS-RAF complex in membrane



Inhibitor target discovery

Scalable coarse grain MD and QMD with UQ

RAS proteins in membrane



Membrane composition effects

Variation of RAS activation pathways

Dynamic RAS-RAF binding models

RAS Inhibitor hit identification computed on CORAL

Signaling process; inhibitor evaluation

Structure and dynamics of RAS in membrane

DOE

NCI

FY16 FY17 FY18

Pilot 3: Population Information Integration and Analysis

Modeling framework and predictive simulations of patient health trajectories

Automated hypothesis generation

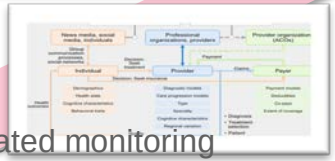
Exascale modeling and simulation of lifelong health trajectories

Scalable platform for multi-modal data integration

Machine learning for deep text comprehension at scale

Graph, visual, and in-memory heterogeneous data analytics and inference methods

Context-sensitive UQ for computational linguistics



Automated monitoring and modeling of disease patterns, care pathways, and outcomes

Data-driven knowledge discovery ecosystem

Text processing and reasoning algorithms

Data and compute infrastructure for optimized selection of precision medicine interventions and prediction of optimal patient care paths

Infrastructure for collection of multi-modal biomarkers to support cancer surveillance

Generalizable algorithms and architectures for unstructured text comprehension at scale

Dynamically adaptive predictive models of cancer progression, recurrence, outcomes

DOE

NCI

FY16 FY17 FY18

The NCI-DOE partnership will extend the frontiers of DOE computing capabilities

In simulation

- Atomic resolution MD simulations of critical protein complex interactions that will require exaflops of floating point performance
- New integrations of QM and multi-timescale methods that enable high accuracy interactions over extended time windows
- Extended theory and tools for UQ in multiple spatial and temporal scales

In data analytics

- Learning dynamic patterns from molecular to population scale data sets on CORAL-class architectures
- Integrated machine learning and simulation systems that bring together mechanistic and probabilistic models

In new computing architectures

- Co-design of architectures integrating learning systems and simulation in new memory memory-intensive hierarchies
- Growth of new computing ecosystems bringing together leadership-class HPC and cloud based data systems
- Integration of beyond Von Neumann architectures into mission workflows

PMI-O, NSCI and the DOE-NCI pilot

Precision Medicine Initiative in Oncology informatics and computational goals

PMI-O: **Informatics Goal**

*Develop a **Cancer Knowledge System**. Establish a national database that integrates genomic information with clinical response and outcomes as a resource.*

PMI-O: **Informatics Goal**

*Develop molecular, imaging, pathology, and clinical **signatures that predict therapeutic response**, outcomes, and tumor resistance*

PMI-O: **Informatics Goal**

*Build **multi-scale, predictive** computational biology **models** for **understanding cancer biology** and informing therapy. Develop detailed **cancer pathway models** to create **targeted combination therapies** in cancer. This approach has transformed HIV therapy and has the potential to do the same in cancer*

