Program

Sunday, November 13, 2021, 8:30 a.m.–12:00 noon Central Time

(All times listed are Central Time)

8:30 a.m.–8:35 a.m.	Welcome — Eighth Computational Approaches for Cancer Workshop (CAFCW22) Eric Stahlberg, PhD, Frederick National Laboratory for Cancer Research
8:35 a.m.–9:15 a.m.	Featured Speaker: Amber Simpson, Queen's University *AI for Generating Real-World Evidence in Cancer* Introduced by Eric Stahlberg, PhD, Frederick National Laboratory for Cancer Research
9:15 a.m. –9:30 a.m.	*Mathematical Discovery and Computational Validation of Two Orthogonal Mechanistically-Driven Whole-Genome Genotype–Survival Phenotype Relationships in Pediatric Neuroblastoma Nerve Cancer* Presenter: Orly Alter, PhD, University of Utah Authors: Orly Alter, PhD, University of Utah and Sri Priya Ponnapalli, PhD, Google LLC Abstract: Prediction, together with understanding and management, of pediatric neuroblastoma (NBL) outcomes, from spontaneous regression to relapse and death, remain limited, and rely mostly on age, stage, and

the one-gene test for MYCN amplification, none of which are NBL specific. Here, we use the generalized singular value decomposition (GSVD), formulated as a multi-tensor decomposition [1], to model whole genomes of patient-matched NBL and blood DNA. The GSVD discovers two orthogonal genome-wide patterns of copy-number alterations (CNAs) in the tumors that are correlated with survival. First, as in previous, experimentally validated, models of, e.g., adult brain astrocytoma [2], one pattern is exclusive to the tumors. Previously unseen is a pattern that is common to both the blood and tumor genomes. Second, both patterns predict survival better than and independent of the existing predictors as well as independent of each other. In both patterns, differential RNA expression consistently map to the DNA CNAs. Third, the GSVD separates these patterns from normal variations that are conserved in the tumors but do not predict outcome, e.g., the male-specific X-chromosome deletion relative to the autosome. We computationally validate both patterns by using – and demonstrating for the first time – the pseudoinverse projection for transfer learning from the ≈3M-bin whole-genome to ≈10K-bin target-capture sequencing profiles of a mutually-exclusive set of patients [3]. We show that the two patterns describe independent, yet complementary cellular mechanisms that transform human normal to tumor cells, predict new personalized therapies, and may predict the response to existing therapies. The tumor-exclusive pattern includes co-occurrence of MYCN amplification with previously unrecognized druggable CNAs, including amplifications of genes encoding for extra-embryonic transcripts, to jointly predict survival. The pattern that is common to the blood and tumor genomes describes an earlier stage in NBL development, where the embryonic program is hijacked toward aneuploidy and where the subsequent tumor development can spontaneously regress via embryonic self-correction.

[1] M. W. Bradley, K. A. Aiello, S. P. Ponnapalli,* H. A. Hanson* and O. Alter, "GSVD- and Tensor GSVD-Uncovered Patterns of DNA Copy-Number Alterations Predict Adenocarcinomas Survival in General and in Response to Platinum," Applied Physics Letters (APL) Bioengineering 3 (3), article 036104 (August 2019); https://doi.org/10.1063/1.5099268

[2] S. P. Ponnapalli, M. W. Bradley, K. Devine, J. Bowen, S. E. Coppens, K. M. Leraas, B. A. Milash, F. Li, H. Luo, S. Qiu, K. Wu, H. Yang, C. T. Wittwer, C. A. Palmer, R. L. Jensen, J. M. Gastier-Foster, H. A. Hanson, J. S. Barnholtz-Sloan and O. Alter, "Retrospective Clinical Trial Experimentally Validates Glioblastoma Genome-Wide Pattern of DNA Copy-Number Alterations Predictor of Survival," Applied Physics Letters (APL) Bioengineering 4 (2), article 026106 (May 2020); https://doi.org/10.1063/1.5142559

[3] O. Alter and G. H. Golub, "Integrative Analysis of Genome-Scale Data by Using Pseudoinverse Projection Predicts Novel Correlation between DNA Replication and RNA Transcription," Proceedings of the National Academy of Sciences (PNAS) USA 101 (47), pp. 16577–16582 (November 2004); https://doi.org/10.1073/pnas.0406767101

9:30 a.m. –9:45 a.m. *Long Document Transformers for Pathology Report Classification*

Presenter: Mayanka Chandrashekar, PhD, Oak Ridge National Laboratory

Authors: Mayanka Chandrashekar, PhD; Isaac Lyngaas, PhD; Shang Gao, PhD; Heidi Hanson, PhD; John Gounley, PhD, Oak Ridge National

