

# Xiao (Katrina) Liu

 [katrina-liu](#) |  [Katrina Liu](#) |  [xkliu0424@gmail.com](mailto:xkliu0424@gmail.com) |  +01 412 315 5028

## EDUCATION

|                      |  |                  |
|----------------------|--|------------------|
| Aug. 2022 - present  | <b>M.S. in Biomedical Informatics, Harvard Medical School</b>              | Boston, MA       |
|                      | Capstone Advisor: Dr. Martha L. Bulyk. Expected graduation date: Dec. 2023 | (GPA: 3.844/4.0) |
| Aug. 2018 - May 2022 | <b>B.S. in Computer Science, Carnegie Mellon University</b>                | Pittsburgh, PA   |
|                      | University Honors. Minor: Mathematical Sciences, Computational Biology     | (GPA: 3.86/4.0)  |

## PUBLICATIONS & PRESENTATIONS

Miguel Lopez de Rodas\*, Zachary Frohock\*, **Xiao Liu**, Noelia Vilarino, Viviana Ahumada, Scott Gettinger, R Krishna Murthy Karuturi, Joshy George, Yi Li, and Kurt Schalper (2023). “Characterization and clinical significance of the intra and inter-tumor spatial immune heterogeneity using 3-dimensional multiplexed phenotyping in non-small cell lung cancer”. In: *Society for ImmunoTherapy of Cancer’s 38th Annual Meeting*. **Rapid Oral Presentation** (Manuscript in Preparation). DOI: [10.1136/jitc-2023-SITC2023.0923](https://doi.org/10.1136/jitc-2023-SITC2023.0923).

Qianwen Wang\*, **Xiao Liu\***, Man Qing Liang\*, Sehi L’Yi, and Nils Gehlenborg (Oct. 2023). “Enabling Multi-modal User Interactions for Genomics Visualization Creation”. In: *2023 IEEE Visualization Conference (VIS)*. DOI: [10.31219/osf.io/5zgye](https://doi.org/10.31219/osf.io/5zgye).

**Xiao Liu**, Luca Mariani, and Martha L. Bulyk (Dec. 2023). “DNA bendability regulates transcription factor pioneer binding to nucleosomes”. In: *Cold Spring Harbor Laboratory Conference on Genome Informatics 2023*. **Poster Presentation** (Manuscript in Preparation). URL: <https://meetings.cshl.edu/posters/info23/virtualposters.aspx>.

## RESEARCH EXPERIENCE

**Brigham & Women’s Hospital and Harvard Medical School** Boston, MA

Graduate Research Assistant, Bulyk Lab

Jan. 2023 - present

- Supervised by Dr. Martha Bulyk for a one-year independent Master’s research capstone studying how mechanistic features of DNA affect transcription factor binding to nucleosomes.
- Designed and implemented R pipelines to pre-process Illumina sequencing reads and performed motif discovery on SELEX-seq experiment data for ~200 transcription factors.
- Employed a published deep learning model, *DNAcycP*, to predict DNA sequence bendability from DNA reads.
- Designed and conducted statistical analyses for evaluating a proposed biophysical models using predicted DNA sequence bendability.
- Processed and analyzed *in vivo* datasets such as MNase-seq, ChIP-seq, and CUT&RUN data. Main tasks include calling nucleosome positions, motif discovery, and motif scanning.

**The Jackson Laboratory**

Bar Harbor, ME

Research Data Analyst Intern, Computer Science Department

May 2023 - August 2023

- Worked on a project using graphical neural networks (GNNs) to study 3D intra/inter-tumor structural heterogeneity in collaboration with the Schalper Lab at Yale University.
- Designed a novel loss function for image registration incorporating convolution within the structural similarity index metric (SSIM) to align morphological structures of tumors.
- Implemented a 3D tumor image registration pipeline in Python to extract images from tumor micro-array images and align vertical slide stacks and cell positions using image transformers.
- Improved a GNN model accuracy by 10% through fine-tuning hyperparameters to predict patient responses to immunotherapy using cell-level tumor micro-environment graphs.

---

\*Co-first-authors

## Harvard Medical School

Boston, MA

Graduate Research Assistant, Gehlenborg Lab

Sep. 2022 - Jun. 2023

- Developed the AutoGosling model which enables multimodal genomic visualizations generation using an object detection neural network framework YOLOv7 and a larger language model GPT-3.5.
- Implemented a JavaScript/Python pipeline to generated ~30,000 genomic visualizations for the training images.
- Trained the YOLOv7 model to predict components of genomic visualizations from images, and subsequently fine-tuned hyperparameters to attain an overall prediction accuracy exceeding 95%.
- Embedded the GPT-3.5 API in AutoGosling and pre-trained it with prompts to allow natural language visualization customization.
- Designed and implemented a web application with Flask and React to allow user interacting with AutoGosling.

