

**Speech Title:** Big Patterns and Genomic Sequence Compression

**Abstract:** Contemporary sequencing platforms and machines have generated massive amounts of genomic sequence data, which have brought up huge challenges to data storage and transmission. High-performance compression is the only way to mitigate this heavy issue. In this talk, I will present: data mining techniques to identify big patterns from genomic sequence data sets, unsupervised clustering methods to group human genome sequences, and lossless compression algorithms to concisely represent the raw data. Here, the data size is in the magnitude of terabytes/gigabytes, and the compression performance can reach to thousands of folds file size reduction.