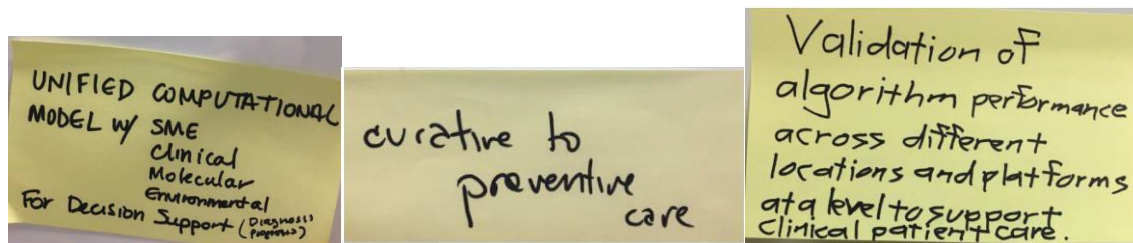# Envisioning Computational Innovations for Cancer Challenges (ECICC) Scoping Meeting Report

Executive Summary, Meeting Overview and Digital Twin Section

March 6-7, 2019
Livermore Valley Open Campus
Lawrence Livermore National Laboratory
Livermore, California

Organized by:

National Cancer Institute
Frederick National Laboratory for Cancer Research
Department of Energy
Lawrence Livermore National Laboratory

Contents

Executive Summary

In March 2019, the National Cancer Institute (NCI) and Department of Energy (DOE), along with the Frederick National Laboratory for Cancer Research (FNLCR) and Lawrence Livermore National Laboratory (LLNL) organized the first *Envisioning Computational Innovations for Cancer Challenges* (*ECICC*) *Scoping Meeting*. The meeting was a unique, two-day, highly interactive event (*see appendix A*) that brought together 74 cancer, bioinformatics, engineering, data and computational scientists at all career levels from research institutions across the United States (*see appendices E & F*).

Participants included:
- Grantees currently funded through related NCI programs including Informatics Technologies for Cancer Research (ITCR) and the Cancer Systems Biology Consortium (CSBC)
- Scientists from nine national laboratories (8 DOE and FNLCR)
- NCI and DOE program staff and personnel
- Industry representatives

Facilitated by Knowinnovation (KI), the meeting format featured a series of dynamic activities to identify compelling computational oncology challenges, problems whose solutions require innovative solutions from both cancer and computational domains—and the cultural shifts required to nurture new collaborations.[1]

Scoping Meeting Goals and Creative Process

There were four overarching meeting goals:
- Identify cancer challenge areas that push the limits of current cancer research, computational practices and compel development of innovative computational technologies;

- Build multi-disciplinary engagement, collaboration and community among cancer, data, and computational scientists to create transformative impact;

- Demonstrate approaches to break down silos and work across domains, disciplines and organizations;

- Define types of cultural and paradigm shifts in cancer research that could be possible through the application of advanced computing and HPC.

Meeting participants were asked to engage in a unique creative process to identify the most critical challenges facing cancer research and computational science-based solutions. Participants engaged in lively discussion and debate, building relationships and potential collaborations. In the end, the meeting participants created an aspirational list of cancer challenges framed to simultaneously accelerate predictive oncology and development of advanced computing approaches. Inspired by new insights resulting from the multi-disciplinary interaction during the meeting, the participants expressed their dedication to focus on these challenges moving forward.

