

# SEMINAR

Fri, 28 Jan 2022 | 10 am | Online Zoom Session

Hosted by Prof Prakash Kumar



## By Lu Yang

Department of Genome Sciences,  
University of Washington, Seattle WA

*Yang Lu is a postdoctoral researcher in Prof. William Noble's group at the University of Washington. Prior to that, he obtained his Ph.D. in Computational Biology and Bioinformatics under the supervision of Prof. Fengzhu Sun from University of Southern California.*

*Before moving to the United States, he received M.S. and B.S. degrees in Computer Science and Engineering from Shanghai Jiao Tong University. Yang Lu's research focuses on developing machine learning and statistical methods for genomics and proteomics data analysis. He is particularly interested in developing interpretation methods to find scientifically interesting and statistically confident hypotheses from complex biological data.*

## Panning for gold: Interpretable and error-controlled pattern discovery from biomedical data

Rapid developments in high-throughput sequencing have enabled biologists to collect large volumes of multi-omics data with unprecedented resolution. However, interpretation of such an increasing amount of heterogeneous biological data becomes highly nontrivial. In my talk, I will present a data-driven research paradigm to discover testable hypotheses directly from biological data in an interpretable and error-controlled fashion. In particular, the talk will mainly focus on three recent works that span the critical components to biomedical research: data collection, hypothesis generation, and hypothesis evaluation: (1) A state-of-the-art peptide identification method in liquid chromatography-mass spectrometry-based Proteomics. This method demonstrated to the community for the first time that protein identification can achieve both sensitivity and reproducibility. (2) An interpretation method that generates testable biological hypotheses from deep learning models. Specifically, I developed an uncertainty-aware method to identify from single-cell RNA-seq data a combinatorial gene set signature to characterize the single-cell type. (3) A statistical method that subjects the hypotheses generated from deep learning models to error control. This method demonstrated to the community for the first time that the interpretation of deep learning models could achieve statistical guarantees.

Join Zoom Meeting | [LINK](#)  
Meeting ID: 884 6821 5497  
Meeting Password: 715191



### Virtual Seminar Etiquette:

- ✓ Please “mute” upon arrival into the meeting room.
- ✓ Questions can be asked after the presentation. You are encouraged to verbally ask questions or submit your questions via chat group.
- ✓ By being present at this meeting, **information presented is a privilege** and you agree that you would **NOT UNDERTAKE** any forms of recording/photo-taking.