Jyotirmoy **Das**

Principal Research Engineer

Befälsgatan 12, 587 50, Linköping, Sweden

he/him/his | +46 (0)76-136 96 451 | jyotirmoy21@gmail.com | jyotirmoydas.netlify.app | 0000-0002-5649-4658 | M-3134-2016 | IMBYOv8AAAAJ | Jyotirmoy-Das-3 | JD2112 | dasjyotirmoy | jyotirmoy21

Computational Biologist, Bioinformaticist

About me

Highly effective computational biologist with doctoral and post-doctoral work focussing on Bioinformatics pipeline development and biological data analysis. Research experience in development and implementation of computational tools for *multi-omics* data analysis, high-throughput image analysis. At present, I am a Principal Research Engineer in the Bioinformatics Core Facility of Faculty of Medicine and Health Sciences at Linkoping University, Sweden. I am engaged with developing scripts to automatise and analysed different biological big data analysis from clinical samples using *Illumina®arrays*, *whole-genome sequencing*, *Nanopore®long-read sequencing*, *Twists' targeted panels* with **R**, **RStudio**, **Shiny**, **Python**, **tidyverse** and visualisation of results.

Area Of Expertise

- **Genomics:** Structural and functional genomics, Whole genome analysis.
- **Epigenomics:** DNA Methylation, microRNAs.
- **Transcriptomics:** RNA microarray, sequence analysis, expression analysis, miRNA microarray.
- **Metabolomics:** Metabolite expression analysis.
- **Proteomics:** Protein expression and protein-protein interaction networks.
- **Downstream analysis:** GO, Pathways, Networks.

Research Experience

Principal Research Engineer — *Linkoping University, Sweden*

Bioinformatics, Core Facility, Facility of Faculty of Medical and Health Sciences — *Sep. 2020 - present*

- Key Responsibilities
- Developing bioinformatics pipeline using Python, R for standardization of different DNA, RNA analyses.
- Developing semi/automated Nextflow workflow to analyze the SARS-CoV-2 whole genome data using Illumina and Nanopore sequencing technologies.
- Developing Shiny based graphical user interface to analyze Illumina array-based DNA methylation and downstream analyses.
- Shiny server setup using CentOS.
- Developing source code on GitHub for general public use.
- Linux system administration, user group maintenance and setup workstation for development and production.
- User and project investigator on different HPCs (Tetralith on NSC, Rackham on Uppmax and NGP).
- Designed course curriculum and carried out course on Bioinformatics for Microbiology at the Medical Microbiology course (8BKG24) at Linköping University, Sweden.
- Key Achievement
- Designed and developed own projects.
- Networking with Clinical Genomics, Sweden and Genomic Medicine, Sweden groups.
- Extensive user of contenarized (Docker, Singularity, Conda) VM on different OS architectures.
- Member and owner of Genomic Medicine, Sweden DockerHub organization.
- Active maintainer of Genomic Medcine, Sweden SARS-CoV-2 (GMS-Artic) pipeline group.
- Member of GMS Microbiology group.
- Published 10 peer-reviewed research articles, and another 2 manuscripts under reviewed on reputed journals.

Postdoctoral Fellow *Linkoping University, Sweden*

Department of Clinical and Experimental Medicine, Linköping University, Sweden *Jun. 2017 - Aug. 2020*

- Key Responsibilities
- Computational analysis of whole genome (WGBS/RRBS)/ array based (450K/850K) DNA methylation analysis from different participants in Sweden and Peru.
- Develop analysis pipelines to identify differential methylation patterns in different groups of dataset.
- Develop pipelines with unsupervised analyses like Multi-Dimensional Analysis (MDA), Principal Component Analysis (PCA), Component Analysis (CA), Multiple Factor Analysis (MFA) to reduce the dimensionality of data.
- Pipeline development to use supervised machine learning algorithms (LDA, GLMNET, RandomForest) to identify biosignature from the study.
- Develop analysis pipeline for transcriptome analysis.
- Use of MATLAB image processing to analyze live cell imaging.
- Key Achievements
- Extensive use of R language.
- Develop packages in R, notes on R Markdown.
- High performance computing analyses.
- Clustering computing with R and Shell.
- Supervising 2 Ph.D. students
- Supervised (main supervisor) two master degree students.
- Co-supervised 4 bachelor and master degree students.
- Published two first authors, one shared first co-authors and two other co-authored research articles in peer-reviewed journals.

Senior Research Fellow, DST-INSPIRE (Govt. of India) *West Bengal, India*

Bioinformatics Centre, Bose Institute *Jan. 2014 - Jan. 2017*

- Key Responsibilities
- Develop of analysis pipelines to identify different microRNAs (miRNAs) in human diseases and their roles in evolutionary perspective.
- Curation of databases to develop dataset to calculate the evolutionary rates in human and orthologs.
- Linux-based approach to search sequence homology.
- Use of different programming languages like Perl, Python to analyze dataset.
- Key achievements
- Extensive statistical software analyses with SPSS.
- Develop algorithm based on shell scripts
- Linux system maintenance

Junior Research Fellow, DST-INSPIRE (Govt. of India) *West Bengal, India*

Bioinformatics Centre, Bose Institute *Jan. 2012 - Jan. 2014*

Institute Fellow *West Bengal, India*

Bioinformatics Centre, Bose Institute *Sep. 2010 - Jan. 2012*

Research Trainee *New Delhi, India*

National Institute of Plant Genome Research, New Delhi *Jul. 2009 - May. 2010*

Skills

Programming	R, Python, Perl, Shell, C/C++, HTML, Rstudio, Shiny, DASH.
OS handling	Linux/Unix, MacOS, Windows, HPC, RedHat, amazon and google cloud.
Reproducible Report	Markdown/Rmarkdown, Quarto, R shiny apps, Jupyter Notebook, LaTeX, Pandoc, lua.
Statistical analysis	SPSS, MATLAB, GraphPad, STATA, OriginLab.
Image processing	ImageJ, CellProfiler, Photoshop, GIMP, Inkscape.
Front-End	HTML/CSS/JS, WordPress, Shiny, DASH.
Containers	Conda/Anaconda, Singularity/Apptainer, Docker, Podman.
Wrokflow manager	NextFlow, Snakemake, CWL
Version control	git, bitbucket.
Office tools	Open office, google docs, Microsoft office, pages, keynote, numbers.
Quantitative Research	t-test, within-/between-subjects/Repeated Measures ANOVAs, Regressions, Clustering, HLM, Factor Analysis, Network Analysis, SEM, PCA, MDS, Unsupervised/Supervised Machine Learning
Editors	vim, nano, gedit, VS Code, VS codium, DataSpell, notepad++.
Mixed Research	Text Mining, Explanatory/Exploratory Data Analysis (EDA)

