Kyle Alford

(313)-409-1230 | kla2122@columbia.edu
www.linkedin.com/in/kyle-alford-987128195/

EDUCATION

Columbia University **New York, NY**
B.S., Computer Science, Biological Sciences, GPA 3.88/4.0 Expected 05/2023

- Awarded Dean's List all eligible semesters
- Chosen as a Gerald E. Thomson Undergraduate Pre-Medical Program (GET-UPP) scholar for the Fall 2021 semester

RESEARCH INTERESTS

My research interests are in developing and applying computational methods to biological questions, particularly in oncology (Breast, ovarian and lung cancer). Additionally interested in the intersection of healthcare and computer science (Biomedical Informatics) such as modeling disease or predicting treatments from clinical data. My career goals will to receive an MD/PhD (BMI) and work on advancing medical decision making research and integrating it in clinical settings.

RESEARCH

Applying Hidden Markov Models to single cell genomics data **New York, NY**
Advised by Drs. Sohrab Shah and Andrew McPherson at MSKCC 05/2021 – 09/2021

Problem	Copy Number Aberrations are mutational events that lead to abnormal numbers of copies in genomic regions implicated in several forms of cancer, so accurately profiling them may lead to understanding early progression in tumors. The lab's algorithm segments WGS read counts into regions of equal copy number then assigns a copy number state to each region using an HMM. The current model depends heavily on the hyperparameters and also misses focal amps/dels.
Data	I used ovarian cancer single cell data, which holds information on the individual cells such as gene names, chromosome position, copy number counts, etc.
Method & Results	I improved the lab's existing software, HMMCopy, using probabilistic programming in Python (Pyro) to replace the algorithm's SGD with stochastic variational inference then benchmarked different emission distributions to find a model robust to the hyperparameters (transition prior, strength, etc.). Other improvements include improving UI by allowing user to choose between multiple models w/ different hyperparams, rearranging the dataset using Pandas so transitions between chromosomes in different cells are no longer modeled (only chrs within single cells) and creating several summary statistic plots using Seaborn and Matplotlib.

Measuring the effect of indoor Volatile Organic Compounds on pulmonary disease **Coral Gables, FL**
Advised by Dr. Naresh Kumar at University of Miami Miller School of Medicine 09/2019 – 02/2021

Problem	Volatile Organic Compounds (VOCs) are sets of organic chemicals found in consumer products, furnitures, paints, etc., and are linked with pulmonary diseases. There was limited literature performing statistical analysis to concretely correlate VOCs and disease.
Data	I compiled data from 200+ peer-reviewed papers from observational clinical studies, including the study population, relevant demographic info and the types of outcomes including lung cancer, asthma, wheezing, throat irritation, etc.
Method & Results	I used R to compute several values such as odds ratios, effect sizes and correlation coefficients to correlate VOCs and pulmonary diseases, with geographic location, age and sex as relevant confounders. I found that VOCs have a medium-sized effect on pulmonary disease, which increased in high-income countries and older populations.

Peer-Reviewed Publication, Journal of Environmental Research & Public Health, Feb 2021
Alford, Kyle L, and Naresh Kumar. "Pulmonary Health Effects of Indoor Volatile Organic Compounds-A Meta-Analysis." International journal of environmental research and public health vol. 18,4 1578. 7 Feb. 2021, doi:10.3390/ijerph18041578

PUBLICATIONS

Alford, Kyle L, and Naresh Kumar. "Pulmonary Health Effects of Indoor Volatile Organic Compounds-A Meta-Analysis." **International journal of environmental research and public health** vol. 18,4 1578. 7 Feb. 2021, doi:10.3390/ijerph18041578

PRESENTATIONS

Online Workshops, Columbia University, Columbia Data Science Society (CDSS) **New York, NY**
- Data Visualization in R using ggplot2 (03/2021)
- Data Ontologies 101 (02/2021)
- Intro to RNA-seq in R using Bioconductor (02/2021)

Online Oral Presentations, Columbia University, Columbia Data Science Society (CDSS) **New York, NY**
- 2021 CDSS 7th Annual Hackathon (09/2021)
- 2020 CDSS 6th Annual Hackathon (09/2020)
- Students of Color in Data Science panel (03/2021)
- Data Science in Biotechnology panel w/ EVQLV (12/2020)

Poster presentation, University of Miami, Undergraduate Research Symposium **Coral Gables, FL**
- Inhibition of Metazoan Tyrosinase in reducing neurological disorders (08/2019)

TEACHING

Workshop Leader & Mentor | Systems Biology Initiative, Columbia University **New York, NY**
Student organization focusing on Computational Biology 09/2020 – Present
- I lead a program to introduce undergraduates to computational biology. I designed and taught workshops covering topics in RNAseq, Bioconductor, Machine Learning and more. I participated on a team to design the annual Systems Biology Career Expo 2020, where I also led a workshop in RNAseq

SCIENTIFIC COMMUNITY ACTIVITIES

Columbia Biomedical Accelerator (BiomedX) **New York, NY**
Project Manager 02/2021 – Present
- Facilitate commercialization of clinical solutions driven by teams of engineers, clinicians and business representatives for Columbia's Biomedical Devices Accelerator

Columbia Data Science Society **New York, NY**
Executive Board Member 08/2020 – Present
- Organized several events throughout the academic year for both Columbia students and other universities in partnership events, including Hackathons that serve 200+ students

Columbia Systems Biology Initiative **New York, NY**
Executive Board Member 08/2020 – Present
- Created several events and computational biology workshops teaching topics in computational genomics, evolutionary biology

CLINICAL EXPERIENCE

Moffitt Cancer Center & Research Institute **Tampa, FL**
Clinical Volunteer 05/2021 – Present
- Escort patients to exam areas, assist with stocking and maintaining inventory, assist with preparation for each patient in exam areas, provide wheelchair transport if needed and support staff as directed.

Tampa General Hospital **Tampa, FL**

Clinical Shadowing 04/2019 – 06/2019

- Shadowed along with several medical teams including Emergency Room, pediatrics, radiology, oncology and pathology. Learned first-hand several procedures used in diagnosing diseases and conditions, and ability to operate several biomedical machines such as CT, MRI, X-Ray, etc.

Skin Health Forever Dermatology Center **Tampa, FL**

Clinical Shadowing 05/2018 – 08/2018

- Shadowed along a Dermatologist and his medical team for the summer of 2018, observing diagnosis, treatment, consultation, etc. in the examination rooms
- Received first-hand training on how to use equipment, as well as several procedures such as punch biopsies, shave biopsies, cryosurgeries and acne surgeries

RELEVANT COURSEWORK

Computational: Introduction to Data Science in Python, Introduction to Programming in Java, Data Structures & Algorithms, Advanced Programming, Computational Linear Algebra, Artificial Intelligence, Calc-Based Statistics and Probability

Biomedical: General Biology, Evolutionary Biology, General Chemistry, Organic Chemistry, Physics

SKILLS

Programming Languages: Python, C/C++, Java, R, HTML, JavaScript, CSS
Tools/Packages: Numpy, Scipy, Matplotlib, PyTorch, Pyro, TensorFlow, Pandas, Scikit-learn
Languages: English, Spanish, Hindi, Urdu