

# Naihui Zhou

---

## Education

- **PhD** Double major, [Iowa State University](#), 2020
  - Bioinformatics and Computational Biology (BCB)  
Advisor: Iddo Friedberg
  - Statistics  
Advisor: Mark S. Kaiser
- **MA** Statistics, [Columbia University](#), 2014
- **BSc** Information and Computing Science, [Beijing Jiaotong University](#), 2013

## Contact

- [ashley.n.zhou@gmail.com](mailto:ashley.n.zhou@gmail.com)
- [ashleyzhou972.github.io](https://github.com/ashleyzhou972)
- [github.com/ashleyzhou972](https://github.com/ashleyzhou972)

## Publications

- **Zhou, N.**, Jiang, Y., Bergquist, T.R., Lee, A.J., Kacsoh, B.Z., Crocker, A.W., Lewis, K.A., Georgiou, G., Nguyen, H.N., Hamid, M.N., Davis, L. et al., 2019. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *bioRxiv*, p.653105. <https://doi.org/10.1101/653105> (Accepted by *Genome Biology*)
- Kacsoh, B.Z., Barton, S., Jiang, Y., **Zhou, N.**, Mooney, S.D., Friedberg, I., Radivojac, P., Greene, C.S. and Bosco, G., 2019. New *Drosophila* long-term memory genes revealed by assessing computational function prediction methods. *G3: Genes, Genomes, Genetics*, 9(1), pp.251-267. <https://doi.org/10.1534/g3.118.200867>
- **Zhou, N.**, Siegel, Z.D., Zarecor, S., Lee, N., Campbell, D.A., Andorf, C.M., Nettleton, D., Lawrence-Dill, C.J., Ganapathysubramanian, B., Kelly, J.W. and Friedberg, I., 2018. Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning. *PLoS computational biology*, 14(7), p.e1006337. <https://doi.org/10.1371/journal.pcbi.1006337>
- Leng, Y., Kou, C., **Zhou, N.**, Li, Q., Liang, Y., Xu, Z. and Chen, S., 2012. Evaluation on transfer efficiency at integrated transport terminals through multilevel grey evaluation. *Procedia-Social and Behavioral Sciences*, 43, pp.587-594. <https://doi.org/10.1016/j.sbspro.2012.04.132>

## Open Source Software

- [PhiMRF](#), an R package for spatial dependency using a Poisson Hierarchical Markov Random Field model for count data.
- [CAFA Assessment Tool](#), assessment of protein function prediction using precision-recall for the [CAFA challenge](#).

# Experience

## Research Assistant, Iowa State University

- **2017- 2019** Exploring the spatial dependency of gene expression using Markov Random Field models
  - Developed a Poisson-lognormal mixed autoregressive model (PhiMRF) to investigate the spatial dependency of RNA-seq data in light of the 3D chromosomal organization in human cells.
  - Implemented a double Metropolis-Hastings MCMC process for the PhiMRF model to derive posterior estimates of spatial dependency parameters.
  - Developed an R package for the PhiMRF model, featuring fast and parallel computation, sparse matrix formats, OpenMP and Lapack routines.
- **2016- 2019** Critical Assessment of protein function prediction
  - Organized the third Critical Assessment of Functional Annotation (CAFA) challenge for the prediction of protein functions, with participants from 69 laboratories across the globe and hundreds of submissions.
  - Evaluated the performance the CAFA structured output predictions using precision-recall based metrics, and other analysis, such as similarity measures based on these predictions.
  - Developed a Python tool for precision-recall evaluation for structured-output prediction tasks.
- **2016- 2018** Crowdsourcing image analysis for plant phenomics
  - Analyzed image labelling performance of different crowdsourced groups using Generalized Linear Mixed Effects model on SAS.

## Intern, New York City Health and Hospitals Corporation

- **2014- 2015** Effects of Beta Blockers on hospital admission rates and cost
  - Analyzed the use of beta blockers in the health records of the New York City MetroPlus insurance population.
  - Applied a zero-inflated Poisson regression model for the effects of co-morbidities, socioeconomic factors on hospital admission rates using SAS

## Grader, Columbia University

- **2014** STAT W1111, Introduction to Statistics

## Conference Presentations

- Using Markov Random Field to Model Gene Expression in the 3D Genome (Talk, Travel Fellowship). **Great Lakes Bioinformatics (GLBIO)**, Madison, WI, USA, 2019.
- Updates on CAFA3 and CAFA3.14 (Talk, Travel Fellowship). Using Markov Random Field to Model Gene Expression in the 3D Genome (Poster). **Intelligent Systems for Molecular Biology (ISMB)**, Chicago, IL, USA, 2018.
- The Critical Assessment of Protein Function Annotation: Preliminary Assessment (Talk). **Intelligent Systems for Molecular Biology (ISMB)/European Conference on Computational Biology (ECCB)**, Prague, Czech Republic, 2017
- The Critical Assessment of Protein Function Annotation: The Road Ahead (Talk, Travel Fellowship). **Biocuration 2017**, Stanford, CA, USA, 2017
- The Critical Assessment of Protein Function Annotation— Improving on the “state-of-the-art” (Poster, Travel Fellowship). **Biological Data Science**, Cold Spring Harbor Laboratories, NY, USA, 2016.
- Effects of Beta Blockers on Hospital Admission Rates and Cost (Talk). **Joint Statistical Meeting**, Seattle, WA, USA, 2015

## Awards

- Lora and Russ Talbot Graduate Fellowship in Veterinary Medicine, College of Veterinary Medicine, Iowa State University, 2019
- F. Wendell Miller Scholarship, Graduate College, Iowa State University, 2015- 2018

## Skills

- Reproducible research, statistical computing, hierarchical models, Markov random fields, Bayesian methods, Markov chain Monte Carlo, generalized linear models, data mining, machine learning, parallel computing, R package development
- RNA-seq data analysis, HiC data analysis, Gene Ontology, Gene Ontology Annotations database
- R, python, C, and shell scripting
- Experience with OpenMP, SAS, Latex and Matlab

## Leadership

- Senator, Graduate and Professional Students Senate, Iowa State University, 2017-2018
- Director of Outreach, BCB Graduate Student Organization, Iowa State University, 2016-2017
- Volunteer, MUMO Education and Orphanage Center, Nairobi, Republic of Kenya, 2011