Cancer, Data Science and Prediction in Medicine

Perspectives for the future

Eric Stahlberg, PhD
Executive Administrative Director
Institute for Data Science in Oncology
MD Anderson Cancer Center
Houston, Texas

Disclaimers and Conflicts of Interest

- **The comments and perspectives reflect my personal experience and do not reflect the views of MD Anderson Cancer Center.**

- **Select content has been drawn from prior public presentations and reports.**

- **Select content has been used with permission from the authors and creators.**

- **Forward looking statements are not commitments.**

Conflicts of Interest:
No conflicts to declare

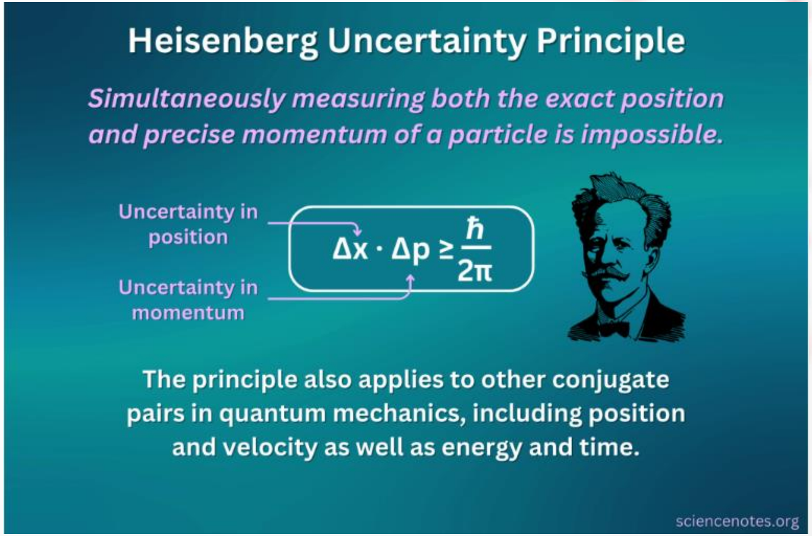
Today’s Presentation

- ***Background and Foundations***
- ***NCI and DOE Collaboration***
- ***ADMIRRAL and Multiscale Modeling***
- ***Clinical Digital Twins in Cancer***
- ***The Whole Patient***
- ***Institute for Data Science in Oncology***
- ***Perspectives for the Future***

Background and Foundations

Starting with Uncertainty

- **The Heisenberg Uncertainty Principle**
- The Schroedinger Equation
- The Hamiltonian Operator



Starting with Uncertainty

- The Heisenberg Uncertainty Principle

- **The Schroedinger Equation**

$$i\hbar\frac{\partial}{\partial t}\left|\Psi\right\rangle = \hat{H}\left|\Psi\right\rangle$$

- The Hamiltonian Operator

Starting with Uncertainty

- The Heisenberg Uncertainty Principle

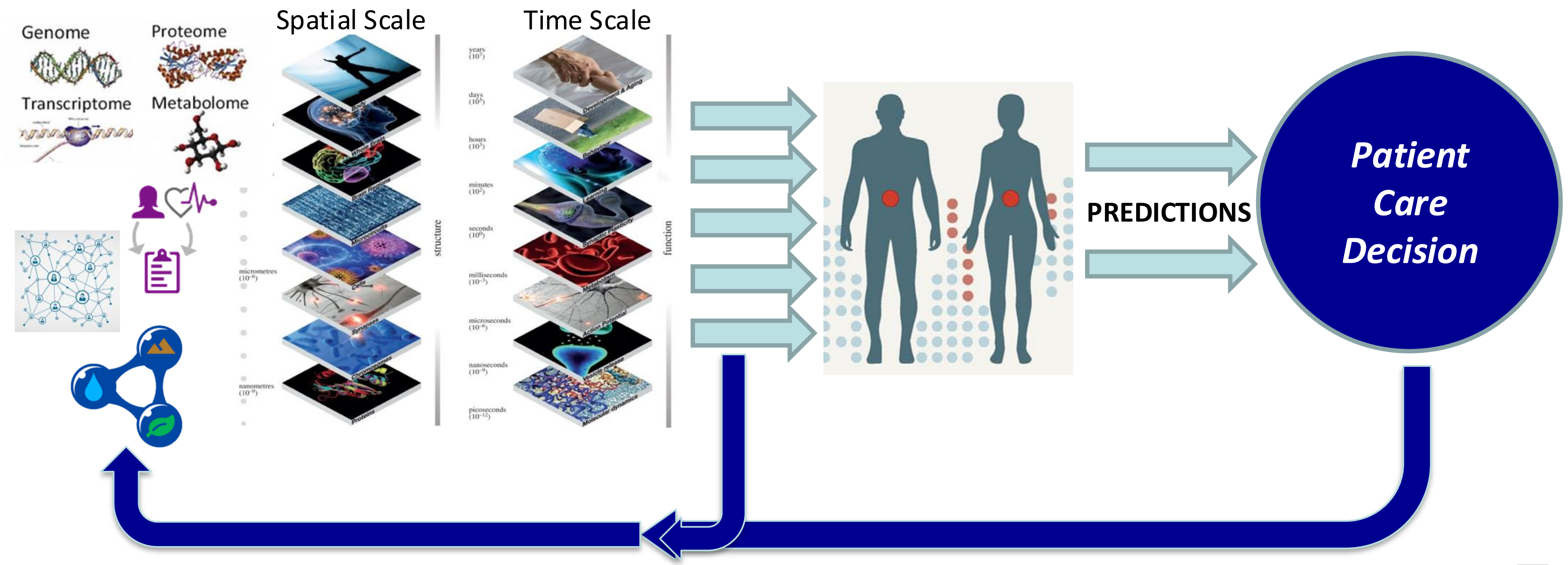
- The Schroedinger Equation

- **The Hamiltonian Operator**

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{n}\nabla_i^2 - \sum_{i=1}^{n}\frac{Z}{R_i} + \sum_{j>i}^{n}\sum_{i=1}^{n}\frac{1}{r_{ij}}.$$

The Challenge: Global Health Learning System

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



Slide content courtesy of Frederick National Laboratory for Cancer Research

National Academies of Science, Engineering and Medicine: 2023 Report on Digital Twins

- US National Academies of Science, Engineering and Medicine Report released December 2023

- Emphasized Research Gaps and Future Directions

Download:
https://www.nationalacademies.org/our-work/foundational-research-gaps-and-future-directions-for-digital-twins

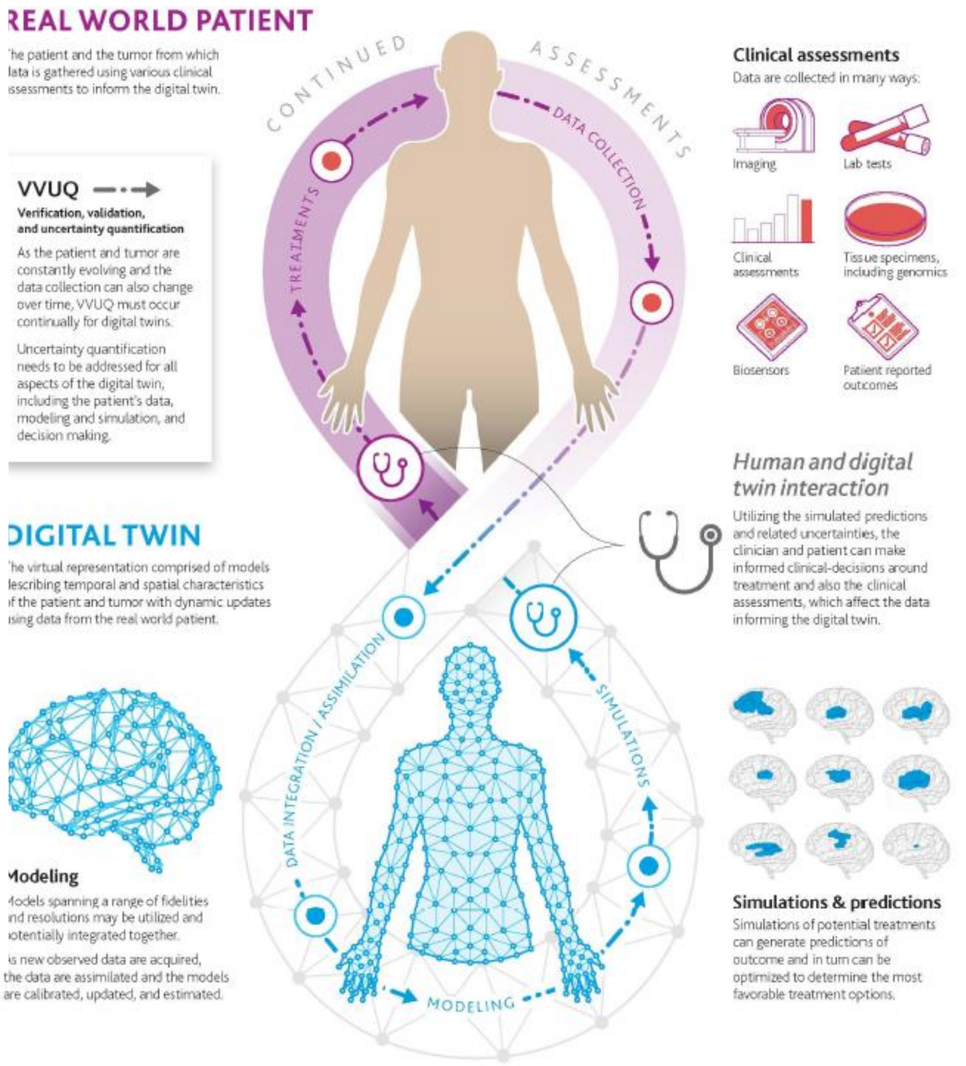
"A digital twin is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value. The bidirectional interaction between the virtual and the physical is central to the digital twin."

Digital Twin Elements

A Cancer Patient Example

- **Virtual representation**: Mechanistic and empirical models representing tumor growth, patient response, etc.
- **Physical counterpart**: Patient data collected from imaging studies, blood tests, and other clinical assessments
- **Bidirectional interaction**
  - updates the virtual representation to reflect characteristics of the individual patient
  - informs clinical decisions: treatments and clinical assessments
  - new clinical assessments inform and update the digital twin

NATIONAL ACADEMIES Sciences Engineering Medicine



MD Anderson Cancer Center

- **1.6 Million outpatient visits**
- **179,399 patients**
- **760 in patient beds**
- **20,986 surgeries**
- **14M pathology/laboratory procedures**
- **637,857 diagnostic imaging procedures**
- **24,498 employees**

MD Anderson is ranked #1 in the nation for cancer care by US News and World Report

BEST HOSPITALS
U.S.News & WORLD REPORT
CANCER
2023-2024

Source: MD Anderson ***FY23*** *Quick Facts*

MD Anderson Cancer Center Research

Jeff Siewerdsen, Ph.D., inside his 'surgineering' lab

Source: MD Anderson FY23 Quick Facts

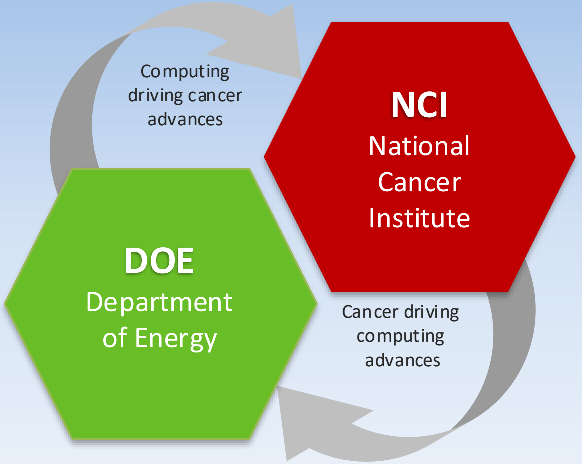
- **$1.2 billion on research**
- **1,568 clinical trials**
- **9,606 patients on clinical trials**
- **90 patents awarded**
- **25 drugs tested at MD Anderson have been FDA approved**

The NCI and the Department of Energy

A Transformative and Foundational Collaboration

NCI-DOE Collaborations Advance Cancer Research Using Computing

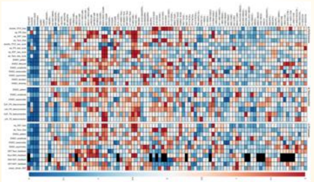
A collaboration between the Department of Energy and the National Cancer Institute



