Mohsen Zakeri

mohsenzakeri.com
mohs.zakeri@gmail.com
https://github.com/mohsenzakeri

EDUCATION

PhD | *Computer Science, Advisor: Prof. Rob Patro*
University of Maryland
Dec. 2021
College Park, MD

MSc | *Computer Science, Advisor: Prof. Rob Patro*
Stony Brook University
May. 2017
Stony Brook, NY

BSc | *Computer Engineering, Major: Software Engineering*
University of Tehran
Jun. 2015
Tehran, Iran

WORK EXPERIENCE

Postdoctoral Fellow
Johns Hopkins University - Langmead lab
2022 - present
Baltimore, MD

- Design and developtment of Efficient data structures for Oxford nanopore read classification and metagenomic quantification. | *C++*

Computational Biologist
Ocean Genomics
2021
Pittsburgh, PA

- I developed Mudskipper which is a software for converting alignments in a SAM file, which is used for storing sequence alignments, with genomic coordinates to transcriptomic coordinated alignments. | *Rust*
- Puffaligner is a fast and efficient aligner for short reads to a collection of reference sequences. I collaborated for debugging steps of developing the softclip feature for this aligner. | *C++*

Software Engineer
Green Silver Leaves Corp.
2017 - 2018
Tehran, Iran

- Resaa is an online system for doctor-patient communications where patients could charge their account to call their doctors. I collaborated for developing various features for the adminstration website and the accounting policy system. | *C#*

PROJECTS AND RESEARCH

Efficient and scalable indexing of pangenomes
Johns Hopkins University
2022 - present

I have implemented Movi which is a data structure based on Move-Structure for indexing pangenomes. Movi computes exact matching queries (pseudo-matching lengths) for short and long reads. The Pseudo-matching lengths computed by Movi is used for real-time classification of Nanopore reads. | *C++*

Efficient and accurate methods for single-cell RNA-seq analysis
University of Maryland
2020 - 2021

I contributed to the design and development of a variation of pseudoalignment with structural constraints (sketch mode) for mapping single-cell RNA-seq reads which is used with Alevin-fry. | *C++ and Rust*

A fast and accurate aligner for short reads
Stony Brook University, University of Maryland
2019 - 2020

Puffaligner is an aligner based on the Pufferfish index which aligns different types of short reads including RNA-seq, DNA-seq and metagenomic reads to a collection of indices. I was involved in the main stages of the development and debugging for this tool. | *C++*

A pipeline for metagenomic abundance estimation
Stony Brook University
2018

I contributed to the design and development of Cedar which is a fast and accurate quantification method for short read metagenomic reads using lightweight alignment methods and optimization with Expected Maximization. | *C++*

Improving the accuracy of lightweight RNA-seq mapping methods
Stony Brook University
2017 - 2019

I contributed to the development of the idea of Selective-alignment for increasing the specificity and sensitivity of lightweight alignment methods which are used in fast RNA-seq quantification approaches. | *C++*

Improving the accuracy of fast RNA-seq quanfitication methods   2016 - 2022
Stony Brook University

I modified the probablistic model of bulk RNA-seq in Salmon to increase the fidelity of the equivalence class-based factorization which led to an improvement of the accuracy of the estimations in challenging cases. | *C++*

Teaching Experience

Needle in a haystack: Finding the origin of shredded sequences (Instructor)   Fall 2023
Johns Hopkins University   Balitmore, MD

Software Engineering in Bio-medical Research (Co-Instructor)   Fall 2022
Johns Hopkins University   Balitmore, MD

Social Networks, Discrete Mathematics, Computer Programming III (TA)   2015 − 2017
Stony Brook University   Stony Brook, NY

Analysis of Algorithms for graduate students (TA)   2016
Stony Brook University   Stony Brook, NY

Advanced Programming, Artificial Intelligence (TA)   Fall 2014 − Spring 2015
University of Tehran   Tehran, Iran

Peer Reviewed Publications

*: co-first authorship

[1] 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* 40.Supplement_1 (2024), pp. i287–i296.

[2] 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.

[3] 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.

[4] 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.

[5] Fatemeh Almodaresi, **Mohsen Zakeri***, and Rob Patro. "Puffaligner: A Fast, Efficient, and Accurate Aligner Based on the Pufferfish Index". In: *Bioinformatics* (2021).

[6] 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.

[7] 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.

[8] **Mohsen Zakeri**, Avi Srivastava, Fatemeh Almodaresi, and Rob Patro. "Improved data-driven likelihood factorizations for transcript abundance estimation". In: *Bioinformatics* 33.14 (2017), pp. i142–i151.

Pre-prints

[1] 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.

[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: *bioRxiv* (2023).

[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] **Mohsen Zakeri**, Nathaniel K Brown, Omar Y Ahmed, Travis Gagie, and Ben Langmead. "Movi: a fast and cache-efficient full-text pangenome index". In: *Recomb-seq, Cambridge, MA*. 2024.

[2] **Mohsen Zakeri**, Nathaniel Brown, Omar Ahmed, Travis Gagie, and Ben Langmead. "Real-time nanopore adaptive sampling with Movi". In: *Genome Informatics, Cold Spring Harbor, NY*. 2023.

[3] Vikram Shivakumar, Omar Ahmed, Sam Kavoka, **Mohsen Zakeri**, and Ben Langmead. "Sigmoni: Efficient Pangenome Multi-Classification of Nanopore Signal". In: *RECOMB-seq, Istanbul*. 2023.

[4] **Mohsen Zakeri**, Zahra Zare Jousheghani, Euphy Wu, Noor Pratap Singh, Ben Langmead, Mike Love, and Rob Patro. "Augmented-data bootstrap reveals missing uncertainty during RNA-seq quantification". In: *RECOMB, San Diego*. 2022.

[5] Hirak Sarkar, Avi Srivastava, **Mohsen Zakeri**, Scott Van Buren, Naim U Rashid, Michael Love, and Rob Patro. "Accurate, efficient, and uncertainty-aware expression quantification of single-cell RNA-seq data". In: 2020. DOI: 10.6084/m9.figshare.13198100.v1.

[6] Fatemeh Almodaresi, **Mohsen Zakeri**, and Rob Patro. "Cedar: scalable, accurate and fast metagenomic abundance estimation". In: *RECOMB, Washington D.C.* 2019.

[7] **Mohsen Zakeri**, Avi Srivastava, Fatemeh Almodaresi, and Rob Patro. "Improved data-driven likelihood factorizations for transcript abundance estimation". In: *ISMB, Prague*. 2017.

Honors and Awards

Best poster award at RECOMB-seq I was a collaborator on "Sigmoni: Efficient Pangenome Multi-Classification of Nanopore Signal"	2023
Travel Fellowship for attending ISMB/ECCB My paper was accepted in the proceedings section of the conference	2017
Special CS Chair Fellowship from Sony Brook University Awarded upon admission for the PhD program	2015
FOE award from the Engineering Department of University of Tehran For ranking amongst top three computer engineering students based on cumulative GPA	2012 and 2014

External Research Support

I collaborated in the writing of the grant with my PI (Ben Langmead)

Efficient and scalable pangenomes with the move structure NIH/NHGRI R21 grant R21HG013433	2/1/2024 - 2/28/2026