

# EPSTEIN INSTITUTE SEMINAR ▪ ISE 651

## Applying Long Read DNA Sequencing for Genome Assembly and Transcriptome Analysis: Existing Solutions and Open Problems

**ABSTRACT** – Third-generation DNA sequencing technologies such as the Pacific Biosciences single molecule, real-time sequencing technology can generate long (10 kb – 80 kb) but less accurate (10 – 15% error rate) reads, presenting a new kind of challenge for applying them to assembling a whole genome or a whole transcriptome. Previous existing algorithms suited for second-generation DNA sequencing technology, which are characterized by short (50 – 400 bases) but accurate ( $\geq 99\%$ ) DNA reads, were inadequate for this new kind of long sequence data. As a result, long read data have spawned a series of new tools for performing sequence alignments, assembling genomes, phasing (for example, separating the two copies of chromosomes in a human), identifying alternatively transcribed genes (transcriptome analysis), and visualizing structural variants that might be disease-causing. This new explosion of data is enabling medical and biological researchers unprecedented insight into human genetic diseases, bacterial infections, crop breeding, animal conservation, and more. I will present the algorithmic challenges, current solutions, and what are still some open problems that may interest the community.



**Dr. Elizabeth Tseng**

Bioinformatics Senior Staff Scientist,  
Pacific Biosciences

**SPEAKER BIO** – Dr. Elizabeth Tseng received her Ph.D. in Computer Science & Engineering from University of Washington in 2012. She received her B.S. in Computer Science from National Taiwan University in 2004. She joined Pacific Biosciences in 2012 as a bioinformatics scientist, with a focus on developing algorithms for applying PacBio's SMRT Sequencing technology for transcriptome analysis. Her algorithm for error correcting transcriptome data, ICE, is part of PacBio's official software that has been used by customers worldwide.

**USC Viterbi**  
School of Engineering  
*Daniel J. Epstein Department of  
Industrial and Systems Engineering*

**TUESDAY, APRIL 4, 2017**

**3:30PM – 4:50PM**

USC ANDRUS GERONTOLOGY CENTER (GER), Room 206