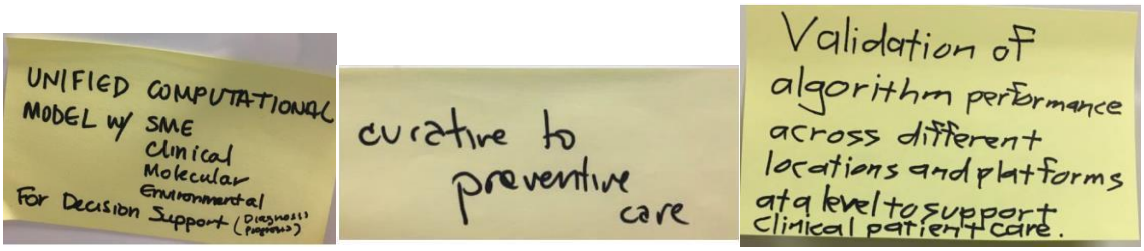
Envisioning Computational Innovations for Cancer Challenges (ECICC) Scoping Meeting Report



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

Table of 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.

Introduction

Envisioning Computational Innovations for Cancer Challenges (ECICC) Origins

The *Envisioning Computational Innovations for Cancer Challenges (ECICC) Scoping Meeting* was inspired by several federal goals and initiatives, including the shared goal to create a National Learning Healthcare System for Cancer.[2] This builds on existing efforts, in particular the Cancer Moonshot℠, to develop predictive cancer models and simulations enabled through a growing volume and breadth of cancer data and application of advanced and high-performance computing (HPC), ultimately employed to provide predictive insights for decisions to improve patient care.

In the past decade, there has been a fundamental shift in cancer research and clinical decision-making, moving from qualitative data to quantitative, digital data. For example, between 2014 and 2018, an estimated 2 exabytes of cancer data—from genomics to diagnostic imaging—were generated in the United States.[3] The explosive growth in data generation has led to a new era of data-driven oncological predictive analytics and clinical application. Yet cancer researchers broadly have had limited access, understanding and resources to fully leverage this data. Researchers are hampered by a combination of data access, reproducibility and sharing constraints as well as lack of access to, and expertise in, High-Performance Computing (HPC) and Artificial Intelligence (AI) technologies.

To begin to address these issues, the National Cancer Institute (NCI) and the U.S. Department of Energy (DOE)—which has the world's largest infrastructure and experience in HPC and advanced computing—have partnered on shared aims to accelerate precision oncology research and shape the future for emerging exascale computing. The *Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)* program, established in June 2016[4], encompasses three co-designed pilot projects that frame forward-looking approaches for integrating and analyzing large, heterogeneous data collections with advanced computational modeling and simulation. The overarching goal of JDACS4C is to collaboratively develop, demonstrate, and disseminate advanced computational capabilities to seek answers to driving scientific questions that increase our understanding in three specific areas, or levels, of cancer research:

1. **Pilot One (cellular-level)** joins deep learning with novel integration and combinations of data to develop computational predictive models for screening tumor drug response to identify promising new cancer treatments.
2. **Pilot Two (molecular-level)** combines experimental data, simulation and AI to provide new insights to understand and explore the biology of heretofore undruggable targets, such as the RAS protein. RAS-related cancers are based on a family of known genes that undergo mutation to oncogenes.
3. **Pilot Three (population-level)** uses AI and clinical information at unprecedented scales to enable real-time precision cancer surveillance and gain critical insights on the drivers of population cancer outcomes that will transform cancer care.

Using co-design principles, each of the pilots in the JDACS4C collaboration is based on, and driven by, team science, which is the hallmark of the collaboration's success. The partnership is also developing new cross-cutting technologies including uncertainty quantification (UQ) methods

to evaluate the level of confidence or certainty in AI model predictions and a scalable, open source deep learning environment (CANDLE).

The three JDACS4C pilot projects have demonstrated there are compelling opportunities to apply HPC and advanced computing to develop machine-learning (ML)-based predictive models and simulations in cancer research across molecular, cellular and population scales. These opportunities can successfully inform hypothesis generation and experimental design as well as lead to new biological insights.

Predictive Oncology Community Building

The success of the JDACS4C collaboration validates that a team science driven, multi-disciplinary approach is critical to moving from descriptive analytics—which is based on analyzing previously collected/observed data—to data-driven predictive analytics—which provides the opportunity to predict future trends and determine the importance of missing data elements. There is an exciting opportunity to support the establishment and growth of an emerging computational oncology community to come together, expand, and identify those research areas that, when focused on by large, motivated, collaborative teams, will move this developing field forward. The Envisioning Computational Innovations for Cancer Challenges (ECICC) community includes scoping meeting participants and others who have participated in follow-on interactive activities. The ECICC community builds on the nascent predictive oncology community cultivated by the Frontiers of Predictive Oncology meetings.[5] The ECICC community is driven by the JDACS4C collaboration and brings together a broader group of cancer, biomedical, bioinformatics, engineering, data and computational scientists at various career stages from government, academia, and industry.

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]

Cancer Challenge Areas

Challenge Area 1: Generating Synthetic Data Sets

Every year in the United States, petabytes of clinical oncology data—from PET scans to biomarkers—are collected from tens of thousands of patients in individual hospitals and research centers.[13] Due to patient privacy and other data sharing issues, however, most of this data is not publicly available or shared and cannot be leveraged by the broader research community. Meeting participants identified this "data jail" dynamic as a "foundational" hurdle to advancing computational oncology.

Computer-generated synthetic data sets are statistically identical to real clinical data sets but are anonymized and are thus not considered protected data. The vision at the heart of this challenge area is an ecosystem in which original "gold standard" data sets remain under the stewardship of the entities that create them, such as cancer registries. The synthetic data set versions created from these gold standard data sets, together with their specific generation rules and metadata, would be made broadly available. These distributable data sets would allow researchers to apply new models of analysis to these synthetic sets, and subsequently offer the products of these analyses, such as new machine learning (ML)-algorithms, back to the original owners of the clinical data sets and the broader research community.

Relevance

Current clinical data sets that are shared are typically of small scale, have significant data attestation rules and requirements, and reflect local inconsistencies in data from source systems.

Access to and sharing of sufficiently large and high-quality comprehensive clinical data sets is inefficient and difficult to incentivize. Gaining access to and managing clinically derived data is time-consuming, expensive, and de-identification is subject to multiple intellectual property (IP) and Institutional Review Board (IRB) restrictions. In addition, de-identification of these data sets varies among institutions. Even when access is achieved, data quality of extracted clinical data sets often suffers due to "over sanitization" in the de-identification process, and inconsistent labeling. This makes it challenging to generalize results to further modeling, training or evaluation needs.

Synthetic data methods and resulting data sets promise to protect patient confidentiality by completely delinking identity. Other benefits include advancing the expectation of transparent and reproducible analysis methods and supporting transferrable methods that can be applied to original data for clinical validity. An ecosystem that supports access to synthetic data sets could both expand the scale of opportunity for education and training on complex clinical data and deepen our understanding of the complexity in healthcare data access and management.

Why Now?