Genomic Data Commons

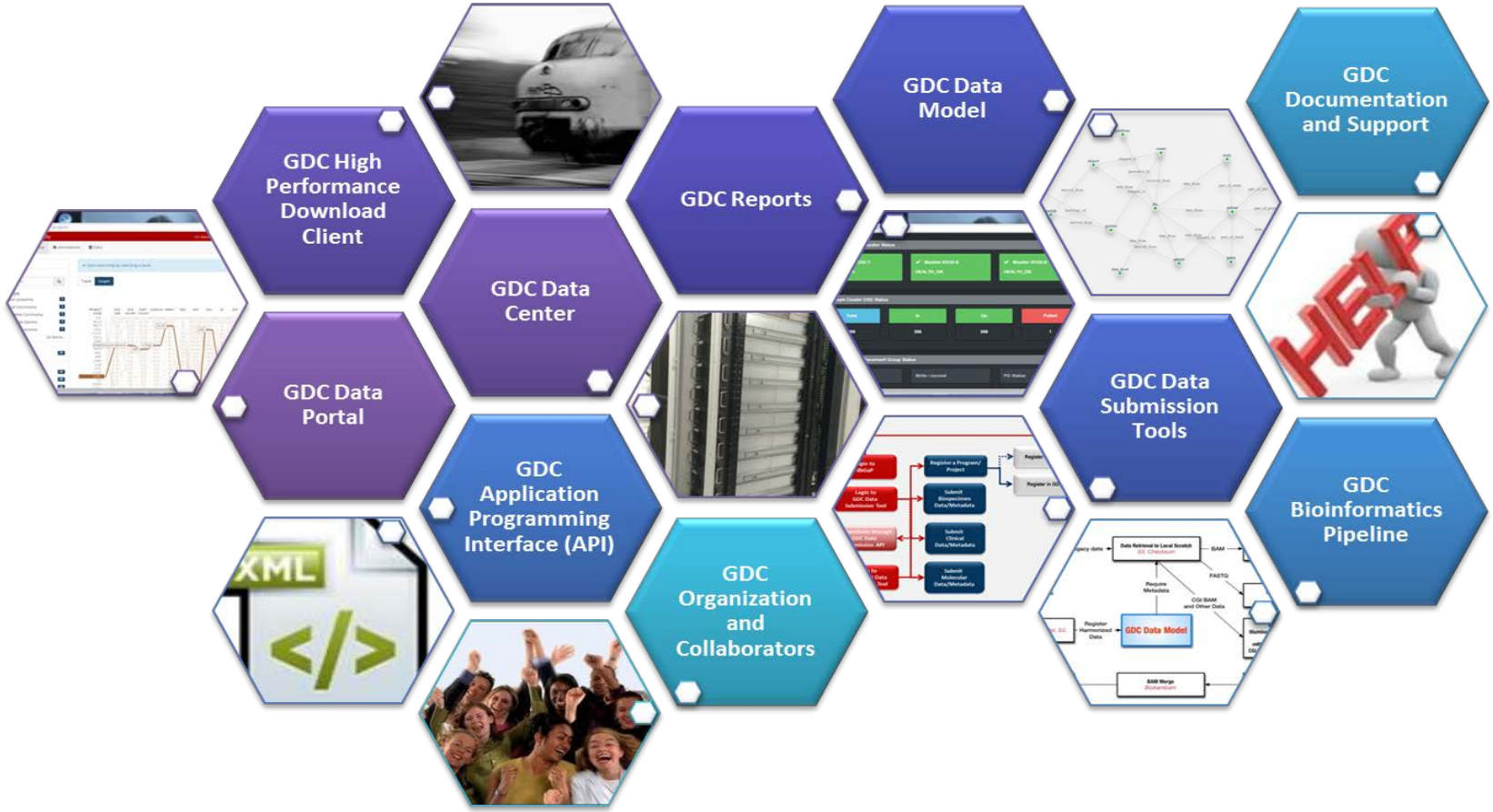
The Cancer Genomic Data Commons (GDC) is an existing effort to standardize and simplify submission of genomic data to NCI and follow the principles of FAIR -Findable, Accessible, Interoperable, Reusable. The GDC is part of the NIH Big Data to Knowledge (BD2K) initiative and an example of the NIH Data Commons

The Genomic Data Commons

Facilitating the identification of molecular subtypes of cancer and potential drug targets

NIH NATIONAL CANCER INSTITUTE

NCI Cancer Genomic Data Commons (GDC)



Genomic Data Commons (GDC) – Rationale

- TCGA and many other NCI funded cancer genomics projects each currently have their own Data Coordinating Centers (DCCs)
  - BAM data and results stored in many different repositories; confusing to users, inefficient, barrier to research
- GDC will be a single repository for all NCI cancer genomics data
  - Will include new, upcoming NCI cancer genomics efforts
  - Store all data including BAMs
  - Harmonize the data as appropriate
    - Realignment to newest human genome standard
    - Recall all variants using a standard calling method
    - Define data sharing standards and common data elements
  - Will be the authoritative reference data set
  - Will need to scale to 200+ petabytes

Genomic Data Commons (GDC)

