

# ANUJ GUPTA

School of Biology, Georgia Tech, Atlanta, Georgia

Contact: [anujg@gatech.edu](mailto:anujg@gatech.edu) / +1-404-579-498

[www.prism.gatech.edu/~agupta399/](http://www.prism.gatech.edu/~agupta399/)

PhD. in **Bioinformatics**, Georgia Institute of Technology

(May'16-Present)

Masters in **Bioinformatics**, Georgia Institute of Technology (GPA: 3.95/4.0)

(Aug'14- Dec'15)

B.Tech in **Biological Sciences and Bioengineering**, Indian Institute of Technology, Kanpur

(July'09-June'13)

## PUBLICATIONS / PRESENTATIONS

- **Gupta A**, I. King Jordan, Lavanya Rishishwar; *stringMLST: a fast k-mer based tool for multilocus sequence typing*. Bioinformatics. 2017 Jan 1;33(1):119-121. doi: 10.1093/bioinformatics/btw586 Citations: 5
- **Gupta A**, Tang Y, Garimalla S, the MaHPIC consortium, Galinski MR, Styczynski MP, Fonseca LL, Voit EO. *Metabolic Modeling Helps Interpret Transcriptomic changes during a Complex Disease*. Elsevier/Molecular Basis of Disease; doi.org/10.1016/j.bbadis.2017.10.023
- **Gupta A.**, (2017, July). *Model-based inferences of disease-related phenotypical alterations from changes in gene expression*. Paper presented at ICMSB 2017, Raitenhaslach, Germany

## AWARDS

- **Computational Biology Faculty Research Award** (School of Biology, Georgia Tech) (Spring'15)
- Recipient of **MCM Scholarship** for continued excellent academic performance (IIT Kanpur) (2009-2013)

## RESEARCH EXPERIENCE

- **Graduate Research Assistant at Voit Lab**, Biomedical Engineering, Georgia Tech and Emory School of Medicine (May'16-Present)
  - Mathematical Modelling of Purine Pathway in *S. enterica* with an aim to understand global effects of AICAR on health, survival and growth of the organism. We are also trying to see bidirectional effects of AICAR and purine pathway
  - Analysing omics especially transcriptomics, metabolomics and metatranscriptomics to explore possibility of dual infection in a euthanized subject of Malaria experiments (MAHPIC)
  - Lipidomic analysis of Whole Blood and Bone Marrow samples from 100 day malaria experiments (MAHPIC)
- **Graduate Research Assistant at Jordan Lab**, Applied Bioinformatics Laboratory (ABiL) (Aug'15-Apr'16)
  - Involved in taking classes and creating teaching material for Bioinformatics class for CDC scientists (over 100 hrs of classroom teaching). ABiL is in partnership with IHRC and one of the aims is workforce development and training
- **Graduate Research Assistant at Jordan Lab**, School of Biology, Georgia Tech (Aug'14-April'15)
  - Created stringMLST, a standalone platform independent MLST typing tool. Used by researchers at GSK among others
  - Upgraded MGIP web platform to include *B. pertussis* and defined a new Genome Based MLST scheme (gtMLST) by finding maximally preserved and minimally conserved loci for high typing resolution and variability. Used by Meningitis and Bordetella Lab at CDC

## PROFESSIONAL EXPERIENCE

- **Bioinformatics Intern, Philips Research, NYC** (May'15-July'15)
  - Studied mutation rate in *E. faecium* to identify nosocomial infection. Predict spread of infection using minimum spanning algorithm
  - Westchester Medical Center(WMC) verified our results. Created demo application for further testing of this methodology
  - Designed a pipeline by implementing Bayesian statistical model on cumulative SNPs to link them to antibiotic resistance. Used feature space reduction and then trained the model to classify antibiotic resistance
- **eGain Communications Pvt. Ltd. , Pune, India** (June'13-July'14)
  - Worked as **Software Engineer** that required coding for both client (HTML, CSS and JavaScript) and server side (JSP, Java and MSSQL)
  - Achievements**
    - Created installer for CISCO adapter and connector for eGain application
    - Created utility features for installation that reduced runtime and failure rate of daily builds
- **Beehyv Software Solutions, Secunderabad, India** (June'12-July'12)
  - Worked with their client (a California based research group) on project related to exploring protein pathway databases
  - Used Reactome and Wikipathways APIs to extract data and form a consensus cleaned network of relevant proteins