Meeting participants noted that computational technologies such as ML and changing community standards (open-access and data sharing policies) combined with pioneering efforts in clinical synthetic data make this an ideal time for a larger oncology-based program.

For example, MIMIC-III is a large, publicly-available database comprising de-identified health-related data associated with approximately 60,000 admissions of patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012.[14] Similarly, a precedent has been set with the sharing of large-scale cancer genomics data via the NCI Genomic Data Commons.[15] These two large-scale data sharing projects are driving expectations for sharing clinical oncology data more broadly.

Innovation Potential for Cancer Research

The ability to produce anonymized or de-linked synthetic data could drive the capture of a wider range of clinical data not currently recorded. Sharing and generating synthetic data and metadata could be critical to improving the reproducibility of clinical data. Different synthetic data sets could be tested against one another using methods developed in other fields to improve clinical data quality in a learning healthcare system.

This ecosystem would also advance the ability to develop and share complex computable phenotypes of patients across medical environments. As well, the envisioned ecosystem would establish new ways to determine missing data or noise though peer reproducibility and validation. In addition, it would help improve and standardize quality measures for source data systems and clinical workflows with provenance provided back to gold-standard data generators.

The access and validation of new or different ML or other analytic models on accessible clinical datasets of any size would advance democratization of data and model transfer. It would also inform how clinically-relevant gold standards of data sets support higher standards of data reproducibility.

Innovation Potential for High-Performance Computing Capabilities

Providing an ecosystem of synthetic data sets based on real, clinically associated data will require new, innovative collaborations in ML, clinical data generators, data managers, data architecture and software engineering.

Anonymizing large data sets (100,000s of patients) and ensuring these cannot be used for re-identification of individual patients will require new advanced algorithms and an expanded expectation for stewardship of the full workflow process involved in large-scale clinical data-set generation.

Challenge Area 2: Machine Learning for Hypothesis Generation and Clinical Decision Support

Meeting participants identified machine learning (ML) as a critical tool for experimental hypothesis generation and guiding experimental design. ML, a type of AI, allows computer systems to automatically learn and improve from experience without being explicitly programmed, and ML algorithms build a mathematical model of sample data to make predictions. Thus, ML is a cornerstone computational tool for the movement from descriptive to predictive approaches and eventually to realize the potential of precision oncology—the right treatment, for the right patient, at the right time.

Relevance

Meeting participants noted that at present there is a cancer data conundrum: researchers and clinicians are inundated with more information than they can handle, while, at the same time, there are sizeable gaps in the biological systems information that is required for research and clinical advance. ML is key to addressing both hurdles. First, ML can provide guidance to experimentalists on what to study and the sequence in which to attack questions. Thus, ML models have the potential to guide experiments and fill out data gaps as efficiently as possible. Second, ML is a critical bridge between large, complex data sets and mining actionable meaning from the data. ML has the unique ability to discover and use algorithms to cluster observations (data), and to do so iteratively with experimentation, in an active-learning process. As such, ML has high potential to help develop tools to support real-time clinical decision making.

Why Now?

In the past decade there has been a fundamental shift from qualitative to quantitative data in cancer research and clinical settings. This ongoing analog to digital transition is producing an extensive pipeline of previously unavailable digital data related to cancer. For example, scanning electron micrograph (SEM) technology has transitioned from photographic data ("*blobology*") to fully digital, near atomic-level resolution scans.[16] Now, however, the volume and types of digital cancer data have outpaced our ability to apply them.

Already, exciting new research has demonstrated that ML and other computational tools can be used to analyze digital histological data—and critically, to do so with greater diagnostic accuracy than clinicians alone.[17] Early results from the JDACS4C cancer research pilots show promise in applying ML for hypothesis generation. Likewise, advances in Deep Learning, a branch of ML based on artificial neural networks, make this an opportune time to apply ML on leadership-class HPC platforms for large-scale analysis of public data sets. DOE's new Summit leadership-class supercomputer has been designed for ML and AI applications and, as of January 2019, is available for cancer research applications.

Innovation Potential for Cancer Research

Meeting participants noted that applying ML would advance cancer research and clinical applications in two key ways. First, it can be used to generate novel hypotheses to determine what additional experiments/data are needed to improve clinical outcomes, understand cancer systems biology or a model itself. Notably, not all ML predictions can be readily tested or validated experimentally. Thus, there will be a collaborative interplay between the need for testable predictions and predictions that will in fact push cancer researchers to explore new experimental terrains.

Second, ML will provide the ability to analyze existing large, complex data sets to move from experimental hypotheses to informing therapy, including identifying druggable targets, dosage strategies and quantifying uncertainty in response to therapy. Notably, due to the enormous size of data involved, the aforementioned cannot be done solely by experiment. One publication recently estimated that for a single patient, this vast "therapy space" includes approximately 10^{40} possibilities. [18] Applying ML on HPC platforms will enable the integration of large sets of molecular and visual data to strengthen research and diagnostic capabilities.

Innovation Potential for High-Performance Computing Capabilities

Meeting participants noted that applying ML to cancer challenges will push HPC and advanced computing in four key ways. First, it will drive advanced computing through the need for new ML models for synthetic, potentially multiscale, data generation (e.g., Generative Adversarial Networks [GAN]). Similarly, massive amounts of extremely heterogeneous data and varying metadata will require advances in modeling and data-integration capabilities. New meta-rules for metadata must also be developed to allow for integration of existing and new technologies into software/API/front-end to enable cancer biologists to easily add, edit and analyze their data.

Second, applying ML to cancer challenges will drive the development of new heterogeneous computational architectures, such as those needed for deep learning AI approaches. Thus, the application of ML to complex cancer data will provide a field for developing and testing advanced architectures on complex biological data.

Third, ML with cancer data will push the field of uncertainty quantification by supporting ML predictions with an associated measure of certainty and reducing input noise of large, heterogeneous real-world data. To be validated, ML outputs will need to be "explainable" to cancer clinicians and researchers. This will require a new level of algorithmic transparency and thus a deeper understanding of ML, especially in applications.

Fourth, this challenge will push the limits of reproducible and comparable data science. As new methods, architectures and datasets become available, the community will need to automatically generate new models and compare outputs to determine the best methods to use and combine. Part of generating new models will be automatic ML (AutoML), such as intelligent/optimized architecture search for both neural networks and ensemble stacking approaches.

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.

Why Now?

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.

Innovation Potential for High-Performance Computing Capabilities

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.

Challenge Area 4: Adaptive Treatments

Meeting participants noted that a key hurdle in the development and implementation of more effective cancer therapies is that cancer comprises many different diseases and is highly heterogeneous within tumors, between tumors and across patients. Tumors are a mix of heterogeneous cell types, can exist with dozens of slightly different genetic variants and can arise through clonal evolution. Cancers are mobile (metastasis), both infiltrating new tissue and triggering distal tissues to recruit cancer cells. Thus, cancer is enormously *adaptive*—hence the need for *adaptive* treatments.

