Avik Sengupta

Ph.D. Scholar in Computational Biology

bo22resch01004@iith.ac.in +91-8972254624 Website Avik Sengupta asgda

Research Profile

Computational biologist focusing on **multi-omics integration**, **machine learning** and **drug-based studies** for cancer theranostics research and development. My Ph.D. at IIT Hyderabad develops end-to-end pipelines and models for colorectal cancer drug repurposing, biomarker identification and deep-learning based cancer subtyping with multi-omic focus, combining bulk transcriptomics, genomics, proteomics and structural omics data with advanced ML and DL (including survival models) techniques, including HPC systems. I am eager to bring this foundation to postdoctoral positions focusing on spatial multi-omics, single-cell analysis, and clinical genomics and transcriptomics translation within world-class cancer research institutes.

Education

Indian Institute of Technology Hyderabad (IIT Hyderabad), India 2022–2026 (expected)
Ph.D. in Biotechnology (Computational Biology); CGPA: 8.88/10
Thesis: *Computer-aided designing of colorectal cancer theranostics using multimodal omics analysis*

National Institute of Technology Rourkela, India 2018–2020
M.Sc. in Life Science; CGPA: 8.75/10
Thesis: *Leaf extract mediated biogenic synthesis of alpha phase iron oxide nanoparticles (IONPs): A comparative evaluation of phytochemical characterization and assessment of its in vitro hemocompatible efficacy*

University of North Bengal, West Bengal, India 2015–2018
B.Sc. (Hons.) in Microbiology; Percentage: 66.5%

Delhi Public School, Siliguri, CBSE 2013–2015
Class XII: 87.0% — Class X: CGPA 9.60/10

Key Skills

Computational & Programming

- Languages & OS: R, Python, Bash, HTML/CSS, MySQL, PHP, Java (applications), Linux (including HPC environments), Apache webserver.
- Machine Learning: Classical ML for classification and regression (scikit-learn, xgboost, etc.), survival models (Cox PH, ML-based survival), deep learning methods for feature selection and optimisation (Graph-based networks, autoencoders, transformers), QSAR modelling, feature selection (WEKA, python-based packages, recurrent feature elimination), hyperparameter optimization (RandomisedSearchCV, GridSearchCV, Optuna), feature interpretation (SHAP-based, intrinsic methods), cross-validation framework.
- High Performance Computing: Running large-scale bioinformatics workflows on SLURM-based HPC clusters using bash scripts, job arrays and for-loops for multi-sample processing, conda environments, and standard monitoring tools.

Bioinformatics & Data Analysis

- Bulk RNA-seq: sra-toolkit, FastQC, fastp, STAR, HISAT2, samtools, featureCounts, DESeq2, GSEA, GSVA.
- Single-cell RNA-seq: Processed data analysis using Seurat in R.
- Genomics & Variants: Use of GATK/Mutect2, VEP, maftools and related tools for structural/functional variant annotation.

- Multi-omics & Cancer: Integrative multi-omics analysis for colorectal cancer and hematological malignancies, clustering and subtype discovery, biomarker identification, survival analysis (coxph, survfit, Kaplan–Meier).
- Hi-C and 3D Genomics: HiCExplorer, cooltools, HiC-Pro, Juicer tools – for chromatin interaction analyses.
- Public Databases Handling: cBioPortal, UCSC Xena Browser, GEO (RNA-seq, microarray, Hi-C, ChIP-seq, ATAC-seq, SV, short-variants, CNV), ArrayExpress, DepMap (CCLE, CRISPR-Cas9), GDSC2.
- Data Visualization: R (ggplot2) and python (matplotlib, seaborn)

Website Development
- Apache based server, MySQL database integration, PHP interface, end-to-end bash scripting, HTML & CSS.

General
- Inkscape, LaTeX, Microsoft Office, Canva, Wordpress.

Experimental (Foundational)
- Molecular biology basics: DNA/RNA extraction, PCR, cloning and transformation, microbial culture; nanoparticle synthesis and characterization (UV-vis spectroscopy, paper chromatography).

Research Experience

Ph.D. Thesis Project, IIT Hyderabad 2022–2026
Computer-aided designing of colorectal cancer theranostics using multimodal omics analysis

- Development of non-apoptotic regulated cell death-based prognostic risk models for colorectal cancer using a machine learning guided two-step framework, integrating transcriptomics and clinical data.
- Development of AI/ML-based QSAR models of colorectal cancer cell lines to identify promiscuous drugs with potential repurposing value.
- Design of multi-omics workflows to identify novel molecular subtypes of colorectal cancer using integrated genomics, transcriptomics, and other structural omics layers **(ongoing)**.
- Implemented scalable HPC workflows for preprocessing, integration, and model training across large patient cohorts.

Masters Thesis Project, NIT Rourkela 2019–2020
Leaf extract mediated biogenic synthesis of alpha phase iron oxide nanoparticles (IONPs): A comparative evaluation of phytochemical characterization and assessment of its in vitro hemocompatible efficacy

- Performed synthesis, characterization, and hemocompatibility assessment of plant-extract-derived iron oxide nanoparticles.

