Tianyu Lu

Bioengineering Department
Stanford University
✉ tianyulu@stanford.edu
⌂ https://tianyu-lu.github.io · ⬤ tianyu-lu

RESEARCH INTERESTS

I am interested how machine learning models can generate hypotheses to probe and understand biology, in particular

- Generative models of protein sequence and structure applied to protein engineering
- Connecting wetlab assay development with probabilistic samplers, model-based optimization
- Joints, conditionals, and marginals of $p(\texttt{sequence}, \texttt{structure}, \texttt{function})$

PUBLICATIONS

Strokach A, **Lu T**, and Kim PM. "ELASPIC2 (EL2): Combining Contextualized Language Models and Graph Neural Networks to Predict Effects of Mutations." Journal of Molecular Biology (2021): 166810.

CONFERENCES

Boucinha A, Kell B, Sheikh F, Diep P, Yeung A, Escobar A, Emond CA, Pierce C, Siddartha K, Chang L, Sadatmousavi P, Stephens S, **Lu T**, Sajtovich VA. "A framework towards transdisciplinary synthetic biology curricula for heterogeneous undergraduate cohorts." Canadian Engineering Education Association Conference (2021).

PREPRINTS

Lu T, Lu AX, and Moses AM. "Random Embeddings and Linear Regression can Predict Protein Function." arXiv preprint arXiv:2104.14661 (2021).

Lu T, and Silva A. "dynUGENE: an R package for uncertainty-aware gene regulatory network inference, simulation, and visualization." bioRxiv (2021).

EXPERIENCE

ProteinQure
Junior Machine Learning Scientist *May 2021 – Aug 2022*
Advisor *Sep 2022 – Present*

- Designed an embedding-based kernel for Gaussian process regression for non-canonical helical and cyclic peptide property prediction. Experimentally validated EC50 and IC50 prediction on various GPCRs and other clinically-relevant targets.
- Modeled 3D atom density preferences with 3D-CNNs for a non-canonical amino acid energy function.
- Building sequence, structure, and surface-based property predictors for nanobody design: immunogenicity, solubility, non-specificity, aggregation propensity, target-agnostic binding score.
- Building nanobody combinatorial libraries guided by sequence and structure-based denoising autoencoders and priors on p(sequence) and p(structure)

University of Toronto
Research Student *Sep 2020 – April 2021*
Supervisor: Prof. Alan Moses

- Modeled gene regulatory network dynamics using neural networks mixed with ODEs.
- Designing gene regulatory networks *de novo* using automatic differentiation through ODE solvers.
- Establish baseline models and datasets protein language model pre-training tasks.

University of Toronto
Research Student *Aug 2019 – Sep 2020*
Supervisors: Prof. Philip Kim, Dr. Pedro Alberto Valiente Flores

- Designing novel protein folds with generative models of protein structures (Transformer, GAN).
- Designing oncoprotein inhibitors using Rosetta and molecular dynamics simulations.
- Implemented code to search the PDB for protein surfaces that mimic DNA.

iGEM Toronto
Co-President *Nov 2020 – Aug 2022*
Drylab Lead *Apr 2020 – Nov 2020*
Drylab Member *Apr 2019 – Apr 2020*
Supervisor: Prof. Radhakrishnan Mahadevan

- Contributed ML-guided designed plastic-degrading enzyme to BioBrick.
- Working on active learning methods for an iterative drylab-wetlab feedback loop.

- Designed a plastic-degrading enzyme using model-based optimization.
- Quantifying benefits of learned protein sequence embeddings on protein function prediction.
- Analyzing PET catalysis dynamics with molecular dynamics simulations.

Canadian Synthetic Biology Education Research Group
Lead Instructor *Oct 2021 – Jun 2022*
Machine Learning Instructor *Sep 2019 – Oct 2021*
Supervisors: Patrick Diep, Brayden Kell

- Created code resource on machine learning for protein design and systems biology.
- Presented seven hours of content, covering both classical (Docking, Rosetta, MD) and recent methods (Sequence-to-function models, generative models, representation learning, active learning).
- Lead meetings on curriculum development based on pedagogical principles.

Education
Stanford University
Ph.D. Bioengineering Year 1 *Sep 2022 – Present*

University of Toronto
B.Sc. Bioinformatics and Computational Biology, Computer Science *Sep 2018 – Jun 2021*
cGPA: 3.94/4.00

McGill University
Biological, Biomedical and Life Sciences *Sep 2017 – May 2018*
cGPA: 4.00/4.00

Skills
Programming Python, PyTorch, NumPy, Bash
Tools PyMOL, VMD, GROMACS, RosettaScripts, Unix, LaTeX

Talks
1. Accelerating Plastic Recycling with PETase, *iGEM Grand Jamboree*, Boston, MA, Oct. 2019.
2. Recurrent Neural Networks for Protein Design, *Ontario iGEM Conference*, Guelph, ON, Jul. 2019.

Awards &
Grants
- Skule Conference Grant ($2000) *2022*
- iGEM Impact Grant ($3000) *2022*
- Skule Endowment Fund ($5000) *2021*
- iGEM Impact Grant ($3000) *2021*
- Institute of Medical Science Grant ($2000) *2021*
- Dean's Student Initiative Fund ($1000) *2021*
- NSERC Undergraduate Research Award ($6000) *2020*
- COVID-19 Student Engagement Grant ($3000) *2020*
- iGEM Gold Medal, Best Manufacturing Project Nomination, Boston MA *2019*
- National Biology Competition, Top 1% ($3000) *2017*
- University of Toronto Entrance Scholarship ($7000) *2017*

Interests
Rachmaninoff piano concertos, Frédéric Chopin, pho, table tennis, cycling, public transport.