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 postdoctoral 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 immunity in diseases. I have developed computational methods to characterize gene regulatory function of disease-associated genetic variants, and also developed interpretable machine learning methods for large-scale single-cell data. I also enthusiastically collaborate with biologists and clinician colleagues to understand immune-related diseases such as cancer and infection.

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
Affiliated Postdoc Member, Broad Institute of MIT and Harvard
- Mentors: Dr. X. Shirley Liu, Dr. Xihong Lin, Dr. Myles Brown

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

RESERACH EXPERIENCE

Dana-Farber Cancer Institute, *Boston, MA*   Jul 2019 - Present
Research Fellow, Department of Data Science
Advisors: Dr. X. Shirley Liu, Dr. Myles Brown
- Elucidated immunity regulation using machine learning on single-cell genomics for diseases like cancer
- Clinical genomics data analysis and interpretation for cancer immunotherapy
Collaboration with Dr. Myles Brown's lab, DFCI
- Studied nuclear receptor gene regulation in tumor immunity using single-cell based gene networks
Collaboration with Dr. Catherine J. Wu's lab, DFCI
- Built RNA analysis pipeline to identify genomic biomarkers in relapsed leukemia in immunotherapy trial
Collaboration with Dr. David Fisher's lab, Massachusetts General Hospital
- Discovered immune transcriptomic switch during treatment of the skin disease nevi and melanoma
Collaboration with Center for Functional Cancer Epigenetics, DFCI
- Developing analytical methods for single-cell multiomics on perturbated immune cells

Harvard University T.H. Chan School of Public Health, *Boston, MA*   Nov 2021- Present
Research Fellow, Department of Biostatistics
Advisor: Dr. Xihong Lin

Yi Zhang, Ph.D.

- Mapping variant to function, NHGRI Impact of Genomic Variation on Function (IGVF) Consortium

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. (10k users since Jul 2018).
- Built online data visualization tools for clustering and dimensional reduction.

Collaboration with Dr. Sushmita Roy's lab, University of Wisconsin-Madison

- Built machine learning models for 3D chromatin structure prediction using epigenomic features.

University of California at Los Angeles, *Los Angeles, CA*   May 2013 – Aug 2013
Undergraduate Research Assistant, Department of Bioengineering.
Advisor: Dr. V. Reggie Edgerton
- Developed algorithms 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.
Advisor: Dr. Guoqiang Bi
- Develop methods for pattern recognition of spatiotemporal neuronal firing in hippocampus cultured on-chip

PUBLICATIONS

Zhang Y, Xiang G, Fan J, Wang C, Lynch A, Zeng Z, Zhang W, Kang J, Jiang A, Gu S, Wan C, Zhang B, Liu XS#, Brown M#, Meyer C#. "MetaTiME: Meta-components of the Tumor Microenvironment." *Under Review*. (2022) bioRxiv: https://doi.org/10.1101/2022.08.05.502989

Penter L, **Zhang Y**, Savell A, Huang T, Cieri N, Thrash EM, Kim-Schulze S, Jhaveri A, Fu J, Ranasinghe S, Li S, Zhang W, Hathaway E, Nazzaro M, Kim H, Chen H, Thurin M, Rodig S, Severgnini M, Cibulskis C, Gabriel S, Livak K, Cutler C, Antin J, Nikiforow S, Koreth J, Ho V, SArmand P, Ritz J, Steicher H, Neuberg D, Hodi F, Gnjatic S, Soiffer R, Liu SX, David M, Bachireddy P, Wu CJ#. "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, Manjunath M, Kim Y, Heintz J, Song JS. "SequencEnG: An interactive knowledge base of sequencing techniques." **Bioinformatics. (2019)**); doi: 10.1093/bioinformatics/bty794. Web resource of 9k 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." **Frontiers in Genetics (2019);** doi:10.3389/fgene.2019.00754

