

Urminder Singh, PhD

Senior Bioinformatics Scientist
Illumina
5200 Illumina way
San Diego, CA 92122, USA.

✉ <first>.<last>.21 AT gmail
📞 [urmi-21](tel:urmi-21)
🐦 [_urminder](https://twitter.com/_urminder)

RESEARCH INTERESTS

De novo gene evolution, Precision medicine, Machine learning, Sociogenomics, Bioinformatics algorithms

EDUCATION

2016 - 2021	Ph.D., Bioinformatics and Computational Biology with Minor in Statistics <i>Iowa State University, Ames, IA, United States</i> Dissertation: “Pan-tissue pan-cancer characterization of novel human orphan genes via analysis of RNA-sequencing data” Supervisor: Prof. Eve Wurtele
2013-2015	M.Tech., Computational and Systems Biology <i>School of Computational and Integrative Sciences</i> Jawaharlal Nehru University, New Delhi, India Thesis: “Computational identification of replication origin like sequences in whole genome data using entropic and machine learning methods” Supervisor: Dr. A. Krishnamachari
2010-2013	M.S. in Computer Science <i>Department of Computer Science</i> South Asian University, New Delhi, India
2007-2010	B.Sc, Applied Physical Sciences <i>S.G.T.B Khalsa College</i> University of Delhi, New Delhi, India

HONORS AND AWARDS

- Elected full member of Sigma Xi, The Scientific Research Honor Society, USA **September 2022**
- Zaffarano Prize (HM) to recognize superior performance in publishable research, Iowa State University **May 2021**
- COVID-19 Exceptional Effort Graduate Student Research Impact Award, Iowa State University **January 2021**
- Research Excellence Award, Iowa State University **Spring 2021**
- International course certificate: Computational Systems Biology of Cancer, Institut Curie, Paris, France **Fall 2019**
- Professional Development Grants, GPSS, Iowa State University **Fall 2019**
- BCB Departmental Award, BCB, Iowa State University **Fall 2019**
- GDCB Departmental Travel Award, GDCB, Iowa State University **Spring 2019**
- ISMB EvolComp COSI award, Intelligent Systems for Molecular Biology (ISMB) **June 2018**
- Sui-Tong Chan Fung Fund, GDCB, Iowa State University **Spring 2018**
- Wendell Miller Trust Graduate Fellowship, Iowa State University **2016 - 2018**
- CALS Presidential Scholarship, Iowa State University **Fall 2016**
- Junior Research Fellowship, University Grants Commission, India **2013 - 2015**
- National Eligibility Test (Computer Science) (Highly competitive and mandatory examination to become a lecturer in India), University Grants Commission, India **2012**

RESEARCH AND ACADEMIC EXPERIENCE

Senior Bioinformatics Scientist (*Illumina*), San Diego

Oct 2022 - Present

I am a senior bioinformatics scientist in the oncology RnD division at Illumina. I contribute to novel cancer biomarker research, developing bioinformatics algorithm for variant detection, and pharmaceutical collaborations for development of new disease assays.

Bioinformatics Scientist (*Illumina*), San Diego

Aug 2021 - Sept 2022

As a bioinformatics scientist in the RnD division at Illumina, I worked closely with oncology, software and assay groups to contribute to cancer biomarker research. I contributed to the validation of Illumina's TruSight Oncology 500 cancer biomarker detection pipeline. My role also involved developing novel computational tools and methods to scale and automate validation of clinical studies.

Research Assistant (*Wurtele Lab*), Iowa State University

Jan 2017 - July 2021

My research is highly inter-disciplinary and collaborative including biological, computational and statistical components. At the core of my research is to computationally identify and characterize species-specific orphan genes in the human genome. I am interested in understanding the evolutionary origins, functions and regulation of these orphan genes, particularly in diseases like cancer. This necessitates developing new statistical models and computational tools for analysis of big heterogeneous RNA-Seq datasets. To enable my research, I have developed *MetaOmGraph*, a public java tool for interactive visualization and exploratory analysis of big omics datasets and *pyrpipe*, a python package allowing easy integration of RNA-Seq tools into a reproducible pipeline. Currently, I am developing and applying deep generative learning approaches to model patterns of orphan gene expression in big heterogeneous RNA-Seq data.

Teaching Assistant (*Fundamentals of Human Anatomy*), Iowa State University

Fall 2017

Lab instructor for Biol 255L, Fundamentals of Human Anatomy at Iowa State University. Taught and demonstrated lab exercises, set up weekly and final exams, and supervised undergraduate TAs.

