ARANYAK GOSWAMI (CV)

CAREER OBJECTIVE: Using computational techniques and bioinformatics methods to understand complex biological problems

Name: Dr Aranyak Goswami

Present Affiliation: Post-Doctoral Fellow, Stanford University, Prof Mark Kay – Department of Paediatrics.

Last Affiliation- Post-Doctoral Associate, Yale University, Yale Centre for Cardiovascular Research.

Supervisor name- Dr. Stefania Nicoli (Associate Professor; Director of the Zebrafish Phenotyping Core for Precision Medicine, Internal Medicine and Genetics, Yale University)

Education

Bachelors Microbiology (Hons) (University of Calcutta), India 2005 (First Class, with highest honors)

Masters in Genetics (University of Calcutta), India 2007 First Class with highest honors)

Ph.D. Bioinformatics (Indian Institute of Chemical Biology, Kolkata), India 2017

Contact details- ag111@stanford.edu, aranyak111@ gmail.com,

Nationality-Indian

Home address- 149 Jackson Apartment 202, Redwood city

Pin code- 94061

Contact number- +1-203-919-4047

Publications

1) Saha, S. K., Goswami, A., & Dutta, C. (2014). Association of purine asymmetry, strand-biased gene distribution and PolC within Firmicutes and beyond: a new appraisal. BMC genomics, 15(1), 1-26.

2) Wendt, F. R., Pathak, G. A., Tylee, D. S., Goswami, A., & Polimanti, R. (2020). Heterogeneity and Polygenicity in Psychiatric Disorders: A Genome-Wide Perspective. Chronic Stress, 4, 2470547020924844.

3) Wendt, F. R., Pathak, G. A., Levey, D. F., Nuñez, Y. Z., Overstreet, C., Tyrrell, C., ... & Polimanti, R. (2021). Sex-stratified gene-by-environment genome-wide interaction study of trauma, posttraumatic-stress, and suicidality. Neurobiology of stress, 14, 100309.

4) Goswami, A., Chowdhury, A. R., Sarkar, M., Saha, S. K., Paul, S., & Dutta, C. (2015). Strand-biased gene distribution, purine asymmetry and environmental factors influence protein evolution in Bacillus. FEBS letters, 589(5), 629-638.

5) Pathak, G. A., Wendt, F. R., De Lillo, A., Nunez, Y. Z., Goswami, A., De Angelis, F., ... & Polimanti, R. (2021). Epigenomic Profiles of African American Transthyretin Val122Ile Carriers Reveals Putatively Dysregulated Amyloid Mechanisms. Circulation: Genomic and Precision Medicine, 14(1), e003011.

6) Wendt, F. R., Pathak, G. A., Levey, D. F., Nuñez, Y. Z., Overstreet, C., Tyrrell, C., ... & Polimanti, R. (2020). Trauma and posttraumatic stress interact with sex-specific risk loci for suicidality and converge on brain extracellular matrix biology and synaptic plasticity. medRxiv.

7) Goswami, A., Wendt, F. R., Pathak, G. A., Tylee, D. S., De Angelis, F., De Lillo, A., & Polimanti, R. (2021). Role of microbes in the pathogenesis of neuropsychiatric disorders. Frontiers in Neuroendocrinology, 62, 100917.

8) De Angelis, F., Wendt, F. R., Pathak, G. A., Tylee, D. S., Goswami, A., Gelernter, J., & Polimanti, R. (2021). Drinking and smoking polygenic risk is associated with childhood and early-adulthood psychiatric and behavioral traits independently of substance use and psychiatric genetic risk. Translational psychiatry, 11(1), 1-12.

9) Pathak, G. A., Wendt, F. R., Goswami, A., Koller, D., De Angelis, F., Polimanti, R., & COVID-19 Host Genetics Initiative. (2021). ACE2 Netlas: In silico Functional Characterization and Drug-Gene Interactions of ACE2 Gene Network to Understand Its Potential Involvement in COVID-19 Susceptibility. Frontiers in genetics, 1523.

10) Exploring The Genome and Proteome Landscapes Of The Genus Bacillus In Quest Of Lineage And Niche-Specific Traits Through In Silico Analysis Aranyak Goswami (2017)

[* I have been involved in all relevant analysis, writing, and designing of figures for all my First Author papers and I have been involved in the conception, writing, and part of analytic execution in my second author papers.]

Published – 3 first author publications and 7-second author publications.

Unpublished work

Yale Post-Doctoral work in Human Genetics under Dr. Renato Polimanti

1) Genome-wide association studies to comprehend the role of Vitamin D on Psychiatric studies on a large-scale population of over 4 hundred thousand European individuals.