Education

Ph.D. (Technology) in Bioinformatics *West Bengal, India*

Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology), India *2017*

- Dissertation Title: "In-silico studies of MicroRNA Regulations in higher eukaryotes from the perspective of Molecular Evolution."
- Supervisor: Prof. Tapash C Ghosh, Bose Institute, India.

Master of Technology in Biotechnology *West Bengal, India*

West Bengal University of Technology, India *2010*

- Title of the Thesis: "Identification and evaluation of Intron Length Polymorphism (ILP) marker in Foxtail millet (Setaria italica) and to study evolutionary relationship between species by cross-species transferability".
- Supervisors: Dr. Manoj Prasad, NIPGR, India. Prof. Nandan Bhattacharyya, Haldia Institute of Technology, India

Master of Science in Biotechnology *Karnataka, India*

Bangalore University *2008*

Bachelor of Science in Biotechnology *West Bengal, India*

University of Kalyani *2006*

- Project 1: "Drug Designing: Evaluation of iron chelation through blood in phytic acid treated thalassemia patients".
- Project 2: "Study of Ecological Diversity in Sunderbans, Especially on Kingfishers"
- Supervisors: Prof. Amit Chakravorty, Dr. Sudipa Chakravorty, IGE, India

Certificates

Research Supervision

Linköping University Pedagogy Course *2020*

How to Lead a Research Group

Certificate *2019*

Becoming a Teacher in Higher Education

Linköping University Pedagogy Course *2019*

Astrobiology and the Search for Extraterrestrial Life The origin and evolution of life and the search for life beyond the Earth.

Coursera e-learning *2015*

Programming for Everybody (Getting Started with Python)

Coursera e-learning *2015*

Publications

∗: Shared first-author publication

Refereed Journal Papers

1. Braian, C., Karlsson, L., **Das, J.**, & Lerm, M. (2023). Selected beta-glucans act as immune-training agents by improving anti-mycobacterial activity in human macrophages-a pilot study. In *Journal of Innate Immunity*.

2. Lundquist, H., Andersson, H., Chew, M. S., **Das, J.**, Turkina, M. V., & Welin, A. (2023). The olfactomedin-4-defined human neutrophil subsets differ in proteomic profile in healthy individuals and patients with septic shock. *Journal of Innate Immunity*, *15*(1), 351–364.

3. Shahin, H., Abdallah, S., **Das, J.**, He, W., El-Serafi, I., Steinvall, I., Sjoberg, F., Elmasry, M., & El-Serafi, A. T. (2023). miRNome and proteome profiling of human keratinocytes and adipose derived stem cells proposed miRNA-mediated regulations of epidermal growth factor and interleukin 1-alpha. *International Journal of Molecular Sciences*, *24*(5), 4956.

4. Verma, D., Kasic, N.-K., Jeppsson, F., Eding, C. B., Lysiak, M., Fekri, S. Z., **Das, J.**, & Enerback, C. (2023). 815 altered methylation of microRNA in the psoriatic epidermis highlights the wnt pathway. *Journal of Investigative Dermatology*, *143*(5), S140.

5. Verma, D., Kasic, N.-K., Jeppsson, F., Eding, C. B., Lysiak, M., Fekri, S. Z., **Das, J.**, & Enerback, C. (2023). Differential DNA methylation of miRNA-encoding genes in psoriatic epidermis highlights the wnt pathway. *The Journal of Investigative Dermatology*, S0022–202X.

6. Volpe, M., & **Das, J.** (2023). methylR: A graphical interface for comprehensive DNA methylation array data analysis.

Bioinformatics, *39*(4), btad184.

7. Huoman, J., Sayyab, S., Apostolou, E., Karlsson, L., Porcile, L., Rizwan, M., Sharma, S., **Das, J.**, Rosen, A., & Lerm, M. (2022). Epigenetic rewiring of pathways related to odour perception in immune cells exposed to SARS-CoV-2 in vivo and in vitro. *Epigenetics*, *17*(13), 1875–1891.

8. Lundquist, H., Andersson, H., Chew, M. S., **Das, J.**, Turkina, M. V., & Welin, A. (2022). The Olfm4-defined human neutrophil subsets differ in proteomic profile in septic shock. *bioRxiv*.

9. Lysiak, M., **Das, J.**, Malmstrom, A., & Soderkvist, P. (2022). Methylation associated with long-or short-term survival in glioblastoma patients from the nordic phase 3 trial. *Frontiers in Genetics*, *13*, 934519.

10. Lysiak, M., **Das, J.**, Soderkvist, P., & Malmstrom, A. (2022). Methylome analysis of short versus long-term GBM survivors from the nordic randomised, phase 3 trial. *Brain Tumor Research and Treatment*, *10*(Suppl).

11. Pehrson, I., Sayyab, S., **Das, J.**, Idh, N., Paues, J., Mendez-Aranda, M., Ugarte-Gil, C., & Lerm, M. (2022). *DNA methylomes derived from alveolar macrophages and alveolar t cells display distinct patterns in tuberculosis–a future precision tool for TB status determination?*

12. Pehrson, I., Sayyab, S., **Das, J.**, Idh, N., Paues, J., Mendez-Aranda, M., Ugarte-Gil, C., & Lerm, M. (2022). The spectrum of tuberculosis described as differential DNA methylation patterns in alveolar macrophages and alveolar t cells. *Clinical Epigenetics*, *14*(1), 1–12.