Four cancer challenge areas were identified that require shared efforts to advance cancer research while simultaneously driving important computational innovations. These four areas are:

- Generation of synthetic data sets for training, modeling and research
- Hypothesis generation using machine learning (ML)
- Creating digital twin technology
- Development of adaptive treatments

The meeting participants also focused on identifying key areas that require additional efforts to enable the computational and cancer research communities to work together more productively. The meeting participants identified six key barriers together with suggested cultural shifts to minimize or overcome the identified hurdles. These barriers and compensating shifts are follows:

Barrier	Compensating Cultural Shift
➢ PI-centric Science	➢ Multi-disciplinary Team Science
➢ Discipline-focused Communication	➢ Cross-Education, Training and Co-Design of Research Studies
➢ The Modeling Paradigm	➢ Co-Design for Computational Oncology
➢ Undefined Confidence Levels in Predictive Models	➢ Uncertainty Quantification (UQ) and Model Validation
➢ Differing Spatial-temporal Perspectives	➢ Integrated System of Systems
➢ Data Access, Sharing and Security	➢ Establishing FAIR Data Principles

Lastly, the meeting participants identified avenues to sustain the energy and interest cultivated in the scoping meeting itself while expanding the community. This includes a series of virtual engagements called 'Microlabs' that provide opportunities for further discussion, development and sharing of the meeting outputs.

This report presents the goals, process and outcomes of the first ECICC meeting. It serves both to capture key insights and provide a reference for the cancer and computing research communities about the tremendous potential for cross-discipline collaboration. Finally, the report conveys the participants' excitement and engagement and sets the stage for expanded or new collaborations across the cancer and computational research communities.

ECICC Scoping Meeting Overview

The ECICC Scoping Meeting was a hands-on, interactive team effort with the goal of identifying challenges of sufficient magnitude to require team-science solutions at the intersection of cancer and computational domains and growing this emerging research community. The meeting included three panel discussions and two keynote presentations that highlighted the opportunities for both cancer and computational scientists working together on predictive oncology challenges. Panel discussions focused on (1) accomplishments and lessons learned from the JDACS4C collaboration; (2) challenges faced by cancer research for both implementing advanced computing solutions (e.g. data sharing/aggregation/quantity; model validation and interpretability) and opportunities for multiscale and multi-omic machine learning (ML)-modeling approaches; and (3) overviews of computational and other unique research capabilities (e.g. advanced light sources) at eight of the DOE's National Laboratories.

The meeting began with an introductory presentation by Emily Greenspan, Ph.D., NCI, and Carolyn Lauzon, Ph.D., DOE Office of Science,[6] who gave agency perspectives on why now is the right time to expand and broaden the NCI-DOE partnership and develop new engagements and collaborations to advance the greater goal of a National Learning Healthcare System for Cancer.[7] They presented a vision based on predictive oncology that identifies the cancer challenges that will compel innovation in computing and the computing innovations that will drive new knowledge and innovation in cancer research.

Peter Nugent, Ph.D., Lawrence Berkeley National Laboratory, provided an aspirational keynote from the high-energy physics and astrophysics perspective[8]. Using his supernova research as an example, Dr. Nugent highlighted how silos between theorists and experimentalists can be bridged. He also explained how computational models have shaped the design of experimental programs.

In a second keynote, Warren Kibbe, Ph.D., Duke University, discussed ways to maximize the value of oncology data now that data generation is no longer a bottleneck in most cancer research.[9] He spoke about recent key changes in oncology, including an emerging systems view of biology and the importance of analyzing both healthy and disease states. Dr. Kibbe also highlighted the opportunity to understand cancer patient trajectories by using advanced computing to enable the shift from observation to prediction and to better link outcomes to care.

During the meeting, participants were asked, individually and in small groups, to identify and discuss cancer challenge areas that push the limits of current cancer research, computational practices and compel development of innovative computational technologies. Cancer researchers were encouraged to think aspirationally about what they would do if they had the right computational tools and resources. Computational scientists gained insight to problems and challenges facing cancer researchers and clinicians and were asked to imagine computational tools that could address these problems, and how solutions might in turn advance computational science.