Laboratory

Abstract: In recent years, deep learning-based models for electronic health records have shown impressive results in many clinical tasks. Deep learning classification models typically require large labeled training datasets and are designed to address specific clinical tasks. Transformers are powerful state-of-art language models designed to learn inherent patterns in unstructured text data in an unsupervised manner. The transformer model's unsupervised training enables generalizability and reusability of the model to various clinical tasks, negating the need for labeled data in the training phase. The trained transformer can then be fine-tuned towards a specific clinical task using a small but task-curated training dataset. In the current work, we build a transformer model that can effectively accommodate the length of typical cancer pathology reports. We use 5.7 million pathology reports from six Surveillance, Epidemiology, and End Results Program's (SEER) cancer registries to train "from scratch" the Big-Bird model. Big-Bird model is a transformer model built for long documents (up to 4096 tokens) compared to popular models such as BERT (up to 512 tokens). As the memory requirement of a transformer model scales quadratically with the sequence length of input text, Big-Bird utilizes sparse attention. In phase one, Big-Bird is built in an unsupervised manner using the pre-training task called masked language prediction. This phase requires the largest amount of computation, and it leverages the secure CITADEL capability for working with protected health information (PHI) data on the Summit supercomputer at the Oak Ridge Leadership Computing Facility. In phase two, we fine-tune the pre-trained Big-Bird model to handle five information extraction tasks: site, sub-site, histology, laterality, and behavior. For fine-tuning, we use data from six SEER registry data with the 10-day window constraint before and after the date of cancer diagnosis, and the ground truth for five tasks is from the manually coded CTC (Cancer/Tumor/Case) report. One advantage of this two-phase approach is the re-usability of the phase one model for any pathology-relevant clinical task in phase two. Our results show that the proposed Big-Bird model fine-tuned with SEER data on five information tasks outperforms the current state-of-the-art deep learning classification model by an average of 2% microF1 score on all tasks and an average 8% macro F1 score on all tasks. In most challenging tasks, subsite has a 4% increase in micro F1 score and histology has a 25% increase in macro F1 score. The results demonstrate the promise of using a single pretrained model on five related clinical tasks. We plan to further test the generalizability and reusability of the model by extending the tasks to other clinically useful tasks such as bio-marker extraction and identification of malignant and metastatic disease.

9:45 a.m. –10:00 a.m. *A Generalized Tumor Segmentation Algorithm for Varying Breast Cancer Subtypes*

Presenter: Imon Banerjee, PhD, Mayo Clinic

Abstract: Background. Automated breast tumor segmentation for dynamic contrast-enhanced magnetic resonance (DCE-MR) is a crucial step to advance and help with the implementation of radiomics for image-based, quantitative assessment of breast tumors and cancer phenotyping. Current studies focus on developing tumor segmentation, which often requires initial seed points from expert radiologists or atlas-based segmentation methods. We develop a robust, fully automated end-to-end segmentation pipeline for breast cancers on bilateral breast MR studies.

Methods. On IRB-approved diverse breast cancer MR cases, a deep learning segmentation algorithm was created and trained. The model's backbone is UNet++, which consists of U-Nets of varying depths whose decoders are densely connected at the same resolution via the skip connections and all the constituent UNets are trained simultaneously to learn a shared image representation. This design not only improves the overall segmentation performance, but also enables model pruning during the inference time. The model was trained on the breast tumors located independently by a radiologist with consensus review by a second radiologist with at least five years of experience. MRI was performed using a 3.0-T imaging system in the prone position with a dedicated 16-channel breast coil and T1 weighted DEC-MR images were analyzed for the study. We used 80:20 random split for training and validation of the model.

Results. A total of 124 breast cancer patients had pre-treatment MR imaging before the start of NST - the cohort comprised 49 HR+HER2-, 37 HR+HER2+, 11 HR-HER2+, and 27 TNBC cases (mean tumor 2.3 cm (+/- 3.1mm).) The model was tested on 2571 individual images. Overall, the model scored 0.85 [0.84 – 0.86, 95% CI] dice score and 0.8[0.79-0.81, 95% CI] IoU score. TNBC tumors scored dice [0.88 – 0.89, 95% CI], HER2 neg and ER/PR positive dice [0.84-0.85, 95% CI] and HER2 positive dice [0.84-0.85, 95% CI]. We observed that model performed equally for the solid tumors and irregular shapes and didn't observe any difference in the segmentation performance between residual and non-residual tumor types - dice score [0.85 – 0.86, 95% CI] and [0.83 – 0.84, 95% CI] respectively.

