YI-YUAN LEE

(412)537-8370 yl759@cornell.edu liyiyuian

EDUCATION

Cornell University — Ithaca, NY
Ph.D. in Computational Biology — Expected 2025

Carnegie Mellon University — Pittsburgh, PA
M.S. in Computational Biology, School of Computer Science — May 2020

National Taiwan University — Taipei, Taiwan
B.S. in Biochemical Science and Technology — June 2014

PUBLICATIONS

- **Yi-Yuan Lee** et. al, "hypoNPAtlas: an atlas of hypothetical natural product for mass spectrometry database search". [in review at Nature Communications, 2022]
- Jung-Lin Wu et. al, including **Yi-Yuan Lee**, "Phosphoproteomics Reveals the Role of Constitutive KAP1 Phosphorylation by B-Cell Receptor Signaling in Chronic Lymphocytic Leukemia", Molecular Cancer Research, 2022.
- Liu Cao, Mustafa Guler, Azat Tagirdzhanov, **Yi-Yuan Lee**, Alexey Gurevich, Hosein Mohimani, "MolDiscovery: Learning Mass Spectrometry Fragmentation of Small Molecules", Nature Chemical Biology, 2021.
- Michelle et. al, including **Yi-Yuan Lee**, "A community resource for paired genomicand metabolomic data mining", Nature Chemical Biology, 2021.
- W.C. Su, S.-F. Hsu, **Y.-Y. Lee**, et al., "A Nucleolar Protein, Ribosomal RNA Processing 1 Homolog B (RRP1B), Enhances the Recruitment of Cellular mRNA in Influenza Virus Transcription", Journal of Virology, 2015.

PATENT AND AWARD

- Behsaz Bahar et. al, including **Yi-Yuan Lee**, "System for Identifying Structures of Molecular Compounds from Mass Spectrometry Data", U.S. Patent application 20220208540, June 30 2022 (pending)
- Part of the DOE award, "DE-SC0021340: Discovery of Signaling Small Molecules (e.g. quorum sensing molecules) from the Microbiome, PI: Mohimani, Hosein. (2021)".

PROJECTS

Generating Novel Food Molecules using VAE-GAN — Carnegie Mellon University
Course: Introduction to Deep Learning (Group Project) — Fall 2019

- Adapted VAE-GAN techniques with PyTorch for *de novo* cancer drug discovery and character RNN to the discovery of food compounds.
- Addressed the challenge of small food compounds data sets by using data augmentation techniques while maintaining novelty, chemical synthesizability and validity.
- Generated food compounds with 99.7% validity and 87.7% novelty.

SKILLS

Programming Language	Python, C, C++, Go, Bash, Rust
Bioinformatics	BLAST, BWA, SAMTools, Scanpy/Seurat, HMMER
Deep Learning	PyTorch, TensorFlow, Scipy, Numpy, Panda, Sklearn, modAL, RDkit
Development Tools	Linux, AWS, Git, Slurm, LaTeX, VIM, tmux

RECENT WORK EXPERIENCE

Department of Computational Biology, Carnegie Mellon University — Pittsburgh, PA
Graduate Research Assistant — Sep 2018 - May 2021

- Developed deep neural networks for discovering novel ribosomally synthesized and post-translationally modified peptides (RiPPs), a class of natural products from microbial genomes. Both models are written in PyTorch and outperform the-state-of-the-art models in similar tasks.
- Implemented a subgraph-isomorphism-based chemical structure predictor, which generates hypothetical structures given a core peptide and a list of tailoring enzymes. Written in C++ and Rust.
- Mentoring three undergraduate students in research of *in silico* natural product discovery.

Genomics Research Center, Academia Sinica — Taipei, Taiwan
Research Assistant — Aug 2015 - May 2018

- Cooperated with research teams to interpret phosphoproteomics data, identify critical kinase in CLL formation and perform drug-resistance study.
- Identified the role of alpha-2,8-sialyltransferase 6 in acute inflammatory response.