Biomedical Digital Twins: A Collaborative Landscape of Incredible Opportunity with Exciting Challenges

Eric Stahlberg – Director, Cancer Data Science Initiatives - Frederick National Laboratory for Cancer Research
Presented at the Purdue University Big Data Training for Cancer Research
May 20, 2022
Notes from the Frontier Agenda

• Foundations for Innovation
• Digital Twins
• Digital Twins and Treatment Optimization
• An Exciting Landscape Ahead
Acknowledgements

• The number of individuals who have contributed to this presentation, both directly and indirectly, represents a large and rapidly growing community.

• Acknowledging specific groups who have contributed
  – Organizations
    • NCI, DOE, FNLCR, LLNL, ORNL, ANL, BNL, LANL
  – Collaborative Teams
    • ATOM Team
    • JDACS4C Team
    • NCI-DOE Collaborations Team
    • Cancer Data Science Initiatives Team
    • NCI-DOE Digital Twin Teams

• The many conversations along the way including with those joining in the multiple workshops, meetings and events along the way
Overview of Frederick National Laboratory for Cancer Research

**FNL’s mission:** To provide a unique national resource for the rapid development of new technologies to address some of the most urgent and demanding problems in the biomedical sciences, including cancer, ongoing unmet challenges in HIV/AIDS, and threats of emerging infectious diseases.

**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. on behalf of the National Cancer Institute

**Two major campuses in Frederick, MD-- located at Fort Detrick and the ATRF**

- Frederick National Laboratory employees co-located with NCI researchers and other contractors
- Additional Frederick National Laboratory scientists at Bethesda and Rockville sites
Overview: Frederick National Laboratory for Cancer Research

- 2300 contract employees
- operations, technical support, and research for NCI and NIAID
  - facilities and animal maintenance at NCI-Frederick
  - bioinformatics and IT support
  - basic, translational, and clinical cancer research
  - basic and applied virus research and development on HIV, Ebola, and other emerging viral threats, including SARS-CoV-2
  - advanced technologies
  - NCI “National Missions”
Expanding Data for Cancer Research

- Genomic data
- Proteomic data
- Metabolomic data
- Clinical Data
  - EMR/EHR
  - Imaging
  - Observations
- Self-reported data
- Social media
- IoT Data
  - Environment
  - Individual wearable devices
- Simulation data
- Research data and observations
- Molecular data
  - ... and more
Exascale Cancer Science – circa 2016

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

Co-Design Efforts

Collaborative Pilot Investigations

Applications: Development, Libraries, Frameworks

Training: Scientists, Developers, Support Personnel

Infrastructure: Networking, Data Transfer, Data Management, HPC Access


Frederick National Laboratory for Cancer Research
Perspectives on HPC – circa 2016

- Big data and exascale are both essential elements of an integrated computing R/D agenda for the future
- Research and development of future algorithms, software and applications are as essential as research in hardware
- The global information technology ecosystem is in flux with transition to low-power mobile, cloud and data analytics
- Market competition and research collaboration are both needed to design, test and deploy the future systems supporting both HPC and big data needs

Source – Reed and Dongarra – Communications of the ACM, July 2015 (vol 57, no 7)
Joint Design of Advanced Computing Solutions for Cancer

**JDACS4C**
- Exascale technologies driving advances
- National Cancer Institute
- DOE Department of Energy

**Supported by NSCI and PMI**
- Argonne National Laboratory
- Oak Ridge National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Frederick National Laboratory for Cancer Research

**Integrated Precision Oncology**
- **Molecular**
  - Pre-clinical Domain – Improved predictive models
    - Computational/hybrid predictive models of drug response
    - Improved experimental design
- **Pre-clinical**
  - Clinical Domain – Precision oncology surveillance
    - Expanded SEER database information capture
    - Modeling patient health trajectories
- **Population**
  - Molecular Domain – Multiscale biological models
    - Models for RAS-RAS complex interactions
    - Insight into RAS related cancers

**CANcer Distributed Learning Environment (CANDLE)**
Scalable Deep Learning for Cancer

JDACS4C established June 27, 2016 with signed MOU between NCI and DOE
Precision Medicine will leverage large volumes and varieties of data to improve insight & outcomes.

>> How do we protect the data, devices, patients?

Many data sources and types...
- Genomic data
- Clinical and fundamental research data
- Clinical care data and observations – image, text, numerical, video, audio, etc.
- Life sciences, environment data
- Connected health and wearables data
- Real World Evidence (RWE) leveraging Unique Device Identifiers (UDI)

© FDHint, LLC
Cancer Research Data Commons

• Data are stored in domain-specific repositories, called Data Nodes (e.g., genomic, proteomic, imaging, etc.).

• Data access is controlled through a common Authentication and Authorization mechanism that secures the data.