Research Rotations (*BCB Program*), Iowa State University

Aug 2016 - Jan 2017

I explored various research topics during my lab rotations. Dr. Walter Moss: developed a computational pipeline for prediction of novel non-coding RNAs expressed during Epstein-Barr virus infection. Dr. Karin Dorman: evaluated statistical methods to estimate errors in Next-Gen sequencing data. Dr. Eve Wurtele: worked on the challenge of orphan gene prediction in humans.

Project Assistant (*Jain Lab*), SCIS, Jawaharlal Nehru University

Nov 2015 - July 2016

Developed a novel machine learning based method (PlncPRO) for the prediction of plant long non-coding RNAs using machine learning methods.

PUBLICATIONS

RESEARCH ARTICLES

1. J. Li*, U. Singh*, P. Bhandary, J. Campbell, Z. Arendsee, A. Seetharam, and E. Wurtele. Foster thy young: enhanced prediction of orphan genes in assembled genomes. *Nucleic Acids Research*, 12 2021. ***Equal first authors**
2. J. Park, J. Foox, T. Hether, ..., U. Singh, et al. System-wide transcriptome damage and tissue identity loss in covid-19 patients. *Cell Reports Medicine*, 3(2):100522, 2022
3. J. T. McDonald, F. J. Enguita, D. Taylor, R. J. Griffin, W. Priebe, M. R. Emmett, M. M. Sajadi, A. D. Harris, J. Clement, J. M. Dybas, U. Singh, et al. Role of mir-2392 in driving sars-cov-2 infection. *Cell reports*, 37(3):109839, 2021

4. **U. Singh**, K. M. Hernandez, B. J. Aronow, and E. S. Wurtele. African americans and european americans exhibit distinct gene expression patterns across tissues and tumors associated with immunologic functions and environmental exposures. *Scientific reports*, 11(1):1–14, 2021
5. **U. Singh**, J. Li, A. Seetharam, and E. S. Wurtele. pyrpipe: a python package for rna-seq workflows. *NAR genomics and bioinformatics*, 3(2):lqab049, 2021
6. **U. Singh** and E. S. Wurtele. orfipy: a fast and flexible tool for extracting ORFs. *Bioinformatics*, 02 2021. btab090
7. J. Li, **U. Singh**, Z. Arendsee, and E. S. Wurtele. Landscape of the dark transcriptome revealed through re-mining massive rna-seq data. *Frontiers in Genetics*, page 1495, 2021
8. **U. Singh** and E. S. Wurtele. Genetic novelty: How new genes are born. *eLife*, 9:e55136, 2020
9. **U. Singh**, M. Hur, K. Dorman, and E. S. Wurtele. MetaOmGraph: a workbench for interactive exploratory data analysis of large expression datasets. *Nucleic Acids Research*, 48(4):e23–e23, 01 2020
10. **U. Singh**, K. Shah, S. Dhar, V. Singh, and A. Krishnamachari. Oris: An interactive software tool for prediction of replication origin in prokaryotic genomes. *Journal of Open Source Software*, 4:1589, 2019
11. Z. Arendsee, J. Li, **U. Singh**, P. Bhandary, A. Seetharam, and E. S. Wurtele. fagin: syntenic-based phylostratigraphy and finer classification of young genes. *BMC bioinformatics*, 20(1):1–14, 2019
12. A. Neogy, T. Garg, A. Kumar, A. K. Dwivedi, H. Singh, **U. Singh**, Z. Singh, K. Prasad, M. Jain, and S. R. Yadav. Genome-wide transcript profiling reveals an auxin-responsive transcription factor, osap2/erf-40, promoting rice adventitious root development. *Plant and Cell Physiology*, 60(10):2343–2355, 2019
13. Z. Arendsee, J. Li, **U. Singh**, A. Seetharam, K. Dorman, and E. S. Wurtele. phylostratr: A framework for phylostratigraphy. *Bioinformatics*, 35(19):3617–3627, 2019
14. **U. Singh**, N. Khemka, M. S. Rajkumar, R. Garg, and M. Jain. Plncpro for prediction of long non-coding rnas (lncrnas) in plants and its application for discovery of abiotic stress-responsive lncrnas in rice and chickpea. *Nucleic acids research*, 45(22):e183–e183, 2017
15. **U. Singh**, S. Chauhan, A. Krishnamachari, and L. Vig. Ensemble of deep long short term memory networks for labelling origin of replication sequences. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–7. IEEE, 2015

