

# Data-Driven Precision Medicine and Translational Research in the Era of

Friday, May 1, 2020 8:00 a.m.- 5:40 p.m.



#### Friday, May 1, 2020 8:00 a.m. - 5:40 p.m.

#### **PROGRAM OVERVIEW**

"Big data" are rapidly shaping the biomedical and clinical research in the new era of precision medicine. In addition to "traditional" big data like genomics, proteomics and neuroimaging data, "novel" types of high-dimensional data are being massively explored. Emerging examples are compositional microbiome biomarkers, health information technology (HIT) including digital biometrics data, real world evidence (RWE) based on electronic health records (EHR), and even a combination of big data from multiple areas. Although big data can result in numerous analytic challenges, they add highly valuable information to the knowledge set essential for translating research efforts to precision medicine practices such as patient screening, disease detection, treatment selection, response monitoring, toxicity or morbidity management, and patient risk stratification.

The symposium intends to cover the following six areas:

- 1. Pediatric Oncology Data Science: Progress, Perspectives, and Challenges
- 2. Precision Medicine and Big Data in Medicine: Challenges and Opportunities
- 3. Emerging Evidence: Advances in Quantitative Microbiome and Wearable Technology Research
- 4. Real-time/Dynamic Disease Risk Prediction
- Electronic Health Records (EHR) based Real World Evidence (RWE)
- 6. Interpretable Machine and Deep Learning: Theory and Applications in Healthcare

This one-day symposium aims to gather renowned data science researchers in emerging big data fields to showcase exciting advances in developing data-driven approaches that help to improve precision medicine and related state-of-the-art technologies involving modern statistical learning, deep learning and artificial intelligence (AI) concepts. The event will highlight applications of big data science and tools to advance precision medicine and translational research.

#### HOST



**Li Tang, PhD** Associate Member, St. Jude Faculty Department of Biostatistics St. Jude Graduate School of Biomedical Science

### SPECIAL GUESTS



**Charles W. M. Roberts, MD, PhD** Member, St. Jude Faculty Executive Vice President Director, Comprehensive Cancer Center Director, Molecular Oncology Division



**Motomi Mori, PhD** Member, St. Jude Faculty Chair, Biostatistics Endowed Chair, Biostatistics Biostatistics Department

### **SPEAKERS**



David Benkeser, PhD Assistant Professor, Department of Biostatistics and Bioinformatics Rollins School of Public Health Emory University

#### Title: Causal inference and the role of machine learning

**Abstract**: Recent years have seen a huge surge of interest in machine learning, and deep learning in particular. In all this hype it is woefully easy to lose sight of the age-old adage that correlation does not equal causation. Causation is at the heart of many questions involving health care policy and clinical decision making -- so what role can machine learning play? In this talk, I will review recent developments towards integrating machine learning and causal inference. I will argue that health researchers absolutely should be excited about machine learning, but must understand exactly what it does (and does not) provide in the context of drawing causal conclusions from data. Several applications across different disease areas will be provided as motivation and illustration.



**Ciprian M. Crainiceanu, PhD** Professor, Department of Biostatistics Johns Hopkins University

#### Title:Wearable and implantable technology (WIT) with biopharmaceutical applications

**Abstract**: Wearable and Implantable Technology (WIT) is rapidly changing the data analytic landscape due to their reduced bias and measurement error as well as to the sheer size and complexity of the recorded signals. In this talk I will review some of the most used and useful sensors in the ever-expanding WIT analytic environment and their potential impact on Biopharmaceutical research. I will describe the use of accelerometers, heart and glucose monitors, as well as their combination with ecological momentary assessment (EMA) for improved patient reported outcomes. Several case studies highlighting the application of WIT in clinical trials will be provided. I will introduce an array of scientific problems that can be answered using WIT and describe methods designed to analyze the WIT data from the micro- (sub-second-level) to the macro-scale (minute-, hour- or day-level) data. Based on a better understanding of the WIT data, I will show how the design of experiments can be improved for specific Biopharmaceutical interventions.



William Hanage, PhD Associate Professor Department of Epidemiology Center for Communicable Disease Dynamics T.H.Chan School of Public Health Harvard University

#### Title: The role of modeling in the COVID-19 pandemic

**Abstract:** For the foreseeable future we will be living with SARS-CoV-2 and the pandemic it has caused. While the pandemic is in its early stages, mathematical modeling offers a way to explore possible futures, and the consequences of different interventions. We will talk about how such models are made and their reliance upon assumptions. We will specifically talk about examples of models of transmission in healthcare, and the role of children in the pandemic, including results of a model of SARS-CoV-2 transmission and how it might impact the non-covid19 cohort.



