

Yang Lu

Curriculum Vitae

3720 15th Ave NE
Seattle, WA 98195
☎ (213) 880-9843
✉ yanglu@uwaterloo.ca
📁 [batmen-lab.github.io](https://github.com/batmen-lab)

RESEARCH INTERESTS

Computational Biology

- Single-cell genomics
- Mass spectrometry-based proteomics
- Metagenomics

Machine Learning

- Interpretable machine learning
- Reproducible (error-controlled) machine learning
- Heterogeneous data integration

EMPLOYMENT

- 2022–Present **Assistant Professor**, *University of Waterloo*, Cheriton School of Computer Science.
- Lead BATMEN (BioinformAtics & Trustworthy Machine LEarNing) Lab .

EDUCATION

- 2013–2017 **Ph.D. in Computational Biology and Bioinformatics.**
University of Southern California, Los Angeles, CA
Thesis: *Big Data Analytics in Metagenomics: Integration, Representation, Management, and Visualization.*
Advisor: Prof. Fengzhu Sun
- 2010–2013 **M.S. in Computer Science and Engineering.**
Shanghai Jiao Tong University, Shanghai, China
- 2006–2010 **B.S. in Computer and Software Engineering.**
Shanghai Jiao Tong University, Shanghai, China

PROFESSIONAL EXPERIENCES

- 2018–2022 **Postdoctoral Fellow**, *University of Washington*, Seattle, WA.
Advisor: Prof. William S. Noble
- Developed a sensitive and reproducible peptide detection method to analyze mass spectrometry-based proteomics data.
 - Developed an uncertainty-aware differential gene expression analysis method to analyze single-cell genomics data.
 - Developed an optimization-based framework to design genome edits that induce a desired functional landscape.
 - Developed a theoretically and practically sound interpretation method for deep neural networks.
 - Developed an error-controlled variable selection method for deep neural networks.

- 2015–2017 **Research Assistant**, *University of Southern California*, Los Angeles, CA.
 Advisor: Prof. Fengzhu Sun
- Developed a general computational framework to group metagenomic contigs into operational taxonomic units.
 - Developed an accelerated and memory-efficient standalone software for the implementation of 28 alignment-free sequence comparison methods.
 - Developed a general metagenomic search engine to learn compact representations of sequences and perform fast comparisons.
 - Developed a data integration framework that learns to rescales features from heterogeneous data.
- 2011/06– **Research Intern**, *Microsoft Research Asia*, Beijing, China.
 2011/11 ○ Developed algorithms to monitor the service failure in the Office365 online service system and diagnose its potential root cause.
- 2010/11– **Software Engineer Intern**, *Microsoft*, Shanghai, China.
 2011/05 ○ Developed software tools to analyze the data produced from the Microsoft e-commerce transaction platform.
- 2010/04– **Software Engineer Intern**, *Tencent*, Shanghai, China.
 2010/08 ○ Developed the user interface of the mobile security manager, which offers various mobile services such as blocking spam calls, filtering unwanted messages, encrypting personal data, etc.
- 2009/08– **Software Engineer Intern**, *Ericsson*, Shanghai, China.
 2009/11 ○ Developed the user interface of a web-based social networking tool.

PUBLICATIONS

* indicates equal contribution. † indicates the corresponding author.

PUBLICATIONS AS LEADING OR CORRESPONDING AUTHOR

1. Z Ma*, YY Lu*†, Y Wang, R Lin, Z Yang, F Zhang, and Y Wang†. "Metric Learning for Comparing Genomic Data with Triplet Network". **Briefings in Bioinformatics**, 23(5):bbac345, 2022.
2. YY Lu and WS Noble. "A wider field of view to predict expression". **Nature Methods**, 18:1155-1156, 2021.
3. YY Lu*, W Guo*, X Xing, and WS Noble. "DANCE: Enhancing saliency maps using decoys". International Conference on Machine Learning (**ICML**), 2021.
4. YY Lu, T Yu, G Bonora, and WS Noble. "ACE: Explaining cluster from an adversarial perspective". International Conference on Machine Learning (**ICML**), 2021.
5. YY Lu, J Bilmes, 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.
6. 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.
7. J Schreiber*, YY Lu*, and WS Noble. "Ledidi: Designing genomic edits that induce functional activity". **ICML Workshop on CompBio**, 2020.
8. 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
9. YY Lu. "Big Data Analytics in Metagenomics: Integration, Representation, Management, and Visualization". **Doctoral Thesis**. University of Southern California, 2017.
10. 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.
11. 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.
12. 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. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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

- 2016 **Contributed to the R01 (5R01GM120624-04) funded by NIH/NIGMS.** .
- Computational Studies of Virus-host Interactions Using Metagenomics Data and Applications
 - PI: Fengzhu Sun and Nathan Ahlgren
 - My contribution: preliminary data generation and writing for one aim.

