

# Mohsen Zakeri

[mohsenzakeri.com](http://mohsenzakeri.com)

[mohs.zakeri@gmail.com](mailto: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 AND PROFESSIONAL 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 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.

## PEER REVIEWED PUBLICATIONS

---

\*: co-first authorship

- [1] Caitlin Guccione, Lucas Patel, Yoshihiko Tomofuji, Daniel McDonald, Antonio Gonzalez, Gregory D Sepich-Poore, Kyuto Sonehara, **Mohsen Zakeri**, Yang Chen, Amanda Hazel Dilmore, et al. “Incomplete human reference genomes can drive false sex biases and expose patient-identifying information in metagenomic data”. In: *Nature Communications* 16.1 (2025), p. 825.
- [2] **Mohsen Zakeri**, Nathaniel K Brown, Omar Y Ahmed, Travis Gagie, and Ben Langmead. “Movi: a fast and cache-efficient full-text pangenome index”. In: *iScience* 27.12 (2024).
- [3] Vikram S Shivakumar, Omar Y Ahmed, Sam Kovaka, **Mohsen Zakeri**, and Ben Langmead. “Sigmoni: classification of nanopore signal with a compressed pangenome index”. In: *Bioinformatics (ISMB Proceedings)* 40.Supplement\_1 (2024), pp. i287–i296.
- [4] Euphy Y Wu, Noor P Singh, Kwangbom Choi, **Mohsen Zakeri**, Matthew Vincent, Gary A Churchill, Cheryl L Ackert-Bicknell, Rob Patro, and Michael I Love. “SEESAW: detecting isoform-level allelic imbalance accounting for inferential uncertainty”. In: *Genome biology* 24.1 (2023), p. 165.
- [5] Dongze He, **Mohsen Zakeri**, Hirak Sarkar, Charlotte Soneson, Avi Srivastava, and Rob Patro. “Alevin-fry unlocks rapid, accurate and memory-frugal quantification of single-cell RNA-seq data”. In: *Nature Methods* 19.3 (2022), pp. 316–322.
- [6] Giorgos Skoufos, Fatemeh Almodaresi, **Mohsen Zakeri**, Joseph N Paulson, Rob Patro, Artemis G Hatzigeorgiou, and Ioannis S Vlachos. “AGAMEMNON: an Accurate metaGenomics And METatranscriptoMics quaNtificatiON analysis suite”. In: *Genome biology* 23.1 (2022), pp. 1–27.
- [7] Fatemeh Almodaresi, **Mohsen Zakeri\***, and Rob Patro. “Puffaligner: A Fast, Efficient, and Accurate Aligner Based on the Pufferfish Index”. In: *Bioinformatics* (2021).
- [8] Avi Srivastava, Laraib Malik, Hirak Sarkar, **Mohsen Zakeri**, Fatemeh Almodaresi, Charlotte Soneson, Michael I Love, Carl Kingsford, and Rob Patro. “Alignment and mapping methodology influence transcript abundance estimation”. In: *Genome biology* 21.1 (2020), pp. 1–29.
- [9] Hirak Sarkar, **Mohsen Zakeri\***, Laraib Malik, and Rob Patro. “Towards selective-alignment: Bridging the accuracy gap between alignment-based and alignment-free transcript quantification”. In: (2018), pp. 27–36.
- [10] **Mohsen Zakeri**, Avi Srivastava, Fatemeh Almodaresi, and Rob Patro. “Improved data-driven likelihood factorizations for transcript abundance estimation”. In: *Bioinformatics (ISMB Proceedings)* 33.14 (2017), pp. i142–i151.

## PRE-PRINTS

---

- [1] Sina Majidian, Stephen Hwang, **Mohsen Zakeri**, and Ben Langmead. “EvANI benchmarking workflow for evolutionary distance estimation”. In: *bioRxiv* (2025), pp. 2025–02.
- [2] Lore Depuydt, Omar Ahmed, Andrej Baláž, Nathaniel K Brown, Adrián Goga, Alessia Petescia, **Mohsen Zakeri**, Jan Fostier, Travis Gagie, Ben Langmead, et al. “r-indexing without backward searching”. In: *arXiv e-prints* (2023), arXiv–2312.
- [3] Avi Srivastava, **Mohsen Zakeri**, Hirak Sarkar, Charlotte Soneson, Carl Kingsford, and Rob Patro. “Accounting for Fragments of unexpected origin improves transcript quantification in RNA-seq simulations focused on increased realism”. In: *bioRxiv* (2021).
- [4] **Mohsen Zakeri**, Avi Srivastava, Hirak Sarkar, and Rob Patro. “A Like-for-Like Comparison of Lightweight-Mapping Pipelines for Single-Cell RNA-seq Data Pre-Processing”. In: *bioRxiv* (2021).

## CONFERENCES AND PRESENTATIONS

---

- [1] **Sequitur: A space-efficient and versatile method for k-mer queries on pangenomes (poster)**. In: *Biological Data Science, Cold Spring Harbor, NY*. 2024.
- [2] **Movi: a fast and cache-efficient full-text pangenome index (talk)**. In: *Recomb-seq, Cambridge, MA*. 2024.
- [3] **Real-time nanopore adaptive sampling with Movi (talk)**. In: *Genome Informatics, Cold Spring Harbor, NY*. 2023.
- [4] **Real-time nanopore adaptive sampling with move data structure (talk)**. In: *Workshop on Emerging Methods for Sequencing Analysis, State College, PA*. 2023.

- [5] **Augmented-data bootstrap reveals missing uncertainty during RNA-seq quantification (poster).** In: *RECOMB, San Diego*. 2022.
- [6] **Cedar: scalable, accurate and fast metagenomic abundance estimation (poster).** In: *RECOMB, Washington D.C.* 2019.

## PEER-REVIEW EXPERIENCE

---

### **PLOS Computational Biology**

Reviewer

### **RECOMB 2019 and 2025**

Sub-Reviewer

### **ISMB 2019 and 2020**

Sub-Reviewer

### **ACM-BCB 2019**

Sub-Reviewer

### **APBC 2017**

Sub-Reviewer

## EXTERNAL RESEARCH SUPPORT

---

I contributed to the writing of the grant:

### **Efficient and scalable pangenomes with the move structure**

NIH/NHGRI R21 grant R21HG013433 (\$377K)

2/1/2024 - 2/28/2026

## TEACHING EXPERIENCE

---

### **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

---

### **Hackathon Judge**

[Hophacks](#), Johns Hopkins University

2023 and 2024

- 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.

## REFERENCES

---

**Ben Langmead**  
langmea@cs.jhu.edu

Associate Professor, Department of Computer Science, Johns Hopkins University

**Rob Patro**  
rob@cs.umd.edu

Associate Professor, Department of Computer Science, University of Maryland

**Travis Gagie**  
travis.gagie@gmail.com

Associate Professor, Faculty of Computer Science, Dalhousie University

**Michael I. Love**  
milove@email.unc.edu

Associate Professor, Department of Biostatistics, University of North Carolina