13. **Das, J.**, Idh, N., Sikkeland, L. I. B., Paues, J., & Lerm, M. (2021). DNA methylome-based validation of induced sputum as an effective protocol to study lung immunity: Construction of a classifier of pulmonary cell types. *Epigenetics*, 1–12.

14. Huoman, J., Sayyab, S., Apostolou, E., Karlsson, L., Porcile, L., Rizwan, M., Sharma, S., **Das, J.**, Rosen, A., & Lerm, M. (2021). *Epigenome-wide DNA methylation profiling of healthy COVID-19 recoverees reveals a unique signature in circulating immune cells*.

15. Huoman, J., Sayyab, S., Apostolou, E., Karlsson, L., Porcile, L., Rizwan, M., Sharma, S., **Das, J.**, Rosen, A., & Lerm, M. (2021). Mild SARS-CoV-2 infection modifies DNA methylation of peripheral blood mononuclear cells from COVID-19 convalescents. *medRxiv*, 2021–2007.

16. Kalsum, S., Andersson, B., **Das, J.**, Schon, T., & Lerm, M. (2021). A high-throughput screening assay based on automated microscopy for monitoring antibiotic susceptibility of mycobacterium tuberculosis phenotypes. *BMC Microbiology*, *21*(1), 1–14.

17.* Karlsson, L., **Das, J.**, Nilsson, M., Tyren, A., Pehrson, I., Idh, N., Sayyab, S., Paues, J., Gil, C. U., Aranda, M. M., & Lerm, M. (2021). A differential DNA methylome signature of pulmonary immune cells from individuals converting to latent tuberculosis infection. *Scientific Reports*, *11*(1), 19418.

18. Pehrson, I., Braian, C., Karlsson, L., Idh, N., Danielsson, E. K., Andersson, B., Paues, J., **Das, J.**, & Lerm, M. (2021). DNA methylation profiling of immune cells from tuberculosis-exposed individuals overlaps with BCG-induced epigenetic changes and correlates with the emergence of anti-mycobacterial "corralling cells." *medRxiv*, 2021–2009.

19. Zhu, G. H., Azharuddin, M., Islam, R., Rahmoune, H., Deb, S., Kanji, U., **Das, J.**, Osterrieth, J., Aulakh, P., Ibrahim-Hashi, H., Manchanda, R., Nilsson, P. H., Mollnes, T. E., Bhattacharyya, M., Islam, M. M., Hinkula, J., Slater, N. K., & Patra, H. K. (2021). Innate immune invisible ultrasmall gold nanoparticles—framework for synthesis and evaluation. *ACS Applied Materials & Interfaces*, *13*(20), 23410–23422.

20. **Das, J.**, Verma, D., Gustafsson, M., & Lerm, M. (2019). Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naive subjects. *Epigenetics*, *14*(6), 589–601.

21. Sen, K., Bhattacharyya, D., Sarkar, A., **Das, J.**, Maji, N., Basu, M., Ghosh, Z., & Ghosh, T. C. (2018). Exploring the major cross-talking edges of competitive endogenous RNA networks in human chronic and acute myeloid leukemia. *Biochimica Et Biophysica Acta (BBA)-General Subjects*, *1862*(9), 1883–1892.

22. **Das, J.**, Podder, S., & Ghosh, T. C. (2014). Insights into the miRNA regulations in human disease genes. *BMC Genomics*, *15*(1), 1–7.

23. **Das, J.**, Chakraborty, S., Podder, S., & Ghosh, T. C. (2013). Complex-forming proteins escape the robust regulations of miRNA in human. *FEBS Letters*, *587*(14), 2284–2287.

24. Gupta, S., Kumari, K., **Das, J.**, Lata, C., Puranik, S., & Prasad, M. (2011). Development and utilization of novel intron length polymorphic markers in foxtail millet (setaria italica (l.) P. beauv.). *Genome*, *54*(7), 586–602.

Manuscript sumitted/in preparation

1. Shahin, H., Belcastro, L., **Das, J.**, Grigoriadi, M. P., Saager, R. B., Steinvall, I., Sjöberg, F., Olofsson, P., Elmasry, M., & El-Serafi, A. T. (2023). MicroRNA-155 mediates multiple gene regulations pertinent to the role of human adipose-derived mesenchymal stem cells in skin regeneration. In *Advances in Wound Care*.

Patents

1. Lerm, M., & **Das, J.** (2022). *Biomarker for detection of mycobacterial exposure and infection*. *WO Patent WO2022169394A1*.

2. Lerm, M., **Das, J.**, & Sayyab, S. (2022). *Method for determining sars-cov-2 exposure with or without remaining symptoms*. *WO Patent WO2022119495A1*.

Grants

- "Travel Grant awarded of 9920 SEK to attend 4th International Conference on Innate Immune Memory, Nijmegen, the Netherlands. 31 October- 1 November", Primary awardee, Oct 2019.
- "Proof-of-Concept Grant from Swedish Research Council (VR) with a funding of 2 MSEK.", Participant Researcher, Dec. 2022 - Dec. 2024.

Poster Presentations and Invited Talks

Das, J. (2023, October). *methylR: a graphical interface for comprehensive DNA methylation array data analysis*. Poster presented at the Scilifelab day, Linköping.

Das, J. (2023, May). *methylR: a graphical interface for comprehensive DNA methylation array data analysis*. Workshop oral presentation organized by Clinical Genomics Linköping.

Das, J. (2023, September). *GMS-Artic: A containerized NextFlow pipeline for detecting Pangolin Typing by analysing SARS-CoV-2 whole genome from short and long sequence reads*. Poster presentation at JIM Symposium Precision medicine in Europe 2022.

Das, J. (2022, October). *gms-artic: features and applications*. oral presentation talk at the GMS-Micro workshop in Karolinska Institute, Sweden.