The vision for adaptive treatments involves the development of biological and nano-device-based, personalized drug treatments that adapt to tumors over time. Creating these treatments will require the use of computational oncology. This approach builds on precision medicine, extending it to a new paradigm for cancer treatment. This challenge imagines direct-to-tumor interactive treatments that:

- Adapt to changing tumor characteristics during treatments;
- Target and attack metastasizing cancer cells, augmenting the immune system;
- Deliver novel therapeutics that fabricate molecules *in-vivo* at the tumor site.

Relevance

The systemic adaptation and multi-variant behavior of cancer must be addressed in future-generation therapies. At present, this is sometimes successful—but often only temporarily—through a sequential or concurrent combination of radiation, chemotherapy, surgery, immunotherapy, proton therapy and other treatments. However, for many cancer sub-types and late-stage cancers, these therapies are largely ineffective. Thus, there is the need to imagine, develop, test and implement a concurrently robust response: adaptive therapies.

Why Now?

This cancer challenge area would not have been addressable ten years ago. At present, however, there is a confluence of theoretical, modeling and applied sciences, infrastructure and tools that offer the possibility to imagine and create adaptive treatments. These tools include synthetic biology, systems biology, genetic engineering tools (ex. CRISPR), ML, nanofabrication, nanorobotics, simulation, and clinical applications in bacterial and viral oncology therapies.

Synthetic biology centers on the design, construction, and characterization of improved or novel biological systems using engineering design principles. At present, the United States is the world-leader in synthetic biology and there have been several national level roadmap exercises.[23]

Innovation Potential for Cancer Research

Adaptive Treatments leverages and extends a resurgence of interest in natural bacterial approaches to cancer therapy.[24] Spontaneous tumor regression has been associated with microbial infection for hundreds of years and, a century ago, inspired American physician, William Coley, M.D. (1862–1936), to pioneer the use of live bacteria as a deliberate cancer treatment. In the past ten years, progress has been made with a variety of bacterial organisms, treating a variety of cancers in cell lines, model systems and a few clinical trials. Similarly, viruses (phage) have been used as experimental oncological treatments. The success of these approaches, albeit limited to date, demonstrates potential utility.[25]

The Adaptive Treatments cancer challenge extends these pioneering bacteria-based approaches by applying computationally driven synthetic biology to the engineering of adaptive biologics and nano-devices. The therapies would monitor the tumor as it develops and respond accordingly to destroy cancer cells. This synthetic bacterial treatment would be a combination sensor (diagnosis) and adaptive treatment factory, able to synthesize dozens of cancer-fighting molecules in response to sensor data. Bacteria and envisioned nano-devices are very small compared to tumor cells and could be used to augment the immune system to destroy metastasizing cells.

Innovation Potential for High-Performance Computing Capabilities

Developing adaptive therapies provides a next-generation challenge for HPC and advanced computing. The first key challenge will be to integrate and scale existing approaches in synthetic biology. This would push both HPC architectures and advanced computing. On a socio-cultural and political level, infecting patients with a synthetic pathogen as a means of cancer treatment might well be a controversial issue both at the patient and physician level. Thus, advanced modeling, simulation and ML will play a key role prior to testing treatments *in vivo* and for guiding *in vitro* and *in vivo* research.

Modeling and ML will drive synthetic biology discovery by using deep learning to define the relationship between the tumor properties and the pathogen properties to infect the tumor and the variation within it.

Collaborative Barriers and Compensating Shifts

The meeting included a discussion of the socio-cultural and organizational factors hindering collaboration between cancer biologists and computational scientists. Participants readily identified a range of perceived cultural barriers to collaboration. They also identified the compensatory cultural shifts—some of which are already in progress—needed to foster and accelerate dynamic multi-disciplinary collaborations in the nascent field of computational oncology.

For many participants, the meeting itself was an important first step in making cultural shifts toward transdisciplinary collaboration. With ongoing engagement, these researchers are poised to advance the field of computational oncology and to become leaders in recruiting others to the community.

Key perceived cultural barriers to both collaboration and idea generation were identified by participants, along with the cultural shifts that are driving the growth of computationally-based predictive oncology. These are listed below.

Barrier 1: PI-centric Science

Computational scientists generally have experience with projects that involve large, complex, multidisciplinary teams. Yet because investigator-initiated research is the predominant way cancer research is funded today, it can be a challenge for cancer researchers, especially those in academia, to think in terms of multidisciplinary science when, as a principal investigator (PI), they are focused on tenure review and grant funding.

Compensating Cultural Shift: Multi-Disciplinary Team Science

There is a growing community of computational oncology researchers focused on precision medicine—a predictive, data-driven healthcare model based on each patient's specific genetic, molecular and environmental factors.[26] *As such, there are role model computational scientists and cancer biologists who are already collaborating on a range of ML, AI and data-driven projects, including models to predict cancer susceptibility, prognosis, and survival. Participants with little to no experience with big team science projects heard first-hand from JDASC4C collaborators about how to shape their thinking and behavior to this approach.*

Barrier 2: Discipline-Focused Communication

Meeting participants stated that one of the fundamental challenges in transdisciplinary collaboration is the inherent barrier concerning scientific language. Cross-disciplinary communication can be hindered by everything from unfamiliar acronyms to the use of a word that means different things in different disciplines. For example, JDACS4C researchers noted that in biology, the term "model" generally refers to an *in vitro or vivo* (or wet lab) replication of a living system, such as 3D cultures of cancer cells. In computation, however, a "model" generally refers to mathematical representations of phenomena that are then simulated on a computational system, for example the supernova simulations presented to meeting participants by Dr. Peter Nugent. Participants

experienced this terminology barrier first-hand when meeting discussions were interrupted for clarification of terms.

Compensating Cultural Shift: Cross-Education, Training and Co-Design of Research Studies

Team science requires that all researchers become educated enough to communicate and work effectively across fields, understand limitations and create cultural linguistic bridges to constructively work together and imagine possibilities. This means learning each other's professional language and ways of approaching problems. This cross-education informs and underpins the co-design of studies that use computational approaches to address cancer challenges and in-turn, how biology might influence advanced computational algorithms, technologies and HPC architectures.

Barrier 3: The Modeling Paradigm

Participants noted that the well-established computational-modeling paradigm found in physics is currently far less established in biology and this difference can be a barrier to collaboration. Computationally experienced scientists are generally comfortable operating from theory and initial observations to create first-principle-based (*ab initio*) computer models and use these to guide future research. In contrast, traditional biological science approaches emphasize empirically-based observations and work with complex living systems in which there are large, fundamental knowledge gaps in how the system operates. These differing approaches further contrast with physics employing mathematical equations leading to computational approaches, while biology, without comparable fundamental laws as are found in physics, employs the modeling paradigm far less so.

Compensating Cultural Shift: Co-Design for Computational Oncology

There is a growing awareness that groups and organizations of both cancer and computational scientists face a common challenge in computational oncology, and that overcoming it could accelerate progress in cancer research. Analogous to the challenges faced in Stockpile Stewardship modeling, cancer is a difficult, multiscale, multi-physics problem that scales from atoms to cells to tumors and ultimately to the entire body.

Many participants also made the case that computational modeling is well-suited for cancer research given the sometimes-prohibitive cost, difficulty in obtaining sufficient sample sizes, and technical and ethical limitations of some wet-lab and human experiments.