• Researchers can bring their own data and tools to the cloud and combine with the data in the CRDC for integrative analysis.
Cancer Data Science Initiatives

• Collaborations with US Department of Energy – Commenced 2016
  – Current Projects:
    • ADMIRRAL – multiscale simulations for Ras-Raf binding
    • MOSSAIC – large scale AI for cancer surveillance
    • IMPROVE – approaches to improve AI predictive models for tumor response
  – Emerging Projects:
    • Cancer Patient Digital Twin & Precision Radiation Oncology
  – Online Cancer Data Science Resource
  – ATOM

• External Collaborations

• Intramural Research Program Support
MoDaC: Central location for annotated computational models and data sets
- Populated currently with results from NCI-DOE collaborations, including JDACS4C and ATOM
- User access
  - User-friendly, public facing web interface
  - Programmatic public REST API interfaces to support workflow integration
- Extensible metadata based search capability for locating models and datasets.
- DOI Support for references
  - Global identifier per asset.
  - Shareable link for citations.
- Multiple data download options including transfer to AWS S3 bucket, Google Drive, Globus and local
CANDLE: CANcer Distributed Learning Environment
Highlights of Deep Learning Framework

- Open source, Deep Learning software platform brings AI acceleration to multiple cancer research areas
  - DOE Exascale Computing Project
  - Brings large scale AI capabilities to NCI-DOE JDACS4C Collaboration for all pilot efforts in tumor response (pilot 1), RAS-membrane biology (pilot 2), and cancer surveillance (pilot 3)
  - Extended applications to multiple other areas including image analysis

- Scalable: Locally runnable while efficiently scaling on the world’s most powerful supercomputers

- Current functionalities
  - Hyperparameter optimization (HPO) of machine/deep learning models using either grid or Bayesian search

- CANDLE source and benchmarks are publicly available on Github:
  - Benchmarks: https://github.com/ECP-CANDLE/Benchmarks
  - Documentation: https://ecp-candle.github.io/Candle/html

CANDLE has been used for hyperparameter optimization of models related to analysis of cellular images, e.g. segmentation of subcellular structures such as mitochondria

CANDLE is also being used for Uncertainty Quantification (UQ) for some of the computational models developed in the pilots
NCI-DOE Collaboration Efforts

Cellular Level Pilot: Predictive Modeling for Pre-Clinical Screening – Pilot 1*

Innovative Methodologies and New Data for Predictive Oncology Model Evaluations (IMPROVE)

Molecular Level: Improving Outcomes for RAS-related Cancers – Pilot 2*

AI-Driven Multi-scale Investigation of RAS/RAF Activation Lifecycle (ADIMRRAL)

Population Level: Population Information Integration, Analysis, and Modeling for Precision Surveillance – Pilot 3*

Modeling Outcomes Using Surveillance Data and Scalable Artificial Intelligence for Cancer (MOSSAIC)

CANDLE – An exascale project providing the deep learning capabilities to each project.

*Resources currently available: https://datascience.cancer.gov/collaborations/nci-doe-capabilities
Joint Design of Advanced Computing Solutions for Cancer

**JDACS4C**
- Exascale technologies driving advances
- DOE National Cancer Institute
- Cancer driving computing advances
- Supported by NSCI and PMI
- NJI National Cancer Institute
- Argonne National Laboratory
- Oak Ridge National Laboratory
- Lawrence Livermore National Laboratory
- Los Alamos National Laboratory
- Frederick National Laboratory for Cancer Research

**Integrated Precision Oncology**

- **Molecular**
  - Pre-clinical Domain – Improved predictive models
    - Computational/hybrid predictive models of drug response
    - Improved experimental design

- **Pre-clinical**
  - Clinical Domain – Precision oncology surveillance
    - Expanded SEER database information capture
    - Modeling patient health trajectories

- **Population**
  - Molecular Domain – Multiscale biological models
    - Models for RAS-RAS complex interactions
    - Insight into RAS related cancers

**CANcer Distributed Learning Environment (CANDLE)**
- Scalable Deep Learning for Cancer

**JDACS4C** established June 27, 2016 with signed MOU between NCI and DOE
Envisioning Computational Innovations for Cancer Challenges (ECICCC) Community

- Dedicated to accelerating computational oncology and developing research collaborations across cancer and computational sciences
- Scientists from over 200 organizations in academia, government, and industry
- Multidisciplinary events to share their ideas and expertise, develop use cases, and explore new research collaborations
- Area dedicated to Cancer Patient Digital Twins

To join, email ECICCC_Community@nih.gov
Digital Twins

Charting a course for applications in cancer
What is a Digital Twin?

A digital twin is a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity.

- Digital twin systems transform business by accelerating holistic understanding, optimal decision-making, and effective action.
- Digital twins use real-time and historical data to represent the past and present and simulate predicted futures.
- Digital twins are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and implemented in IT/OT systems.

Source: The digital twin consortium – December 3, 2020
Use cases for digital twins

• Transportation
  – Planes, trains and automobiles

• Manufacturing
  – Supply chain, production lines, reactors

• Systems
  – Patient flow, water systems, treatment systems, power plants, traffic, finance

• Defense, aerospace

• Life science
Growing Use of Digital Twins in Medicine

The Potential of a Digital Twin in Surgery

The personal digital twin, ethical considerations

Towards the Development of Digital Twins for the Bio-manufacturing Industry

A CFD Digital Twin to Understand Miscible Fluid Blending

Single-cell Digital Twins for Cancer Preclinical Investigation

Digital Twin Technology: The Future of Predicting Neurological Complications of Pediatric Cancers and Their Treatment
Why the Digital Twin for Predictive Oncology?

