

Yi Zhang

+1(217)-419-9634 • yzhan201@illinois.edu • 2132 IGB, Urbana, IL 61801

EDUCATION

Ph.D. candidate, Bioengineering and Biomedical Engineering Expected May 2019
University of Illinois at Urbana-Champaign (UIUC) Urbana-Champaign, IL

- Track: Bioinformatics and computational biology
- Dissertation: Functional interpretation of cancer-associated genetic variants.

Bachelor of Science, Biosciences June 2014
University of Science and Technology of China (USTC) Hefei, China

RESEARCH EXPERIENCE

Research Assistant, University of Illinois at Urbana-Champaign 2015 - Now

Functional interpretation of breast cancer-associated genetic variants.

- Developed integrative algorithms for large-scale and multi-omics NGS data analyses of cancer patients.
- Designed haplotype imputation-based statistics on RNA-seq data to reveal variants' allele-specific effects.
- Integrated resources such as TCGA, GTEx, 1000 Genomes, dbSNP, ENCODE for variant interpretation.
- Discovered biological functions of breast cancer-associated non-coding variants as modulators of transcription factor binding activity through *cis*-regulation by enhancers.

Machine learning models for gene regulation and cancer subtype prediction.

- Built Random Forest models for 3D chromosome regulation prediction using epigenomic data.
- Implemented Convolutional Neural Networks for breast cancer subtype prediction using genotype data.

Interactive web resource for machine learning analysis of biomedical data, and NGS techniques.

- Built web tools for clustering, dimensional reduction, and data visualization (Docker, Javascript).
- Built an interactive knowledge base of ~70 NGS techniques and analysis (3000+ views in 3 months).

Undergraduate Summer Researcher, University of California at Los Angeles 2013

- Developed signal detection algorithms and interfaces for neuromuscular data analysis and visualization.

PUBLICATIONS

Y. Zhang*, M. Manjunath*, S. Zhang, D. Chasman, S. Roy, and J.S. Song. "Integrative genomic analysis predicts causative cis-regulatory mechanisms of the breast cancer-associated genetic variant rs4415084." **Cancer Research**, 78(7), 1579-1591, (2018). *co-first authors.

Y. Zhang, M. Manjunath, Y. Kim, J. Heintz, and J.S. Song. "**SequencEnG**: an Interactive Knowledge Base of Sequencing Techniques." **Bioinformatics**, bty794, (2018).

M. Manjunath, **Y. Zhang**, Y. Kim, S.H. Yeo, O. Sobh, N. Russell, C. Followell, C. Bushell, U. Ravaoli, and J.S. Song. "**ClusterEnG**: an interactive educational web resource for clustering and visualizing high-dimensional data." **PeerJ Comput. Sci.**, 4:e155, (2018).

CONFERENCE AND POSTERS

Y. Zhang, M. Manjunath, S. Zhang, D. Chasman, S. Roy, and J.S. Song. "Integrative genomic analysis discovers the causative regulatory mechanisms of a breast cancer-associated genetic variant." **AACR Annual Meeting**, Chicago, IL (2018)

Y. Zhang, M. Manjunath, and J.S. Song. "Identifying causal mechanisms of germline risk variants in breast cancer." **Bioengineering Graduate Student Symposium**, UIUC (2017)

Y. Zhang, P. Gad, V. R. Edgerton. "Burst detection in EMG of stepping spinal cord injured rats." **Cross-disciplinary Scholars in Science and Technology summer research**, No.69, Los Angeles, CA (2013)

SKILLS

Programming: Python, Linux/Bash, R, MATLAB, C, HTML/CSS, JavaScript, TensorFlow

Bioinformatics: Biomedical data analysis, NGS data analysis, cancer genomics, statistics, machine learning

RELATED COURSES

| | | |
|------------------------------|----------------------|-------------------------|
| Computational Cancer Biology | Machine Learning | Multivariate Analysis |
| Stochastic Processes | Statistical Learning | Mathematical Statistics |

COURSE PROJECTS

Statistical Learning: Prediction of Movie votes and campus shooting with SVM and gradient boosting models.

Machine Learning: Classification of NGS technique articles using natural language processing and SVM models.

Stochastic Processes: Segmentation of copy number using HMM based on genotype data in cancer cells.

Statistical Data Analysis in Physics: A cross-entropy method for graph clustering and Cheeger constant estimation.

PROFESSIONAL EXPERIENCE

| | |
|--|-----------|
| Invited Speaker, Bioengineering Graduate Seminar, <i>UIUC</i> | 2018 |
| NHGRI Short Course on NGS: Technology & Statistical Methods, <i>Birmingham, AL</i> | 2016 |
| Big Data to Knowledge (BD2K) KnowEnG center at UIUC | 2015-2018 |
| Student study group organized and participated: | |
| • Deep Learning Study Group | 2018 |
| • Statistical Learning Study Group | 2017 |

TEACHING EXPERIENCE

| | |
|--|------------|
| Invited Speaker, Bioinformatics Seminar, <i>UIUC</i> | 2016, 2018 |
| • Presented bioinformatics lectures and lead discussions among graduate students | |
| Teaching Assistant, Biomedical Instrumentation Lab, Bioengineering, <i>UIUC</i> | 2014-2015 |
| • Lead lab and discussions among graduate students, graded lab reports | |

HONORS AND AWARDS

| | |
|--|-----------|
| Best Poster Award, Bioengineering Graduate Student Symposium, <i>UIUC</i> | 2017 |
| Outstanding Student Scholarship, School of Life Sciences, <i>USTC</i> | 2013-2014 |
| Outstanding Undergraduate Student Honor, <i>USTC</i> | 2013 |
| UCLA Cross-disciplinary Scholars in Sci. and Tech. (CSST) Summer Research Scholarship, <i>USTC</i> | 2013 |
| Aegon-Industrial Foundation Scholarship, <i>USTC</i> | 2012 |
| Panasonic Elite Scholarship, <i>USTC</i> | 2011 |

UNIVERSITY SERVICE

| | |
|---|-----------|
| Volunteer, Genome Day, Carl R. Woese Institute for Genomic Biology, <i>UIUC</i> | 2017 |
| Vice-chairman of Student Union, School of Life Sciences, <i>USTC</i> | 2012-2013 |
| Organizer, School Photography Exhibition | 2012 |