# ALEX LU

lualex@microsoft.com Senior Researcher Microsoft Research New England

## **RESEARCH INTERESTS**

My research interests center around machine learning methods for the discovery of new hypotheses in cellular and molecular biology.

My specific research interests include:

- Unsupervised methods for rapid and unbiased exploration of data without labels or prior knowledge
- Robust methods that can cope with novel phenotypes and conditions not seen during training
- Methods for biological images, which are understudied compared to other modalities

My research website and projects can be found at alexluresearch.com

#### PUBLICATIONS

## Peer Reviewed Publications

- Wu KE, Yang KK, Berg RV, Zou JY, Lu AX, Amini AP. Protein structure generation via folding diffusion. Nat Commun 15, 1059. 2024 Feb 5.
- Yang KK, Fusi N, Lu AX. Convolutions are competitive with transformers for protein sequence pretraining. Cell Systems. Feb 2024.
- Desai A, Berger L, Minakov FO, Milan V, Singh C, Pumphrey K, Ladner RE, Daumé III H, Lu AX, Caselli N, Bragg D. ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Recognition. NeurIPS 2023. 2023 Dec.
- Falahkheirkhah K, Lu AX, Alvarez-Melis D, Huynh G. Domain adaptation using optimal transport for invariant learning using histopathology datasets. MIDL 2023. 2023 Jul.
- Hua SB, Chen I, Lu AX, Erdman L. Supervised contrastive learning for improved view labeling in pediatric renal ultrasound videos. 2023 April 18. ISBI 2023.
- Lin A, Lu AX. Incorporating knowledge of plates in batch normalization improves generalization of deep learning for microscopy images. In Machine Learning in Computational Biology 2022 Dec 19 (pp. 74-93). PMLR.
- Lu AX, Lu AXP, Pritisanac I, Zarin T, Forman-Kay JD, Moses AM. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. PLOS Computational Biology. 2022 Jun 29.
- Wang A, Amini AP, Lu AX\*, Yang KK. Learning from physics-based features improves protein property prediction. MLSB 2022 workshop.
- Hua SB, Lu AX, Moses AM. CytoImageNet: A large-scale pretraining dataset for bioimage transfer learning. LMRL at NeurIPS 2021.
- Yang K, Goldman S, Jin W, Lu A, Barzilay R, Jaakkola T, Uhler C. Improved Conditional Flow Models for Molecule to Image Synthesis. CVPR 2021.

- Lu AXP, Lu AX, Moses, AM. Evolution Is All You Need: Phylogenetic Augmentation for Contrastive Learning. MLCB 2020.
- Lu AX, Lu AXP, Schormann W, Andrews DW, Moses AM. The Cells Out of Sample (COOS) dataset and benchmarks for measuring out-of-sample generalization of image classifiers. NeurIPS 2019. 2019 December.
- Lu AX, Kraus OZ, Cooper S, Moses AM. Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. PLOS Computational Biology. 2019 Sept 3;15(9):e1007348.
- Lu AX, Zarin T, Hsu IS, Moses AM. YeastSpotter: accurate and parameter-free web segmentation for microscopy images of yeast cells. Bioinformatics. 2019 May 16.
- Lu AX, Chong YT, Hsu IS, Strome B, Handfield LF, Kraus O, Andrews BJ, Moses AM. Integrating images from multiple microscopy screens reveals diverse patterns of change in the subcellular localization of proteins. Elife. 2018 Apr 5;7:e31872.
- Lu AX, Handfield LF, Moses A. Extracting and integrating protein localization changes from multiple image screens of yeast cells. Bio-Protocol. 2018 Apr;8(18).
- Huisman E, Lu AX, Jamil S, Mousavizadeh R, McCormack R, Roberts C, Scott A. Influence of repetitive mechanical loading on MMP2 activity in tendon fibroblasts. Journal of Orthopaedic Research. 2016 Nov;34(11):1991-2000.
- Lu AX, Moses AM. An unsupervised kNN method to systematically detect changes in protein localization in high-throughput microscopy images. PloS one. 2016 Jul 21;11(7):e0158712.
- Mousavizadeh R, Scott A, Lu AX, Ardekani GS, Behzad H, Lundgreen K, Ghaffari M, McCormack RG, Duronio V. Angiopoietin-like 4 promotes angiogenesis in the tendon and is increased in cyclically loaded tendon fibroblasts. The Journal of physiology. 2016 Jun 1;594(11):2971-83.
- Grewal N, Thornton GM, Behzad H, Sharma A, Lu AX, Zhang P, Reid WD, Granville DJ, Scott A. Accumulation of oxidized LDL in the tendon tissues of C57BL/6 or apolipoprotein E knockout mice that consume a high fat diet: potential impact on tendon health. PLoS One. 2014 Dec 12;9(12):e114214.
- Huisman E, Lu AX, McCormack RG, Scott A. Enhanced collagen type I synthesis by human tenocytes subjected to periodic in vitro mechanical stimulation. BMC musculoskeletal disorders. 2014 Dec;15(1):386.
- Debruin EJ, Hughes MR, Sina C, Lu AX, Cait J, Jian Z, Lopez M, Lo B, Abraham T, McNagny KM. Podocalyxin regulates murine lung vascular permeability by altering endothelial cell adhesion. PloS one. 2014 Oct 10;9(10):e108881.
- Behzad H, Sharma A, Mousavizadeh R, Lu AX, Scott A. Mast cells exert pro-inflammatory effects of relevance to the pathophyisology of tendinopathy. Arthritis research & therapy. 2013 Dec;15(6):R184.

