Yi Zhang, Ph.D.

+1(217)419-9634 • yiz@ds.dfci.harvard.edu • yi-zhang.github.io

Department of Data Science, Dana-Farber Cancer Institute, Harvard University, Boston, MA, USA

SUMMARY

I am a computational biologist as a Research Fellow in the Department of Data Science at Dana-Farber Cancer Institute and Harvard University T.H. Chan School of Public Health. My research interest falls at the intersection of biomedical data science, machine learning, single-cell genomics, and cancer immunotherapy. I have developed computational methods to study gene regulatory role of disease-associated human genetic variants, and currently focus on developing large-scale single-cell data integration method to learn about tumor immunity.

ACADEMICS AND EDUCATION

Dana-Farber Cancer Institute, Harvard University, *Boston, MA*   Jul 2019 - Present
Research Fellow,
Department of Data Science, Dana-Farber Cancer Institute
Department of Biostatistics, Harvard University T.H. Chan School of Public Health

University of Illinois at Urbana-Champaign, *Urbana, IL*   Aug 2014 - May 2019
Ph.D. Bioengineering and Biomedical Engineering
- Dissertation: Computational methods for functional interpretation of cancer-associated genetic variants
- Committee: Dr. Jun S. Song, Dr. Saurabh Sinha, Dr. Dave S. Zhao, and Dr. Pablo Perez-Pinera

University of Science and Technology of China, *Hefei, China*   Aug 2010 - Jun 2014
Bachelor of Natural Science, Biological Science, Talent Program in Life Sciences

PROFESSIONAL EXPERIENCE

Dana-Farber Cancer Institute, Harvard University, *Boston, MA*   Jul 2019 - Present
Research Fellow, Department of Data Science
Advisor: Dr. X. Shirley Liu
- Develop computational methods for integrating single cell genomics to elucidate tumor immunity regulation.
- Clinical genomics data analysis and interpretation for cancer immunotherapy.
Collaboration with Dr. Myles Brown's lab, DFCI
- Nuclear receptor gene regulation in tumor immunity and immunotherapy efficacy.
Collaboration with Dr. Catherine J. Wu's lab, DFCI
- Transcriptomics analysis for an immunotherapy trial to identify genomic biomarkers in relapsed leukemia.
Collaboration with Dr. David Fisher's lab, Massachusetts General Hospital
- Transcriptomics analysis for discovering immune switch during melanoma timecourse treatment
Collaboration with Dr. Xihong Lin's lab, Harvard University
- Human genetics associated with rare variants using epigenetic data

University of Illinois at Urbana-Champaign, *Urbana, IL*   Jan 2015 - May 2019
Research Assistant, NIH Big Data to Knowledge (BD2K) U54 Center
Advisor: Dr. Jun S. Song
Functional interpretation of breast cancer-associated genetic variants.
- Developed integrative methods for large-scale and multi-omics Next Generation Sequencing (NGS) data analyses for cancer genomics.

- Designed haplotype imputation-based statistics on RNA-seq data to reveal variants' allele-specific effects.
- Discovered non-coding variants as modulators of transcription factor binding activity in enhancers.

Machine learning models to predict 3D chromatin structure.
- Built machine learning models using random forest to predict chromatin interaction from epigenetics.

Interactive web resource for sequencing techniques and machine learning analysis of biomedical data.
- Built an interactive educational web resource of ~70 sequencing techniques and analysis pipelines (9k users and 18k views since Jul 2018).
- Built online data visualization tools for clustering and dimensional reduction.

University of California at Los Angeles, *Los Angeles, CA*   May 2013 – Aug 2013
Undergraduate Research Assistant, Department of Bioengineering.
- Developed an algorithm for burst detection in electromyography signals of stepping spinal cord injured rats.

University of Science and Technology of China, *Hefei, China*   May 2013 – May 2014
Undergraduate Research Assistant, Neurophysics Lab.
- Develop methods for pattern recognition of spatiotemporal neuronal firing in hippocampus cultured on-chip

PUBLICATIONS

Penter L, **Zhang Y**, Savell A, Huang T, Cieri N, Thrash EM, et al. Molecular and cellular features of CTLA-4 blockade for relapsed myeloid malignancies after transplantation. **Blood. (2021)** https://pubmed.ncbi.nlm.nih.gov/33720354/

Zhang Y, Liu Y, Liu XS. Neural network architecture search with AMBER. **Nature Machine Intelligence. (2021)** *#News & Views.* 3:372–3. https://www.nature.com/articles/s42256-021-00350-x