• Traditional approaches rely on many individuals to develop general predictions
• Results take time to achieve
• Imprecise conditions
• Explorations limited by available physical models, samples, data
“The Digital Twin Approach”

- Digital twin approaches involve many models for precise predictions
- Explore many possible treatments
- More rapid explorations
- Set specific conditions
- Progressively integrate understanding and insights

Goal: Provide critical insights for the individual cancer patient
A Digital Twin in a Learning Cancer Health System

- 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
Cancer Patient Digital Twin: New insights and approaches from molecular to patient scale!

New observables
New data
New approaches

Multiple perspectives
Integrated objectives

Reaching the Patient for Impact

Multiple starting points
team collaborations

Prescriptive cancer models
Patient specific disease ensembles

Multi-scale
Multi-cell models
Multi-mechanism
Multi-ensemble
Multi-scale
Single cell models
Multi-mechanism
Single ensemble
Single cancer ensemble
Multiple distinct cancer models

Single mechanism
Single ensemble
Single mechanism
Single trial

Integrated objectives:
- New observables
- New data
- New approaches

Multiple perspectives:
- Single mechanism
- Multi-mechanism
- Multi-scale

Prescriptive cancer models:
- Patient specific disease ensembles

Reaching the Patient for Impact:
- Multiple starting points
- Team collaborations
The Challenge: Digital Twin for Predictive Oncology

Patient-tailored models incorporating multi-omic, clinical, environmental and social data that can evaluate and predict the most effective prevention and therapeutic plans.
Cancer Patient Digital Twin Paradigm

Cancer Patient Digital Twin based care brings together data, models, computing, and technology together with researchers, patients and physicians to improve the care options for each individual.

Cancer Patient Digital Twins – Implications and Opportunities

• Individual Implications
  – Provide insights to best predicted treatment combinations
  – Improve decision making during treatment

• Implications beyond the individual
  – Accumulated trajectory and outcomes data provide insight on successful treatments
  – Enable health systems to better prepare to respond to real-time health situations and health disparities

• Realization of potential
  – Requires contributions from experimental, clinical and computational communities

• Key Challenges
  – Data challenges – high-quality and high-volume multiscale data, healthy and diseased state, diverse population coverages
  – Modeling and integration challenges – harmonizing data, integrating models, HPC access, standards
  – Ethical and community challenges – broad stakeholder involvement, bias and privacy, governance of data

From Hernandez-Boussard et al, Digital twins for predictive oncology will be a paradigm shift for precision cancer care, Nature Medicine, November 2021
A Learning Health System for Cancer Patients

- Technical pathways emerging to support digital twin
  - FAIR ecosystem (data, models, software)
  - Cloud computing
  - Affordable computing
  - AI, trusted and sustainable AI
  - Data sharing and data security
  - 5G
  - Blockchain
  - Medical IoT
  - Exascale computing
  - Quantum computing

Image From 2019 Panel at SC19 on Edge to Exascale
Digital Twin Innovation Ideas Lab - 2020

• Think beyond the immediate....
• What would be possible if...?
• What will be possible when...?
• What will be different when current efforts finish?
• Where will technology be in ten years?
• As current barriers are surpassed, what follows?

• How to get started?
• What are the steps to move ahead?

Guiding long-term objective: Deliver a digital twin that provides critical, current and responsive insight for the *individual* cancer patient!
## Cancer Patient Digital Twin Teams

<table>
<thead>
<tr>
<th>University</th>
<th>Project Aim</th>
<th>PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEORGETOWN UNIVERSITY</td>
<td><strong>Simulating One Million Pancreatic Cancer Patient Digital Twins to Plan Precision Medicine Treatment Strategies and Improve Long-term Survival.</strong> PI: Matthew McCoy, PhD</td>
<td></td>
</tr>
<tr>
<td><strong>Project Aim:</strong></td>
<td>Simulate one million pancreatic Cancer Patient Digital Twins (CPDT) in a models repository by parameterizing input from real patient trajectories for drug sensitivity and resistance</td>
<td></td>
</tr>
<tr>
<td>Stanford University</td>
<td><strong>An Adaptive Digital Twin Approach for Monitoring Treatment Response and Resistance</strong> PI: Olivier Gevaert, Ph.D., Assistant Professor in Medicine (Biomedical Informatics)</td>
<td></td>
</tr>
<tr>
<td><strong>Project Aim:</strong></td>
<td>Advance the hypothesis that optimal pathways for a specific cancer patient can be selected by exploring the treatment pathway space through a dynamical, multiscale digital twin derived by harnessing patient’s own data and leveraging data from similar patients in the population.</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td><strong>Dynamic Multiscale Digital Twin for a Lung Cancer Patient.</strong> PI: Qi Wang, Professor of Mathematics</td>
<td></td>
</tr>
<tr>
<td><strong>Project Aim:</strong></td>
<td>Build the first prototype of an individualized digital twin of non-small cell lung cancer for identifying the optimal treatment pathway and adaptive treatment monitoring, leveraging the population information and new test results.</td>
<td></td>
</tr>
<tr>
<td>UMass Amherst</td>
<td><strong>Prototyping a Self-learning Digital Twin Platform for Personalized Treatment in Melanoma Patients</strong> PI: Paul Macklin, PhD, Associate Professor of Intelligent Systems Engineering</td>
<td></td>
</tr>
<tr>
<td><strong>Project Aim:</strong></td>
<td>Rapidly prototype a 3D multiscale model of melanoma metastases that interact with the host immune system with or without treatment (autologous vaccine immunotherapy)—and verify that it recapitulates a broad variety of clinically relevant patient trajectories.</td>
<td></td>
</tr>
<tr>
<td>UMass Amherst</td>
<td><strong>My Virtual Cancer.</strong> PI: Leili Shahriyari, PhD., Assistant Professor, Mathematics and Computer Science</td>
<td></td>
</tr>
<tr>
<td><strong>Project Aim:</strong></td>
<td>Combine mechanistic, machine learning, and stochastic modeling approaches to create a DT platform that utilizes biological, biomedical, and EHR data sets. Will focus on one common cancer—breast cancer—and one rare cancer—uveal melanoma—to evaluate the performance of the DT for both common and rare cancers.</td>
<td></td>
</tr>
</tbody>
</table>