BOOK CHAPTERS

1. N. K. Khemka, **U. Singh**, A. K. Dwivedi, and M. Jain. Machine learning-based annotation of long non-coding rnas using plncpro. In *Legume Genomics*, pages 253–260. Humana, New York, NY, 2020

PRE-PRINTS

1. **U. Singh** and E. S. Wurtele. A pan-tissue pan-cancer compendium of human orphan genes. *Pan-tissue pan-cancer characterization of novel human orphan genes via analysis of RNA-Sequencing data*, page 121, 2021
2. Z. Arendsee, A. Wilkey, **U. Singh**, J. Li, M. Hur, and E. Wurtele. synder: inferring genomic orthologs from syntenic maps. *BioRxiv*, page 554501, 2019

THESIS AND DISSERTATIONS

1. **U. Singh**. *Pan-Tissue Pan-Cancer Characterization of Novel Human Orphan Genes via Analysis of RNA-Sequencing Data*. PhD thesis, Iowa State University, 2021
2. **U. Singh**. Computational identification of replication origin like sequences in whole genome data using entropic and machine learning methods. Master's thesis, Jawaharlal Nehru University, 2015
3. **U. Singh**. Java based software tool for finding origin of replication in genomic sequence data. Master's thesis, South Asian University, 2013

PROFESSIONAL ACTIVITIES

GRANTS

2019	Co-PI	<i>The prediction of plant & animal orphan genes with comparative genomics & data mining</i> Agency: Extreme Science and Engineering Discovery Environment (XSEDE) Status: Awarded (\$15,284) Award ID: TG-MCB190098
2017	Co-PI	<i>The prediction of plant & animal orphan genes with comparative genomics & data mining (start-up allocation)</i> Agency: Extreme Science and Engineering Discovery Environment (XSEDE) Status: Awarded (\$3,376) Award ID: TG-BIO170064

SYNERGISTIC ACTIVITIES

- **Journal editor:** Guest topic editor for a special issue in *Frontiers in Genetics* journal
- **Peer reviewing:** Peer reviewed for *BMC Genomics*, *Scientific Reports*, *Bioinformatics Advances*, *Frontiers in Molecular Biosciences*, and *Frontiers in Earth Science*
- **Book review:** Technical reviewer for the book titled “Bioinformatics with python cookbook”, Packt publishing

INVITED LECTURES AND WORKSHOPS

1. *Exploring career paths in biological sciences* **29 April, 2022**
Invited speaker and panelist for a discussion on career advancement with postdocs and graduate students. **Iowa State University, Iowa, USA.**
2. *MetaOmGraph for Interactive Exploratory Omics Data analysis* **13 February, 2020**
Talk and workshop on omics data analysis using the MetaOmGraph software. **University of Houston, Texas, USA.** Prepared workshop and taught an audience of 15 people.
3. *Exploratory data analysis with MetaOmGraph* **10 October, 2019**
Talk and workshop on exploratory data analysis using MetaOmGraph, a software which I developed. **School of Computational and Integrative sciences, Jawaharlal Nehru, New Delhi.** Prepared workshop and taught an audience of 35 people.
4. *Methods and Tools for Integrative Analysis of Big Heterogeneous RNA-Seq Datasets* **24 September, 2019**
My application was selected by a scientific committee for this competitive workshop **2nd Course on Computational Systems Biology of Cancer**, Institut Curie, Paris, France.
5. *Guest Lecturer* **Fall, 2018**
Developed and taught three guest lecture/labs (three hours each) on interactive exploratory data analysis of big data for **Biol 444, Bioinformatic Analysis, Iowa State University, USA.**
6. *Characterization of human orphan genes and its functions using annotated and unannotated genomic and massive RNA-seq data and metadata* **9 July, 2018**
The 26th Annual Conference on Intelligent Systems for Molecular Biology (ISMB), Illinois, USA.
7. *Data Science with Python workshop instructor* **19 October, 2018**
I led the data-science with Python workshop organized by **BCBGSO at Iowa State University, USA.** Prepared the workshop and taught an audience of 50 people.
8. *Python workshop instructor* **23-24 March, 2018**
Instructor for the Python workshop organized by **BCBGSO at Iowa State University, USA.** Prepared the workshop material and taught an audience of 60 people.

