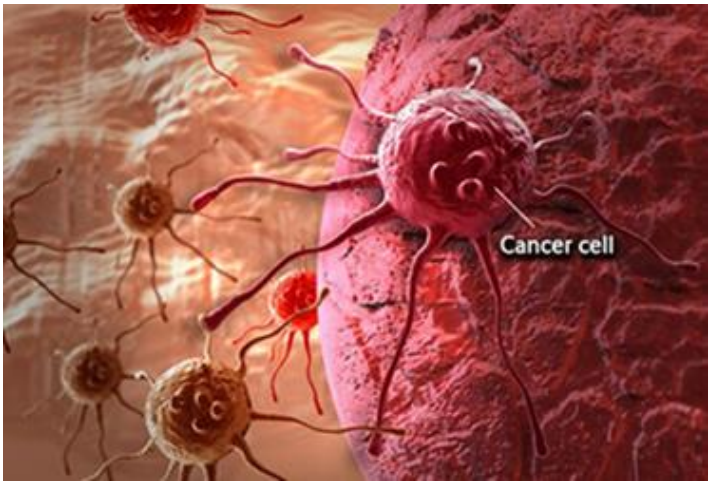


FRONTIERS OF PREDICTIVE ONCOLOGY AND COMPUTING II

ABSTRACT

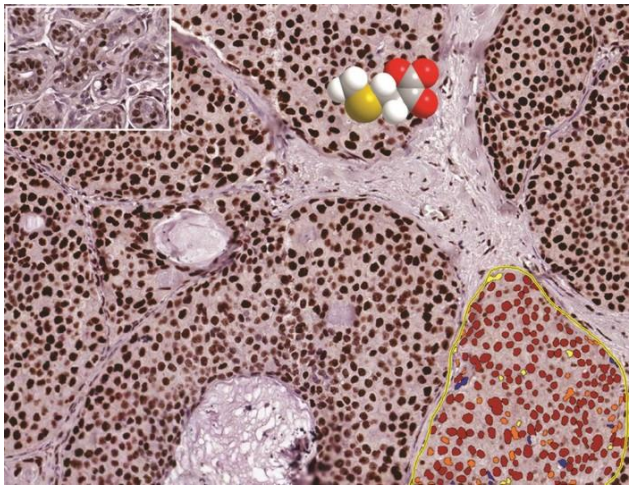
Over 80 of the nation's experts in oncology, computational and data science, and information technology met in New York City in October 2017 for the second annual Frontiers of Predictive Oncology and Computing meeting. Focusing foremost on accelerating impact to cancer patients, the cross-domain, multi-disciplinary group shared findings and insights about the convergence of advanced computing and precision oncology in cancer research. Building on the outcomes from the first meeting in 2016, they focused on how emerging capabilities and innovative computational approaches are revolutionizing cancer research, diagnostic and treatment options, and what's on the horizon for the future. This report offers key insight into the ground-breaking developments and outcomes of that meeting.

October 17-19, 2017

Frontiers of Predictive Oncology and Computing II
October 2017, New York, NY.

When citing, please reference as 'The Frontiers of Predictive Oncology and Computing II Meeting, New York, NY, October 2017

Contents



Preface

The annual Frontiers of Predictive Oncology and Computing (FPOC) meetings are an outgrowth of the Biological Applications of Advanced Strategic Computing (BAASIC) program initiated in 2015 by the Department of Energy's (DOE) Lawrence Livermore National Laboratory.

The BAASIC program is itself a component of DOE's Computational Predictive Biology program and focuses on exploring the opportunities and challenges involved in bringing together advanced computing and the life sciences. The goal of this synthesis is to apply the power of extreme computing, big-data analysis, and the explosion of knowledge in life sciences to make possible predictive simulations of human biology and, within that framework, transform the promise of predictive oncology into a reality guiding the clinical care of cancer patients. Realizing this goal requires the creation of a new generation of simulation tools and analytical approaches that can address research challenges of unprecedented complexity.

The FPOC meetings are an example of the convergence of interests driven to address challenges in cancer and offer a unique opportunity for thought leaders from leading public and private-sector organizations to share ideas and cross-domain expertise on new approaches to advance predictive oncology and computing.

Through a series of presentations, interactive sessions, and informal discussions, the participants engage in a multidisciplinary exploration of critical issues, challenges, and opportunities for accelerating the broader impact—and patient benefit—from both predictive oncology and computing technology.

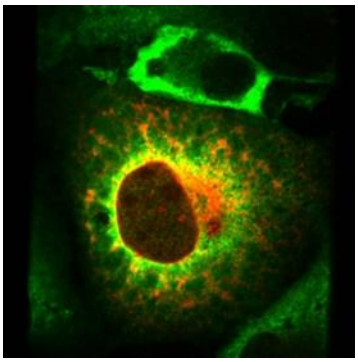
Overarching themes have included emerging avenues for patient impact, the use of advanced computing technologies, access and aggregation of data, and frontiers of describing, probing and measuring the disease in its many forms and stages.

The **first annual Frontiers of Predictive Oncology and Computing (FPOC)** meeting was held in July 2016 in Washington, DC. At this meeting over 100 experts from clinical practice, industry, government, and academia shared insights, knowledge, and a vision for the future of computationally predictive oncology. Experts from the NCI Cancer Moonshot℠ initiative, five national laboratories, the NCI, the FDA, leading universities, major cancer centers, and industry leaders from information technology, clinical diagnostics, and pharma participated.

The first meeting focused on the NCI-DOE collaboration, Joint Design of Advanced Computing Solutions for Cancer (JDACS4C), whose chief objective is to develop scalable and exascale-ready tools, algorithms, and capabilities to enable predictive methods for precision oncology.

At the first meeting, participants emphasized the benefit and importance of bringing together experts from different disciplines—in both oncological science and exascale computing—across the public and private sectors. Intel Corporation compiled the summary report for that meeting.

The **second annual Frontiers of Predictive Oncology and Computing meeting (FPOC II)**, and the basis of this report, was held in October 2017 in New York City and focused on challenges and opportunities in "computational pathology." Collectively participants explored the frontiers of approaches for observing, describing, and probing cancer in the context of predictive research and clinical applications. The attendees concentrated on the broader application of technology, computation, and domain expertise to accelerate understanding and treatment of cancer as a disease. They also explored the growth and impact of several new developments affecting the landscape of cancer research and discovery, including:



- New discoveries in genetics and genomic science, oncological science, and clinical and translational research
- New techniques in molecular and radiological imaging and pathology informatics
- Cutting-edge computational techniques in predictive analytics for cancer, including supercomputing technologies developed through the NCI-DOE collaboration, Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) and related joint efforts

The fact that 96% of clinical data for an individual patient is only available at a single institution was noted at both meetings. Participants emphasized that creation of a national cancer data ecosystem will ensure that cancer researchers, clinicians, and patients can access, share, and contribute data in the data analysis and annotation pipelines.

Several attendees participated in both the 2016 and 2017 meetings. Leaders from the following organizations (to name a few) participated in the 2017 meeting:

- Cancer centers such as the Albert Einstein Cancer Center, Stony Brook University Cancer Center, and Memorial Sloan Kettering Cancer Center
- Oncology research, bioinformatics, and high-performance computing centers such as the National Cancer Institute and Frederick National Laboratory for Cancer Research
- Computational centers of excellence such as the New York Center for Computational Sciences, the US Department of Energy National Nuclear Security Administration and the DOE national laboratories
- Private-sector technology innovators such as Intel Corporation and General Electric
- Pharmaceutical companies such as GlaxoSmithKline
- Medical schools such as Duke University School of Medicine, Albert Einstein College of Medicine, Harvard Medical School, and the Icahn School of Medicine at Mount Sinai
- Teaching hospitals such as Massachusetts General Hospital, the University of Pittsburgh Medical Center, and the University of California San Francisco Medical Center

Section 1. Introduction

The goal of the **Frontiers of Predictive Oncology and Computing (FPOC)** community is twofold: 1) to advance new capabilities with the potential to provide a greater understanding of cancer biology, improved diagnosis, and the discovery and development of molecularly targeted therapeutic agents; and 2) to make these new capabilities widely available to oncologists and patients.

The second **Frontiers of Predictive Oncology and Computing (FPOC)** meeting expanded the explorations begun in 2016 to develop shared concepts and insights to extend the collaborative development of a shared extreme-scale information-technology ecosystem and advanced analytical algorithms for cancer research. The second meeting broadened the focus to developments in computational pathology that can accelerate cancer research and expand the favorable impact to a greater number of cancer patients more effectively.

Implicit in predictive oncology is the need to effectively characterize current and future states of the cancer system. Experts in computing, applications of predictive oncology, pathology, instrumentation and imaging, informatics initiatives, and clinical applications explored and shared insights at the meeting about the current state and future directions in computational pathology and related areas. Participants examined how these rapidly developing areas are converging to create a common understanding of priorities for future collaborations.

The broad objectives of the **second Frontiers of Predictive Oncology and Computing** meeting were as follows:

- **Develop shared insight** into existing efforts and challenges about opportunities to impact cancer patients in multi-scale imaging, predictive oncology, and computing
- **Develop broader understanding and awareness of the important role of 'computational pathology' in predictive oncology—**and the insights it offers across multiple scales of time and space, from research to clinical applications
- **Identify and characterize new opportunities for collaboration, cross-domain education, and shared efforts** to address challenges and accelerate cancer research and clinical applications

The meeting included a variety of keynote presentations, plenary sessions, panel discussions, and interactive break-out groups. Reflecting FPOC priority areas of patient engagement and impact; data collection, analysis, and distribution; cultural change within organizations and multi-organization efforts; and rapid expansion of technology, breakout sessions were organized around the following three areas:

1. **The role of computational predictive oncology to inform cancer treatments**
2. **Predictive oncology software and algorithms**: present and future opportunities and challenges
3. **The evolving role of pathology, tissue, and biospecimen data** in predictive oncology

Guided by lessons and insights gained from the first BAASIC meeting in 2015 and the **first Frontiers of Predictive Oncology and Computing meeting (FPOC) in 2016, the second annual FPOC meeting** also focused on how computing and predictive oncology can accelerate discovery; how to address the challenges in achieving a shared understanding across the cancer research community; and the promise of cross-organizational and interdisciplinary collaborations. This meeting summary captures these elements, discussions, priorities, and recommended actions.

Predictive Oncology

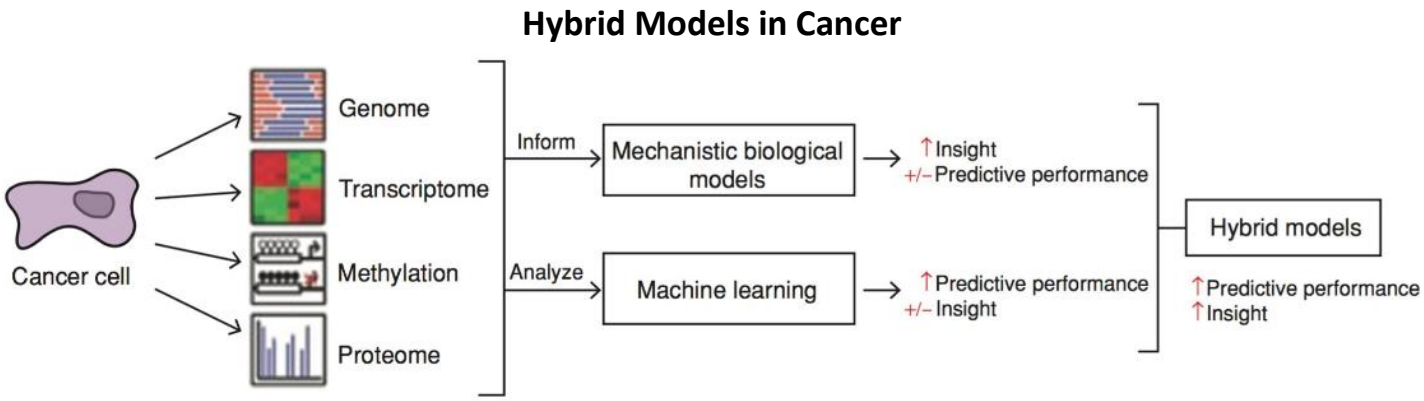
*Predictive oncol*ogy builds on the concept of *precision oncology* but represents a step beyond. The NCI defines *precision oncology* as follows:

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

This model includes an implicit element of predictive analytics to guide treatment selection, and it is this element along with concomitant advanced computational requirements that form the focus of predictive oncology.

The extreme-scale capabilities required for predictive oncology include management of data exceeding the petascale level—more than a million gigabytes— (necessary for attaining sufficient statistical power for probabilistic reasoning), integrative big-data analysis, machine learning, pattern recognition, modeling of complex systems, and predictive simulations at scale.

The full meeting agenda is available on pages 43-54.



The model, above, is one example of models being used for drug discovery.
Numerous algorithms and computational models are used across the spectrum of predictive oncology.

Section 2. Executive Summary

Intel, Lawrence Livermore National Laboratory, a Department of Energy National Laboratory, the Frederick National Laboratory for Cancer Research, and the National Cancer Institute welcomed leaders from academia, oncology research and clinical practice, government, technology industry, and pharma to the second Frontiers of Predictive Oncology and Computing meeting which was held on October 17-19 at the SUNY Global Center in New York City.

Over 80 scientists, clinicians, and computational experts gathered for two and one-half days to share perspectives and expertise. Given the variety of domains represented—including disciplines ranging from physics to artificial intelligence to cancer clinical care—the meeting provided a rare opportunity to continue building on the goals and outcomes of the first Frontiers of Predictive Oncology and Computing (FPOC) meeting in 2016; namely, a collaborative, multidisciplinary community around high-performance computing and predictive oncology.

