Sam Park

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

Education

University of Virginia, Charlottesville, Virginia

Master of Science (G. 2026). Program: Data Science
Bachelor of Science (G. 2022). Major: Biomedical Engineering | Minor: Computer Science

Experience

Data Visualization Engineer, Vedanta Biosciences (Contract) **2025 – Current**

- **Process Dashboard:**
  - Developed a probiotics manufacturing process dashboard using a React frontend and FastAPI backend
  - Implemented Mosaic and vgplot to enable real-time cross-filtering and linked selections across multiple coordinated views
  - Built interactive time-series and clustering visualizations tracking fermentation titer, projected recovery rates, and metabolite concentrations across dozens of batches

Bioinformatician, University of Virginia **2024 – Current**

- **Software:**
  - Contributed to React + TS frontend and Python FastAPI for lab software published by Sheffield Lab
  - Built PEPhub schema editor with versioning, tagging, and metadata management to allow for more structured validation of PEPs
  - Integrated Apple's Embedding Atlas into BEDbase for interactive exploration of genomic region embedding space, enabling users to visually explore thousands of BED files or upload their own for comparison
  - Developed Refget SCOM interface to visualize comparisons between dozens of genomic sequence collections via Vega-Lite
  - Built Refgenie frontend interface to allow for browsing hundreds of reference genomes and associated assets visually
  - Wrote R bindings for gtars, a high-performance toolkit for genomic tools and algorithms written in Rust
- **Research:**
  - Analyzed ATACseq data of mouse HSC-myeloid cells to investigate how differentiation and aging impact gene accessibility
  - Performed sex-stratified GWAS study across TOPMed cohorts to investigate how sex contributes to COPD phenotypes
  - Colocalized above GWAS with sex-stratified pQTL results to identify causal SNPs linked to protein expression and COPD
  - Performed QC, clustering, and differential expression analyses on single cell RNA lung tissue samples using Seurat
  - Explored the interactive visualization of scRNAseq embeddings using Mosaic + vgplot and Embedding Atlas

Rotational Automation Engineer, Merck & Co. (Contract) **2022 – 2025**

- **Data Visualization:**
  - Helped lead automation community project incorporating R Shiny, plumber API, AWS, and PI Web API to automate manual continuous historian report writing process, eliminating the workload of 500+ reports written annually
  - Developed initial proof of concept for above tool that served as foundation for codebase
  - Built interactive Shiny dashboards incorporating Plotly and timevis plots to track concurrent batch events across 3 facilities
  - Worked on automated data pipelines that use R Markdown and PI Web API to manage factory data on AWS S3 and RDS
  - Designed real-time PI ProcessBook displays to help with real-time monitoring and investigations of manufacturing areas

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

Skills

Interfaces

- **Frontend:** React, React Router, TanStack Query, Zustand, TypeScript, JavaScript, HTML, CSS, Bootstrap
- **Data Visualization:** Vega-Lite, Mosaic/vgplot, Observable Plot, D3, Plotly, Plotly Dash, ggplot, timevis, Shiny

Infrastructure

- **Backend:** FastAPI, plumber, SQL, DuckDB, AWS, Posit Connect
- **Tools:** Git, GitHub, Jupyter, R Markdown, Slurm
- **Data Analysis:** tidyverse, data.table, Seurat, DESeq2, GENESIS, SeqArray, coloc, susieR, Pandas, NumPy