

Anupama Jha anupamaj@uw.edu

Department of Genome Sciences, University of Washington, Seattle

Education

- Doctor of Philosophy in Computer and Information Science
University of Pennsylvania, Philadelphia, PA, USA *2014–2020*
- Master of Science in Informatics
Technische Universität München, Munich, Bavaria, Germany *2012–2014*
- Bachelor of Technology in Information Technology
Guru Gobind Singh Indraprastha University, New Delhi, India *2006–2010*

Research and Professional Experience

UW Data Science Postdoctoral Fellow at [Noble Lab](#), University of Washington
Advisor: [Dr. William Stafford Noble](#) *March 2021–Present*
My research projects involve developing machine learning methods for imputing high-throughput genomic data sets, particularly with respect to 3D genome architecture.

Postdoctoral Fellow at [Biociphers Lab](#), University of Pennsylvania
Advisor: [Dr. Yoseph Barash](#) *January 2021–February 2021*

Graduate Research Assistant at [Biociphers Lab](#), University of Pennsylvania
Advisor: [Dr. Yoseph Barash](#) *August 2014–December 2020*
My research projects have involved predicting splicing differences between tissues, regulatory networks between RNA-binding proteins with deep learning and developing EIG, an interpretation method for splicing code and other deep learning models based on genomic data. In parallel, I have contributed to research towards reliable identification and quantification of splicing events from RNA-Seq data and understanding the role of RNA-binding proteins in post-transcriptional regulation.

Master Thesis Student at [Model-Based Systems and Qualitative Reasoning Group](#), TU Munich
Advisor: [Dr. Peter Struss](#) *November 2013–August 2014*
Conceptualization and implementation of a generic tool for selection of learning goals for a knowledge-based machine learning system, application of the tool for the fitness training domain.

Graduate Research Assistant for [Human Brain Project](#), [Neurorobotics](#)
with [Dr. Florian Röhrbein](#) *November 2013–March 2014*
Surveyed on the availability of robot simulation tools and game engines for the neurorobotics platform of the Human Brain Project.

Graduate Research Assistant at [Rostlab](#), TU Munich
Supervisor: [Dr. Tobias Hamp](#) *March 2013–September 2013*
Project: [Prediction of Interaction sites in Proteins using only sequences](#). Developed neural network models for prediction of interaction hot spots in proteins using amino acid sequences and evolutionary information.

Systems Engineer at [Infosys](#)
Supervisor: [Saravjit Singh](#), [Raminder Singh](#) *June 2010–September 2012*
Trainings: Android, ASP.NET, ADO.NET, C#, C, UNIX, Oracle, SQL, Geographical Information Systems, PERL
Responsibilities included development of UI and back end for GPS enabled embedded navigation device with Android for [Tom-Tom Int. BV](#). (Eindhoven, Netherlands)

Awards and Honors

- Travel Fellowship, [ISMB/ECCB 2023](#), July 2023, Lyon, France.
2022 Parker Travel Award, [GS Retreat](#), September 2022, University of Washington.
UW Data Science Postdoctoral Fellow, [eScience Institute](#), April 2021, University of Washington.
Selected student abstract for oral presentation, [SAGES 2019](#), June 2019, Philadelphia, Pennsylvania, USA.
GHC 18 Scholarship, [Grace Hopper Celebration 2018](#), September 2018, Houston, Texas, USA.
Travel Fellowship, [ISMB/ECCB 2017](#), July 2017, Prague, Czech Republic.
Best Poster Award, [Symposium on Advances in Genomics, Epidemiology, and Statistics](#), June 2017, University of Pennsylvania.
Best Poster Award, [RNA Biology & Cancer 2017 Symposium](#), May 2017, University of Pennsylvania.

Publications (* represents equal contribution)

Anupama Jha*, Stephanie C Bohaczuk*, Yizi Mao, Jane Ranchalis, Benjamin J Mallory, Alan T Min, Morgan O Hamm, Elliott Swanson, Connor Finkbeiner, Tony Li, Dale Whittington, William Stafford Noble, Andrew B Stergachis, Mitchell R Vollger*. Fibertools: fast and accurate DNA-m6A calling using single-molecule long-read sequencing. *In review, Nature Genetics* (2023).

Di Wu, **Anupama Jha**, San Jewell, Natalie Maus, Jacob R Gardner, Yoseph Barash. Generative modeling for RNA splicing code predictions and design. *OpenReview* (2023).

Tangqi Fang, Yifeng Liu, Addie Woicik, Minsi Lu, **Anupama Jha**, Xiao Wang, Gang Li, Borislav Hristov, Zixuan Liu, Hanwen Xu, William Stafford Noble, Sheng Wang. Enhancing Hi-C contact matrices for loop detection with Capricorn, a multi-view diffusion model. *bioRxiv* (2023).