INVITED PRESENTATIONS

- 2022/10 **HSE University, Moscow, Russia.**
- Invited talk, School of Data Analysis and Artificial Intelligence
- 2022/05 **University of Texas MD Anderson Cancer Center, Houston, TX.**
- Invited talk, Data Science and Modeling Forum
- 2022/04 **Chinese Academy of Sciences, Beijing, China.**
- Invited seminar, Academy of Mathematics and Systems Science
- 2022/03 **University of Toronto, ON, Canada.**
- Invited seminar, Donnelly Centre for Cellular and Biomolecular Research
- 2022/02 **University of Waterloo, ON, Canada.**
- Invited seminar, Cheriton School of Computer Science
- 2022/02 **University of Oregon, Eugene, Oregon.**
- Invited seminar, Department of Computer and Information Science
- 2022/02 **Georgia Institute of Technology, Atlanta, GA.**
- Invited seminar, School of Computational Science and Engineering

- 2022/02 **University of Notre Dame, Notre Dame, IN.**
- Invited seminar, Department of Computer Science and Engineering
- 2022/02 **University of Texas MD Anderson Cancer Center, Houston, TX.**
- Invited seminar, Department of Bioinformatics and Computational Biology
- 2022/02 **Pennsylvania State University, State College, PA.**
- Invited seminar, Department of Computer Science and Engineering
- 2022/02 **University of Washington, Seattle, WA.**
- Invited guest lecture, Paul G. Allen School of Computer Science and Engineering
- 2022/02 **Drexel University, Philadelphia, PA.**
- Invited seminar, College of Computing and Informatics
- 2022/01 **National University of Singapore, Singapore.**
- Invited seminar, Department of Biological Sciences
- 2022/01 **Indiana University School of Medicine, Indianapolis, IN.**
- Invited seminar, Herman B Wells Center for Pediatric Research
- 2022/01 **University of Florida, Gainesville, FL.**
- Invited seminar, Department of Health Outcomes and Biomedical Informatics
- 2021/12 **University of Hong Kong, Hong Kong.**
- Invited seminar, Department of Computer Science
- 2021/12 **University of Miami, Miami, FL.**
- Invited seminar, College of Engineering
- 2021/09 **Bristol Myers Squibb.**
- Invited talk, Data Science
- 2021/08 **University of Washington, Seattle, WA.**
- Invited talk, Department of Genome Sciences
- 2021/08 **Machine-learning Methods for Single-cell Analysis Workshop, ACM-BCB 2021.**
- Contributed talk
 - Most Innovative Presentation Award
- 2021/07 **International Conference on Machine Learning (ICML).**
- Two contributed (Spotlight) talks
- 2021/07 **International Conference on Intelligent Systems for Molecular Biology (ISMB).**
- Two contributed talks
- 2021/04 **Probabilistic Modeling in Genomics Meeting, Cold Spring Harbor Laboratory.**
- Contributed poster
- 2020/07 **International Conference on Intelligent Systems for Molecular Biology (ISMB).**
- Contributed poster
- 2020/06 **ASMS Conference on Mass Spectrometry and Allied Topics.**
- Contributed poster
- 2019/07 **Cascadia Proteomics Symposium, Seattle, WA.**
- Contributed talk
- 2019/06 **ASMS Conference on Mass Spectrometry and Allied Topics, Atlanta, GA.**
- Contributed poster
- 2018/12 **Conference on Neural Information Processing Systems (NeurIPS), Montreal, CA.**
- Contributed poster
- 2016/08 **Joint Statistical Meetings (JSM), Chicago, IL.**
- Contributed talk

- 2015/04 **RECOMB Satellite Workshop on Massively Parallel Sequencing (RECOMB-Seq)**,
Santa Monica, CA.
- Contributed talk
- 2014/11 **University of Southern California**, *Los Angeles, CA.*
- Invited talk, MCB Retreat

TEACHING

- Fall 2015 **Teaching Assistant**, *BISC305 – Statistics for the Biological Sciences.*
- Department of Biological Science, University of Southern California
 - Designed quizzes and exams, graded assignments, and hosted review sessions for the whole class.
- Spring 2012 **Teaching Assistant**, *Introduction to Artificial Intelligence.*
- Department of Computer Science and Engineering, Shanghai Jiao Tong University
 - Graded assignments and taught students during office hours.
- Fall 2011 **Teaching Assistant**, *Introduction to Bioinformatics.*
- Department of Computer Science and Engineering, Shanghai Jiao Tong University
 - Graded assignments and taught students during office hours.

MENTORING

- 2019 – present **Winston Chen**, *undergraduate student.*
- School of Computer Science and Engineering, University of Washington, Seattle, WA
 - 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
 - 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
 - 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.

AWARDS AND SCHOLARSHIPS

- 2021/08 **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

SERVICE

- 2022/11–present **Associate Editor.**
 - ACM Transactions on Probabilistic Machine Learning
- 2022/11–present **Review Editor.**
 - Frontiers in Bioinformatics, Specialty Section of Single Cell Bioinformatics
- 2021/11 **Program Committee.**
 - Machine Learning in Computational Biology (MLCB), 2021
- 2021/08 **Program Committee.**
 - Machine-learning Methods for Single-cell Analysis Workshop, ACM-BCB 2021
- 2021/07 **Program Committee.**
 - ISMB/ECCB 2021 conference for the Computational Mass Spectrometry (CompMS)
- 2020–present **Member.**
 - American Society for Mass Spectrometry (ASMS)
- 2020–present **Member.**
 - US Human Proteome Organization (US HUPO)
- All time **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, Journal of Proteome Research, Nature Communications, Nature Machine Intelligence, Nature Methods, Nucleic Acids Research, PLOS Computational Biology
- All time **Conference Reviewer.**
 - AAAI, AISTATS, APBC, ICLR, ICML, ISMB, NeurIPS, RECOMB