Frederick National Laboratory for Cancer Research
Simulating one million pancreatic cancer patient digital twins to plan Precision Medicine treatment strategies and improve long-term survival

**Project Aim:** Simulate one million pancreatic Cancer Patient Digital Twins (CPDT) in a models repository by parameterizing input from real patient trajectories for drug sensitivity and resistance
Simulating one million pancreatic cancer patient digital twins to plan Precision Medicine treatment strategies and improve long-term survival

Project Team

Matthew McCoy PhD
Assistant Professor, Department of Oncology
Lombardi Comprehensive Cancer Center and ICBI
Georgetown University Medical Center

Robert A. Beckman, MD
Professor of Oncology and of Biostatistics
Bioinformatics, and Biomathematics
Lombardi Comprehensive Cancer Center and ICBI
Georgetown University Medical Center
Scientific Advisor, Office of the Senior Vice President for Research
Georgetown University

Subha Madhavan, PhD, FACMI
Adjunct Faculty, Department of Oncology
Georgetown University Medical Center

Samir Gupta PhD
Research Faculty, Informatics Innovation Center for Biomedical Informatics
Georgetown University Medical Center
Simulating one million pancreatic cancer patient digital twins to plan Precision Medicine treatment strategies and improve long-term survival

Cancer Patient Digital Twins (CPDT) rely on incomplete patient data to parameterize imperfect computational models, so an iterative approach is needed where model outcomes inform additional data collection. This leads to refined models and directs further data collection and so on.

During each round of modeling, patient derived data is used to constrain ranges on each of the model parameter values. A population of individual DTs, each corresponding to a unique combination of values sampled from those parameter ranges, is used to inform therapeutic strategies and additional data collection.
Simulating one million pancreatic cancer patient digital twins to plan Precision Medicine treatment strategies and improve long-term survival.

An Adaptive Digital Twin Approach for Monitoring Treatment Response and Resistance

**Project Aim:** Advance the hypothesis that optimal pathways for a specific cancer patient can be selected by exploring the treatment pathway space through a dynamical, multiscale digital twin derived by harnessing patient’s own data and leveraging data from similar patients in the population.
An Adaptive Digital Twin Approach for Monitoring Treatment Response

Goal: Develop an adaptive digital twin (DT) that will

- Predict response to initial therapy
- Monitor and assess resistance mechanisms during the maintenance phase
- Enable rapid and effective treatment reassignment

Approach:

- Integrating baseline multimodal data with
- Repeated, ongoing clinical measurements using
- Dynamic models with
- Continuous training and model updating to enable
- Forward simulation of anticipated future clinical indicators and outcomes

Outcome:

This DT approach will help physicians make initial treatment determinations, monitor treatment response and effectiveness, and decide when to discontinue or change treatment.
Project Team

**Olivier Gevaert** - Expertise in multi-omics, multi-modal data fusion

**Sepideh Dolatshahi** - Expertise in data-driven and mechanistic modeling, Systems Immunology

**David Stracuzzi** - AI/Machine learning, uncertainty quantification, quality of the models & prediction

**Talayeh Razzaghi** - Data science & machine learning, big data

**Jinhua Wang** - Bioinformatician, genome informatics, sequence analysis & single cell genomics

**Eva Katsoulakis** - Radiation oncologist, experience with radiomics, genomics, informatics
Lung Variational Auto-Encoder to dynamically model tumor size

Workflow

- Lung CTs
- Auto-Encoder
- Embedding
- Downstream tasks
- Treatment response

Results

- Original
- Reconstructed

Ongoing work

- Dynamic model of tumor size

Applicable for multiple tasks

- Treatment response
- Disease progression
- ...
Dynamic Multiscale Digital Twin for a Lung Cancer Patient

**Project Aim:** Build the first prototype of an individualized digital twin of non-small cell lung cancer for identifying the optimal treatment pathway and adaptive treatment monitoring, leveraging the population information and new test results.
Multimodal Fusion Across Scales for Building Digital Twin Models

Team Members:
Qi Wang, University of South Carolina
Jun Deng, Yale University
Pamela Jackson, Mayo Clinic
Eva Katsoulakis, Veterans Affairs James A. Haley Tampa VA/University South Florida/Moffitt Cancer Center
Marieke Kuijjer, University of Oslo
Leili Shahriyari, University of Massachusetts Amherst
Yi Sun, University of South Carolina
Tanveer Syeda-Mahmood, IBM Research
In a complex disease such as cancer, the interactions between the tumor and host can exist at the molecular, cellular, tissue, and organism levels.

Evidence for the disease and its evolution may be present in multiple modalities across scale such as clinical, genomic, molecular, pathological and radiological imaging.