Leveraging the more than 200 individual cancer challenge ideas identified through scoping exercises (*see Appendix A*), meeting participants identified nine (9) overarching computational oncology challenges. [10] Breakout groups were created for each of the nine challenge areas

Participants in each group wrote draft summaries and presented the summary write-ups to all meeting participants. The nine cancer challenge areas were:

1. ML for hypothesis generation
2. Studying the mechanisms of cancer across scales
3. Collaboration between communities
4. Adaptive drug and immunotherapy treatment
5. Defining optimal treatments
6. Why cancer kills
7. Bridging spatial-temporal scales
8. Simulating care pathways
9. Synthetic data

With overlap present across these nine original ideas, four broad cancer challenge areas were identified that integrate the nine original ideas and provide the organizational framework for this report and the follow-on activities. [11] Each of these four areas requires multi-disciplinary approaches and:

- Includes predictive oncology challenges and opportunities for substantial HPC and AI technology implementation with the potential to advance data and computational science, as well as cancer research;

- Was proposed by a group(s) comprised of computational scientists and cancer researchers, with notable excitement, debate and varying perspectives on feasibility;

- Incorporates elements identified at the meeting as addressing barriers to collaboration.

The four broad cancer challenges ultimately provide a progression from the critical function of data collection and analysis, to hypothesis generation and finally to cancer technology ideas intended to ultimately support predictive oncology goals for improved cancer patient treatment. They are:

- **Generation of synthetic data sets for training, modeling and research.** *Cancer researchers need large-scale, validated and statistically realistic synthetic data sets (for clinical patient data, other PII data, areas of sparse data, etc.) with which to develop solutions, evaluate models and test hypotheses.*

- **Hypothesis generation using machine learning (ML).** *ML has the potential to efficiently guide hypothesis generation and experimental design in cancer research and enable the analysis of existing large, complex data sets to guide, among other things, clinical trials and decision making.*

- **Creating digital twin technology.** *A patient's digital or virtual twin would be a holistic, in-silico model of preventive, diagnostic and care trajectories encompassing space and time that would ultimately be used in clinical settings to inform treatment decisions.*

- **Development of adaptive treatments.** *Adaptive treatments are imagined as nanoscale devices or biologically-based agents, enabling precision treatments that adapt to the changing nature of a tumor over time.*

Cultural Barriers and Organizational Shifts

In addition to the cancer challenge areas, participants also identified key cultural barriers and ongoing organizational shifts needed to fully implement advanced computing in cancer research, including:

- Expanding PI-centric cancer research to include *multidisciplinary team science;*

- Broadening discipline-specific expertise to incorporate *cross-education, training and co-design of research studies;*

- Transforming undefined confidence levels in predictive models to *quantified uncertainty and model validation;* and

- Eliminating barriers to access cancer data and encouraging sharing and security to *establish cancer data based on Findable, Accessible, Interoperable, Reusable (FAIR) principles.*[12]

###

NOTE: The Full Scoping Meeting Report is available at:
https://ncihub.org/groups/cicc/File:/uploads/ECICC_Scoping_Meeting_Report.pdf

Challenge Area 3: Creating Digital Twin Technology

Today, cancer care teams cannot offer patients a personalized view of their health trajectories, particularly when faced with various treatment options. Meeting participants envisioned a future in which a patient's digital twin (aka, avatar or virtual patient) could be used as a holistic *in-silico* model in cancer wet lab research, clinical trials and in clinical settings to guide more effective and personalized treatment choices. Digital twins would incorporate models of relevant biological processes, as well as disparate kinds of data unique to a patient. The technology would not just be used to stratify patients, but to predict the dynamics of their disease trajectories. This would expand precision medicine to *predictive* medicine.

Creating digital twin technology would be a grand challenge in HPC and oncology. It involves bridging spatiotemporal scales as never before—from the molecular, cellular, and tissue levels to the individual, population, and environmental levels. At each scale, agents interact with each other, and it will be necessary to identify the multitude of variables – many not currently captured systematically – that allow scales to be bridged and connected.

Many participants were enthusiastic about the potential of this systems-based approach and agreed that the digital twin is the ultimate multi-scale model. They also agreed that creating digital twin technology could only be accomplished through a dynamic, large-scale, multidisciplinary collaboration. As such, it is a major opportunity for HPC and oncology co-design efforts.

