Aymen Maqsood Mulbagal

(857) 296-2150 | aymenmaqsoodmulbagal@gmail.com | linkedin.com/in/aymen-maqsood | github.com/mulbagalamaq

EDUCATION

Master of Science in Bioinformatics   September 2022 – December 2024
Northeastern University, Boston, MA   *GPA: 3.7/4.0*
Coursework: Bioinformatics Computational Methods, Statistics for Bioinformatics, Machine Learning and Data Mining, Bioinformatics Programming, Genomics in Bioinformatics, Biotech Enterprise and Visual Analytics
Leadership: Head of Bioinformatics - Graduate Biotech and Bioinformatics Association , Graduate Student Ambassador (COS)

TECHNICAL SKILLS

Programming: Python (NumPy, Pandas, Matplotlib, scikit-learn), R, SQL, C, C++, Java, MATLAB, Bash
Deep Learning & AI Architectures: PyTorch, TensorFlow, Transformers, Graph Neural Networks (GNNs), Attention Mechanisms
Bioinformatics & Computational Biology: Scanpy, scRNA-seq Analysis, Differential Expression, Cell-Type Annotation, Sequence Analysis, Dimensionality Reduction (PCA, UMAP, t-SNE)
Software Engineering & Cloud: Git, GitHub, Docker, AWS (EC2, S3), CI/CD, Unit Testing (pytest), Code Documentation, SLURM
Tools: GATK,GSEA, STAR, BWA, Bowtie, SAMtools, BEDTools, DESeq2, Seurat, Scanpy, IGV, UCSC Genome Browser

PROFESSIONAL EXPERIENCE

Bioinformatics Analyst   March 2025 – Present
JustLabs - VC Studio   *Remote*
- Analyzed large-scale single-cell multi-omics datasets, including methylation, CHIP-seq, ATAC-seq, and RNA-seq
- Designed ML models using TensorFlow to forecast energy consumption patterns of lab equipment and GPUs, capturing runtime/cost metrics and using them to model computational efficiency
- Shipped a full-stack emissions analytics dashboard to detect emission spikes across HPC clusters; reduced GPU usage by 15%
- Tools & Technologies: **Python, TypeScript, React, Node.js, PostgreSQL, Docker, TypeScript, RAG, Docker**

Quantitative Biology & Bioinformatics Analyst Co-op   July 2024 – February 2025
Vor Bio   *Boston, MA*
- Built production-ready CRISPR-WGS analysis pipelines using Nextflow and AWS (EC2/S3), enabling discovery of 24 new off-target edit sites and improving turnaround time by ~7 hours per sample
- Applied statistical and ML methods to assess precision genome editing outcomes from NGS data, delivered reports that accelerated gRNA selection
- Led analysis of single-cell multiomics data (CITE-seq + scRNA-seq) of clinical-stage drug product for acute myeloid leukemia
- Tools & Technologies: **R, Python, BASH, Nextflow, AWS (EC2/S3), R Shiny, Bitbucket, Jira**

Research Associate, Bioinformatics   November 2023 – June 2024
Institute for Experiential AI, Northeastern University   *Boston, MA*
- Streamlined ENCODE eCLIP preprocessing on HPC (SLURM) using Snakemake pipeline to identify RNA-binding motifs
- Developed a sequence search algorithm for enhanced motif discovery and visualization
- Generated ML-ready feature sets for alternative splicing prediction; trained XGBoost models on RNA-binding protein datasets
- Tools & Technologies: **Python, Snakemake, SLURM, XGBoost, BASH**

Research Intern   February 2022 – July 2022
Cancer Systems Biology LAB, BSSE, Indian Institute of Science   *Bangalore, India*
- Engineered a systems biology framework to analyze RNA-Seq data, characterizing metabolic plasticity and hybrid E/M states in cancer metastasis under Dr. Mohit Kumar Jolly
- Constructed and validated a gene regulatory network (GRN) involving ELF3, miR-200, SNAIL, ZEB-1, and SLUG, identifying key topological drivers of the Mesenchymal-Epithelial Transition (MET)
- Simulated phenotypic distributions to elucidate the context-dependent role of lncRNA HOTAIR in modulating EMT dynamics, revealing its dual function in promoting hybrid cellular states and tumor progression
- Applied statistical modeling to quantify correlations between gene expression signatures and phenotypic aggressivity, providing molecular targets for reducing metastatic potential

RELEVANT PROJECTS

BioGraphRAG: GraphRAG for Biomedical Q&A (AWS Open Data Knowledge Graph Hackathon) | Git Repository Link
- Built a graph-native RAG pipeline that leverages the TCGA (129,375 nodes, 4M+ relationships)) and PubMed Knowledge Graphs to accurately transform complex biomedical questions into grounded answers, citing specific PMIDs/experiment IDs.
- Implemented a GNN+LLM fact ranking system that reduced hallucination and improved inference speed by ~15%

NK Cell Differentiation (CITE-seq) Analysis | [Git Repository Link](#)
- Built a GPU-accelerated multimodal pipeline (TOTALVI/scvi-tools) on HPC to jointly model CITE-seq RNA (4,000 HVGs) and 28-protein surface panel from 19,443 NK cells, performing batch correction across donors and generating spatially-aware AnnData objects with denoised protein expression for downstream biological interpretation
- Trained balanced ensemble classifiers (XGBoost, Balanced Bagging Classifier) on the integrated latent space to classify four NK subtypes, achieving 0.89 F1-score with stratified cross-validation

Stack Foundation Model Analysis on Perturb-Sapiens | [Git Repository Link](#)
- Applied Arc Institute's Stack-Large foundation model (trained on 150M single-cell samples) to generate 1,600-dimensional embeddings for 974K cells from the Perturb-Sapien's human perturbation atlas, enabling zero-shot cell representation on A100 GPU
- Performed comparative UMAP analysis of ADSF cytokine (513K cells) vs PBS control (460K cells) perturbation conditions across 75 tissues and 20+ cell types, visualizing embedding structure by cell type and tissue origin

LEADERSHIP

Head of Bioinformatics   August 2024 - August 2025
Graduate Bioinformatics and Biotech Association (Northeaster University)
- Organized a bioinformatics hackathon with 120+ participants, tackling industry-relevant challenges and leading workshops on command-line techniques and version control and launched a monthly journal club
- Spearheaded partnerships with industry professionals, arranging guest lectures to bridge the gap between academic learning and real-world applications

PUBLICATIONS

Qiu, H., Wang, R., Wang, S., Mulbagal, A. M., Anderson, E., & Ge, H. G. (2025). *SANTON: Sequencing Analysis Toolkits for Off-target Nomination*. bioRxiv. https://doi.org/10.1101/2025.05.09.653082

Peyton, M. P., Chung, C., Mohanty, S., Truong, V. Q., Scouten, A., Shukla, S., Lazard, C., Masood, D., Murphy, J. D., Ketter, A., Ahooyi, T. M., Amin, V. R., Arora, A., Baron, J. A., Bazzi, F., Bondali, A., Bristy, N. A., Carter, E. K., Chen, Y., . . . Busby, B. (2025). A Blueprint for Open Science: How Transatlantic Teams Built and Deployed Knowledge Graphs to Enable Biological (AI) Models. BioHackrXiv. https://doi.org/10.37044/osf.io/g4rk2_v1