## **Pre-prints**

- Li FZ, Amini AP, Yue Y, Yang KK, Lu AX. Feature reuse and scaling: Understanding transfer learning with protein language models. 2024 Feb 7.
- Kedzierska KZ, Crawford L, Amini AP, Lu AX. Assessing the limits of zero-shot foundation models in single-cell biology. bioRxiv preprint. 2023 Oct 17.
- Alamdari S, Thakkar N, van den Berg R, Lu AX, Fusi N, Amini AP, Yang KK. Protein generation with evolutionary diffusion: sequence is all you need. bioRxiv preprint. 2023 Sep.

## **Book Chapters**

• Lu AX, Moses AMM. Localization velocity: visualizing changes in protein localization detected in high-throughput microscopy. May 2024. Contributed to Methods in Molecular Biology - Imaging Cell Signaling (Springer Nature).

# CONFERENCE PRESENTATIONS AND INVITED TALKS

## Invited Talks

- Towards Meaningful Pretrained Models for Biology Models, Algorithms and Inference, Broad Institute Broad Institute, Cambridge, MA, October 24th, 2023
- Understanding big biological data with self-supervised machine learning EWSC Initiative, Broad Institute Broad Institute, Cambridge, MA, March 17th, 2022
- Understanding big biological data with self-supervised machine learning University of Rochester Virtual Seminar, March 10th, 2022
- Understanding biomedical images with self-supervised machine learning Spring Discovery Virtual Seminar, October 13th, 2021
- Learning biology through puzzle-solving: unbiased automatic understanding of microscopy images through self-supervised learning Microscopy & Microanalysis 2020 Virtual Conference, August 4th, 2020
- Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting

Cell Painting Group, Broad Institute Virtual Seminar, April 6th, 2020

## **Contributed Oral Presentations**

- Discovering interpretable features of the intrinsically disordered dark proteome with contrastive learning ISMB 2022 Madison, Wisconsin, July 2022
- Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting MLCB 2019 Vancouver, Canada, December 14th, 2019
- Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting

Biological Data Science 2018 Cold Spring Harbor Laboratory, Long Island, New York, November 10th, 2018 • An unsupervised kNN method to systematically detect changes in protein localization in high-throughput microscopy images GLBIO/CCBC 2016 Toronto, Canada, May 16th, 2016

#### **Poster Presentations**

- Transfer Learning vs. Batch Effects: what can we expect from neural networks in computational biology?
  MLCB 2019
  Vancouver, Canada, December 14th, 2019
- The Cells Out of Sample (COOS) dataset and benchmarks for measuring out-ofsample generalization of image classifiers. NeurIPS 2019 Vancouver, Canada, December 11th, 2019
- Learning representations of biology with small and homogeneous training datasets NeurIPS 2019 (Learning Meaningful Representations of Life Workshop) Vancouver, Canada December 13th, 2019
- Paired cell inpainting: self-supervised multiple-instance learning for bioimage analysis ICML 2019 (Self-Supervised Learning Workshop) Long Beach, California, June 16th, 2019

## **Campus Talks**

- Exploring big image data with self-supervised deep learning OICR Seminar Series Toronto, Canada, November 11th, 2019
- Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting TorBUG Seminar Series Toronto, Canada, November 28th, 2018
- Discovering biological insights from large-scale microscopy datasets with unsupervised deep learning Sunnybrook Research Institute Toronto, Canada, November 26th, 2018

## EDUCATION

## PhD (Computer Science)

University of Toronto Thesis: Unsupervised machine learning for hypothesis discovery and representation learning in biological image and sequence analysis.

# MSc (Computer Science)

University of Toronto

Thesis: An unsupervised kNN method to systematically detect changes in protein localization in high-throughput microscopy images.

2017 - 2021

2016 - 2017

# **BSc** (Honors Specialization in Bioinformatics)

University of Western Ontario

Thesis: Observations in extracting and parsing subcellular relations using natural language processing systems in biomedical publications.