Relevance

The creation of digital twins could completely alter basic, translational and clinical cancer research, treatment, and population health by providing an advanced, *in-silico* modeling environment across the oncology spectrum. Researchers and clinicians need to understand the inter-relationships of spatiotemporal scales, in both healthy and disease states, to predict the impact of molecularly targeted treatment for the individual, and how the individual's environment, behavior, etc. impacts molecular, cellular, and overall physical level response. The digital twin would provide researchers with a computational tool to formulate predictions based on hypotheses and approximations that would improve over time with recourse to finer-scale calculations and observations.

A digital twin would enable iterative and ensemble "what-if" evaluations of proposed interventions. This would allow physicians to not just better select the most effective treatment, but help patients weigh their treatment choices against their personal priorities and constraints. The digital twin population could identify high-risk populations and allow policymakers to evaluate different screening practices and guidelines. The digital twin capability has the potential to significantly impact policy and population health. In a clinical setting, a digital twin would also be a powerful tool for patient-physician communication to facilitate informed patient choice and shared decision making.

This cancer challenge builds on existing, but uncoordinated, efforts to create the first proto-digital twins. These pioneering models are far from being whole-patient representations. For example, German researchers are using a very rudimentary virtual model to select the best treatments for melanoma patients.[19] There are also extensive examples of advanced computing models of individual cells and organs.[20]

Meeting participants noted that now is the time to harness new and emerging HPC resources to combine existing specific computational oncology models, such as those for tumor growth and vascularization, into a holistic, multi-scale model that can even produce population-level models. Notably, creating this integration with uncertainty-aware calculations requires massively parallel ensembles of simulations and analysis tasks that utilize emerging HPC systems. Collaborative efforts across disciplines are underway, and there is a need to coordinate efforts to deal with rapidly evolving data streams with various quality and time-scale issues.[21]

Innovation Potential for Cancer Research

Digital twins promise to greatly increase resolution and decrease uncertainty in cancer research. A multi-scale framework will incorporate genomic, molecular, cellular, and population models that are consistent across space and time scales. These models can incorporate social, behavioral and environmental factors such as diet and pollution exposure.

One suggested biological framework for the digital twin-model presented by meeting participants is the Hallmarks of Cancer[22]. These are defined as phenotypic changes at the cellular level that are shared by most, and possibly all, cancer types. However, these hallmarks of cancer are mostly studied in isolation and have proven to be of limited predictive utility at clinical scales. A digital twin program to computationally integrate these disparate hallmarks of cancer into one coherent model could be a major step toward understanding, predicting, and reducing cancer lethality.

At point of care, digital twins could provide personalized evidence to guide treatment decisions. Patients would be able to see their virtual twin across multiple treatment scenarios, providing personalized information of their cancer progression, treatment related side-effects, and quality of life. The use of a population of digital twins, combined with leadership-class computing power, could augment the gold standard randomized clinical trial and enable rapid virtual clinical trials.

These might be able to quickly and efficiently identify potential treatment failures and opportunities. The time, resources and cost of conducting current clinical trials make this a compelling alternative, including potentially saving billions of dollars in the development of new drugs.

The step-wise development of digital twins would broadly drive innovation in both HPC architectures and advanced computing. The complexity of multi-scale, high-resolution, predictive models is anticipated to be more difficult than pure-physics models, thus pushing the state-of-the- art in HPC predictive science. For example, while most physics-based models involve proximate interactions over short time scales (femtoseconds to seconds), a digital twin would involve modeling across the time frames of molecular interaction to multiple years in a patient's life. Similarly, digital twin models must include cancer's ability to metastasize, thus, to act at a distance. Population models for virtual prevention trials would take place over decades, over the entire space of the U.S.

The biological models will need to be validated and doing so will push the art of verifying models in complex systems. Similarly, a digital twin model would push the frontier of uncertainty quantification and error estimation that reflect both computational and oncological sources of error.