Student Presentations

Carter Allen
“SPRUCE: Bayesian Multivariate Mixture Models for High Throughput Spatial Transcriptomics Data”

High throughput spatial transcriptomics (HST) is a rapidly emerging experimental technology that allows for spatially resolved gene expression profiling at the single cell level. With HST data, we seek to identify sub-populations within a tissue sample that reflect biological cell types or states. Existing methods ignore the spatial dependence in gene expression, fail to account for features such as skewness and heavy-tails, or are heuristic-based methods that lack the benefits of statistical models. To address this gap, we develop SPRUCE: a Bayesian spatial multivariate finite mixture model based on multivariate skew-normal and skew-t distributions, to identify sub-populations in HST data. We implement a novel combination of Pólya-Gamma data augmentation and spatial random effects to infer spatially correlated mixture component membership probabilities. We evaluate the performance of SPRUCE through comprehensive simulation studies and its application to mouse brain HST data. The R package spruce, an efficient R implementation of the proposed models, is currently available from our research group GitHub repository (https://dongjunchung.github.io/spruce/).

Emily Hoskins
“Detecting and characterizing gene fusions utilizing large data”

Gene fusions result in a chromosomal rearrangement involving two distinct genes to form a new chimeric product. This genomic alteration is known to drive cancer, including by enhancing oncogenic activity or inhibiting tumor suppressors. Gene fusions are present in nearly 5% of reported cancer cases. Because most fusions are absent in healthy tissue, they are one of the most clinically relevant biomarkers in cancer events that can be targeted and treated. Although next generation sequencing and bioinformatic innovations have led to the discovery of many gene fusions that can be targeted with specific cancer therapies, the full landscape of gene fusions remains incomplete. Therefore, we are using large sources to identify and characterize gene fusion events in cancer. In our collaboration with the Oncology Research Information Exchange Network (ORIEN), we have access to RNA-seq, whole exome, and clinical data from 11,000 tumors across several participating institutions, including Ohio State. We propose to utilize this large data source to improve gene fusion detection methods in RNA-seq data by characterizing the clinical and genomic associations of gene fusion events, as well as build a machine learning model to select for true positive cases. We anticipate that this study will lead to 1) the discovery of novel gene fusion events, 2) enhance understanding of the role of gene fusions in cancer, and ultimately 3) help patients harboring gene fusions match to available targeted therapies.

Jacqueline Penaloza
“Constraint on single nucleotide variants disrupting long non-coding RNA secondary structure”
Long non-coding RNAs (lncRNAs) are the most abundant non-coding RNA and has been implicated in cancer, neurological, and cardiac diseases. Next Generation Sequencing (NGS) technologies have revolutionizing patient diagnosis, especially in the case of yet to be diagnosed or rare genetic conditions. Falling sequencing costs have enabled whole genome sequencing (WGS) to become routine, allowing for exploration of the genome beyond just the protein-coding regions. However, variants in non-coding regions are difficult to interpret and there is a critical need for lncRNA variant annotation and classification tools.

The secondary structure of lncRNA plays an essential role in the molecular function of the molecule. As such we hypothesize that, even on a local scale, genetic variation impacting lncRNA secondary structure could be an indicator of pathogenicity. To test this hypothesis, we took advantage of big-data technology to model lncRNA structures for the entire transcriptome. We evaluated single nucleotide variants (SNVs) at every position in the transcriptome and discovered that variants impacting lncRNA structure are under significant constraint in the human population. We are adapting our previously developed machine learning classifier to rank lncRNA variants by their stability disruption. Using our cohorts of 3,000 genome sequenced Congenital Heart Disease (CHD) cases, our goal is to identify lncRNA that are potentially linked to CHD due to loss of integrity in the lncRNA secondary structure.

Nida Viquar  
“Informing Development of a Mobile Health Application for Outpatient Parenteral Antibiotic Therapy”

The purpose of this paper is twofold: First, to examine the role of needs analysis in the development of mobile health apps, both generally, and specifically in the areas of cardio-oncology and infectious disease. This is in order to inform future work in these areas. These specific topics were chosen in consultation with a broader project team who is interested in developing mobile phone apps for cardio-oncology and outpatient parenteral antibiotic therapy (OPAT). Second, to conduct a needs analysis of infectious disease providers to inform development of a mobile health app on OPAT. I conducted interviews with healthcare providers to better understand their perceptions and preferences of what features would be beneficial to have on an OPAT mobile app. After interviews were conducted and transcribed, coding of the transcriptions was done to analyze the interviews qualitatively. Cohen's Kappa was calculated to determine agreement between coders. The IRB at The Ohio State University approved this study. The Cohen's Kappa was calculated to be 72.5%. All three participants had similar responses to questions when completing their interviews. Following the interviews, coding, and calculating the agreement variable, it was clear to see the emergence of five major, overall themes. These themes are the following: (1) the description of OPAT and the providers’ role, (2) patient and disease factors, (3) positive application features, (4) challenges to integration, and (5) the role of the EHR. These themes greatly informed development of an app for OPAT patients by providing valuable findings. In this study I was able to understand provider needs for a mobile health application for OPAT patients. These findings can potentially be used to inform development for other fields of medicine such as cardio-oncology. Future work should also include the perspectives of patients to design and develop a mobile health application.

Friday, April 16th, 11:00am-12:00pm  
Carmen Zoom