Naihui Zhou

Curriculum Vitae

CONTACT

Email	ashley.n.zhou@gmail.com
Phone	+1 (781) 864 7096
Github	`https://github.com/ashleyzhou972`

EDUCATION

PhD dual-major
Bioinformatics and Computational Biology
Statistics 2015-present
Iowa State University
Ames, IA, USA
Advisors:
Iddo Friedberg, Department of Veterinary Microbiology and Preventive Medicine, ISU
Mark S. Kaiser, Department of Statistics, ISU

MA. Statistics 2013-2014
Columbia University
New York, NY, USA

BSc. Information and Computing Science 2009-2013
Beijing Jiaotong University
Beijing, China

PUBLICATIONS

- **At Iowa State University:**
  - **Zhou, N.**, Friedberg, I., and Kaiser, M. S. (2019). Hierarchical Markov Random Field model captures spatial dependency in gene expression, demonstrating regulation via the 3D genome. *bioRxiv.*
    `https://doi.org/10.1101/2019.12.16.878371`
  - **Zhou, N.**, Jiang, Y., Bergquist, T. R., Lee, A. J., Kacsoh, B. Z., Crocker, A. W., ... and Friedberg, I. (2019). The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome Biology*, 20(1), 1-23.
    `https://doi.org/10.1186/s13059-019-1835-8`
  - 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`

- **At Beijing Jiaotong University:**
  - 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`

RESEARCH

Connecting gene expression and the 3D genome with spatial statistics 2017-present
Iowa State University

- Developed a Hierarchical Markov Random Field (MRF) model for RNA-seq data. A MRF model is a probabilistic model for spatial dependency through auto-correlation. The hierarchical MRF model is able to detect spatial dependency in gene expression, demonstrating regulation mechanisms through the 3D structure of the genome.
- Processed and analyzed spatial genome interactions from HiC experiments, as well as RNA-seq data, for human cell lines.
- Inference from the hierarchical model follows a Bayesian framework, using a double Metropolis-Hastings algorithm.
- Developed an R package with full documentation and vignettes for the model. The core computation is implemented in C with parallelization.

Critical assessment of protein function prediction 2016-2019
Iowa State University

- One of the organizers of the third Critical Assessment of Functional Annotation (CAFA) challenge. CAFA is a community challenge evaluating computational algorithms for predicting protein function.
- Wrote evaluation tool in Python for assessing the accuracy of protein function predictions. The software tool handles gigabytes of genomic data such as sequence and annotation, and computes precision-recall evaluation for a multi-label prediction task.
- Communicated with participants from around the world throughout the submission process and results release.
- Performed auxiliary analysis of the evaluation results such as keyword frequency, methods similarity, etc.

Crowdsourcing image analysis for plant phenomics 2016-2018
Iowa State University

- Analysis of image labelling performance of different crowdsourced groups, such as Amazon Mechanical Turk. By labelling images of corn field with corn tassels, the study could potentially speed up the lengthy process of quantifying plant phenotypic traits.
- Performed evaluation of the image labelling results using a linear mixed-effects model on SAS.

Effects of Beta Blockers on hospital admission rates and cost 2014-2015
New York City Health and Hospitals Corporation

- Examined the use of beta blockers in the New York City MetroPlus insurance population
- Applied zero-inflated regression models and gamma regression to model the effects of co-morbidities and socioeconomic factors on hospital admission of patients.

ACADEMIC CONFERENCES

Genome Informatics 2019

Cold Spring Harbor, NY, USA

- Oral presentation: "Exploring the 3D spatial dependency of gene expression using Markov Random Fields."

Great Lakes Bioinformatics (GLBIO) 2019

Madison, WI, USA

- Oral presentation: "Using Markov Random Field to Model Gene Expression in the 3D Genome"
- 2019 GLBIO Travel Fellowship Award from the International Society for Computational Biology

Intelligent Systems for Molecular Biology (ISMB) 2018

Chicago, IL, USA

- Oral presentation: "Updates on CAFA3 and CAFA3.14"
- Poster presentation: "Using Markov Random Field to Model Gene Expression in the 3D Genome"
- 2018 ISMB Travel Fellowship Award from the International Society for Computational Biology

Intelligent Systems for Molecular Biology (ISMB) /
European Conference on Computational Biology (ECCB) 2017

Prague, Czech Republic

- Oral presentation: "The Critical Assessment of Protein Function Annotation: Preliminary Assessment"

Biocuration 2017 2017

Stanford, CA, USA

- Oral presentation: "The Critical Assessment of Protein Function Annotation: The Road Ahead"
- Student Travel Fellowship from the International Society for Biocuration

Joint Statistical Meeting 2015

Seattle, WA, USA

- Oral presentation: "Effects of Beta Blockers on Hospital Admission Rates and Cost"

AWARDS

Lora and Russ Talbot Graduate Fellowship in Veterinary Medicine 2019

College of Veterinary Medicine, Iowa State University

F. Wendell Miller Scholarship 2015-2018

Graduate College, Iowa State University

WORK EXPERIENCE

Intern 2014-2015

New York City Health and Hospitals Corporation

- Cleaned and prepared gigabytes of patient data including demographic, prescription and diagnostics, with an extensive use of SAS
- Applied generalized linear models to analyze the correlation between drug use, diagnoses and hospital admission and provide insights in hospital service and health policy

LEADERSHIP

Senator, Graduate and Professional Students Senate 2017-2018
Iowa State University

- Represented the graduate students in the department of Veterinary Microbiology and Preventive Medicine.
- Discussed and voted on issues pertaining to the life and study of graduate and professional students at ISU.
- Served on the Elections Committee.

Director of outreach, BCB Graduate Student Organization 2016-2017
Iowa State University

- Lead and organized outreach activities in cooperation with Women in Science and Engineering (WISE) of ISU, the Iowa Science Center and others to engage and educate the public about bioinformatics research.
- Organized the annual BCB symposium, a one-day research-focused event with external speakers and student presentations.

SKILLS

- Reproducible research, hierarchical models, random fields, Bayesian methods, Markov chain Monte Carlo, generalized linear models, data mining, machine learning, statistical computing, parallel computing.
- Gene Ontology, Gene Ontology Annotations database, RNA-seq data analysis, HiC data analysis.
- Fluent in C, R, Python and Unix/Linux shell scripting.
- Experience with SAS, LaTeX and Matlab.