CONFERENCE POSTERS

1. E. Wurtele, **U. Singh**, P. Bhandary, J. Li, and A. S. Seetharam. Targets of opportunity: The birth of orphan genes, exemplified in yeast, arabidopsis, maize and humans. *Poster presented at Plant and Animal Genomes XXVIII conference, San Diego, California.* January 11-15, 2020
2. **U. Singh**, K. Dorman, and E. Wurtele. Methods and tools for integrative analysis of big heterogeneous rna-seq datasets. *Poster presented at 2nd Course on Computational Systems Biology of Cancer, Institut Curie, Paris, France, .* September 23-27, 2019
3. **U. Singh**, Z. Arendsee, M. Hufford, J. Walley, and E. Wurtele. De novo gene evolution in maize. *Poster presented at Society for Molecular Biology Evolution (SMBE) , Manchester, United Kingdom.* July 21-25, 2019
4. **U. Singh**, A. Seetharam, Z. Arendsee, and E. Wurtele. Characterization of human orphan genes and its functions using annotated and unannotated genomic and massive rna-seq data and metadata. *Poster presented at The 26th Annual Conference on Intelligent Systems for Molecular Biology (ISMB), Illinois, USA.* July 6-10, 2018
5. E. Wurtele, M. Hur, and **U. Singh**. Digital tools and resources workshop: Metaomgraph for 'omics data: Nosql-enabled big data visualization and analysis. *Poster presented at Plant and Animal Genomes XXVI conference, San Diego, California.* January 13-17, 2018

MENTORING

Graduate Mentor

Summer, 2020-Spring, 2021

Guided one Computer Science Masters student's thesis project. The project involved analysis of single-cell data.

Graduate Mentor

Summer, 2020-Spring, 2021

Guided one Computer Science Masters student's thesis project. The project involved developing interactive single-cell data visualization methods.

Undergraduate Mentor

Summer, 2020

Guided one bioinformatics undergraduate student

Undergraduate Mentor

Spring, 2019

Guided one Bioinformatics undergraduate student on his research working with Ribo-Seq data to identify novel ORFs with evidence of translation.

Undergraduate Mentor

Fall, 2018

Guided two Statistics undergraduate students in research-based project on statistical methods for analysis and interactively visualize of [big] sports data.

Graduate Mentor

Summer, 2018

Guided a Computer Science Masters student in his creative component project. This project was focused on identifying orphan genes in *Oryza sativa* and find their expression patterns using large amounts of RNA-Seq data.

MEMBERSHIPS IN HONORARY AND PROFESSIONAL SOCIETIES

Sigma Xi, USA

Society for Molecular Biology and Evolution

International Society for Computational Biology

ADMINISTRATIVE EXPERIENCE

Director of IT for BCB-GSO

August, 2017 - July, 2018

Served as the Director of IT for the Bioinformatics and Computational Biology Graduate Student Organization (BCB-GSO). My duties included planning and organizing BCB-GSO's educational workshops, planning and organizing our annual BCB symposium and assisting BCB students with IT related issues.

SOFTWARE TOOLS

1. **MetaOmGraph** A java software for interactive exploratory analysis of big omics datasets.
2. **pyrpipe** A python package for writing RNA-Seq workflows in an object oriented manner.
3. **orfiy** A fast and flexible ORF caller for big transcriptomic and metagenomic sequence data.
4. **community-detection-java** A Java library for community detection in networks using ensemble learning (in development).
5. **PLncPRO** A machine learning based tool written in python for prediction of lncRNAs using transcriptomic datasets.
6. **ORIS** A java tool for interactive exploratory analysis of bacterial and archaeal genomes and prediction of replication origin sites.

TECHNICAL SKILLS

- **Programming languages:** Python, Java, R, C, C++
- **Databases:** MySQL
- **Operating Systems:** Unix/Linux, Windows, macOS.

HOBBIES

- **Music:** I am a self-taught musician and I have been playing guitar for 10 years along with keyboards and drums. I love learning and exploring new musical instruments.
- **Athletics:** Basketball is my favourite sport and I played competitive basketball in school and college. I also enjoy playing tennis, swimming, kayaking and hiking.

REFERENCES

Dr. Eve Syrkin Wurtele

Professor, GDCB Dept.,
Iowa State University,
601 Science II,
Ames, IA 50011, USA.

✉ mash@iastate.edu
☎ (515) 708-3232
🐦 EveSyrkin