Michael R. Kosorok, PhD W.R. Kenan, Jr. Distinguished Professor and Chair, Department of Biostatistics Professor, Department of Statistics and Operations Research University of North Carolina at Chapel Hill

#### Title: Recent developments and future possibilities in precision health

**Abstract:** Precision health is the science of data-driven decision support for improving health at the individual and population levels. This includes precision medicine and precision public health and circumscribes all health-related challenges which can benefit from a precision operations approach. This framework strives to develop study design, data collection and analysis tools to discover empirically valid solutions to optimize outcomes in both the short and long term. This includes incorporating and addressing heterogeneity across persons, time, location, institutions, communities, key stakeholders, and other contexts. In this lecture, we review some recent developments in the area, imagine what the future of precision health can look like, and outline a possible path to achieving it. Several applications in a number of disease areas will be examined.



**Michael LeBlanc, PhD** Member, Biostatistics, Public Health Sciences Division, Fred Hutchinson

#### Title: Experiences in building sequential and platform precision medicine trials

**Abstract:** Recent developments in biologically targeted therapies and the rapidly increasing number of successful immunotherapies have fundamentally changed the treatment strategies for many cancers. Current evaluation of new treatments for cancer utilizes designs that enrich or target the population of patients who are thought to show the maximum treatment effect. Challenges for oneat-time or sequential precision medicine trials include accrual issues with associated with targeting biomarkers or genetic abnormalities which often occur at a low frequency. We explore the concept of master or platform protocols which use a single infrastructure, overall trial design, and protocols to simultaneously evaluate multiple drugs and/or disease sub-populations in sub-studies. We present two examples of large-scale precision medicine studies primarily funded by the National Cancer Institute and by public private partnerships. The first, Lung-MAP, is a precision medicine platform trial for advanced lung cancer. The second is DART (Dual Anti-CTLA-4 & Anti-PD-1 blockade Trial), an innovative "basket" design trial, which allows for testing the drug combination simultaneously in approximately 50 rare tumor type cohorts. Both studies are conducted by the SWOG Cancer Research Network.



Hongzhe Li, PhD Perelman Professor of Biostatistics, Epidemiology and Informatics Professor of Biostatistics and Statistics Vice Chair of Integrative Research Director, Center for Statistics in Big Data Perelman School of Medicine, University of Pennsylvania

## Title: Interrogating the gut microbiome: estimation of growth dynamics and prediction of biosynthetic gene clusters

**Abstract:** The gut microbiome plays an important role in maintenance of human health. Highthroughput shotgun metagenomic sequencing of a large set of samples provides an important tool to interrogate the gut microbiome. Besides providing footprints of taxonomic community composition and genes, these data can be further explored to study the bacterial growth dynamics and metabolic potentials via generation of small molecules and secondary metabolites. In this talk, I will present several computational and statistical methods for estimating the bacterial growth dynamics and for predicting Biosynthetic Gene Clusters (BGCs) based on shotgun metagenomic data, including optimal permutation recovery based on low-rank projection and deep learning methods to improve prediction of BGCs. I will demonstrate the application of these methods using several ongoing microbiome studies of inflammatory bowel disease at the University of Pennsylvania.



**Lei Liu , PhD** Professor, Division of Biostatistics, School of Medicine Washington University in St. Louis

#### Title: Precision medicine: subgroup Identification in longitudinal pharmacogenetic studies

**Abstract:** In clinical studies, the treatment effect may be heterogeneous among patients. It is of interest to identify subpopulations which benefit most from the treatment, regardless of the treatment's overall performance. In this study we are interested in subgroup identification in longitudinal studies when nonlinear trajectory patterns are present. Under such a situation, evaluation of the treatment effect entails comparing longitudinal trajectories while subgroup identification requires a further evaluation of differential treatment effects among subgroups induced by moderators. To this end, we propose a tree-structured subgroup identification method, termed "interaction tree for longitudinal trajectories", which combines mixed effects models with regression splines to model the nonlinear progression patterns among repeated measures. Extensive simulation studies are conducted to evaluate its performance and an application to an alcohol addiction pharmacogenetic trial is presented.