The second Frontiers of Predictive Oncology and Computing (FPOC) meeting focused on extending and expanding the themes established at the first meeting in four main areas:

Patient-centric:	Patient engagement and impact
Organizational:	Cultural change within organizations and multi-organizational efforts
Data:	Data quality, access, intake, analysis, curation, aggregation, integration, sharing, and management
Technology:	Rapid expansion, development, and application of technological capabilities in computing, instrumentation, automation, and sensors

Like the 2016 meeting, sessions explored multiple initiatives, findings, and computational tools and methods that can accelerate patient outcomes and increase understanding of cancer biology. New and expanded opportunities and challenges were explored.

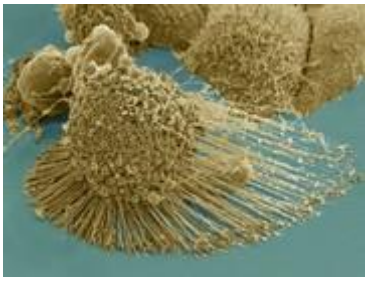
Collectively, this context is summarized with these main points:

- **Cancer researchers have identified the need for computationally intensive multi-scale approaches** to explore new treatments, guide active treatment decisions, and increase understanding of tumor heterogeneity.

- ***Pathomics* and *radiomics* are now mainstream capabilities** with the importance of radiology, pathology and omics integration being broadly recognized, even within its relatively early developmental stage.

- **A major Cancer Moonshot goal is to achieve broadly available and interoperable data, accessible by researchers, clinicians, as well as patients.** (There are 7,400 public and private project activities connected to the Cancer Moonshot.)

- **There is a need for a US national learning healthcare system for cancer; a system that focuses on patients' health, not a disease, and delivers care holistically.**



- **Through these efforts, many of which include public/private partnerships,** new cancer research techniques and new patient therapies are being developed.

Common Themes in the Keynote and Plenary Sessions

Several notable themes emerged across the keynote and plenary sessions of the meeting. These are briefly summarized as:

- **Aggregation and integration of information from multiple sources** is a common need. Whether this requires the fusion of similar data to reach levels required for effective deep learning, or the integration of multiple types of data to bring together multiple complementary observations, integrating data is a common requirement.

- **The amount of biologic data is growing exponentially.** The proliferation of data and computational tools have created emerging fields of study, such as single cell analytics, which can analyze enormous amounts of data points.

- **Longitudinal and time-dependent data have application and represent a critical need across multiple domains of predictive oncology**. This includes outcome and response data for patients, patient-derived models or molecular interactions. **The role of time-dependent data is critical**.

- **New technologies for observing different aspects of cancer and oncology across scales are being developed and applied in key areas—and at greater resolution and refinement.**

    - New software is being developed to manage the workflow (analysis pipeline intersecting imaging and *omics*).

    - Advanced analytics and computational tools are being used across domains to probe fundamental molecular functions, cellular compositions and makeup, and the dynamics of multiple cells.

        - New imaging techniques also provide analysis of data variability and multiple aspects of tumors over time. Automation delivers a broader integration of cell biology.

        - Multi-modal data integration is needed to identify tumor imaging signatures.

    - In the clinic, leading-edge technology is revolutionizing how medical data is collected, interpreted, and shared with diagnosticians and patients. New imaging techniques such as radiomics use quantitative data and provide insight beyond what can be seen from biopsied tissue alone. For example, computational data can be used to characterize the tumor phenotype and is adaptable for dynamic changes in tumor biology, including patient characteristics and differences in the biological effects of treatment.

- **Computing is emerging as an important integrating element for predictive biology**. Development of new platforms is leading to new computing capabilities. To ensure the right information is collected and used, a partnership between computational experts and scientists is needed. This will result in co-development of believable codes and actionable simulations (codes that deliver results we believe in) in predictive oncology models.

- **Rapid changes in technology have resulted in a broad gap in understanding among cancer researchers and clinicians** about the new capabilities and applications for research and clinical practice. As well, a legacy culture of limited data sharing capabilities and practices has created a lag in adoption of emerging technologies.

Panel discussions focused on two pivotal areas of predictive oncology and computing: 1) computational pathology and 2) longitudinal and multiscale data. Across the panels, numerous opportunities and challenges were identified and discussed, such as the integration of functional imaging with molecular assays and new technology tools and methods in machine learning that have significantly increased capabilities for computational pathology to accelerate the diagnostic process and understanding of cancer cells.

As well, common themes to accelerate patient impact included focusing on health holistically, not disease, the need to make predictive modeling more transparent or understandable across the lifecycle for patients, the importance of providing clear information to patients when they need to decide about treatment, and opportunities to leverage the influential patient-doctor relationship.

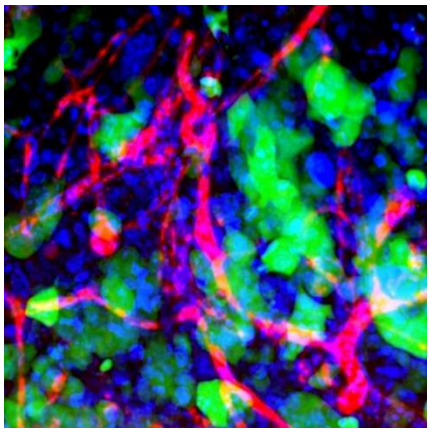
Three **breakout sessions** were held during the meeting, each one focusing on a distinct priority area of computational pathology in the context of predictive oncology.

Breakout I: Informing Cancer Treatments with Computational Predictive Oncology. Participants in this group focused on patients throughout their journey of care, from screening, biopsy, and diagnosis to treatment and survival. Potential strategies to improve data capture and curation, disparity in the amount of data for different types of cancer, and the ability of patients to use their own data (portability) were discussed, such as creation of a National Cancer Data Ecosystem that is inclusive, open, and interoperable.

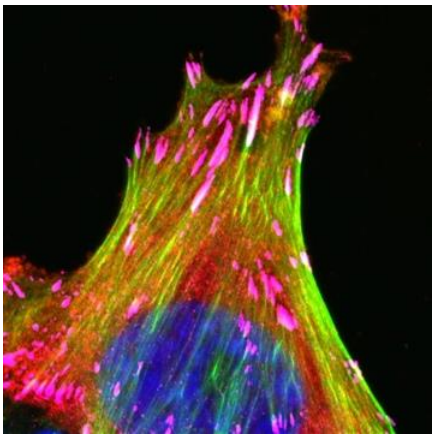
Breakout II: Predictive Oncology Algorithm and Software: Challenges, Opportunities, and Paths Forward. Discussion in this breakout highlighted the significant progress in the development of algorithms, marketing, data, and workflow. Other topics included current software and computational tools and resources, and strategies to substantiate the efficacy of machine learning tools, such as a standards body for quality assurance, scalability, and support for the tools, linking tool development to clinical outcomes, and adding tools into clinical workflows.

Breakout III: Evolving Role of Pathology, Tissue, and Biospecimen Data in Predictive Oncology and Analytics. Participants in this group discussed progress made since last year in three areas: 1) the use and computational analysis of data, 2) the development of new algorithms, and 3) the impact of the FPOC community's contributions to date on patient care.

Throughout the meeting, considerations and actions were proposed and examined to advance computational pathology and predictive oncology. Consensus on specific recommendations and actionable next steps was reached during the breakout sessions. Forward looking actions focused on improving the data available for predictive oncology; creating and expanding multi-disciplinary collaborations that include oncologists and patients; the tremendous potential for accelerated cancer research due to the growth and broader use of machine learning and other artificial intelligence tools; and the importance of cultural change within and across organizations to change mindsets and practices about accessing, using, and sharing data and adopting new technology in research and clinical practice.



Section 3. Current Efforts and Initiatives



Since the first Frontiers of Predictive Oncology and Computing meeting, several notable developments have occurred that created a new context for the meeting. These include the funding and launch of multiple efforts as part of the Cancer Moonshot program, the launch of the DOE Exascale Computing Project, new efforts for the integration of informatics and pathology data, and the meteoric rise in the level of industry support for applications of deep learning, to name just a few. The keynote presentations captured this updated context and served as anchors for the balance of the meeting. A recap appears below.

Keynote Presentations

The following national experts in cancer research and exascale computing attended the FPOC II meeting and delivered keynote presentations on current initiatives, findings, and challenges related to predictive oncology (listed alphabetically):

Speakers

Michael Becich, MD, PhD – Associate Vice-Chancellor for Informatics in the Health Sciences, Chairman, and Distinguished University Professor, Department of Biomedical Informatics, Associate Director, University of Pittsburgh Medical Center Hillman Cancer Center

Michael Idelchik – Vice President, Advanced Technology Programs, General Electric

Dimitri Kusnezov, PhD – Chief Scientist and Senior Advisor to the Secretary, National Nuclear Security Administration (NNSA), Department of Energy

Jerry S.H. Lee, PhD – Deputy Director, Center for Strategic Scientific Initiatives, National Cancer Institute, Deputy Director for Cancer Research and Technology, Cancer Moonshot℠ Task Force

Joel Saltz, MD, PhD – Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Presentations

Dr. Joel Saltz (Stony Brook University) opened the meeting with an overarching summary of the cancer research landscape. **Drs. Jerry S.H. Lee (NCI)** and **Dimitri Kusnezov (DOE/NNSA)** provided additional context with status updates on numerous programs that comprise the Cancer Moonshot℠. **Dr. Michael Becich (University of Pittsburgh Medical Center)** offered insights on computational approaches for pathology. **Michael Idelchik (General Electric)**, provided a future trajectory for predictive oncology as it parallels the role of predictive modeling in other industry transformations.

Collectively, this context is summarized with these main points:

- **Cancer researchers have identified the need for computationally intensive multi-scale approaches** to explore new treatments, guide active treatment decisions, and increase understanding of tumor heterogeneity.

- **Pathomics and radiomics are now mainstream capabilities** with the importance of radiology, pathology and omics integration being broadly recognized, even within its relatively early developmental stage.

- **A major Cancer Moonshot goal is to achieve broadly available and interoperable data, accessible by researchers, clinicians, as well as patients**. (There are 7,400 public and private project activities connected to the Cancer Moonshot.)

- **There is a need for a US national learning healthcare system for cancer; a system that focuses on patients' health, not a disease, and delivers care holistically**.

- **Through these efforts, many of which include public/private partnerships, new cancer research techniques and new patient therapies are being developed.**

Selected major points from each of the keynote presentations are summarized below.

Opening Remarks*: Context and History for the Frontiers of Predictive Oncology and Computing (FPOC) Meeting—Dr. Joel Saltz*

- In 2016, the Department of Energy and the National Cancer Institute established a three-year pilot collaboration to leverage DOE's supercomputing capabilities and computational expertise for cancer research and discovery.

- Today pathologists are using new analytical tools such as digital pathology, that have potential to generate multiple exabytes of data. There are at least eight major NCI supported projects that involve digital pathology, with several presentations in the meeting bringing more specifics and details into the discussion.

US Federal Cancer Moonshot: One Year Later—Dr. Jerry S.H. Lee and Dr. Dimitri Kusnezov

- New cancer research techniques and new patient therapies are being developed using artificial intelligence and exascale computing. Joint efforts, such as the DOE-NCI collaboration Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) and Accelerating Therapeutics for Opportunities in Medicine (ATOM), include contributions from multiple sources, including pharma, technology, and government.

- Heterogeneous computing represents a frontier in supercomputing, with different elements and different characteristics solving different problems. Examples of heterogeneous computing include memory intensive, neuromorphic learning systems, and quantum computing.[1]

- Multi-scale integrative analysis and multi-scale simulations offer tremendous promise to predict treatment outcomes, select and monitor treatments, explore new classification schemes, and reduce human variability in diagnosis.

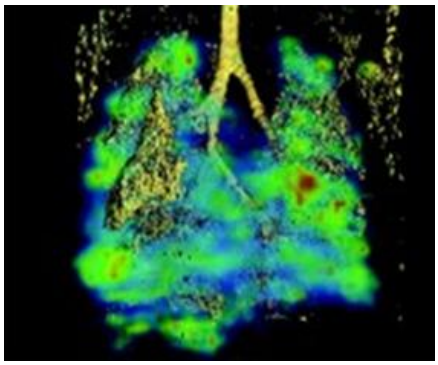
Towards a Digital Pathology Commons: Why this is the Right Time to Forge a Path to Predictive Oncology?—Dr. Michael Becich

- Digital methods and tools such as computational pathology, large-scale computing, and Whole Slide Imaging (WSI) are being used in predictive oncology. Whole slide imaging is a new technique approved by the FDA for use beginning in the spring of 2018 that provides digitized images of biopsied tissue, eliminating the need to evaluate tissue samples from a microscope. WSI requires specialized hardware and software.

- Intra-tumoral spatial heterogeneity complicates accurate diagnosis and prognosis. A demonstration showed how computationally annotated WSI can be used to quantify tumor heterogeneity and provide insight on a host response to a tumor.

- FDA-approved drugs could help a greater number of patients more expeditiously if researchers could find ways to work more closely with industry to reduce barriers to pre-clinical validations.

- The cloud-based information commons pilot between the University of Pittsburgh School of Medicine and Harvard University, called PIC-SURE. (https://pic-sure.org/) offers a concrete example of the future direction for computational pathology. As a biomedical Information Commons, the resource combines genetic, environmental, imaging, behavioral, and clinical data of individual patients from multiple sources into integrated data sets. FAIR (Findable, Accessible, Interoperable, and Re-Usable) principles will shape future informatics environments. (https://www.force11.org/group/fairgroup/fairprinciples)



[1] Heterogenous computing refers to systems that use more than one kind of processor core. These systems gain performance or energy efficiency not just by adding the same type of processors, but by adding dissimilar co-processors, usually incorporating specialized processing capabilities to handle tasks.

