

# Xin Yin

3219 Snedecor Hall  
Iowa State University  
Ames, IA 50011

Phone: (917)-214-8086  
Email: xinyin@iastate.edu  
<https://xydrolase.github.io>

---

## Research Interests

Computational genomics, data mining, machine learning, statistical computing, high-performance computing

## Education

Ph. D. in **Bioinformatics & Computational Biology (BCB)** 2010 - Expected Spring 2016  
Co-major Ph.D. in **Statistics**  
*Iowa State University* Ames, IA, United States

**Advisors:** Dr. Karin Dorman, Dr. Gregory Phillips

**Research Area:** Computational Genomics, Statistical Computing, Error-correction and Base-calling of Next-Generation Sequencing Data, Metagenomics

B. Eng. in **Biomedical Engineering** 2006 - 2010  
*University Of Electronic Sci. & Tech. Of China (UESTC)* Chengdu, China

**Advisor:** Dr. Feng-biao Guo

**Thesis:** Re-annotation of protein-coding genes in *S. cerevisiae* genome using support vector machines

## Research and Professional Experience

**Research Assistant** Apr. 2011 - Present / Nov. 2010 - Feb.2011  
*Karin Dorman Lab* Iowa State University

- Probabilistic modeling of next-generation sequencing base-calling and error-correction.
  - Developed a flexible, probabilistic and reference-free method for correcting substitution, insertion and deletion errors for prevalent high-throughput DNA sequencing platforms (*e.g.* Illumina, Ion Torrent).
  - Developed an integrated-approach to base-calling and error-correction for Illumina sequencing data, that models nucleotide dependence in genomic sequence and sequencer-specific error-characteristics.
- Statistical analysis of MeDIP-Chip and MeDIP-Seq data to identify DNA Methylation.

**Summer Intern, Computational Biologist** Jul. 2014 - Sept. 2014  
*Boehringer Ingelheim - Computational Biology Group* Ridgefield, CT

- Statistical analysis of gene expression (micro-array and RNA-seq) data.
- Automation of gene pathway analysis with task queues and web management console.
- Web-based interactive visualization of pathway analysis and differential expression.

**Research Assistant, Lab Rotation** Feb. 2011 - Apr. 2011  
*Robert Jernigan Lab* Iowa State University

Mixed atomistic and coarse-grained Elastic Network Modeling of protein structural dynamics.

**Research Assistant, Lab Rotation** Sept. 2010 - Nov. 2010  
*Vasant Honavar Lab* Iowa State University

Identification of gene pathways specific to retinal cell types using network alignment.

<b>Research Assistant</b> <i>Fengbiao Guo Lab</i>	Sept. 2009 - Aug. 2010 UESTC
Reannotation of protein-coding genes in <i>S. cerevisiae</i> genome using support vector machines.	
<b>Software Developer</b> <i>Dormforce, Student Association at UESTC</i>	2007 - 2010 Chengdu, China

## Teaching Experience

<b>Teaching Assistant</b> STAT 330: Probability and Statistics for Computer Science <i>Department of Statistics, Iowa State University</i>	Fall 2011
<b>Teaching Assistant</b> BCB 568: Advanced Genomic Informatics <i>Bioinformatics &amp; Computational Biology Program, Iowa State University</i>	Spring 2014, Spring 2015

## Awards

<b>Cornette Fellowship</b> <i>Bioinformatics &amp; Computational Biology Program, Iowa State University</i>	Jun. 2014
<b>1st Place, Iowa State University Team 1, Data Mining Cup 2014</b> <i>Prudsys AG, Berlin, Germany</i>	Mar. 2014 - May 2014
<b>Bancroft Award</b> <i>Department of Statistics, Iowa State University</i>	Sept. 2014
<b>People's Scholarship</b> (Prize for outstanding contribution, 3rd Class, 2nd Class) <i>School of Life Science and Technology, UESTC</i>	2007, 2008, 2009

## Skills

**Programming skills**  
C/C++, Python, R/Bioconductor, HTML/CSS, Javascript, Shell scripting

## Publication

### Peer-reviewed Journals

- **Yin X.**, Dorman K. S. (journal article in preparation). PREMIER-bc: Probabilistic Framework for Integrated Illumina Sequencing Base-calling and Error-correction.
- **Yin X.**, Noorzi V., Dorman K. S., Ramamoorthy, A. (journal article in preparation). Pindel: Probabilistic Insertion, Deletion and Substitution Error Correction using Markov Inference in Next Generation Sequencing Reads.
- **Yin X.**, Dorman K. S., Ramamoorthy, A. (journal article in preparation). PREMIER: Probabilistic Error-correction using Markov Inference in Errored Reads.
- Lin, D., **Yin, X.**, Wang, X., Zhou, P., Guo, F.B. (2013). Re-Annotation of Protein-Coding Genes in the Genome of *Saccharomyces cerevisiae* Based on Support Vector Machines. *PloS One*.

