SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

CENTER FOR BIOINFORMATICS RESEARCH TALK

Michael Schatz

Johns Hopkins University

Thursday, March 28, 2019

3:00 pm

State Room East (IMU)

In pursuit of perfect genomes

Abstract: On April 14, 2003, after more than ten years of work and billions of research dollars spent, the human genome project was declared completed. With it the complete set of genes and all other genetic information for our species was known for the first time. In the years since, countless studies have benefited from this tremendous resource to study human evolution, human biology, and disease.

As important and wondrous this map has been, it suffers from two major shortcomings. Firstly, the human reference genome is not actually complete, and millions of nucleotides still remain undetermined and other regions are incorrectly represented. Secondly, the reference human genome doesn't actually represent any specific human, which can distort the interpretation of other genome sequences since we all carry millions of differences from the reference.

Fortunately, new single molecule DNA sequencing and mapping biotechnologies are beginning to make it possible for us to move away from reference genomes and move towards personalized genomes for everyone. In this presentation, I'll describe the biotechnologies and algorithmic approaches required to assemble these personal genomes. This will begin with the signal processing that is needed to convert raw electrical and optical signals produced during DNA sequencing into their underlying nucleotide sequences. Next the individual nucleotide sequences are compared to each other and compared to the reference genome to identify and phase the variants present in each person. Here we will focus on the intrinsic data characteristics, especially sequencing error rates, sequencing lengths, and oversampling needed to assemble high-quality personal genomes. Finally, I'll discuss some of the approaches and open challenges for interpreting genome sequences using comparative and functional genomics, with a focus on improved analysis of cancer genomes.

Biography: Michael Schatz is the Bloomberg Distinguished Associate Professor of Computer Science and Biology at Johns Hopkins University. His research is at the intersection of computer science, biology, and biotechnology, and focuses on development of novel algorithms and systems for comparative genomics, human genetics, and personalized medicine. In 2015, Schatz received the Alfred P. Sloan Foundation Fellowship to develop computational methods to probe the genetic components of autism and cancer, and in 2014 Schatz received the NSF CAREER award to develop computational methods to study plant and animal genomes using new long-read single molecule DNA sequencing technologies. Schatz joined JHU in 2016, after spending 6 years at Cold Spring Harbor Laboratory where he remains an Adjunct Associate Professor of Quantitative Biology. Schatz received his Ph.D. and M.S. in Computer Science from the University of Maryland in 2010 and 2008, his B.S. in Computer Science from Carnegie Mellon University in 2000, and spent 5 years at the Institute for Genomic Research (TIGR) in between. More information is available on his lab website: http://schatz-lab.org

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING