Department of Biostatistics and Medical Informatics Seminar

Anshul Kundaje, PhD

Assistant Professor
Department of Genetics and of Computer Science
Stanford University

Friday, April 22, 2022
12:00-1:00 pm
Via Zoom:
https://uwmadison.zoom.us/j/99912142921?pwd=QS9lWUpNanR0WnFJNFNYTnFLZktpQT09

Deep learning oracles for genomic discovery

Abstract: The human genome sequence contains the fundamental code that defines the identity and function of all the cell types and tissues in the human body. Genes are functional sequence units that encode for proteins. But they account for just about 2% of the 3 billion long human genome sequence. What does the rest of the genome encode? How is gene activity controlled in each cell type? Where do the regulatory control elements lie and what is their sequence composition? How do variants and mutations in the genome sequence affect cellular function and disease? These are fundamental questions that remain largely unanswered. The regulatory code that controls gene activity is encoded in the DNA sequence of millions of cell type specific regulatory DNA elements in the form of functional sequence syntax. This regulatory code has remained largely elusive despite exciting developments in experimental techniques to profile molecular properties of regulatory DNA. To address this challenge, we have developed high performance neural networks that can learn de-novo representations of regulatory DNA sequence to map genome-wide molecular profiles of protein DNA interactions and chromatin state at single base resolution across diverse cellular contexts. We have developed methods to interpret DNA sequences through the lens of the models and extract local and global predictive syntactic patterns revealing many insights into the regulatory code. Our models also serve as in-silico oracles to predict the effects of natural and disease-associated genetic variation i.e. how differences in DNA sequence across healthy and diseased individuals are likely to affect molecular mechanisms associated with common and rare diseases. These models enable optimized design of genome perturbation approaches to decipher functional properties of DNA and variants and serve as a powerful lens for genomic discovery.

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN–MADISON