MIT Bioinformatics Seminar

Deep learning approaches to computational mass spectrometry

Speaker: Haixu Tang, Indiana University

Date: Monday, Dec. 12, 2022

Zoom URL: `https://mit.zoom.us/j/92936688687`

As a high throughput technology for molecular analyses, tandem mass spectrometry can rapidly generate the spectra of fragment ions from peptides or small molecules, featuring distinctive patterns and signals associated with molecular structures. As such, computational methods play a critical role in identifying the molecules from their tandem mass (MS/MS) spectra, in particular in large-scale proteomics and metabolomics studies. In the past decade, deep learning algorithms have demonstrated their power for discovering complex rules and patterns automatically from massive observational or experimental scientific data. In my talk, I will introduce several deep learning approaches to deciphering the complex patterns and rules in MS/MS spectra of molecules, which are subsequently used for the molecular identification. I will first present PredFull, a deep neural network for predicting the full MS/MS spectra from peptide sequences, and will then present PepNet to tackle the inverse problem,

which uses a fully convolutional neural network for de novo peptide sequencing from their MS/MS spectra. Finally, I will introduce Mol3DNet, a point-based deep neural network to predict the MS/MS spectra of small molecules from their three-dimensional (3D) conformations. These tools were trained and evaluated on massive experimental MS/MS spectra and achieved state-of-the-art performance on the respective tasks. These results again supported that deep learning algorithms can discover complex patterns and rules from massive experimental data, and thus are very helpful for scientific research.