Learning from Industry Challenges in Multiscale Analytics and Relevance to Cancer Research and Imaging—Michael Idelchik

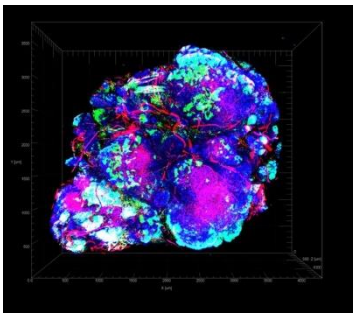
- Lessons from industry challenges in multiscale analytics and imaging are relevant to cancer research. Key steps in GE's approach to modeling were delineated, starting with problem definition. A jet engine is a good example of modeling at the microscale. For example, defining every problem on an engine offers predictive information on the engine's life expectancy.

- There is a critical need to understand measurement error and uncertainty in the model to understand the limits of applicability, as well as characterize areas for improvement. Model representation and validation are also key steps. With limited data, an effective approach is to simplify models by evaluating key aspects, including parameters, measures, and variances.

- In a parallel application to cancer research, GE developed a new multiplexed imaging platform technology called MultiOmyx™ that conducts multiplexed tissue analysis to quantify cell types, function, and spatial location. With this capability, the cancer community could increasingly understand core behaviors of cancer cells.

Plenary Sessions

The plenary sessions for the meeting provided added detail and context for specific applications at the frontiers of predictive oncology and computing. While the multi-disciplinary group shared insights on several different areas, the overriding focus was on patient impact and outcomes.



Experts highlighted their research and clinical findings and the group discussed how to expand the frontiers of computing and predictive oncology to improve diagnostic understanding of cancer, develop new treatments, or more accurately characterize underlying cancer biology.

Updates on the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) collaboration between the NCI and the DOE also provided insight into the progress, challenges, and opportunities at the convergence of exascale computing and predictive oncology.

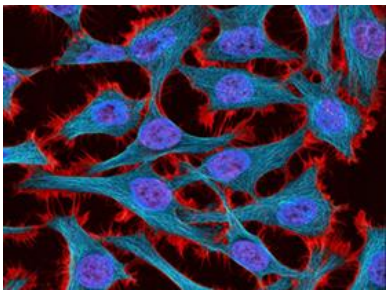
Each plenary session focused on a different frontier area of predictive oncology, listed below. Presentation summaries follow.

Plenary Session I: **Drivers Impacting Computational Pathology: Patients, Treatments, and Improving Outcomes**

Plenary Session II: **Technologies to Probe Biology: Unlocking Frontiers of Computational Pathology**

Plenary Session III: **Updates on the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C): Frontier Collaborations in Predictive Oncology and Computing**

Plenary Session I: Drivers Impacting Computational Pathology: Patients, Treatments, and Improving Outcomes



The presentations in the initial plenary session highlighted the specific ways in which computational approaches in cancer may translate into the development of improved outcomes for cancer patients. These impacts range from improved diagnostics to new treatments and a better understanding of individual cancers.

Predictive Modeling for Pre-Clinical Screening—Dr. Janet Eary, Deputy Associate Director, Cancer Imaging Program, National Cancer Institute

- The presentation focused on applications in cancer imaging. The issue of tumor heterogeneity was highlighted in the discussion. There is tremendous genetic diversity both across tumors and within a single tumor. Cancer is a dynamic biologic process, with continuous changes; therefore, the element of time in the biologic process is the new frontier.

- Of note, the evolution of the cancer cells within a tumor can be seen at early stages. Advanced molecular imaging and radiomics offer new ways to characterize tumor phenotypes and discover relationships between genotypes and phenotypes. Incorporating different imaging techniques and computational data is critical to advancing data analysis for clinical use and patient treatment.

Scale and Challenges of Available Data—Dr. John Baldoni, Senior Vice President of Artificial Intelligence Discovery, GlaxoSmithKline (GSK)

- The new Accelerating Therapeutics for Opportunities in Medicine (ATOM) public-private consortium leverages a large scale of data (https://atomscience.org/). The vision for ATOM is to integrate high-performance computing, diverse biological data, and emerging new biotechnologies to transform drug discovery from a slow, sequential, iterative, high-failure process to a rapid, integrated, and patient-centric model.

- In the ATOM program, deep learning technologies (DL) have been applied in optical biopsies and multi-modal spectroscopy. Deep learning methodologies offer tremendous potential to see how molecules interact on the biopsied tissue.

- Using a mathematical approach and computational techniques applied to MALDI (Matrix Assisted Laser Desorption Ionization Mass Spectrometry) data, GSK's researchers can generate data consistently across multiple studies. It is critical to monitor data longitudinally. Integrating ATOM's findings into a pre-competitive platform will be critical to the program's success. Additional collaboration with other pharmaceutical companies, government agencies, and technology companies is being sought.

Integrative Genomics Approach for Precision Medicine in Cancer—Dr. Kun Huang, Assistant Dean for Data Sciences, IUSM PHI Chair for Genomics Data Sciences, Professor of Medicine, Indiana University

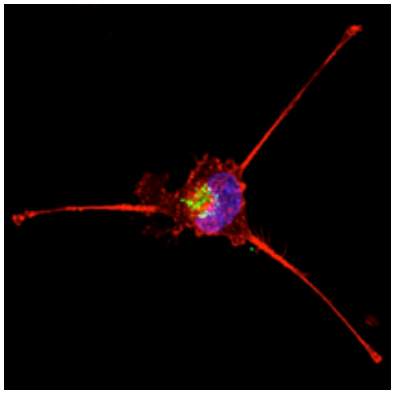
- The central theme of this presentation was the value of an integrative genomics and trans-omics approach for precision medicine in cancer. Identifying the correct images as well as discovering and expressing variety are among the key challenges in data integration.

- Various imaging techniques, such as 3D shape characterizations (3D nuclei), were highlighted, along with the integrated role of spatial pattern analysis, segmentation, visualization, quantification, computing, and histopathology features for future cancer prognosis. A pipeline overview using whole slide images (WSI) to conduct feature extraction leads to insight into feature distribution and topological characteristics.

Plenary Session II: New Technologies to Probe Biology: Unlocking Frontiers of Computational Pathology

The second plenary session explored the frontier of developing technologies with the potential to provide even greater insight into the behavior of cancer. Several technologies were presented that offer various capabilities for exploring approaches to characterize individual cells across dimensions of time and space.

Dr. Fiona Ginty, Biosciences Technical Operations Lead and Principal Investigator, GE Global Research Center

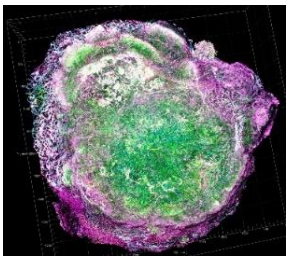


- In this presentation, GE's work on *in situ* single cell analysis and spatial cell analysis were highlighted. Single cell analytics is a fast-emerging field which generates tremendous amounts of information leading to an analysis of millions of data points.

- GE has partnered with NCI-supported investigators and pharma organizations on multiplex analysis. A new research focus is on tumor heterogeneity (within and across patients). GE's fluorescence-based imaging platform (MultiOmyx™/Cell Dive) can multiplex up to 60 proteins in the same sample and—unlike traditional homogenic methods—can provide visibility into the precise location of the biomarkers.

- Because tumors undergo dynamic changes, multiplexing at any level provides insights on the tumor microenvironment. Two points were emphasized: the importance of longitudinal data and the need to understand what is required to characterize the unique biologic features of each individual patient.

Multiphoton Intravital Imaging at Single Cell Resolution Reveals Mechanisms of Cancer Dissemination—Dr. John Condeelis, Director, Tumor Microenvironment and Metastasis Program, Scientific Director, Analytical Imaging Facility, Albert Einstein Cancer Center

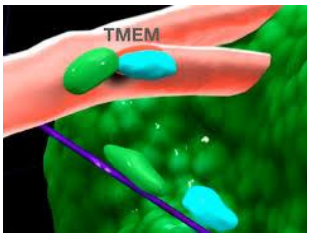


- This presentation focused on techniques for understanding the dissemination of tumor cells. The dispersion of tumor cells is what ultimately leads to the death of the patient, not the isolated dynamics of the primary tumor.

- An emerging technology—computational intravital imaging—enables real-time analysis of single cell phenotypes in very large tumor volumes. A multi-photon, multi-channel instrument, coupled with a computationally directed algorithm delivers a dynamic 3-dimensional view of tissue cells. These capabilities provide insight and approaches to improve understanding of tumor heterogeneity and to identify the relevant microenvironments and mechanisms driving the phenotype.

- These technologies provide insights into the role of the tumor microenvironment of metastasis (TMEM). A complex of cells, specifically an invasive carcinoma cell, a vascular macrophage and its associated endothelial cell, called TMEM, is, in effect, a door for the tumor cells to spread to blood vessels—and is also a prognostic marker of outcome. The drug, Rebastinib, which inhibits TMEM has been validated.[2]

The Impact of Chemotherapy on TMEM-Mediated Cancer Cell Dissemination—Dr. Maja Oktay, Professor, Department of Pathology, Department of Anatomy and Structural Biology, Albert Einstein Cancer Center



- Building on the discovery of TMEM's role in cancer dissemination, studies on how to counteract the effects of chemotherapy on TMEM function are underway, with a focus on the microenvironment for cell activity.

- In pre-clinical and clinical studies, results have been troubling, with increased TMEM scores after chemotherapy. Higher TMEM scores mean there are more pathways for the tumor cells to spread to blood vessels. In addition, the role of race appears to be a significant potential factor in the risk for developing distant metastasis.

- A working model was developed that produced the same relapse and survival rates results as the other studies. To mitigate the effects of chemotherapy on TMEM function, the combination of chemotherapy with the drug Rebastinib is more effective in improving the long-term outcome for the patient.

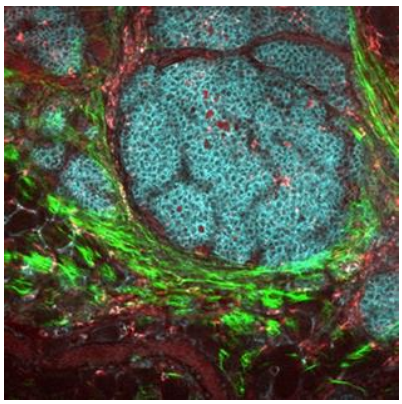
[2] *Neoadjuvant chemotherapy induces breast cancer metastasis through a TMEM-mediated mechanism,* Copyright © 2017 Science Translational Medicine 05 Jul 2017: Vol. 9, Issue 397, eaan0026, DOI: 10.1126/scitranslmed. aan0026 http://stm.sciencemag.org/content/9/397/eaan0026

- When evaluating patient treatment, unique physical and environmental factors (such as age, race, and co-morbidities) contribute to systemic response in patients that can change in a microenvironment. This tumor microenvironment can affect the growth of cancer cells; therefore, treatment must also target the interaction of the tumor microenvironment with cancer.

Investigating the Tumor-Immune Microenvironment using Cancer Genomics and Computational Pathology—Dr. Vesteinn Thorsson, Senior Research Scientist at the Institute for Systems Biology

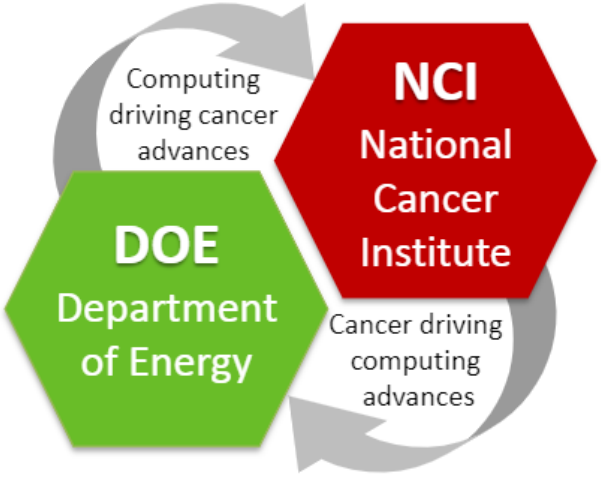
- The presentation focused on the progress made to date in understanding the tumor microenvironment. The use of hematoxylin and eosin stain (H&E) imaging, various scales—including spatial, molecular, cellular, tissue, organism, H&E, immunohistochemistry (IHC) and cancer genomics—along with the inhibition of checkpoint proteins in immunotherapy offers tremendous promise.

- The integration of molecular and cellular scales can help identify the proportion of immune cells on large tumors, which provides key diagnostic information. The combination of the NCI cloud resources (https://cbiit.cancer.gov/ncip/cloudresources) and data annotation help make relevant pieces of data available, such as The Cancer Genome Atlas (TCGA) data. The benefits of using the cloud and extreme scale computing for quick scale-up in the pipeline and assessment of the quality of the tumor's vital characteristics were also discussed.



Plenary Session III: DOE-NCI Collaboration: Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)—Frontier Collaborations in Predictive Oncology and Computing

Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) *is a three-year pilot collaboration created in 2016 between the National Cancer Institute (NCI) and the Department of Energy (DOE)* to advance the missions of both agencies. *(*https://cbiit.cancer.gov/ncip/hpc/jdacs4c*)*



Now starting the second year of the pilot, this session highlighted the progress made in the first year. The leaders of the three pilots from the DOE National Laboratories, the National Cancer Institute, and the Frederick National Laboratory for Cancer Research presented updates on their efforts at the convergence of the frontiers of predictive oncology and computing. Their presentation summaries appear below.