Effective patient-tailored therapeutic guidance and planning through digital twins in the future will require bridging spatiotemporal scales through novel multimodal fusion formalisms.
Multimodality DT Modeling to Predict RT Efficacy and Toxicity for Cancer Patients
Prototyping a Self-learning Digital Twin Platform for Personalized Treatment in Melanoma Patients

**Project Aim:** Rapidly prototype a 3D multiscale model of melanoma metastases that interact with the host immune system with or without treatment (autologous vaccine immunotherapy)—and verify that it recapitulates a broad variety of clinically relevant patient trajectories.
Prototyping a self-learning digital twin platform for personalized treatment in melanoma patients

PIs:
Tina Hernandez-Boussard (Stanford University)
Paul Macklin (Indiana University)
Ilya Shmulevich (Institute for Systems Biology)
Approach:

- Create a generalized model of melanoma micrometastases, immune interactions, and immunotherapy
- Use HPC to broadly explore space of trajectories
- Use ML to cluster trajectories into patient templates
Models with different parameters and initial conditions exhibit highly diverse clinical outcomes.

- **immune escape**
- **immune equilibrium**
- **immune elimination**
My Virtual Cancer

**Project Aim:** Combine mechanistic, machine learning, and stochastic modeling approaches to create a DT platform that utilizes biological, biomedical, and EHR data sets. Will focus on one common cancer—breast cancer—and one rare cancer—uveal melanoma—to evaluate the performance of the DT for both common and rare cancers.
My Virtual Cancer - Personnel

Computational & Mathematical Scientists

Leili Shahriyari, PHD (UMass Amherst) computational oncology, QSP modeling, machine learning, stochastic agent-based modeling.

Navid Mohammad Mirzaei, PHD (UMass Amherst), PDE modeling.

Pamela Jackson, PHD (Mayo Clinic) image processing.

Wenrui Hao, PHD (Penn State) scientific computing, PDE, biophysics modeling, sensitivity analysis.

Alireza Asadpoure, PHD (UMass Dartmouth) structural engineering, topology optimization (TO), stochastic modeling.

Young Hwan Chang, PHD, (OHSU) biomedical engineering, image processing, deep learning.

Biological & biomedical Scientists

Ioannis Zervantonakis (University of Pittsburgh) biomedical engineering, animal models, metastasis, breast and ovarian cancer.

Adrian Lee, PHD (University of Pittsburgh) biochemistry, breast cancer, breast biology.

Colleen Cebulla, MD, PHD, (Ohio State Univ) ocular oncology, vitreoretinal surgery, cancer biology.

Mohamed Abdel-Rahman, MD, PHD, (Ohio State Univ) pathology, clinical genetics and molecular genetics.

Daniel G. Stover, MD (Ohio State Univ) breast cancer, clinical computational oncology.

Output of Pilot Phase

- Mathematical model of breast tumors progression based on their immune infiltration (*Personalized Medicine*, doi: 10.3390/jpm12050807)

- Investigating key cell types and molecules dynamics in PyMT mice model of breast cancer through a mathematical model (*PLOS CB*, doi: 10.1371/journal.pcbi.1009953)

- A PDE Model of Breast Tumor Progression in MMTV-PyMT Mice (*Personalized Medicine*, doi: 10.3390/jpm11101031)

**Challenge:** Lack of time course data for cancer patients

**Parameter estimation:**
- Found some of parameters by literature review.
- Assumed large tumors are at the steady state.
- Made some assumptions about the relationships among some of parameters.

**Sensitivity analysis:** Performed a global gradient-based sensitivity analysis on non-dimensionalized system by changing each assumption 5000 times.

Analyzing TCGA & METABRIC data
A Broader Community for Digital Twins
A Broader Community for Digital Twins

- CompBioMed: European effort focusing on developing the virtual human

- Digital Twin Consortium: providing domain agnostic resources and sector specific working groups

- With more groups emerging in life science
Digital Twins and Treatment Optimization

ATOM: Community Effort for Integrated Modeling to Develop New Therapeutics
The Challenge: Digital Twin for Predictive Oncology

*Patient-tailored models incorporating multi-omic, clinical, environmental and social data that can evaluate and predict the most effective prevention and therapeutic plans*
New Molecular Entity Discovery

The opportunity for transformation...

- **Target ID & Selection**
  - Reductionist & functional screening of millions of molecules to inform selection

- **Design, make, & test**
  - 1000s of new molecules
  - Optimize empirically by sequential evaluation

- **Lengthy in-vitro and in-vivo experiments**
  - Continuous tradeoffs

- **Lead Discovery**
  - 1.5 yrs

- **Lead Optimization**
  - 3 yrs

- **Preclinical**
  - 1.5 yrs

- **~5 years**

- **~33% of total cost of medicine development**

- **~12% clinical success**

- **~$40B 2020 NIH Budget**

- **1/3 of ~$189B 2020 Private sector spend in Pharma R&D**

- **~$40B 2020 NIH Budget**

Resources:
“2017” ATOM target-to-clinical trial roadmap
Active learning approaches to accelerate timeline and reduce experimentation