Barrier 4: Undefined Confidence Levels in Predictive Models

Especially in a clinical context, oncologists and researchers may be unfamiliar with, mistrust and often face institutional and policy hurdles in the explicit application of computational tools and predictive models for clinical decision support and research. While many instruments in common use employ computational approaches internally for the acquisition and generation of data, the subsequent use of the resulting data for decision support presents a challenge for adoption.

Compensating Cultural Shift: Uncertainty Quantification (UQ) & Model Validation

Uncertainty quantification (UQ) and statistical model validation were discussed as methods computational scientists can use to build consensus among clinicians and researchers in computational oncology- based, clinical decision-making. Overcoming this barrier would involve educating cancer researchers on the basics of the methods commonly used in computational science and ML and adopting them in research and clinical care. Meeting participants also noted that there is a need to make computational models and algorithms interpretable and explainable.

Barrier 5: Varied Spatial-temporal Perspectives

Meeting participants noted that many computational scientists are familiar with modeling systems in physics and chemistry in which the interactions are largely proximate and often occur on time scales of milliseconds or faster. In cancer modeling, interactions must often be considered over longer time periods (days to years) and must also consider non-linear, distant interactions throughout the entire body, notably in tumor metastasis.

Compensating Cultural Shift: Integrated System-of-Systems

Participants noted that multi-scale integration is a central challenge in cancer modeling. There was widespread enthusiasm among cancer researchers for using HPC and advanced computing resources to achieve this system-of-systems understanding. Similarly, computational scientists noted that cancer modeling represents a frontier of multi-scale complexity that will drive advanced computing and novel HPC architectures.

Barrier 6: Data Access, Sharing and Security

Computational scientists and cancer researchers face key differences in data access and sharing regulations. Computational scientists operate in an environment in which data is often rapidly and broadly shared with the entire community, likely stemming from its non-human origins. However, cancer researchers often have difficulty accessing the data needed and must gain approval from Institutional Review Boards (IRBs) to use and access data. The perception is that many cancer researchers, particularly those in academia, are unaccustomed to sharing data that would be required by the kind of big-data projects participants hope to pursue. Moreover, as discussed previously, current funding mechanisms reward individual PIs, or a small number of co-PIs, rather than large groups, which often results in data silos.

Compensating Cultural Shift: Establishing Findable, Accessible, Interoperable and Reusable (FAIR) Data Principles

Meeting participants noted that advancing computational oncology requires a shift toward sharing data in a safe and timely manner. Future discussions will be needed to address ethical and security issues. For example, technologies such as blockchain are being explored to maintain data integrity. It was also suggested that the volume of data already being collected on cancer has yet to be fully utilized. Participants discussed extending the circle of trust being built among cancer and computational scientists to those who collect cancer data on the frontlines: cancer registrars. Collaboration with registrars could potentially allow this real-world data to be included in future projects.

Post Meeting Community-Building Activities

The vision for the ECICC community includes expanded connections among researchers, multidisciplinary collaborations, and increased knowledge and access to supercomputing and HPC capabilities and resources at DOE labs that will benefit the cancer research community. Due to the great enthusiasm and excitement among the scoping meeting participants and in keeping with the meeting goal to build a community, the organizers recognized the need for continued dialogue and engagement across a broader research community.

After the Scoping Meeting

A follow-on virtual, interactive event referred to as a Microlab was held on June 11, 2019. Over 200 attendees from 50 organizations participated, including several Scoping Meeting participants.

A Microlab is a 60-90 minute, highly interactive virtual event. Unlike webinars which are focused on disseminating information, the MicroLabs facilitate stimulating scientific discussions in smaller, more intimate, virtual breakout groups. The Microlab was facilitated by Knowinnovation, which also facilitated the ECICC Scoping Meeting.[27]

The goals of the initial Microlab were to 1) foster dialogue among cancer and computational scientists beyond the attendees at the Scoping Meeting, 2) deepen connections among people with similar research interests, 3) explore potential multi-disciplinary collaborations, and 4) expand the computational oncology community.

The team leads who led development of the challenge areas at the Scoping Meeting presented an overview of their respective cancer challenge area and hosted self-selected, virtual breakout group discussions to further develop ideas. The presenters included the following researchers: *Nick Anderson*, University of California, Davis; *Tina Hernandez-Boussard*, Stanford University; *Jeremy Goecks*, Oregon Health and Science University; *Paul Macklin*, Indiana University; *John McPherson*, University of California, Davis; *William G. Richards*, Brigham and Women's Hospital/Harvard University; *Ilya Shmulevich*, Institute for Systems Biology*; Amber Simpson*, Queens University; *Rick Stevens*, Argonne National Laboratory/University of Chicago; and *Tanveer Syeda-Mahmood*, IBM. Ongoing dialogue among participants is fostered through the community Hub site, which currently has over 156 members and is growing.[28]

After the First Microlab

One week after the microlab, requests for presentations were received from:

- Indiana University's Pervasive Technology Institute for a presentation at the International Supercomputing Conference in Frankfurt, Germany (http://isc-hpc.com); and

- NIH Interagency Modeling & Analysis Group (IMAG), National Institute of Biomedical Imaging & Bioengineering. (https://www.nibib.nih.gov/research-funding/interagency-modeling-and-analysis-group-imag)

Second Microlab

The second microlab was held on September 25, 2019, also facilitated by Knowinnovation (KI). Over 167 people registered from numerous organizations, including people new to the ECICC community.[29]

Building on the breakout discussions from the first microlab, participants worked in small multidisciplinary virtual groups of five-six people to develop use cases based on a persona developed by the team leads of the four cancer challenge areas.[30] (The team leads also provided input on the questions for the use case template). Participants were excited about identifying the critical next steps to help shape more well-defined research project ideas in computational oncology.

Meeting organizers and the presenters are currently assimilating and assessing the data from the use cases developed at the second microlab. When complete, microlab participants (and all members of the ECICC community) will be invited to provide feedback and rank the top challenges, barriers, and potential actions needed from the broad multidisciplinary cancer research community.

A digest of all input, recommendations, resources, and action ideas from both microlabs are posted on the Hub site.[31] The ideas from the community are too numerous to act on en masse but serve as an important reference for setting future priorities.

Engagement Across NCI/NIH and Other Organizations

Meeting organizers are in discussions about how best to foster and support the growing community to move the ideas from the Scoping Meeting and the Microlabs forward as well as coordinate with related efforts across NCI/NIH and other government and non-government organizations.

For example, the Interagency Modeling and Analysis Group (IMAG) is particularly interested in the ECICC Digital Twin challenge area. Members of the JDACS4C program have joined with other programs at NCI to present a cohesive webinar prior to the annual IMAG Machine Learning and Multiscale Modeling meeting in October 2019. The ECICC planning team is in discussions with IMAG officials about developing a Digital Twin track for the IMAG annual meeting in Spring 2020.

As well, the planning team is working with DOE representatives to create a user-friendly, collection of DOE computational resources and tutorials with guided explanations that can be used by cancer and biomedical researchers and will be available on the website in the coming weeks and months.

