Shubham Choudhury

+91-9040468955 | shubhamc@outlook.in | shubhamc-iiitd.github.io

shubham-choudhury-b22677120 | shubhamc-iiitd | Shubham Choudhury

Berhampur, Odisha - 760010, India

Research Interests

Seeking a postdoctoral position to advance research in RNA biology, with a focus on subcellular localization, RNA-protein interactions, and their role in disease conditions. Aiming to bridge fundamental RNA research with therapeutic applications, particularly in RNA-based drug development.

Experience

- **CSIR-Institute of Microbial Technology [🔗]** *July 2019 - July 2020*
  Project Assistant Chandigarh, India
  - Developed a web resource dedicated to COVID-19
  - Performed metagenomic sequencing and its data analysis for investigating changes in virome of sewage samples across Chandigarh
  - Worked on a drug repurposing project involving virus-based diseases using a drug-target network analysis

Education

- **Indraprastha Institute of Information Technology [🔗]** *January 2021 - Ongoing*
  Doctor of Philosophy New Delhi, India
  - CGPA: 8.29/10.00
  - Thesis Title: Unraveling RNA Functionality: Predictive Models for Subcellular Localization and Protein Interactions
- **Indian Institute of Science Education and Research[🔗]** *July 2014 - July 2019*
  BS-MS Dual Degree Kolkata, India
  - CGPA: 7.11/10.00
  - Thesis Title: Transcriptomic meta-analysis of Human sepsis and other related illnesses

Academic Research Projects

- **Master Thesis: Transcriptomic meta-analysis of Human sepsis and other related illnesses** *May 2018 - June 2019*
  Tools: R, RStudio, Bash, HPC Computing
  - Used publically available Sepsis microarray datasets from GEO
  - Used R and RStudio to analyse the data
  - Validated whether the upregulation of the Osteoclast Differentiation Pathway is specific to Sepsis or it is a generic response to inflammation.
- **PhD Project 1: MRSLpred - Prediction of subcellular localization of human mRNA** *January 2022*
  Tools: Python, ML Algorithms(scikit-learn), Webserver Development(Linux, Apache, PHP) [🐙]
  - Used subcellular localization data from RNALocate database
  - Implemented machine learning models on features generated using NFeature
  - Improved prediction acurracy using MERCI motifs
  - Achieved an average AUROC (for 6 locations) of 0.742
- **PhD Project 2: CytoLNCpred - Prediction of subcellular localization of lncRNA in 15 cell-lines** *January 2023*
  Tools: Python, ML Algorithms(scikit-learn), Webserver Development(Linux, Apache, PHP), DNABERT-2 (LLM) [🐙]
  - Data acquired from lncAtlas database
  - Extracted composition/correlation based features as well as embeddings
  - Deployed machine learning as well as LLMs for prediction
  - Achieved an average AUROC (for 15 cell-lines) of 0.709
- **PhD Project 3: A LLM for Predicting PDAC patients from Blood Exosomal Transcriptomics Data** *June 2024*
  Tools: Python, PeptideBERT, ProtBERT, ESM2, BLAST, MERCI, Scikit-learn, PyTorch [🐙]
  - Converted gene expression profiles of 50 genes from 284 PDAC and 217 non-PDAC patients into sequence representations
  - Fine-tuned LLMs (PeptideBERT, ProtBERT, ESM2) on protein sequence data for classification
  - Evaluated models on an independent dataset, with esm2-t6_8M_UR50D achieving an AUC of 0.95
  - Developed an ensemble model combining LLM-based predictions with BLAST and MERCI alignment-based approaches