Conclusion. The proposed segmentation model can perform equally well on various clinical breast cancer subtypes. The model has high false positive rate towards biopsy clip and high background enhancement, which we plan to solve by adding annotation of the clip and high non-cancer enhancement in future training data. We will release the trained

model with open-source license to increase the scalability of the radiomics studies with fully automated segmentation. Given the importance of breast cancer subtypes as prognostic factors in women with operable breast cancer, automated segmentation of varying breast tumor subtypes will help to analyze imaging biomarkers embedded within the standard of care imaging studies in a larger scale study, which will potentially help radiologists, pathologists, surgeons, and clinicians understand features driving breast cancer phenotypes and pave the way for developing digital twin for breast cancer patients.

10:00 a.m.–10:30 a.m. CAFCW22 Morning Break

10:30 a.m.–10:45 a.m. *GPU-Accelerated Differential Dependency Analysis of Single-Cell Transcriptomics Data*

Presenter: Gil Speyer, PhD, Arizona State University

Authors: Gil Speyer, PhD, Arizona State University, Xishuang Dong, PhD, and Seungchan Kim, PhD, Prairie View A&M University

Abstract: Complex diseases such as cancer and neurological disorders require a systemic approach to understand underlying causes and identify therapeutic targets to help patients. More comprehensive analyses, however, often bring significant computational challenges. EDDY (Evaluation of Differential DependencY) is a computational method to identify rewiring of biological pathways between biological conditions such as drug responses or subtypes of disease [1]. Through its probabilistic framework with resampling and permutation, aided by the incorporation of annotated gene sets, EDDY demonstrated superior sensitivity to other methods. Further development integrated prior knowledge into these interrogations [2]. However, the considerable computational cost for this statistical rigor limited its application to larger datasets. Fortunately, ample and independent computation coupled with manageable memory footprint positioned EDDY as a strong candidate for graphical processing unit (GPU) implementation. With custom kernels to decompose the independence test loop, network construction, network enumeration, and Bayesian network scoring to accelerate the computation. GPU-accelerated EDDY consistently benchmarked at two orders of magnitude in performance enhancement [3]. EDDY has been applied to the determination of rewired pathways controlling differing small molecule responses in cancer cell lines [4]. Further investigations extended this to pathways associated with pulmonary hypertension [5].

Recent emergence of single cell transcriptomic and spatial transcriptomic data raises additional computational challenges, mainly due to an order of magnitude increase in sample size, compared to bulk cell transcriptomic data, often bringing the number of samples to

analyze to hundreds of thousands of cells (samples). This called for additional optimization of the existing EDDY-GPU codes. By working with a NVIDIA team through Princeton Hackathon 2022, we were able to dramatically increase the computational speed of the EDDY-GPU. New sampling strategies has been implemented to adjust to samples counts at this scale. In addition, the latest code development phase identified various performance bottlenecks, which not only improved acceleration but allowed for the incorporation of even larger gene sets, such as immune pathways. Hence, EDDY's statistical rigor can now be brought to bear in the inference of specific diagnostic and treatment strategies for the individual patient, and with an implementation that allows this data analysis to be run on a physician's desktop within reasonable time. We will present preliminary results using this newly improved EDDY-GPU with single cell transcriptomic data from cancer, Alzheimer's disease, and pulmonary hypertension.

10:45 a.m.–11:00 a.m. *Genetic Algorithm Mutations for Molecules with a hybrid Language Model-based GAN architecture*

Presenter: Debsindhu Bhowmik, PhD, Oak Ridge National Laboratory

Authors: Debsindhu Bhowmik, PhD, Oak Ridge National Laboratory; Andrew Blanchard, PhD, Amgen Inc.; Isaac Lyngaas, PhD; Xiao Wang, PhD; Stephan Irie, PhD; and John Gounley, PhD, Oak Ridge National Laboratory

Abstract: Drug discovery is a time-consuming process with successive stages, often taking ~10 to ~15 years to develop candidate molecules into molecular therapeutics. In the computer aided drug discovery, new technologies are being developed to shorten the first stage of the drug discovery process: screening candidates for hit molecules. Given the large size of chemical space from which a new drug molecule has to be selected, this screening step is a challenge and reducing the number of costly experiments required is a priority.

A desirable solution for accelerating this process while keeping the cost under control is to generate drug molecules with desired properties via virtual design-build-test cycle. AI methods and HPC resources have shown potential for leveraging widely available small molecule libraries to generate new optimized molecules.

Recent progress has demonstrated advantages of using generative models, specifically Transformer-based language models (LM) that have been successfully implemented to predict desired chemical properties from sequence data (1, 2). These LMs are applied as powerful automated mutation operator, learning from commonly occurring chemical sequences available in the database. This calculated shift towards chemical-sequence for model training points to a revolution in

moving away from the time-consuming feature engineering and curation that has long relied on molecular properties and fingerprints. As an example, our recent work illustrated a possible LM-based efficient strategy for creating generalizable models for small target molecules and protein sequences (3).

