

# Yang Lu

## Curriculum Vitae

Cheriton School of Computer Science, University of Waterloo  
200 University Avenue W. Waterloo, Ontario N2L 3G1

☎ (213) 880-9843

✉ [yanglu@uwaterloo.ca](mailto:yanglu@uwaterloo.ca)

📁 [batmen-lab.github.io](https://github.com/batmen-lab)

## RESEARCH INTERESTS

- 1 **Machine Learning.**
  - Interpretable machine learning
  - Error-controlled machine learning
  - AI for science/healthcare
- 2 **Computational Biology and Bioinformatics.**
  - Computational genomics/proteomics/microbiomics
  - Multi-omics integration
  - Bioinformatics analysis using large-language models

## EDUCATION

- 2017 **Ph.D. in Computational Biology and Bioinformatics.**  
University of Southern California, Los Angeles, CA, USA  
Thesis: "Big Data Analytics in Metagenomics: Integration, Representation, Management, and Visualization".  
Advisor: Dr. Fengzhu Sun
- 2013 **M.S. in Computer Science and Engineering.**  
Shanghai Jiao Tong University, Shanghai, China
- 2010 **B.S. in Computer and Software Engineering.**  
Shanghai Jiao Tong University, Shanghai, China

## PROFESSIONAL APPOINTMENTS

- 01/2023– **University of Waterloo**, Ontario, Canada.  
Present ◦ Assistant Professor at Cheriton School of Computer Science
- 01/2018– **University of Washington**, Seattle, WA, USA.  
11/2022 ◦ Postdoctoral Fellow at Department of Genome Sciences  
◦ Advisor: Dr. William S. Noble
- 09/2015– **University of Southern California**, Los Angeles, CA, USA.  
08/2017 ◦ Research Assistant at Department of Quantitative and Computational Biology  
◦ Advisor: Dr. Fengzhu Sun
- 06/2011– **Microsoft Research Asia**, Beijing, China.  
11/2011 ◦ Research Intern

## PUBLICATIONS

\* indicates equal contribution. † indicates the corresponding author.

## PUBLICATIONS AS LEADING OR CORRESPONDING AUTHOR

1. W Chen, Y Jiang, WS Noble, and YY Lu<sup>†</sup>. "Error-controlled non-additive interaction discovery in machine learning models". Under review, 2024.
2. Y Jiang, D Liao, Q Zhu, and YY Lu<sup>†</sup>. "PhyloMix: Enhancing microbiome-trait association prediction through phylogeny-mixing augmentation". bioRxiv:10.1101/2024.08.26.609661, 2024.
3. YY Lu, WS Noble, and U Keich. "A BLAST from the past: revisiting BLAST's E-value". bioRxiv:10.1101/2024.07.16.603405, 2024.
4. Z Dong, V Zhong, and YY Lu<sup>†</sup>. "BioMANIA: Simplifying bioinformatics data analysis through conversation". bioRxiv:10.1101/2023.10.29.564479, 2023.
5. Y Jiang, M Aton, Q Zhu<sup>†</sup>, and YY Lu<sup>†</sup>. "MIOSTONE: Modeling microbiome-trait associations with taxonomy-adaptive neural networks". bioRxiv:10.1101/2023.11.04.565596, 2023.
6. H Zhou, K Cao, and YY Lu<sup>†</sup>. "SONATA: Disambiguated manifold alignment of single-cell data". bioRxiv:10.1101/2023.10.05.561049, 2023.
7. W Chen, WS Noble<sup>†</sup>, and YY Lu<sup>†</sup>. "DeepROCK: Error-controlled interaction detection in deep neural networks". arXiv:2309.15319, 2023.
8. Z Ma\*, YY Lu\*<sup>†</sup>, Y Wang, R Lin, Z Yang, F Zhang, and Y Wang<sup>†</sup>. "Metric Learning for Comparing Genomic Data with Triplet Network". **Briefings in Bioinformatics**, 23(5):bbac345, 2022.
9. YY Lu and WS Noble. "A wider field of view to predict expression". **Nature Methods**, 18:1155-1156, 2021.
10. YY Lu\*, W Guo\*, X Xing, and WS Noble. "DANCE: Enhancing saliency maps using decoys". International Conference on Machine Learning (**ICML**), 2021.
11. YY Lu, T Yu, G Bonora, and WS Noble. "ACE: Explaining cluster from an adversarial perspective". International Conference on Machine Learning (**ICML**), 2021.
12. YY Lu, J Billes, RA Rodriguez-Mias, J Villen, and WS Noble. "DIAMeter: Matching peptides to data-independent acquisition mass spectrometry data". International Conference on Intelligent Systems for Molecular Biology (**ISMB**), 2021.
13. YY Lu\*, J Bai\*, Y Wang, Y Wang, and F Sun. "CRAFT: Compact genome Representation towards large-scale Alignment-Free daTabase". **Bioinformatics**, 37(2):155-161, 2020.
14. J Schreiber\*, YY Lu\*, and WS Noble. "Ledidi: Designing genomic edits that induce functional activity". **ICML Workshop on CompBio**, 2020.
15. YY Lu\*, Y Fan\*, J Lv, and WS Noble. "DeepPINK: reproducible feature selection in deep neural networks". Advances in Neural Information Processing Systems (**NeurIPS**), 8676-8686, 2018.
16. YY Lu. "Big Data Analytics in Metagenomics: Integration, Representation, Management, and Visualization". **Doctoral Thesis**. University of Southern California, 2017.
17. YY Lu, J Lv, JA Fuhrman, and F Sun. "Towards enhanced and interpretable clustering/classification in integrative genomics". **Nucleic Acids Research**, 45(20):e169, 2017.
18. YY Lu, K Tang, J Ren, JA Fuhrman, MS Waterman, and F Sun. "CAFE: aCcelerated Alignment-FrEe sequence analysis". **Nucleic Acids Research**, 45(W1):W554-W559, 2017.
19. YY Lu, T Chen, JA Fuhrman, and F Sun. "COCACOLA: binning metagenomic contigs using sequence COMposition, read CoverAge, CO-alignment and paired-end read LinkAge". **Bioinformatics**, 33(6):791-798, 2017.

