# Sanya Badole

Atlanta, GA | 470-301-6866 | sbadole6@gatech.edu | linkedin.com/in/sanya-badole | github.com/sanyabadole | sanyabadole.com

## Education

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

*Master of Science in Bioinformatics | GPA: 3.77/4.0* — *Expected: May 2026*

- Relevant Coursework: Machine Learning in Computational Biology, Computational Genomics, Applied Bioinformatics & Genomics, Biostatistics, Drug Discovery, Health Informatics, Reproducible Bioinformatics, Database Systems, Human Evolutionary Genomics, Data & Visual Analytics

**Dr. D.Y. Patil Biotechnology & Bioinformatics Institute** — Pune, India

*B.Tech/B.Sc. in Biotechnology | GPA: 4.0/4.0* — *Awarded: Jun. 2024*

- Relevant Coursework: Cancer Biology, Immunology, Molecular Modelling & Drug Design, Structural Biology, rDNA Technology, Stem Cell Technology, Cell Biology, Biochemistry, Microbiology — with hands-on wet lab training across all core disciplines

## Experience

**Bioinformatics Software Engineer · Digital Transformation Intern** — May 2025 – Nov. 2025

*Boehringer Ingelheim* — *Ridgefield, CT*

- Built end-to-end bioinformatics pipelines in Python processing large-scale multi-modal NGS data with 95% test coverage and automated QA protocols using pytest and CI/CD workflows.
- Deployed production tools on OpenShift and AWS cloud infrastructure; contributed to Agile/Scrum development cycles with 200+ Git commits and 30+ code and design reviews via Bitbucket.
- Processed and curated large-scale single-cell sequencing datasets to support cross-functional collaboration between computational scientists and immunologists across pharma R&D teams.
- Built interactive D3.js dashboards and technical reports to communicate results to 15+ stakeholders; implemented Django/React platform with SQL backend deployed with nginx at 99% uptime.

**Graduate Research Assistant** — Aug. 2024 – May 2025

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

- Built scalable single-cell pipelines integrating scRNA-seq, scATAC-seq, and spatial transcriptomics across hundreds of samples on HPC/SLURM infrastructure using Python, Scanpy, and Seurat.
- Applied ensemble ML methods (Harmony, scVI, Seurat) for automated batch effect correction and cell type classification, uncovering novel tumor-immune interactions across large-scale datasets.
- Applied dimensionality reduction (UMAP, PCA, t-SNE), clustering, and gene regulatory network inference (SCENIC) to characterize immune cell populations across large-scale single-cell atlases.

**Bioinformatics Analyst** — Jun. 2023 – Aug. 2024

*Eminent Biosciences* — *Indore, India*

- Engineered high-throughput cheminformatics workflows processing large compound libraries for virtual screening, molecular docking, and in-silico ADMET analysis for drug candidate prioritization.
- Developed automated RNA-seq analysis pipelines using Python and R (DESeq2, edgeR) with standardized QC protocols across 100+ samples for biomarker discovery applications.
- Built multimodal deep learning models (PyTorch, scikit-learn) integrating genomic and imaging features, achieving 89% prediction accuracy for therapeutic target identification.

## Technical Skills

**Single-Cell & Spatial Genomics**: scRNA-seq, scATAC-seq, VDJ-seq, CITE-seq, Spatial Transcriptomics (10x Visium, Visium HD), bulk RNA-seq, multi-modal integration, variant annotation, NGS data analysis

**Core Bioinformatics Tools**: FastQC, MultiQC, GATK, PLINK, STAR, BWA, SAMtools, BEDtools, Trimmomatic, CellRanger, featureCounts, HISAT2

**Analysis Frameworks**: Seurat, Scanpy, Harmony, scVI, MuData, DESeq2, Monocle3, Slingshot, CellTypist, ClusterProfiler, Immcantation Framework

**Programming & Software Development**: Python, R, SQL, Java, JavaScript, C++, Linux/BASH/Shell scripting

**Immunology & Biology**: Tumor immune microenvironment, cancer immunology, B-cell biology & differentiation, immune cell heterogeneity, ligand-receptor signaling

**Workflow & Reproducibility**: Nextflow, Snakemake, nf-core, Docker, Singularity, CI/CD pipelines, HPC/SLURM, OpenShift, Git/GitHub/Bitbucket

**Cloud & Machine Learning**: AWS (S3, EC2), GCP, Microsoft Azure, PyTorch, scikit-learn, pandas, statistical modeling, QC & technical variability evaluation
**Web & Visualization**: Django, React, RESTful APIs, D3.js, HTML/CSS, Tableau, publication-quality figures
**Health Informatics & Standards**: FHIR, HL7, SNOMED-CT, ICD-10, LOINC, RxNorm, Athena/ATLAS, dbSNP, ClinVar, dbVar, dbGaP
**Wet Lab**: PCR, RFLP, ELISA, chromatography, analytical techniques, SOP development & documentation

## Projects & Presentations

**Targeting GBM Recurrence: ErbB4 as a Therapeutic Node** | *scRNA-seq, Python, Drug Discovery* Feb. 2026

- Analyzed 128,670 single cells across 76 GBM patients to identify a 28-gene recurrence signature enriched in NPC_neuronallike cells (6.4-fold expansion at relapse); demonstrated TMZ chemotherapy directly selects for ERBB4-NRG autocrine signaling ($p = 9.9\times10^{-42}$).
- Proposed combination therapy of Afatinib (ERBB4 inhibitor) + Prexasertib (CHK1/2 inhibitor) as a two-hit strategy to eliminate chemotherapy-resistant GBM survivors.

**Integrated Single-Cell Analysis & Visualization Platform** | *Python, Scanpy, Dash, UMAP, PHATE* 2025

- Built a unified, scalable scRNA-seq analysis platform with automated QC, dimensionality reduction (UMAP, t-SNE, PHATE, DiffMap, TriMap), and an interactive Dash dashboard for gene expression visualization.
- Processed 100,000+ cells end-to-end on a breast cancer dataset; designed for accessibility to non-programmer researchers with interactive web-based reporting.

**Bacterial Virulence Prediction via Protein Structure & ML** | *Python, AlphaFold2, PyTorch, XGBoost* 2025

- Developed a pipeline combining AlphaFold2 structural embeddings with an ensemble classifier (Random Forest, XGBoost, CatBoost, LightGBM) to predict bacterial virulence across diverse species.
- Achieved 87.4% test accuracy and AUC of 0.94; identified structural stability (pLDDT) and hydrogen-bond density as key virulence predictors, significantly outperforming sequence-only models.

**RxBiome — Pharmacomicrobiomics Drug Discovery Platform** | *Python, Multi-omics, Microbiome* 2025

- Designed a microbiome-based drug discovery platform integrating multi-omics data to explore how gut microbiome composition influences drug metabolism, efficacy, and toxicity.
- Developed computational framework to identify microbiome-derived therapeutic targets and optimize drug response predictions using pharmacomicrobiomics principles.

**High-Affinity Small Molecule CDK1 Inhibitors for Glioblastoma** | *Molecular Modelling, Drug Design* 2024

- B.Tech thesis investigating CDK1 as a therapeutic target in GBM through computational drug discovery, including virtual screening, molecular docking, and small molecule inhibitor design.
- Identified high-affinity CDK1 inhibitor candidates through in-silico ADMET analysis and structure-based drug design; guided by Dr. Anuraj Nayarisseri, Eminent Biosciences.