

Faculty Research Interests

Clinical Biostatistics

- ❖ [Rick Chappell](#) My research is on the analysis and design of clinical trials, link estimation in generalized linear models, bivariate and nonparametric survival analysis, predicting the course of Alzheimer's Disease / prodromal AD, and models in radiobiology.
- ❖ [Guanhua Chen](#) My expertise lies in the development of statistical and machine learning methods for clinical and biomedical research. I focus on uncovering intricate patterns in high-dimensional data from various sources, including genomics and electronic health records, to promote precision medicine. My areas of technical interest include causal inference, reinforcement learning, and empirical process.
- ❖ [Tom Cook](#) Specific interests include analysis of recurrent events in the presence of competing risks, composite outcomes, sample size calculations accounting for potential missing data, principal stratification and related topics in randomized clinical trials.
- ❖ [Ron Gangnon](#) My research interests include spatial and spatio-temporal epidemiology, high-dimensional variable selection with applications to spatial cluster detection, small area estimation including ranking and clustering, age-period-cohort modeling, modeling complex chemical mixtures and epidemiologic applications of multi-state models. I have active collaborations in cancer, infectious disease, environmental epidemiology, infectious disease, asthma and allergic diseases and geriatrics.
- ❖ [Tianyuan Lu](#) Our research focuses on developing and implementing rigorous genetic epidemiology and statistical genetics methods to improve the prevention, diagnosis, and treatment of complex diseases, and translating research findings into new medical care approaches and therapies. We strive to conduct research from which all members of society can benefit.
- ❖ [Lu Mao](#) My research is in causal inference, missing or censored data, semiparametric theory, clinical trials, statistical learning in medical studies, diagnostic radiology, and cardiovascular disease.
- ❖ [Mary Ryan Baumann](#) My research interest is in the analysis of correlated data and study design. I have particular interests in pragmatic and cluster randomized trial methodology, group sequential trial monitoring, and trial issues in the Alzheimer's disease space.
- ❖ [Jiwei Zhao](#) My overarching research is to use statistical methods and machine learning techniques to analyze data with massive structures in various biomedical studies. He has broad interests in clinical trial, missing data analysis, causal inference, semiparametrics, robustness, high-dimensional statistical inference, semi-supervised learning, transfer learning, patient-reported outcomes, electronic health records, health services research, health disparity/equity.

Clinical Informatics

- ❖ [Anoop Mayampurath](#) I co-lead the ICU Data Science Lab with Drs. Churpek and Afshar. We develop models to predict outcomes in patients using electronic health record data and machine learning. We use structured, unstructured, and image data for our research and touch upon a number of clinical problems. My particular interest is developing explainable models for predicting outcomes in hospitalized children.

Computational Biology

- ❖ [Mark Craven](#) The focus of my research program is to develop and apply machine-learning methods to inferring models of, and reasoning about, networks of interactions among genes, proteins, clinical and environmental factors, and phenotypes of interest. Of particular interest are machine-learning problems that involve sequences, time series, networks, natural language, and high-dimensional data. The work in my group emphasizes interpretable machine learning, active learning, and learning with weak supervision.
- ❖ [Colin Dewey](#) My research focuses on the computational analysis of biological sequence data, particularly large-scale nucleotide sequence sets, such as the output of the latest DNA sequencers and the genome sequences that may be assembled with them. This area of research is crucial to the advancement of biology, which has become increasingly dependent on high-throughput sequencing for a wide variety of applications. My projects generally fall into one of three main lines of research: (i) transcriptome analysis with RNA sequencing (RNA-seq) data, (ii) evolutionary analysis of whole genomes, and (iii) automated standardization of metadata to enable training of machine learning models on large sequence databases. In all lines of my research, I focus on the development of new computational methodology that is both efficient and statistically-grounded.
- ❖ [Huy Dinh](#) Our research is to understand how tumor heterogeneity changes toward finding immunotherapeutic biomarkers and targets by leveraging computational biology approaches.
- ❖ [Anthony Gitter](#) My research is in computational approaches to reconstruct signaling pathways and transcriptional regulatory networks from multiple types of high-throughput data. In addition, we use machine learning to prioritize biological experiments for drug discovery, protein engineering, and other biomedical challenges.
- ❖ [Irene Ong](#) My research interests lie in data mining, artificial intelligence, machine learning, probabilistic methods, dynamical models, inductive logic programming, and statistical relational learning with applications to biological and medical data. She is particularly interested in the integration and analysis of clinical, genomics, transcriptomics, proteomics, immunome, metagenomics, metabolomics, and other molecular data especially as it pertains to precision medicine.
- ❖ [Sushmita Roy](#) My research focuses on developing and applying computational methods based on statistical machine learning for understanding gene regulatory networks that control gene expression patterns in living cells. I am interested in identifying networks under different environmental, developmental, disease, population and species contexts, comparing these

networks across contexts, and understanding how changes in the regulatory network cause differences across contexts. Our approaches harness the increasingly available repertoires of bulk and single cell genomic measurements and are applicable to diverse plant and mammalian systems. Our methods can be used to study gene regulatory processes underlying different normal conditions, e.g., cell fate specification and host microbe interactions, and, disease processes such as cancer and neurodevelopmental diseases. Three major directions of research in my group are to develop methods for: (a) Genome-scale transcriptional regulatory network inference from bulk and single cell omic datasets, (b) Evolutionary analysis of gene regulatory networks and regulatory genomic datasets, (c) Understanding 3D Genome organization.