Future Goal

Notably, enthusiasm among the original cancer challenge area leads from the scoping meeting continues to grow. Each one of the team leads is a distinguished cancer or computational researcher from a leading university across the United States and Canada. These team leads continue to volunteer their time, expertise, and energy to co-develop strategy and next steps for community building, future events and beyond.

A week-long meeting, known as an Ideaslab is envisioned for Q2 or Q3 of FY20 that would be based on the accumulation of ideas from the community and would include mentors who are senior leaders in cancer research and computational science. The goal of the Ideaslab would be to create projects likely to receive seed funding, and the planning team and mentors would provide guidance to selected projects about obtaining funding. Knowinnovation would be the facilitator. KI has facilitated many successful Ideaslab meetings for other NCI and NIH programs, the FDA, National Science Foundation, NASA, and industry.

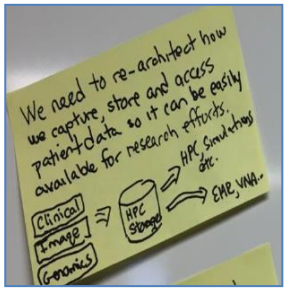
Appendices

Appendix A: Scoping Meeting Process

To drive innovation and community building among cancer, data and computational science researchers, the meeting's organizers hired Knowinnovation (KI), a facilitation firm focused on accelerating scientific innovation, to design and deliver a high-energy, highly interactive and engagement-focused style of event, which they refer to as a scoping meeting.[32] A scoping meeting maximizes diverse expertise to break down a broad topic into actionable challenges and opportunities.

The two-day meeting included content-driven plenary sessions and a variety of small-group sessions to foster cross-disciplinary education, conversation, collaboration and identification of cancer challenges. As well, KI facilitators orchestrated several fast-paced, brief, interdisciplinary interactions designed to push participants to form novel connections. For example, participants wore different colored name-badge lanyards, identifying themselves primarily as either a computational or cancer researcher. This enabled participants to quickly partner with someone from a different discipline.

Central to the scoping process, individuals and groups were asked to write one candidate challenge per sticky note. In a whole-group event, meeting participants grouped 200-plus sticky notes into thematic cancer research areas.



Finally, based on these cancer challenge areas, participants joined in groups to write draft versions of the challenge areas reflected in this report. The challenge areas identified are also now being used to identify topics for follow-up virtual education and collaboration events for this growing community.

Appendix B: Participant Feedback

The 2019 ECCIC Scoping Meeting powerfully engaged an emergent community of oncological, data and computational scientists who conduct cancer research. Feedback from an anonymous online survey, completed by 51 of 74 participants, indicates that the meeting lay the foundation for extensive, novel multi-disciplinary collaborations. Based on an anonymous online survey, 92-percent of respondents reported engaging with one or more potential collaborators at the meeting. Eighty-six percent of respondents said they were inspired to think in new ways about their research. The excitement and enthusiasm reflected in participant feedback from the meeting represents a belief in the promise of building this multidisciplinary, cross-organizational community to advance shared interests and impact cancer research in a significant way.

"If there is the ability to develop synthetic data sets that allows original data sets to remain managed with the security and privacy conditions under which they were created –such as clinical trial or other patient data – but remove those restrictions from synthetic data, then those new derivative data sets can traverse the world along with their specific methods, metadata and labeling and allow a broad range of users and uses."

-- Nick Anderson, Ph.D., UC Davis

"What is needed are computational models that can generate a hypothesis from the existing data that then can feed back into the experimental loop making cancer research and clinical trials (e.g. adaptive clinical trials) more efficient."

-- Jeremy Goecks, Ph.D., Oregon Health & Science University

"That's a really big challenge: to connect many different models with different pieces and get something coherent out when you're done. So, it's a fascinating computer engineering, computer science and software mathematics problem that's going to take state-of-the-art computer facilities to actually build, train and explore these models."

-- Paul Macklin, Ph.D., Indiana University

"There's tremendous potential to drive innovation in cancer research at a variety of scales."

-- Paul Macklin, Ph.D. Indiana University

"Data is pervasive. We can think about things in different ways than we did 10 years ago."

-- Warren Kibbe, Ph.D., Duke University

"Engaging in this JDACS4C pilot has been a paradigm shift."

-- Fred Streitz, Ph.D.
Lawrence Livermore National Laboratory

"Very different! I may never think the same again. I LOVED suspending disbelief and ignoring hurdles, if even for just 2 days. We do not do enough of that."

-- Anonymous Feedback

"It was great to see the meshing and discomfort as the different disciplines came together."

-- Anonymous Feedback

"*We had to learn each other's vocabulary*."

-- Yvonne A. Evrard, Ph.D., Frederick National Laboratory for Cancer Research

"Engineering this (creation of adaptive treatments) requires the ability to first imagine how it might work, get it working in simulation while we try to get it working in an actual system and do essentially exhaustive virtual clinical trials."

-- Rick Stevens, Argonne National Laboratory

"What we're talking about doing is programming a bacterium or programmable nano-device to change its behavior in a tumor micro-environment in response to the specifics of a patient. Think about a bacterium or nano-device as a mini, portable chemistry factory. Plus, the fact that it's going to synthesize the drug at the site, using materials that it has in the environment."

-- Rick Stevens, Argonne National Laboratory

"How can we simulate a cancer patient's care trajectory from pre-diagnosis to survivorship?"

-- Tina Hernandez-Boussard, Ph.D., Stanford University

"It was great to meet folks from other disciplines. It was a wonderful research community building opportunity."

-- Anonymous Feedback

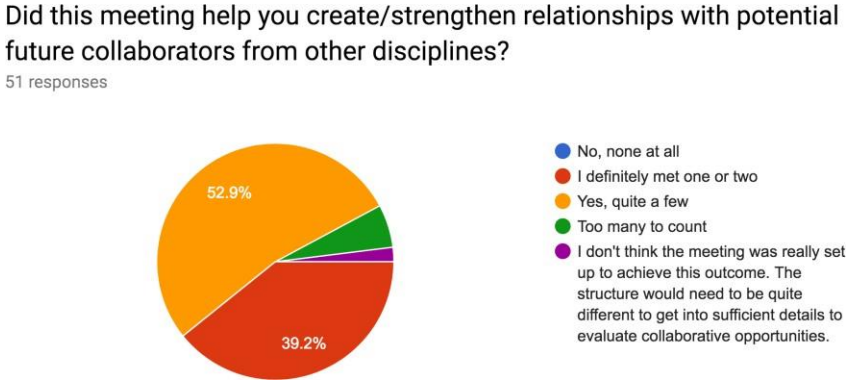
"This was much more interactive and collaborative (compared to other meetings)."

-- Anonymous Feedback

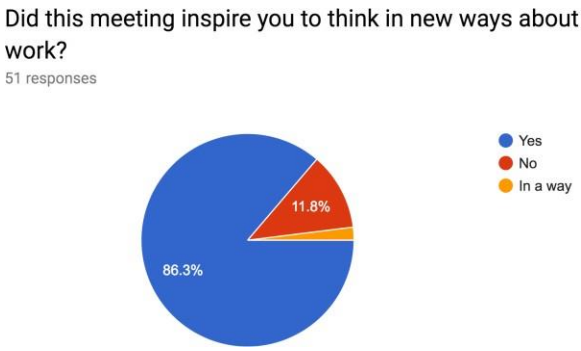
"*The cultural differences we identified helped me think about ways I can help our community focus on the self-imposed barriers to team science we face*."