Methods used-To conduct such studies Heritability(h^2) and Genetic Correlation(r_g), estimates have been first computed between different Psychiatric disorders with Vitamin D, time spent outside summer and winter taken from the Psychiatric Genome Consortium Website for over four hundred thousand individuals of European descent. Then genetic correlation and heritability estimates were computed for all dietary phenotypes taken from UK Biobank. Hierarchical clustering has been performed between the phenotypes involving diet with psychiatric disorders which survived multiple testing corrections. Similar calculations were computed between Vitamin D with those Psychiatric disorders which survived multiple testing corrections. Mendelian randomization was then performed for trait pairs that have high bootstrap values to find the principal causal variables associated with the disease.
Based on early Mendelian Randomization studies it was seen that Psychiatric disorders like depression have little effect due to exposure to Vitamin D or Vitamin D supplements.

Techniques used- LD score regression to compute Genetic Correlation and Heritability., Hierarchal Clustering, Mendelian Randomization.

Work in progress

Yale Postdoctoral experience in transcriptomics and systems biology under Dr. Nicoli Stefania

Genetic Compensation is a phenomenon by which organisms despite having deleterious mutations can function optimally due to the compensatory action of paralogous genes, metabolic modulation, or transcriptional adaptation. This study involves large-scale computational analysis of RNA seq data. Hence as a part of the study, I have established a pipeline that is used to study bulk scale RNA seq data and micro-RNA sequencing for gene assembly, annotation, alignment, differential gene expression, and visualization.

Results-After identification of a set of differentially expressed genes, functional annotation, and network analysis of gene sets in question was done through GO, KEGG, and Reactome pathway analysis. As a result of such studies genes potentially involved in biological pathways critical for genetic compensation have been identified.

Packages used – FastQC for assessing read quality, STAR and Bowtie for assembly and annotation, HTseq for reading counting and quantification. DeSEQ2 for differential Gene expression, GO enrichment, and KEGG annotation for novel biological pathway identification using topGO and ClusterProfiler.

Published Work underlying the genetic underpinnings of the human microbiome

Gut microbiota plays an important role in the bidirectional communication between the gut and the central nervous system. Mounting evidence suggests that gut microbiota can influence brain function via neuroimmune and neuroendocrine pathways as well as the nervous system. Advances in sequencing techniques help in facilitating the investigation of the underlying relationship between gut microbiota and psychiatric disorders. I wrote a review article utilizing Meta-analysis of existing showing the less explored area of the relationship of Microbiome with Neuropsychiatric disorders.

Journal Service:

Journal Reviewer- Journal Nucleus

Postdoctoral work in Bose Institute India under National Post-Doctoral Fellowship-Department of Science and Technology Govt of India

Fellowship Amount – 100000 Indian rupees (13,320 USD)

Project period- 2018 April to 2020 April

Host Virus Interaction work using system biology approaches to identify novel therapeutic targets

Pathogens manipulate cellular mechanisms of host organisms via pathogen-host interactions (PHIs) so that pathogens can take advantage of the capabilities of host cells, leading to infections. The crucial role of these interspecies molecular interactions in initiating and sustaining infections demands a thorough understanding of the corresponding mechanisms from diverse perspectives.

Unlike the traditional approach of considering the host or pathogen separately, a systems-level approach, considering the Pathogenic host Interaction (PHI) system is a prerequisite to elucidate the mechanisms of infection. Systems biology-based methods for the inference and analysis of PHI (Pathogen Host Interactions) regulatory metabolic and protein-protein networks to shed light on infection mechanisms are gaining increasing demand thanks owing to the availability of omics data.

The knowledge derived from the PHIs may largely contribute to the identification of new and more efficient therapeutics to prevent or cure infections. The goal of the project is thus to understand host-pathogen interactions from new computational perspectives.

The three broad goals pursued

1) Investigate evolutionary signatures in host-pathogen interacting proteins through their relative conservation and functional diversity**.**
2) To study compositional features in pathogen-targeted protein domains concerning non-targeted protein domains**.**
3) To study the role of alternative splicing in understanding host-pathogen dynamics
4) To identify the regulatory aspects of host-pathogen interactions to design therapeutics in the future.

Goals achieved

1) The role of alternative splicing in establishing host-virus competition has been achieved through multi-omics level computational system biology approaches.
2) Using programming scripts in R, Python, and Shell in Linux hub proteins have been identified**.**
3) Characterization of these hub proteins may help in the identification of important viral drug targets in the future.

Some other relevant working experience

Cancer Biology experience- Familiarity with TCGA, Oncomine database leveraging data from metabolomic, proteomic, and exome sequencing profiles for biomarker development. Machine learning classification techniques to infer differences in tumor grades and variation in malignant v non-malignant cells.