Here we present a first-of-its-kind comparative study between LM and a novel architecture on where LM can be efficiently deployed on Generative Adversarial Network (GAN) platform, to perform different specific optimizations tasks using genetic algorithm-based mutations. Fundamentally this hybrid architecture (LM-GAN) uses traditional generator and discriminator but takes advantage of pre-trained LM while predicting new molecules. During training, the mutation rate is varied from 10% to 100% in four different set of population size ranging from 5K to 50K. Random mutations were considered to select μ parents from population and to generate new molecules with 5 top predictions for a given set of masks. Thus, implemented genetic algorithm has (μ+5μ) survivor selection scheme where only novel unique molecules are retained in population.

Our results show that LM-GAN performs better with smaller size population (up to 10K) in generating molecules both in terms of better optimized properties and with a greater number of atoms, but this trend reverses as the population size increases. On the other hand, LM performs better in terms of generating more novel molecules. Finally, when estimating the ratio of accepted molecules to the generated novel molecules with desired optimized properties, LM-GAN performs consistently better in all population size.

Apart from drug or molecules discovery, in terms of HPC and AI, this work paves the way for further study in understanding the necessity of pre-training and fine-tuning of population data (type, sampling, diversity and size) requirements, the effect of GAN framework on LM models with variation in mutation rate, the effect of LM in replacing CNNs to capture non-local, long-range dependencies and addressing the problem of mode collapse.

11:00 a.m.–11:15 p.m. *Machine Learning in 4DCT Lung Stereotactic Body Radiotherapy (SBRT) Treatment Planning*

Presenters: David Gonzalez, Ignacio Bartol and Sam Taylor, Georgia Institute of Technology

Authors: David Gonzalez, Ignacio Bartol and Sam Taylor, Georgia Institute of Technology; Anees Dhabaan, PhD and Mohammad Khan, PhD, MD, Emory University; Shaheen Dewji, PhD, Georgia Institute of Technology

Abstract: The current project seeks to enhance current 4DCTlung cancer

patient image processing, image guidance, and adaptive radiotherapy verification through the integration of machine learning (artificial intelligence - AI) methods in an existing clinical radiation oncology framework. The current state-of-the-art for Lung SBRT treatment planning begins with the accurate delineation of target organ volumes and their surrounding structures, which is usually done using semi-automatic methods, mixing computer-assisted tools, and dedicated physicians. When it comes to 4DCT scans, what is usually done is to compute a visual average of the images across the different respiratory phases and the contours for those organs (in one specific phase) are delineated. In the last few years, deformable image registration (DIR) techniques have been developed and used in this field to propagate the contour delineation from one specific phase to the rest of the respiratory phases in the CT. Results of the target region delineation are then used by physicians and clinicians to select an optimal treatment phase. Rather than the mostly manual and slow/iterative process introduced above, our current project seeks to create more accurate and more robust delineations through improved machine learning models, decreasing time spent per patient plan, and applying a more mathematically rigorous and objective manner of selecting the optimal radiation treatment gating window, while enhancing image resolution, enhancing target definition, and treatment delivery. The project is subsequently divided in various phases. One phase consists of deriving the deformation parameters to describe the three-dimensional movement of the patient's target treatment region through deformation propagation. The second phase involves a surrogate model for fast reconstruction of the dose distribution in the gross tumor volume(s), GTV(s), and organs at risk, OAR(s), across all the phases, accounting for the deformed target region in time. With these dose profiles, the physicians will have the bounds (within a confidence interval) for the absorbed dose by different organs and tumors across all the respiratory cycle and they will be able to determine if the treatment plan for that patient is accurate and appropriate or if it needs to be replanned. Additional phases of the algorithm use AI approached to enhance this step. The project's algorithm is unique and captures a higher degree of individualization based on the patient's specific organ movement compared with prior non-AI algorithms. Although other studies in the past have explored integrating AI for image segmentation for auto-contouring, our project's novelty lies in the manner of initialized parameters and the specific operations performed. Our project furthers patient-specific treatment planning while adopting a more streamlined approach and helps make more informed decisions using AI, to arrive at improved radiation treatment plans for lung cancer patients undergoing SBRT. We have tested our algorithm in several

patients and have seen encouraging improvements.