-- Anonymous Feedback

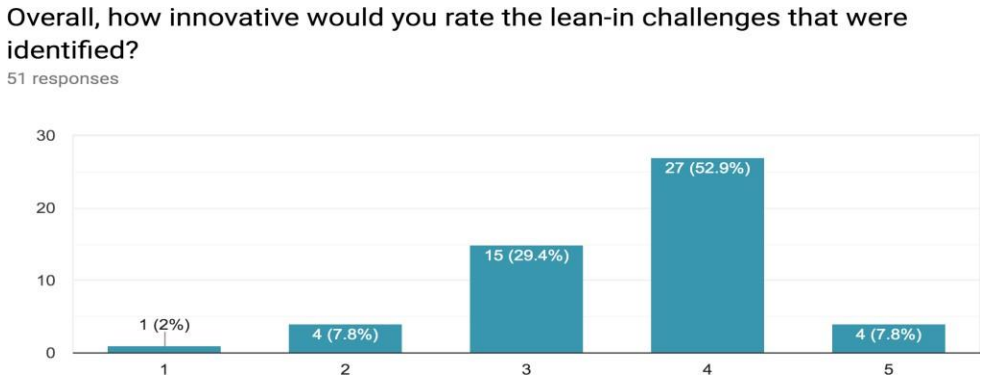
Meeting Potential Collaborators. Overall, 92.1 percent of respondents felt they had met one or more potential collaborators at the meeting (*see below*). On a scale of 1-to-5, respondents found the meeting appropriately multi-disciplinary (84.4 percent answered 4-or-5).



Thinking Differently. Shifting the culture of cancer research begins with changing the way researchers think. In that regard, 86.3 percent of respondents said they were inspired to think in new ways about their research (*see below).*



Innovation. More than half of the respondents (60.7 percent) rated the challenge areas either a 4 or a 5 on a five-point scale for innovative ideas. As one presenter pointed out, people fall in love with ideas. It was important, therefore, to have participants *experience* the excitement and the possibility that comes from the generation and sharing of ideas through multi-disciplinary collaboration.



Engagement. Respondents found the meeting fun (84.4 percent answered either 4 or 5 on a scale of 1 to 5) and a good use of their time (92.2 percent).

Appendix C: Bibliography

Generating Synthetic Data

Dahmen J, & Cook D. (2019). SynSys: A Synthetic Data Generation System for Healthcare Applications. *Sensors*. *19*, 5,1181.

Chen J, Chun D, Patel M, Chiang E, James J. (2019). The validity of synthetic clinical data: a validation study of a leading synthetic data generator (Synthea) using clinical quality measures. – *Med. Inform. Decis. Mak*. *19*,1, 230.

Hypothesis Generation Using Machine Learning

Wozniak, J. M. *et al*. (2018) CANDLE/Supervisor: A workflow framework for machine learning applied to cancer research. *BMC Bioinformatics*. 19,

Azuaje, F. (2019). Artificial intelligence for precision oncology: beyond patient stratification. *Precis. Oncol*. *3*, 1–5.

Eric J. Topol. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7

Creating Digital Twin Technology

Eunjung Kim, Vito W. Rebecca, Keiran S.M. Smalley, Alexander R.A. Anderson. (2016). Phase *i* trials in melanoma: A framework to translate preclinical findings to the clinic. *Eur. J. Cancer*. *67*:213–222. http://dx.doi.org/10.1016/j.ejca.2016.07.024.

J. Ozik, N. Collier, R. Heiland, G. An, and P. Macklin. (2019) Learning-accelerated Discovery of Immune-Tumor Interactions. *Molec. Sys. Design Eng.* (*in review*).
Preprint: https://dx.doi.org/10.1101/573972

Development of Adaptive Treatments

Kramer MG, Masner M, Ferreira FA and Hoffman RM (2018) Bacterial
Therapy of Cancer: Promises, Limitations, and Insights for Future Directions. *Front. Microbiol*. *9*:16. doi: 10.3389/fmicb.2018.00016

Nikalje AP. (2015). Nanotechnology and its Applications in Medicine. *Med. Chem. 5*, 081-089. doi:10.4172/2161-0444.1000247

Appendix D: Organizing Committee

Lynn Borkon	Frederick National Laboratory for Cancer Research
Michael Cooke, PhD	Department of Energy, Office of Science
Janelle Cortner, PhD	National Cancer Institute
Tony Dickherber, PhD	National Cancer Institute
Emily Greenspan, PhD	National Cancer Institute
Amy Gryshuk, PhD	Lawrence Livermore National Laboratory
Sean Hanlon, PhD	National Cancer Institute
Shannon Hughes, PhD	National Cancer Institute
Roxanne Jensen, PhD	National Cancer Institute
Miles Kimbrough	Frederick National Laboratory for Cancer Research
Carolyn Lauzon, PhD	Department of Energy, Office of Science
Christopher Miller, MD	Department of Energy, Office of Science (Fellow)
Stefanie Nelson, PhD	National Cancer Institute
Eric Stahlberg, PhD	Frederick National Laboratory for Cancer Research

Appendix E: Agenda

Envisioning Computational Innovations for Cancer Challenges: Scoping Meeting

AGENDA

March 6-7, 2019

Livermore Valley Open Campus

High Performance Computing Innovation Center

Building 6475 - 7000 East Ave - Livermore, CA 94550

March 6

8:00 am – 8:45 am **ARRIVAL AND CHECK-IN**

8:45 am – 9:45 am **WELCOME AND ORIENTATION**

Why Now?
Emily Greenspan, *National Cancer Institute (NCI)*
Carolyn Lauzon, *DOE Office of Science*

HPC in Cancer Research Brief Overview
Eric Stahlberg, *Frederick National Lab (FNLCR)*
Amy Gryshuk, *Lawrence Livermore National Lab*

Knowinnovation Introduction
Stavros Michailidis, *Knowinnovation (KI)*
Andy Burnett, *Knowinnovation (KI)*

9:45 am – 10:45 am **SESSION 1: WHAT'S POSSIBLE COLLABORATING ACROSS DISCIPLINES AND ORGANIZATIONS**

JDACS4C Panel:
Jessica Boten, *NCI*
Yvonne Evrard, *FNLCR*
Dwight Nissley, *FNLCR*
Rick Stevens, *Argonne National Lab*
Fred Streitz, *Lawrence Livermore National Lab*
Gina Tourassi, *Oak Ridge National Lab*

