

# Guojie Zhong

PhD in Systems Biology, Columbia University

1130 St. Nicholas Avenue, New York, NY, USA | Phone: (+01)7187305993 | Email: gz2294@cumc.columbia.edu

Website: <https://zhongguojie1998.github.io/GuojieZhong/>

---

## RESEARCH INTEREST

I build biologically-inspired machine learning models to bridge the gap between complex diseases and effective treatments, deciphering disease mechanisms from high-throughput genomics data to identify new therapeutic targets.

---

## EDUCATION

08/2019 – 11/2024 PhD in Systems Biology, Integrated program in cellular, molecular and biomedical sciences  
Department of Systems Biology, **Columbia University**, New York, NY, USA.

Thesis Advisor: [Yufeng Shen, PhD](#)

Committee: [Wendy Chung, MD, PhD](#); [Mohammad AlQurashi, PhD](#); [David Knowles, PhD](#)

---

08/2017 – 12/2017 Exchange Student in Berkeley Bioscience Study Abroad Program (BBSA)  
Department of Integrated Biology, **University of California, Berkeley**, CA, USA.

---

09/2015 – 07/2019 Bachelor of Science, Integrated Science Program  
Yuanpei College, **Peking University**, Beijing, China.

Thesis Advisor: [Zemin Zhang, PhD](#)

Thesis: Reconstruction of cell spatial organization from single-cell RNA sequencing data  
based on ligand-receptor mediated self-assembly

---

## PUBLICATIONS

(# = co-first author)

### *Deep Learning and missense variant effect prediction*

- ◆ **Zhong, G.**, Zhao, Y., Zhuang, D., Chung, W. K. & Shen, Y. PreMode predicts mode-of-action of missense variants by deep graph representation learning of protein sequence and structural context. *Nat Commun* **16**, 7189 (2025).
- ◆ Zhao, Y., Lan, T., **Zhong, G.**, Hagen, J., Pan, H., Chung, W.K., and Shen, Y. A probabilistic graphical model for estimating selection coefficients of nonsynonymous variants from human population sequence data. *Nat Commun* **16**, 4670 (2025).

### *Statistical genetics and rare genetic diseases analysis*

- ◆ **Zhong, G.**, Choi, Y.A. and Shen, Y. VBASS enables integration of single cell gene expression data in Bayesian association analysis of rare variants. *Commun Biol* **6**, 774 (2023).
- ◆ **Zhong, G.**, and Shen, Y. Statistical models of the genetic etiology of congenital heart disease. *Curr Opin Genet Dev* **76**, 101967 (2022).
- ◆ **Zhong, G.**, Ahimaz, P., Edwards, N.A., Hagen, J.J., Faure, C., Lu, Q., Kingma, P., Middlesworth, W., Khlevner, J., El Fiky, M., et al. Identification and validation of candidate risk genes in endocytic vesicular trafficking associated with esophageal atresia and tracheoesophageal fistulas. *HGG Adv* **3**, 100107 (2022).
- ◆ Edwards, N., **Zhong, G.**, Ahimaz, P., Kenny, A., Kingma, P., Wells, J., Shen, Y., Chung, W., and Zorn, A. Discovering the Developmental Basis of Trachea-Esophageal Birth Defects: Evidence for Endosome-opathies. *The FASEB Journal* **36** (2022).

### *Single cell and spatial transcriptomics*

- ◆ Su, J., Reynier, J.B., Fu, X., **Zhong, G.**, Jiang, J., Escalante, R.S., Wang, Y., Aparicio, L., Izar, B., Knowles, D.A., and Rabadan, R. Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. *Genome Biol* **24**, 291(2023).
- ◆ Ren, X.#, **Zhong, G.**#, Zhang, Q., Zhang, L., Sun, Y., and Zhang, Z. Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res* **30**, 763-778 (2020).
- ◆ Zhang, Q., He, Y., Luo, N., Patel, S.J., Han, Y., Gao, R., Modak, M., Carotta, S., Haslinger, C., Kind, D., Peet G.W., **Zhong, G.**, Lu, S., Zhu, W., Mao, Y., Xiao, M., et al. Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* **179**, 829-845 e20 (2019).

---

## CONFERENCES

- ◆ **Zhong, G.**, and Shen, Y. (2023). Predicting mode of action of missense variants by graph representation of protein structural context. *American Society of Human Genetics 2023 Annual Meeting*, Washington, D.C.
- ◆ **Zhong, G.**, and Shen, Y. (2022). Representation of missense variants for predicting modes of action. *Machine Learning in Structural Biology, Workshop at the 36th Conference on Neural Information Processing Systems*, New Orleans, LA.
- ◆ **Zhong, G.**, Choi, Y.A., and Shen, Y. (2022). Integration of gene expression data in Bayesian association analysis of rare variants. *American Society of Human Genetics 2022 Annual Meeting*, Los Angeles, CA.
- ◆ **Zhong, G.**, Wang, J., He, S. and Fu, X. (2021). Towards better understanding of developmental disorders from integration of spatial single-cell transcriptomics and epigenomics. *The 2021 ICML Workshop on Computational Biology*, Virtual meeting.