Das, J. (2021, October) *Visualization, front-end solution.* lightening-fast oral presentation of group discussion in Clinical Genomics Platform retreat at Sigtuna, Sweden.

Das, J. (2021, September). *gms-artic: the pipeline*. Workshop presentation organized by Genomic Medicine Sweden.

Das, J. (2021, June). *How to run a pipeline on NGP server: with a demonstration of gms-artic*. Workshop oral presentation organized by Genomic Medicine Sweden.

Das, J. (2019, October). *Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naïve subjects*. Poster presentation at 4th International Conference on Innate Immune Memory organized by Roudbound University, the Netherlands.

Das, J. (2019, May). *Altered methylation of microRNA (miRNA) in the psoriatic epidermis* Poster presentation in Society for Investigative Dermatology (SID), Chicago, USA.

Das, J. (2019, April). *Epigenetic patterns predicting the response to tuberculosis vaccination*. One-minute presentation to market your research organized by Mucosal Infection and Inflammation Center (MIIC), Linköping University. *Best Presenter Award*.

Das, J. (2018, November). *Advanced image analysis with MATLAB*. Oral presentation organized by NDPIA and Linköping University.

Das, J. (2018, October). *Identification of DNA methylation patterns predisposing for a trained immunity response to BCG vaccination in healthy BCG-naïve subjects*. Poster presentation on TB research seminar organized by Centre for Tuberculosis Research, Karolinska Institute.

Das, J. (2017, August). *High-content screening of live-cell imaging and analysis of Mycobacterium marinum-infected Dictyostellim discoideum*. Poster presentation on International TB conference in Stockholm.

Software Developments and Publications

Open-Source Project on GitHub

TranscriptR: shiny-based transcriptome analysis and visualisation *https://github.com/JD2112/transcriptr*

Shiny, R, Snakemake, Docker, Singularity *2024*

- Developer of the application

MethylR: a single shiny solution from sequencer data to pathway analysis *https://github.com/JD2112/methylr*

Shiny, R, Docker, Singularity *2023*

- Developer of the application

ComplexPCA: a ggplot PCA with complex annotation *https://github.com/JD2112/ComplexPCA*

R, Rstudio *2021*

- Developer of the application

Alveolar-Cell-Type-Deconvolution: DNA methylome-based validation of induced sputum as an effective protocol to study lung immunity: construction of a classifier of pulmonary cell types *https://github.com/JD2112/Alveolar-Cell-Type-Deconvolution*

R, Rstudio *2019*

- Developer of the application

Image-Processing-MATLAB: Acquired time lapse images from the IncuCyte® and processed with MATLAB to create a time-lapse video of one particular aggregate of bacteria *https://github.com/JD2112/Image-Processing-MATLAB*

MATLAB *2020*

- Developer of the application

JASEN: Epitypification pipeline for clinical NGS data. Written in NextFlow, Python & Bash. *https://github.com/JD2112/JASEN*

NextFlow, Python, Conda, Docker, Singularity, Bash *2021*

- Co-developer

gms-artic: A Nextflow pipeline for running the ARTIC network's fieldbioinformatics tools *https://github.com/JD2112/gms-artic*

NextFlow, Python, Conda, Docker, Singularity, Bash, R *2020*

- Co-developer

Teaching Experience

Bioinformatics for Microbiology Course (8BKG24) *Linköping University, Sweden*

Course lecture *2021 - 2023*

- Teaching (with designing the curriculum, examination) Bioinformatics for Microbiology to the course Medical Microbiology, Linköping University, Sweden

3 weeks R workshop for PhD students *Linköping University, Sweden*

R Workshop *Sep. 2019*

- Organise the R fundamental workshop for the PhD students (total 18 participants) to understand the data analysis, management, statistical analysis and graphical representation of multi-variate data in R in 3 weekly sessions. The workshop is supported by Forum Scientium, LiU

Random Forest model to biomarker identification *Linköping University, Sweden*

Lecture *May. 2019*

- One lecture to visiting undergraduate students from Purdue University, USA.

Live imaging of intracellular infections *Linköping University, Sweden*

NDPIA workshop *Dec. 2018*

- Mentoring and co-working with a group of students at NDPIA one week's workshop at Department of Clinical and Experimental Medicine.

Expert reviewer of master thesis *Linköping University, Sweden*

Examiner *May. 2018*

- Interactome link between Alzheimer's disease and diabetes mellitus type II

Supervise two masters' projects in Bioinformatics *Linköping University, Sweden*

Supervisor *2018 - 2019*

- to guide two master's theses on the bioinformatics part with DNA methylation analysis in R and fundamental of Linux.

Co-supervise one master's project on Image Analysis *Linköping University, Sweden*

Co-supervisor *2018*

- guide one master's thesis with image analysis and statistical calculations with live imaging techniques.

Bioinformatics for Microbiology *Bidhannagar Govt. College, WB, India*

Teaching *Jan. 2015 - May. 2015*

- One semester Bioinformatics course curriculum design, deliver lectures and set examination for a group of masters (microbiology) students.

Supervision

Main supervisor: master degrees project student *Linköping University, Sweden*

Supervisor *Jan. 2019 - Jun. 2019*

- Total number of students - 2

Co-supervisor: master degrees project student *Linköping University, Sweden*

Co-supervisor *2019 - 2020*

- Total number of students - 9

Co-supervisor: doctoral degrees student *Linköping University, Sweden*

Co-supervisor *2020 - 2024*

- Total number of students - 2

Awards and Honors

Research Grant Award *Bose Institute, Kolkata*

Doctoral Fellowship *Jan. 2012 - Jan. 2017*

- Awarded by DST-INSPIRE, Government of India for pursuing doctoral research

Institute Scholarship *Bose Institute, Kolkata*

Doctoral Fellowship *Sep. 2010 - Jan. 2012*

- National Scholarship awarded to pursue PhD program by Bose Institute, Kolkata - an autonomous body undertaken by Govt of India.

University Gold Medal *MAKAUT, WB*

University Award *2010*

- Awarded University Gold Medal for the topper of the batch, M.Tech (BT), 2010.