Moderators:
Michael Cooke, DOE Office of Science
Betsy Hsu, NCI

10:45 am – 11:30 am	**A Coffee Break with a Purpose** - Generating ideas for cancer challenge areas (part 1)
11:30 am – 12:30 pm	**KEYNOTE PRESENTATION** **DOE Success Story: Leading with Science in a Computational Context** Peter Nugent, *Lawrence Berkeley National Laboratory*
12:30 pm – 2:00 pm	**Working Lunch** - Generating ideas for cancer challenge areas (part 2)
2:00 pm – 3:00 pm	**SESSION 2: CHALLENGE AREAS IN CANCER RESEARCH** **Panel:** Gregory Cooper, *University of Pittsburgh* Tina Hernandez-Boussard, *Stanford University* Paul Macklin, *Indiana University* John Quackenbush, *Harvard University* Amanda Randles, *Duke University* William Richards, *Brigham and Women's Hospital* Ilya Shmulevich, *Institute for Systems Biology* Amber Simpson, *Queens University* **Moderators:** Roxanne Jensen, *National Cancer Institute* Amy Gryshuk, *Lawrence Livermore National Laboratory*
3:00 pm – 3:45 pm	**SPEED NETWORKING** - Generating ideas for cancer challenge areas (part 3)
3:45 pm – 4:00 pm	**BREAK**
4:00 pm – 5:00 pm	**SESSION 3: DOE CAPABILITIES AND RESEARCH** **DOE National Laboratories Panel:** Frank Alexander, *Brookhaven* Silvia Crivelli, *Lawrence Berkeley* John Feddema, *Sandia* Sarah Michalak, *Los Alamos* Ana Paula de Oliveira Sales, *Lawrence Livermore* Robert Rallo, *Pacific Northwest* Rick Stevens, *Argonne* Gina Tourassi, *Oak Ridge* **Moderators:** Carolyn Lauzon, *DOE Office of Science* Eric Stahlberg, *Frederick National Lab*
5:00 pm – 5:30 pm	**Closing plenary | ADJOURN DAY**
6:30 pm – 8:00 pm	**Dinner in self organized groups**

March 7

8:00 am – 8:45 am	**ARRIVAL AND CHECK-IN**
8:45 am – 9:00 am	**RECAP DAY 1 | OVERNIGHT IDEAS**
9:00 am – 10:00 am	**KEYNOTE PRESENTATION** **Blue Sky Possibilities at the Intersection of Oncology and Computing** Warren Kibbe, *Duke University*
10:00 am – 11:00 am	**BREAK OUT GROUPS** - Generating ideas for challenge areas (part 4)
11:00 am – 12:00 pm	**SYNTHESIS OF CANCER CHALLENGE AREAS**
12:00 pm – 1:00 pm	**LUNCH**
1:00 pm – 1:30 pm	**PRIORITIZATION OF CANCER CHALLENGE AREAS & WRITING GROUP FORMATION**
1:30 pm – 3:30 pm	**WRITING GROUPS** For each challenge area, small writing teams will address: • An introduction to the challenge • Why is this a relevant and important challenge? • Why is now the right time? • How will it drive innovation in cancer research? What isthe impact for cancer research? • How will it drive innovation in high-performance computing (HPC)? What is the impact for HPC? • What are the key and historical challenges? • What cultural shifts are required?
3:30 pm – 5:30 pm	**WRITING GROUP PRESENTATIONS WITH FEEDBACK**
5:30 pm – 5:45 pm	**WRAP-UP**
6:00 pm	**ADJOURN DAY 2**

Appendix F: List of Participants

Name	Affiliation
Alexander, Frank	Brookhaven National Laboratory
Anderson, Nick	University of California, Davis
Basu, Amrita	University of California, San Francisco
Behera, Madhusmita	Winship Cancer Institute, Emory University
Bian, Jiang	University of Florida
Borkon, Lynn	Frederick National Laboratory for Cancer Research
Boten, Jessica	National Cancer Institute
Bouchard, Kris	Lawrence Berkeley National Laboratory
Bremer, Timo	Lawrence Livermore National Laboratory
Buluc, Aydin	Lawrence Berkeley National Laboratory
Chung, Arlene	UNC School of Medicine
Clyde, Austin	Argonne National Laboratory /University of Chicago
Coles, Theresa	Duke University
Cooke, Michael	Department of Energy
Cooper, Gregory	University of Pittsburgh
Cortner, Janelle	National Cancer Institute
Crivelli, Silvia	Lawrence Berkeley National Laboratory
De Oliveira Sales, Ana Paula	Lawrence Livermore National Laboratory
De Val Alda, Natalia	Frederick National Laboratory for Cancer Research
Evrard, Yvonne	Frederick National Laboratory for Cancer Research
Feddema, John T.	Sandia National Laboratories
Goecks, Jeremy	Oregon Health and Science University
Greenspan, Emily	National Cancer Institute
Gryshuk, Amy	Lawrence Livermore National Laboratory
Hernandez-Boussard, Tina	Stanford University
Hinkson, Izumi	Frederick National Laboratory for Cancer Research
Hoang, Thuc	Department of Energy/National Nuclear Security Administration
Hossain, M. Zubaer	University of Delaware
Hsu, Betsy	National Cancer Institute
Jacobs, Sam	Lawrence Livermore National Laboratory
Jankowska, Marta	University of California, San Diego
Jensen, Roxanne	National Cancer Institute
Jha, Shantenu	Brookhaven National Laboratory
Kaplan, Alan	Lawrence Livermore National Laboratory
Kibbe, Warren	Duke University
Kimbrough, Miles	Frederick National Laboratory for Cancer Research
Knutsdottir, Hildur	Johns Hopkins University

Name	Affiliation
Lanman, Nadia Atallah	Purdue University
Lau, Mai Chan	Dana-Farber Cancer Institute
Lauzon, Carolyn	Department of Energy
Li, Jerry	National Cancer Institute
Macklin, Paul	Indiana University
Madduri, Ravi	Argonne National Laboratory
Manion, Frank	University of Michigan Cancer Center
Matasci, Naim	University of Southern California
McManus, Michael	Intel
McPherson, John	University of California, Davis
Michalak, Sarah	Los Alamos National Laboratory
Miller, Christopher	Department of Energy
Narayan, Kedar	Frederick National Laboratory for Cancer Research
Nissley, Dwight	Frederick National Laboratory for Cancer Research
Nowak, Jonathan	Brigham and Women's Hospital / Harvard Medical
Nugent, Peter	Lawrence Berkeley National Laboratory
Paragas, Jason	Lawrence Livermore National Laboratory
Peterson, Andrea	Department of Energy
Quong, Andrew	Frederick National Laboratory for Cancer Research
Rallo, Robert	Pacific Northwest National Laboratory
Randles, Amanda	Duke University
Richards, William	Brigham and Women's Hospital / Harvard Medical
Ropelewski, Alexander	Carnegie Mellon University
Sahinalp, Cenk	Indiana University, Bloomington
Shmulevich, Ilya	Institute for Systems Biology
Simpson, Amber	Queens University
Stahlberg, Eric	Frederick National Laboratory for Cancer Research
Stevens, Rick	Argonne National Laboratory
Streitz, Fred	Lawrence Livermore National Laboratory
Syeda-Mahmood, Tanveer	IBM Almaden Research Center
Templeton, Jeremy	Sandia National Laboratories
Tourassi, Georgia	Oak Ridge National Laboratory
Webb-Robertson, Bobbie-Jo	Pacific Northwest National Laboratory
Wen, NIng	Henry Ford Health System
Whalen, Dawn	Lawrence Livermore National Laboratory
Wozniak, Justin	Argonne National Laboratory
Wu, Shandong	University of Pittsburgh
Zaki, George	Frederick National Laboratory for Cancer Research
Zhang, Min	Purdue University
Zhu, Yitan	Argonne National Laboratory

Appendix G: Citations

[1] Perez, R. K., Kang, R., Chen, R., Castellanos, J. G., Milewski, A. R., & Perez, A. R. (2018). Computational Oncology. *Journal of oncopathology and clinical research*, *2*(1), 3.