Zhang Y, Xiang G, Fan J, Wang C, Lynch A, Zeng Z, Zhang W, Kang J, Jiang A, Gu S, Wan C, Zhang B, Brown M, Meyer C, Liu XS. TSMeC: interpretable meta-components learned from large-scale single cells in tumor microenvironment. *#Manuscript in preparation (2021).*

Wang X*, Tokheim C*, Wang B*, Gu SS*, Tang Q, Li Y, Traugh N, **Zhang Y**, et al. In Vivo CRISPR Screens Identify E3 Ligase Cop1 as a Modulator of Macrophage Infiltration and Cancer Immunotherapy Target. **Cell. (2021)** *Accepted*. https://www.biorxiv.org/content/10.1101/2020.12.09.418012v1

Zeng Z, Wong C, Yang L, Ouardaoui N, Li D, Zhang W, Gu S, **Zhang Y**, Liu Y, Wang X, Fu J, Zhou L, Zhang B, Kim S, Yates K, Brown M, Freeman G, Uppaluri R, Manguso R, Liu XS. TISMO: syngeneic mouse tumor database to model tumor immunity and immunotherapy response. Nucleic Acids Research. https://doi.org/10.1093/nar/gkab804

Gu SS, Zhang W, Wang X, Jiang P, Traugh N, Li Z, et al. Therapeutically Increasing MHC-I Expression Potentiates Immune Checkpoint Blockade. **Cancer Discovery (2021)** 11:1524–41. https://cancerdiscovery.aacrjournals.org/content/11/6/1524

Manjunath M, Yan J, Youn Y, Drucker KL, Kollmeyer TM, McKinney AM, Zazubovich V, **Zhang Y**, et al. Functional analysis of low-grade glioma genetic variants predicts key target genes and transcription factors. **Neuro-Oncology. (2021)** 23:638–49. https://academic.oup.com/neuro-oncology/article/23/4/638/5948532

Qiu X, Boufaied N, Hallal T, Feit A, Polo A de, Luoma AM, et al. MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. bioRxiv. (2021).04.23.441016. https://www.biorxiv.org/content/10.1101/(2021).04.23.441016v1

Baur B, Schreiber J, Shin J, Zhang S, **Zhang Y**, Manjunath M, et al. Leveraging epigenomes and three-dimensional genome organization for interpreting regulatory variation. **bioRxiv** (2021).08.29.458098. https://www.biorxiv.org/content/10.1101/(2021).08.29.458098v1

Penter L, **Zhang Y**, Savell A, Ranasinghe S, Huang T, Cieri N, et al. Local and Systemic Effects of Immune Checkpoint Blockade on Relapsed Myeloid Malignancies Following Allogeneic Hematopoietic Stem Cell Transplantation. **Blood. (2020);** 136:34–5. https://doi.org/10.1182/blood-2020-139487

Gu SS, Wang X, Hu X, Jiang P, Li Z, Traugh N, et al. Clonal tracing reveals diverse patterns of response to immune checkpoint blockade. **Genome Biology. (2020)** 21:263. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02166-1

Manjunath M, **Zhang Y**, Zhang S, Roy S, Perez-Pinera P, Song JS. ABC-GWAS: Functional Annotation of Estrogen Receptor-Positive Breast Cancer Genetic Variants. **Frontiers in Genetics. (2020)**;0:730.

Zhang Y, Manjunath M, Kim Y, Heintz J, Song JS. SequencEnG: An interactive knowledge base of sequencing techniques. **Bioinformatics. (2019**); doi: 10.1093/bioinformatics/bty794. NGS web resource of 8.5k users and 17.7k view.

Zhang Y, Manjunath M, Yan J, Baur BA, Zhang S, Roy S, et al. The cancer-associated genetic variant rs3903072 modulates immune cells in the tumor microenvironment. **Front Genet. (2019);** doi: 10.3389/fgene.2019.00754

Zhang Y. Functional interpretation of cancer-associated genetic variants. **University of Illinois at Urbana-Champaign.** (**2019**) https://www.ideals.illinois.edu/handle/2142/105017

Zhang Y*, Manjunath M*, Zhang S, Chasman D, Roy S, Song JS. Integrative Genomic Analysis Predicts Causative Cis-Regulatory Mechanisms of the Breast Cancer–Associated Genetic Variant rs4415084. **Cancer Research. (2018)** 78:1579–91. https://cancerres.aacrjournals.org/content/78/7/1579

Manjunath M, **Zhang Y**, Kim Y, Yeo SH, Sobh O, Russell N, et al. ClusterEnG: An interactive educational web resource for clustering and visualizing high-dimensional data. **PeerJ Computer Science.** (**2018**); doi: 10.7717/peerj-cs.155