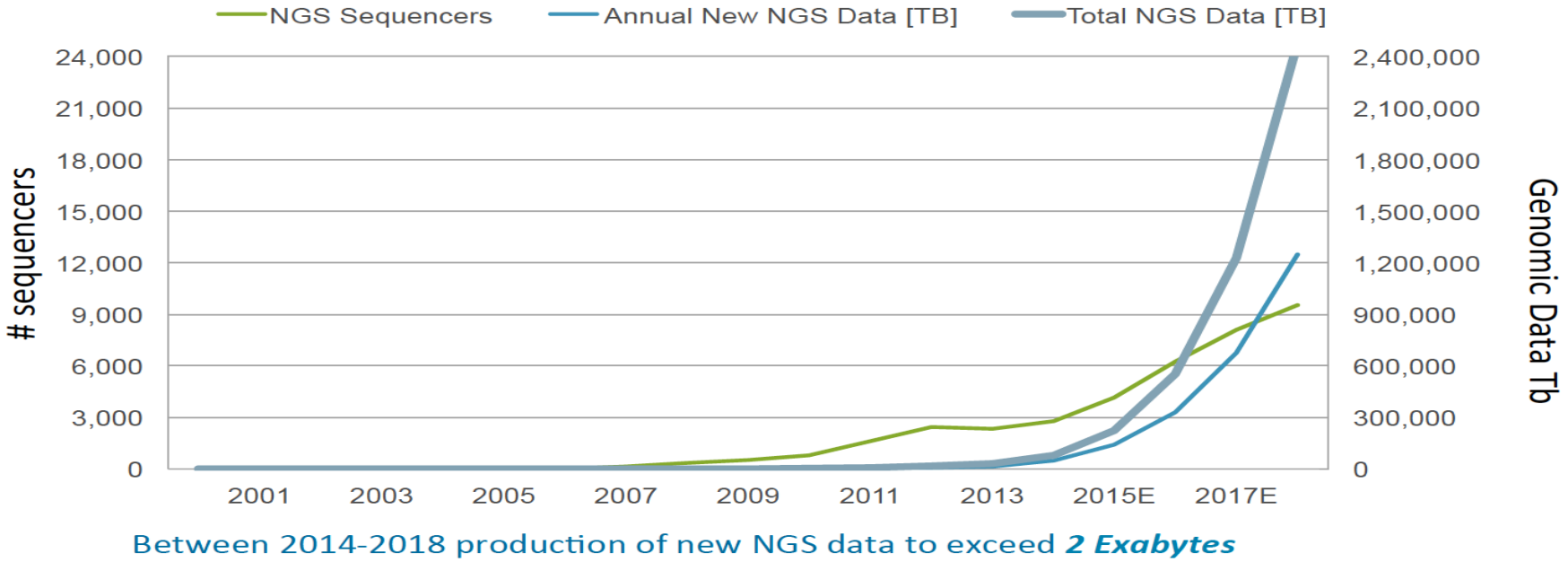
- First step towards development of a knowledge system for cancer
- Foundation for a genomic precision medicine platform
- Consolidate all genomic and clinical data from:
  - TCGA, TARGET, CGCI, Genomic NCTN trials, future projects
- Project initiated Spring of 2014
  - Contract awarded to University of Chicago
  - PI: Dr. Robert Grossman
  - Go live date: Mid 2016
  - Not a commercial cloud
- Data will be freely available for download subject to data access requirements

The NCI Cancer Genomics Cloud Pilots

Understanding how to meet the research community's need to analyze large-scale cancer genomic and clinical data