[2] Hsu E.R., Klemm J.D., Kerlavage A.R., Kusnezov D. & Kibbe W.A. (2017). Cancer moonshot data and technology team: enabling a national learning healthcare system for cancer to unleash the power of data. *Clin. Pharmacol. Ther*. *101*, 613–615.

[3] "Intel's Cancer Cloud: Complementing NCI Collaboration" (Retrieved: May 20, 2019, from http://www.bio-itworld.com/2016/07/11/intels-cancer-cloud-complementing-nci-collaboration.aspx)

[4] Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) https://datascience.cancer.gov/collaborations/joint-design-advanced-computing

[5] The first *Frontiers of Predictive Oncology and Computing* meeting was held in 2016, with a second and third annual meeting in 2017 and 2018. Participants included nationally recognized thought leaders from government, academia and industry with expertise in cancer, science, computing, and data—and a passion to explore the convergence of predictive oncology and computing. The 2016 and 2017 meeting reports are available on the Intel website. (https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/predictive-oncology-and-computing-2.pdf)

[6] ECICC Scoping Meeting Presentation: "Welcome: Why Now?" Emily Greenspan, National Cancer Institute, and Carolyn Lauzon, Department of Energy (https://ncihub.org/groups/cicc/pastmeetings/ecicc)

[7] Hsu E.R., Klemm J.D., Kerlavage A.R., Kusnezov D. & Kibbe W.A. (2017). Cancer moonshot data and technology team: enabling a national learning healthcare system for cancer to unleash the power of data. *Clin. Pharmacol. Ther*. *101*, 613–615.

[8] ECICC Scoping Meeting Keynote Presentation: "DOE Success Story: Leading with Science in a Computational Context" Peter Nugent, *Lawrence Berkeley National Laboratory* (https://ncihub.org/groups/cicc/pastmeetings/ecicc)

[9] ECICC Scoping Meeting Keynote Presentation: "Blue Sky Possibilities at the Intersection of Oncology and Computing" Warren Kibbe, *Duke University* (https://ncihub.org/groups/cicc/pastmeetings/ecicc)

[10] Full writeups of the nine cancer challenge areas may be found at this link: https://docs.google.com/document/d/1hblQP4IVjZ1YGRHjILnjHhiIJelUMbTQq2gQB_H4ZU/edit

[11] Of the nine cancer challenges: Synthetic Data and ML for Hypothesis Generation are stand-alone; Adaptive Treatments was combined with Defining Optimal Treatments; and Collaboration

between Communities was integrated into the Cultural Barriers section. Four challenges that focused on bridging the spatio-temporal scales of cancer and the hallmarks of cancer were integrated into the Digital Twin challenge area.

[12] "Fair Principles." https://www.go-fair.org/fair-principles/ Retrieved from https://www.go-fair.org/go-fair-initiative/

[13] Hinkson, I. V., Davidsen, T. M., Klemm, J. D., Kerlavage, A. R., & Kibbe, W. A. (2017). A Comprehensive Infrastructure for Big Data in Cancer Research: Accelerating Cancer Research and Precision Medicine. *Frontiers in cell and developmental biology*, 5, 83. doi:10.3389/fcell.2017.00083

[14] Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*. *3*, 160035. doi: 10.1038/sdata.2016.35

[15] Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, Louis M. Staudt. 2016. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med*, *375*,1109-1112. doi: 10.1056/NEJMp1607591

[16] Gerard J. Kleywegt, Sameer Velankar, Ardan Patwardhan. (2018). Structural biology data archiving–where we are and what lies ahead. *FEBS Letters*, *592*,12, 2153-2167. doi.org/10.1002/1873-3468.13086

[17] Derek Wong & Stephen Yip. (2018). Machine Learning Classifies Cancer. *Nature* 555, 446-447. doi: 10.1038/d41586-018-02881-7

[18] Berry, Donald A., (2015). The Brave New World of clinical cancer research: Adaptive biomarker-driven trials integrating clinical practice with clinical research. *Molecular Oncology*, *9*, 951-959. doi: 10.1016/j.molonc.2015.02.011.

[19] "Need a Doctor? Send in your digital twin," *Cancerworld*, Sept. 24, 2018. (Retrieved: May 20, 2019 from https://cancerworld.net/spotlight-on/need-a-doctor-send-in-your-digital-twin/)

[20] Cranford, J.P., O'Hara, T.J., Villongco, C.T. et al. (2018). "Efficient Computational Modeling of Human Ventricular Activation and Its Electrocardiographic Representation: A Sensitivity Study. Cardiovasc Eng. Tech. *9*: 447–467. doi.org/10.1007/s13239-018-0347-0

[21] J. Ozik, N. Collier, R. Heiland, G. An, & P. Macklin. 2019. Learning-accelerated Discovery of Immune-Tumor Interactions. *Molec. Sys. Design Eng.* (*in review*). Preprint: https://dx.doi.org/10.1101/573972

[22] Hanahan, Douglas et al. (2000). The Hallmarks of Cancer. *Cell*, *100*,1:57-70.

[23] Si, T., & Zhao, H. (2016). A brief overview of synthetic biology research programs and roadmap studies in the United States. *Synthetic and systems biotechnology*, *1*(4), 258–264. doi:10.1016/j.synbio.2016.08.003

[24] Kramer MG, Masner M, Ferreira FA and Hoffman RM (2018) Bacterial Therapy of Cancer: Promises, Limitations, and Insights for Future Directions. *Front. Microbiol*. *9*:16. doi: 10.3389/fmicb.2018.00016

[25] Ibid.

[26] Karolak A, Markov DA, McCawley LJ, Rejniak KA. (2018). Towards personalized computational oncology: from spatial models of tumour spheroids, to organoids, to tissues. *J. R. Soc. Interface. 15*. doi.org/10.1098/rsif.2017.0703

[27] Link to the 1st Microlab: https://ncihub.org/groups/cicc/pastmeetings/june_11th_microlab

[28] ECICC Community Hub Site: https://ncihub.org/groups/cicc/overview

[29] Link to the 2nd Microlab: https://ncihub.org/groups/cicc/pastmeetings/sept25thmicrolab

[30] List of Personae for the 2nd Microlab: file:///C:/Users/borkonll/Downloads/Sept_25_Microlab_-_List_of_Personae_%20(1).pdf

[31] ECICC Community Hub site: https://ncihub.org/groups/cicc/overview

[32] http://knowinnovation.com/