11:15 a.m.–11:55 a.m. **Panel:** ***Accelerating Drug Discovery with AI Panel***

Moderator: Sally Ellingson, PhD, University of Kentucky

Panelists:

Jonathan Allen, PhD, Lawrence Livermore National Laboratory
Orly Alter, PhD, University of Utah
Debsindhu Bhowmik, PhD, Oak Ridge National Laboratory
Stephen Litster, PhD, Amazon Web Services
Amber Simpson, Queen's University

11:55 a.m.–12:00 p.m. **Wrap Up**

Featured Speaker

Amber Simpson, Queen's University

Amber Simpson is the Canada Research Chair in Biomedical Computing and Informatics, and Associate Professor jointly appointed in the Department of Biomedical and Molecular Sciences and School of Computing at Queen's University. She is an Affiliate of the Vector Institute for AI as well as a Senior Investigator at the Canadian Cancer Trials Group. Dr. Simpson is the Director of the Centre for Health Innovation, a joint venture with Kingston Health Sciences Centre and Queen's. She received her PhD in Computer Science from Queen's and was a postdoctoral fellow in the Department of Biomedical Engineering at Vanderbilt University. Recently recruited from a faculty position at Memorial Sloan Kettering Cancer Center in New York, she holds research funding from the National Institutes of Health as well funding from all three Canadian research councils. Dr. Simpson is an American Association of Cancer Research and Pancreatic Cancer Action Network award holder and a charter member of NIH study section, which recognizes her innovations in biomedical research. She specializes in biomedical data science with a focus on developing novel computational strategies for improving human health.

Posters

Breast Cancer Patient-Specific Reaction-Diffusion from Spectral Analysis of Immunohistochemistry, Stefano Pasetto, PhD, Moffitt Cancer Center

Comparison of Radiomics from Prostate Bi-parametric MRI and Pharmacokinetic Parameters from Dynamic Contrast Enhanced MRI for Risk Stratification of PI-RADS=3 Prostate Cancer Lesions, Aaron Ng, Michael Sobota, Ansh Roge and Amogh Hiremath, PhD, Case Western Reserve University; Nathaniel Braman, PhD, Picture Health Inc.; Sree Harsha Tirumani, MD, Leonardo Kayat

Bittencourt, MD, PhD, and Lee Ponsky, MD, University Hospitals Cleveland Medical Center; Anant Madabhushi, PhD, and Rakesh Shiradkar, PhD, Emory University

Deep Learning in Cervical Cancer: Searchable Catalogs and Smart Data Curation, Daniela Ushizima, PhD, Lawrence Berkeley National Laboratory, University of California, Berkeley, University of California, San Francisco; Andrea Campos Bianchi, PhD, Federal University of Ouro Preto; Fatima Sombra Medeiros, PhD, Federal University of Ceara, and Claudia Carneiro, PhD, Federal University of Ouro Preto

Developing a Deep Learning Pipeline to Infer Outcomes from Whole Slide Images and Genomic Data for Diffuse Large B Cell Lymphoma, Swaminathan Iyer, MD, MD Anderson

Ensemble Learning of Attention-based Models for Whole Slide Imaging Comprehension, Adam Saunders, University of Dayton; Jacob Hinkle, PhD, Aristeidis Tsaris, PhD, Hong-Jun Yoon, PhD, Folami Alamudun, PhD, Sajal Dash, PhD, Oak Ridge National Laboratory

HPC Pipeline for Calculating Polygenic Risk Scores in Cancer, Mark Xiao, Alex Rodriguez, PhD, and Ravi Madduri, Argonne National Laboratory

The Importance of High Speed Storage in Deep Learning Model's Training, David Apostal and Solene Bechelli, PhD, University of North Dakota

Mapping Phenotypic Heterogeneity in Melanoma onto the Epithelial-hybrid-mesenchymal Axis, Dev Barbhaya, Indian Institute of Technology (IIT), Kanpur

Mitigating Biases in Deep Learning Models for Clinical Document Classification, Mohammed Alawad, PhD, Wayne State University

Pure Seminoma Subtyping Using Computational Approaches, Kirill E. Medvedev, PhD, Anna V. Savelyeva, PhD, Aditya Bagrodia, MD, Liwei Jia, MD, PhD, and Nick V. Grishin, PhD, University of Texas Southwestern Medical Center

Quantum Computing Approach Using Medicinal Plants Anticancer Properties, Amit Saxena, PhD, Centre for Development of Advanced Computing, India, and Akshay Seetharam, Open Health Systems Laboratory

Supporting a Community of Cancer Models with the CANDLE Checkpoint Module, Rajeev Jain, Justin M. Wozniak, PhD, Jamaludin Mohd-Yusof, PhD, Argonne National Laboratory

Temporal Stability of Immuno-Phenotype Radiomic Score in Melanoma, Nizam Ahamed, PhD, Evan Porter, MD, Baher Elgohari, MD, Mohamed Abdelhakiem, MD, John Kirkwood, MD, Diwakar Davar, MD, Zaid Siddiqui, MD, University of Pittsburgh Medical Center Hillman Cancer Center

Transfer Learning for Language Model Adaptation: A Case-study with Hepatocellular Carcinoma, Amara Tariq, PhD, Mayo Clinic; Omar Kallas, MD, Patricia Balthazar, MD, and Scott Lee, MD, Emory University; Terry Desser, MD, and Daniel Rubin, MD, Stanford University; Judy Wawira Gichoya, MD, Emory University; and Imon Banerjee, PhD, Mayo Clinic

