# HAO CHENG

Hao.Cheng@osumc.edu

Department of Biomedical Informatics, College of Medicine

The Ohio State University

310 Lincoln Tower, 1800 Cannon Dr.

(+1) 614-592-5668

GitHub: https://github.com/chthub

## Education

**Undergraduate, ShanDong University**                                            2018.09 - 2022.07

Major in Data science and Big data technology, School of Software

Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR)

**Ph.D. student, The Ohio State University**                                        2022.08 - present

Biomedical Sciences Graduate Program

**Advisor**: Prof. Qin Ma

## Research Experience

**Undergraduate Researcher**                                                        2019.09 - 2019.12

Research Center of Human-Computer Interaction and Virtual Reality, SDU, Zhifang Jiang

- Psychological sand table data collection and analysis; Use python spiders to crawl public psychoanalytic data; Use natural language processing technology to analyze user sand table data and generate primary analysis reports. Design a primary analysis reports generator based on rules.

**Undergraduate Researcher**                                                        2020.07 - 2021.08

Research Center of Software and Data Engineering, SDU, Leyi Wei

- Peptide detectability is an important factor in shotgun proteomics; Using the deep learning method, the pepformer model is proposed. The transformer and siamese network architectures are used to predict the detectability of peptides only based on peptide sequences and obtain the state of art. The relevant results are published in Journal of Analytical Chemistry.
- De novo sequencing of peptides is a method to determine the amino acid sequence of peptides by tandem mass spectrometry; Try to do de novo sequencing of peptides using deep learning method on public mass spectrometry data.

**Undergraduate Researcher**                                                        2021.09 – 2022.07

Bioinformatics and Mathematical Biosciences Lab, OSU, Qin Ma

- The Task 3 (Joint Embedding) in NeurIPS 2021 Competition, Open Problems in Single-Cell Analysis. A joint embedding from multiple modalities should be learned by algorithms. An auto-encoder framework was designed to generate the low dimensional representation vectors. The

heterogeneous graph neural networks were used to model the multiple modalities data and capture the heterogeneity. Finally, ranked fifth in this competition.

**Graduate Research Assistance**                                    2022.08 – present

Bioinformatics and Mathematical Biosciences Lab, OSU, Qin Ma

- Single cell multi-omics data analysis.
- Deep learning framework design. Design a new framework for the discovery of senescent cell and marker genes.
- Summarize and present graph representation learning related works and slides.

## Publications

1. **Hao Cheng**, Anjun Ma, Cankun Wang, Ana L. Mora, Mauricio Rojas, Dongjun Chung, Qin Ma. **DeepSAS: Identifying cell-type-specific senescent cells and signature genes using heterogeneous graph contrastive learning**. In preparation.
2. Anjun Ma, Xiaoying Wang, Jingxian Li, Cankun Wang, Tong Xiao, Yuntao Liu, **Hao Cheng**, Juexin Wang, Yang Li, Yuzhou Chang, Jinpu Li, Duolin Wang, Yuexu Jiang, Li Su, Gang Xin, Shaopeng Gu, Zihai Li, Bingqiang Liu, Dong Xu & Qin Ma (2023). **Single-cell biological network inference using a heterogeneous graph transformer**. Nature Communications, 14(1), 964. https://doi.org/10.1038/s41467-023-36559-0.
3. Haocheng Gu, **Hao Cheng (co-first author)**, Ma, A., Li, Y., Wang, J., Xu, D., & Ma, Q. (2022). **scGNN 2.0: a graph neural network tool for imputation and clustering of single-cell RNA-Seq data**. Bioinformatics, 38(23), 5322-5325. https://doi.org/10.1093/bioinformatics/btac684.
4. **Hao Cheng** et al., **PepFormer: End-to-End Transformer-Based Siamese Network to Predict and Enhance Peptide Detectability Based on Sequence Only**, Analytical Chemistry 2021 93 (16), 6481-6490, DOI: 10.1021/acs.analchem.1c00354.

## Poster

1. **Identifying cell-type-specific senescent cells and signature genes using heterogeneous graph contrastive learning**. Hao Cheng, Anjun Ma, Cankun Wang, Ana L. Mora, Mauricio Rojas, Dongjun Chung, Qin Ma.
   Poster presentation at the TriState SenNet Annual Conference, the Ohio State University.
   Received the second-place poster award.                                    2023.04.10

## Skills

1. Data analysis and model building using Python and R.
2. Develop web server for bioinformatics applications.
3. Train and fine-tune large deep learning models.
4. Use Adobe Illustrator or Photoshop to illustrate the paper.
5. Use deep learning methods to study bioinformatics problems.