First MOU June 2016 between NCI and DOE

IMPROVE

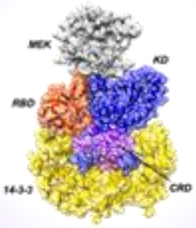
Innovative Methodologies and New Data for Predictive Model Evaluation



Rick Stevens (ANL)
Jeff Hildesheim (NCI)
Ryan Weil (FNLCR)

ADMIRRAL

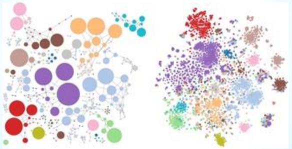
AI-Driven Multiscale Investigation of RAS-RAF Activation Lifecycle



Dwight Nissley (FNLCR)
Fred Streitz (LLNL)

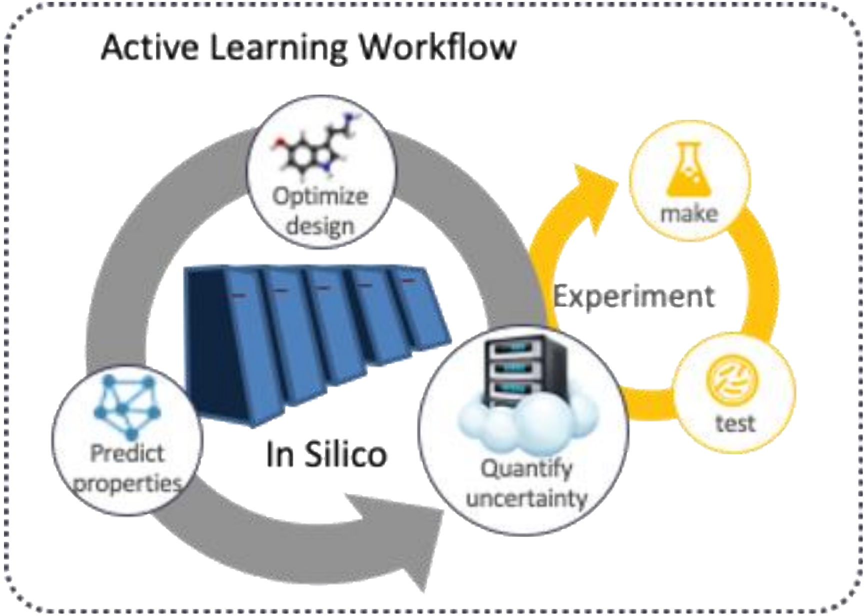
MOSSAIC

Modeling Outcomes using Surveillance data and Scalable AI for Cancer



Betsy Hsu (NCI)
Heidi Hanson (ORNL)

Model Driven Drug Discovery:
Accelerating Therapeutics for Opportunities in Medicine (ATOM)



Supported by LLNL, FNL, UCSF and the community

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

Open source on github
Models at modac.cancer.gov

Personal Health Monitoring

- **Emerging Opportunities**
  - Cloud computing
  - Affordable computing
  - AI, trusted and sustainable AI
  - Data sharing and data security
  - 5G
  - Blockchain
  - Medical IoT
  - Specialized computing
  - Mobile phones and apps

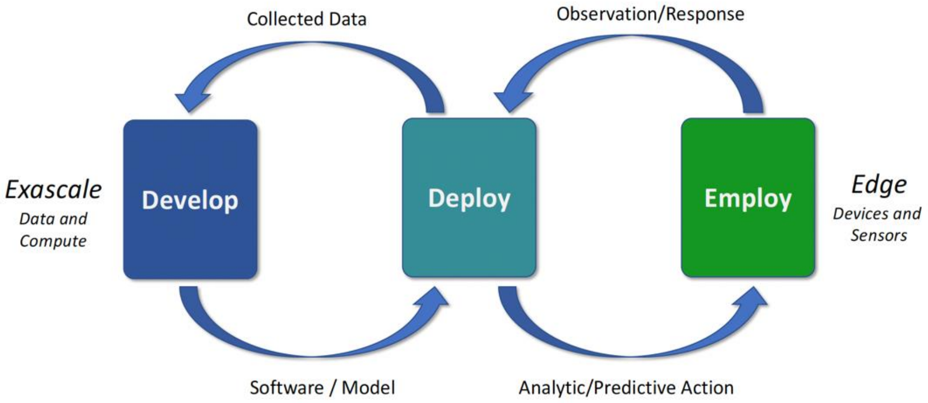
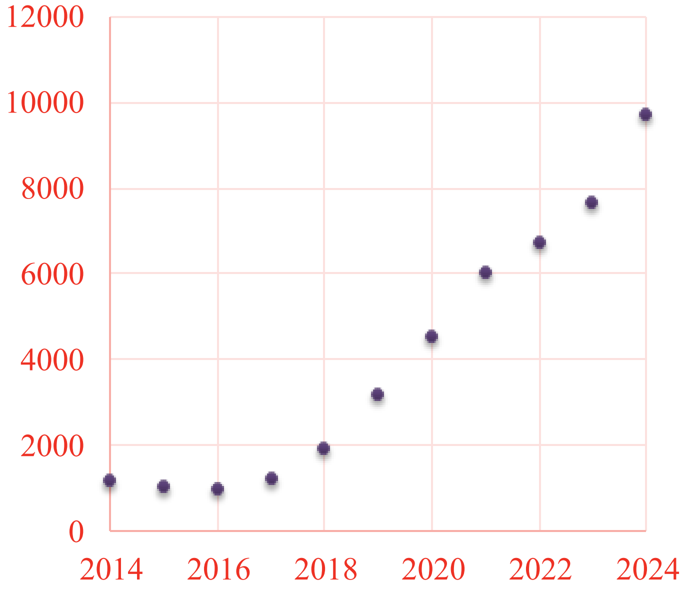


Image From 2019 Panel at SC19 on "Edge to Exascale"

AI and Machine Learning

- More AI and cancer publications added in 2024 than for the first 50 years combined
- 14% increase 2022 to 2023 (pre ChatGPT4)
  27% increase 2023 to 2024 (post ChatGPT4)
- Health AI publications between 5x and 6x cancer AI pubs
- FDA guidance for AI in software as a medical device
- NIH issued report on Ethical AI in 2024 focusing on transparency
- Advanced computing, neural networks and LLMs

New AI and Cancer Publications

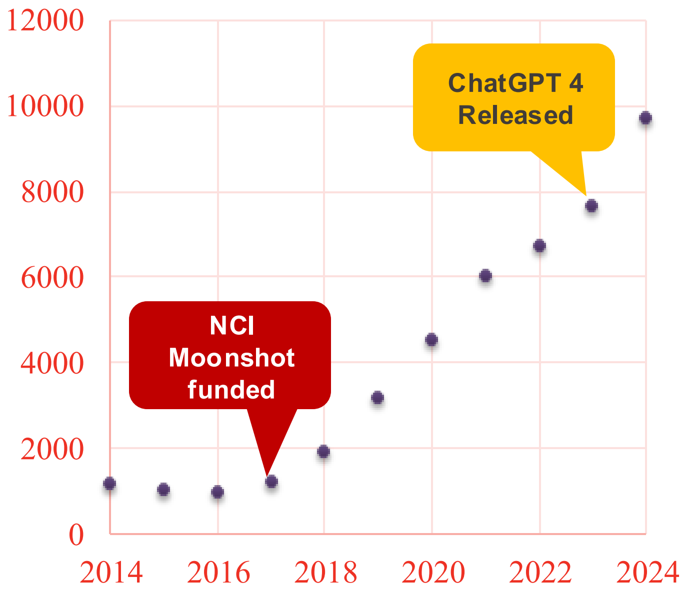


[8222 total publications 1961 to 2014]

AI and Machine Learning

- More AI and cancer publications added in 2024 than for the first 50 years combined
- 14% increase 2022 to 2023 (pre ChatGPT4)
  27% increase 2023 to 2024 (post ChatGPT4)
- Health AI publications between 5x and 6x cancer AI pubs
- FDA guidance for AI in software as a medical device
- NIH issued report on Ethical AI in 2024 focusing on transparency
- Advanced computing, neural networks and LLMs

New AI and Cancer Publications



[8222 total publications 1961 to 2014]

ADMIRRAL

Taking on the Multiscale Modeling Challenge

The ADMIRRAL Project

AI-Driven Multiscale Investigation of RAS-RAF Activation Lifecycle

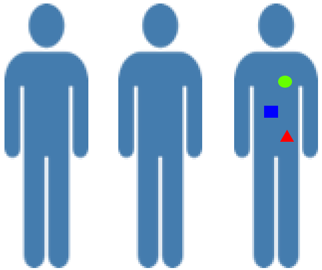
NCI-DOE Collaboration Symposium and Executive Committee Meeting
Sep 15th, 2025

Dwight V. Nissley
Frederick National Laboratory for Cancer Research
US National Institutes for Health

Helgi I. Ingólfsson and
Felice C. Lightstone
Lawrence Livermore National Laboratory
US Department of Energy

