

Cole Phalen
Bioinformatics Analyst: Single Cell Genomics Specialist
Seattle, WA
Phone: (425)-444-2633
Email: colephalen@gmail.com

PROFESSIONAL SUMMARY

Data Science student with a strong foundation in bioinformatics, machine learning, and single-cell genomics. Currently completing a Master of Information and Data Science (MIDS) at UC Berkeley. Proven ability to deliver end-to-end data science solutions, from experimental design and data wrangling to deep learning modeling and results interpretation. Skilled in Python, R, Machine Learning, SHAP explainability, and multimodal integration of high-dimensional biological data. Published co-author and experienced cross-functional collaborator with a focus on driving biological insights and improving human health outcomes.

TECHNICAL SKILLS

Programming Languages: Python, R, Bash

Tools & Frameworks: Jupyter, Pandas, ggplot2, Plotly Dash, bedtools, Bowtie2, Apache Spark, Hadoop

Machine Learning: TensorFlow, Random Forest, SHAP, Convolutional Neural Networks (CNN), Transformers, Hugging Face, LLM Fine-tuning, Deep Learning

Bioinformatics: Seurat, Scanpy, scVI, Scrublet, Celltypist, MRVI,

Data Science: NLP, Multimodal Modeling, Data Visualization, Statistical Modeling, Model Interpretation

PROFESSIONAL EXPERIENCE

Allen Institute for Immunology – Seattle, WA

Bioinformatics Analyst: Single Cell Genomics Specialist

September 2023 – Present

- Lead data analyst and co-author on CryoSCAPE (J Transl Med, 2025), validating cryopreserved whole blood for immune profiling.
- Developed deep learning model using CNN to classify single-cell gene expression by patient type and stimulation status; achieved over 97% accuracy.
- Applied SHAP explainability to identify key gene drivers, informing biological validation and follow-up experiments.
- Designed targeted gene panels for spatial transcriptomics using 10X Genomics Xenium platform.
- Trained wet-lab scientists on computational workflows, improving reproducibility and team efficiency.
- Presented data analyses to scientific and executive audiences to support data-driven decision making.

Research Associate II

December 2021 – August 2023

- Created visualization tools for TEA-seq (RNA + CITE + ATAC) datasets used by international collaborators.
- Conducted comparative analysis of 10X Genomics v3.1 vs. Flex chemistry; enabled cost-saving protocol changes.
- Led 4-person team to develop new intracellular antibody staining and RNA-seq assay.
- Beta-tested and troubleshooted new single-cell assays, increasing data yield by 24× through barcode optimization.
- Analyzed gene expression across stimulation conditions to identify pathways relevant to rheumatoid arthritis.

Allen Institute for Brain Science – Seattle, WA

Research Associate I

July 2018 – December 2021

- Evaluated RNA-seq pipelines (Scanpy, scVI, Wishbone) for Alzheimer's research team.
- Performed 1000+ genotyping reactions weekly with over 98% success rate.

Seattle Reproductive Medicine / Overlake Reproductive Health – Seattle, WA

Andrology Technician

September 2017 – July 2018

- Processed and prepared samples for IVF and IUI procedures.
- Conducted endocrinology testing and cryopreservation to support reproductive health treatments.

Infectious Disease Research Institute – Seattle, WA

Intern, Tuberculosis Drug Discovery Group

June 2015 – August 2015

- Designed and validated assay for high-throughput screening of TB drug candidates.

Matrix Genetics – Seattle, WA

Intern, Synthetic Biology Group

June 2014 – August 2014

- Engineered CRISPRi system for targeted transcriptional repression in cyanobacteria.

EDUCATION

University of California, Berkeley

Master of Information and Data Science (MIDS) | GPA: 3.93

August 2023 – December 2025

Relevant Courses: Applied Machine Learning, NLP with Deep Learning, Generative AI, Machine Learning at Scale, Data Engineering

Projects:

- Built multimodal CNN to predict wine scores from descriptions and metadata (>70% test accuracy).
- Fine-tuned FLAN-T5 on KELM dataset for knowledge graph triple generation; outperformed GPT-4o-mini, Llama 3.2, Mistral 7B.
- Used Spark to conduct longitudinal analysis to predict flight delays using 5 years of flight data covering millions of flights.

University of Washington – Seattle, WA

Certificate in Python Programming

February 2019 – November 2019

Whitman College – Walla Walla, WA

BA in Biochemistry, Biophysics, and Molecular Biology

August 2012 – May 2016