## PUBLICATIONS AS SUPPORTING AUTHOR

1. C Ma, H Zhao, L Zheng, J Xin, Q Li, L Wu, Z Deng, YY Lu, Q Liu, and L Kong. "Retrieved Sequence Augmentation for Protein Representation Learning". bioRxiv:10.1101/2023.02.22.529597, 2023.
2. A Kertesz-Farkas, FL Nii Adoquaye Acquaye, K Bhimani, JK Eng, WE Fondrie, C Grant, MR Hoopmann, A Lin, YY Lu, RL Moritz, MJ MacCoss, and WS Noble. "The Crux Toolkit for Analysis of Bottom-Up Tandem Mass Spectrometry Proteomics Data". **Journal of Proteome Research**, 22(2):561-569, 2023.
3. AB Dincer, YY Lu, D Schweppe, S Oh, and WS Noble. "Reducing peptide sequence bias in quantitative mass spectrometry data with machine learning". **Journal of Proteome Research**, 21(7):1771-1782, 2022.
4. Z Wang, Z Wang, YY Lu, F Sun, and S Zhu. "SolidBin: Improving Metagenome Binning with Semi-

- supervised Normalized Cut". **Bioinformatics**, 35(21):4229-4238, 2019.
5. DF Read, K Cook, YY Lu, KG Le Roch, and WS Noble. "Predicting gene expression in the human malaria parasite *Plasmodium falciparum* using histone modification, nucleosome positioning, and 3D localization features". **PLoS Computational Biology**, 15(9):e1007329, 2019.
  6. A Hu, YY Lu, J Bilmes, and WS Noble. "Joint precursor elution profile inference via regression for peptide detection in data-independent acquisition mass spectra". **Journal of Proteome Research**, 18(1):86-94, 2018.
  7. J Ren, X Bai, YY Lu, K Tang, Y Wang, G Reinert, and F Sun. "Alignment-free sequence analysis and applications". **Annual Review of Biomedical Data Science**, 1:93-114, 2018.
  8. K Tang, YY Lu, and F Sun. "Background Adjusted Alignment-free Dissimilarity Measures Improve the Detection of Horizontal Gene Transfer". **Frontiers in Microbiology**, 9:711, 2018.
  9. Y Wang, K Wang, YY Lu, and F Sun. "Improving contig binning of metagenomic data using  $d_2^S$  oligonucleotide frequency dissimilarity". **BMC Bioinformatics**, 18(1):425, 2018.
  10. J Ren, NA Ahlgren, YY Lu, JA Fuhrman, and F Sun. "VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data". **Microbiome**, 5(1):69, 2017.
  11. NA Ahlgren, J Ren, YY Lu, JA Fuhrman, and F Sun. "Alignment-free  $d_2^*$  oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences". **Nucleic Acids Research**, 45(1):39-53, 2017.