Pilot I: Molecular Scale Predictive Oncology—Dr. Yvonne Evrard (Frederick National Laboratory for Cancer Research) and Rick Stevens (Argonne National Laboratory)

Goal: *Advance high-performance computing technologies and capabilities in conjunction with accelerating efforts to improve and expedite the development of new targeted therapies for cancer patients*.

Access to relevant datasets is key to achieving the pilot's three objectives:

1) Build machine learning models to predict the results of pre-clinical screening
2) Apply experimental design principles to improve decision-making about future experiments
3) Provide predictive and explanatory information

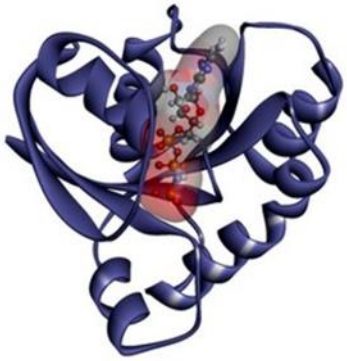
- This pilot currently focuses on developing new computational approaches, using available information to identify novel ways to perform the model transfer, and using cell line data to make quantitative predictions to another domain, such as newly emerging patient-derived xenograft data (PDX), a novel technique to grow patient tumors in a mouse.

- The broadest set of available data for capability development is found in the NCI 60 cell lines which have been screened for years on hundreds of thousands of compounds. Predictive modeling efforts are constrained, however, by small sample sizes relative to the complex biology of cancer (including heterogeneity within a patient tumor) and lack of fully annotated clinical cases.

- New data are anticipated using patient-derived xenografts (PDX). While PDX models take a long time to grow compared to traditional models, the goal of NCI's Patient-Derived Models Repository (PDMR) (https://pdmr.cancer.gov) is to develop at least 1,000 PDX models, fully characterized, clinically annotated, and sequenced for public availability. Modeling drug response in traditional preclinical models has been an ongoing challenge due to limited clinical data (molecular profiles, treatment, and outcomes).

Pilot II: Pre-clinical Scale Predictive Oncology— Dr. Dwight Nissley (Frederick National Laboratory for Cancer Research) and Dr. Fred Streitz (Lawrence Livermore National Laboratory)

Goal: *Produce an unprecedented scale of predictive molecular dynamic model simulation for the RAS oncogene system to facilitate understanding leading to drug discovery and development. Machine learning serves a vital role in developing computational approaches for this pilot.*



- Increasing understanding of how the RAS protein works in the cell membrane is fundamental because the membrane is the only place in the cell in which RAS is active. Very large-scale simulations made possible by DOE supercomputing capabilities have provided early insight about the physical RAS-membrane interaction at the atomistic level.

- Atomistic simulations comprised of millions of atoms can be performed on a single membrane. Machine learning allows the use of very fine-scale simulation (thousandths of micro-seconds) for greater understanding of cell activity and interactions.

- Deep learning is envisioned to drive the multi-scale simulations necessary for greater insight. Molecular Dynamics (MD) and Coarse-Grained Molecular Dynamics (CGMD) are being used to develop predictive models that can span timescales and enable atomistic molecular simulations to reach biologically relevant timeframes.

- A close collaboration between experimentalists and theorists is needed to build the predictive models, with experimental insight providing data and information to develop models and confirm predictions. The experimentally supported simulations will start to answer some of the unknowns.

Pilot III: Population Scale Predictive Oncology—Dr. Paul Fearn (National Cancer Institute), and Dr. Georgia Tourassi (Oak Ridge National Laboratory)[3]

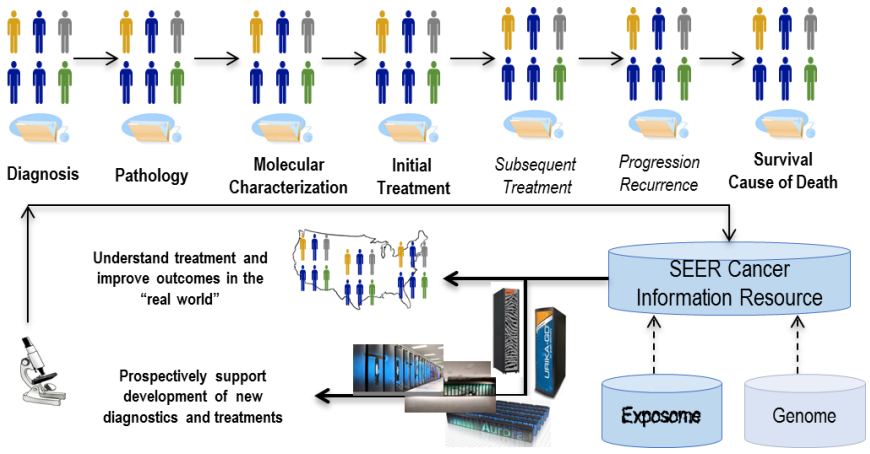
Goal: *Transform cancer care by applying advanced computational capabilities to population-based cancer data to understand the impact of new diagnostics, treatments, and patient factors in real-world patients.*

- This research pilot works productively with industry, the population science community, and the clinical community. Scalable machine learning for natural language processing (NLP) is used to automate the extraction of key patient information to extend the depth and breadth of the Surveillance, Epidemiology and End Results (SEER) database.

- Using a large dataset of pathology reports from the Louisiana Tumor Registry, an experimental pipeline was developed based on a dynamic "plug and play" approach; one that allows for the addition of new systems and various new algorithms (such as rule-based, machine learning, deep learning, and different feature representation). This plug and play ecosystem is much different from the static, controlled ecosystem used in clinical trials, and can be used by the broader SEER community.

- Three different types of deep learning methods were developed which offer a broad framework and a toolbox for the research community to use. Preliminary results have been published—including in the Journal of American Medical Informatics Association (JAMIA)[4] and the IEEE Journal of Biomedical and Health Informatics[5]—and members of this pilot have conducted workshops with other researchers on using natural language processing and deep learning.

- Moving forward, the broader applicability of algorithms across multiple registry sites is a major question to be addressed. Another key issue is the ultimate translation of results from predictive models, so the information is usable for the patient and favorably impactful on patient care.



[3] *Steve Friedman, a senior advisor to the SEER program at NCI and a member of the Pilot III team, substituted for Dr. Paul Fearn, the Pilot III Co-Lead, for this presentation.*

[4] Gao, Shang, Michael T. Young, John X. Qiu, Hong-Jun Yoon, James B. Christian, Paul A. Fearn, Georgia D. Tourassi, and Arvind Ramanthan. 2017. "Hierarchical Attention Networks for Information Extraction from Cancer Pathology Reports." *Journal of the American Medical Informatics Association: JAMIA*, November. https://doi.org/10.1093/jamia/ocx131.

[5] Qiu, John, Hong-Jun Yoon, Paul A. Fearn, and Georgia D. Tourassi. 2017. "Deep Learning for Automated Extraction of Primary Sites from Cancer Pathology Reports." *IEEE Journal of Biomedical and Health Informatics*, May. https://doi.org/10.1109/JBHI.2017.2700722.

Computing Frontiers—Cross-Cutting Technologies in the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)

At the frontiers of predictive oncology and computing, new technical approaches are being developed with broad applicability across multiple domains. Summaries of presentations regarding new technology approaches being used across all three JDACS4C pilots appear below.

Uncertainty Quantification: Dr. Tanmoy Bhattacharya, Los Alamos National Laboratory

- This presentation provided context for uncertainty quantification (UQ) in deep learning (DL) projects; specifically, that UQ provides an understanding and allows division of work between machines and humans. Some key elements that affect predictive model uncertainty include the data distributions, the overall reliability of the data, and the structure of the deep learning.

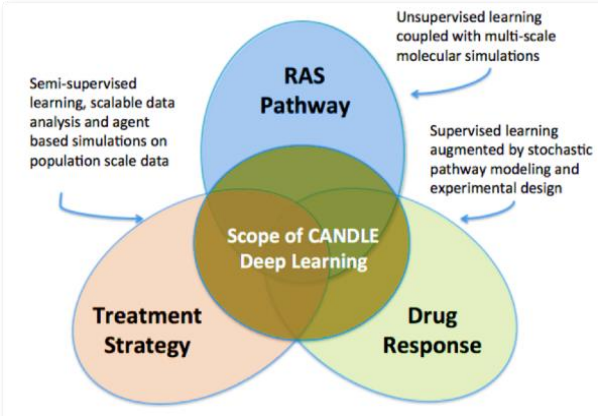
- UQ in deep learning is in the early stages of development and currently has several known limitations, including accounting for sampling bias imparted by the choices of data used and empirical error related to errors inherent in the measurement of data used to develop new models. A framework was highlighted to address some of the limitations, noting the important role of designing theories about the learning capacity of DL.

- Uncertainty quantification (UQ) offers tremendous benefits that can reduce human workload and accelerate the rate of discovery. For example, understanding and measuring UQ allows the improved design of experiments, provides insight into statistical errors, and adds confidence parameters employed in specific model predictions. UQ can also quantify the effects of sampling biases.

- As more data becomes available and models become more complex, there will be a greater need for computing across domains to handle multi-modal uncertainties. UQ methods draw upon the expertise developed previously by DOE, and the JDACS4C pilots are adding to the toolkit, which will benefit other scientific deep learning projects as well.

CANDLE – CANcer Distributed Learning Environment - Rick Stevens, Argonne National Laboratory

- The context and capabilities of the new cross-cutting exascale computing program known as CANDLE (CANcer Distributed Learning Environment) were explained in this presentation. Each of the three JDACS4C pilots has a need for machine learning and this cross-cutting effort provides infrastructure supporting the deep learning needs in all the projects. Deep neural networks inherent in CANDLE can already utilize a huge volume of data with which to create various deep learning models—with dimensions ranging into the millions.

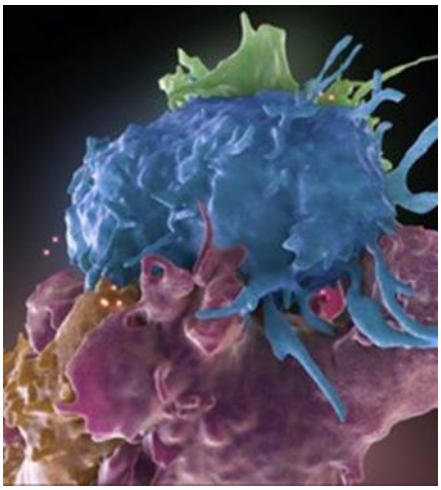


- The goals for CANDLE are:
  - Support all three JDACS4C pilot programs
  - Build scalable open source platforms on top of the exascale infrastructure
  - Optimize for CORAL computers[6]
    - Provide a basis for collaboration with vendors that work on the exascale systems

- Seven CANDLE benchmarks include representative cancer problems from across the three JDACS4C pilots, are freely available, and may be downloaded on GitHub. (https://github.com/ECP-CANDLE/Benchmarks)

- Ultimately, neural networks may enable computers to self-design new deep neural networks and will expedite problem-solving significantly.

- The DOE/NCI collaboration is influencing the evolution of supercomputing. The future capabilities computational hardware will offer at exascale and generally were also noted.



[6] CORAL is a first-of-its-kind U.S. Department of Energy (DOE) collaboration between the National Nuclear Security Administration's (NNSA's) ASC Program and the Office of Science's Advanced Scientific Computing Research program (ASCR) that will culminate in three ultra-high performance supercomputers at Lawrence Livermore, Oak Ridge, and Argonne national laboratories. The systems, delivered in the 2017 timeframe, will be used for the most demanding scientific and national security simulation and modeling applications, and will enable continued U.S. leadership in computing. https://asc.llnl.gov/coral-info

Common Themes in the Plenary Sessions

Several notable themes emerged across the plenary sessions of the meeting. These are briefly summarized as:

- **Aggregation and integration of information from multiple sources** is a common need. Whether this requires the fusion of similar data to reach levels required for effective deep learning, or the integration of multiple types of data to bring together multiple complementary observations, integrating data is a common requirement.

- **The amount of biologic data is growing exponentially**. The proliferation of data and computational tools have created emerging fields of study, such as single cell analytics, which can analyze enormous amounts of data points.

- **Longitudinal and time-dependent data have application and are a critical need across multiple domains of predictive oncology**. This includes outcome and response data for patients, patient-derived models or molecular interactions. **The role of time-dependent data is critical**.

- **New technologies for observing different aspects of cancer and oncology are being developed and applied in key areas—and at greater resolution and refinement.**

    - New software is being developed to manage the workflow analysis pipeline that intersects imaging and omics.

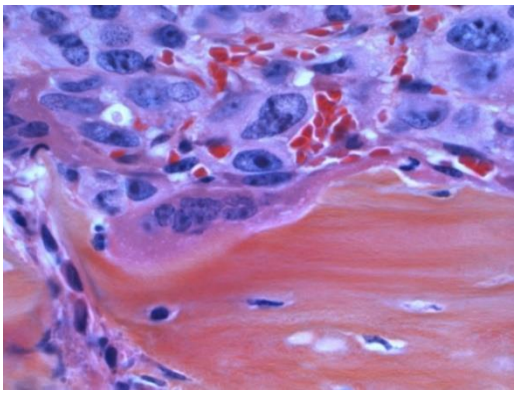
    - Advanced analytics and computational tools are being used across domains to probe fundamental molecular functions, cellular compositions and makeup, and the dynamics of multiple cells.

        - Advanced imaging techniques such as radiomics (which uses quantitative data to enhance visual observations) enable characterization of the phenotype, which is not possible with biopsied tissue.
        - New imaging techniques also provide analysis of data variability and multiple aspects of tumors over time. Automation delivers a broader integration of cell biology.
        - Multi-modal data integration is needed to characterize tumor imaging signatures.

