

Kyuho Kim

McLean, VA | (516) 289-3255 | Kyuhokim2@gmail.com | <https://www.linkedin.com/in/kyuho-kim>

## EXPERIENCE

Associate Bioinformatics Engineer / Data Scientist

AstraZeneca Pharmaceuticals, March 2023 - Present

- Developed and maintained scalable data processing pipelines using Python, Bash, and Java, aggregating and analyzing multi-omics data from 10,000+ samples.
- Collaborated cross-functionally with scientists and analysts to develop automated model QC workflows integrating LLMs and AWS cloud services for chat-based querying.
- Designed and implemented a production-grade database adhering to FAIR principles, used for sample quality inference and analytics.
- Led containerization and orchestration of 20-step analysis workflows using batch job scheduling and centralized logging.
- Regularly contributed to internal GitHub repositories (>200 commits), ensuring maintainability and reproducibility of code.
- Maintained internal gene signature database with versioned updates, similar to MSigDB.
- Built cost analytics pipelines for sequencing using SQL, cloud metrics, and financial modeling.
- Used CI/CD principles to ensure quality pipeline releases in a high-performance computing environment.

Data Scientist / Bioinformatician

Psomagen Inc., March 2022 - March 2023

- Developed multi-modal genomic analysis pipelines in Java, Python, R, and Bash.
- Employed Sun Grid Engine and UNIX-based clusters for distributed computing on NGS datasets.
- Created automated systems for sequencing QC and reporting using shell scripting and job schedulers.
- Applied tools such as Cellranger, BWA-GATK, and Seurat in production-grade workflows.
- Spearheaded team adoption of Agile methodology and version control with Git and Jira.

## EDUCATION

MS in Bioinformatics

Johns Hopkins University - Baltimore, MD

Sept 2020 - May 2023

BS in Biochemistry

Union College - Schenectady, NY

Sept 2012 - June 2016

## TECHNICAL SKILLS

- Languages: Java, Python, Bash, R, SQL, Typescript, JavaScript
- Frameworks/Tools: Spring Boot, Gradle, REST, Git, Jenkins, Sonar, Postman, IntelliJ, JSON, YAML
- Data/Modeling: Monte Carlo simulations, regression analysis, model QA, pandas, NumPy
- Systems: Linux/Unix, RedHat Tomcat, Autosys, Informatica, KDB/q
- Cloud: AWS (Certified Solutions Architect - Associate), distributed compute/grid systems
- Agile/CI-CD: Jira, GitHub, Jenkins, Bitbucket