## GRANTS

- 2023 – 2028 **NSERC Discovery Grant**, *Natural Sciences and Engineering Research Council of Canada*.
- Title: Uncertainty-aware integrative pattern discovery for single-cell analysis
  - Role: PI
  - Amount: \$167,500

## INVITED LECTURES

- 2023
- Amgen, Data Sciences
  - Western University, Bioinformatics Seminar
  - University of Ottawa, Technologies for Microbiome Science and Engineering seminar
- 2022
- HSE University, School of Data Analysis and Artificial Intelligence
  - University of Texas MD Anderson Cancer Center, Data Science and Modeling Forum
  - Chinese Academy of Sciences, Academy of Mathematics and Systems Science
  - University of Toronto, Donnelly Centre for Cellular and Biomolecular Research
  - University of Waterloo, Cheriton School of Computer Science
  - University of Oregon, Department of Computer and Information Science
  - Georgia Institute of Technology, School of Computational Science and Engineering
  - University of Notre Dame, Department of Computer Science and Engineering
  - University of Texas MD Anderson Cancer Center, Department of Bioinformatics and Computational Biology
  - Pennsylvania State University, Department of Computer Science and Engineering
  - University of Washington, Paul G. Allen School of Computer Science and Engineering
  - Drexel University, College of Computing and Informatics
  - National University of Singapore, Department of Biological Sciences
  - Indiana University School of Medicine, Herman B Wells Center for Pediatric Research
  - University of Florida, Department of Health Outcomes and Biomedical Informatics

- 2021
  - University of Hong Kong, Department of Computer Science
  - University of Miami, College of Engineering
  - Bristol Myers Squibb, Data Science
  - University of Washington, Genome Sciences Seminar
  - ACM-BCB, Machine-learning Methods for Single-cell Analysis Workshop
  - International Conference on Machine Learning (ICML)
  - International Conference on Intelligent Systems for Molecular Biology (ISMB)
- 2019
  - Cascadia Proteomics Symposium
  - ASMS Conference on Mass Spectrometry and Allied Topics
- 2016
  - Joint Statistical Meetings (JSM)

## TEACHING

- CS482/682 **Computational Techniques in Biological Sequence Analysis**, *University of Waterloo*.
- Winter 2023, 2024
  - Enrollment: 60 students
- CS341 **Algorithms**, *University of Waterloo*.
- Winter 2024
  - Enrollment: 180 students

## STUDENT MENTORING

- 2023 – present **Yifan Jiang**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario
  - Project: Modeling microbiome-trait associations with taxonomy-adaptive neural networks.
- 2023 – present **Han Zhou**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Project: Disambiguated manifold alignment of single-cell data.
- 2023 – present **Zhengyuan Dong**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Project: Simplifying bioinformatics data analysis through conversation.
- 2023 – present **Fatemeh Alipour**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Co-supervised with Prof. Lila Kari
- 2023 – 2024 **Zhenbo Li**, *MMath student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Co-supervised with Prof. Bin Ma
- 2019 – 2022 **Winston Chen**, *Undergraduate student*.
- School of Computer Science and Engineering, University of Washington, Seattle, WA, USA
  - Project: Error-controlled interaction detection in deep neural networks.
- Fall 2020 **Timothy Yu**, *Rotational graduate student*.
- Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA
  - Project: Explaining single-cell cluster from an adversarial perspective.
- Fall 2020 **Chris Hsu**, *Rotational graduate student*.
- Department of Genome Sciences, University of Washington, Seattle, WA, USA
  - Project: Designing sequence edits to change RNA splicing.
- 2018 – 2020 **Jiaxing Bai**, *Master student*.
- Department of Automation, Xiamen University, Xiamen, China
  - Project: Learn the compact representations of metagenomic sequences.

