

Mohsen Zakeri

mohsenzakeri.com

mohs.zakeri@gmail.com

<https://github.com/mohsenzakeri>

CURRENT POSITION

Postdoctoral Fellow, Computer Science Department

Johns Hopkins University, Baltimore, MD

Advisor: Ben Langmead

2022 - present

EDUCATION

PhD | *Computer Science, Advisor: Prof. Rob Patro*

University of Maryland

Dec. 2021

College Park, MD

MS | *Computer Science, Advisor: Prof. Rob Patro*

Stony Brook University

May. 2017

Stony Brook, NY

BS | *Computer Engineering, Major: Software Engineering*

University of Tehran

Jun. 2015

Tehran, Iran

RESEARCH EXPERIENCE

Scalable full-text pangenome indexes

Postdoc fellow at Johns Hopkins University

2022 - present

- Designed and developed [Movi](#), a data structure based on the Move-Structure for indexing pangenomes. Movi is a full-text index which scales very well for highly similar references while being very fast to query. | *C++*

Oxford Nanopore reads classification

Postdoc fellow at Johns Hopkins University

2022 - present

- developed Movi Color for multi-class and taxonomic classification of metagenomic samples.
- Developed a method to compute exact matching queries (pseudo-matching lengths) with [Movi](#) for real-time classification of Oxford Nanopore reads and fast host depletion with pangenomes. | *C++*
- Contributed to the design of [Sigmoni](#) for multi-class classification of Nanopore signals using a compressed full-text index (r-index).

Efficient and accurate methods for single-cell RNA-seq preprocessing

Research assistant at University of Maryland

2020 - 2021

- Contributed to the design and development of a variation of pseudoalignment with structural constraints (sketch mode) for mapping single-cell RNA-seq reads, used in [Alevin-fry](#). | *C++ and Rust*

A fast and accurate aligner for short reads

Research assistant at Stony Brook University and University of Maryland

2019 - 2020

- Collaborated on the design and development of [Puffaligner](#), an aligner based on Pufferfish (a ccdbg index) that aligns various types of short reads to a collection of references. | *C++*

A pipeline for abundance estimation of metagenomics sequencing reads

Research assistant at Stony Brook University

2018 - 2019

- Collaborated on the design and development of [Cedar](#), a fast and accurate method for quantifying metagenomics sequencing reads. Cedar focuses on eliminating spurious reads in an iterative manner during the EM optimization. | *C++*

Improving the accuracy of fast RNA-seq quantification methods

Research assistant at Stony Brook University

2016 - 2019

- Enhanced the probabilistic model of bulk RNA-seq in [Salmon](#) to improve the fidelity of the equivalence class-based [factorization](#) models for RNA-seq data. This new model improved the accuracy of abundance estimations, particularly in challenging cases, i.e., paralogous genes. | *C++*
- Collaborated on the design and development of [Selective-alignment](#) to enhance the specificity and sensitivity of lightweight alignment methods used in fast RNA-seq quantification approaches. | *C++*
- Improved the uncertainty estimation of missing data in RNA-seq quantification by using augmented-data bootstrap. This approach is specifically useful in the presence of allelic imbalance.

*: co-first authorship

- Majidian, S., Hwang, S., **Mohsen Zakeri**, & Langmead, B. (2025). Evani benchmarking workflow for evolutionary distance estimation. *Briefings in Bioinformatics*, 26(3), bba267
- Guccione, C., Patel, L., Tomofuji, Y., McDonald, D., Gonzalez, A., Sepich-Poore, G. D., Sonehara, K., **Mohsen Zakeri**, Chen, Y., Dilmore, A. H., et al. (2025). Incomplete human reference genomes can drive false sex biases and expose patient-identifying information in metagenomic data. *Nature Communications*, 16(1), 825
- **Mohsen Zakeri**, Brown, N. K., Ahmed, O. Y., Gage, T., & Langmead, B. (2024). Movi: A fast and cache-efficient full-text pangenome index. *IScience*, 27(12)
- Shivakumar, V. S., Ahmed, O. Y., Kovaka, S., **Mohsen Zakeri**, & Langmead, B. (2024). Sigmoni: Classification of nanopore signal with a compressed pangenome index. *Bioinformatics (ISMB Proceedings)*, 40(Supplement_1), i287–i296
- Wu, E. Y., Singh, N. P., Choi, K., **Mohsen Zakeri**, Vincent, M., Churchill, G. A., Ackert-Bicknell, C. L., Patro, R., & Love, M. I. (2023). Seesaw: Detecting isoform-level allelic imbalance accounting for inferential uncertainty. *Genome biology*, 24(1), 165
- He, D., **Mohsen Zakeri**, Sarkar, H., Soneson, C., Srivastava, A., & Patro, R. (2022). Alevin-fry unlocks rapid, accurate and memory-frugal quantification of single-cell rna-seq data. *Nature Methods*, 19(3), 316–322
- Skoufos, G., Almodaresi, F., **Mohsen Zakeri**, Paulson, J. N., Patro, R., Hatzigeorgiou, A. G., & Vlachos, I. S. (2022). Agamemnon: An accurate metagenomics and metatranscriptomics quantification analysis suite. *Genome biology*, 23(1), 1–27
- Almodaresi, F., **Mohsen Zakeri***, & Patro, R. (2021). Puffaligner: A fast, efficient, and accurate aligner based on the pufferfish index. *Bioinformatics*
- Srivastava, A., Malik, L., Sarkar, H., **Mohsen Zakeri**, Almodaresi, F., Soneson, C., Love, M. I., Kingsford, C., & Patro, R. (2020). Alignment and mapping methodology influence transcript abundance estimation. *Genome biology*, 21(1), 1–29
- Sarkar, H., **Mohsen Zakeri***, Malik, L., & Patro, R. (2018). Towards selective-alignment: Bridging the accuracy gap between alignment-based and alignment-free transcript quantification, 27–36
- **Mohsen Zakeri**, Srivastava, A., Almodaresi, F., & Patro, R. (2017a). Improved data-driven likelihood factorizations for transcript abundance estimation. *Bioinformatics (ISMB Proceedings)*, 33(14), i142–i151