1) **Single-cell RNA sequencing experience**- Experienced with single-cell RNA sequencing data, especially with microRNA mutants which shows a transition from hematopoietic to stem cell development. I am also working with PBMC (peripheral blood mononuclear cell) datasets using packages like **Seurat and Monocle**.
2) Datasets mainly used 10X, high throughout Dropseq data.
3) Experienced in using UMAP architecture to understand trajectories in single-cell datasets. Apart from that Yale computer Scientists have developed a technique called **PHATE** ((Potential of Heat-diffusion for Affinity-based Trajectory Embedding) a python-based tool for visualizing high dimensional data. PHATE uses a novel conceptual framework for learning and visualizing the manifold to preserve both local and global distances that has advantages over usual trajectory analysis with **UMAP**.
4) Performed RNA velocity analysis to find a proper trajectory for epithelial to hematopoietic transitions in microRNA mutants to see dynamic differences between spliced and unsliced isoforms.
5) Significant data wrangling experience with ATAC seq, Chip-seq, Hi-seq, Mi-seq data and using multi-omics Cancer genome datasets taken from TCGA, Oncomine datasets to make inferences about malignant v non-malignant cells. Used machine learning techniques like supervised, unsupervised learning, and deep learning techniques to classify different grades of tumor data.

Research Fields

Broad areas of research – **(A)** Human Microbiome, 16S metagenomics, Shotgun sequencing, Computational systems biology investigating host-pathogen interactions, Next-Generation sequencing analysis using Hiseq, Miseq, data from Illumina Platform, Metagenomics and Bacterial taxonomical diversity analysis using Nanopore Minion platform, Genome-wide Association Studies (GWAS), Phenome-wide, Association Studies (PheWas), Gene set enrichment analysis, Complex trait mapping, Bulk scale RNA seq analysis, microRNA analysis, Single-cell RNA seq analysis using Seurat, UMAP, and PHATE. Hidden Markov models, deep-neural nets, saliency map construction to make sense of biological image data.

(B) Network biology approaches- Gene ontology Pathway & Network analysis KEGG, Geo-Ontology Blast2GO, WEBMGA Cytoscape, Bingo, Ingenuity Pathways Studio, topGO, Clusterprofiler.

(C) Experience in Microbial genomics in domains of Genome Assembly, De novo and Reference using Kmer Centroid approach implementing DBruijn Algorithms, Gene Orthology analysis),

(D) Specific areas of research-Metabolomics, and Transcriptomics, Bacterial and Mammalian Evolutionary Bioinformatics, Genome-wide association analysis for complex genetic diseases, bulk scale, and single-cell transcriptomics.

Technical skills

Operating Systems: Windows, Mac, Linux, UNIX, Shell, Virtual Box, and Docker.

1) Built a pipeline for large-scale RNA sequencing studies of mRNA and small RNA sequences (microRNA, si-RNA, snoRNA) involving alignment, annotation, quantification, Differential Gene Expression, and Data visualization using high-performance Yale Clusters in a Unix-based environment using a combination of shell, Python, and R scripts to understand the mechanism of Genetic Compensation in zebrafish.
2) Using Supervised and Unsupervised machine learning techniques to dissect the modalities of complex Genetics in Psychiatric Disorders.
3) **Programming languages-** R, Python, Perl, C, and bash programming in a UNIX environment. Bioconductor and all other standard statistical packages. Experience with Git version control, R markdown, and Jupyter Notebook. **Workflows used** – Snake make, Cromwell for automated bioinformatics pipelines.
4) **Machine Learning techniques used-** Linear and Logistic Regression. Random Forest, Convolution Neural Networks, and Hierarchical clustering to make sense of complex polygenic multimodal biological data involved in large-scale population Genetics.

   Using Markov models and Generative Adverse Neural Networks to classify single-cell and bulk RNA seq data.
5) **Deep learning**- Using convolutional neural networks, U-Net architecture to solve image segmentation problems using tools and interfaces like Cellpose.
6) Workflows deplored- Snakemake and Nextflow,GATK.