<table>
<thead>
<tr>
<th>Timeline</th>
<th>Target</th>
<th>Molecule for Clinical trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>1.5 yrs Lead Discovery</td>
<td>6 yrs</td>
</tr>
</tbody>
</table>
| 2020     | 0.5 yrs 1.5 yrs Lead Optimization | 3.5 yrs  
                        | 1.5 yrs Preclinical | 6 yrs                       |
| 2021     | 1 yr 1.5 yrs            | In-silico approaches to improve cycle time  |
|          | 1.5 yrs Lead Optimization | Active learning, empirical integration  |
|          | 2.5 yrs Preclinical | ATOM Platform Proof of Concept (POC) initiated: Target to Candidate Molecule in < 1 year  |
| 2024     | 1.5 yrs                  | Additional computational improvements after POC  |
|          | 1 yr                    | Automated, reduced preclinical chemical synthesis  |
|          |                        | Reduced number of animal studies  |
| 2025-2030| 1 yr                     | ATOM 2.0 platform – personalized predictive models & precision medicine  |
|          |                        | Animal models replaced by computational and in vitro models  |
|          |                        | Regulatory agreement  |
ATOM Network Today

Lawrence Livermore National Laboratory
UCSF
Frederick National Laboratory for Cancer Research
Argonne National Laboratory
Oak Ridge National Laboratory
Brookhaven National Laboratory

Tech

Pharma

Cancer Centers

Gov’t Labs

Academic

Partners

High-performance computing

Diverse biological data

Emerging experimental capabilities

ATOMIC
Emerging computing and experimental technologies can enable a new approach

1. **Active learning** — Predictive computational models incorporating AI and high-performance simulation specify exactly which experimental to do.

2. **Multiparameter molecular design** simultaneously optimizes efficacy, safety, pharmacokinetics, and manufacturability.

3. **Human relevant models** — both computational systems models and experimental human organoids – in the design loops to improve success rates in human testing.
The Foundational ATOM Molecular Design Workflow

- **Working Compound Library**
- **Property Prediction Models**
  - Efficacy
  - Safety
  - PK
  - Developability
- **Multi-Parameter Optimization Loop**
- **Design Criteria**
- **Generative Molecular Design**
  - Proposes new molecules with optimized properties
- **Simulation**
- **Active learning** decides if/when a simulation or experiment is needed to improve or validate models
- **Experiment**
  - Human-relevant assays, complex in vitro models
  - Chemistry Design & Synthesis
- **A Drug Candidate**
  - Retrain property prediction models

**Lawrence Livermore National Laboratory**
The Foundational ATOM Molecular Design Workflow

- **Working Compound Library**
- **Property Prediction Models**
  - Efficacy
  - Safety
  - PK
  - Developability
  - Machine learning parameter models
  - Human-level systems models
- **Digital Twin Informed Models**
- **Generative Molecular Design**
  - Proposes new molecules with optimized properties
- **Simulation**
- **Experiment**
- **Multi-Parameter Optimization Loop**
  - Design Criteria
  - Active learning decides if/when a simulation or experiment is needed to improve or validate models

**Digital Twin Informed Assays**
- Human-relevant assays, complex in vitro models
- Chemistry Design & Synthesis

**Digital Twin Informed Systems Biology Models**
- Retrain property prediction models

**A Drug Candidate**
Driving Technology Development Innovation through Challenge
Project #1: AURK A/B Selectivity Improvement
Development of AMPL and Integrated ATOM Workflow

Validation data shows general effectiveness of models

>200 new potent, compounds with favorable other properties

>200 new compounds with desired efficacy

Candidate Panel

Efficacy
AURK B (pIC50 > 9)
AURK B/A Selectivity > 1000
Safety
BSEP (pIC50 < 4)
HERG (pIC50 < 4)
PK
Solubility (>10μM)
Cl\text{app} (< 3 ml/min/g)
Developability
Solubility (>10μM)
SAS: Synthetic Accessibility Score
QED: Quantitative Estimation of Drug Likeness

AURK A, B

AURK B vs. AURK A pIC50

>200 new compounds with desired efficacy
Project #2: Neurocrine H1 inhibitor design

- H1-antihistamines are a class of drugs used to relieve allergic symptoms.
- First generation H1-antihistamines have undesirable side effects mainly due to off-target activities against muscarinic receptors and hERG channel.
- Neurocrine developed a selective H1 antagonist (Ki=4nM) with > 1000-fold selectivity versus the muscarinic receptors and 350-fold selectivity vs the hERG channel, but it was hard to synthesize due to chiral resolution.
- We want to identify new selective H1 antagonists through the GMD loop that don’t have such synthetic difficulties.

Driving development of ADI and managed generative molecular design
Project # 2: Receptor Antagonist Binding Optimization

Overall timescale: ~3 months

Project Characteristics
- Deeply understood target
- Abandoned pharma project
- Pharmco data donated to ATOM
- Pharma ran all experiments to confirm hits
- Novel molecules found meeting all criteria
- IP filing in progress