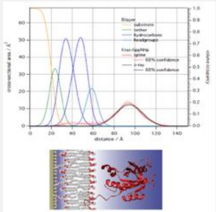
NCI-DOE Collaboration - ADMIRRAL

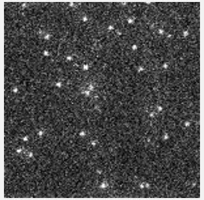
20-30% of cancers have mutated RAS



Current therapies ineffective against RAS-driven cancer

Biological and mechanistic insights to facilitate discovery of therapeutics



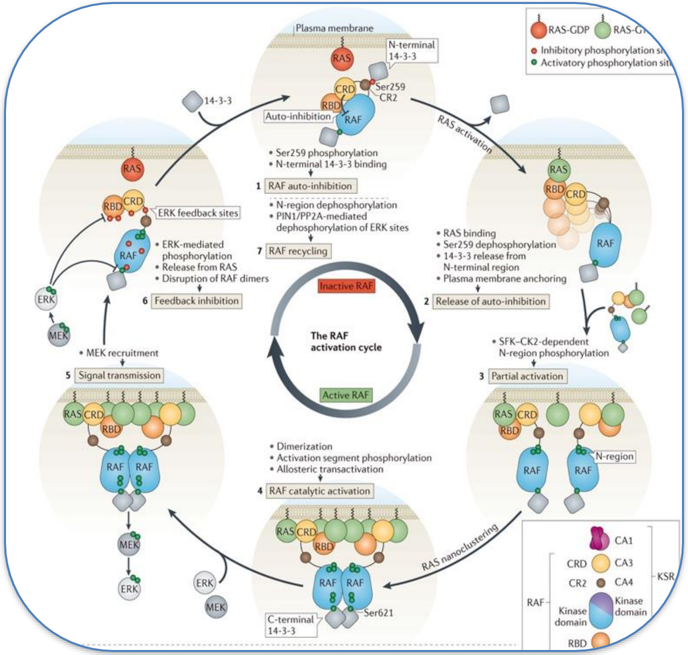


FNLCR

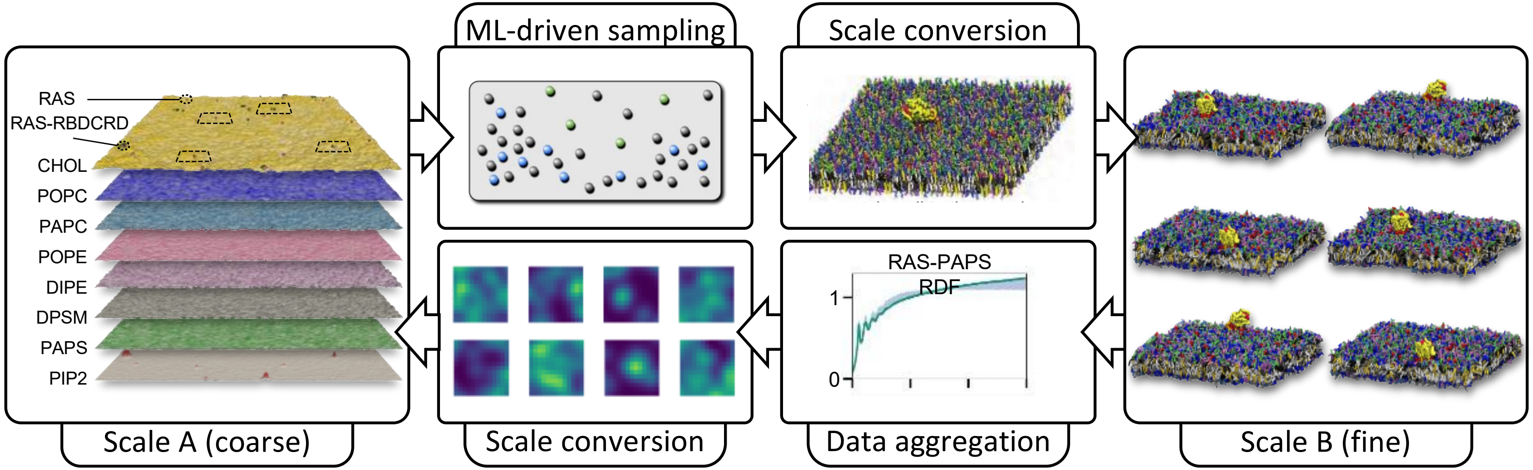
AI & Molecular Dynamics Modeling

LLNL

RAS/RAF Activation Lifecycle



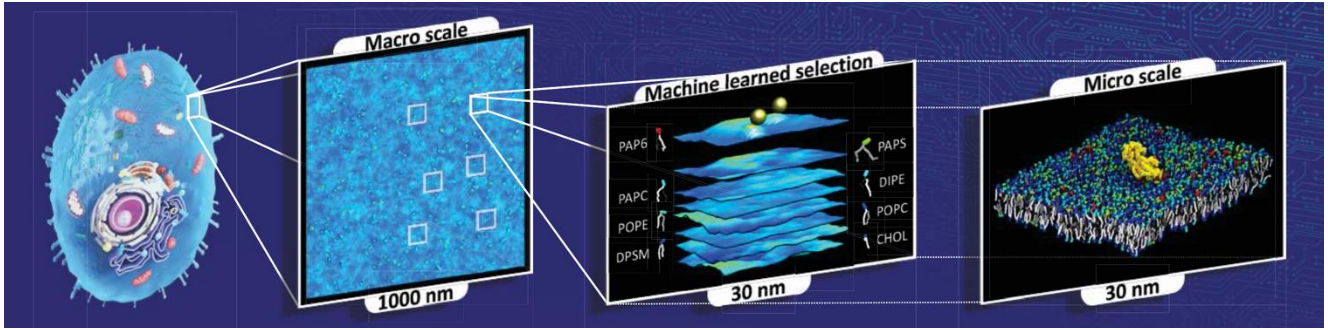
MuMMI, a Multiscale Machine-learned Modeling Infrastructure



MuMMI enables ML-driven ensemble-based coupled multiscale simulations

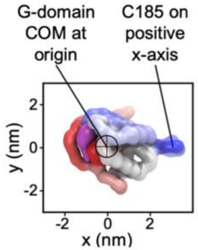
Bhatia, H.; Aydin, F.; Carpenter, T. S.; et al. 2023. A Convergence of Machine Learning and Multiscale Simulations.*COSB*

MuMMI multiscale simulations to highlight RAS plasma membrane dynamics

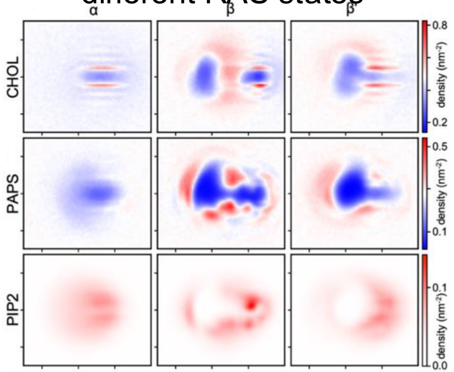


- Executed a very-large and well sampled simulation ensemble
- Revealed strong RAS-lipid coupling
- Lipid composition dictating RAS aggregation and membrane configuration

RAS



Averaged lipid densities for different RAS states



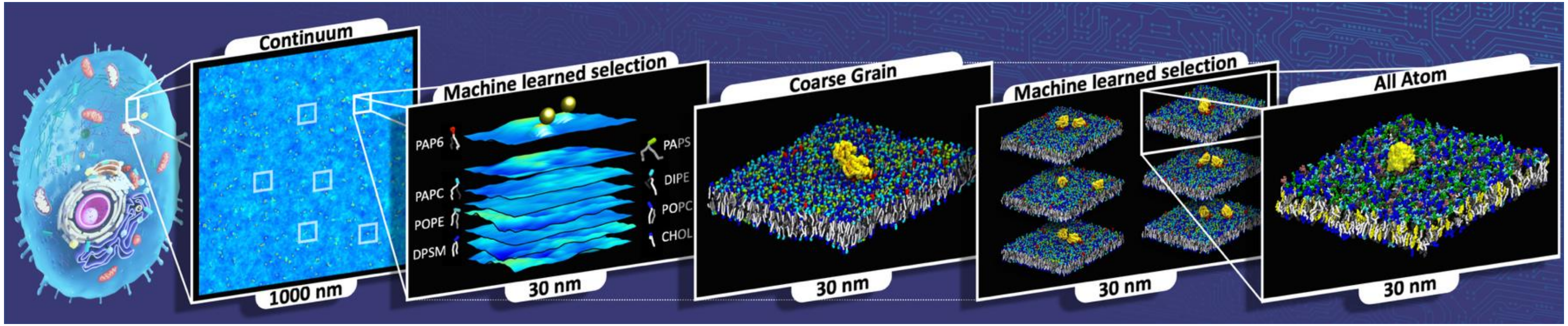
ML RAS state prediction from lipid snapshots

Actual \ Predicted	α	β	β'
α	91.4%	1.8%	6.8%
β	12.2%	71.8%	16.0%
β'	5.7%	10.3%	84.0%

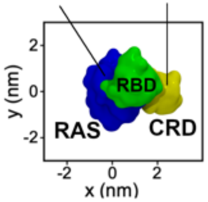
Ingólfsson, Neale, Carpenter, et al. 2022. Machine learning-driven multiscale modeling reveals lipid-dependent dynamics of RAS signaling proteins. *PNAS*

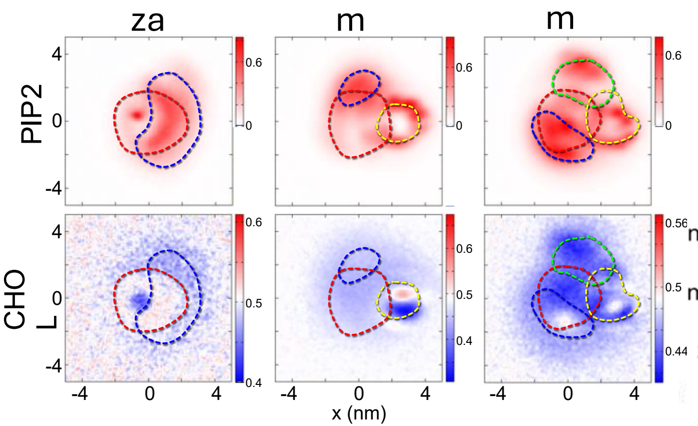
Di Natale et al. 2019. A massively parallel infrastructure for adaptive multiscale simulations: modeling RAS initiation pathway for cancer. In *SC19*.

Three-scale MuMMI, used to resolve RAS-RBDCRD membrane dependent dynamics



- A large three-sale simulations of RAS and RAS-RBDCRD demonstrated:
  - Strong lipid mediated protein association of RAS-RBDCRD
  - Lipid dependent orientational state
  - CRD secondary structure adaptation upon membrane insertion





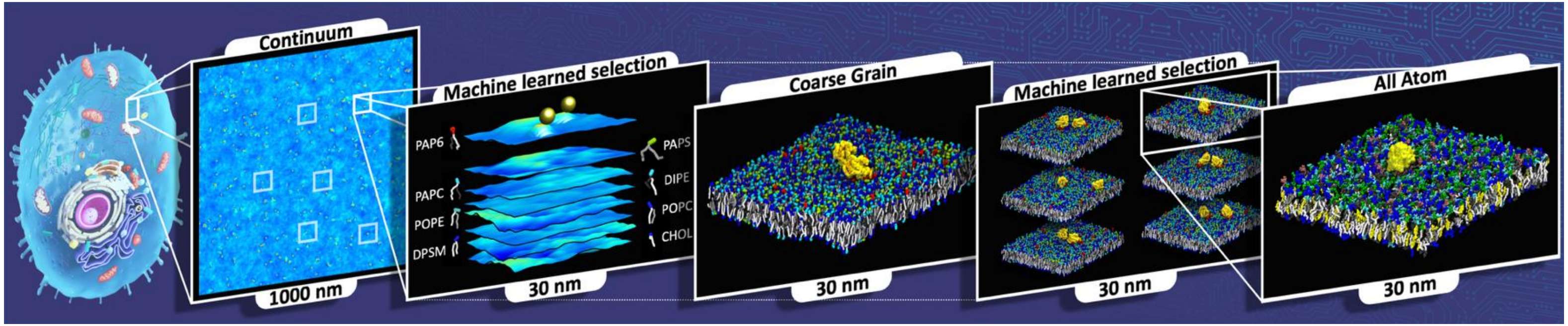
	α	β	β′	ma	mb	za
α	67.2	13.8	9.5	5.5	2.1	1.9
β	11.2	81.3	4.3	1.5	0.9	0.8
β′	38.4	16.4	36.3	4.9	2.2	1.8
ma	3.1	0.9	0.8	90.7	4.1	0.4
mb	1.1	0.6	0.4	5.5	91.8	0.6
za	23.5	15.9	5.6	5.9	10.9	38.2

Predicted State

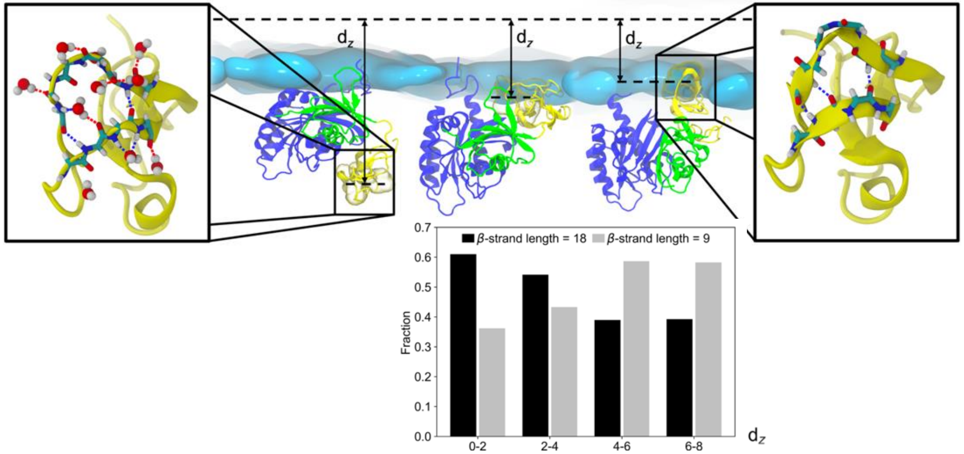
Carpenter et al. Dynamics and lipid membrane coupling of the RAS-RAF complex revealed via multiscale simulations. *Biophysical Journal,* Accepted (2025).

U.S. DEPARTMENT OF ENERGY
NIH NATIONAL CANCER INSTITUTE

Three-scale MuMMI, used to resolve RAS-RBDCRD membrane dependent dynamics



- A large three-sale simulations of RAS and RAS-RBDCRD demonstrated:
  - Strong lipid mediated protein association of RAS-RBDCRD
  - Lipid dependent orientational state
  - CRD secondary structure adaptation upon membrane insertion

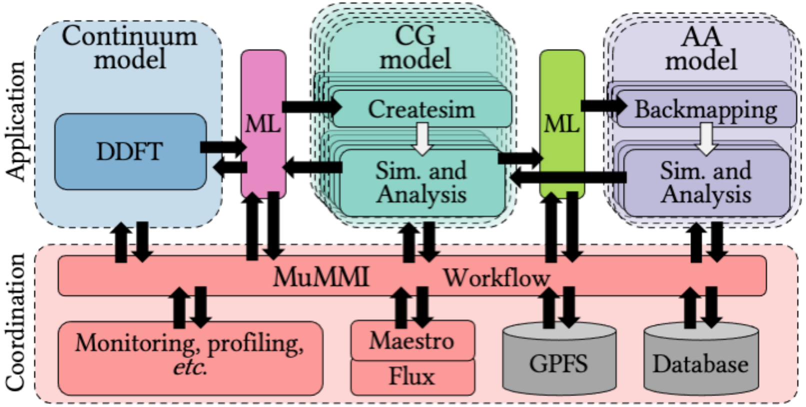


Shrestha, R., Carpenter, T.S., Van, Q.N. et al. Membrane lipids drive formation of KRAS4b-RAF1 RBDCRD nanoclusters on the membrane. *Commun Biol* 7, 242 (2024).

