

Taleb Ahsan

+1 (408) 669-7052 • taleb.ahsan@gmail.com • [LinkedIn.com/in/tahs](https://www.linkedin.com/in/tahs)

EXPERIENCE

Data Scientist - Fred Hutch

June 2023 - Current

- Successfully distinguished cancerous cells from thousands of others by processing, standardizing, and evaluating scRNA-seq data across more than 50,000 cells. This involved handling an average of 2 million reads per cell and efficiently managing a dataset totaling over 100 billion data points
- Leveraged an advanced deep learning framework built on scikit-learn to classify dimensionality reduced feature vectors (cells), employing the Tabula Sapiens dataset for effective training (yet to be published)
- Built simulation of surface protein expression of 40 million bone marrow cells using data augmentation
- Created visualizations to allow bioengineers to intuitively understand workflow and key findings

Data Science Intern - Giachelli Lab, University of Washington

July 2021 - June 2023

- Identified a key gene regulatory network involved in the onset of calcific aortic valve disease (CAVD)
- Directed comprehensive differential expression analysis in RNA-seq datasets using edgeR, from initial preprocessing to normalization and statistical modeling. Utilized generalized linear models to dissect complex gene expression patterns and online graph-based databases to find relevant biological applications
- Executed broad array of methodologies, including data cleaning (identification and imputation of missing values, outlier detection), exploratory data analysis (PCA for dimensionality reduction, unsupervised clustering for pattern detection) and rigorous statistical testing (ANOVA, t-test)
- Used a fast approximation algorithm to align human RNA-sequencing data to validate murine models

Data Science Intern - Luscombe Group, University of Washington

February 2021 - June 2021

- Manually annotated hundreds of rows of data while working on the feature engineering team to help train an ML model designed to predict complex chemical reactions; used python to web scrape chemistry textbooks
- Learned firsthand model development using scikit-learn from postdocs, built alphabet identification model

PROJECTS

Poster Presentation @ Vasculata 2022, Duke University

- Distilled high-dimensional data into clear, accessible visual formats to cater to academic researchers
- Provided novel understanding of calcific aortic valve disease (CAVD) onset on a molecular level through employing statistical models and basic machine learning algorithms (findings to be published)

Identification of Surface Markers in Endothelial-Mesenchymal Transition

- Found cell surface markers used to discern where a cell is in the endothelial-mesenchymal transition
- Utilized Seurat R package for comprehensive scRNA-seq data analysis, focusing specifically on endothelial to mesenchymal transition (EndMT) stages. Applied UMAP dimensionality reduction and clustering
- Employed advanced pseudotime analysis techniques to delineate the progression and transitional hierarchy between distinct cell states, and strategically binned cells based on their developmental stages (inspired by the methodology outlined by Simeonov, 2021)

EDUCATION

University of Washington, Seattle, Washington

Fall 2019 - June 2023

- **B.S. in Biochemistry**
Algorithms, Data Structures, Web Programming, Mathematical Reasoning, Genomics, Multivariable Calculus, Matrix Algebra, Differential Equations

SKILLS

Python, pandas, SciPy, NumPy, scikit-learn, TensorFlow, Torch, R, MySQL, Matplotlib, Prism, seaborn, GGplot2, Git