

**University of California, Irvine
Statistics Seminar**

***Effective Representation Learning to Dissect the Gene
Regulatory Grammar***

**Jing Zhang
Assistant Professor of Computer Science
UC Irvine**

**4 p.m., Thursday, December 9, 2021
Join via Zoom: <https://tinyurl.com/4azdp5zh>**

The recent advances in sequencing technologies provide unprecedented opportunities to decipher the multi-scale gene regulatory grammars at diverse cellular states. Here, we will introduce our computational efforts on cell/gene representation learning to extract biologically meaningful information from high-dimensional, sparse, and noisy genomic data. First, we proposed a deep generative model, named SAILER, to learn the low-dimensional latent cell representations from single-cell epigenetic data for accurate cell state characterization. SAILER adopted the conventional encoder-decoder framework and imposed additional constraints for biologically robust cell embeddings invariant to confounding factors. Then at the network level, we developed TopicNet using latent Dirichlet allocation (LDA) to extract latent gene communities and quantify regulatory network connectivity changes (network “rewiring”) between diverse cell states. We applied our TopicNet model on 13 different cancer types and highlighted gene communities that impact patient prognosis in multiple cancer types.