PUBLICATIONS

1. Aggarwal, Suchet, Anjali Dhall, Sumeet Patiyal, **Shubham Choudhury**, Akanksha Arora, and Gajendra P. S. Raghava. "An Ensemble Method for Prediction of Phage-Based Therapy against Bacterial Infections." *Frontiers in Microbiology* **14** (2023): 1148579.
2. Bajiya, Nisha, **Shubham Choudhury**, Anjali Dhall, and Gajendra P. S. Raghava. "AntiBP3: A Method for Predicting Antibacterial Peptides against Gram-Positive/Negative/Variable Bacteria." *Antibiotics* **13**(2) (2024): 168.
3. **Shubham Choudhury**, Nisha Bajiya, Sumeet Patiyal, and Gajendra P. S. Raghava. "MRSLpred—a Hybrid Approach for Predicting Multi-Label Subcellular Localization of mRNA at the Genome Scale." *Frontiers in Bioinformatics* **4** (2024): 1341479.
4. **Shubham Choudhury**, Anand Singh Rathore, and Gajendra P. S. Raghava. "Compilation of Resources on Subcellular Localization of lncRNA." *Frontiers in RNA Research* **2** (2024): 1419979.
5. **Shubham Choudhury**, Naman Kumar Mehta, and Gajendra P. S. Raghava. "CytoLNCpred: A computational method for predicting cytoplasm associated long non-coding RNAs in 15 cell-lines." *Frontiers in Bioinformatics*, Sec. RNA Bioinformatics, **5** (2025): 1585794.
6. Dhall, Anjali, Sumeet Patiyal, **Shubham Choudhury**, Shipra Jain, Kashish Narang, and Gajendra P. S. Raghava. "TNFepitope: A Webserver for the Prediction of TNF-α Inducing Epitopes." *Computers in Biology and Medicine* **160** (2023): 106929.
7. Gahlot, Pushpendra Singh, **Shubham Choudhury**, Nisha Bajiya, Nishant Kumar, and Gajendra P. S. Raghava. "Prediction of Plant Resistance Proteins Using Alignment-Based and Alignment-Free Approaches." *Proteomics* **25**(5–6) (2025): e202400261.
8. Gupta, Amit Kumar, Md Shoaib Khan, **Shubham Choudhury**, Adhip Mukhopadhyay, Sakshi, Amber Rastogi, Anamika Thakur, Pallawi Kumari, Manmeet Kaur, and Shalu. "CoronaVR: A Computational Resource and Analysis of Epitopes and Therapeutics for Severe Acute Respiratory Syndrome Coronavirus-2." *Frontiers in Microbiology* **11** (2020): 1858.
9. Kumar, Nishant, **Shubham Choudhury**, Nisha Bajiya, Sumeet Patiyal, and Gajendra P. S. Raghava. "Prediction of Anti-Freezing Proteins From Their Evolutionary Profile." *Proteomics* **25**(3) (2025): e202400157.
10. Kumar, Nishant, Sumeet Patiyal, **Shubham Choudhury**, Ritu Tomer, Anjali Dhall, and Gajendra P. S. Raghava. "DMPPred: A Tool for Identification of Antigenic Regions Responsible for Inducing Type 1 Diabetes Mellitus." *Briefings in Bioinformatics* **24**(1) (2023): bbac525.
11. Mukhopadhyay, Adhip, **Shubham Choudhury**, and Manoj Kumar. "Metaviromic Analyses of DNA Virus Community from Sediments of the N-Choe Stream, North India." *Virus Research* **330** (2023): 199110.
12. Rajput, Akanksha, Anamika Thakur, Amber Rastogi, **Shubham Choudhury**, and Manoj Kumar. "Computational Identification of Repurposed Drugs against Viruses Causing Epidemics and Pandemics via Drug-Target Network Analysis." *Computers in Biology and Medicine* **136** (2021): 104677.
13. Rathore, Anand Singh, **Shubham Choudhury**, Akanksha Arora, Purva Tijare, and Gajendra P. S. Raghava. "ToxinPred 3.0: An Improved Method for Predicting the Toxicity of Peptides." *Computers in Biology and Medicine* **179** (2024): 108926.
14. Rathore, Anand Singh, Nishant Kumar, **Shubham Choudhury**, Naman Kumar Mehta, and Gajendra P. S. Raghava. "Prediction of Hemolytic Peptides and Their Hemolytic Concentration." *Communications Biology* **8**(1) (2025): 176.

SKILLS

- **Programming Languages:** Python, R, Bash, PHP
- **Web Technologies:** Apache, PHP, HTML5, REST APIs
- **Database Systems:** MySQL
- **Data Science & Machine Learning:** scikit-learn, TensorFlow, Keras, PyTorch, pandas, NumPy, Bioconductor
- **Cloud Technologies:** Docker, Singularity
- **DevOps & Version Control:** Git, GitHub, Bash Scripting, Docker
- **Specialized Area:** Computational Biology, lncRNA Analysis, lncRNA–Protein Interaction Prediction, Metagenomic Sequencing, Oxford Nanopore Sequencing, Viral RNA Extraction, Environmental Viromics, Machine Learning for Genomics, Deep Learning, Large Language Models (LLMs)
- **Mathematical & Statistical Tools:** R (tidyverse, ggplot2), SciPy, MATLAB (basic), Excel
- **Basic Biological Laboratory Techniques:** Sample Collection, RNA/DNA Extraction, PCR, Gel Electrophoresis, Centrifugation, Viral Enrichment, Library Preparation for Sequencing, Cell Culture, Western Blot, Cloning
- **Research Skills:** NGS Data Analysis, RNA-seq & Metagenomic Pipeline Development, Viral Genome Assembly, Environmental Sampling, Nucleic Acid Extraction, Scientific Writing, Literature Review, Open-Source Tool Development, PyPI Package Publishing, GitHub Repository Management

Honors and Awards

- **DBT-JRF (Category I)** *June 2020*
  Department of Biotechnology, Government of India [🌐]
  - Awarded Category I under the DBT-Junior Research Fellowship program based on national-level Biotechnology Eligibility Test (BET) merit.
  - Recognized among the top candidates eligible for Ph.D. research in reputed Indian institutions with fellowship support from DBT.
- **DST-INSPIRE Fellowship** *January 2021*
  Department of Science and Technology (DST), Government of India [🌐]
  - Awarded the prestigious DST-INSPIRE Fellowship for pursuing doctoral research, based on outstanding academic performance at the undergraduate and postgraduate levels.
  - Provides financial support and recognition to pursue Ph.D. in basic and applied sciences at recognized institutions across India.

Additional Information

Languages: English (C1 – Advanced), Hindi (B2 – Upper-Intermediate), Odiya (C2 – Proficient / Native-like)

References

1. **Gajendra P.S. Raghava**
   Professor, Department of Computational Biology
   Indraprastha Institute of Information Technology, New Delhi, India
   Email: raghava@iiitd.ac.in
   Phone: +91-11-26907444
   Relationship: PhD Supervisor
2. **Jaspreet Kaur Dhanjal**
   Assistant Professor, Department of Computational Biology
   Indraprastha Institute of Information Technology, New Delhi, India
   Email: jaspreet@iiitd.ac.in
   Phone: +91-11-26907522
   Relationship: PhD Monitoring Committee Member
3. **Manoj Kumar**
   Professor, Department of Pharmacoinformatics
   National Institute of Pharmaceutical Education & Research, S.A.S. Nagar, India
   Email: manojk@niper.ac.in
   Relationship: Project Supervisor