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

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

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

Choi Y*, Erlich T*, Franque M, Rachmin I, Flesher J, Schiferle E, **Zhang Y**, Silva M, Jiang A, Dobry A, Lacher S, Freund O, Feder E, Cortez J, Ryu S, Samuels Y, Zakka L, Azin M, Burd C, Shapless NN, Liu XS, Meyer C, Austen W, Bojovic B, Certrulo C, Mihm M, Hoon D, Demehri S, Hawryluk E, Fisher DE#. "Topical therapy for regression and melanoma prevention of congenital giant nevi." **Cell (2022)**

Qiu X*, Boufaied N*, Hallal T, Feit A, Polo A de, Luoma AM, Alahmadi W, Larocque J, Zadra G, Xie Y, Gu S, Tang Q, **Zhang Y**, Syamala S, Seo JH, Bell C, O'Connor E, Liu Y, Schaeffer E, Karnes R, Weinmann S, Davicioni E, Morrissey C, Cejas P, Ellis L, Loda M, Wucherpfennig K, Pomerantz M, Spratt D, Corey E, Freedman M, Liu XS, Brown M, Long H#, Labbé D#. "MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets." **Nature Communications. (2022)** https://www.biorxiv.org/content/10.1101/2021.04.23.441016v1

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. (2022)** https://doi.org/10.1093/nar/gkab804

Zhou L*, Zeng Z*, Egloff AM, Zhang F, Guo F, Campbell KM, Du P, Fu J, Zolkind P, Ma X, Zhang Z, **Zhang Y**, Wang X, Gu S, Riley R, Nakahori Y, Keegan J, Haddad R, Schoenfeld J, Griffith O, Manguso RT, Lederer JA, Liu XS#, Uppaluri R#. "Checkpoint blockade-induced CD8+ T cell differentiation in head and neck cancer responders." **Journal for ImmunoTherapy of Cancer. (2022)** doi:10.1136/jitc-2021-004034

Wang X*, Tokheim C*, Wang B*, Gu SS*, Tang Q, Li Y, Traugh N, **Zhang Y**, Li Z, Zhang B, Fu J, Xiao T, Li W, Meyer C, Chu J, Jiang P, Cejas P, Lim K, Long H, Brown M#, Liu XS#. "In Vivo CRISPR Screens Identify E3 Ligase Cop1 as a Modulator of Macrophage Infiltration and Cancer Immunotherapy Target." **Cell. (2021)** https://pubmed.ncbi.nlm.nih.gov/34582788/

Gu SS*, Zhang W*, Wang X*, Jiang P*, Traugh N, Li Z, Meyer C, Stewig B, Xie Y, Bu X, Manos M, Front-Tello A, Gjini E, Lako A, Lim K, Conway J, Tewari A, Zeng Z, Sahu A, Tokleim C, Weirather J, Fu J, **Zhang Y**, Kroger B, Liang J, Cejas P, Freeman G, Rodig S, Long H, Gewurz B, Hodi F, Brown M#, Liu XS#. "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**, Costello J, Eckel-Passow J, Selvin P, Jenkins R, Song JS#. "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

Baur B, Schreiber J, Shin J, Zhang S, **Zhang Y**, Manjunath M, Song JS, Noble WS, Roy S#. "Leveraging epigenomes and three-dimensional genome organization for interpreting regulatory variation." *Under Review.* bioRxiv (2021).08.29.458098. https://www.biorxiv.org/content/10.1101/(2021).08.29.458098v1

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.

Gu SS, Wang X, Hu X, Jiang P, Li Z, Traugh N, Bu X, Tang Q, Wang C, Zeng Z, Fu J, Meyer C, Zhang Y, Cejas P, Lim K, Wang J, Zhang W, Tokheim C, Sahu A, Xing X, Kroger B, Ouyang Z, Long H, Freeman G, Brown M, Liu XS.

