Urminder Singh

PhD Candidate
Iowa State University
415 Bessey Hall
Ames, IA 50011, USA.

✉ usingh AT iastate.edu
📞 (515) 817-3304
�github urmi-21
🐦 _urminder

Research Interests

Integrative omics analysis, *de novo* evolution, Machine learning, Interactive *Big Data* visualization

Education

2016 - present	**Ph.D. Candidate in Bioinformatics and Computational Biology with Minor in Statistics** *Iowa State University, Ames, IA, United States* **Supervisor**: Prof. Eve Wurtele **GPA**: 3.69/4.00
2013-2015	**M.Tech., Computational and Systems Biology** *School of Computational and Integrative Sciences* 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 **GPA**: 8.33/9.00
2010-2013	**M.S. in Computer Science** *Department of Computer Science* South Asian University, New Delhi, India **Thesis**: "SOFT: Software based Origin Finder Tool" **GPA**: 8.12/9.00
2007-2010	**B.Sc, Applied Physical Sciences** *S.G.T.B Khalsa College* University of Delhi, New Delhi, India

Honors and Awards

- **Elected associate member of Sigma Xi, The Scientific Research Honor Society**, USA   **February 2020**
- **International course certificate: Computational Systems Biology of Cancer**, Institut Curie, Paris, France **Fall 2019**
- **Professional Development Grants**, GPSS, Iowa State University, Ames, IA   **Fall 2019**
- **BCB Departmental Award**, BCB, Iowa State University, Ames, IA   **Fall 2019**
- **GDCB Departmental Travel Award**, GDCB, Iowa State University, Ames, IA   **Spring 2019**
- **ISMB EvolComp COSI award**, Intelligent Systems for Molecular Biology (ISMB)   **June 2018**
- **Sui-Tong Chan Fung Fund**, GDCB, Iowa State University, Ames, IA   **Spring 2018**
- **Wendell Miller Trust Graduate Fellowship**, Iowa State University, Ames, IA   **2016 - 2018**
- **CALS Presidential Scholarship**, Iowa State University, Ames, IA   **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

Research Assistant(*Wurtele Lab*), Iowa State University   **Jan 2017 - Present**

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 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. **U. Singh** and E. S. Wurtele. Genetic novelty: How new genes are born. *eLife*, 9:e55136, 2020

2. **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

3. **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

4. Z. Arendsee, J. Li, **U. Singh**, P. Bhandary, A. Seetharam, and E. S. Wurtele. fagin: synteny-based phylostratigraphy and finer classification of young genes. *BMC bioinformatics*, 20(1):1–14, 2019

5. 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

6. 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

7. **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

8. **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. A. S. Seetharam, **U. Singh**, J. Li, P. Bhandary, Z. Arendsee, and E. S. Wurtele. Maximizing prediction of orphan genes in assembled genomes. *bioRxiv*, 2019

2. J. Li, Z. Arendsee, **U. Singh**, and E. S. Wurtele. Recycling rna-seq data to identify candidate orphan genes for experimental analysis. *bioRxiv*, page 671263, 2019

3. Z. Arendsee, A. Wilkey, **U. Singh**, J. Li, M. Hur, and E. Wurtele. synder: inferring genomic orthologs from synteny maps. *BioRxiv*, page 554501, 2019

PROFESSIONAL ACTIVITIES

GRANTS

2019	**Co-PI**	*The prediction of plant & animal orphan genes with comparative genomics & data mining* **Agency**: Extreme Science and Engineering Discovery Environment (XSEDE) **Status**: Awarded ($6,461)
2017	**Co-PI**	*The prediction of plant & animal orphan genes with comparative genomics & data mining (start-up allocation)* **Agency**: Extreme Science and Engineering Discovery Environment (XSEDE) **Status**: Awarded ($3,376)

INVITED LECTURES AND WORKSHOPS

1. *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.

2. *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.

3. *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.

4. *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**.

5. *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.

6. *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.

7. *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

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 **2020-2022**
Society for Molecular Biology and Evolution **2020-2023**
International Society for Computational Biology **2018**

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

1. **community-detection-java** A Java library for community detection in networks using ensemble learning (in development).

   Source Code: https://github.com/urmi-21/community-detection-java

2. **pyrpipe** A python package for writing RNA-Seq workflows in an object oriented manner.

   Source Code: https://github.com/urmi-21/pyrpipe

3. **MetaOmGraph** A java software for interactive exploratory analysis of big omics datasets.

   Source Code: https://github.com/urmi-21/MetaOmGraph

4. **PLncPRO** A machine learning based tool written in python for prediction of lncRNAs using transcriptomic datasets.

   Source Code: https://github.com/urmi-21/PLncPRO

5. **ORIS** A java tool for interactive exploratory analysis of bacterial and archaeal genomes and prediction of replication origin sites.

   Source Code: https://github.com/urmi-21/ORIS

TECHNICAL SKILLS

- **Programming languages::** Python, Java, R, C, C++, PHP
- **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. My musical style is mostly inspired and influenced by psychedelic, progressive, and heavy metal genres.
- **Athletics:** Basketball is my favourite sport and I played competitive basketball in school and college. I also enjoy playing tennis, swimming, kayaking and hiking.
- **Motorcycles and Bikes:** I have a passion for motorcycles and bikes. Riding my motorcycle through the Himalayan regions of North India is one of the most pleasurable experiences of my life. I occasionally ride my bike through the scenic trails of Iowa, USA.

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

Dr. Karin Dorman
Professor, Dept. of Statistics,
Iowa State University,
2411 Snedecor,
Ames, IA 50011, USA.

✉ kdorman@iastate.edu
📞 (515) 294-6078

Dr. Mukesh Jain
Professor, School of Computational and Integrative Sciences,
Jawaharlal Nehru University,
New Delhi, 100671, India.

✉ mjain@jnu.ac.in
📞 +91-011-26704686

Dr. Annangarachari Krishnamachari
Asst. Prof., School of Computational and Integrative Sciences,
Jawaharlal Nehru University,
New Delhi, 100671, India.

✉ chari@mail.jnu.ac.in
📞 +91-011-26738703