

# Joseph Min

Pasadena, CA | ☎(626) 240-9851 | ✉phoenixterg6@gmail.com | 🌐lioscro | inlioscro

## EDUCATION

### MS, Computer Science

MIT | Cambridge, MA

### BS, Biology and Computer Science

Caltech | Pasadena, CA

## SKILLS

### Programming

- **Python** (including Cython)
- **Javascript**, Typescript
- Ruby on Rails
- SQL
- C# (.NET Framework)
- C, C++
- R

### Software

- Git, CI/CD
- Docker
- Apache Airflow
- Scanpy, AnnData
- samtools, bcftools
- AWS (Batch, EC2, S3, Lambda)
- Nextflow
- React.js
- Snowflake
- Kubernetes

### Other

- Software design & architecture
- Orchestration & distributed computing
- Bioinformatics/Computational biology (NGS)
- Technical communication

## PUBLICATIONS

- [1] Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. in press, 2024
- [2] Dynast: Inclusive and efficient quantification of metabolically labeled transcripts in single cells. MIT, 2022
- [3] Modular and efficient pre-processing of single-cell RNA-seq. Nature Biotechnology, 2021.

## EXPERIENCE

### Senior Software Engineer

Aug 2022 – Present | Ginkgo Bioworks, Boston, MA

- Developed backend software solutions for synthetic biology applications in a highly collaborative and Agile environment.
- Served as technical lead for a software engineering team embedded in a major pharmaceutical project.
- Owned the development and maintenance of multiple bioinformatic pipelines, including those for variant calling and copy number estimation.
- Led cross-functional efforts involving multiple engineers to enhance the reliability, scalability, and throughput of computational biology workload orchestration.
- Mentored junior engineers and interns in software engineering best practices.

### Software Developer

Aug 2020 – Jun 2022 | MIT, Cambridge, MA

- Developed software tools for cancer evolution and synthetic biology research.
- Processed terabyte-scale multi-modal (single-cell) sequencing datasets using LSF in a distributed, parallelized environment.
- Designed and implemented computational algorithms for lineage tracing, preprocessing and spatial transcriptomics.

## PROJECTS

### Genomic Edit Tracking

Apr 2024 - Mar 2025 | Ginkgo Bioworks

- Built a system with Git-like version control to track and verify millions of genomic edits across 100k samples, enabling branching, merging, and rebasing of genome edits.
- Owned the bioinformatic pipeline to verify genomic edits, including optimizations that reduced runtime by 50% and costs by 90%.
- Developed a high-throughput Nextflow pipeline to detect unintended edits (variants) and predict their impact, processing over 5k samples per hour.
- Designed and implemented an ETL process with Apache Airflow to load data into a Snowflake warehouse for user access and downstream analysis.

### Distributed Analysis Orchestration Pipeline

Sept 2022 - Apr 2024 | Ginkgo Bioworks

- Led the development of the first company-wide analysis registry, improving analysis discoverability and tracking. Reduced onboarding time for new analyses by 50% and mentored an intern through its successful delivery.
- Optimized resource allocation and job scheduling in cloud infrastructure (AWS Batch) by implementing auto-scaling, cutting costs by 30% with no performance impact.
- Enhanced distributed computing in cloud infrastructure for an in-house biosecurity NGS pipeline, increasing throughput by 1000% and reducing costs by 50%.

### Scientific Research Software

Aug 2020 - Jun 2022 | MIT

- Developed Spateo 🌀, a Python package for high-resolution spatial transcriptomics analysis, pioneering mRNA-based cell segmentation.
- Developed Dynast 🌀, an advanced preprocessing pipeline for biological sequencing, delivering more comprehensive results at a fraction of the computational cost of existing methods.
- Enhanced Cassiopeia 🌀, a single-cell lineage reconstruction tool, reducing runtime by 30% and implementing novel graph reconstruction algorithms.