---

## AWARDS

2024	Chinese Government Award for Outstanding Self-Financed Students Abroad
2019	Dean's fellowship, Graduate School of Arts and Science, Columbia University
2015	2014-2015 academic year Outstanding Freshman Scholarship, Peking University
2014	The 28th National Olympiad in Chemistry in Provinces, 1st prize, Zhejiang Province, China

---

## OPEN-SOURCE SOFTWARES

2022-2024	PreMode	<a href="https://github.com/ShenLab/PreMode">https://github.com/ShenLab/PreMode</a> Pretrained SE(3)-equivariant graph neural network model for missense variant effect representation that can accurately predict mode-of-action of missense variants using protein language models and structural context.
2020-2022	VBASS	<a href="https://github.com/ShenLab/VBASS">https://github.com/ShenLab/VBASS</a> Semi-supervised machine learning algorithm that combines neural network with Bayesian statistical models to identify disease risk genes for rare genetic conditions jointly from scRNA-seq data and rare variants data.
2018-2020	CSOmap	<a href="https://github.com/zhongguojie1998/CSOmap">https://github.com/zhongguojie1998/CSOmap</a> Unsupervised machine learning algorithm to reconstruct single cell spatial information and infer cellular interactions in tumor microenvironments from scRNA-seq data.

---

## MENTORSHIP EXPERIENCES

05/2025 – 08/2025	Di Liu, summer rotation PhD student Ren Lab, New York Genome Center 101 6th Ave, New York, NY, USA
06/2023 – 08/2023 06/2022 – 08/2022	Demi Zhuang, summer visiting undergraduate student  Shen Lab, Department of Systems Biology Columbia University Irving Medical Center, New York, NY, USA Developed protein family specific benchmarking tasks and datasets for PreMode. Applied PreMode to those datasets and discovered several novel likely G/LoF variants.
06/2021 – 08/2021	Andrew Lee, summer visiting high school student Shen Lab, Department of Systems Biology Columbia University Irving Medical Center, New York, NY, USA Developed an R app to visualize the scRNA-seq data of autism risk genes in brain cell types from public datasets. Applied it to discovery of novel autism risk genes and identification of disease associated cell types.

---

## PROFESSIONAL EXPERIENCES

03/2025 – present	Postdoctoral Research Associate Ren Lab, New York Genome Center 101 6th Ave, New York, NY, USA
01/2024 – 02/2025	Postdoctoral Research Scientist Shen Lab, Department of Systems Biology Columbia University Irving Medical Center, New York, NY, USA
06/2024 – 12/2024	Graduate Research Assistant

Shen Lab, Department of Systems Biology  
Columbia University Irving Medical Center, New York, NY, USA

06/2024 – 09/2024

Data Science & Machine Learning Intern  
Insitro, Inc.  
South San Francisco, CA, USA

03/2017 – 07/2019

Research Assistant  
Zhang Lab, Biomedical Pioneering Innovation Center (BIOPIC)  
Peking University, Beijing, China

---

## ACADEMIC AND VOLUNTEER SERVICE

*Peer Review Service*

PLOS Computational Biology; Molecular Medicine; Scientific Reports; PeerJ;  
Machine Learning in Structural Biology Workshop (NeurIPS 2022).

---

## REFERENCES

Yufeng Shen, Ph.D.,  
Associate Professor, Department of Systems Biology, Columbia University Irving Medical Center  
Associate Professor, Department of Biomedical Informatics, Columbia University Irving Medical Center  
Associate Director, JP Sulzberger Columbia Genome Center  
Tel: 917-543-9716  
Email: ys2411@cumc.columbia.edu  
Relationship: PhD Advisor

Wendy K. Chung, M.D., Ph.D.,  
Chief, Department of Pediatrics, Boston Children's Hospital  
Mary Ellen Avery Professor of Pediatrics, Harvard Medical School  
President, The Children's Hospital Pediatric Associates  
Tel: 617-355-6394  
Email: Wendy.Chung@childrens.harvard.edu  
Relationship: PhD Thesis Committee Chair

Mohammed AlQuraishi, Ph.D.,  
Assistant Professor, Department of Systems Biology, Columbia University Irving Medical Center  
Assistant Professor, Department of Computer Science, Columbia University  
Email: ma4129@cumc.columbia.edu  
Relationship: PhD Thesis Committee Member

David A. Knowles, Ph.D.,  
Assistant Professor, Department of Computer Science, Columbia University  
Assistant Professor, Department of Systems Biology, Columbia University Irving Medical Center  
Affiliate Member, Data Science Institute, Columbia University  
Core Faculty Member, New York Genome Center  
Email: dak2173@columbia.edu  
Relationship: PhD Thesis Committee Member