Bhavay Aggarwal

Ph.D. Candidate in Bioinformatics | AI/ML for Spatial & Multimodal Biology
236 16th Street NW, Atlanta, GA 30363 Email: baggarwal9@gatech.edu
bhavaygg.github.io | github.com/bhavaygg | linkedin.com/in/bhavay-aggarwal Mobile: +1-(404)-747-8290

EXPERIENCE

- **Georgia Institute of Technology** Atlanta, GA

  Graduate Research Assistant – Sinha Lab *January 2023 – Present*

  - Developed SCOUBI, a signaling-informed deep learning framework for analyzing extrasomatic mRNA in brain spatial transcriptomics, enabling systematic exploration of synaptic biology.
  - Developed CellSP, an open-source tool that extracts gene-cell programs from high-resolution spatial transcriptomics to resolve and interpret subcellular gene organization.
  - Designed and evaluated deep learning approaches for spatial transcriptomics, including vision transformers and graph neural networks, to model tissue context, spatial neighborhoods, and subcellular structures.
  - Drove a major overhaul of InSTAnT, parallelizing core workflows to reduce analysis runtime by over 20x, expanding analytical capabilities, and validating the framework across spatial transcriptomics technologies.
  - Developed computational analyses for collaborative cell-manufacturing studies, integrating imaging and lipidomic measurements to characterize MSC potency and building cross-validated models linking potency-assay measurements to patient outcomes.
  - Guide computational research strategy across lab projects, advising on analytical approaches, methodological decisions, and validation plans.

- **GlaxoSmithKline** Collegeville, PA

  Computational Biology Intern *May 2023 – August 2023*

  - Built an interpretable perturbation-modeling framework for MegaMap, using large-scale longitudinal gene-knockdown data to identify combinations of perturbations that recapitulate disease expression signatures.
  - Identified compact gene panels that captured differences among disease states, creating a practical basis for comparing perturbations across indications.
  - Connected experimental findings to clinical evidence through rigorous validation of prioritized gene–disease relationships.

- **Indian Institute of Science (IISC)** Remote, India

  Research Intern – ATCG Lab *May 2021 – December 2021*

  - Expanded approximate nearest-neighbor search methods to support genomic distance measures and large sequence-derived datasets.
  - Cut genomic indexing time by 90% through a distributed Python/C++ workflow built with Ray and Dask.

- **Wikilimo** Remote, India

  Machine Learning Intern *February 2021 – April 2021*

  - Built data pipelines for cleaning, normalizing, and interpolating environmental time series used in wildfire modeling.
  - Trained wildfire forecasting models while using SHAP to identify the environmental factors driving predictions.

- **Indraprastha Institute of Information Technology Delhi** Delhi, India

  Research Intern – Complex Systems Lab *May 2020 – July 2020*

  - Applied language models to uncover contextual relationships among ingredients, translating recipe text into features for computational cuisine analysis.

- **All India Institute of Medical Sciences** Delhi, India

  Research Intern *February 2020 – October 2020*

  - Created machine learning models to segment brain and neck tumors from X-ray images using super voxels.

PUBLICATIONS

1. **Aggarwal, B.**, Jain, S. and Sinha, S., 2026. **Revealing the transcriptomic organization of extrasomatic interfaces using SCOUBI.** *Under review.*
2. **Aggarwal, B.**, Barbasch, T., Bell, A., Dan, U., Kim, Y., Sinha, S. and Zhao, D. 2026. **A novel framework for testing the social decision-making network hypothesis using single-nucleus and spatial transcriptomics data.** *Manuscript in preparation.*

3. Hao, Y., Kim, Y., **Aggarwal, B.** and Sinha, S., 2026. **MANTIS: Analytics toolkit for spatial metabolomics with matching spatial transcriptomics data.** bioRxiv, pp.2026-01.

4. Kumar, A., Moctezuma, F.R., **Aggarwal, B.**, Zhang, N., Coskun, A.F. and Sinha, S., 2026. **CellWHISPER disentangles direct cell–cell communication from structural proximity.** bioRxiv, pp.2026-01.

5. **Aggarwal, B.** and Sinha, S., 2025. **CellSP enables module discovery and visualization for subcellular spatial transcriptomics data** . Communications Biology, 8(1), p.1528.

6. Schneider, R.S., Nieves, E.B., **Aggarwal, B.**, Bowles-Welch, A.C., Stevens, H.Y., Kippner, L.E., Boden, S.D., Mautner, K., Drissi, H., Roy, K. and Lam, W.A., 2025. **On-chip 3D potency assay for prediction of clinical outcomes for cell therapy candidates for osteoarthritis** . Nature Communications, 16(1), p.4915.

7. Kumar, A., Schrader, A.W., **Aggarwal, B.**, Boroojeny, A.E., Asadian, M., Lee, J., Song, Y.J., Zhao, S.D., Han, H.S. and Sinha, S., 2024. **Intracellular spatial transcriptomic analysis toolkit (InSTAnT)** . Nature Communications, 15(1), p.7794.

8. Priyadarshani, P., Van Grouw, A., Liversage, A.R., Rui, K., Nikitina, A., Tehrani, K.F., **Aggarwal, B.**, Stice, S.L., Sinha, S., Kemp, M.L. and Fernández, F.M., 2024. **Investigation of MSC potency metrics via integration of imaging modalities with lipidomic characterization.** Cell Reports, 43(8).

9. **Aggarwal, B.** and Ray, A., 2021. **AgAnt: A computational tool to assess Agonist/Antagonist mode of interaction.** bioRxiv, pp.2021-11.

10. Sharma, T., Upadhyay, U., Kalra, J., Arora, S., Ahmad, S., **Aggarwal, B.** and Bagler, G., 2020, April. **Hierarchical clustering of world cuisines.** In 2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW) (pp. 98–104). IEEE.

EDUCATION

- **Georgia Institute of Technology** — Atlanta, GA
  Ph.D. in Bioinformatics (Expected May 2027) | M.S. in Bioinformatics, 2023 | GPA: 3.91 — *August 2022 – Present*
  Coursework: ML for Computational Biology, ML for Chemistry, Data Science for Epidemiology, Computational Genomics

- **Indraprastha Institute of Information Technology** — Delhi, India
  Bachelor of Engineering in Computer Science and Biosciences; GPA: 7.91/10.0 — *August 2018 – June 2022*
  Coursework: Advanced Machine Learning, Deep Learning, Information Retrieval, Software Development using Open Source

TECHNICAL SKILLS

Languages	Python, R, C/C++, SQL, Java
Libraries & Frameworks	PyTorch, Keras, Scikit-learn, Hugging Face, Scanpy, Seurat, scvi-tools, AnnData, NumPy, Pandas
Measurements	spatial, single-cell/single-nucleus, bulk
Modalities	transcriptomics, metabolomics, lipidomics, imaging
Tools & Platforms	Git, Linux, Slurm, Snakemake, Ray, Dask, Conda, AWS, GCP

PROJECTS

- **AgAnt:** Developed deep-learning models for receptor–ligand binding prediction, using GNNs and Transformers to learn protein representations through transfer learning.

- **Outbreak analysis of *Escherichia coli O157*:** Contributed to a bioinformatics interface combining genome assembly, functional annotation, and comparative genomics for rapid outbreak analysis of *E. coli* strains.

- **Disease Symptom Search Trend Analysis:** Modeled spatiotemporal Google search trends for disease symptoms to forecast regional COVID-19 cases.

- **pyLFDA:** Developed an open-source Python library for pairwise lipid force-distribution analysis in molecular-dynamics simulations.