## PROJECTS

- **Mathematical Modelling of Purine Pathway – Mentor: E.O.Voit** (Sep'17-Present)
  - Used machine learning techniques(LDA,PCA,ICA) to clean irregular metabolite(MS) data involving multiple replicates and experiments

- Set up a GMA power law model for Purine Metabolism to understand role of AICAR in salvaging purines
- Parameter set was estimated for various control experiments and knockouts in various media. Estimated flux and concentration in various condition gave insights into role of AICAR in Purine metabolism
- **Multi-omics analysis to find signs of co-infection in a euthanized subject in Malaria experiment (MAHPIC) –** *(Aug'17-Present)*  
*Mentor: E.O.Voit*
  - Performed a pilot study based on transcriptomics and metabolomics to show that the subject that had to be euthanized due to severe malarial infection was actually different even before conceiving sporozoite infection suggesting possibility of co-infection
  - As a fishing expedition, performed various pathway and disease enrichment analysis to explore the possibilities
  - Performed metatranscriptomic analysis to narrow down on pathogenic causes of the coinfection
- **Lipidomic characterization of malarial infection on M. mulatta subjects (MAHPIC) –** *(Jan'17-Apr'17)*  
*Mentor: E.O.Voit*
  - Performed comparative analysis of lipidomics of severe and non-severe M. mulatta subjects infected with P.coatneyi and P.cynomolgi
  - Bridged the functional characteristic of glycerophospholipids and sphingolipids to their profile during the course of infection
  - Conducted in depth side-chain and saturation analysis to gain better insights about the functionality of various lipid species
- **Model-based inferences of disease-related phenotypical alterations from changes in gene expression (MAHPIC) –** *(Apr'16-Dec'16)*  
*Mentor: E.O.Voit*
  - Combined information about metabolic state with transcription in a GMA model to infer metabolic and flux changes
  - Model revealed an interesting finding that higher release of inosine, hypoxanthine and xanthine which may be related to their function as immune-regulators or to the parasite's high demand for purines
- **stringMLST: a fast k-mer based tool for multilocus sequence typing –** *(Jan'15-Dec'15)*  
*Mentor: I.K.Jordan*
  - Developed MLST based typing tool which is alignment and assembly free. This tool is standalone and platform independent
  - Started the project from scratch and completed it along with exhaustive testing for over 2000 whole genome sequences
  - Further work on the project led to filing of patent (pending) by Jordan Lab
- **VisualAIDS: An Interactive visualization of HIV/AIDS data -** *(Oct'17-Nov'17)*  
*Mentor: Alex Endert*
  - An interactive visualization designed in d3.js to investigate and explore HIV/AIDS data for various countries over time. More information at [www.prism.gatech.edu/~agupta399/InfoViz-Project/](http://www.prism.gatech.edu/~agupta399/InfoViz-Project/)
- **Acute Myeloid Leukemia Outcome Prediction (Machine Learning in Biomedicine) -** *(Jan'17-Feb'17)*  
*Mentor: Peng Qiu*
  - Defined, selected and created features from flow-cytometry data to be able to differentiate normal from AML patients
  - Performance evaluation helped compare multiple classifiers and pick the best (The classifier was ranked best in class)
- **Comparison of classification approaches to identify factors associated with learning in a mouse model of Down Syndrome (Machine Learning in Biomedicine) -** *(Mar'17-Apr'17)*  
*Mentor: Peng Qiu*
  - Use protein expression data to create a multi-layer classification model to predict genotype based on treatment and behaviour of the mice subjects
- **Computational Genomics –** *(Jan'15-Apr'15)*  
*Mentor: I.K.Jordan*
  - Conducted comparative analysis of 53 WGS of N. meningitidis, H. influenzae and H. haemolyticus provided by CDC
  - Created independent pipeline for best assembly after comparing assemblies from multiple assemblers and mapping to reference
  - Integrated GBrowse, blast and MLST typing tool on web application for the genomes (<http://gbrowse2015.biology.gatech.edu>)
  - Presented the web platform as wholistic database for the provided genome (gene annotation and prediction) with an inbuilt typing tool along with genome browser and blast functionality to collaborators at CDC and Boozallen Hamilton
- **SNP calling pipeline –** *(Aug'14-Dec'14)*  
*Mentor: I.K.Jordan*
  - Aligned genomic reads to reference genome using BWA and sorted it using SAMTOOLS
  - Used GATK tools for improvement steps like reducing effects of analysis artefacts produced by sequencing machines
  - Used SAMTOOLS and HTS package for processing and calling variants
- **Modelling rTCA cycle using simultaneous Michaelis–Menten Kinetics and simulating its constants–** *(Jan'15-Apr'15)*  
*Mentor: J.Weitz*
  - Simplified complex rTCA cycle for carbon fixation as simplified Michaelis Menten model for reversible reactions
  - Estimated variables involved in intermediate steps by simulating reactions of the cycle through eight dimensional variable space
  - The model highlighted the dampening effect of enzyme concentration and effect of input substrates on final products
- **Common set of Differentially Expressed Genes associated with some cancers -** *(Aug'14-Dec'14)*  
*Mentor: Dr J. Choi*
  - Used RNAseq data for lung, liver and prostate cancer from TCGA and GDAC databases for patients
  - Normalized and filtered the data and identified Differentially Expressed Genes using Gene Pattern
  - Determined significant pathways using Gene Set Enrichment Analysis tool GSEA
  - Found genes that were common in the cancers and verified our results through various publications
- **Variant analysis of an African Caribbean male from Barbados -** *(Aug'14-Dec'14)*  
*Mentor: Dr J. Choi*
  - Analyzed raw exome sequencing data (1000 Genome) and analyzed probable disease causing SNVs
  - FASTQC was used to access the quality of the paired-end reads. Used BWA to align the sequence to reference genome
  - Used Picard tools and GATK toolkit for post-processing of the alignment. Variant calling step was performed using Samtools
  - Used Annovar to annotate the variants. The SNV's were filtered on SNP location, type, function, frequency and conservation