- 2018 – 2020 **Yiwen Wang**, *Master student*.
- Department of Automation, Xiamen University, Xiamen, China
  - Project: Automatic taxonomy embedding and categorization by Siamese triplet network.
- 2016 – 2017 **Kun Wang**, *Master student*.
- Department of Automation, Xiamen University, Xiamen, China
  - Project: Improving metagenomic binning by using alignment-free similarity measures.

## THESIS COMMITTEES

- 2024 **Jingjing Wang**, *Master student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Thesis: "Multivariate Triangular Quantile Maps for Novelty Detection".
- 2024 **Shaokai Wang**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Thesis: "Deep Learning Methods for Novel Peptide Discovering and Function Prediction".
- 2023 **Pablo Millan Arias**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Thesis: "Deep Learning for Unsupervised Clustering of DNA Sequences".
- 2023 **Johra Moosa**, *PhD student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Thesis: "Improving Peptide Identification in Proteomics Data Analysis through Repeat-Preserving Decoy and Decoy-Free Retraining".
- 2023 **Shuyang Zhang**, *Master student*.
- Cheriton School of Computer Science, University of Waterloo, Ontario, Canada
  - Thesis: "Spectrum and Retention Time Prediction for N-Glycopeptides Using Deep Learning".

## AWARDS AND SCHOLARSHIPS

- 08/2021 **Most Innovative Presentation Award**.
- Machine-learning Methods for Single-cell Analysis Workshop, ACM-BCB 2021
- 2013–2015 **Provost's Fellowship**.
- University of Southern California, Los Angeles, CA
- 2010–2013 **Graduate Fellowship**.
- Shanghai Jiao Tong University, Shanghai, China
- 2008 **National Endeavor Scholarship**.
- Shanghai Jiao Tong University, Shanghai, China

## SERVICES

- SERVICE TO PROFESSION**
- Associate Editor.**
- PLOS Computational Biology, 2023–present
  - ACM Transactions on Probabilistic Machine Learning, 2023–present
- Review Editor.**
- Frontiers in Bioinformatics, Specialty Section of Single Cell Bioinformatics, 2022
- Area Chair.**
- ACM Conference on Bioinformatics, Computational Biology and Biomedicine (ACM-BCB), 2023
  - AI4Science Workshop, International Conference on Machine Learning (ICML), 2023

**Program Committee.**

- Machine Learning in Computational Biology (MLCB), 2021
- Machine-learning Methods for Single-cell Analysis Workshop, ACM-BCB, 2021
- ISMB/ECCB Conference for the Computational Mass Spectrometry (CompMS), 2021, 2022

**Journal Reviewer.**

- Annals of Applied Statistics, Bioinformatics, BMC Bioinformatics, Briefings in Bioinformatics, Frontiers in Genetics, Genome Biology, IEEE/ACM Transactions on Computational Biology and Bioinformatics, IEEE Transactions on Neural Networks and Learning Systems, Journal of Proteome Research, Nature Biotechnology, Nature Communications, Nature Machine Intelligence, Nature Methods, Nucleic Acids Research, PLOS Computational Biology

**Conference Reviewer.**

- AAAI, AISTATS, APBC, ICLR, ICML, ISMB, NeurIPS, RECOMB

SERVICE TO  
UNIVERSITY

**Graduate Committee.**

- Member, 2023–2024

**Nominating Committee for Director of School.**

- Member, 2023–2024

**Graduate Recruiting Committee.**

- Member, 2022–2023

---

## REFERENCES

Prof. **William S. Noble** (Postdoc advisor)  
Professor  
Department of Genome Sciences and  
School of Computer Science and Engineering  
University of Washington  
send.Noble.D3EF3353F6@interfoliodossier.com

Prof. **Uri Keich**  
Associate Professor  
School of Mathematics and Statistics  
University of Sydney  
send.Keich.E4268BFC98@interfoliodossier.com

Prof. **Ming Li**  
University Professor Emeritus  
School of Computer Science  
University of Waterloo  
mli@uwaterloo.ca

Prof. **Fengzhu Sun** (PhD advisor)  
Professor  
Department of Quantitative and Computational Biology and  
Mathematics  
University of Southern California  
send.Sun.41F38EBF06@interfoliodossier.com

Prof. **Michael S. Waterman**  
University Professor Emeritus  
Department of Quantitative and Computational Biology, Math-  
ematics, and Computer Science  
University of Southern California  
send.Waterman.A13F563250@interfoliodossier.com