

SAGORIKA NAG

Sagorika.nag.phe19@itbhu.ac.in | (+91) 8866062293 | [Linkedin](#) | [Website](#) | [Twitter](#)

EDUCATION

Indian Institute Of Technology (BHU) Varanasi

Integrated Dual Degree in Pharmaceutical Engineering and Technology

Aug'19 - Present

CPI: 9.36/10, Class Rank: 2

RESEARCH EXPERIENCE

Dmitry Frishman Lab, TU Munich, School Of Life Sciences, Germany (Intern)

May'23-

TCR-epitope binding prediction by exploiting AI and pre-trained protein language model (pLM) embeddings

- Collected binding and non-binding TCR CDR3 β - Epitope pairs from publicly available datasets.
- Generated fixed-size sequence embeddings for each TCR and epitope sequence using a state-of-art pLM.
- Embeddings for each pair were merged and used as an independent feature in the prediction model.
- Using Random forest, we achieved an accuracy of 96%, significantly outperforming other models.

Outcome/Impact: The model aids in the identification of antigens for precision immunotherapy.

Tools/Skills Used: Python, Deep Learning, Machine learning, Immunology

Supervisors: Prof. Dr. Dmitry Frishman, Dr. Barnali Das,

Intelligent Data Ecosystems Department, Rothamsted Research, England (Remote Intern) *Jun'22-April'23*

Pangenomic analysis in wheat to Identify Genetic factors controlling frost tolerance (FroT)

- A FroT-related Quantitative Trait Loci (QTL) on the 5A chromosome was examined in 17 cultivars.
- After extracting the QTL using Scree and BLAST, Structural Variation (SV) was studied using dot plots.
- Homologous gene relationships were identified using OMA and visualized by Circos.
- SVs and mutations in protein sequences were correlated with the phenotypic data provided by RAGT.

Outcome/Impact: Identified potential FroT-controlling genes, Pipeline to study QTLs in a pangenome developed.

Tools/Skills Used: Python, Linux, SQLite, MUMmer4, Scree, OMA, BLAST, Circos, Comparative genomics

Supervisors: Dr. Dan Smith, Dr. Keywan Hassani-Pak

DasSarma Lab, University Of Maryland Baltimore, US (Visiting Scholar, remote)

Oct'21-May'23

Genomic Analysis of Haloarchaea Reveals Diversity of UV Survival and Photolyase Gene Variants

- Examined the UV-resistant properties of nine diverse Haloarchaea in presence of photo-reactivating light.
- Photolyase gene, responsible for photorepair, was analyzed for phylogeny, synteny, and structural studies.
- Diversity in UV resistance phenotypes was correlated with the identified mutations and isolation location.
- Key photolyase residues essential for function were found in strains isolated from the surface brines.

Outcome/Impact: The study aids in understanding mechanisms of UV survival on Earth and other planets, like Mars.

Tools/Skills Used: SWISS-MODEL, DeepView, Clustal Omega, Inkscape, Needleman Wunsch Alignments

Supervisors: Prof. Shiladitya DasSarma, Priya DasSarma

Khattri Lab, IIT-BHU (Master's student)

Feb'22-

1. Identifying Somatic Mutations in Indian Oral Cancer Patients Using RNA-Seq Data

- Analyzed WES and RNA-seq data of Indian Oral Cancer Patients to identify and compare somatic mutations.
- Generated critical visualizations, including oncoplots, lollipop plots, etc., for comparative studies.
- Established RNA-seq's capability to detect novel somatic mutations complementing WES studies.

Outcome/Impact: The study suggests using RNAseq for variant detection to ease patients' financial burdens.

Tools/Skills Used: R programming, Linux, Maftools, cBioPortal, IGV

2. Integrated gene expression analysis to identify targets for Immunotherapy in HNSCC

- Examined the role of 14 *SIGLEC* genes in Head and Neck Cancer using RNA-seq and Microarray data.
- DGE analysis revealed higher expression of *SIGLECs* in tumor as compared to the matching normal.
- Enrichment of immune cells, clinical parameter association studies, etc., were conducted on subgroups.

Outcome/Impact: The study suggests *SIGLECs* as potential candidates for ICB immunotherapy.

Tools/Skills Used: R programming, Microarray data analysis, PTM Biology.