*co-first authors

CONFERENCES AND POSTER PRESENTATIONS

L. Penter, **Y. Zhang**, A. Savell, S. Ranasinghe, T. Huang, N. Cieri, S. Gohil, W. Zhang, S. Li, R. Zeiser, K. J. Livak, J. Ritz, H. Kim, R. J. Soiffer, S. X. Liu, M. S. Davids, P. Bachireddy, C. J. Wu. "Local and systemic effects of immune checkpoint blockade on relapsed myeloid malignancies following allogeneic hematopoietic stem cell transplantation." **American Society of Hematology Annual Meeting**, *Washington, DC* (2020)

Y. Zhang, M. Manjunath, J. Yan, Britany Baur, S. Zhang, S. Roy, and J.S. Song. "A breast cancer-associated genetic variant modulates immune cells in the tumor microenvironment." **Variant to Function Symposium**, *Broad Institute of MIT and Harvard, Boston, MA* (2020).

Y. Zhang, M. Manjunath, J. Yan, Britany Baur, S. Zhang, S. Roy, and J.S. Song. "A breast cancer-associated genetic variant modulates immune cells in the tumor microenvironment." **Program in Quantitative Genomics**, *Harvard T.H. Chan School of Public Health, Boston, MA* (2019).

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." **American Association for Cancer Research 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), **Best Poster Award**.

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

TECHNICAL SKILLS

Programming: Python, R, scikit, Pytorch, Git, HTML, CSS, JavaScript, Unix, AWS, Julia, MATLAB, C.
Bioinformatics: RNA-seq, Single-cell, Genotyping array, ChIP-seq, Hi-C, Multi-omics.
Statistics: Regularized regression, Bayesian statistics, MCMC, multivariate analysis.
Machine/Deep learning: Supervised (CNN, boosted tree, random forest, SVM), Unsupervised (clustering, autoencoder, low-dimensional embedding).

TALKS

"Functional mechanisms of cancer-associated human variants." *Graduate Student Conference, UIUC.* (2018)
"Prioritizing functional cancer-associated human germline variants." *UIUC BD2K center talk.* (2017)

PEER REVIEW

Journals: Nature Machine Intelligence, Nature Medicine, Nature Metabolisms, Cell Reports, Biochimica et Biophysica Acta, Frontiers in Immunology, Frontiers in Genetics, PeerJ.

TEACHING EXPERIENCE

Mentorship
- Jiajinlong Kang, *Master student, Harvard University, Computational Biology and Quantitative Genetics*
- Yijia Jiang, *Master student, Harvard University, Master of Biomedical Informatics*
- Yingxiao Shi, *Ph.D. student, Harvard University, Biological and Biomedical Sciences*
- Evan Arnold, *Undergraduate student, Harvard University*

Teaching Assistant, Harvard University   2020
- Developed GWAS series course videos used in STAT115/215 at Harvard University
- Course materials for High Impact Cancer Research Course at Harvard Medical School

Invited Speaker, Bioinformatics Seminar, UIUC   2016, 2018
- Presented bioinformatics lectures and led discussions among graduate students

Teaching Assistant, Biomedical Instrumentation Lab, Bioengineering, UIUC   2014 - 2015
- Led labs and discussions among graduate students.

FUNDING AND AWARDS

DFCI-Novartis Drug Discovery and Translational Research Grant, main contributor   2021
Best Poster Award, Bioengineering Graduate Student Symposium, UIUC   2017
National Student Innovation Research Grant, USTC   2014
- Pattern recognition of spatiotemporal neuronal firing of cultured hippocampus on-chip

Outstanding Student Scholarship, School of Life Sciences, USTC   2013 - 2014
Outstanding Undergraduate Student Honor, USTC   2013
UCLA Cross-disciplinary Scholars in Sci. and Tech. (CSST) Research Scholarship, USTC   2013
Aegon-Industrial Foundation Scholarship, USTC   2012

COURSES AND TRAINING

Related Courses

Mathematical Statistics	Machine Learning	Statistical Data Analysis in Physics
Stochastic Processes	Statistical Learning	Computational Cancer Biology

Course Projects
- *Statistical* Learning: Prediction of Movie votes and campus shooting with gradient boosting models.
- *Machine Learning*: Classification of NGS technique articles using 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 Cheeger constant estimation.

Study Groups: Deep Learning (2018), Statistical Learning (2017).

Workshops: NHGRI Short Course on NGS: Technology & Statistical Methods, *Birmingham, AL* (2016)
New England Future Faculty Workshop, Northeastern University, *Boston, MA* (2020)