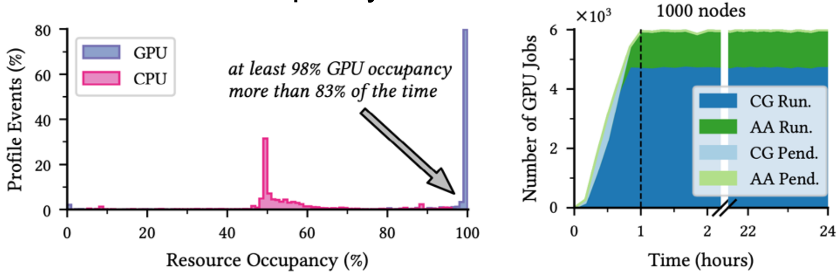
MuMMI efficiently utilizes heterogenous resources and scales across some of the worlds largest supercomputers

- Automatic and generalizable framework for coupling scales
- Ran full machine runs on *Sierra* and *Summit* (simultaneous use of 16,000 and 24,000 GPUs respectively)
- Using an internal Flux scheduler and optimized placement of simulation modules, MuMMI makes use of all CPU and GPU resources

The MuMMI workflow manages the simulation modules uses ML-guided selection



Resource occupancy and restart time on *Summit*

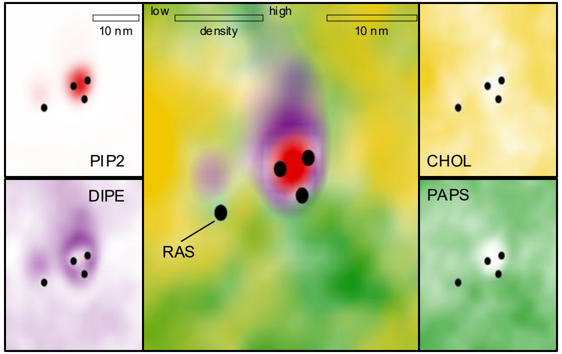


Di Natale et al. 2019. A massively parallel infrastructure for adaptive multiscale simulations: modeling RAS initiation pathway for cancer. In *SC19*.

Bhatia et al. 2021. Generalizable Coordination of Large Multiscale Ensembles: Challenges and Learnings at Scale. In *SC21*.

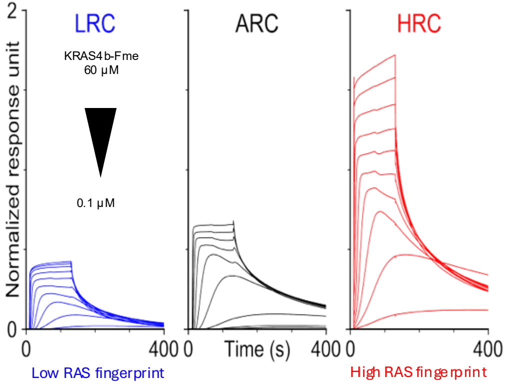
Computational model predicts lipid-RAS dynamics that influence clustering of RAS (and RAF) to initiate signaling

Lipid fingerprints associated with RAS clusters



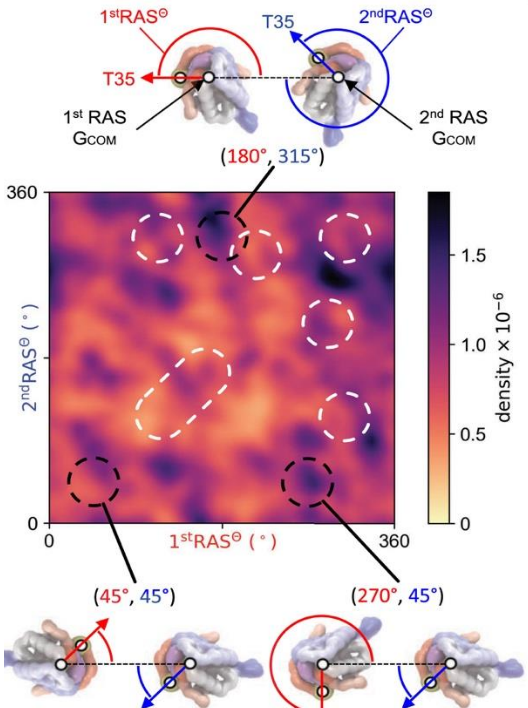
PIP2 + DIPE + CHOL + PAPS

RAS binding correlates with fingerprint

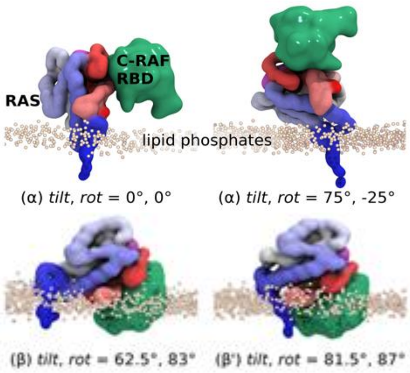


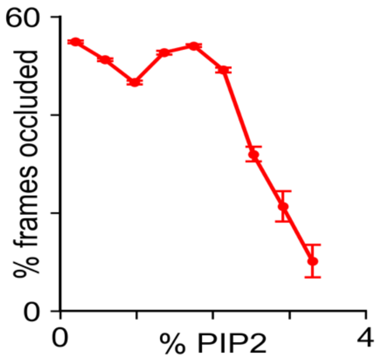
No preferential RAS interface

- Multimers transiently associate via a broad ensemble
- Multimer formation driven by lipids not specific interfacial contacts

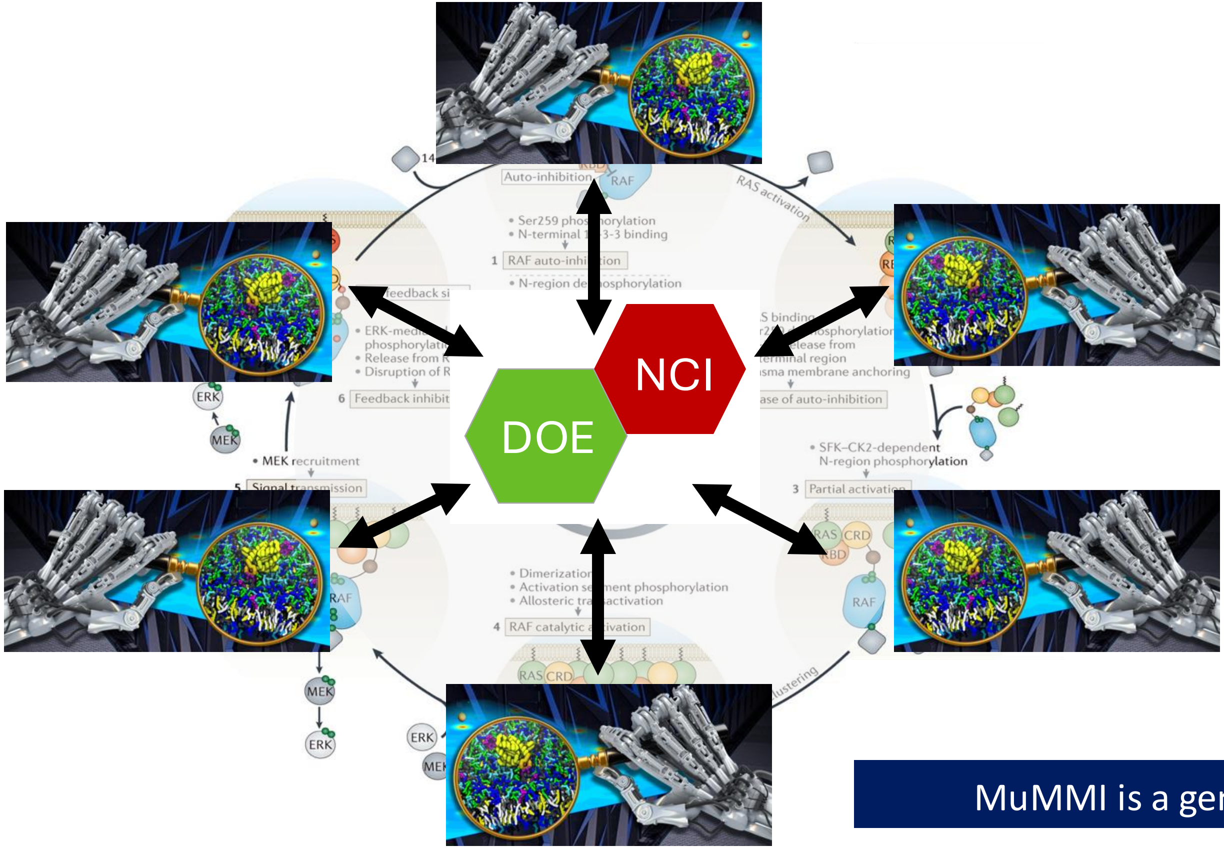


Lipid fingerprints influence RAS orientations (states) and clustering





Mini-MuMMI resource can be used for other projects

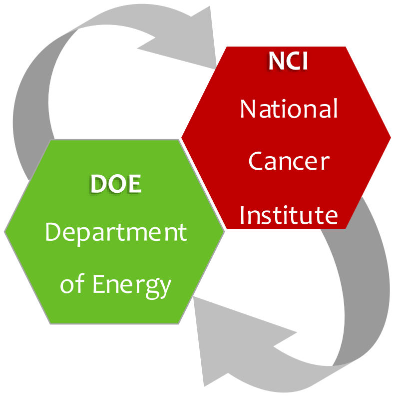


NCI/DOE collaboration(s) with new targets

- Leverage MuMMI's evolving generalizability and usability
- Collaboration with funded NCI partners
- Small "campaigns" using mini-MuMMI
- NCI/DOE leadership

MuMMI is a general computational biology tool

Connections spurred by the Pilot 2 and ADMIRRAL NCI-DOE collaboration projects



NIST National Institute of Standards and Technology

National Magnetic Resonance Facility at Madison

Argonne National Laboratory

Los Alamos National Laboratory

Oak Ridge National Laboratory

IBM

accenture

Harvey Mudd College Clinic Program

Molecular modeling

HPC utilization

HPC workflows

Accessible multiscale simulations

Medical College of Wisconsin

Different structure function assays

NIH National Cancer Institute

Lawrence Livermore National Laboratory

Macro scale modeling

San José State University 1857

NIH National Institutes of Health
Turning Discovery Into Health

Drug development

ML analysis

Force field parameters

UCSF University of California San Francisco

bridgebio

University of Calgary

university of groningen

cnrs

NOVA Universidade Nova de Lisboa

IIBEAA

Aarhus University

The NCI-DOE Collaboration Pilot 2 Team

Argonne National Laboratory : Arvind Ramanathan

Oak Ridge National Lab: Debsindhu Bhowmik, Chris Stanley

Lawrence Livermore National Lab: Fikret Aydin, Harsh Bhatia, Timo Bremer, Tim Carpenter, Joseph Chavez, Gautham Dharuman, Tim Hsu, Brian Van Essen, Jeremy Tempkin, Jim Glosli, Helgi Ingólfsson, Konstantia Georgouli, Loic Pottier, Sam Jacobs, Felice Lightstone, Robert Stephany, Adam Moody, Joseph Moon, Francesco Di Natale, Tomas Oppelstrup, Fred Streitz, Jayram Thathachar, Xiaohua Zhang

NCI's Frederick National Laboratory for Cancer Research: Dominic Esposito, Debanjan Goswami, Gulci Gulten, Dwight Nissley, Rebika Shrestha, Andy Stephen, Tommy Turbyville, Que Van

Los Alamos National Lab: Nick Hengartner, Christoph Jungans, Cesar Lopez, Chris Neale, Kien Nguyen, Sandrasegaram Gnanakaran, Sumantra Sarkar

In collaboration with:

IBM: Bruce D'Amora, Changhoan Kim, Claudia Misale, Lars Schneidenbach, Sara Schumacher; **SJSU:** Liam Stanton; **NIH:** David Durrant, Deborah Morrison; **NIST:** Frank Heinrich; **UCSF:** Frank McCormick

The NCI-DOE Collaboration ADMIRRAL Team

FNLCR
- Andy Stephen
- Que Van
- Suzanne Sandin
- Erik Larsen
- Sophie Krahnke
- Tommy Turbyville
- Rebika Shrestha
- Pedro Andrade Bonilla
- Rodrigo Caceres
- Dom Esposito
- Matt Drew
- Simon Messing
- Jenna Hull
- Kelly Snead
- Jen Mehalko
- Billy Burgan
- Dhirendra Simanshu
- Daniel Bonsor
- Caroline Dehart
- Robert D'Ippolito
- Grace Scheidemantle
- Alexandria Sohn
- Dwight Nissley

NCI
- Debbie Morrison
- David Durant
- Dan Ritt

UCSF
- Frank McCormick

NMRFAM
- Marco Tonelli

MCW
- Jason Sidabras

NIST, CMU
- Frank Heinrich

LLNL
- Fikret Aydin
- Harsh Bhatia
- Timo Bremer
- Tim Carpenter
- Tim Hsu
- Joseph Chavez
- Brian Van Essen
- Jeremy Tempkin
- Jim Glosli
- Mike Jones
- Konstantia Georgouli
- Mark Andrew Heimann
- Loic Eric Pottier
- Felice Lightstone
- Christa Manning
- Joseph Moon
- Francesco Di Natale
- Tomas Oppelstrup
- Claudio Santiago
- Xiaohua Zhang
- Fred Streitz

SJSU
- Liam Stanton

Cornell
- Robert Stephany

Unv. Groningen
- Siewert Jan Marrink
- Kasper Busk Pedersen

ENS de Lyon
- Paulo C. T. de Souza

Unv. Calgary
- Mariia Borbuliak
- Peter Tilemann

U.S. DEPARTMENT OF ENERGY | NIH NATIONAL CANCER INSTITUTE

Computing support: LLNL computing and Oak Ridge Leadership Computing Facility

Resources availability

MuMMI Framework https://github.com/mummi-framework
Includes mummi-core and mummi-ras codebases.

