Arjun Baghela

Bioinformatician & Data Scientist, PhD Candidate
arjunsbaghela.com | arjunsbaghela@gmail.com

I work at the intersection of immunology, statistics, and data science as a PhD Candidate in Bioinformatics. My dissertation involves the application of statistical/machine learning (ML) methods to identify molecular biomarkers with prognostic value in complex conditions like sepsis. I have 6+ years of experience analyzing large biological and clinical datasets using state-of-the-art sequencing technologies and statistical methods, where I have proposed solutions to society's most challenging health problems.

Education

Doctor of Philosophy, Bioinformatics, University of British Columbia (UBC), CA
September 2016 – October 2021 (expected); GPA: 4.33/4.33
Dissertation: Identification of gene expression biomarkers to triage early sepsis patients.
Supervisors: Drs. REW Hancock (Immunology) and G Cohen-Freue (Statistics).

Bachelor of Science (w/ Distinction), Biochemistry, Simon Fraser University (SFU), CA
September 2011 – December 2015; GPA: 3.80/4.33

Highlights

Programming Languages – R (expert), Python, SQL (proficient)
Data Engineering & Applied ML –Tidyverse, Shiny, glmnet, caret (R); Pandas, Scikit-learn (Python)
Data Science Workflows – Git, Make, Jupyter, R Studio, Anaconda
Bioinformatics – Experimental design, high-throughput data analysis, Bioconductor software
Teaching/Leadership – Data science TA, intro R workshop leader, 3x undergrad research mentor
Project Management – Leading one of the largest sepsis transcriptomics study (500+ patients) in the world from five global cohorts, helped acquire $400K+ CAD in funding for related projects

Work Experience

Bioinformatics Scientist, Center for Microbial Diseases and Immunity Research, UBC, CA
September 2016 – Present

- Applying differential abundance, functional enrichment, clustering, and predictive models for biomarker discovery in sepsis and Covid-19 sepsis. Collaborate with non-bioinformatician colleagues to plan and analyze various genetics experiments (two publications).
- I have mentored three undergraduate students working on parallel projects. Helped students apply ML pipelines and led the effort to centralize published biomarkers of sepsis (SeptiSearch; manuscript in progress).

Teaching Assistant, Masters of Data Science (MDS), UBC, CA
August 2018 – Present

- Leading tutorials and labs for 30+ MDS students in various courses. I have taught courses related to experimental and causal inference, regression, data management and databases, reproducible workflows, and machine learning.

- Organizing and leading introduction R and Python workshops to graduate students wanting to expand their data science toolkit.

Learning and Writing Peer Educator, SFU, CA
September 2015 – August 2016

- Worked with fellow university students to improve their writing, study, and note taking strategies, primarily through one-on-one advising sessions.

Undergraduate Researcher, SFU, CA
January 2015 – September 2015

- Introduction to research and bioinformatics; led a mutational screen of *Caenorhabditis elegans* to identify essential genes for survival. I applied key genetics and inheritance principles to design screens.

Publications

Also available on Google Scholar.

A Baghela et al. Early gene expression signatures triage hospital patients with prospective sepsis into mechanistic endotypes, predict severity, and are relevant to Covid-19 sepsis. *Manuscript in progress.*

A An … A Baghela et al. An overview of biological and computational methods for designing mechanism-informed anti-biofilm agents. *Manuscript in review.*

D Pletzer … A Baghela et al. The stringent stress response controls proteases and global regulators under optimal growth conditions in *Pseudomonas aeruginosa*. *mSystems*, 2020.

B Dhillon … A Baghela et al. Systems Biology Approaches to Understanding the Human Immune System. *Frontiers in Immunology*, 2020.

M Alford, A Baghela et al. NtrBC regulates invasiveness and virulence of *Pseudomonas aeruginosa* during high-density infection. *Frontiers in Microbiology*, 2020.

B Grande, A Baghela et al. Hackathon-driven tutorial development. *F1000 Research*, 2018

Select Talks & Poster Presentations

Mechanistic insights into human diseases, e.g. sepsis, from large omics datasets (Talk). UBC Department of Immunology Summer Series. August 2020.

Using Gene Expression and Clinical Data Profiles to Predict Sepsis at ER Admission (Poster). *Intelligent Systems for Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB)*. Basel, Switzerland. July 2019.

Machine learning approaches for classifying emergency room patients progressing to sepsis (Talk). *Casa Matemática Oaxaca (CMO) Workshop on Statistical and Computational Challenges in High-Throughput Genomics*. Oaxaca, Mexico. November 2018.

Applying biclustering to uncover disease endotypes within the sepsis syndrome (Poster). ECCB. Athens, Greece. September 2018.

Scholarships & Awards

UBC Centre for Blood Research (CBR) Travel Award, December 2019
UBC CBR Graduate Award Program, September 2017
NSERC-CREATE ECOSCOPE Trainee (2-year funding), September 2017
Golden Key Society Inductee, December 2015