# ACADEMIC WORK EXPERIENCE

# Senior Researcher Microsoft Research New England Leads research program at the intersection of machine learning and big biological data.

# Intern

Phenomic AI, Toronto Internship in a start-up. Research and development for deep learning methods for image segmentation.

**Teaching Assistant** Department of Computer Science, University of Toronto Teaching assistant for upper year artificial intelligence and computational linguistic courses.

# **Bioinformatics Undergraduate Research Assistant**

Robert HN Ho Research Center, University of British Columbia Undergraduate summer student funded by NSERC USRA. Investigated tendon pathology using computational analyses of immunohistochemistry images.

# TEACHING AND SERVICE

# **Committees**

- Communications Chair NeurIPS Communications chair for NeurIPS 2023
- Thesis Dissertation Committee University of Washington PhD thesis dissertation member for Aashaka Desai.
- June 2022 October 2022 • Microsoft PhD Fellowship Committee Microsoft Research Jointly led Health and Life Sciences committee for awarding PhD fellowships - coordinated reviews for over 100 applications.
- TorBUG Committee Member TorBUG, University of Toronto Committee member for organizing monthly bioinformatics seminar series.

November 2021 - Present

Jan 2019 - May 2019

Sept 2015 - May 2016

May 2015 - August 2015

Jan 2024 - Present

Feb 2023 - Present

June 2019 - June 2021

# Teaching

- Collaborator on Bioinformatics Training Platform January 2024 to Present Collaborator on Bioinformatics.ca program intended to aggregate, scale, and link training communities across Canada.
- Genome Biology and Bioinformatics Program: Deep Learning for Biology Seminars November 2019 to January 2020 Designed and taught three-part talk series for the Genome Biology and Bioinformatics program at the University of Toronto.
- Introduction to Deep Learning for Biologists September 17th and 24th, 2018 Designed and taught introductory workshops on building deep neural networks to graduate students in the Department of Cells and Systems Biology at the University of Toronto.

## Service

• Advisor and Keynote Speaker for AIBIDS Program at Gallaudet/University of Pittsburg

Dec 2022 to June 2023 - advised biomedical informatics program for underrepresented students.

- Speaker for ISMB/ECCB Advancing Accessibility in Bioinformatics Training July 2023 - speaker on how to advance diversity and inclusion in relation to disability in bioinformatics.
- Speaker for Accessibility Services at University of Toronto March 29 2023 - advised students on navigating disability and work in panel discussion.
- Dean's Advisory Council for Arts and Science Faculty at University of Toronto April 2022 to present - advised Dean Melanie Woodin on academic programming.
- Peer Review

Peer reviewer for Neurips 2022 workshops, NeurIPS 2021, Nature Communications, Genome Biology, PLOS Computational Biology, MLCB 2020, Bioinformatics, Cell Systems, CIBB 2019, IEEE TM, and BMC Supplements. Area chair for LMRL 2022.

• Community Engagement

Speaker for Skype a Scientist for high school classes in Providence, RI and Englewood, NJ.

• Conference Organizing Scientific committee for DAC2021.

# **Open-Source Software**

# • Reverse Homology

October 2021 to Present Developer and maintainer of Reverse Homology web application, for the discovery of IDR function (http://reversehomology.eastus.cloudapp.azure.com/)

# • YeastSpotter

December 2018 to Present Developer and maintainer of open-source segmentation web application YeastSpotter (http://yeastspotter.csb.utoronto.ca)

# HONORS AND AWARDS

| NSERC MSFSS   | Dec 2019   |
|---|------------|
| Award to fund visiting research position; \$6,000 for three months                  |            |
| Ray Reiter Graduate Award in Computer Science                                       | Dec 2019   |
| Departmental award with value of \$1,000  |            |
| MLCB Travel Award   | Dec 2019   |
| Conference award at MLCB, awarded on basis of merit to five outstanding papers      |            |
| NSERC CGS-D   | May 2017   |
| National research fellowship: awarded \$35,000 per year over three years            |            |
| Ranked 1st of 192 applicants in cellular and molecular biology                      |            |
| Best Paper: GLBIO/CCBC 2017   | May 2016   |
| Regional conference award for best full paper                                       |            |
| NSERC CGS-M   | (Declined) |
| National research fellowship: awarded \$13,000. Declined due to institution choice. |            |
| NSERC URSA  | May 2015   |
| National undergraduate research fellowship: awarded \$4,500 for summer research     |            |
| Western Gold Medal for Bioinformatics   | May 2015   |
| Institutional: Awarded to top graduating student from program                       | Ū          |
| NSERC URSA  | May 2013   |
| National undergraduate research fellowship: awarded \$4,500 for summer research     | -          |