- ❖ [Daifeng Wang](#) My lab focuses on developing machine learning approaches and bioinformatics tools to analyze multimodal data for (1) improving genotype-phenotype prediction and (2) understanding functional genomics and gene regulation in the human brains and brain diseases. The current research topics include bio-inspired machine learning, single-cell functional genomics, multimodal integration and imputation.

Imaging

- ❖ [Juan Caicedo](#) Our work is focused on creating computational models of phenotypic variation using microscopy imaging and machine learning. Microscopy imaging is the best strategy to observe the complexity of cellular structures and tissue organization, and to capture complex phenotypes in space and time. We use machine learning to model the biological information captured in microscopy images and to enable quantitative inference with complex phenotypes. These computational models enable us to study the impact of diseases and the effects of treatments at the cellular level, bringing precision and sensitivity to biological research.
- ❖ [Moo Chung](#) My research interest is functional data analysis on manifolds, topological data analysis (TDA), shape analysis, network analysis and computational neuroimaging.
- ❖ [Yin Li](#) Computer vision and its applications in healthcare; Mobile Health and wearable Computing
- ❖ [Vikas Singh](#) My overarching research goal is to develop novel models and algorithms to solve *real* image (and data) analysis problems in scientific, biomedical, engineering or industrial applications. We often focus on developing algorithms (by extending or adapting ideas from statistics, geometry or optimization) whose mathematical properties can be analyzed and then providing open-source tools to benefit the scientific or industrial application that initiated the work.
- ❖ [Andreas Velten](#) The computational optics group develops imaging and pattern recognition methods that use ultra-fast and single photon cameras. We build imaging systems that can see the propagation of light, and use fast fluorescence response of materials to infer their molecular makeup. This allows our cameras to diagnose tissue conditions, provide crucial contrast during surgery, image around corners, see through fog.

Machine Learning

- ❖ [Junjie Hu](#) has a broad interest in natural language processing and machine learning. In particular, he works on multilingual NLP, transfer learning, multimodal learning, and their applications to support healthcare and communications. His research goal is to build robust intelligent systems that evolve with changes in the environment and extract information from natural language text.
- ❖ [Daniel Pimentel-Alarcón](#) My research focuses on developing theory and algorithms to learn from messy data — that is, data with outliers, missing values, noise, sparsity patterns, unbalanced classes, small sample size, etc. I investigate both, the fundamental limits of how messy can data be so that we can still learn from it, and the practical algorithms that we can use to learn from such messy data. To these ends, I use techniques from optimization, signal processing, algebraic geometry, machine learning, and more. This research is motivated by a wide range of applications ranging from biomedicine to astronomy.

Statistical Genomics

- ❖ [Karl Broman](#) My research focuses on statistical problems in genetics and genomics, particularly methods for detecting and identifying genes contributing to variation in complex traits in experimental organisms.
- ❖ [Sunduz Këlës](#) Our group develops statistical and computational data science methods for problems in high dimensional genomic and biomedical data. We exploit, leverage, and integrate high-throughput functional, genomic, and genetic data to understand the fundamentals of gene regulation in development and disease.
- ❖ [Christina Kendzierski](#) My research focuses primarily on statistical methods and software for computational biology and genomics.
- ❖ [Qiongshi Lu](#) Current research focuses on developing statistical methods for human genetics research. Areas of expertise in Qiongshi's group include genome-wide association studies, gene-environment interaction, and genetic risk prediction.
- ❖ [Michael Newton](#) My research concerns statistical inference problems from the biological sciences, especially the use and development of stochastic models, nonparametric and empirical Bayesian methods, and high-dimensional inference. In addition to research on data-science methodology, I work extensively in collaborative projects, especially where the analysis of high-dimensional molecular, genetic, or genomic measurements is required.
- ❖ [Zhengzheng Tang](#) Our group focuses on the development and application of statistical methods and computational tools for high-dimensional “omics” data, arising from modern high-throughput technologies. Our current research interests include developing novel methods for integrating functional information in genetic association studies, meta-analysis for sequencing studies, and analysis of microbiome & metagenomics data.