Posters can be viewed at https://cafcw22.virtualpostersession.org/

Thank you to our CAFCW22 Program Committee:

Boris Aguilar, PhD, Institute for Systems Biology
Orly Alter, PhD, University of Utah
Jane Bai, PhD, US Food and Drug Administration
Kristy Brock, PhD, MD Anderson
Jeff Buchsbaum, MD, PhD, National Cancer Institute
Hsun-Hsien Shane Chang, PhD, Novartis
Caroline Chung, MD, MD Anderson
Michael Difilippantonio, PhD, National Cancer Institute
Keyvan Farahani, PhD, National Cancer Institute
James Glazier, PhD, Indiana University
Emily Greenspan, PhD, National Cancer Institute
Ryuji Hamamoto, PhD, National Cancer Center Japan/RIKEN, Tokyo
Christopher Hartshorn, PhD, National Cancer Institute
David Hormuth, PhD, University of Texas at Austin
Florence Hudson, Northeast Big Data Innovation Hub at Columbia University in the City of New York
Ai Kagawa, PhD, Brookhaven National Laboratory
Ho-Joon Lee, PhD, Yale University
Ernesto Lima, PhD, University of Texas at Austin
Amanda Paulson, PhD, University of California, San Francisco
Thomas Radman, PhD, National Institutes of Health
Katarzyna (Kasia) Rejniak, PhD, Moffitt Cancer Center
Russell Rockne, PhD, City of Hope
Gundolf Schenk, PhD, University of California, San Francisco
Ilya Shmulevich, PhD, Institute for Systems Biology
Amber Simpson, PhD, Queen's University
Thomas Steinke, PhD, Zuse Institute Berlin
Kristin Swanson, PhD, Mayo Arizona
Gina Tourassi, PhD, Oak Ridge National Laboratory
Thomas Yankeelov, PhD, Oak Ridge National Laboratory
George Zaki, PhD, Frederick National Laboratory

CAFCW21 Organizing Committee:

Eric Stahlberg, PhD, Frederick National Laboratory for Cancer Research
Sean Hanlon, PhD, National Cancer Institute
Sally Ellingson, PhD, University of Kentucky
Patricia Kovatch, Icahn School of Medicine, Mount Sinai
Lynn Borkon, Frederick National Laboratory for Cancer Research
Petrina Hollingsworth, Frederick National Laboratory for Cancer Research

Presenters

Jonathan Allen, PhD, Lawrence Livermore National Laboratory

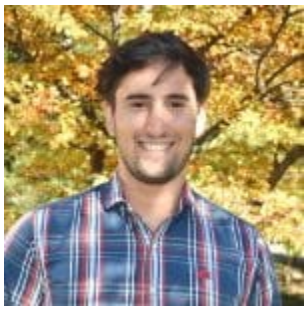
Jonathan E. Allen has worked in bioinformatics research under the Science and Technology and Global Security programs for 15 years at Lawrence Livermore National Laboratory (LLNL). He leads an informatics team focused on modeling and managing data for complex biomolecular systems. He has extensive experience developing new software tools for pathogen characterization from complex biological samples, transcript analysis in host response and machine learning for small molecule drug discovery. Currently, he serves as a technical director for computing in the Accelerating Therapeutics for Opportunities in Medicine (ATOM) consortium. Dr. Allen received his Ph.D. in Computer Science from Johns Hopkins University, after conducting his thesis work at The Institute for Genomic Research (now The J. Craig Venter Institute).

Orly Alter, PhD, University of Utah

Orly Alter is a Utah Science, Technology, and Research associate professor of bioengineering and human genetics at the Scientific Computing and Imaging Institute and the Huntsman Cancer Institute at the University of Utah, the principal investigator of a National Cancer Institute's Physical Sciences in Oncology project, and the chief technology officer and a co-founder of Eigengene, Inc. Alter received her PhD in applied physics at Stanford University and her BSc *magna cum laude* in physics at Tel Aviv University. Her PhD thesis on "Quantum Measurement of a Single System," which was published by Wiley, is recognized as crucial to gravitational wave detection. Inventor of the "eigengene," Alter formulates physics-inspired multi-tensor generalizations of the singular value decomposition to (i) compare and integrate any data types, of any number and dimensions, and (ii) scale with data sizes. Her models (iii) are interpretable in terms of known biology and batch effects and (iv) correctly predict previously unknown mechanisms. By validating a genome-wide pattern of DNA copy-number alterations in brain tumors as the best predictor of survival, her retrospective clinical trial proved that the models (v) discover accurate, precise, and actionable genotype-phenotype relationships, (vi) are relevant to populations based upon whole genomes of small cohorts, and (vii) can be validated. She discovered this, and patterns in lung, nerve, ovarian, and uterine tumors, in public data. Such alterations were recognized in cancer, yet attempts to associate them with outcome failed, demonstrating that Alter's algorithms are uniquely suited to personalized medicine.