- **In the clinic, leading-edge technology is revolutionizing how medical data is collected, interpreted, and shared with diagnosticians and patients.** Computational data can be used to characterize the tumor phenotype adaptable for dynamic changes in tumor biology, such as patient characteristics and differences in the biological effects of treatment.

- **Computing is emerging as an important integrating element for predictive biology.** Development of new platforms is leading to new computing capabilities. To ensure the right information is collected and used, a partnership between computational experts and scientists is needed. This will result in co-development of believable codes and actionable simulations (codes that deliver results we believe in) in predictive oncology models.

- **Rapid changes in technology have resulted in a broad gap in understanding among oncology researchers and clinicians** about the new capabilities and applications for research and clinical practice. As well, a legacy culture of limited data capabilities and practices has created a lag in adoption of emerging technologies.



- **Collaboration among imaging scientists, clinicians, and computational scientists offers additional insight and a more holistic view of cancer as a disease.** A cross-discipline collaboration is the best approach for determining the appropriate treatment options for patients.

Section 4. Opportunities and Challenges at the Frontier of Predictive Oncology and Computing

Throughout the meeting, participants identified key opportunities to expedite predictive oncology and computing, along with the related challenges. Informal sessions, including panel discussions and breakout sessions, offered opportunities for dialogue and deep dives among attendees at the meeting.

Panel Discussions

Panel discussions focused on two pivotal areas of predictive oncology and computing: 1) computational pathology and 2) longitudinal and multiscale data. Across the panels, common areas of opportunities and challenges emerged in the following areas:

Patient-centric:	Patient engagement and impact
Organizational:	Cultural change within organizations and dynamics of multi-organizational efforts
Data:	Data quality, access, intake, analysis, curation, integration, sharing, and management
Technology:	Rapid expansion of technological capabilities

Highlights of the opportunities and challenges discussed appear below.

Patient-centric Opportunities and Challenges—Patient engagement and impact:

- **Establish a national learning healthcare system for cancer to:**
    - Enable sharing knowledge and experience of every cancer patient
    - Unleash the power of data to enhance, improve, and inform the journey of every cancer patient from the point of diagnosis through survivorship
    - Increase efficiency to reduce costs to patients
- **Focus on the patient's health, not just the disease.** There is a need to change thinking and perspective on what defines 'normal.' There is no normal, all health is a trajectory of some kind—people are becoming either healthier or increasingly sick. It is key to **translate data into knowledge and manage knowledge to help patients.**
    - Turn sick care into healthcare by understanding causation
    - New paradigm: combine computational (molecular data) and systems pathology

- **Provide the information patients need at the time they must decide about treatment and in terms they embrace and understand** (e.g., common language instead of jargon, as well as implications of undergoing or foregoing a treatment). Patients are best positioned having and understanding information when they are ready to make decisions about their treatment, not afterward.
- **Local treatment and relationship with the doctor.** Most patients receive cancer care and treatment in their local community; above all else, the relationship and trust between a patient and his/her doctor is paramount. One study found that 80% of patients were more compliant with prescribed treatment just by monitoring them.
    - At present, there is a large gap between research findings and local treatment, particularly outside large metropolitan areas**. Closing this information and treatment gap is a priority.**

- **Make predictive modeling more transparent or understandable across the lifecycle for patients.** Treatment is not a one-time event, it is a series of interactions and therapies at various stages of the disease. **Find ways to connect with the patient such as in those areas mentioned below:**

    - **What do patients need to know so they feel involved?** For example, computer tools can be provided at the patient's bedside for instantaneous re-assessment and transparency to the patient.[7] **Make it possible for patients to access, use, contribute to and move their own data.**

    - **What does the broader community of oncologists need to know?** Doctors need to understand outcome, so they can recommend a path for progress.

Data Opportunities and Challenges—data collection, intake, analysis, aggregation, and curation; and sharing data, information, and knowledge:

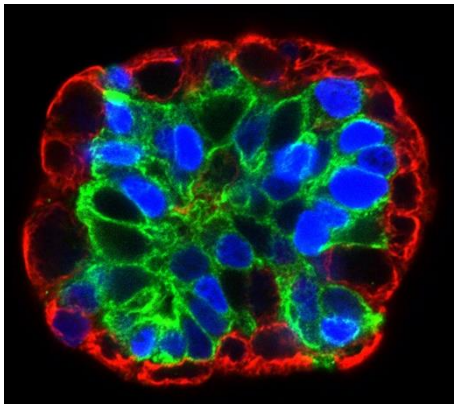
- **Amount of available data is still a big issue.**

    - **Make research data open and available in real time.**

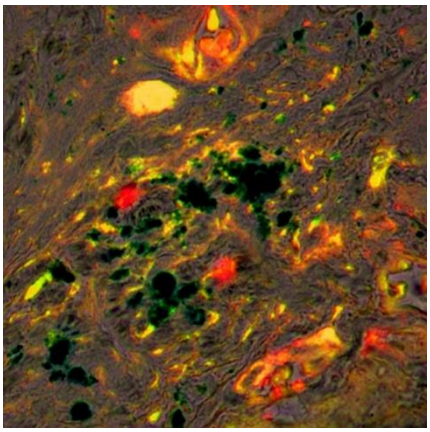
    - **Create more shareable clinical data and make it open to researchers, patients, and the public. This includes molecular profiles, treatment, and outcome data.** Training data is also very limited and is needed to create predictive models for survivorship curves. Machine learning tools exist today that can develop these models, but there is not enough data to do so. Explore the role of machine learning and AI to support data sharing.

    - **Through machine learning, an explosion of biodata is becoming available along with new techniques.** In combination with new machine learning algorithms, it is now possible to do things that have never been possible before.

- **Develop data standards to facilitate sharing data across systems and across studies.**



[7] Dr. Rachael Calcutt presented a video from the UCSF Center for Digital Health Innovation (CDHI) *Smarter Health* illustrating the use of new computational assessment tools at the UCSF Medical Center. (http://centerfordigitalhealthinnovation.org/our-portfolio/#portfolio)



- **Provide additional mechanisms to expand minority participation in clinical trials.** Minority populations are significantly under-represented in clinical trials. As well, it was reported that only five percent of patients (mostly Caucasian) currently participate in clinical trials. Increase participation among all populations.

- **Concerns about privacy** and other issues must be overcome. Confidential data can remain at the institution at which data are collected to help allay privacy concerns. The algorithm is the only thing that must be moved, not the data itself. Technology offers methods to move datasets, so they cannot be traced—including across international borders—and to remove personally identifiable health information.

- **Reduce data silos and increase interoperability. Create a National Cancer Data Ecosystem** including use cases, SEER data groupings, NCI's National Clinical Trials Network (NCTN), and other sources. All components and data should be interoperable.

- **Take the Technology to the Data.** New analytical and computational tools make it easier for researchers to use new technology and create a **new paradigm: send technology tools to the data for local analysis** rather than asking researchers to send their data somewhere else for analysis. Networks of clinical trials operate on this model today.

- **Integrate functional imaging with molecular assays/data.** Molecular data is not routinely collected, but imaging data is ubiquitous. Incorporate datasets from both and provide funding for projects that integrate different types of data and methodologies (such as radiomics, genomics, and clinical data).

- **Understand more about the dissemination of tumor cells.** The spread of tumor cells is what leads to patient death; not the primary tumor in isolation.

Technological Opportunities and Challenges—rapid expansion of technology capabilities at different scales and in different realms, from data acquisition to clinical decision:

- **Computing and computational pathology provide a different way to view medical research.** Supporting research at scale with infrastructure is a challenge, but emerging information commons allow using more data. In biomedical research, 75% of future methods will come from technology or microtechnology. Explore new approaches to funding grants.

- **Computational pathology accelerates the diagnostic process.** Bioinformatics provides information about relationships. It is important to use the tools holistically, which is not easy to do; computational tools must be effective and agile. The field is moving from clinical decision support (CDS) algorithms to analytics as devices become increasingly capable to respond to data.

Bioinformatics tools/capabilities include:

- **Whole slide imaging (WSI) technologies—offer tremendous opportunity for both researchers and clinicians. Several key points are listed below.**

    - Using the new WSI technologies, pathologists can perform computational analytics in minutes instead of taking 10-12 hours to analyze.

    - Hundreds of thousands to millions of features can be seen in a whole slide tissue sample. In addition, spatial tumor heterogeneity can be seen in transmitted light images of tissue sections.

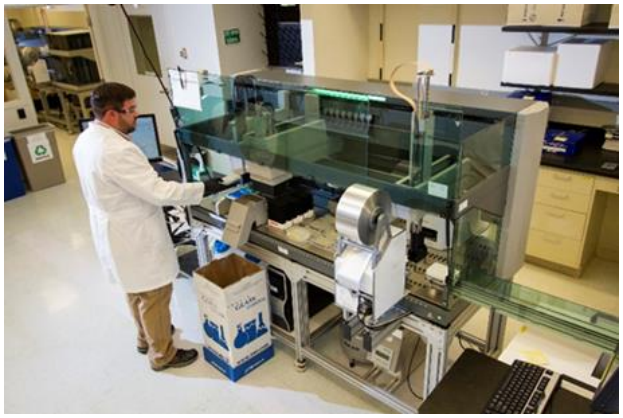
    - One WSI can be used to build:
        - Pathology-radiology correlations and molecular/anthropology correlations
        - Multi-dimension analysis for genotyping and phenotyping

    - Histology: today we can cut through an entire tumor (not just a section) and make tissue samples. This tells where in the tumor the important area is.

    - Protein multiplexing at a microanatomical detail is opening a new frontier of information available from slides.

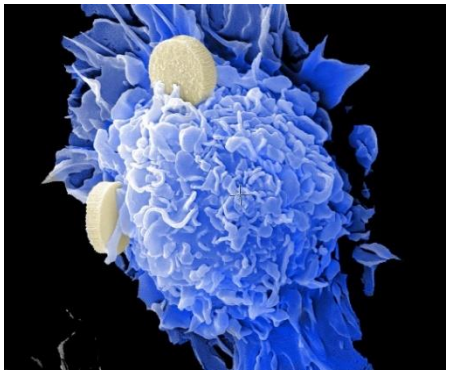
- **Next generation computing will be different.** Rapid technological changes are driving changes in computing technologies, applications, and point of deployment. New capabilities on the horizon include different memory and processing technologies, such as quantum computing and computing with neuromorphic elements built in.

- **Cancer is not one disease, but many diseases which respond very differently to treatment.** Many sub-types of cancer add to the complexity of the disease. There is neither one path nor one set of problems. It is important to target different abnormalities now.



- **In the future, molecular classification rather than anatomic will be more prevalent.** There is enormous opportunity to scale machine learning and leverage predictive models in oncology.

- **Biologic unknowns are a broad problem. Computational approaches such as machine learning and deep learning can help solve the biological unknowns.**



- **Uncertainty must be incorporated into predictive models.** These include systematic and measurement error, biological heterogeneity, and lack of a definition and measure of health. Current and future needs include:
    - The ability to build, analyze, and validate predictive models
    - Well-annotated and appropriate data sets

- **Machine learning is expanding the reach of new technologies hyper-exponentially.** Neural network-based diagnosis has been used for 10 years and has worked extremely well.
    - **Build a learning system to impact decision-making**, for example in biology, specifically for areas in which models are not currently prevalent. The knowledge and tools to build models is now readily accessible.
    - **Machine learning algorithms are much more efficient than they were before**. Machines can run continuously and predict when there will be problems with the code – and the machine can fix problems by itself. These capabilities will lead to automatic hypothesis generation and dynamic validation. This is very powerful for biology. Soon, a set of simulations can get the right answer.
    - **Adopt common conventions to facilitate the interchange of data-driven models.**

Organizational Opportunities and Challenges—cultural change within organizations and dynamics of multi-organizational efforts:

- **Team Science is Critical.** The prevalence and speed of changes in science, technology, computing, scale, and other factors demand inter-disciplinary collaboration.
    - **Use a team science approach to help connect research findings** to patient care.
    - **Report findings**, make more data available, and involve teams that include oncologists and pathologists. All involved must understand the impact on a patient.

- **Collaborations across disciplines are more important than ever for cancer research**. While data and technology are critical to making progress, one of the most vital challenges is the need for culture change within—and across—organizations. There is a need for organizations to transform from a data and intellectual property focus to a shared mindset community. In today's environment, sharing helps everyone, including those providing the data.

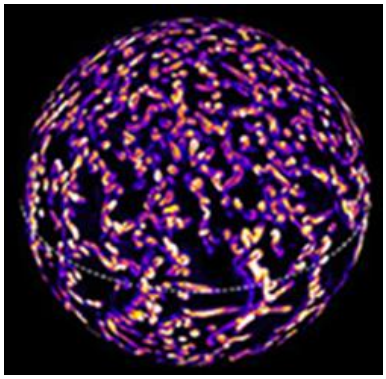
- **Interdisciplinary groups** of expert oncologists, researchers, technologists, pharma, and technology who are committed to making a difference for patients can engage in dialogue about needs, priorities, capabilities, costs, and best practices.

- **Interdisciplinary dialogues** can result in immediate progress on a tool, system, protocol or treatment that will affect patients.

- **Holding several interdisciplinary problem-solving sessions** will drive progress faster, better, and for less cost than any isolated effort.

- **The message to all involved is that the synergy of the group is the most expedient way to develop solutions for the problems they are focusing on.** Participation, at least in multi-disciplinary discussions, is no-risk with a high potential for gain.

- **New collaborative projects such as ATOM and the Apollo Network (APOLLO) and resources such as CANDLE and the VA's Million Veteran database[8] are good examples.** Each of these projects has an actionable vision and ambitious, impactful work that began with a multi-disciplinary discussion.