**Sheng Luo, PhD** Professor, Department of Biostatistics and Bioinformatics Duke University

#### Title: Functional data analysis: novel statistical methods and applications in medical research

**Abstract:** Modern technology increasingly collects data whose units of observations are functions recorded continuously during a time interval or intermittently at several discrete time points. These functions can be one-dimensional curves (e.g., electroencephalogram or EEG, physical activity data measured by accelerometers, and stock price of Amazon), two-dimensional images (e.g., a slice of MRI), three-dimensional images (e.g., voxel-based whole-brain image), or four-dimensional object (e.g., functional MRI). Functional data analysis (FDA) is the statistical methodology for analyzing such data. In the first part of this presentation, I will give a brief introduction of functional data and the analysis methods. In the second part, I will give two detailed case-studies of our recent research work. The first example investigates the effects of brain atrophy measured by MRI on the cognitive function and risk of developing Alzheimer's disease while the second example investigates the association between physical activity data and physical performance among aged individuals.



Arzu Onar-Thomas, PhD Member, St. Jude Faculty Biostatistics Department St. Jude Graduate School of Biomedical Sciences St. Jude Children's Research Hospital

#### Title: Precision medicine in pediatric brain tumors: challenges and opportunities

**Abstract:** This talk will focus on some of the unique challenges and opportunities that exist in implementing molecular classification into pediatric clinical trials and the implications of the use of precision medicine in ultrarare diseases such as pediatric central nervous system (CNS) tumors. The talk will focus on two vignettes: one in SHH-driven medulloblastoma and the other in pediatric Low-Grade Glioma. We will summarize a decade long effort through several molecularly-driven clinical trials in each of these two tumor-types highlighting roadblocks and surprises along the way. Specific emphasis will be placed on demonstrating the uniqueness of pediatric populations in the context of these examples and the level of collaboration that is required within the pediatric cancer community in order to fully evaluate molecularly targeted agents. We will conclude with a summary of some of the current efforts to accelerate progress in pediatric CNS tumors.



**Peter Song, PhD** Professor Associate Chair,Research Department of Biostatistics School of Public Health University of Michigan

## Title: An epidemiological forecast model to assess the effect of social distancing on flattening the coronavirus curve in the USA

**Abstract:** We develop a health informatics toolbox that enables us to project time-course dynamics of the COVID-19 epidemics in the USA. This toolbox is built upon a hierarchical epidemiological forecast model for observed daily proportions of infected and removed cases that are generated from an underlying Markov process of evolving Susceptible, Infectious and Removed (SIR) compartments of the COVID-19 infectious disease. We extend the classical SIR model to incorporate various types of time-varying social distancing protocols, which allows us to assess the effect of social distancing on flattening the coronavirus curve in the US. Some possible extensions of the epidemiological model to predict county-level risk are discussed. Such regional risk information is of critical importance for business reopening in the near future.



**Jeremey Weiss, MD, PhD** Assistant Professor of Health Informatics Heinz College Carnegie Mellon University

#### Title: Machine learning amidst health record data irregularity: subgrouping in dimensions of space and time

**Abstract:** Real world evidence (RWE) in the form of electronic health records (EHRs) presents an opportunity and a challenge for health analysts. In this talk I will navigate EHR challenges of scale and passive data collection that lead to techniques for clustering, visualization, and risk stratification as tools for RWE users. I will describe one finding: that in recurrent event settings, likelihood optimization gives disproportionate attention to those at high risk and leads to comparatively underwhelming results in low risk individuals. We propose an approach by introducing an adjusted likelihood formulation as an objective for point process neural networks and apply it to identifying mental status changes in the critical care setting.



Jinghui Zhang, PhD Member, St. Jude Faculty Chair, Department of Computational Biology St. Jude Endowed Chair in Bioinformatics St. Jude Graduate School of Biomedical Sciences St. Jude Children's Research Hospital

