Hirak Sarkar

54 Garden View Ter, Unit 8
East Windsor, NJ-08520
☎ *(631) 520 8131*
✉ *hiraksarkar.cs@gmail.com*
🖱 *www.hiraksarkar.com*

Education

2020 **Ph.D**, *Department of Computer Science*, University of Maryland, Maryland, USA.

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

2013 **M.Tech**, *Indian Statistical Institute*, Calcutta, India.
Research Area: Theoretical Computer Science

2011 **B.Tech**, *West Bengal University of Technology*, Calcutta, India.
Computer Science

Experience

2023 - present **Princeton Ludwig Scholar**, *Princeton University*, New Jersey, USA.
Advisor: Prof. Yibin Kang, Prof. Ben Raphael

2022 - 2023 **Research Associate**, *Rutgers Cancer Institute of New Jersey*, New Jersey, USA.
Advisor: Prof. Eileen White

2021 - 2022 **Postdoctoral Fellow**, *Harvard Medical School*, Harvard University, Massachusetts, USA.
Advisor: Prof. Peter Kharchenko

2019-2020 **CBCB, UMIACS, University of Maryland**, *College Park*, Maryland, USA.
Research Assistant

2018 (Summer) **Facebook Inc.**, *Menlo Park*, San Francisco, California, USA.
Ph.D Data Scientist Intern

Stony Brook University, *Stony Brook*, NY, USA.
2016 - 2018 *Research Assistant*
2014 - 2016 *Teaching Assistant*

Publications

Computational Biology (Published / In press /)

2025 *“Joint imputation and deconvolution of gene expression across spatial transcriptomics platforms”*, by Hongyu Zheng*, Hirak Sarkar*, and Ben Raphael. [***RECOMB’25/Genome Research (under review)***]

2024 *“Comparative Single-Cell Transcriptomics of Human Neuroblastoma and Preclinical Models Reveals Conservation of an Adrenergic Cell State”*, by Bethel Tesfai Embaie*, Hirak Sarkar* et al. [***Cancer Research’24***]

2024 *“Deciphering normal and cancer stem cell 65 niches by spatial transcriptomics: opportunities and challenges*, by Hirak Sarkar, Eunmi Lee, J, Sereno Lopez-Darwin, and Yibin Kang [***Gene & Development’24***]

2024 *“A count-based model for delineating cell–cell interactions in spatial transcriptomics data”*, by Hirak Sarkar*, Uthsav Chitra*, Julian Gold, and Ben Raphael [***ISMB ’24***, *Bioinformatics’24*]

2023 *“Mapping the topography of spatial gene expression with interpretable deep learning”*, by Uthsav Chitra, Brian J Arnold, Hirak Sarkar Cong Ma, Sereno Lopez-Darwin, Kohei Sanno, Benjamin J Raphael [***Nature Methods’24***]

2023 *“Single-cell and spatial transcriptomic analysis reveals alterations that correlate with human prostate cancer progression”*, by Taghreed Hirz, Shenglin Mei, Hirak Sarkar, . . ., Peter Kharchenko, David B. Sykes [***Nature Communications’23***]

2023 *“Haplotype-enhanced inference of somatic copy number profiles from single-cell transcriptomes”*, by Teng Gao, Ruslan Soldatov, Hirak Sarkar, Adam Kurkiewicz, Evan Biederstedt, Po-Ru Loh, Peter Kharchenko [***Nature Biotech’22***]

2022 *“Airpart: Interpretable statistical models for analyzing allelic imbalance in single-cell datasets”*, by Wancen Mu, Hirak Sarkar, Avi Srivastava, Kwangbom Choi, Rob Patro, Michael I Love [***Oxford Bioinformatics’22***]

2022 *“Alevin-fry unlocks rapid, accurate, and memory-frugal quantification of single-cell RNA-seq data”*, by Dongze He, Mohsen Zakeri, Hirak Sarkar, Charlotte Soneson, Avi Srivastava, Rob Patro [***Nature Methods’22***]

2021 *“Compression of quantification uncertainty for scRNA-seq counts”*, by Scott Van Buren, Hirak Sarkar, Avi Srivastava, Naim U Rashid, Rob Patro, Michael I Love [***Oxford Bioinformatics’21***]

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 *“An Efficient, Scalable and Exact Representation of High-Dimensional Color Information Enabled via de Bruijn Graph Search”*, 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***]

Consortium Papers

2023 *“Epigenetic regulation during cancer transitions across 11 tumour types”* [***Nature ’23***]

2023 *“Best practices for single-cell analysis across modalities”* [***Nature Reviews Genetics ’23***]

Other

2020 *“ Social Media Attributions in the Context of Water Crisis* by Rupak Sarkar, Hirak Sarkar, S Mahinder and AR KhudaBukhsh. [***EMNLP’21***]

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

2019/20,2024	ISMB
2018	ACM-BCB, Washington DC, USA
2013	WABI, IIT-kharagpur, India

Awards

2024	ISCB Travel Award
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

Review Editor	Frontiers in Genetics
Reviewer	Oxford Bioinformatics, BMC Genomics, Genome Biology, ISMB, RECOMB, WABI, ACM-BCB
Program Chair	ISMB'22

Skills

C++, Rust, Python

Teaching Assistant

CSE219	Game Programming, 2014-2015
CSE549	Computational Biology, 2016-2017
CSE548	Analysis of Algorithms, Spring 2018

References

Ph.D Advisor

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

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

Dr. Hector Corrada Bravo , *Senior Scientist*, Research and Early Development, Genentech.
corrada-bravo.hector@gene.com

Prof. Peter Kharchenko , *Associate Professor*, Department of Biomedical Informatics, Harvard Medical School, MA, USA.
Peter_Kharchenko@hms.harvard.edu