The DOE-NCI collaboration continues to make progress, publish results after one year of work, drive other initiatives, and help others—nationwide and across the globe—learn how to use the algorithms and other analytical tools being developed. ATOM is a direct offshoot of the DOE-NCI collaboration, Joint Design of Advanced Computing Solutions for Cancer (JDACS4C).

- **Increase visibility and expand collaborations.** Celebrities can serve as champions and patient advocacy groups can help generate interest in the new technologies, computing, methods, funding, research findings, treatment, and opportunities for collaboration**.**

- **Use predictive modeling to learn how to monitor therapy and diagnosis or manage the impact on entire populations.** The NCI Surveillance Research Program offers tools to collect information and lay the foundation for a scalable modeling framework in the research and clinical communities. These tools also can be used to predict general trajectory, which is a key area for patient care, but predicting tumor progression and heterogeneity is not currently part of standard clinical practice.

[8] The Department of Veterans Affairs' Million Veteran Program is the largest genomic database in the world. https://www.va.gov/opa/pressrel/pressrelease.cfm?id=2806

- **Expedite the path from research findings to patient care**. Physicians are unlikely to use new treatments unless the safety and efficacy are well established. FDA approved drugs could help patients today, but the infrastructure to test the treatments is too expensive to invest in:

    - **Report new insights on cancer biology** to the broader community
    - **Pay attention to the needs of those who will be using the new capabilities**
    - **Develop collaborations** among pharma, the FDA, and computing technology vendors
    - **Improve the ability to work with industry to expedite the use of new treatment for patients. NCI has several programs with industry.** The University of Pittsburgh is one model resource for sharing data and information (such as tissue samples, bioinformatics, and genomics data).
    - **Develop guidelines for physicians** about when to start using new markers and when to stop using the old ones**—and promote awareness of these guidelines** (through medical associations and other methods)

- **Major issues are uncertainties about what the technology can do,** why it is needed, and what effect it may have on stakeholders (such as balancing intellectual property concerns with the need to be used by patients as quickly as possible). In addition:

    - **Oncologists and technologists must come together to understand the needs and capabilities of their counterparts in other disciplines.**

    - **The oncology community may not have a picture of what is easy or difficult for technology to do** and technologists may not have enough insight into what oncologists and researchers need.

- **Understanding human motivation and behavior is key to adoption of any new tool, methodology or collaboration,** including a new paradigm that unites people from different disciplines and incentivizes collaboration plus data and information sharing.

    - **The message to pathologists and clinicians is that leveraging healthcare analytics in conjunction with clinical judgment is essential for patient care.**

    - No artificial intelligence tool or method can do what human beings can do, which is to ask why.

Breakout Sessions

Three breakout sessions focused on the following cutting-edge areas of predictive oncology:

Breakout I: **Informing Cancer Treatments with Computational Predictive Oncology**

Breakout II: **Predictive Oncology Algorithms and Software: Challenges, Opportunities, and Paths Forward**

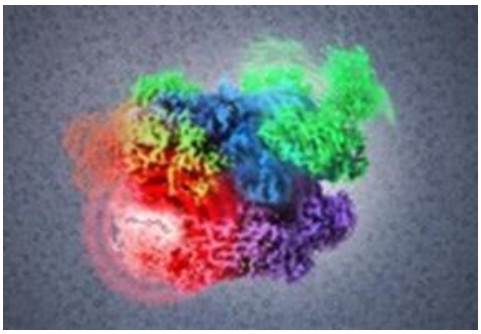
Breakout III: **Evolving Role of Pathology, Tissue, and Bio-specimen Data in Predictive Oncology and Analytics**

Participants in the breakouts discussed current resources, key opportunities, and challenges and roadblocks in their area of focus. Then each group identified compelling next steps that are expected to advance predictive oncology and accelerate patient impact. The groups presented their summaries and recommendations to the full group on the final day of the meeting. Highlights of the discussions appear below.

Breakout I

Informing Cancer Treatments with Computational Predictive Oncology

Discussion in this breakout prioritized a focus on patients throughout their journey of care, from screening, biopsy, and diagnosis to treatment and survivorship. Considerations include:



- **Complexity** of the disease remains an ongoing challenge for diagnosis, treatment, and survivorship. In addition, because cancer is biologically dynamic, there are millions of sub-types. As a result, there is neither one set of problems nor a universal approach to all cancers.

- **Data Collection and Availability**. There was broad agreement that capturing and curating data is still a big challenge. As well, there is a huge disparity in the amount of data for different types of cancer. New data sharing policies and practices are needed to help patients use and guide their own data. A National Cancer Data Ecosystem could be established that is inclusive, open, and interoperable. Cooperation across public research programs such as the NCI Informatics Technology for Cancer Research (ITCR), the NCI-DOE collaboration, Joint Design of Advanced Solutions for Cancer (JDACS4C), NCI Center for Bioinformatics and Information Technology (CBIIT), NCI's Quantitative Imaging Network (QIN), and others must continue to grow.

- **Data Analytics. Currently, researchers do not have good models to capture phenotyping, imaging, and other critical cancer data.** One issue is how to integrate quantitative imaging and functional imaging with molecular assays. Most patients get imaged, but not everyone has molecular assays. Additional funding is needed to determine how to integrate radiomics, pathomics and machine learning.

- **Registries and Infrastructure. Leveraging SEER and other registries may also help fill gaps**; adding digitized pathology images to SEER would expand the population base. Infrastructure could also be shared to extract and harmonize data across new projects.

- **Disparities in Investment, Scale of Data, and Level of Understanding.** There is a wide disparity across populations in financial investment as well as available data. Additional funding is needed to recruit minority populations for clinical trials.

Breakout II
Predictive Oncology Algorithms and Software: Challenges, Opportunities, and Paths Forward

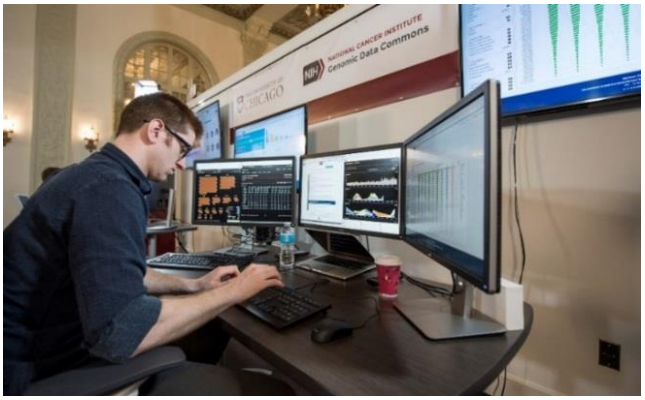
Discussion in this breakout session began with an assessment of what was new this year at FPOC II.

- **Algorithms, Marketing, Data, and Workflows.** There has been significant progress in the development of algorithms, marketing, data, and workflows. The group was particularly pleased that a problem identified last year at FPOC is being addressed; namely, the creation of a community that is finding different ways to share data and create a common approach to define and solve problems.

- **An investment in marketing to raise visibility of available tools and technologies could be very valuable to ensure extramural researchers are aware of these resources,** both as consumers and contributors.

- **Software and Computational Tools.** Current resources include the NCI Informatics Technology for Cancer Research (ITCR), the NCI Cancer Research Data Commons, the DOE CANcer Distributed Learning Environment (CANDLE), NCI's Quantitative Imaging Network (QIN), Dockstore (https://Dockstore.org) – genomic related containers, workflow containers, Openslide, The Cancer Imaging Archive (TCIA), NCI Cloud Resources, medical imaging resources, and others. Recommendations include maintaining an up-to-date curated list of available tools (which ITCR provides for its resources) and leveraging DOE non-cancer tools such as Computational Fluid Dynamics (CFD) and other scaled code from the national laboratories.

- **Machine Learning Tools, Frameworks, and Standards.** The group discussed several strategies to raise the level of confidence of machine learning-based tools. These include:
  - **Link tool development to clinical outcomes and bring tools into clinical workflows**
  - **Create a standards body** for quality assurance, scalability, and support for the tools
  - **Map hospital electronic medical records (EHR/EMR) data into a standard form** to facilitate input into CANDLE



- **Workflows in the Cloud.** The quality of available workflows is inconsistent and there may be low awareness of workflows already available in the cloud. Moving from desktop to cloud-based tools could facilitate greater consistency and confidence in computational analytics.

- **Synthetic Data.** To address data disparities, researchers are exploring the feasibility of synthetic data for machine learning. The task of benchmarking datasets is helpful in determining if it is possible to generate enough synthetic data to use for training.

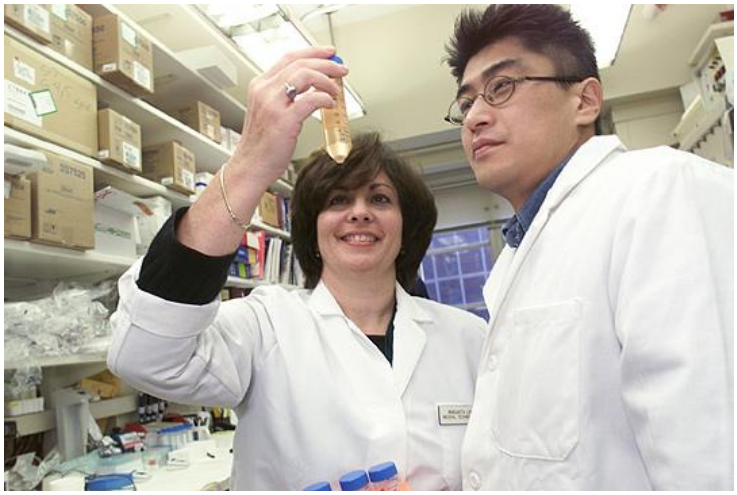
Breakout III
Evolving Role of Pathology, Tissue, and Biospecimen Data in Predictive Oncology and Analytics

Participants in this breakout group also discussed progress made since last year in three areas: 1) the use and computational analysis of data, 2) the development of new algorithms, and 3) the impact of the FPOC community's contributions to date on patient care. Key points include:

- **Data Quality and Accessibility remains the top issue limiting research and progress.** Work is underway within the public data registries to consolidate data and increase availability. The overarching goal is to get the right data in the right amounts to the right place at the right time.

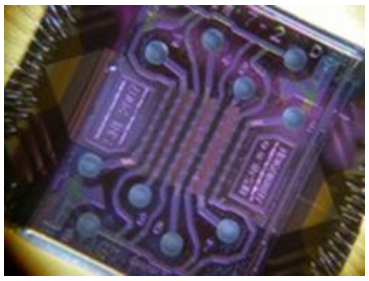
- **Within the private and academic healthcare systems, a vast amount of data is available, yet remains difficult to access universally.** Data in the private healthcare system(s) may be more accessible to researchers and clinicians than government data. Efforts are needed to raise visibility of available data for research.

- **Clinical Data, Standards, and Annotation.** Making biospecimen data easier to use would expedite advancements in computational pathology, discovery, and treatment for patients. Priority goals include development of common definitions within an interchange standard; standardization of data formatting, annotation, and collection; and development of extensible data annotation.



Section 5. Moving Forward

A brief retrospective of contrasts between the first and second annual Frontiers of Predictive Oncology and Computing is also valuable. In the first meeting held in July 2016 in Washington DC, the priority focus was on technology, interdisciplinary collaboration and the role of data. The common themes for recommendations at the first meeting included the following:



- Convening groups including patients and patient representatives together with life scientists and computational scientists to address issues including data access and quality; and

- Investing in cross-disciplinary efforts that will inform, train, and educate stakeholders on priority areas in cancer and new computational technologies ranging from analytics, imaging, computational modeling, and health monitoring.

Participants at the second meeting echoed these themes. In addition, as noted previously, the landscape of new oncology research findings and new technology tools has changed significantly since the first meeting. For example, there is a much broader, more in-depth use of machine learning and other artificial intelligence tools. These tools enable integration of molecular and imaging data analysis and powerful new open-source technology platforms, such as the CANcer Distributed Learning Environment (CANDLE), that are being used in cancer research across a wide range of challenging applications.

There was an impassioned sense of urgency to overcome the roadblocks and challenges that stand in the way of improved, early diagnosis and treatment of cancer patients.

Following the breakout sessions at the end of the meeting, strong momentum was established to act on the following, even prior to the next Frontiers of Predictive Oncology and Computing meeting:

- **Identify calls to action across the predictive oncology community; engage volunteer champions**
- **Expand the predictive oncology community by involving more oncologists**
- **Convene working sessions on an ongoing basis**
- **Determine how to leverage the new technology tools to accelerate development of computational models at an unprecedented scale**
- **Increase data sharing, access, and portability**
- **Coalesce and increase collaboration and participation across the entire multi-disciplinary community: academia, research, IT industry, pharma, clinical practice, and patients**

Participants in each breakout session emphasized the importance—and challenge—of improving the data available for predictive oncology, including the type, breadth, and suitability of the data for use in predictive oncology modeling. Also, within each breakout session, attendees highlighted their desire to work collaboratively across disciplines, bringing biologists, clinicians, and other scientists together with computational, data and technology experts. Participants agreed overwhelmingly that multidisciplinary, cross-organizational collaboration is essential for developing vital data resources and making those resources widely available.