PRE-PRINTS

- Brown, N. K., Depuydt, L., **Mohsen Zakeri**, Alhadi, A., Allam, N., Begleiter, D., Karpagavalli, N. B. K., Khajjayam, S. S., Wahed, H., Gage, T., & Langmead, B. (2025). Kebab: k -mer based breaking for finding super-maximal exact matches. *arXiv preprint arXiv:2502.20338*
- Tan, S., Majidian, S., Langmead, B., & **Mohsen Zakeri**. (2025). Movi color: Fast and accurate long-read classification with the move structure. *bioRxiv*, 2025–05
- Srivastava, A., **Mohsen Zakeri**, Sarkar, H., Soneson, C., Kingsford, C., & Patro, R. (2021). Accounting for fragments of unexpected origin improves transcript quantification in rna-seq simulations focused on increased realism. *bioRxiv*
- **Mohsen Zakeri**, Srivastava, A., Sarkar, H., & Patro, R. (2021). A like-for-like comparison of lightweight-mapping pipelines for single-cell rna-seq data pre-processing. *bioRxiv*

CONFERENCES AND PRESENTATIONS

- **Movi Color: fast and accurate long-read classification with the move structure (poster).** In: WABI Conference on Algorithms in Bioinformatics, College Park, MD, 2025
- **Sequitur: A space-efficient and versatile method for k-mer queries on pangenomes (poster).** In: Biological Data Science, Cold Spring Harbor, NY, 2024
- **Movi: a fast and cache-efficient full-text pangenome index (talk).** In: RECOMB-seq, Cambridge, MA, 2024
- **Real-time nanopore adaptive sampling with Movi (talk).** In: Genome Informatics, Cold Spring Harbor, NY, 2023
- **Real-time nanopore adaptive sampling with move data structure (talk).** In: Workshop on Emerging Methods for Sequencing Analysis, State College, PA, 2023
- **Augmented-data bootstrap reveals missing uncertainty during RNA-seq quantification (poster).** In: RECOMB, San Diego, CA, 2022
- **Cedar: scalable, accurate and fast metagenomic abundance estimation (poster).** In: RECOMB, Washington D.C., 2019

PROFESSIONAL EXPERIENCE

Postdoctoral Fellow Johns Hopkins University - Langmead lab	2022 - present Baltimore, MD
• Involved in projects focused on design and development of efficient data structures for pangenomes and Oxford Nanopore read classification. <i>C++</i>	
Computational Biologist Intern Ellumigen (formerly: Ocean Genomics)	Summer 2021 Pittsburgh, PA
• Developed Mudskipper , a software tool that converts alignments in SAM files - commonly used for storing sequence alignments - from genomic coordinates to transcriptomic coordinates. <i>Rust</i>	
Software Engineer Green Silver Leaves Corp.	2017 - 2018 Tehran, Iran
• Resaa is an online platform for doctor-patient communications, allowing patients to call their doctors using their available credit. I collaborated in developing various features for the administration website and the accounting policy system. <i>C#</i>	

PEER-REVIEW EXPERIENCE

PLOS Computational Biology, PeerJ, PeerJ Life & Environment Reviewer
RECOMB 2019 and 2025, ISMB 2019 and 2020, ACM-BCB 2019, APBC 2017 Sub-Reviewer

EXTERNAL RESEARCH SUPPORT

I contributed to the writing of the grant:

Efficient and scalable pangenomes with the move structures NIH/NHGRI R21 grant R21HG013433 (\$377K)	2/1/2024 - 2/28/2026
---	----------------------

TEACHING AND MENTORING

Mentor for Undergraduate Research Project Johns Hopkins University, Baltimore, MD Student: Steven Tan	2024–2025
Instructor of a HEART course Johns Hopkins University, Baltimore, MD Needle in a haystack: Finding the origin of shredded sequences	Fall 2023
Co-Instructor of a HEART course Johns Hopkins University, Baltimore, MD Software Engineering in Bio-medical Research	Fall 2022

Teaching assistant of undergraduate courses Stony Brook University, Stony Brook, NY Social Networks, Discrete Mathematics, Computer Programming III	2015 - 2017
Teaching assistant of a graduate course Stony Brook University, Stony Brook, NY Analysis of Algorithms	Fall 2016
Teaching assistant of undergraduate courses University of Tehran, Tehran, Iran Advanced Programming, Artificial Intelligence	2014 - 2015

COMMUNITY INVOLVEMENT

Conference Session Chair WABI Conference on Algorithms in Bioinformatics at University of Maryland <ul style="list-style-type: none"> Moderated presentations, leading discussion and Q&A 	August 2025
Conference Scientific Committee Johns Hopkins Annual Postdoc Conference <ul style="list-style-type: none"> Provided volunteer support for postdoc conference organization and logistics 	May 2025
Hackathon Judge Hophacks , Johns Hopkins University <ul style="list-style-type: none"> Volunteered as a judge at an annual health-focused hackathon, where participants developed and presented applications over three days in a science fair format. Assessed projects on creativity, usefulness, polish, and technical difficulty, contributing to the development and recognition of innovative health technologies. 	2023 and 2024