Supervisor: Dr. Arun Khattri

PUBLICATIONS AND CONFERENCE PRESENTATIONS

1. **Nag, S.,** DasSarma, P., Crowley, D.J., Hamawi, R., Tepper, S., Anton, B.P., Guzmán, D. and DasSarma, S. (2023). Genomic Analysis of Haloarchaea from Diverse Environments, including Permian Halite, Reveals Diversity of Ultraviolet Radiation Survival and DNA Photolyase Gene Variants. *Microorganisms*. [Link](#)
2. **Nag, S.,** Mandal, A., Joshi, A., Jain, N., Srivastava, R.S., Singh, S. and Khattri, A. (2022). Sialyltransferases and Neuraminidases: Potential Targets for Cancer Treatment. *Diseases*. [Link](#)
3. **Nag, S.,** Baidya, A. T. K., Mandal, A., Mathew, A. T., Das, B., Devi, B., & Kumar, R. (2022). Deep learning tools for advancing drug discovery and development. *3 Biotech*. [Link](#)
4. Bhattarai, S., Kumar, R., **Nag, S.,** & Namasivayam, V. (2022). Big Data in Drug Discovery. *Machine Learning and Systems Biology in Genomics and Health*, 17–48. [Link](#)
5. Tiwari, J.K., **Nag, S.,** Nizamuddin, S., Singh, S., Hess, J., Seiwert, T.Y., and Khattri, A. Integrated gene expression analysis reveals SIGLECs as new targets for Immunotherapy in Head and Neck Squamous Cell Carcinoma. [\[Manuscript in Preparation\]](#)
6. Negi, S., **Nag, S.,** and Khattri, A. Utility of RNASeq data as an alternative to detect Somatic mutations in Oral cancer patients. [\[Manuscript in Preparation\]](#)
7. ProtTCR: TCR-epitope binding prediction by exploiting AI and pre-trained protein language model embeddings. [\[Manuscript in Preparation\]](#)
8. DasSarma, S., DasSarma, P., Kimmance, O., **Nag, S.,** Soto, L. M., Anton, B. P., & Crowley, D. J. (2022, May 18). Bioinformatic Approaches to Understanding the Evolution and Survival of Extremophiles. *Agu.confex.com; AGU*. [Link](#)
9. **Nag, S.,** DasSarma, P., & DasSarma, S. (2023, Feb 4). Bioinformatic Approaches to Understanding Photorepair Capabilities of Diverse Halophilic Archaea. Inter IIT Tech Meet 11.0 's STUDENTS' ACADEMIC CONFERENCE (SAC). [Link](#)

TEACHING AND VOLUNTEERING EXPERIENCE

- Presented at **BlueSciCon'2021**, a scientific seminar series conducted by Blue Marble Space, Institute Of Science (BMSIS), on the topic: “**DNA repair in extremophiles.**” [Link](#) (27th Aug 21)
- Delivered a presentation on “**Scope of Bioinformatics and Interdisciplinary Research**” to high school students at Jamnabai Narsee School, Gujarat, India. (27th July 23)

HONORS AND ACADEMIC ACHIEVEMENTS

- **IBGAA GSAS 2023** - Graduate Study Application Scholarship by the IIT-BHU Alumni Association
- **DAAD-WISE'23 Scholarship**, 68-day fully funded internship at TU Munich, School Of Life Sciences.
- **Mitacs Globalink Research Internship'23** - 12-week fully funded opportunity at Laval University.
- Selected for **Young Scientist Program'21**, conducted by Blue Marble Space, Institute of Science (BMSIS).
- Qualified **IIT (JEE) Advance 2019**, amongst the top 0.05% out of 1.2 million applicants.

EXTRA-CURRICULARS

- **Member of Genelab Multi-Omics AWG** - Networking with researchers working on astrobiological multi-omics and building impactful projects in less discussed areas.
- **Volunteer at IIT-BHU Research Community** - Served as Social Media head and later as AMA (Ask Me Anything) sessions head, aiding students in securing Indian and foreign research internships.

RELEVANT COURSES

- **Academic Courses:**
Fundamentals of Bioinformatics (Ongoing), Biology, Computer Programming, Human Physiology, Probability and statistics, Microbiology and biotechnology, Principles and techniques in Molecular Biology, Structural Biology, Computational Chemistry, Computational Drug Design.
- **Online courses:**
Machine Learning Specialization (Andrew Ng), Machine Learning A-M: Hands-On Python & R In Data Science (Udemy), Data Visualization and SQL (Kaggle), Genome Sequencing I&II (UCSD), Python for Genomic Data Sciences (JHU), Python skills for handling biological data (Stepik), Writing in the Sciences (Stanford).