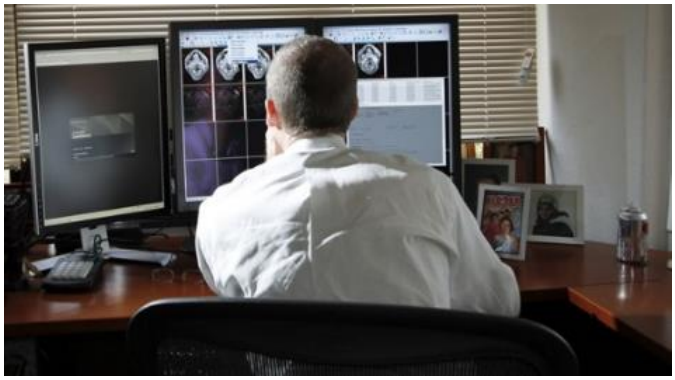
While the meeting brought forward many ideas, opportunities and challenges, the following provides a consolidated and prioritized perspective of key next steps to be pursued by the predictive oncology community, organized by the four focus areas of the meeting: Patient-centric, Organizational, Data and Technology.

Patient Engagement and Impact

- Define ways to share long-term data, create linkages to nonlocal data, and obtain outcome data
- Study potential incentives (e.g., financial) to better annotate and share data
- Convene a working meeting with oncologists, researchers, pharma, and vendors to understand what oncologists need and collectively develop a roadmap for the future
- Identify key drivers for development of a strategy to integrate applications and support development of a cohesive predictive oncology community
- Make computational diagnostic efforts transparent for patients throughout the lifecycle and increase patient involvement
- Motivate individual behavior to maintain compliance with treatment and improve health

Cultural Change Within Organizations and Multi-Organizational Efforts

- Pay attention to human factors (such as skepticism, cultural change, and other barriers to acceptance) in efforts to accelerate patient impact
- Conduct discipline-specific workshops
- Improve participation (in collaborations) from Electronic Medical Records (EMR) practitioners and vendors
- Establish teams with broad representation of pathologists and EMR vendors



Data Generation, Collection, Analysis, and Distribution

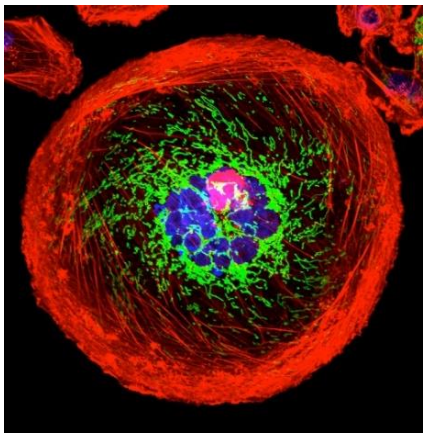
- Integrate molecular data with imaging
- Define resources and processes to access data across systems and improve access to private data sets
- Clarify and communicate tools such as The Cancer Imaging Archive (TCIA), NCI Cloud Resources, and the availability of various Data Commons (such as the NCI Cancer Research Data Commons)
- Share more data by reducing data silos and increasing interoperability
- Integrate different types of data and combine with computational resources to provide analytic and predictive insights in the complexities of cancer

Technology: Rapid Expansion, Development and Application

- Integrate radiomics, pathomics, and machine learning
- Improve the verification and validation of software and models
- Incorporate expertise in complex software systems and integrations

Future developments in technology will offer even greater analytical tools, provide additional, critical insight into cancer biology, and advance precision diagnosis, treatment, and survival.

Due to advances in computational pathology, and the documented impact to patients at leading universities and cancer centers, the future holds exceptional potential. As multi-disciplinary, cross-organizational collaborations develop further—with the inclusion of cancer patients and their oncologists—extraordinary new discoveries are on the horizon.



Appendix – Participant List – Alphabetical by last name

Francis Alexander
Brookhaven National Laboratory/Department of Energy

Jonas Almeida
Stony Brook University

John Baldoni
GlaxoSmithKline

Muthu Baskaran
Reservoir Labs

Michael Becich
University of Pittsburgh Medical Center Hillman Cancer Center

Aviv Bergman
Albert Einstein College of Medicine

Tanmoy Bhattacharya
Los Alamos National Laboratory

Lynn Borkon
Frederick National Laboratory for Cancer Research

Thomas Brettin
Argonne National Laboratory

Ashley Bucholz
Oak Ridge National Laboratory

Rachael Callcut
University of California San Francisco

S. Chakra Chennubhotla
University of Pittsburgh Medical Center

Yoon Choi
Blueprint Medicines

John Condeelis
Albert Einstein Cancer Center

Blair Christian
Oak Ridge National Laboratory

Claudine Conway
Intel Corporation

Carlos Cordon-Cardo
Icahn School of Medicine at Mount Sinai

Ana Paula de Oliveira Sales
Lawrence Livermore National Laboratory

Synho Do
Harvard Medical School

Paul Dotson
Los Alamos National Laboratory

Eric Durbin
University of Kentucky

Janet Eary
National Cancer Institute

Yvonne Evrard
Frederick National Laboratory for Cancer Research

Paul Fearn
National Cancer Institute

Michael Fonstein
Argonne National Laboratory

Steve Friedman
National Cancer Institute

Walt Gall
Saffron Technology

Mike Gann
Intel Corporation

John Gilbertson
Harvard Medical School

Fiona Ginty
General Electric Global Research Center

James Glosli
Lawrence Livermore National Laboratory

Gnana Gnanakaran
Los Alamos National Laboratory

Emily Greenspan
National Cancer Institute

Amy Gryshuk
Lawrence Livermore National Laboratory

Rajarsi (Raj) Gupta
Stony Brook University Medical Center

Scott Hammond
University of California San Francisco

Sean Hanlon
National Cancer Institute

Stephanie Harmon
National Cancer Institute

Robert Harrison
Stony Brook University/Brookhaven National Laboratory

David Heimbrook
Frederick National Laboratory for Cancer Research

Debra Hope
Frederick National Laboratory for Cancer Research

Elizabeth Hsu
National Cancer Institute

Heng Huang
University of Pittsburgh

Kun Huang
Indiana University

Michael Idelchik
General Electric

Stephen Jett
National Cancer Institute

Shantenu Jha
Rutgers University

Warren Kibbe
Duke University School of Medicine

Miles Kimbrough
NetImpact Strategies

Patricia Kovatch
Mount Sinai School of Medicine

Tahsin Kurc
Stony Brook University

Dimitri Kusnezov
National Nuclear Security Administration/Department of Energy

Maryl Lambros
Albert Einstein College of Medicine

Jerry S.H. Lee
National Cancer Institute

Quanzheng Li
Harvard Medical School

Michael Liebman
IPQ Analytics, LLC

Felice Lightstone
Lawrence Livermore National Laboratory

Dwight Nissley
Frederick National Laboratory for Cancer Research

Maja Oktay
Albert Einstein Cancer Center

Diane Palmieri
National Cancer Institute

Arvind Ramanathan
Oak Ridge National Laboratory

Joel Saltz
Stony Brook University Cancer Center

Mary Saltz
Stony Brook University Cancer Center

Romeil Sandhu
Stony Brook University

Brion Sarachan
General Electric

Mark Seager
Intel Corporation

Ashish Sharma
Emory University School of Medicine

Maulik Shukla
Argonne National Laboratory

Amber Simpson
Memorial Sloan Kettering Cancer Center

Eric Stahlberg
Frederick National Laboratory for Cancer Research

Rick Stevens
Argonne National Laboratory

Frederick Streitz
Lawrence Livermore National Laboratory

Vesteinn Thorsson
Institute for Systems Biology

Georgia Tourassi
Oak Ridge National Laboratory

Nadejda Tsankova
Mount Sinai Hospital

Eric Tucker
General Electric Global Research Center

Daifeng Wang
Stony Brook University

Fusheng Wang
Stony Brook University

Dawn Whalen
Lawrence Livermore National Laboratory

Fangfang Xia
Argonne National Laboratory

George Zaki
Frederick National Laboratory for Cancer Research

Meeting Agenda

Frontiers of Predictive Oncology and Computing II

From Pathology to Computation – The Path to Dynamic Models for Cancer

October 17-19, 2017
SUNY Global Center, New York, NY
Limited Capacity: Participation by Invitation Only

Primary Goals for the Meeting

- Bring together experts from industry, government, and academia working across the combined frontiers of pathology, radiology (multi-scale imaging), predictive oncology and computing
- Provide insight into existing challenges and efforts to address challenges where multi-scale imaging, predictive oncology and computing share common opportunities
- Provide opportunities to share in discussion of new opportunities arising from new ideas for collaborations, cross-disciplinary education, and shared efforts to accelerate cancer research and clinical application of research advances
- Bring focus to the role of "computational pathology" across multiple time and length scales and areas of application ranging from digital pathology to opportunities in drug discovery and integrated multiscale modeling
- Share future visions from multiple perspectives to develop a common appreciation for the integrated role domain knowledge, technology, and information will play in the future for computationally predictive oncology

Meeting History

The Frontiers of Predictive Oncology and Computing meeting is an annual event tying its origins to the original Biological Applications of Advanced Strategic Computing meetings initiated by Lawrence Livermore National Laboratory. Bringing a specific focus to the challenges and opportunities for cancer, the first Frontiers of Predictive Oncology and Computing meeting was held July 2016 in Washington DC. At this meeting over 100 thought leaders from industry, government and academia converged to share insights, knowledge and vision for the future of computationally predictive oncology. The overview of this meeting is available online at http://www.cvent.com/d/05qn91.

This Year's Meeting

- The second Frontiers of Predictive Oncology and Computing meeting brings focus to the topic of "computational pathology", discussing the broader application of technology, computation and domain expertise to understand and describe the specifics of cancer as a disease. With origins in digital pathology, extended in recent years to include molecular level signatures through sequencing and other forms of enhanced observation, the concept of "computational pathology" embraces the dynamic range of options from virtual microscopy to molecular to probe cancer and capture observations of disease behaviors across space and time scales. The Frontiers of Predictive Oncology and Computing meeting brings context to these methods of observation, providing insight into the key role the collected information plays in the development of computationally predictive oncology models and methods.

Specific topical areas to be discussed include:

- Longitudinal multi-modal data in predictive oncology - Pre-diagnosis, detection, and post-diagnosis monitoring
- Multiscale data in predictive oncology – From molecular, cellular, and tumor, to organ, tissue, body, and population
- Clinical and commercial applications - Predictive oncology applied (metastasis, treatment decisions, treatment development, etc.)
- Computational frontiers - HPC, sensors, edge computing

Frontiers of Predictive Oncology and Computing

Biological Applications of Advanced Strategic Computing (BAASiC) Meeting

October 17-19, 2017

SUNY Global Center
116 E 55th St, New York, NY 10022
www.suny.edu/about/nyc

Day One – Tuesday, October 17, 2017

8:00 AM **Arrival and check-in** at the SUNY Global Reception Center to receive badge

Registration and continental breakfast Global Classroom

9:00 AM **Welcome & Introductory Remarks** Global Classroom

Emily Greenspan, PhD
Program Director, Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute (NCI)

Dimitri Kusnezov, PhD
Chief Scientist & Senior Advisor to the Secretary, National Nuclear Security Administration (NNSA), Department of Energy (DOE)

Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Robert Harrison, PhD
Professor in Applied Mathematics and Statistics, Director of the Institute for Advanced Computational Science, State University of New York (SUNY) Stony Brook University Director, Computational Science Center and New York Center for Computational Sciences, Brookhaven National Laboratory

---

9:15 AM **Meeting Overview – Computational Pathology and Predictive Oncology**

Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

---

9:30 AM **Keynote – Cancer Moonshot – One Year Later**

Jerry S.H. Lee, PhD
Deputy Director, Center for Strategic Scientific Initiative, National Cancer Institute (NCI)
Deputy Director, for Cancer Research and Technology, Cancer Moonshot Task Force

---

10:15 AM ***Break – networking***

10:30 AM **Plenary Session – Drivers for Predictive Oncology Impacting Computational Pathology – Patients, Treatments, and Improving Outcomes**

Moderators: Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Janet Eary, MD
Deputy Associate Director, Cancer Imaging Program, National Cancer Institute (NCI)

John Baldoni, PhD
Senior Vice President of Platform Technology and Science, GlaxoSmithKline

Kun Huang, PhD
Assistant Dean for Data Sciences, Professor of Genomics Data Sciences, Professor of Medicine, Indiana University

---

12:00 PM ***Lunch*** Global Classroom

1:30 PM **Plenary Session – Frontier Technologies to Probe Biology – Unlocking Frontiers of Computational Pathology**

Moderator: Robert Harrison, PhD

Professor in Applied Mathematics and Statistics, Director of the Institute for Advanced Computational Science, State University of New York (SUNY) Stony Brook University, Director, Computational Science Center and New York Center for Computational Sciences, Brookhaven National Laboratory

Fiona Ginty, PhD

Biosciences Technical Operations Leader & Principal Investigator, GE Global Research Center

John Condeelis, PhD

Professor & Co-Chair of Anatomy & Structural Biology, The Judith and Burton P. Resnick Chair in Translational Research, Co-Director, Gruss Lipper Biophotonics Center, Co-Director, Integrated Imaging Program, Director, Tumor Microenvironment and Metastasis Program, Scientific Director, Analytical Imaging Facility, Albert Einstein Cancer Center

Maja Oktay, MD, PhD

Professor, Department of Pathology, Department of Anatomy & Structural Biology, Albert Einstein Cancer Center

Vesteinn Thorsson, PhD

Senior Research Scientist, Institute for Systems Biology

3:00 PM ***Break – networking***

3:30 PM **Panel Session – Exploring the Frontiers of Computing and the Future of Computational Pathology**

Moderator: Mark Seager, PhD

Intel Fellow, Chief Technology Officer for the High-Performance Computing (HPC) Ecosystem, Intel Corporation

Tahsin Kurc, PhD

Vice Chair and Research Associate Professor, Department of Biomedical Informatics, Stonybrook University

