

Yifan Yang

(437) 267-2198 | yifan19970911@outlook.com | [linkedin.com/in/yifan-yang-biostat/](https://www.linkedin.com/in/yifan-yang-biostat/) | yifanyang0810.github.io/

HIGHLIGHTS OF QUALIFICATIONS

Bioinformatician with **3+ years of experience** in RNA-seq analysis, pipeline development, and statistical modeling. Developed a **novel algorithm** for DGE analysis, outperforming conventional tools in small-sample settings, and gained in-depth familiarity with sequencing and analysis techniques. This expertise enables the delivery of **optimal pipelines** tailored to diverse sequencing data (e.g., bulk RNA-seq, single-cell RNA-seq, Ribo-seq). Designed scalable RNA-seq pipelines using Nextflow, improving processing efficiency by 30%. Skilled in Python, R, and HPC tools, with a proven ability to collaborate across disciplines to drive scientific discovery.

EXPERIENCE

Bioinformatician, Medical Genetics

07/2024 – Present

Cumming Medical School, University of Calgary

Calgary, AB

- Designed and deployed an automated scalable RNA-seq analysis pipelines using Nextflow on high-performance computing (HPC) clusters, reducing processing time by 30% and minimizing manual intervention for large-scale genomic datasets
- Developed pipelines for diverse data types, including LC-MS/MS and NGS sequencing (Bulk-Seq, RIP-Seq, polysome profiling, and Ribo-Seq), enabling the identification of 10+ disease-associated genes for gene therapy and drug discovery
- Conducted exploratory data analysis and functional enrichment (GO, KEGG) using Python and R, uncovering key biological pathways and contributing to 2+ peer-reviewed publications
- Collaborated with interdisciplinary teams (biologists, technicians) to translate bioinformatics findings into actionable experimental strategies, streamlining workflows and enabling faster decision-making in disease research

Researcher, Bioinformatics Tool Development

09/2020 – 03/2024

Biology (Bioinformatics) Department, University of Waterloo

Waterloo, ON

- Developed a novel algorithm for DGE analysis, achieving 70% statistical power with a <5% false-positive rate, outperforming traditional tools (e.g., edgeR, DESeq2, Limma-voom) in small-sample settings
- Validated the algorithm on 10+ simulated and real-world datasets across diverse species, demonstrating its robustness for real-world applications
- Contributed to the development of an open-source R package package, enhancing accessibility and usability for the research community to integrate the tool into their analyses

Machine Learning Research Intern (Computational Biology)

02/2022 – 07/2022

Sinovation Ventures

China

- Designed and implemented a data cleaning and preprocessing pipeline in Python, reducing processing time by 20% and ensuring data integrity for downstream analysis
- Managed a MySQL database for large-scale datasets, improving data quality by 15% and enabling efficient retrieval for machine learning workflows
- Applied feature engineering and scaling techniques, integrating critical biological insights to develop algorithms for drug-protein binding affinity predictions, achieving a 5% improvement in prediction accuracy.

TECHNICAL SKILLS

Bioinformatics Tools & NGS Assays: Bulk RNA-seq, single-cell RNA-seq, RIP-seq, Ribo-seq, BWA, Bowtie2, Kallisto, CellRanger, Scanpy, scVI-tools, edgeR, DESeq2

Pipeline Development & Programming: Python (Numpy, Pandas), R (R Shiny, ggplot2), Nextflow (HPC pipeline development), MySQL

Platforms: Linux (Bash scripting), Git version control, HPC clusters (ARC), Docker

Statistical & Analytical Techniques: Statistical modeling, hypothesis testing, machine learning (lasso, ridge, elastic net), dimensionality reduction (PCA, UMAP)

EDUCATION

University of Waterloo

Master of Science (Thesis-based) in Biology, Bioinformatics

Waterloo, ON

05/2024

University of Waterloo

Bachelor of Science (Honors) in Material and Nano Science, Minor in Mathematics

Waterloo, ON

06/2020

PROJECTS

Single-Cell RNA-Seq Analysis of Human Lung Cells Post-COVID-19 Infection	12/2023
<ul style="list-style-type: none">• Conducted scRNA-seq analysis, including data preprocessing, dimensionality reduction (UMAP), and Leiden clustering, to reveal cellular heterogeneity, aligning with findings from the COVID-19 Lung Atlas• Applied trajectory inference and differential expression analysis to identify key pathways and markers associated with post-infection responses, consistent with published studies on COVID-19 lung pathology• Utilized Scanpy and scVI-tools for scalable integration, probabilistic modeling, and annotation refinement, improving the detection of rare cell populations and enhancing biological insights	

CERTIFICATIONS

Machine Learning Specialization, Coursera	10/2022
Clinical Trials: Good Clinical Practice Specialization, Coursera	03/2024
SAS Certified Professional: Advanced Programming Using SAS 9.4, SAS	03/2024
SAS Certified Specialist: Base Programming Using SAS 9.4, SAS	02/2024