

# Wanwen Zeng

318 Campus Drive,  
Stanford, CA, 94305

Page: [wanwenzeng.github.io](https://wanwenzeng.github.io)  
Email: [wanwen@stanford.edu](mailto:wanwen@stanford.edu)

## Academic Appointment

---

<b>Stanford University</b> , Department of Statistics <i>Postdoctoral Scholar</i>	06/2021 – Present Advisor: <b>Wing H. Wong</b>
<b>Nankai University</b> , College of Software Engineering <i>Assistant Professor</i>	11/2019 – 05/2021

## Education

---

<b>Tsinghua University</b> , Beijing, China Ph.D., Department of Automation Thesis: <i>Construction and analysis of regulatory networks based on machine learning approaches</i>	09/2014 – 10/2019 Advisor: <b>Rui Jiang</b>
<b>Stanford University</b> , California, USA Visiting Scholar, Department of Statistics	09/2017 – 09/2018 Advisor: <b>Wing H. Wong</b>
<b>Tianjin University</b> , Tianjin, China B.S., Department of Computer Science and Technology, Rank: 1/107 Thesis: <i>Integrating multiple genomic data to prioritize disease-causal SNVs</i>	09/2010 – 07/2014
<b>Nankai University</b> , Tianjin, China BEc., Department of Finance (Double Major), Rank: 3/89	09/2011 – 07/2014

## Publications

---

\* *co-first authors*; # *co-corresponding authors*

*Preprints or ongoing works:*

- [1] **Wanwen Zeng**, Hanmin Guo, Qiao Liu, Wing H. Wong<sup>#</sup>. How to improve polygenic prediction from increasingly prevalent whole-genome sequencing data? (Under Review in *Proceedings of the National Academy of Sciences*)
- [2] Xuejian Cui, Qijin Yin, Zijing Gao, Zhen Li, Xiaoyang Chen, Shengquan Chen, Qiao Liu, **Wanwen Zeng**<sup>#</sup>, and Rui Jiang<sup>#</sup>. CREATE: cell-type-specific cis-regulatory elements identification via discrete embedding. (Under Review in *Nature Communications*)
- [3] **Wanwen Zeng**<sup>\*</sup>, Shuang Chen<sup>\*</sup>, Yuti Liu<sup>\*</sup>, Qiao Liu, Wing H. Wong<sup>#</sup>. bpBERT: base-resolution BERT models reveal DNA sequence regulatory syntax and variants. (In submission)
- [4] Qiao Liu<sup>\*</sup>, **Wanwen Zeng**<sup>\*</sup>, Lexin Li<sup>#</sup>, Wing H. Wong<sup>#</sup>. Associating genotype to imaging and clinical phenotypes of Alzheimer's disease by leveraging genomic large language model. (In submission)
- [5] Xixi Yan<sup>\*</sup>, Yang Li<sup>\*</sup>, **Wanwen Zeng**, Wen Wang<sup>#</sup>, Yu Wang<sup>#</sup>. Decoding the cellular and molecular basis of cattle horn development using single-cell multi-omics data. (Under Review in *Proceedings of the National Academy of Sciences*)
- [6] Zijing Gao, Qiao Liu, **Wanwen Zeng**, Rui Jiang, Wing Hung Wong. EpiGePT: a Pretrained Transformer model for epigenomics. (Minor revision in *Genome Biology*)

*Peer-reviewed works:*

- [8] Wei Shao, Yuti Liu, Shuang Zhang, Shuang Chen, Qiao Liu, **Wanwen Zeng**. Inferring gene regulatory networks based on genetically perturbed scATAC-seq data. *International Conference on Bioinformatics and Biomedicine (BIBM)*, 2024
- [9] Wenran Li<sup>\*</sup>, **Wanwen Zeng**<sup>\*</sup>, Wing H. Wong<sup>#</sup>. Modeling the causal mechanism between genotypes and phenotypes using large-scale biobank data and context-specific regulatory networks. *Cybernetics and Intelligence*, 2024.
- [10] Qijin Yin, Rui Fan, Xusheng Cao, Qiao Liu<sup>#</sup>, Rui Jiang<sup>#</sup>, **Wanwen Zeng**<sup>#</sup>. Deepdrug: a general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. *Quantitative Biology*, 2023, 11(3): 260-274.
- [11] Shuang Zhang<sup>\*</sup>, Shuang Chen<sup>\*</sup>, Yuti Liu<sup>\*</sup>, Qiao Liu, **Wanwen Zeng**<sup>#</sup>. Applications of transformer-based language models in bioinformatics: a survey. *Bioinformatics Advances*, 2023, 3(1): vbad001.
- [12] Qiao Liu<sup>\*</sup>, **Wanwen Zeng**<sup>\*</sup>, Wei Zhang, Sicheng Wang, Hongyang Chen, Rui Jiang<sup>#</sup>, Mu Zhou<sup>#</sup>, Shaoting Zhang<sup>#</sup>. Deep generative modeling and clustering of single cell Hi-C data. *Briefings in Bioinformatics*, 2023, 24(1): bbac494.

