SAGNIK BANERJEE

Ames, IA 50010

725-333-0696 | sagnik@iastate.edu | https://www.linkedin.com/in/sagnikbanerjee15/ | https://github.com/sagnikbanerjee15

PROFESSIONAL SUMMARY

Efficient self-driven graduate research assistant with success in independent and collaborative research environments. Proven public speaking skills and presented scientific results at conferences both in-person and virtually. Over 5 years of experience developing softwares, automating data analysis pipelines and brainstorming innovative approaches to mine unique next-generation sequencing (NGS) data primarily using python, R and C. Highly adept in organizing scientific meets and leading peers in their academic endeavors.

KEY SKILLS

• Data Visualization • Predictive Analysis • Statistical Modeling • Clustering & Classification • Data Analytics
• Data Mining • Quantitative Analysis • ML algorithm • Model Development • Pipeline Development

TECHNICAL SKILLS

NGS: RNA-Seq, Ribo-Seq, CAGE-Seq, RAMPAGE, ATAC-Seq, Yeast two-hybrid NGS, Iso-Seq, ChIP-Seq, DNA-Seq
Programming languages: Python, R, Jupyter, Unix, C, C++, Prolog, Java, JavaScript, PHP, MySQL, JSP, Perl
Machine learning: Regression, Clustering, Classification, Cross Validation, Feature Selection, Feature Extraction
Deep learning: Recurrent Neural Networks, Bi-directional Recurrent Neural Networks
Predictive applications: Gene function prediction, Protein secondary structure prediction, Protein post-translational modification prediction, Protein disordered region prediction
Statistics: Gene count normalization, Changepoint analysis, Hypothesis testing, Differential gene expression
Software design: Version control, GitHub deployment, Code maintenance
Genomics: Short-read alignment, Long-read alignment, Long-read generation, PacBio read error correction, Genome annotation, Genome assembly, Variant calling, Epigenetic analysis (MEME)
Transcriptomics: Transcriptome assembly, Gene co-expression analysis
Proteomics: Protein-protein interaction, Effector mining, post-translational modification
Genome browsers: IGV, UCSC genome browser, jbrowse, Apollo
Databases: NCBI-SRA, NCBI-GEO, UniProt, Swiss-Prot, Ensembl, PDB, PhosphoELM, SCOP
Miscellaneous skills: High performance computing (SLURM), Pipeline construction

EXPERIENCE

Oak Ridge Institute for Science and Education, USDA-ARS, Ames, IA   **January 2018-Present**
Bioinformatics Plant Pathology Graduate Research Assistant
• Improved the structural annotation of >1200 genes in *Arabidopsis thaliana* by employing changepoint detection from short-read coverage data
• Collaborated with researchers to find >100 temporally significant novel fungal effectors from RNA-Seq data using computational approaches in 3 different pathosystems
• Illustrated the relevance of gene counts to infer regulatory properties of ~20000 genes in *Arabidopsis thaliana* using machine learning through a collaborative venture
• Explored over 5 assemblers to assemble fungal genomes from short-read data and evaluated their performance
• Expedited and improved the extent of compression of short-read alignments by over 30%

Department of Plant Pathology and Microbiology, ISU, Ames, IA   **May 2017-July 2019**
Bioinformatics Graduate Research Assistant
• Modelled phasiRNAs expression using hypergeometric distribution leading to discovery of >400 phasing loci in barley

- Discovered >100 novel protein interactions from NGS data by designing an optimized computational pipeline, called **NGPINT**
- Assessed the efficacy of various normalization methods for analyzing NGIS data and participated in designing the pipeline **Y2H-SCORES** to prioritize candidate interactors for secondary validations
- Identified over 200 novel genes involved in regulatory activity in the barley-blumeria patho-system

Department of Computer Science and Engineering, Jadavpur University, India   **August 2013-January 2015**
Machine Learning Graduate Research Assistant
- Advanced the understanding of post-translational modifications by incorporating 40 high quality physico-chemical properties of amino acids as features
- Constructed classifiers to predict disordered regions of proteins

EDUCATION

Iowa State University (ISU), Ames, IA   **August 2016-July 2022**
Doctor of Philosophy, Ph.D.
Major: Bioinformatics and Computational Biology, Minor: Statistics

Jadavpur University, Kolkata, India   **August 2012-July 2014**
Master of Engineering, M.E.
Major: Computer Science and Engineering

Institute of Engineering and Management, Kolkata, India   **August 2008-July 2012**
Bachelor of Technology, B.Tech.
Major: Computer Science and Engineering

LEADERSHIP SKILLS AND SERVICE

Graduate and Professional Students Senate (GPSS), ISU, Ames, IA   **August 2018-July 2019**
Senator, **Department of Plant Pathology and Microbiology**
- Participated in annual senate elections and volunteered in organizing the annual conference

Bioinformatics Graduate Student Organization, ISU, Ames, IA   **August 2017-July 2018**
Vice President
- Reviewed venue and optimized expenditure while organizing the 4th annual Bioinformatics and Computational Biology (BCB) student symposium

Lead Judge, Ames, IA   **2019,2020,2021**
State Science and Technology Fair of Iowa
- Examined the performance of 15 projects based on novelty, presentation and response to questions

SELECT HONORS (2 of 5 honors)

- First place, Flash talk, Student seminar, Bioinformatics & Computational Biology Program, ISU   **2020**
- Student poster award, 3rd position for Student Poster Presentation at the 4th Annual BCB Symposium   **2018**

SELECT PUBLICATIONS (2 of 16)
- **Banerjee S.**, Bhandary P., Woodhouse M., Sen T.Z., Wise R.P., Andorf C.M. (**2021**). FINDER: An automated software package to annotate eukaryotic genes from RNA-Seq data and associated protein sequences. **BMC Bioinformatics**
- **Banerjee S.**, Velásquez-Zapata V., Fuerst G., Elmore J. M., Wise R. P. (**2020**). NGPINT: A Next-generation protein-protein interaction software.