#### Title: Real-time analysis of the expanding pediatric cancer genomics data on St. Jude Cloud

Abstract: Sharing of cancer genomics data and analysis tools is essential to facilitate scientific discovery and thereby to improve outcomes for pediatric cancer patients. To support the worldwide pediatric cancer community, we developed St. Jude Cloud (https://stjude.cloud), a platform that includes genomics data from over 10,000 pediatric cancer patients, analytical genomics tools and user-friendly visualizations. Since its debut in 2018, access to St. Jude Cloud datasets has been granted to 159 research groups in 16 countries. Our latest release includes approximately 2,000 RNA-seg pediatric acute lymphoblastic leukemia (ALL) samples from a pan-ALL subtype classification study. Additionally, results from our three-platform clinical sequencing of whole genome, whole exome and transcriptome are periodically uploaded prior to publication as part of the real-time clinical genomics (RTCG) initiative. This regular deposition of clinical data is enabled by a rigorous and largely automated process including confirmation of patient consent, sequence quality, sample deidentification, remapping to the latest genome build and manual quality checking. From March through December 2019, we uploaded 1,798 WGS, 2,304 WES and 1,109 RNA-seq RTCG samples. Altogether, RTCG uploads have provided genomics sequencing data for 51 types of pediatric cancer, including 11 rare cancers not represented in our prior release of research data. Additionally, we have focused on developing applications that allow users to upload their own data and perform an integrated analysis with the data hosted on St. Jude Cloud. The latest addition is an explorable, interactive t-SNE plot where a user's uploaded RNA-seq data is plotted amidst a pre-computed landscape of pediatric cancers. The St. Jude Cloud tumors are annotated with defined diagnoses and, when available, driver mutations, fusions and subtypes. Users have the option to restrict the analysis to data generated from brain tumors, solid tumors or leukemias. The interactive t-SNE plot, along with the delivery of HTSeq raw counts for RNA-seq data in St. Jude Cloud, will become an important resource for improving classification of pediatric cancers in the research enterprise. Through the above real-time provision and analysis of pediatric clinical genomics data on the St. Jude Cloud Genomics Platform, we aim to facilitate rapid advances for diagnosis and therapeutic decision making for children with catastrophic disease.

## AGENDA

TIME	EVENT
8:00 - 8:10 am	<b>Opening remarks – Charles Roberts</b> , Member, St. Jude Faculty; Executive Vice President; Director, Comprehensive Cancer Center; Director, Molecular Oncology Division
Breaking Session I: COVID-19	
8:10 - 8:55 am	An Epidemiological Forecast Model to Assess the Effect of Social Distancing on Flattening the Coronavirus Curve in the USA <b>Peter Song</b> , University of Michigan
Session 1: Pediatric Oncology Data Science: Progress, Perspectives, and Challenges	
8:55 - 9:35 am	Real-Time Analysis of the Expanding Pediatric Cancer Genomics Data on St. Jude Cloud <b>Jinghui Zhang,</b> St. Jude Children's Research Hospital
9:35 - 10:15 am	Precision Medicine in Pediatric Brain Tumors: Challenges and Opportunities <b>Arzu Onar-Thomas,</b> St. Jude Children's Research Hospital,
10:15 - 10:25 am	Session 1 Discussion and Break
Session 2: Precision Medicine and Big Data in Medicine: Challenges and Opportunities	
10:25 - 11:05 am	Experiences in Building Sequential and Platform Precision Medicine Trials Michael LeBlanc, Fred Hutchinson
11:05 - 11:45 am	Recent Developments and Future Possibilities in Precision Health Michael Kosorok, University of North Carolina-Chapel Hill
11:45 - 11:55 am	Session 2 Discussion and Break
Breaking Session II: COVID-19	
12:00 - 12:45 pm	The Role of Modeling in the COVID-19 Pandemic William Hanage, Harvard University
Session 3: Emerging Evidence: Advances in Quantitative Microbiome and Wearable Technology Research	
1:00 - 1:40 pm	Interrogating the Gut Microbiome: Estimation of Growth Dynamics and Prediction of Biosynthetic Gene Clusters <b>Hongzhe Li,</b> University of Pennsylvania
1:40 - 2:20 pm	Wearable and Implantable Technology (WIT) with Biopharmaceutical Applications <b>Ciprian M. Crainiceanu,</b> Johns Hopkins University
2:20 - 2:30 pm	Session 3 Discussion and Break
Session 4: Real-time/Dynamic Disease Risk Prediction	
2:30 - 3:10 pm	Functional Data Analysis: Novel Statistical Methods and Applications in Medical Research <b>Sheng Luo,</b> Duke University
3:10 - 3:50 pm	Precision Medicine: Subgroup Identification in Longitudinal Pharmacogenetic Studies Lei Liu, Washington University in St. Louis
3:50 - 4:00 pm	Session 4 Discussion and Break
Session 5: Electronic Health Records (EHR) based Real World Evidence (RWE)	
4:00 - 4:40 pm	Machine Learning Amidst Health Record Data Irregularity: Subgrouping in Dimensions of Space and Time <b>Jeremy Weiss,</b> Carnegie Mellon University
Session 6: Interpretable Machine and Deep Learning: Theory and Applications in Healthcare	
4:40 - 5:20 pm	Causal Inference and the Role of Machine Learning David Benkeser, Emory University
5:20 - 5:30 pm	Session 5 and 6 Discussion
5:30 - 5:40 pm	Concluding Remarks - Motomi Mori, Member, St. Jude Faculty, Endowed Chair, St. Jude Biostatistics