Target design criteria for novel inhibitor case study

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-target binding affinity</td>
<td>$K_i \leq 1 \text{ nM}$</td>
</tr>
<tr>
<td>Off-target binding affinity</td>
<td>$K_i \geq 1 \text{ \mu M}$</td>
</tr>
<tr>
<td>hERG inhibition</td>
<td>IC50 $\geq 10 \text{ \mu M}$</td>
</tr>
<tr>
<td>Synthetic accessibility score</td>
<td>SAS $\leq 3$</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>$180 \leq MW \leq 500$</td>
</tr>
<tr>
<td>Lipophilicity (octanol-water partition coefficient, predicted)</td>
<td>-0.4 $\leq$ SlogP $\leq$ 5.6</td>
</tr>
<tr>
<td>Topological polar surface area</td>
<td>TPSA $\leq 140$</td>
</tr>
<tr>
<td>Hydrogen bond donors</td>
<td>HBD $\leq 5$</td>
</tr>
<tr>
<td>Hydrogen bond acceptors</td>
<td>HBA $\leq 10$</td>
</tr>
<tr>
<td>Rotatable bonds</td>
<td>RB $\leq 10$</td>
</tr>
</tbody>
</table>

2nd GMD Run Results

Predicted off-target vs on-target pKi values
Project 3: Enzyme Inhibitor with High Isoform Selectivity and Selective PK/PD

ATOM POC project

Characteristics of Project

• Well validated target, multiple approved drugs
• Limited efficacy in a large patient subpopulation
• Non-desired effects
• ARA/ATOM assembled a SME team to address the design flaws exhibited in current drugs.
• Intent to design an IND enabled molecule

GMD 1

• 12 parameter optimization (see figure)
• Other parameters will be brought in as the models become available
• Synthesis and experimental measurements of proposed molecules next
• Additional models to be brought in for next GMD

Driving development of extended systems model integration
Accelerating Open Drug Discovery
The Foundational ATOM Molecular Design Workflow

Property Prediction Models
- Efficacy
- Safety
- PK
- Developability
  - Machine learning parameter models
  - Human-level systems models

Multi-Parameter Optimization Loop
Design Criteria

Generative Molecular Design
Proposes new molecules with optimized properties

Simulation
Molecular Feature Simulations

Experiment
- Human-relevant assays, complex in vitro models
- Chemistry Design & Synthesis

Working Compound Library
Retrain property prediction models

A Drug Candidate

Active learning decides if/when a simulation or experiment is needed to improve or validate models
Open Predictive Models - AMPL

Property Prediction Models
- Efficacy
- Safety
- PK
- Developability

Multi-Parameter Optimization Loop
- Design Criteria
- Active learning decides if/when a simulation or experiment is needed to improve or validate models

Generative Molecular Design
- Proposes new molecules with optimized properties

Simulation
- Molecular Feature Simulations

Experiment
- Human-relevant assays, complex in vitro models
- Chemistry Design & Synthesis

Work Compound Library
- Retrain property prediction models

A Drug Candidate
AMPL provides:

- Easy integration of diverse datasets
- Integration with scalable data and model services environment
- High-performance hyperparameter optimization
- Rapid evaluation of model architecture
- Seamless HPC integration using world-class compute systems
- Ensemble integration of models from multiple sources

https://github.com/ATOMScience-org/AMPL
‘QBAR’ modeling combines chemical and biological features

In vivo BBB penetrance defined by logBB: \[ \log\left(\frac{[cmpd]_{CSF}}{[cmpd]_{PB}}\right) \]

Calculate molecular descriptor features for 1181 compounds with logBB data

QSAR: use chemical descriptors alone to predict logBB

‘QBAR’: add true & predicted PAMPA & transporter assay data as bio features

Model building and hyperparameter optimization via grid search

Compare final model performance

Hypothesis: biological data from cheaper, higher throughput experiments as inputs will improve prediction of complex in vivo phenotypes over chemical descriptors alone
Fingerprint splits and QBAR modeling improve generalization of logBB predictions

Wilcoxon Signed-Rank test for significance
Effect sizes (ΔMAD): Scaff valid = 0.28; FP valid = 0.13; FP test = 0.25

ATOM  
Slides courtesy of Amanda Paulson (UCSF)
Generative Molecular Design - GMD

Property Prediction Models
- Efficacy
- Safety
- PK
- Developability

Multi-Parameter Optimization Loop
Design Criteria

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

Working Compound Library
Retrain property prediction models

Simulation
Molecular Feature Simulations

Experiment
Human-relevant assays, complex in vitro models
Chemistry Design & Synthesis

A Drug Candidate
Generative Molecular Design (GMD)

Integrating multiple components

- A **generative model** for proposing new chemical structures, often involving a latent space representation.

- **Predictive models** to infer the properties of compounds given their structures. Can use either mechanistic or machine learning (ML) approaches.

- An **optimizer** to guide the generative model to produce molecules with favorable predicted properties.