Graduate Aptitude Test in Engineering (GATE) *HIT, WB*

Master Fellowship *2008 - 2010*

- Awarded by Ministry of Human Resources Development, GOI to pursue Master of Technology (Biotechnology).

Projects

Current Projects

1. GMS-Artic: A containerized NextFlow pipeline for detecting Pangolin Typing by analysing SARS-CoV-2 whole genome from short and long sequence reads GitHub link

2. Developing a nextflow pipeline for DNA methylation from Nanopore sequencing

1. methylR: a single shiny solution from sequencer data to pathway analysis
GitHub link

We introduce methylR, a complete pipeline for the analysis of both 450K and EPIC Illumina arrays which not only offers data visualization and normalization but also provide additional features such as the annotation of the genomic features resulting from the analysis, pairwise comparisons of DMCs with different graphical representation plus functional and pathway enrichment as downstream analysis, all packed in a minimal, elegant and intuitive graphical user interface which brings the analysis of array DNA methylation data.

Related publication: Volpe, M & **Das, J.** methylR: a graphical interface for comprehensive DNA methylation array data analysis. 2022 (Under review in Bioinformatics, Oxford)

2. Process-engineered synthesis of ultrasmall gold nanoparticles
Collaborators: Dr. Hirak Patra
Concept & design:

Related publication: Zhu GH, Azharuddin M, Islam R, Rahmoune H, Deb S, Kanji U, **Das, J.**, Osterrieth J, Aulakh P, Ibrahim-Hashi H, Manchanda R. Innate Immune Invisible Ultrasmall Gold Nanoparticles—Framework for Synthesis and Evaluation. ACS Applied Materials & Interfaces. 2021 May 12;13(20):23410-22. Link to the publication

3. Methylome analysis for prediction of long and short-term survival in glioblastoma patients
Collaborator: Prof. Peter Söderkvist
Concept & design:

Related publication: Łysiak M, **Das, J.**, Malmström A and Söderkvist P (2022) Methylation associated with long- or short-term survival in glioblastoma patients from the Nordic phase 3 trial. Front. Genet. 13:934519. doi: 10.3389/fgene.2022.934519 Link to the publication

4. Identification of miRNAs related to psoriasis from RRBS data
Collaborators: Dr. Deepti Verma, Prof. Charlotta Enerbeck
Concept & design:

Related publication:

5. Proteomic profiling of Olfactomedin-4 marked human neutrophil subset
Collaborator: Dr. Amanda Welin
Concept & design:

Related publication: Lundquist H, Andersson H, Chew MS, **Das, J.**, Turkina MV, Welin A. The Olfm4-defined human neutrophil subsets differ in proteomic profile in septic shock. bioRxiv. 2022 Link to the publication

1. Identification of DNA methylation patterns predisposing for an efficient response to BCG vaccination in healthy BCG-naïve subjects

The protection against tuberculosis induced by the Bacille Calmette Guérin (BCG) vaccine is unpredictable. In our previous study, altered DNA methylation pattern in peripheral blood mononuclear cells (PBMCs) in response to BCG was observed in a subgroup of individuals, whose macrophages killed mycobacteria effectively ('responders'). These macrophages also showed production of Interleukin-1β (IL-1β) in response to mycobacterial stimuli before vaccination. Here, we hypothesized that the propensity to respond to the BCG vaccine is reflected in the DNA methylome. We mapped the differentially methylated genes (DMGs) in PBMCs isolated from responders/non-responders at the time point before vaccination aiming to identify possible predictors of BCG responsiveness. We identified 43 DMGs and subsequent bioinformatic analyses showed that these were enriched for actin-modulating pathways, predicting differences in phagocytosis. This could be validated by experiments showing that phagocytosis of mycobacteria, which is an event preceding mycobacteria-induced IL-1β production, was strongly correlated with the DMG pattern.

Key responsibilities
✓ Computational pipeline development to Illumina 450K methylation data analysis
✓ Gene-gene interaction network analysis
✓ High-throughput statistical data analysis with Multiple Factorial analysis (MFA)
✓ Manuscript preparation

Key achievements
✓ Extensive use of R and Bioconductor packages
✓ Gene and Protein Interaction network analysis
✓ Performing factorial analysis
✓ Collaboration with different departments and faculty members

2. Advanced image tracking of bacterial biofilms from live cell imaging analysis using MATLAB image processing

GitHub link

Related publication: Kalsum S, Andersson B, Das J, Schön T, Lerm M. A high-throughput screening assay based on automated microscopy for monitoring antibiotic susceptibility of Mycobacterium tuberculosis phenotypes. BMC microbiology. 2021 Dec;21(1):1-4.

3. Machine Learning algorithm to identify bio-signatures from tuberculosis-exposed and non-exposed individuals

Biology is a complex system and human with their more than three billion base-pairs is of course a too complex system. The story doesn't end with the number of base-pairs or nucleotides, within it there is another large more complexity society where everybody has got some works to support those three billion system.

The story of Epigenetics not only tells the action of biological macromolecules, it also learns and modifies itself with the surroundings. In the computer system, the advanced algorithm helps to learn from various kind of data and apply them to another set of the data to generate or validate a hypothesis.

With the Big data, human minds need years to read a set of data whereas the computer can read it faster and along with the algorithm it can also understands the pattern of the data to validate the hypothesis. Artificial Intelligent after all can thus reduce the work load for biologists with the proper design of the algorithm.

Why do this?

Tuberculosis, with the long co-evolutionary history with mankind and more than one million death worldwide per year, is one of the infectious diseases which has only one known vaccine till date. Identification of new bio-signature can show pave to the development of a new vaccine.

Machine learning is used to develop the biosignature from the epigenetics data.