NIH NATIONAL CANCER INSTITUTE

Amount of genomic data will exceed available resources



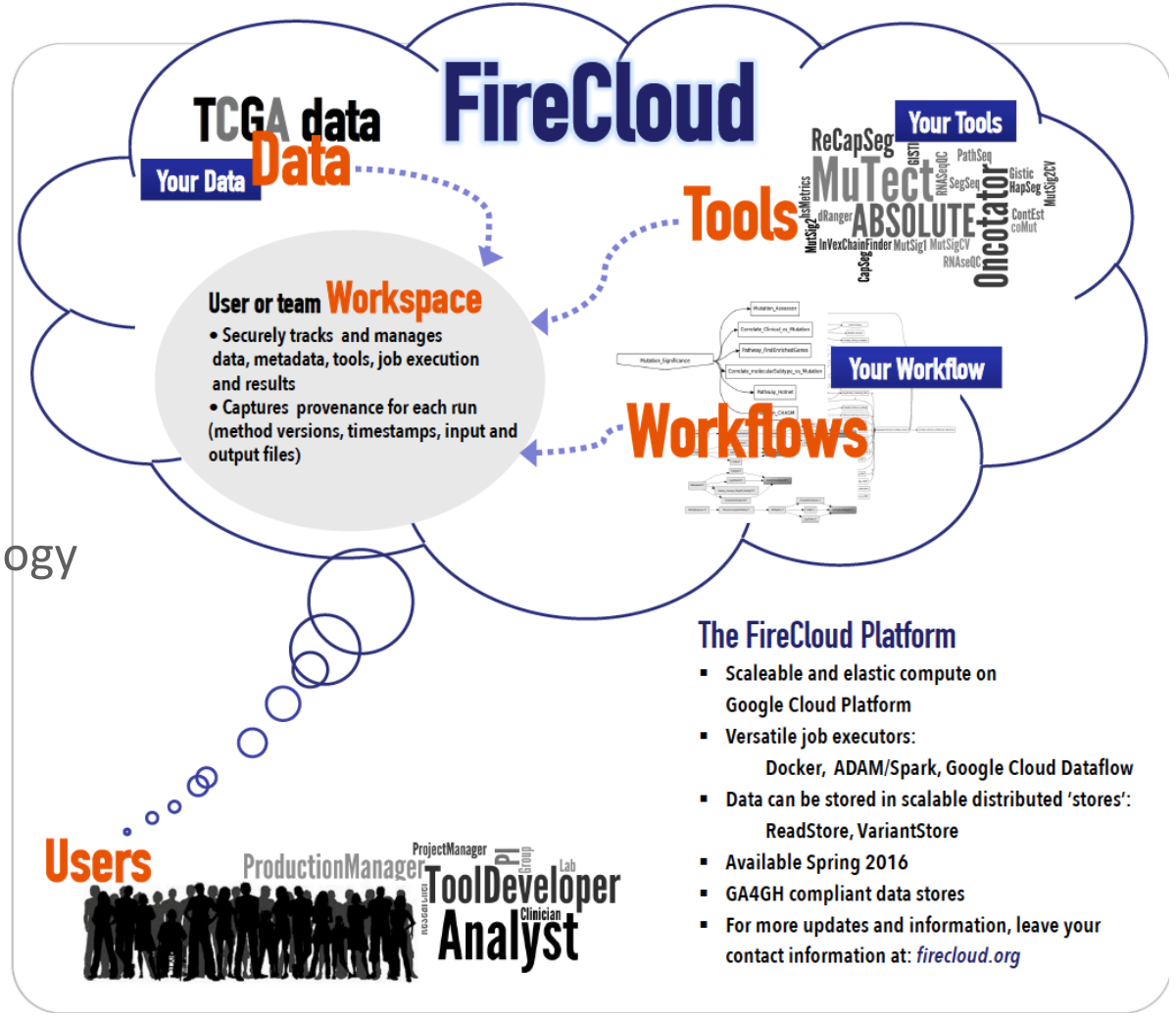
Between 2014-2018 production of new NGS data to exceed ***2 Exabytes***

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.

Slide courtesy of Deniz Kural, Seven Bridges Genc

NCI Cloud Pilots

The Broad
PI: Gad Getz
Institute for Systems Biology
PI: Ilya Shmulevich
Seven Bridges Genomics
PI: Deniz Kural



NCI GDC and the Cloud Pilots

- Working together to build **common APIs**
- Working with the Global Alliance for Genomics and Health (**GA4GH**) to **define** the next generation of **secure, flexible, meaningful, interoperable, lightweight interfaces**
- Competing on the **implementation**, collaborating on the **interface**
- Aligned with **BD2K** and serving as a part of the **NIH Commons** and working toward shared goals of **FAIR** (Findable, Accessible, Interoperable, Reusable)
- Exploring and defining **sustainable precision medicine information infrastructure**

Information problem(s) we intend to solve with the Precision Medicine Initiative for Oncology

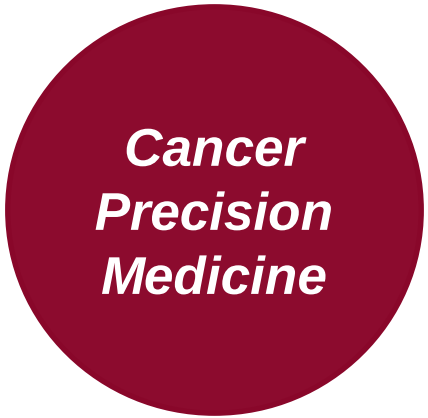
- **Establish** a sustainable infrastructure for cancer genomic data – through the **GDC**
- **Provide** a data integration platform to allow multiple data types, multi-scalar data, temporal data from cancer models and patients
  - Under evaluation, but it is likely to include the GDC, TCIA, Cloud Pilots, tools from the ITCR program, and activities underway at the Global Alliance for Genomics and Health
- **Support** precision medicine-focused clinical research