*Slides courtesy of Kevin McLoughlin (LLNL)*
<table>
<thead>
<tr>
<th>Description</th>
<th>CostFunction</th>
<th>Target_min</th>
<th>Target_max</th>
<th>Weight</th>
<th>Scale</th>
<th>Allow_neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-target 1 pIC50</td>
<td>exp</td>
<td>8</td>
<td></td>
<td>3</td>
<td>0.72</td>
<td>TRUE</td>
</tr>
<tr>
<td>Off-target 2 pIC50</td>
<td>exp</td>
<td>6</td>
<td></td>
<td>3</td>
<td>0.41</td>
<td>FALSE</td>
</tr>
<tr>
<td>log brain/blood ratio</td>
<td>exp</td>
<td>0.3</td>
<td></td>
<td>2</td>
<td>0.44</td>
<td>FALSE</td>
</tr>
<tr>
<td>is p-GP substrate</td>
<td>binary</td>
<td>0</td>
<td></td>
<td>1</td>
<td>0.46</td>
<td>FALSE</td>
</tr>
<tr>
<td>is BCRP substrate</td>
<td>binary</td>
<td>0</td>
<td></td>
<td>1</td>
<td>0.36</td>
<td>FALSE</td>
</tr>
<tr>
<td>Total polar surface area</td>
<td>exp</td>
<td>90</td>
<td></td>
<td>0.5</td>
<td>25.54</td>
<td>FALSE</td>
</tr>
<tr>
<td>Synthetic accessibility score</td>
<td>exp</td>
<td>3</td>
<td></td>
<td>0.5</td>
<td>0.90</td>
<td>FALSE</td>
</tr>
<tr>
<td>Number of rotatable bonds</td>
<td>exp</td>
<td>4</td>
<td></td>
<td>0.3</td>
<td>2.47</td>
<td>FALSE</td>
</tr>
<tr>
<td>pKa (acidity coefficient)</td>
<td>exp</td>
<td>6</td>
<td>10.5</td>
<td>0.3</td>
<td>3.13</td>
<td>FALSE</td>
</tr>
<tr>
<td>Number of H-bond donors</td>
<td>exp</td>
<td>3</td>
<td></td>
<td>0.5</td>
<td>1.19</td>
<td>FALSE</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>exp</td>
<td>180</td>
<td>450</td>
<td>0.5</td>
<td>88.45</td>
<td>FALSE</td>
</tr>
<tr>
<td>logP (lipophilicity)</td>
<td>exp</td>
<td>2</td>
<td>4</td>
<td>0.3</td>
<td>1.90</td>
<td>FALSE</td>
</tr>
</tbody>
</table>

**Exponential cost function term:**

\[
\text{cost} = \text{weight} \times \max\left(0, \exp\left(\frac{\max(\text{value} - \text{target}_{\text{max}}, \text{target}_{\text{min}} - \text{value})}{\text{scale}}\right) - 1\right)
\]

**Binary cost function term:**

\[
\text{value} = \text{weight if value} \neq \text{target}_{\text{min}}, 0 \text{ otherwise}
\]

Slides courtesy of Kevin McLoughlin (LLNL)
In Silico Validation Using Simulation

Property Prediction Models
- Efficacy
- Safety
- PK
- Developability
  - Machine learning parameter models
  - Human-level systems models

Multi-Parameter Optimization Loop

Design Criteria

Generative Molecular Design
- Proposes new molecules with optimized properties

Simulation
- In silico validation and evaluation

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

Experiment
- Human-relevant assays, complex in vitro models
- Chemistry Design & Synthesis

A Drug Candidate
- Retrain property prediction models

Working Compound Library
In Silico Simulation, Evaluation and Validation

- Utilizes multiple in silico models for “validating" generated compounds and generating critical insight and data
  - Molecular Dynamics
  - Docking
  - PBPK modeling
  - Pharmacodynamics (PD) with Mechanistic Models

- Addresses multiple key questions
  - Reasonable parameters for target optimization properties
  - Determination of binding likelihood
  - Will new chemical entity (NCE) reach the intended tissue?
  - Extend existing data
  - Bootstrap data for molecules in new domains
Human systems biology models are used to set molecular property targets.

**Pharmacokinetics**
- Absorption
- Distribution
- Metabolism
- Elimination

**Pharmacodynamics**
- Venous
- Arterial

Slide courtesy of Jim Brase, LLNL

Digital Twin Informed Systems Biology Models
An Exciting Landscape Ahead

Converging to Individual Precision Medicine
Building the **Global** Community for Digital Twins

---

**Envisioning Computational Innovations for Cancer Challenges**
Visit: [https://ncihub.org/groups/cicc/overview](https://ncihub.org/groups/cicc/overview)
Email: ECICC_Community@nih.gov

*Special Digital Twins Section*
*Ideas Labs and other opportunities*
NCI lead, Emily Greenspan

---

**Workshops and Conferences**
BioITWorld
(Upcoming) Fifth HPC Applications of Precision Medicine (HAPM22) June 2, 2022, Hamburg Germany
(Upcoming) Hood College Symposium, September 2022, Frederick, MD
(Upcoming) Eighth Computational Approaches for Cancer Workshop, November 13, 2022, Dallas, TX

---

**Collaborations with Frederick National Laboratory for Cancer Research**
Memos of Understanding
Workshops
Education and Training
Cooperative R&D Agreements (CRADA)
The challenge consists, in part, of the need to interrogate the enormous search space for determining the mapping across levels, which constitutes a many-to-many probabilistic problem. Computational innovation will be a key effort to help close these gaps. Portion of figure adapted with permission from ref. 67, Elsevier. Also shown are some of the ways in which QC can aid in the interrogation of these levels.
Key Takeaways

• The community is embracing biomedical digital twins at an accelerating rate
• Early biomedical digital twin efforts are drawing from successful use of digital twins in other disciplines
• There are several levels of biomedical digital twins in development and even use in healthcare
• Biomedical digital twins for patients is a collaborative effort
  – Across disciplines, communities, organizations, nations, cultures and demographics
• A combination of technologies are making biomedical digital twins feasible
• Advances will help make biomedical digital twins for human patients practical
Contact Information

- Cancer Data Science Initiatives
  - Eric Stahlberg – Eric.Stahlberg@nih.gov