

Research Interests

- Biological sequence > structure ensemble > function > evolution
- Large language model for biological sequences
- Functional analysis of genetic mutations in disease
- Biomolecular interactions and subcellular localization
- Biomolecular designation and effect prediction

- Postdoc Research Scientist Columbia University Sep 2023 - Present
 - ✧ Advisor: [Dr. Yufeng Shen](#)

- Ph.D. in Biomedical Informatics Peking University Sep 2020 - Jul 2023
 - ✧ Advisor: [Dr. Tingting Li](#)
- Bachelor of Medicine and Economics Peking University Sep 2015 - Jul 2020

(# indicates co-first author, * indicates co-corresponding author)

1. *MotifAE Reveals Functional Motifs from Protein Language Model: Unsupervised Discovery and Interpretability Analysis.*
Hou C*, Di L, Shen Y*. BioRxiv. Nov 2025. (submitted)
2. *Understanding Language Model Scaling on Protein Fitness Prediction.*
Hou C*, Di L, Zafar A, Shen Y*. BioRxiv. Apr 2025. (under review)
3. *Learning Biophysical Dynamics with Protein Language Models.*
Hou C, Zhao H, Shen Y*. BioRxiv. Oct 2024. (under review)
4. *Investigating phase separation properties of chromatin-associated proteins using gradient elution of 1,6-hexanediol.*
Zhu P#, **Hou C**#, Liu M, Chen T, Li T*, Wang L*. BMC Genomics. Aug 2023.
5. *PhaSepDB in 2022: annotating phase separation-related proteins with droplet states, co-phase separation partners and other experimental information.*
Hou C#, Wang X#, Xie H#, Chen T, Zhu P, Xu X, You K, Li T*. Nucleic Acids Res. Jan 2023.
6. *Systematic prediction of degrons and E3 ubiquitin ligase binding via deep learning.*
Hou C, Li Y, Wang M, Wu H, Li T*. BMC Biology. Jul 2022.
7. *Screening membraneless organelle participants with machine-learning models that integrate multimodal features.*
Chen Z#, **Hou C**#, Wang L#, Yu C, Chen T, Shen B, Hou Y, Li P*, Li T*. Proc Natl Acad Sci U S A. Jun 2022.

8. [Serum Antigenome Profiling Reveals Diagnostic Models for Rheumatoid Arthritis](#). Han P, **Hou C**, Zheng X, Cao L, Shi X, Zhang X, Ye H, Pan H, Liu L, Li T*, Hu F*, Li Z*. Front Immunol. Apr 2022.
9. [MloDisDB: a manually curated database of the relations between membraneless organelles and diseases](#). **Hou C**#, Xie H#, Fu Y#, Ma Y#, Li T*. Brief Bioinform. Jul 2021.

Other Publications

10. [PhaSeDis: A Manually Curated Database of Phase Separation–Disease Associations and Corresponding Small Molecules](#). Chen T, Tang G, Li T, Yanghong Z, Hou C, Du Z, Ma L, Li T. Genomics Proteomics Bioinformatics. Mar 2025.
11. [Investigating variant impact on phosphorylation events driving carcinogenesis](#). Xu X, Li Y, Chen T, Hou C, Yang L, Zhu P, Zhang Y, Li T. Brief Bioinform. Dec 2023.
12. [Distinctive Network Topology of Phase-Separated Proteins in Human Interactome](#). Yu C, Lang Y, Hou C, Yang E, Ren X, Li T. J Mol Biol. Jan 2022.
13. [Quantifying the phase separation property of chromatin-associated proteins under physiological conditions using an anti-1,6-hexanediol index](#). Shi M, You K, Chen T, Hou C, Liang Z, Liu M, Wang J, Wei T, Qin J, Chen Y, Zhang MQ, Li T. Genome Biol. Aug 2021.

Oral Presentations

- 2025.10 [Understanding Language Model Scaling on Protein Fitness Prediction](#). Retreat of Department of Systems Biology of Columbia University. NY, US.
- 2024.09 [Learning Representation of Protein Dynamic Properties with a Language Model](#). Retreat of Department of Systems Biology of Columbia University. PA, US.
- 2023.07 [The degradation regulation of phase separating proteins](#). The student symposium in Fudan international summer school of life science, Shanghai, China.
- 2023.06 [The degradation regulation of phase separating proteins](#). Excellent graduate students symposium of Peking University School of Basic Medical Sciences, Beijing, China.
- 2023.04 [The degradation regulation of phase separating proteins](#). Multidisciplinary Conference on New ideas and New Technologies at Peking University, Beijing, China.
- 2022.12 [Targeting disordered degrons on phase separating proteins](#). Silk Road International Symposium for Distinguished Young Scholars at Xi'an Jiaotong University, Online meeting.

Poster Presentations

- 2025.02 [Learning Biophysical Dynamics with Protein Language Models](#). BPS 2025. Los Angeles, US.
- 2024.11 [Predicting missense mutation effects with biophysics-informed protein language model](#). ASHG 2024. Denver, US.
- 2024.07 [Learning Representation of Protein Dynamic Properties with a Language Model](#). ISMB 2024. Montreal, Canada.

Honors and Awards

- 2023 First prize in the student symposium in Fudan international summer school of life science
- 2023 Peking University Excellent Graduate
- 2022 Peking University President Scholarship

2021 Peking University Doctoral Innovation Scholarship
2020 Peking University Junior Scholar
2016 National Scholarship
2014 Silver Medal in National Physics Olympiad for high school students

Journal Referee

Ebiomedicine, Plos Computational Biology, Protein Science

Teaching

PhD: Di Liu (Columbia), Aziz Zafar (Columbia).

Undergrad: Zixuan Tao (UConn), Jason Xie (Columbia), Tianhao Li (PKU), Xinyi Li (PKU), Haotai Xie (PKU).

High school student: Kevin Liu, Ludovic Pham.