NCI Precision Medicine Informatics Activities

- As we receive additional funding for Precision Medicine, we plan to:
  - **Expand** the GDC to handle additional data types
  - **Include** the learning from the Cloud Pilots into the GDC
  - **Scale** the GDC from 10PB to hundreds of petabytes
  - Include imaging by interoperating between the GDC and the **Quantitative Imaging Network TCIA** repository
  - **Expand** clinical trials tooling from NCI-MATCH to NCI-MATCH Plus
  - **Strengthen** the ITCR grant program to explicitly include precision medicine-relevant proposals

Bridging Cancer Research and Cancer Care

- Making clinical research relevant in the clinic
- Supporting the virtuous cycle of clinical research informing care, and back again
- Providing decision support tools for precision medicine



Cancer Genomics Project Teams

CGC Pilot Team Principal Investigators

- **Gad Getz, Ph.D** - Broad Institute - http://firecloud.org
- **Ilya Shmulevich, Ph.D** - ISB - http://cgc.systemsbiology.net/
- **Deniz Kural, Ph.D** - Seven Bridges – http://www.cancergenomicscloud.org

NCI Project Officer & CORs

- Anthony Kerlavage, Ph.D –Project Officer
- Juli Klemm, Ph.D – COR, Broad Institute
- Tanja Davidsen, Ph.D – COR, Institute for Systems Biology
- Ishwar Chandramouliswaran, MS, MBA – COR, Seven Bridges Genomics

GDC Principal Investigator

- Robert Grossman, Ph.D - University of Chicago

Center for Cancer Genomics Partners

- JC Zenklusen, Ph.D.
- Daniela Gerhard, Ph.D.
- Zhining Wang, Ph.D.
- Liming Yang, Ph.D.
- Martin Ferguson, Ph.D.

NCI Leadership Team

- Doug Lowy, M.D.
- Lou Staudt, M.D., Ph.D.
- Stephen Chanock, M.D.
- George Komatsoulis, Ph.D.
- Warren Kibbe, Ph.D.

Thank you

Questions?