MuMMI protein sampling version + mini-MuMMI
open-source review underway.

ddcMD https://github.com/LLNL/ddcMD
A fully GPU-accelerated molecular dynamics program for the Martini force field.

ddcMDconverter https://github.com/LLNL/ddcMDconverter
Converts between ddcMD and GROMACS for files.

MemSurfer https://github.com/LLNL/MemSurfer
MemSurfer is a tool to compute and analyze membrane surfaces found in a wide variety of large-scale molecular simulations.

MDAnalysis https://github.com/XiaohuaZhangLLNL/mdanalysis
MDAnalysis modified for use with ddcMD output.

Martini parameters http://cgmartini.nl/index.php/force-field-parameters/lipids and https://github.com/Martini-Force-Field-Initiative/M3-Lipid-Parameters
Refined Martini 2 and 3 lipid parameters that reproduce experimental phase diagram have been uploaded for public use.

Maestro https://github.com/LLNL/maestrowf
Maestro Workflow Conductor is a Python tool and library for specifying and automating multi-step computational workflows both locally and on supercomputers.

Flux https://github.com/flux-framework
Flux is a next-generation resource and job management framework.

DataBroker https://github.com/IBM/data-broker
The Data Broker (DBR) is a distributed, in-memory container of key-value stores enabling applications in a workflow to exchange data through one or more shared namespaces.

Dynim https://github.com/LLNL/dynim
Framework for dynamic diversity sampling used for coupling pairs of scales in MuMMI.

PytarIdx https://github.com/LLNL/pytaridx
Framework for creating robust archives of data to allow MuMMI to handle up to billions of files.

Macro model open-source review just finished.

RAS-RAF models
Maide available with each publication on https://bbs.llnl.gov/data

Simulation data
Pilot 2 data three campaigns totaling >350TB. Has been open-sourced can be accessed on MoDac and later ORNL Constellation.
ADMIRRAL data, three campaigns totaling >200TB. Under open-source review will be available on ORNL Constellation.

Clinical Digital Twins in Cancer

Patient Level Models for Patient Level Decisions

Digital Twins in Radiation Oncology and Beyond

Heiko Enderling, PhD, FSMB
Department of Radiation Oncology
Institute for Data Science in Oncology

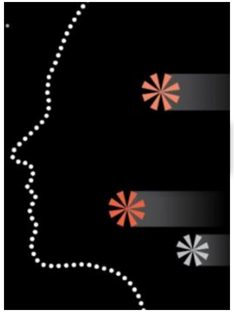
henderling@mdanderson.org

PRISM
PERSONALIZED RADIOTHERAPY WITH INTEGRATED SCIENTIFIC MODELING

THE UNIVERSITY OF TEXAS MD Anderson Cancer Center
Making Cancer History®

Digital Twin

Dr. Caroline Chung



"*A digital twin is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value. The bidirectional interaction between the virtual and the physical is central to the digital twin.*

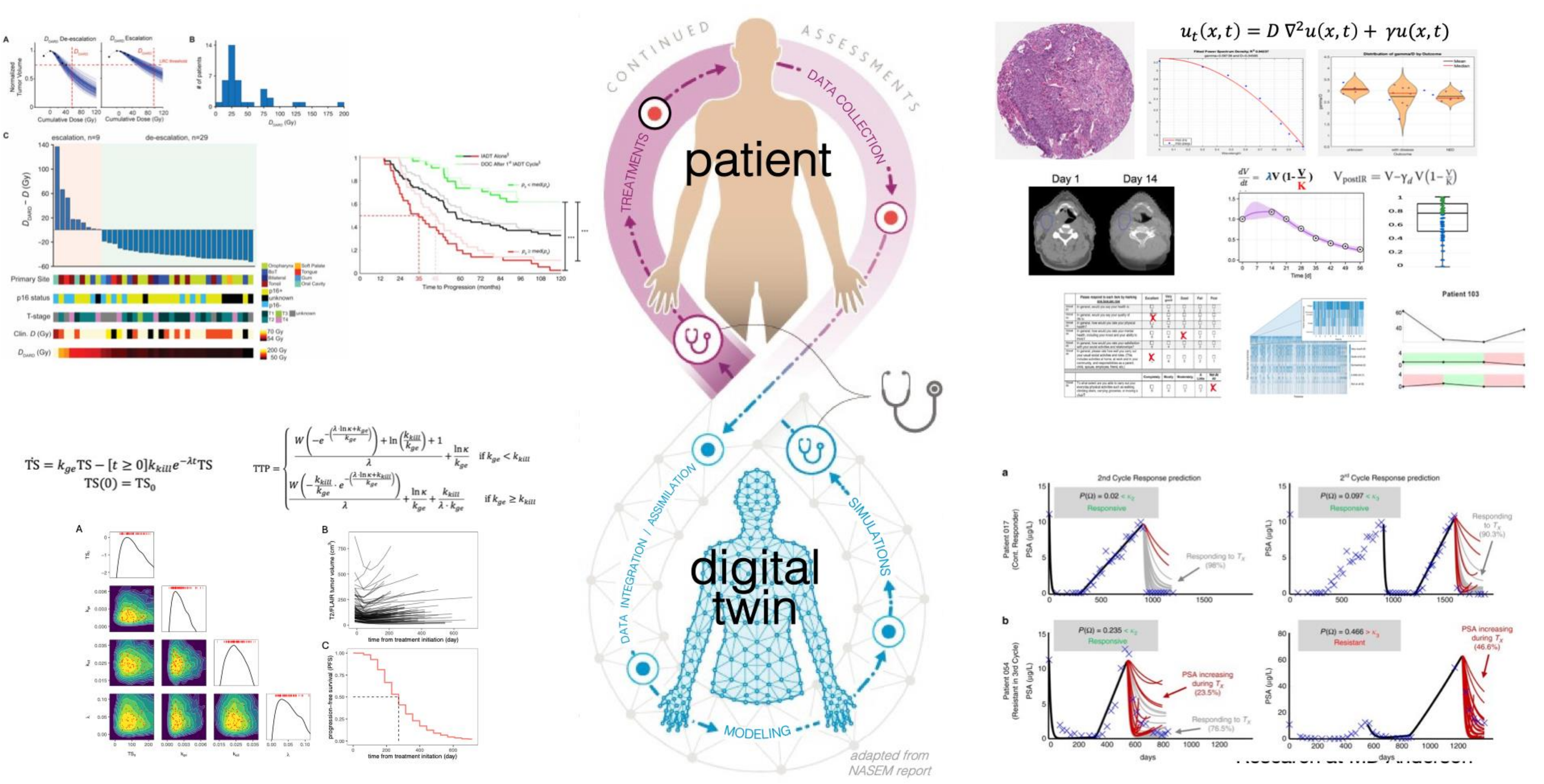
https://nap.nationalacademies.org/catalog/26922/opportunities-and-challenges-for-digital-twins-in-biomedical-research-proceedings

NATIONAL ACADEMIES Sciences Engineering Medicine

Committee's definition builds on a definition from an AIAA and AIA Position Paper (2020)

Courtesy of Dr. Caroline Chung

Digital Twins – The future of personalized cancer care



Dr. Jimmy Caudell
Moffitt Cancer Center

Can we use a digital twin

- to simulate head and neck cancer radiotherapy ?
- to predict patient-specific radiation responses ?
- to personalize radiotherapy dose and dose fractionation ?

PRiSM
PERSONALIZED RADIOTHERAPY WITH INTEGRATED SCIENTIFIC MODELING

Research at MD Anderson

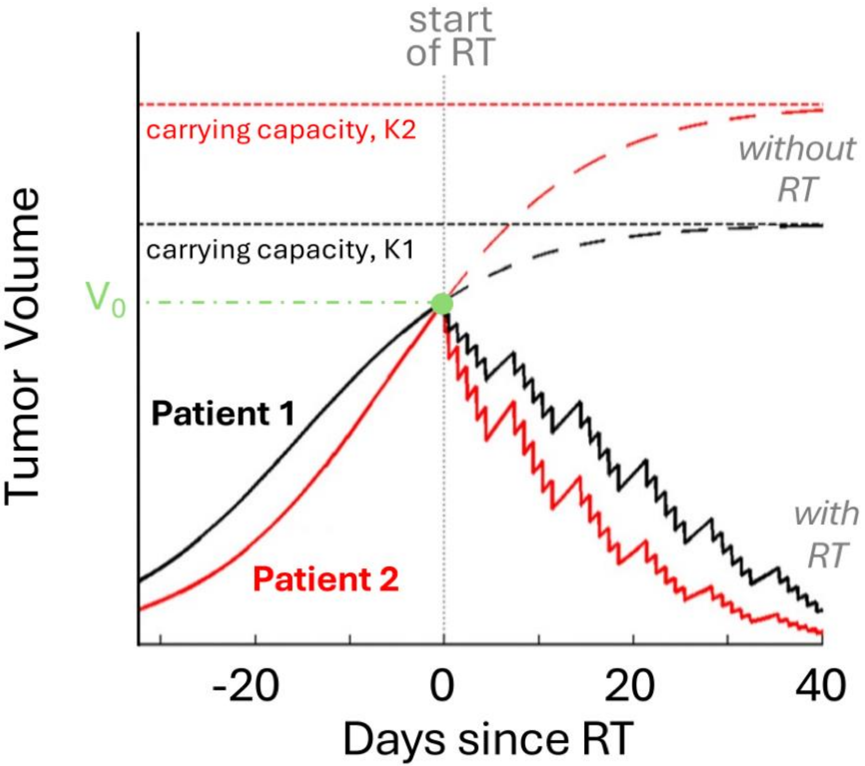
Proliferation Saturation Index, PSI

Dr. Sotiris Prokopiou

$$\frac{dV}{dt} = \lambda V\left(1 - \frac{V}{K}\right)$$

$$V_{postIR} = V - \gamma_d V\left(1 - \frac{V}{K}\right)$$

$$\textbf{PSI} = \frac{V_0}{K}$$



Prokopiou et al. Radiation Oncology (2015) 10:159
DOI 10.1186/s13014-015-0465-x

RADIATION ONCOLOGY

RESEARCH Open Access

A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation

Sotiris Prokopiou[1], Eduardo G. Moros[2,3], Jan Poleszczuk[1], Jimmy Caudell[2], Javier F. Torres-Roca[2], Kujtim Latifi[2], Jae K. Lee[4], Robert Myerson[5], Louis B. Harrison[2] and Heiko Enderling[1*]

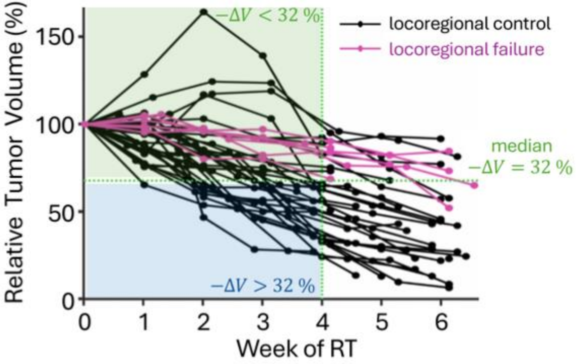
Research at MD Anderson

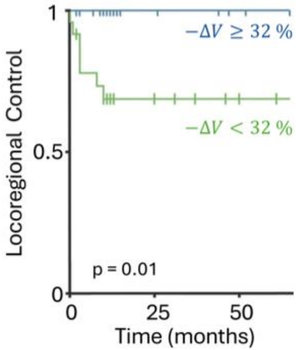
PSI determines response to radiation fractionation