Imon Banerjee, PhD, Mayo Clinic

Imon Banerjee is an associate professor and lead AI scientist in the Department of Radiology, Mayo Clinic Arizona. Dr. Banerjee is an affiliated faculty in Computer Engineering, Arizona State University. Her technical training is in computer science, particularly in artificial intelligence and data mining. Her current research is focused on unstructured medical data analysis and integration of multisource medical data from varying hospital systems for building predictive model to benefit diagnosis and treatment. Dr. Banerjee successfully led multiple federal and non-federal research projects related to machine and deep learning-based research projects, e.g., histopathologic model development for PRISM, prediction of CT imaging outcome using current and historic EMR data, longitudinal clinical outcome prediction, early-stage diagnosis of bone and prostate cancer, automated cancer staging.



Ignacio Bartol, Georgia Institute of Technology

Ignacio Bartol is a PhD student in the NRE/MP Program in the RED2 Laboratory at Georgia Tech. His work focuses on using computational fluid and particle dynamics (CFPD) to model the particle deposition of inhaled radioactive aerosols, under the advising of Dr. Dewji and Dr. Tano. In this work he used computer vision techniques to obtain the 3D models of the human respiratory tract. This area mixes high-accuracy numerical modeling with machine learning methods to improve the current state-of-the-art in internalized particle deposition. He is working with Emory university on a project that combines the latest technologies in medical imaging and radiation therapy by using machine learning to optimize treatment planning for Stereotactic Radiation Body Therapy (SBRT) for early-stage lung cancer. Ignacio graduated as a nuclear engineer (i.e., master of engineering degree equivalent) from the Balseiro Institute, a prestigious university in this field located in Bariloche, Argentina. He has previously done two years of physics at the University of Buenos Aires (UBA), a requirement to be admitted to his undergraduate program. Their undergraduate thesis was titled “Optimization of F-18 administered dose for neurological PET/MR studies,” performed at the Institute of Nuclear technologies for health (INTECNUS).

Debsindhu Bhowmik, PhD, Oak Ridge National Laboratory

Debsindhu Bhowmik is an expert in artificial intelligence (AI) technologies especially focusing on the domain of quantitative biophysics for integrative molecular approach. He is presently a computational scientist in Advanced Computing for Health (ACH) Sciences Section of Computational Sciences and Engineering Division (CSED) and Health Data Sciences Institute (HDSI) at the Oak Ridge National laboratory (ORNL). His current work lies on the interface of implementing AI techniques, deploying multi-scale high performance accelerated simulations and performing scattering experiments

(especially neutron and X-Ray) for problems related to soft matter systems especially biomedical and biological sciences, and making new drug molecules with desired properties.

Mayanka Chandrashekar, PhD, Oak Ridge National Laboratory

Mayanka Chandra Shekar is a research scientist in Biomedical NLP at Oak Ridge National Laboratory. At ORNL, she works on natural language processing for free-text electronic health records under the Veteran Affairs' MVP prostate cancer exemplar, NCI-DOE MOSSAIC, and Cincinnati Children's Hospital Medical Center's Mental Health Trajectory projects. She received her PhD from the University of Missouri-Kansas City in transfer learning and domain adaptation in image and text datasets. During her PhD, she did a research internship in Pacific Northwest National Lab on natural language processing, specifically worked on the DARPA Modeling Adversarial Activity (MAA) project on graph generation.

Sally Ellingson, PhD, University of Kentucky

Sally Ellingson is a computational scientist working at the intersection of computational biology and high performance computing. She has undergraduate degrees in computer science and mathematics from Florida Institute of Technology. She obtained her doctoral degree at the University of Tennessee and Oak Ridge National Laboratory under a fellowship funded by the National Science Foundation in computational biology. She is an assistant professor in the Division of Biomedical Informatics at the University of Kentucky College of Medicine. In her additional role as the manager for High Performance Computing Services for the Markey Cancer Center's Cancer Research Informatics Shared Resource Facility, she facilitates high throughput genomics and big data processing for precision medicine resulting in targeted cancer therapies. Recently she has been combining simulations, big data, and machine learning to increase the accuracy of drug binding predictions. With her passion for high performance computing, her research goals lie in harnessing computational power for discoveries otherwise not possible in biomedical areas of high societal importance.

Dr. Ellingson engages in mentoring and outreach, especially for underrepresented groups in computational sciences. She has been actively engaged in the organization of the Students@SC program at Supercomputing since 2014. She has helped organize the Broader Engagement program, spent several years helping with the Student Volunteer program, ran the Mentor-Protégé program, organized Student Programming and School Tours. This year she is the Vice and Deputy Chair for Students@SC.