Warren A. Kibbe

warren.kibbe@nih.gov

NATIONAL CANCER INSTITUTE

www.cancer.gov

www.cancer.gov/espanol

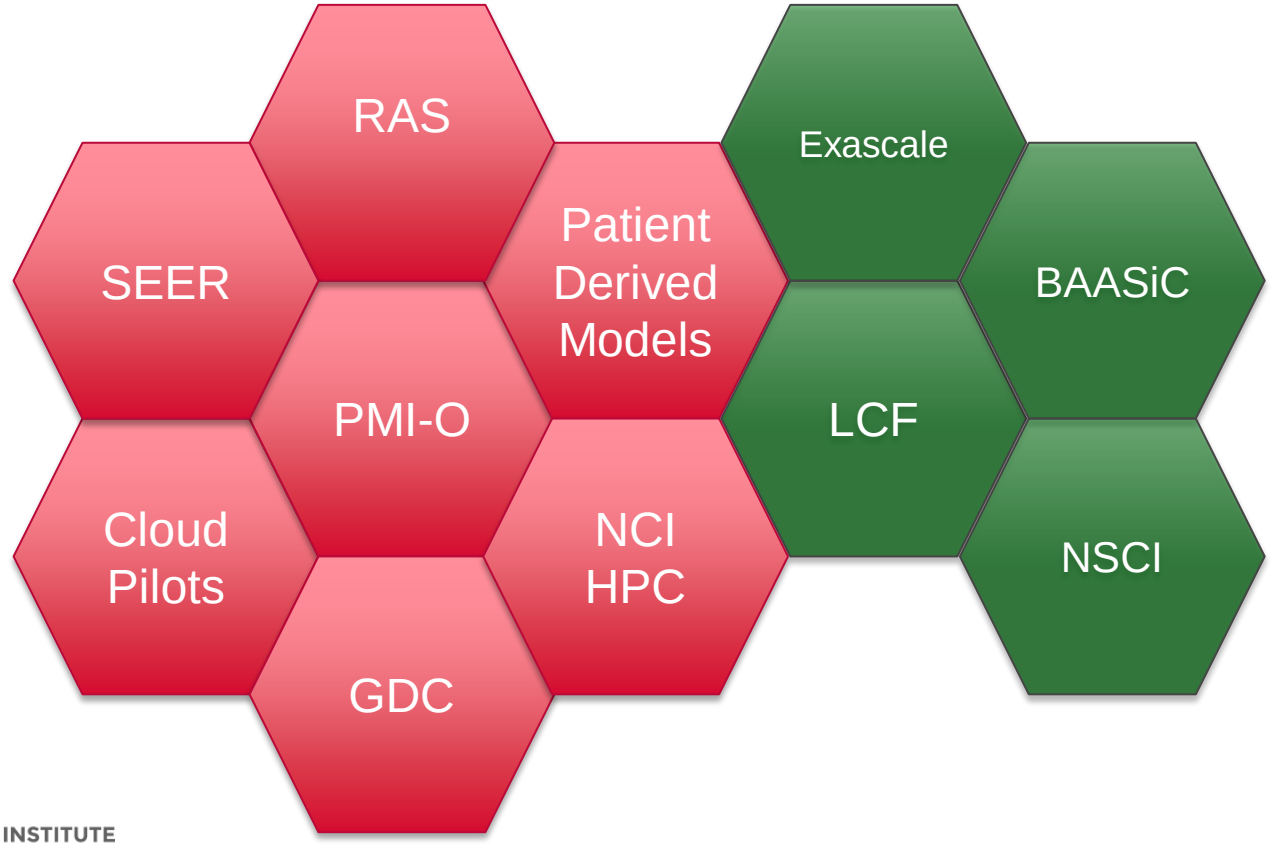
Thank you

Questions?

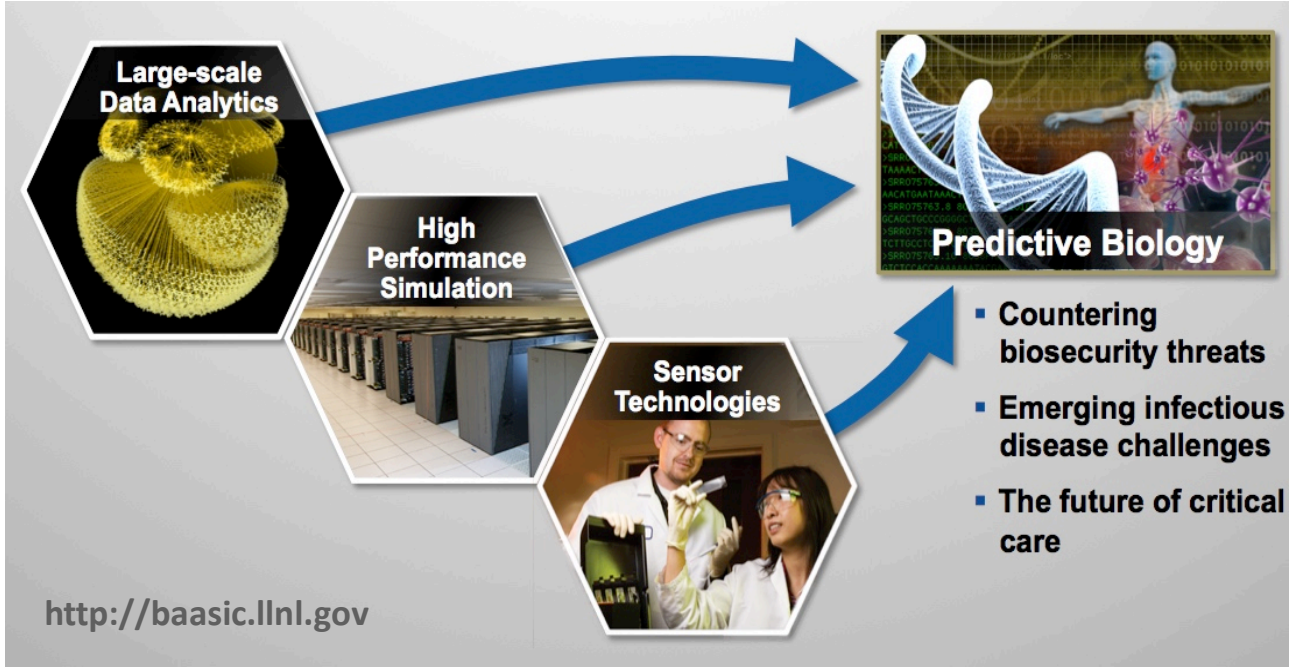
Warren A. Kibbe

warren.kibbe@nih.gov

Expanding Collaborations



BAASiC - Biological Applications of Advanced Strategic Computing



- Livermore led consortium
- Driving DOE Exascale advances in computing
- Specifically interested in cancer applications

- NCI/FNLCR target roles
- Cancer expertise and essential data
- Models, frameworks, “collaboratorium”