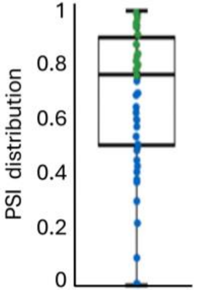
Dr. Mohammad Zahid

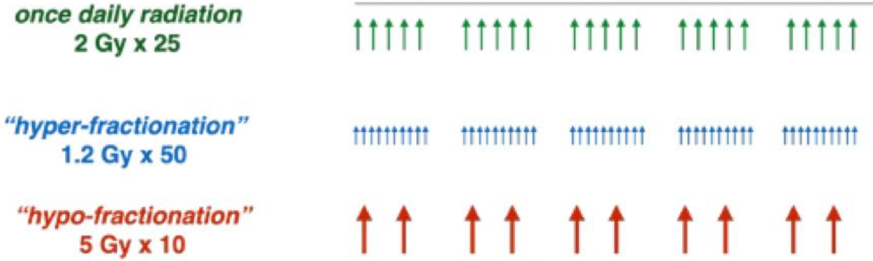
Dr. Jimmy Caudell
Moffitt Cancer Center

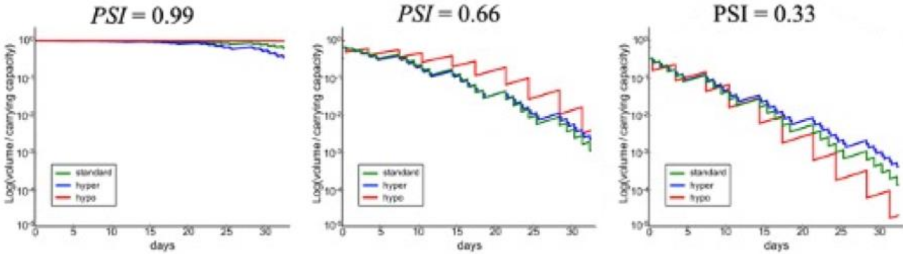
Dr. C. Dave Fuller









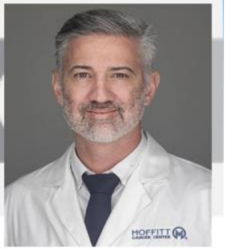


Zahid et al., *Red Journal, 2021*
Zahid et al., *Appl. Rad. Oncol.*, 2021
Zahid et al., *J. Pers. Med.*, 2021

Poleszczuk et al., *Bull. Math. Biol.*, 2018
Zahid et al., *Appl. Rad. Oncol.*, 2021

Kutuva et al., *Front. Oncol.*, 2023
Mohsin et al., *J. Theor. Biol.*, 2024

Worldwide first mathematics-guided prospective radiation fractionation clinical trial



Dr. Jimmy Caudell
Moffitt Cancer Center

Patient population	Math-model guided stratification	Primary Endpoint
HPV+ OPC T1-2, N0, M1	PSI < 0.75: 2 Gy q.d. *total dose: 66 Gy*	Tumor volume reduction at week 4 *by cone beam CT or MRI*
	PSI ≥ 0.75: 1.2 Gy b.i.d. (> 6 hours apart) *total dose: 60 Gy*	


ESTRO 2023 12-16 May 2023 Vienna, Austria

Prospective Trial of Personalized Fractionation in Low-risk HPV Positive Oropharyngeal Cancerroph

Jimmy Caudell, *USA*

Presentation Number: OC-0106

Abstract

Abstract Title: Prospective Trial of Personalized Fractionation in Low-risk HPV Positive Oropharyngeal Cancerroph

Authors: Jimmy Caudell[1], Michelle Echevarria[1], George Yang[1], Youngchul Kim[2], Kedar Kirtane[3], Julie Kish[4], Jameel Muzaffar[3], Christine Chung[3], Heiko Enderling[5]

Conclusion

Selection of personalized radiotherapy fractionation using the PSI model appears to be a promising approach, with a 12% improvement in the percentage of patients achieving a mid-treatment imaging response.

NIH U.S. National Library of Medicine
ClinicalTrials.gov

NCT03656133. Use of a Proliferation Saturation Index to Determine Personalized Radiotherapy for HPV+ Oropharyngeal Cancers (OPC).

The Whole Patient Challenge

A Frontier for Multiscale Integrated Modeling

Digital Twin Vision

•Virtual humans are *digital twins* of the human body that reproduce the way the body works.

Why it is significant?

- Current medicine is not predictive, save in the limited sense that doctors expect "Patient X" should respond like similar patients who have been studied in the past.
- Will be predictive and personalized

BUT

- Requires education and training of medics

Whole Patient Predictions

Slide courtesy of Peter Coveney, UCL

Health Digital Twins

- Improve health, wellness, and care for each person
- Represent the next frontier for precision medicine
- Focuses research on key gaps in translation
- Translation of research to clinical application
- Improve collaborative patient involvement
- An avenue to address medical disparities
- The low technical barrier to entry

Source: 2024 VHGS Report

Digital Twins in **Radiology**

Cardiovascular care with digital twin technology in the era of generative artificial intelligence

Radiation Dosimetry, Artificial Intelligence and Digital Twins: Old Dog, New Tricks

How artificial intelligence is transforming **nephrology**

Digital Twins for **Multiple Sclerosis**

Digital Twins of **human organs** are here. They're set to transform medical treatment.

Children's views on artificial intelligence and digital twins for the daily management of their **asthma:** a mixed-method study

Digital Twins and artificial intelligence in **metabolic disease** research

Recent Advances in Artificial Intelligence to Improve **Immunotherapy** and the Use of Digital Twins to Identify Prognosis of Patients with Solid Tumors

Generative artificial intelligence empowers digital twins in **drug discovery and clinical trials**

Title Source: PubMed January 2025 Search; MIT Review 2025

Uniting Global Perspectives Around Medical Digital Twins

- Research
- Infrastructure
- Industry
- Clinical Translation
- Community Health
- Government
- Global Collaboration

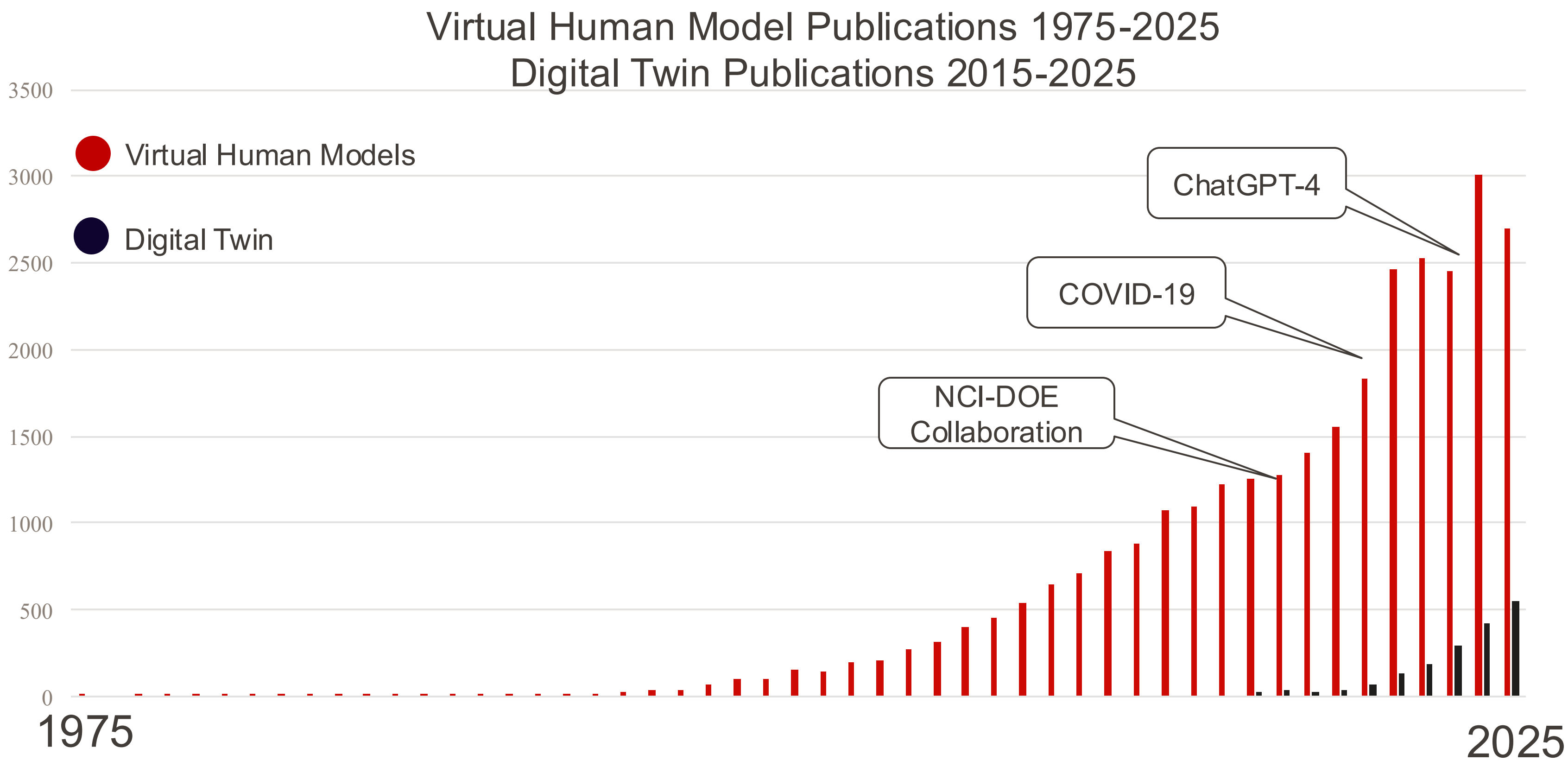
Second Virtual Human Global Summit in planning…

Medical Digital Twins: Looking Ahead (2023)

- **Focus on personal health and wellness, not just treatment**
- **Organize the global community**
- **Engagement and communication**
  - Communicate across stakeholders and expand participation
  - Engage with patients and patient advocacy groups
  - Build awareness of benefits
  - Identify problems to be addressed in the clinic
  - Develop training opportunities for stakeholders
  - Build trust in technologies
- **Technical foundations**
  - Create framework to develop individual patient health baselines
  - Provide structure for mechanistic models, AI models, and data to move ahead together
  - Identify approaches for sustainable and equitable access to data

Source: The First Virtual Human Global Summit: Prepublication Meeting Report, https://www.osti.gov/biblio/2428904