Publications

First-author and thesis-related
- **A. Sengupta**, S.K. Singh, R. Kumar. "Support Vector Machine-Based Prediction Models for Drug Repurposing and Designing Novel Drugs for Colorectal Cancer." *ACS Omega*, 2024, 9(16):18584–18592. DOI: 10.1021/acsomega.4c01195.
- **A. Sengupta**, S.S. Kar, R. Kumar. "Non-apoptotic regulated cell death based prognostic risk model for colorectal cancer using a machine learning guided two-step framework." *Briefings in Bioinformatics*, 2025, 26(6): bbaf639. DOI: 10.1093/bib/bbaf639.

Collaborative research articles

- K. Bhattacharjee, **A. Sengupta**, R. Kumar, A. Ghosh. "Identification of key hub genes in pancreatic ductal adenocarcinoma: an integrative bioinformatics study." *Frontiers in Bioinformatics*, 2025, 5:1536783. DOI: 10.3389/fbinf.2025.1536783.
- K. Roffin, A. Kumar, K. Vinod Kumar, **A. Sengupta** et al. "AagingBase: a comprehensive database of anti-aging peptides." *Database*, 2024, baae016. DOI: 10.1093/database/baae016.
- S.S. Roy, S. Bagri, S. Vinayagamurthy, **A. Sengupta**, C.R. Then, R. Kumar, S. Sridharan, S. Chowdhury. "Artificially inserted strong promoter containing multiple G-quadruplexes induces long-range chromatin modification." *eLife*, 2024, 13:RP96216. DOI: 10.7554/eLife.96216.3.
- V.K. Kumar, A. Kumar, K. Kundal, **A. Sengupta**, R. Kumar et al. "AMLdb: a comprehensive multi-omics platform to identify biomarkers and drug targets for acute myeloid leukemia." *Briefings in Functional Genomics*, 2024, elae024. DOI: 10.1093/bfgp/elae024.
- A. Kumar, K.V. Kumar, K. Kundal, **A. Sengupta** et al. "MyeloDB: a multi-omics resource for multiple myeloma." *Functional & Integrative Genomics*, 2024, 24:17. DOI: 10.1007/s10142-023-01280-0.
- A. Viswanathan, K. Kundal, **A. Sengupta**, A. Kumar, K.V. Kumar, A.B. Holmes, R. Kumar. "Deep learning-based classifier of diffuse large B-cell lymphoma cell-of-origin with clinical outcome." *Briefings in Functional Genomics*, 2023, 22(1):42–48. DOI: 10.1093/bfgp/elac038.

Book chapters

- A. Kumar, K. Vinod Kumar, K. Roffin, K. Kundal, **A. Sengupta**, R. Kumar. "Advancement of in silico tools for stem cell research." In: *Computational Biology for Stem Cell Research*. Academic Press, 2024, pp. 3–16. ISBN 9780443132223.
- B.T. Korra, K. Roffin, S. Singh, S.S. Kar, K. Kundal, **A. Sengupta**, R. Kumar. "Drug Repurposing-1." In: Leszczynski, J. (ed.) *Springer Handbook of Chem- and Bioinformatics*. Springer, 2025. DOI: 10.1007/978-3-031-81728-1_41.

Awards and Scholarships

- Institute Research Excellence Award, Department of Biotechnology, IIT Hyderabad (Apr 2025) – for excellence in Ph.D. research in Biotechnology.
- University Silver Medalist, University of North Bengal (2018).

Academic Outreach and Development

Conferences & Presentations

- Data-Driven Cancer Research 2026, Edinburgh, UK (2026) – Poster: Deciphering the Prognostic Landscape of Regulated Cell Death in Colorectal Cancer using a Two-Step Machine Learning Framework.
- ICGA 2025, IIT Madras (Nov 2025) – Poster: Non-apoptotic regulated cell death based prognostic risk model for colorectal cancer using a machine learning guided two-step framework.
- HySci 2024, IIT Hyderabad (Apr 2024) – Poster: Support Vector Machine-based prediction models for drug repurposing and designing novel drugs for colorectal cancer.
- IACR 2024, IISER Pune (Jan 2024) – Poster: Support Vector Machine-based prediction models for drug repurposing and designing novel drugs for colorectal cancer.
- Foundation Day 2023, Dept. of Biotechnology, IIT Hyderabad (Nov 2023) – Oral Talk: SVM-based prediction models for drug repurposing in colorectal cancer.

Workshops

- "Computational Methods in Advanced and Precision Oncology", IIT Hyderabad (Workshop, Apr 2025).
- "Theoretical approaches in cancer progression and treatment", TIFR-ICTS Bangalore (Workshop, Mar 2024).
- "Medical Image Computing", IISc Bangalore (Workshop, Feb 2023).

Key Positions of Responsibility

- Overall Student Coordinator, Biotech Industry Meet, Dept. of Biotechnology, IIT Hyderabad (Sep 2025) – Industry-Academia connect.
- Overall Mentor, HySci 2024 Conference, IIT Hyderabad (Apr 2024) – Annual Hyderabad Science conference "For the Students, By the Students" for life science and biology.
- President, BioTech Society, Dept. of Biotechnology, IIT Hyderabad (Aug 2023–Apr 2024).
- Placement Coordinator, NIT Rourkela (May 2019–Apr 2020) – Part of a team of 22 coordinators responsible for placements in the 2019–20 season.

Interests and Hobbies

- Sports: Cricket, football, badminton, volleyball, gym.
- Other: English poem and article writing; reading non-fiction and fiction; self-help literature; travelling.