Related publication: Das, J., Idh, N., Pehrson, I., Paues, J., & Lerm, M. (2021). A DNA methylome biosignature in alveolar macrophages from TB-exposed individuals predicts exposure to mycobacteria. medrxiv. Link to the publication

Lerm, M., **Das, J.** (2022) Biomarker for detection of mycobacterial exposure and infection. WO patent. WO2022119495A1

4. Identification of Differential Methylation patterns of tuberculosis patients, household contacts and healthy participants

DNA methylation is one of the epigenetic changes that regulates the function of the genes in human as well as other living organisms. The cytosine base of the DNA molecules sometimes get methylated with one methyl (-CH3) group at its 5′ carbon molecule. In a particular position, the methylation varies due to their different presence in different cells.

Array-based methylation using Illumina or genome-wide methylation analysis generates a lot of data which requires strong computational approaches to identify the differential methylation patterns between the sample groups. The array-based methylation value is calculated mainly using the β value which is a ratio of methylation value over the unmethylation value with a static coefficient. The mathematical equation to calculate the β value is:

where M = methylated value; UM = Unmethylated value, c= constant =100; $-1 \leq \beta \leq 1$ With other epigenetic modifiers like Histone modifications, microRNAs, the main role of the DNA methylation is to regulate the gene expression levels and that also reflects to the protein expressions.

The cytosine with methyl base and a guanine molecule creates a stretch of similar CpG base-pairs that is known as the CpG islands. The stretch can be short or long to several hundred base-pairs together. These CpG islands dispersed over transcription start sites (TSS), promoters, intergenic regions and also gene body regions.

Why do this?

DNA methylation is variable and also depends on the external environments. Exposure to different environmental conditions changes the pattern of the methylation. In this project, the hypothesis is that there is a different methylation pattern between individuals when they are either exposed or non-exposed to tuberculosis. The difference in the methylation pattern among differently exposed individuals can lead to the identification of the responsible genes or pathways that helps to the fast tuberculosis diagnosis as well as the probable treatment procedures.

Related publication: Pehrson, I., Braian, C., Karlsson, L., Idh, N., Danielsson, E.K., Andersson, B., Paues, J., **Das, J.** and Lerm, M., 2021. DNA methylation profiling of immune cells from tuberculosis-exposed individuals overlaps with BCG-induced epigenetic changes and correlates with the emergence of anti-mycobacterial'corralling cells'. medRxiv. Link to the publication

Pehrson, I., **Das, J.**, Idh, N., Karlsson, L., Rylander, H., af Segerstad, H.H., Reuterswärd, E., Marttala, E., Paues, J., Méndez-Aranda, M. and Ugarte-Gil, C., 2021. DNA methylomes derived from alveolar macrophages display distinct patterns in latent tuberculosis-implication for interferon gamma release assay status determination. MedRxiv. Link to the publication

5. Analysis of differential methylation patterns between tuberculosis-exposed and non-exposed individuals using different illumina platforms

6. Reduced Representation of Bisulfite Sequencing (RRBS) data analysis from tuberculosis-exposed samples in different cell types.

A number of methods are available to identify the genome-wide DNA methylation in human and other organisms. Reduced representation of bisulfite sequencing (RRBS) is one of these methods to identify the genome-wide DNA methylation using the MspI enzyme library preparation and whole-genome sequencing.

Bisulfite converted DNA preserves the cytosine bases which are methylated in the genome, but converts the non-methylated cytosine residues to thymine residue and that procedure results to the easy identification of the methylated sites in the genome. The standard procedure follows the library preparation of bisulfite converted DNA from a small input samples (100 ng) and then use the library in the sequencer with the nucleotide molecules to generate the small stretch of sequences saved in the fastq format.

Why do this?

Genome-wide methylation analysis reveals the intergenic regions along with the genic regions and also identifies methylation patterns other than CpG islands. These illustrates huge scope to identify the new features of the causing disease patterns among the individuals.

Related publication: Karlsson L, Das J, Nilsson M, Tyrén A, Pehrson I, Idh N, Sayyab S, Paues J, Ugarte-Gil C, Méndez-Aranda M, Lerm M. A differential DNA methylome signature of pulmonary immune cells from individuals converting to latent tuberculosis infection. Scientific reports. 2021 Sep 30;11(1):1-3. Link to the publication

7. RNA sequencing data analysis using genome-wide transcriptome data from tuberculosis-exposed samples in high endemic country

8. Identification of exosomal-derived microRNAs from Mycobacterium tuberculosis infected cells

9. SARS-CoV-2 and DNA Methylation
GitHub link

Related publication: Huoman J, Sayyab S, Apostolou E, Karlsson L, Porcile L, Rizwan M, Sharma S, **Das, J.**, Rosén A, Lerm M. Epigenetic rewiring of pathways related to odour perception in immune cells exposed to SARS-CoV-2 in vivo and in vitro. Epigenetics. 2022 Jun 26:1-7. Link to the publication

Maria Lerm, Jyotirmoy Das, Shumaila Sayyab. Method for determining sars-cov-2 exposure with or without remaining symptoms. 2022. WO2022119495A1

In-silico studies of MicroRNA Regulations in higher eu- karyotes from the perspective of Molecular Evolution
Supervisor: Prof. Tapash C.Ghosh
(2012-01 to 2017-01)

The evolution of miRNA genes, identification of numerous new miRNAs and miRNA-targeted genes, expansion of miRNAs during animal evolution, functions of miRNAs – are the central questions for molecular biologists for the last decade. How these 22-28 nucleotides change the progression of cellular and physiological developments; helps to correlate with the protein complexes or in the advancement of the diseases in case of human. In this thesis, I have studied the association of miRNAs comparing with the protein complexes resided in the interaction network and analyzed their evolutionary rates from the perspective of miRNA regulation. I found that these "tiny" miRNAs weakly regulate the proteins that are present in the protein complexes and they also target the duplicated genes that are essential in human protein complexes to avoid dosage imbalance. The role of miRNAs in human diseases have been well established, so I have taken the opportunity to study which of the human diseases is targeted more likely by the miRNAs, obtaining the cancer related diseases as the most addressed disease class in human disease gene categories. In addition, I have also encountered the question by asking the role of newly emerged and old miRNAs in regulation of human disease genes. Our study suggested that the dominating role of old miRNAs in human disease progressions over new miRNAs. This study also demonstrated the early evolutionary miRNAs are retained in the organisms to do specific functions despite more number of target genes by new miRNAs.