Conferences and Poster presentations

1) Invited talk on genome and protein architecture in *Bacillus* genomes shedding light on evolutionary adaptive strategies used by these microorganisms in the **Department of Genetics (University of Calcutta)-2017.**

2) Oral presentation on the application of in silico approaches to explore microbial genomics and evolution in the Indian Institute of Chemical Biology (IICB)-**2018.**

4) NGS training cum workshop at **Assam Agricultural University**-**2019**.
5) International (European Molecular Biology Organization) **EMBO workshop** and symposium on Human Microbiome and its applications on Next-Generation Therapeutics Kalyani West Bengal-**2019**.
6) R and Python workshops organized by **Yale Centre for Research Computing and Yale statistical services -2020.**
7) Attended Workshop – "**Introduction to Biomedical Data Science and Health Informatics Summer Course**" by Yale Centre for Biomedical Data Science-**2020.**
8) **World Congress of Psychiatric Genetics** virtually at Yale-**2020**
9) Poster presented at 4th Annual Post-Doctoral symposium at Yale **-2021**.
10) Presentation on computational approaches to study Genetic Compensation in Yale Postdoc bimonthly presentation event-**2021**.

11) **Machine learning for single-cell data analysis** organized by Yale Genetics and Yale School of Applied and Engineering Sciences-**2021**.

Mentorship – Coordinating for a team of postdoctoral fellows to achieve career dependence under Yale faculty Nikhil Joshi at Yale-**2020-present**.

The organizer of faculty trainee lunch sessions acting as a career development program for Students interacting with Faculty as a part of Yale School of Medicine and Department of Genetics-**2020**

Intersections Fellow Symposium 2021 organized by Yale. Acting as an application reviewer and judge for postdocs trying to transition as a faculty, especially in the field of Bioinformatics-**2021**.

Scholarships and Awards obtained

2021-Present – Post-Doctoral Fellowship at Yale Centre for Cardiovascular Research

2019-2020- Post-Doctoral Fellowship at Yale School of Medicine-Human Genetics, Department of Psychiatry.

2018- DST (Department of Science and Technology –Govt of India) National Post-Doctoral Fellowship > (10% of applicants obtain the grant.)

2010- (Senior Research Fellowship) (NET) (UGC- CSIR) University Grants Commission, Council of Scientific and Industrial Research India. [**Fellowship amount equivalent to -7000 US dollars per year for 3 years].**

2008 – JRF (Junior Research Fellowship) [[University Grants Commission, Council of Scientific and Industrial Research India] **Fellowship amount equivalent to 5000 US dollars per year for 3 years].**

National Eligibility Test (NET) (UGC- CSIR) University Grants Commission, - Council of Scientific and Industrial Research India. (> 5% of examinees obtain the fellowship).

2002- Eligible for National scholarship (Govt of West Bengal, India)

2000- Nalandasree award for obtaining more than 85% in science subjects.

Language skills- Proficient in reading, writing, and communicating in English.

Teaching and Mentoring students

Involved in mentoring M-tech students in finishing their final thesis. One of such theses presented was entitled "**Genomic and Proteome Signatures in *Deinococcus* Exploring Causes Responsible for Its Extreme Radioactive Tolerance**"

Mentored 5 master's level students in Bioinformatics to complete their master's thesis.

Taught Bioinformatics Courses, Basic Bioinformatics data analysis to Ph.D. students as a part of the teaching and Modern Instructor Workshop 2022.

Extra-Curricular Activities- Writing for Indian Science Congress in Frontiers areas of Bioinformatics, Genomics, Biochemistry, Microbiology**.**

Indian Science Congress is the official body that recognizes the contributions of Young to Eminent Indian Scientists by providing them with memberships to prestigious academies like Fellow of National Academy of Sciences India.

Grant writing experience- National Post-Doctoral fellowship Govt of India. Obtained a grant of **1 million Indian Rupees (13,389 dollars)** in a total period of 2 years.

Project name- **Exploring the evolutionary signatures of host-pathogen interaction using in-silico studies**

Acting as a Grant advisor for the National Science Foundation (NSF) project on the documentary "My mom the Scientist "spearheaded by acclaimed documentary filmmaker and Yale Professor Thomas Allen Harris.

Professional Membership

Member –Calcutta Consortium of Human Genetics, India.

Member- American Society of Human Genetics.

Member- World Congress of Psychiatric Genetics.

Member of American Society of Gene & Cell Therapy

References

1) Dr. Stefania Nicoli- Associate Professor; Director of the Zebrafish Phenotyping Core for Precision Medicine, Internal Medicine, and Genetics email -stefania.nicoli@yale.edu Phone- 203.737.6480

2) **Dr.Renato Polimanti**- Assistant Professor of Psychiatry, email- renato.polimanti@yale.edu Phone- 203

3) Dr Mark Kay – Professor Departments of Pediatrics and Genetics Director, Program in Human Gene Therapy Stanford University School of Medicine, email- markay@stanford.edu Phone-: (650) 498-6531

4 Dr. Chitra Dutta- Chief Scientist (Indian Institute of Chemical Biology) email id- **cdutta.iicb@gmail.com** Phone- +91 9831356342.