New probabilistic models for comparing genome organization and function

Yang Yang
School of Computer Science, Carnegie Mellon University

In human and other eukaryotes, chromosomes are spatially folded in the cell nucleus. The 3D genome organization is closely related to vital genome functions such as DNA replication timing (RT) and transcription. However, algorithms that facilitate cross-species pattern recognition of 3D genome organization are underexplored, limiting our understanding on how higher-order genome structure has evolved. In this talk, I will introduce two new probabilistic models: Phylo-HMGP and Phylo-HMRF, both of which jointly model the temporal dependencies across species and the spatial dependencies among genomic loci. The methods provide the first generic frameworks to identify genome-wide evolutionary patterns on 1D and 3D multi-species continuous genomic signals, respectively. Real data applications to multi-species RT data and Hi-C data offered high-resolution characterization of evolutionary patterns of 3D genome organization and function. Together, the methods have the potential to help reveal genomic regions with conserved or species-specific structural and regulatory roles, and provide key insights into genome organization and function through cross-species comparisons.