Key responsibilities
✓ Projects handling,
✓ Pipeline developing,
✓ Manuscript preparation and publications

Key achievements
✓ Linux system administration and maintenance in Workstation, Servers,
✓ Shell scripts, Perl, Python scripting languages
✓ Molecular phylogenetic analyses
✓ Statistical softwares handling

1. Complex-forming proteins escape the robust regulations of miRNA in human.

Most of the proteins carry out their functions by participating in protein complexes. Recently, miRNAs are identified as promising post-transcriptional regulators that influence a large proportion of genes in higher eukaryotes. We aim to understand the role of miRNA in the regulation of human proteins that are present in protein complexes. Here, we show that robust regulations of miRNA are absent in human complex forming proteins. Moreover, the numbers of miRNA hits cannot direct the evolutionary fate of complex-forming proteins independently. However, the duplicated complex-forming proteins having a severe effect on organismal fitness are profoundly targeted by miRNA, probably for reducing the chances of dosage imbalance.

Key responsibilities ✓ Design analysis pipeline
✓ ENSEMBL, UCSC database handling
✓ Evolutionary rate calculations
✓ Analysis of gene duplicates
✓ Data analysis using Linux shell and in-house Perl scripts
✓ Statistical analysis through SPSS, GraphPad
✓ Wrote manuscript

Key achievements
✓ Learn Perl scripting
✓ Learn Linux terminal and shell scripting
✓ Orthologous, homologous gene data analysis
✓ Evolutionary rate calculation from gene sets
✓ Learn statistical analysis in SPSS, STATA, Tanagra
✓ API handling of databases
✓ Microarray gene expression data handling
✓ Published research article

Softwares, Databases and Statistics
✓ **Softwares:** Linux OS, BLASTP, Perl, Shell
✓ **Databases:** CORUM, TargetScan, miRwalk, BioGPS, Ensembl, OGEE,
✓ **Statistics:** Spearman rank correlation, Mann-Whitney U test, Partial correlation

Related publication:
Das, J., Chakraborty, S., Podder, S., Ghosh, TC. Complex-forming proteins escape the robust regulations of miRNA in human. FEBS letters. 2013 Jul 11;587(14):2284-7. Link to the publication

2. Insights into the miRNA regulations in human disease genes.

MicroRNAs are a class of short non-coding RNAs derived from either cellular or viral transcripts that act post-transcriptionally to regulate mRNA stability and translation. In recent days, increasing numbers of miRNAs have been shown to be involved in the development and progression of a variety of diseases. We, therefore, intend to enumerate miRNA targets in several known disease classes to explore the degree of miRNA regulations on them which is unexplored till date. Here, we noticed that miRNA hits in cancer genes are remarkably higher than other diseases in human. Our observation suggests that UTRs and the transcript length of cancer related genes have a significant contribution in higher susceptibility to miRNA regulation. Moreover, gene duplication, mRNA stability, AREScores and evolutionary rate were likely to have implications for more miRNA targeting on cancer genes. Consequently, the regression analysis has confirmed that the AREScores plays most important role in detecting miRNA targets on disease genes. Interestingly, we observed that epigenetic modifications like CpG methylation and histone modification are less effective than miRNA regulations in controlling the gene expression of cancer genes. The intrinsic properties of cancer genes studied here, for higher miRNA targeting will enhance the knowledge on cancer gene regulation.

Key responsibilities
✓ Design study pipeline
✓ Data collection and analysis through Perl scripting
✓ Statistical analysis of data using SPSS, GraphPad, OriginLab
✓ Manuscript preparation

Key achievements
✓ Deep learning of Perl scripts
✓ Learning Python script
✓ Intensive use of shell programming in Linux
✓ Handling of clusters, servers, workstations
✓ Linux system administration
✓ Handling of Next-Generation Sequencing data
✓ Published research article

Softwares, Databases and Statistics
✓ **Softwares:** Linux OS, BLASTP, Perl, Shell
✓ **Databases:** ICGC, HGMD, GAD, TargetScan, TarBase, UCSC, Ensembl, BioGPS, AREScores, HHMD, NGSmethDB, CancerMiner
✓ **Statistics:** Spearman rank correlation, Mann-Whitney U test, Partial correlation, Shapiro-Wilk test, McCullum proportion test

Related publication: Das, J., Podder, S., Ghosh, TC. Insights into the miRNA regulations in human disease genes. BMC genomics. 2014 Dec;15(1):1-7. Link to the publication

3. Explicating the role of old miRNAs in human disease progression

Most of the microRNAs (miRNAs) are mainly described as conserved in the evolutionary timescales, although some miRNAs are species-specific in nature. The conserved “old” miRNAs and “newly” emerged miRNAs may have played some differential roles in human disease progression. However, modulating the level of expression of human disease genes by new and old miRNAs has not been testified yet. Here, we identified 1,190 new and 97 old miRNAs in humans and showed that these old miRNAs play more influential roles in human disease gene expression levels than the new ones. Additionally, we also performed our analyses on four other mammals to find whether a similar pattern of miRNA regulation is followed there as that showed by old miRNAs in humans. Furthermore, our analysis indicated that miRNAs targeting is dominant over other genomic regulators in controlling human disease gene expression

levels. Our study demonstrated that the early evolutionary miRNAs are retained in the organisms to do some specific functions despite of the fact that more number of genes are being targeted by new miRNAs.