Jorge Vaquero-Garcia, Joseph K Aicher, Paul Jewell, Matthew R Gazzara, Caleb M Radens*, **Anupama Jha***, Christopher J Green, Scott S Norton, Nicholas F Lahens, Gregory R Grant, Yoseph Barash. RNA splicing analysis using heterogeneous and large RNA-seq datasets. *Nature Communications* (2023).

Anupama Jha*, Mathieu Quesnel-Vallieres*, Andrei Thomas-Tikhonenko, Kristen W Lynch, Yoseph Barash. Identifying common transcriptome signatures of cancer by interpreting deep learning models. *Genome Biology* (2022).

Xinjun Ji, **Anupama Jha**, Jesse Humenik, Louis R. Ghanem, Andrew Kromer, Christopher Duncan-Lewis, Elizabeth Traxler, Mitchell J. Weiss, Yoseph Barash, Stephen A. Liebhaber RNA binding proteins PCBP1 and PCBP2 are critical determinants of murine erythropoiesis. *Molecular and Cellular Biology* (2021).

Barry Slaff, Caleb Matthew Radens, Paul Jewell, **Anupama Jha**, Nicholas Lahens, Gregory R. Grant, Andrei Thomas-Tikhonenko, Kristen W. Lynch, Yoseph Barash MOCCASIN: A method for correcting for known and unknown confounders in RNA splicing analysis. *Nature Communications* (2021).

William P. Bone, Katherine M. Siewert, **Anupama Jha**, Derek Klarin, Scott M. Damrauer, the VA Million Veteran Project, Kyong-Mi Chang, Philip S. Tsao, Themistocles L. Assimes, Marylyn D. Ritchie, Benjamin F. Voight Multi-trait association studies discover pleiotropic loci between Alzheimers disease and cardiometabolic traits. *Alzheimer's research & therapy* (2020).

Anupama Jha, Joseph K Aicher, Matthew R Gazzara, Deependra Singh, and Yoseph Barash. Enhanced Integrated Gradients: improving interpretability of deep learning models using splicing codes as a case study. *Genome Biology* (2020).

Anupama Jha, Matthew R. Gazzara, and Yoseph Barash. Integrative Deep Models for Alternative Splicing. *Bioinformatics* (2017).

Matthew R. Gazzara, Michael J. Mallory, Renat Roytenberg, John Lindberg, **Anupama Jha**, Kristen W. Lynch, and Yoseph Barash. Ancient antagonism between CELF and RBFOX families tunes mRNA splicing outcomes. *Genome Research* (2017).

Pinaki Chakraborty, Shweta Taneja, Nandita Anand, **Anupama Jha**, Diksha Malik, and Ankit Nayar. An Optimizing Compiler for Turing Machine Description Language, *Journal of Computer Sciences, ICFAI University Press, Hyderabad, India* (2011).

Posters (+ represents best poster award, * represents equal contribution)

Predicting trans Hi-C contacts from DNA sequence with TwinC MLCB 2023, 4DN Annual Meeting 2023
Anupama Jha, Borislav Hristov, Xiao Wang, Sheng Wang, William J. Greenleaf, Anshul Kundaje, Erez Lieberman Aiden, William Stafford Noble.

Chrome-Zoo: cross-species chromatin profile prediction using DNA Zoo data UW GS Retreat 2022+, ISMB/ECCB 2023
Anupama Jha, Jacob Schreiber, Olga Dudchenko, Georgi K. Marinov, Anshul Kundaje, William J. Greenleaf, Erez S. Lieberman Aiden, William Stafford Noble.

Identifying universal cancer transcriptome patterns by interpreting deep learning models ISMB/ECCB 2021
Anupama Jha*, Mathieu Quesnel-Vallieres*, Andrei Thomas-Tikhonenko, Kristen Lynch, Yoseph Barash.
Presented by co-author Mathieu Q.V.

RBP-Pokedex: Prediction of RBP knockdown effect via DNN experiment modeling ISMB 2020
Anupama Jha, Matthew R. Gazzara, Caleb M. Radens, Paul R. Jewell, Yoseph Barash.

Interpretation of deep learning models in genomics: splicing codes as a case study SAGES 2019⁺, ISMB/ECCB 2019
Anupama Jha, Joseph K. Aicher, Deependra Singh, Yoseph Barash.

Assembling the Building Blocks for a Unified Splicing Code ISMB 2018
Anupama Jha, Matthew R. Gazzara, Yoseph Barash.

Integrative Deep Models for Alternative Splicing SAGES 2017⁺, ISMB/ECCB 2017, RNA Bio & Cancer Symposium 2017⁺
Anupama Jha, Matthew R. Gazzara, Yoseph Barash.

