

Sam Park

703-966-8310 sp5fd@virginia.edu www.sanghoon.io

Education

University of Virginia, Charlottesville, Virginia

Master of Science, Program: Data Science

Bachelor of Science, Major: Biomedical Engineering | Minor: Computer Science

Experience

Bioinformatician, University of Virginia

2024 – Current

- Contributed to React + TS frontend and Python FastAPI for lab software and maintained R libraries published by Sheffield Lab
- Performed sex-stratified GWAS study of COPD patients to investigate how sex and ancestry contribute to COPD severity
- Analyzed ATACseq data of mouse HSC-myeloid cells to investigate how differentiation and aging impacts gene accessibility

Rotational Automation Engineer, Merck & Co. (Contractor)

2022 – Current

- **Data Science:**
 - Wrote R scripts and functions incorporating PI Web API to track batch data and events internally
 - Helped develop Shiny dashboards incorporating interactive Plotly and timevis plots to provide detailed factory batch tracking
 - Produced visually appealing ggplot2 figures to present any accompanying data analysis results
 - Worked on automated data pipelines that use R Markdown to manage dashboard data on AWS servers
 - Performed exploratory machine learning analysis using correlation matrices, PCA, logistic regression, and random forest models on batch data to determine feasibility of predicting batch failures across time
- **Automation Support:**
 - Authored continuous historian reports with PI DataLink to support technical operations
 - Developed real-time PI ProcessBook displays of manufacturing areas to support investigations

R&D Intern, Predictiv Care, Inc.

2022 – 2024

- **Pharmacogenomics (PGx) and VCF Annotation:**
 - Used published CPIC guidelines to identify customer star alleles and drug metabolism phenotypes from customer variant call format files (WES and WGS)
 - Performed additional VCF file annotation with levels of pathogenicity and accompanying levels of clinical evidence for each matched variant with published ACMG and ClinGen guidelines
- **Polygenic Risk Scores (PRS):**
 - Wrote R scripts that incorporate published PRS models to calculate risk scores with customer VCF files
 - Reviewed LDpred2 and other algorithms to study established PRS model development practices
- **Additional Projects:**
 - Explored using 23andMe and AncestryDNA output files to identify important variants for customer Wellness reports

Projects and Awards

Automated Continuous Historian Tool

2023

- Helped lead Merck automation community project incorporating R Shiny, plumber API, HTML/CSS, AWS, and PI Web API to automate our continuous historian report writing process, eliminating the workload of 500+ reports needing to be manually written per year
- Developed initial proof of concept that served as foundation for codebase
- Used Git for version control and Posit Connect to deploy frontend and backend components of the tool
- Explored GMP and regulatory processes needed to document and test the tool before product release

Identifying Patterns of Metabolic Dysfunction in Inflammatory Bowel Disease

2021

- Received the UVA Harrison Undergraduate Research Award (funded research grant; \$4,000) for a proposal involving modeling cellular metabolism in inflammatory bowel disease using patient gene expression data
- Research incorporates published RNAseq data with genome-scale metabolic modeling to identify trends in disease
- Used Kallisto and DESeq2 to quantify transcript abundances and GIMME to constrain metabolic models
- Presented findings at the 2021 Annual BMES Conference and the 2022 UVA Undergraduate Research Symposium

Skills

Programming and Data Analysis

- R, Python, HTML/CSS/JavaScript, TypeScript, React.js, Git, Jupyter, SQL
- R packages: tidyverse, data.table, ggplot2, Plotly, timevis, Shiny, plumber

Technical

- AWS, PowerBI, ImageJ, Adobe Photoshop, MS Office