Key responsibilities
✓ Designed the study pipelines and analyzed the data
✓ Molecular phylogenetic analysis using PHYLIP, MEGA6, UPGMA
✓ Single Nucleotide Polymorphism (SNP), gene expression levels, gene regulation analysis
✓ DNA methylation, transcription factor binding sites (TFBS), histone modifications analysis along with miRNA dataset
✓ Gene orthologous, homologous data analysis and calculation of evolutionary rates
✓ Prediction analysis of microRNA birth and death

Key achievements
✓ Learn phylogenetic analysis through different species
✓ In-depth scripting analysis,
✓ Learn R as a scripting language
✓ Intensive high-dimensional statistical data analysis
✓ Prepared manuscript

Softwares, Databases and Statistics
✓ **Softwares:** Linux OS, BLASTP, Perl, Shell, MEGA6, Phylopat 52, COUNT,
✓ **Databases:** miRbase, TargetScan, miRDB, miRNAMap2, MicroCosm, HGMD, GAD, GeneCards, PolymiRTS, mircancer, UCSC, TRANSFAC, HHMD, Ensembl,
✓ **Statistics:** Spearman rank correlation, Mann-Whitney U test, Partial correlation, Shapiro-Wilk test, McCullum proportion test, Regression analysis

1. Identification and evaluation of Intron Length Polymorphism (ILP) marker in Foxtail millet (Setaria italica) and to study evolutionary relationship between species by cross-species transferability
Supervisor: Prof. Manoj Prasad, Scientist
(2009-06 to 2010-05)

Introns are noncoding sequences in a gene that are transcribed to precursor mRNA but spliced out during mRNA maturation and are abundant in eukaryotic genomes. The availability of codominant molecular markers and saturated genetic linkage maps have been limited in foxtail millet (*Setaria italica (L.) P. Beauv.*). Here, we describe the development of 98 novel intron length polymorphic (ILP) markers in foxtail millet using sequence information of the model plant rice. A total of 575 nonredundant expressed sequence tag (EST) sequences were obtained, of which 327 and 248 unique sequences were from dehydration- and salinity-stressed suppression subtractive hybridization libraries, respectively. The BLAST analysis of 98 EST sequences suggests a nearly defined function for about 64% of them, and they were grouped into 11 different functional categories. All 98 ILP primer pairs showed a high level of cross-species amplification in two millets and two nonmillets species ranging from 90% to 100%, with a mean of ⊠97%. The mean observed heterozygosity and Nei's average gene diversity 0.016 and 0.171, respectively, established the efficiency of the ILP markers for distinguishing the foxtail millet accessions. Based on 26 ILP markers, a reasonable dendrogram of 45 foxtail millet accessions was constructed, demonstrating the utility of ILP markers in germplasm characterizations and genomic relationships in millets and nonmillets species.

Key responsibilities
✓ Cloning of bacterial genomes
✓ Microbiology and molecular biology techniques, bacterial culture, cloning, PCR, agarose-gel running,
✓ Polyacrylamide gel separation, microRNA isolation and identification, sequencing, primer designing
✓ Looked after of plantlets in germination chamber, field
✓ Designing of ILP markers

Key achievements
✓ Different wet lab techniques
✓ How to develop study pipelines
✓ Publication

1. Drug designing: evaluation of iron chelation through blood in phytic acid treated thalassemia patients
Supervisor: Prof. Sudipa Chakraborty
2006-01 to 2006-05

The main objective of the project is to use rice bran as a source of phytic acid for the thalassemia patients. Thalassemia is one of genetic blood disorders characterized by decreased hemoglobin production. In India, β-thalassemia is more common. The autosomal recessive disorder has reported insertion of 5 bp in the β-globin gene. The oxygen carrying capacity of hemoglobin molecules reduces and hence iron chelation is one of treatment measures for thalassemic patients. The iron overload occurs very rapidly in the patients and the toxicity level reaches high within 2-3 weeks after blood transfusion, even some patients need to get the transfusion 2 times in a week. Therefore, the chelation of iron shortens the number of transfusions for the patients. However, the iron chelation agents for thalassemia patients are costly and it is quite a burden for the patients' family. Researches indicated that rice bran is a good natural source for the phytic acid which is often considered as one of the well-known references as iron-chelating agents. In this project, we worked as a team of four members whose objectives are collection of patients' data, extraction of phytic acid from rice bran, identifying the chemical compositions of the rice bran and test the production of phytic acid. As a part of the group, my job was to collect blood samples from patients, extract the phytic acid from rice bran and test the concentration of phytic acid. Handling of lot of patients' data, extraction and purification of phytic acid from rice bran, measurement of phytic acid concentration using calorimeter and spectrophotometer were part of the project.

Key responsibilities:
✓ Handling of patients, aged less than one year to 34 years,
✓ Study the iron chelation technique,
✓ Sample handling and data collection,
✓ Statistical analysis of data,
✓ Colorimetry and spectrophotometry analysis

Key achievements
✓ learning new techniques like color estimation from samples,
✓ spectrophotometry analysis of samples
✓ collection of patients' data
✓ biochemical analysis of phytic acid

2. Study of ecological diversity in the Sundarbans, especially on kingfishers
Supervisor: Prof. Amit Chakraborty
2006-01 to 2006-05

Sundarban is the largest mangrove delta of the world and home for more than 150 species of fish, 270 species of birds, 100 species of mammals, reptiles, amphibians, nearly 350 species of trees. The word "sundar" came from the local name of the plant species, *Heritiera fomes* (Sundari). While the pneumatophores are the main attraction of these delta that made almost impossible for people to enter the area, the area is also the house of great Royal Bengal Tigers. The distinct ecological diversity with rich flora and fauna in the South part of West Bengal, India also inhabited by more than fifty thousand human beings.

Kingfisher, the brightly colored birds has a cosmopolitan distribution all over the world and more than 110 species are found. In India, nearly 19 species are observed over the country. My objective was to observe and documented the species of kingfishers found in Sundarbans. With the help of locals and forest services, the area was visited during the spring of 2006 in a group of fifty people. From official forest records and observations made in the forest area, I was successfully identified 13 species of kingfishers in the area. The collection of data, photographs, characteristic details were done with the help of local habitants and forest personnel.

Key responsibilities
✓ Data collection from locals and forest personnel
✓ Detail characteristic analysis of kingfishers
✓ Observe and photographed kingfishers

Key achievements
✓ Study of ecological niche
✓ Field work
✓ Sample collection and characteristic observation