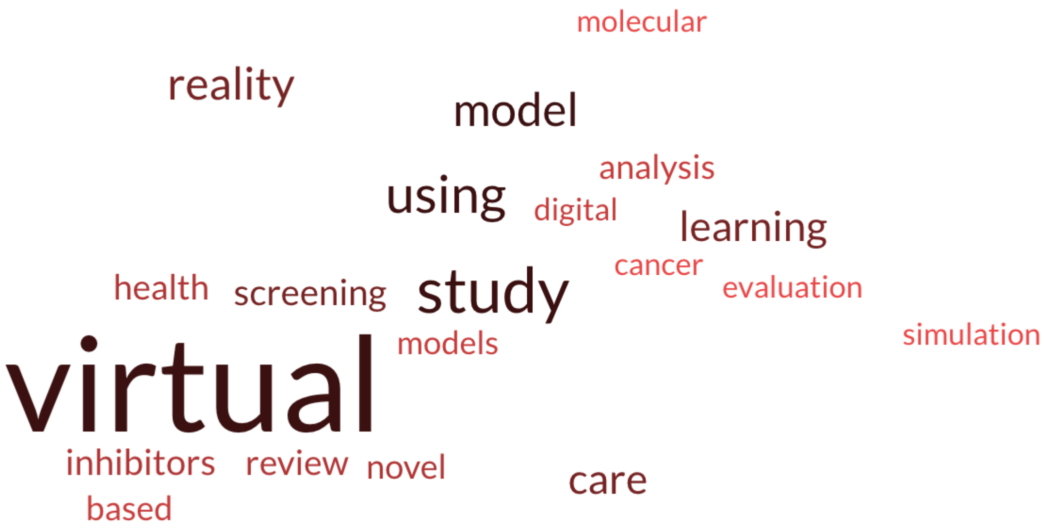
50 Years of Virtual Human Models



Source: Pubmed query October 18, 2025

Publication Activity in Virtual Human Models 2024-2025

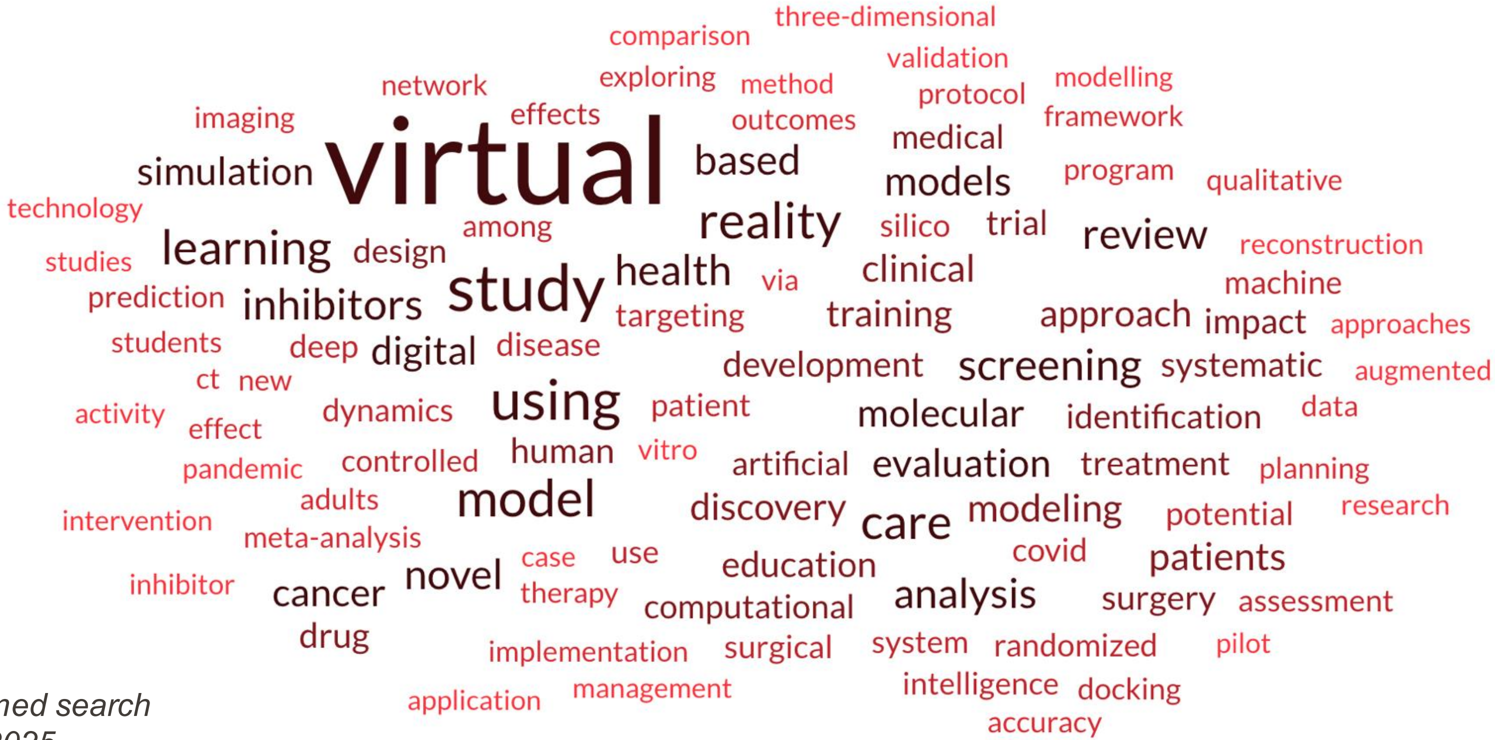
Top 20 words from 5710 titles virtual human models query



Source: Pubmed search October 18, 2025

Publication Activity in Virtual Human Models 2024-2025

Top 100 words from 5710 titles - virtual human models query



Source: Pubmed search
October 18, 2025

January 2023

NASEM Workshop on Biomedical Digital Twins

July 2023

EDITH Draft Roadmap Released

October 2023

First Virtual Human Global Summit

In 2023…

- **January 2023**: NASEM Workshop on Biomedical Digital Twins
- **July 2023**: EDITH Draft Roadmap Released
- **October 2023**: First Virtual Human Global Summit
- **November 2023**: First Digital Twin Workshop at SC23
- **December 2023**: NASEM Digital Twin Study Released
- **December 2023**: NCI Releases first NOSI for Cancer Digital Twins
- **January 2024**: First EDITH Ecosystem Public Meeting
- **March 2024**: NSF, NIH, FDA release first RFA for Biomedical Digital Twins
- **June 2024**: MD Anderson hosts first Digital Twin Summit for Cancer
- **July 2024**: Second EDITH Ecosystem Public Meeting
- **October 2024**: ARPA-H Launches CATLYST Program
- **December 2024**: EDITH: VHT Ethics and Code of Conduct
- **May 2025**: Second NSF, NIH, FDA RFA in BDTs closes
- **October 2025**: Second MD Anderson Cancer DT Summit
- **October 2025**: Second Virtual Human Global Summit

And in just two years…

Virtual Human Global Summit II

October 23-24, 2025

- 16 countries represented 3 continents: Austria, Belgium, France, Germany, Greece, India, Italy, Luxembourg, the Netherlands, New Zealand, Norway, Poland, Spain, Switzerland, UK, USA
- Two venues: Barcelona and New York City
- Over 120 attendees
- More than 7 hospital systems
- Industry including major pharma, cloud computing, small businesses and entrepreneurs
- Multiple international research institutes, centers and universities

Virtual Human Global Alliance Vision

- Community development and organization worldwide
- FAIR digital assets for virtual human model applications
- Trust, exchange and interoperability
- Collaborations and cross-organizational efforts
- Education, advocacy and policy
- Access, affordability, and cooperation

Contact: vhga@virtualhumansummit.org

Wednesday, September 17, 2025

NIH launches landmark project on whole-person health and function

The National Institutes of Health (NIH) has launched an effort to advance research on whole-person health and create an integrated knowledge network of healthy physiological function. Whole person health involves looking at the whole person—not just separate organs or body systems—and considering multiple factors that promote health. For example, a multicomponent lifestyle intervention including healthy diet, physical activity and stress management may improve multiple and interconnected aspects of health including cardiovascular (e.g. blood pressure), metabolic (e.g. glucose metabolism) and musculoskeletal function (e.g. muscle strength).

"Biomedical research is largely organized around the study of specific organs and diseases. In contrast, we do much less research on health itself, which is an integrated process involving the whole person," said Helene M. Langevin, M.D., director of NIH's National Center for Complementary and Integrative Health, which leads the NIH-wide program.

The five-year research initiative will proceed in several stages, drawing from existing scientific knowledge to develop a complete, working model of healthy human physiology. It will build on the NIH Human Reference Atlas and the Human BioMolecular Atlas Program (HuBMAP) to connect the complex anatomy and function of the body's different organs and systems into a single "map."

Future stages of the project will link common clinical measures, such as blood pressure, blood glucose and cholesterol, to major physiological functions. This initiative will also populate the framework with existing human data and ultimately build and test an interactive model of whole-person health.

"By organizing healthy physiological function into a whole-body knowledge network, researchers will be able to explore scientific questions about health in a new way," said Dr.

The Institute for Data Science in Oncology

Catalyzing and Driving Translational Cancer Data Science

Unlocking the power of data and enabling impactful change through team data science

Intelligent Systems for Quality and Safety:

Safer and more cost effective.
Improve cancer patient safety for our patients and beyond

Streamlining Operations for Access:

Continuous monitoring and modelling of our operational performance for improved access/scheduling

Predicting Outcomes for our Patients:

Finding the optimal combination of our current and emerging treatments.

And so much more…

The Talent:
Attract, Engage, and Build the Best Minds and Partnerships

+

The Culture:
Enabling Team Data Science

+

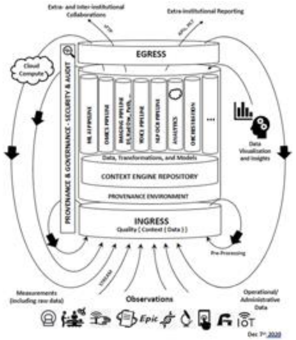
The Ecosystem:
Engineering a World-leading Data Science Capability

IDSO Enabling & Elevating the MD Anderson Data Science Ecosystem

- Bringing the power of data science to every decision we make means a ***massive transformation*** to MD Anderson.
- ***Every part of MD Anderson has a role to take in realizing the transformation.***

Talent - New skills; New career paths; Human Resources

Culture of Recognition - Team Data Science; Promotion and Tenure



Context Engine - People, Processes & Technology; Faculty; Staff; TDI; Compliance, and many more!

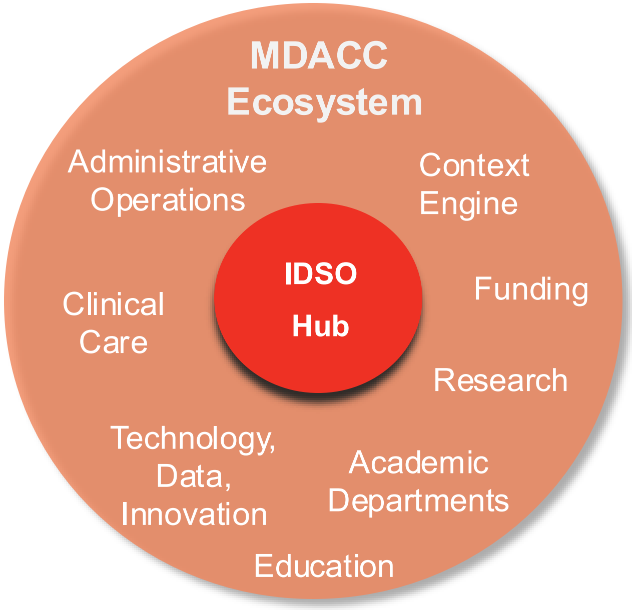
IDSO exists as a Hub within the MD Anderson Ecosystem

IDSO brings focus to impactful cancer data science

Foster collaborations and bridge communities

A ‘Hub’ for learning about cancer data science

Catalyst for moving cancer data science forward



IDSO Incubates Focus Areas for Translational Impact

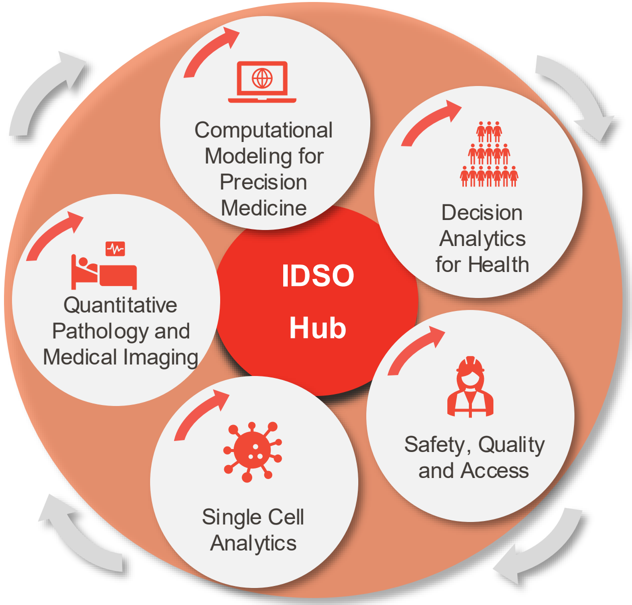
IDSO works to incubate data science advances in oncology

Focus Areas are the driving "flywheels" building momentum for organizational transformation

- Seeking opportunities
- Developing approaches
- Creating innovative solutions
- Translating for impact

Spin out new advances and capabilities

IDSO integrates FAs and helps build momentum on these areas across the institution



IDSO Coordinates Closely with the Institution

IDSO

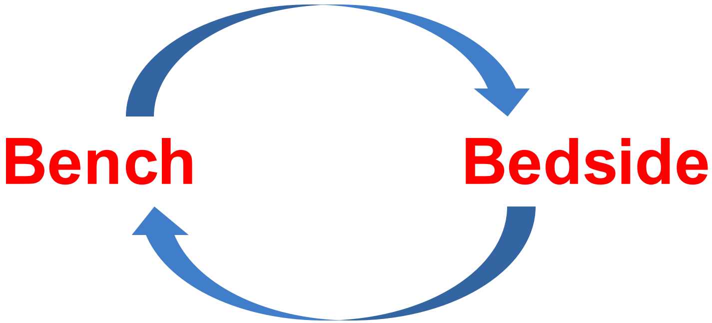
- Enables catalytic activities in data science
- Fosters transformative efforts
- Inspires innovative approaches
- Drives our data science transformation through:
  - Focus Areas, Fellowships, Affiliates, Partnerships

Critical Interdependencies for IDSO Success

- Data Ecosystem (internal and external)
- Clinical, Research, and Education Operations
- Administrative Operations
- Strong coordination with the TDI portfolio

Anticipation and Informing Strategy

- Maintaining awareness of changes and opportunities outside MDACC



Impactful data science will create new research questions in the fields of oncology and data science.