David Gonzalez, Georgia Institute of Technology

David Gonzalez is a senior mechanical engineering student in the Woodruff School of Mechanical Engineering at the Georgia Institute of Technology with a minor in economics. David conducts research in radiotherapy treatment planning through computational methods and machine learning with the Radiation Engineering, Detection, and Dosimetry Laboratory under Dr. Shaheen Dewji. This research area contributes to the state-of-the-art methods to improve external beam radiation by leveraging statistical properties of medical imaging captures and numerical methods of quantifying internal movement of organs to produce more robust solutions for cancer treatment. Upon graduation, David will matriculate into his PhD program for medical physics at Georgia Tech where he will continue his research.

Sean E. Hanlon, PhD, National Cancer Institute

Sean E. Hanlon is Acting Deputy Director of the NCI Center for Strategic Scientific Initiatives (CSSI) where he provides leadership in the planning, developing, and implementing initiatives with a focus on emerging areas of science with potential impact across the cancer research continuum. Dr. Hanlon provides scientific leadership to collaborative transdisciplinary programs and innovative partnership, including the NCI-CRUK Cancer Grand Challenges program, the NCI's Human Tumor Atlas Network, and the NIH Common Fund's 4D Nucleome program. He also serves as a representative on NCI, NIH, and inter-agency committees, including Cancer Moonshot Implementation teams, the trans-NCI Data Sharing working group, and the trans-NCI Artificial Intelligence working group.

Patricia Kovatch, Icahn School of Medicine at Mount Sinai

Patricia Kovatch is the Senior Associate Dean for Scientific Computing and Data Science at the Icahn School of Medicine at Mount Sinai, founding the division in October 2011. She is also an associate professor for the Department of Genetics and Genomic Sciences, the Icahn Institute for Data Science and Genomic Technology and Pharmalogical Sciences.

Stephen Litster, PhD, Amazon Web Services

Steve Litster is Senior Manager of Health Care and Life Sciences Compute Specialists at AWS.After obtaining his PhD in chemical crystallography in 1992, he continued his research career at the University of Calgary in the area of protein crystallography. During this time, he was exposed to his first super computer, the Fujitsu VPX240, and realized the impact HPC could have on scientific discovery. In 1998, he was awarded an HHMI Medical Fellowship working under the supervision of Professors Don Wiley and Steve Harrison in the area of macromolecular and computational crystallography at Harvard University. In 2004, he joined the Novartis Institutes for Biomedical Research as the global lead for Scientific Computing, specializing in the application of high performance, data analytics and cloud computing to advance drug discovery. After 14 years in the role, he moved to the Markley Group as CTO and subsequently joined AWS in 2020 as Principal Business Development Lead for HPC in Healthcare and Life Science.

Gil Speyer, PhD, Arizona State University

Upon receiving his BS in electrical engineering from MIT, Dr. Speyer worked for Xilinx, Inc. in San Jose, CA, on programmable logic. Subsequently, he earned his MS and PhD in electrical engineering at Arizona State University (ASU) researching transport in molecular devices. Recent computational biology efforts as research faculty at the Translational Genomics Research Institute (TGen) led to the GPU accelerated implementation of the EDDY package. Now as the Director of the Computational Research Accelerator out of Knowledge Enterprise at ASU, Dr. Speyer investigates and promotes cutting-edge hardware and software solutions for the computational research community.

Eric Stahlberg, PhD, Frederick National Laboratory for Cancer Research

Eric Stahlberg serves as the director of Cancer Data Science Initiatives at the Frederick National Laboratory for Cancer Research (FNLCR). Joining the team at Frederick in 2011 to establish and lead the bioinformatics core supporting the NCI Center for Cancer Research, Dr. Stahlberg shifted his attention in 2014 to lead a new NCI Center for Biomedical Informatics and Information Technology-supported initiative to accelerate cancer research through applications of high performance computing. Working collaboratively with NCI leadership, Dr. Stahlberg helped established the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) collaboration between the NCI and the US Department of Energy as well as Accelerating Therapeutics for Opportunities in Medicine (ATOM), a public-private collaboration to

dramatically increase the pace and success of new treatments. He has further brought his collaborative insight to the formation of IMPROVE, now being co-led by colleagues from FNL, NCI, and DOE. Driven to forge advances at the intersection of leading-edge science and computing, Dr. Stahlberg continues to build the cross-disciplinary community through efforts with the Computational Approaches for Cancer and HPC Applications of Precision Medicine workshops at SC, precision medicine workshops at ISC, and other community efforts including advancing biomedical digital twins. In 2017, he was recognized as one of FCW's Federal 100. Stahlberg holds a PhD in computational chemistry from the Ohio State University and bachelor's degrees in computer science, chemistry, and mathematics.