Mihir Bafna

mbafna@gatech.edu | 858-603-6580 | LinkedIn ⬈ | GitHub ⬈ | Google Scholar ⬈ | Website ⬈

EDUCATION

Georgia Institute of Technology   Aug 2020 - Dec 2023 (*Graduation*)
B.S. Computer Science (Senior Standing) - GPA: 4.0

Relevant Coursework - Deep Learning, Machine Learning, Machine Learning for Bioscience, Advanced Linear Algebra, Applied Combinatorics, Probability Theory & Statistics, Design & Analysis of Algorithms, Graduate Algorithms, Automata & Complexity Theory, Data Structures, Computer Architecture, Discrete Mathematics, Objects and Design, Multivariable Calculus, Differential Equations, Intro to AI, Robotics/Perception.

Extracurriculars - Undergraduate Research Ambassador (URA), TA for Deep Learning (CS 4644)

PUBLICATIONS/PATENTS

- **M. Bafna**, H. Li, X. Zhang. (2022). CLARIFY: Cell-cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *To Appear in Proceedings of the 31st Conference on Intelligent Systems for Molecular Biology* (**ISMB'23**). In press in journal ***Bioinformatics***

- Rajkumar U., Javadzadeh S., **Bafna M.**, Wu D., Yu R., Shang J., Bafna V. (2022). DeepViFi: detecting oncoviral infections in cancer genomes using transformers. *In Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (**ACM-BCB'22**). Article 2, 1−8. (Fully Refereed Conference) DOI: 10.1145/3535508.3545551

- Nguyen, Mora-Blanco, Turner, Christiansen, **Bafna**. 2021. Computer-implemented methods for quantitation of features of interest in whole slide imaging. PCT/US2021/022308, filed March 15, 2021. **(Provisional patent)**.

PRESENTATIONS/POSTERS

- **Mihir Bafna**, Hechen li, Xiuwei Zhang. CLARIFY: Cell-cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Atlanta Workshop for Single-cell Omics (AWSOM 2023)*. **Best Poster Award.**

- **Mihir Bafna**, Vikranth Keerthipati, Subhash Kanaparthi. DiffRNAFold: Generating RNA Structures & Conformations with Latent Space Diffusion. *Deep Learning (CS 7643) Final Project Symposium & Poster Presentation*. **Best Overall Project & Poster (1st out of 200)**.

- **Mihir Bafna**, Xiuwei Zhang. Benchmarking and Refining Cell-Cell Interactions with Spatial Transcriptomics and Graph Neural Networks. *St. Jude's National Symposium for Undergraduate Research (NSUR 2022)*. **Poster Presentation (52/352 selected)**.

EXPERIENCE

Bonnie Berger Lab | **Massachusetts Institute of Technology**   Boston, MA
Research Assistant   *May 2023 – Present*

- Working on diffusion models for genome structure (via scHi-C) and protein structure conformational change.
- Accepted into the Broad Summer Research Program (BSRP) starting June

Jian Ma Lab (ML & CompBio) | **Carnegie Mellon University**   Pittsburgh, PA
Research Intern   *Dec 2021 – Present*

- Implemented a Hypergraph Graph Neural Network model architecture (hyperSAGNN) for understanding genetic interactions, specifically trigenic, and enrichment in protein complex formation
  - Predicted 400,000 new trigenic interactions in yeast and bounded each prediction with model uncertainty using Gaussian Processes
  - Using AlphaFold2 and ESMFold2 Protein Sequence Embeddings to guide model in predicting trigenic interactions that are enriched in protein complex formation.
  - **Preparing journal submission** for aforementioned work
- Creating novel Graph Autoencoder architecture to obtain Hierarchical and Task Agnostic Whole Graph Embeddings
  - Incorporating LSTMS, Attention, and Transformers instead of pooling to obtain whole graph embeddings that preserve hierarchy ordering and can be decoded into the original graph.

Zhang Lab (ML & Bioinformatics) | **Georgia Institute of Technology**   Atlanta, GA
Student Researcher   *Nov 2021 – Present*

- Developed *Clarify*: a graph autoencoder based tool for refining extracellular AND intracellular interactions (cell-cell inference, gene regulatory network prediction, subnetwork comparison) with SubGraph Neural Networks and Spatial Transcriptomics data
  - * **Paper accepted to ISMB**
  - * **Best Poster Award at AWSOM**
- Creating a centralized data resource by benchmarking existing ST-CCI inference methods on 240 simulated and 2 real datasets
- Analysis of cell clusterings from the learned latent factors of non-negative matrix factorization (scMoMat) for identifying marker features (genes, regions, etc)

Boundless Bio Inc   San Diego, CA
SWE/Bioinformatics Research Intern   *Aug 2020 – Dec 2021*

- **Provisional patent** for creating metaDetect: a computer vision algorithm for identifying metaphase spreads in stained whole slide images of cancer cells using image analysis filtering/techniques.
- Created automated pipeline for whole slide imaging → metaDetect (CV) → ecDNA quantification (CNN)
- **Writing publication** for Whole Slide Imaging and the metaDetect algorithm (above)
- Implemented better workflow and cloud in ecDNA Analytics
- Used machine learning (precision recall / roc curves) for optimization of image analysis parameters
- Frameworks|Languages: scikit-learn, OpenCV, Jupyter Notebooks, Google Cloud | (Python)

