

GOVINDA KAMATH

Research Scientist / ML Engineer — Computational Biology & Life Sciences

Pleasanton, CA | 510-260-1933 | govinda.kamath@gmail.com | govinda-kamath.github.io | [Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

Applied ML researcher/engineer with 5+ years turning algorithms into production systems used by biologists and engineers, spanning fundamental academic research (PhD, Stanford) and biotech industry practice (10x Genomics). I work in tight collaboration with experimental biologists to design analyses and build pipelines they can depend on — across single-cell, spatial transcriptomics, and imaging platforms — with a career-long focus on making biological inference rigorous: avoiding false discoveries and building methods scientists can actually trust.

EXPERIENCE

Senior Scientist — 10x Genomics

Jul 2021 – Present | Pleasanton, CA

- **Visium HD spatial transcriptomics:** worked in a tight loop with biologists on assay/protocol development, designing and optimizing metrics for sensitivity, specificity, and false-positive detection; algorithms I wrote to extract spatial signal from data — effectively a 30,000-sparse-channel image — ship in Visium HD, the industry-leading spatial transcriptomics platform, and run in production for every Visium HD customer.
- **Production ML pipelines:** led end-to-end development of ML models for cell-type annotation and nucleus segmentation, from data collection and labeling through training, evaluation, and production deployment; worked across all three 10x Genomics platforms — Visium, Xenium, and Chromium (single-cell).
- **Engineering infrastructure:** own production-quality pipelines end-to-end within 10x Genomics' large shared code-base (primarily Python and Rust, also Go and C/C++), shipped in released software and used daily by scientists and engineers; set up reproducible builds with Bazel.
- **Agentic AI for single-cell workflows:** building a tool-use agentic environment that automates single-cell analysis workflows, using eval-driven development — designed evaluation suites to measure agent performance, correctness, and failure modes, and used them to iteratively drive agent and model improvements, with attention to reliability and responsible use.
- Led a team of 3 ML scientists/engineers building ML algorithms from data collection through production.

Post-doctoral Researcher — Microsoft Research

Aug 2019 – Jul 2021 | Cambridge, MA

- Research on knowledge distillation, bandit-based rank-1 models, and connections between sequence alignment and compression; published at ICLR and NeurIPS.

EDUCATION

Ph.D., Electrical Engineering — Stanford University

May 2019 | Stanford, CA

Advisor: David Tse. Thesis: *Almost Linear Time Algorithms for Problems from Computational Genomics*. GPA: 4.0/4.0

TECHNICAL SKILLS

- **Languages & tooling:** Python, RustLang, GoLang, C/C++, Bazel, MATLAB, Shell, R
- **ML & agents:** production model training/deployment, eval-suite design, agentic workflow automation, knowledge distillation, deep learning
- **Domain:** single-cell and spatial transcriptomics, genome assembly, sequence alignment, computational genomics
- **Bioinformatics databases & tooling:** extensive use of TCGA and CZ CELLxGENE; NCBI databases and tools including BLAST
- **Publications:** peer-reviewed research in ML for biology — ICLR, NeurIPS, RECOMB, ICML, AISTATS, Genome Research; full list on [Google Scholar](#)

SELECTED PHD RESEARCH & PUBLICATIONS

- **Bacterial genome assembly:** graph algorithm to resolve all resolvable repeats from long reads; implemented an assembler in C++/Python used by biologists. ([Genome Research 2017](#))
- **Haplotype assembly via stochastic block models:** spectral algorithm proven theoretically optimal. ([ICML 2016](#))
- **Near-linear-time nearest neighbors / k-means:** framework computing nearest neighbors in $O(n \log d)$ and Lloyd's k-means iterations in $O(nk \log d)$. ([AISTATS 2018](#))
- **Low-rank models for sequence alignment:** spectral estimators for seed-and-extend alignment algorithms. ([NeurIPS 2020](#); [RECOMB 2020](#) / [Cell Patterns 2020](#))
- **Valid post-clustering differential analysis:** statistically valid inference procedure for differential expression testing after clustering in single-cell RNA-seq. ([RECOMB 2019](#) / [Cell Systems 2019](#))
- **Knowledge distillation & semiparametric inference:** cast knowledge distillation as a semiparametric inference problem, deriving guarantees for the prediction error of standard distillation. ([ICLR 2021](#))