Affiliates

- ❖ [Majid Afshar](#) research interests include predictive analytics (diagnostics and prognostics) using electronic health record data and public health data for individuals with critical illness and substance misuse. His focus is on early detection and monitoring using methods in natural language processing and machine learning.
- ❖ [Elizabeth Burnside](#) is interested in using computational techniques to improve the early detection of breast cancer. Her work centers on the development of an expert system that can accurately assess the probability of breast cancer using patients' demographic risk factors and mammography findings. Dr. Burnside and colleagues have developed a Bayesian Network that is designed to assist radiologists in the post-discovery aspects of mammography: interpretation and decision-making. Currently, she is investigating how novel informatics algorithms can improve the performance of this system and therefore impact breast cancer care.
- ❖ [Matthew Churpek](#) Dr. Churpek's data science laboratory utilizes electronic health record data and machine learning techniques, such as natural language processing and deep learning, to identify patients at risk for clinical deterioration, sepsis, acute kidney injury, and other syndromes of critical illness. His research has been supported by a K08 from NHLBI, an ATS Foundation Recognition Award for Early Career Investigators, and an R01 from NIGMS.
- ❖ [AnHai Doan](#) My work has charted new directions or bet on emerging directions that I believe would become fundamental for data management. Solving problems in these directions often requires a combination of machine learning, scalable data management, effective human-data interaction, and cloud technologies.
- ❖ [Kevin Eliceiri](#) develops new biomedical imaging and analysis methods to understand the roles of cells in disease. Our projects include developing measures that can characterize the progression of cancer, building instruments to collect richer data than traditional methods, using machine learning to classify biomedical images, and developing open-source software for use by the image analysis community around the world.
- ❖ [Felix Elwert](#) is a scholar of social inequality, social demography, and applied statistics. He develops methods of causal inference and conducts research on the contextual drivers of inequality, income, education, and health. Ever on the lookout for ways to coax causality from data, Elwert collaborates with teams around the globe to conduct large-scale randomized experiments, mine population registers, and analyze surveys.
- ❖ [John Garrett](#) Dr. Garrett has extensive expertise in evaluating and translating modern deep learning and machine learning techniques for medical imaging applications. A significant area of interest lies in developing and assessing these tools within multi-site frameworks, particularly through approaches like federated learning. With a robust background in x-ray and CT physics and substantial experience in high-performance and parallel computing, Dr. Garrett's research emphasizes leveraging machine learning for opportunistic screening in abdominal CT, clinical evaluation of deep learning tools for computer vision and image analysis, and advancing methodologies such as federated learning to enable secure and large-scale computational

solutions in medical imaging.

- ❖ [Hyunseung Kang](#) my research is focused on developing methods to analyze causal relationships by using instrumental variables, econometrics, semi/nonparametric methods, network analysis, and machine learning. I am interested in applications to genetics, epidemiology, infectious diseases, health policy, education, and applied microeconomics.
- ❖ [Benjamin Lengerich](#) We develop ML methods and AI systems to advance precision medicine.
- ❖ [Brian Patterson](#) research interests are in clinical informatics and geriatric emergency medicine. His work aims to use routinely collected clinical data to generate actionable insights to improve the quality and safety of emergency care for older adults. His current work focuses on automated risk scoring to identify older adults at high risk for falling after ED visits, and automating referral to interventions to prevent future falls.
- ❖ [Guilherme J.M. Rosa](#) Dr. Rosa's research applications encompass the analysis of farm-level operational data to optimize management practices, high-throughput phenotyping techniques for real-time monitoring of individual animals and disease surveillance, as well as quantitative genetics/genomics and breeding.
- ❖ [Margie Rosenberg](#) Rosenberg's research interests are in the application of statistical methods to health care, and applying her actuarial expertise to cost and policy issues in health care. She publishes in actuarial, medical, dental, and other relevant journals. Her recent work has to analyze the linkage oral health to overall health, reflecting differences among various population subgroups.
- ❖ [Adam Rule](#) My research spans the areas of health informatics and human-computer interaction. I study how people interact with electronic health records (EHRs), how we can use traces of EHR activity to study collaborative workflows in healthcare, and how people tell stories with data.
- ❖ [David Schneider](#) research group is broadly interested in improving decision-making for surgical patients. Specifically, they use informatics tools to automate and improve decision-making at the point of care. This includes machine learning and natural language to provide additional knowledge and predictive capability for clinical decisions.
- ❖ [Claudia Solis-Lemus](#) Our work produces a collection of new statistical methods with solid theoretical guarantees and efficient computational implementations that are adaptable to analyze the complex characteristics of modern big biological data.
- ❖ [Christine Sorkness](#) is ICTR Senior Associate Executive Director and provides leadership for ICTR's workforce development and mentorship training initiatives, as well as the Pilot Awards program. She serves as an investigator with ICTR's Collaborative Center for Health Equity. She has a special interest in advancing health equity in the management of asthma in children residing in urban settings and serves as a co-investigator with the NIAID-funded Childhood Asthma in Urban Settings (CAUSE) multisite intervention trials.

- 
- ❖ [Hannah Wayment-Steele](#) Our research seeks a predictive understanding of biomolecular dynamics and mechanisms that underlie biological function and shape evolution.
 - ❖ [Jun Zhu](#) The main components of my research activities are statistical methodological development and scientific collaborative research. My statistical methodological research concerns developing statistical methodology for analyzing spatially referenced data (spatial statistics) and spatial data repeatedly sampled over time (spatio-temporal statistics) that arise in the biological, physical, and social sciences.