CompBio Lab | **UC San Diego**   San Diego, CA
Student Researcher   *July 2021 – Sep 2021*

- Developed Machine Learning (Random Forest) model to classify Human Papillomavirus genomes into their four broad classes (Alpha, Beta, Gamma, Other).
- Created a faster and more accurate classifier with machine learning gradient boosting techniques (LightGBM & XGBoost)
- **Paper accepted** for DeepViFi: Detecting Novel Viral Integration in Cancer Genomes using a transformer architecture and LightGBM for further family classification
- Frameworks|Languages: LightGBM, XGBoost, scikit-learn, Pandas, NumPy | (Python, JS)

Adibi Lab (Medical Imaging) | **Georgia Institute of Technology**   Atlanta, GA
Student Researcher   *Jan 2021 – Jun 2022*

- Using deep learning to detect lung diseases from X-ray images (Segmentation/Classification)
- Created saliency maps (GRAD-CAM method) to understand and visualize our CNN model's predictions.
- Frameworks|Languages: TensorFlow, PyTorch, Flask, React, Azure | (Python, JS)

AWARDS/HONORS

Georgia Tech Faculty Honors| (4.0 GPA)	Awarded Every Semester
Atlanta Workshop for Single-cell Omics (AWSOM)	Best Poster Award	April 2023
Broad Summer Research Program (BSRP)| Accepted	March 2023
Harvard Summer Institute for Biomedical Informatics (SBMI)| Accepted	March 2023
CS 7643 Best Project Award	First Place Winner	Dec 2022
Georgia Tech PURA	Presidential Undergraduate Research Award ($1500 Stipend)	May 2022
HealthTech Hacks @ GT	First Place Winner	March 2021
Grand Challenges @ GT	First Place Winner for best project idea	May 2021
Westview HS Honors	Top Math Student in HS Graduating Class (600)	June 2020
Westview HS Honor Chain	Top 3% GPA in HS Graduating Class (600)	June 2020

PROJECTS

DiffRNAFold | *Flask, Python, Javascript*   GitHub Repository 🗗

- Latent space diffusion denoising model for RNA conformation generation.
  - * Implementing conditional generation aspect with pretrained DNABERT language model.
  - * Implementing rotational and translational equivariant point cloud autoencoder using the recently published Frame-Averaging technique.
- Winner of Best Project Award (CS 7643)
- Quick visualization of RNA molecule point cloud DiffRNAFold 🗗

CovidAutoEncoder | *Python, Pytorch*   GitHub Repository 🗗

- Created and tested various autoencoder architectures (Convolutional, Variational, Linear) to learn latent representations of SARS-Cov-2 genomes and most importantly: the latent differences between the variant genomes
- Once the lower dimensional latent representations were learned by training the autoencoders on the Alpha and Delta (more prevalent) strains, we encoded Omicron genomes into these representations and were able to clearly distinguish the variants with a simple PCA.

scAtlasVis | *Python, Flask*   GitHub Repository 🗗

- Low Dimensional Manifold Visualization of Pan Cancer Single B-Cell Atlas
- Visualization tool will be included in **journal submission** of the Pan Cancer B-Cell Atlas.

ecDNA Analytics | *Flask, Python, Javascript*   GitHub Repository 🗗

- Web platform that integrates the metaDetect and ecSeg tools with respective visualizations for the purpose of extrachromosomal DNA analysis and annotations.
- View at ecDNA Analytics 🗗

MatrixOperationsLibrary | *Python*   GitHub Repository 🗗

- Created module that contains functions for theoretical interpretations in Linear Algebra
- Essentially a bootleg NumPy that helped me deeply understand the purpose of efficient matrix computations and their use in computer science

LANOS | *Javascript, React, Express, Node.js, MongoDB*   GitHub Repository 🗗

- Startup platform for connecting HS students with career/volunteer opportunities.
- Developed full stack matching platform as part of the Startup Lab and Idea 2 Prototype Create-X programs at Tech.

VetAssist | *React Native, MongoDB, Alexa Skills*   GitHub Repository 🗗

- Mental health support platform that allows veterans to connect with others nearby and integrates sentiment analysis of their daily virtual journals to assess mental health
- Virtual journaling was created with the Alexa Skills library so that the veterans could speak to a virtual "friend" to journal their day
- Winner of HealthTech Hacks GT

AtlantaEnvJustice | *Flask, Python, Javascript*   GitHub Repository 🗗

- Interactive map website to highlight the disproportionate modes of quality and environmental injustice in Atlanta.
- Uses the mapbox api and an aesthetic UI to bring to light certain injustices (racial, financial, etc.) regarding environmental inequality in counties around the Atlanta metropolitan area.

SKILLS

Programming Languages: Python, Java, Javascript, HTML5/CSS, C, Matlab, R

Tools | **Frameworks:** PyTorch, TensorFlow, OpenCV, Jupyter, numpy, Scikit, Flask, Git, MERN Stack, Google Cloud, Raspberry Pi, NVIDIA Jetson, Arduino, pysam, Biopython