## Carnegie Mellon University

Pittsburgh, PA

Undergraduate Research Assistant, Mohimani Lab

Feb 2021 - Dec. 2022

- Worked on a project using genomic sequences to predict type II polyketide structures.
- Collected post-translational modifications in type II polyketides synthesis pathways from ~100 research papers.
- Implemented a Python pipeline to extract domain sequences of polyketide backbone structures.
- Applied machine learning models to predict polyketide backbone features from the domain sequences.
- Implemented a Python program to reconstruct polyketide structures based on the predicted backbones and post-translational modifications.

## Acuity Diagnostics

Pittsburgh, PA

Machine Learning Research Intern

Jun. 2022 - Sep. 2022

- Worked on a project using deep learning models and H&E histopathology slides to predict prostate cancer patient responses to chemotherapy.
- Implemented a Python pipeline to pre-process whole slide images into slide patches and perform semantic and nucleus segmentation to identify tumor regions and lymphocyte cells.
- Implemented a Python pipeline to detecting muscle-invading tumor regions from whole slide images.

## AWARDS AND FELLOWSHIPS

---

Summer 2021

**Summer Undergraduate Research Fellowship**, Carnegie Mellon University  
Awarded \$3500.

Fall 2018-Spring 2022

**Dean's List**, Carnegie Mellon University

## SELECTED COURSE PROJECTS

---

### Exploring the Impact of Aging in Cell-State Transitions Using Single-Cell Transcriptomics and Chromatin Accessibility

Massachusetts Institute of Technology, 6.8701, Group Project Leader

Sep. 2023 - Dec. 2023

In this project, we explored the impact of transcription factor (TF) perturbations on cell-state transitions in samples collected from both young and old individuals. Using a published tool, *CellOracle*, we conducted *in silico* TF perturbations across diverse biological contexts including hematopoiesis, epithelial-mesenchymal transition, and motor neuron development. Our analyses focused on discerning the variations in cell-state transitions of different age groups after perturbing the same TFs. Additionally, we built a similar model to predict changes in cell-level chromatin accessibility after TF perturbations.

### Exploring Variational Autoencoders in Predicting Prostate Adenocarcinoma Subtypes from Transcriptome Profiles

Harvard Medical School, BMI 707, Group Project Leader

March. 2023 - May. 2023

We employed variational autoencoders (VAEs) to generate latent representations of transcriptomic profiles from prostate cancer patients. This involved 500 sample gene expression profiles encompassing 66,000 genes from the TCGA-PRAD project. Additionally, we assessed VAE's efficacy as a means to counter class imbalance issues in classifying rare prostate cancer subtypes. We compared the subtype classification accuracy of the encoded latent expressions against the original gene expressions and discovered an improvement. Furthermore, we benchmarked VAE against alternative dimension reduction methods like PCA and found that VAE encoded representations is more efficient in capturing differences in gene expressions across subtypes.

## TEACHING EXPERIENCE

---

### Teaching Fellow, Harvard Medical School

Boston, MA

BMI713 Computing Skills for Biomedical Sciences

Aug. 2023 - Oct. 2023

BMI715 Computational Statistics for Biomedical Sciences

Oct. 2023 - Dec. 2023

- Led in-class group activity sessions for 50 students.
- Drafted three weekly problem sets, held weekly office hours, and graded problem sets and exams.

### Teaching Assistant, Carnegie Mellon University

Pittsburgh, PA

17-214/514 Principles of Software Construction

Aug. 2021 - May. 2022

- Led weekly 1-hour recitation for 40 students and gave feedback to student presentations.
- Prepared recitation materials, held weekly office hours, and graded labs and exams.

### Peer Tutor, Carnegie Mellon University

Pittsburgh, PA

15-150 Functional Programming

May. 2020 - Jul. 2020

- Tutored two sections of 6 students every week to help them review lectures and complete homework assignments.

### Teaching Assistant, Carnegie Mellon University

Pittsburgh, PA

21-120 Differential and Integral Calculus

Aug. 2019 - Dec. 2019

- Led 1-hour recitation for 20 students twice a week, held weekly office hours, and grade homework and exams.

## SKILLS

---

|                    |  |
|--------------------|--|
| Programming Skills | Python, Java, JavaScript/Typescript, C/C++, Linux, R, HTML/CSS, L <sup>A</sup> T <sub>E</sub> X                |
| Development Skills | Django, Agile Development, Object Orient Programming, PyTorch\TensorFlow                                       |
| Laboratory Skills  | Gel Electrophoresis, PCR, Bacteria Transformations, Protein Purification, Western Blotting, Coomassie Staining |
| Languages          | Chinese/Mandarin (First Language), English (Professional Level)  |

## ACTIVITY AND OUTREACH

---

2023      Member, **Association for Women in Science**

2018-2019    Problem Writer, **Carnegie Mellon Informatics and Mathematics Competition**