

# Hirak Sarkar

## Education

- 2020 **Ph.D**, Department of Computer Science, University of Maryland, Maryland, USA.  
Advisor: Prof. Rob Patro, Thesis: Efficient Processing and Statistical Inference in RNA-seq data
- 2014-  
transferred **MS**, Department of Computer Science, Stony Brook University, New York, USA.  
Advisor: Prof. Rob Patro, Thesis: Improving storage and alignment methodologies for RNA-seq data
- 2011-2013 **M.Tech**, Indian Statistical Institute, Calcutta, India.  
Research Area: Theoretical Computer Science
- 2007-2011 **B.Tech**, West Bengal University of Technology, Calcutta, India.  
Major: Computer Science and Engineering

## Experience

- CBCB, UMIACS, University of Maryland, College Park, Maryland, USA.**
- 2019-2020 *Research Assistant*  
**Facebook Inc., Menlo Park, San Francisco, California, USA.**
- 2018(Summer) *Ph.D Data Scientist Intern*  
**Stony Brook University, Stony Brook, NY, USA.**
- 2016 - 2018 *Research Assistant*
- 2014 - 2016 *Teaching Assistant*  
**Cold Spring Harbor Lab, Cold Spring Harbor, NY, USA.**
- 2016(Summer) *Research Intern, Siepel Lab*

## Publications

### Bioinformatics

- 2020 “Terminus enables the discovery of data-driven, robust transcript groups from RNA-seq data”, by Hirak Sarkar, Avi Srivastava, Hector Corrada Bravo, Michael I. Love and Rob Patro. [ISMB’ 20]
- 2020 “A Bayesian framework for inter-cellular information sharing improves dscRNA-seq quantification”, by Avi Srivastava, Laraib Malik, Hirak Sarkar, Rob Patro. [ISMB’ 20]
- 2020 “Alignment and mapping methodology influence transcript abundance estimation”, by Avi Srivastava, Laraib Malik, Hirak Sarkar, Mohsen Zakeri, Charlotte Soneson, Michael I. Love, Carl Kingsford, Rob Patro. [Genome Biology]
- 2019 “Minnow: A principled framework for rapid simulation of dscRNA-seq data at the read level”, by Hirak Sarkar, Avi Srivastava and Rob Patro [ISMB’19]
- 2018 “Towards Selective-Alignment: Producing Accurate And Sensitive Alignments Using Quasi-Mapping”, by Hirak Sarkar\*, Mohsen Zakeri\*, Laraib Malik and Rob Patro [ACM BCB’ 18]
- 2018 “A space and time-efficient index for the compacted colored de Bruijn graph”, by Fatemeh Almodaresi\*, Hirak Sarkar\*, Avi Srivastava and Rob Patro [ISMB’ 18]

- 2017 “*Quark enables semi-reference-based compression of RNA-seq data*” by Hirak Sarkar and Rob Patro [*Oxford Bioinformatics’ 17*]
- 2016 “*Fast, Lightweight Clustering of de novo Transcriptomes using Fragment Equivalence Classes*” by A Srivastava\*, Hirak Sarkar\*, Laraib Malik and Rob Patro [*RECOMB-seq’ 16*]
- 2016 “*RapMap: A Rapid, Sensitive and Accurate Tool for Mapping RNA-seq Reads to Transcriptomes*” by A Srivastava, Hirak Sarkar, Nitish Gupta and Rob Patro [*ISMB’ 16*]

### Other

- 2020 “*Social Media Attributions in the Context of Water Crisis* by Rupak Sarkar, Hirak Sarkar, S Mahinder and AR KhudaBukhsh. [*ArXiv’20*]
- 2016 “*Voronoi Game on Graphs*” (Extended version) by S. Bandyapadhyay, A. Banik, S. Das and H. Sarkar (in alphabetical order of surnames) [*Theoretical Computer Science’15*]

### Posters

- 2017 “*Pufferfish: A fast graph-based indexing and query strategy for large genomic sequences* by Fatemeh Almodaresi\*, Hirak Sarkar\*, and Rob Patro, Poster presented in [*WABI’17*]
- 2016 “*Joint probabilistic model for multiple steps of gene regulation* by Hirak Sarkar”, Yi-Fei Huang and Adam Siepel, Poster presented in *BioData’16*

## Talks

- 2019 ISMB, Basel, Switzerland
- 2018 ACM-BCB, Washington DC, USA
- 2013 WABI, IIT-kharagpur, India

## Awards

- 2014 Special CS Chair Fellowship, Stony Brook University
- 2014 NUS Research Scholarship - discontinued
- 2011-2013 Post-graduate Scholarship by, Govt. of India.
- 2013 First Prize for Software Competition (IEM), Calcutta

## Open Source Tools Contributed to

- Terminus **Terminus enables the discovery of data-driven, robust transcript groups from RNA-seq data**, <https://github.com/COMBINE-lab/terminus>
- Minnow **Minnow: A principled framework for rapid simulation of dscRNA-seq data at the read level**, <https://github.com/COMBINE-lab/minnow>
- Pufferfish **Pufferfish: A fast graph-based indexing and query strategy for large genomic sequences**, <https://github.com/COMBINE-lab/pufferfish>
- RapMap **RapMap: A Rapid, Sensitive and Accurate Tool for Mapping RNA-seq Reads to Transcriptomes**, <https://github.com/COMBINE-lab/rapmap>
- Salmon **Towards Selective-Alignment: Producing Accurate And Sensitive Alignments Using Quasi-Mapping**, <https://github.com/COMBINE-lab/salmon>

## Grant Support

- 2018-2019 **Efficient tools for quantifying and simulating transcript-level abundance in single-cell RNA-seq**, SVCF 182752, PI: Rob Patro, Role: Research Assistant
- 2018-2020 **Approximate Membership Query Data Structures in Computational Biology and Storage**, NIH R01 HG009937, PI: Rob Patro, Role: Research Assistant

---

## Professional Activities

Reviewer Oxford Bioinformatics  
Sub-reviewer ISMB, RECOMB, WABI, ACM-BCB  
Deligated Genome Biology

---

## Skills

C++, Rust, Python

---

## Teaching Assistant

CSE219 Game Programming, 2014-2015  
CSE549 Computational Biology, 2016-2017

---

## References

Ph.D Advisor **Prof. Rob Patro**, *Assistant Professor*, Department of Computer Science, University of Maryland, MD, USA.  
rob@cs.umd.edu

**Prof. Michael Love**, *Department of Biostatistics*, Department of Genetics, University of North Carolina-Chapel Hill, NC. USA.  
michaelisaiahlove@gmail.com

**Prof. Héctor Corrada Bravo**, *Associate Professor*, Department of Computer Science, University of Maryland, MD, USA.  
hcorrada@umiacs.umd.edu