Predictive modeling framework for splice factor knockdown experiments ISMB 2016
Anupama Jha, Matthew R. Gazzara, Yoseph Barash.

Talks

Predicting trans Hi C contacts from DNA sequence with TwinC *November 2023, December 2023*
Talk at MLCB, November 2023.
Talk at 4D-Nucleome Monthly Meeting, December 2023.

Bidirectional translation between Hi-C and ATAC-seq data across species using Chrome-Zoo *2022, 2023*
Talk at ISMB/ECCB, July 2023.
UW Combi Seminar, May 2023.
Talk at the UW GS Retreat, Sep 2022.
Talk at the 4D-Nucleome monthly meeting, August 2022.
Talk at the DNA-Zoo monthly meeting, March 2022.
Talk at the ENCODE Consortium Meeting, Jan 2022.

Identifying common transcriptome signatures of cancer by interpreting deep learning models *July 2021*
Talk at the iRNA COSI of ISMB/ECCB 2021. Presented by co-author Mathieu Q.V.

RBP-Pokedex: Prediction of RBP knockdown effect via DNN experiment modeling *July 2020*
Talk at the iRNA COSI of ISMB 2020.

Improving interpretation of deep learning models: splicing codes as a case study *June 2019, July 2019*
Talk at the MLCSB COSI of ISMB/ECCB 2019.
Selected student abstract for oral talk at SAGES 2019.

Integrative Deep Models for Alternative Splicing *April 2017, July 2017, October 2017*
Talk at the iRNA COSI of ISMB/ECCB 2017.
Talk at PRiML Group at Upenn.
Talk at Penn RNA Group.

Teaching Experience

Software Carpentry Workshop, eScience Institute, UW *May 2021, October 2021*
Teaching Assistant for Python software carpentry workshop. Undertook Carpentry instructor training in Jan 2023.

GCB 537: Advanced Computational Biology *Spring 2016-2018*
TA with Dr. Yoseph Barash, UPenn
Teaching Assistant for the Ph.D. level course with three components: statistical data analysis and machine learning techniques for computational biology, discussion on current topics in genomics and computational biology, and hands on experience in data analysis, coding and evaluation of computational biology tools/algorithms.

GCB 537 Guest Lecture: Support Vector Machine *April 2016-2019*
Guest lecture on Support Vector Machine for Ph.D. level course at UPenn.

Deep Learning Reading Group *Summer 2016, Spring 2018*
Co-organizer with Dr. Yoseph Barash, UPenn
Reading group to cover deep learning book and interpretation methods for deep learning models.

Skills

Languages, Systems, and Tools

Proficient: Python, Tensorflow, PyTorch, High Performance Computing, UNIX, LaTeX.

Past work experience: R, Perl, MATLAB, Java, Android, C#, .NET Framework, C++, Oracle, SQL.

Relevant Graduate Courses

Machine Learning, Bayesian Statistics, Mathematical Statistics, Deep Learning, Advanced Computational Biology, RNA World, High-throughput Datasets for Biologists, Interpretation of Deep Learning Models, Adversarial and Secure Machine Learning, Computational Linguistics.

Service and outreach

Team Leader, UW Genome Sciences Hackathon

September 2022, September 2023

Led a team of 8 graduate students to make an m6A caller for Fiber-seq using PacBio long reads. The project led to Fibertools, a tool used for m6A calling by ENCODE, Human Pangenome Research Consortium (HPRC), All of Us Network, Undiagnosed Diseases Network (UDN), and the GREGoR Consortium.

Guest Lecture: **Genome Hackers**, Python camp for high school students

August 2023

Teach Python to high school students who identify as a gender minority.

Panel: **Is Graduate School for me?**

March 2019

Panel at [CAPWIC 2019](#) to discuss entry requirements and graduate school experience.

Guest Lecture: **AI and Computational Biology**

March 2018-2019

Guest lecture with Dr. Yoseph Barash at the [WICS High School Day for Girls at UPenn](#).

Mentoring Experience

Mentored four PhD students and two undergraduate students in my PhD and postdoctoral tenure.

Mentored undergraduate students at University of Pennsylvania in their final year project in Fall 2017 and Spring 2018.

Reviewer

Independent reviewer for Genome Biology, Nature Communications, PLOS Computational Biology, Briefings in Bioinformatics, Bioinformatics, RECOMB 2022-2024, ISMB 2022-2023, ACM-BCB 2022, TEAMC-2018, Nature Scientific Reports.

Co-reviewer with Prof. William Stafford Noble for Genome Research and Nature Genetics.

Co-reviewer with Prof. Yoseph Barash for NeurIPS 2015-2020, ISMB/ECCB 2016-2019, ICLR 2018-2020, ICML 2019-2020, PLOS Computational Biology, Bioinformatics, Nucleic Acids Research.

References are available on request.