- [13] **Wanwen Zeng**<sup>\*</sup>, Qiao Liu<sup>\*</sup>, Qijin Yin<sup>\*</sup>, Rui Jiang<sup>#</sup>, Wing H. Wong<sup>#</sup>. HiChIPdb: a database of HiChIP regulatory interactions. *Nucleic Acids Research*, 2022.
- [14] **Wanwen Zeng**<sup>\*</sup>, Jingxue Xin<sup>\*</sup>, Rui Jiang<sup>#</sup>, Yong Wang<sup>#</sup>. Reusability report: Compressing regulatory networks to vectors for interpreting gene expression and genetic variants. *Nature Machine Intelligence*, 2021: 1-5.
- [15] **Wanwen Zeng**<sup>\*</sup>, Shengquan Chen<sup>\*</sup>, Xuejian Cui<sup>\*</sup>, Xiaoyang Chen, Zijing Gao, Rui Jiang. SilencerDB: a comprehensive database of silencers. *Nucleic Acids Research*, 2021, 49(D1): D221-D228.
- [16] **Wanwen Zeng**, Yong Wang<sup>#</sup> and Rui Jiang<sup>#</sup>. Integrating distal and proximal information to predict gene expression via a densely connected convolutional neural network. *Bioinformatics*, 2020, 36(2): 496-503.
- [17] **Wanwen Zeng**<sup>\*</sup>, Xi Chen<sup>\*</sup>, Zhana Duren<sup>\*</sup>, Yong Wang, Rui Jiang<sup>#</sup> and Wing Hung Wong<sup>#</sup>. DC3 is a method for deconvolution and coupled clustering from bulk and single-cell genomics data. *Nature Communications*, 2019, 10(1): 1-11.
- [18] **Wanwen Zeng**, Xu Min and Rui Jiang. EnDisease: a manually curated database for enhancers-diseases associations. *Database*, 2019, 2019.
- [19] **Wanwen Zeng**, Mengmeng Wu and Rui Jiang. Prediction of enhancer-promoter interactions via natural language processing. *BMC Genomics*, 2018, 19(2): 13-22.
- [20] Xu Min<sup>\*</sup>, **Wanwen Zeng**<sup>\*</sup>, Shengquan Chen, Ning Chen, Ting Chen<sup>#</sup> and Rui Jiang<sup>#</sup>. Predicting enhancers with deep convolutional neural networks. *BMC Bioinformatics*, 2017, 18(13): 35-46.
- [21] Qijin Yin, Qiao Liu, Zhuoran Fu, **Wanwen Zeng**, Boheng Zhang, Xuegong Zhang, Rui Jiang, Hairong Lv. scGraph: a graph neural network-based approach to automatically identify cell types[J]. *Bioinformatics*, 2022, 38(11): 2996-3003.
- [22] Jianyu Zhou, Pan Li, **Wanwen Zeng**, Wenxiu Ma, Zhipeng Lu, Rui Jiang, Qiangfeng Cliff Zhang and Tao Jiang. IRIS: A method for predicting in vivo RNA secondary structures using PARIS data[J]. *Quantitative Biology*, 2020, 8(4): 369-381.
- [23] Zhana Duren<sup>\*</sup>, Xi Chen<sup>\*</sup>, Mahdi Zamanighomi<sup>\*</sup>, **Wanwen Zeng**, Ansuman T. Satpathy, Howard Y. Chang, Yong Wang, and Wing Hung Wong. Integrative analysis of single-cell genomics data by coupled nonnegative matrix factorizations. *Proceedings of the National Academy of Sciences*, 2018, 115(30): 7723-7728.
- [24] Mengmeng Wu, **Wanwen Zeng**, Wenqiang Liu, Hairong Lv, Ting Chen and Rui Jiang. Leveraging multiple gene networks to prioritize GWAS candidate genes via network representation learning. *Methods*, 2018, 145: 41-50.
- [25] Xu Min, **Wanwen Zeng**, Ning Chen, Ting Chen<sup>#</sup> and Rui Jiang<sup>#</sup>. Chromatin accessibility prediction via convolutional long short-term memory networks with  $k$ -mer embedding. *Bioinformatics*, 2017, 33(14): i92- i101.
- [26] Mingxin Gan, Wenran Li, **Wanwen Zeng**, Xiaojian Wang and Rui Jiang. Mimvec: a deep learning approach for analyzing the human phenome. *BMC Systems Biology*, 2017, 11(4): 3-16.
- [27] Jiaxin Wu, Mengmeng Wu, Lianshuo Li, Zhuo Liu, **Wanwen Zeng** and Rui Jiang. dbWGFP: a database and web server of human whole-genome single nucleotide variants and their functional predictions. *Database*, 2016, 2016.

## Teaching Experience

---

Lecturer, Computer Networking (undergraduate), Nankai University	2020
Lecturer, Cloud Computing and Application (undergraduate), Nankai University	2020, 2021
Lecturer, Introduction to Artificial Intelligence (undergraduate), Nankai University	2019, 2020
Teaching Assistant, Statistical Inference (graduate), Tsinghua University	2015, 2016

## Honors and Awards

---

- 2024, Invited talk at Biostatistics Seminar, Stanford University, CA, USA
- 2023, Invited talk at Biostatistics Seminar, UC Berkeley, CA, USA
- 2021, Winner of NeurIPS Competition, Multimodal Data Integration, rank 1<sup>st</sup>/2611 submissions in two Joint Embedding tasks, rank 2<sup>nd</sup>/1324 submissions in one Modality Prediction task
- 2017, CSC scholarships, China Scholarship Council
- 2014, Outstanding Graduates, Tianjin University, China
- 2013, Meritorious Winner in the Mathematical Contest in Modeling
- 2012, National Scholarship, Ministry of Education of China

## Professional Services

---

- Reviewer for *Bioinformatics*, *Briefing in Bioinformatics*, *Communications Biology*, *IEEE/ACM Transactions on Computational*

*Biology and Bioinformatics, BMC Bioinformatics, Frontiers in Genetics, Frontiers of Computer Science, Journal of Clinical Medicine, International Journal of Molecular Sciences, Cancers, Complex & Intelligent Systems, Engineering Applications of Artificial Intelligence.*

- Review Editor for *Frontiers in Genetics*