

JAWAHAR MAHENDRAN

mahendranjawahar@gmail.com | (612) 296-3150 | [Jawahar Mahendran](#) | [jawa23bio](#)

SUMMARY

I'm a curious and driven data scientist who translates complex datasets into meaningful insights. With hands-on experience building data pipelines, predictive models, dashboards, and GenAI tools, I'm at my best in research-driven environments where data quality, and real-world impact are paramount. Passionate about advancing digital innovation to improve efficiency and outcomes in the pharmaceutical space.

PROJECTS

NGS Data Analysis ([Github](#)) Jan 2024 – May 2024

Developed automated, scalable Snakemake pipelines to process and analyze multi-omics datasets, boosting efficiency by 40%. Created advanced visualizations to distill complex genomic data into definitive conclusions for research decision-making.

DrosaGenXplorer ([Github](#)) Nov 2023

Built and deployed a self-service analytics platform using R Shiny, enabling scientists and non-technical users to perform exploratory data analysis (PCA, t-SNE) across 100+ genomic datasets and independently extract key biological insights.

Diseasome Network Analysis ([Github](#)) Mar 2024

Architected a high-throughput pipeline to mine disease-gene-phenotype relationships from public databases like ORPHANET and Open Targets, optimizing HPC workflows to boost analytical throughput by 80% and visualizing the resulting network in Cytoscape.

PROFESSIONAL EXPERIENCE

Data Science and Machine Learning Co-op Sep 2024 – Sep 2025

GlaxoSmithKline, Collegeville, PA

- Designed and automated Spotfire dashboards for real-time equipment utilization, ELN compliance, and user-engagement metrics, providing lab scientists and R&D stakeholders with centralized, actionable insights that increased R&D operational efficiency by 15%.
- Developed a novel AI-enabled Flask web app to generate, automate, and sync technical dashboard documentation with GitHub, cutting manual effort by over 99%, drastically reducing turnaround times, and streamlining R&D collaboration.
- Implemented a production-grade Spotfire comments feature with ETL/ELT pipelines and database integration to enable GenAI-driven reporting, allowing scientists to annotate visualizations and enhance collaborative insights.
- Collaborated with technical and scientific stakeholders to optimize and validate a standardized data framework, enabling 75% adoption across R&D teams and supporting scalable, consistent, and analytics-ready data integration.
- Scoped mixing and stability batch processes to identify automation opportunities across lab workflows and data management, driving improvements in operational efficiency and product quality within Drug Product Development teams.

Data Science Intern – Bioinformatics & Multi-Omics ([Github](#)) ([Poster](#)) Jun 2024 – Aug 2024

Boston University Medical Campus, Boston, MA

- Programmed a high-throughput ChIP-seq analysis pipeline using Nextflow, enabling scalable, modular, and reproducible processing of large-scale raw sequencing data for transcription factor binding site discovery.
- Integrated proprietary and public ChIP-seq datasets, applying statistical overlap analysis to identify 200+ co-occupied genomic regions and uncover potential patterns of transcription factor cooperation in stress-response pathways.
- Harmonized multi-omics data using correlation analysis, and signal enrichment profiling to model GPS2-ATF transcriptional regulation.
- Produced publication-ready visualizations, including heatmaps and UCSC genome browser tracks, to translate high-dimensional ChIP-seq and epigenomic data into interpretable biological understandings for therapeutic target exploration.

Bioinformatics Research Analyst ([Github](#)) Jan 2024 – Aug 2024

Boston University Medical Campus, Boston, MA

- Engineered and maintained a relational database (SQL) for TF-gene-CRE networks in Alzheimer's disease, and developed an interactive web interface using Python, JavaScript, CGI, AJAX, and HTML to enable queryable access to regulatory relationships.
- Processed and integrated multi-omics datasets (scRNA-seq, ATAC-seq, eQTL) to identify transcription factors and 170+ cis-regulatory elements associated with SNPs, supporting feature extraction for modeling regulatory influence.
- Designed a two-stage predictive model using Pecanpy for graph embedding and TensorFlow (Keras), achieving 60% accuracy in forecasting gene expression and TF binding affinity.

Bioinformatics Research Assistant ([Publication](#)) Dec 2023 – Feb 2024

Boston University, Boston, MA

- Built pipelines for ATAC-seq and CUT&RUN to analyze gene regulation around the Dlk1-Dio3 ncRNA cluster using CRISPR knockout data.
- Conducted motif enrichment (GSEA), differential binding (DiffBind), and pathway analysis (ChIPSeeker), integrating histone modification data to identify key regulators of mitochondrial function and transcriptional dynamics in muscle development.

Junior Validation Engineer Jun 2022 – Aug 2023

ValGenesis, Chennai, India

- Automated documentation workflows for computer system validation, ensuring compliance with 21 CFR Part 11 and cGxP standards.
- Leveraged Azure DevOps and Team Foundation Server (TFS) to streamline test case execution, issue tracking & audit trail management.

Research and Development Officer Jun 2021 – Jun 2022

Matrix Labs Pvt. Ltd, Chennai, India

- Developed 12+ diagnostic kits (CLIA, ELISA, FIA) with >95% accuracy, reducing R&D time by ~1.5 months via in-silico design.
- Established reproducible validation workflows and performance testing pipelines, improving product reliability for point-of-care use.

SKILLS

Programming & Tools: R, Python, SQL, PostgreSQL, JavaScript, Bash, Git, REST APIs, DevOps, Databricks, ETL/ELT, Docker, Flask, CI/CD.

DS, ML & Cloud: Predictive Modeling, Machine Learning (Supervised/Unsupervised), Scikit-Learn, Statistical Inference, Database Design & Management, Microsoft Azure, AWS, GenAI (prompt design, LLM-based summarization), Data Modelling, Pipeline Orchestration.

Data Visualization: Spotfire, Power BI, R Shiny, Tableau, Cytoscape.

Bioinformatics: RNA-seq, ATAC-seq, ChIP-seq, scRNA-seq, eQTL, CRISPR-Cas9, WES, Nextflow, Snakemake.

EDUCATION

Boston University Boston, MA

Master of Science in Bioinformatics | Merit Scholarship | (Grade: 3.8 / 4.0)

Sep 2023 - Sep 2024

Anna University Chennai, India

Bachelor of Technology in Industrial Biotechnology | (Grade: 3.4 / 4.0)

Jun 2017 - May 2021