IDSO will stimulate and support cancer data science innovation into the future.

IDSO Program Impact Areas

Focus Area Impact
- Ongoing support for developed capabilities (e.g. CanSAR drug discovery tool)
- Incremental support for project resources (e.g. software, datasets, data scientists, etc.)
- Pilot project support (e.g. seed projects, prototypes, project management, etc.)

Education and Training
- Workshops and training opportunities (e.g. data science topics, guest speakers, hackathons, etc.)
- Fellowships and student support (e.g. individual fellowships, first year graduate students, etc.)
- Sustainable program support (e.g fellowship leadership, data literacy program, etc.)

Ecosystem and Community
- Collaboration and event support (e.g. JCCO fellowships, Rice University CORC, etc.)
- Essential software, licenses, and equipment
- Human-AI interactions (e.g. AI literacy, novel consents, decision-making with data, etc.)

IDSO Connects Across the Ecosystem

IDSO impacts MD Anderson.

- Culture and Education
- Model Lifecycle Management
- Data Management and Governance
- Responsible Data Science
- Human Experience

With a focus on bringing impact for our patients.

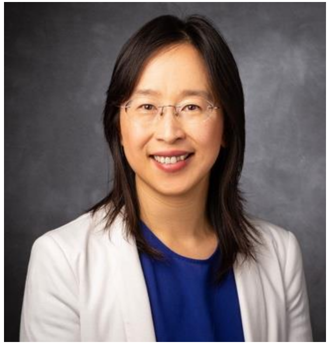
Purpose & Focus

IDSO integrates the tremendous power of data science with the institution's unparalleled scientific and clinical expertise to transform research and care to impact patients in ways only possible at MD Anderson.

To advance this collaborative, industry-leading approach, our teams have prioritized five focus areas:

- Quantitative Pathology and Medical Imaging
- Single Cell Analytics
- Computational Modeling for Precision Medicine
- Decision Analytics for Health
- Safety, Quality and Access

Focus Area 1 - Quantitative Pathology and Medical Imaging

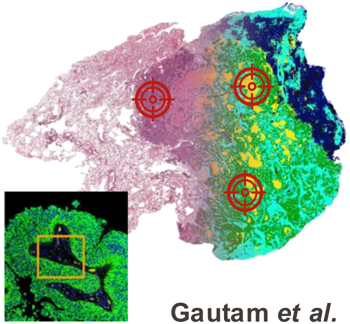


Yinyin Yuan, Ph.D.
Professor,
Translational Molecular Pathology
FA#1 Co-Lead

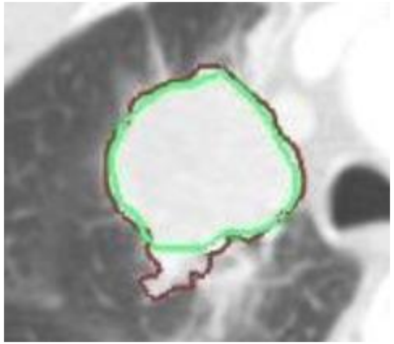
Caroline Chung, M.D.
Associate Professor,
Radiation Oncology and Diagnostic Radiology
TMI Initiative

VISION

Harness enormous amounts of rich, medical imaging and pathology data to provide accelerated, rigorous and actionable insight for patients, from diagnosis to intervention and follow-up.



Gautam *et al.*
2024 MICCAI

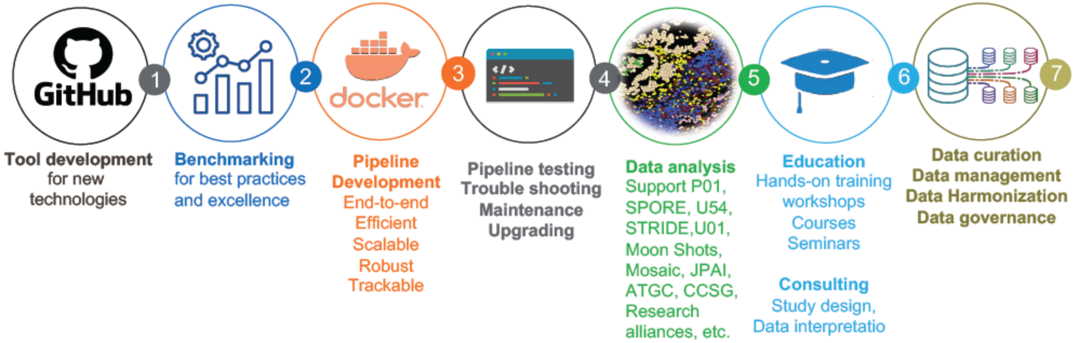


Focus Area 2 - Single Cell Analytics and Spatial Multi-omics

Linghua Wang, M.D., Ph.D.
Associate Professor,
Genomic Medicine
FA#2 Co-Lead

VISION

Build a world-class data science hub for single-cell and spatial multi-omics studies to accelerate the discovery and translation of cancer research at MD Anderson



Focus Area 2 is designed to work in close collaboration with the JPAI

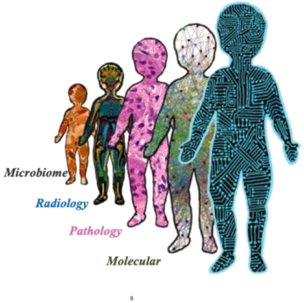
Focus Area 3 - Computational Modeling for Precision Medicine

Heiko Enderling, Ph.D.
Professor,
Radiation Oncology
FA#3 Co-Lead

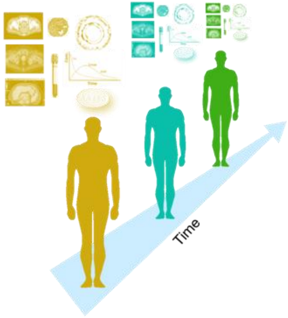
Bissan Al-Lazikani, Ph.D.
Professor,
Genomic Medicine
FA#3 Co-Lead

VISION

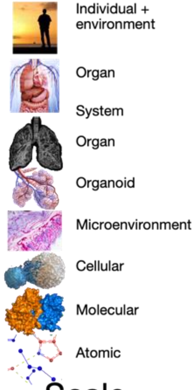
Create and leverage multiscale models of cancers from atomic to patient level to develop adaptive digital twins to identify optimal therapies for individuals that maximize response and minimize adverse events.



Factors



Time



Scale

Focus Area 5 - Safety, Quality, and Access

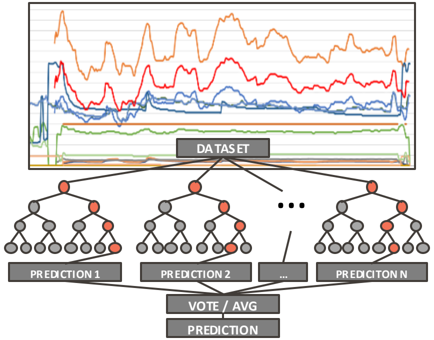
Jeffrey Siewerdsen, Ph.D.
Professor,
Imaging Physics
FA#5 Co-Lead

VISION

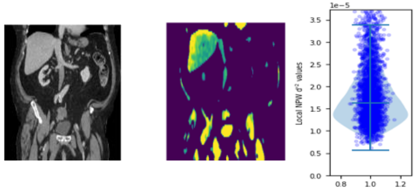
Develop automated approaches to increase access, safety and quality in cancer care.

ANESTHESIA DATA SCIENCE

Predicting:
Adverse Events, Complications,
Length of Stay

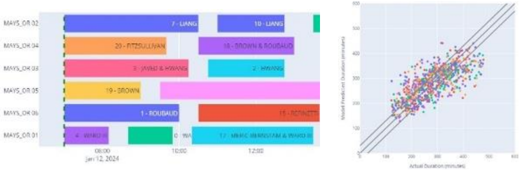


MEDICAL IMAGING QA



Conventional (“Noise”) → Task-Based Performance

OR SCHEDULE OPTIMIZATION



Case Duration Prediction

Acknowledgements

Cancer Team Data Science

IDSO Founding Co-Directors

David Jaffray, PhD
--*Senior VP and CTO*
Caroline Chung, MD, MS
--*VP and CDAO*

IDSO Co-leads

Yinyin Yuan, PhD
Caroline Chung, MD, MS
Linghua Wang, MD, PhD
Heiko Enderling, PhD
Bissan Al-Lazikani, PhD
Jeff Siewerdsen, PhD
Stephanie Schmidt, PhD
Amy Moreno, MD

IDSO Affiliates

Over 50 affiliates

IDSO Fellows

Over 40 fellows

IDSO External Advisory Board

IDSO Interal Advisory Committee

IDSO Supported SciTech Staff

More than 30 staff supported

IDSO Internal Parters

IDSO External Collaborators

Many organizations including:
Texas Advanced Computing Center
Texas Medical Center
Rice University
University of Texas Oden Institute
Texas A&M University
US Department of Energy
And many more…

IDSO Administrative Staff

Andrea Hawkins-Daarud, PhD
Shuhan Yang
Tara Blaylock
Nicholas Bisase
Renu Nargund, PhD

Perspectives for the Future

Observations, Challenges and Opportunities

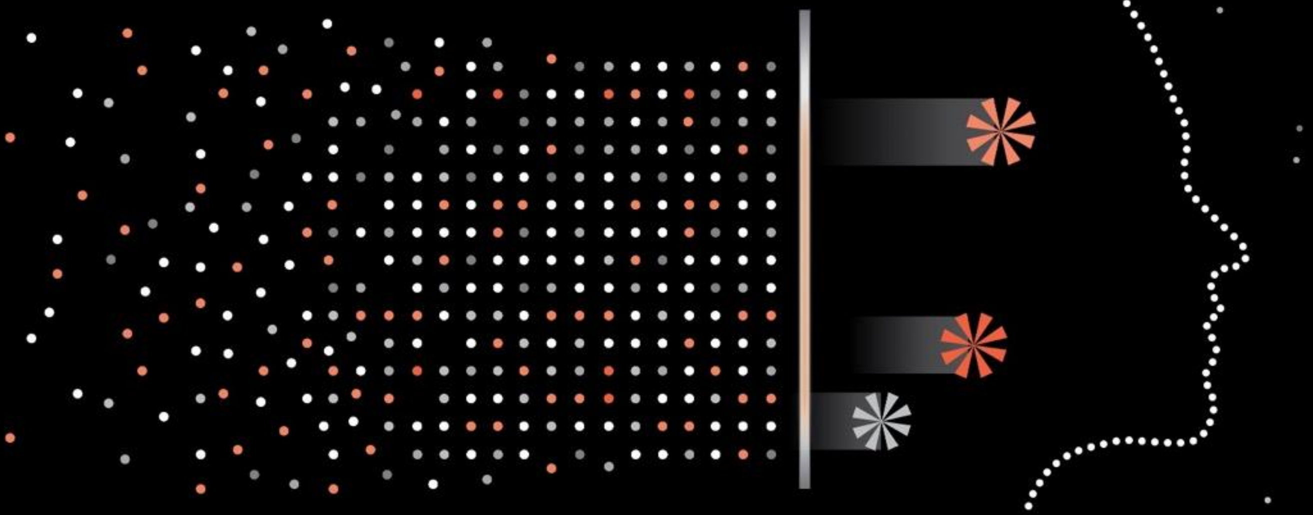
Some Takeaways

- Uncertainty is inherent in what is observed, modeled or predicted
- Models are used to provide insight and support decisions
- Balance insight and confidence required with complexity of solution

- Opportunities are rapidly expanding for mathematical and statistical innovation in health
- Innovations for integrating models across physiology and scale
- Methods for trust, reliability and availability for n=1 precision medicine practice
- Interoperability among computational models

- Novel collaborations across domains are hyper-catalysts for innovation
- Modeling human health and physiology is of global interest and value

Institute for Data Science in Oncology



THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center