### Conference Proceedings

- **Yin X.**, Noorzi V., Dorman K. S., Ramamoorthy, A. (2015) Pindel: Probabilistic Insertion, Deletion and Substitution Error Correction using Markov Inference in Next Generation Sequencing Reads. (*submitted to RECOMB2016*).
- **Yin X.**, Song Z., Dorman K. S., Ramamoorthy, A. (2013) PREMIER Turbo: Probabilistic Error-correction using Markov Inference in Errored Reads using the Turbo principle. In *IEEE Global Conf. on Sig. and Info. Proc. (GlobalSIP)*.

- **Yin X.**, Song Z., Dorman K. S., Ramamoorthy, A. (2013) PREMIER: Probabilistic Error-correction using Markov Inference in Errored Reads. In *IEEE Intl. Symposium on Info. Th. (ISIT)*.
- **Yin, X.**, Wang, X., Dorman, K. S. (2011). Hidden Markov Model of Genome-scale Methylated DNA Immunoprecipitation (MeDIP) data. In *Proceedings of Statistical, Computational and Visualization Methods in Medical Informatics (SCVMMI11)*.

## Talks and posters

- **Yin X.** An Introduction to AWS CodeDeploy. *Cloud Computing Working Group*, Department of Statistics, Iowa State University, October 15, 2015.
- **Yin X.** Amazon Web Services: IAM, S3, Spot Instances and more. *Cloud Computing Working Group*, Department of Statistics, Iowa State University, October 01, 2015.
- **Yin X.**, Dorman, K.S. PREMIER-bc: Integrated Base-calling and Error-correction Approach Using Hidden Markov Model. *SAMSI Bioinformatics Transition Workshop*, Raleigh, NC, May 11, 2015.
- **Yin X.**, Dorman, K.S. An Update to PREMIER-bc, the Hybrid Base-caller and Error-corrector: Modeling choices, Derivations and Algorithms. *SAMSI, Bio: Next Generation Sequencing Errors*, April 26, 2015.
- **Yin X.**, Dorman, K.S. Probabilistic Error-correction and Base-calling using Markov Inference. *SAMSI, Bio: Next Generation Sequencing Errors*, April 13, 2015.
- **Yin X.** Error corrections for Illumina sequencers, *SAMSI, Bio: Next Generation Sequencing Errors*, November 10, 2014.
- Noroozi V.\*, **Yin X.**, Dorman K. S., Ramamoorthy A. Probabilistic Error Correction Using Markov Inference in Erroneous Reads with Insertions, Deletions, and Substitutions. *NSF/DTRA Algorithms for Threat Detection Workshop*, Boulder, CO, March 10, 2014.
- **Yin X.**, Dorman K. S., Ramamoorthy, A. (poster) PREMIER: Probabilistic Error-correction using Markov Inference in Errored Reads. *NSF/DTRA Algorithms for Threat Detection Workshop*, Boulder, CO, March 11, 2014.
- **Yin X.**, Dorman K. S., Ramamoorthy, A. Hidden Markov Modeling of Next Generation Sequencing Error Correction. *NSF/DTRA Algorithms for Threat Detection Workshop*, San Diego, CA, November 27, 2012.
- **Yin X.** Practices to Optimize Numeric Code: using Hidden Markov Model As An Example *Computational Statistics Working Group*, Department of Statistics, Iowa State University, September 12, 2012.
- **Yin, X.**, Wang, X., Dorman, K. S. Hidden Markov Model of Genome-scale Methylated DNA Immunoprecipitation (MeDIP) data. *Biostatistics Working Group*, Department of Statistics, Iowa State University, October 21, 2011.
- **Yin, X.**, Wang, X., Dorman, K. S. Hidden Markov Model of Genome-scale Methylated DNA Immunoprecipitation (MeDIP) data. *Iowa State University - University of Iowa Bioinformatics Meeting*, Iowa State University, August 27, 2011.

## Reference

- available upon request