Scott Hammond, MD
*Strategic Director Outlier Initiative/**S**marter**Heal**th*
CDHI Expert in Residence, University of California San Francisco (UCSF)

Fred Streitz, PhD
Chief Computational Scientist, Physical and Life Sciences Directorate, Director, High Performance Computing Innovation Center (HPCIC), Lawrence Livermore National Laboratory

Warren Kibbe, PhD
Chief of Translational Biomedical Informatics, Department of Biostatistics and Bioinformatics, Chief Data Officer, Duke University School of Medicine

---

5:00 PM **Adjourn - Social Networking Opportunity**

Evening on own – dinner on own

Day Two – Wednesday, October 18, 2017

8:00 AM **Arrival and check-in** at the SUNY Global Reception Center to receive badge

Registration and continental breakfast Global Classroom

8:45 AM **Welcome & Recap** Global Classroom

Emily Greenspan, PhD
Program Director, Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute (NCI)

Dimitri Kusnezov, PhD
Chief Scientist & Senior Advisor to the Secretary, National Nuclear Security Administration (NNSA), Department of Energy (DOE)

Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Robert Harrison, PhD
Professor in Applied Mathematics and Statistics, Director of the Institute for Advanced Computational Science, State University of New York (SUNY) Stony Brook University Director, Computational Science Center and New York Center for Computational Sciences, Brookhaven National Laboratory

---

9:00 AM **Keynote – Towards a Digital Pathology Commons**

Michael Becich, MD, PhD
Associate Vice-Chancellor for Informatics in the Health Sciences, Chairman and Distinguished University Professor, Department of Biomedical Informatics, Associate Director, University of Pittsburgh Medical Center (UPMC) Hillman Cancer Center

---

10:00 AM ***Break – networking***

10:15 AM **Plenary Session – Joint Design of Advanced Computing Solutions for Cancer (JDACS4C): Frontier Collaborations in Predictive Oncology and Computing**

Moderators: Amy Gryshuk, PhD
Director, Strategic Engagements & Alliance Management, Physical & Life Sciences Directorate (PLS), Biosciences & Biotechnology Division (BBTD), Lawrence Livermore National Laboratory

Eric Stahlberg, PhD
Director, Strategic and Data Science Initiatives, Biomedical Informatics and Data Science Directorate, Frederick National Laboratory for Cancer Research

Molecular Scale Predictive Oncology

Dwight Nissley, PhD
Director, Cancer Research Technology Program, Frederick National Laboratory For Cancer Research

Fred Streitz, PhD
Chief Computational Scientist, Physical and Life Sciences Directorate, Director, High Performance Computing Innovation Center (HPCIC), Lawrence Livermore National Laboratory

Pre-clinical Scale Predictive Oncology

Yvonne Evrard, PhD
Operations Manager, NCI Patient-Derived Models Repository
Frederick National Laboratory for Cancer Research

Rick Stevens
Associate Laboratory Director, Argonne National Laboratory

Population Scale Predictive Oncology

Paul Fearn, PhD
Chief, Division of Cancer Control and Population Sciences, Surveillance Informatics Branch, National Cancer Institute (NCI)

Georgia Tourassi, PhD
Director, Health Data Sciences Institute, Oak Ridge National Laboratory

---

12:00 PM ***Lunch***

1:00 PM **Computing Frontiers: JDACS4C Cross-cutting Technologies**

Uncertainty Quantification

Tanmoy Bhattacharya, PhD
External Professor and Scientist, Los Alamos National Laboratory

CANDLE – CANcer Distributed Learning Environment

Rick Stevens
Associate Laboratory Director, Argonne National Laboratory

1:45 PM ***Break***

2:00 PM **Panel Session – Longitudinal and Multiscale Data: Challenges and Opportunities for Computational Pathology**

Moderators: Rachael Calcutt, MD, MSPH
Associate Professor of Surgery Trauma, Critical Care & General Surgery, Zuckerberg San Francisco General Hospital, University of California San Francisco (UCSF)
Director of Data Science, UCSF Center for Digital Health
Program Director, UCSF SmarterHealth Artificial Intelligence Initiative

Scott Hammond, MD
Strategic Director Outlier Initiative/SmarterHealth
CDHI Expert in Residence, University of California San Francisco (UCSF)

Carlos Cordon-Cardo, MD, PhD
Professor and System Chair, Pathology, Professor, Genetics and Genomic Sciences, Oncological Sciences, Mount Sinai School of Medicine

Chakra Chennubhotla, PhD
Associate Professor, Department of Computational and Systems Biology, University of Pittsburgh Medical Center

John Gilbertson, MD
Associate Professor, Harvard Medical School
Associate Chief of Informatics, Director of Pathology Informatics, Massachusetts General Hospital

---

3:00 PM ***Break – networking (Global Classroom to be reset for Breakout Sessions)***

3:15 PM **Breakout Sessions**

	Session Name	Room Location
Session I	Informing Cancer Treatments with Computational Predictive Oncology	Global Classroom, Side 1
Session II	Predictive Oncology Algorithms and Software – Challenges, Opportunities and Paths Forward	Global Classroom, Side 2
Session III	Evolving Role of Pathology, Tissue and Biospecimen Data in Predictive Oncology and Analytics	Multipurpose Room – 2nd Floor

4:45 PM ***Break – networking***

5:00 PM **Adjourn**

Social Event

7:00 PM **Meeting Dinner, *Angus Club Steakhouse,* 135 E 55 Street, Manhattan, NY 10022**
**Reservation under 'FPOC'. Business casual attire. Open seating format. Cash bar.*

Day Three – Thursday, October 19, 2017

8:00 AM **Arrival and check-in** at the SUNY Global Reception Center to receive badge

Registration and continental breakfast Global Classroom

8:45 AM **Welcome & Introductory Remarks** Global Classroom

Emily Greenspan, PhD
Program Director, Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute (NCI)

Dimitri Kusnezov, PhD
Chief Scientist & Senior Advisor to the Secretary, National Nuclear Security Administration (NNSA), Department of Energy (DOE)

Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Robert Harrison, PhD
Professor in Applied Mathematics and Statistics, Director of the Institute for Advanced Computational Science, State University of New York (SUNY) Stony Brook University
Director, Computational Science Center and New York Center for Computational Sciences, Brookhaven National Laboratory

---

9:00 AM **Keynote – Learning from Industry Challenges in Multiscale Analytics and Relevance to Cancer Research and Imaging**

Michael Idelchik
Vice President, Advanced Technology Programs, General Electric

---

9:30 AM ***Break – networking***

10:00 AM **Individual Breakout Session Conclusion and Preparation**

10:30 AM **Facilitated Discussion on Breakouts**

Moderator: Mike Gann
Director, Global Healthcare, Intel Corporation

11:30 AM **Next Steps and Meeting Wrap-up**

Emily Greenspan, PhD
Program Director, Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute (NCI)

Dimitri Kusnezov, PhD
Chief Scientist & Senior Advisor to the Secretary, National Nuclear Security Administration (NNSA), Department of Energy (DOE)

Joel Saltz, MD, PhD
Cherith Professor and Founding Chair, Department of Biomedical Informatics, Vice President for Clinical Informatics, Stony Brook Medicine, Associate Director, Stony Brook University Cancer Center

Robert Harrison, PhD

Professor in Applied Mathematics and Statistics, Director of the Institute for Advanced Computational Science, State University of New York (SUNY) Stony Brook University
Director, Computational Science Center and New York Center for Computational Sciences, Brookhaven National Laboratory

---

12:00 PM **Meeting Adjournment**

Image Credits

Cover Page

Davis, Charles Patrick. *Understanding Cancer: Metastasis, Stages of Cancer, and More*. Digital Image. *Medicinenet.com.* Jul. 2016. Web. https://www.medicinenet.com/cancer_101_pictures_slideshow/article.htm

Page 3

Gardner, Kevin. *Model of Small Molecule Drug*. Digital image. *National Cancer Institute Visuals Online*. National Cancer Institute, 2013. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10156

Page 5

Krawczyk, Ewa. *HPV-16 E5 Oncoprotein.* Digital image. *National Cancer Institute Visuals Online*. National Cancer Institute, Apr. 2008. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10626

Page 7

Stevens, Rick and Yvonne Evrard. *Pilot 1: Predictive Modeling for Pre-Clinical Screening.* Digital Image. *Frontiers of Predictive Oncology and Computing II Meeting*, New York, Oct. 2017. https://wiki.nci.nih.gov/display/HPC/FPOC+II+Presentations?preview=/356521732/357696737/FPOC%20II%20Day%202%20Plenary%20-%20Evrard%2C%20Stevens.pdf

Page 9

Deerinck, Tom. *HeLa Cells.* Digital image. *National Cancer Institute Visuals Online*. National Cancer Institute, Sept. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11870

Page 11

Wyatt, Emily and Mark Davis. *Nanoparticles in Brain Metastases.* Digital image. *National Cancer Institute Visuals Online*. National Cancer Institute, Jul. 2016. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11170

Page 12

Clevenger, Charles and Katherine Harrington. *Focal Adhesions in Breast Cancer*. Digital image. *National Cancer Institute Visuals Online*. National Cancer Institute, Dec. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10506

Page 13

Cancer Moonshot Implementation. Digital image. *Cancer Moonshot*. National Cancer Institute, Oct. 2017. Web. https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative

Page 14

Bhang, Hyo-eun and Martin Pomper. *Novel Multimodality Imaging Approaches to Target Metastatic Cancers*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Sept. 2012. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10885

Page 15

Lee, Steve. *Tumor Hypoxia*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Aug. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10586

Page 16

Deerinck, Tom. *HeLa Cells*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Apr. 2011. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11867

Page 17

Qian, Wei. *Treating Triple-Negative Breast Cancer*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, May 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10605

Page 18

Lee, Steve Seung-Young. *Spatial Heterogeneity in the Tumor Microenvironment*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jul. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10584

Study Uncovers Previously Unrecognized Effect of Chemotherapy. Digital image. *Cancer Currents Blog*. National Cancer Institute, Aug. 2017. Web. https://www.cancer.gov/news-events/cancer-currents-blog/2017/chemotherapy-effect-metastasis

Page 19

Szulczewski, Joseph, David Inman, Devin Eliceiri, and Patricia Keely. *Breast Tumor Microenvironment.* Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jan. 2016. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10573

Page 20

Joint Design of Advanced Computing Solutions for Cancer. Digital image. *DOE-NCI Pilots Presentation at the Frederick National Laboratory Advisory Committee (FNLAC)*. National Cancer Institute, Jun. 2016. Web. http://www.slideshare.net/WarrenKibbe/doenci-pilots-presentation-at-the-frederick-national-laboratory-advisory-committee-fnlac

Page 21

KRAS Protein Structure. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Nov. 2014. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11166

Page 22

Machine-learning curve. Digital image. *ASCR Discovery*. US Department of Energy, Office of Science, Nov. 2016. Web. http://ascr-discovery.science.doe.gov/2016/11/machine-learning-curve/

Page 23

Russell, John. *Deep Learning Thrives in Cancer Moonshot.* Digital Image. *HPCWire*, Aug. 2017. Web. https://www.hpcwire.com/2017/08/08/deep-learning-thrives-cancer-moonshot/

Page 24

Bliss, Donald and Sriram Subramaniam. *HIV Infected and Uninfected Immune Cells Interact*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Sept. 2014. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10401

Page 25

The NIH Clinical Center Data Center: Then and Now. Digital image. *Department of Clinical Research Informatics CIO Newsletter*, Oct. 2015. Web. https://cris.cc.nih.gov/cionews/newsletter/2015/oct2015/newsletter.htm

Page 26

Mohammad, Khalid and Theresa Guise. *Cancer Spreading to the Bone*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jan. 2016. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10583

Page 27

Sone, Daniel. *Doctors and Patients*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jun. 2013. Web. https://visualsonline.cancer.gov/details.cfm?imageid=9524

Page 28

Wahl, Geoffrey and Christopher Dravis. *Mobile Cancer Factories*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Sept. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10556

Page 29

Zhou, Hongyu and Lily Yang. *Nanoparticles in Pancreatic Tumor*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jan. 2017. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11238

Page 30

Sone, Daniel. *DNA Genotyping and Sequencing*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Aug. 2016. Web. https://visualsonline.cancer.gov/details.cfm?imageid=11172

Page 31

Melendez, Brenda and Rita Serda. *Developing Novel Vaccine Delivery Systems for Cancer Therapy*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Aug. 2012. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10884

Page 32

Introduction to Transformative Research. Digital image. *National Science Foundation Transformative Research*. Web. https://www.nsf.gov/about/transformative_research/

Page 33

Sone, Daniel. *Laboratory Researcher*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Sept. 2014. Web. https://visualsonline.cancer.gov/details.cfm?imageid=9787

Page 34

Falconieri, Veronica and Sriram Subramaniam. *Cryo-Electron Microscopy.* Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Dec. 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10547

Page 35

Kozloff, Robert. *Genomic Data Commons.* Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Jun. 2016. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10838

Page 36

Branson, Bill. *Scientists Examine Vial.* Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Dec. 2000. Web. https://visualsonline.cancer.gov/details.cfm?imageid=3711

Page 37

Osterfeld, Sebastian and Shan Wang. *Multiplexed Blood Protein Profiling Via Magneto-nano Chip In Vitro Diagnostics.* Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Feb. 2008. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10847

Page 38

Sone, Daniel. *Reviewing Brain Scans*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, Mar. 2010. Web. https://visualsonline.cancer.gov/details.cfm?imageid=8813

Page 39

Qian, Wei. *Hurricane in a Cell*. Digital image. N*ational Cancer Institute Visuals Online*. National Cancer Institute, May 2015. Web. https://visualsonline.cancer.gov/details.cfm?imageid=10606

END OF REPORT