"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**, 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 # co-corresponding authors

TALKS

"Meta-components of the tumor immune microenvironment." *Invited Seminar Talk, Emerging Scholars in Genome Sciences Symposium, 6 selected scholars*, *Center for Public Health Genomics, University of Virginia* (2022)

"MetaTiME: single-cell derived gene expression signatures for interpreting the tumor microenvironment." *Invited Seminar Talk*, *Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute* (2022)

"Functional mechanisms of cancer-associated human variants." *Invited talk, UIUC Bioinformatics Seminar* (2018)

"Functional mechanisms of cancer-associated human variants." *Graduate Student Conference, UIUC.* (2018)

"Prioritizing functional cancer-associated human germline variants." *UIUC BD2K center talk.* (2017)

CONFERENCES AND POSTER PRESENTATIONS

Zhang Y, Xiang G, Fan J, Wang C, Lynch A, Zeng Z, Zhang W, Kang J, Jiang A, Gu S, Wan C, Zhang B, Liu XS#, Brown M#, Meyer C#. "MetaTiME: Meta-components of the Tumor Microenvironment." IGVF Annual Meeting 2022

Liu S, **Zhang Y**, Brown M. "Immune population changes in mammary adipose tissue during the development of obesity and their influence on triple negative breast cancer progression." Cancer Research 15 June 2022; 82 (12_Supplement): 6147.

Gu SS, Zhang W, Wang X, Jiang P, Traugh N, Li Z, Meyer C, Stewig B, Xie Y, Bu X, Manos M, Front-Tello A, Gjini E, Lako A, Lim K, Conway J, Tewari A, Zeng Z, Sahu A, Tokleim C, Weirather J, Fu J, **Zhang Y**, Kroger B, Liang J, Cejas P, Freeman G, Rodig S, Long H, Gewurz B, Hodi F, Brown M, Liu XS. "Increasing MHC-I expression to potentiate immune checkpoint blockade therapy", The Society for Immunotherapy of Cancer (SITC) Annual Meeting, *Washington, DC* (2021)

Penter L, **Zhang Y**, Savell A, Ranasinghe S, Huang T, Cieri N, Gohil S, Zhang W, Li S, Zeiser R, Livak KJ, Ritz J, Kim H, Soiffer RJ, Liu XS, Davids MS, Bachireddy P, Wu CJ. "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)

Zhang Y, Manjunath M, Yan J, Baur B, Zhang S, Roy S, and Song JS. "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).

Zhang Y, Manjunath M, Yan J, Baur B, Zhang S, Roy S, and Song JS. "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).

Zhang Y, Manjunath M, Zhang S, Chasman D, Roy S, and Song JS. "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).

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

Zhang Y, Gad P, Edgerton VR. "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).

TEACHING EXPERIENCE

Student Mentorship
- Jack Kang, *Master student, Harvard University, Computational Biology and Quantitative Genetics*
- Alva 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
- Developed course materials for High Impact Cancer Research Course at Harvard Medical School

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

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

FUNDING AND AWARDS

EECS Rising Star Workshop Awardee, University of Texas at Austin	2022
Emerging Scholars in Genome Sciences Symposium, Center for Public Health Genomics, University of Virginia	2022
DFCI-Novartis Research Grant, major contributor in grant application	2021

- DFCI-Novartis Drug Discovery & Translational Research Program, $350k for 3 years
- High-throughput Characterization of Human Immune Response under Cytokine Stimulation

Best Poster Award, Bioengineering Graduate Student Symposium, UIUC	2017
National Student Innovation Research Grant, USTC	2014

- One-year research grant awarded to undergraduate student to finish thesis research

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

- International exchange program for top undergraduate student to perform funded summer research

Aegon-Industrial Foundation Scholarship, USTC	2012
Panasonic Elite Scholarship, USTC	2011

MEMBERSHIPS

American Association for Cancer Research (AACR)	2022
International Society for Computational Biology (ISCB)	2022
The Society for Immunotherapy of Cancer (SITC)	2019-2020

SKILLS

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

COURSES AND TRAINING

Related Courses

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

Course Projects, UIUC

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

ACTIVITIES

Organizer and Host, HMS-CSSA Harvard Life Science Annual Symposium (virtual, 500 registrants)	2022
Host, HMS-CSSA Harvard Life Science Monthly Symposium	2022
Wildlife Photography	2015-present
UIUC Genomics Day	2017
USTC Student Traditional Orchestra	2014
National Physics Olympiad Competition, Shandong Province, Silver	2008

PEER REVIEW

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