- **Comparison of basic Machine Learning techniques** - *Mentor: C. Isbell* (Jan'15-Apr'15)
  - Implemented and tested of learning algorithms (Decision trees, NN, KNN, SVM) on data sets and comparison of performance
  - Used local search algorithms to find weights for Neural Network to obtain best search for our problem
  - Implemented and tested unsupervised learning algorithms for clustering and dimensionality reduction
- **Review Analysis for Google Play Store** - *Mentor: Duen Horng (Polo) Chau* (Aug'15-Nov'15)
  - Analyze textual reviews for Google Play Store to generate topics and features deemed important by customer
  - Use sentiment analysis, topic modelling with NLP to generate key features of the application
  - Extraction and clustering of these topics into time series view to analyze trends for particular apps
  - Created an interactive visualization for features as function of time showing what stood out for various versions of apps
- **Engineering the Osteochondral Interface Using Simultaneous Transdifferentiation of Fibroblasts** - *Mentor: Dr. D S Katti* (Jan'13-Apr'13)
  - Studied state of the art techniques and suggested improvements for treatment of osteoarthritis
  - Proposed osteochondral interface using fibroblasts to transdifferentiate to osteocytes and chondrocytes
  - Proposed its fabrication using Agarose-hydroxyapatite composite while maintaining a gradient of transdifferentiaion factors

## TECHNICAL SKILLS

- **Programing Languages:** Python, R, Perl, Shell, Java, C++, C, PHP, SQL, HTML, JavaScript, MATLAB
- **Data Analysis:** Quality Control, Gene Expression, Prediction, Annotation and Variant Analysis, D3, Hadoop, Spark, Pig, Tableau, Tibco
- **Bioinformatics Tools:** NCBI-Toolkit, KEGG, BioModels, SBML, GATK, APIs for NCBI, KEGG etc., Bowtie2, BED tools, SAM tools, BWA, Gene Pattern, GSEA, Genome Browsers (Jbrowse, Gbrowse etc.)

## RELEVANT COURSES

Genomics and Applied Bioinformatics  
 Structural Biology and Bioinformatics with Lab  
 Biochemistry and Biochemical Engineering with Lab  
 Molecular Biology and Lab

Programming for Bioinformatics  
 Computational Genomics  
 Machine Learning  
 Data and Visual Analytics

Deterministic and Linear Optimization  
 Intermediate Statistics  
 Probability and Statistics  
 Calculus and Linear Algebra